

CONSTRUCCIÓN Y EXPLOTACIÓN DE UN ALMACÉN DE DATOS PARA EL ANÁLISIS DE INFORMACIÓN SOBRE EL TRÁNSITO RODADO DE VEHÍCULOS

Jesús Gutiérrez de la Vega
ETIG

Carles Llorach Rius
6 de Enero de 2014

DEDICATORIA Y AGRADECIMIENTOS

Antes de nada, quisiera agradecer y dedicar este trabajo a todas aquellas personas que durante estos años han estado apoyándome, guiándome en mis estudios, colaborando conmigo mediante los grupos de trabajo, así como a todas aquellas personas a las que sin cuya paciencia y ánimos recibidos, no podría haber conseguido esta meta.

RESUMEN Y PALABRAS CLAVES

Este documento, contiene la entrega final del Trabajo de Fin de Carrera¹ adscrito al área de almacenes de datos de la Universidad Oberta de Catalunya², en la titulación Ingeniería Técnica en Informática de Gestión³.

En dicho documento se pretende detallar el trabajo realizado en la asignatura durante este semestre resumiendo y unificando en la medida de lo posible todo el trabajo ya realizado durante el proyecto.

El objetivo principal del TFC es la construcción y explotación de un almacén de datos, utilizando para ello las técnicas aprendidas durante el transcurso de la titulación así como las herramientas software existentes en el mercado.

El proceso de trabajo se basa en la ejecución de un caso práctico, en el cual se presenta un escenario en el que se necesita desarrollar un almacén de datos para la *Fundació d'Estudis per a la Conducció Responsable*⁴ la cual desea estudiar la evolución del número de desplazamientos en vehículo de motor en Cataluña así como analizar las posibles correlaciones entre medios de locomoción, perfiles de conductores y algunas variables de seguridad vial.

Las palabras claves del TFC serían:

Business Intelligence, Data Warehouse, almacén de datos, estrella, hechos, dimensión, jerarquía, ETL, Pentaho Data Integration, Oracle,

¹ En adelante TFC

² En adelante UOC

³ En adelante ITIG

⁴ En adelante FECRES

ÍNDICE DE CONTENIDOS

| | | |
|--------|--|----|
| 1. | INTRODUCCIÓN | 1 |
| 1.1 | Objetivos del TFC..... | 1 |
| 1.2 | Enfoque y método seguido..... | 1 |
| 1.3 | Planificación..... | 3 |
| 1.4 | Productos obtenidos..... | 3 |
| 2. | ANÁLISIS DEL SISTEMA | 4 |
| 2.1 | Objetivo | 4 |
| 2.2 | Análisis de la Información | 4 |
| 2.2.1 | Fichero de Datos de Conductores..... | 4 |
| 2.2.2 | Fichero de Datos de Municipios..... | 6 |
| 2.2.3 | Fichero de Datos de Vehículos | 6 |
| 2.2.4 | Fichero de Datos de Radares | 7 |
| 2.3 | Datos Adicionales | 8 |
| 2.4 | Modelo Multidimensional de Datos | 8 |
| 2.4.1 | Identificación de Dimensiones y Jerarquías | 9 |
| 2.4.2 | Identificación de Hechos y Medidas | 10 |
| 2.5 | Requisitos Funcionales del Sistema | 10 |
| 2.5.1 | Requisitos de Carga de Información | 10 |
| 2.5.2 | Requisitos de Explotación de los datos..... | 11 |
| 2.5.3 | Requisitos de Informes y Cuadros de Mandos Predefinidos..... | 11 |
| 2.6 | Requisitos no Funcionales del Sistema..... | 11 |
| 2.6.1 | Requisitos de Interfaz | 12 |
| 2.6.2 | Requisitos de Prestaciones..... | 12 |
| 2.6.4 | Requisitos de Operación | 13 |
| 2.6.5 | Requisitos de Pruebas de Aceptación..... | 13 |
| 2.6.6 | Requisitos de Documentación..... | 13 |
| 2.6.7 | Requisitos de Portabilidad | 14 |
| 2.6.8 | Requisitos de Calidad | 14 |
| 2.6.9 | Requisitos de Fiabilidad | 15 |
| 2.6.10 | Requisitos de Instalación | 15 |
| 3. | DISEÑO DEL SISTEMA..... | 16 |
| 3.1 | Objetivo | 16 |
| 3.2 | Diagrama de Arquitectura | 16 |
| 3.3 | Modelo Físico de Base de Datos | 17 |
| 3.3.1 | Tabla H_FECRES..... | 18 |
| 3.3.2 | Tabla D_TIPVEH | 18 |

| | | |
|-------|---|----|
| 3.3.3 | Tabla D_TIPGEN | 19 |
| 3.3.4 | Tabla D_LOCALI | 19 |
| 3.3.5 | Tabla D_TIPPER..... | 19 |
| 3.3.6 | Tabla D_ANUALI | 19 |
| 3.4 | Proceso ETL | 20 |
| 3.4.1 | Proceso de Trabajo | 20 |
| 3.4.2 | Carga de Datos de Municipios..... | 22 |
| 3.4.3 | Carga de Datos de Vehículos | 22 |
| 3.4.4 | Carga de Datos de Radares..... | 23 |
| 3.4.5 | Carga de Datos de Conductores..... | 23 |
| 3.4.6 | Carga de Datos de la Estrella | 24 |
| 4. | APLICACIÓN ESTADÍSTICA | 25 |
| 4.1 | Introducción..... | 25 |
| 4.2 | Descripción y Configuración de la Aplicación..... | 25 |
| 4.3 | Acceso al Sistema..... | 25 |
| 4.4 | Ventana Principal | 26 |
| 4.4.1 | Menú Principal | 27 |
| 4.5 | Ventana de Detalle de un Indicador | 27 |
| 4.6 | Operaciones del Sistema..... | 29 |
| 4.6.1 | Filtrar..... | 29 |
| 4.6.2 | Navegar | 30 |
| 4.6.3 | Exportar | 30 |
| 4.6.4 | Agregar / Eliminar Favorito | 31 |
| 5. | INFORMES Y GRÁFICAS | 32 |
| 5.1 | Indicador de Datos de Conductores..... | 32 |
| 5.2 | Indicador de Datos de Conductores / Población | 33 |
| 5.3 | Indicador de Datos de Conductores / Radares | 35 |
| 5.4 | Indicador de Datos de Conductores / Vehículos | 36 |
| 5.5 | Indicador de Datos de Población | 38 |
| 5.6 | Indicador de Datos de Radares / Vehículos..... | 39 |
| 5.7 | Indicador de Datos de Vehículos | 41 |
| 5.8 | Indicador de Datos de Vehículos / Población..... | 43 |
| 5.9 | Indicador de Datos de Vehículos / Radares..... | 45 |
| 5.10 | Indicador de Densidad de Población | 46 |
| 5.11 | Indicador de Densidad de Tránsito | 48 |
| 6. | CONCLUSIONES | 50 |
| 7. | LINEAS DE EVOLUCIÓN FUTURAS | 51 |
| 8. | BIBLIOGRAFIA..... | 52 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Ilustración 1: Comparativa entre datos e información | 2 |
| Ilustración 2: Esquema ETL / Data Warehouse / Explotación..... | 2 |
| Ilustración 3: Diagrama de Gantt | 3 |
| Ilustración 4: Modelo conceptual de datos de la estrella | 9 |
| Ilustración 5: Diagrama de arquitectura del sistema | 16 |
| Ilustración 6: Modelo físico de datos | 17 |
| Ilustración 7: Distribución del modelo de datos | 20 |
| Ilustración 8: Proceso de ETL..... | 21 |
| Ilustración 9: Carga de datos de municipios | 22 |
| Ilustración 10: Carga de datos de vehículos..... | 23 |
| Ilustración 11: Carga de datos de radares | 23 |
| Ilustración 12: Carga de datos de conductores..... | 24 |
| Ilustración 13: Pantalla de logado de la aplicación estadística | 26 |
| Ilustración 14: Ventana principal de la aplicación estadística..... | 26 |
| Ilustración 15: Menú principal de la aplicación estadística..... | 27 |
| Ilustración 16: Ventana detalle de un indicador..... | 28 |
| Ilustración 17: Tabla detalle de un indicador..... | 28 |
| Ilustración 18: Menús visibles desde el detalle de un indicador | 29 |
| Ilustración 19: Pantalla de filtros del indicador | 29 |
| Ilustración 20: Pantalla de navegación entre indicadores..... | 30 |
| Ilustración 21: Pantalla de exportación de datos | 31 |
| Ilustración 22: Informe PDF generado por la aplicación..... | 31 |
| Ilustración 23: Gráfica de evolución de conductores en la comunidad autónoma | 32 |
| Ilustración 24: Grafica de evolución de los conductores por provincia | 32 |
| Ilustración 25: Evolución de los conductores por comarca | 33 |
| Ilustración 26: Evolución de los conductores respecto a habitantes en la comunidad autónoma | 33 |
| Ilustración 27: Evolución de los conductores respecto a habitantes por provincia | 34 |
| Ilustración 28: Evolución de los conductores respecto a habitantes por comarca | 34 |
| Ilustración 29: Evolución de los conductores respecto a radares en la comunidad autónoma | 35 |
| Ilustración 30: Evolución de los conductores respecto a radares por provincias | 35 |
| Ilustración 31: Evolución de los conductores respecto a radares por comarcas..... | 36 |
| Ilustración 32: Evolución de los conductores respecto a vehículos en la comunidad autónoma | 36 |
| Ilustración 33: Evolución de los conductores respecto a vehículos por provincia | 37 |

| | |
|---|----|
| Ilustración 34: Evolución de los conductores respecto a vehículos por comarca | 37 |
| Ilustración 35: Evolución de la población en la comunidad autónoma | 38 |
| Ilustración 36: Evolución de la población por provincias | 38 |
| Ilustración 37: Evolución de la población por comarcas | 39 |
| Ilustración 38: Evolución de radares respecto a vehículos en la comunidad autónoma | 39 |
| Ilustración 39: Evolución de radares respecto a vehículos por provincia | 40 |
| Ilustración 40: Evolución de radares respecto a vehículos por comarca | 40 |
| Ilustración 41: Evolución de los vehículos en la comunidad autónoma | 41 |
| Ilustración 42: Evolución de los vehículos por provincias | 41 |
| Ilustración 43: Evolución de los vehículos por comarcas | 42 |
| Ilustración 44: Distribución de vehículos por provincias | 42 |
| Ilustración 45: Distribución de vehículos por comarcas | 43 |
| Ilustración 46: Evolución de vehículos respecto a población en la comunidad autónoma | 43 |
| Ilustración 47: Evolución de vehículos respecto a la población por provincia | 44 |
| Ilustración 48: Evolución de vehículos respecto a la población por comarcas | 44 |
| Ilustración 49: Evolución de vehículos respecto a radares en la comunidad autónoma | 45 |
| Ilustración 50: Evolución de vehículos respecto a radares por provincia | 45 |
| Ilustración 51: Evolución de vehículos respecto a radares por comarca | 46 |
| Ilustración 52: Evolución de la densidad de población en la comunidad autónoma | 46 |
| Ilustración 53: Evolución de la densidad de población por provincias | 47 |
| Ilustración 54: Evolución de la densidad de población por comarcas | 47 |
| Ilustración 55: Evolución de la densidad de tránsito en la comunidad autónoma | 48 |
| Ilustración 56: Evolución de la densidad de tránsito por provincias | 48 |
| Ilustración 57: Evolución de la densidad de tránsito por comarcas | 49 |

1. INTRODUCCIÓN

1.1 Objetivos del TFC

El objetivo del proyecto es la construcción y explotación de un almacén de datos para el análisis de la información sobre el tránsito rodado de vehículos. Como se ha comentado anteriormente, el proyecto desarrolla el Trabajo de Fin de Carrera de la Ingeniería Técnica en Informática de Gestión y el mismo, está adscrito al área de almacenes de datos de dicha carrera.

En el proyecto se han realizado las tareas de planificación, análisis, diseño, implementación y explotación de un almacén de datos en bases de datos transaccionales a partir de las especificaciones y datos recibidos por parte del cliente.

Como fin principal, el proyecto trata de instruir en el diseño y construcción de almacenes de datos, con la realización de todas las fases que conlleva la realización de un sistema de tal complejidad.

1.2 Enfoque y método seguido

El proyecto se ha dividido en cuatro etapas (hitos de entrega) correspondientes a las cuatro entregas de la evaluación continua de la asignatura. En cada una de ha entregado a los consultores de la UOC el trabajo realizado.

Según se describe en el plan docente⁵, las fases del proyecto son las siguientes:

- **PLAN DE TRABAJO:** *Lo primero que se pide al estudiante es la creación de un plan de trabajo donde se deberá indicar con un cierto nivel de detalle las tareas que se deberán realizar, así como un análisis de riesgos y un diagrama de Gantt.*
- **ANÁLISIS y DISEÑO:** *En este segundo apartado del TFC, se deberá de realizar el análisis de la problemática propuesta en el enunciado y se deberá de realizar el diseño de la solución generando el modelo multidimensional de datos, así como toda la información necesaria para poder conseguir los objetivos solicitados.*
- **IMPLEMENTACIÓN:** *En esta fase se deberá crear la BDD tal y como se ha diseñado en el apartado anterior, y también se deberá cargar los datos que se disponen. En este proceso de carga deberá haber un proceso de transformación y adecuación a lo que se solicita. Una vez cargada la BDD se deberá generar los informes necesarios para poder resolver lo que se solicita. Estos informes se han de generar con la herramienta indicada por el consultor.*
- **MEMORIA Y PRESENTACIÓN VIRTUAL:** *Al final del semestre, y una vez entregadas las fases anteriores, se deberá crear la memoria del TFC donde se explicará el trabajo realizado (habitualmente será una suma y adecuación de los entregables anteriores). También se deberá crear una presentación donde se explique el trabajo realizado. Esta presentación será la que permitirá la defensa el trabajo delante del tribunal.*

Cada una de estas fases ha sido evaluada por los consultores de la UOC, y estos han realizado anotaciones y puntualizaciones para mejorar el resto de entregas y en consecuencia, el trabajo final presentado.

5

http://cv.uoc.edu/tren/trenacc/web/GAT_EXP.PLANDOCENTE?any_academico=20131&cod_a_signatura=75.134&idioma=CAS&pagina=PD_PREV_SECRE&cache=S

Durante el desarrollo del proyecto, se hará uso de las técnicas de desarrollo de Data Warehouse, el uso de motores de análisis, así como la creación y/o uso de sistemas de extracción, transformación y carga de información (ETL).

El uso de las técnicas mencionadas, permiten convertir simples datos (ficheros de texto, bases de datos inconexas, ect...) en Información, es decir que los mismos estén procesados y transformados para que éstos tengan sentido y puedan ser explotables:

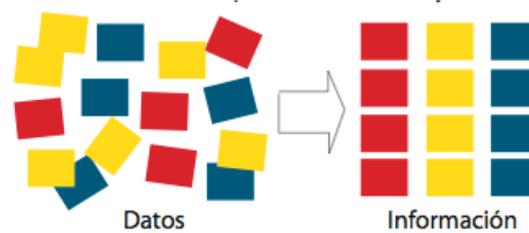


Ilustración 1: Comparativa entre datos e información

Los almacenes de datos son un paso más en el área del almacenamiento de datos. No orientados a la manipulación de datos sino a almacenarlos para hacer búsquedas y consultas:⁶

- Orientados a temas: Los datos de un mismo acontecimiento esta relacionados entre sí.
- Variable en el tiempo: Mantiene un histórico de los datos.
- No volátil: La información no se elimina ni modifica. Es de solo lectura.
- Integrado: Contiene datos de todos los Sistemas de Información.

