

VERSION 1.0

JUNIO, 2018



TRABAJO FINAL DE GRADO DE INGENIERÍA INFORMÁTICA

BUSINESS INTELLIGENCE

FERNANDO CORTÉS HIERRO

XAVIER MARTÍNEZ FONTES
ATANASI DARADOUMIS HARALABUS



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>Análisis del impacto de la climatología en los precios de los alimentos básicos en países subdesarrollados y en vías de desarrollo.</i>
Nombre del autor:	<i>Fernando Cortés Hierro</i>
Nombre del consultor:	<i>Xavier Martínez Fontes</i>
Nombre del PRA:	<i>Atanasi Daradoumis Haralabus</i>
Fecha de entrega (mm/aaaa):	<i>06/2018</i>
Titulación:	<i>Grado de Ingeniería Informática</i>
Área del trabajo final:	<i>Business Intelligence</i>
Idioma del trabajo:	<i>Español</i>
Palabras Clave	<i>Business Intelligence, World Food Programme, Big Data, Weather</i>

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.*

Existe una creciente preocupación entre la comunidad internacional debido a la paulatina inestabilidad del precio de los alimentos básicos en países subdesarrollados o en vías de desarrollo que tiene como consecuencia hambrunas y migraciones. Los diferentes actores implicados intentan paliar la situación en la medida de lo posible, pero existen multitud de causas que pueden influir en esta fluctuación de precios: meteorológicas, plagas, costes asociados o situaciones políticas y económicas.

Todos los factores causantes de estas fluctuaciones en el precio generan ingentes cantidades de datos que son susceptibles de ser analizados y comparados entre sí para encontrar correlaciones y modelos predictivos que ayuden a las organizaciones y países a implementar medidas paliativas que eviten dramas humanitarios.

Este trabajo explora las opciones que ofrecen las herramientas de análisis de negocio actuales para trabajar con grandes cantidades de datos y extraer información valiosa, mediante la creación de un prototipo que permita analizar la relación entre la variación del precio de los alimentos y la climatología del país objeto de estudio.

El resultado obtenido demuestra la idoneidad del uso de este tipo de herramientas para trabajar con grandes volúmenes de datos, además de un prototipo funcional para analizar la inestabilidad de los precios en función del clima. Sin bien es cierto que se puede observar cierta causa-efecto entre precio y clima, es necesario añadir más causas a la ecuación para

poder obtener resultados más precisos y que podrían ayudar a predecir cambios en los precios si se evoluciona el prototipo a una solución completa.

Abstract (in English, 250 words or less):

There is a growing concern among the public and international community about the fluctuating food prices and global food security on the regions throughout the developing world that cause hunger and regional migrations. The different actors involved are trying to mitigate the situations as much as possible, but there are many different inter-linkages causes that can impact into this price fluctuation like the weather, pests, logistics costs or political and economic instabilities.

All these factors causing these fluctuating food prices generate huge amounts of data that can be analysed and compared to be able to find correlations and create models to predict these fluctuations, helping organizations and countries to mitigate the possible disastrous effects in affected countries.

This academic work explores the different possibilities offered by current business analysis tools on the cloud to work with large amounts of information and extract insights, developing a prototype to try to find the relationship between the fluctuating food prices and the countries data-weather.

The results obtained shown the suitability of these type of tools to work with large volumes of data, as well as a functional prototype to analyse the instability of prices according to the weather. It was discovered a small cause-effect between price fluctuations and weather, but it is mandatory to increase the number of causes to be able to obtain accurate results and which could help predict price fluctuations if the prototype is evolved to a complete solution.

TABLA DE CONTENIDOS

FICHA DEL TRABAJO FINAL	2
TABLA DE CONTENIDOS	4
REGISTRO DE CAMBIOS	6
1.- INTRODUCCIÓN	7
1.1.- CONTEXTO Y JUSTIFICACIÓN DEL TRABAJO	7
1.2.- OBJETIVOS DEL TRABAJO.....	8
1.3.- ALCANCE DEL TRABAJO.....	8
1.4.- METODOLOGÍA DE TRABAJO	9
1.5.- PLANIFICACIÓN DEL TRABAJO.....	9
1.6.- BREVE RESUMEN DE PRODUCTOS OBTENIDOS.....	11
2.- LA INESTABILIDAD DEL PRECIO DE LOS ALIMENTOS BÁSICOS	12
2.1.- CAUSAS	12
2.1.1.- <i>METEOROLÓGICAS: TEMPERATURA Y PRECIPITACIONES</i>	12
2.1.2.- <i>PESTES Y ENFERMEDADES EN LAS COSECHAS</i>	13
2.1.3.- <i>COSTES LOGÍSTICOS</i>	13
2.1.4.- <i>COSTES LABORALES</i>	14
2.1.5.- <i>SITUACIONES POLÍTICAS Y ECONÓMICAS</i>	14
2.2.- CONSECUENCIAS	14
2.2.1.- <i>MIGRACIONES</i>	15
2.2.2.- <i>HAMBRUNAS</i>	15
2.3.- PLANES DE CONTINGENCIA.....	15
3.- ANÁLISIS Y PREDICCIÓN DEL PRECIO DE ALIMENTOS MEDIANTE BI	17
3.1.- COMPARATIVA DE HERRAMIENTAS DEL MERCADO	17
3.2.- HERRAMIENTAS DE ANÁLISIS EN MICROSOFT AZURE	19
3.3.- FUENTES DE DATOS.....	20
3.4.- ANÁLISIS DE DATOS HISTÓRICOS	21
3.5.- ANÁLISIS PREDICTIVO	22
3.6.- REPRESENTACIÓN DE LA INFORMACIÓN	23
4.- IMPLEMENTACIÓN DEL PROTOTIPO DE ANALISIS DE PRECIOS EN FUNCIÓN DE FACTORES METEOROLÓGICOS	24
4.1.- ANALISIS FUNCIONAL	24
4.1.1.- <i>ESTUDIO DE FUENTES DE DATOS</i>	25
4.1.2.- <i>ESTUDIO DE HERRAMIENTAS BUSINESS INTELLIGENCE EN MICROSOFT AZURE</i>	27
4.1.3.- <i>DEFINICIÓN DE INDICADORES</i>	30
4.1.4.- <i>DEFINICIÓN DE INFORME</i>	31
4.2.- DISEÑO	33
4.2.1.- <i>ESTRUCTURA DE DATOS</i>	33
4.2.2.- <i>EXTRACCIÓN Y TRANSFORMACIÓN DE DATOS</i>	34
4.2.3.- <i>CARGA DE DATOS</i>	38
4.3.- IMPLEMENTACIÓN	39
4.3.1.- <i>CONFIGURACIÓN DE HERRAMIENTAS MICROSOFT AZURE</i>	40
4.3.2.- <i>TRANSFORMACIÓN DE DATOS (ETL)</i>	42

4.3.3.- CARGA DE DATOS.....	50
4.3.4.- IMPLEMENTACIÓN INFORME.....	52
4.4.- RESULTADOS.....	58
4.4.1.- ALGORITMOS DE INFORMACIÓN DE INTERÉS.....	60
5.- LINEAS DE TRABAJO FUTURO.....	63
6.- CONCLUSIONES.....	64
7.- GLOSARIO.....	65
8.- BIBLIOGRAFÍA.....	66

REGISTRO DE CAMBIOS

A continuación, se presenta el registro de cambios realizados en este documento.

Fecha	Autor	Descripción	Versión
17/03/2018	fcortesh	Estructura y formato.	0.1
31/03/2018	fcortesh	Introducción e índice inicial	0.2
16/04/2018	fcortesh	Primera versión para revisión	0.3
21/05/2018	fcortesh	Segunda versión para revisión	0.4
26/05/2018	fcortesh	Revisión después comentarios de Xavier Martínez en PEC2	0.5
28/05/2018	fcortesh	Resumen del trabajo	0.6
04/06/2018	fcortesh	Revisión después comentarios de Xavier Martínez en PEC3	0.7
18/06/2018	fcortesh	Versión final	1.0

1.- INTRODUCCIÓN

1.1.- CONTEXTO Y JUSTIFICACIÓN DEL TRABAJO.

La motivación de este trabajo surge de la necesidad que pueden tener los organismos humanitarios enfocados en paliar las hambrunas y éxodos masivos de población, en países en vías de desarrollo y subdesarrollados por el efecto devastador de la escasez de alimentos básicos.



ILUSTRACIÓN 1 - RELACIÓN DE PAÍSES OBJETO DE ESTUDIO (FUENTE: PROPIA USANDO POWER BI)

Si bien es cierto que existen diferentes factores que influyen en el precio final de los alimentos denominados como básicos en las diferentes áreas del globo, como pueden ser: guerras, conflictos sociales puntuales, desastres naturales, etc., existe un común denominador que siempre hará fluctuar los precios. La climatología es por tanto uno de los factores más influyentes en el precio final de los alimentos básicos debido a su impacto en las cosechas.

Según un informe del Programa Mundial de Alimentos, *“Uno de los impactos más significativos del cambio climático es el aumento potencial de la inseguridad alimentaria y la malnutrición”* por lo que factores climatológicos como la lluvia y la temperatura a lo largo de los últimos años, servirán para estudiar el impacto sobre las plantaciones y, por lo tanto, sobre el precio final de alimentos.

La razón fundamental de este trabajo es dotar de herramientas de análisis y prevención en la lucha por las hambrunas y migraciones en áreas especialmente sensibles a los factores climatológicos.

1.2.- OBJETIVOS DEL TRABAJO.

La finalidad de este trabajo es dotar a las organizaciones no gubernamentales, de un sistema de análisis de datos que permita visualizar la evolución de los precios de alimentos básicos en diferentes países en función de los factores climatológicos.

Para ello, se desarrollará un prototipo funcional con información obtenida del programa mundial de alimentos y datos climatológicos mensuales del NOAA, donde se podrán visualizar la evolución de precios por país, de los alimentos básicos que se puedan ver afectados por la climatología, y su relación con factores climatológicos.

1.3.- ALCANCE DEL TRABAJO.

En función de los antecedentes antes descritos y con los objetivos marcados para este trabajo de fin de grado, el proyecto contará con los siguientes aspectos:

Configuración de la plataforma de análisis y transformación de datos en Microsoft Azure

Debido a la experiencia profesional en cuanto a tecnologías Microsoft y la posibilidad de acceder a subvenciones en el uso de herramientas en Azure para construir el prototipo, se opta por usar Microsoft Azure y sus tecnologías de análisis de datos.

Este entregable formará parte de la memoria y consistirá en la configuración, y puesta a punto de las herramientas en una cuenta en Microsoft Azure.

Definición de indicadores que permitan analizar la evolución de precios y los factores meteorológicos asociados.

Se deberán examinar los datos disponibles previamente, tanto de países, tipos de alimentos, precios y sus monedas, y fechas, así como, los datos relativos a la climatología que se puedan encontrar. Una vez analizados los datos disponibles se podrán definir los indicadores relativos a los datos que sean de interés para cumplir el objetivo del proyecto.

Este entregable formará parte de uno de los apartados de la presente memoria.

Implementación de los procesos de Análisis y transformación de los datos disponibles según país.

Realizar todos los procesos de análisis y transformación de los datos para obtener un conjunto de información de precios y climatología que permita aplicar futuros procesos predictivos, así como, mediante el uso de herramientas de visualización, mostrar los indicadores definidos.

Este entregable formará parte de la documentación final del proyecto y consistirá en el proyecto de Visual Studio 2017 con los scripts de transformación U-SQL.

Visualizadores de los datos procesados.

Utilizando la potencia de la herramienta de visualización de datos *Microsoft Power BI*, se construirán informes que permitan analizar visualmente la evolución de precios y climatología, así como su posible interrelación. Dentro de este entregable, y previamente a la implementación de los informes se definirá que informes y con qué información se crearán.

Este entregable será accesible mediante una dirección web descrita en la presente memoria.

1.4.- METODOLOGÍA DE TRABAJO.

Debido a la necesidad de gestionar los posibles cambios de alcance y problemas en la implementación del piloto, se hace necesaria el seguimiento de una metodología de gestión ágil como Scrum, mediante la cual se seguirá un modelo de entregas al finalizar cada sprint de 2 semanas que permitirá minimizar posibles desviaciones de alcance y un mayor control sobre la calidad del producto final.

1.5.- PLANIFICACIÓN DEL TRABAJO.

Se adjunta a continuación la descomposición de tareas asociadas al proyecto y su cronología.

Grupo	Tarea	Comienzo	Fin
Definición del Proyecto		mie 21/02/18	sáb 03/03/18
	Búsqueda de información	mie 21/02/18	vie 02/03/18
	Videollamada	sáb 03/03/18	sáb 03/03/18
Plan de Proyecto		lun 05/03/18	vie 09/03/18
	Análisis de situación	lun 05/03/18	mar 06/03/18
	Estudio fuentes de datos y herramientas	mie 07/03/18	mie 07/03/18
	Plan de Trabajo	jue 08/03/18	vie 09/03/18
Evaluación Continua		lun 12/03/18	lun 12/03/18
	Entrega de PEC 1	lun 12/03/18	lun 12/03/18
Análisis		mar 13/03/18	vie 30/03/18
	Herramientas BI en Microsoft Azure	mar 13/03/18	mie 14/03/18
	Definición de indicadores	jue 15/03/18	mar 20/03/18
	Proveedores de datos de precios y climatología	mie 21/03/18	jue 22/03/18
	Modelado de datos	vie 23/03/18	mar 27/03/18
	Definición de contenido de informes	mie 28/03/18	vie 30/03/18

Grupo	Tarea	Comienzo	Fin
Diseño		lun 02/04/18	jue 19/04/18
	Estructura de datos	lun 02/04/18	mie 04/04/18
	Descarga de datos actualizados	jue 05/04/18	jue 05/04/18
	Extracción de información	vie 06/04/18	lun 09/04/18
	Limpieza y Transformación de información	mar 10/04/18	jue 12/04/18
	Maquetación de Informes	vie 13/04/18	jue 19/04/18
Evaluación Continua		lun 16/04/18	lun 16/04/18
	Entrega de PEC 2	lun 16/04/18	lun 16/04/18
Implementación		vie 20/04/18	jue 17/05/18
	Carga de datos en Microsoft Azure	vie 20/04/18	mie 25/04/18
	Configuración de herramientas BI	jue 26/04/18	lun 30/04/18
	Transformación de datos (ETL)	mar 01/05/18	lun 07/05/18
	Implementación de Informes	mar 08/05/18	jue 17/05/18
Evaluación Continua		lun 21/05/18	lun 21/05/18
	Entrega de PEC 3	lun 21/05/18	lun 21/05/18
Cierre del Proyecto		vie 18/05/18	vie 15/06/18
	Memoria	vie 18/05/18	mar 05/06/18
	Auto Informe de competencias transversales	mie 06/06/18	jue 07/06/18
	Presentación virtual	vie 08/06/18	vie 15/06/18
Evaluación Continua		lun 18/06/18	lun 18/06/18
	Entrega de PEC 4 / Producto Final	lun 18/06/18	lun 18/06/18
Gestión de Cambios		vie 16/03/18	mar 26/06/18
Gestión de Proyecto		vie 16/03/18	mar 26/06/18

1.6.- BREVE RESUMEN DE PRODUCTOS OBTENIDOS.

Una vez finalizado el trabajo de fin de grado se obtendrán los siguientes productos principales.

Producto	Descripción
Memoria de Proyecto	Descripción detallada del trabajo realizado durante el desarrollo del presente TFG.
Presentación Virtual	Presentación del trabajo realizado para la consecución del presente TFG desde sus etapas iniciales hasta el prototipo obtenido.
Prototipo	Informe interactivo accesible a través de una dirección web donde analizar la relación clima y precio de alimentos por país y periodo temporal. En la presente memoria se muestran y analizan ejemplos de los cuadros de indicadores más relevantes.

2.- LA INESTABILIDAD DEL PRECIO DE LOS ALIMENTOS BÁSICOS

La inestabilidad del precio de los alimentos en los últimos 20 años genera una preocupación creciente entre los diferentes actores implicados que intentan paliar la situación en la medida de lo posible. El Banco Mundial de Alimentos tiene en marcha desde hace años diferentes programas de ayuda a la agricultura y a los sectores conexos, así como a los productores y consumidores para hacer frente a las diferentes causas de esta inestabilidad en el precio de los alimentos básicos que provocan hambrunas, conflictos sociales y migraciones masivas.

A continuación, se evalúan las causas principales e históricamente más influyentes en la fluctuación del precio de los alimentos básicos, sin embargo, no son las únicas causas ya que los mercados especulativos de cereales, el incremento de la demanda de los países asiáticos o el auge del biocombustible basado en cereales también está impactando sobre estos precios.

2.1.- CAUSAS

Existen numerosas causas que pueden afectar al precio de los alimentos en los diferentes países en vías de desarrollo y en general en todo el planeta. Debido a que estamos en un mundo globalizado, los precios fluctúan en función de cambios en la climatología, malas cosechas debido a pestes o enfermedades, aumento de los costes logísticos ya sea por el precio del carburante, o conflictos laborales, así como inestabilidades políticas y económicas en las diferentes fases de producción de los alimentos básicos.

2.1.1.- METEOROLÓGICAS: TEMPERATURA Y PRECIPITACIONES

El factor meteorológico es una de las causas más influyentes en la inestabilidad de los precios de numerosos alimentos básicos como el trigo, arroz, sorgo, etc. Sin ir más lejos, los efectos de la sequía junto a problemas de producción a causa de diferentes condiciones climáticas están causando problemas en los mercados mundiales. Sin embargo, no es solo la falta de precipitaciones y el incremento de las temperaturas los principales causantes de estos problemas de producción, sino que el incremento de las precipitaciones en ciertas zonas también afecta a números cultivos como ocurrió en Brasil en 2016, cuando según la FAO, los precios del azúcar subieron cerca de un 15 por ciento como consecuencia de las lluvias torrenciales sufridas que afectaron a la cosecha y mermaron los rendimientos productivos¹.

Estos incrementos en las precipitaciones tienen en las emisiones de gases de efecto invernadero un catastrófico aliado debido a que al atrapar el calor y aumentar las temperaturas del aire que absorben más humedad, causan un detrimento de las lluvias, una mayor evaporación del agua de los ríos y lagos, y por tanto una tierra más seca que favorece el incremento de las inundaciones con su consecuente destrozo de las cosechas.

¹ Anónimo. (2016, julio). "El índice de precios de los alimentos de la FAO señala su mayor aumento mensual en cuatro años". Organización de las Naciones Unidas para la Alimentación y la Agricultura [artículo en línea]. <<http://www.fao.org/news/story/es/item/423208/icode>>

La temporada de crecimiento en cualquier cosecha es la parte más crítica por lo que cualquier cambio en el clima afectará este periodo tan importante, es decir, los cultivos no podrán prosperar durante esta etapa si la temporada es muy fría o demasiado caliente, demasiado húmeda o demasiado seca.

En definitiva, que escasee un cultivo concreto afectará a su precio incrementándolo, a la vez que, si una temporada de cultivo es excelente, los precios de los alimentos podrían bajar. Esto tendrá una relación directa con indicadores meteorológicos como: temperatura mensual máxima, mínima o media, precipitaciones mensuales en forma de agua, etc.

2.1.2.- PESTES Y ENFERMEDADES EN LAS COSECHAS

Las cosechas no están exentas de enfermar y por tanto morir o reducir su capacidad de producción afectando a una comunidad local o incluso a todo un país. Existen diferentes elementos como malas hierbas, insectos, virus o bacterias que pueden infectar frutas, vegetales u otras plantas y alimentos durante el periodo de crecimiento, previo a la cosecha o incluso durante el almacenaje o transporte, llegándose a dispersar por todo el mundo.

En las etapas de crecimiento y previo a la cosecha se pueden minimizar los riesgos de infección y realizar tratamientos para el exterminio del agente infeccioso en caso de estar ya presente en el cultivo. Es en este periodo en el que por causas naturales debido al cambio climático se alteran los patrones normales de crecimiento de insectos y enfermedades, generándose estas epidemias que afectan al cultivo y que son poco menos que inevitables, pero si tratables.

Sin embargo, la globalización y el comercio internacional de productos agrícolas básicos de alimentación aumentan los riesgos de que plagas de plantas circulen a través de las diferentes fronteras y regiones en el contexto del comercio mundial, causando un serio problema de seguridad alimentaria para millones de agricultores y por tanto afectando seriamente al precio del alimento a todos los niveles.

Los indicadores más importantes para esta causa estarían relacionados con el porcentaje de plagas por cultivo o toneladas de productos contaminados detectados por frontera comercial.

2.1.3.- COSTES LOGÍSTICOS

La logística y el transporte es un coste que no se puede evitar y afecta siempre al precio final de los alimentos tanto en periodos de precios bajos como en periodos cuando el precio es mayor. Según el Programa Mundial de alimentos, entre el 18 y el 32 por ciento del precio de entrega del grano importado viene de los costes de logística, estando en niveles similares los costes para otros alimentos como vegetales o frutas. Esto representa entre el 15 y el 25 por ciento del PIB de estos países.

