



Analítica de datos en el entorno portuario

Ismael Torres Boigues

Máster Universitario en Ciencia de Datos

Área 3 del trabajo final

Tutora: Elisenda Bonet Carné

Fecha: 03 de enero de 2021



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Analítica de datos en el entorno portuario</i>
Nombre del autor:	<i>Ismael Torres Boigues</i>
Nombre del consultor/a:	<i>Nombre y dos apellidos</i>
Nombre del PRA:	<i>Elisenda Bonet Carné</i>
Fecha de entrega	01/2021
Titulación:	Máster Universitario en Ciencia de Datos
Área del Trabajo Final:	TFM - Área 3 aula 1
Idioma del trabajo:	Castellano
Palabras clave	<i>Integración Datos Abiertos, Puerto Analítico, Modelos predictivos.</i>
Resumen del Trabajo:	
<p>El sector portuario es uno de los sectores industriales más competitivos a nivel internacional y es el encargado de transportar más del 80% de las mercancías a nivel mundial. Pese al gran volumen de negocio, los márgenes económicos son muy bajos y no todos los puertos son rentables económicamente hablando.</p> <p>Actualmente, los puertos españoles disponen de herramientas de Business Intelligence(BI) que les permiten analizar lo ocurrido en base a datos históricos del propio puerto. Estas herramientas, no incorporan información de otros puertos nacionales/internacionales para poder compararse. Además, carecen de técnicas predictivas que les ayuden a predecir tendencias futuras y planificarse en base a dichas predicciones.</p> <p>Este proyecto proporciona a través de una herramienta de BI, cuadros de mando con información relevante de todos los puertos españoles proveniente de orígenes de datos públicos.</p> <p>Gracias a esta información y al uso de diferentes técnicas analíticas predictivas, se obtienen predicciones futuras a corto y medio plazo.</p> <p>Con esta solución BI, los puertos españoles disponen de una potente solución analítica de ayuda a la toma de decisiones en base al histórico de datos y a las predicciones futuras.</p>	

Abstract:

The port sector is one of the most competitive industrial sectors at international level and is responsible for transporting more than 80% of goods worldwide. Despite the high volume of business, economic margins are very low and not all ports are economically profitable.

Currently, Spanish ports have Business Intelligence (BI) tools that allow them to analyse what has happened based on historical data from the port itself. These tools do not incorporate information from other national/international ports to be able to compare with them. In addition, they do not incorporate predictive techniques to predict future trends.

This project provides through a BI tool, dashboards with relevant information of all Spanish ports from public data sources.

Thanks to this information and the use of different predictive analytical techniques, future predictions are obtained in the short and medium term.

With this BI solution, the Spanish ports have a powerful analytical solution to assist in decision making based on historical data and future predictions.

Índice

Contenido

1.	Introducción.....	10
1.1.	Contexto y justificación del Trabajo.....	10
1.2.	Objetivos del Trabajo.....	11
1.3.	Enfoque y método seguido.....	11
1.4.	Planificación del Trabajo.....	15
1.5.	Breve resumen de productos obtenidos.....	15
1.6.	Breve descripción de los otros capítulos de la memoria.....	16
2.	Estado del Arte.....	17
2.1.	Business Intelligence (BI).....	17
2.2.	Uso de la IA en el entorno Portuario.....	18
2.2.1.	Uso de la IA desde el punto de vista Smart Port.....	19
2.2.2.	Uso de AI en Industria 4.0 (aplicada al sector portuario-logístico).....	19
2.2.3.	AI para predicciones del volumen de negocio en puertos.....	20
2.2.4.	Aplicación de AI el presente trabajo.....	26
2.3.	Fuentes de datos abiertas.....	26
2.3.1.	Datos de puertos españoles.....	26
2.3.2.	Datos de puertos Internacionales.....	28
3.	Alcance del proyecto.....	31
4.	Carga de Datos e Integración.....	32
4.1.	Carga anual de datos portuarios españoles históricos.....	33
4.2.	Carga mensual de datos portuarios españoles históricos.....	38
4.3.	Carga datos internacionales.....	44
4.4.	Conclusiones sobre la carga de datos.....	45
5.	Análisis Series Temporales.....	46
5.1.	ARIMA.....	47
5.1.1.	Implementación.....	47
5.1.2.	Resultados aplicación ARIMA datos Anuales Puerto Valencia.....	48
5.2.	SARIMA.....	50
5.3.	ARIMAX.....	51
5.4.	Redes neuronales recurrentes (RNN).....	54
5.5.	Facebook Prophet.....	55
5.6.	Resumen del Análisis de Datos.....	56
5.6.1.	Series Anuales.....	56
5.6.2.	Series mensuales.....	57

5.6.3. Conclusiones y limitaciones sobre los análisis realizados.....	57
6. Cuadros de Mando (BI).....	59
6.1. Conclusiones sobre Qlik	63
7. Conclusiones y Trabajos Futuros	64
8. Glosario	66
9. Bibliografía	67
10. Anexos.....	70
10.1. Anexo Recursos Públicos del Proyecto.	70
10.2. Business Intelligence (BI).....	71
10.3. Anexo Carga de datos.....	84
10.3.1. Scripts de creación de Base de datos.....	84
10.4. Anexo Detalle Análisis de Series Temporales	89
10.4.1. Resultados aplicación ARIMA datos anuales Puerto Valencia	89
10.4.2. Resultados aplicación ARIMA datos mensuales Puerto Valencia	91
10.4.3. Resultados aplicación ARIMA datos Anuales Puerto Barcelona	93
10.4.4. Resultados aplicación ARIMA datos mensuales Puerto Barcelona	94
10.4.5. Resultados aplicación ARIMA datos anuales Puerto Algeciras	96
10.4.6. Resultados aplicación ARIMA datos mensuales Puerto Algeciras.....	98
10.4.7. Resultados aplicación SARIMA datos mensuales Puerto Valencia.....	101
10.4.8. Resultados aplicación SARIMA datos Mensuales Puerto Barcelona	103
10.4.9. Resultados aplicación SARIMA datos Mensuales Puerto Algeciras.....	105
10.4.10. Resultados aplicación ARIMAX datos Anuales Puerto Valencia.....	107
10.4.11. Resultados aplicación ARIMAX datos Anuales Puerto Barcelona	109
10.4.12. Resultados aplicación ARIMAX datos Anuales Puerto Algeciras	111
10.4.13. Resultados aplicación RNN datos Anuales Puerto Valencia	113
10.4.14. Resultados aplicación RNN datos Anuales Puerto Barcelona	113
10.4.15. Resultados aplicación RNN datos Anuales Puerto Algeciras.....	114
10.4.16. Resultados aplicación RNN datos Mensuales Puerto Valencia	114
10.4.17. Resultados aplicación RNN datos Mensuales Puerto Barcelona.....	115
10.4.18. Resultados aplicación RNN datos Mensuales Puerto Algeciras	115
10.4.19. Resultados aplicación Prophet Mensuales Valencia	116
10.4.20. Resultados aplicación Prophet Mensuales Barcelona	118
10.4.21. Resultados aplicación Prophet Mensuales Algeciras.....	120
10.4. Anexo Qlik Sense	122

Lista de Figuras

Figura 1: Diagrama GANTT del proyecto	15
Figura 2: Esquema módulos a desarrollar	16
Figura 3: Ejemplo regresión lineal	21
Figura 4: Ejemplo árbol de regresión	22
Figura 5: Estadística Mensual de Puertos del Estado	27
Figura 6: Ejemplo de Resumen Mensual de Puertos del Estado	28
Figura 7: Volumen de contenedores por país y total de Europa (2009 al 2018).....	29
Figura 8: Volumen de contenedores por puerto - Top 20 (2009-2018)	29
Figura 9: Lista de los 50 puertos más importantes - Wikipedia	30
Figura 10: Cantidad de Mercancías Transportadas por Regiones.....	30
Figura 11: Atributos de las Diferentes Tablas de la BD	33
Figura 12: Trabajo para la carga de los datos de la tabla spanish_historic_annual	37
Figura 13: Trabajo para la carga de los datos de la tabla spanish_historic_contenedores	38
Figura 14: Ejemplo de Volumen de TEUS Mensual de Puertos del Estado	39
Figura 15: Resumen Registros Almacenados en BD	42
Figura 16: Registros Almacenados en Tabla Spanish_month_contenedores	43
Figura 17: Registros Almacenados en Tabla Spanish_historic_annual	43
Figura 18: Datos Internacionales Cargados	44
Figura 19: PIB Español (x106) y mundial(x109) de 1980 a 2019	45
Figura 20: Gráfica ARIMA Anual Predicción - Puerto de Valencia	49
Figura 21: Gráfica ARIMA Anual Comparativa - Puerto de Valencia	50
Figura 22: Descomposición Serie en tendencia y estacionalidad	51
Figura 23: : Gráfica Volumen TEUS histórico Puerto de Valencia	52
Figura 24: Gráfica PIB Mundial 1980-2019.....	52
Figura 25: : Gráfica PIB Español 1980-2019.....	53
Figura 26: Predicciones del PIB mundial desde el 2020 al 2025	53
Figura 27: Predicción de la Tendencia	55
Figura 28: Predicción de la estacionalidad	55
Figura 29: Predicciones Puerto de Valencia de 2019 a 2024 usando Prophet.....	56
Figura 30: Esquema de datos cargados en Qlik	59
Figura 31: Volumen de TEUS Mensuales	60
Figura 32: Volumen de TEUS Mensuales Filtrado	60
Figura 33: Volumen de TEUS Anuales.....	61
Figura 34: Volumen de TEUS anuales Filtrados	61
Figura 35: Volumen de Total de Mercancías Anuales.....	62
Figura 36: Detalle Tráfico total desglosado por tipo de Mercancías Anuales.....	62
Figura 37: Detalle Tráfico Total Desglosado Filtrado por Puerto y Año	63
Figura 38: Cuadrante Mágico de Plataformas de Inteligencia de Negocio y Analíticas	72
Figura 39: Ejemplo de Cuadro de Mandos con Tableau	74
Figura 40: Ejemplo de Cuadro de Mandos con Power BI.....	74
Figura 41: Ejemplo de Cuadro de Mandos con Qlik.....	75
Figura 42: Ejemplo de Cuadro de Mandos con ThoughtSpot	76
Figura 43: Ejemplo de Cuadro de mando con CARTO.....	77
Figura 44: Ejemplo de Cuadro de mando con Kibana.....	78
Figura 45: Ejemplo de Cuadro de Mando con Grafana.....	79

Figura 46: Esquema Funcionalidades Alfatec	80
Figura 47: Ejemplo de Cuadro de mando con Edisa	81
Figura 48: Ejemplo de Cuadro de mandos a diferentes niveles con INTEGRA2	83
Figura 49: Gráfica ARIMA Anual Predicción - Puerto de Valencia	89
Figura 50: Gráfica ARIMA Anual Comparativa - Puerto de Valencia	90
Figura 51: Gráfica ARIMA Mensual Predicción - Puerto de Valencia	91
Figura 52: Gráfica ARIMA Mensual Comparativa - Puerto de Valencia	92
Figura 53: Gráfica ARIMA Anual Predicción - Puerto de Barcelona	93
Figura 54: Gráfica ARIMA Anual Comparativa - Puerto de Barcelona	94
Figura 55: Gráfica ARIMA Mensual Predicción - Puerto de Barcelona	95
Figura 56: Gráfica ARIMA Mensual Comparativa - Puerto de Barcelona	96
Figura 57: Gráfica ARIMA Anual Predicción - Puerto de Algeciras.....	97
Figura 58: Gráfica ARIMA Anual Comparativa - Puerto de Algeciras.....	98
Figura 59: Gráfica ARIMA Mensual Predicción - Puerto de Algeciras	99
Figura 60: Gráfica ARIMA Mensual Comparativa - Puerto de Algeciras	100
Figura 61: Gráfica SARIMA Mensual Predicción - Puerto de Valencia	101
Figura 62: Gráfica SARIMA Mensual Comparativa - Puerto de Valencia	102
Figura 63: Gráfica SARIMA Mensual Predicción - Puerto de Barcelona.....	103
Figura 64: Gráfica SARIMA Mensual Comparativa - Puerto de Barcelona	104
Figura 65: Gráfica SARIMA Mensual Predicción - Puerto de Algeciras	105
Figura 66: Gráfica SARIMA Mensual Comparativa - Puerto de Algeciras.....	106
Figura 67: Gráfica ARIMAX Anual Predicción - Puerto de Valencia	107
Figura 68: Gráfica ARIMAX Anual Comparativa - Puerto de Valencia.....	108
Figura 69: Gráfica ARIMAX Anual Predicción - Puerto de Barcelona	109
Figura 70: Gráfica ARIMAX Anual Comparativa - Puerto de Barcelona	110
Figura 71: Gráfica ARIMAX Anual Predicción - Puerto de Algeciras.....	111
Figura 72: Gráfica ARIMAX Anual Comparativa - Puerto de Algeciras.....	112
Figura 73: Gráfica RNN Anual Predicción - Puerto de Valencia	113
Figura 74: Gráfica RNN Anual Predicción - Puerto de Barcelona.....	113
Figura 75: Gráfica RNN Anual Predicción - Puerto de Algeciras.....	114
Figura 76: Gráfica RNN Mensual Predicción - Puerto de Valencia.....	114
Figura 77: Gráfica RNN Mensual Predicción - Puerto de Barcelona	115
Figura 78: Gráfica RNN Mensual Predicción - Puerto de Algeciras.....	115
Figura 79: Gráfica Prophet Mensual Predicción - Puerto de Valencia	116
Figura 80: Gráfica Prophet Mensual Comparativa - Puerto de Valencia	117
Figura 81: Gráfica Prophet Mensual Predicción - Puerto de Barcelona.....	118
Figura 82: Gráfica Prophet Mensual Comparativa - Puerto de Barcelona	119
Figura 83: Gráfica Prophet Mensual Predicción - Puerto de Algeciras	120
Figura 84: Gráfica Prophet Mensual Comparativa - Puerto de Algeciras.....	121

Lista de tablas

Tabla 1: Resumen herramientas de visualización revisadas	18
Tabla 2: Datos Históricos Generales Puertos Españoles	27
Tabla 3: Descripción Tablas de BD creadas para la importación	32
Tabla 4: Atributos del Fichero resumengeneral.xls	34
Tabla 5: Atributos del Fichero resumenmercanciassegunpresentacion.xls	35
Tabla 6: Atributos del Fichero resumenotrasmercancias.xls	35
Tabla 7: Atributos del Fichero resumenotrosdatos.xls	36
Tabla 8: Atributos del Fichero contenedorporunidades.xls	36
Tabla 9: Atributos del Fichero contenedoressegunnavegacion.xls	37
Tabla 10: Modelos aplicados a las series anuales y mensuales	46
Tabla 11: Datos de entrenamiento - Validación	46
Tabla 12: Correspondencia entre las libretas Jupyter y los análisis realizados	47
Tabla 13: ARIMA Valores Observados & Predicción Valencia	50
Tabla 14: Predicción PIB Mundial próximos años	54
Tabla 15: Error obtenido por los diferentes algoritmos con las series anuales	57
Tabla 16: Error obtenido por los diferentes algoritmos con las series anuales	57
Tabla 17: ARIMA Valores Anuales Observados & Predicción Valencia	90
Tabla 18: ARIMA Valores Mensuales Observados & Predicción Valencia	92
Tabla 19: ARIMA Valores Anuales Observados & Predicción Barcelona	94
Tabla 20: ARIMA Valores Mensuales Observados & Predicción Barcelona	96
Tabla 21: ARIMA Valores Anuales Observados & Predicción Algeciras	97
Tabla 22: ARIMA Valores Mensuales Observados & Predicción Algeciras	100
Tabla 23: SARIMA Valores Observados & Predicción Valencia	102
Tabla 24: SARIMA Valores Observados & Predicción Barcelona	104
Tabla 25: SARIMA Valores Observados & Predicción Algeciras	106
Tabla 26: ARIMAX Valores Observados & Predicción Valencia	108
Tabla 27: ARIMAX Valores Observados & Predicción Barcelona	110
Tabla 28: ARIMAX Valores Observados & Predicción Algeciras	112
Tabla 29: Prophet Valores Observados & Predicción Valencia	117
Tabla 30: Prophet Valores Observados & Predicción Barcelona	119
Tabla 31: Prophet Valores Observados & Predicción Algeciras	121

1. Introducción

1.1. Contexto y justificación del Trabajo

Actualmente la mayoría de las autoridades portuarias españolas utilizan como herramienta de gestión un producto llamado INTEGRA2, que podría considerarse un ERP portuario, desarrollado por Prodevelop y financiado por Puertos del Estado (Puertos del Estado, s.f.) , que es el organismo público que se encarga de la coordinación y control de la eficiencia del sistema Portuario español.

INTEGRA2 incorpora un módulo de Business Intelligence (BI) que muestra el estado de un puerto desde diferentes vertientes, y que permite a los diferentes roles del puerto controlar el estado del puerto desde diferentes puntos de vista (estratégico y operacional).

Al estar en un mercado tan competitivo, disponer de información centralizada de calidad es clave para la toma de decisiones.

Las soluciones actuales de “Business Intelligence” portuarias no incorporan información de otros puertos nacionales/internacionales para poder compararse. Además, también carecen de técnicas predictivas que les ayuden a predecir tendencias futuras propias o de la competencia a varios años vista. Es decir, los puertos ven una foto/instantánea de su situación actual e histórica, no contextualizada por no contener información de otros puertos. La toma de decisiones, actualmente la realizan los gestores de los puertos en base a los datos del propio puerto e informaciones que obtienen de otras fuentes (no integradas con sus sistemas de BI) y a la experiencia de las personas encargadas de la toma de decisiones.

Este proyecto proporciona a través de una herramienta de BI, cuadros de mando con información relevante de todos los puertos españoles proveniente de orígenes de datos públicos (Puertos del Estado y Eurostat (EuroStat, 2020)).

Gracias a esta información y al uso de diferentes técnicas analíticas predictivas, los puertos podrán conocer la evolución de las diferentes áreas de negocio, compararse con la competencia y obtener predicciones futuras “cantidad de contenedores, beneficio económico global y por tipo de carga,” a corto y medio plazo.

Con esta solución BI, los puertos españoles dispondrán de una potente solución analítica de ayuda a la toma de decisiones, que le acercará al concepto de “puerto analítico” y cuyas decisiones estén basadas en datos.

Los resultados de este TFM, en un futuro se podrán integrar en la herramienta INTEGRA2 de Prodevelop, ya que son necesidades demandadas por los clientes y permitirán dar un salto de calidad en las funcionalidades ofrecidas por dicho producto.

1.2. Objetivos del Trabajo

Los objetivos de trabajo se dividen en objetivos generales (principal y secundario) y objetivos personales.

El objetivo principal del proyecto es dotar a los puertos españoles de una herramienta analítica, que permita compararse con otros puertos a través de la integración de datos públicos históricos existentes y de predicciones futuras en base a los datos existentes. Para alcanzar el objetivo se aplicarán diferentes técnicas estadísticas y analíticas de series temporales.

El objetivo secundario del proyecto sería la integración de los resultados del TFM a una solución empresarial real.

A nivel personal, el TFM tiene como objetivo el poner en práctica los conocimientos adquiridos en el máster para el desarrollo de un caso real completo.

1.3. Enfoque y método seguido

La metodología seguida para el desarrollo del TFM, está basada en una metodología tradicional para resolver proyectos de I+D y que está muy implantada en Prodevelop, que trata por un lado de minimizar los riesgos que conlleva la ejecución de proyectos innovadores, y por otro lado, trata de conseguir un producto innovador.

Los trabajos a desarrollar se agruparán en paquetes de trabajo (PT) y estos se descomponen en tareas (T). Además por cada paquete de trabajo se realizarán una serie de entregas software (SW) o documentación (DOC) como justificación de los mismos.

PT1: Gestión del proyecto

Este paquete de trabajo aglutina todas las tareas de gestión que conlleva la ejecución del proyecto

T1.1- Definir metodología de trabajo

T1.1	Definir metodología de trabajo
Fechas	16/09/2020 - 01/10/2020
Desc.	Definir y validar la metodología de trabajo a seguir durante todo el proyecto
Entregables	DOC 1- PAC1 – Definición y planificación del trabajo final

T1.2- Reportes periódicos

T1.1	Reportes periódicos (cada 2 semanas)
Fechas	01/10/2020 - 10/01/2021
Desc.	Cada dos semanas se realizará un informe de las tareas realizadas y se planificarán las tareas para el siguiente periodo.
Entregables	DOC 2.X- Informes periódicos con las tareas realizadas y el plan para el siguiente periodo.

PT2- Análisis de mercado y toma de requisitos

En este paquete de trabajo se realizará un estudio de los diferentes productos de BI existentes en el ámbito portuario y las diferentes fuentes de datos interesantes a usar en el proyecto.

Por otro lado, en base al análisis del mercado, fuentes de datos existentes y estado del producto Posidonia BI, se concretará el alcance real del presente TFM.

T2.1	Análisis del mercado
Fechas	28/09/2020 - 18/10/2020
Desc.	En esta tarea se realizará un estudio de los diferentes productos de BI existentes en el ámbito portuario y las diferentes fuentes de datos interesantes a usar en el proyecto.
Entregables	DOC 3.1- Estado del arte - BI portuario DOC 3.2- Fuentes de datos abiertas - BI portuario DOC 3.3- PAC2 - Estado del arte y análisis de mercado

T2.2	Requisitos del proyecto
Fechas	28/09/2020 - 18/10/2020
Desc.	En base al análisis del mercado, fuentes de datos existentes y estado del producto Posidonia BI, se concretará el alcance real del presente TFM.

Entregables	DOC 3.4- Alcance técnico del TFM
--------------------	----------------------------------

PT3 Desarrollos técnicos

En este paquete de trabajo se realizarán los diferentes desarrollos técnicos del proyecto, incluyendo la adquisición de datos desde fuentes externas y la realización de las diferentes analíticas.

T3.1	Adquisición de datos puertos del estado
Fechas	18/10/2020 - 15/11/2020
Desc.	Realizar la adquisición de datos desde el “open data” de Puertos del estado
Entregables	SW 1- Módulo de adquisición de datos de puertos nacionales

T3.2	Adquisición de datos puertos europeos
Fechas	18/10/2020 - 15/11/2020
Desc.	Realizar la adquisición de datos de puertos europeos desde diferentes fuentes.
Entregables	SW 2- Módulo de adquisición de datos de puertos internacionales.

T3.3	Analíticas
Fechas	15/11/2020-11/12/2020
Desc.	Uso de técnicas analíticas para realización de predicciones portuarias
Entregables	SW 3- Módulo de predicciones portuarias

PT4 Piloto

Probar los desarrollos realizados utilizando un puerto español

T4.1	Analíticas
Fechas	01/12/2020-20/12/2020
Desc.	Desarrollo de un cuadro de mandos aplicando las predicciones realizadas en el PT3
Entregables	DOC 4- PAC 3 – Diseño e implementación del trabajo

PT5 Diseminación

Preparar la documentación y la presentación para la defensa del proyecto

T5.1	Memoria TFM
Fechas	21/12/2020-03/01/2021
Desc.	Escribir memoria del proyecto
Entregables	DOC 5- PAC 4 – Redacción de la memoria

T5.2	Presentación TFM
Fechas	04/01/2021-10/01/2021
Desc.	Preparar defensa del proyecto
Entregables	DOC 6- PAC 5 – Presentación y defensa del proyecto

Aunque el trabajo a desarrollar en este proyecto se puede dividir en varios módulos independientes que de forma conjunta pueden constituir un producto viable por sí mismo. El objetivo futuro es integrarlos en el producto de Prodevelop "INTEGRA2" que es el producto líder de BI el mercado nacional portuario. De este modo se complementará este producto, ya implantado en el mercado nacional, con algunas de las carencias más demandadas en los puertos hoy en día.

Los módulos a desarrollar a lo largo del trabajo son:

- Adquisición de datos de actividad portuaria española (origen: puertos del estado)
- Adquisición de datos de actividad portuaria europea (origen: Eurostats, revistas especializadas del sector marítimo y otros organismos europeos)
- Módulo de predicciones basado en diferentes técnicas de analíticas de datos

1.4. Planificación del Trabajo

GANTT

PT	Tarea	Inicio	Fin	Septiembre		Octubre				Noviembre				Diciembre				Enero	
				S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
1	T1.1. Definir metodología de trabajo	16/09/2020	01/10/2020	█	█														
1	T1.2. Reportes periódicos	01/10/2020	10/01/2021	█	█	█	█	█	█	█	█	█	█	█	█	█	█	█	
2	T2.1. Análisis del mercado	29/09/2020	18/10/2020		█	█	█												
2	T2.2. Requisitos del proyecto	28/09/2020	18/10/2020		█	█													
3	T3.1. Adquisición de datos españoles	18/10/2020	15/11/2020				█	█	█										
3	T3.2. Adquisición de datos internacional	18/10/2020	15/11/2020				█	█	█										
3	T3.3. Análíticas	15/11/2020	11/12/2020							█	█	█							
4	T4.1 Piloto	01/12/2020	20/12/2020								█	█	█						
5	T5.1. Memoria TFM	21/12/2020	03/01/2021									█	█	█					
5	T5.2. Presentacion TFM	04/01/2021	10/01/2021														█	█	

Figura 1: Diagrama GANTT del proyecto

1.5. Breve resumen de productos obtenidos

Como resultado del Proyecto se desarrollarán varios módulos de adquisición de datos (ETL), una Base de datos para almacenar la información, un módulo de análisis predictivo que permitirá usar diferentes técnicas analíticas con el propósito de poder hacer predicciones de la evolución de un puerto a corto/medio plazo, y un conjunto de cuadros de mando, dotando a los gestores el puerto de una herramienta completa para facilitar la toma de decisiones.

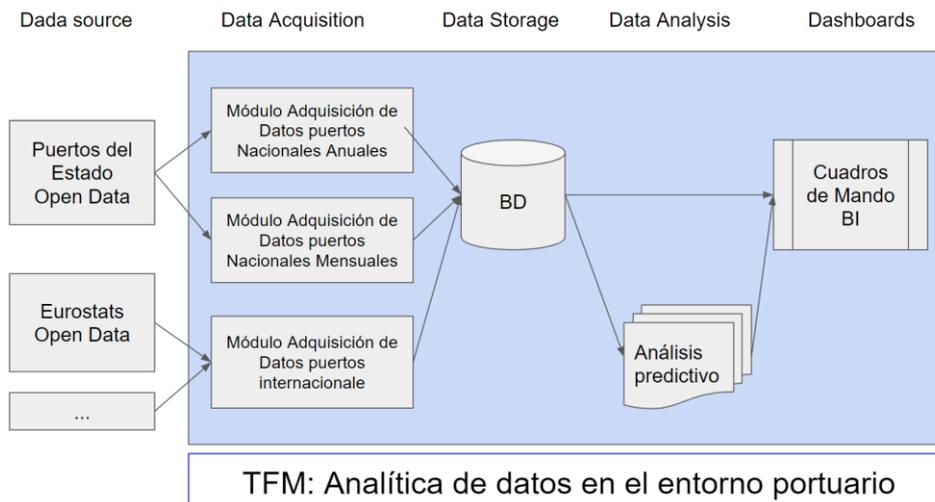


Figura 2: Esquema módulos a desarrollar

1.6. Breve descripción de los otros capítulos de la memoria

Estado del arte y alcance del proyecto. En este capítulo se ha realizado un repaso a las diferentes soluciones de “Business Intelligence”, alternativas relacionadas con predicciones de series temporales y se han identificado posibles fuentes de datos útiles para el TFM.

Alcance del proyecto. En este capítulo se ha definido el alcance del proyecto en base a las averiguaciones realizadas en el capítulo anterior.

Carga de datos e integración. En este capítulo se explica cómo se han realizado las diferentes cargas de datos tanto de puertos nacionales como internacionales.

Análisis de Series temporales. En este capítulo se han analizado 5 alternativas para predecir volúmenes futuros en base a series temporales.

Cuadros de Mando. En este capítulo se explican los diferentes cuadros de mando creados para visualizar la información del proyecto

Conclusiones y Trabajos futuros. En este capítulo se recorren las principales conclusiones del trabajo y futuras ampliaciones que se podrían hacer en un futuro.

2. Estado del Arte.

2.1. Business Intelligence (BI)

En este apartado se hablará de soluciones BI disponibles en el mercado, algunas de ellas son soluciones BI generalistas multipropósito, en cambio, otras son soluciones de propósito más específico como por ejemplo, para el análisis de datos en tiempo real y para visualizar información geográfica.

Debido a la extensión de este apartado, se ha movido a un anexo el detalle de las herramientas analizadas.

A modo resumen se puede afirmar que las soluciones basadas en productos BI genéricos están evolucionando mucho en relación a la gestión de los datos y a la inclusión de analíticas. Las 4 herramientas estudiadas en profundidad son todas igual de válidas. La elección de una u otra dependerá de los cuadros de mandos a realizar y del conocimiento que tenga la empresa de la misma.

Herramienta	Pros	Contras
Librería RShiny	Adecuada para crear visualizaciones en R.	Solo funciona con R. Solo visualizaciones, no cuadros de mandos.
D3.js	La librería JS más usada, contiene muchos tipos de visualización.	Requiere conocimientos de programación. No está pensada para cuadros de mando.
Freeboard	Va más allá de las librerías de visualizaciones y permite crear tus propios cuadros de mando de manera sencilla. No se requiere conocimientos de programación.	No es una solución completa de BI.
Tableau	Una de las herramientas de BI más usadas. Muy potente e intuitiva. Capacidades de descubrimiento y analíticas.	Funcionalidades avanzadas complejas.
Power BI	Herramienta líder del mercado, integrada con Office. funcionalidad para descubrimiento de datos, cuadros de mandos interactivos y analíticas avanzadas.	Integrada sólo con Azure, la versión cloud es más limitada que la versión local.
Qlik Sense	Capacidades analíticas avanzadas. Solución on-premise/cloud. Documentación extensa.	No existe versión gratuita y es costosa.
ThoughtSpot	Funcionalidad de Búsqueda basada en lenguaje natural y capacidad para procesar grandes cantidades de datos.	No incluye una herramienta de preparación de datos tipo ETL.
CARTO	Especializada en análisis espacial.	Solución muy cara, sólo apropiada para grandes proyectos y requiere de grandes conocimientos y esfuerzos de programación.

Elastic	Especializada en series temporales, Incorpora una de las bases de datos NoSQL. La versión gratuita sirve en la mayoría de los casos.	La realización de cuadros de mandos no es tan intuitiva como en herramientas BI. Solo se integra con elasticsearch. Algunos plugins básicos son de pago.
Grafana	Es la alternativa a Elastic para series temporales. Se integra con muchas fuentes de datos, aunque algunas son de pago. Incluye un buen sistema de alarmas.	La realización de cuadros de mandos requiere de conocimientos avanzados.

Tabla 1: Resumen herramientas de visualización revisadas

En cuanto a las soluciones BI para el sector portuario, se han revisado diferentes soluciones utilizadas en el sector portuario español.

Las empresas más relevantes en este sector son: **Alfatec Sistemas, Edisa , Proyecto BrainPort Analytics, Puerto de Valencia y Prodevelop**. Para más información sobre las empresas y sus productos, se puede consultar el anexo correspondiente.

A modo resumen, podemos decir que las soluciones de BI en el sector portuario nacional están basadas en su mayoría en Qlik y muestran la situación histórica y actual de un puerto en concreto.

Limitaciones de las soluciones actuales:

- Muestran una foto actual del puerto y su evolución en el tiempo. No incluyen capacidades predictivas sobre el futuro comportamiento del puerto.
- No comparan la situación del puerto con otros puertos nacionales y extranjeros para tener un mejor contexto de cómo se comporta el puerto en relación a su sector.

Estas dos limitaciones son las que trata de abordar el presente trabajo. Con dicha información, por un lado, podremos comparar la evolución de un puerto respecto al sector y su competencia, y por otro lado se podrá predecir la evolución futura del puerto por medio de técnicas analíticas predictivas basadas en Inteligencia artificial. En el siguiente punto se analizarán las diferentes técnicas predictivas existentes para realizar cálculos predictivos en el sector portuario.

2.2. Uso de la IA en el entorno Portuario

Un puerto a grandes rasgos puede verse como una ciudad industrial y desde hace unos años se está produciendo una voraz revolución digital basada en dos paradigmas tecnológicos.

- Smart cities/ Smart ports - Donde se trata al puerto como una entidad global comprometida con el medioambiente, sostenible y resiliente.
- Industria 4.0. Trata de optimizar los procesos portuarios con el fin de hacerlos más eficientes y eliminar los cuellos de botella existentes.

La presente sección, por un lado, explica a nivel general las tendencias en IA que se están investigando en la comunidad portuaria desde dos aproximaciones: “Smart ports” e “Industria 4.0”. Por otro lado, detalla las tendencias en IA para predecir el volumen de negocios en Puertos, que es el problema que trata de resolver el presente trabajo.

2.2.1. Uso de la IA desde el punto de vista Smart Port

El concepto de smart port o puerto inteligente se basa en el uso de la tecnología para digitalizar y optimizar procesos, facilitar la conexión y el intercambio de información entre los diferentes actores de la cadena logística y ganar eficiencia; donde el análisis de estos datos recopilados permite la toma de mejores decisiones. Avanzar hacia un modelo de “Smart Port” es, en definitiva, evolucionar hacia un puerto más eficiente, sostenible (Transeop, 2020) .

Internet de las cosas (IoT) tiene un papel clave en el concepto de “Smart Port”, ya que gracias al despliegue de sensores, se puede captar información y monitorizar en tiempo real los parámetros relevantes. Estos sensores generan gran cantidad de datos (Big Data) y su análisis permite llevar un control óptimo de la actividad portuaria.

A groso modo, un puerto puede entenderse como una ciudad industrial cuyas necesidades son similares a las de una ciudad a nivel global. La mayor preocupación de los puertos radica en proporcionar un porvenir a corto y medio plazo de la comunidad portuaria. Esta visión está liderada siempre por las Autoridades portuarias, que velan por el buen funcionamiento del puerto, su sostenibilidad y resiliencia.

Desde esta perspectiva, los “Smart Port” usan técnicas de IA para monitorizar el impacto ambiental, seguridad, consumo energético, gestión de vehículos/ buques y optimización de las actividades portuarias.

Algunos de los verticales / aplicaciones que usan IA en este ámbito son:

- Identificación de vehículos a través de cámaras inteligentes que leen la matrícula de los vehículos o mediante dispositivos GPS.
- Ocupación real de parkings y recomendaciones para transporte.
- Detección de intrusiones en el perímetro del puerto a través de cámaras.
- Detección de embarcaciones sospechosas en el puerto mediante cámaras, radares y detección de patrones de comportamientos anómalos.
- Gestión de citas previas de entrada al puerto de camiones en base a predicciones para minimizar congestiones.
- Monitorización del impacto ambiental (Contaminación del aire, ruido, luminosidad, calidad del agua) producido por las actividades portuarias mediante el uso de sensores y modelos de dispersión de aire, ruido y luz.
- Gestión inteligente de energía (smart Grid) para la realización de las operaciones y para abastecer a los buques atracados.
- Monitorización y detección temprana de vertidos en el agua y localización del causante de la misma o bien accidentalmente o bien intencionadamente.
- La predicción de la llegada al puerto de los buques “Estimated Time of Arrival (ETA)”, es sin duda alguna la aplicación que más impacto tiene en los puertos, porque condiciona toda la actividad portuaria.

2.2.2. Uso de AI en Industria 4.0 (aplicada al sector portuario-logístico)

Desde la perspectiva de la Industria 4.0, vemos al puerto como un hub logístico, cuyo objetivo es la optimización de la cadena de suministro, tratando de alcanzar el sistema Just-in-time, que es un sistema de organización de la producción que usan muchas industrias (Monden, 1996,

pág. 22) . En el caso de la logística portuaria, aún se está lejos de dicho objetivo, debido principalmente a la falta de estándares para el intercambio de datos entre todos los actores involucrados en el proceso logístico (IMO, 2020).

Los actores principales en este caso son las terminales que se encargan de la carga y la descarga de mercancías en los buques. Actualmente existe mucha diversidad en las terminales y se pueden encontrar desde terminales 100% manuales a terminales automatizadas con apenas operarios y maquinaria automatizada. Esto se debe a que cualquier modernización de una terminal requiere de inversiones millonarias y la amortización de las inversiones, incluidas las grúas es de varias decenas de años. Por eso es frecuente ver grúas de más de 20 años de antigüedad totalmente operativas.

Dada la gran diversidad de terminales y sus diferentes niveles de automatización y digitalización, los objetivos de modernización difieren considerablemente. A continuación, se listan una conjunto de usos de la IA para la optimización de la cadena de suministro:

- Optimización de las operaciones de cargas y descarga para minimizar cuellos de botella.
- Determinar la demanda futura.
- Minimizar el movimiento de contenedores.
- Optimizar rutas de vehículos.
- Automatización de vehículos/ maquinarias y mantenimiento preventivo.
- Identificación de contenedores y vehículos por medio de visión artificial e IoT.
- Optimización de las colas de vehículos.
- Planificación de RRHH en función de la demanda.
- Gestión de citas previas de entrada al puerto y a terminales de camiones en base a predicciones de colas y a disponibilidad de mercancía/buque/almacén.
- Determinar la mejor hora de carga y descarga, considerando las previsiones meteorológicas para minimizar el impacto ambiental.
- Predicción ETA, La llegada/salida de un buque de puerto es el elemento fundamental en el hub logístico y condiciona el resto de las operaciones (Pani, Vanelslander, Fancello, & Cannas, 2020).

2.2.3. AI para predicciones del volumen de negocio en puertos

En esta sección se explican las tendencias/aproximaciones actuales para el cálculo de predicciones del volumen de negocio en puerto, que es el objetivo del presente trabajo.

Dado lo rápido que evoluciona el comercio mundial y la importancia del sector marítimo, la previsión del volumen de negocio de un puerto es de vital importancia para adecuar sus infraestructuras a la futura demanda. Existen muchos métodos de previsión de series cronológicas que pueden utilizarse para pronosticar el tráfico de contenedores de un puerto.

Comenzaremos el apartado explicando las diferentes alternativas que tenemos a la hora de realizar análisis predictivos, para a continuación seguir con su aplicación para obtener predicciones del volumen de negocio en puertos. Finalizaremos la sección con las lecciones aprendidas y con las posibles técnicas y parámetros a estudiar en este trabajo.

2.2.3.1. Modelado y Análisis Predictivo

El Análisis Predictivo es una disciplina del análisis de datos que usa técnicas de estadística, como aprendizaje computacional o minería de datos, para desarrollar modelos que predicen eventos futuros o conductas. Estos modelos predictivos permiten aprovechar los patrones de comportamiento encontrados en los datos actuales e históricos para identificar riesgos y oportunidades. Estos modelos se pueden utilizar para predicciones de todo tipo, desde predicciones económicas a audiencias televisivas. En el presente trabajo se usarán técnicas predictivas para predecir el volumen de negocio en los puertos españoles.

El análisis predictivo se fundamenta en la identificación de relaciones entre variables en eventos pasados, para luego explotar dichas relaciones y predecir posibles resultados en futuras situaciones.

Los enfoques y técnicas utilizados para realizar el análisis predictivo pueden agruparse de una manera muy general en técnicas de regresión y técnicas de aprendizaje computacional.

- **Técnicas de regresión.** Los modelos de regresión son el pilar de la analítica predictiva. El enfoque se basa en el establecimiento de una ecuación matemática como modelo para representar las interacciones entre las diferentes variables en consideración. Dependiendo de la situación, hay una gran variedad de modelos que se pueden aplicar durante la realización del análisis predictivo.
- **Modelo de regresión lineal.** Este modelo analiza la relación de la variable dependiente con un conjunto de variables independientes. Esta relación se expresa como una ecuación lineal.