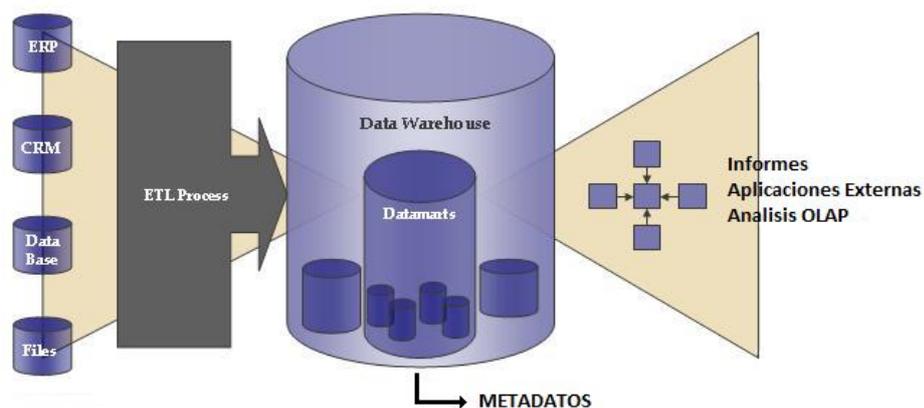


Ilustración 2: Esquema ETL / Data Warehouse / Explotación

⁶ Jesús Gutiérrez de la Vega. AXSO 2013. UOC

1.3 Planificación

A continuación se muestra el plan de actividades desglosado que se ha seguido para la realización del TFC. En él se han añadido los hitos más importantes del proceso, así como los dos actores que han tenido relación directa con el estudiante:

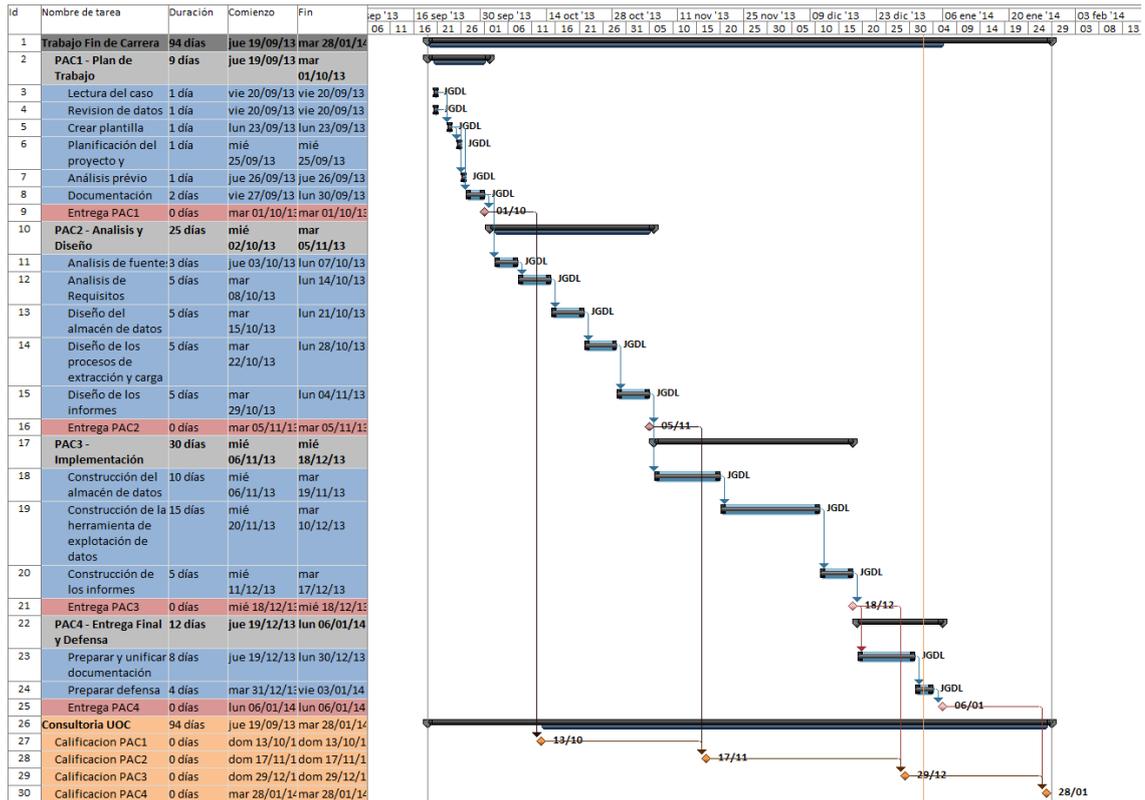


Ilustración 3: Diagrama de Gantt

1.4 Productos obtenidos

Los productos que se han obtenido durante el desarrollo del TFC han sido:

- Documentación de las diferentes PAC de la asignatura así como la presente memoria final del proyecto
- Modelo de datos del *Data Warehouse*
- Proceso de extracción, transformación y carga usando *Pentaho Data Integration* y scripts *Oracle*
- Configuración de la herramienta de análisis para la generación y visualización de informes sobre los indicadores solicitados

2. ANÁLISIS DEL SISTEMA

2.1 Objetivo

El objetivo del análisis es por un lado definir los requisitos que el cliente ha establecido, incluyendo todas las necesidades detectadas, así como la realización de un estudio minucioso de los ficheros aportados, con el fin de crear la base necesaria para la posterior fase de diseño de la aplicación.

Según nos ha transmitido el cliente, actualmente desean explotar los siguientes indicadores:

- Total de vehículos.
- Total de conductores.
- % de vehículos respecto a la población.
- Densidad de población (habitantes/km²) y densidad de tránsito (vehículos/km²).
- Número de vehículos/ Número de radares.
- % de conductores per radar.
- Indicador de conductores contra habitantes per género.
- Indicador de radares contra vehículos.
- Ratio de vehículos por conductor.
- Cantidad de vehículos por superficie del territorio.

Todos estos indicadores se pretenden explotar agregados por comarca/provincia, tipo de vehículo y tipo de permiso de conducción.

2.2 Análisis de la Información

Para poder obtener los datos estadísticos que se nos pide, FECRES nos ha proporcionado una serie de ficheros, los cuales se describen y analizan a continuación.

2.2.1 Fichero de Datos de Conductores

Se trata de un conjunto de cinco ficheros en formato TXT, que abarcan los años comprendidos entre 2007 y 2011, los cuales contienen el número de permisos y licencias de conducción distribuidos por sexo y municipio de residencia.

Todos los ficheros tienen una serie de líneas cabecera las cuales deben ser ignoradas para el tratamiento (desde la línea 1 a la línea 7, ambas incluidas). A partir de la línea 8 el formato que sigue el fichero es el siguiente:

- Provincia (Texto variable entre 6 y 9 caracteres)
- Carácter separador "-" que delimita la provincia y el municipio

- Municipio (Texto variable entre 4 y 44 caracteres)
- Nº de permisos de mujeres (número decimal delimitado por el carácter doble comilla)
- Nº de permisos de hombres (número decimal delimitado por el carácter doble comilla)
- Nº de licencias de mujeres (número decimal delimitado por el carácter doble comilla)
- Nº de licencias de hombres (número decimal delimitado por el carácter doble comilla)

El fichero contiene al final del mismo, 8 líneas que también deben ser descartadas (líneas en blanco, fuente y copyright).

En el caso del fichero de 2011, no cumple exactamente el formato mencionado con anterioridad, debido a que contiene una serie de caracteres delimitadores extra, debido a que cada campo está antecedido y sucedido por dobles comillas, teniendo como carácter separador un punto y coma. De este fichero se puede observar que la DGT ofrece en un mismo campo, la información de Provincia y Municipio, aunque para nuestro análisis es recomendable la separación con el objetivo de poder explotar esta información. El formato detallado se especifica a continuación:

- Carácter doble comilla que indica el inicio de una línea
- Provincia (Texto variable entre 6 y 9 caracteres).
- Carácter separador "-" que delimita la provincia y el municipio.
- Municipio (Texto variable entre 4 y 44 caracteres).
- Carácter delimitador de campo ";"
- Nº de permisos de mujeres (número decimal delimitado por dos caracteres doble comilla).
- Carácter delimitador de campo ";"
- Nº de permisos de hombres (número decimal delimitado por dos caracteres doble comilla).
- Carácter delimitador de campo ";"
- Nº de licencias de mujeres (número decimal delimitado por dos caracteres doble comilla).
- Carácter delimitador de campo ";"
- Nº de licencias de hombres (número decimal delimitado dos caracteres doble comilla)
- Carácter delimitador de campo ";"
- Carácter doble comilla que indica el fin de una línea.

Hay que tener en cuenta que en dichos ficheros se encuentran líneas como esta:

```
Barcelona - Barcelona (municipio sin especificar) "0.0" "1.0" "18.0"
"84.0"
```

En este caso, FECRES nos deberá de aclarar que se pretende realizar con los datos inconsistentes, pudiendo eliminar el registro, asignar los valores a la provincia, etc...

2.2.2 Fichero de Datos de Municipios

Se trata de un fichero en formato Microsoft Excel el cual contiene la población de los municipios catalanes entre los años 2007 y 2012. Además en el mismo se incluye el código INE (Código identificativo del Instituto Nacional de Estadística) del municipio, así como la extensión que tiene el mismo.

El fichero contiene una cabecera en la línea 1 la cual detalla el valor de cada columna. A continuación se detalla el contenido del mismo:

- La columna A esta denominada es "Nombre Municipio o Total Provincial y/o CC.AA", pero después de analizar los datos, solo se ha encontrado datos de municipios en la misma.
- La columna B es "Código INE" y se trata de un código numérico de cinco dígitos los cuales identifican unívocamente el municipio.
- Desde la Columna C hasta la Columna H se almacenan la población del municipio entre los años 2012 y 2007. Hay que remarcar, que existen registros (por ejemplo la fila 99 del fichero mencionado) donde para el municipio correspondiente, no existen datos poblacionales para algunos años.
- La última columna del fichero a tratar es la I "Extensión (km2)", la cual contiene un numero entero con el valor de la extensión del municipio.

2.2.3 Fichero de Datos de Vehículos

Se trata de un fichero en formato Microsoft Excel el cual sigue una estructura parecida al fichero de municipios. En él aparte del nombre del municipio y su código INE, existen varios conjuntos de columnas, algunas de ellas clasificadas por años, las cuales detallan el número de vehículos de cada categoría censados en un determinado año en el municipio. Se ha comprobado que el número de registros de este fichero y del fichero de licencias es el mismo. La estructura del fichero se detalla a continuación:

- La columna A esta denominada es "Nombre Municipio o Total Provincial y/o CC.AA", pero después de analizar los datos, solo se ha encontrado datos de municipios en la misma.
- La columna B es "Código INE" y se trata de un código numérico de cinco dígitos los cuales identifican unívocamente el municipio.
- Desde la Columna C hasta la Columna G se almacenan el número de Vehículos de motor desde los años 2011 hasta 2007. Esta columna corresponde a la suma de

otras columnas del mismo fichero [Automóviles + Camiones y Furgonetas + Otros vehículos de motor].

- Desde la Columna H hasta la Columna M, se corresponde al número de Automóviles desde el año 2012 hasta 2007.
- Desde la Columna N hasta la Columna S, se corresponde al número de Camiones y furgonetas desde el año 2012 hasta 2007.
- Desde la Columna T hasta la Columna Y, se corresponde al número de Otros vehículos de motor desde el año 2012 hasta 2007.
- La columna Z corresponde al número de Motocicletas (esta columna no aparece clasificada por años).
- La columna AA corresponde al número de Autobuses (esta columna no aparece clasificada por años).
- La columna AB corresponde al número de Tractores Industriales (esta columna no aparece clasificada por años).
- La columna AC corresponde al número del Resto de vehículos de motor (esta columna no aparece clasificada por años).

En este fichero se ha encontrado municipios como podría ser el registro 99 del mismo (el cual pertenece a "Canonja (La)") cuyos datos no están completos (faltan los datos anteriores a 2012).

Según aclaraciones realizadas por FECRES, este fichero contiene una errata en la columna "Otros vehículos de motor 2012" ya que los datos son los mismos del año 2011. En su defecto se debe tomar la suma de las Columnas Z, AA, AB, y AC, es decir, de Motocicletas, Autobuses, Tractores Industriales y Resto de Vehículos de Motor ya que estas columnas corresponden a los datos de 2012.

Del mismo modo, nos han sugerido que los datos de esos cuatro tipos de vehículos para los que no existen datos por años, se pueden estimar según su proporción entre el año 2011 y los años anteriores.

2.2.4 Fichero de Datos de Radares

Se trata de un fichero TXT, el cual contiene un listado de los radares situados en las carreteras de Cataluña. En este fichero se detalla la vía en la que se encuentra cada uno de ellos, el municipio al que pertenece la vía, la comarca y la provincia.

De este fichero se debe descartar las 12 primeras líneas que corresponden al detalle de la estructura y denominación del mismo. A continuación se detalla la estructura que sigue:

- Nombre de la Vía.

- Carácter tabulador.
- Nombre del Municipio.
- Carácter tabulador.
- Nombre de la Comarca.
- Carácter tabulador.
- Nombre de la Provincia.

2.3 Datos Adicionales

Para poder cumplir que las necesidades que FECRES nos ha descrito, se necesitan datos adicionales a los que nos ha suministrado en un primer momento, debido a que por ejemplo no existe una relación de los municipios con la comarca a la que pertenece (se podría extraer parte de esta información del fichero de radares, pero la misma quedaría incompleta). Para ello se ha recurrido al IDESCAT⁷ de donde se ha obtenido un listado de todas las comarcas catalanas así como una relación de los municipios con la comarca a la que pertenecen.

2.4 Modelo Multidimensional de Datos

Los modelos multidimensionales son modelos lógicos de los datos para organizar de una forma óptima los datos de cara al uso de los mismos en un sistema de *Business Intelligence*.

A diferencia de los modelos relacionales usados sobre todo en aplicaciones de gestión, en los modelos multidimensionales, la información y los datos están orientados a la consulta de estos agregándolos por una serie de conceptos que también se desea medir.

Los modelos dimensionales se suelen crear usando los modelos de estrella o copo de nieve.

En el modelo de estrella, sitúa la tabla principal, denominada tabla de hechos, donde se ubican los datos a medir y explotar en el centro del modelo. De ella, y solo de ella, parten una serie de tablas denominadas dimensiones, que contienen los datos por los que se mide la tabla de hechos.

El modelo de copo de nieve se basa en el mismo principio del modelo de estrella, pero permite jerarquías a nivel de tablas en las dimensiones.

La elección de un tipo u otro depende básicamente del rendimiento y de la necesidad de almacenamiento que se necesite. El modelo en estrella penaliza el almacenamiento al tener las dimensiones normalmente desnormalizadas, mientras que el modelo en copo de nieve penaliza el rendimiento al tenerlas normalizadas (obliga a realizar un mayor procesamiento para mostrar los datos al tener que hacer más uniones las consultas).

⁷ IDESCAT. Instituto de Estadística de Cataluña. <http://www.idescat.cat/es>

En nuestro caso, se ha optado por el uso de un modelo en estrella, donde una de las dimensiones se ha desnormalizado para recoger todos los datos de localización. A continuación se muestra el diagrama del modelo de datos que se ha generado tras el análisis de la información aportada:

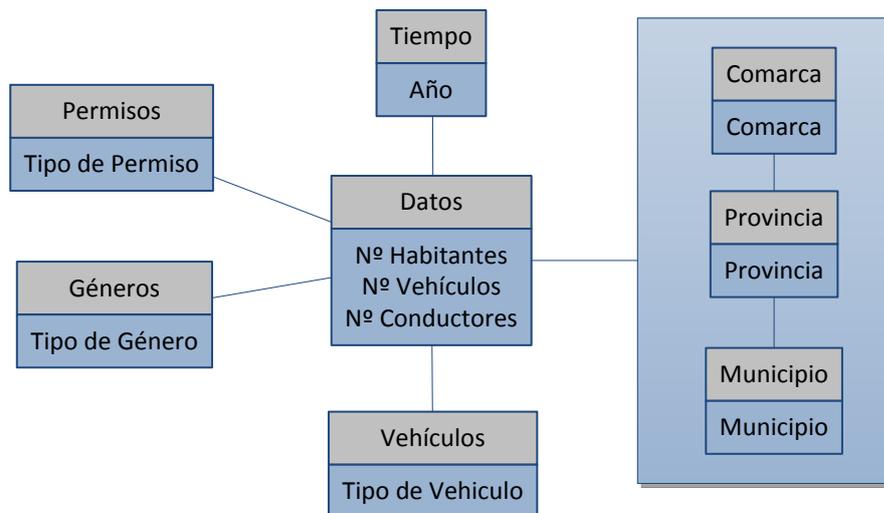


Ilustración 4: Modelo conceptual de datos de la estrella

2.4.1 Identificación de Dimensiones y Jerarquías

Se han identificado cuatro dimensiones alrededor de la tabla de hechos:

- Dimensión Tiempo, que permite agrupar y filtrar los datos de la tabla de hechos entre los diferentes años que se hubiesen cargado en el sistema.
- Dimensión Tipo de Vehículo, que permite realizar un estudio de los datos referentes al número de vehículos agrupando o filtrando los mismos entre los diferentes tipos de vehículos que se han definido anteriormente:
 - Automóviles.
 - Camiones y furgonetas.
 - Motocicletas.
 - Autobuses.
 - Tractores industriales.
 - Resto vehículos de motor.
- Dimensión Género, que permite agrupar y filtrar los datos referentes al número de conductores agrupando los mismos entre hombres y mujeres.

- Dimensión Tipo de Permiso, que permite agrupar y filtrar los datos referentes al número de conductores agrupando los mismos entre poseedores de licencias de conducción o permisos de conducción.
- Dimensión Localización, que permite agrupar y filtrar los datos de conductores y vehículos por la provincia, comarca y municipio al que pertenecen. Esta dimensión se encuentra jerarquizada y desnormalizada, agrupando en la misma las provincias, comarcas y municipios

2.4.2 Identificación de Hechos y Medidas

La tabla de hechos que se ha identificado, tendría solo dos medidas:

- Número de Vehículos
- Número de Conductores

Estas medidas estarían disgregadas por todas las dimensiones que se han descrito antes, es decir, tipos de vehículo, tipo de permiso, tiempo, localización y género.

Aunque en la tabla de hechos se podría haber optado por realizar un precálculo de todos los indicadores que nos solicitan desde FECRES, ello podría ocasionar problemas de incongruencia de datos a la hora de realizar agrupaciones de los mismos (porque dependerían del total de resultados mostrados). Este problema, acompañado del número de registros previstos que contenga el sistema (de donde se ve que no habrá problemas de rendimiento al realizar las consultas), hace que se tome esta opción.