Esto es debido en gran parte a la ineficiente, costosa y lenta manera de cargar y descargar cereales en diferentes países subdesarrollados o en vías de desarrollo como consecuencia del uso de maquinaria obsoleta, y al coste de los carburantes usados en el transporte y la maquinaria que carga y descarga los productos.

Además, habría que añadir los aranceles en el precio del intercambio comercial, que a pesar de los diferentes acuerdos bilaterales y regionales de libre comercio en todo el mundo, existen costos como el flete marítimo que puede aumentar el precio final al consumidor hasta un 30 o 40 por ciento.

Si bien es cierto que la ineficiencia en la carga y descarga es difícilmente medible de forma sistemática en los diferentes países y mercados mundiales, sí que podemos usar indicadores basados en precios de fletes marinos y aranceles en el precio del intercambio comercial, junto al precio de los combustibles.

2.1.4.- COSTES LABORALES

Agricultores, recolectores, procesadores, minoristas, etc., son parte fundamental para que el sistema agroalimentario funcione y, por tanto, se hace necesaria mucha mano de obra que proporcione buenas oportunidades laborales en los diferentes países involucrados. Sin embargo, el coste de la mano de obra tendrá impacto en los precios en función de los salarios asignados.

Para medir esta causa basaríamos los indicadores en el precio por hora y por perfil en función del país.

2.1.5.- SITUACIONES POLÍTICAS Y ECONÓMICAS

El sistema agroalimentario se sitúa como clave en muchas economías de países desarrollados por lo que existe cierto grado de proteccionismo alrededor de agricultores domésticos y el contexto asociado, en las importaciones y exportaciones de alimentos básicos. Posibles conflictos políticos causarán fluctuaciones en los precios de los alimentos afectando a países en vías de desarrollo o subdesarrollados como consecuencia de su debilidad a la hora de influir y contrarrestar esta inestabilidad. Además, las revueltas sociales y conflictos armados en países dependientes de la producción local de cereales verán incrementado el precio de los alimentos básicos debido a la destrucción de los campos de cultivo, el secuestro de la producción o la imposibilidad de comprar producto en los mercados externos como consecuencia de la devaluación de la moneda local.

Encontrar indicadores para esta causa se antoja complicado, sin embargo, existen posibles indicadores asociados a la economía de los países como el PIB o indicadores asociados a revueltas y fenómenos violentos proporcionados por el proyecto GDELT.

2.2.- CONSECUENCIAS

Como consecuencia de la fluctuación de los precios de los alimentos básicos como los cereales, tan importantes en la alimentación de las personas de los países pobres o en vías de desarrollo, se generan tensiones sociales a nivel local y regional incitados por el instinto de supervivencia del propio ser humano.

2.2.1.- MIGRACIONES

Una de las principales consecuencias de la fluctuación de los precios de los alimentos básicos son las migraciones forzadas como consecuencia de la incapacidad de conseguir alimento. Este tipo de flujos migratorios son continuos y se ven agravados por periodos de hambruna o conflictos armados latentes en los diferentes países y regiones de África, Asia o América del sur.

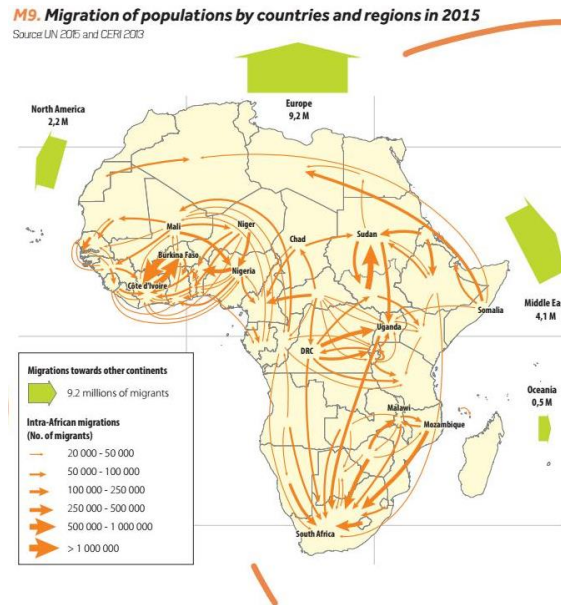


ILUSTRACIÓN 2- MIGRACIÓN POBLACIONAL EN ÁFRICA EN 2015

(FUENTE: [HTTPS://ISSUU.COM/CIRAD/DOCS/ATLAS_NEPAD_EN_FINAL-OPTIMISE](https://issuu.com/cirad/docs/atlas_nepad_en_final-optimize))

2.2.2.- HAMBRUNAS

Existe un grave problema de transmisión del incremento del precio internacional de los alimentos básicos a los mercados locales en países con grandes núcleos de población urbana y rural pobre que dependen de la importación de productos, que tiene que gastar la mayor parte de sus ingresos en comida. Como consecuencia se producen problemas de abastecimiento y muchas personas no pueden tener acceso a una alimentación normal.

2.3.- PLANES DE CONTINGENCIA

Definidas las causas y expuestas las consecuencias se constata como el mundo se enfrenta a un gran problema humanitario a todos los niveles. Las grandes sequías que alteran la producción mundial de cereales y otros alimentos básicos² usados para la alimentación primordial de

² Anónimo. (2017, febrero). "La sequía provoca una fuerte subida de los precios alimentarios en África oriental". Organización de las Naciones Unidas para la Alimentación y la Agricultura. [artículo en línea]. <<http://www.fao.org/news/story/es/item/470646/icode>>

millones de personas y que son recurrentes en diferentes zonas del mundo, así como las graves inundaciones que también causan enormes daños a los cultivos con la consecuente pérdida de toneladas de cereales son los factores más importantes de este aumento de los niveles de precios y la volatilidad asociada.

Como plan de contingencia a corto plazo la comunidad internacional y en concreto el WFP inyecta ingentes cantidades de dinero en paliar los problemas derivados de la climatología para evitar las consecuencias de hambruna y migraciones descritas anteriormente. Sin embargo, el WFP destaca que por cada 10 por ciento de aumento en la cesta de alimentos básicos significa un aumento de 200 millones de dólares anuales en asistencia alimentaria.

El riesgo por tanto es muy grande y más aún con los 80 millones de nuevas bocas que alimentar al año. Se hace necesario buscar la forma de optimizar y predecir cómo, cuándo y dónde se producirán estos fenómenos para prever las ayudas económicas necesarias para evitar el hambre en las zonas más sensibles a estos aumentos de precio, y no actuar tarde.

3.- ANÁLISIS Y PREDICCIÓN DEL PRECIO DE ALIMENTOS MEDIANTE BI

El problema a solucionar en resumen es dotar al Banco Mundial, WFP y diferentes organizaciones no gubernamentales, de una herramienta que les permita analizar como todos y cada uno de los indicadores de las diferentes causas que alteran al alza el precio de los alimentos básicos se interrelacionan para poder anticiparse a la hora de crear planes de contingencia y acción que limiten o palien las consecuencias en los países en vías de desarrollo y pobres, concretamente en las diferentes zonas locales afectadas con mayor frecuencia.

Por tanto, se propone la creación de un sistema de análisis de grandes volúmenes de datos heterogéneos basados en los indicadores de las diferentes causas de volatilidad de los precios para su estudio mediante informes predefinidos y sistemas de predicción de la evolución de los precios de los alimentos básicos.

3.1.- COMPARATIVA DE HERRAMIENTAS DEL MERCADO

Para desarrollar este tipo de aplicaciones donde la capacidad de computación de la enorme cantidad de datos es vital para reducir los tiempos de análisis y proporcionar resultados en tiempos razonables, se hace necesario el uso de la Infraestructura como servicio en la nube. Hoy en día, y como se puede ver en la siguiente imagen, los principales actores que lideran el mercado de este tipo de servicios son *Microsoft Azure*, *Amazon Web Services* y *Google Cloud Platform*.

Figure 1. Magic Quadrant for Cloud Infrastructure as a Service, Worldwide



ILUSTRACIÓN 3 – MAGIC QUADRANT FOR CLOUD IAAS, WORLDWIDE. (FUENTE: GARTNER ABRIL 2018)

Para la implementación de la solución completa que se describe en esta memoria, la elección de un proveedor concreto de servicios Cloud IaaS, dependerá del tipo de trabajos que finalmente se vayan a ejecutar en la nube. Cabe destacar, que empresas que implementan este tipo de soluciones pueden hacer uso de diferentes proveedores Cloud para ejecutar diferentes partes de su operativa en función de las necesidades y características propias de los proveedores en la nube.

Los tres proveedores están apostando fuertemente por mejorar sus servicios en la nube por lo que en todos los casos disponen de diferentes herramientas para solucionar problemas específicos actuales. Por ejemplo, los tres disponen de soporte para clusters de tipo Hadoop: AWS con *Elastic Map Reduce*, Azure con *HDInsight* y Google con *DataProc*, además de soporte para servicios de *Machine Learning*, *IoT* o *Dockers* por ejemplo.

Además de que los tres proveedores disponen de soporte para las últimas tecnologías de análisis de datos, inteligencia artificial y aprendizaje automatizado, también son líderes en proporcionar servicios tradicionales de computación, almacenamiento, bases de datos y redes. Es por ello, que las tres plataformas compiten con servicios muy parecidos, donde el precio de estos servicios también se asemeja en términos generales. Aunque no es menos cierto, que la comparativa en términos de precio, es más complicada debido entre otras cosas a los diferentes planes de precio, descuentos y actualizaciones.

Como consecuencia de lo anterior, una comparativa al uso de cada uno de los servicios que ofrecen estos proveedores no aporta valor a la decisión de optar por uno u otro, por lo que a continuación se describirá a alto nivel cuales son los pros y contras de cada plataforma:

Google Cloud Platform

Google fue una de las pioneras en apostar por la innovación de empresas nativas en la nube y tiene como punto fuerte la amplia experiencia en aprendizaje automático e inteligencia artificial utilizada durante los últimos años de forma interna en la compañía usando su herramienta *TensorFlow* como referente en la que basar su estrategia. Además, cabe destacar como *Google* para evitar ser calificado de un mero imitador de AWS o *Azure*, ha sabido innovar en el área del análisis de datos con *BigQuery*, en bases de datos distribuidas con *Cloud Spanner* o en el aprendizaje automatizado.

Por el contrario, *Google* no ha podido consolidarse aún como socio estratégico en el mundo empresarial relegándose a pruebas en pequeños proyectos innovadores de grandes empresas, a pesar de contar con una buena posición en la comunidad *open source*.

Amazon Web Services

La principal ventaja de este proveedor es la nutrida oferta de servicios que ofrece AWS que construyó de forma pionera a principios de 2006 y que consiguió atraer a responsables del área de sistemas de la información de las organizaciones y a desarrolladores. Además, cabe destacar la amplia red de socios estratégicos que se integran perfectamente con los servicios ofertados por AWS y los servicios de software de terceros.

Por el contrario, el concepto de nube híbrida sigue siendo un punto en contra para apostar por *Amazon Web Services*. Actualmente, diferentes organizaciones tienden a mantener sus datos confidenciales en sus propios centros de datos y conectar sus infraestructuras locales con su infraestructura en la nube. Es aquí donde AWS no alcanza a dar el servicio que buscan estas empresas. Es reseñable como la enorme variedad de servicios que ofrece AWS llega a ser contraproducente debido a cómo los consumidores experimentan cierta dificultad en entender qué características de las que se ofrecen encajarían mejor en su modelo de negocio. Y habría que añadir que ciertas empresas ven AWS como un proveedor complejo para su administración.

Microsoft Azure

La principal ventaja de *Microsoft Azure* es la penetración en el mundo empresarial que dispone Microsoft facilita gracias a su sólida base en las empresas, ese paso natural de la infraestructura local a la nube. Al compartir fabricante, la relación entre los sistemas clave como *Windows Server* o *Active Directory* y *Microsoft Azure* es de entendimiento total. Además, se puede saltar cómo *Microsoft Azure* destaca entre los otros dos proveedores en la implantación de nubes híbridas, lo que facilita la transición de organizaciones con reticencia a mover a la nube sus datos sensibles, manteniéndolos en sus centros de datos locales.

Otro de sus puntos fuertes es la facilidad de administración, que supone una ventaja competitiva con respecto a AWS o *Google Cloud* a la hora de facilitar la adopción por parte de pequeñas empresas con pocos recursos.

Por el contrario, *Azure* es más restrictivo en la posibilidad de proporcionar a los usuarios el uso de otras plataformas que no estén basadas en tecnologías *Microsoft*, aunque en el último año *Microsoft Azure* ha hecho un gran avance al dar soporte a sistemas operativos Linux y está dispuesto a adoptar plataformas de código abierto.

3.2.- HERRAMIENTAS DE ANÁLISIS EN MICROSOFT AZURE

Tradicionalmente el proceso de análisis de negocio para grandes cantidades de datos seguía unos pasos predeterminados. Se comenzaba identificando en los requisitos de usuario los datos e informes deseados para el proceso de negocio a analizar, para posteriormente definir el correspondiente esquema de base de datos y las consultas a realizar en la misma. Se buscarían las fuentes de datos internas y externas necesarias para la ingesta en el sistema y se crearían los procesos ETL necesarios para la extracción de los datos y su transformación en el esquema definido previamente. Finalmente se crearían los informes necesarios definidos en las etapas iniciales y se procedería a analizar los datos.

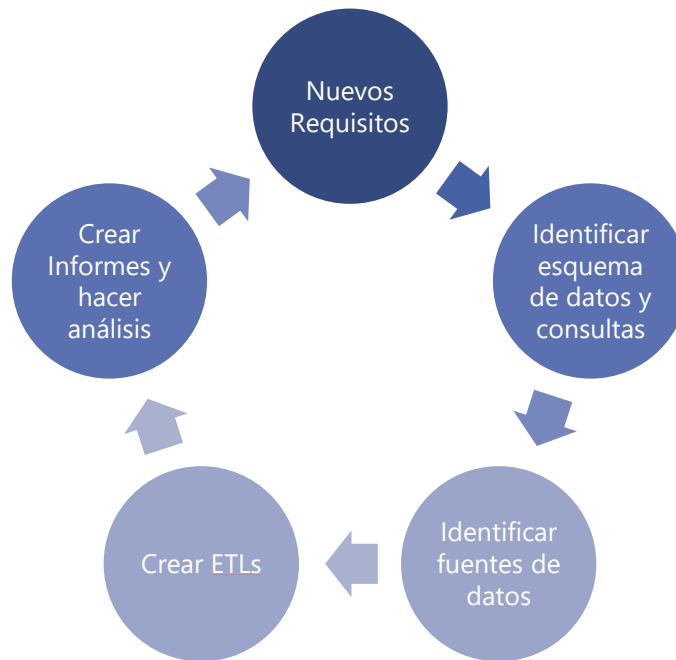


ILUSTRACIÓN 4 - PROCESO TRADICIONAL DE ANÁLISIS DE NEGOCIO (FUENTE: PROPIA)

Este proceso perfectamente funcional y que generalmente se utilizaban en plataformas basadas en tecnología Microsoft como *SQL Server Integration Services* tienen la principal limitación en cómo se tratan los datos, ya que toda información que no es requerida se descarta o archiva quedando fuera de los procesos analíticos posteriores.

Sin embargo, para el caso que nos ocupa toda la información tiene valor potencial y las fuentes de datos son muy heterogéneas, por lo que la opción tecnológica actual pasa por un sistema de Big Data que permita mediante la carga de datos heterogéneos alimentar los procesos de análisis para su tratamiento y presentación en los informes predeterminados.

Basándonos en tecnologías Microsoft, y haciendo uso de la capacidad de proceso y almacenamiento que proporciona la nube de Microsoft Azure, usaremos *Microsoft Azure Data Lake Store* para el almacenamiento y *Microsoft Azure Data Lake Analytics* para el proceso de análisis de la información.

3.3.- FUENTES DE DATOS

El análisis de la evolución de los precios de los alimentos se tiene que basar en multitud de factores e indicadores que causan esta fluctuación y aumento. Para ello, se hace necesario un repositorio en la nube sin límite de tamaño y que permita alojar datos no estructurados, semiestructurados y estructurados. *Azure Data Lake Store* es la pieza clave en este sistema que servirá para alojar toda la información presente y futura para el análisis de forma masiva.

En función de las causas y sus indicadores que hacen variar los precios de los alimentos, podemos encontrar distintas fuentes de datos públicos por parte de diferentes organismos a nivel mundial:

PRECIO DE ALIMENTOS

La parte fundamental con los precios mensuales de los diferentes alimentos se puede obtener del WFP que dispone de esta información gracias a las diferentes fuentes de cada país.

CLIMATOLOGÍA

Para obtener el histórico de datos meteorológicos mundiales y futuros, se propone el uso de la base de datos pública del NOAA que proporciona diferentes indicadores que hacen referencia a temperaturas, precipitaciones, días de sol, vientos, etc.

COSTES LOGÍSTICOS

Para los costes logísticos se deberán usar diferentes fuentes de datos. Una de ellas la proporciona el Banco Mundial con su base de datos *de índice de rendimiento de logística a nivel mundial*, que contiene información relativa a la eficiencia del transporte aéreo, marítimo y terrestre, tasas, aranceles, etc. Además, se deberán incluir otras fuentes de datos del mismo proveedor relativas al coste del combustible por litro y país, así como datos relativos al coste del flete marítimo.

COSTES LABORALES

Gracias al programa ILC de la oficina de estadísticas laborales de los Estados Unidos de América, es posible acceder a una base de datos a nivel mundial de los costes laborales por país con indicadores como el coste por hora, PIB per cápita y por hora, productividad, etc.

SITUACIONES POLÍTICAS Y ECONÓMICAS

El proyecto GDELT proporciona información sobre eventos observados en diferentes noticias los cuales son analizados y categorizados en función del país, fechas, tipo de evento, etc.

Mediante procesos que corren en *Azure Data Factory*, se obtendrá la información directamente desde las fuentes donde sea posible a través de FTP y cuando no sea posible se proporcionarán contenedores en la nube donde alojar los ficheros con la información para ser consumida y clasificada en el repositorio común en *Azure Data Lake Store*.

3.4.- ANÁLISIS DE DATOS HISTÓRICOS

Debido a la gran cantidad de datos a analizar se propone el uso de *Azure Data Lake Analytics* como servicio de análisis en la nube donde ejecutar el procesamiento y la transformación de grandes cantidades de datos en paralelo, con las siguientes ventajas:

- No necesita administrar una infraestructura compleja ya que no hay servidores, máquinas virtuales ni clústeres.
- Permite procesar datos a petición con el consiguiente ahorro de costes de proceso.
- Permite un escalado rápido de las unidades de análisis facilitando el crecimiento del sistema en función de los datos a analizar.

La mayor complejidad estriba en la unificación de los datos maestros de los países en función de las fuentes de datos.

3.5.- ANÁLISIS PREDICTIVO

Para añadir valor a la solución propuesta haciendo uso de las tecnologías actuales y que forman parte del ecosistema en *Microsoft Azure*, se propone la creación de soluciones de análisis predictivo que permitan en la medida de lo posible anticiparse a fluctuaciones dramáticas de precios de alimentos para minimizar con acciones concretas las consecuencias explicadas en puntos anteriores.

Se propone el uso de *Microsoft Azure Machine Learning Studio* ya que permite crear, implementar y compartir fácilmente soluciones de análisis predictivo usando los datos y modelos ya creados para esta solución propuesta. Los resultados de estas soluciones podrían ser consultados a través de un servicio web y basado en los indicadores definidos que causan estos cambios al alza de los precios de los alimentos básicos.

La solución propuesta para el análisis predictivo tendrá como parámetros de entrada los valores de los indicadores de las diferentes causas:

Climatología

Temperatura Mínima, Temperatura Máxima, Temperatura media y Precipitación en forma de lluvia.

Costes Logísticos

Precio combustible por litro, aranceles, precio de flete marítimo.

Costes Laborales

Precio por hora.

Situaciones Políticas y Económicas

Número de eventos desestabilizadores.

Estas entradas junto al país, permitirá al modelo de análisis predictivo dar una tendencia de precio para los diferentes productos básicos del país por el que se pregunta.

3.6.- REPRESENTACIÓN DE LA INFORMACIÓN

Dentro de las posibilidades existentes basadas en tecnologías Microsoft para la visualización de todos los datos analizados en esta solución, se opta por el uso de *Microsoft Power BI*. Esta herramienta permite la creación de informes y cuadros de mando conectados con las fuentes de análisis de datos creadas y descritas en los puntos anteriores para estudiar de forma detallada la información sobre las tendencias y los indicadores clave de rendimiento.

