



Proyecto Fin de Carrera

**DESARROLLO DE UNA APLICACIÓN WEB SOBRE
DEMOGRAFÍA Y POBLACIÓN DE SANTANDER**
(Development of a web application on Demographic data of
Santander)

Para acceder al Título de
INGENIERO EN INFORMÁTICA

Autor: Rafael Pereda Fernández

Director: Carlos Blanco Bueno

Junio - 2017

Agradecimientos

A mi familia y compañeros de estudios
por su apoyo durante toda la carrera.

A Carlos Blanco por su dirección.

RESUMEN

El objetivo de este Proyecto Fin de Carrera es el diseño y desarrollo de una aplicación web que permita la consulta de estadísticas demográficas de los datos de población histórica del Ayuntamiento de Santander a partir de los ficheros del Open Data Santander. Implementando una solución de Business Intelligence (BI), se ofrecerán cuadros de mando interactivos que permitan al usuario realizar consultas rápidas, atractivas visualmente en forma de gráficos & tablas, obteniendo análisis de la evolución de la población en términos demográficos y también sobre las corrientes migratorias que se hayan producido durante estos años objeto de la investigación.

El Ayuntamiento de Santander proporciona a través de la web <http://datos.santander.es/> una serie de conjuntos de datos en abierto. Este proyecto se centrará en el sector demografía, que recoge unos ficheros históricos de Población y Bajas de Población del Ayuntamiento de Santander. Se trata de dar más valor añadido a esa información histórica con la realización de una aplicación que implementa una solución de Business Intelligence.

Se utilizará Pentaho como plataforma Open Source para implementar la solución BI, y se desarrollarán todas las fases típicas de este tipo de proyectos: Análisis de los datos de partida, Modelo de datos, Procesos ETL (Extracción, Transformación y Carga de datos), creación de Cubos OLAP y de los Cuadros de Mando que permitirán un análisis multidimensional de la información.

Palabras Clave: Business Intelligence, análisis multidimensional, proceso ETL, cubo OLAP, Pentaho, Cuadro de Mando, demografía, población histórica, estadísticas.

ABSTRACT

This Master Thesis aims to design and develop an application web that allows the user to make statistical queries based on the historical demographic data of the city of Santander, thanks to the data available from the Open Data Santander website. By implementing a Business Intelligence (BI) solution, the user will be able to use interactive dashboards, making quick, attractive queries in the form of graphics & tables, obtaining through the process an analysis of the evolution of the population of Santander, as well as the different migration flows that may have taken place during this past few years.

Santander's City Hall offers a series of data sets through the website <http://datos.santander.es/>. This project will focus on the demographic area, whose files contain historical records of the people registered in the city, as well as the changes on the census of Santander's City Hall. The objective here is to increase the value of these historical records by developing an application that implements a Business Intelligence solution.

Pentaho will serve as an Open Source tool to implement said BI solution, setting up all the associated tasks typical of this kind of projects: an analysis of the data, a data model, ETL process (Extracting, Transforming and Data Loading), the creation of OLAP cubes, and finally the appropriate dashboards which will allow a multidimensional analysis of the information.

Keywords: Business Intelligence, multidimensional analysis, ETL process, OLAP cubes, Dashboards, Demography, Population statistics.

ÍNDICE GENERAL

1. Introducción.....	- 1 -
1.1. Motivación.....	- 1 -
1.2. Objetivos.....	- 2 -
2. Herramientas y método.....	- 2 -
2.1. Selección de software BI.....	- 2 -
2.2. Herramientas.....	- 6 -
2.2.1. Pentaho BI.....	- 6 -
2.2.2. Plugins para Pentaho BI.....	- 6 -
2.2.3. Otras herramientas.....	- 6 -
2.3. Metodología.....	- 7 -
3. Análisis y captura de requisitos.....	- 7 -
4. Diseño e implementación de la solución	- 8 -
4.1. Estudio de fuentes de datos y modelo de datos.....	- 8 -
4.1.1. Esquemas en estrella.....	- 9 -
4.2. Procesos ETL.....	- 11 -
4.2.1. Procesos ETL para las Bajas de Población.....	- 11 -
4.2.2. Procesos ETL Población.....	- 16 -
4.3. Creación Cubos OLAP.....	- 21 -
4.4. Diseño y Desarrollo de Cuadros de Mando.....	- 26 -
4.4.1. Diseño de Mockups.....	- 26 -
4.4.2. Desarrollo de Cuadros de Mando.....	- 32 -
4.4.2.1. Creación Cuadro de Mando Bajas de Población.....	- 32 -
4.4.2.2. Creación Cuadro de Mando de Población Histórica.....	- 36 -
4.4.2.3. Creación Cuadro de Mando de Flujo Migratorio.....	- 41 -
4.5. Pruebas de los Cuadros de Mando.....	- 43 -
4.5.1. Prueba del Cuadro de Mando de Bajas de Población.....	- 43 -
4.5.2. Prueba del Cuadro de Mando de Población Histórica.....	- 52 -
4.5.3. Prueba del Cuadro de Mando de Flujo Migratorio.....	- 57 -
4.6. Integración de Cuadros de Mando.....	- 58 -
5. Conclusiones y trabajos futuros.....	- 61 -
6. Bibliografía.....	- 62 -

(LISTADO DE FIGURAS)