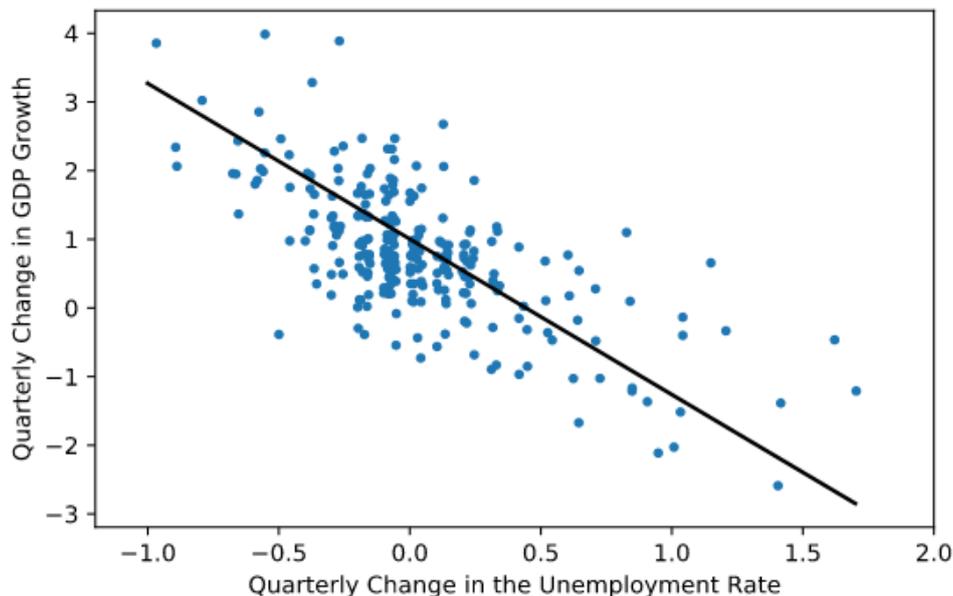


Figura 3: Ejemplo regresión lineal

Fuente: https://commons.wikimedia.org/wiki/File:Okun%27s_Law.svg

- **Árboles de clasificación o regresión “Classification And Regression Trees (CART)”**. Son una alternativa a los modelos de regresión, en este caso se pretende predecir el valor de una variable categórica dependiendo del valor de otras variables

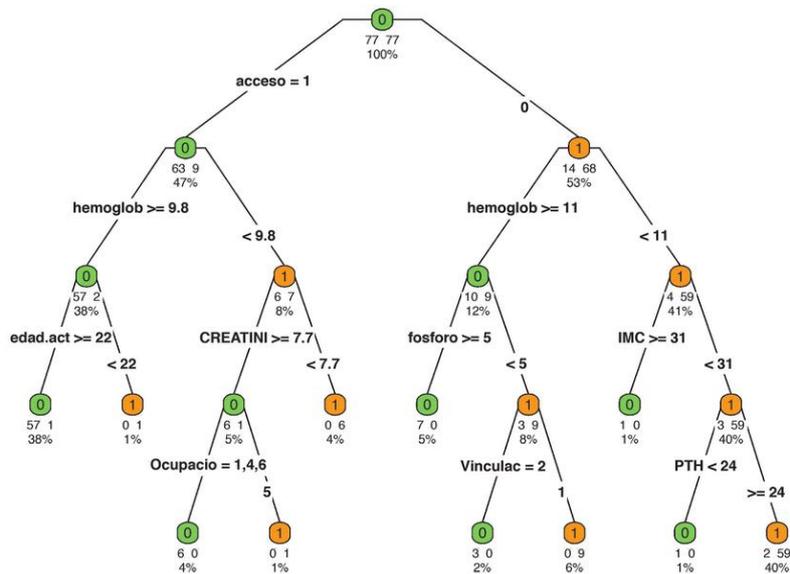


Figura 4: Ejemplo árbol de regresión

Fuente: <https://www.researchgate.net/publication/321073328/figure/fig1/AS:560916036501504@1510744212189/Figura-1-Arboles-de-clasificacion-y-regresion-Classification-and-Regression-Tree-CART.png>

- **Técnicas de aprendizaje automático.** Permiten predecir la variable dependiente sin tener en cuenta las relaciones subyacentes entre variables. En este caso el aprendizaje automático trata de emular el proceso de aprendizaje humano y aprende a través del entrenamiento para poder llegar a predecir eventos futuros.
 - **Redes neuronales.** Esta técnica trata de imitar el funcionamiento de las redes neuronales cerebrales. Con el entrenamiento, las neuronas van creando y reforzando ciertas conexiones para "aprender" algo. Una vez entrenada, la red se puede usar para hacer predicciones o clasificaciones. Dentro de las redes neuronales existen multitud variantes que pueden funcionar mejor para problemas determinados, como puede ser el procesamiento de imágenes o el procesamiento de textos.
 - **Support Vector Machines "SVM".** Son un conjunto de algoritmos de aprendizaje supervisado indicados para problemas de clasificación y regresión. Un modelo SVM representa las muestras separadas mediante hiperplanos que nos permiten clasificar nuevas muestras.
 - **Naïve Bayes.** un clasificador Naïve Bayes es un clasificador probabilístico fundamentado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales

Análisis de Series temporales

En este punto se realiza una pequeña introducción a las series temporales, que son fundamentales en el campo de la economía para calcular predicciones en base a observaciones pasadas.

Una serie temporal es una sucesión de observaciones de una/s variable/s realizadas a lo largo del tiempo, a intervalos regulares (Parra, 2019).

Existen diferentes tipos de series temporales según cómo y cuándo realicemos las mediciones:

- Continuas o discretas, dependiendo del intervalo en el que se recojan las muestras.
- Flujo o Stock. Se dice que es de flujo si comprende un periodo de tiempo (un año, un mes) y en caso de ser un momento puntual se dice que es de stock.
- Periodicidad, cada cuanto tiempo se obtienen las muestras.

Para que el análisis de las series sea de utilidad debemos tener cierta estabilidad en los datos y estos deben de ser homogéneos en el tiempo. El objetivo del análisis de las series es determinar el comportamiento de la variable a lo largo del tiempo, para poder predecir el comportamiento futuro suponiendo que no se producen cambios estructurales significativos. Actualmente, con la crisis del COVID-19 que estamos viviendo, no se podrían aplicar modelos de series temporales para determinar el comportamiento futuro.

La calidad de las previsiones dependerá de si la variable a predecir sigue algún patrón más o menos fijo o recurrente.

Dentro de los métodos de predicción cuantitativos (en base a patrones de conducta seguidos en el pasado), distinguimos dos aproximaciones:

- Análisis univariante de series temporales, las predicciones de los valores de una variable se realizan en base a la información contenida en los valores de la propia serie temporal. Dentro de esta aproximación se incluyen los métodos de descomposición y la familia de modelos ARIMA univariantes.
- Análisis multivariante o de tipo causal, para la realización de las predicciones de las variables se usa información de otras variables adicionales.

Modelos econométricos

Como en el presente trabajo vamos a desarrollar modelos predictivos para calcular el volumen de negocio de un puerto, describiremos los modelos econométricos (Caridad y Ocerin, 2008). La econometría se puede considerar como una rama de la teoría económica, en la que se utilizan métodos y técnicas de estadística matemática en la estimación de relaciones económicas.

La teoría económica postula una serie de relaciones causales entre diversas magnitudes económicas. En cambio, un modelo econométrico está formado por una o varias ecuaciones en las que la variable explicada o endógena depende de una o varias variables explicativas. Por ejemplo, el volumen de negocio de una empresa de consumo podría depender del PIB y de la inversión en publicidad.

$$V_t = A + B_1X_t + B_2P_t + E_t$$

Un modelo econométrico está formado por :

- Una o varias ecuaciones o relaciones estructurales
- variables explicativas y explicadas
- parámetros (A+ B) a estimar
- un conjunto de observaciones o datos para el proceso de estimación.

2.2.3.1. *Análisis predictivo del volumen de negocio en Puertos*

En este punto abordaremos cómo se está tratando de predecir el volumen de negocio según los diferentes estudios publicados. Nos centraremos en el cálculo del volumen de contenedores movidos, ya que es la medida más común / estándar a la hora de establecer los diferentes rankings de puertos nacionales e internacionales.

Tradicionalmente se usaban modelos de regresión para realizar predicciones, estos modelos no incluyen inteligencia y se basan en ecuaciones matemáticas. Por contra, los modelos basados en aprendizaje automático aprenden patrones no lineales a partir de los datos sin necesidad de tener una fórmula preestablecida.

En el artículo **“A comparison of time series methods for forecasting container throughput”** (Hing Kai, Shuojiang, & Xiaoguang, 2019) se introducen diferentes modelos basados en series temporales para la realización de predicciones y se comparan seis de ellos (“Moving Average”, “Multivariate Adaptive Regression Splines”, “Autoregressive Integrated Moving Average”, “Grey Model”, “Artificial Neural Network” y “Support Vector Regression-SVR”) aplicados a los mismos datos para predecir el volumen de negocio de un puerto.

El modelo SVR fue el que dio mejores resultados, pero se tiene que considerar que la serie de datos era corta y no se consideraron datos socio-económicos. Según este estudio los modelos basados en Machine Learning dan mejores resultados que los métodos tradicionales.

En el artículo **“A comparison of Methods for Forecasting the Container Throughput in North Adriatic Ports”** (Dragam, Kramberger, & Intihar, 2014), se analizan diferentes aproximaciones para obtener la predicción de contenedores en puertos del mar Báltico usando tres técnicas analíticas (“Holt-Winters exponential smoothing model”, “classical decomposition model” y “AutoRegressive Integrated Moving-Average -ARIMA”). Los modelos fueron entrenados con 10 años de datos (divididos en trimestres). Según el experimento el modelo ARIMA dio mejores resultados. Hay que considerar que la serie de datos abarca el periodo del 2002 al 2012 y durante dicho periodo hubo una gran crisis económica que afecta directamente a la corrección de las predicciones que afecta principalmente a modelos simples como son los “classical decomposition model”.

En el artículo **“The impact of macroeconomic indicators of forecasting a total cargo throughput in the Adriatic seaport”** (Dragan, Lisec, Intihar, & Kramberger, 2017) se realizó un estudio para evaluar el impacto de los indicadores macroeconómicos para predecir el volumen de negocio en puertos. El análisis se basó en una combinación de los modelos “Dynamic Factor Model- DFM” y “ARIMAX”. Se usó el modelo DFM para extraer información obtenida de variables macroeconómicas. Esta información fue usada como entradas para el modelo ARIMAX, el cual se encargó de obtener las predicciones. El resultado del estudio concluye que al introducir algunos indicadores macroeconómicos se mejoran las predicciones en cuanto al total de carga de un puerto.

La inclusión de variables macroeconómicas es normalmente usada para mejorar las predicciones en diversos ámbitos. La elección de estas variables no es un tema sencillo. En caso de usar muchas variables macroeconómicas, para no aumentar mucho la dimensionalidad de los

cálculos, es recomendable usar técnicas de reducción de dimensionalidad. En estudio contempla variables de dos aspectos diferentes, por un lado variables macroeconómicas (por ejemplo: Producto Interior Bruto- PIB, Exportaciones) y por otro lado variables geográficas (por ejemplo: Países o regiones de influencia).

En el artículo **“A comparison of univariate methods for forecasting container throughput volumes”** (Wen & Ching, 2009) se analizan 6 modelos (“Classical Decomposition Model”, “Trigonometric Regression Model”, “Regression model with seasonal dummy variables”, “Grey model”, “Hybrid grey model”, y “SARIMA”) predictivos para predecir la cantidad de contenedores a corto plazo, para ello se usan datos mensuales solo de 3 años y se prueban los modelos en los puertos Keelung, Taichung y Kaohsiung en Taiwán. Tal como los propios autores reconocen, les faltan más datos para entrenar mejor los modelos. Tras las pruebas concluyen que el “Classical Decomposition” es el que mejores predicciones proporciona a corto plazo (1-2 años vista), y que no siempre un modelo complejo, proporciona mejores resultados.

En el artículo **“Container port volume forecasting framework with the implications on dredging projects”** (Sang-Yoon & Hyunwoo, 2017) se introducen variables relacionadas con la mejora de las infraestructuras portuarias para realizar las predicciones. La mayoría de los estudios sobre la predicción del volumen en puertos se realiza extrapolando datos históricos junto con las predicciones de crecimiento económicos global. En cambio hay pocos estudios que tratan la predicción de crecimiento considerando la mejora en las infraestructuras (mejores máquinas, más calado, nueva terminal, nuevas rutas o ampliación de un canal.) y estrategias comerciales (nuevas rutas, cadenas de suministros, negociaciones con navieras,), debido principalmente a la falta de datos. El artículo trata de determinar el crecimiento de un puerto en base a una operación de dragado que va a permitir que el puerto aumente el tamaño de los buques con los que puede operar, pasado de operar con barcos de 4000 TEUS a barcos de 8000 TEUS. Esta nueva información es importante porque los criterios principales de navieras u operadores logísticos para elegir un puerto son la distancia, tiempo de navegación, costes de navegación, costes del puerto, servicios portuarios (eficiencia y fiabilidad en operaciones). Para realizar el estudio se usó el modelo “ARIMA” y se pasaron encuestas a navieras para preguntarles su interés en cambiar de puerto en base a sus nuevas capacidades. Los resultados son prometedores pero las predicciones dependen principalmente de la calidad de los datos de la encuesta

En el artículo **“Short-term forecast of container throughput: An ARIMA-intervention model for the port of Antwerp”** (Rashed, Meersman, Van de Voorde, & Vanelslander, 2017) se usan los modelos ARIMA y ARIMAX junto con indicadores económicos para predecir el volumen de contenedores a corto plazo. Para la realización del estudio se emplean datos mensuales correspondientes al periodo de 1995 a 2015. Se observa una fuerte correlación de la actividad portuaria con la economía global y se ha evaluado el impacto de la crisis del 2008 en la predicción del volumen de contenedores tras la crisis. El modelo ARIMAX proporcionó mejores predicciones.

2.2.4. Aplicación de AI el presente trabajo

Tras estudiar las diferentes alternativas que tenemos a la hora de realizar análisis predictivos y ver cómo se aplican para obtener predicciones del volumen de negocio en puertos, se han acotado los posibles modelos a aplicar y se ha clarificado las posibles variables a usar

A la hora de la selección de modelos para predecir el volumen de contenedores en puertos, se evaluarán modelos basados en Redes Neuronales, SVM y ARIMA.

A la hora de definir la información a utilizar, parece apropiado usar variables macroeconómicas y geográficas, ya que queda demostrado que estas influyen en el volumen de negocio. Algunas de las variables macroeconómicas a probar son: PIB, importaciones y exportaciones. Estas variables se referirán a los países y áreas geográficas más influyentes para el sector portuario nacional, que en este caso son “EEUU”, “China” y “Europa”.

Se descarta el uso de información referida a las infraestructuras portuarias, por la dificultad de obtener series temporales sobre las mismas.

2.3. Fuentes de datos abiertas

2.3.1. Datos de puertos españoles

Puertos del Estado es una entidad pública que depende del Ministerio de Transportes, Movilidad y Agenda Urbana de España, y que se encarga de velar por el buen funcionamiento del sistema portuario español. Puertos del estado contiene un histórico de datos del volumen de carga de los diferentes puertos. Estos datos se pueden obtener de la web de Puertos del Estado.

El trabajo hará uso de datos públicos de los diferentes tipos de carga, haciendo hincapié en el volumen de contenedores, que es la medida más estándar a la hora de realizar comparaciones y rankings entre puertos.

Los datos disponibles se pueden agrupar en dos tipos, dependiendo de la antigüedad:

Datos antiguos (1962-2009)

Los datos antiguos están disponibles desde el menú “Estadística Histórica” y recopila datos del volumen de carga de los puertos desde 1962. En este caso tenemos los datos agrupados anualmente y estos se pueden descargar en una Excel.

Este conjunto de datos generales está formado por 1373 registros de 11 atributos y contiene los registros desde 1962 hasta 2009. A continuación se muestra un subconjunto de datos a modo de ejemplo:

Año	Autoridad Portuaria	Graneles Líquidos	Graneles Sólidos	Mercancía General	Pesca Fresca	Avituallamiento	Tráfico Local	TOTAL TRAFICO	TEUS	Buques
2009	A Coruña	6.820.497	3.215.589	1.460.292	43.108	93.570	283.941	11.916.997	7.778	1.231
2009	Alicante	114.588	1.111.169	1.260.064	511	24.400	3	2.510.735	132.059	935
2009	Almería	1.502	3.291.672	542.994	4.037	117.334	0	3.957.539	1.425	1.999
2009	Avilés	673.268	2.293.135	984.041	13.590	35.981	0	4.000.015	2.575	700
2009	Bahía de Algeciras	20.142.781	1.743.479	42.316.996	2.168	3.375.642	2.329.889	69.910.955	3.043.268	24.852

Tabla 2: Datos Históricos Generales Puertos Españoles

Datos recientes (2009- actualmente)

Los datos recientes están disponibles desde la opción “Estadística Mensual”, y comprenden datos mensuales desde enero de 2009 hasta la fecha actual.

The screenshot shows the 'Puertos del Estado' website interface. At the top, there is a header with the logo and navigation options. Below the header, there is a search bar and a navigation menu. The main content area is titled 'Estadística mensual' and features a dropdown menu for selecting the year, currently set to '2020'. Below this, there are two columns of download links for 'Formato Excel' and 'Formato PDF'. Each column lists months from January to August 2020, along with the file size in KB. For example, '01 Enero 2020' has a file size of 1396 KB in Excel and 921 KB in PDF. At the bottom left, there is a logo for 'Logistics made perfect'.

Figura 5: Estadística Mensual de Puertos del Estado

Para cada mensualidad se puede descargar la información de los diferentes puertos. A diferencia de la estadística histórica, en este caso se dispone de muchos más atributos y esta viene desglosada por meses en vez de anualmente.

La información de las diferentes mensualidades se puede descargar en dos formatos: Excel y PDF, ambos formatos contienen la misma información

A modo resumen se presenta una primera tabla que contiene la información general a nivel nacional

CONCEPTO			Mes Enero		Acumulado desde Enero		Variación	
			2019	2020	2019	2020	Diferencia	%
MERCANCÍAS SEGÚN SU PRESENTACIÓN	GRANELES	LÍQUIDOS	15.006.504	16.106.749	15.006.504	16.106.749	1.100.245	7,33
		SÓLIDOS	9.115.627	7.102.858	9.115.627	7.102.858	-2.012.769	-22,08
	MERCANCÍA GRAL.	CONVENCIONAL	5.927.745	5.927.538	5.927.745	5.927.538	-207	0,00
		EN CONTENEDORES	15.689.208	15.585.857	15.689.208	15.585.857	-103.351	-0,66
		TOTAL	21.616.953	21.513.395	21.616.953	21.513.395	-103.558	-0,48
TOTAL			45.739.084	44.723.002	45.739.084	44.723.002	-1.016.082	-2,22
OTRAS MERCANCÍAS	PESCA	TOTAL PESCA	12.910	9.218	12.910	9.218	-3.692	-28,60
		COMBUSTIBLES LÍQUIDOS	695.376	590.309	695.376	590.309	-105.067	-15,11
	AVITUALLAMIENTO	OTROS	135.530	135.624	135.530	135.624	94	0,07
		TRÁFICO INTERIOR	TOTAL TRÁFICO INTERIOR	237.767	149.156	237.767	149.156	-88.611
	TOTAL			1.081.583	884.307	1.081.583	884.307	-197.276
TOTAL TRÁFICO PORTUARIO (*)			46.820.667	45.607.309	46.820.667	45.607.309	-1.213.358	-2,59
OTRAS INFORMACIONES	MERCANCÍAS	EN TRÁNSITO	12.110.017	13.063.297	12.110.017	13.063.297	953.280	7,87
		EN CONTENEDORES EN TRÁNSITO	9.770.100	9.793.634	9.770.100	9.793.634	23.534	0,24
	TRÁFICO RO-RO	TOTAL TRÁFICO RO-RO	4.896.154	4.913.214	4.896.154	4.913.214	17.060	0,35
		REMOLQUES SEMIRR. Y PLATAF. (uds.)	123.806	119.711	123.806	119.711	-4.095	-3,31
	CONTENEDORES	TOTAL CONTENEDORES (TEUS)	1.426.519	1.412.743	1.426.519	1.412.743	-13.776	-0,97
		EN TRÁNSITO (TEUS)	788.159	786.579	788.159	786.579	-1.581	-0,20
		ENTRADAS-SALIDAS NACIONAL (TEUS)	156.916	158.636	156.916	158.636	1.720	1,10
		IMPORT-EXPORT EXTERIOR (TEUS)	481.444	467.529	481.444	467.529	-13.915	-2,89
		NACIONAL Y EXTERIOR (TEUS)	638.360	626.165	638.360	626.165	-12.195	-1,91
	PASAJEROS	REG. TRANSPORTE Y DE CRUCERO (nº)	2.010.982	2.083.912	2.010.982	2.083.912	72.930	3,63
		DE CRUCERO (nº)	585.984	565.428	585.984	565.428	-20.556	-3,51
	AUTOMÓVILES	EN RÉGIMEN DE PASAJE (uds)	347.568	375.873	347.568	375.873	28.305	8,14
		EN RÉGIMEN DE MERCANCÍA (**)	224.240	219.635	224.240	219.635	-4.605	-2,05
	BUQUES MERCANTES	TOTAL BUQUES MERCANTES (uds)	12.456	12.319	12.456	12.319	-137	-1,10
		UDS. ARQUEO BRUTO (G.T.)	185.778.970	183.483.598	185.778.970	183.483.598	-2.295.372	-1,24
CRUCEROS (uds)		218	214	218	214	-4	-1,83	

Figura 6: Ejemplo de Resumen Mensual de Puertos del Estado

En las siguientes pestañas de la Excel, se desglosa la información de cada concepto mostrado en el resumen.

2.3.2. Datos de puertos Internacionales.

2.3.2.1. Datos puertos europeos

A través del portal estadístico europeo **Eurostat** (EuroStat, 2020) se pueden obtener datos agregados por países y también datos de los puertos europeos más importantes. En este caso solo se han encontrado datos anualizados.

- Volumen de contenedores por país y total de Europa (2009 al 2018) (Eurostat Volumen Contenedores por país, s.f.)

↑	TIME	2009 ↓	2010 ↓	2011 ↓	2012 ↓	2013 ↓
REP_MAR	↓					
European Union - 27 countries (from 2020)		62 410	69 455	74 962	77 236	78 314
European Union - 28 countries (2013-2020)		69 790	77 679	83 107	85 244	86 560
European Union - 27 countries (2007-2013)		69 638	77 534	82 952	85 100	86 430
European Union - 25 countries (2004-2006)		68 917	76 846	82 147	84 250	85 590
European Union - 15 countries (1995-2004)		66 907	74 180	78 896	80 662	81 589
Belgium		8 490	9 601	9 511	9 165	9 188
Bulgaria		136	142	152	174	180
Denmark		637	734	782	764	752
Germany		11 919	13 092	15 240	15 290	15 563
Estonia		129	152	198	228	254
Ireland		818	773	744	732	726
Greece		1 025	1 184	2 047	3 219	3 615
Spain		11 677	12 424	13 858	14 059	13 550
France		4 028	4 247	4 257	3 981	4 183
Croatia		152	145	154	144	130
Italy		7 223	8 466	8 480	9 298	9 563

Figura 7: Volumen de contenedores por país y total de Europa (2009 al 2018)

- volumen de contenedores por puerto (2009-2018) (Eurostat Volumen de contenedores por Puerto, s.f.)

↑	TIME	2009 ↓	2010 ↓	2011 ↓	2012 ↓	2013 ↓	2014 ↓
REP_MAR	↓						
Antwerpen		7 014	8 144	8 317	8 174	8 256	8 812
Zeebrugge		1 467	1 437	1 157	930	880	880
Bremerhaven		4 552	4 858	5 911	6 111	5 822	5 731
Hamburg		7 031	7 906	9 035	8 891	9 302	9 775
Peiraia		667	850	1 681	2 815	3 199	3 493
Algeciras		2 947	2 773	3 593	4 113	3 988	4 555
Barcelona		1 846	1 928	2 006	1 745	1 717	2 056
Bilbao		443	532	573	610	606	630
Las Palmas		1 006	1 118	1 284	1 208	1 017	977
Valencia		3 654	4 211	4 332	4 471	4 328	4 407
Le Havre		2 257	2 369	2 222	1 997	2 186	2 433
Marseille		943	1 031	1 095	1 147	1 197	1 190
Genova		1 311	1 020	1 277	1 578	1 546	2 014
Gioia Tauro		2 725	3 897	3 307	3 725	3 652	3 708
La Spezia		840	1 181	1 205	1 181	1 207	1 262
Rotterdam		9 579	11 017	11 340	11 418	11 021	11 634

Figura 8: Volumen de contenedores por puerto - Top 20 (2009-2018)

2.3.2.1. Datos puertos internacionales

A nivel internacional no existe una fuente clara que proporcione en abierto los datos históricos de los puertos. Aunque sí se pueden consultar la lista de los 100 puertos más importantes producida todos los años por la empresa Lloyd (One Hundred Ports, 2020) .

Aunque el informe es muy completo, no se facilitan los datos en un formato para que se puedan importar fácilmente.

Otras referencias para obtener información histórica de puertos serían:

- **Wikipedia.** Lista de los 50 puertos más importantes (List of Busiest container ports, 2020)

Container Traffic (in thousand TEUs):											
#	Port	Country	Region	Location	2018 ^{[1][2]}	2017 ^[3]	2016 ^[4]	2015 ^[5]	2014 ^[6]	2013 ^[6]	2012 ^[7]
1	Shanghai	China	East Asia	Yangtze Delta	42,010	40,233	37,133	36,537 ^[16]	35,268	33,617	32,529
2	Singapore	Singapore	Southeast Asia	Singapore Strait	36,660	33,666	30,904 ^[17]	30,922 ^[17]	33,869	32,240	31,649
3	Shenzhen	China	East Asia	Pearl River Delta	27,740	25,208	23,979 ^[18]	24,204 ^[18]	23,798	23,280	22,940
4	Ningbo-Zhoushan	China	East Asia	Yangtze Delta	26,350	24,607	21,560 ^[19]	20,620 ^[19]	19,450	17,351	16,670
5	Guangzhou	China	East Asia	Pearl River Delta	21,870	20,370	18,858 ^[20]	17,625	16,160	15,309	14,744
6	Busan	South Korea	East Asia	Korean Strait	21,660	20,493	19,850 ^[21]	19,469 ^[21]	18,423	17,690	17,046

Figura 9: Lista de los 50 puertos más importantes - Wikipedia

- World Bank Open Data.** Proporcionan acceso libre y gratuito a diversos datos sobre el desarrollo mundial. Esta publicación contiene el tráfico portuario de contenedores (TEU: unidades equivalentes a 20 pies) de todos los países del mundo. Los datos están disponibles desde el año 2000. El tráfico portuario de contenedores mide el flujo de contenedores de los modos de transporte terrestre a los marítimos y viceversa, en unidades equivalentes a veinte pies (TEU), un contenedor de tamaño estándar.
 - Volumen de contenedores anuales a nivel mundial y por países 2000- 2018 (Container port Traffic (TEU), 2020).
- United nations conference on trade and development.** Volumen de mercancías movidos por puertos a nivel global (económico)
 - Cantidad de mercancías transportada desde 1970 hasta 2018, a nivel global y por regiones (en este caso la serie temporal tiene menos histórico) y también por tipo de mercancía (crudo, dry cargo, No por TEUS) (Movimiento de mercancías por tipo y region, 2020)

YEAR		2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
ECONOMY	CARGO TYPE													
World	Crude oil loaded	1 783	1 814	1 785	1 710	1 785	1 751	1 785	1 738	1 707	1 761	1 832	1 875	1 886
	Other tanker trade loaded ⁽¹⁾	915	934	957	931	968	1 034	1 055	1 091	1 118	1 171	1 226	1 272	1 308
	Dry cargo loaded	5 004	5 289	5 489	5 216	5 656	5 990	6 355	6 685	7 018	7 091	7 237	7 570	7 811
	Total goods loaded	7 702	8 036	8 231	7 857	8 408	8 775	9 195	9 513	9 842	10 023	10 295	10 716	11 005
	Crude oil unloaded	1 932	1 996	1 943	1 875	1 939	1 897	1 930	1 882	1 850	1 910	1 986	2 034	2 048
	Other tanker trade unloaded ⁽¹⁾	891	902	933	923	971	1 039	1 056	1 096	1 128	1 188	1 235	1 289	1 322
	Dry cargo unloaded	5 055	5 242	5 410	5 035	5 468	5 860	6 204	6 523	6 859	6 919	7 056	7 379	7 632
	Total goods unloaded	7 878	8 140	8 286	7 832	8 378	8 797	9 190	9 501	9 837	10 018	10 277	10 702	11 002
Developing economies	Crude oil loaded	1 527	1 554	1 518	1 453	1 509	1 501	1 525	1 478	1 433	1 467	1 504	1 515	1 525
	Other tanker trade loaded ⁽¹⁾	537	531	515	503	503	540	555	588	618	661	730	739	757
	Dry cargo loaded	2 767	2 935	3 052	2 843	3 049	3 247	3 449	3 707	3 856	3 847	3 924	4 059	4 188
	Total goods loaded	4 832	5 020	5 085	4 799	5 061	5 288	5 529	5 774	5 907	5 974	6 159	6 313	6 470
	Crude oil unloaded	644	743	685	746	777	807	834	874	885	915	998	1 054	1 102
	Other tanker trade unloaded ⁽¹⁾	353	374	405	388	421	454	495	529	605	653	735	790	821
	Dry cargo unloaded	2 646	2 956	3 099	3 231	3 466	3 748	4 013	4 342	4 589	4 658	4 686	4 982	5 170
	Total goods unloaded	3 643	4 073	4 189	4 366	4 665	5 009	5 342	5 745	6 080	6 226	6 419	6 826	7 093
Developing economies: Africa	Crude oil loaded	354	363	379	354	344	338	364	328	299	294	271	291	289
	Other tanker trade loaded ⁽¹⁾	86	82	83	83	82	69	70	82	74	59	59	70	74
	Dry cargo loaded	282	288	304	271	308	317	324	405	384	403	363	379	404
	Total goods loaded	722	732	767	708	733	724	758	815	758	755	693	741	767
	Crude oil unloaded	41	46	45	45	42	38	33	37	37	39	39	41	42
	Other tanker trade unloaded ⁽¹⁾	39	45	44	40	39	46	51	65	71	72	81	94	94
	Dry cargo unloaded	269	290	288	303	318	294	310	331	362	374	374	363	380
	Total goods unloaded	350	380	377	387	400	379	394	433	470	486	493	497	516
Developing economies: America	Crude oil loaded	251	252	235	226	231	254	253	240	233	223	232	225	219
	Other tanker trade loaded ⁽¹⁾	94	91	93	74	73	83	86	70	76	84	76	72	78
	Dry cargo loaded	685	724	781	730	825	902	943	954	983	1 020	1 028	1 075	1 106
	Total goods loaded	1 031	1 067	1 108	1 030	1 130	1 239	1 283	1 264	1 292	1 328	1 337	1 372	1 404
	Crude oil unloaded	50	76	74	64	69	71	75	69	65	66	52	48	52
	Other tanker trade unloaded ⁽¹⁾	60	64	70	74	76	74	84	89	100	102	128	141	149
	Dry cargo unloaded	264	276	293	234	262	363	389	411	407	422	386	428	452
	Total goods unloaded	373	416	437	372	407	508	547	569	572	590	566	617	653
Developing economies: Asia	Crude oil loaded	921	938	903	872	933	908	906	909	899	948	999	997	1 014

Figura 10: Cantidad de Mercancías Transportadas por Regiones

3. Alcance del proyecto

Tras realizar el estudio de los datos disponibles, interesantes para la realización del proyecto, y las diferentes analíticas que se están usando para resolver el problema se ha refinado el alcance del proyecto.

El trabajo se centrará en el cálculo de las predicciones de los puertos españoles solo para predecir la cantidad de contenedores “TEUS” movidos por los tres puertos españoles más importantes Valencia, Barcelona y Algeciras. En caso de que se obtengan resultados satisfactorios, se ampliarán las predicciones a otras mercancías y a otros puertos.

Puertos del estado facilita datos históricos de los puertos españoles. La serie de datos histórica sólo contempla datos anualizados, en cambio las series más recientes contienen los datos mensualmente. Se realizarán los análisis tanto para la serie histórica anualizada como para la serie reciente mensual y se compararán resultados. Seguramente la estacionalidad no afecte igual a todos los puertos y puede que dependiendo de las características de cada puerto interese más los datos anuales o mensuales.

A la hora de definir la información a utilizar, parece apropiado usar variables macroeconómicas y geográficas, ya que queda demostrado que estas influyen en el volumen de negocio. Algunas de las variables macroeconómicas a probar son Producto Interior Bruto “PIB”, importaciones y exportaciones. Estas variables se referirán a los países y áreas geográficas más influyentes para el sector portuario nacional, que en este caso son “EEUU”, “China” y “Europa”.

Existen dos problemas con los datos de PIB y de comercio de mercancías para algunas regiones o mercancías. El primero es, que no es fácil encontrar series temporales mensuales, y el segundo es, que las series temporales no siempre son extensas e incluyen datos antiguos.

En cuanto a la selección de las técnicas a emplear para predecir el volumen de contenedores en puertos, se evaluarán modelos ARIMA, ARIMAX, SARIMA, Redes neuronales recurrentes y Facebook Prophet.

4. Carga de Datos e Integración

El objetivo de la carga de datos es tener en un único lugar, todos datos para poder realizar los posteriores análisis predictivos. Para ello se ha creado una base de datos dónde se va a almacenar toda la información considerada en puntos anteriores.

A medida que se han ido estudiando los orígenes de datos existentes, se han ido incorporando tablas y atributos a la base de datos.

Aunque para el análisis de los datos se ha usado solamente información de los TEUS, en el proceso de carga se ha cargado toda la información disponible en Puertos del Estado para de este modo poder usar la información para poder comparar toda esta información a través de los cuadros de mando que se han realizado.

Para el diseño de la base de datos se ha usado la herramienta MySQL Workbench (MySQL Workbench, s.f.), que nos permite diseñar de manera visual la base de datos. Una vez finalizado el proceso de diseño, se utiliza la función de exportar que nos genera el script oportuno para crear la base de datos. Para el presente proyecto se ha usado como base de datos MySQL (MySQL, s.f.). El proyecto MySQL Workbench y los diferentes scripts de SQL se pueden consultar en el repositorio del proyecto (Anexo 10.1) subdirectorio “Modelo_bd”.

Se ha optado por esta solución por ser una solución conocida que no requiere aprendizaje, aunque para integrarlo finalmente en el producto INTEGRA2 habrá que exportar los datos de MySQL e importarlos a Oracle, que es la base de datos de dicho producto.

La base de datos está compuesta por 4 tablas:

Nombre Tabla	Descripción
spanish_historic_annual	Almacena la información general histórica anualizada.
spanish_historic_contenedores	Almacena la información mensual de los puertos españoles del 2006 a 2020
spanish_month_contenedores	Almacena la información mensual de los puertos españoles del 2006 a 2020
international_data	Almacena la información internacional necesaria para realizar los diferentes análisis. Los tipos de datos que almacena son los TEUS de las regiones y puertos internacionales, volumen de mercancías transportadas y PIB de los países más relevantes.

Tabla 3: Descripción Tablas de BD creadas para la importación

En la siguiente figura se puede ver como quedarían las 4 tablas con sus correspondientes atributos

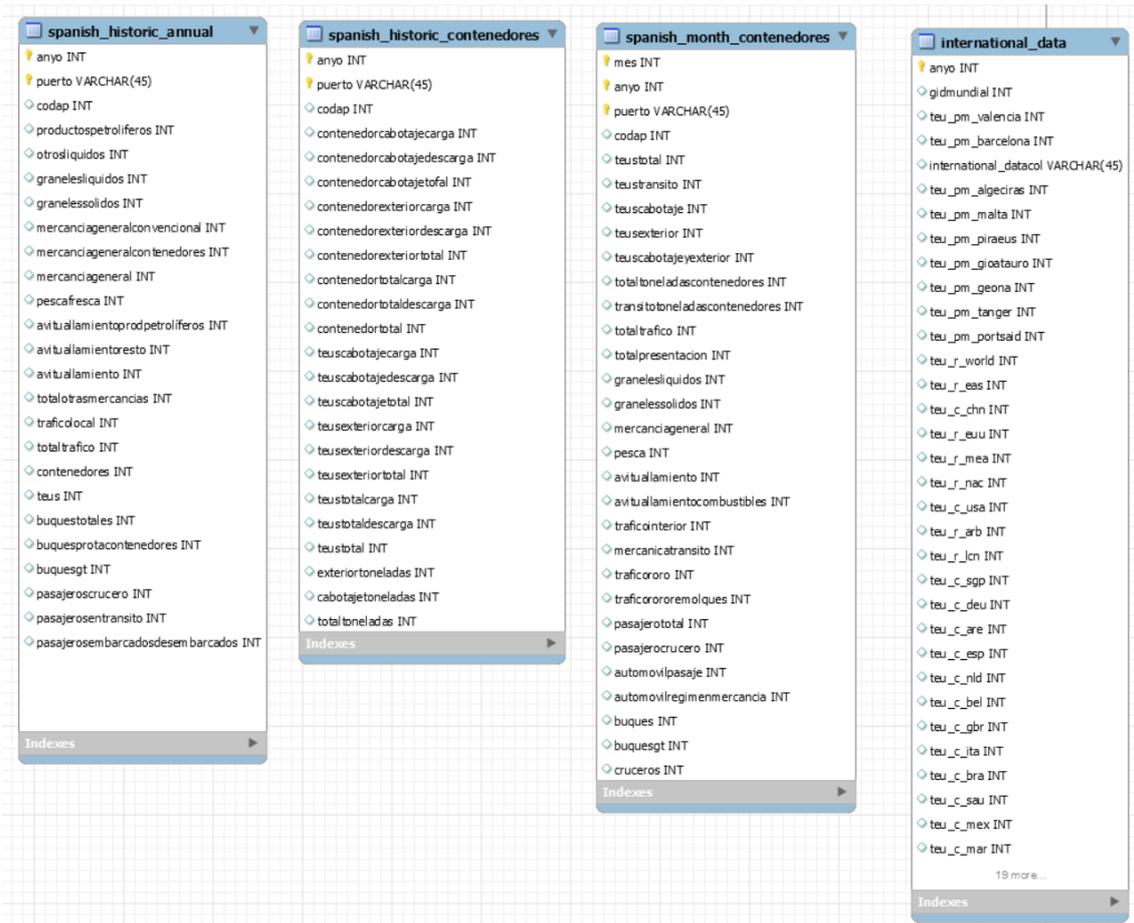


Figura 11: Atributos de las Diferentes Tablas de la BD

El proceso de carga se ha dividido en 3 subprocesos: carga anual de datos históricos españoles, carga mensual de datos mensuales españoles y carga de datos internacionales.

4.1. Carga anual de datos portuarios españoles históricos

Estos datos representan, la serie de datos anuales de 1962 a 2009 (ref. 2.3.1. Datos de puertos españoles), que se obtienen de la web puertos del estado y accesibles a través del repositorio del proyecto (Anexo 10.1), subcarpeta datos_historicos_españa.