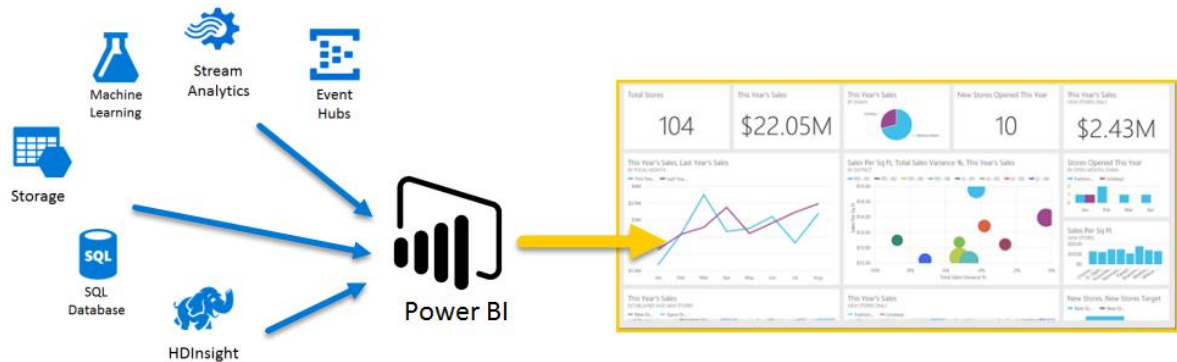


ILUSTRACIÓN 5 - MICROSOFT AZURE Y POWER BI

(FUENTE: [HTTPS://DOCS.MICROSOFT.COM/ES-ES/POWER-BI/SERVICE-AZURE-AND-POWER-BI](https://docs.microsoft.com/es-es/power-bi/service-azure-and-power-bi))

4.- IMPLEMENTACIÓN DEL PROTOTIPO DE ANALISIS DE PRECIOS EN FUNCIÓN DE FACTORES METEOROLÓGICOS

Como consecuencia del tiempo disponible a la hora de desarrollar esta solución para este trabajo académico de final de grado, se propone trabajar en un subconjunto de la funcionalidad de la solución propuesta, en la que se analizará la posible correlación entre el precio de los alimentos históricos y variables climatológicas en los diferentes países para los que se dispone de información, a través de un informe interactivo.

Se ha optado por basar el prototipo en tecnologías Microsoft mediante el uso de Microsoft Azure. Una de las razones de mayor peso, es la mejor integración de herramientas IA y *Machine Learning*, que permitirán evolucionar el prototipo a un sistema de predicción más robusto y totalmente integrable con la base del prototipo a realizar. Otra razón es el precio asociado a las necesidades del sistema de análisis que resulta sencillo calcularlo mediante las herramientas de estimación que proporciona el proveedor y, por ende, se presupuesta un menor coste, además se ha tenido en cuenta la facilidad de las diferentes herramientas para los estudiantes que proporciona Microsoft. Finalmente, teniendo en cuenta la amplia experiencia profesional en tecnologías Microsoft del autor del presente TFG en herramientas similares y lenguajes de la compañía, se estima un ahorro muy importante en tiempo de autoformación que resultará de suma importancia para concluir con garantías de éxito el trabajo de fin de grado.

4.1.- ANALISIS FUNCIONAL

La arquitectura del prototipo a desarrollar se basará en un sistema con varias partes diferenciadas en *Microsoft Azure* en el que se podrá consultar la información analizada a través de una dirección web.

El sistema será provisionado con ficheros de datos no estructurados (CSV y ficheros de texto) de las diferentes fuentes seleccionadas: climatología y precios de alimentos. Estos ficheros serán tratados mediante la ejecución puntual para este prototipo de diferentes procesos de análisis para finalmente ser alojados en una base de datos relacional que permitirá ser consumidos por el sistema de informes.

La ejecución del proceso de análisis será puntual, pero es configurable para ser ejecutado de forma recurrente para la actualización de los datos.

A continuación, se muestra una imagen que ilustra cual es la arquitectura que será implementada en el presente proyecto.

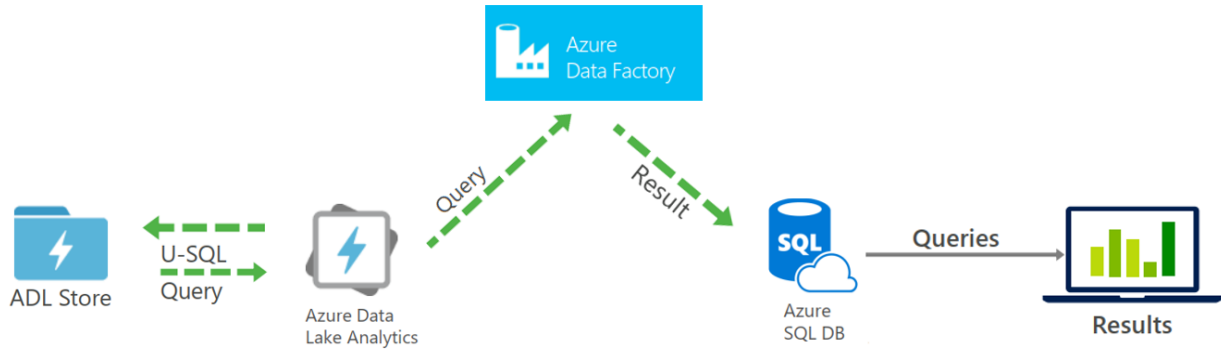


ILUSTRACIÓN 6- ARQUITECTURA PROTOTIPO (FUENTE: PROPIA)

4.1.1.- ESTUDIO DE FUENTES DE DATOS

Como se ha descrito en apartados anteriores para este prototipo se utilizan datos de fuentes reconocidas como el WFP y el GHCN que depende del NOAA. Cada proveedor de datos proporciona la información en diferentes formatos, pero siempre basados en ficheros de texto o CSVs.

WFP

La última versión disponible con información de diferentes países en vías de desarrollo y subdesarrollados se puede encontrar con fecha 5 de diciembre de 2017, del que se extraerá un listado de países con información de precios, un listado de monedas, un listado de mercados desde donde se han recogido datos de precios, un listado de tipos de mercado donde se han obtenido estos datos, un listado de medidas, un listado de alimentos/mercancía para los que hay precio, la fuente que proporciona el dato, el mes y el año donde ha sido registrada la información.

La estructura de este fichero CSV es la siguiente:

Campo	Descripción
adm0_id	Identificador de país
adm0_name	Nombre del país
adm1_id	Identificador de localidad
adm1_name	Nombre de localidad
mkt_id	Identificador de mercado
mkt_name	Nombre de mercado
cm_id	Identificador de mercancía
cm_name	Nombre de la mercancía
cur_id	Identificador de moneda

Campo	Descripción
cur_name	Nombre de moneda
pt_id	Identificador del tipo de mercado
pt_name	Nombre del tipo de mercado (Al por menor, por mayor, granja, productor)
um_id	Identificador del tipo de medida de la mercancía registrada
um_name	Nombre de la medida usada para la mercancía registrada
mp_month	Mes de registro
mp_year	Año de registro
mp_commoditysource	Fuente que proporciona la información del precio

GHCN

NOOA al ser un organismo público proporciona la información del GHCN de forma más estructurada. Además, posibilita el acceso continuado a información actualizada de las diferentes estaciones meteorológicas desde donde se obtiene la información. Para este prototipo se ha obtenido información actualizada para el periodo comprendido entre enero del año 2000 hasta enero del año 2018.

GHCN proporciona varios ficheros de texto con información relativa a las tablas maestras de los datos que proporciona, así como la información propiamente dicha en ficheros CSV (uno por año).

Las tablas maestras que usaremos en este prototipo serán las relativas a países y estaciones meteorológicas, con un formato específico en cada fichero con columnas de tamaño fijo.

ghcnd-countries.txt

De este fichero se extraerá un listado completo de países para los que se tiene información climatológica y que será cruzado con el listado de países obtenido de la fuente de datos de los precios de los alimentos. El fichero sigue la siguiente estructura:

Campo	Columna	Descripción
CODE	1-2	Código de identificación del país
NAME	4-50	Nombre del país

ghcnd-stations.txt

Este fichero contiene un listado de estaciones meteorológicas por país con información adicional que será extraída para poder relacionar las mediciones de cada una de estas estaciones con el país en el que ha sido registrado la información. El fichero sigue la siguiente estructura:

Campo	Columna	Descripción
ID	1-11	Identificador de la estación con un formato donde los dos primeros caracteres son el código de país
LATITUDE	13-20	Latitud (en grados decimales)
LONGITUDE	22-30	Longitud (en grados decimales)
ELEVATION	32-37	Elevación (en metros)
STATE	39-40	Código postal estadounidense para estaciones de EEUU
NAME	42-71	Nombre de la estación
GSN FLAG	73-75	Indica si la estación pertenece a la red de superficie GCOS
HCN/CRN FLAG	77-79	Indica si la estación pertenece a la red de estaciones históricas de EEUU
WMO ID	81-85	Número de la World Meteorological Organization

Finalmente, cada fichero con información meteorológica por estación donde se ha registrado el dato está nombrado por el año al que pertenecen las mediciones y sigue el siguiente formato.

Campo	Descripción
ID	Identificador de la medición
YEAR/MONTH/DAY	Fecha de la medición en formato YYYYMMDD
ELEMENT	Tipo de elemento a medir: <ul style="list-style-type: none"> • PREP = Precipitaciones en milímetros • TAVG = Temperatura media en grados • TMIN = Temperatura mínima en grados • TMAX = Temperatura máxima en grados
DATA VALUE	Valor del elemento
M-FLAG	Identifica más información de medición
Q-FLAG	Identifica más información sobre la calidad de la medición
S-FLAG	Identifica más información sobre la fuente de la medición
OBS-TIME	Hora y minutos de la medición

4.1.2.- ESTUDIO DE HERRAMIENTAS BUSINESS INTELLIGENCE EN MICROSOFT AZURE

Para la implementación del prototipo de análisis basado en Microsoft Azure en la nube, se utilizará parte del potencial proporcionado por las siguientes herramientas.

Data Lake Store

Es un repositorio de gran escala para cargas de trabajo de análisis de macro datos. Es decir, permitirá capturar cualquier tipo de dato, de cualquier tamaño y velocidad de ingestión, desde un único lugar para realizar análisis. Además, permite acceder al repositorio usando Hadoop y las API REST compatibles.

ADLS está diseñado y optimizado para el almacenamiento y análisis de grandes cantidades de datos heterogéneos, incluyendo las capacidades empresariales esperadas para una herramienta de esta envergadura: seguridad, administración centralizada, escalabilidad, disponibilidad y confiabilidad.

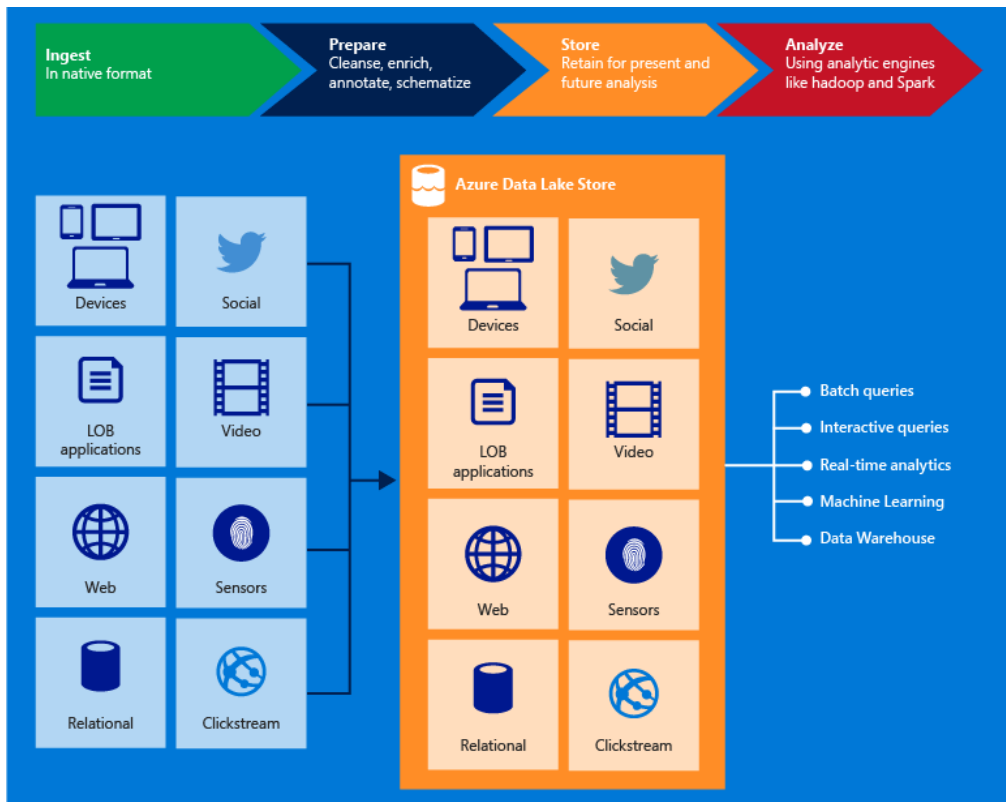


ILUSTRACIÓN 7- AZURE DATA LAKE. (FUENTE: DOCS.MICROSOFT.COM)

Data Lake Analytics

Es un servicio bajo demanda de análisis de datos que simplifica mucho los trabajos Big Data, ya que permite escribir consultas y transformar datos para buscar información sensible, en lugar de desplegar, configurar y ajustar hardware costoso. Los trabajos de análisis pueden ser de cualquier tamaño y complejidad, pagándose únicamente por el servicio cuando el trabajo de análisis está en ejecución, lo que reduce el coste de uso y facilita su adopción en empresas de cualquier tamaño.

La ventaja principal de ADLA, radica en el escalado dinámico que, en función del tamaño de los datos a analizar, ya sean gigabytes, terabytes o petabytes, la herramienta provisionará los recursos de computación necesarios para procesar la información de forma transparente y sin la necesidad de reescribir código.

Además, dispone de una integración total con Visual Studio, por lo que permite utilizar las herramientas con las que los desarrolladores están familiarizados para ejecutar, depurar y optimizar los scripts de análisis. Y todo esto, usando un nuevo lenguaje U-SQL que está basado en SQL con expresiones de C#, lo que hace que la curva de aprendizaje para desarrolladores de SQL y .NET sea corta debido a que podrán usar conocimientos que ya tienen.

Data Factory

Es un servicio de integración de datos basado en la nube que permite la creación de flujos de trabajo orientados a datos heterogéneos con el fin de orquestar y automatizar tanto el movimiento de información como su transformación desde distintas fuentes de datos, mediante el uso de diferentes servicios de proceso para Big Data como *Azure HDInsight Hadoop, Spark, Azure Data Lake Analytics o Azure Machine Learning*, y finalmente publicando los resultados en distintos almacenes de datos para ser consumidos por aplicaciones de BI.



ILUSTRACIÓN 8- FLUJO DE TRABAJO ORIENTADO A DATOS EN DATA FACTORY. (FUENTE: DOCS.MICROSOFT.COM)

Data Factory permite el uso de diferentes tipos de desencadenadores que inician la ejecución del flujo de trabajo orientado a datos en función de diferentes tipos de eventos.

Azure SQL Database

SQL Database es un servicio de base de datos relacional administrado en la nube que facilita el uso de diferentes tipos de estructuras de datos relacionales, JSON y XML. La principal ventaja de este servicio en la nube es la escalabilidad dinámica que permite un incremento del rendimiento de la base de datos en función de la necesidad del negocio. Además, permite que la administración de la infraestructura necesaria para bases de datos SQL Server desaparezca y sea transparente ya que es Microsoft quien se encarga de la aplicación de revisiones y actualizaciones del motor de base datos en el que se basa *SQL Database*.

Cabe destacar la escalabilidad dinámica que se ofrece por defecto para este tipo de bases de datos que permite evitar los tiempos de inactividad. Además, dispone de servicios de valor añadido aplicables a optimizar de forma automática e inteligente la base de datos y aplicar medidas de seguridad avanzadas, reduciendo el tiempo de administración de dicha base de

datos y permitiendo poner el foco en el desarrollo de las aplicaciones por lo que se reduce el tiempo de publicación de aplicaciones en el mercado.

Microsoft Power BI

Power BI es un conjunto de aplicaciones de análisis de datos de negocio que permite analizar datos de diferentes fuentes y compartir información sensible relativa a los datos procesados. Permite, mediante el uso de paneles de visualización de diferentes tipos, ofrecer una vista general de las diferentes dimensiones de las que se componen los datos de negocio analizados, facilitando la actualización en tiempo real de la información y siendo accesible desde diferentes dispositivos.

Cualquier analista de datos puede construir informes complejos mediante el uso de datos de bases de datos, archivos y servicios web diversos con diferentes herramientas visuales que facilitarán la comprensión de grandes cantidades de datos.

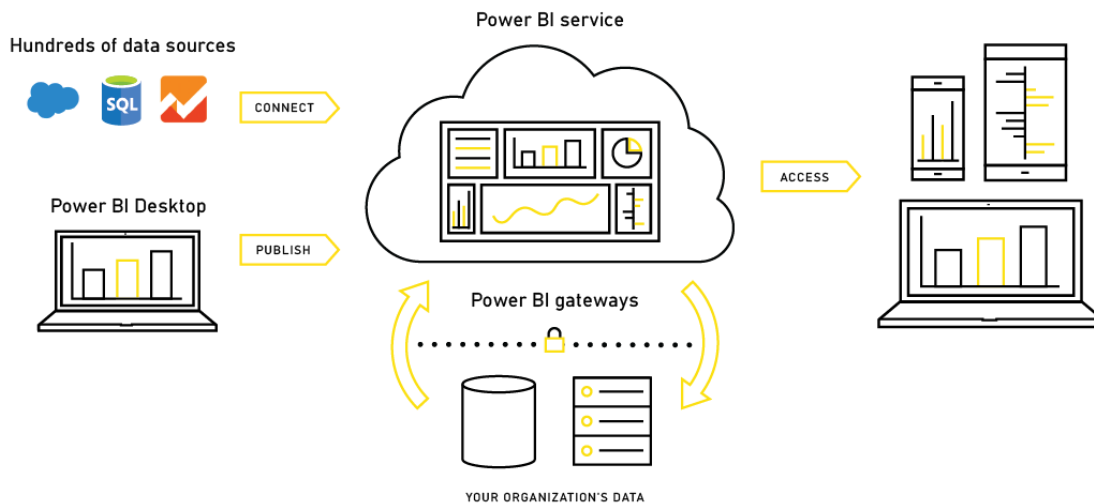


ILUSTRACIÓN 9- RESUMEN DEL SERVICIO POWER BI. (FUENTE: POWERBI.MICROSOFT.COM)

4.1.3.- DEFINICIÓN DE INDICADORES

Encontrar una causa-efecto entre las variaciones de precipitaciones y temperatura de un área con el precio de los alimentos básicos no es tarea sencilla debido a la multitud de factores que pueden afectar al precio, como ya se ha explicado en anteriores apartados. Sin embargo, es posible realizar una aproximación mediante una serie de indicadores, a la respuesta que se busca con este prototipo, ¿Afectan los factores climatológicos al precio de los alimentos básicos en países en subdesarrollados o en vías de desarrollo?

Relación de precio y precipitaciones por país, alimento y periodo dado.

Estudiar las diferencias del precio para un alimento en función del aumento o de la falta de precipitaciones en un periodo, nos dará pistas de cómo la temporalidad de las precipitaciones en función de las estaciones de lluvia afecta o no al precio del alimento en el país.

Relación de precio y temperaturas (medias, mínimas y máximas) por país, alimento y periodo dado.

Al igual que el indicador anterior, es posible estudiar las diferencias y evolución del precio de un alimento en un periodo dado. En función del alimento podremos estudiar cómo temperaturas extremadamente bajas o extremadamente altas tienen impacto sobre el precio de alimento.

Desviación estándar muestral de la temperatura y precipitaciones medias en función del precio para un periodo dado.

Con este indicador se pretende el estudio de la dispersión de los datos relativos a las temperaturas y precipitaciones que afectan al precio, y si existe desviaciones significativas de la media en un periodo dado.

Evolución de precios medios, temperatura media y precipitaciones medias en un periodo dado.

Tres indicadores distintos para el estudio de la evolución de la temperatura media y su tendencia a lo largo del tiempo. Un periodo de tiempo amplio permite ver mejor junto al resto de indicadores de evolución de temperatura y precipitación, cuál es la tendencia de los precios.

Simetría de distribución de los datos relativos a precios medios por alimento en un periodo.

Estos indicadores facilitan el estudio de la dispersión de los puntos con la mediana, los percentiles y los valores mínimos y máximos. Permitirá detectar si la distribución es simétrica o no, así como la posible presencia de valores atípicos.

4.1.4.- DEFINICIÓN DE INFORME

Para facilitar la interacción se toma la decisión de añadir filtros dinámicos al informe mediante los cuales se podrá acotar la información a analizar por país, periodo temporal y producto. Esto facilitará en gran medida el análisis de la información de forma visual debido a la gran cantidad de datos que se tratan en este prototipo.

Microsoft Power BI facilita la visualización de datos mediante diferentes tipos de visualizaciones de forma que una buena combinación de gráficos proporcionará información relevante a primera vista.