2.5 Requisitos Funcionales del Sistema

Los requisitos funcionales del sistema se tratan de características requeridas del sistema que expresan una capacidad de acción de los mismos así como una funcionalidad, la cual está expresada generalmente de forma verbal.

A continuación se detallan cada uno de ellos agrupados por el ámbito de los mismos:

2.5.1 Requisitos de Carga de Información

| Código | Descripción |
|------------|--|
| FU-CIN-001 | Se creará un sistema que permita cargar ficheros de datos en la base de datos del sistema de información. |
| FU-CIN-002 | El sistema permitirá cargar los ficheros de ofrecidos por FECRES: datos de conductores, datos de vehículos, datos de municipios y radares. |
| FU-CIN-003 | El sistema permitirá cargar también los ficheros adicionales necesarios para la correcta manipulación de los datos, correspondientes a la relación de municipios y comarcas. |

| | |
|------------|--|
| FU-CIN-004 | El sistema controlará la carga de una forma adecuada para que en ningún momento un proceso de carga fallido deje sin datos el sistema. |
|------------|--|

| | |
|------------|---|
| FU-CIN-005 | El proceso de extracción de datos desde los ficheros origen, procesará los mismos, tratando posibles fallos de datos en los mismos, resolviéndolos siempre que sea posible. También deberá transformar esos datos para convertirlos al modelo multidimensional definido a tal efecto. |
|------------|---|

2.5.2 Requisitos de Explotación de los datos

| Código | Descripción |
|------------|--|
| FU-EDA-001 | El sistema permitirá explotar los siguientes datos: <ul style="list-style-type: none"> • Total de vehículos. • Total de conductores. • % de vehículos respecto a la población. • Densidad de población (habitantes/km²) y densidad de tránsito (vehículos/km²). • Número de vehículos/ Número de radares. • % de conductores per radar. • Indicador de conductores contra habitantes per género. • Indicador de radares contra vehículos. • Ratio de vehículos por conductor. • Cantidad de vehículos por superficie del territorio. |
| FU-EDA-002 | Todos los datos se podrán explotar agregando los mismos por comarca / provincia, tipo de vehículo, tipo de permiso de conducción, género y año. |

2.5.3 Requisitos de Informes y Cuadros de Mandos Predefinidos

| Código | Descripción |
|------------|--|
| FU-ICM-001 | Se definirá un conjunto de informes predefinidos que cubra las necesidades descritas por FECRES. Al menos se definirá un informe para cada uno de los datos que se desea explotar. |
| FU-ICM-002 | El cuadro de mandos a realizar permitirá navegar por cada uno de los informes realizados, permitiendo exportar los datos a diferentes formatos para su procesamiento externo. |

2.6 Requisitos no Funcionales del Sistema

Los requisitos no funcionales del sistema se tratan de características requeridas del sistema, del proceso de desarrollo, del servicio prestado o de cualquier otro aspecto del desarrollo, que señala una restricción al propio sistema.

A continuación se detallan cada uno de ellos agrupados por la naturaleza de los mismos:

2.6.1 Requisitos de Interfaz

Herramienta de Extracción, Transformación y Carga

| Código | Descripción |
|--------------|--|
| NF-ETL-001 | El proceso ETL tendrá una interfaz gráfica amigable que permita realizar la carga de una forma fácil y cómoda para los administradores del sistema |
| NF- ETL -002 | La interfaz gráfica del proceso ETL guiará al usuario solicitando todos los datos necesarios para realizar la carga del sistema. |

Acceso y explotación de la información

| Código | Descripción |
|------------|--|
| NF-AEI-001 | El cuadro de mandos tendrá una interfaz clara y simple que permita de un vistazo consultar los datos deseados. |
| NF-AEI-002 | El sistema será multiusuario, donde cada persona podrá ver una serie de consultas personalizadas. |

2.6.2 Requisitos de Prestaciones

Requisitos de capacidad

| Código | Descripción |
|------------|--|
| NF-CAP-001 | El sistema estará preparado para la posible ampliación de datos en años venideros (nuevos datos temporales, nuevos tipos de permisos, etc...) así como la posible expansión del sistema a otras comunidades. |
| NF-CAP-002 | El sistema deberá manejar el acceso de múltiples usuarios concurrentes a la aplicación. |

Tiempo de respuesta

| Código | Descripción |
|------------|--|
| NF-TDR-001 | El tiempo de respuesta de las consultas al sistema debe ser el menor posible, asegurando que el sistema está optimizado según los recursos proporcionados. |
| NF-TDR-002 | El tiempo de procesado del proceso ETL debe ser el menor posible, usando el menor número de recursos posibles. |

2.6.4 Requisitos de Operación

Acceso de usuarios

| Código | Descripción |
|------------|--|
| NF-AUS-001 | El acceso de los usuarios se realizará mediante usuario y clave. El sistema validará si el acceso está permitido o no. |
| NF-AUS-002 | Se registrará la fecha del último acceso del usuario. |

2.6.5 Requisitos de Pruebas de Aceptación

Pruebas de Aceptación de Sistema

| Código | Descripción |
|------------|---|
| NF-PAS-001 | FECRES realizará las pruebas de aceptación del sistema de acuerdo con la documentación entregada. |
| NF-PAS-002 | La aceptación del sistema se producirá si se cumplen todos los requisitos solicitados por FECRES. En caso contrario se rechazara el producto. |

Pruebas de Aceptación de Negocio

| Código | Descripción |
|------------|---|
| NF-PAN-001 | Cada componente del sistema, se evaluará de forma individual. Todos ellos deben de funcionar de forma adecuada y siguiendo las especificaciones previstas. |
| NF-PAN-002 | La aceptación del sistema se producirá si cada componente del sistema cumple todos los requisitos solicitados por FECRES. En caso contrario se rechazara el producto. |

2.6.6 Requisitos de Documentación

Memoria

| Código | Descripción |
|------------|---|
| NF-MEM-001 | Se realizará una memoria del proyecto que englobará todas las fases del proyecto. |
| NF-MEM-002 | Se incluirá una presentación multimedia que clarifique el trabajo realizado. |

Presentación del producto

| Código | Descripción |
|------------|---|
| NF-PPR-001 | El producto se presentará a FECRES mediante una presentación multimedia. |
| NF-PPR-002 | Para la presentación se usarán los recursos técnicos necesarios (diapositivas, audio, video...) |

Modelo de Datos

| Código | Descripción |
|------------|--|
| NF-MDA-001 | Se realizará una documentación detallada sobre el modelo de datos. |
| NF-MDA-002 | Se proporcionará todos los datos necesarios para entender y modificar el modelo por FECRES o terceras personas si el cliente lo considerase necesario. |

2.6.7 Requisitos de Portabilidad

Plataformas

| Código | Descripción |
|------------|---|
| NF-PLA-001 | El proceso de extracción, transformación y carga de datos se realizará usando tecnologías de software libre que permita funcionar en entornos multiplataforma. |
| NF-PLA-002 | El sistema gestor de base de datos que se utilizara para la implantación del sistema será Oracle. Esto no exime que en un futuro se desee modificar el mismo, por lo que el código que se realice debe ser lo más portable posible. |

2.6.8 Requisitos de Calidad

Pruebas de sistema

| Código | Descripción |
|------------|---|
| NF-PSI-001 | El sistema será probado por parte de los desarrolladores antes de realizar la entrega para comprobar que se entrega un software que cumple con los estándares necesarios. |
| NF-PSI-002 | El sistema deberá pasar por pruebas que comprueben su robustez, su seguridad y su tolerancia a fallos. |

Calidad de la información en el sistema

| Código | Descripción |
|------------|---|
| NF-CIS-001 | Se asegurará que la información que esté almacenada en el sistema no contiene errores, que pudieran producir falsos resultados en las estadísticas y consultas realizadas por FECRES. |
| NF-CIS-002 | Se asegurará que los datos almacenados son los mínimos necesarios para la ejecución del sistema, no almacenándose datos temporales, ni inconsistentes tras los procesos de carga. |

2.6.9 Requisitos de Fiabilidad

Tolerancia y recuperación a fallos

| Código | Descripción |
|------------|--|
| NF-TRF-001 | El sistema será capaz de recuperarse ante caídas de alguno de los componentes del mismo. |
| NF-TRF-002 | Si la base de datos no estuviese accesible, el sistema debe proporcionar un error amigable, no dando nunca un error fatal no controlado. |

2.6.10 Requisitos de Instalación

Control del versionado

| Código | Descripción |
|------------|--|
| NF-CVE-001 | El sistema tendrá un correcto sistema de versionado, que evitará posibles problemas de entregas ante correcciones del mismo. Se usará un sistema de tres dígitos, donde el primer dígito indicará grandes evolutivos del sistema, el segundo modificaciones que no afectan a la totalidad del sistema ni pone en riesgo la integridad del mismo, y el tercer dígito que estará reservado a las correcciones del sistema. |
| NF-CVE-002 | Se realizará versionado tanto del software como de los script de base de datos que son necesarios lanzar para que el sistema funcione correctamente. |
| NF-CVE-003 | Si se tuviese que realizar modificaciones en el sistema que originase cambios en base de datos, se proporcionará tanto scripts aumentativos como scripts completos con el objeto de poder reinstalar el sistema sin tener que lanzar todos los anteriores o poder actualizar el sistema desde una versión previa. |

3. DISEÑO DEL SISTEMA

3.1 Objetivo

El objetivo de la fase de diseño es obtener un modelo del software a realizar más cercano a la realidad, que permita realizar la fase de implementación de una forma adecuada. Del mismo modo, se presenta una aproximación de la globalidad del sistema mediante los diagramas de arquitectura y del proceso ETL.

3.2 Diagrama de Arquitectura

El diagrama de arquitectura se trata de una imagen que representa de una forma gráfica y simple el sistema que se pretende desarrollar, además de tener una idea global de todos los actores y componentes que intervienen en el.

A continuación se muestra el diagrama correspondiente a nuestro sistema:

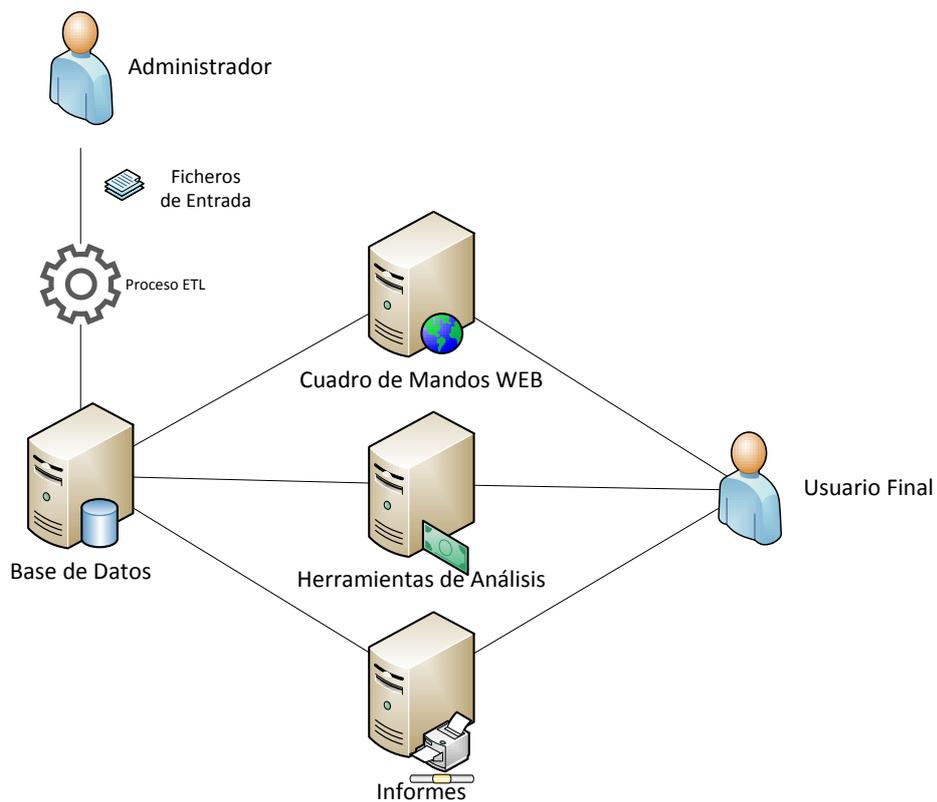


Ilustración 5: Diagrama de arquitectura del sistema

Los componentes del sistema serían el proceso ETL, la base de datos y las diferentes herramientas de explotación de la información que se usen. En este caso, la base de datos

en si se considera un componente ya que parte de la lógica del procesamiento de la información residirá en la propia BBDD.

3.3 Modelo Físico de Base de Datos

El modelo físico de base de datos ayuda a definir detalladamente la estructura que tendrá la base de datos, pasando de un modelo conceptual a un modelo más cercano en interpretable por un motor de base de datos.

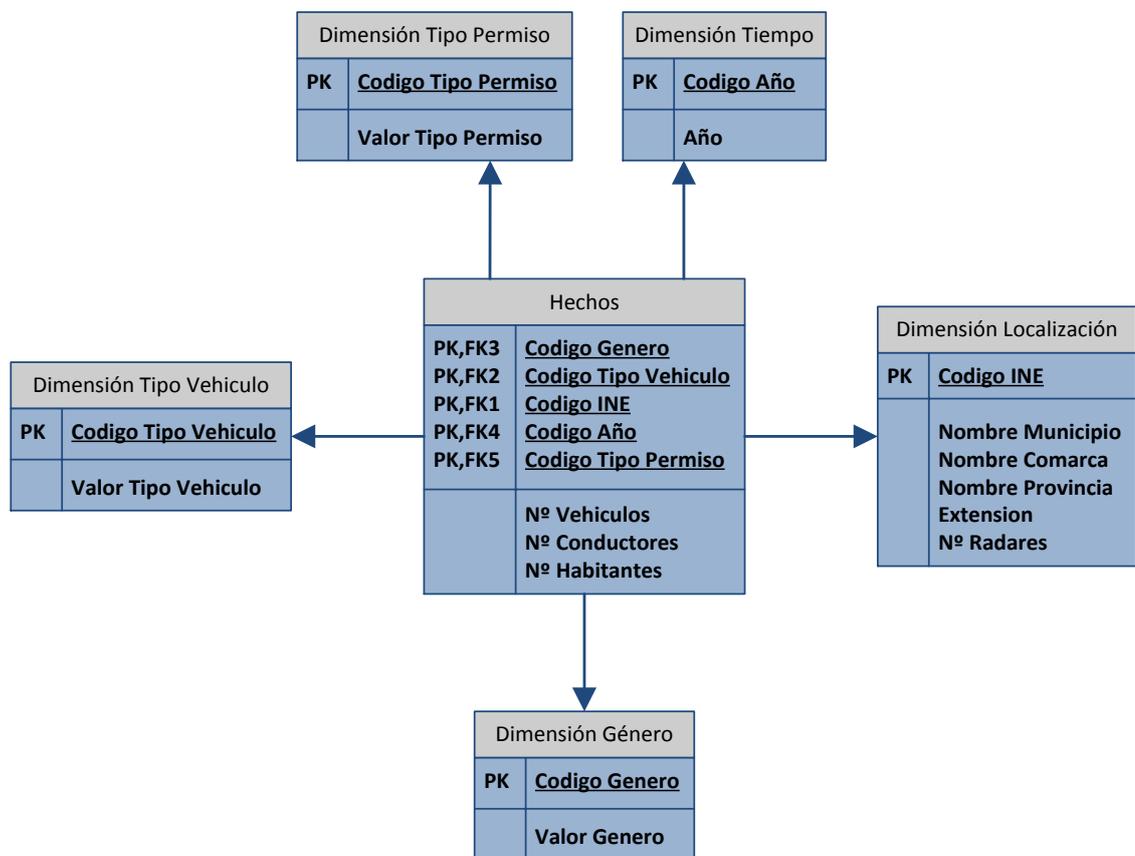


Ilustración 6: Modelo físico de datos

En nuestro caso, el modelo conceptual que previamente habíamos definido, se ha convertido en un conjunto de seis tablas, las cuales se pasan a describir:

3.3.1 Tabla H_FECRES

Descripción: Tabla de hechos de la estrella. Contiene los valores a medir agregados por las dimensiones por las que se desea medir.

| Nombre | Tipo | Restricciones | Descripción |
|-------------|-------------|----------------------|--|
| X_FECR | NUMBER(10) | Clave Primaria | Clave primaria de la tabla de hechos de la estrella. |
| TIVE_X_TIVE | NUMBER(2) | Clave Ajena, No Nulo | Clave ajena de la tabla Tipos de Vehículos. |
| TIGE_X_TIGE | NUMBER(1) | Clave Ajena, No Nulo | Clave ajena de la tabla Tipos de Géneros. |
| LOCA_C_LOCA | VARCHAR2(5) | Clave Ajena, No Nulo | Clave ajena de la tabla Localidades. |
| TIPE_X_TIPE | NUMBER(1) | Clave Ajena, No Nulo | Clave ajena de la tabla Tipos de Permisos. |
| ANUA_X_ANUA | NUMBER(2) | Clave Ajena, No Nulo | Clave ajena de la tabla Anualidades. |
| N_VEHI | NUMBER(8) | No Nulo | Valor del número de vehículos agregado por las dimensiones relacionadas. |
| N_COND | NUMBER(8) | No Nulo | Valor del número de conductores agregado por las dimensiones relacionadas. |
| N_HABI | NUMBER(8) | No Nulo | Valor del número de habitantes agregado por las dimensiones relacionadas. |