Tras revisar los datos disponibles se han exportado los siguientes datos:

- Resumen general - 1373 registros - (resumengeneral.xls)

Columna	Descripción	unidades
Año	año al que pertenecen los datos	número entero (año)
Autoridad Portuaria	nombre de la autoridad portuaria	texto

Graneles líquidos	cantidad de graneles líquidos transportados por puerto en dicho año	número entero (t)
Graneles Sólidos	cantidad de graneles sólidos por puerto en dicho año	número entero (t)
Mercancía General	cantidad de mercancía transportada por puerto en dicho año	número entero (t)
Pesca Fresca	cantidad de pescado fresco por puerto en dicho año	número entero (t)
Avituallamiento	cantidad de avituallamiento por puerto en dicho año	número entero (t)
Tráfico local	total tráfico local por puerto en dicho año	número entero (t)
Total Trafico	Total mercancía transportada por puerto en dicho año	número entero (t)
TEUS	Nº de contenedores por puerto en dicho año	número entero (TEUS)
Buques	Nº de buques por puerto en dicho año	número entero (Buques)

Tabla 4: Atributos del Fichero resumengeneral.xls

- Resumen mercancías según presentación - 1373 registros - (resumenmercanciassegunpresentacion.xls)

Columna	Descripción	unidades
Año	año al que pertenecen los datos	número entero (año)
Autoridad Portuaria	nombre de la autoridad portuaria	Texto
Productos Petrolíferos	Cantidad de productos petrolíferos	número entero (t)
Otros Líquidos	Cantidad de productos líquidos no petrolíferos	número entero (t)
TOTAL G. LIQUIDOS	Total graneles líquidos	número entero (t)
Graneles Sólidos	total graneles sólidos	número entero (t)
M. General en Contenedores	Mercancía general movida en contenedores	número entero (t)
M. General Convencional	Mercancía general convencional	número entero (t)
TOTAL M. GENERAL	Total mercancía general	número entero (t)

TOTAL MERCANCIAS SEGUN SU PRESENTACION	Total mercancías	número entero (t)
---	------------------	-------------------

Tabla 5: Atributos del Fichero resumenmercanciassegunpresentacion.xls

- Resumen - otras mercancías - 1373 registros - (resumenotrasmercancias.xls)

Columna	Descripción	unidades
Año	año al que pertenecen los datos	número entero (año)
Autoridad Portuaria	nombre de la autoridad portuaria	texto
Pesca Fresca	cantidad de pescado fresco por puerto en dicho año	número entero (t)
Avituallamiento Prod. Petrolíferos	cantidad de avituallamiento productos petrolíferos por puerto en dicho año	número entero (t)
Avituallamiento Resto	cantidad de avituallamiento resto por puerto en dicho año	número entero (t)
TOTAL AVITUALLAMIENTO	cantidad de avituallamiento total por puerto en dicho año	número entero (t)
Tráfico local	total tráfico local por puerto en dicho año	número entero (t)
TOTAL OTRAS MERCANCIAS	total tráfico otras mercancías por puerto en dicho año	número entero (t)
TOTAL TRAFICO PORTUARIO	Total tráfico portuario por puerto en dicho año	número entero (t)

Tabla 6: Atributos del Fichero resumenotrasmercancias.xls

- Resumen - otros datos - 1373 registros - resumenotrosdatos.xls

Columna	Descripción	unidades
Año	año al que pertenecen los datos	número entero (año)
Autoridad Portuaria	nombre de la autoridad portuaria	Texto
Contenedores Número	Nº de contenedores por puerto en dicho año	número entero (Contenedores)

Contenedores TEUS	Nº de TEUS por puerto en dicho año	número entero (TEUS)
Buques	Nº de buques por puerto en dicho año	número entero (Buques)
Buques GT	Nº de toneladas por puerto en dicho año	número entero (t)
Pasajeros de Crucero	Nº de pasajeros de Crucero por puerto en dicho año	número entero (Pasajeros)
Pasajeros en Tránsito	Nº de pasajeros en tránsito por puerto en dicho año	número entero (Pasajeros)
Pasajeros Embarcados Y Desembarcados	Nº de pasajeros total por puerto en dicho año	número entero (Pasajeros)

Tabla 7: Atributos del Fichero resumenotrosdatos.xls

- Contenedores por unidades - contenedorporunidades.xls

Columna	Descripción	Unidades
Año	año al que pertenecen los datos	número entero (año)
Autoridad Portuaria	nombre de la autoridad portuaria	Texto
Cabotaje Carga	Contenedores tráfico interior cargados	número entero (Contenedores)
Cabotaje Descarga	Contenedores tráfico interior descargados	número entero (Contenedores)
Cabotaje TOTAL	Contenedores tráfico interior totales	número entero (Contenedores)
Exterior Carga	Contenedores tráfico exterior cargados	número entero (Contenedores)
Exterior Descarga	Contenedores tráfico exterior descargados	número entero (Contenedores)
Exterior TOTAL	Contenedores tráfico exterior total	número entero (Contenedores)
TOTAL Carga	Contenedores cargados total	número entero (Contenedores)
TOTAL Descarga	Contenedores descargados total	número entero (Contenedores)
TOTAL	Contenedores totales	número entero (Contenedores)

Tabla 8: Atributos del Fichero contenedorporunidades.xls

- Contenedores por TEUS - contenedoresporteus.xls

Los campos son exactamente igual que en la tabla anterior pero en vez de usar como medidas el contenedor se usa TEUS

- Contenedores según navegación - contenedoressegunnavegacion.xls

Columna	Descripción	unidades
Exterior Toneladas	Total toneladas movidas en contenedores a nivel internacional	número entero (t)
Cabotaje Toneladas	Total toneladas movidas en contenedores a nivel nacional	número entero (t)
TOTAL Toneladas	Total toneladas movidas en contenedores	número entero (t)

Tabla 9: Atributos del Fichero contenedoressegunnavegacion.xls

Una vez descargadas las diferentes Excel de Puertos del estado se utilizó la herramienta Pentaho Data Integration (PDI) v7.1 “Spoon” para realizar el proceso ETL e importar los datos a la base de datos (Data Integration, s.f.).

Para cargar la información genérica anualizada almacenada en los ficheros resumengeneral, resumenmercanciassegunpresentacion, resumenotrasmercancias y resumenotrosdatos, se desarrolló un trabajo en Spoon para poblar la tabla “spanish_historic_annual”

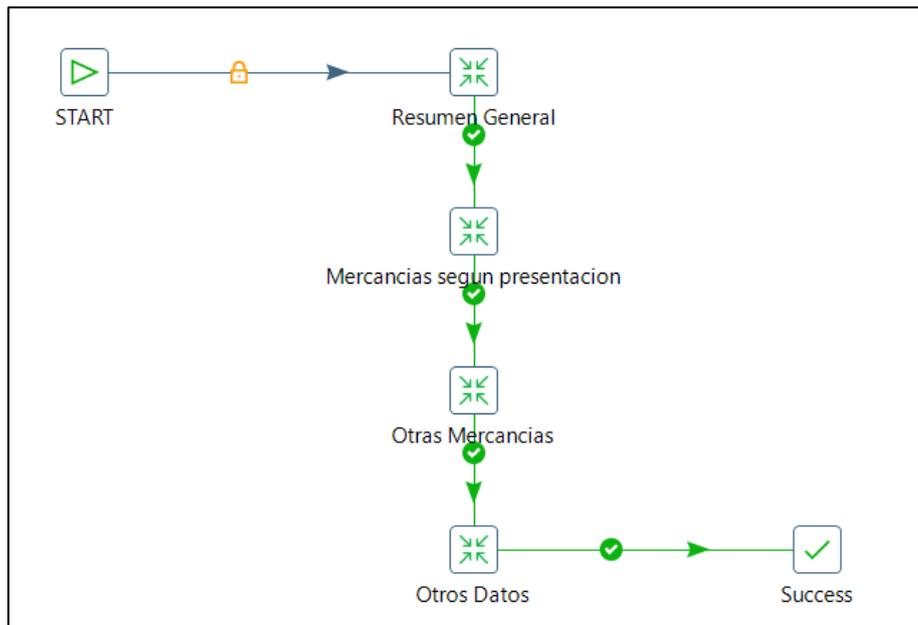


Figura 12: Trabajo para la carga de los datos de la tabla spanish_historic_annual

Para cargar la información específica de contenedores almacenada en (Contenedores por unidades, Contenedores por TEUS y Contenedores según navegación), se desarrolló un trabajo en Spoon para poblar la tabla “spanish_historic_contenedores”.

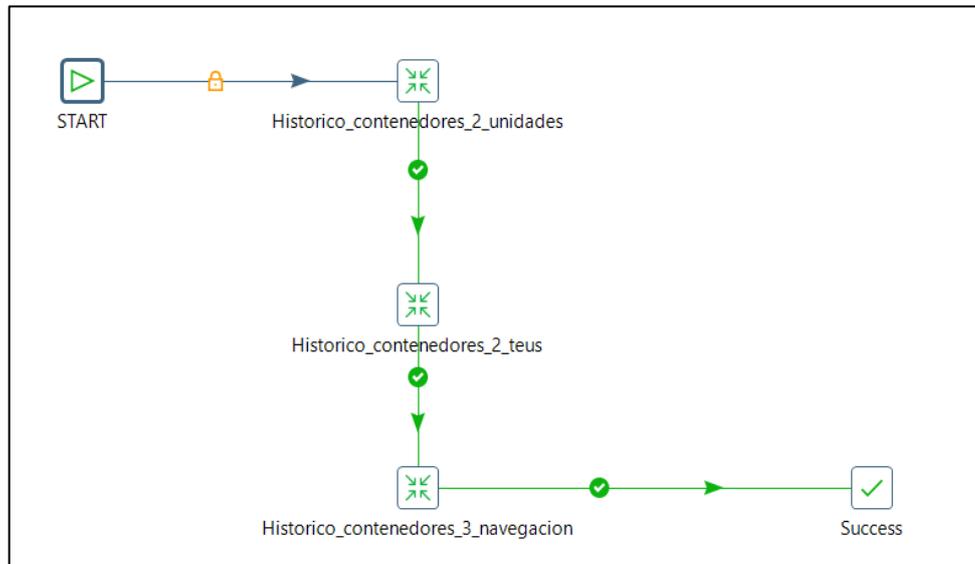


Figura 13: Trabajo para la carga de los datos de la tabla spanish_historic_contenedores

En la carpeta “spoon_project” del repositorio del proyecto se pueden consultar todos los trabajos y transformaciones realizados.

Conclusiones sobre la carga de datos anuales

En la serie de datos históricos de los puertos españoles, la exportación a Excel genera un fichero Excel que no es válido y se tuvo que abrir y guardar en formato Excel 1997-2003.

La herramienta Pentaho data Integration, por defecto no puede crear conexiones con MySQL y se tuvo que descargar una librería adicional con el driver de MySQL y copiar en la carpeta lib de PDI.

4.2. Carga mensual de datos portuarios españoles históricos

Inicialmente se intentó realizar la carga de las series de datos mensual por medio de PDI al igual que se hizo la carga de datos anuales. Pero debido a la estructura de los ficheros Excel a cargar con celdas combinadas, no fue posible llevar a cabo la carga mediante PDI.

Tras descartar esta opción, se probó a realizar la carga mediante un programa en Excel y tras verificar que la librería de java POI API (POI API Apache, s.f.) podía leer los Excel que contenían los datos mensuales de los puertos, se desarrolló un programa java para tal efecto.

A continuación se muestra, a modo de ejemplo, el contenido de la pestaña TEUS de uno de los ficheros mensuales

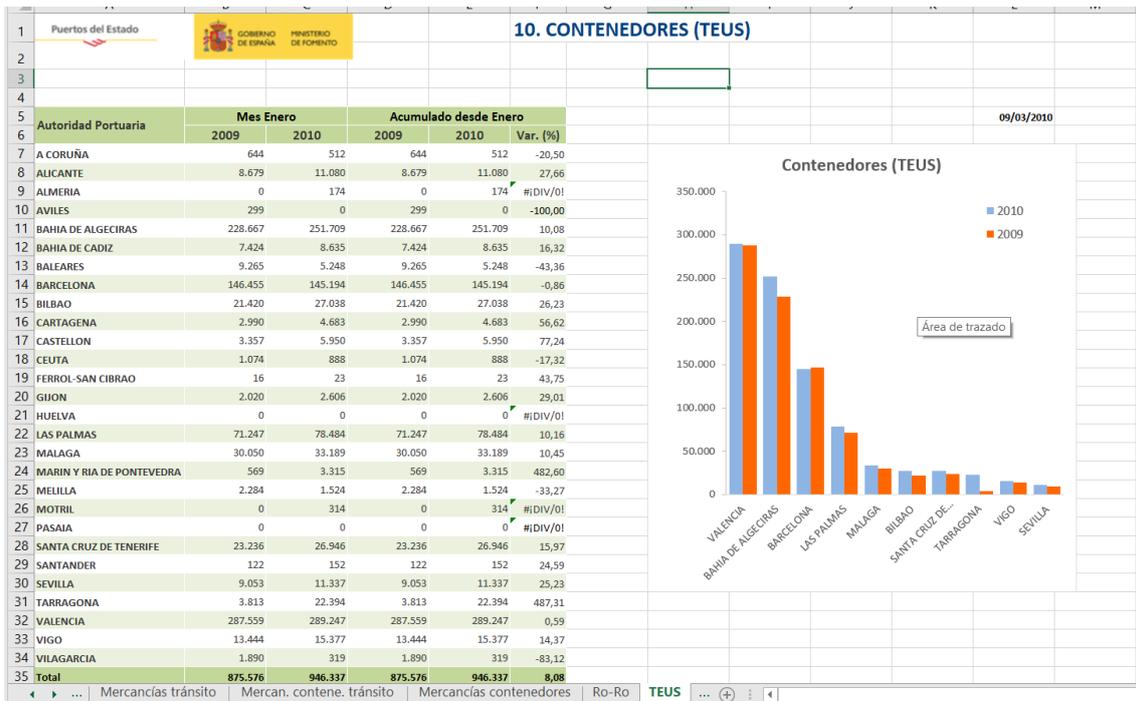


Figura 14: Ejemplo de Volumen de TEUS Mensual de Puertos del Estado

En el repositorio del proyecto (ver Anexo 10.1) se encuentra tanto el programa Java (TFM_java_workspace) como la documentación del mismo (Carpeta TFM_javadoc).

A modo resumen, el programa realiza las siguientes acciones:

1) Carga datos mensuales

Lee todos los ficheros mensuales (177 ficheros) y rellena la tabla "spanish_month_contenedores" que almacena los datos mensuales. Para ello lee todos los tabs o pestañas del fichero Excel y recupera el dato mensual de todos los puertos. Una vez recuperada toda la información de un mes, realiza las inserciones correspondientes en la base de datos y pasa al siguiente mes.

```
public void ETL_spanish_month_contenedores() throws Exception{
    createConnection();
    for (int i =0; i < fileNames.length;i++) {
        String fileName =fileNames[i];
        HSSFWorkbook wb = new HSSFWorkbook(new FileInputStream(fullFilePath));
        GetProperties properties = new GetProperties();
        ExtractIni(wb, listContenedores, anyo, mes_numero, false);
        ExtractTab(wb, properties.getPropValues("tab_totaltrafico"),
            "totaltrafico",listContenedores, false);
        ...
        ExtractTab(wb, properties.getPropValues("tab_cruceros"), "cruceros",
            listContenedores, false);
    }
}
```

```

        SQLInsertMonltyData(listContenedores);
    }
    closeConnection();
}

```

```

private void SQLInsertMonltyData(Spanish_month_contenedores_row[] rows) throws
SQLException{
    for(int i=0; i <rows.length; i++) {
        Spanish_month_contenedores_row row = rows[i];
        PreparedStatement insertMontltydata = con.prepareStatement("INSERT INTO
spanish_month_contenedores (mes,anyo,puerto,teustotal,teustransito,teuscabotaje,teusext
erior,teuscabotajeyexterior,totaltoneladascontenedores,transitotoneladascontenedores,t
otaltrafico,totalpresentacion,granelesliquidos,granelessolidos,mercanciageneral,pesca,
avituallamiento,avituallamientocombustibles,traficointerior,mercanicatransito,traficor
oro,traficorororemolques,pasajerototal,pasajero crucero,automovilpasaje,automovilregime
nmercancia,buques,buquesgt,cruceiros) VALUES
(?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?)");

        insertMontltydata.setInt(1, row.mes);
        insertMontltydata.setInt(2, row.anyo);
        insertMontltydata.setString(3, row.puerto);
        insertMontltydata.setInt(4, row.teustotal);
        insertMontltydata.setInt(5, row.teustransito);
        insertMontltydata.setInt(6, row.teuscabotaje);
        insertMontltydata.setInt(7, row.teusexterior);
        insertMontltydata.setInt(8, row.teuscabotajeyexterior);
        insertMontltydata.setInt(9, row.totaltoneladascontenedores);
        insertMontltydata.setInt(10, row.transitotoneladascontenedores);
        insertMontltydata.setInt(11, row.totaltrafico);
        insertMontltydata.setInt(12, row.totalpresentacion);
        insertMontltydata.setInt(13, row.granelesliquidos);
        insertMontltydata.setInt(14, row.granelessolidos);
        insertMontltydata.setInt(15, row.mercanciageneral);
        insertMontltydata.setInt(16, row.pesca);
        insertMontltydata.setInt(17, row.avituallamiento);
        insertMontltydata.setInt(18, row.avituallamientocombustibles);
        insertMontltydata.setInt(19, row.traficointerior);
        insertMontltydata.setInt(20, row.mercanicatransito);
    }
}

```

```

insertMontltydata.setInt(21, row.traficororo);
insertMontltydata.setInt(22, row.traficorororemolques);
insertMontltydata.setInt(23, row.pasajerototal);
insertMontltydata.setInt(24, row.pasajerocrucero);
insertMontltydata.setInt(25, row.automovilpasaje);
insertMontltydata.setInt(26, row.automovilregimenmercancia);
insertMontltydata.setInt(27, row.buques);
insertMontltydata.setInt(28, row.buquesgt);
insertMontltydata.setInt(29, row.cruceros);
insertMontltydata.executeUpdate();
    if (insertMontltydata != null) {
    try {
        insertMontltydata.close();
    } catch (SQLException e) { /* ignored */
    }
}

```

2) Carga datos generales anuales

Lee todos los ficheros de los meses de diciembre para recuperar el total anual y así poder completar la serie anual desde 2006 a 2019. No todos los campos de la tabla anual se pueden completar, ya que la información mensual contiene muchos menos campos que la información mensual.

```

public void ETL_spanish_historic_annual() throws Exception{
    createConnection();
    for (int i =0; i < fileNames.length;i++)
    {
        String fileName =fileNames[i];
        if (!mes_numero.equals("12")) continue;
        HSSFWorkbook wb = new HSSFWorkbook(new FileInputStream(fullFilePath));
        GetProperties properties = new GetProperties();
        ExtractIni(wb, listContenedores, anyo, mes_numero, true);
        ExtractTab(wb, properties.getPropValues("tab_totaltrafico"), "totaltrafico",
        listContenedores, true);
        ...
        ExtractTab(wb, properties.getPropValues("tab_cruceros"), "cruceros",
        listContenedores, true);
        SQLInsertHistoricAnual(listContenedores);
    }
}

```

```
closeConnection() ;  
}
```

3) Completar los campos código autoridad portuaria

En los ficheros no se incluye un código de autoridad portuaria, pero la aplicación INTEGRA2 en la cual se van a insertar los datos si se dispone de dicho dato. Por tanto, se recorren todas las tablas y se rellena el campo código de autoridad portuaria con el código específico

```
private void SQLUpdateTableWithCodAP(String tableName, CodAutoridadPortuaria[]  
ArrayCodAP ) throws SQLException, IOException{  
  
    for(int i=0; i <ArrayCodAP.length; i++)  
  
    {  
  
        if (con == null) SQLCreateConnection();  
  
        PreparedStatement updateCodAP = con.prepareStatement("UPDATE "+ tableName + " SET  
codap="+  
  
        ArrayCodAP[i].codap + " WHERE puerto = '"+ArrayCodAP[i].puerto+"'");  
  
        if (updateCodAP != null) {  
  
            try {  
  
                updateCodAP.close();  
  
            }catch (SQLException e) {    }  
  
        }  
  
    }  
  
}
```

Tras la ejecución del programa, que tarda unos pocos minutos, se rellena la tabla con datos mensuales y se completa la tabla de datos anuales con los datos de 2006 a 2020. A continuación, se muestra una captura de las tablas de la base de datos con un resumen del número de registros almacenados

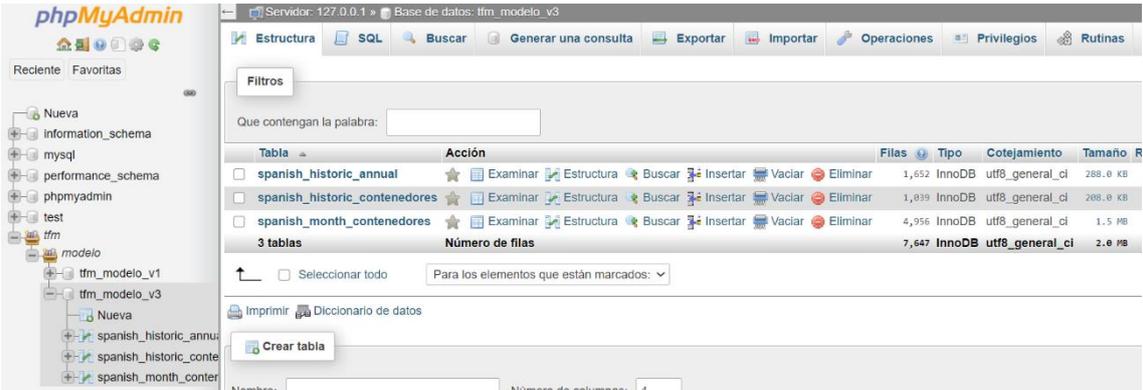


Figura 15: Resumen Registros Almacenados en BD

Ejemplo de datos almacenados en la tabla Spanish_month_contenedores

Mostrando filas 0 - 24 (total de 4956, La consulta tardó 0,0004 segundos.) [año: 2020... - 2020...]

SELECT * FROM `spanish_month_contenedores` ORDER BY `spanish_month_contenedores`.`año` DESC

Perfilando [Editar en línea]

Número de filas: 25 Filtar filas: Buscar en esta tabla Sort by key: Ninguna

+ Opciones

	mes	año	puerto	codap	teustotal	teustransito	teuscabotaje	teusexterior	teuscabotajeyexterior	totaltoneladascontenedores
<input type="checkbox"/>	9	2020	VILAGARCIA	24	3020	4	2849	167	3016	25389
<input type="checkbox"/>	9	2020	VIGO	26	17180	1088	1021	15071	16092	259274
<input type="checkbox"/>	9	2020	VALENCIA	63	477194	274235	19258	183701	202959	5482484
<input type="checkbox"/>	9	2020	TARRAGONA	71	3369	2	745	2622	3367	33109
<input type="checkbox"/>	9	2020	SEVILLA	33	9984	0	9895	89	9984	72151
<input type="checkbox"/>	9	2020	SANTANDER	13	2209	0	0	2209	2209	26669
<input type="checkbox"/>	9	2020	SANTA CRUZ DE TENERIFE	94	31144	1626	27024	2494	29518	264055
<input type="checkbox"/>	9	2020	PASAIA	11	0	0	0	0	0	0
<input type="checkbox"/>	9	2020	MOTRIL	52	5	0	0	5	5	32
<input type="checkbox"/>	9	2020	MELILLA	56	908	0	837	71	908	6450
<input type="checkbox"/>	9	2020	MARIN Y RIA DE PONTEVEDRA	25	4322	247	227	3847	4075	57297
<input type="checkbox"/>	9	2020	MALAGA	50	6381	4188	1010	1183	2193	32561
<input type="checkbox"/>	9	2020	LAS PALMAS	90	92696	51840	34993	5863	40856	1021353
<input type="checkbox"/>	9	2020	HUELVA	30	7426	0	5481	1945	7426	58423
<input type="checkbox"/>	9	2020	GIJON	14	4870	2	1044	3824	4868	60748

Figura 16: Registros Almacenados en Tabla Spanish_month_contenedores

Ejemplo de datos almacenados en la tabla Spanish_historic_annual

Mostrando filas 0 - 24 (total de 1652, La consulta tardó 0,0004 segundos.) [año: 2019... - 2019...]

SELECT * FROM `spanish_historic_annual` ORDER BY `spanish_historic_annual`.`año` DESC

Perfilando [Editar en línea]

Número de filas: 25 Filtar filas: Buscar en esta tabla Sort by key: Ninguna

+ Opciones

	año	puerto	codap	productospetroliferos resumen mercancías según presentación	otrosliquidos resumen mercancías según presentación	granelesliquidos en la hora resumen mercancías por presentación se l...	granelessolidos	mercanciageneralconvencional
<input type="checkbox"/>	2019	VILAGARCIA	24	939	NULL	182544	479972	311925
<input type="checkbox"/>	2019	VIGO	26	70229	NULL	46770	297778	1160394
<input type="checkbox"/>	2019	VALENCIA	63	252059	NULL	3120013	2190118	14585870
<input type="checkbox"/>	2019	TARRAGONA	71	15831	NULL	21210546	9718750	1324310
<input type="checkbox"/>	2019	SEVILLA	33	11800	NULL	293337	2294024	780335
<input type="checkbox"/>	2019	SANTANDER	13	9931	NULL	323613	3646528	2408150
<input type="checkbox"/>	2019	SANTA CRUZ DE TENERIFE	94	521561	NULL	4811479	421960	3895454
<input type="checkbox"/>	2019	PASAIA	11	8868	NULL	0	1031321	2166602
<input type="checkbox"/>	2019	MOTRIL	52	19506	NULL	1324304	618405	789304
<input type="checkbox"/>	2019	MELILLA	56	0	NULL	75935	7100	637489
<input type="checkbox"/>	2019	MARIN Y RIA DE PONTEVEDRA	25	17927	NULL	7	888504	676064
<input type="checkbox"/>	2019	MALAGA	50	38661	NULL	83774	1464546	505852
<input type="checkbox"/>	2019	LAS PALMAS	90	2275822	NULL	8070275	453302	4661610

Figura 17: Registros Almacenados en Tabla Spanish_historic_annual

4.3. Carga datos internacionales

La carga de datos internacionales permitirá en los análisis posteriores realizar mejores predicciones en base a dichos datos internacionales que ayudan a explicar las series actuales.

Los datos internacionales cargados han sido:

- TEUS movidos por los puertos del Mediterráneo más importantes.
- TEUS movidos en las regiones y países más importantes para los puertos analizados.
- Mercancías movidas en las regiones y países más importantes para los puertos analizados.

La carga de datos internacionales ha consistido en la elaboración de una Excel de manera manual, dado que no existe una fuente única que provea dichos datos y el inicio temporal de dichos datos es muy variable, no se han podido conseguir series temporales muy largas (2000-2018) ni con periodicidad menor a un año. Sólo para el caso de algunos indicadores económicos como el PIB se pueden conseguir datos trimestrales para algunas regiones.

En el repositorio del proyecto (Anexo 10.1), carpeta datos_internacionales se puede encontrar el fichero *excel_internacional_manual.xlsx* que contiene el conjunto de datos considerado.

- **Puertos del Mediterráneo:** teu_pm_valencia, teu_pm_barcelona, teu_pm_algeciras, teu_pm_malta, teu_pm_piraeus, teu_pm_gioatauro, teu_pm_geona, teu_pm_tanger y teu_pm_portsaid)
- **Regiones y países del mundo:** teu_r_world, teu_r_eas, teu_c_chn, teu_r_euu, teu_r_mea, teu_r_nac, teu_c_usa, teu_r_arb, teu_r_lcn, teu_c_sgp, teu_c_deu, teu_c_are, teu_c_esp, teu_c_nld, teu_c_bel, teu_c_gbr, teu_c_ita, teu_c_bra, teu_c_sau, teu_c_mex, teu_c_mar, teu_c_dza
- **Movimiento de mercancías:** m_TO, m_AG, m_AGFO, m_MI, m_MIFU, m_MA, m_MAIS, m_MACH, m_MACHPH, m_MAMT, m_MAMTOF, m_MAMTOTEP, m_MAMTOTTL, m_MAMTOTIC, m_MAMTTE, m_MAMTAU, m_MATE, m_MACL)

En la siguiente figura se muestran las primeras columnas de la Excel a modo de ejemplo.

	A	N	O	P	Q	R	S	T	U	V	W	X	Y
1	anyo	teu_r_euu	teu_r_mea	teu_r_nac	teu_c_usa	teu_r_arb	teu_r_lcn	teu_c_sgp	teu_c_deu	teu_c_are	teu_c_esp	teu_c_nld	teu_c_bel
22	2000	41051632	16651385	31227942	28300000	14221727	15860053	17100000	7695688	5055801	5789693	6407162	505757
23	2001	42672228	17414481	30197964	27307576	14873478	16129381	15572677	8426520	5081964	6156351	6227321	510969
24	2002	46785653	19516960	32984200	29676890	16669178	15090170	16986010	9252670	5872240	6664100	6797530	582561
25	2003	51632029	22001588	36320556	32689484	18791266	19556512	18441000	10943452	6955202	7363964	7293790	648244
26	2004	59654992	25071303	38827775	34901628	21413069	22294663	21329100	12479419	8661636	8270437	8482190	727963
27	2005	63373218	27874090	42661263	38497839	23806892	24816487	23192200	13598987	9851709	9170557	9471829	788999
28	2006	67710907	30873147	45226905	40896742	26368347	28018282	24792400	15009691	10967048	10033089	10046503	870829
29	2007	81299855	36285930	49253317	44839390	30991333	32720909	28767500	16644222	13182412	13346028	11290260	1025751
30	2008	81986205	41994621	47132433	42411770	35945776	35627767	30891200	17183042	14756127	13461304	11362089	1093713
31	2009	70521473	42545025	41545143	37353575	36475109	32630750	26592800	13296300	14425039	11803192	10066374	970149
32	2010	76777620	48941446	46838790	42058000	41658986	37085353	29147000	14715340	15177436	12604419	11503660	1105574
33	2011	85836387	52078711	47424659	42579150	44299275	42245973	30647000	18744981	16866912	13959826	12129700	1094000
34	2012	87842352	55140383	48964166	43740123	47818575	44869839	32347000	18931700	18120112	14087029	12204336	1067119
35	2013	90410158	55904010	49932584	44683127	48931469	45835852	33388000	19261033	18693112	13947724	11891300	1067420
36	2014	94182776	58394003	53481796	48055621	51236112	46088775	34688000	20128033	20223612	14263337	12559200	1106140
37	2015	92737116	58469105	55636809	49842957	51312470	47034727	31710200	19139033	21233200	14326512	12497016	1123760
38	2016	96558447	59416551	56323768	50485284	51707543	46444438	32668000	19364033	20613200	15372581	12746016	1148990
39	2017	102134169	60501907	58510434	52132844	51962238	49148866	33667000	19718233	19128300	17763587	14134867	1195770
40	2018	105634851	61999449	61352043	54688353	54002417	51659086	36600000	19597633	19054000	17189759	14825967	1268210

Figura 18: Datos Internacionales Cargados

Además de estos datos, se han cargado en un fichero (gdpmundial_anual_español.csv) los datos del PIB Mundial y español desde 1980 a 2019.

A	B	C
anyo	gdp_world	gpd españa
1980	11227,7874	159.100
1981	11624,0372	173.339
1982	11514,7257	190.291
1983	11747,2826	182.781
1984	12180,065	207.776
1985	12793,5269	226.288
1986	15118,7514	243.382
1987	17201,255	263.138
1988	19244,4107	302.674
1989	20087,6893	357.882
1990	22626,7366	401.686
1991	23966,8855	443.715
1992	25453,1892	463.260
1993	25858,2104	425.936
1994	27771,1922	425.089
1995	30887,2814	459.337

Figura 19: PIB Español (x106) y mundial(x109) de 1980 a 2019

Hubiera sido interesante poder disponer de series de datos más largas y con menos periodicidad para poder incluir estos datos en los diferentes análisis. Finalmente, solo los datos del PIB Español y Mundial se incluyeron en las predicciones del método ARIMAX.

4.4. Conclusiones sobre la carga de datos

El proceso de carga ha sido incluso más costoso “en tiempo” que el proceso de análisis de los mismos.

Inicialmente se pensaba usar únicamente Pentaho Data Integration para realizar la carga, ya que es una herramienta ETL muy potente y fácil de utilizar. Esta herramienta se ha utilizado para cargar exitosamente las dos tablas con datos españoles anualizadas.

También se empezó utilizando esta herramienta para la carga de los datos mensuales, pero debido a la estructura de los ficheros Excel que contenían celdas combinadas, no se ha podido realizar la carga de los datos mensuales. Al no poder encontrar una solución en tiempo razonable y para no retrasar la carga de estos datos y poner en peligro las fechas de entrega del proyecto, se optó por realizar una pequeña prueba en Java para validar la viabilidad de cargar estas Excels. Tras la prueba satisfactoria, se completó el programa en java que cargó la serie de datos mensuales.

Se ha cargado en la Base de datos toda la información disponible en la web de Puertos del Estado, para de este modo poder realizar comparación de datos entre los diferentes puertos a través de cuadros de mando que se explicarán más adelante.

Los datos internacionales no han llegado a guardarse en BD, y tan solo se ha realizado el proceso manual de rellenar un par de ficheros Excel con los campos requeridos.

5. Análisis Series Temporales

Tras la carga de datos, se procederá al análisis de los datos de los TEUS de los puertos de Valencia, Barcelona y Algeciras, con el objetivo de ver qué modelo funciona mejor a la hora de calcular el volumen de TEUS futuro.

Se han analizado 5 modelos/algoritmos de series temporales para tratar de determinar si hay un modelo que obtiene mejores resultados para todos los puertos o en cambio, cada puerto se comporta mejor con un modelo u otro. Algunos modelos se han probado tanto para series mensuales como para series anuales, en cambio otros modelos se han usado solo para una de las dos series, ya que están más pensados para un determinado tipo de serie.

En la siguiente tabla, se muestran los análisis que se han realizado para cada una de las series:

Modelo	Serie Anual	Serie Mensual
ARIMA	X	X
SARIMA		X
ARIMAX	X	
RNN	X	X
FB Prophet		X

Tabla 10: Modelos aplicados a las series anuales y mensuales

Para todos los modelos se usará una parte de la serie temporal para entrenar el modelo y la segunda parte para la validación (ver Tabla 11). Se usarán los mínimos cuadrados de los errores para determinar qué modelo se ajusta mejor para los datos de validación.

Serie	Datos Entrenamiento	Datos Validación	Datos Predichos
Serie Anual	1962-2010	2010-2019	2020-2024
Serie Mensual	01/2006-07/2018	08/2018-12/2019	01/2020-12/2022

Tabla 11: Datos de entrenamiento - Validación

Por otro lado, como el año 2020 ha sido un año anómalo debido al COVID-19, se estimará cuál hubiera sido el volumen de los puertos en caso de que no hubiera habido COVID.

Para cada modelo/puerto se han definido una serie de libretas de Jupyter (Jupyter Notebooks, s.f.) . El código completo de libretas se puede consultar en el repositorio del proyecto (Anexo 10.1 directorio Notebook) y los detalles de cada ejecución (desviación/error y gráficas) se pueden consultar en el anexo correspondiente.

El nombre de las diferentes libretas con los diferentes análisis para cada puerto se muestra en la siguiente tabla:

Modelo	Valencia	Barcelona	Algeciras
ARIMA	ARIMA_VA_ValenciaAnual ARIMA_VM_ValenciaMensual	ARIMA_BA_BarcelonaAnual ARIMA_BM_BarcelonaMensual	ARIMA_AA_AlgecirasAnual ARIMA_AM_AlgecirasMensual
SARIMA	SARIMA_VM_ValenciaMensual	SARIMA_BM_BarcelonaMensual	SARIMA_AM_AlgecirasMensual
ARIMAX	ARIMAX_VA_ValenciaAnual	ARIMAX_BA_BarcelonaAnual	ARIMAX_AA_AlgecirasAnual
RNN	RNN_VA_ValenciaAnual RNN_VM_ValenciaMensual	RNN_BA_BarcelonaAnual RNN_BM_BarcelonaMensual	RNN_AA_AlgecirasAnual RNN_AM_AlgecirasMensual
Prophet	Prophet_VM_valencia	Prophet_BM_Barcelona	Prophet_AM_Algeciras

Tabla 12: Correspondencia entre las libretas Jupyter y los análisis realizados

5.1. ARIMA

El modelo ARIMA (ARIMA-AutoRegressive Integrated Moving Average) es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción futura. Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.

5.1.1. Implementación

El modelo necesita identificar los coeficientes y número de regresiones que se utilizarán. Este modelo es muy sensible a la precisión con que se determinan sus coeficientes. A la hora de usar el modelo ARIMA, existen 3 parámetros a configurar. Estos parámetros nos permiten definir el modelo para nuestra serie.

La invocación al modelo es ARIMA (p,d,q):

- **p**: Es la parte autorregresiva del modelo. Determina el efecto (grado de asociación lineal) de valores anteriores en el modelo.
- **d**: Es la parte integrada del modelo. Que nos ayuda a aplanar la serie y quita tendencias. A través del parámetro “d-diferenciación, se adapta la serie para poder usar el modelo, ya que este no funciona con series estacionarias y con tendencia. Es decir que tengan diferentes medias y varianzas.
- **q**: Es la parte de la media móvil del modelo. Establece el error de los modelos en combinación con valores anteriores.

A la hora de determinar los parámetros de una serie, hay 2 Aproximaciones / maneras de hacerlo

1. Manualmente. Analizando la serie, viendo las peculiaridades de la misma y determinando los parámetros de manera manual .
2. Creación de una matriz que prueba las combinaciones más usuales y se queda con la que genera un menor error.

En las libretas ARIMA_***. se puede ver la aplicación de ambos métodos para los tres puertos, tanto con datos mensuales como con datos anuales.

A modo de ejemplo se muestra el detalle de la ejecución del método ARIMA a los datos anuales del puerto de Valencia.

Nota: El detalle del resto de modelos y puertos se puede consultar en el anexo correspondiente.

5.1.2. Resultados aplicación ARIMA datos Anuales Puerto Valencia

- Parámetros seleccionados: ARIMA(2, 2, 1) MSE=1188.866
- Gráfica comparativa de datos reales y estimados y la predicción para los próximos años

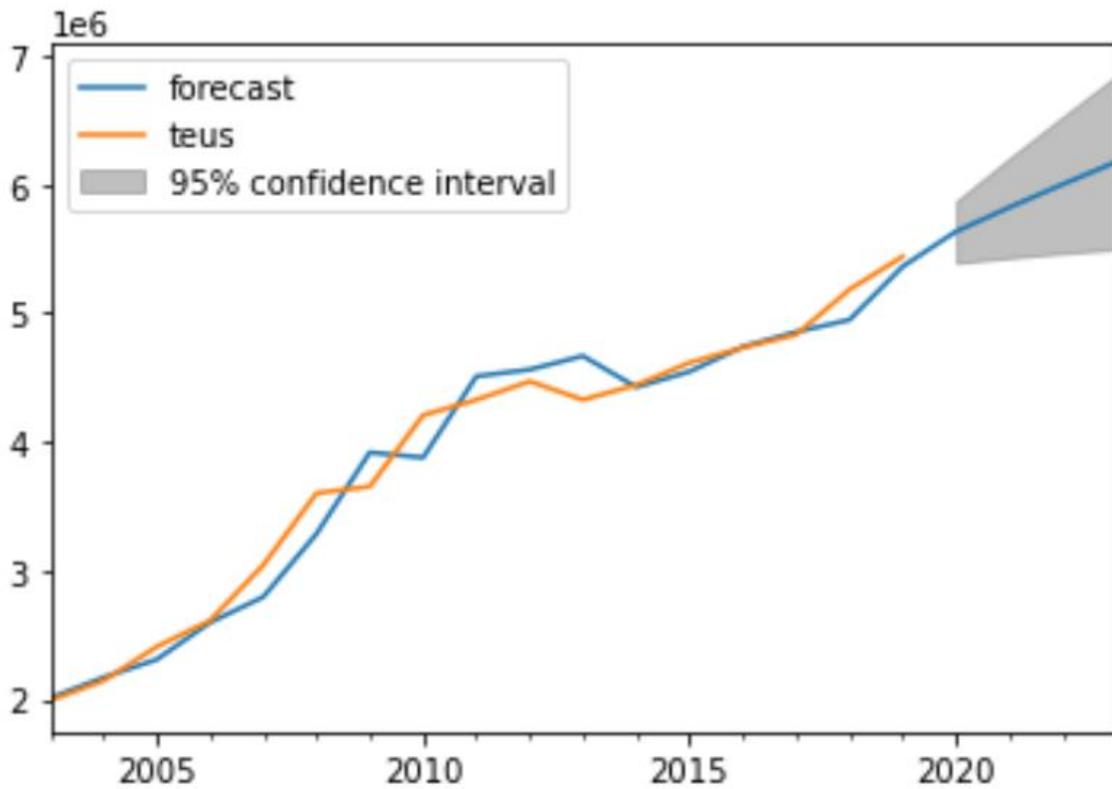


Figura 20: Gráfica ARIMA Anual Predicción - Puerto de Valencia

- Evaluación de la predicción. usando un 80% de datos para entrenamiento y un 20% para la validación

Año	Observado (Real)	Predicción
2010	4206937	3958023
2011	4327371	4723643
2012	4469754	4545867
2013	4327838	4965581
2014	4441949	4404216
2015	4615196	4416448
2016	4732136	4765173
2017	4832156	4891204

2018	5182665	4960328
2019	5439827	5356852

Tabla 13: ARIMA Valores Observados & Predicción Valencia

- Total error: 270799
- Comparativa Predicciones y observaciones en los datos de validación

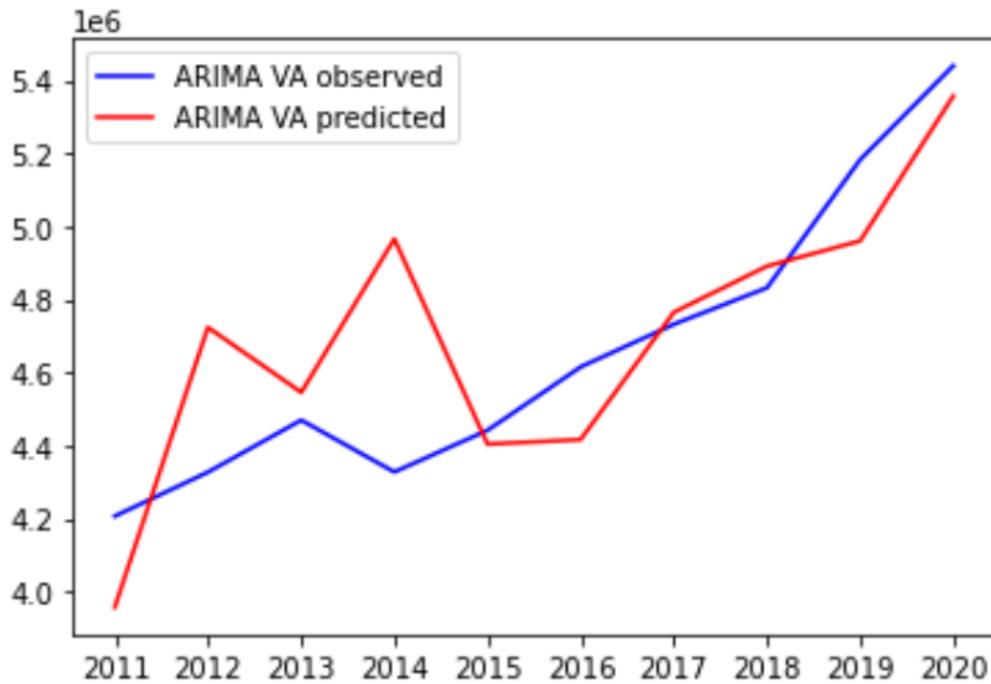


Figura 21: Gráfica ARIMA Anual Comparativa - Puerto de Valencia

5.2. SARIMA

El modelo ARIMA puede generalizarse aún más para considerar el efecto de la estacionalidad. En ese caso, se habla de un modelo SARIMA (Seasonal AutoRegressive Integrated Moving Average).