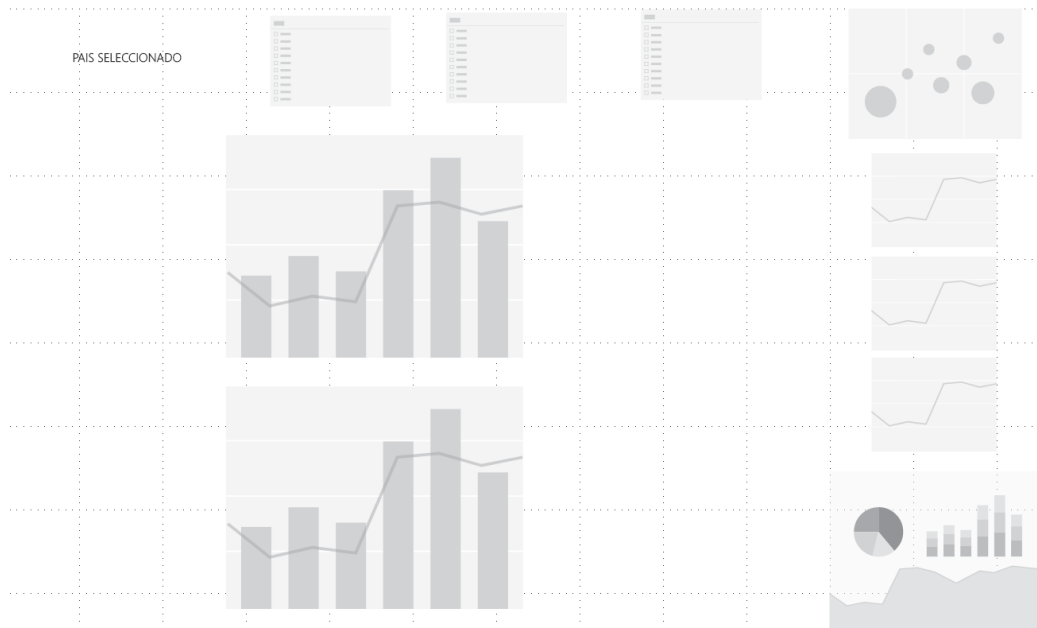


ILUSTRACIÓN 10- BORRADOR DEL DISEÑO DEL INFORME FINAL (FUENTE: PROPIA)

El informe estará compuesto por una cabecera donde se mostrará el nombre del país seleccionado para dar contexto de donde se realiza el estudio, junto a tres filtros que afectarán a todos los gráficos de la página:

Filtro por país

Se podrá seleccionar un país únicamente.

Filtro por fecha

Permitirá establecer un periodo para los que mostrar la información del país y producto.

Filtro por producto

Para facilitar la visualización se podrá filtrar por uno o varios productos de los que hay datos en función de los filtros previos por país y periodo.

Ocupando gran parte de la página del informe, dos gráficos de columnas y líneas mostrarán la relación de precio y precipitaciones o temperaturas para el periodo y productos seleccionados. Gracias a estos informes se podrá ver de forma muy visual cómo evoluciona el precio en función de las precipitaciones medias, las temperaturas medias, temperaturas máximas y mínimas.

En el margen derecho del informe se mostrarán las evoluciones de precio, precipitación y temperatura en tres gráficos de líneas distintos. Junto a estos implementaremos un gráfico con la desviación estándar de las temperaturas y precipitaciones en función del precio, que mostrará con mayor o menor degradado de color la desviación estándar de precipitación, y con un tamaño mayor o menor del círculo para la temperatura.

Finalmente, el informe mostrará en un diagrama de cajas el precio medio por producto en el periodo seleccionado.

4.2.- DISEÑO

La solución propuesta incluye a alto nivel, un repositorio de datos no estructurados, donde alojaremos los ficheros obtenidos de las diferentes fuentes de datos, un sistema de extracción y análisis de datos que aprovecha las ventajas del nuevo lenguaje de consulta de datos no estructurados U-SQL, un sistema de volcado de información y una base de datos relacional desde la que se realizarán las consultas para el informe.

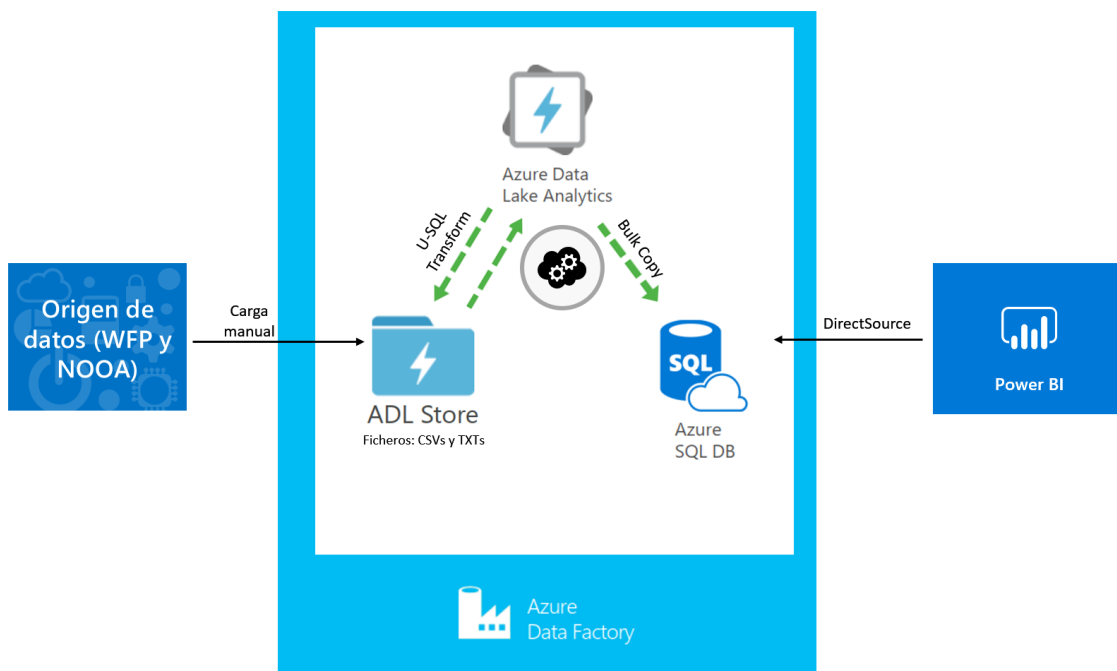


ILUSTRACIÓN 11- SOLUCIÓN PROPUESTA. (FUENTE: PROPIA)

En los siguientes apartados se detalla la estructura de datos que tendrá la información a medida que es procesada.

4.2.1.- ESTRUCTURA DE DATOS

Una vez procesados los datos, se hace necesaria la consulta desde herramientas de visualización como *Microsoft Power BI* o *Microsoft Excel* que permitan acceder a toda la información disponible que relacione precios de alimentos con indicadores climatológicos. Para ello se decide usar una base de datos relacional Azure SQL Server con el siguiente modelo de datos.

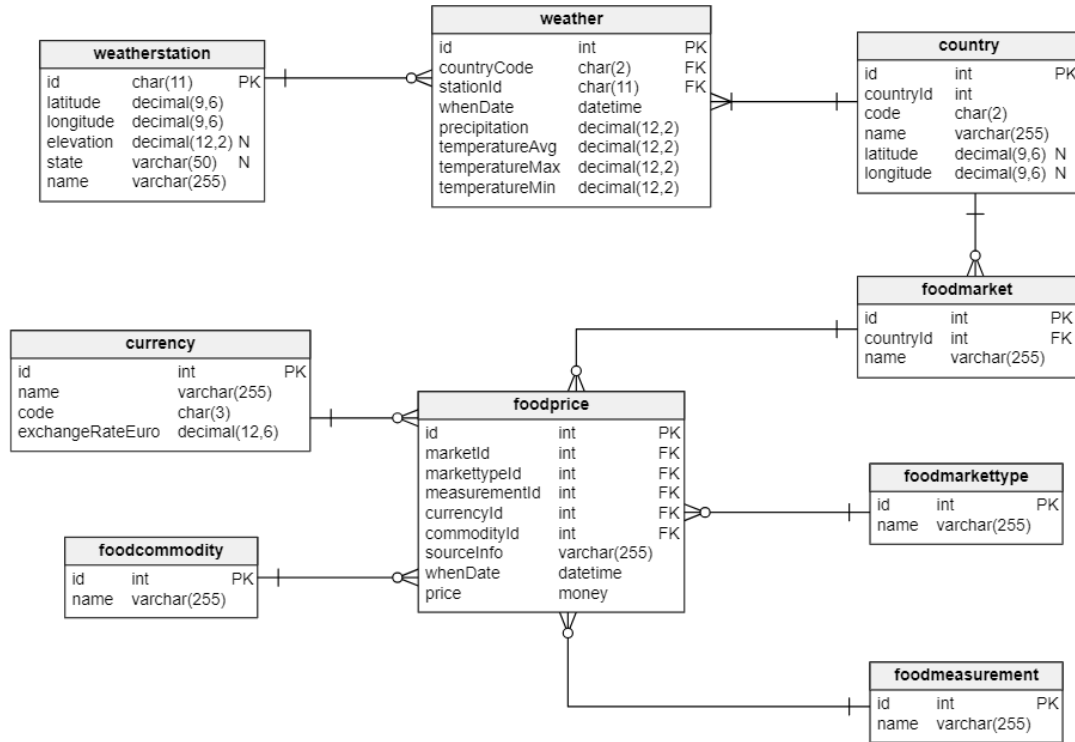


ILUSTRACIÓN 12- MODELO DE DATOS DEL PROTOTIPO (FUENTE: PROPIA)

Este modelo de datos mantiene una relación directa con las entidades de los ficheros disponibles que contienen la información de precios y clima.

4.2.2.- EXTRACCIÓN Y TRANSFORMACIÓN DE DATOS

La cantidad ingente de información disponible en los ficheros de precios y clima hace necesario el uso de herramientas de análisis y proceso de datos en la nube como los que dispone *Microsoft Azure*. En este caso, necesitamos ser capaces de procesar cada uno de los ficheros y extraer la información de forma que se puedan relacionar las diferentes entidades de datos para su posterior análisis. Para ello utilizaremos las bondades del lenguaje U-SQL que permite, mediante el uso de un lenguaje muy similar a SQL combinado con C#, procesar grandes ficheros y realizar tareas de transformación de datos de forma sencilla y escalarlo fácilmente.

La carga de datos inicial, de la que se extraerán datos maestros de las diferentes fuentes y que no son susceptibles de cambiar en el tiempo, será de forma manual como se describe más adelante. Sin embargo, para las actualizaciones del sistema de forma automática, será necesario dotar al conjunto con trabajos de chequeo de las fuentes de datos en busca de nuevos ficheros. Esto será posible gracias a la orquestación con Azure Data Factory, que permite crear trabajos que descargarán, transformarán mediante U-SQL y cargarán datos actualizados a la base de datos relacional. Un esquema del proceso de actualización que seguirá el sistema se muestra en la siguiente imagen:

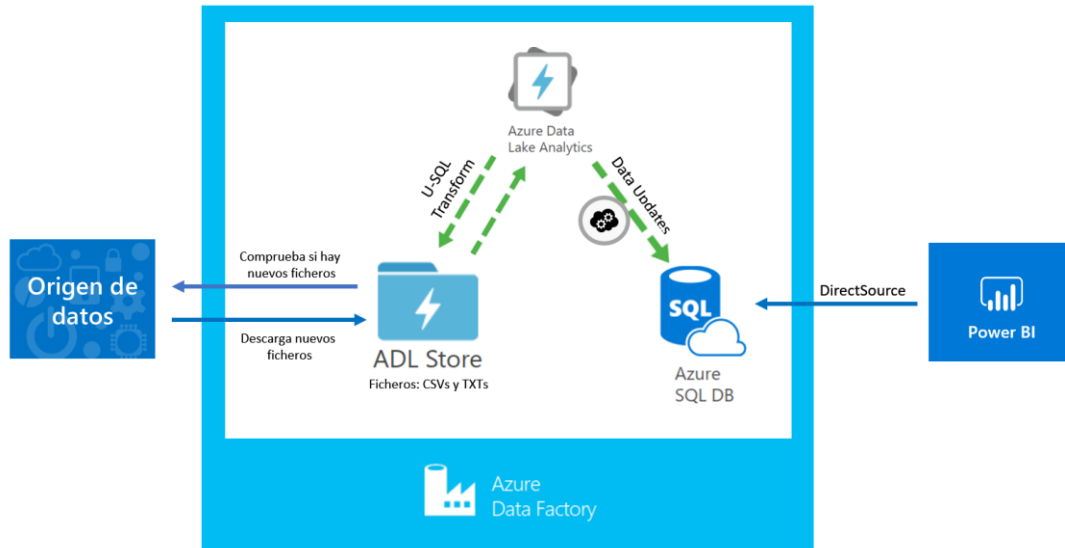


ILUSTRACIÓN 13- PROCESO DE COMPROBACIÓN DE NUEVOS DATOS Y ACTUALIZACIÓN DEL SISTEMA (FUENTE: PROPIA)

Previamente a la carga de datos maestros e iniciales, se proponen ficheros intermedios de datos como resultado de los procesos de extracción y transformación:

commoditiesWFP_output.csv

Contendrá la lista de productos para los que hay precios en el fichero de la WFP.

Campo	Descripción
ID	Identificador del producto en el fichero WFP
NAME	Nombre del producto

marketTypesWFP_output.csv

Contendrá la lista de los tipos de mercado para los que hay precios en el fichero de la WFP.

Campo	Descripción
ID	Identificador del tipo de mercado en el fichero WFP
NAME	Nombre del tipo de mercado

measurementsWFP_output.csv

Contendrá la lista de las diferentes unidades de medida para los que hay precios en el fichero de la WFP. En este caso y debido a la disparidad de unidades de medida en los diferentes productos según las diferentes fuentes de información, habrá que unificarlas para facilitar la visualización posterior, y conseguir el siguiente listado unificado: Kilogramos, Litros, Unidades.

Campo	Descripción
ID	Identificador de la unidad de medida en el fichero WFP
NAME	Nombre de la unidad de medida

[currenciesWFP_output.csv](#)

Contendrá la lista de las diferentes monedas que representan el precio de los productos en el fichero WFP. En este caso se hace necesario obtener el tipo de cambio a EUROS para facilitar la visualización del precio final.

Campo	Descripción
ID	Identificador de la moneda
NAME	Nombre de la moneda
RATE	Tipo de cambio de la moneda a EUROS
CODE	Código internacional de tres caracteres de la moneda

[marketsWFP_output.csv](#)

Contendrá la lista de los diferentes mercados donde se ha realizado el estudio de precios según WFP. En este fichero añadiremos el identificador del país para facilitar la manipulación de los datos a posteriori.

Campo	Descripción
ID	Identificador del mercado en el fichero WFP
NAME	Nombre del mercado
COUNTRYID	Identificador del país donde está el mercado

[countriesWFP_output.csv](#)

Contendrá la lista de los diferentes países donde se han realizado estudios de precios por WFP.

Campo	Descripción
ID	Identificador del país en el fichero WFP
NAME	Nombre del país

[foodpricesWFP_output.csv](#)

Contendrá los precios de los alimentos por mercado según los datos obtenidos por WFP.

Campo	Descripción
ID	Identificador generado para la fila
MARKETID	Identificador del mercado
MARKETTYPEID	Identificador del tipo de mercado
CURRENCYID	Identificador de la moneda
COMMODITYID	Identificador del producto
MEASUREMENTID	Identificador del tipo de medida
WHENDATE	Fecha en la que se obtiene el dato
PRICE	Precio en la moneda local
COMMODITYSOURCE	Fuente del dato a nivel local

[countriesGHCND_output.csv](#)

Contendrá la lista de los diferentes países donde se ha monitorizado el clima por el GHCND.

Campo	Descripción
CODE	Código del país
NAME	Nombre del país

[stationsGHCND_output.csv](#)

Contendrá la lista de las diferentes estaciones meteorológicas donde se han realizado mediciones por el GHCND.

Campo	Descripción
STATIONID	Identificador de la estación meteorológica
COUNTRYCODE	Código identificador del país
LATITUDE	Latitud donde se encuentra la estación
LONGITUDE	Longitud donde se encuentra la estación
ELEVATION	Elevación en metros donde se encuentra la estación
STATE	Estado cuando la estación es norteamericana
NAME	Nombre de la estación
GSNFLAG	Etiqueta GSN
HCNFLAG	Etiqueta HCN
WMOID	Identificador WMO

weather_output.csv

Contendrá la lista de las mediciones de las estaciones meteorológicas.

Campo	Descripción
STATIONID	Identificador de la estación meteorológica
COUNTRYCODE	Código identificador del país
WHENDATE	Fecha cuando se realizó la medición
PRCP	Precipitación en milímetros del día
TAVG	Temperatura media del día
TMAX	Temperatura máxima del día
TMIN	Temperatura mínima del día

4.2.3.- CARGA DE DATOS

Una vez obtenidos los ficheros intermedios como resultado del procesado de la información disponible de precios y clima, se deberán cargar en una base de datos relacional para su posterior consulta. En este punto, el prototipo a desarrollar para este trabajo solo incluye la carga manual de los datos, sin embargo, como línea de mejora futura, sería posible automatizar este proceso para que, una vez se haya terminado de procesar y transformar nuevos ficheros de información de precios o climatológica, se actualice la base de datos relacional de la que se utilizará como fuente de datos para los informes.

Como herramienta para la carga de esta información se optará por usar Azure Data Factory que permite crear, programar y manejar la integración de datos de forma escalable. Será necesaria la creación de 9 diferentes trabajos para cada una de las entidades de las que tenemos que provisionar con la información transformada.

Trabajo	Descripción
1_Bulk_Copy_Countries	Trabajo para la carga de países
2_Bulk_Copy_Currencies	Trabajo para la carga de monedas
3_Bulk_Copy_FoodCommodities	Trabajo para la carga de tipos de producto
4_Bulk_Copy_FoodMarket	Trabajo para la carga de mercados
5_Bulk_Copy_FoodMarketType	Trabajo para la carga de tipos de mercado
6_Bulk_Copy_FoodMeasurements	Trabajo para la carga de tipos de mediciones
7_Bulk_Copy_FoodPrices	Trabajo para la carga de los precios de los productos
8_Bulk_Copy_WeatherStations	Trabajo para la carga de las estaciones meteorológicas
9_Bulk_Copy_Weather	Trabajo para la carga de los datos de climatología

4.3.- IMPLEMENTACIÓN

Para la implementación de este prototipo será necesaria una suscripción a Microsoft Azure donde podremos configurar las diferentes herramientas para alojar los datos no estructurados, realizar las transformaciones y cargar los datos en una base de datos relacional.

La configuración y puesta en marcha de los diferentes servicios son posibles a través del panel de administración propio de Microsoft Azure en <https://portal.azure.com/> pero, además serán necesarias otras herramientas que facilitarán la construcción de los procesos de transformación de datos y la gestión de la base de datos relacional.

En primer lugar, se ha descargado la versión 2017 de Microsoft Visual Studio, que permite la ejecución en local y en Azure de los scripts de transformación U-SQL. Microsoft Visual Studio proporciona un interfaz de desarrollo amigable e intuitivo desde el que se pueden desarrollar diferentes productos en diferentes lenguajes para Móviles, Escritorio, Web, etc., además de soluciones de Inteligencia de Negocio y base de datos.

Para poder trabajar con Azure Data Lake, es necesaria la instalación de la extensión *Azure Data Lake and Stream Analytics Tools* que permite la ejecución en local y en Azure de scripts U-SQL.

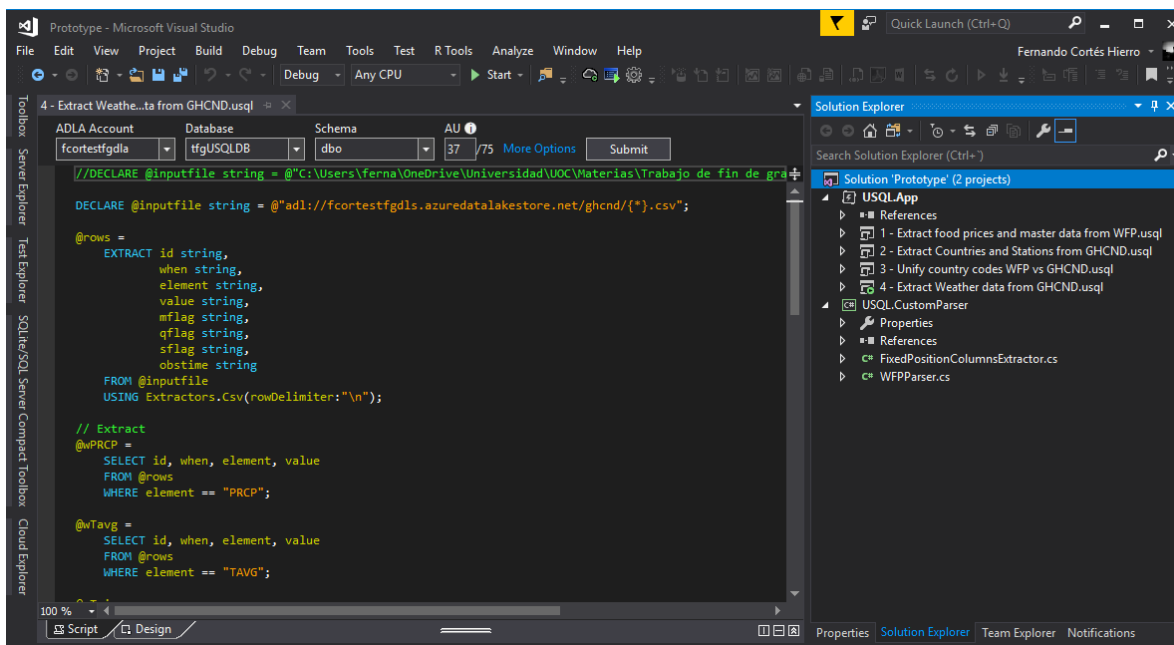


ILUSTRACIÓN 14- VISUAL STUDIO 2017 CON LOS SCRIPTS U-SQL (FUENTE: PROPIA)

Finalmente, para poder trabajar con la base de datos relacional en Azure SQL Server, se ha instalado y configurado el interfaz gráfico de gestión de bases de datos Microsoft SQL Server Management Studio 17 que permitirá conectarse a la base de datos, editar esquemas, consultar y actualizar información, o incluso revisar el rendimiento de las consultas para los informes.