3.3.2 Tabla D_TIPVEH

Descripción: Tabla de la dimensión Tipo de Vehículo.

| Nombre | Tipo | Restricciones | Descripción |
|--------|--------------|----------------|---|
| X_TIVE | NUMBER(2) | Clave Primaria | Clave primaria de la tabla de tipos de vehículos. |
| D_TIVE | VARCHAR2(30) | No Nulo | Descripción del tipo de vehículo. |

3.3.3 Tabla D_TIPGEN

Descripción: Tabla de la dimensión Tipo de Género.

| Nombre | Tipo | Restricciones | Descripción |
|--------|--------------|----------------|---|
| X_TIGE | NUMBER(1) | Clave Primaria | Clave primaria de la tabla de tipos de géneros. |
| D_TIGE | VARCHAR2(30) | No Nulo | Descripción del tipo de género. |

3.3.4 Tabla D_LOCALI

Descripción: Tabla de la dimensión Localización.

| Nombre | Tipo | Restricciones | Descripción |
|---------|--------------|----------------|---|
| C_LOCA | VARCHAR2(5) | Clave Primaria | Clave primaria de la tabla de localizaciones. |
| D_MUNI | VARCHAR2(50) | No Nulo | Denominación del municipio. |
| D_COMA | VARCHAR2(50) | No Nulo | Denominación de la comarca a la que pertenece el municipio. |
| D_PROVI | VARCHAR2(20) | No Nulo | Denominación de la provincia a la que pertenece el municipio. |
| N_EXTE | NUMBER(5) | No Nulo | Extensión del municipio. |
| N_RADA | NUMBER(2) | No Nulo | Número de radares del municipio. |

3.3.5 Tabla D_TIPPER

Descripción: Tabla de dimensión Tipo de Permiso.

| Nombre | Tipo | Restricciones | Descripción |
|--------|--------------|----------------|--|
| X_TIPE | NUMBER(1) | Clave Primaria | Clave primaria de la tabla de tipos de permisos. |
| D_TIPE | VARCHAR2(30) | No Nulo | Descripción del tipo de permiso. |

3.3.6 Tabla D_ANUALI

Descripción: Tabla de la dimensión A anualidades.

| Nombre | Tipo | Restricciones | Descripción |
|--------|--------------|----------------|---|
| X_ANUA | NUMBER(2) | Clave Primaria | Clave primaria de la tabla de anualidades |
| D_ANUA | VARCHAR2(30) | No Nulo | Descripción de la anualidad. |

3.4 Proceso ETL

3.4.1 Proceso de Trabajo

El proceso de trabajo por un lado ha consistido en la generación del modelo de base de datos necesario para la carga de todos los datos, basando el mismo en la fase anterior de análisis, añadiendo además la *Staging Area* y tablas temporales (copias de las tablas finales donde se carga en un momento intermedio los mismos).

El modelo contiene las siguientes tablas:



Ilustración 7: Distribución del modelo de datos

Una vez que el modelo de datos estaba completo se realizó usando Pentaho Data Integration⁸ un sistema de que procesaba los ficheros origen proporcionado por FECRES, realizando con esta herramienta tareas como la limpieza de datos, cálculo de datos no existentes, separación de datos, etc... Toda esta información finalmente se cargó a la *Staging Area* la cual se procesaría más tarde mediante consultas DML.

Una vez cargados los datos que se disponían de FECRES, se realizó una segunda carga de información de datos adicionales que no ofrecía FECRES (datos de comarcas, y la asociación de las mismas con localidades). Debido a que el formato que tenían estos datos no era fácil de explotar ni transformar, se realizó la misma manualmente mediante la aplicación Excel, mediante la cual se generaron una serie de sentencias Inserts de base de datos para cargar el resto de tablas de la *Staging Area* que faltaban.

Con todas las tabla de *la Staging Area* cargadas, ya se podía realizar el proceso de carga que llenase las tablas de dimensión y hechos. Según los requisitos, se solicitaba que en ningún momento el sistema pudiera dejar vacías por error el almacén de datos, por lo que se creó un mecanismo por el cual la información que se está cargando se inserta en una "replica" del sistema final a usar, y si el proceso ha ido correctamente, el sistema intercambia la réplica con las tablas finales en un proceso atómico. Esto evita que si durante la carga de las tablas de dimensión y hechos se produce un error de claves, duplicidades, etc... no se quede vacío el almacén de datos final.

⁸ En adelante PDI

El proceso queda explicado en el siguiente diagrama:

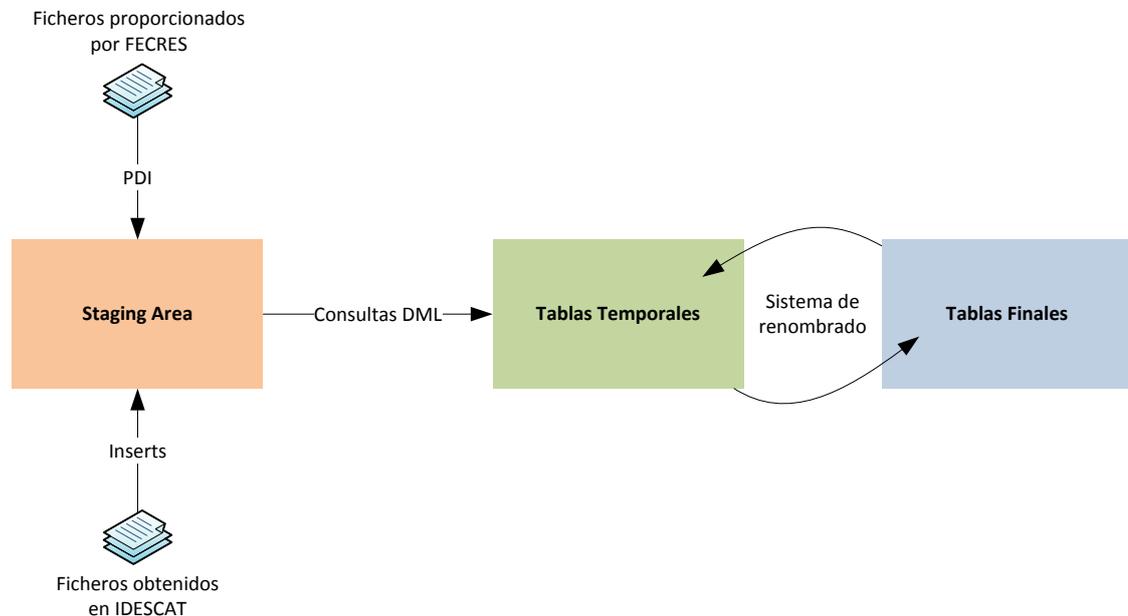


Ilustración 8: Proceso de ETL

Una vez que el sistema de tablas finales estaba cargado se pasó a realizar todos los informes de los indicadores que se habían analizado.

Para ello se usó la herramienta *Resumen Ejecutivo*. Esta herramienta permite la creación de informes y graficas basándose en modelos de base de datos en estrella (aunque por su naturaleza, puede acceder a cualquier otro tipo de modelo, como relacional o copo de nieve, aunque el procesamiento y rendimiento puede no ser el mismo).

Para configurar la herramienta se desarrollaron todas las consultas SQL que eran necesarios para cada uno de los informes y con ellas se creó los registros necesarios en el modelo de datos de esta herramienta (graficas, agrupaciones, indicadores, filtros, listas de valores...)

En el desarrollo de toda la solución se han tenido varios problemas que a priori no se habían detectado. Entre ellos se puede citar:

- No se había detectado que faltaban municipios en los datos que FECRES había proporcionado. Después de realizar algunas consultas se vio que solo se había proporcionado información de los municipios que tenían más de 10,000 habitantes. La solución que se adoptó fue incluir en todos los informes una leyenda advirtiendo de este posible descuadre de datos con la realidad.

- La aplicación *Resumen Ejecutivo* tenía diversos bug's en la visualización de los datos, los cuales tuvieron que ser solventados (tenía fallos en el cálculo de ciertos tipos de gráficas así como en la representación de algunos tipos de datos).
- Se tuvo problemas con la máquina virtual que se proporcionó para el desarrollo. Debido a una limitación de licencias, estas se apagan a las 8 horas de uso, lo que al principio provocó varias pérdidas de datos inesperados. Tras varias consultas, se optó por reiniciar la máquina cuando estuviese a punto de agotarse las 8 horas de uso para evitar problemas.

3.4.2 Carga de Datos de Municipios

Para realizar la carga de municipios en primer lugar se lee el fichero origen proporcionado por FECRES y se le sustituyen los valores no definidos ("n.d.") con 0. Una vez que se tienen los datos formateados correctamente ya se pueden pasar a base de datos, pero como el fichero tiene los datos en columnas y para cargarlo en la tabla del *Staging Area* nos interesa en filas, creamos filas con los valores necesarios usando el componente calculadora, de cuya salida tenemos filas preparadas para insertar en BBDD.

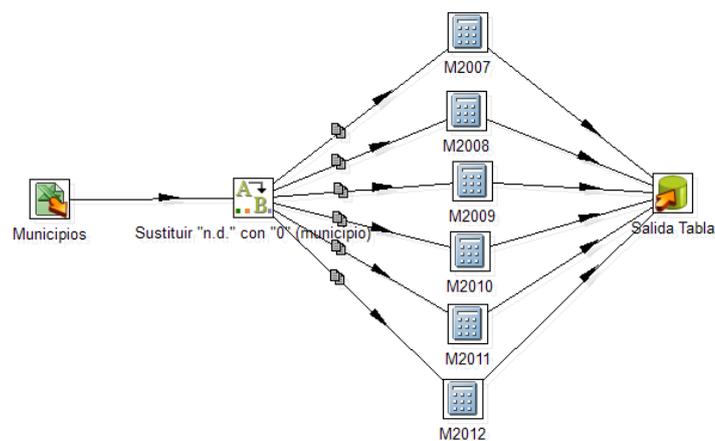


Ilustración 9: Carga de datos de municipios

3.4.3 Carga de Datos de Vehículos

Del mismo modo que se realiza en los municipios, para la carga de vehículos se lee el fichero Excel suministrado por FECRES como origen y se sustituyen los valores no definidos por 0. En este caso se detectó que algunos campos tenían espacios en blancos, por lo que se vuelve a aplicar otro filtro de cadenas, para eliminarlos. Posteriormente, una vez que todos los datos informados tienen valores numéricos, se cambia el tipo de campo a Integer para poder realizar cálculos con ellos, y utilizando el componente de fórmula se calculan todos los valores faltantes (anualidad anteriores, recálculo de columnas, totales...). Una vez se tiene

todos los datos, se realiza la conversión de columnas a filas de los datos para insertarlos en BBDD.

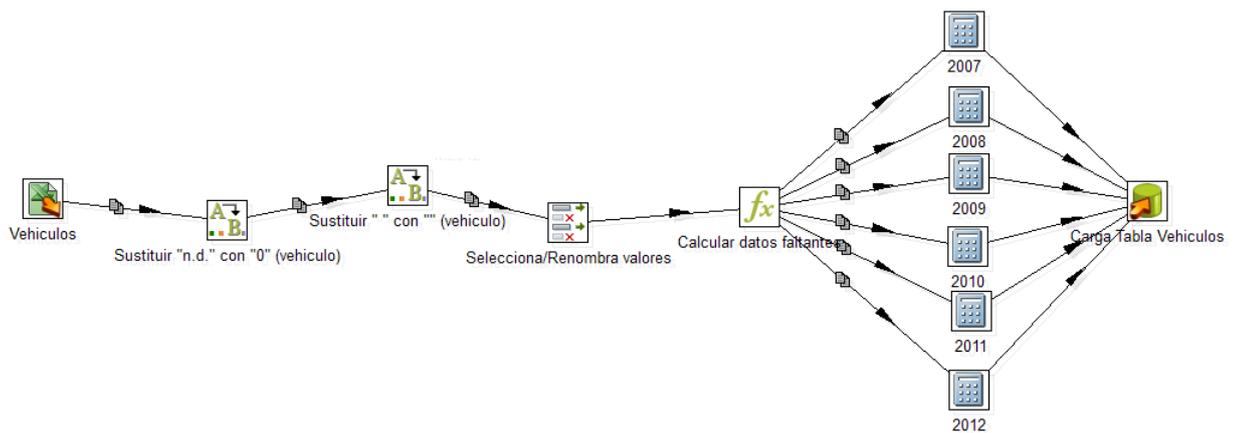


Ilustración 10: Carga de datos de vehículos

3.4.4 Carga de Datos de Radares

La carga de radares es el proceso más simple de todos los ficheros, ya que de fichero se puede extraer los campos simplemente dividiéndolo por el carácter tabulador, y directamente se puede cargar en BBDD

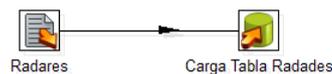


Ilustración 11: Carga de datos de radares

3.4.5 Carga de Datos de Conductores

Para el fichero de conductores es necesario aplicar dos tratamientos ya que el fichero de 2011 tiene un formato diferente (tiene un separador de campo diferente y caracteres comillas para delimitar los campos también). Una vez que se leen los ficheros, se separa el campo de provincia y municipio para poder tratarlo de una mejor forma, se eliminan las comillas si es que las tienen, se añade el año al que pertenece el fichero y se insertan los datos en BBDD.

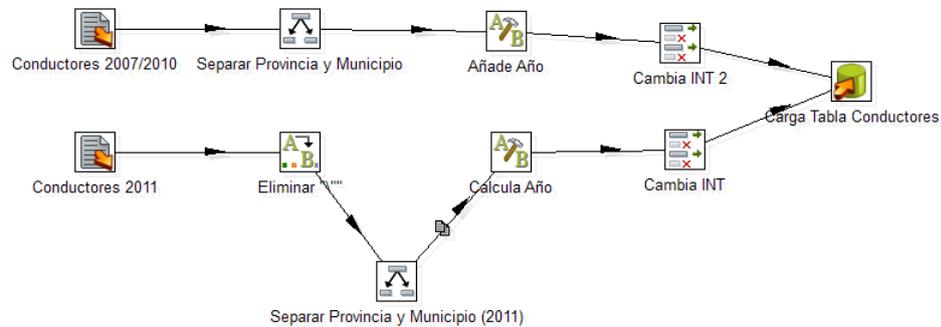


Ilustración 12: Carga de datos de conductores

3.4.6 Carga de Datos de la Estrella

Una vez lanzado el proceso de carga de datos, se lanza los scripts siguientes:

- *03 - CARGA NEW_D.sql*: Script encargado de cargar las tablas de dimensión con los datos estáticos y la dimensión de Localidades con los datos procedentes de la *Staging Area*
- *04 - CARGA NEW_H.sql*: Script encargado de cargar la tabla de hechos de la estrella. Utilizando la *Staging Area* así como la nuevas tablas de dimensiones cargadas, crea la nueva estrella que se pondrá a disposición del usuario
- *05 - RENAME.sql*: Script encargado de renombrar las tablas una vez que estas han sido cargadas
- *06 - LIMPIA.sql*: Script encargado de vaciar las tablas temporales para dejar el esquema con el menor consumo de espacio necesario.

4. APLICACIÓN ESTADÍSTICA

4.1 Introducción

Para la explotación estadística del sistema se ha usado la herramienta *Resumen Ejecutivo*, aplicación que permite la generación y exportación de informes y graficas las cuales se han predefinido anteriormente. Se ha elegido esta opción debido a la facilidad de configuración que tiene, la potencia de la herramienta en el uso de tipos de gráficas, filtros, etc... aparte de la experiencia en el uso de la misma al haberla desarrollado tiempo atrás.

4.2 Descripción y Configuración de la Aplicación

La aplicación Resumen Ejecutivo se trata de una aplicación web desarrollada usando el framework Struts de Java. Para su ejecución se necesita el uso de un contenedor de aplicaciones web como Tomcat o Jboss y una máquina virtual Java igual o superior a la versión 5. El fichero desplegable de la aplicación se distribuye en la entrega para su instalación en otros entornos.

Antes del uso de la misma es necesario configurar los ficheros de propiedades de la aplicación (situados dentro de la carpeta WEB-INF bajo la ruta de despliegue). En esta carpeta se debe de modificar los ficheros *hibernate.cfg.xml* y *log4j.properties* con los datos de acceso a base de datos y con los datos donde se almacena los ficheros de log de la aplicación.

Previamente a la ejecución de la aplicación se debe lanzar los scripts de creación del modelo propio de base de datos suministrado, así como los datos iniciales para el mismo

4.3 Acceso al Sistema

Para acceder al sistema, el usuario deberá acceder al él a través de su navegador web, simplemente accediendo a la dirección que el Servicio de Informática le indique. En la actualidad el sistema se encuentra en la dirección: <http://<IP-AMAZON>:8080/TFC/>

El sistema nos presentará una ventana de acceso para identificarnos donde se debe de introducir el usuario y contraseña que nos han proporcionado previamente.