La aplicación del modelo sólo sirve para datos con una clara estacionalidad, en nuestro caso se puede aplicar a los datos mensuales que presentan mucha estacionalidad tal y como se puede observar en la gráfica "Seasonality" de la siguiente figura.

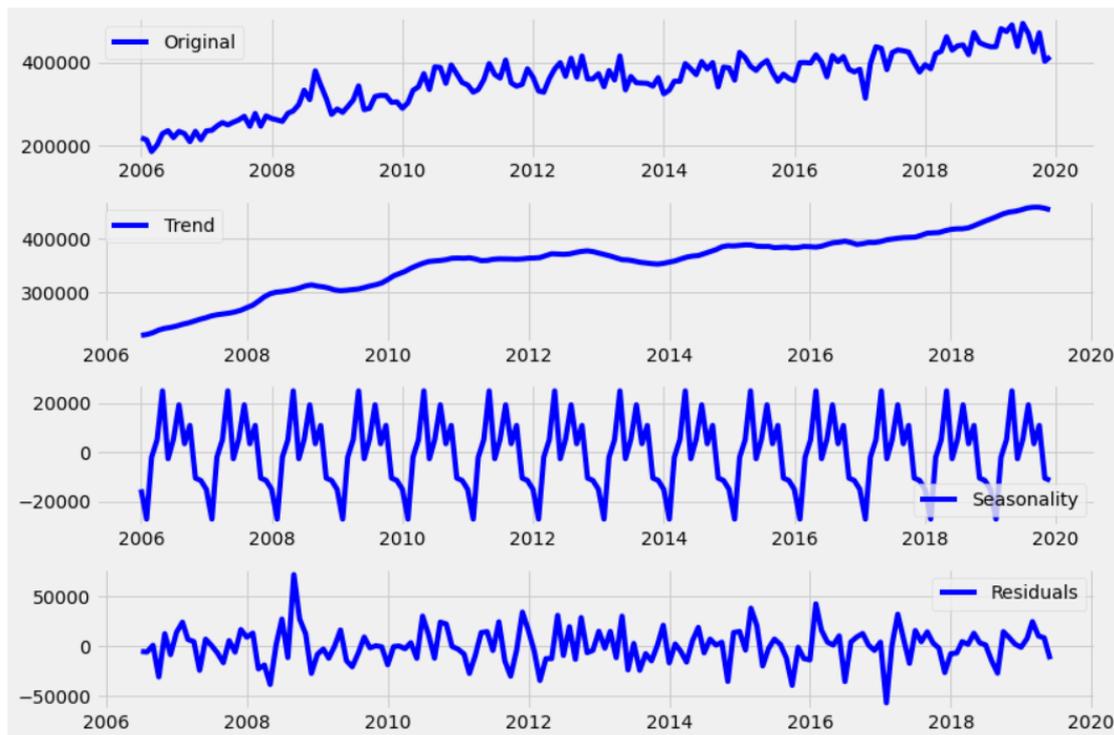


Figura 22: Descomposición Serie en tendencia y estacionalidad

Se usará este modelo para realizar predicciones sobre las series temporales mensuales de los 3 puertos sobre los que se está realizando en análisis.

5.3. ARIMAX

El modelo ARIMAX es similar al modelo ARIMA, pero en vez de considerar únicamente la variable del estudio (o endógenas), en este caso el número de TEUS movidos por año, se consideran adicionalmente variables externas (o exógenas) que están correlacionadas con la variable del estudio y pueden ayudar mejorar ligeramente la comprensión de la serie temporal.

En la figura siguiente se muestra la serie temporal anual correspondiente al puerto de Valencia

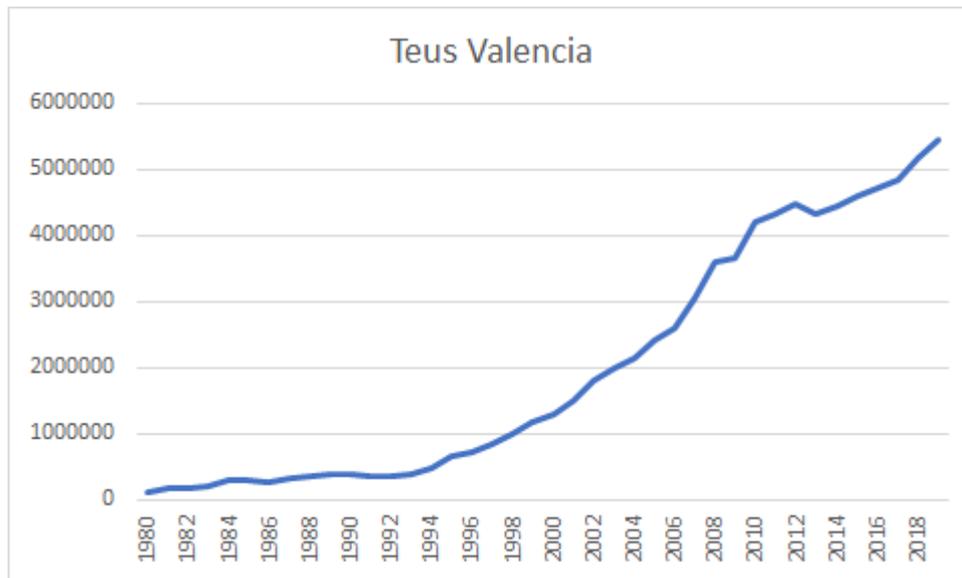


Figura 23: : Gráfica Volumen TEUS histórico Puerto de Valencia

Si comparamos la serie de TEUS con el crecimiento de la economía mundial (PIB Mundial) o economía nacional (PIB español), podemos ver que las gráficas se asemejan y es algo que entra dentro de lo esperado, ya que a una mayor economía está relacionada con un mayor comercio y movimiento de mercancías.

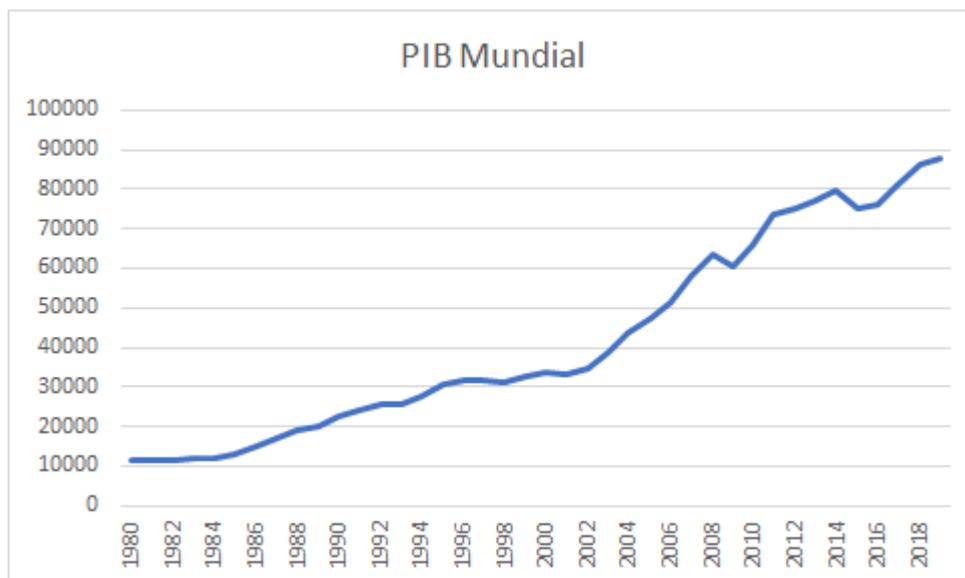


Figura 24: Gráfica PIB Mundial 1980-2019

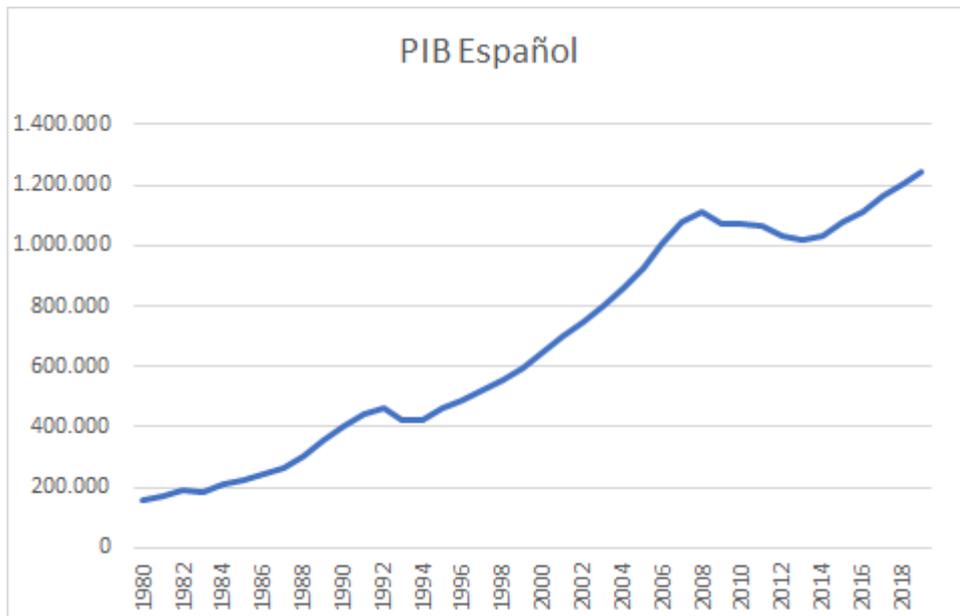


Figura 25: : Gráfica PIB Español 1980-2019

La gráfica de TEUS Valencia se asemeja un poco más a la gráfica del PIB Mundial que a la gráfica del PIB Español, y eso es debido a que los grandes puertos como el de Valencia (5to de Europa en 2019), operan a nivel global.

Para poder usar variables exógenas para realizar predicciones, se tiene que añadir las predicciones de la serie exógena. A través de la web del “International Monetary Fund” (International Monetary Fund, 2020) se obtienen las predicciones del PIB mundial desde el 2020 al 2025. Por tanto completamos la serie PIB Mundial con los valores estimados de 2019 a 2025

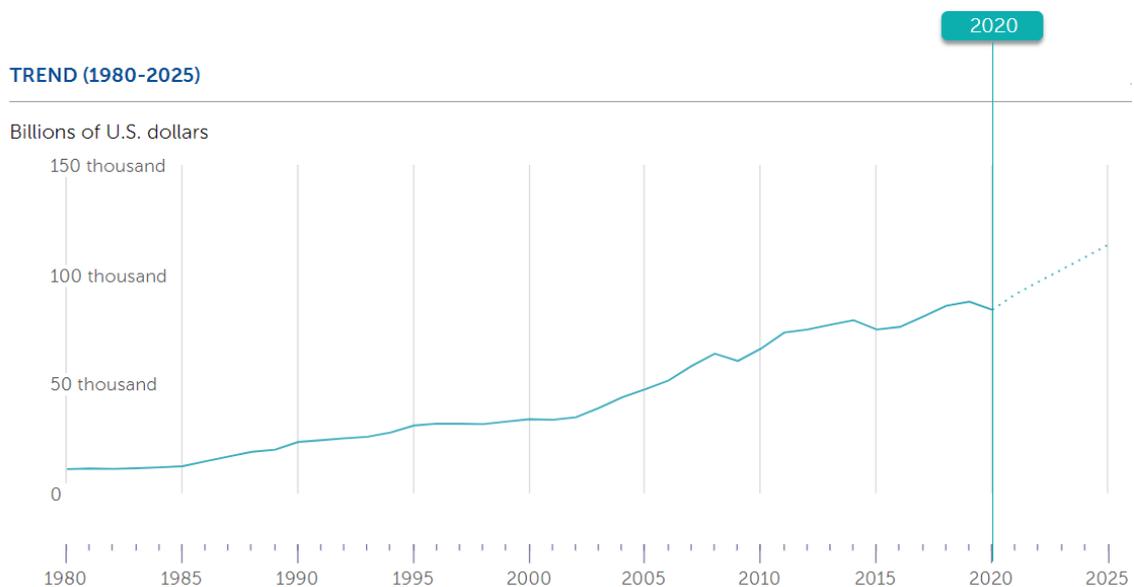


Figura 26: Predicciones del PIB mundial desde el 2020 al 2025

Año	Predicción PIB Mundial
2019 (real)	87.55 x 1000M
2020 (predicción)	84.84 x 1000M
2021 (predicción)	91.03 x 1000M
2022 (predicción)	96.7 x 1000M
2023 (predicción)	102.25 x 1000M
2024 (predicción)	107.81 x 1000M
2025 (predicción)	113.48 x 1000M

Tabla 14: Predicción PIB Mundial próximos años

Durante el proceso de carga de datos internacionales se han cargado hasta 47 variables endógenas agrupadas en varias categorías:

- Volumen de TEUS movido por puertos mediterráneos
- Volumen de TEUS movido por países y regiones influyentes para el transporte español
- PIB de Regiones y países influyentes para el transporte nacional
- Toneladas transportadas por tipo de mercancía

Estos datos son anualizados. Ha sido imposible conseguir estos datos en series mensuales, ante esta imposibilidad, este modelo se aplicará a los datos anualizados del puerto de Valencia, Barcelona y Algeciras. La serie abarca datos de 1980 a 2019.

Para los tres puertos se ha generado el modelo ARIMAX usando datos PIB Mundial, PIB Español y una Mezcla de PIB Español y Mundial

5.4. Redes neuronales recurrentes (RNN)

Mediante este análisis se pretende probar cómo funcionan las redes neuronales para la predicción del cálculo de TEUS, para ello se ha partido de los apuntes “Anàlisi de sèries temporals (TSA) emprant xarxes neuronals recurrents (RNN) amb Keras” de la asignatura de Deep learning.

Mediante este análisis se descomponen las series temporales en tendencia y en parte estacional y se crea la red neuronal para la parte estacional. El resultado predicho es la suma de las predicciones en base a la tendencia y a la estacionalidad.

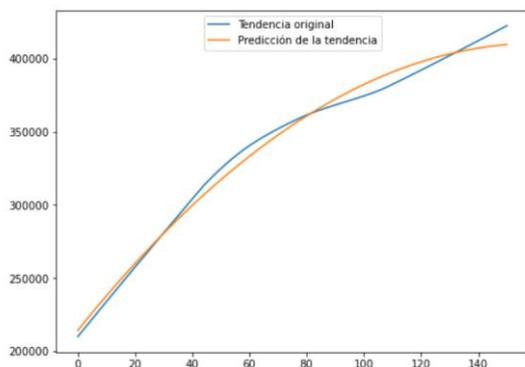


Figura 27: Predicción de la Tendencia

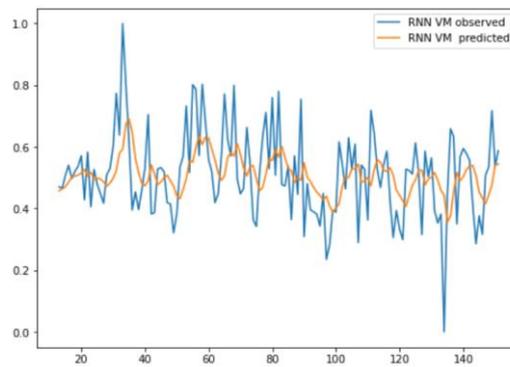


Figura 28: Predicción de la estacionalidad

Este modelo se ha podido aplicar tanto para las series analizadas como mensuales de los 3 puertos. De todos los análisis realizados, el análisis con RNN ha dado los peores resultados

5.5. Facebook Prophet

Prophet es un algoritmo desarrollado por Facebook en el 2017 y que permite hacer predicciones de series temporales y desde su lanzamiento está teniendo mucha aceptación por parte de los usuarios. El algoritmo dispone de una información muy completa y muchas opciones de configuración (Facebook Prophet, s.f.).

En el presente estudio, se ha utilizado este algoritmo para predecir las series temporales mensuales de los tres puertos bajo estudio. Se ha configurado el algoritmo para indicarle que se usan datos mensuales, ya que por defecto está pensado para series temporales con datos diarios.

En la Figura 29 se observa las predicciones obtenidas por Prophet para el puerto de Valencia de 2020 a 2024.

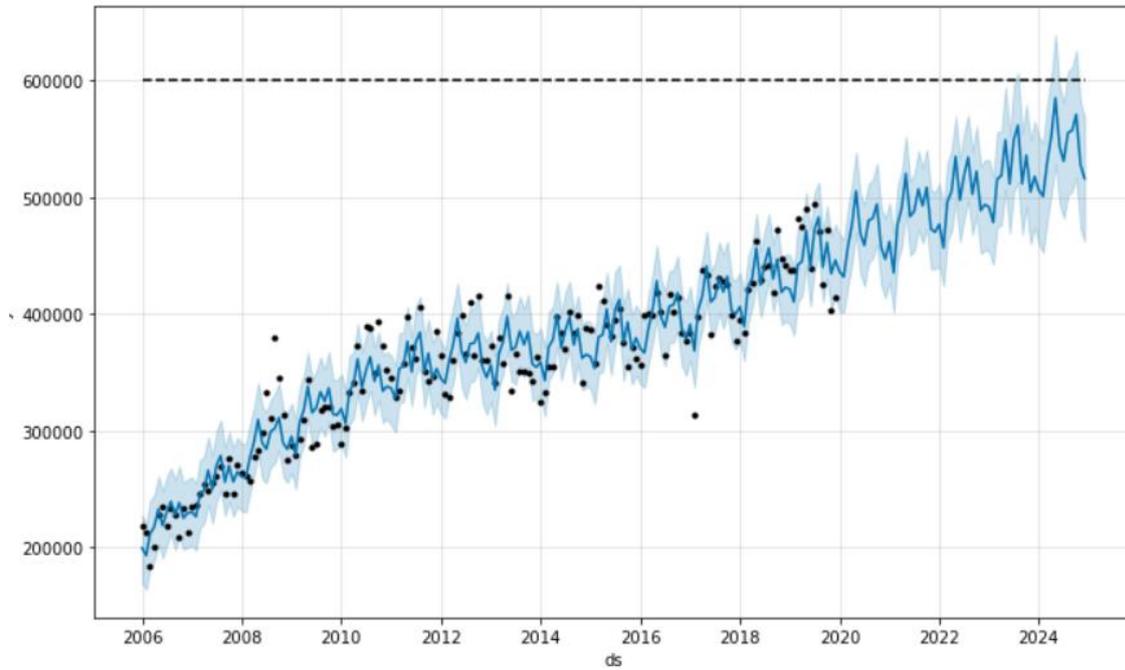


Figura 29: Predicciones Puerto de Valencia de 2019 a 2024 usando Prophet

5.6. Resumen del Análisis de Datos

Tras la realización de los análisis se muestra una tabla resumen con los resultados, en ella se muestra el error obtenido tras la realización de las predicciones. A un menor error, mejor se ajusta el modelo a la serie

5.6.1. Series Anuales

Tras aplicar 5 métodos distintos a las series anuales de los 3 puertos nacionales más importantes, se puede observar que las mejores predicciones varían para cada puerto. No hay un mismo método que funcione en general mejor para todos los casos. En el caso de Valencia, el método que mejor funciona es el ARIMAX usando como datos Exógenos el PIB Español. Para el puerto de Barcelona, funciona mejor el método ARIMAX usando una combinación de PIB Español y Mundial, aunque el PIB español también funciona de manera similar. En cambio, para Algeciras funciona mejor el método ARIMA.

Modelo	Valencia	Barcelona	Algeciras
ARIMA	270799	478966	402894
ARIMAX PIB Mundial	217097	281362	417743
ARIMAX PIB ESPAÑOL	190169	334595	464829

ARIMAX PIB MUNDIAL+ESPAÑOL	190541	277246	447108
RNN	465672	1263340	926267

Tabla 15: Error obtenido por los diferentes algoritmos con las series anuales

5.6.2. Series mensuales

Para analizar las series mensuales se han usado 4 métodos, dos de los cuales ya se habían usado con las series anuales “ARIMA y RNN” y dos nuevos métodos, específicos para series con tendencias cíclicas, se han incorporado al análisis. El método SARIMA ha sido el que ha realizado mejores predicciones portado para Valencia y Algeciras. Para el Puerto de Barcelona el método ARIMA ha sido el que mejor ha funcionado, aunque el método SARIMA también ha obtenido resultados satisfactorios.

Modelo	Valencia	Barcelona	Algeciras
ARIMA	29430	18836	22785
SARIMA	25323	20249	20167
RNN	43956	28358	27370
Prophet	33916	40782	28200

Tabla 16: Error obtenido por los diferentes algoritmos con las series anuales

5.6.3. Conclusiones y limitaciones sobre los análisis realizados

Debido a la falta de experiencia con los modelos usados, se realizaron algunas pruebas que no tenían mucho sentido, como son las pruebas de modelos SARIMA con datos anuales sin estacionalidad.

También se dedicó un esfuerzo importante de tiempo en la búsqueda y carga de datos internacionales que finalmente no se usaron en el análisis o bien porque las series no eran lo bastante completas o tras un primer análisis, no ayudaban a explicar bien la serie de TEUS. Por otro lado, no se ha podido aplicar el método ARIMAX a las series mensuales, al no disponer de datos mensuales de las variables exógenas a comparar.

Tenía esperanza que el modelo ARIMAX diera mejores resultados y que ayudará a explicar la evolución del movimiento de TEUS. Es evidente que la evolución económica y el transporte de

mercancías afecta al número de contenedores movidos a nivel global. También es evidente que el volumen de mercancías movidas en el Mediterráneo es bastante constante en los últimos 20 años y al haber dos grandes puertos (Tánger y Pireos) en el Mediterráneo que crecen más de un 20% anual en los últimos años y que tienen previsión de seguir esta tendencia hasta doblar su tráfico en 5 años. El gran crecimiento de estos dos puertos repercutirá negativamente en el resto de los puertos del Mediterráneo, que verán mermado su tráfico futuro. Este análisis no se ha podido llevar a término.

No hay un algoritmo que funcione mejor o peor para todos los casos, por tanto para poder realizar mejores predicciones, para cada serie habría que probar los diferentes modelos y con diferentes parámetros. Estos modelos son muy sensibles a los cambios de tendencias no esperados.

De todos los modelos, el más fácil de usar y con mejor documentación es el algoritmo Prophet. A priori si se trabajan más las diferentes opciones de configuración (hiperparámetros), previsiblemente se mejoren los resultados obtenidos, llegando a ser el mejor modelo para realizar predicciones en base a series temporales mensuales.

Las tendencias de las series temporales de los puertos suelen servir para predecir el futuro pero existen ciertas variables que pueden condicionar mucho el futuro y que no están explicadas por las series temporales. Por ejemplo, la capacidad máxima anual del Puerto de Valencia es de 7M de contenedores y se estima que en 4-5 años, el puerto alcanzará este límite. Por ese motivo, desde hace varios años, existe una propuesta para construir una nueva terminal con capacidad de 5M de TEUS que aumentaría la capacidad hasta 12M de TEUS. Con casi total probabilidad, cuando se abra la nueva terminal, el crecimiento del volumen de contenedores aumentará sustancialmente y esto no se puede estimar analizando los datos pasados. También existen otros factores que pueden afectar al número de contenedores movidos, como por ejemplo, un mayor calado del puerto permite el acceso a buques más grandes, cambio en el concesionario de una terminal, adquisición de nueva maquinaria,...

6. Cuadros de Mando (BI)

Tras la carga de datos y el análisis realizado, se han usado la herramienta de BI Qlik para crear cuadros de mandos que nos permiten analizar los datos. Qlik es la herramienta usada por INTEGRA2 para desarrollar su BI. De este modo el trabajo realizado para la creación de cuadros de mando se podrá reutilizar e integrar fácilmente en INTEGRA2.

La licencia disponible actualmente “Qlik Desktop” no permite la publicación de los cuadros de mando.

Inicialmente se han cargado en Qlik las tablas:

- spanish_historic_annual y spanish_historic_contenedores que contiene la información anualizada del 1973 a 2019.
- spanish_month_contenedores que contiene información mensual de enero de 2006 a septiembre de 2020

En la siguiente figura se puede ver el esquema de datos cargados. En la sección de anexos se incluyen los scripts completos de carga de datos.



Figura 30: Esquema de datos cargados en Qlik

Tras la carga de los datos se han realizado los siguientes cuadros de mando:

Cuadro de Mandos 1 : TEUS Mensuales ESP

A través de este cuadro de mando se puede observar la tendencia mensual de los puertos desde el principio de la serie hasta septiembre de 2020 (Componente TEUS Mensual). Además, también muestra los TEUS agregados por mes, para observar cómo se comportan los puertos cada mes (Componente TEUS Agregado por Mes). Para facilitar el análisis, se incluye un filtro que permite filtrar por mes, año y puerto.

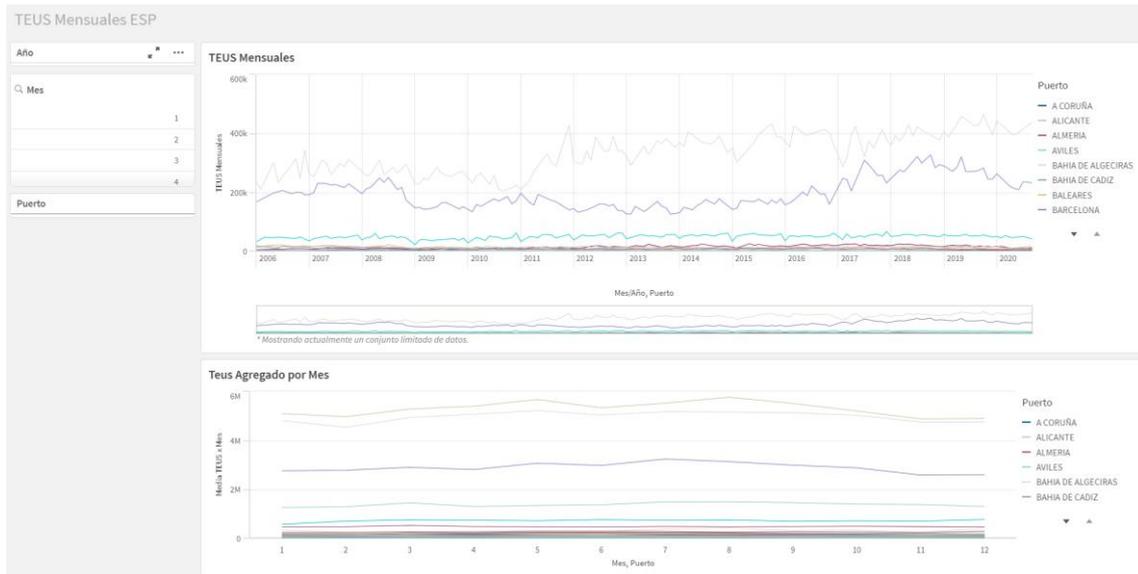


Figura 31: Volumen de TEUS Mensuales

Por ejemplo, si filtramos las anualidades de 2019 y 2020, se puede ver como el primer semestre de 2020 se ha visto afectado el volumen de contenedores por el Cobib-19 y como se aprecia su recuperación desde Julio de 2020. Los puertos de Algeciras y Barcelona han recuperado el volumen previo a la crisis. En cambio el puerto de Barcelona se ha visto más afectado y no ha recuperado el volumen previo a la crisis (ver Figura 32)

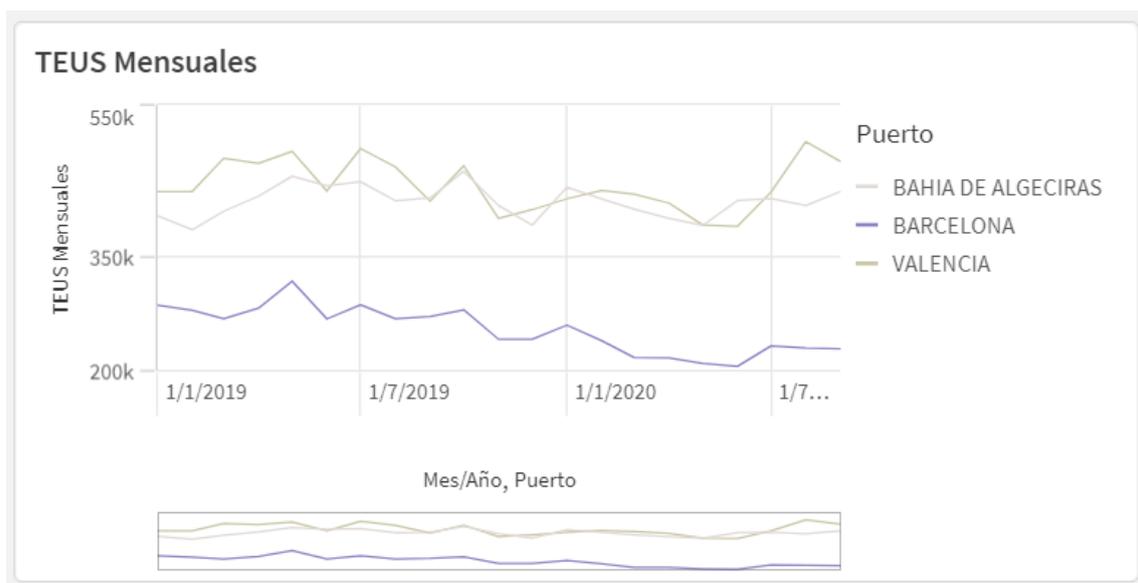


Figura 32: Volumen de TEUS Mensuales Filtrado

Cuadro de Mandos 2 : TEUS Anuales ESP

A través de este cuadro de mando se puede observar la tendencia anual del movimiento de TEUS de los puertos desde el principio de la serie hasta 2019 (Componente TEUS Anuales). Además también muestra los TEUS acumulados desde el principio de la serie (Componente TEUS Acumulados). Para facilitar el análisis, se incluye un filtro que permite filtrar por año y puerto.

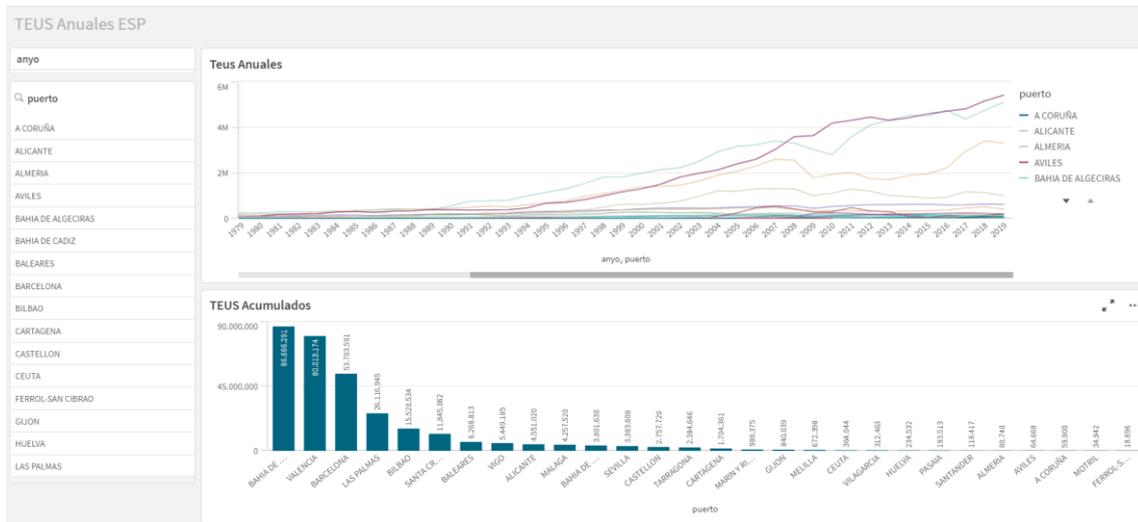


Figura 33: Volumen de TEUS Anuales

Gracias al Filtro, podemos ver qué puertos ocupaban las primeras posiciones respecto al movimiento de contenedores hasta el 1980.

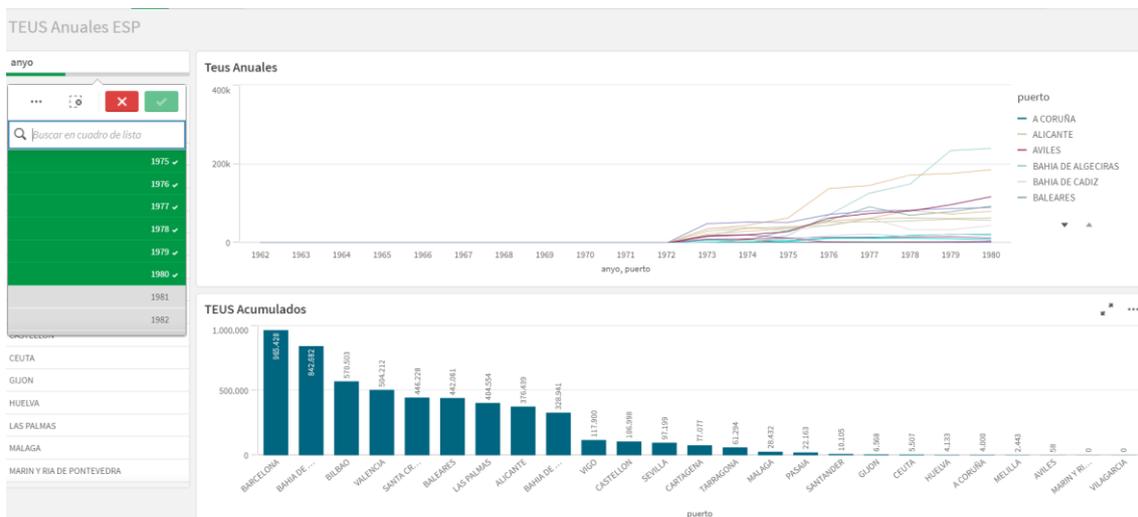


Figura 34: Volumen de TEUS anuales Filtrados

Cuadro de Mandos 3 : Tráfico Total ESP

A través de este cuadro de mando se puede observar la tendencia anual del movimiento total de mercancías de los puertos desde el principio de la serie hasta 2019 (Componente Mercancías Totales Anuales). Además también muestra la mercancías transportadas acumuladas desde el principio de la serie (Componente Mercancía Totales Acumuladas). Para facilitar el análisis, se incluye un filtro que permite filtrar por año y puerto.

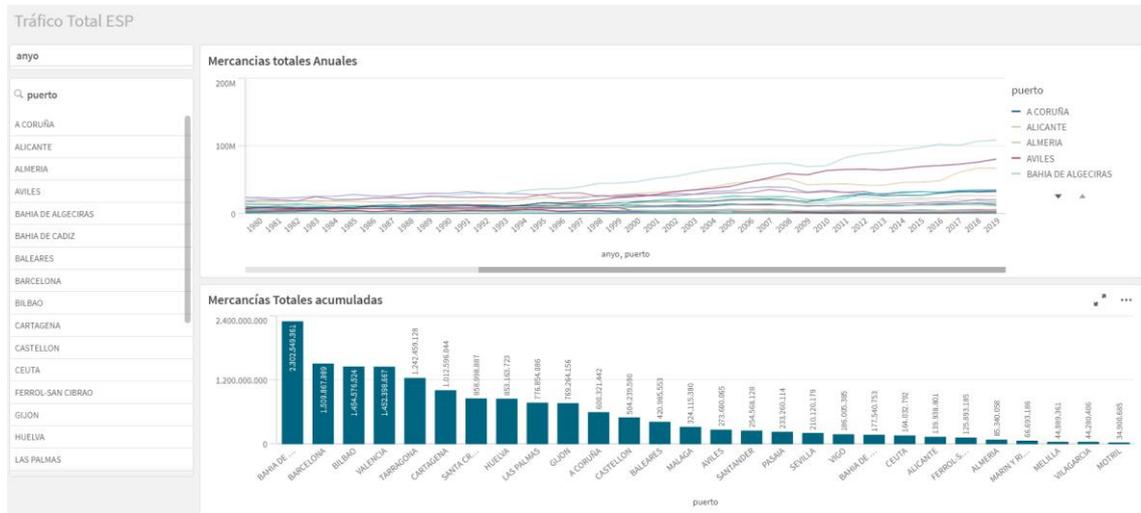


Figura 35: Volumen de Total de Mercancías Anuales

Cuadro de Mandos 4 : Tráfico Total Detalles ESP

A través de este cuadro de mando se puede observar una tabla con los detalles del movimiento de mercancías desglosado por tipo de mercancías.

Tráfico Total Detalles ESP							
puerto	Valores						
	Graneles Líquidos	Graneles Sólidos	Mercancía General	Pesca Fresca	Avituallamiento	Trafico Local	Total Trafico Calculado
A CORUÑA	401168327	134469638	29285305	4143383	15925400	13734285	-
ALICANTE	24497355	52821031	59896049	489001	2235210	64	-
ALMERIA	257785	73767262	9689786	65324	1495472	64429	-
AVILES	27138428	160095595	82022586	745369	3564705	33379	-
BAHIA DE ALGECIRAS	908801980	69650361	1177678006	1513211	81732851	44315285	-
BAHIA DE CADIZ	20828083	99951152	83192881	1861613	11246845	460379	-
BALEARES	82184158	57378187	266482174	148738	13385797	1406499	-
BARCELONA	439017724	233626834	798301407	390962	36766356	1764702	-
BILBAO	738264875	281539170	336352816	58375	16897310	79245631	-
CARTAGENA	821082667	143566340	36937374	97330	9846611	735579	-
CASTELLON	355622772	86399221	58862469	572209	2539876	243041	-
CEUTA	83283933	2569286	36253014	75482	40825123	1043955	-
FERROL-SAN CIBRAO	24101491	93677856	7873446	2826	165591	22938	-
GIJON	69034562	646373318	44167541	343412	7519948	1825145	-
HUELVA	562354431	253906563	26589627	1547589	7050066	1488950	-
LAS PALMAS	198567566	36906227	407130406	8708244	125541066	575	-
MALAGA	207709092	54887487	54694043	528574	6287069	8508	-
MARIN Y RIA DE PONTEVEDRA	393049	28777865	29824220	861341	3815857	3020837	-
MELILLA	3128068	13154403	27117390	156594	1432546	0	-
MOTRIL	19134556	9731534	5624375	34794	375426	0	-
PASAJA	33224513	87159995	104460965	2098019	6034011	282612	-
SANTA CRUZ DE TENERIFE	528127238	44276616	218311930	1322781	66797883	156195	-
SANTANDER	45003199	145356836	58529417	460745	5187536	30394	-

Figura 36: Detalle Tráfico total desglosado por tipo de Mercancías Anuales

La tabla con la información de las mercancías permite el filtrado por puertos y años. En la siguiente figura se puede observar el detalle de las mercancías transportadas por los puertos de Algeciras, Barcelona y Valencia para las anualidades de 2017,2018 y 2019. Gracias a esta tabla podemos ver que Valencia es el puerto que mueve más mercancías generales (TEUS), pero que a nivel de Graneles y Avituallamiento está muy lejos de Algeciras y Barcelona.

Tráfico Total Detalles ESP								
puerto Q año Q		Valores						
		Graneles Líquidos	Graneles Sólidos	Mercancía General	Pesca Fresca	Avituallamiento	Trafico Local	Total Trafico Calculado
● BAHIA DE ALGECIRAS		91115120	4795176	207846135	2829	7773829	6798958	-
2017		28774987	2102907	65452560	1014	2833641	2390861	101555970
2018		31763060	1718439	69062429	878	2541866	2274355	107361027
2019		30577073	973830	73331146	937	2398322	2133742	109415050
● BARCELONA		45914191	12794045	133215105	9707	4883531	0	-
2017		14483623	4465637	41120874	2961	1293845	0	61366940
2018		15298349	4257025	46339750	3171	1857961	0	67756256
2019		16132219	4071383	45754481	3575	1731725	0	67693383
● VALENCIA		8233192	7013090	215156480	6325	835485	0	-
2017		3203487	2278857	67767129	1937	308467	0	73559877
2018		1909692	2544075	71971743	2894	192697	0	76621101
2019		3120013	2190118	75417608	1494	334321	0	81063554

Figura 37: Detalle Tráfico Total Desglosado Filtrado por Puerto y Año

6.1. Conclusiones sobre Qlik

Se han realizado varios cuadros de mando que muestran la información relevante de los puertos españoles de manera mensual y anualizada. Faltaría integrar en dichos cuadros de mando las predicción obtenidas con los diferentes algoritmos predictivos utilizados en el proyecto o al menos con el que da mejores resultados.