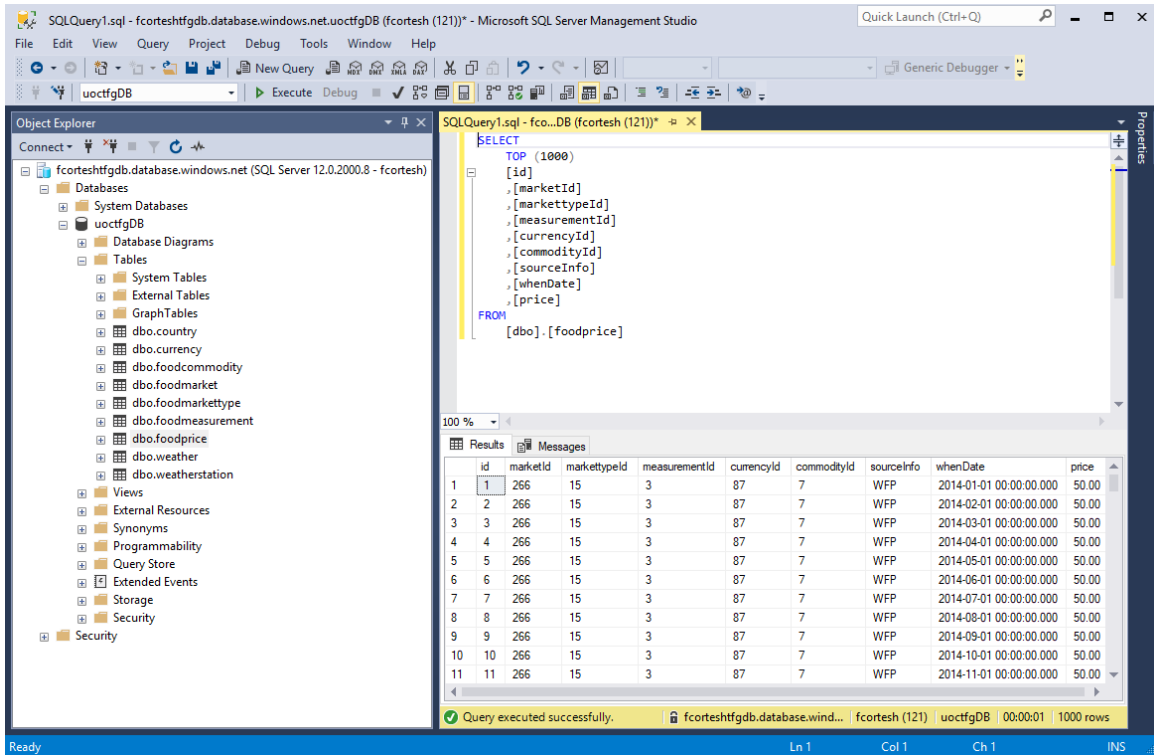


ILUSTRACIÓN 15- MICROSOFT SQL MANAGEMENT STUDIO MOSTRANDO LA BASE DE DATOS DEL PROTOTIPO (FUENTE: PROPIA)

4.3.1.- CONFIGURACIÓN DE HERRAMIENTAS MICROSOFT AZURE

Para configurar los diferentes servicios y herramientas necesarias para el prototipo que se está describiendo, en primer lugar, habrá que comenzar creando un *Azure Data Lake Store* donde alojar tanto los ficheros de precios de alimentos del WFP y del clima del GHCND, como los ficheros resultado de la ejecución de los scripts de procesado y transformación de datos.

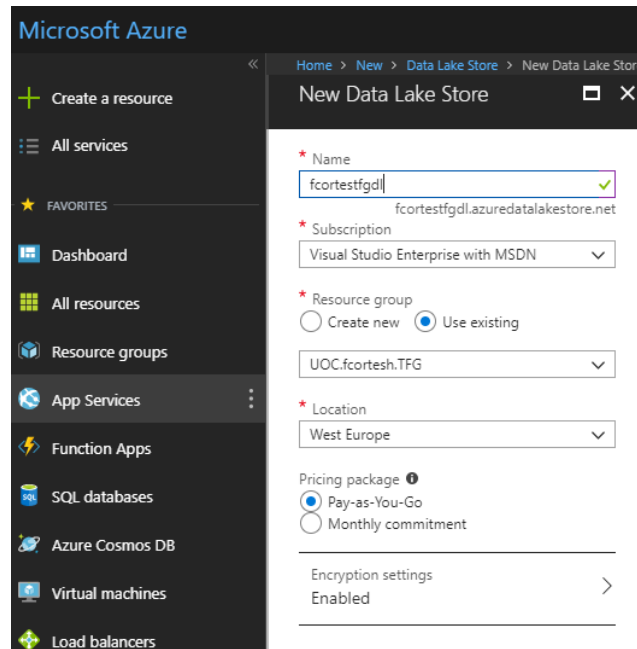


ILUSTRACIÓN 16- PROCESO DE CREACIÓN DE UN NUEVO DATA LAKE STORE (FUENTE: PROPIA)

Una vez creado, se configuran tres carpetas:

wfp

Donde se guardarán los ficheros asociados a los precios de los alimentos

ghcnd

Donde se guardarán los ficheros asociados al clima

output

Donde se generarán los ficheros una vez procesados y realizadas las transformaciones pertinentes de los ficheros de origen.

A continuación, se crea un servicio de tipo *Azure Data Lake Analytics* que permite la ejecución de scripts de procesamiento de datos mediante U-SQL contra los datos no estructurados alojados en un *Azure Data Lake Store* asociado.

Se configura *Azure Data Factory* que permitirá la ejecución de trabajos para el volcado de los datos procesados y tratados en la base de datos relacional. Esta base de datos relacional será una *Azure SQL Database*.

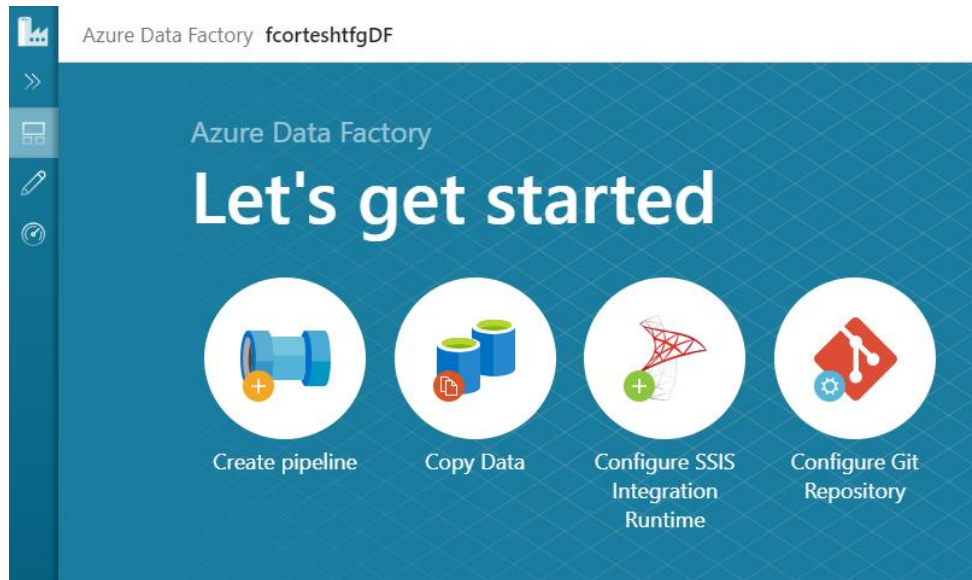


ILUSTRACIÓN 17- PORTAL DE CREACIÓN DE TRABAJOS EN DATA FACTORY (FUENTE: PROPIA)

Por último y no por ello menos importante se configura una base de datos relacional de tipo *Azure SQL Server* con una configuración de tipo *Standard*, con un tamaño máximo de 250GB, suficiente para el prototipo que se está desarrollando.

4.3.2.- TRANSFORMACIÓN DE DATOS (ETL)

Previamente a la transformación de los datos obtenidos de las diferentes fuentes citadas en apartados anteriores, se deberán subir los ficheros a procesar a las carpetas del *Azure Data Lake Store* creadas para este cometido.

Select files to upload to: adl://fcortestfgdls.azuredatalakestore.net/ghcnd

"2017.csv"

Allow overwrite existing files

Add selected files

FILE NAME	SIZE	ACCOUNT	PATH	MANAGE	STATUS	PROGRESS
2000.csv	1.06 GiB	fcortestfgdls	/ghcnd		Uploading	8%
2001.csv	1.08 GiB	fcortestfgdls	/ghcnd		Uploading	7%
2002.csv	1.09 GiB	fcortestfgdls	/ghcnd		Uploading	9%
2003.csv	1.12 GiB	fcortestfgdls	/ghcnd		Uploading	8%
2004.csv	1.14 GiB	fcortestfgdls	/ghcnd		Uploading	7%
2005.csv	1.13 GiB	fcortestfgdls	/ghcnd		Uploading	3%
2006.csv	1.13 GiB	fcortestfgdls	/ghcnd		Uploading	6%
2007.csv	1.15 GiB	fcortestfgdls	/ghcnd		Uploading	2%
2008.csv	1.2 GiB	fcortestfgdls	/ghcnd		Uploading	5%

ILUSTRACIÓN 18- PROCESO DE SUBIDA DE LOS FICHEROS CSV CON INFORMACIÓN DEL CLIMA (FUENTE: PROPIA)

Una vez los ficheros con los datos a tratar están disponibles en el contenedor de datos no relacionales de Azure, comenzamos la ejecución de los diferentes scripts de proceso y transformación.

1 - Extract food prices and master data from WFP

Este script lee el fichero de precios de alimentos de WFP y genera los ficheros maestros de datos relacionados como países, monedas, productos, mercados, tipos de mercados y finalmente, los precios de los alimentos.

```
USE DATABASE [tfgUSQLDB];

DECLARE @inputfileWFP string = @"adl://fcortestfgdls.azuredatalakestore.net/wfp/WFPVAM_FoodPrices_05-12-2017.csv";
DECLARE @inputfileExchange string = @"adl://fcortestfgdls.azuredatalakestore.net/wfp/CurrencyExchange.csv";
DECLARE @inputfileSymbols string = @"adl://fcortestfgdls.azuredatalakestore.net/wfp/CurrencySymbols.csv";

REFERENCE ASSEMBLY [USQL.CustomParser];

@rows =
    EXTRACT adm0_id int,
            adm0_name string,
            adm1_id int,
            adm1_name string,
            mkt_id int,
            mkt_name string,
            cm_id int,
            cm_name string,
            cur_id int,
            cur_name string,
            pt_id int,
            pt_name string,
            um_id int,
            um_name string,
            mp_month int,
            mp_year int,
            mp_price double,
            mp_commoditysource string
    FROM @inputfileWFP
    USING Extractors.Csv(skipFirstNRows : 1);

// Extract Market Types
@marketTypes =
    SELECT DISTINCT pt_id,
                   pt_name
    FROM @rows;

OUTPUT @marketTypes
TO @"/Output/marketTypesWFP_output.csv"
USING Outputters.Csv(outputHeader : false, quoting : false);
```

ILUSTRACIÓN 19- MUESTRA DEL SCRIPT U-SQL DE PROCESADO DE DATOS MAESTROS DE WFP. (FUENTE: PROPIA)

Como consecuencia de la disparidad de fuentes dentro de la propia WFP, se hace necesaria la homogeneización de cierta información para aportar valor a los informes finales. Gracias, a U-SQL es posible desarrollar pequeñas funciones en C# que facilitan la manipulación de los datos y su posterior transformación.

Para el caso de los tipos de medidas, en el fichero de precios de alimentos se podrán encontrar valores como: gramos que hay que pasar a kilogramos, libras que hay que pasar a kilogramos,

mililitros que hay que pasar a litros, galones o metros cúbicos que hay que pasar a litros, etc. Además, en muchos casos el precio no representa el valor de una unidad, sino que existen casos de 500 gramos de harina de trigo, que hay que unificar en precio por Kilogramo.

Otro caso importante que tratar en este script son las monedas. Todos los precios del fichero del WFP están en la moneda local del país donde está monitorizado el precio, por lo que es necesario obtener el tipo de cambio asociado en función del EURO. Para esta tarea, se utiliza la herramienta gratuita para pequeñas consultas, <https://fixer.io> que permite obtener un fichero JSON con el tipo de cambio de todas las monedas.

```
{
  "success": true,
  "timestamp": 1519296206,
  "base": "EUR",
  "date": "2018-05-18",
  "rates": {
    "AUD": 1.566015,
    "CAD": 1.560132,
    "CHF": 1.154727,
    "CNY": 7.827874,
    "GBP": 0.882047,
    "JPY": 132.360679,
    "USD": 1.23396,
    [...]
  }
}
```

ILUSTRACIÓN 20- EJEMPLO DEL FICHERO JSON CON LOS TIPOS DE CAMBIO. (FUENTE: PROPIA)

Una vez descargado y subido a ADLS, es procesado y asociado junto al listado de monedas extraído del fichero de precios de WFP para generar un fichero final con la información de las monedas. Como ya se apunta en otros apartados, queda fuera del prototipo la gestión de estos datos externos para la actualización periódica que mantendría el tipo de cambio al día. Sin embargo, su implementación no sería muy complicado mediante el uso de los trabajos en *Azure Data Factory* que serán usados más adelante.

Una vez ejecutado este script como resultado disponemos de nueve nuevos ficheros en la carpeta */Output* con el formato explicado en apartados anteriores.

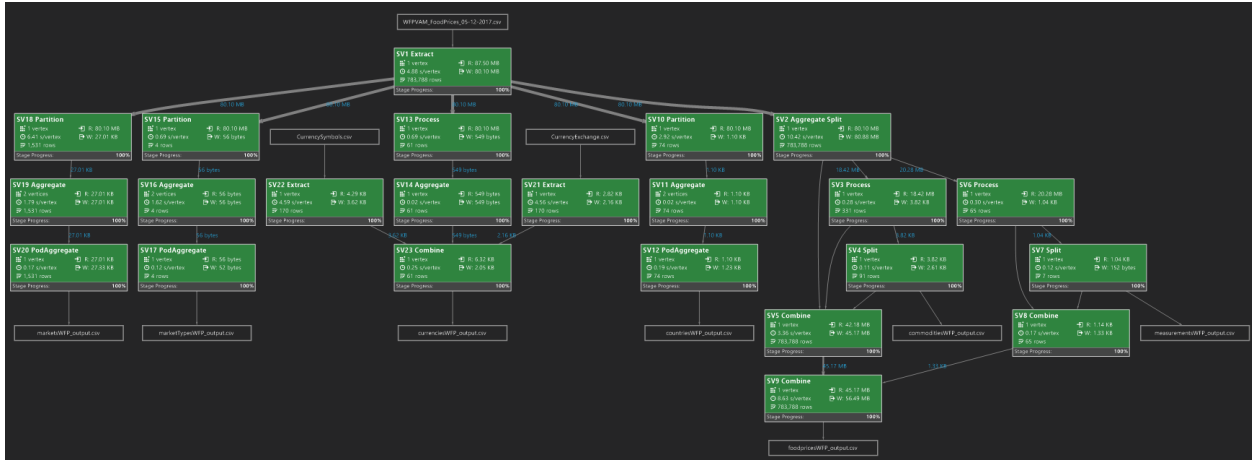


ILUSTRACIÓN 21- EJECUCIÓN DEL SCRIPT U-SQL DE PROCESO DE DATOS MAESTROS DE WFP. (FUENTE: PROPIA)

2 - Extract Countries and Stations from GHCND

Este segundo script extrae directamente a ficheros intermedios el listado de países y de estaciones meteorológicas obtenidas de GHCND.

Para poder procesar este tipo de ficheros donde cada columna tiene un tamaño prefijado, es necesaria la creación de un extractor específico para este tipo de ficheros, ya que los que vienen por defecto no encajan con el formato.

Se desarrolla un extractor de campos con un tamaño fijo que consiste principalmente en leer una línea del fichero e ir extrayendo las columnas en función del tamaño y la posición marcadas en la definición del fichero.

```
USE DATABASE [tfqUSQLDB];

DECLARE @inputfileCountries string = @"adl://fcortestfgdls.azuredatalakestore.net/ghcnd/ghcnd-countries.txt";
DECLARE @inputfileStations string = @"adl://fcortestfgdls.azuredatalakestore.net/ghcnd/ghcnd-stations.txt";

REFERENCE ASSEMBLY [USQL.CustomParser];

@rows =
    EXTRACT
        code string,
        name string
    FROM
        @inputfileCountries
    USING new USQLCustomParser.FixedPositionColumnsExtractor(
        new SQL.MAP<string, string> {
            {"0", "2"},
            {"3", "0"} });

@countries =
    SELECT code.Trim() AS code,
           name.Trim() AS name
    FROM @rows;

OUTPUT @countries
    TO @"/Output/countriesGHCND_output.csv"
    USING Outputters.Csv(outputHeader:false, quoting:true);
```

ILUSTRACIÓN 22- MUESTRA DEL SCRIPT U-SQL DE PROCESO DE PAÍSES Y ESTACIONES DE GHCND. (FUENTE: PROPIA)

Una vez ejecutado el script obtendremos dos nuevos ficheros intermedios para su posterior uso.

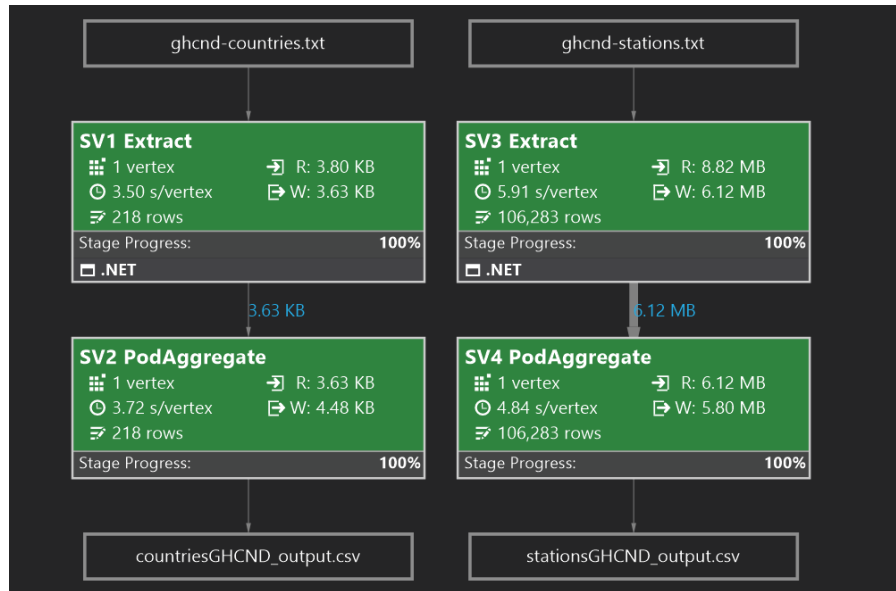


ILUSTRACIÓN 23- EJECUCIÓN DEL SCRIPT U-SQL DE PROCESO DE PAÍSES Y ESTACIONES DE GHCND. (FUENTE: PROPIA)

3 - Unify country codes WFP vs GHCND

Este tercer script es uno de los más relevantes en el prototipo ya que es el encargado de establecer una relación entre los datos de los precios de los alimentos y de la climatología usando como nexo los países en donde se ha registrado la información.

Toman como fuente de datos los resultados de los dos scripts anteriores que ya han sido ejecutados para poder aglutinar y obtener un fichero final de países que pueda ser usado para establecer una relación entre los precios y las variables climatológicas.