Ilustración 13: Pantalla de logado de la aplicación estadística

Una vez cumplimentados ambos datos y pinchando sobre el botón ENTRAR, accedemos a la ventana principal del sistema.

4.4 Ventana Principal

La ventana principal que muestra el sistema es la siguiente:



Ilustración 14: Ventana principal de la aplicación estadística

Sobre la parte superior derecha de la ventana se puede observar el nombre del usuario conectado. También aparece el botón "Cerrar sesión" que sirve para salir del sistema y volver a la pantalla inicial de autenticación.

En la parte superior izquierda se encuentra el menú principal del sistema. Este menú se presenta de forma horizontal y más adelante detallaremos su funcionamiento.

Y finalmente en el centro de la ventana se muestra el contenedor de agrupaciones favoritas del usuario conectado. Una agrupación no es más que una forma de presentar los datos correspondientes a un determinado indicador.

4.4.1 Menú Principal

El menú contiene dos apartados:

- **Inicio:** mediante el uso del mismo se puede volver a la ventana principal desde cualquier parte de la aplicación.
- **Indicadores Estadísticos:** contiene un conjunto de indicadores de diversas temáticas predefinidas en la aplicación.

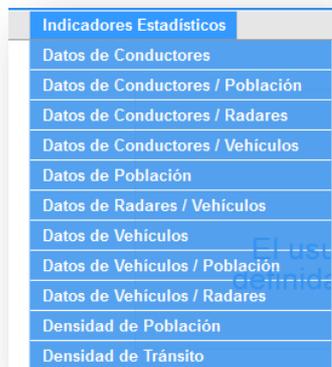


Ilustración 15: Menú principal de la aplicación estadística

4.5 Ventana de Detalle de un Indicador

Si desde el menú anterior se pulsa sobre algún indicador la ventana de detalle del mismo que se nos presenta es la siguiente:



Ilustración 16: Ventana detalle de un indicador

Todo indicador tiene asociado una agrupación por defecto y es ésta la que siempre se muestra si el acceso a esta ventana es a través del menú. Como se verá más adelante existen dos formas más de acceder a esta pantalla de detalle.

| Año | Hombres | Mujeres |
|------|-----------|-----------|
| 2007 | 2.349.892 | 1.594.609 |
| 2008 | 2.352.742 | 1.610.924 |
| 2009 | 2.359.801 | 1.619.272 |
| 2010 | 2.367.621 | 1.643.558 |
| 2011 | 2.372.260 | 1.664.106 |

Ilustración 17: Tabla detalle de un indicador

La imagen superior muestra la tabla de datos del indicador desplegada:

4.6 Operaciones del Sistema

4.6.1 Filtrar

Los datos que se muestran de forma gráfica o tabular son el resultado de una consulta que se realiza al almacén de datos del sistema. Esta consulta ejecutada sobre una determinada agrupación puede contener o no filtros. Un filtro no es más que una condición que debe cumplir los datos resultantes.

Para acceder a la ventana de filtrar hay que pulsar el botón "Filtrar".

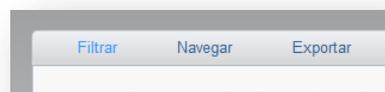


Ilustración 18: Menús visibles desde el detalle de un indicador

La ventana emergente muestra un conjunto de filtros y cada filtro puede contener un conjunto de valores.

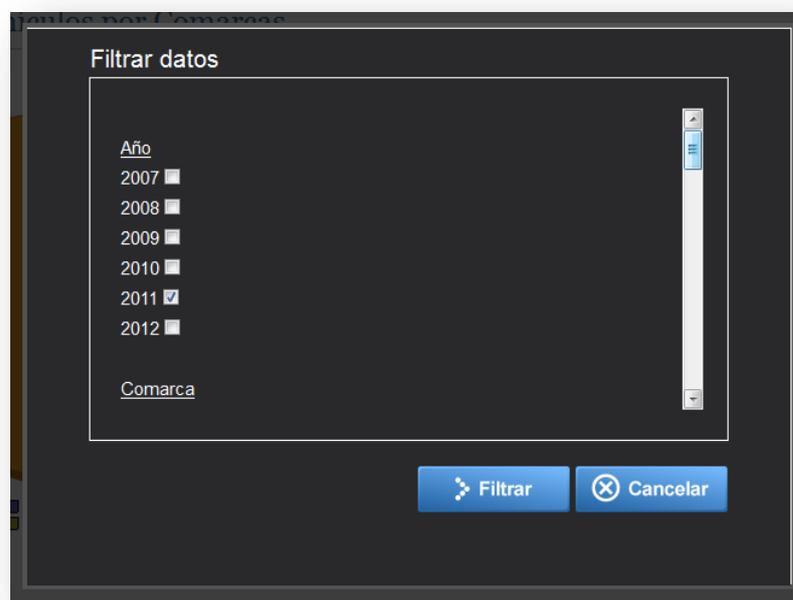


Ilustración 19: Pantalla de filtros del indicador

En el ejemplo anterior se puede ver como existe en este caso el filtro "Año". El valor inicialmente marcado para este filtro es el año 2007.

Sin embargo si seleccionamos otros valores de la lista, pulsamos en "Filtrar" y volvemos a la ventana de detalle mostrada anteriormente se ve como ahora los datos son distintos.

4.6.2 Navegar

Como se comentó anteriormente para llegar a la ventana de detalle solo bastaba con pulsar en el correspondiente indicador del menú principal. Pero navegando de esta forma solo se mostraba la agrupación por defecto del indicador.

Por ejemplo el indicador "Datos de Conductores" tiene asociado una agrupación por defecto llamada "Evolución de los Conductores en la Comunidad Autónoma". Si se quiere ver las otras agrupaciones existentes será necesario realizar la operación de "Navegar".

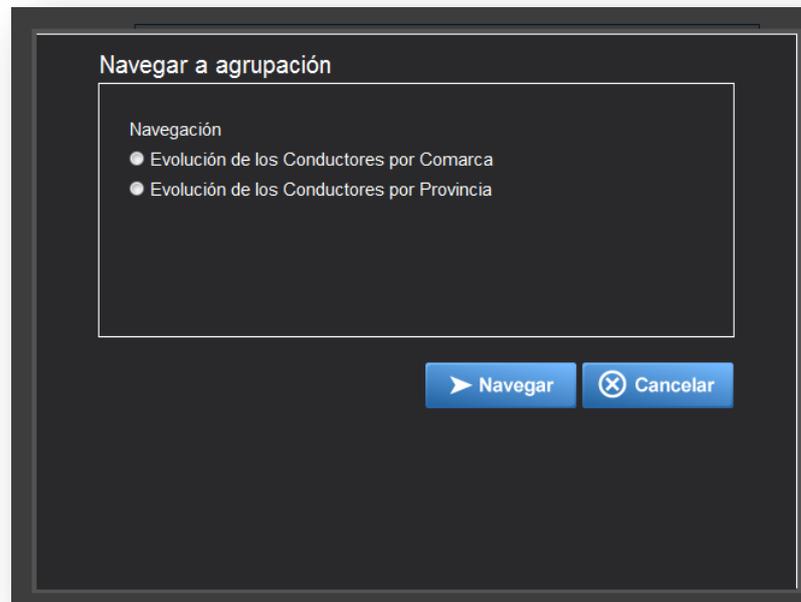


Ilustración 20: Pantalla de navegación entre indicadores

Seleccionando del listado de opciones anteriores "Evolución de los Conductores por Provincia" y pulsando en "Navegar" se mostraría los datos anteriores con un nivel de agrupación diferente.

4.6.3 Exportar

Mediante el uso de esta tercera operación se puede obtener los datos mostrados en la pantalla en un documento. El sistema ofrece la posibilidad de descargar el documento en diferentes formatos para poder visualizarlo en diferentes aplicaciones.



Ilustración 21: Pantalla de exportación de datos

Una vez que aparece la pantalla emergente se elige formato de exportación y se pulsa el botón exportar, apareciendo el informe que se desea en el formato seleccionado:



Ilustración 22: Informe PDF generado por la aplicación

4.6.4 Agregar / Eliminar Favorito

Mediante la ejecución de esta última operación se puede agregar agrupaciones favoritas al usuario conectado. Para ello hay que pulsar botón "Agregar Favorito" en la ventana de los indicadores. Si la gráfica ya está agregada el botón que aparece será "Eliminar Favorito".

Las gráficas favoritas permiten en acceso a las mismas directamente desde la pantalla inicial del sistema y permiten tener de un vistazo, algunos datos para la consulta rápida.

5. INFORMES Y GRÁFICAS

5.1 Indicador de Datos de Conductores

Evolución de los Conductores en la Comunidad Autónoma

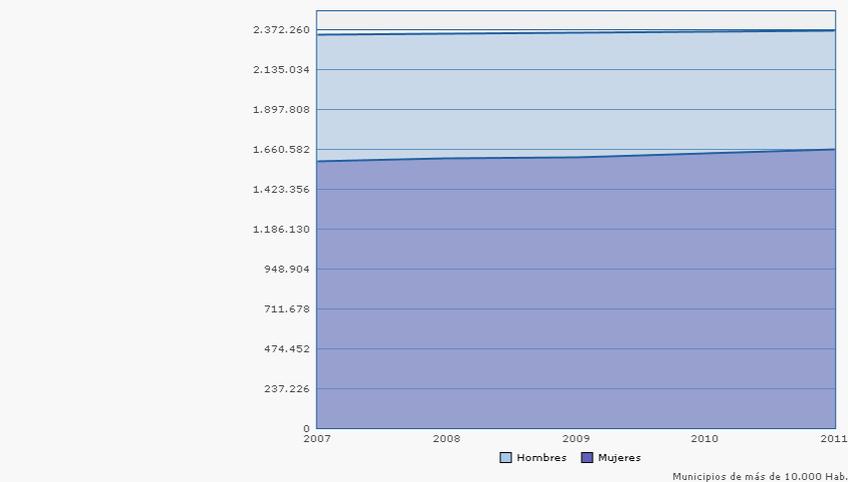


Ilustración 23: Gráfica de evolución de conductores en la comunidad autónoma

Evolución de los Conductores por Provincia

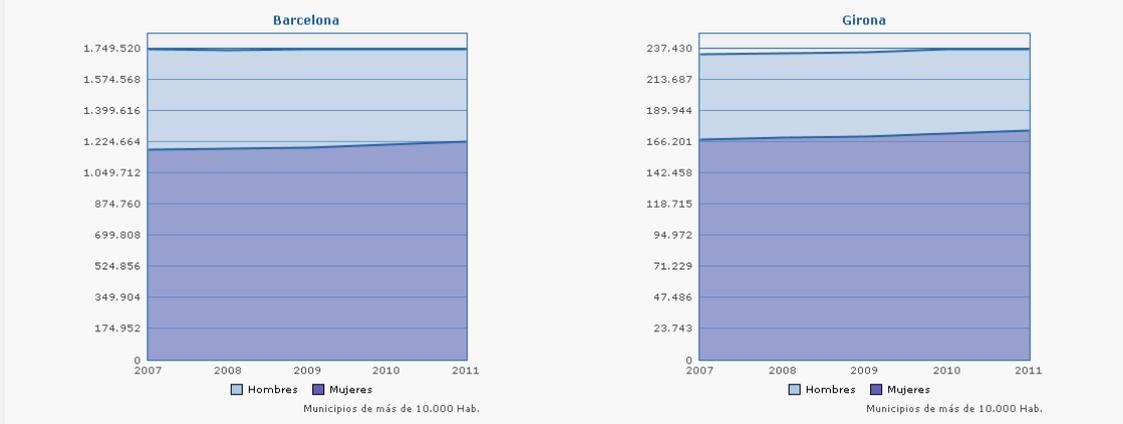


Ilustración 24: Gráfica de evolución de los conductores por provincia

Evolución de los Conductores por Comarca

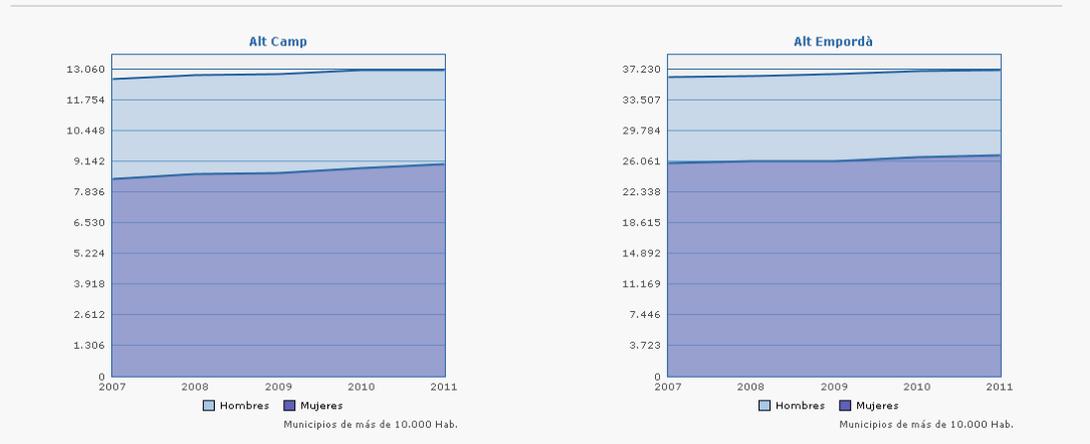


Ilustración 25: Evolución de los conductores por comarca

5.2 Indicador de Datos de Conductores / Población

Evolución de los Conductores respecto a Habitantes en la Comunidad Autónoma

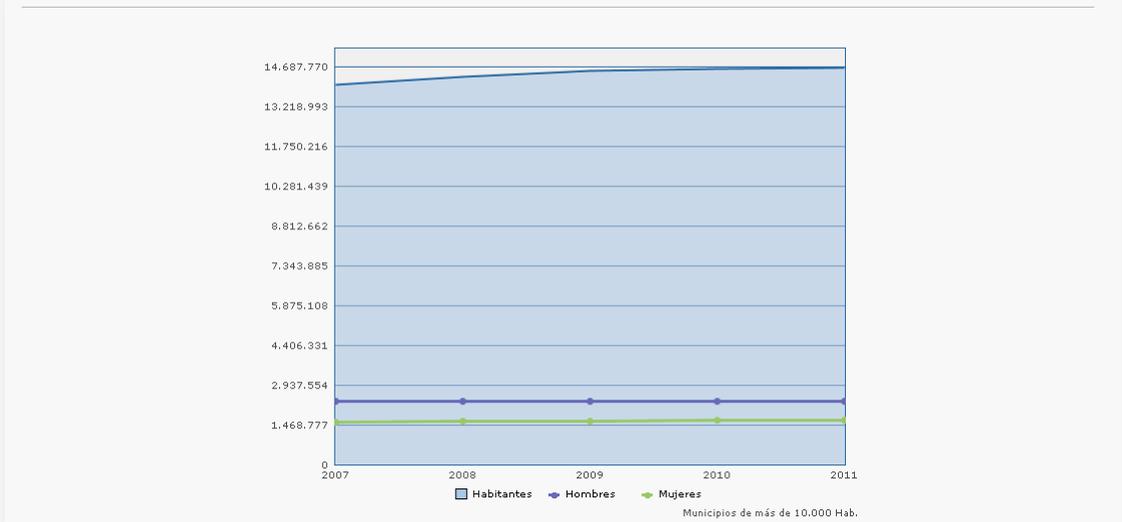


Ilustración 26: Evolución de los conductores respecto a habitantes en la comunidad autónoma