Se ha elegido Qlik, por ser la herramienta de BI utilizada por INTEGRA2 y disponer licencia de uso. Como no se había trabajado con dicha herramienta hasta el comienzo del proyecto, se han tenido que realizar unos tutoriales de iniciación de la página oficial de la herramienta. Tras dichos tutoriales se han podido realizar los cuadros de mando definidos en el proyecto.

La experiencia con Qlik sense ha sido buena, ya que es bastante intuitiva y potente. El problema es que con la licencia disponible no es posible publicar los cuadros de mando públicamente y solo se pueden explotar internamente en la organización.

7. Conclusiones y Trabajos Futuros

Se han analizado las series temporales mensuales (de enero de 2006 a septiembre de 2019) y las anuales (de 1963 a 2019) de los tres puertos españoles más importantes. Tras la realización del trabajo, se puede afirmar que no hay un modelo que funcione mejor para todas las series. Cada serie se ajusta mejor a uno de los métodos utilizados. Dado que el entrenamiento de los modelos es muy rápido, sería recomendable utilizar todos los métodos para ver cual se ajusta mejor a un nuevo puerto. En caso de no querer probar más de un método, los métodos ARIMAX y SARIMA ajustan bastante bien las series temporales de TEUS anuales y semestrales respectivamente.

Al basarse en datos pasados, las predicciones funcionan mucho mejor si el puerto sigue una tendencia más o menos continua de varios años. En cuando sucede algún acontecimiento externo (exógeno) que rompe la tendencia, las predicciones no son muy fiables, tal y como se observa en las anualidades de 2009-2010 cuando ocurrió la crisis económica mundial o actualmente con la crisis del COVID.

A parte de estas variables externas, también hay variables dependientes de cada puerto que pueden considerar notablemente el volumen futuro. Por ejemplo, las mejoras en infraestructuras o cambios de adjudicatarios en las concesiones condicionan notablemente el volumen futuro. Estos cambios suelen estar planificados y de algún modo se deberían considerar en los modelos, para tratar de predecir el comportamiento considerando estas variables. Por ejemplo, el Puerto de Valencia tiene un crecimiento bastante constante, pero se espera que cuando entre en funcionamiento la nueva terminal, se espera que este crecimiento aumente considerablemente. Y si no se quiere errar en las primeras predicciones tras la puesta en funcionamiento de la terminal, se debería considerar dicha información.

Pese a que el modelo ARIMAX, ha sido el que mejor ha funcionado para dos de las series anuales, no se ha podido experimentar con muchas variables exógenas. De hecho, sólo se han incorporado al modelo los datos relativos al PIB Mundial, PIB Español y una combinación de ambos. El motivo de no haber probado más variables ha sido principalmente por no disponer de series temporales más completas (en años y periodicidad) de dichas variables, como por ejemplo las series temporales de los puertos del Mediterráneo más grandes.

El proceso de carga de los datos ha sido el más costoso de desarrollar, ya que no existía una única fuente de datos con todos los datos necesarios y se desconocía la herramienta "Pentaho" utilizada para cargar los datos anuales.

El método por medio de Facebook Prophet, se incorporó a los análisis a última hora y no dio tiempo a experimentar en profundidad utilizando diferentes configuraciones de hiperparámetros. Al ser un método más actual y con mucho recorrido, esperaba obtener mejores resultados.

Objetivos

En cuanto a los objetivos planteados al inicio del proyecto. Tras la finalización del mismo, se puede afirmar que se ha conseguido el objetivo principal del proyecto, que consistía en tener unificados en un único lugar los datos históricos de todos los puertos españoles para poder compararse entre ellos e incorporar a dichos datos, predicciones sobre volúmenes futuros esperados.

En relación al objetivo secundario de integrar los resultados del TFM en una solución real, no se ha hecho efectivo aun, pero se han presentado los resultados del TFM internamente en Prodevelop, y la idea ha gustado mucho. Se espera que durante el primer semestre de 2021, se añada tanto la información de todos los puertos españoles y las predicciones a la solución de Business Intelligence de INTEGRA2.

A nivel personal, se ha conseguido con creces el objetivo de desarrollar un proyecto analítico de principio a fin y el aprender a utilizar métodos de series temporales que no se habían estudiado a lo largo del máster.

Trabajos futuros

Cómo trabajo futuro, queda pendiente el integrar los resultados del TFM en la herramienta INTEGRA2. No será una integración automática, ya que se requerirá de adaptar los datos a la estructura de la base de datos de la herramienta y habrá que extender las predicciones a más mercancías. También habrá que revisar los cuadros de mandos y adaptarlos a la apariencia de los cuadros de mandos existentes. No se descarta publicar dichos cuadros de mando de manera pública para que todo el mundo interesado en estos datos pueda consultarlos.

Por otro lado, sería interesante seguir profundizando con los métodos ARIMAX y Prophet para tratar de conseguir mejorar las predicciones obtenidas en este TFM.

8. Glosario

ACRÓNIMOS	DESCRIPCIÓN
AAPA	American Association of Port Authorities
ARIMA	AutoRegressive Integrated Moving-Average
ARIMAX	seasonal auto regressive integrated moving average
BI	Business Intelligence
CORS	Cross-origin resource sharing
DFM	Dynamic Factor Model
DOC	DOCumentación
ELK	Elastic Search, Logstash and Kibana
ETL	Extract,Load and Transform
FB	FaceBook
IA	Inteligencia Artificial
JS	JavaScript
KPI	Key Performance Indicator
PIB	Producto interior bruto
PT	Paquetes de Trabajo
RNN	Recurrent Neural Networks
SARIMA	Seasonal AutoRegressive Integrated Moving-Average
SVM	Support Vector Machine
SVR	Support Vector Regression
T	Tarea
TEUS	Twenty-foot Equivalent Unit

9. Bibliografía

- Alfatec*. (s.f.). Obtenido de <https://alfatecsistemas.es/>
- AYESA*. (s.f.). Obtenido de <https://www.ayesa.com/es/sectores/smart-life/smart-grid/87-brainport-analytics>
- Brainport*. (s.f.). Obtenido de AP Algeciras: <https://innovacion.apba.es/brainport-analytics/>
- Caridad y Ocerin, J. (2008). *Econometria: Modelos Económicos Y Series Temporales Vol 1*. Ed Reverte.
- CARTO*. (s.f.). Obtenido de <https://carto.com/>
- Chart.js*. (s.f.). Obtenido de <https://www.chartjs.org/>
- Container port Traffic (TEU)*. (2020). Obtenido de world bank: <https://data.worldbank.org/indicator/IS.SHP.GOOD.TU>
- D3.js*. (s.f.). Obtenido de <https://d3js.org/>
- Data Integration*. (s.f.). Obtenido de Hitachi: <https://community.hitachivantara.com/s/article/data-integration-kettle>
- Dragam, D., Kramberger, T., & Intihar, M. (2014). A comparison of Methods for Forecasting the Container Throughput in North Adriatic Ports. *Norfolk*. VA-USA.
- Dragan, D., Lisec, A., Intihar, M., & Kramberger, T. (2017). The impact of macroeconomic indicators of forecasting a total cargo throughput in the adriatic seaport. *Internationa Symposium of Logistics*.
- EDISA*. (s.f.). Obtenido de www.edisa.com/
- Elastic*. (s.f.). Obtenido de <https://www.elastic.co/es/>
- EuroStat*. (2020). Obtenido de Eurostat: <https://ec.europa.eu/eurostat/data/database>
- Eurostat Volumen Contenedores por país*. (s.f.). Obtenido de Eurostat: https://ec.europa.eu/eurostat/databrowser/view/mar_mg_am_cvh/default/table?lang=en
- Eurostat Volumen de contenedores por Puerto*. (s.f.). Obtenido de Eurostat: https://ec.europa.eu/eurostat/databrowser/view/mar_mg_am_pvh/default/table?lang=en
- Facebook Prophet*. (s.f.). Obtenido de Facebook: <https://facebook.github.io/prophet/>
- Freeboard*. (s.f.). Obtenido de <https://freeboard.io/>
- Grafana*. (s.f.). Obtenido de <https://grafana.com/>

Hing Kai, C., Shuojiang, X., & Xiaoguang, Q. (2019). A comparison of time series methods for forecasting container throughput. *International Journal of Logistics Research and Applications*, 294-303.

IMO. (2020). *Just in Time Arrival Guide*. International Maritime Organization. Obtenido de <http://www.imo.org/en/OurWork/PartnershipsProjects/Documents/GIA-just-in-time-hires.pdf>

International Monetary Fund. (2020). Obtenido de IMF: <https://www.imf.org/external/datamapper/NGDPD@WEO/WEOWORLD>

Jupyter Notebooks. (s.f.). Obtenido de Jupyter: <https://jupyter.org/>

List of Busiest container ports. (01 de 10 de 2020). Obtenido de Wikipedia: https://en.wikipedia.org/wiki/List_of_busiest_container_ports

Monden, Y. (1996). *El Just in Time hoy en Toyota*. Bilbao: Deusto.

Movimiento de mercancías por tipo y región. (2020). Obtenido de UNCTAD: <https://unctadstat.unctad.org/wds/TableViewer/tableView.aspx?ReportId=32363>

MySQL. (s.f.). Obtenido de MySQL: <https://www.mysql.com/>

MySQL Workbenck. (s.f.). Obtenido de MySQL Workbenck: <https://www.mysql.com/products/workbench/>

(2020). *One Hundred Ports*. Lloyds List. Obtenido de <https://lloydslist.maritimeintelligence.informa.com/one-hundred-container-ports-2020>

Pani, C., Vanelslander, T., Fancello, G., & Cannas, M. (2020). Prediction of late/aerly arrivals in container terminals - A qualitative approach. *European Journal of Transport and Infrastructure Research*, 536-550.

Parra, F. (2019). *Estadística y Machine Learning con R*.

POI API Apache. (s.f.). Obtenido de Apache: <https://poi.apache.org/apidocs/dev/overview-summary.html>

PowerBI. (s.f.). Obtenido de Microsoft: <https://powerbi.microsoft.com/>

Prodevelop. (s.f.). Obtenido de <https://www.prodevelop.es/>

Puerto de Valencia. (s.f.). Obtenido de <https://www.valenciaport.com/>

Puertos del Estado. (s.f.). Obtenido de Puertos del Estado: <http://www.puertos.es/es-es>

Qlik. (s.f.). Obtenido de <https://www.qlik.com/>

Rashed, Y., Meersman, H., Van de Voorde, E., & Vanelslander, T. (2017). Short-term forecast of container throughput: An ARIMA-intervention model for the Port of Antwerp. *Maritime Economics & Logistics*, 749-764.

RShiny. (s.f.). Obtenido de <https://shiny.rstudio.com>

Sang-Yoon, L., & Hyunwoo, L. (2017). Containre port volume forecasting framework with the implications of dredging projects. *Maritime Economics & Logistincs* .

Tableau. (s.f.). Obtenido de <https://www.tableau.com/>

Thoughtspot. (s.f.). Obtenido de <https://www.thoughtspot.com/>

Transeop. (15 de octubre de 2020). Obtenido de Transeop:
<https://www.transeop.com/blog/smart-ports/369/>

Wen, Y., & Ching, W. (2009). A comparison of univariate methods for forecasting container throughput volumes. *Mathematical and computer Modelling*, 1045-1057.

10. Anexos

10.1. Anexo Recursos Públicos del Proyecto.

Todos datos y desarrollos del proyectos están públicos a través de una carpeta pública ubicada en Google Drive y accesible a cualquier persona con el siguiente enlace.

URL Carpeta Pública de Google Drive
https://drive.google.com/drive/folders/1Zyi984gEsEngOFEvvMTTcbzteRbCByCf?usp=sharing

Dentro de esta carpeta se han agrupado los recursos en diferentes subcarpetas:

Nombre Carpeta	Descripción
datos_historicos_espanya	Contiene datos históricos anuales de los puertos españoles
datos_internacionales	Contiene los datos históricos anuales internacionales
datos_mensuales_espanya	Contiene los datos mensuales de los puertos españoles
datos_mensuales_espanya_xlsx	contiene los datos mensuales de los puertos españoles en formato xlsx.
Modelo_bd	Modelo definido mediante MySQL Workbench para almacenar los datos cargados
Notebooks	Libretas con la ejecución de todos los modelos para los 3 puertos analizados
spoon_project	proyecto de Pentaho Data Integration que permiten cargar los datos históricos anuales de España
TFM_java_workspace	proyecto en java que permite cargar los datos históricos mensuales de España.
TFM_javadoc	Documentación en formato javadoc del proyecto "TFM_java_workspace" que se encarga de la carga de los datos mensuales.

10.2. Business Intelligence (BI)

SOLUCIONES BI

Visualización y generación de Informes

La visualización de datos es el proceso de búsqueda, interpretación, contrastación y comparación de datos que permite un conocimiento en profundidad y detalle de éstos de tal forma que se transformen en información comprensible para el usuario. El modo más usual de presentar estas visualizaciones es mediante cuadros de mando o paneles de monitorización que muestran los KPIs de negocio más relevantes.

En este punto se listan y se describen algunos de los principales frameworks para la construcción de cuadros de mando y visualización de datos que gozan de mayor aceptación en la actualidad. A la hora de decantarse por una solución u otra, se debe tener en cuenta:

- La calidad de la solución y curva de aprendizaje.
- Soporte, comunidad y número de usuarios.
- Restricciones de licencia (aceptación de soluciones libres y de código abierto).
- Tecnología y facilidad de integración: orientado a la web, que sean gráficos atractivos, dinámicos y con interacción.

Librerías JavaScript “JS” para crear gráficos dinámicos

Las siguientes librerías son referentes a la hora de crear visualizaciones y objetos SVG (gráficos vectorizados escalables), para su utilización se requiere de conocimientos de programación. Estas librerías están pensadas para crear visualizaciones aisladas y no están pensadas para crear cuadros de mando. Entre las más conocidas y empleadas se encuentran:

- **RShiny** (RShiny, s.f.). Shiny está pensada para crear gráficos como salida de análisis hechos con el lenguaje de programación R. Resulta muy útil si se ha utilizado R para realizar el análisis.
- **D3.js** (D3.js, s.f.) es un librería JavaScript de código abierto para la visualización de datos. Permite enlazar datos arbitrarios a un modelo de documento de datos, y luego aplicar transformaciones basadas en datos al documento. Es compatible con grandes conjuntos de datos y comportamientos dinámicos de interacción y animación. D3 se ha convertido en un estándar de facto para la visualización de datos. También hay bibliotecas construidas sobre D3 (dimple, NVD3, xCharts, Rickshaw, C3.js, etc.). Como alternativa a D3, existen muchas librerías JS, pero no están tan implantadas, como son amcharts, echarts, highcharts,
- **Chart.js** (Chart.js, s.f.) es una pequeña biblioteca de código abierto que, aunque restringida en cuanto a disponibilidad de componentes se refiere (acotado número de opciones disponibles) es perfecta para proyectos cuyo alcance es acotado.

Además, de las librerías mencionadas anteriormente, existen herramientas pensadas para crear cuadros de mandos:

- **Freeboard** (Freeboard, s.f.) es un software open source gratuito de construcción de cuadros de mandos, no es una solución BI completa, pero es fácil de integrar con múltiples fuentes de datos. Por defecto ya tiene cargada la visualización de datos meteorológicos de Open Weather Map, y tiene preparados gráficos para alimentarse de datos JSON a partir de su URL. Está configurado para recuperar datos de plataformas IoT como Citrix, BuildingLink, Orion, PubNub, Xively. De cada fuente de datos se le especifican parámetros y tipos de datos, así como frecuencias de actualización. Resuelve el problema de CORS (recursos compartidos de dominio cruzado).

Las ventajas de Freeboard frente a las librerías JS son que:

- Permite crear cuadros de mandos de manera intuitiva, ordenados, diseño atractivo y altamente flexible.
- Muy fácil de utilizar, permitiendo crear cuadros de mando arrastrando componentes "Drag and drop" por lo que no se requieren conocimientos de programación.

Soluciones BI Genéricas

Esta sección se centrará en soluciones específicas para BI y para ello nos basaremos en el informe "Magic Quadrant for Analytics and Business Intelligence Platform" elaborado anualmente por la consultora Gartner [REF BI-1]) y cuya última versión es de febrero de 2020



Figura 38: Cuadrante Mágico de Plataformas de Inteligencia de Negocio y Analíticas

Según Gartner, la mayoría de las herramientas BI modernas han cambiado su enfoque centrado en capacidades de generación de informes, hacia un enfoque de análisis de autoservicio que permite a los usuarios la construcción de informes y análisis de manera intuitiva y ágil.

Un elemento común de las soluciones BI actuales es que se integran con decenas de fuentes de datos y su gran facilidad de uso, proporcionando asistentes avanzados que ayudan en todo el proceso de creación de los cuadros de mandos, desde la preparación de los datos a la exploración visual.

En cambio, las principales diferencias entre las soluciones BI se centran en las siguientes capacidades:

- Informes integrados que facilitan la modernización/unificación de los informes empresariales tradicionales.
- Analíticas avanzadas basadas en inteligencia artificial y aprendizaje automático son actualmente el elemento diferenciador principal de las herramientas BI

Además de los dos elementos diferenciadores principales vistos anteriormente, existen otras características diferenciadoras relevantes como son la seguridad, servicio en la nube, conectividad de datos, preparación de datos, modelos de datos complejos, generación de catálogos, visualizaciones avanzadas, consultas basadas en lenguaje natural, generación de infografías, informes avanzados.

Aunque todas las herramientas propuestas por Gartner son interesantes, en el presente trabajo, por acotar el alcance, vamos a centrarnos en las 4 herramientas consideradas líderes según Gartner.

- **Tableau**

Tableau (Tableau, s.f.) es una de las herramientas de Business Intelligence más potentes e innovadoras del mercado. La incógnita principal de este producto/empresa, viene relacionada con la dirección que va a tomar tras haber sido adquirida por Salesforce en 2019.

Tableau permite analizar grandes volúmenes de datos, visualizando de forma ágil y satisfaciendo las necesidades de BI moderno, como son el descubrimiento y la gobernanza de los datos. Además, incorpora una cuidada interfaz de usuario fácil e intuitiva. En los últimos años, Tableau ha incorporado nuevas capacidades analíticas basadas en aprendizaje automático y el uso de consultas basadas en lenguaje natural.

Tableau puede usarse tanto en on-premise como en la nube.

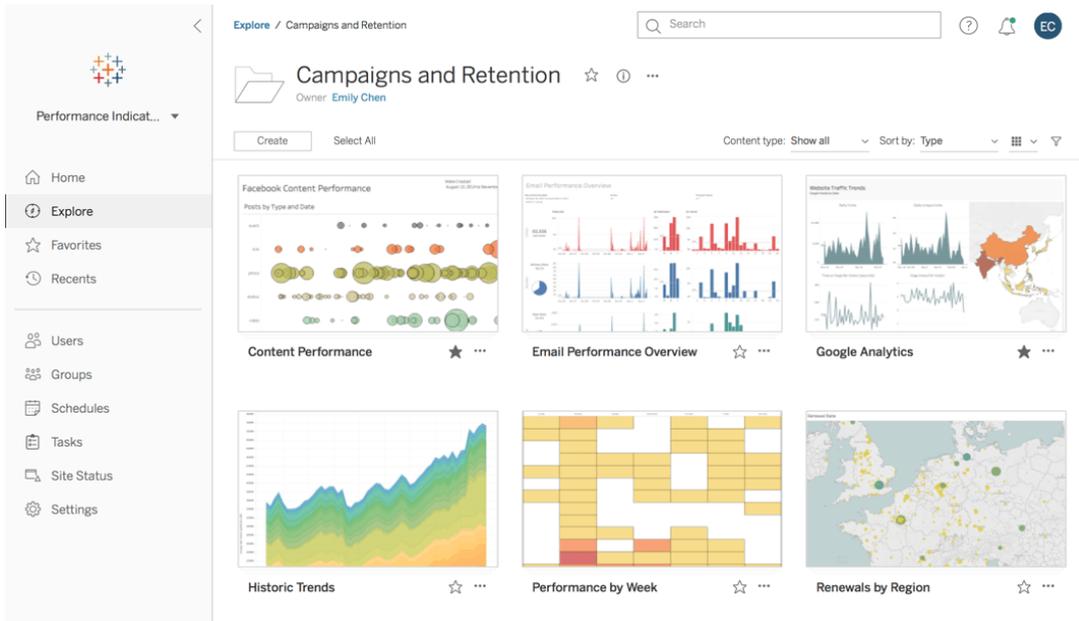


Figura 39: Ejemplo de Cuadro de Mandos con Tableau

- **Power BI (Microsoft)**

Power BI (PowerBI, s.f.) es la herramienta líder del mercado gracias a la visibilidad que le da Microsoft a la integración del producto en su suite Microsoft office y a su prometedor Roadmap.

Power BI ofrece una potente funcionalidad para la preparación y descubrimiento de datos, cuadros de mandos interactivos y analíticas avanzadas. Además, es compatible con Azure la plataforma Cloud de Microsoft. Power BI Pro está en constante evolución y ofrece actualizaciones y mejoras semanalmente. Actualmente su versión Power BI Pro es de las que ofrece funcionalidades más completas incluyendo funcionalidades analíticas innovadoras. Por contra, la versión on-premise no está tan cuidada y no incluye todas las funcionalidades que tiene la versión en la nube.

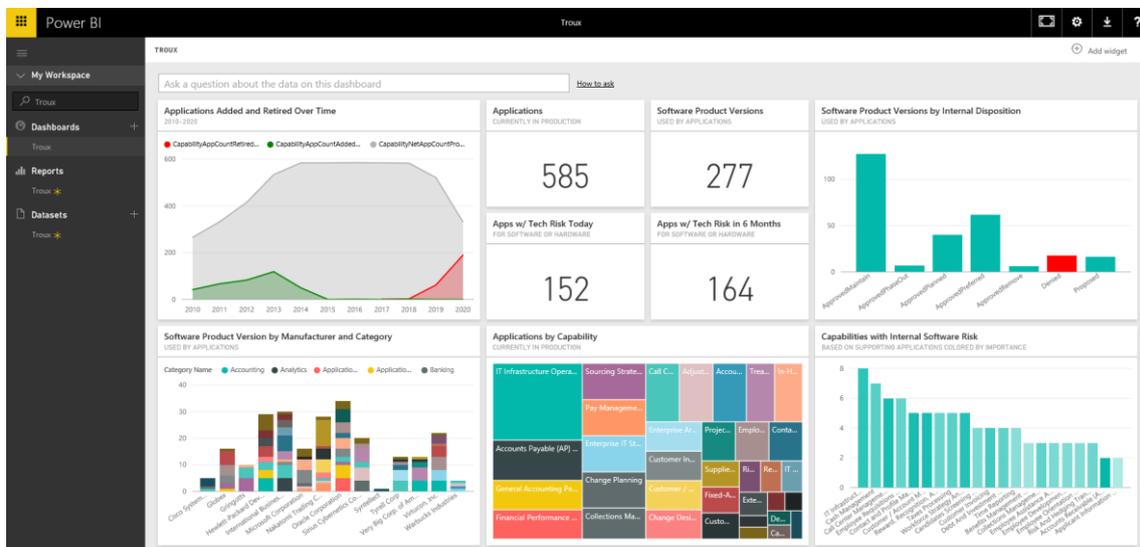


Figura 40: Ejemplo de Cuadro de Mandos con Power BI

- Qlik sense

Qlik (Qlik, s.f.) es otra de las empresas que históricamente viene apareciendo como líder del mercado y que está invirtiendo mucho en mejorar las capacidades analíticas de sus productos.

Qlik Sense es una plataforma completa de análisis de datos que establece estándares para una nueva generación de análisis. Con su motor asociativo de análisis único, inteligencia artificial sofisticada y una potente plataforma en la nube, puede brindar a todos en su organización la capacidad de tomar mejores decisiones cada día, creando una empresa verdaderamente basada en datos.

Qlik Sense se ofrece como producto on-premise y en la nube. La ventaja frente a Power BI, es que funciona en cualquier plataforma en la nube y no solo en Azure.

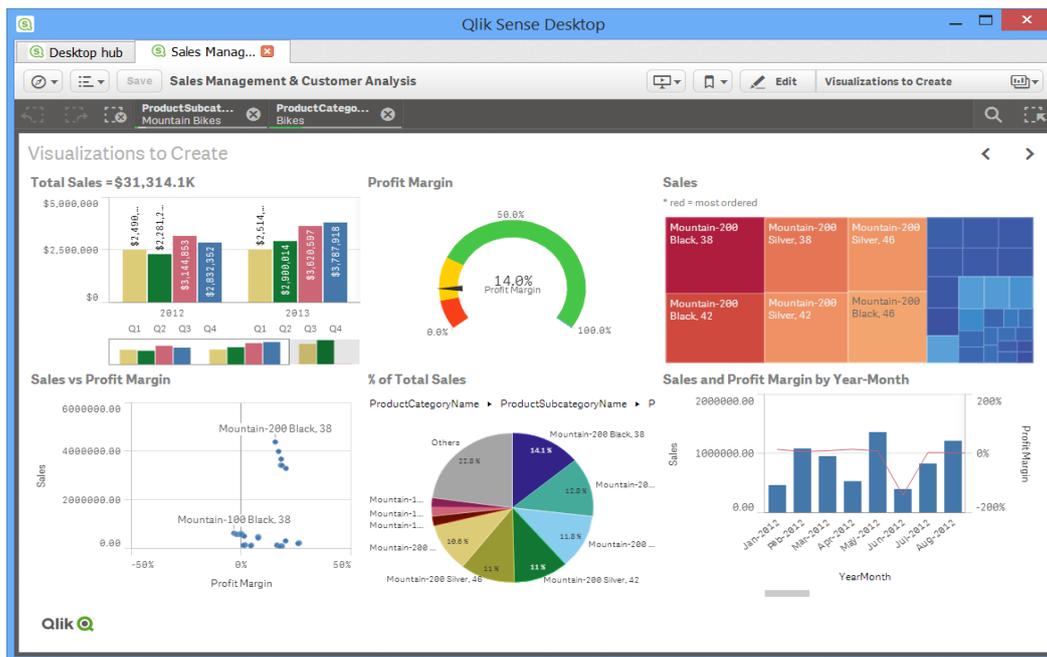


Figura 41: Ejemplo de Cuadro de Mandos con Qlik

- ThoughtSpot

ThoughtSpot (Thoughtspot, s.f.) no es una empresa muy grande en comparación a sus competidores y aunque haya obtenido rondas de financiación importantes en 2019 sigue careciendo de una importante presencia a nivel global.

ThoughtSpot es otra de las herramientas líderes y su ventaja competitiva diferencial es su innovadora funcionalidad de búsqueda basada en lenguaje natural y capacidad de procesar grandes cantidades de datos, sin olvidarnos de sus avanzadas capacidades analíticas.

Por contra, ThoughtSpot tiene carencias a la hora de la preparación de los datos y sus usuarios deben apoyarse en otras herramientas para este propósito.

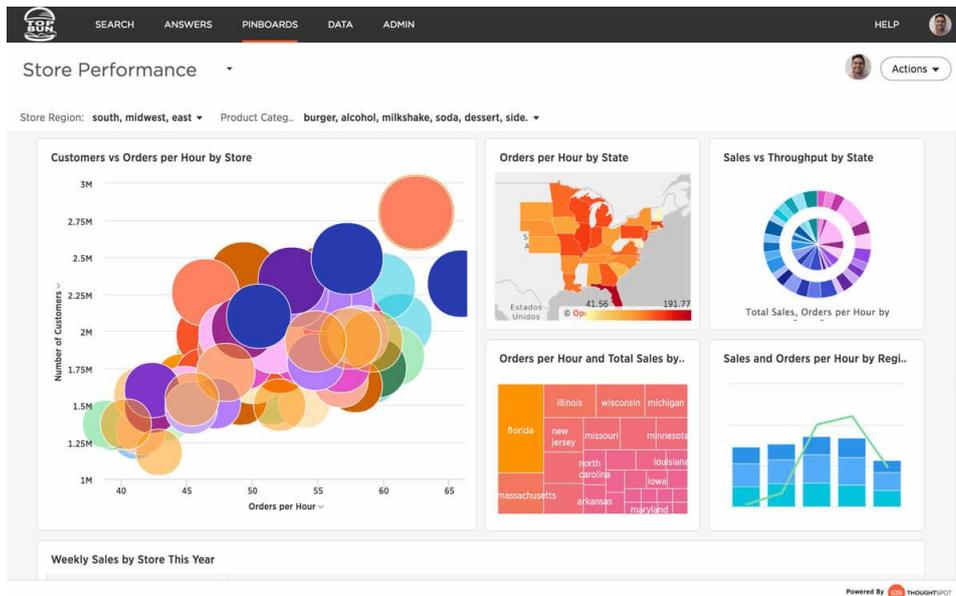


Figura 42: Ejemplo de Cuadro de Mandos con ThoughtSpot

Soluciones BI Especializadas

En el punto anterior se ha hablado de soluciones BI genéricas multipropósito, sin embargo, hay ocasiones en las que se demandan herramientas especializadas en alguna faceta concreta, que no está tan bien resuelta por estas herramientas genéricas. Por ejemplo, a la hora de realizar análisis avanzados relacionados con información geográfica o información en tiempo real, hay otras herramientas especializadas que deberían considerarse a la hora de la toma de decisiones.

A continuación, describiremos algunas de las herramientas especializadas:

- CARTO

CARTO (CARTO, s.f.) es una empresa especializada en proporcionar soluciones centradas en el análisis espacial (Información geográfica) y es una de las empresas líderes en este sector.

Carto proporciona visualizaciones y cuadros de mando para una completa herramienta de análisis de datos, se trata de una solución fácil de usar que aprovecha los datos abiertos y los filtros, los aumentos y los análisis para obtener información más profunda en el análisis de BI.

CARTO Builder es el componente que se utiliza para crear los cuadros de mando y visualizaciones, es una herramienta de análisis de "arrastrar y soltar" basada en la web que viene con una variedad de hojas de datos de plataformas públicas. Es diferente de sus versiones anteriores debido a dos características: widgets y funciones de predicción. Ninguno de ellos requiere el conocimiento de programación para ser utilizado.

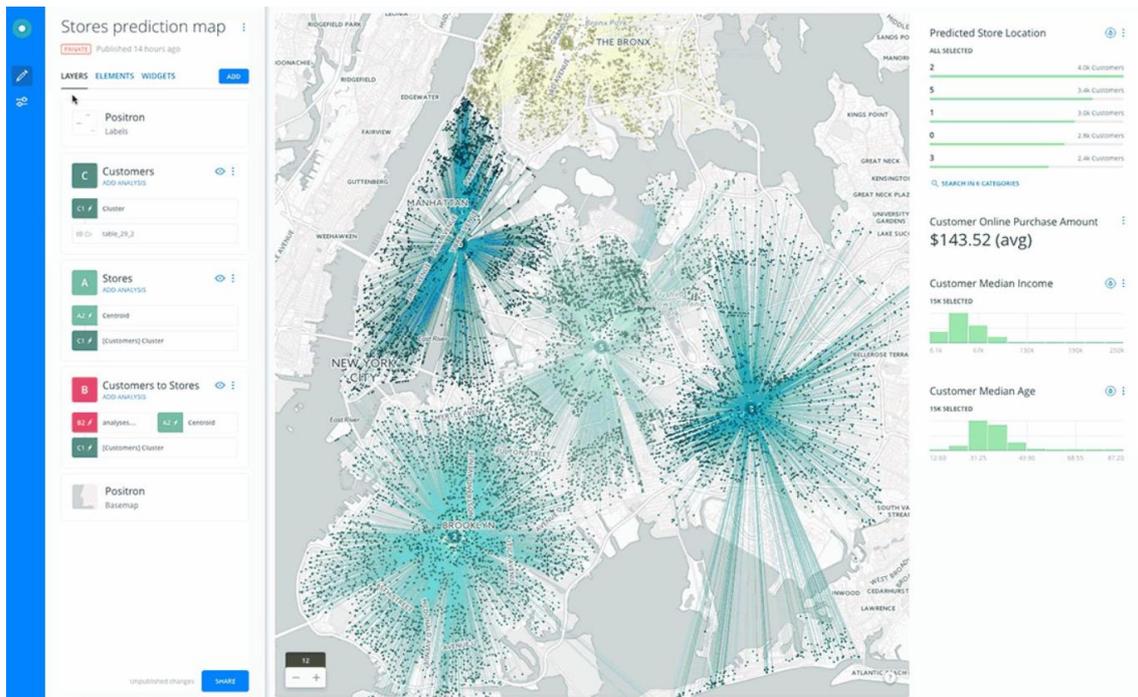


Figura 43: Ejemplo de Cuadro de mando con CARTO

CARTO ofrece herramientas de análisis espacial como el software gis tradicional: superposiciones, calcular clusters, cálculo de áreas de servicio por tiempo, proximidad, etc. Para usuarios avanzados con experiencia en programación, ofrece poderosas capacidades mediante consultas SQL.

- **Elastic**

Elastic (Elastic, s.f.) proporciona una suite completa de productos “ELK (ElasticSearch + Logstash* + Kibana)” para monitorizar grandes cantidades de información en tiempo real. Existe una versión gratuita sin limitaciones y que puede ser útil para la mayoría de los propósitos. La versión de pago incluye ciertas funcionalidades extra, relacionadas con la seguridad y AI que son requeridos en caso de querer abrir la solución fuera de la Intranet corporativa y querer tener predicciones.

ELK es una solución que inicialmente se pensó para la monitorización centralizada de servidores y aplicaciones en producción, pero que actualmente sirve para monitorizar cualquier tipo de información relacionada con series temporales. Funciona tanto el local como en la nube

Kibana (Dashboard) es una plataforma de análisis y visualización que permite explorar, visualizar y representar los datos contenidos en un clúster Elasticsearch utilizando gráficos, tablas y mapas. La creación de visualizaciones y cuadros de mandos no es tan sencilla como en otras herramientas y requiere de ciertos conocimientos, pero sus grandes posibilidades de configuración la dotan de una potencia que no disponen otras herramientas.

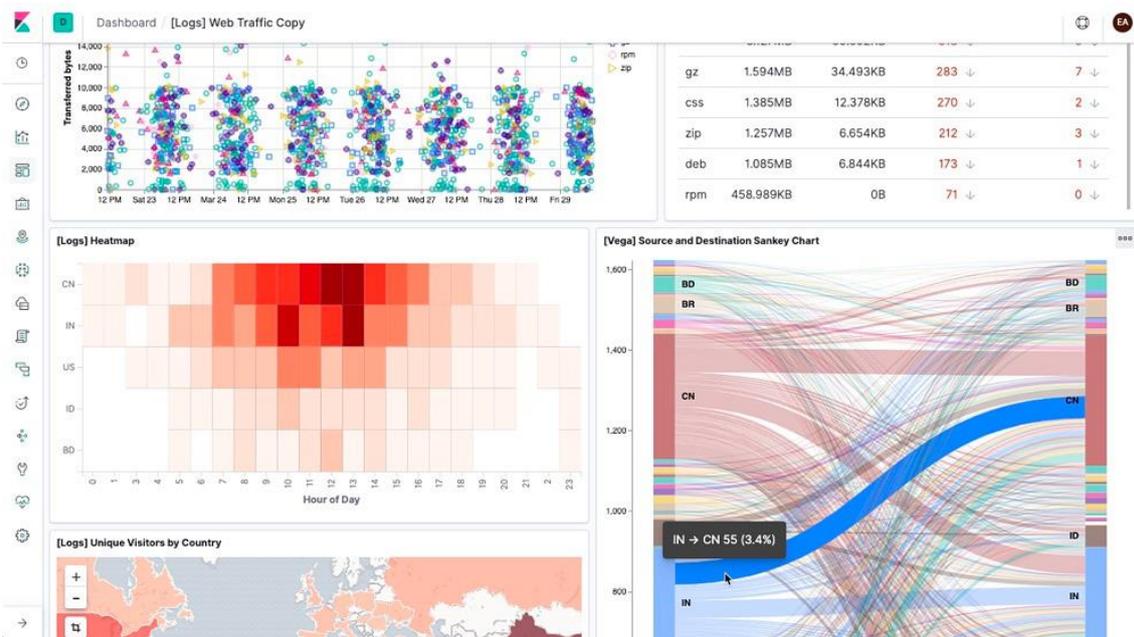


Figura 44: Ejemplo de Cuadro de mando con Kibana

Elasticsearch (Base de Datos no-SQL). Motor de búsqueda avanzado extremadamente rápido. Con Elasticsearch, puede buscar y filtrar todo tipo de datos a través de una simple API. La API es RESTful, por lo que no sólo puede utilizarla para el análisis de datos, sino también utilizarla en la producción para aplicaciones basadas en la Web. La ventaja competitiva de Elasticsearch es que está pensada para grandes volúmenes de datos y su tiempo de respuesta es muy pequeño aun haciendo consultas contra millones de datos.

Logstash (Adquisición y transformación de datos). Herramienta encargada de alimentar la base de datos con los datos a analizar que previamente se pueden filtrar y limpiar. El mayor inconveniente de esta herramienta es que requiere de conocimientos de programación y no dispone de una interfaz visual como disponen algunas herramientas ETL.

- **Grafana**

Grafana (Grafana, s.f.) es la alternativa a Elastic, tiene las ventaja de que funciona con cualquier origen de datos, pero por contra solo se centra en la visualización, sin incluir mecanismos de almacenamiento ni de adquisición de datos.

Grafana es una herramienta de código abierto para visualizar datos de serie temporales. A partir de una serie de datos recolectados obtendremos un panorama gráfico de la situación de un sistema u organización.

Admite más de 30 fuentes de código abierto, así como bases de datos/fuentes de datos comerciales que incluyen MySQL, PostgreSQL, Graphite, Elasticsearch, OpenTSDB, Prometheus e InfluxDB.

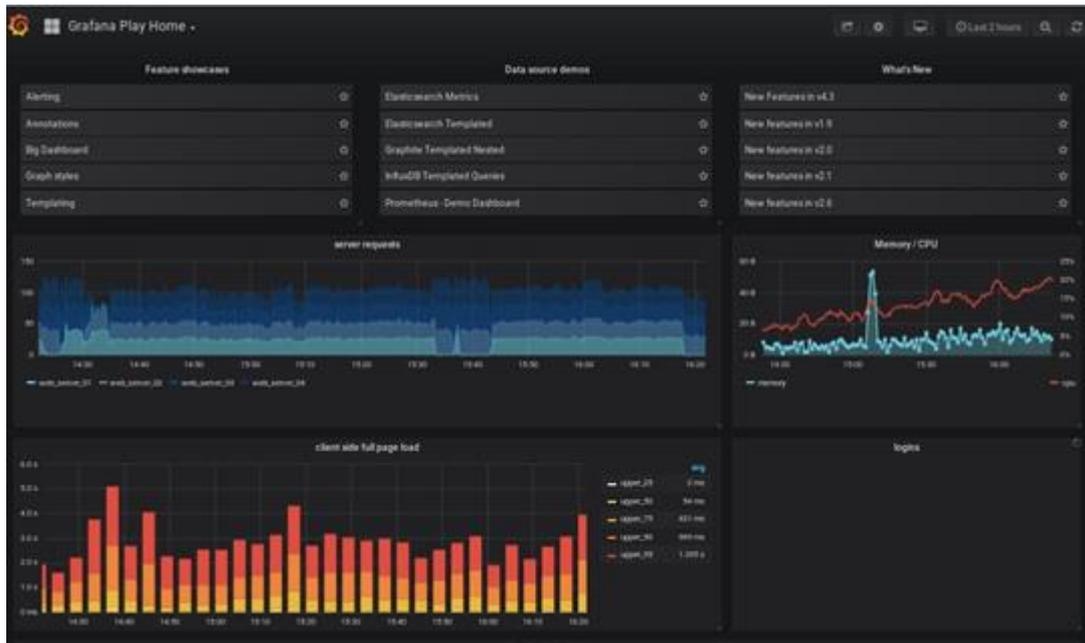


Figura 45: Ejemplo de Cuadro de Mando con Grafana

SOLUCIONES BI PARA EL SECTOR PORTUARIO

En este punto se revisarán las diferentes soluciones de BI adaptadas al sector portuario.

El negocio principal de las autoridades portuarias es velar por la sostenibilidad y el buen funcionamiento de la operativa portuaria que se lleva a cabo en sus instalaciones que concesionan a largo plazo para diferentes actividades. La autoridad portuaria debe controlar todos los accesos al puerto, garantizar la seguridad, proveer de servicios auxiliares para que se lleven a cabo las operaciones portuarias, gestionar las instalaciones y asegurar la viabilidad económica del puerto. Para lograr su fin, las autoridades suelen disponer de herramientas de BI de propósito general para controlar el buen funcionamiento del puerto y asegurar la viabilidad económica del mismo.