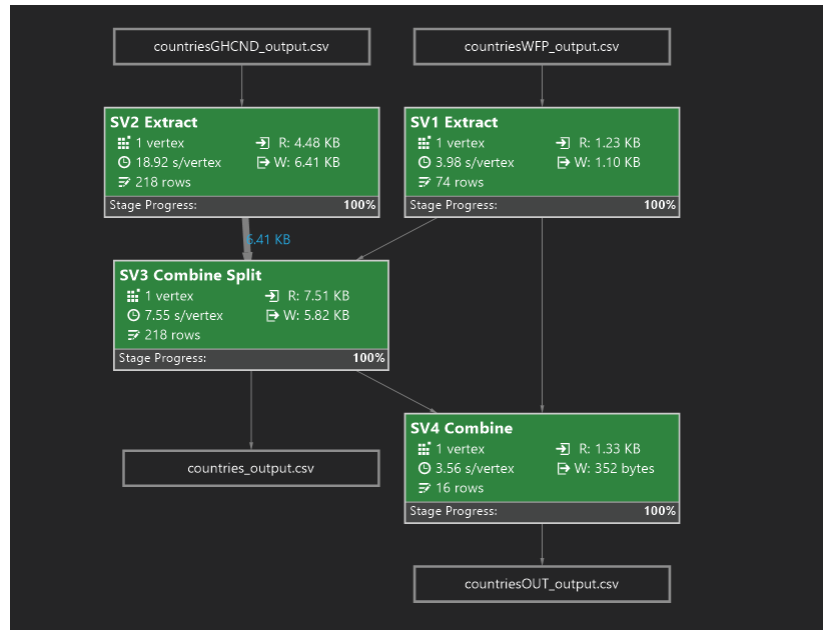


ILUSTRACIÓN 24- EJECUCIÓN DEL SCRIPT U-SQL DE UNIFICADO DE PAÍSES. (FUENTE: PROPIA)

Debido a pequeños detalles tipográficos en los nombres de los países, el proceso no ha podido encontrar de forma automática una relación directa para todos, por lo que se han obtenido 16 países que no eran enlazables que tuvieron que ser modificados manualmente.

File preview
countriesOUT_output.csv

Format Download Rename file Access

0	1
31	Bhutan
59	Congo
108	Haiti
68	Democratic Republic of the Congo
70	Djibouti
117	Iran (Islamic Republic of)
90	Gambia
139	Lao People's Democratic Republic
171	Myanmar
226	Somalia
257	United Republic of Tanzania
238	Syrian Arab Republic
269	Yemen
242	Timor-Leste
999	State of Palestine
70001	South Sudan

ILUSTRACIÓN 25- RESULTADO DE PAÍSES NO UNIFICADOS. (FUENTE: PROPIA)

4 - Extract Weather data from GHCND

Finalmente, este script permite procesar los 19 ficheros de información meteorológica por año para generar un fichero único de información meteorológica que pueda ser relacionado con el resto de información procesada.

Es destacable como en aproximadamente 7 minutos y gracias a las posibilidades de procesamiento y escalado que ofrece *Microsoft Azure* y más concretamente *Azure Data Lake Analytics*, se han procesado cerca de 120 Gb de información y transformado en un fichero único con la información relevante para este prototipo, pero que podría ser distinta en función de las necesidades futuras del trabajo.

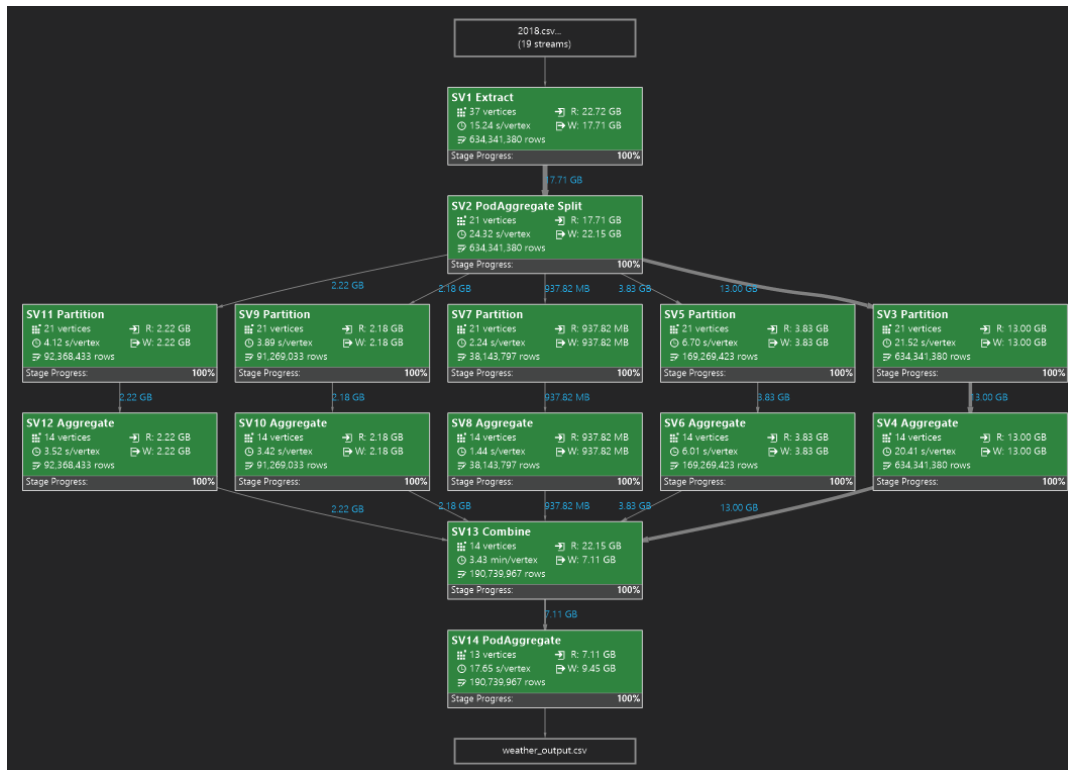


ILUSTRACIÓN 26- RESULTADO DE LA EJECUCIÓN DEL SCRIPT U-SQL DEL PROCESO DE DATOS CLIMATOLÓGICOS. (FUENTE: PROPIA)

Como resultado obtenemos los ficheros que servirán como fuente de datos para los informes que permitirán analizar la información para buscar una posible conexión entre la temperatura o precipitaciones y los precios de los alimentos en determinados países.

The screenshot shows a file explorer window titled 'fcortestfgdls Data Lake Store'. The interface includes a top navigation bar with icons for Filter, New folder, Upload, Access, Rename folder, Folder properties, Delete folder, and More. Below this, the current path is 'fcortestfgdls > Output'. A table lists the files in the folder:

NAME	SIZE	LAST MODIFIED
commoditiesWFP_output.csv	1.08 KB	5/7/2018, 12:00:16 PM
countries_output.csv	5.59 KB	5/7/2018, 12:09:27 PM
countriesGHCND_output.csv	4.48 KB	5/7/2018, 12:05:08 PM
countriesOUT_output.csv	352 bytes	5/7/2018, 12:09:27 PM
countriesWFP_output.csv	1.23 KB	5/7/2018, 12:00:16 PM
currenciesWFP_output.csv	2.05 KB	5/7/2018, 12:00:16 PM
foodpricesWFP_output.csv	56.5 MB	5/7/2018, 12:00:17 PM
marketsWFP_output.csv	27.3 KB	5/7/2018, 12:00:17 PM
marketTypesWFP_output.csv	52 bytes	5/7/2018, 12:00:16 PM
measurementsWFP_output.csv	55 bytes	5/7/2018, 12:00:16 PM
stationsGHCND_output.csv	5.8 MB	5/7/2018, 12:05:08 PM
weather_output.csv	9.45 GB	5/7/2018, 3:24:09 PM

ILUSTRACIÓN 27- RESULTADO FINAL DE LA EJECUCIÓN DE LOS SCRIPTS U-SQL. (FUENTE: PROPIA)

4.3.3.- CARGA DE DATOS

Una vez procesados y transformados los datos iniciales obtenidos de WFP y GHCND, para facilitar el análisis visual mediante los informes a crear con *Microsoft Power BI*, se realizará una carga de la información en la base de datos en Azure.

Para ello se crean diferentes trabajos que serán ejecutados de forma manual, los cuales leen la información en los ficheros resultantes de los scripts U-SQL y los vuelcan en la base de datos relacional creada para el análisis de la información de forma visual en los informes.

El proceso de creación de los trabajos de volcado de datos pasa por seleccionar el origen y el destino de la información, así como el mapeo necesario a aplicar. Un ejemplo de este proceso de creación de los trabajos se puede ver en las siguientes imágenes.

La fuente de datos para estos trabajos serán los ficheros CSV alojados en la carpeta /Output de ADLS, que son procesados por extractores de datos asociados a cada uno de los trabajos según la naturaleza del contenido de cada fichero.

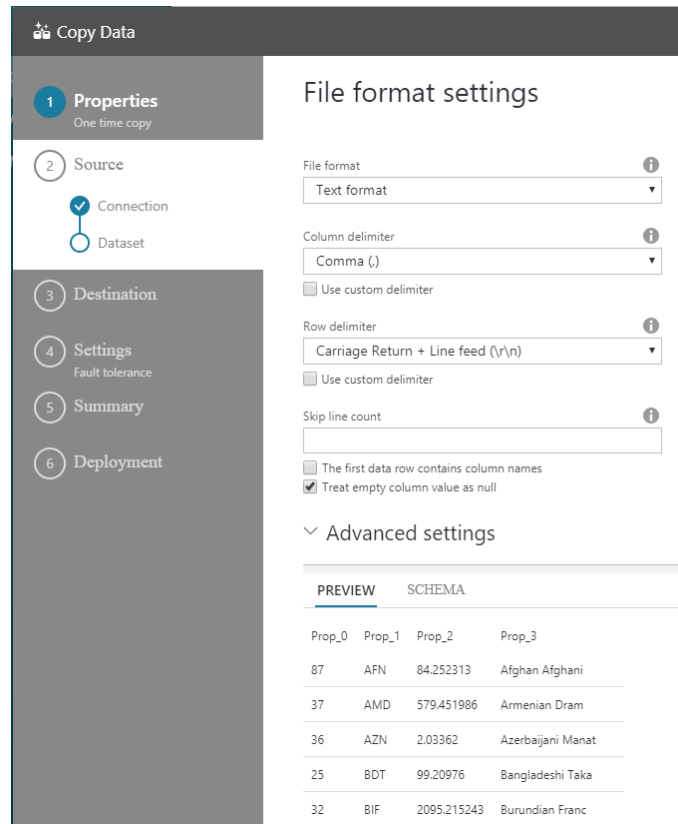


ILUSTRACIÓN 28- CONFIGURACIÓN DE LA FUENTE DE DATOS DEL TRABAJO DE VOLCADO. (FUENTE: PROPIA)

El mapeo de la información es la parte más importante de cada uno de los trabajos para el volcado de datos, ya que relaciona la información de cada campo en los ficheros de origen con los campos de las tablas de destino.

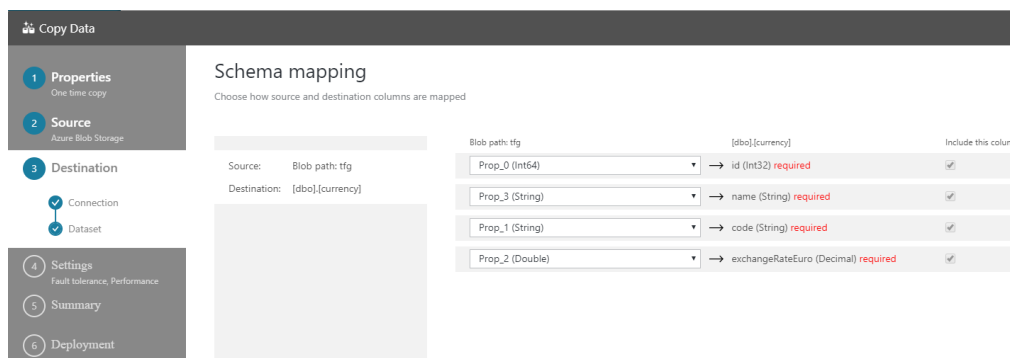


ILUSTRACIÓN 29- CONFIGURACIÓN DEL MAPEO DE DATOS DEL TRABAJO DE VOLCADO. (FUENTE: PROPIA)

Una vez ejecutados los trabajos de forma manual, la base de datos queda preparada para su explotación analítica mediante informes.

fcorteshtfgDF | Monitor Pipeline Runs

Refresh Alerts Metrics

Last 24 Hours 05/07/2018 2:04 PM - 05/08/2018 2:04 PM Time Zone (UTC+01:00) Brussels, Copenhagen, Madrid, Paris

All Succeeded In Progress Failed Cancelled

Pipeline Name	Actions	Run Start	Duration	Triggered By	Status
9_Bulk_Copy_Weather		05/07/2018, 8:14:26 PM	10:26:28	Manual trigger	Succeeded
8_Bulk_Copy_WeatherStations		05/07/2018, 8:02:46 PM	00:00:31	Manual trigger	Succeeded
7_Bulk_Copy_FoodPrices		05/07/2018, 6:52:38 PM	00:01:40	Manual trigger	Succeeded
6_Bulk_Copy_FoodMeasurements		05/07/2018, 6:47:01 PM	00:00:22	Manual trigger	Succeeded
5_Bulk_Copy_FoodMarketType		05/07/2018, 6:45:24 PM	00:00:23	Manual trigger	Succeeded
4_Bulk_Copy_FoodMarket		05/07/2018, 6:42:55 PM	00:00:24	Manual trigger	Succeeded
3_Bulk_Copy_FoodCommodities		05/07/2018, 6:40:35 PM	00:00:20	Manual trigger	Succeeded
2_Bulk_Copy_Currencies		05/07/2018, 6:13:31 PM	00:02:55	Manual trigger	Succeeded
1_Bulk_Copy_Countries		05/07/2018, 5:08:19 PM	00:00:21	Manual trigger	Succeeded

ILUSTRACIÓN 30- RESULTADO DE LA EJECUCIÓN DE LOS TRABAJOS DE VOLCADO DE DATOS. (FUENTE: PROPIA)

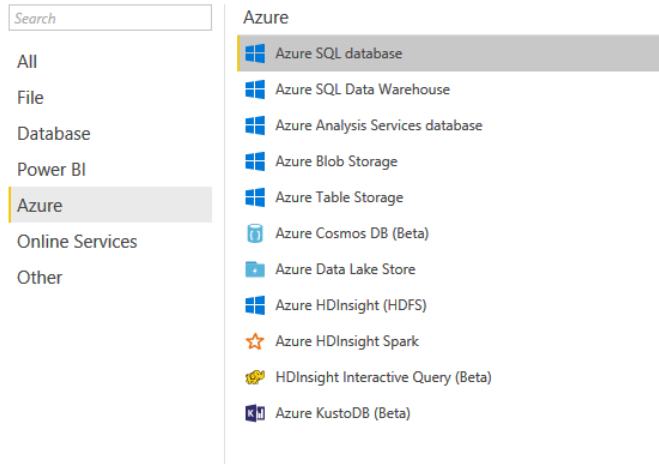
Como se puede observar en la imagen anterior, la ejecución del volcado de datos del clima tarda unas 10 horas y media de proceso. Sin embargo, esta es una carga inicial para el prototipo, pero sería posible realizar cargas incrementales con la nueva información que se vaya obteniendo de clima o precios.

4.3.4.- IMPLEMENTACIÓN INFORME

Las posibilidades que ofrece *Microsoft Power BI* satisfacen completamente las características del presente prototipo y facilita la evolución de este para futuras nuevas fuentes de datos.

Microsoft Power BI permite la conexión con una gran cantidad heterogénea de fuentes de datos como es Azure SQL database, donde se aloja la información relativa a los precios de alimentos y climatología procesados en anteriores etapas.

Get Data



SQL Server database

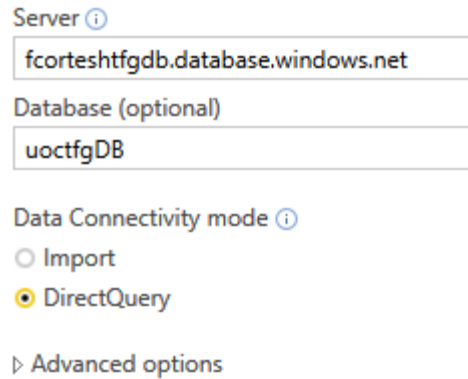


ILUSTRACIÓN 31- CONFIGURACIÓN DE LA FUENTE DE DATOS PARA CREAR EL INFORME. (FUENTE: PROPIA)

Una vez conectados a la base de datos correspondiente, es necesario comprobar que los datos en los que se basan los diferentes gráficos tienen el formato correcto y no sean tratados por la herramienta como datos susceptibles de ser pasados por una función de agregados como suele ocurrir habitualmente con los identificadores numéricos de cada entidad.

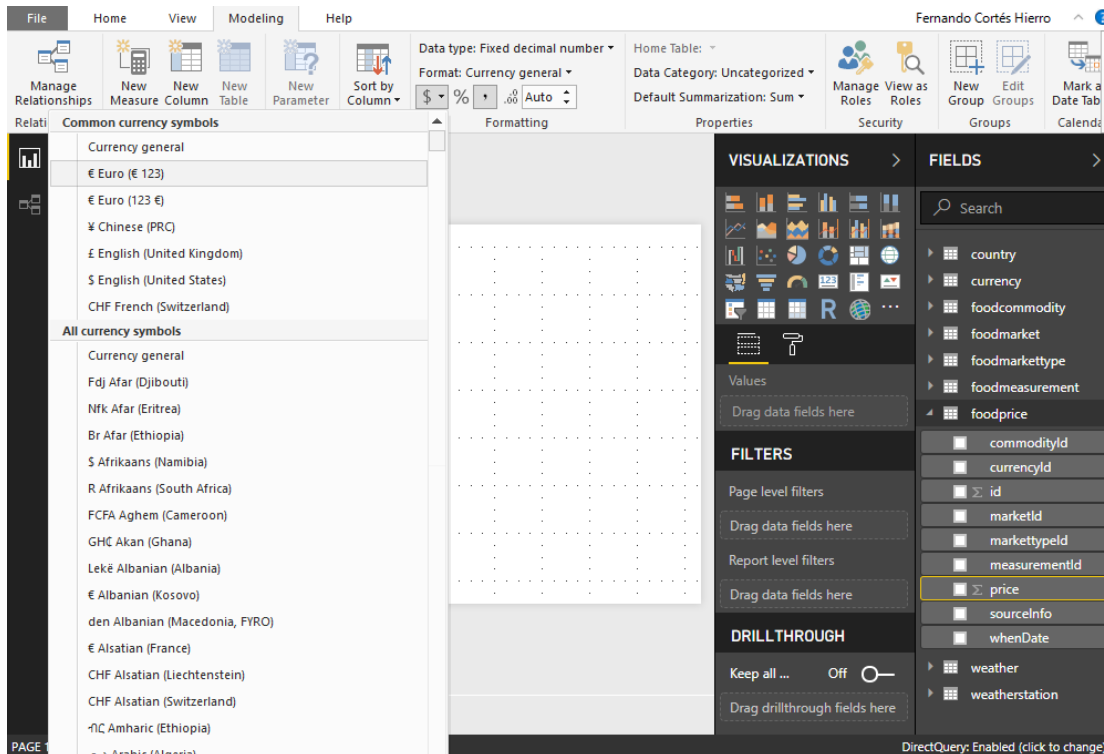


ILUSTRACIÓN 32- PROCESO DE ADAPTACIÓN DE DATOS EN FUNCIÓN DEL TIPO Y FORMATO. (FUENTE: PROPIA)

Finalmente, y siguiendo el diseño inicial propuesto en apartados anteriores, se procede a construir el informe con los diferentes elementos visuales previstos para este prototipo.

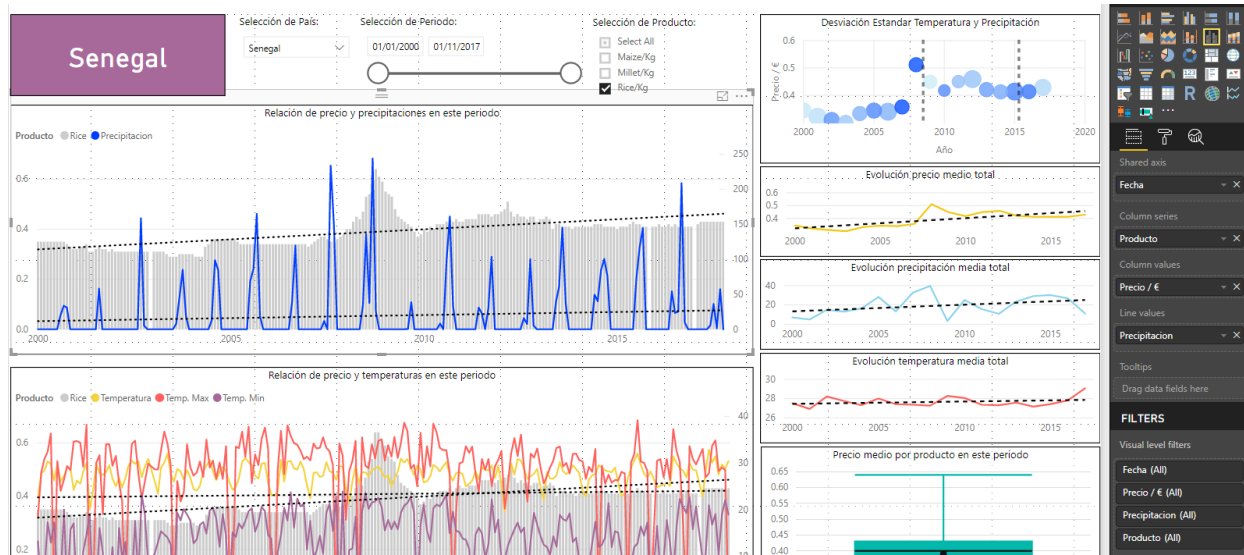


ILUSTRACIÓN 33- CONFIGURACIÓN DE LOS VISUALIZADORES SEGÚN DISEÑO INICIAL. (FUENTE: PROPIA)

MALI – Sorgo – Enero 2000 a Noviembre 2017

Como muestra de las conclusiones que se pueden extraer gracias al informe y sus diferentes componentes visuales, se presenta un ejemplo en el que se estudiará el precio del sorgo en Mali en el periodo comprendido entre enero del 2000 y noviembre del 2017.