Evolución de los Conductores respecto a Habitantes por Provincia

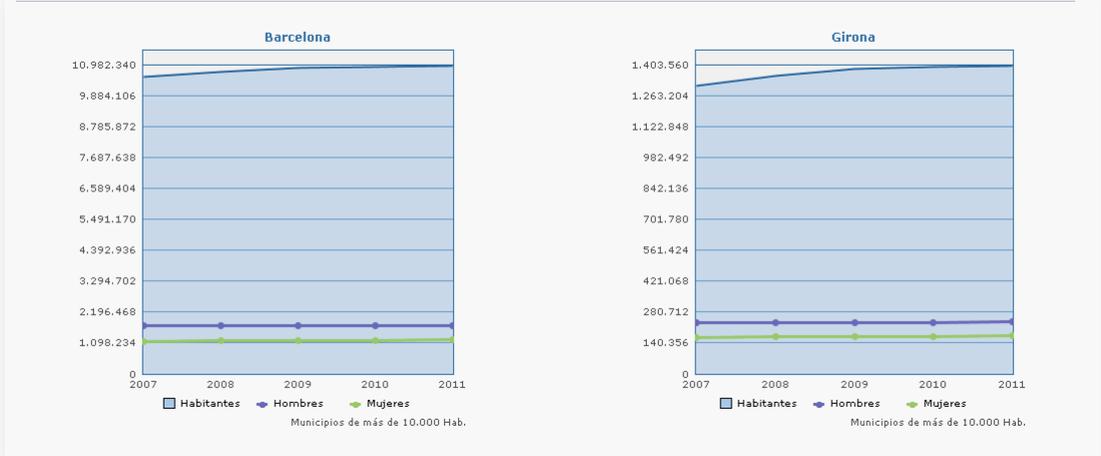


Ilustración 27: Evolución de los conductores respecto a habitantes por provincia

Evolución de los Conductores respecto a Habitantes por Comarca

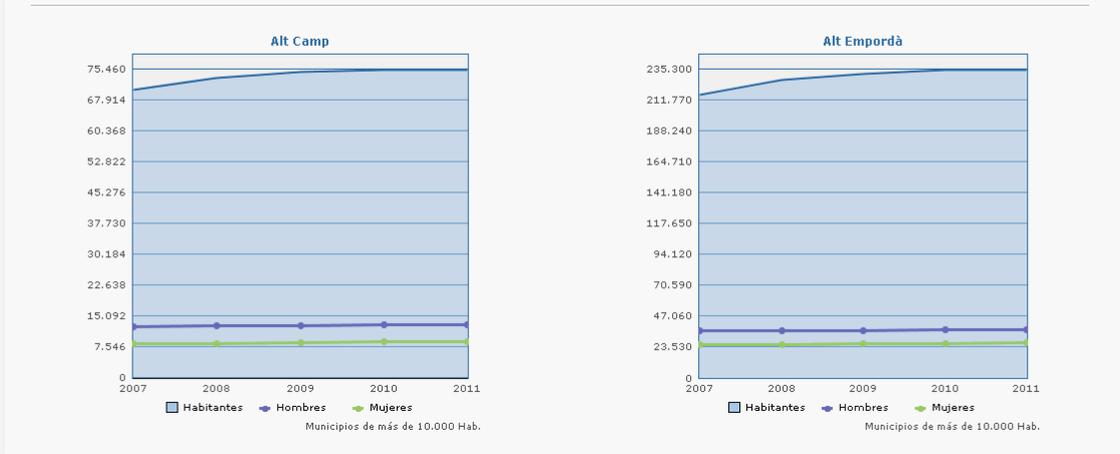


Ilustración 28: Evolución de los conductores respecto a habitantes por comarca

5.3 Indicador de Datos de Conductores / Radares

Evolución de Conductores respecto a Radares en la Comunidad Autónoma

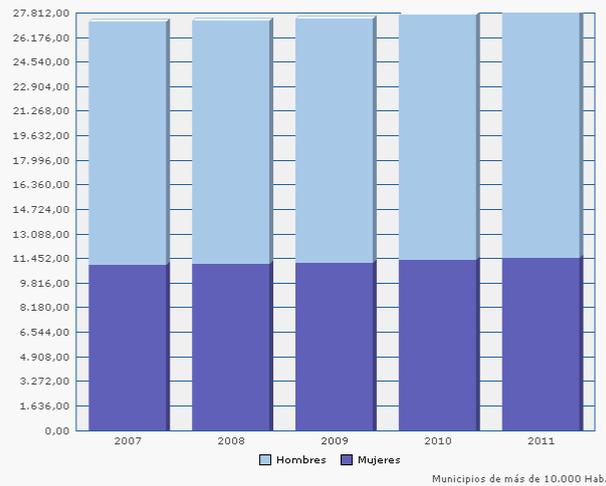


Ilustración 29: Evolución de los conductores respecto a radares en la comunidad autónoma

Evolución de Conductores respecto a Radares por Provincias

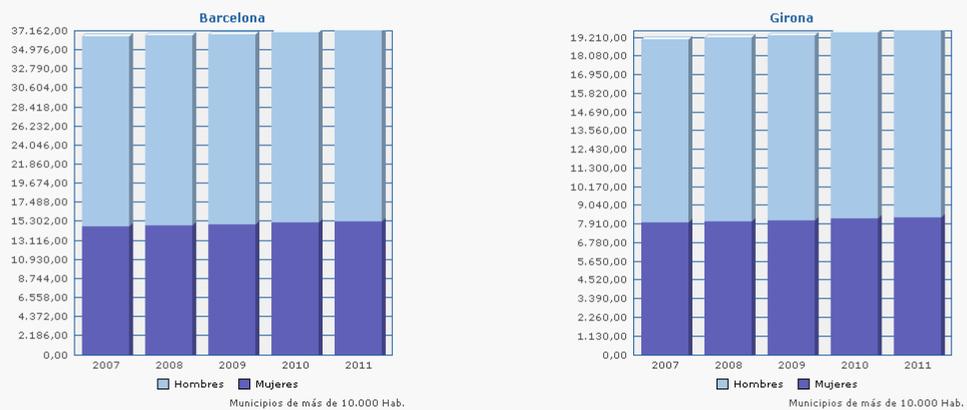


Ilustración 30: Evolución de los conductores respecto a radares por provincias

Evolución de Conductores respecto a Radares por Comarcas

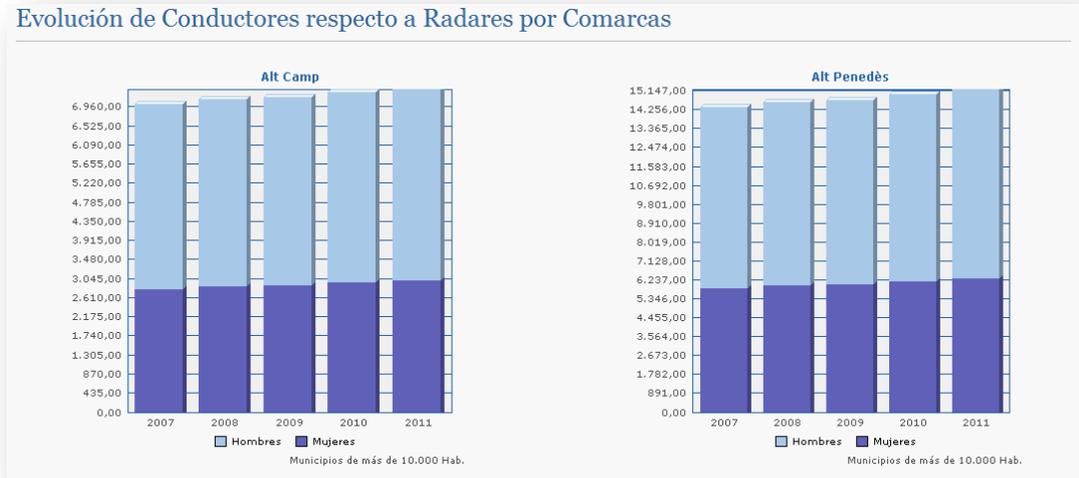


Ilustración 31: Evolución de los conductores respecto a radares por comarcas

5.4 Indicador de Datos de Conductores / Vehículos

Evolución de Conductores respecto a Vehículos en la Comunidad Autónoma

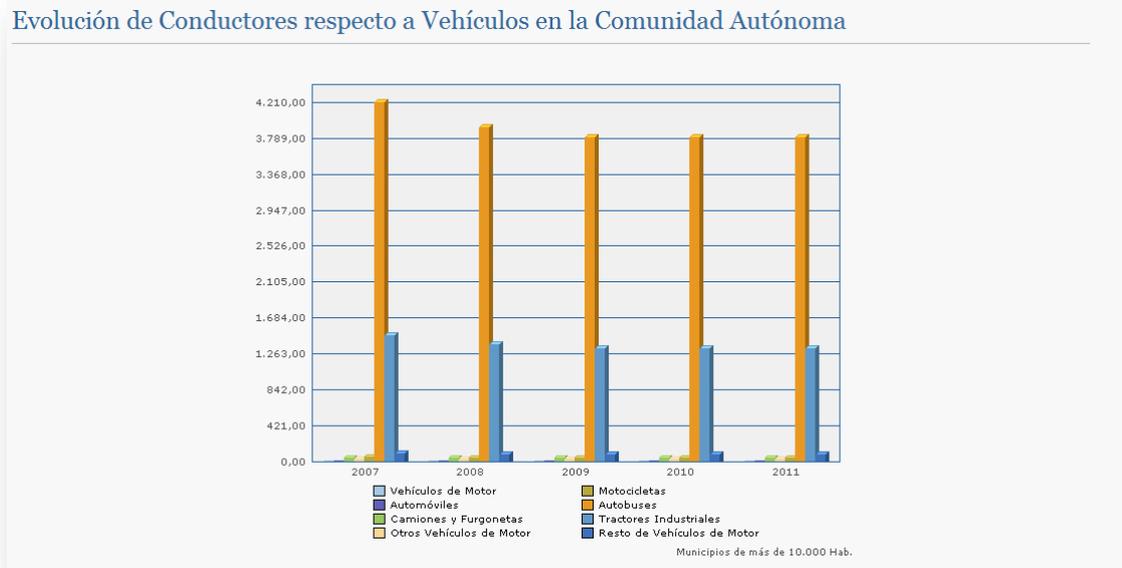


Ilustración 32: Evolución de los conductores respecto a vehículos en la comunidad autónoma

Evolución de Conductores respecto a Vehículos por Provincia

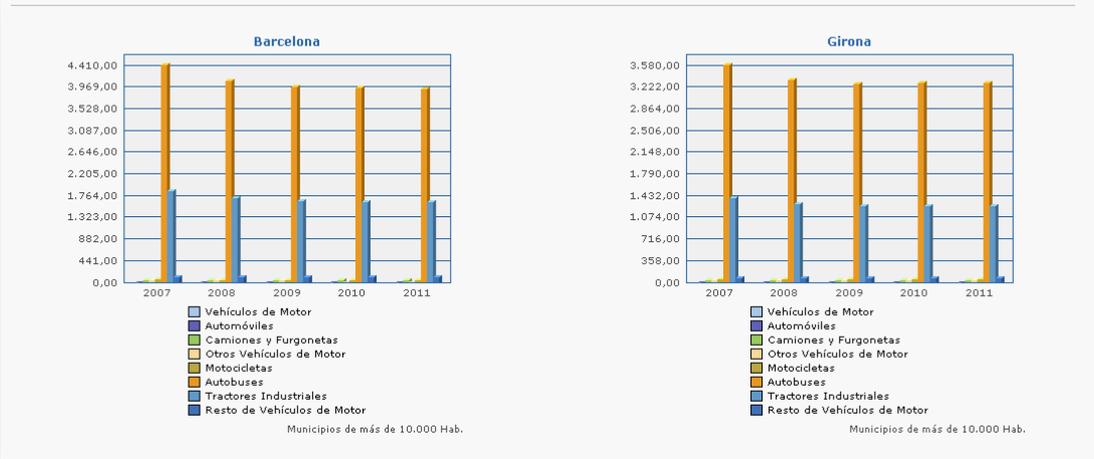


Ilustración 33: Evolución de los conductores respecto a vehículos por provincia

Evolución de Conductores respecto a Vehículos por Comarca

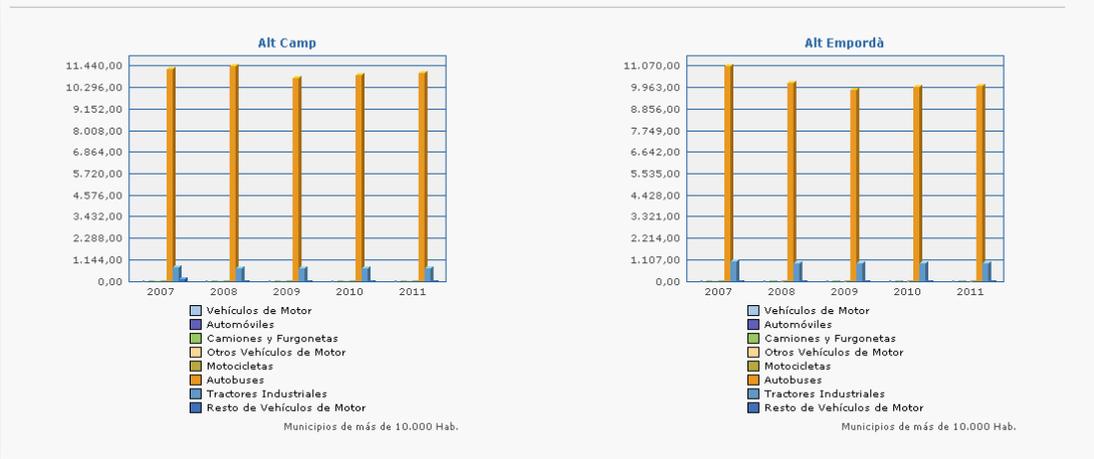


Ilustración 34: Evolución de los conductores respecto a vehículos por comarca

5.5 Indicador de Datos de Población

Evolución de la Población en la Comunidad Autónoma

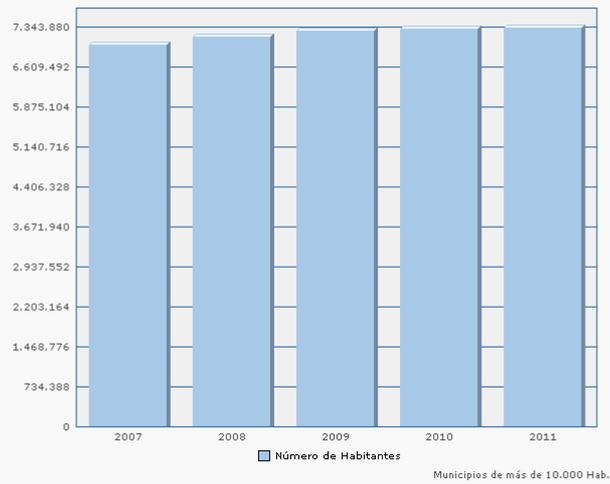


Ilustración 35: Evolución de la población en la comunidad autónoma

Evolución de la Población por Provincias

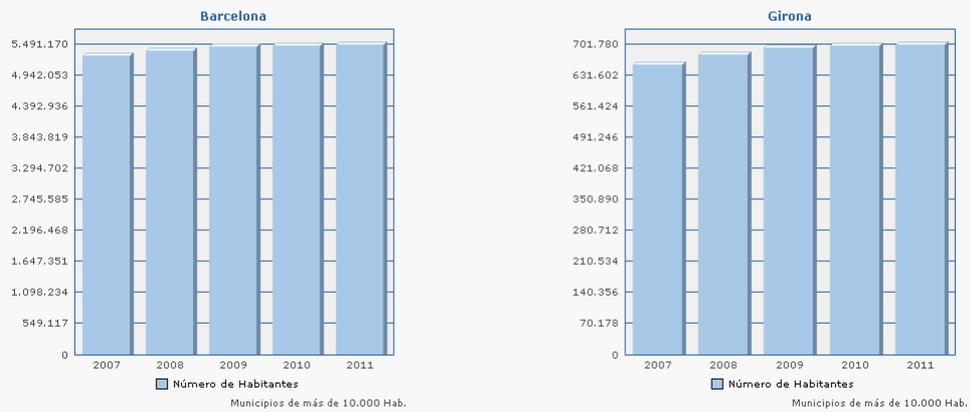


Ilustración 36: Evolución de la población por provincias

Evolución de la Población por Comarcas

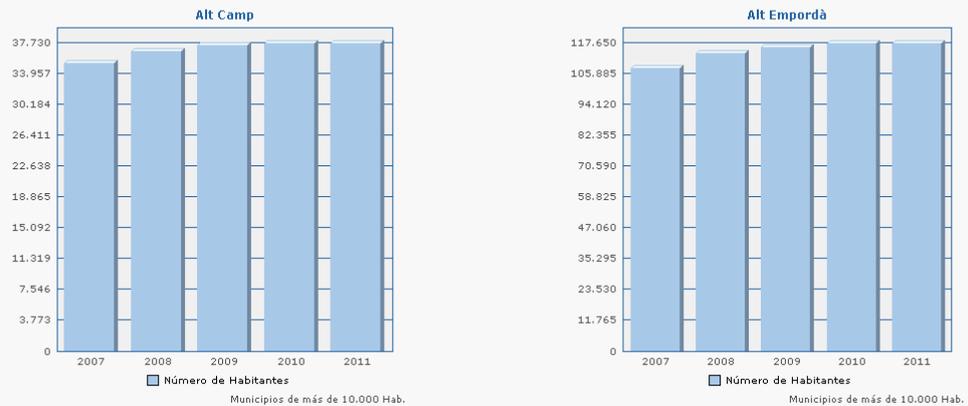


Ilustración 37: Evolución de la población por comarcas

5.6 Indicador de Datos de Radares / Vehículos

Evolución de Radares respecto a Vehículos en la Comunidad Autónoma

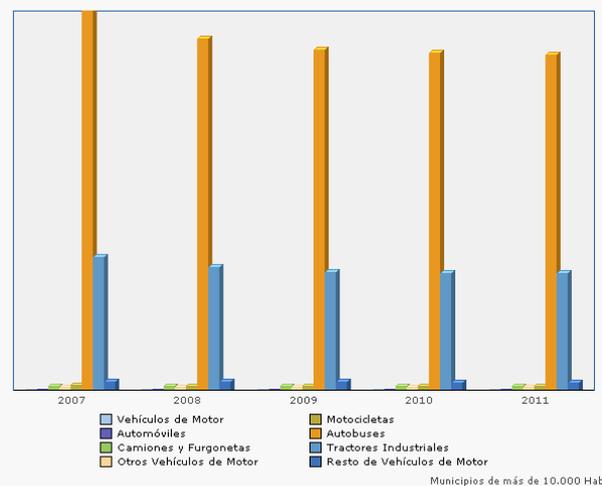


Ilustración 38: Evolución de radares respecto a vehículos en la comunidad autónoma