En cambio, el negocio de las terminales es la importación y exportación de mercancías. Estas funcionan como grandes centros logísticos que mueven grandes cantidades de mercancías por mar/río, su principal objetivo es maximizar su productividad y mejorar su rendimiento. En este caso, los objetivos y el control requerido se asemejan más al de la Industria 4.0, cuyo objetivo es monitorizar la actividad de la terminal en tiempo real por medio de sistemas SCADA y series temporales, para minimizar los cuellos de botella y aumentar su productividad.

El negocio de la digitalización de las autoridades portuarias está copado por empresas nacionales, ya que al ser entes públicos y emplear licitaciones, el mercado de proveedores es prácticamente nacional. En cambio, las terminales, al estar gestionadas por conglomerados de empresas privadas, los proveedores son a nivel internacional y es normal que una empresa que gestione varias terminales por todo el mundo implante una misma solución a nivel global.

En el presente trabajo nos vamos a focalizar en soluciones de BI para autoridades portuarias españolas. Aun así, no es fácil obtener una visión clara de los competidores ya que no hay mucha información pública al respecto. En el sector portuario, cuando un puerto tiene una necesidad, al ser entidades públicas, se presenta una licitación y cualquier empresa que cumpla los requisitos y esté interesada, se puede presentar a la misma. Entre 2018-2020 solo ha habido una licitación exclusiva sobre “Business Intelligence” en el sector portuario. A dicha licitación se presentaron 7 empresas y solamente Prodevelop y Alfatec pasaron el corte de criterios cualitativos, esto refuerza la idea de que, para ser competitivo en este sector, no sólo se necesita ser experto en BI, sino que se debe de conocer muy bien el sector. De este modo se pueden personalizar las diferentes herramientas de BI a las necesidades portuarias, aportando un valor diferencial.

A continuación, se describen las empresas más relevantes que ofrecen servicios de BI en puertos españoles.

- **Alfatec Sistemas**

Alfatec (Alfatec, s.f.) es una empresa dedicada a la consultoría, desarrollo y mantenimiento de Sistemas de información en varios sectores (sanidad, industria, puertos y automoción). Alfatec tiene a nivel global más de 100 clientes que usan sus soluciones de BI, con dichas soluciones permiten descubrir el valor de los datos gracias a cuadros de mando interactivos y KPIs que recogen toda la información relevante y ayudan en la toma de decisiones. Las soluciones BI se construyen usando Qlik (Qlikview y QlikSense) y Microsoft Power BI.

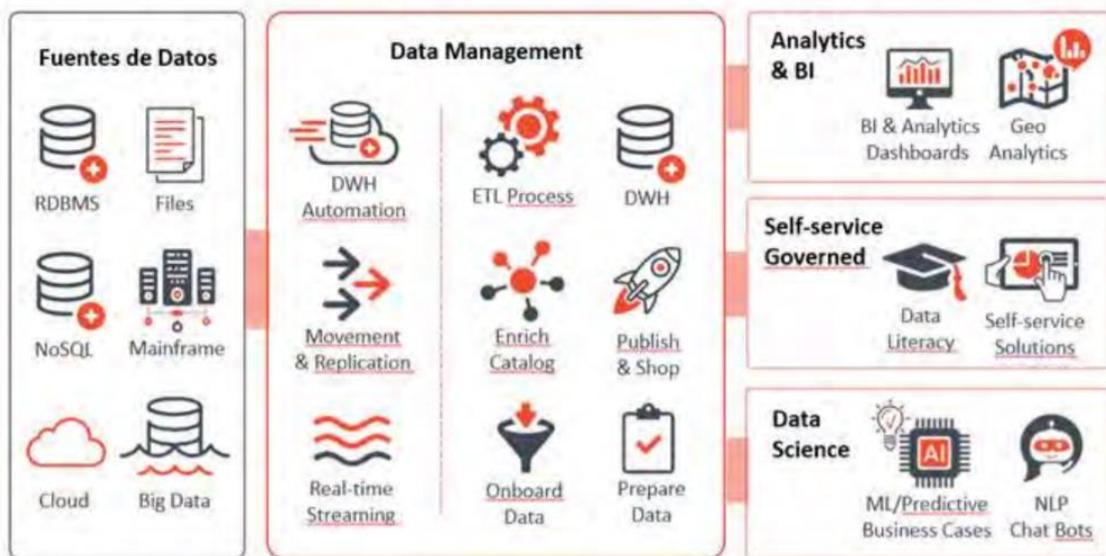


Figura 46: Esquema Funcionalidades Alfatec

En cuanto al sector portuario, Alfatec tiene más de 10 años de experiencia y su solución AtlantisPort permite una gestión integral de autoridades portuarias. Su solución de BI para puertos cuenta con cuadros de mandos de Qlikview y PowerBI integrados en la funcionalidad de NAV (Microsoft Dynamics 265). Esto permite cubrir las necesidades que surgen en el análisis y toma de decisión de la organización para perfiles que pueden no necesitar hacer uso de NAV necesariamente.

- **Edisa**

Edisa (EDISA, s.f.) es una multinacional española fabricante de software de gestión empresarial ERP desde hace 35 años.

Edisa Business Intelligence, es una solución BI genérica que se puede implantar en diferentes sectores. Edisa cuenta con una experiencia contrastada de más de 15 años en soluciones analíticas de negocio y más de 1200 proyectos de BI, es una compañía experta en cuadros de mando estratégico y operativo para todas las áreas críticas de la organización. Sus desarrollos están basados soluciones Qlik

Edisa BI facilita una mayor eficiencia y productividad son los retos actuales a los que se enfrentan las autoridades portuarias, encargadas de la gestión de los puertos. Edisa BI cuenta con cuadros de mando para controlar los indicadores relevantes en las siguientes áreas: Comercialización, Gestión de operaciones, Procesos de carga y descarga, Almacenaje y transporte, Gestión de obras e infraestructuras, Control financiero y de la deuda, Aduanas y policía portuaria. Esta solución se ha usado en el Puerto de Gijón, A Coruña o Baleares.



Figura 47: Ejemplo de Cuadro de mando con Edisa

- **Proyecto BrainPort Analytics (proyecto de investigación aplicada)**

El objetivo del proyecto BrainPort Analytics (Brainport, s.f.) desarrollado por Ayesa (AYESA, s.f.) es construir una solución para monitorizar, a partir de una red de sensores eficiente, los procesos operativos (físicos y administrativos) que tienen lugar en los puertos gestionados por la Autoridad Portuaria Bahía de Algeciras e implementar un sistema analítico basado en herramientas BigData y Business Intelligence (BI) para procesar grandes volúmenes de datos.

BrainPort Analytics es una plataforma Smart City aplicada a puertos con sensorización en tiempo real, Gestión integral de la Operación Paso del Estrecho, KPIs, Gestión de tráfico.

Los cuadros de mando para el análisis y la toma de decisiones se han desarrollado usando Qlikview.

A diferencia de otras soluciones, BrainPort integra información proveniente de sensores en tiempo real y no está tan focalizada en aspectos de gestión o económicos.

- **Puerto de Valencia**

El Puerto de Valencia (Puerto de Valencia, s.f.) es el puerto que mueve más contenedores a nivel nacional y es de los más importantes de Europa. Dado su tamaño, el puerto tiene su propio departamento informático y además cuenta con la Fundación Valencia Port que es un centro de Investigación Aplicada, Innovación y Formación, al servicio del clúster logístico portuario.

El puerto de Valencia usa las soluciones de Qlik para la optimización de su inteligencia de negocio y la generación de informes estratégicos.

- **Prodevelop**

Prodevelop (Prodevelop, s.f.) es una empresa española, altamente especializada en la aplicación de las Tecnologías de la Información en el sector portuario, tanto a nivel nacional como internacional, donde lleva más de 25 años aportando valor. Actualmente 27 de las 28 autoridades portuarias españolas usan sus productos portuarios. Prodevelop proporciona una serie de soluciones portuarias bajo la marca comercial Posidonia que incluye productos para las diferentes necesidades de los puertos nacionales e internacionales. Posidonia Business Intelligence es su solución de Business Intelligence en el ámbito portuario. Actualmente las soluciones de BI de Prodevelop están siendo usadas por 17 autoridades portuarias españolas.

Posidonia BI, utiliza tecnología Qlik para la elaboración de cuadros de mandos dirigidos a la dirección y otros perfiles de la Autoridad Portuaria que necesitan monitorizar y conocer de manera sencilla la situación operativa del puerto en tiempo real, sin necesidad de formación ni de conocer los diferentes sistemas de información que deben alimentar el cuadro de mando. Posidonia BI se integra de forma nativa con el resto de los productos Posidonia y también tiene la posibilidad de integrarse con otros productos existentes en el ámbito portuario.

A partir del cuadro de mando se accede a información de segundo nivel con indicadores sobre un área de gestión: Escalas, Mercancía, Pasaje, Dominio Público, Pesca, Embarcaciones Deportivas, Ocupación de superficie, Servicios Diversos y Facturación.

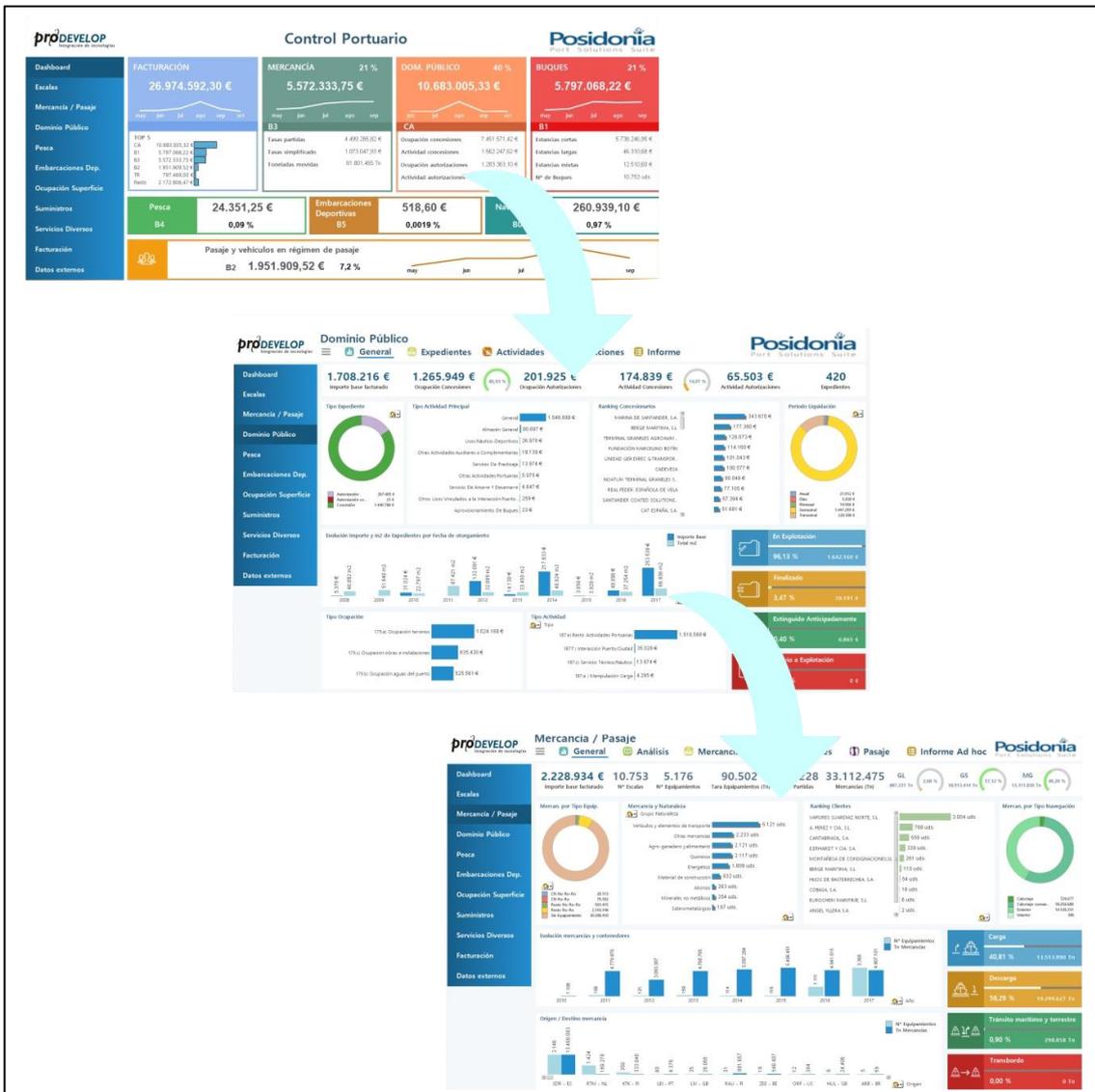


Figura 48: Ejemplo de Cuadro de mandos a diferentes niveles con INTEGRA2

10.3. Anexo Carga de datos

10.3.1. Scripts de creación de Base de datos

Script de creación de base de datos

```
-- MySQL Script generated by MySQL Workbench
-- Sun Dec 27 11:36:11 2020
-- Model: New Model    Version: 1.0
-- MySQL Workbench Forward Engineering

SET @OLD_UNIQUE_CHECKS=@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@SQL_MODE,
SQL_MODE='ONLY_FULL_GROUP_BY,STRICT_TRANS_TABLES,NO_ZERO_IN_DATE,NO_ZERO_DATE,ERROR_FO
R_DIVISION_BY_ZERO,NO_ENGINE_SUBSTITUTION';

-----

-- Schema tfm_modelo_v3
-----

-----

-- Schema tfm_modelo_v3
-----

CREATE SCHEMA IF NOT EXISTS `tfm_modelo_v3` DEFAULT CHARACTER SET utf8 ;
USE `tfm_modelo_v3` ;

-----

-- Table `tfm_modelo_v3`.`spanish_historic_annual`
-----

CREATE TABLE IF NOT EXISTS `tfm_modelo_v3`.`spanish_historic_annual` (
  `anyo` INT NOT NULL,
  `puerto` VARCHAR(45) NOT NULL,
  `codap` INT NULL,
  `productospetroliferos` INT NULL DEFAULT NULL COMMENT 'resumen mercancias según
presentación',
  `otrosliquidos` INT NULL COMMENT 'resumen mercancias según presentación',
  `granelesliquidos` INT NULL DEFAULT NULL COMMENT 'en la hora resumen mercancias por
presentación se llama :\nTOTAL G. LIQUIDOS',
  `granelessolidos` INT NULL DEFAULT NULL,
```

```

`mercanciageneralconvencional` INT NULL DEFAULT NULL,
`mercanciageneralcontenedores` INT NULL DEFAULT NULL,
`mercanciageneral` INT NULL DEFAULT NULL,
`pescafresca` INT NULL DEFAULT NULL,
`avituallamientoprodpetrolíferos` INT NULL,
`avituallamientoresto` INT NULL,
`avituallamiento` INT NULL DEFAULT NULL,
`totalotrasmercancias` INT NULL,
`traficolocal` INT NULL DEFAULT NULL,
`totaltrafico` INT NULL DEFAULT NULL,
`contenedores` INT NULL,
`teus` INT NULL DEFAULT NULL,
`buquestotales` INT NULL DEFAULT NULL COMMENT 'buques en la Excel',
`buquesprotacontenedores` INT NULL DEFAULT NULL COMMENT 'buques según tipo y
nacionalidad',
`buquesgt` INT NULL,
`pasajeroscruzero` INT NULL,
`pasajerosentransito` INT NULL,
`pasajerosembarcadosdesembarcados` INT NULL,
PRIMARY KEY (`anyo`, `puerto`))
ENGINE = InnoDB;

-----
-- Table `tfm_modelo_v3`.`spanish_historic_contenedores`
-----

CREATE TABLE IF NOT EXISTS `tfm_modelo_v3`.`spanish_historic_contenedores` (
  `anyo` INT NOT NULL,
  `puerto` VARCHAR(45) NOT NULL,
  `codap` INT NULL,
  `contenedorcabotajecarga` INT NULL,
  `contenedorcabotajedescarga` INT NULL,
  `contenedorcabotajetofal` INT NULL,
  `contenedorexteriorcarga` INT NULL,
  `contenedorexteriordescarga` INT NULL,
  `contenedorexteriortotal` INT NULL,
  `contenedortotalcarga` INT NULL,
  `contenedortotaldescarga` INT NULL,
  `contenedortotal` INT NULL,
  `teuscabotajecarga` INT NULL,

```

```

`teuscabotajedescarga` INT NULL,
`teuscabotajetotal` INT NULL,
`teusexteriorcarga` INT NULL,
`teusexteriordescarga` INT NULL,
`teusexteriortotal` INT NULL,
`teustotalcarga` INT NULL,
`teustotaldescarga` INT NULL,
`teustotal` INT NULL,
`exteriortoneladas` INT NULL,
`cabotajetoneladas` INT NULL,
`totaltoneladas` INT NULL,
PRIMARY KEY (`anyo`, `puerto`))
ENGINE = InnoDB;

-----
-- Table `tfm_modelo_v3`.`spanish_month_contenedores`
-----

CREATE TABLE IF NOT EXISTS `tfm_modelo_v3`.`spanish_month_contenedores` (
  `mes` INT NOT NULL,
  `anyo` INT NOT NULL,
  `puerto` VARCHAR(45) NOT NULL,
  `codap` INT NULL,
  `teustotal` INT NULL,
  `teustransito` INT NULL,
  `teuscabotaje` INT NULL,
  `teusexterior` INT NULL,
  `teuscabotajeyexterior` INT NULL,
  `totaltoneladascontenedores` INT NULL,
  `transitotoneladascontenedores` INT NULL,
  `totaltrafico` INT NULL COMMENT '(Tn)',
  `totalpresentacion` INT NULL COMMENT '(Tn)',
  `granelesliquidos` INT NULL COMMENT '(Tn)',
  `granelessolidos` INT NULL COMMENT '(Tn)',
  `mercanciageneral` INT NULL COMMENT '(Tn)',
  `pesca` INT NULL COMMENT '(Tn)',
  `avituellamiento` INT NULL COMMENT '(Tn)',
  `avituellamientocombustibles` INT NULL COMMENT '(Tn)',
  `traficointerior` INT NULL COMMENT '(Tn)',
  `mercanicatransito` INT NULL COMMENT '(Tn)',

```

```

`traficororo` INT NULL COMMENT '(Tn)',
`traficorororemolques` INT NULL COMMENT 'unidades',
`pasajerototal` INT NULL,
`pasajerocrucero` INT NULL,
`automovilpasaje` INT NULL,
`automovilregimenmercancia` INT NULL,
`buques` INT NULL,
`buquesgt` INT NULL COMMENT 'unidades de arqueo bruto',
`cruceros` INT NULL,
PRIMARY KEY (`mes`, `anyo`, `puerto`))
ENGINE = InnoDB;

-----
-- Table `tfm_modelo_v3`.`international_data`
-----

CREATE TABLE IF NOT EXISTS `tfm_modelo_v3`.`international_data` (
  `anyo` INT NOT NULL,
  `gidmundial` INT NULL,
  `teu_pm_valencia` INT NULL,
  `teu_pm_barcelona` INT NULL,
  `international_datacol` VARCHAR(45) NULL,
  `teu_pm_algeciras` INT NULL COMMENT 'pm puerto mediterraneo\n',
  `teu_pm_malta` INT NULL,
  `teu_pm_piraeus` INT NULL,
  `teu_pm_gioatauro` INT NULL COMMENT 'puerto italia\n',
  `teu_pm_geona` INT NULL,
  `teu_pm_tanger` INT NULL COMMENT 'puerto marruecos\n',
  `teu_pm_portsaid` INT NULL COMMENT 'puerto egipto\n',
  `teu_r_world` INT NULL COMMENT 'world\n',
  `teu_r_eas` INT NULL COMMENT 'East asia & pacific',
  `teu_c_chn` INT NULL COMMENT 'country china',
  `teu_r_euu` INT NULL COMMENT 'european union',
  `teu_r_mea` INT NULL COMMENT 'Middle East & North Africa',
  `teu_r_nac` INT NULL COMMENT 'north america',
  `teu_c_usa` INT NULL COMMENT 'EEUU',
  `teu_r_arb` INT NULL COMMENT 'arab world',
  `teu_r_lcn` INT NULL COMMENT 'latin america y caribe',
  `teu_c_sgp` INT NULL COMMENT 'ingapore',
  `teu_c_deu` INT NULL COMMENT 'gemany',

```

```

`teu_c_are` INT NULL COMMENT 'united ara emirated\n',
`teu_c_esp` INT NULL COMMENT 'españa',
`teu_c_nld` INT NULL COMMENT 'netherlands',
`teu_c_bel` INT NULL COMMENT 'belgium\n',
`teu_c_gbr` INT NULL COMMENT 'United Kingdom',
`teu_c_ita` INT NULL COMMENT 'italy',
`teu_c_bra` INT NULL COMMENT 'brazil\n',
`teu_c_sau` INT NULL COMMENT 'saudi arabia',
`teu_c_mex` INT NULL COMMENT 'Mexico',
`teu_c_mar` INT NULL COMMENT 'morocco\n',
`teu_c_dza` INT NULL COMMENT 'Algeria',
`mercancia_to` INT NULL COMMENT 'total',
`mercancia_ag` INT NULL COMMENT 'agricultural products',
`mercancia_agfo` INT NULL COMMENT 'food',
`mercancia_mi` INT NULL COMMENT 'fuels and mining',
`mercancia_mifu` INT NULL COMMENT 'fuels',
`mercancia_ma` INT NULL COMMENT 'manufactures',
`mercancia_mais` INT NULL COMMENT 'iron and steel',
`mercancia_mach` INT NULL COMMENT 'chemicals',
`mercancia_machph` INT NULL COMMENT 'pharma',
`mercancia_mamt` INT NULL,
`mercancia_mamtof` INT NULL,
`mercancia_mamtotep` INT NULL,
`mercancia_mamtottl` INT NULL,
`mercancia_mamtotic` INT NULL,
`mercancia_mamtte` INT NULL COMMENT 'transport equipment\n',
`mercancia_mamtau` INT NULL COMMENT 'automotive products',
`mercancia_mate` INT NULL COMMENT 'textiles',
`mercancia_macl` INT NULL COMMENT 'clothing',
PRIMARY KEY (`anyo`))
ENGINE = InnoDB;

SET SQL_MODE=@OLD_SQL_MODE;
SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS;
SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS;

```

10.4. Anexo Detalle Análisis de Series Temporales

En este anexo se muestran los resultados de la aplicación de los 5 métodos predictivos (ARIMA, SARIMA, ARIMAX, RNN, PROHPET) a los 3 puertos analizados (Valencia, Barcelona y Algeciras)

10.4.1. Resultados aplicación ARIMA datos anuales Puerto Valencia

- Parámetros seleccionados: ARIMA(2, 2, 1) MSE=1188.866
- Gráfica comparativa de datos reales y estimados y la predicción para los próximos años

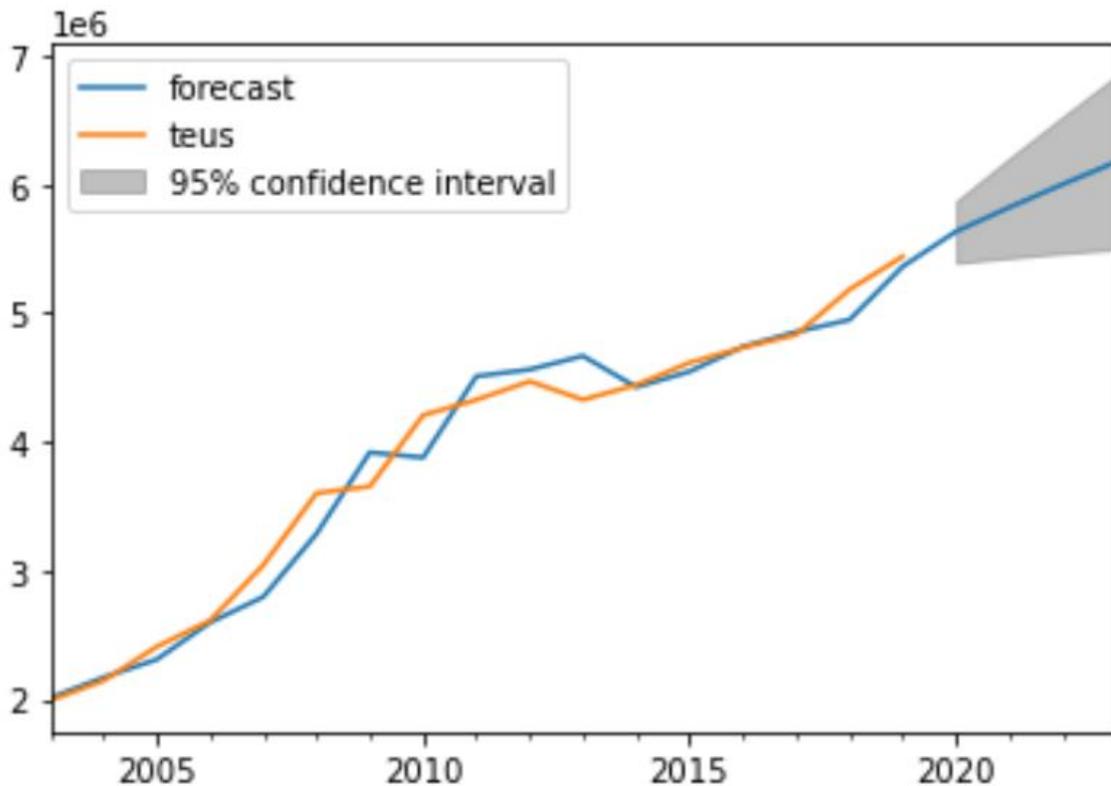


Figura 49: Gráfica ARIMA Anual Predicción - Puerto de Valencia

- Evaluación de la predicción. usando un 80% de datos para entrenamiento y un 20% para la validación

Año	Observado (Real)	Predicción
2010	4206937	3958023
2011	4327371	4723643

2012	4469754	4545867
2013	4327838	4965581
2014	4441949	4404216
2015	4615196	4416448
2016	4732136	4765173
2017	4832156	4891204
2018	5182665	4960328
2019	5439827	5356852

Tabla 17: ARIMA Valores Anuales Observados & Predicción Valencia

- Total error: 270799
- Comparativa Predicciones y observaciones en los datos de validación

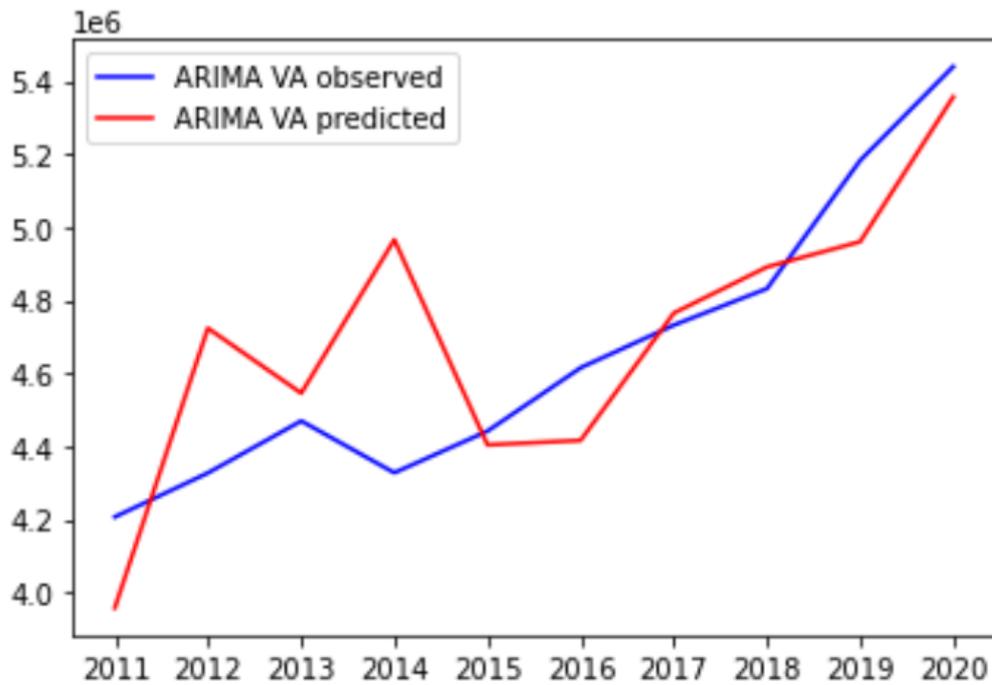


Figura 50: Gráfica ARIMA Anual Comparativa - Puerto de Valencia

10.4.2. Resultados aplicación ARIMA datos mensuales Puerto Valencia

- Parámetros seleccionados: ARIMA(4, 2, 2) MSE=3864.540
- Gráfica comparativa de datos reales y estimados y la predicción para los próximos meses

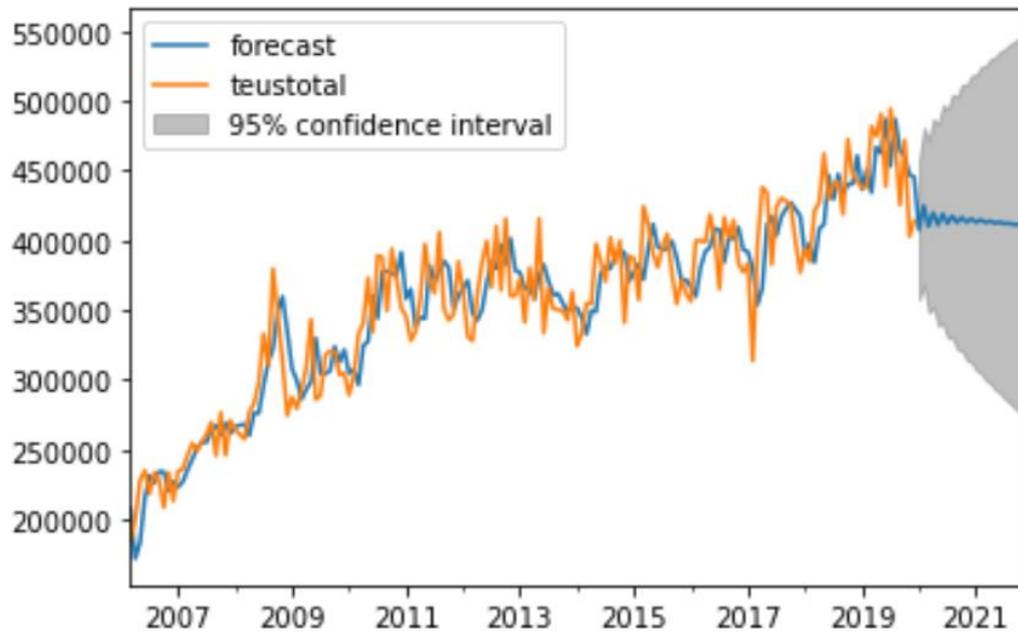


Figura 51: Gráfica ARIMA Mensual Predicción - Puerto de Valencia

- Evaluación de la predicción. usando un 90% de datos para entrenamiento y un 10% para la validación

Mes	Observado (Real)	Predicción
2018-08	442335	435944
2018-09	418714	443089
2018-10	472232	434055
2018-11	447903	452965
2018-12	442029	454269
2019-01	437958	447782
2019-02	438024	443110
2019-03	481607	444265

2019-04	474644	464418
2019-05	490587	471245
2019-06	438593	478881
2019-07	494262	460063
2019-08	470218	476704
2019-09	425314	481776
2019-10	471730	455734
2019-11	402816	460364
2019-12	414074	443574

Tabla 18: ARIMA Valores Mensuales Observados & Predicción Valencia

- Total error : 29430
- Comparativa Predicciones (rojo) y observaciones en los datos de validación

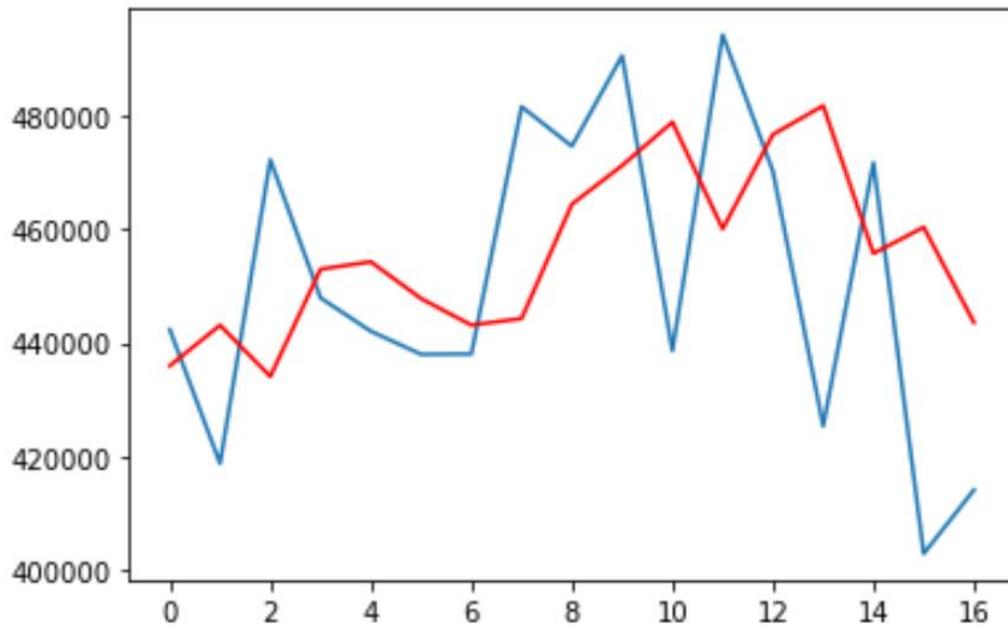


Figura 52: Gráfica ARIMA Mensual Comparativa - Puerto de Valencia

10.4.3. Resultados aplicación ARIMA datos Anuales Puerto Barcelona

- Parámetros seleccionados: ARIMA(0, 2, 2) MSE=1228.041
- Gráfica comparativa de datos reales y estimados y la predicción para los próximos años

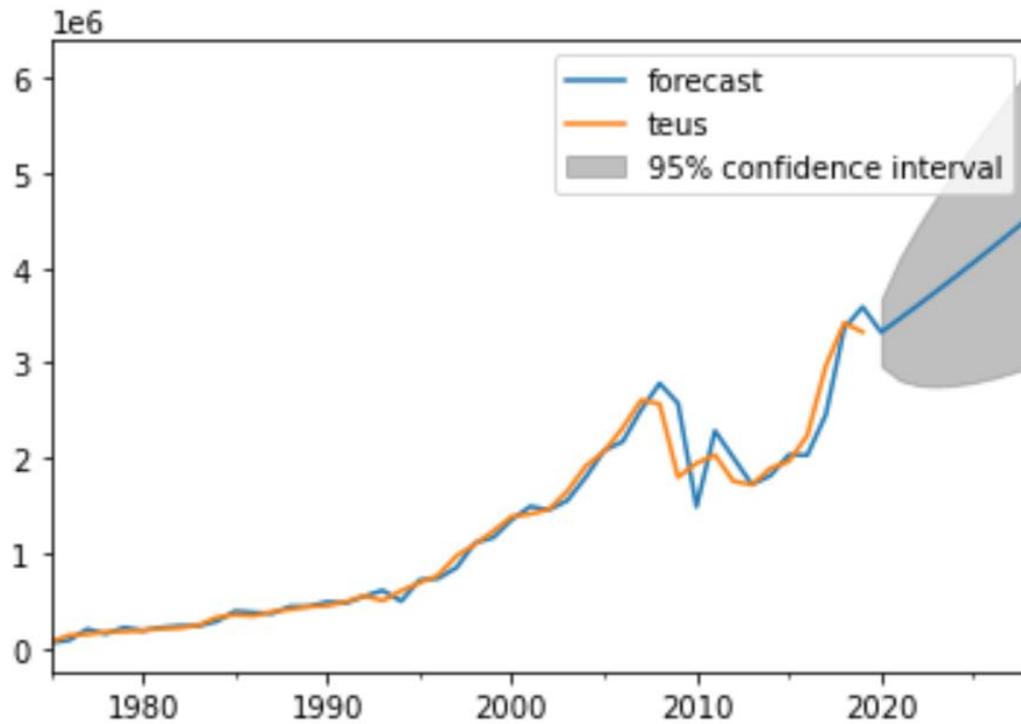


Figura 53: Gráfica ARIMA Anual Predicción - Puerto de Barcelona

- Evaluación de la predicción. usando un 80% de datos para entrenamiento y un 20% para la validación

Año	Observado (Real)	Predicción
2010	1945734	850979
2011	2033549	2397596
2012	1756429	1647559
2013	1720384	1576922
2014	1893300	1959468
2015	1965241	1831525
2016	2236961	1915784

2017	2968757	2401541
2018	3422978	3291828
2019	3324651	3617691

Tabla 19: ARIMA Valores Anuales Observados & Predicción Barcelona

- Total Error: 478966
- Comparativa Predicciones y observaciones en los datos de validación.

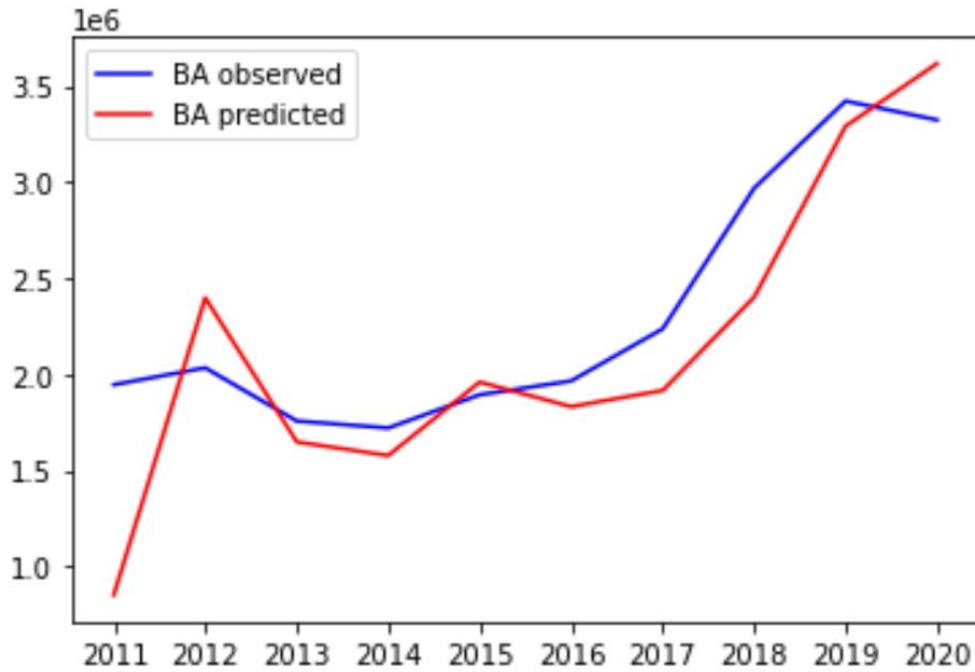


Figura 54: Gráfica ARIMA Anual Comparativa - Puerto de Barcelona

10.4.4. Resultados aplicación ARIMA datos mensuales Puerto Barcelona

- Parámetros seleccionados: ARIMA(3, 1, 2) MSE=3718.350
- Gráfica comparativa de datos reales y estimados y la predicción para los próximos meses

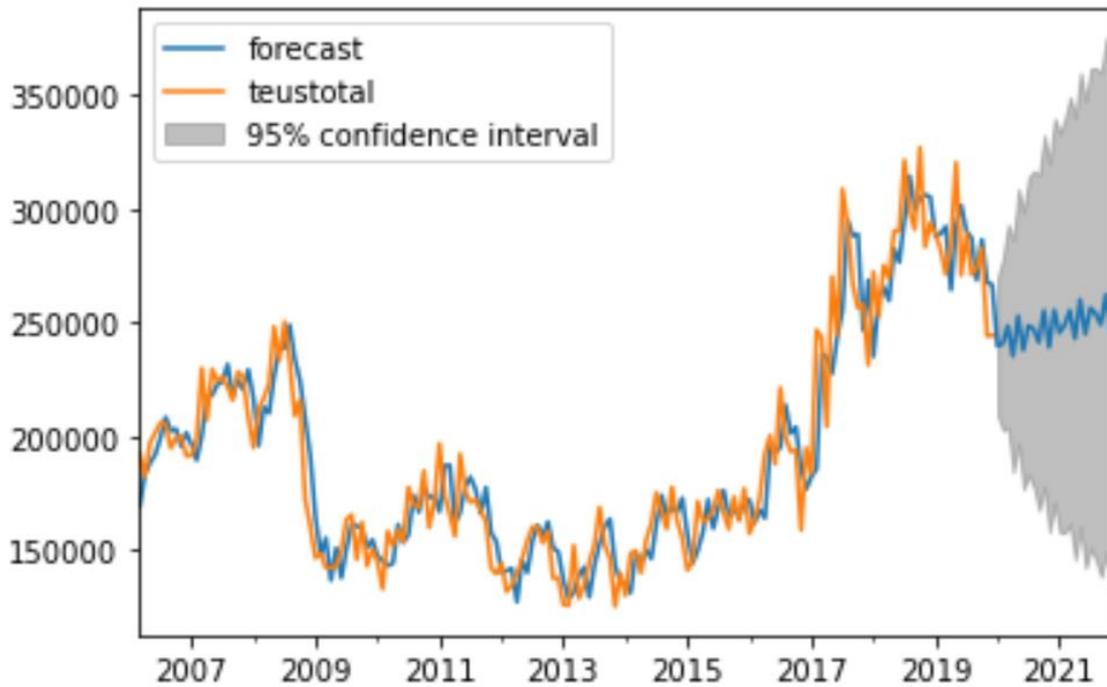


Figura 55: Gráfica ARIMA Mensual Predicción - Puerto de Barcelona

- Evaluación de la predicción. usando un 90% de datos para entrenamiento y un 10% para la validación

Mes	Observado (Real)	Predicción
2018-08	299089	314377
2018-09	291004	302809
2018-10	326985	302279
2018-11	283262	307499
2018-12	293923	306292
2019-01	289017	288157
2019-02	282162	288765
2019-03	271205	293132
2019-04	284893	264797

2019-05	320322	291985
2019-06	270787	302886
2019-07	289426	290972
2019-08	270965	287866
2019-09	274172	269720
2019-10	282753	286166
2019-11	244190	269086
2019-12	244306	268928

Tabla 20: ARIMA Valores Mensuales Observados & Predicción Barcelona

- Total Error : 18836
- Comparativa predicciones y observaciones

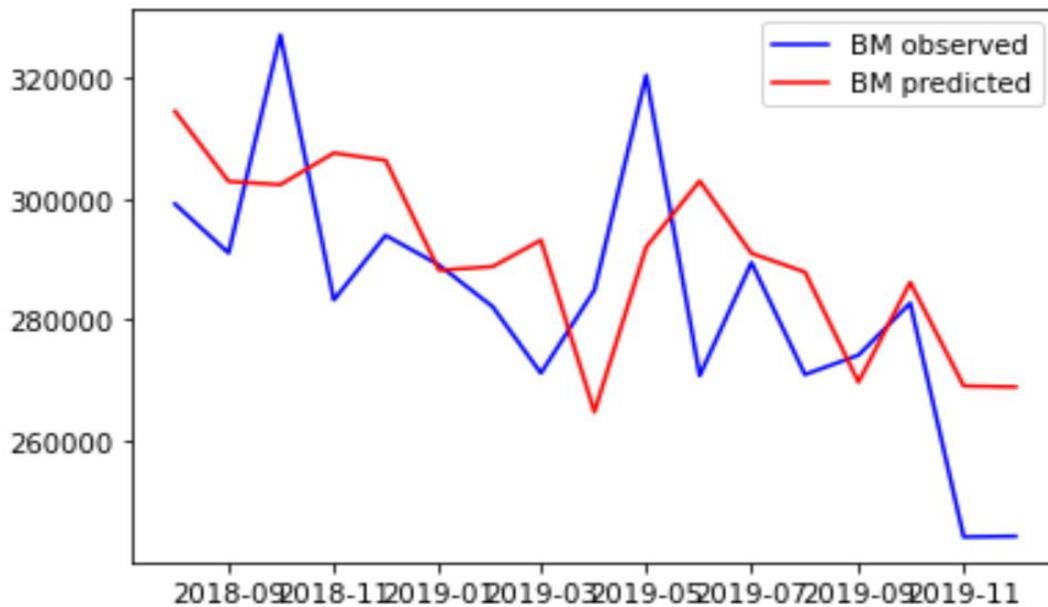


Figura 56: Gráfica ARIMA Mensual Comparativa - Puerto de Barcelona

10.4.5. Resultados aplicación ARIMA datos anuales Puerto Algeciras

- Parámetros seleccionados: ARIMA(1, 2, 2) MSE=1230.047
- Gráfica comparativa de datos reales y estimados y la predicción para los próximos años.