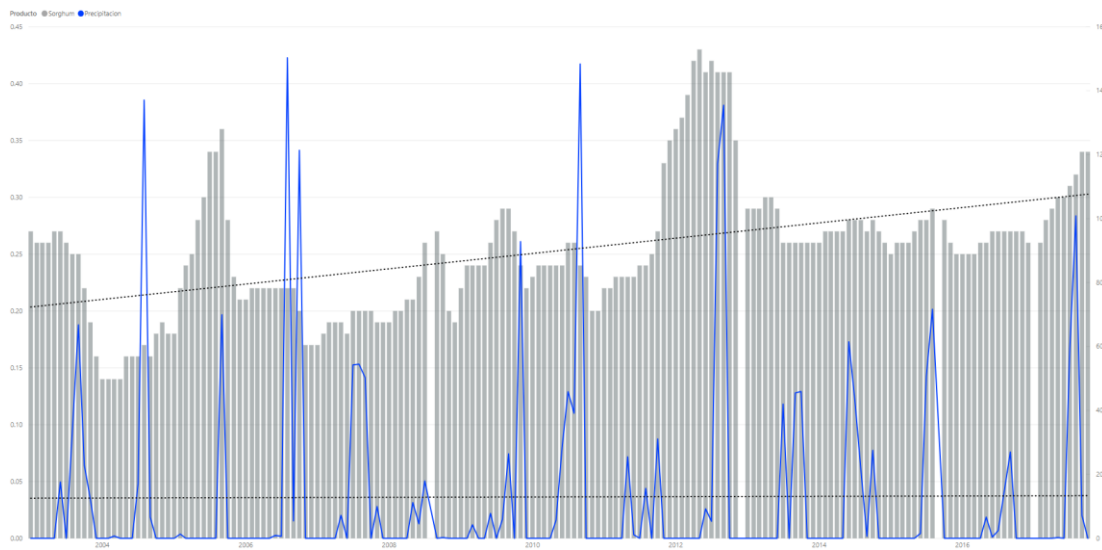


ILUSTRACIÓN 34- RELACIÓN DEL PRECIO DEL SORGO Y PRECIPITACIONES EN MALI EN EL PERIODO SELECCIONADO.

(FUENTE: PROPIA)

Gracias al gráfico anterior en el que se puede ver la relación del precio del sorgo en este caso con las precipitaciones para el periodo seleccionado en Mali, podemos ver una clara tendencia de precios ascendentes en los periodos de falta de precipitaciones y como después de los picos de precipitaciones hay una bajada de precios. Además, se puede ver una tendencia ascendente de precios en el periodo total seleccionado, marcado por la línea oblicua de puntos que recorre el gráfico de izquierda a derecha. En este caso concreto, se puede apreciar solamente con las precipitaciones una relación clara entre la falta de precipitaciones y el aumento de precios, así como las precipitaciones que favorecen la cosecha y permiten la bajada de los precios de forma casi inmediata, como se puede ver en el gráfico.

En el siguiente gráfico, se aprecia la relación entre las temperaturas mínimas, máximas y medias, con el precio del sorgo en Mali. Como hemos visto en el gráfico anterior, los picos descendentes del precio del sorgo acontecen tras periodos de precipitación, que junto los datos de este gráfico podemos ver como tiene relación en determinados casos con periodos de temperaturas altas. Esto tiene sentido en cuanto a que la planta del sorgo tiene su máximo nivel producto con temperaturas que rozan los 35 grados³.

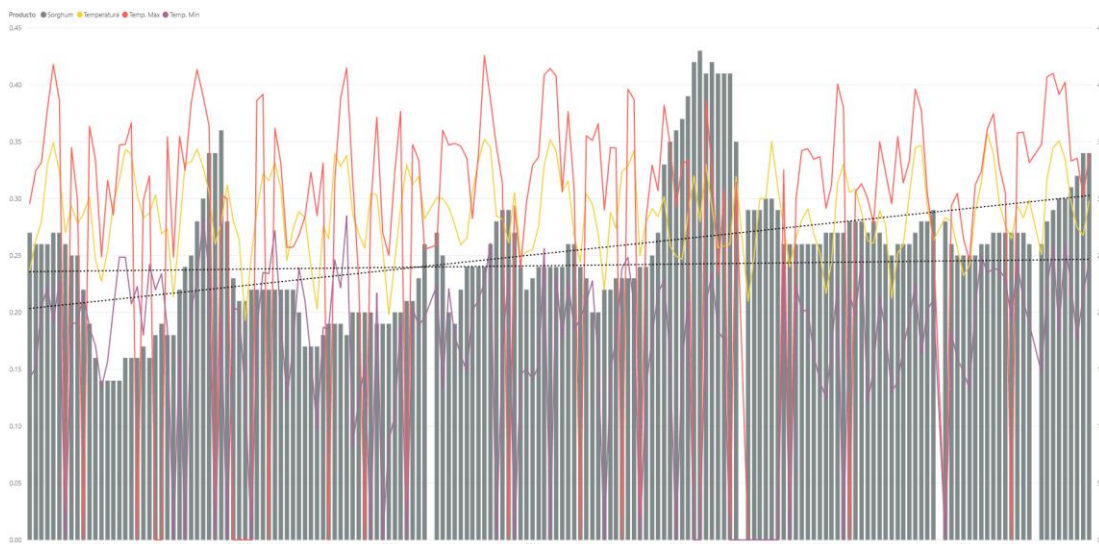


ILUSTRACIÓN 35- RELACIÓN DEL PRECIO DEL SORGO Y TEMPERATURAS EN MALI EN EL PERIODO SELECCIONADO.

(FUENTE: PROPIA)

A continuación, se muestra la desviación estándar de la temperatura y precipitación en función del precio y el periodo seleccionado. Gracias a este tipo de gráficos, es posible interpretar en función del tamaño de los círculos y el degradado del color usado, cuál es la dispersión del conjunto de datos con respecto a la media. En el caso de Mali, una desviación estándar muy alta

³ **Varios Autores.** (2018, abril). "Sorghum – Exigencias del cultivo". Wikipedia [documento en línea]. <
https://es.wikipedia.org/wiki/Sorghum#Exigencias_del_cultivo>

en la media de la precipitación y baja en la temperatura, tiene relación con un precio alto del sorgo en el año 2012. Ocurre lo contrario en el año 2004 donde el precio bajo del sorgo tiene relación con una desviación estándar baja en la precipitación, pero alta en la temperatura.

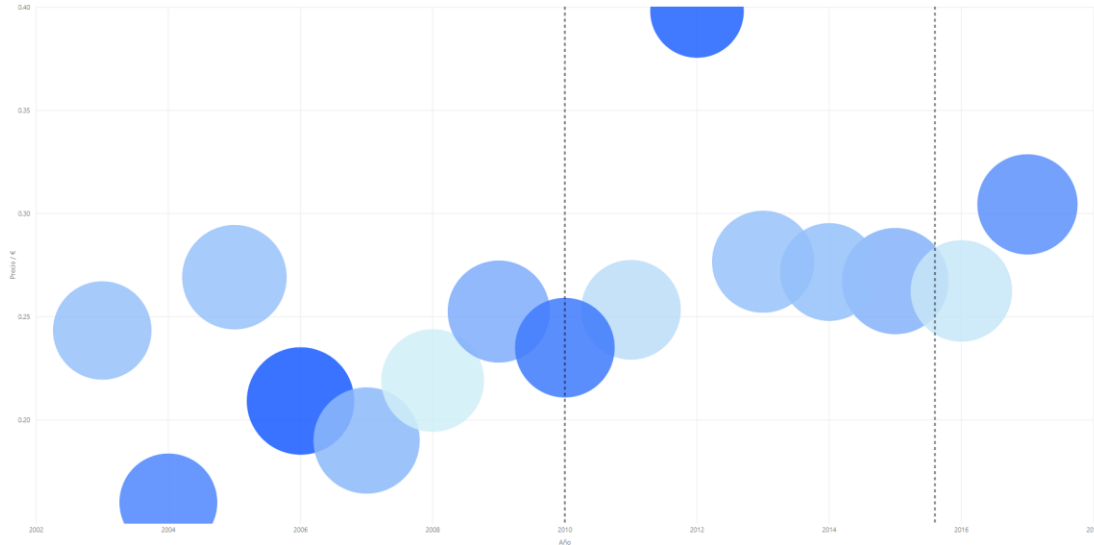


ILUSTRACIÓN 36- DESVIACIÓN ESTÁNDAR DE TEMPERATURA Y PRECIPITACIÓN EN MALI EN EL PERIODO SELECCIONADO.

(FUENTE: PROPIA)

El siguiente gráfico permite observar la evolución media del precio del producto seleccionado, en este caso se observa un aumento progresivo del precio del sorgo en Mali en el periodo indicado. El gráfico muestra la tendencia que se aprecia claramente ascendente.

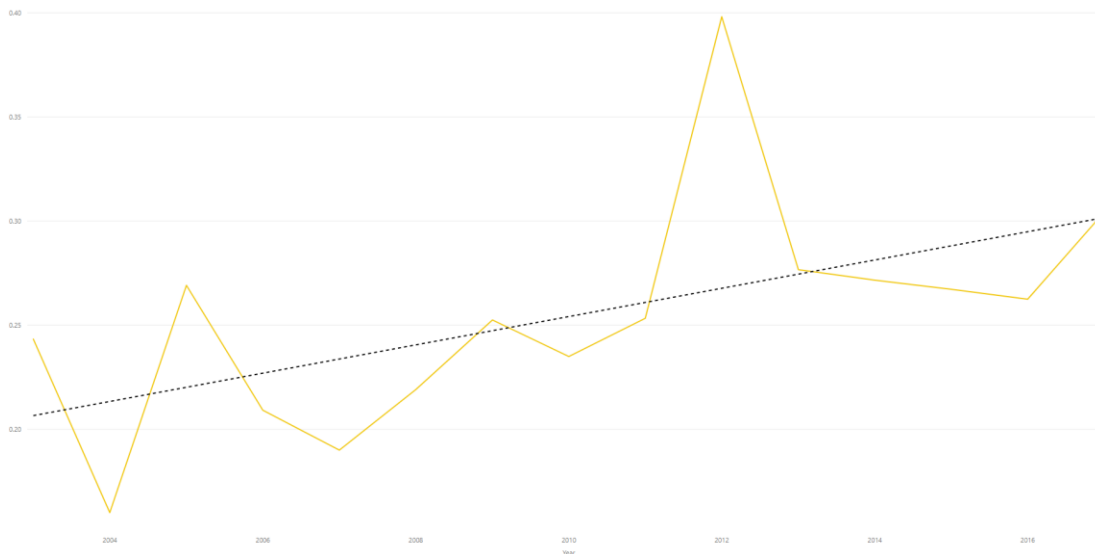


ILUSTRACIÓN 37- EVOLUCIÓN DEL PRECIO MEDIO TOTAL EN MALI EN EL PERIODO SELECCIONADO.

(FUENTE: PROPIA)

Otro de los gráficos que se pueden interpretar en el informe tiene que ver con la evolución de las precipitaciones en el país y periodo seleccionado. En este caso para Mali, se puede apreciar la variación de los periodos de precipitaciones con unos años más lluviosos que otros. La tendencia según el gráfico y para este periodo es de estabilidad.

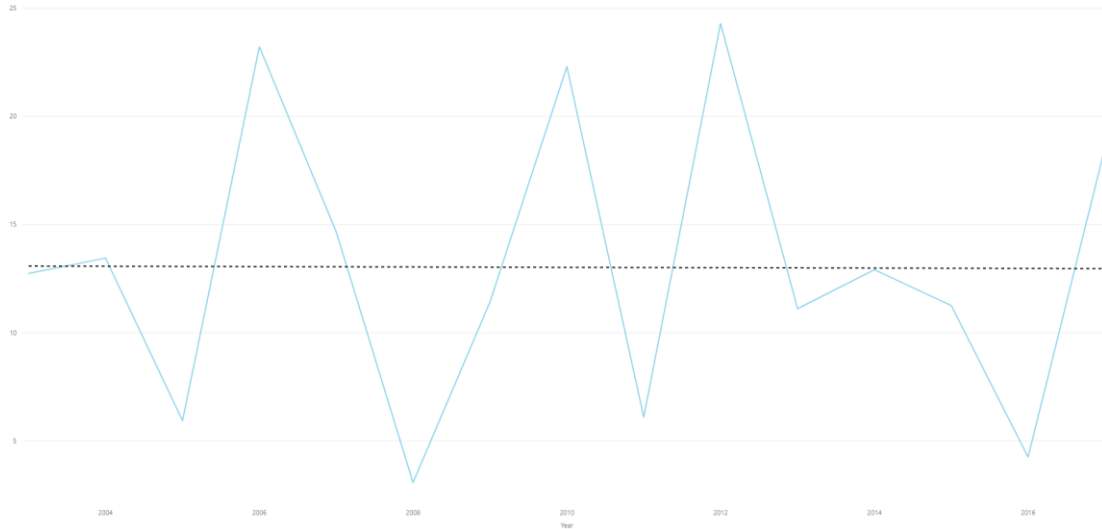


ILUSTRACIÓN 38- EVOLUCIÓN DE LA PRECIPITACIÓN MEDIA TOTAL EN MALI EN EL PERIODO SELECCIONADO.

(FUENTE: PROPIA)

Como el anterior gráfico, éste muestra la evolución de las temperaturas medias del país y periodo seleccionado. En este caso, en Mali se puede observar cómo en los últimos años los picos de temperatura se han acrecentado, siendo el año 2013 el comienzo de una subida de las temperaturas medias. A pesar de la acentuada subida del último periodo, la tendencia mostrada por el gráfico mediante la línea de puntos discontinua es ligeramente ascendente.

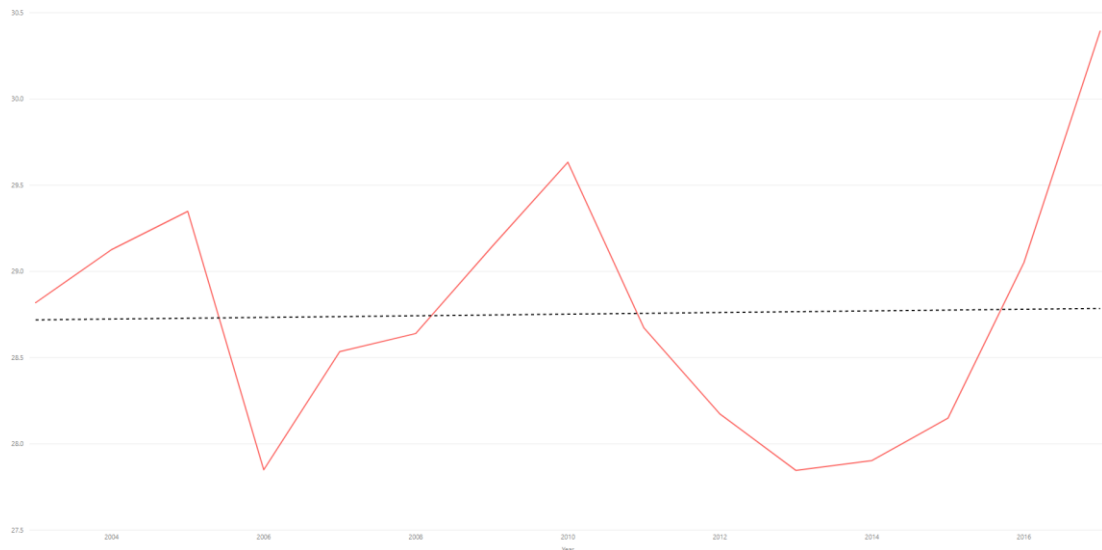


ILUSTRACIÓN 39- EVOLUCIÓN DE LA TEMPERATURA MEDIA TOTAL EN MALI EN EL PERIODO SELECCIONADO.

(FUENTE: PROPIA)

Por último, el siguiente gráfico permite estudiar la existencia de valores atípicos en el precio medio de un producto, en este caso del sorgo en Mali, así como la simetría de la distribución. Se muestran los valores mínimos, máximos, los cuartiles y la mediana relativos a los datos de precios medios.

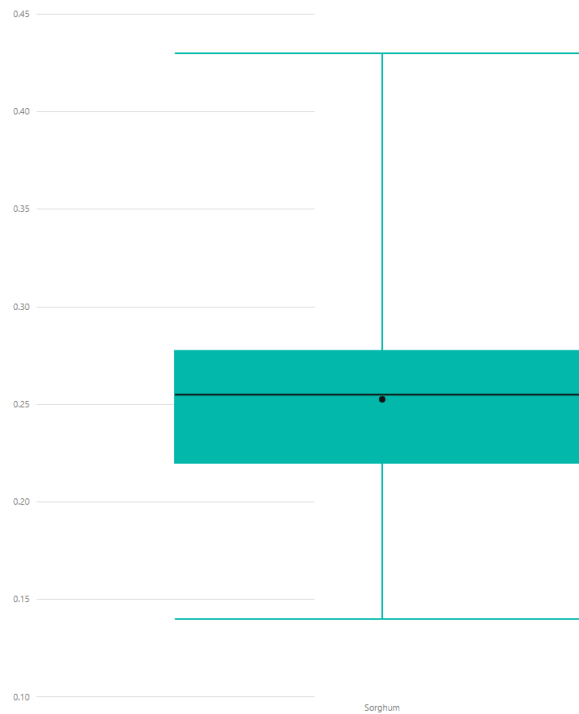


ILUSTRACIÓN 40- DIAGRAMA DE CAJAS DEL PRECIO POR PRODUCTO EN MALI EN EL PERIODO SELECCIONADO.

(FUENTE: PROPIA)

4.4.- RESULTADOS

Sin bien es cierto que el prototipo obtenido no aportaría un valor significativo como tal al estudio de la correlación entre las variables estudiadas, es un primer paso de sistematización de un prototipo que incluya más variables asociadas a otras causas que influyen el precio, y que permitirían ir afinando el peso de cada una de ellas para el análisis de la fluctuación de precios de alimentos.

Sin embargo, es posible obtener ciertas conclusiones a nivel general, y sin entrar en casuísticas específicas por país. Se puede observar que la desviación estándar de temperatura media en la gran mayoría de los países analizados se desvía significativamente de la media y tiene una relación causa-efecto en la desviación estándar del precio medio. Como consecuencia, podemos inferir que existe mayor dispersión de datos relativos a temperaturas que afectan al precio que los datos relacionados con la precipitación, aunque claramente también tiene impacto en estos

precios. Sin embargo, y como se explica durante todo el trabajo, es tremendamente complicado concluir que únicamente los precios de los alimentos básicos se ven influidos solamente por temperaturas y precipitaciones.

Como ventaja del uso de *Microsoft Power BI*, existe la posibilidad de compartir el trabajo realizado a través de diferentes medios telemáticos, a saber: desde incrustar el informe en páginas web, exportarlo a *Microsoft PowerPoint*, insertarlo como parte de un sitio web colaborativo para su consulta, hasta finalmente compartirlo entre un grupo de trabajo para realizar modificaciones conjuntas al informe.