Evolución de Radares respecto a Vehículos por Provincia

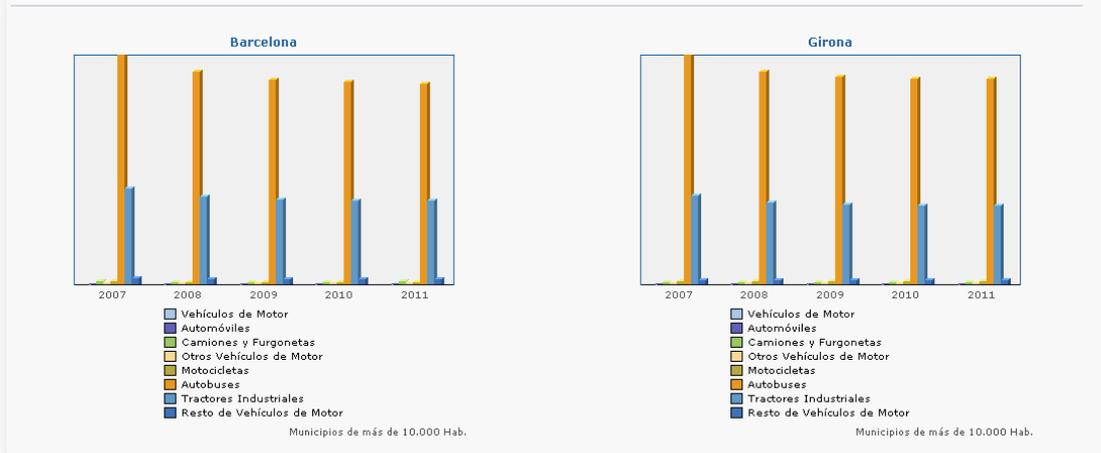


Ilustración 39: Evolución de radares respecto a vehículos por provincia

Evolución de Radares respecto a Vehículos por Comarca

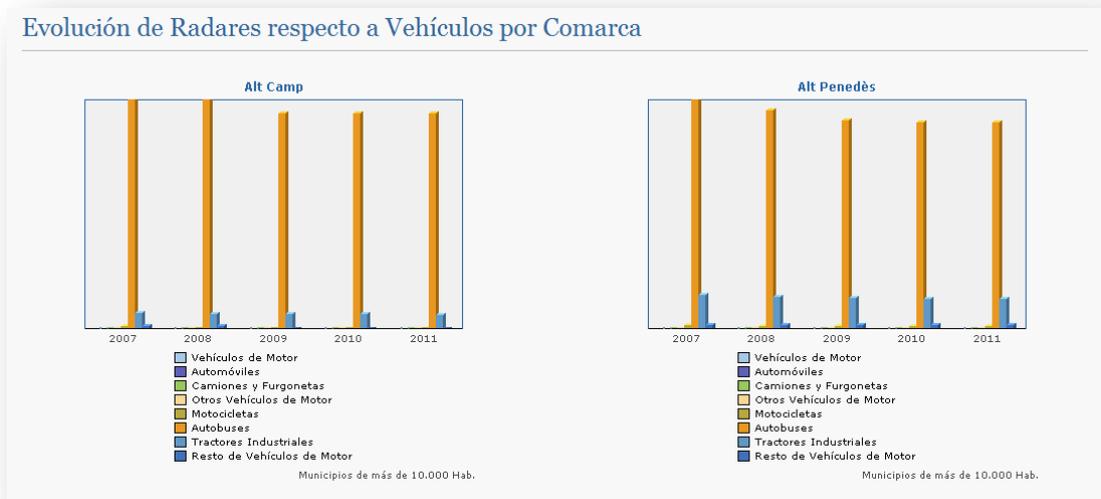


Ilustración 40: Evolución de radares respecto a vehículos por comarca

5.7 Indicador de Datos de Vehículos

Evolución de los Vehículos en la Comunidad Autónoma

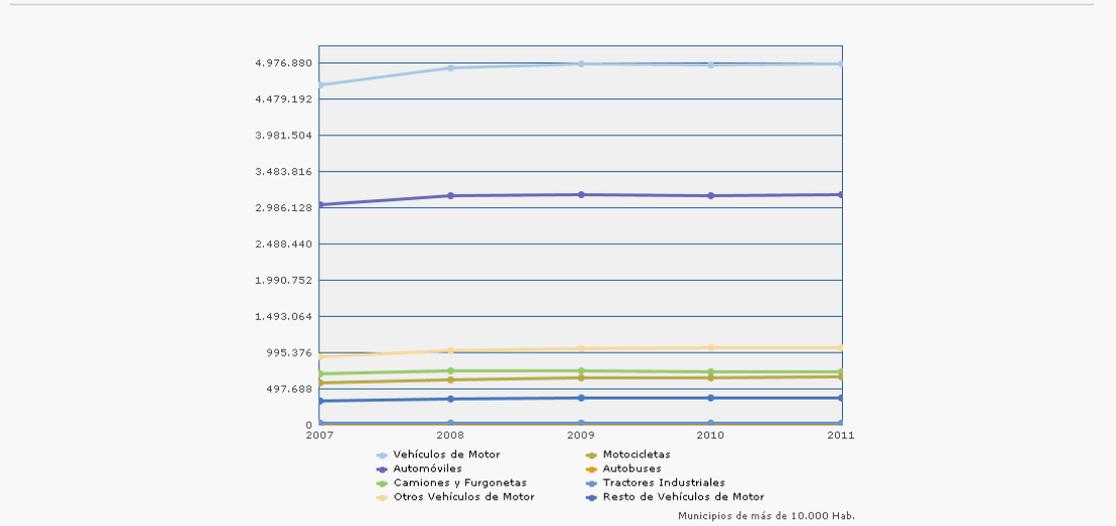


Ilustración 41: Evolución de los vehículos en la comunidad autónoma

Evolución de los Vehículos por Provincias

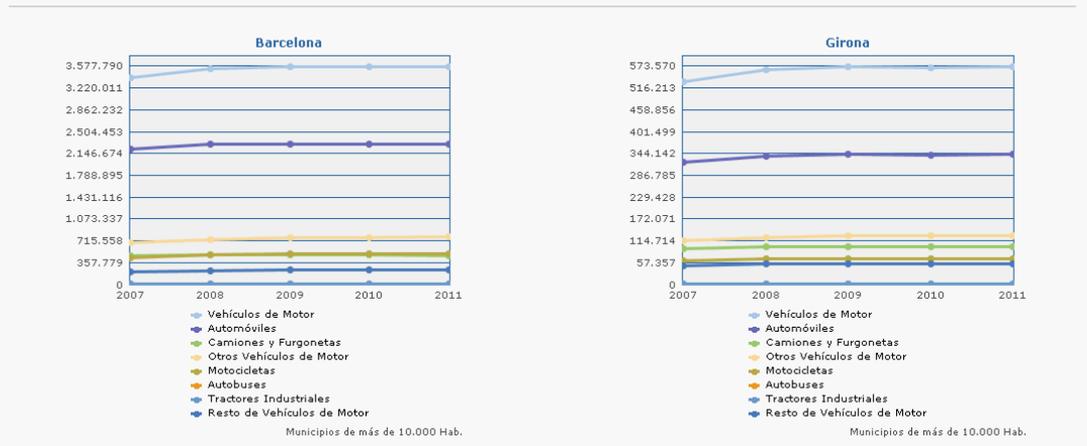


Ilustración 42: Evolución de los vehículos por provincias

Evolución de los Vehículos por Comarcas

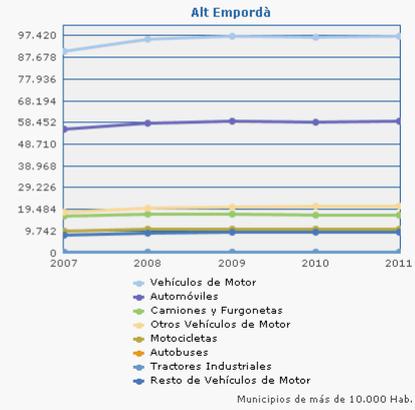


Ilustración 43: Evolución de los vehículos por comarcas

Distribución de Vehículos por Provincias

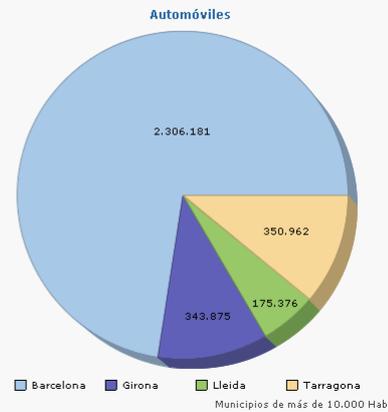
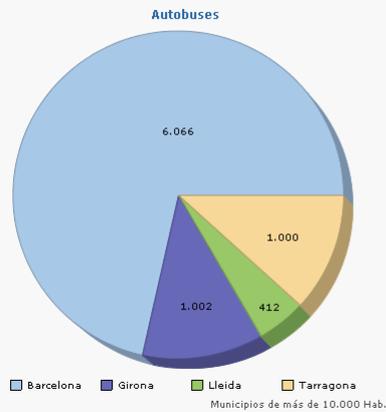


Ilustración 44: Distribución de vehículos por provincias

Distribución de Vehículos por Comarcas

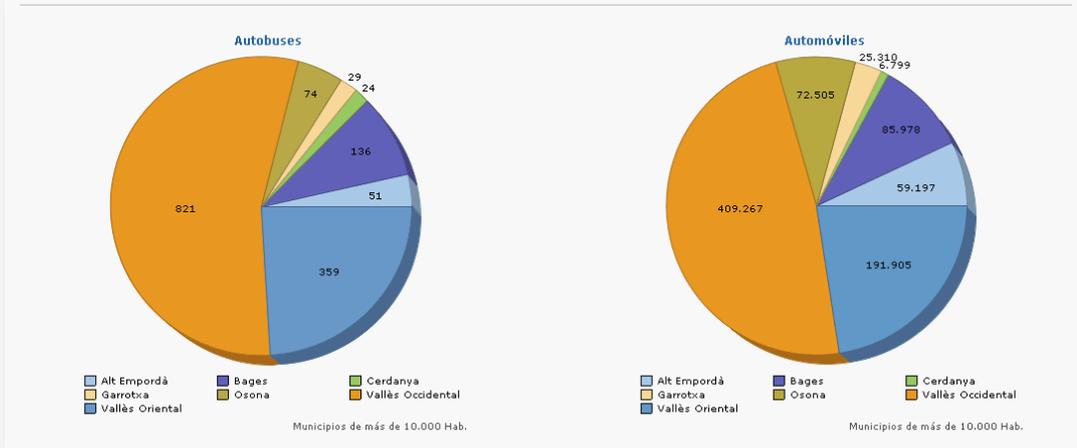


Ilustración 45: Distribución de vehículos por comarcas

5.8 Indicador de Datos de Vehículos / Población

Evolución de Vehículos respecto a la Población en la Comunidad Autónoma

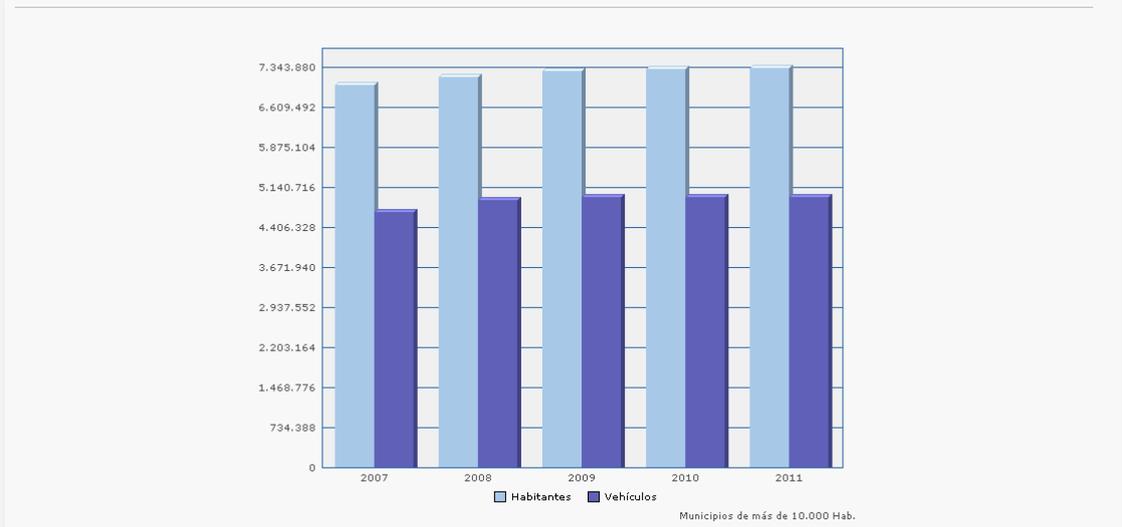


Ilustración 46: Evolución de vehículos respecto a población en la comunidad autónoma

Evolución de Vehículos respecto a la Población por Provincia

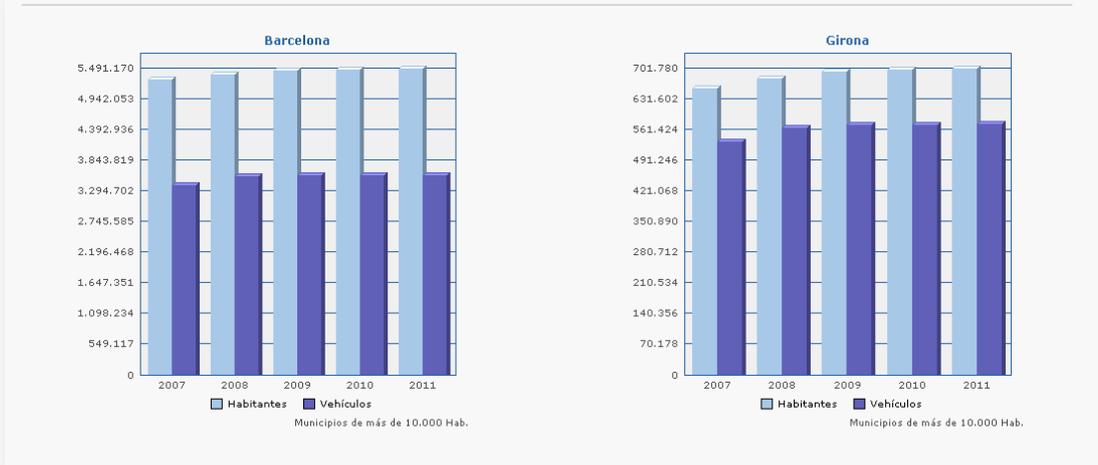


Ilustración 47: Evolución de vehículos respecto a la población por provincia

Evolución de Vehículos respecto a la Población por Comarca

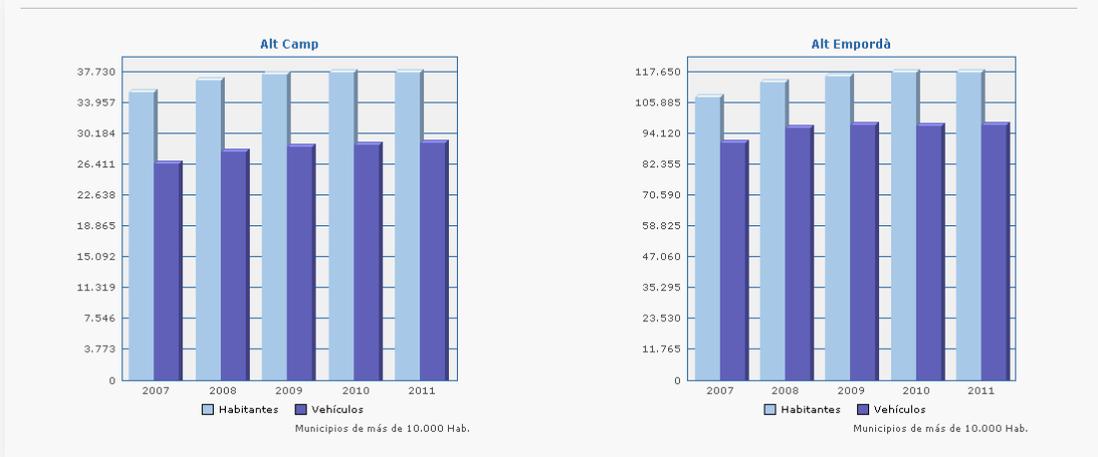


Ilustración 48: Evolución de vehículos respecto a la población por comarcas

5.9 Indicador de Datos de Vehículos / Radares

Evolución de Vehículos respecto a Radares en la Comunidad Autónoma

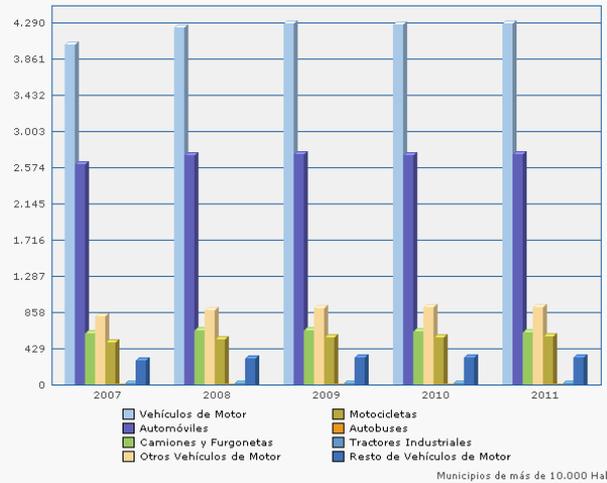


Ilustración 49: Evolución de vehículos respecto a radares en la comunidad autónoma