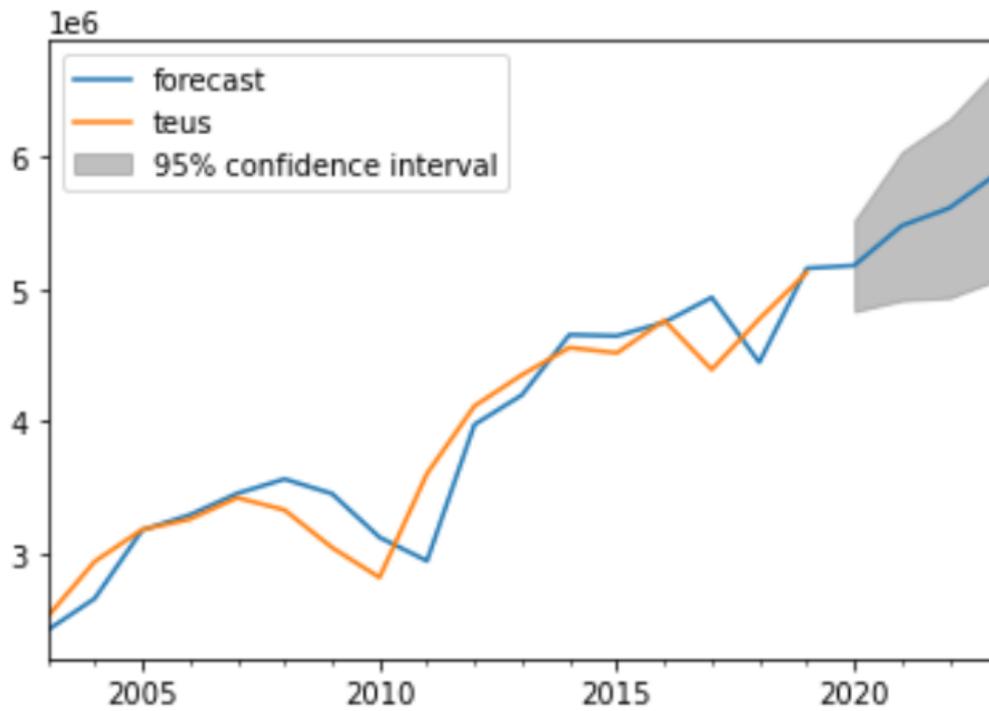


Figura 57: Gráfica ARIMA Anual Predicción - Puerto de Algeciras

- Evaluación de la predicción. usando un 80% de datos para entrenamiento y un 20% para la validación.

Año	Observado (Real)	Predicción
2010	2816556	2859662
2011	3602631	2586695
2012	4114231	4079288
2013	4349755	3927611
2014	4556492	4397827
2015	4515768	4921357
2016	4761444	4500833
2017	4389851	5040201
2018	4773158	4198177
2019	5125385	4820788

Tabla 21: ARIMA Valores Anuales Observados & Predicción Algeciras

- Total error 481387.786
- Comparativa Predicciones y observaciones en los datos de validación

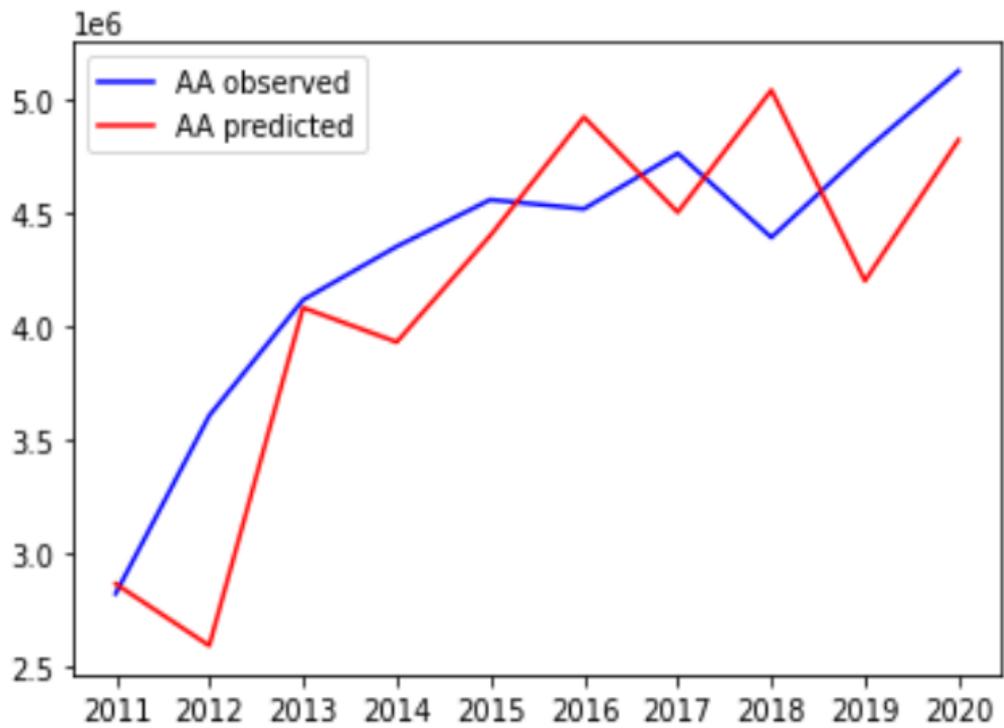


Figura 58: Gráfica ARIMA Anual Comparativa - Puerto de Algeciras

10.4.6. Resultados aplicación ARIMA datos mensuales Puerto Algeciras

- Parámetros seleccionados: ARIMA(0, 1, 1) MSE=3915.283
- Gráfica comparativa de datos reales y estimados y la predicción para los próximos meses

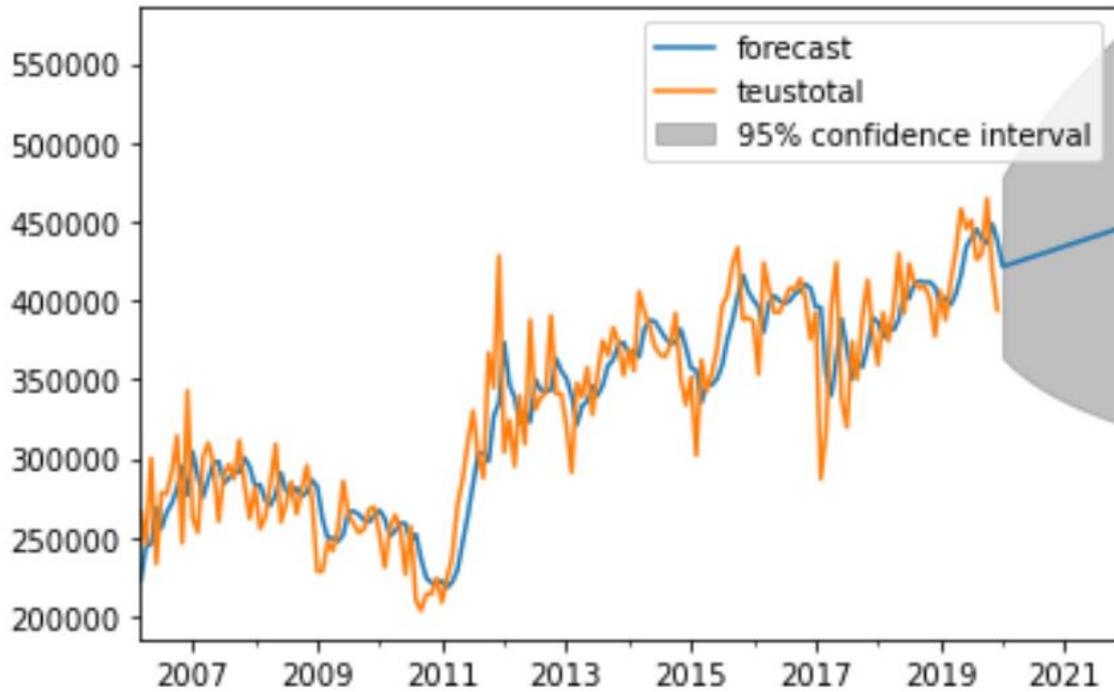


Figura 59: Gráfica ARIMA Mensual Predicción - Puerto de Algeciras

- Evaluación de la predicción. usando un 90% de datos para entrenamiento y un 10% para la validación

Mes	Observado (Real)	Predicción
2018-08	411208	407681
2018-09	407779	413325
2018-10	409251	409805
2018-11	400781	414936
2018-12	377618	407560
2019-01	406064	398312
2019-02	387693	402701
2019-03	412163	400184
2019-04	431747	403254
2019-05	458003	418072

2019-06	445637	431565
2019-07	450763	439149
2019-08	425925	440054
2019-09	429553	437672
2019-10	464455	435598
2019-11	419485	454903
2019-12	393897	440426

Tabla 22: ARIMA Valores Mensuales Observados & Predicción Algeciras

- Total error : 22785
- Comparativa Predicciones y Observaciones.

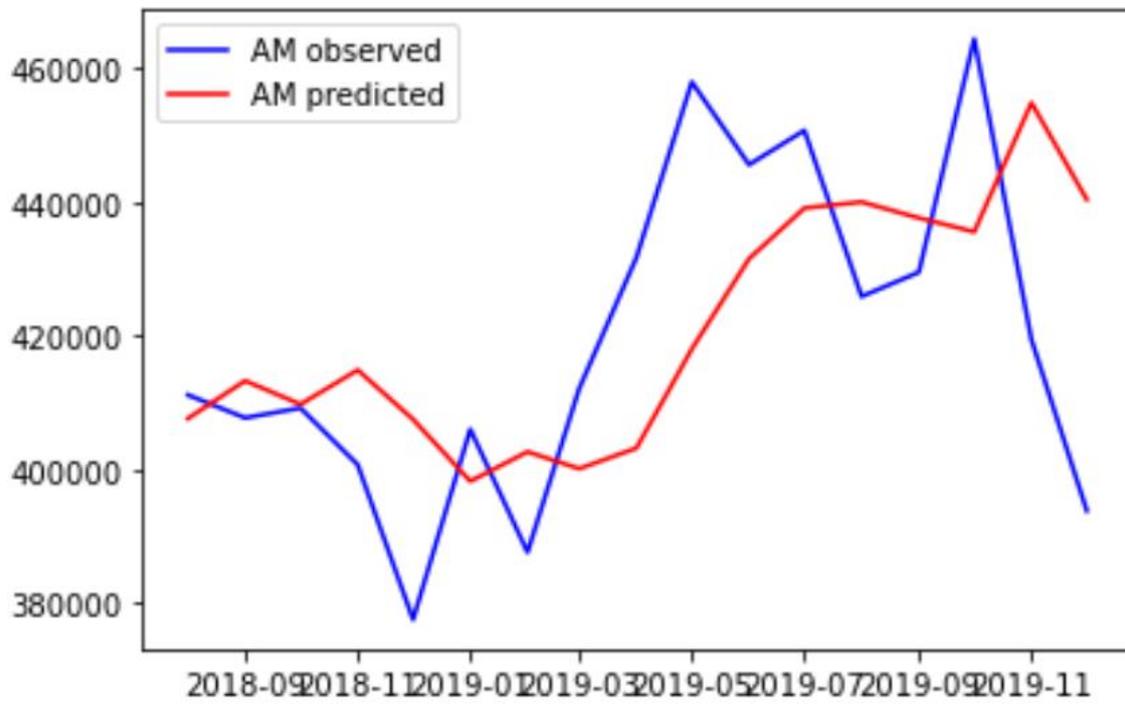


Figura 60: Gráfica ARIMA Mensual Comparativa - Puerto de Algeciras

10.4.7. Resultados aplicación SARIMA datos mensuales Puerto Valencia

- Parámetros seleccionados: SARIMA(0, 1, 1, 1, 1, 1, 12) MSE=3281.848
- Gráfica predicciones futuras

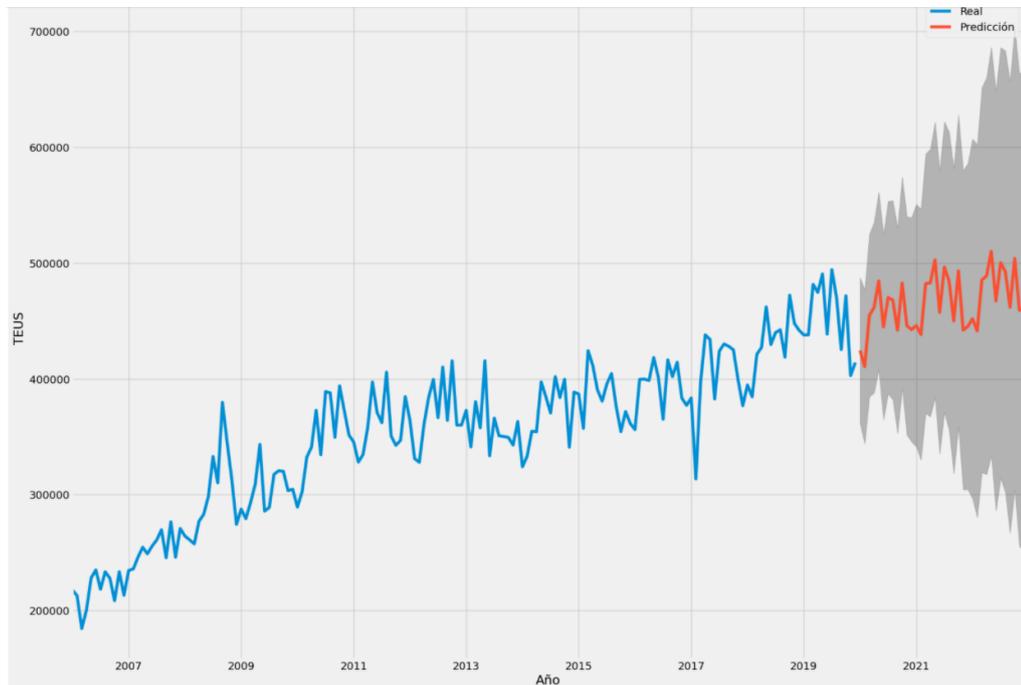


Figura 61: Gráfica SARIMA Mensual Predicción - Puerto de Valencia

- Evaluación de la predicción. usando un 90% de datos para entrenamiento y un 10% para la validación

Mes	Observado (Real)	Predicción
2018-08	442335	456408
2018-09	418714	439625
2018-10	472232	437185
2018-11	447903	423071
2018-12	442029	420291
2019-01	437958	435852

2019-02	438024	397392
2019-03	481607	471904
2019-04	474644	498164
2019-05	490587	500085
2019-06	438593	457674
2019-07	494262	470964
2019-08	470218	490758
2019-09	425314	473050
2019-10	471730	471751
2019-11	402816	446874
2019-12	414074	416772

Tabla 23: SARIMA Valores Observados & Predicción Valencia

- Total error : 25323.518
- Comparativa Predicciones y observaciones

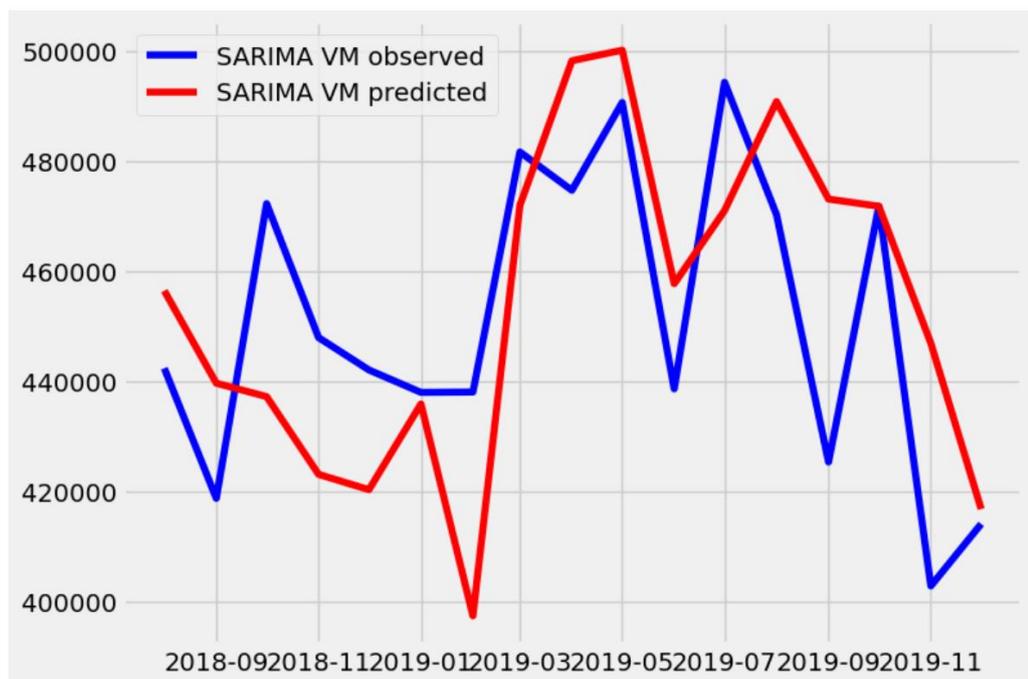


Figura 62: Gráfica SARIMA Mensual Comparativa - Puerto de Valencia

10.4.8. Resultados aplicación SARIMA datos Mensuales Puerto Barcelona

- Parámetros seleccionados: Best ARIMA(0, 1, 1, 0, 1, 1, 12) MSE=3166.039
- Gráfica predicciones futuras

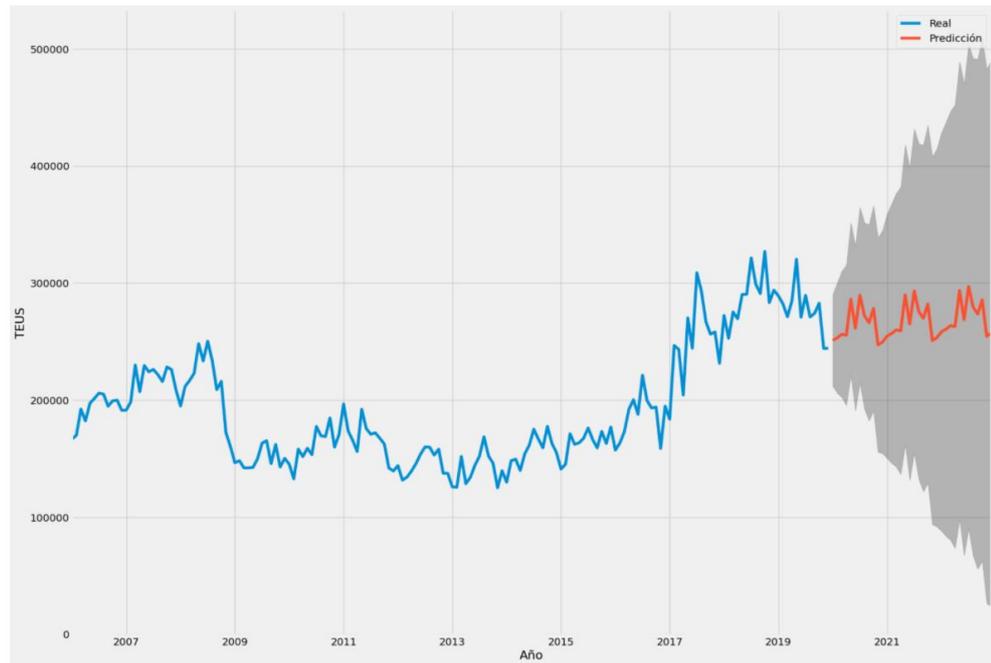


Figura 63: Gráfica SARIMA Mensual Predicción - Puerto de Barcelona

- Evaluación de la predicción. usando un 90% de datos para entrenamiento y un 10% para la validación

Mes	Observado (Real)	Predicción
2018-08	299089	308232
2018-09	291004	287665
2018-10	326985	289489
2018-11	283262	299055
2018-12	293923	293559
2019-01	289017	299791
2019-02	282162	309424

2019-03	271205	302320
2019-04	284893	268376
2019-05	320322	309577
2019-06	270787	307601
2019-07	289426	322092
2019-08	270965	282559
2019-09	274172	261203
2019-10	282753	281929
2019-11	244190	259065
2019-12	244306	251901

Tabla 24: SARIMA Valores Observados & Predicción Barcelona

- Total error : 20249.471
- Comparativa Predicciones y Observaciones

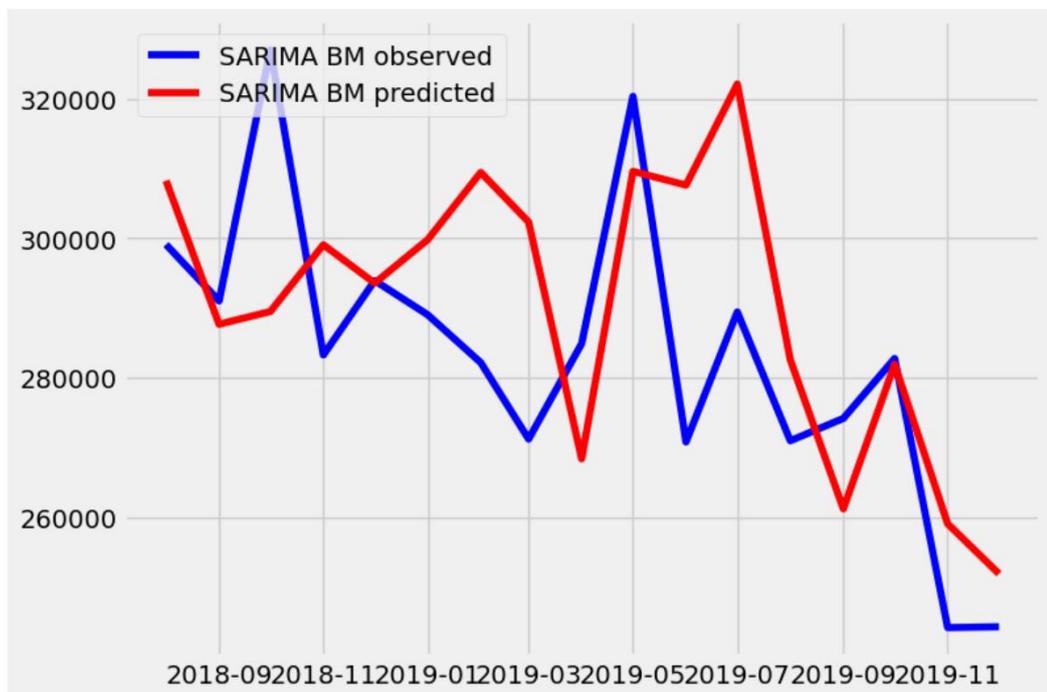


Figura 64: Gráfica SARIMA Mensual Comparativa - Puerto de Barcelona

10.4.9. Resultados aplicación SARIMA datos Mensuales Puerto Algeciras

- Parámetros seleccionados: ARIMA(0, 1, 1, 0, 1, 1, 12) MSE=3322.765
- Gráfica predicciones futuras

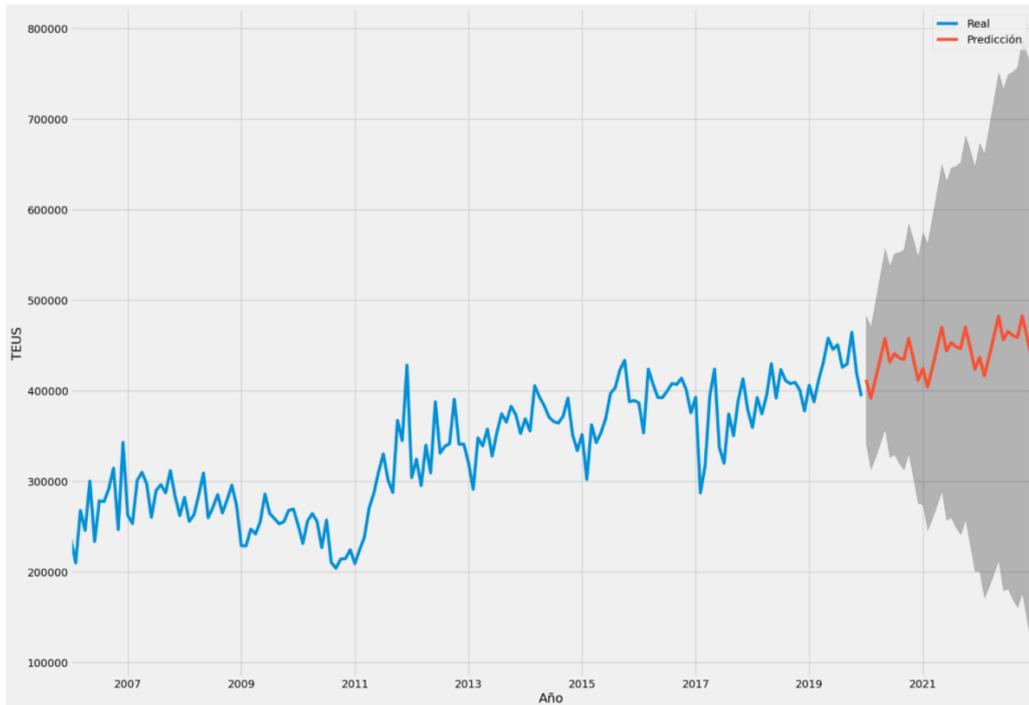


Figura 65: Gráfica SARIMA Mensual Predicción - Puerto de Algeciras

- Evaluación de la predicción. usando un 90% de datos para entrenamiento y un 10% para la validación

Mes	Observado (Real)	Predicción
2018-08	411208	430010
2018-09	407779,	415362
2018-10	409251	433821
2018-11	400781	413935
2018-12	377618	386089
2019-01	406064	380609

2019-02	387693	358727
2019-03	412163	396310
2019-04	431747	431434
2019-05	458003	451470
2019-06	445637	415050
2019-07	450763	439663
2019-08	425925	459315
2019-09	429553	435951
2019-10	464455	449504
2019-11	419485	452214
2019-12	393897	412980

Tabla 25: SARIMA Valores Observados & Predicción Algeciras

- Total error : 20167
- Comparativa Predicciones y Observaciones

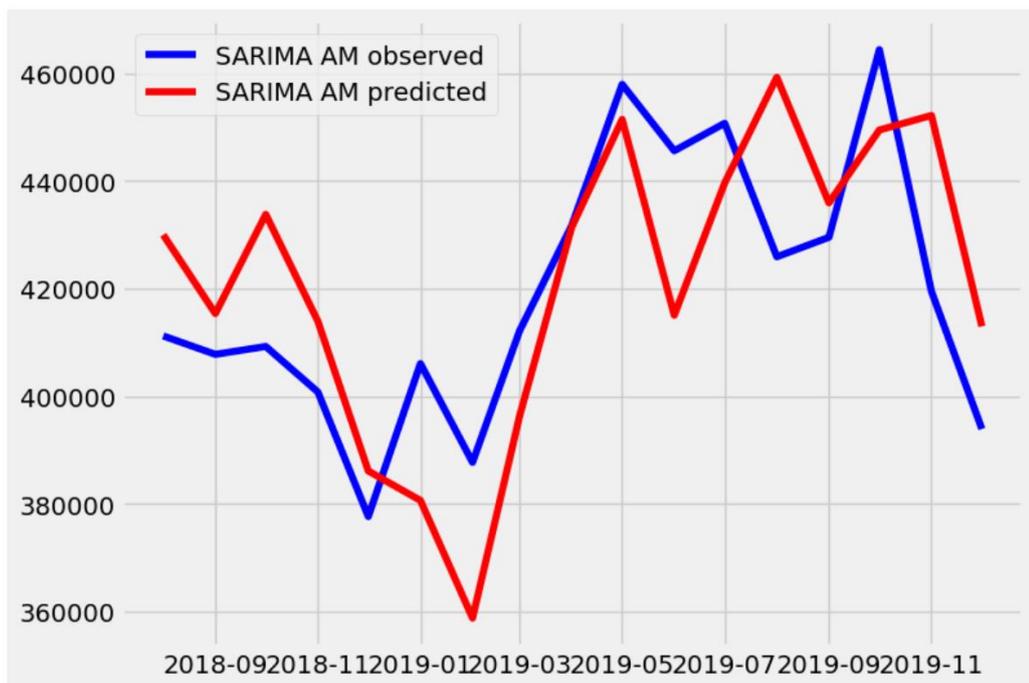


Figura 66: Gráfica SARIMA Mensual Comparativa - Puerto de Algeciras

10.4.10. Resultados aplicación ARIMAX datos Anuales Puerto Valencia

- Parámetros seleccionados: ARIMAX(2,2,1)
- Gráfica predicciones futuras usando PIB Mundial

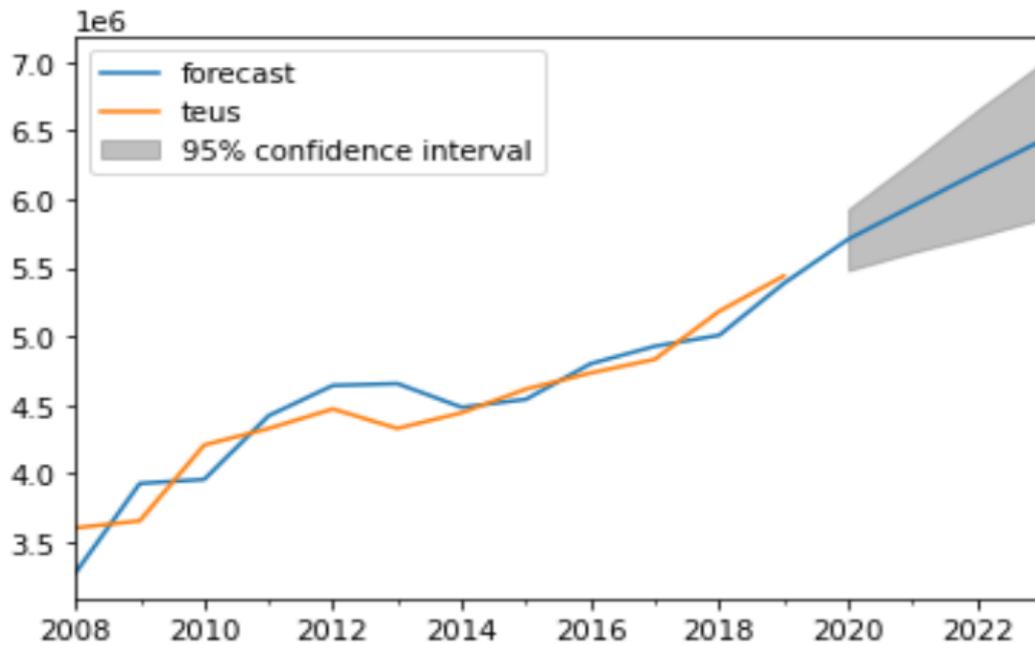


Figura 67: Gráfica ARIMAX Anual Predicción - Puerto de Valencia

- Evaluación de la predicción. Usando una parte de los datos para el entrenamiento y otros para la validación

Mes	Observado (Real)	ARIMAX Predicción con PIB Mundial	ARIMAX Predicción con PIB Español	ARIMAX Predicción con PIB Mundial + español
2010	42069370	40183659	38990113	39364276
2011	43273710	47340530	45532559	46058604
2012	44697540	45200943	45651973	45290382
2013	43278380	46414927	46994762	46546468
2014	44419490	43427461	44015120	43758192

2015	46151960	43001754	45835482	44272903
2016	47321360	47400940	47486020	47519138
2017	48321560	49273392	48719025	49382851
2018	51826650	49926097	49401367	49891218
2019	54398270	53168352	53672352	53248666
Error		217097	190169	190541

Tabla 26: ARIMAX Valores Observados & Predicción Valencia

- Gráfica Comparativa Predicciones y Observaciones usando como variable exógena el PIB Español

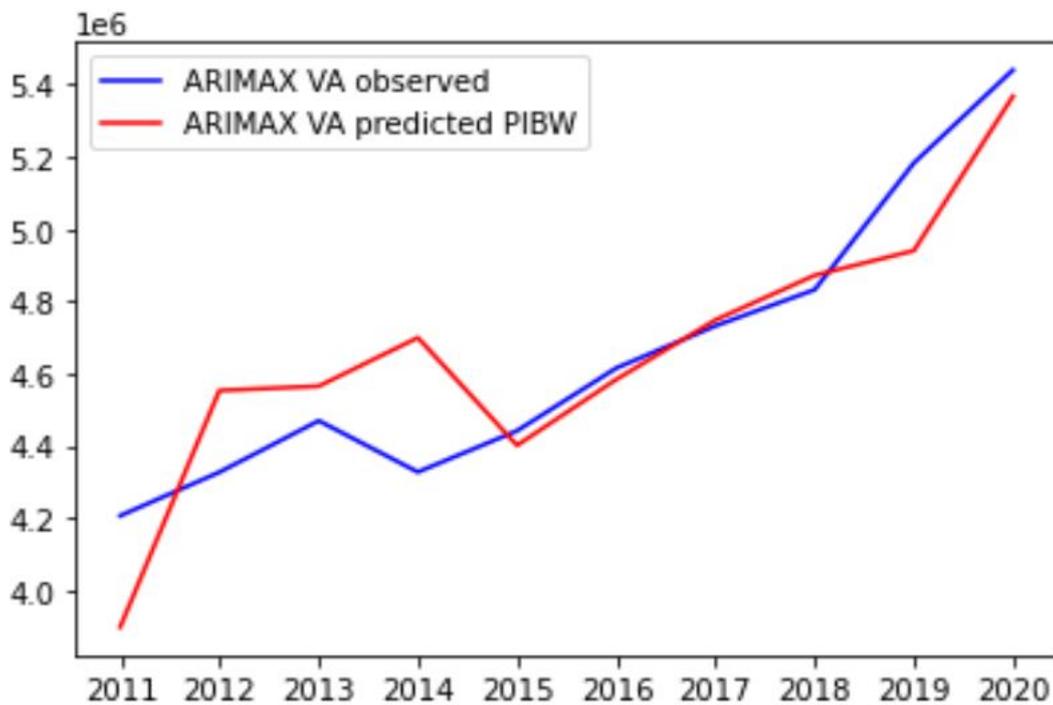


Figura 68: Gráfica ARIMAX Anual Comparativa - Puerto de Valencia

10.4.11. Resultados aplicación ARIMAX datos Anuales Puerto Barcelona

- Parámetros seleccionados: ARIMAX(0,2,2)
- Gráfica predicciones futuras usando PIB Mundial

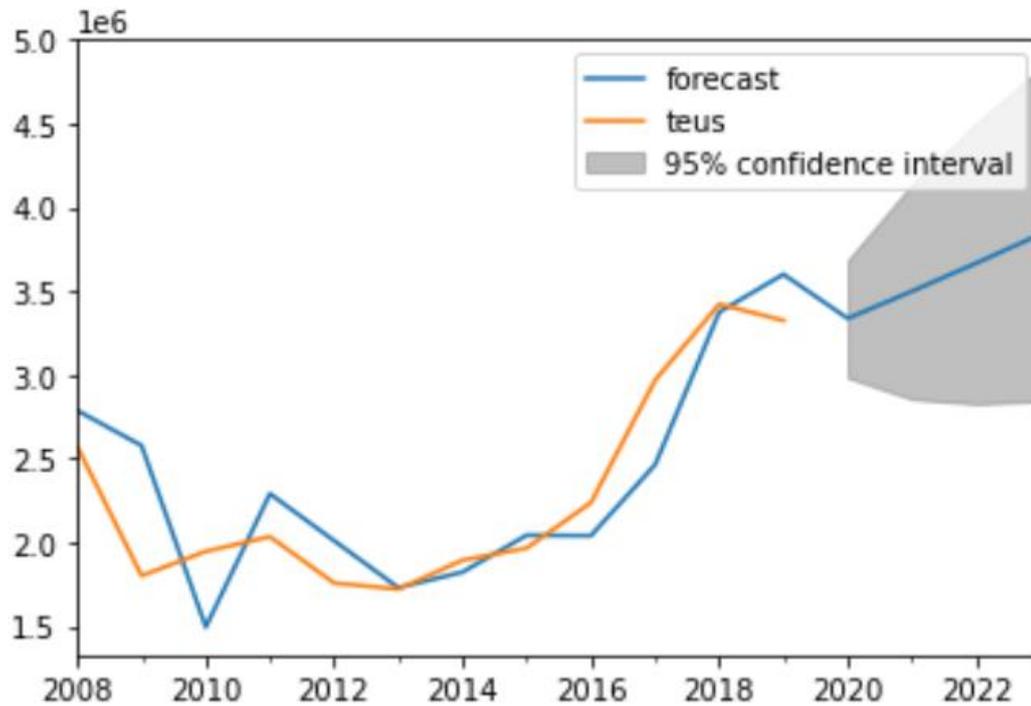


Figura 69: Gráfica ARIMAX Anual Predicción - Puerto de Barcelona

- Evaluación de la predicción. Usando una parte de los datos para el entrenamiento y otros para la validación

Mes	Observado (Real)	ARIMAX Predicción con PIB Mundial	ARIMAX Predicción con PIB Español	ARIMAX Predicción con PIB mundial + español
2010	1945734	1532554	1158891	1331778
2011	2033549	2217995	1983634	2162076
2012	1756429	1700818	1906624	1938496
2013	1720384	1528169	1772290	1669883