Para el caso que ocupa este trabajo de fin de Grado, y como parte de los entregables, se facilita una dirección web desde la que profesores y miembros del tribunal de la *Universitat Oberta de Catalunya*, pueden interactuar con el resultado de este trabajo.⁴

DIRECCIÓN WEB PARA ACCEDER AL INFORME INTERACTIVO RESULTADO DEL PRESENTE PROTOTIPO

<https://app.powerbi.com/view?r=eyJrIjoiMjg0ZGRhNWZkZDk1ZS1hOGZlTQyODdkY2U2ZmU2OSIsImMiOiJh9>

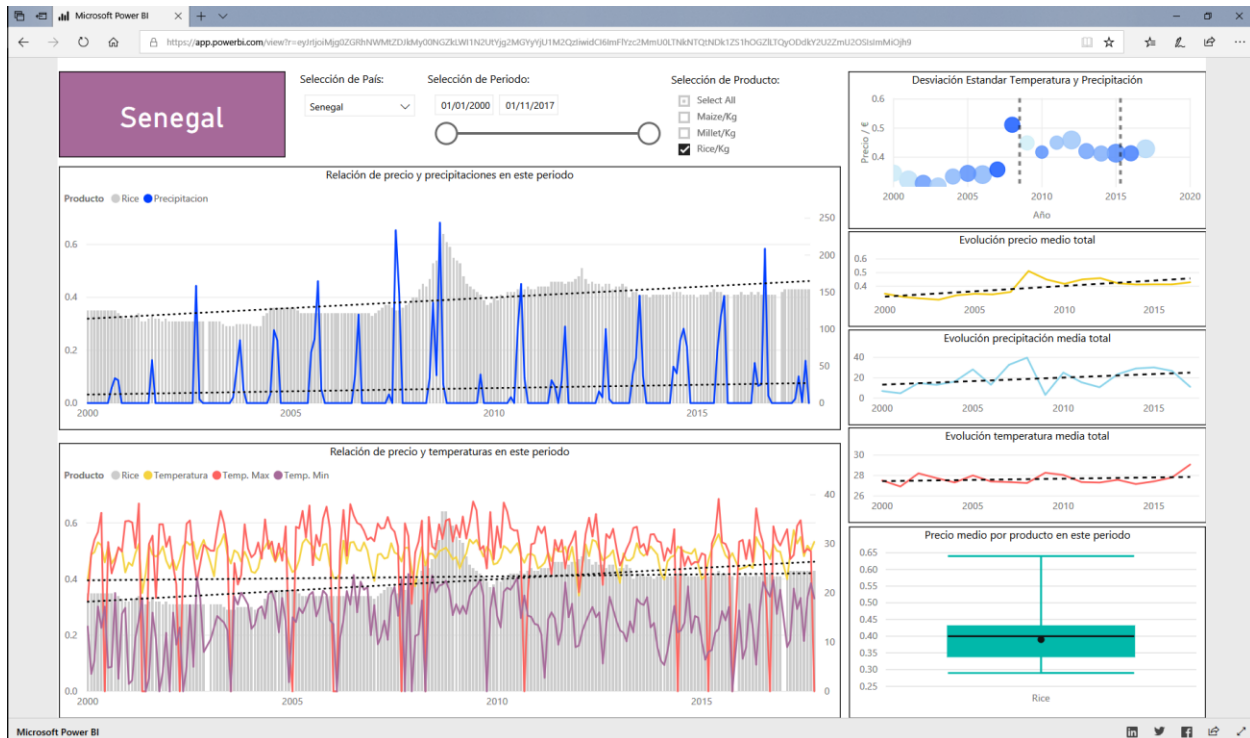


ILUSTRACIÓN 41- VISUALIZACIÓN DEL INFORME INTERACTIVO VÍA WEB. (FUENTE: PROPIA)

⁴ Esta dirección web pública estará disponible hasta el 14 de Julio de 2018 debido a la caducidad del periodo de prueba de la suscripción premium para *Microsoft Power BI*.

4.4.1.- ALGORITMOS DE INFORMACIÓN DE INTERÉS

Gracias a la potencia de *Microsoft Power BI*, es posible aplicar automáticamente de forma general diferentes algoritmos de búsqueda de información clasificada como interesante, que permite el análisis con detalle de elementos que resulten relevantes para el estudio de los datos de precios y clima en los que se basa este prototipo. Estos algoritmos son ejecutados automáticamente con los datos que dispone *Microsoft Power BI* al conectar el informe a la fuente de datos de precios de alimentos y datos climatológicos, por lo que no todos los resultados arrojan información relevante debido a que es una inteligencia artificial la que propone y busca correlaciones sobre los datos que tiene.

Es por esto por lo que, aunque los algoritmos usados son variados y no todos arrojan información relevante para el estudio de los datos que tienen que ver con este prototipo, si son destacables los siguientes algoritmos y graficas asociadas que proporcionan interesantes puntos de partida para el estudio de la información.

Como línea de mejora, entrenar esta inteligencia artificial para que encuentre correlaciones significativas en todos los datos procesados, mejorará los estudios comparativos a nivel mundial entre productos y países.

Algoritmo de Tendencias generales de la serie temporal

Este algoritmo detecta las tendencias más significativas hacia arriba o debajo de los datos de la serie temporal.

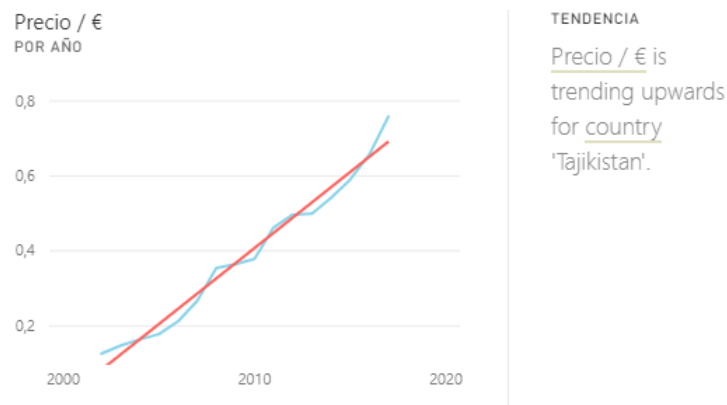
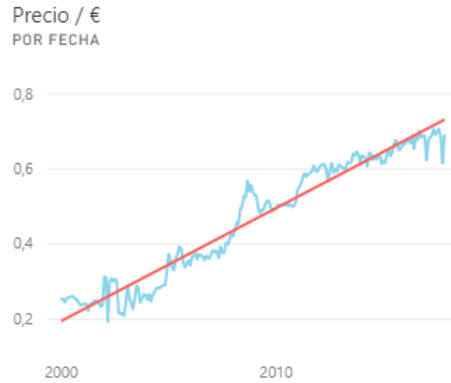


ILUSTRACIÓN 42- TENDENCIA DEL PRECIO/€ ANUAL EN TAYIKISTÁN. (FUENTE: PROPIA)



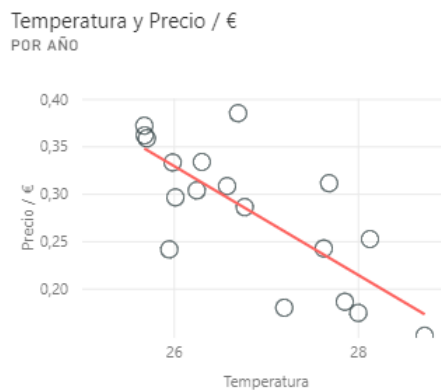
TENDENCIA
 Precio / € is
 trending upwards
 for Producto 'Rice'.

ILUSTRACIÓN 43- TENDENCIA DEL PRECIO/€ ANUAL MUNDIAL PARA EL ARROZ. (FUENTE: PROPIA)

En los ejemplos anteriores, se muestran dos tendencias detectadas por los algoritmos donde se destaca cómo para un producto tan importante a nivel mundial para la alimentación básica de millones de personas, el precio mundial sigue una tendencia ascendente en los últimos 18 años. Además, otra de las conclusiones de estas tendencias detectadas, es como el precio medio de los alimentos básicos en Tayikistán, ha experimentado un incremento en los últimos años que actualmente continúa al alza.

Algoritmo de Correlación

Este algoritmo detecta los casos en los que varias medidas muestran una correlación entre sí al trazarse en una dimensión específica del conjunto de datos.



CORRELACIÓN
 There is a
 correlation between
Temperatura and
Precio / € for
Producto
 'Sorghum'.

ILUSTRACIÓN 44- CORRELACIÓN ENTRE TEMPERATURA Y PRECIO/€ PARA EL SORGHUM. (FUENTE: PROPIA)

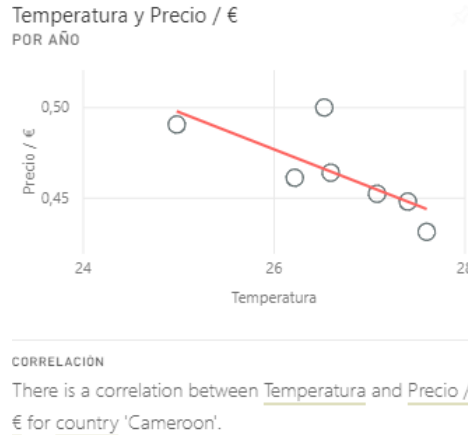


ILUSTRACIÓN 45- CORRELACIÓN ENTRE TEMPERATURA Y PRECIO/€ EN CAMERÚN. (FUENTE: PROPIA)

En las anteriores imágenes, se muestra que mediante la aplicación del algoritmo de correlación se ha detectado una evidencia de cómo a nivel mundial, existe una relación directa entre la temperatura media y el precio del sorgo, otro de los cereales más usados en el mundo para consumo humano y animal.

Finalmente, se destaca otra relación directa entre la temperatura media y el precio medio de los alimentos en un país como Camerún. La gráfica muestra como el precio medio de los alimentos en Camerún decrece a medida que aumenta la temperatura como norma general. Esto puede ser debido al tipo de cultivos que se utilizan en este país y que tienen su nivel óptimo de explotación con altas temperaturas, lo que repercute en el precio. Es decir, con altas temperaturas hay más producción por lo que el precio baja.

5.- LINEAS DE TRABAJO FUTURO

La principal ventaja del éxito en la consecución del prototipo desarrollado a lo largo del presente trabajo de final de grado es asegurar la viabilidad del diseño inicial. Gracias a esto, las posibilidades que se abren de cara al futuro para el desarrollo de una solución completa en cuanto a nuevos orígenes de datos son muchas desde el punto de vista del análisis y la predicción de la fluctuación de precios de alimentos.

Disponer de una solución en la nube, y más concretamente como muestra el prototipo en Microsoft Azure, garantiza una capacidad casi ilimitada en cuanto a escalabilidad, ya que facilita procesar multitud de fuentes de datos distintas que permitan optimizar la solución y conseguir establecer una correlación fiable entre precios de alimentos básicos y sus causas.

Otra de las garantías de éxito basando la solución en las herramientas seleccionadas para el prototipo, es la posibilidad de alojar en ADLS cualquier tipo de dato para ser usado como fuente de información. Una de las líneas de mejora es la posibilidad de utilizar trabajos en Data Factory para obtener datos actualizados mediante APIs como podría ser el caso de tipos de cambio actualizados, o datos actualizados de información meteorológica mediante FTP del NOAA. Esto abre un sinfín de posibilidades para que la solución final disponga de los datos más actualizados posibles que mejoren la exactitud del análisis de la información.

Explotar las posibilidades de los trabajos en Data Factory será una gran fuente de mejoras para la solución de análisis con todo tipo de fuentes de datos. Desde Data Factory, se pueden mantener actualizadas las fuentes de información como se ha explicado en el párrafo anterior, pero, además permite ejecutar trabajos periódicos de transformación de datos mediante scripts U-SQL como los usados en el prototipo de este TFG, con lo que la base de datos relacional que contiene la información que será consumida por el informe, quedará automáticamente actualizada. Esto, unido a una orquestación adecuada de los procesos de actualización, transformación y volcado, permitirá disponer de una solución completa que facilitará el análisis de la correlación existente entre los precios y las diferentes causas.

Como posibilidad de mejora futura, está la traducción de los literales en las fuentes de datos actuales y futuras para mostrar una información completa en un idioma común.

Finalmente, una vez dispuesta toda la infraestructura para alojar datos no estructurados, actualizarlos, transformarlos y alojar el resultado para su posterior análisis en una plataforma como *Microsoft Azure*, es un paso natural aprovecharse de las posibilidades que ofrece en *Machine Learning*. La solución estaría completa proporcionando un servicio web que fuese consultable desde *Microsoft Power BI*, *Microsoft Excel* o cualquier aplicación de terceros, que permitiese predecir el precio de un alimento concreto: arroz, trigo, sorgo, etc., en un país concreto en función de una serie de parámetros dados como: precipitación y temperaturas medias del último mes, precio medio del combustible del último mes, etc. Esta opción, permitirá por fin predecir fuertes fluctuaciones del precio de los alimentos con la consiguiente posibilidad de implementar medidas paliativas de ayuda a la población.

6.- CONCLUSIONES

El trabajo de fin de grado es la culminación de los conocimientos adquiridos a lo largo del grado de Ingeniería Informática y resulta en ser capaz de entender, planear y afrontar con éxito, el desarrollo de un proyecto BI relativo a un caso real.

Al comiendo de este TFG se planeó como objetivo principal el proporcionar un sistema de análisis de datos que permitiese visualizar la evolución de los precios de los alimentos básicos en diferentes países en función de factores climatológicos. El resultado es muy satisfactorio desde el punto de vista de la ejecución técnica ya que se ha cumplido con el objetivo inicial de proporcionar un prototipo 100 por cien funcional y accesible desde cualquier navegador web donde poder analizar los datos, sin embargo, desde el punto de vista cualitativo no se ha llegado obtener los resultados esperados que permitieran evidenciar una correlación directa entre las variables analizadas, y es aquí donde se abren nuevas líneas de mejora para añadir más variables al análisis de precios.

Este trabajo ha permitido continuar con el desarrollo de habilidades y competencias clave de la gestión e implementación de las TIC en el área de BI además de permitir profundizar en las nuevas herramientas de BI en la nube disponibles hoy en día y que son muy demandadas actualmente.

No es menos cierto que la consecución con éxito de este trabajo ha tenido momentos complicados debido principalmente al tiempo disponible y el desconocimiento de varias herramientas, pero en conjunto, la descomposición de tareas de la planificación inicial ha sido bastante precisa, permitiendo una gestión realista del progreso junto al tutor del proyecto y una vez finalizado el trabajo, los conocimientos adquiridos de las nuevas herramientas son de gran valor para aplicarlos a la vida profesional.

Finalmente, y en relación con el producto final, cabe destacar que se ha cumplido la reflexión inicial que ponía de manifiesto que los precios de alimentos básicos en los diferentes países no solo estaban condicionados por un único factor, como el climatológico, sino que existen multitud de causas que interfieren en la fluctuación de los precios. A pesar de esto, si es cierto que es posible encontrar correlaciones de cambio de precios y situaciones de temperatura o precipitación extrema, en determinados países y para determinados productos.

7.- GLOSARIO

Relación de términos y acrónimos más relevantes utilizados en esta memoria.

Término	Descripción
TFG	Trabajo de fin de grado
BI	Business Intelligence (<i>Inteligencia de Negocio</i>)
PIB	Producto interior bruto
GDELT	Proyecto de base de datos global de eventos, lenguaje y tono con acontecimientos sociales, políticos y geopolíticos
NOOA	National Oceanic and Atmospheric Administration
GHCN	Global Historical Climatology Network
ETL	Extract-Transform-Load (Extraer, transformar y cargar)
SISS	SQL Server Integration Services
WFP	World Food Programme
ILC	International Labor Comparisons
ADLS	Azure Data Lake Store
ADLA	Azure Data Lake Analytics
API	Application Programming Interface
JSON	Javascript Object Notation
XML	eXtensible Markup Language
AWS	Amazon Web Services
AI	Artificial Intelligence
IAAS	Infrastructure as a service
FAO	Food and Agriculture Organization

8.- BIBLIOGRAFÍA

Anónimo. (2012, agosto). "*High Food Prices: 10 Questions Answered*". World Food Programme [artículo en línea]. <<https://www.wfp.org/stories/rising-food-prices-10-questions-answered>>

Anónimo. (2014). "*5 Factors that Affect Food Prices*". All About Food [artículo en línea]. <<http://allaboutfood.aitc.ca/article/5-factors-that-affect-food-prices.php>>

Anónimo. (2014, mayo). "*Factors Affecting Food Security*". Food Security: Concepts & Issues [artículo en línea]. <<http://www.masterhdfs.org/masterHDFS/wp-content/uploads/2014/05/4-FS-CI-Ad.2.pdf>>

Anónimo. (2012, septiembre). "*What are the facts about rising food prices and their effect on the region?*". The World Bank [artículo en línea]. <http://www.worldbank.org/en/news/feature/2012/09/13/america_latina_crisis_precio_alimento>

Amadeo, K. (2018, abril). "*Why Food Prices Are Rising, Trends, and 2018 Forecast*". The Balance [artículo en línea]. <<https://www.thebalance.com/why-are-food-prices-rising-causes-of-food-price-inflation-3306099>>

Gustafson, D. (2013, septiembre). "*Rising food costs & global food security: Key issues & relevance for India*". IJMR [artículo en línea]. <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3818609/>>

Anónimo. (2012, julio). "*La inestabilidad del precio de los alimentos genera una preocupación creciente, y el Banco Mundial está preparado para responder*". The World Bank [artículo en línea]. <<http://www.bancomundial.org/es/news/press-release/2012/07/30/food-price-volatility-growing-concern-world-bank-stands-ready-respond>>

Schartz, J. (2011, junio). "*Food Prices: Eating the Cost of Logistics*". The World Bank [artículo en línea]. <<http://blogs.worldbank.org/latinamerica/food-prices-eating-the-cost-of-logistics>>

Varios autores. (2017, julio). "*Modelling the impacts of pests and diseases on agricultural systems*". Agricultural Systems [artículo en línea]. <<https://www.sciencedirect.com/science/article/pii/S0308521X1730104X>>

Anónimo. (2015, marzo). "*Keeping plant pests and diseases at bay: experts focus on global measures*". Food and Agriculture Organization of the United Nations [artículo en línea]. <<http://www.fao.org/news/story/en/item/280489/icode/>>

Smyth, S. (2017, diciembre). "*The Economics of Crop Damage*". Sustainable Agricultural Innovations & Food [artículo en línea]. <<http://saifood.ca/crop-damage-pests-and-disease/>>

The GDELT Project [sitio web]. <<https://www.gdeltproject.org/>>

Relatos del hambre [sitio web]. <<http://www.projectefam.cc/relatosdelhambre/>>

Varios autores. (2016, octubre). "*El hambre de la migración*". eldiario.es [artículo en línea]. <https://www.eldiario.es/desalambre/hambre-migracion_0_571143070.html>

- Anónimo.** (2012, agosto). "What Rising Food Prices Mean For The Fight Against Hunger". World Food Programme [artículo en línea]. <<http://www.wfp.org/aid-professionals/podcast/what-rising-food-prices-mean-fight-against-hunger>>
- Graziano, J.** (2012, septiembre). "Tackling The Root Causes Of High Food Prices And Hunger". World Food Programme [artículo en línea]. <<http://www.wfp.org/news/news-release/tackling-root-causes-high-food-prices-and-hunger>>
- Anónimo.** (2016). "Trade Logistics Global Rankings 2016". The World Bank [artículo en línea]. <<https://ipi.worldbank.org/international/global>>
- Shepherd, B.** (2011). "Logistics Costs And Competitiveness: Measurement And Trade Policy Applications". Department for International Development [documento en línea]. <<https://openknowledge.worldbank.org/bitstream/handle/10986/26724/815750WP0Trans00Box379836B00PUBLIC0.pdf?sequence=1&isAllowed=y>>
- Gerard.** (2017, noviembre). "Les unités de mesures les plus courantes en Haïti". Association Enfants-Soleil [artículo en línea]. <<https://www.enfants-soleil.org/spip.php?article122>>
- Varios autores.** (2018, mayo). "Azure y Power BI". Microsoft Docs [documento en línea]. <<https://docs.microsoft.com/es-es/power-bi/service-azure-and-power-bi>>
- Varios autores.** "U-SQL Examples and Issue Tracking". Github Azure USQL [sitio web]. <<https://github.com/Azure/usql>>
- Smith, B.** (2016, octubre). "A Fixed-Width Extractor for Azure Data Lake Analytics". Microsoft Developer Blogs [artículo en línea]. <https://blogs.msdn.microsoft.com/data_otaku/2016/10/27/a-fixed-width-extractor-for-azure-data-lake-analytics/>
- Varios autores.** (2017, junio). "Get started with U-SQL". Microsoft Docs [documento en línea]. <<https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-u-sql-get-started>>
- Varios autores.** (2018, mayo). "Develop U-SQL scripts by using Data Lake Tools for Visual Studio". Microsoft Docs [documento en línea]. <<https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-data-lake-tools-get-started>>
- Varios autores.** (2016, noviembre). "Test and debug U-SQL jobs by using local run and the Azure Data Lake U-SQL SDK". Microsoft Docs [documento en línea]. <<https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-data-lake-tools-local-run>>
- Varios autores.** (2017, diciembre). "Introduction to the U-SQL Tutorial (versión 0.7)". U-SQL Tutorial [artículo en línea]. <<https://saveenr.gitbooks.io/usql-tutorial/content/>>
- Carlo, M.** (2016, noviembre). "Grouping and Improved Date Slicer". Power Bi Tips And Tricks [artículo en línea]. <<https://PowerBI.tips/2016/11/grouping-and-improved-date-slicer/>>

Varios autores. (2017, octubre). "*Create and publish a Power BI organizational content pack (tutorial)*". Microsoft Docs [documento en línea]. <<https://docs.microsoft.com/en-us/power-bi/service-organizational-content-pack-create-and-publish>>

Varios autores. (2017, octubre). "*Organizational content packs: Copy, refresh, and get access*". Microsoft Docs [documento en línea]. <<https://docs.microsoft.com/en-us/power-bi/service-organizational-content-pack-copy-refresh-access>>

Varios autores. (2018, marzo). "*Publish to web from Power BI*". Microsoft Docs [documento en línea]. <<https://docs.microsoft.com/en-us/power-bi/service-publish-to-web>>

Varios autores. (2018, abril). "*Sorghum*". Wikipedia [artículo en línea]. <<https://es.wikipedia.org/wiki/Sorghum>>