Evolución de Vehículos respecto a Radares por Provincia

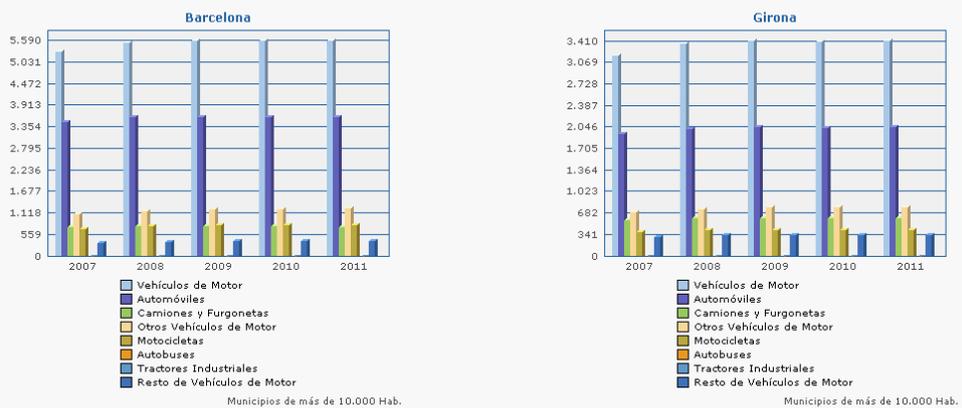


Ilustración 50: Evolución de vehículos respecto a radares por provincia

Evolución de Vehículos respecto a Radares por Comarca

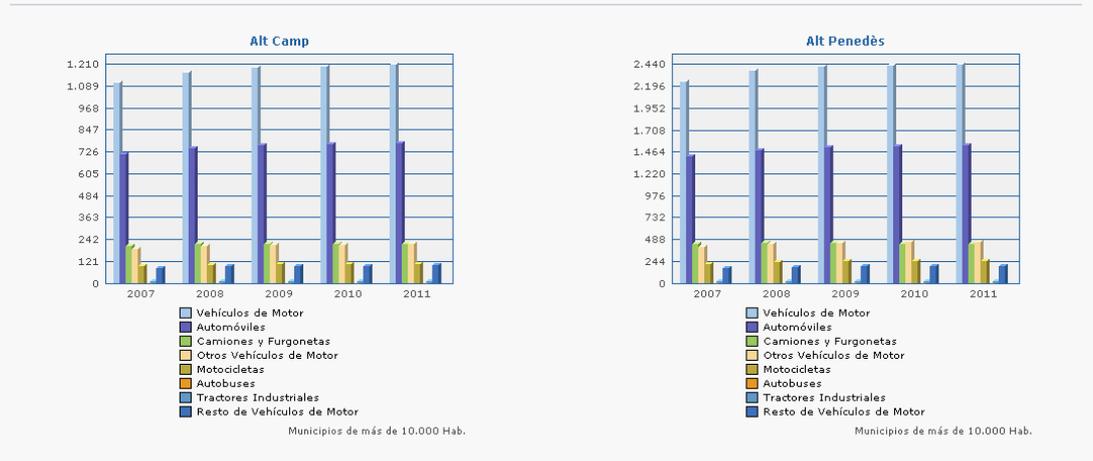


Ilustración 51: Evolución de vehículos respecto a radares por comarca

5.10 Indicador de Densidad de Población

Evolución de la Densidad de Población en la Comunidad

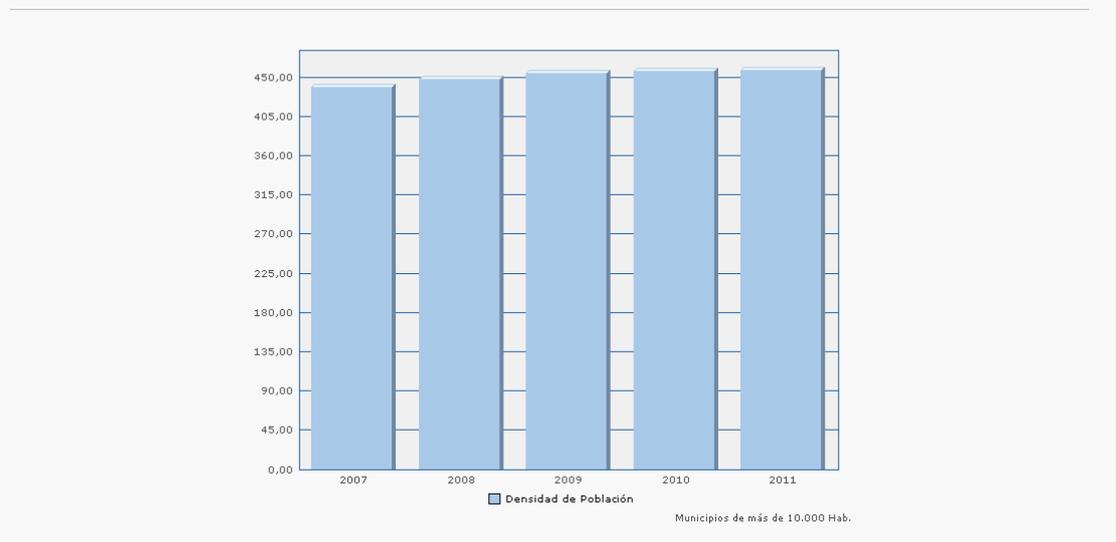


Ilustración 52: Evolución de la densidad de población en la comunidad autónoma

Evolución de la Densidad de Población por Provincias

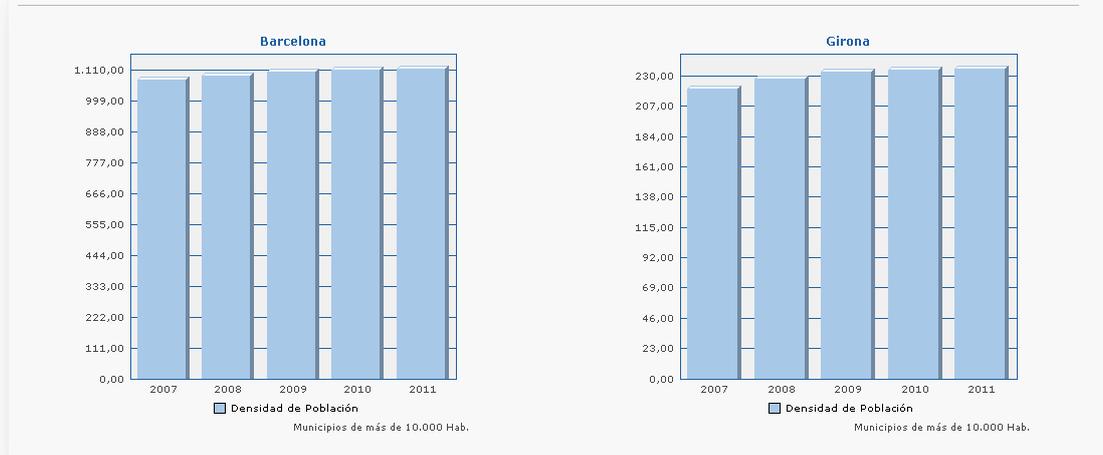


Ilustración 53: Evolución de la densidad de población por provincias

Evolución de la Densidad de Población por Comarcas

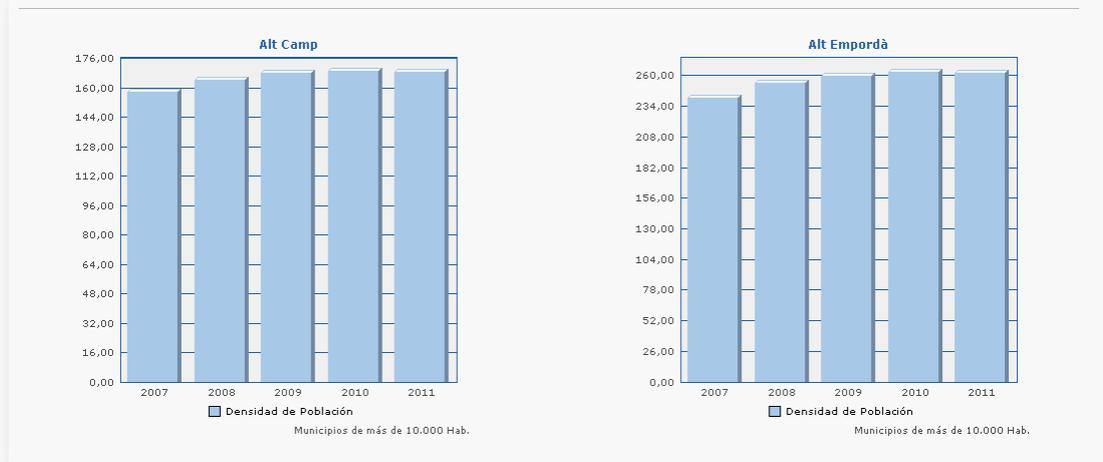


Ilustración 54: Evolución de la densidad de población por comarcas

5.11 Indicador de Densidad de Tránsito

Evolución de la Densidad de Tránsito en la Comunidad Autónoma

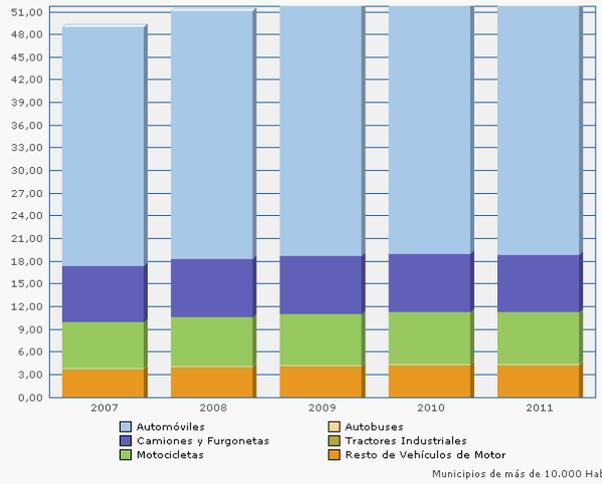


Ilustración 55: Evolución de la densidad de tránsito en la comunidad autónoma

Evolución de la Densidad de Tránsito por Provincias

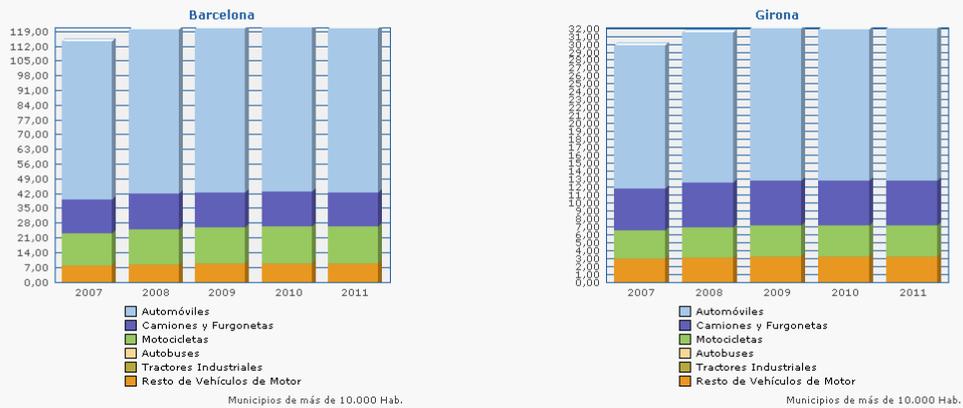


Ilustración 56: Evolución de la densidad de tránsito por provincias

Evolución de la Densidad de Tránsito por Comarcas

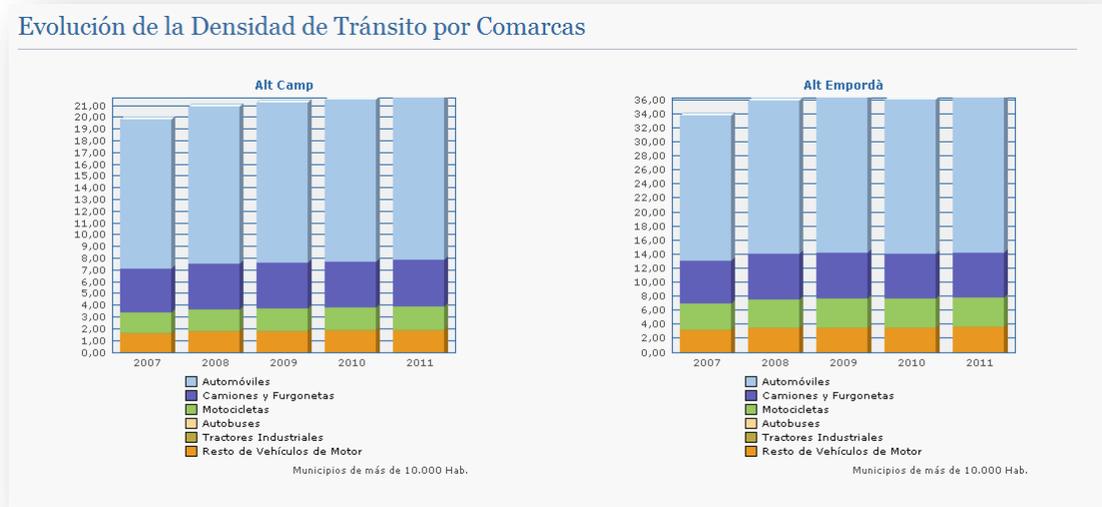


Ilustración 57: Evolución de la densidad de tránsito por comarcas

6. CONCLUSIONES

Las principales conclusiones que se han obtenido una vez finalizado el proyecto fin de carrera son las siguientes:

- Se ha conseguido cumplir los objetivos planteados por la asignatura así como los requisitos solicitados por FECRES
- Se ha seguido la metodología de trabajo orientada desde la asignatura
- Se ha comprobado que es necesario la realización de un análisis de requisitos detallado así como un exhaustivo análisis de los datos aportados desde el principio, para evitar encontrarnos con problemas de última hora
- Se ha cumplido la planificación propuesta sin que haya habido desviaciones considerables de la planteada inicialmente.
- He ampliado mis conocimientos de PDI, ya que las veces que lo había usado laboralmente, solo había realizado transformaciones muy simples
- Se han diseñado un conjunto de graficas sobre los indicadores que se solicitaban en el caso práctico. Se han añadido además algunas graficas que podrían ser interesantes para el usuario.
- Finalmente, considero que el TFC propuesto en la asignatura me ha ayudado a adquirir más conocimientos y experiencia en proyectos relacionados con *Business Intelligence*, lo que me ayudara en mi carrera profesional.

7. LINEAS DE EVOLUCIÓN FUTURAS

A continuación se exponen varias líneas de desarrollo que se podrían seguir en el futuro para la mejora de la aplicación y de los procesos de carga:

- Se podría realizar optimizar el sistema para que no fuera necesario realizar en dos pasos el proceso ETL sino que la propia ejecución de la transformación en PDI realizase el resto de pasos.
- Se podría controlar que si no existe el modelo de datos, el mismo proceso controlase y crease las tablas necesarias.
- Se podría crear un sistema de auditoria tanto para el acceso a la aplicación de explotación estadística como para el sistema de carga de datos, donde se pudiese monitorizar el acceso de los usuarios y las cargas que se realizan.
- Se podría mejorar la aplicación de explotación estadística añadiendo un sistema de cubos OLAP Mondrian (basado en *Pentaho*)
- Se podría mejorar la aplicación de explotación estadística añadiendo una zona de administración donde poder crear nuevos indicadores, navegaciones y filtros
- Se podría mejorar la aplicación de explotación estadística para añadir filtros *LOV* dinámicos basados en subconsultas

8. BIBLIOGRAFIA

Bases de datos I - XP05/75002/00493 - Jaume Sistac Planas. UOC

Bases de datos II - XP03/75053/02082 - Jaume Sistac Planas. UOC

Técnicas de desarrollo de software - XP06/75049/00155 - Fatos Xhafa. UOC

Ingeniería del software - PID_00161209. UOC

Trabajo final de carrera - P08/89018/00446 - Roser Beneito Montagut. UOC

Trabajo final de carrera - P08/89018/00445. UOC

Plan docente de la asignatura:

http://cv.uoc.edu/tren/trenacc/web/GAT_EXP.PLANDOCENTE?any_academico=20131&cod_a_signatura=75.066&idioma=CAS&pagina=PD_PREV_SECRE&cache=S

Wiki Pentaho: <http://community.pentaho.com/>

Wiki de desarrollo de software de Ayesa Advanced Technologies