2014	1893300	1638639	1753511	1722021
2015	1965241	1543071	2083252	1839301
2016	2236961	2422906	2035542	2130369
2017	2968757	2699689	2433892	2592178
2018	3422978	3703004	3253473	3466385
2019	3324651	3668748	3614177	3697802
Error		281362	334595	277246

Tabla 27: ARIMAX Valores Observados & Predicción Barcelona

En este caso una combinación entre las variables exógenas PIB mundial y PIB español han obtenido un mejor resultado

- Gráfica comparativa predicciones y observaciones usando como variable exógena el PIB mundial + PIB español

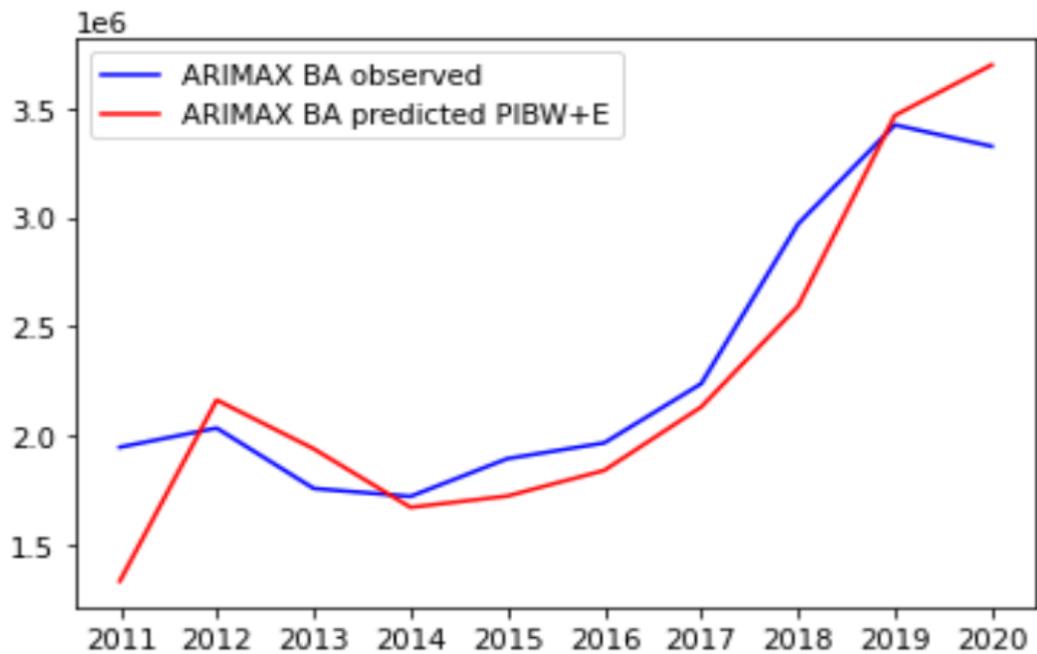


Figura 70: Gráfica ARIMAX Anual Comparativa - Puerto de Barcelona

10.4.12. Resultados aplicación ARIMAX datos Anuales Puerto Algeciras

- Parámetros seleccionados: ARIMAX(1,2,2)
- Gráfica predicciones futuras usando PIB Mundial

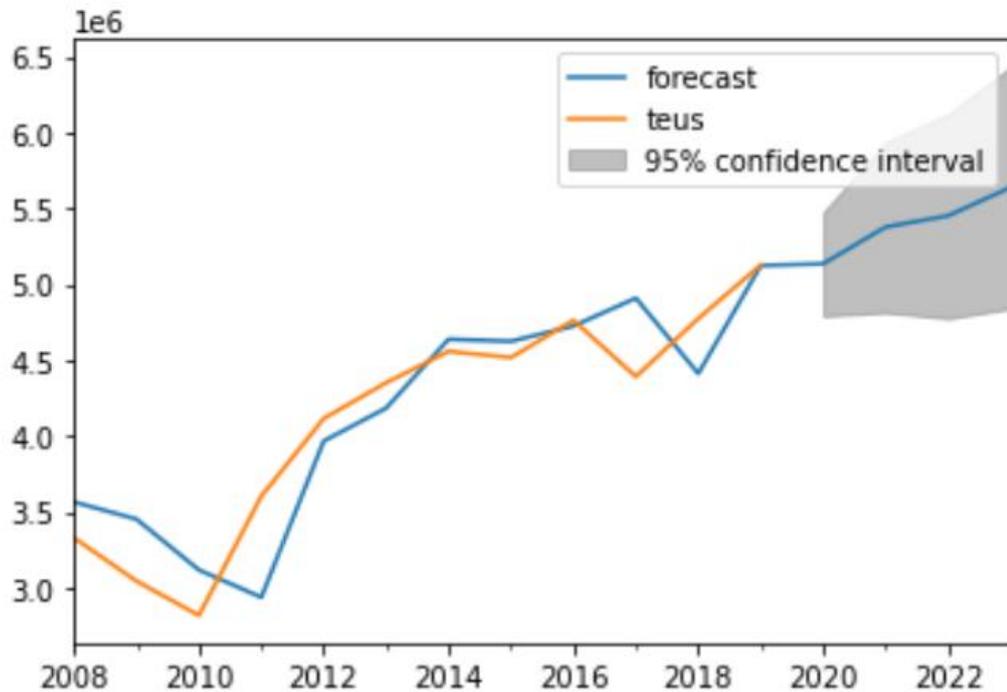


Figura 71: Gráfica ARIMAX Anual Predicción - Puerto de Algeciras

- Evaluación de la predicción. Usando una parte de los datos para el entrenamiento y otros para la validación

Mes	Observado (Real)	ARIMAX Predicción con PIB mundial	ARIMAX Predicción con PIB español	ARIMAX Predicción con PIB mundial + español
2010	2816556	3092492	3016810	3108437
2011	3602631	2659319	2679447	2692336
2012	4114231	3749469	3430980	3473743
2013	4349755	4304965	4169843	4246599

2014	4556492	4506755	4456763	4459681
2015	4515768	4569897	4702906	4588588
2016	4761444	4619735	4478730	4572480
2017	4389851	5039961	4960572	5006135
2018	4773158	4385177	4249228	4349773
2019	5125385	4912294	4915679	4891042
Error		417743	464829	447108

Tabla 28: ARIMAX Valores Observados & Predicción Algeciras

En este caso el modelo que mejor funciona es el que usar la variable exógenas PIB mundial

- Gráfica Comparativa Predicciones y Observaciones usando como variable exógena el PIB Mundial

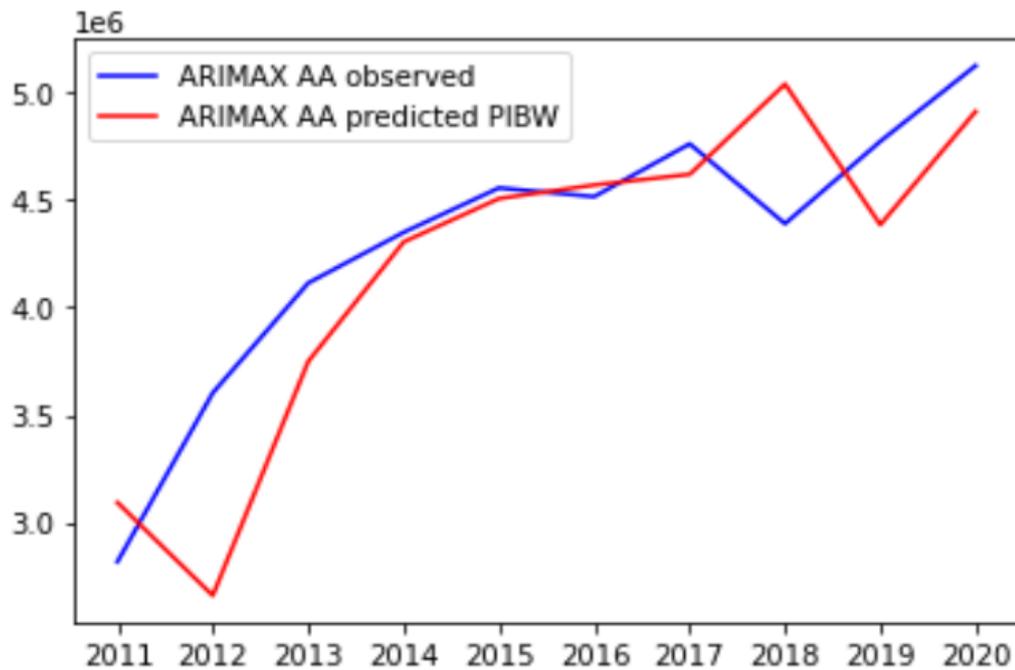


Figura 72: Gráfica ARIMAX Anual Comparativa - Puerto de Algeciras

10.4.13. Resultados aplicación RNN datos Anuales Puerto Valencia

- Total error: 461499
- Gráfica comparativa predicciones y observaciones usando RNN

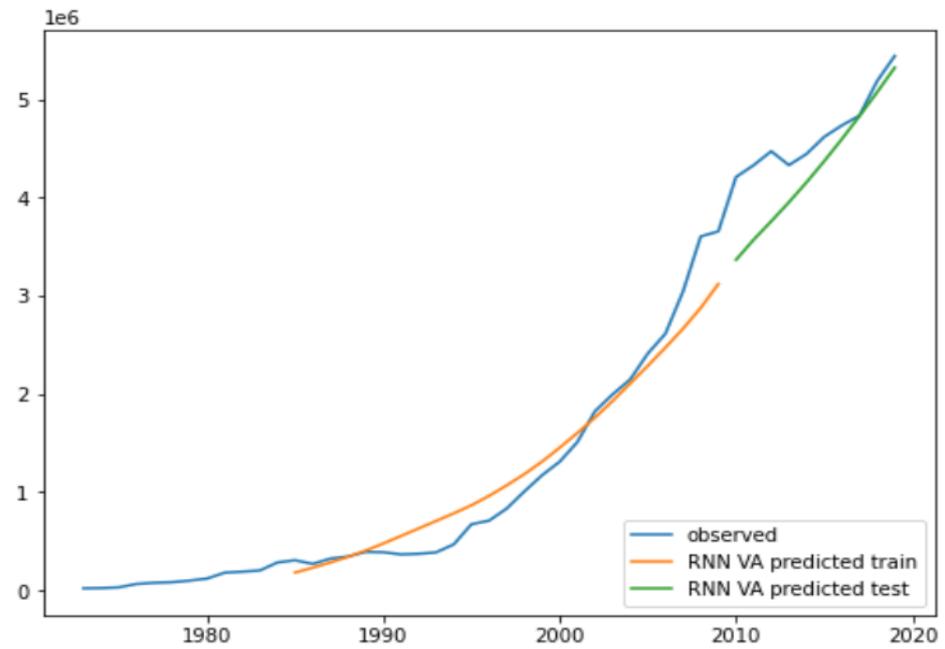


Figura 73: Gráfica RNN Anual Predicción - Puerto de Valencia

10.4.14. Resultados aplicación RNN datos Anuales Puerto Barcelona

- Total error: 1267159
- Gráfica comparativa predicciones y observaciones usando RNN

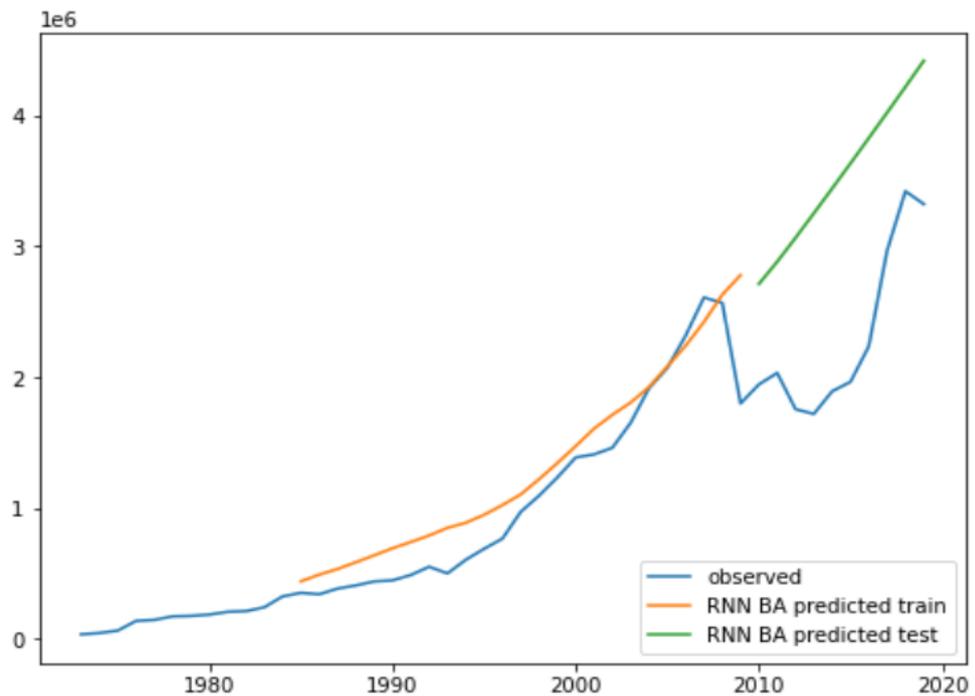


Figura 74: Gráfica RNN Anual Predicción - Puerto de Barcelona

10.4.15. Resultados aplicación RNN datos Anuales Puerto Algeciras

- Total error: 911019
- Gráfica comparativa predicciones y observaciones usando RNN

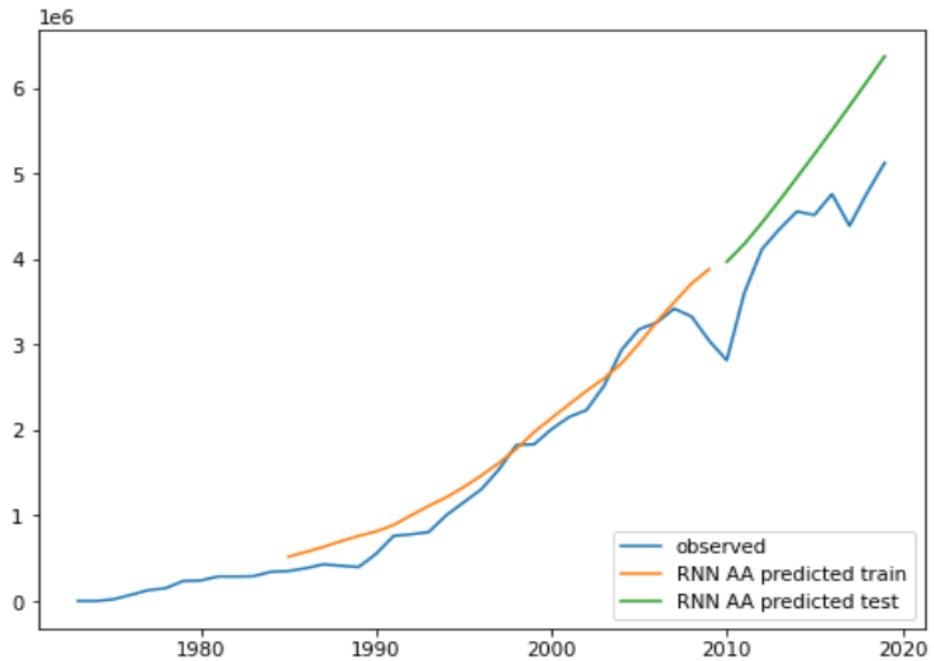


Figura 75: Gráfica RNN Anual Predicción - Puerto de Algeciras

10.4.16. Resultados aplicación RNN datos Mensuales Puerto Valencia

- Total error: 46698
- Gráfica comparativa predicciones y observaciones usando RNN

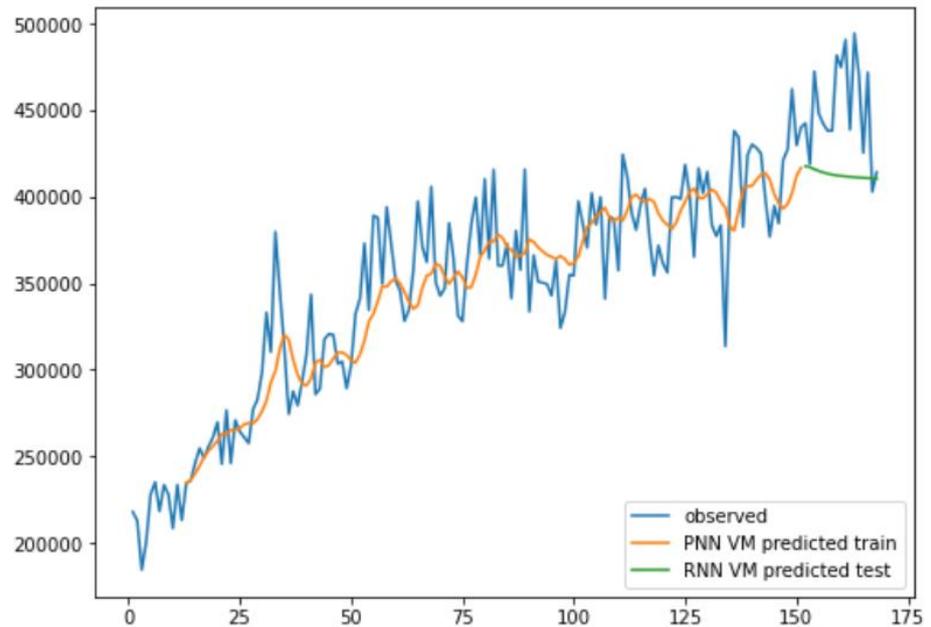


Figura 76: Gráfica RNN Mensual Predicción - Puerto de Valencia

10.4.17. Resultados aplicación RNN datos Mensuales Puerto Barcelona

- Total error: 28125
- Gráfica comparativa predicciones y observaciones usando RNN

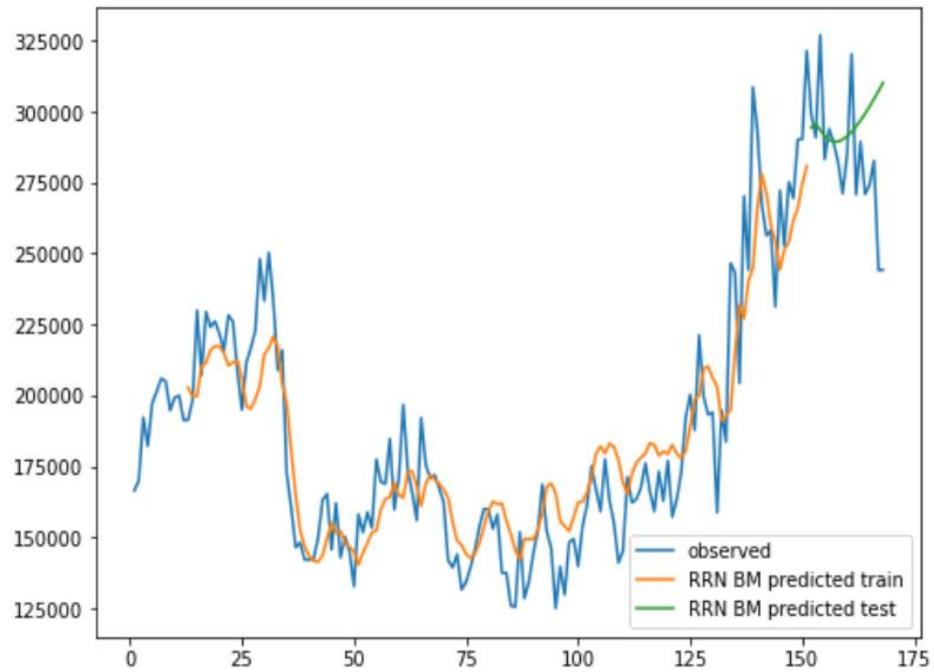


Figura 77: Gráfica RNN Mensual Predicción - Puerto de Barcelona

10.4.18. Resultados aplicación RNN datos Mensuales Puerto Algeciras

- Total error: 29110
- Gráfica comparativa predicciones y observaciones usando RNN

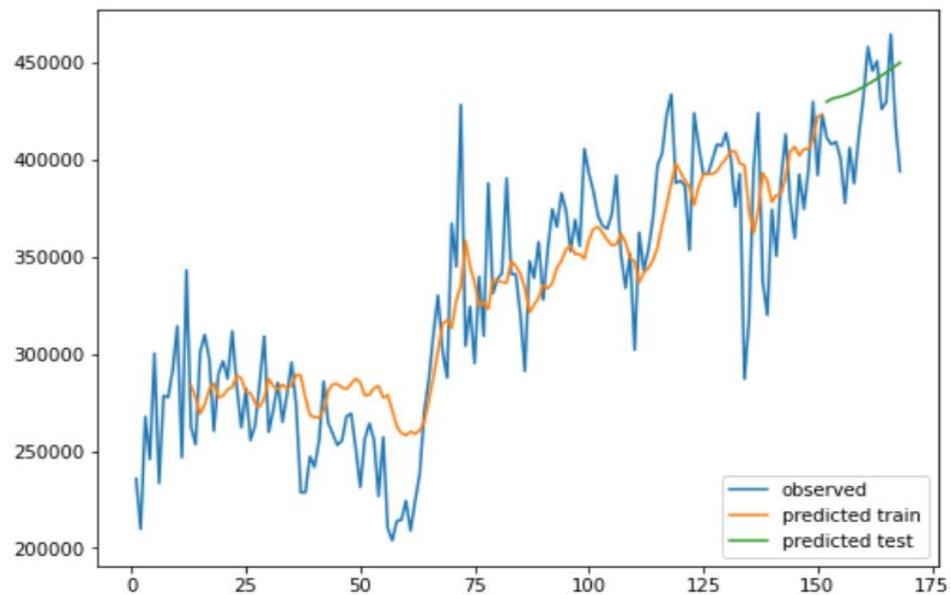


Figura 78: Gráfica RNN Mensual Predicción - Puerto de Algeciras

10.4.19. Resultados aplicación Prophet Mensuales Valencia

- Predicción de contenedores para los próximos años

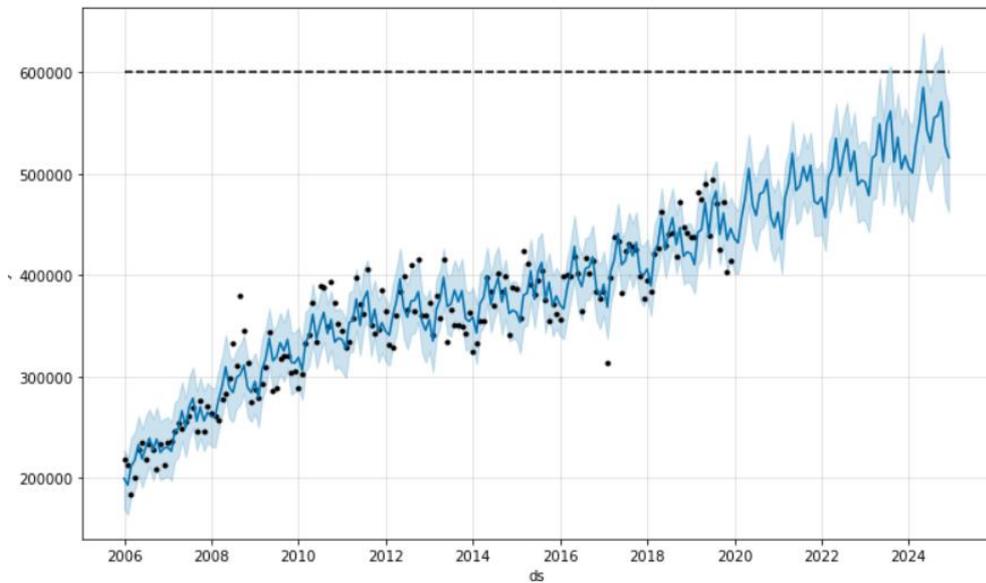


Figura 79: Gráfica Prophet Mensual Predicción - Puerto de Valencia

- Evaluación de la predicción. Usando una parte de los datos para el entrenamiento y otros para la validación.

Mes	Observado (Real)	Predicción
2018-08	442335	448847
2018-09	418714	422197
2018-10	472232	427078
2018-11	447903	408205
2018-12	442029	414628
2019-01	437958	405571
2019-02	438024	394829
2019-03	481607	419903
2019-04	474644	425527
2019-05	490587	454330

2019-06	438593	427391
2019-07	494262	449647
2019-08	470218	470100
2019-09	425314	427395
2019-10	471730	429019
2019-11	402816	420038
2019-12	414074	435176

Tabla 29: Prophet Valores Observados & Predicción Valencia

- Error: 33916
- Gráfica Comparativa Predicciones y Observaciones.

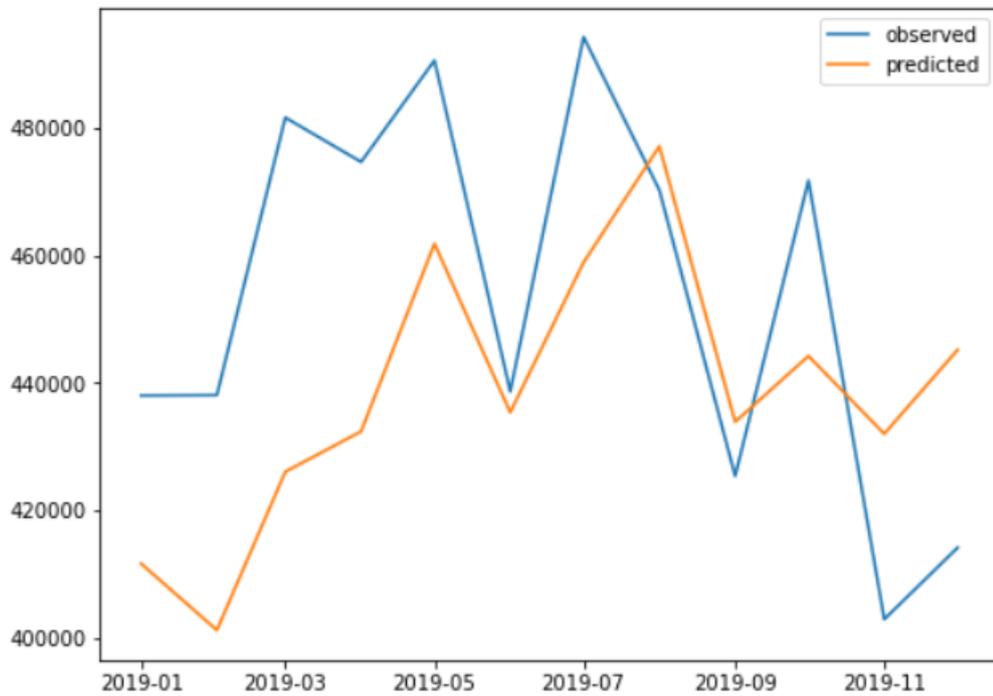


Figura 80: Gráfica Prophet Mensual Comparativa - Puerto de Valencia

10.4.20. Resultados aplicación Prophet Mensuales Barcelona

- Predicción de contenedores para los próximos años

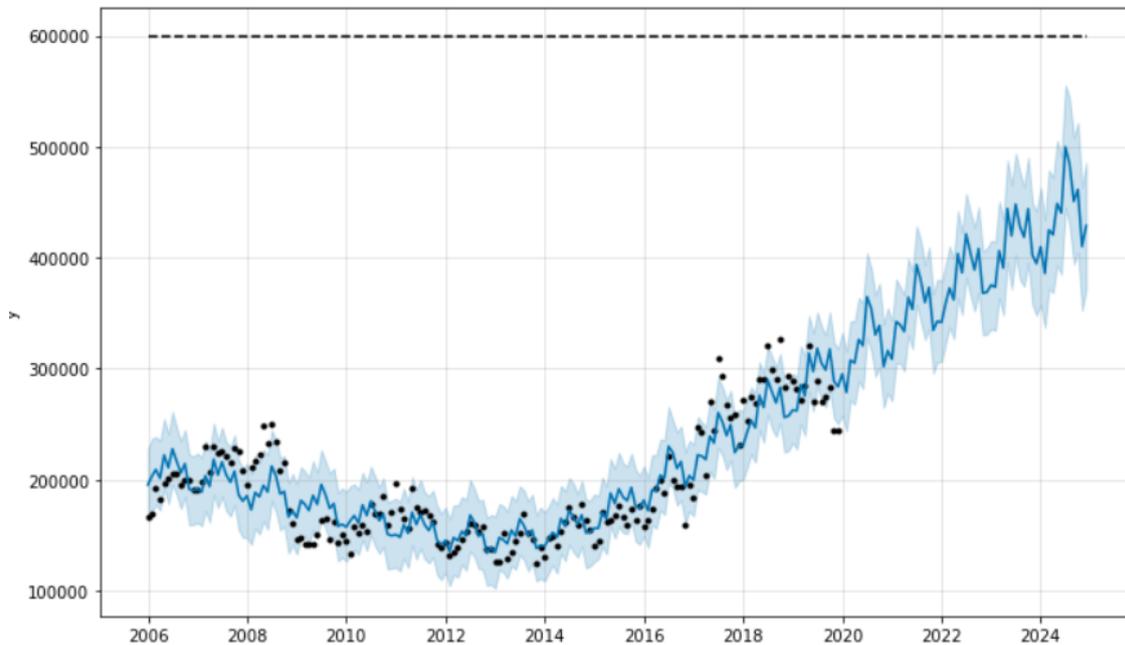


Figura 81: Gráfica Prophet Mensual Predicción - Puerto de Barcelona

- Evaluación de la predicción. Usando una parte de los datos para el entrenamiento y otros para la validación.

Mes	Observado (Real)	Predicción
2018-08	299089	298950
2018-09	291004	283628
2018-10	326985	296064
2018-11	283262	275999
2018-12	293923	275768
2019-01	289017	270643
2019-02	282162	273515

2019-03	271205	310490
2019-04	284893	287142
2019-05	320322	327452
2019-06	270787	326374
2019-07	289426	344993
2019-08	270965	333152
2019-09	274172	317256
2019-10	282753	335101
2019-11	244190	318619
2019-12	244306	312784

Tabla 30: Prophet Valores Observados & Predicción Barcelona

- Error: 40782
- Gráfica comparativa predicciones y observaciones.

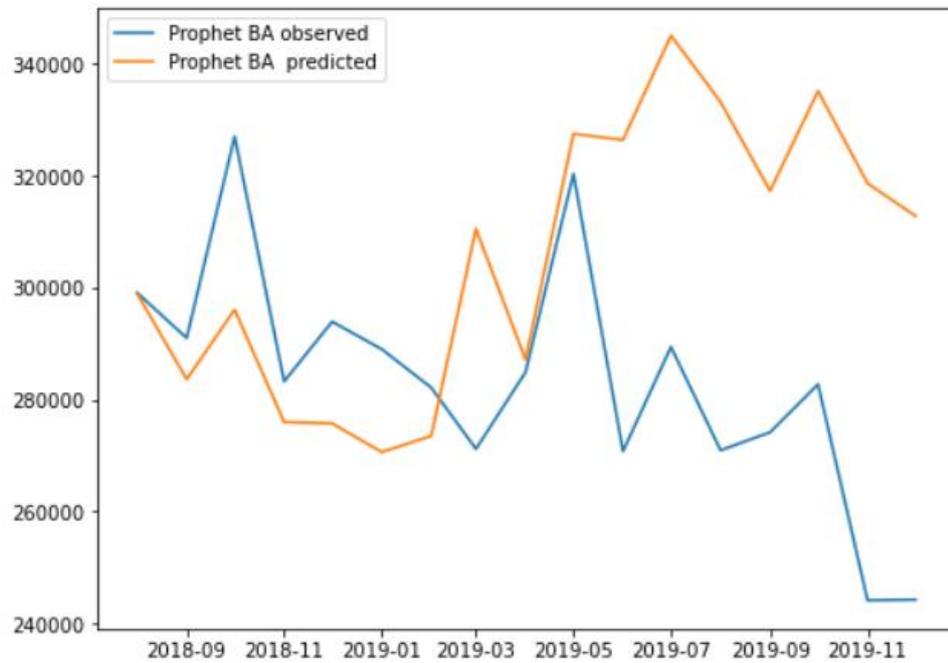


Figura 82: Gráfica Prophet Mensual Comparativa - Puerto de Barcelona

10.4.21. Resultados aplicación Prophet Mensuales Algeciras

- Predicción de contenedores para los próximos años

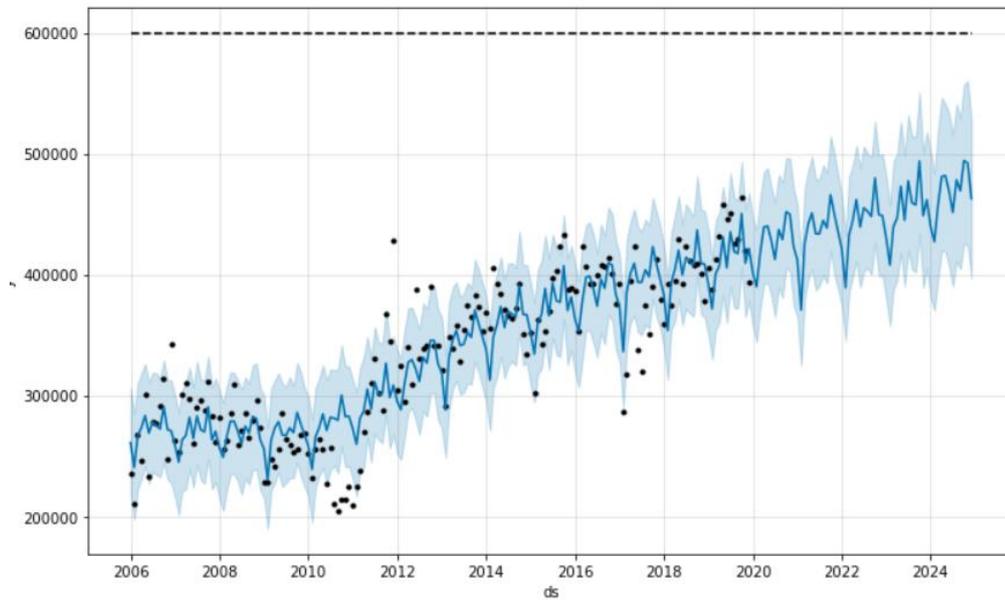


Figura 83: Gráfica Prophet Mensual Predicción - Puerto de Algeciras

- Evaluación de la predicción. Usando una parte de los datos para el entrenamiento y otros para la validación

Mes	Observado (Real)	Predicción
2018-08	411208	401749
2018-09	407779	397953
2018-10	409251	429234
2018-11	400781	398444
2018-12	377618	411008
2019-01	406064	380589
2019-02	387693	360803
2019-03	412163	390824
2019-04	431747	394788

2019-05	458003	418327
2019-06	445637	390489
2019-07	450763	426666
2019-08	425925	406626
2019-09	429553	403575
2019-10	464455	439422
2019-11	419485	393196
2019-12	393897	426447

Tabla 31: Prophet Valores Observados & Predicción Algeciras

- Error: 28200
- Gráfica Comparativa Predicciones y Observaciones.

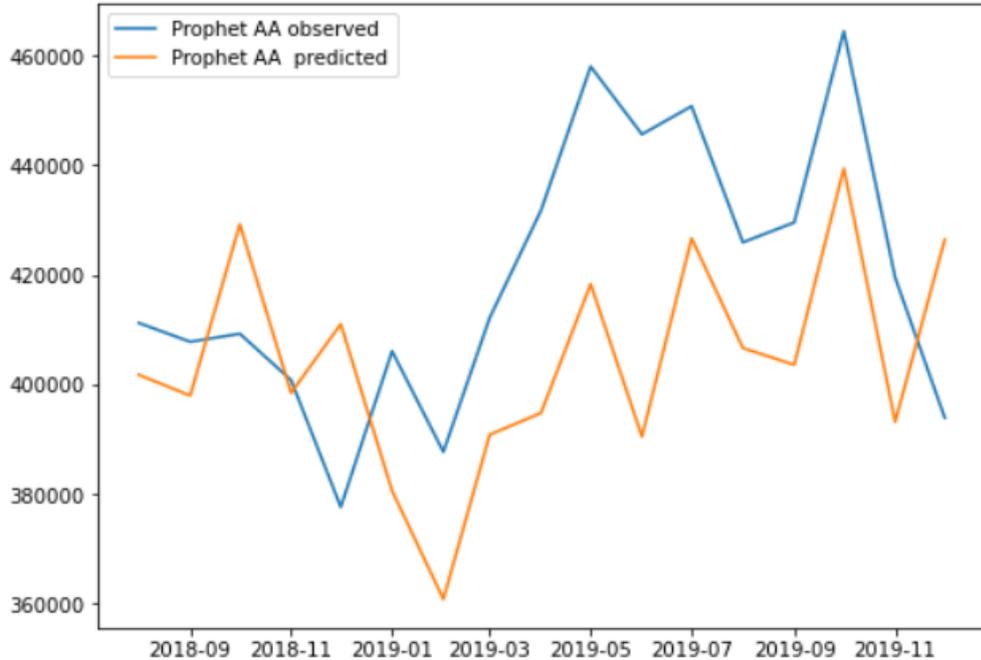


Figura 84: Gráfica Prophet Mensual Comparativa - Puerto de Algeciras

Nota: para llevar a cabo el estudio se han tenido que modificar ligeramente las series temporales, ya que el algoritmo funciona usando tan solo 2 columnas con nombres especificados.

10.4. Anexo Qlik Sense

Scripts de carga de datos

Para la realización de la carga de datos se ha partido de los siguientes ficheros CSV.

- **spanish_historic_annual.csv** que contiene la información general anualizada desde 1973 y que ha sido exportada desde la tabla “spanish_historic_annual” de la base de datos. El fichero contiene 1491 registros y 25 campos.

Script de importación
<pre>qualify *; unqualify anyo, puerto; LOAD anyo, puerto, codap, productospetroliferos, otrosliquidos, granelesliquidos, granelessolidos, merc999999999anciageneralconvencional, mercanciageneralcontenedores, mercanciageneral, pescafresca, avituallamientoprodpetroliferos, avituallamientoresto, avituallamiento, totalotrasmercancias, tráfico local, totaltrafico, contenedores, teus, buquestotales, buquesprotacontenedores, buquesgt, pasajeroscruceiro, pasajerosentransito, pasajerosembarcadosdesembarcados</pre>

```
FROM [lib://Downloads/spanish_historic_annual.csv]
(txt, utf8, embedded labels, delimiter is ',', msq);
```

- **spanish_historic_contenedores.csv** que contiene la información de contenedores anualizada desde 1973 y que ha sido exportada desde la tabla “spanish_historic_contenedores” de la base de datos. El fichero contiene 936 registros y 25 campos.

Script de importación

```
qualify *;

unqualify anyo, puerto;

LOAD
    anyo,
    puerto,
    codap,
    contenedorcabotajecarga,
    contenedorcabotajedescarga,
    contenedorcabotajetofal,
    contenedorexteriorcarga,
    contenedorexteriordescarga,
    contenedorexteriortotal,
    contenedortotalcarga,
    contenedortotaldescarga,
    contenedortotal,
    teuscabotajecarga,
    teuscabotajedescarga,
    teuscabotajetotal,
    teusexteriorcarga,
    teusexteriordescarga,
    teusexteriortotal,
    teustotalcarga,
    teustotaldescarga,
    teustotal,
    exteriortoneladas,
```

```

cabotajetoneladas,
totaltoneladas
FROM [lib://Downloads/spanish_historic_contenedores.csv]
(txt, utf8, embedded labels, delimiter is ',', msq);

```

- **spanish_month_contenedores.csv** que contiene la información mensual de los puertos desde enero de 2006 y que ha sido exportada desde la tabla “spanish_month_contenedores” de la base de datos. El fichero contiene 4957 registros y 30 campos.

Script de importación

```

LOAD
    pem_mes,
    pem_año,
    pem_año&Month(pem_mes) as año_mes,
    MakeDate(pem_año,pem_mes) as mes_año,
    pem_puerto,
    pem_codap,
    pem_teustotal,
    // If(pem_año<2010,'Real','Predicción') as campo,
    pem_teustransito,
    pem_teuscabotaje,
    pem_teusexterior,
    pem_teuscabotajeyexterior,
    pem_totaltoneladascontenedores,
    pem_transitotoneladascontenedores,
    pem_totaltrafico,
    pem_totalpresentacion,
    pem_granelesliquidos,
    pem_granelessolidos,
    pem_mercanciageneral,
    pem_pesca,
    pem_avituellamiento,
    pem_avituellamientocombustibles,
    pem_traficointerior,
    pem_mercanicatransito,

```

```
pem_traficororo,  
pem_traficorororemolques,  
pem_pasajerototal,  
pem_pasajerocrucero,  
pem_automovilpasaje,  
pem_automovilregimenmercancia,  
pem_buques,  
pem_buquesgt,  
pem_cruceros  
// 1 as camp01,  
FROM [lib://Downloads/pem_contenedores.csv]  
(txt, utf8, embedded labels, delimiter is ',', msq);
```