Figura 1: Gartner Magic Quadrant BI año 2017.....	3
Figura 2: Características plataforma Pentaho	4
Figura 3: Arquitectura plataforma Pentaho	5
Figura 4: Esquema en estrella Bajas de Población.....	10
Figura 5: Esquema en estrella Población Histórica.....	10
Figura 6: ETL_ Bajas_ pasos inicio fichero entrada json Bajas 2010	11
Figura 7: ETL_ Bajas_ visualización fichero entrada json Bajas 2010	12
Figura 8: ETL_ Bajas_ paso Renombra campos bajas 2010	13
Figura 9: ETL_ Bajas_ paso Añadir año2010 identifier.....	13
Figura 10: ETL_ Bajas_ paso Unión Ordenada.....	13
Figura 11: ETL_ Bajas_ paso crear dimensión Pais_Prov_Destino	14
Figura 12: ETL_ Bajas_ Transformación para crear todas dimensiones y hechos Bajas	14
Figura 13: ETL_ Bajas_ paso Selecciona campos hechos Bajas	15
Figura 14: ETL_ Bajas_ paso Filas Unicas claves dim	15
Figura 15: ETL_ Bajas_ paso Salida a Tabla baj_hechos.....	16
Figura 16: ETL_ Poblac_ visualización fichero entrada json Pob 2014	17
Figura 17: ETL_ Poblac_ paso Unir fich. entrada y crear dimensión Pais_Prov_Proced.....	18
Figura 18: ETL_ Poblac_ paso fichero entrada Pob2010	18
Figura 19: ETL_ Poblac_ paso Add anypadron 2010	18
Figura 20: ETL_ Poblac_ Transformación para crear todas dimensiones y hechos Población.....	20
Figura 21: ETL_ Job_ crear tablas y dimensiones Bajas y Población	20
Figura 22: OlapCube_ Bajas_ Asociar Tabla hechos	22
Figura 23: OlapCube_ Bajas_ Crear dimensión Motivos Baja	22
Figura 24: OlapCube_ Bajas_ Definir Jerarquía en Motivos Baja	22
Figura 25: OlapCube_ Bajas_ Añadir nivel en Motivos Baja	23
Figura 26: OlapCube_ Bajas_ Definir la medida	23
Figura 27: Olap_Cube Bajas de Población	24
Figura 28: OlapCube_ Población.....	25
Figura 29: Mockup_ Bajas Población _Pantalla inicio	27
Figura 30: Mockup_ Bajas Población _pestaña Bajas por País Nac. y Nacionalidad	28
Figura 31: Mockup_ Bajas Población _pestaña Bajas por Año y Mes	28
Figura 32: Mockup_ Población Histórica _Pantalla inicio	29
Figura 33: Mockup_ Población Histórica _pestaña Por País Nac. y Nacionalidad.....	30
Figura 34: Mockup_ Población Histórica _pestaña Altas por Año y Mes	30
Figura 35: Mockup_ Población Histórica _pestaña Comparativa por Años	30
Figura 36: Mockup_ Flujo Migratorio	31
Figura 37: Cuadro_Mando_Bajas de Población.....	33
Figura 38: CM_ Bajas_ Seleccionar Motivo Baja	34
Figura 39: CM_ Bajas_ Layout_ Seleccionar Motivo Baja	34
Figura 40: CM_ Bajas_ Componente Bootstrap_Panel_ Seleccionar Motivo Baja.....	35
Figura 41: CM_ Bajas_ Componente Pie_Chart_ Seleccionar Motivo Baja	35
Figura 42: código_ Script_clikAction_CM_ Bajas_ Pie_Chart.....	35
Figura 43: código_ Script_modif_Tooltip_CM_ Bajas_ Pie_Chart	35
Figura 44: CM_ Bajas_ Datasource _ Seleccionar Motivo Baja	36
Figura 45: código_ QueryMdx_CM_ Bajas_ Seleccionar Motivo Baja	36
Figura 46: CM_ Bajas_ Popup Seleccionar Motivo Baja	36
Figura 47: Cuadro_Mando Población Histórica Año Padrón 2010	37
Figura 48: código_ QueryMdx_CM_ Población_ Pestaña por Tramo Edad y Sexo.	38
Figura 49: código_ QueryMdx_CM_ Población_ Pestaña por Provincia de Nacimiento.	38
Figura 50: CM_ Población año2011_Vista Tabla Nacionalidades	39
Figura 51: CM_ Población_ Pestaña Altas Mensuales	40
Figura 52: código_CM_ Población_queryMdx_Altas Mensuales.	40
Figura 53: Cuadro Mando Flujo Migratorio, Año: 2014, Nacionalidad: Colombia	41
Figura 54: Código_queryMdx_AñosAlta_CM_Flujo Migratorio.	42
Figura 55: CM_ Bajas_Error_Motivo Baja_Cambio Residencia España	43
Figura 56: CM_ Bajas_Error_Motivo Baja_Defunción	44
Figura 57: CM_ Bajas_Top 15 Paises_ Cambio Residencia al Extranjero	44
Figura 58: CM_ Bajas_Top 15 Paises_ Cambio Residencia Extranjero_Vista Tablas	45
Figura 59: CM_ Bajas_País_Destino: Reino Unido_Pestaña_Por Edad, Sexo y Estudios.	46
Figura 60: CM_ Bajas_País_Destino: Reino Unido_Pestaña_Por País Nac. y Nacionalidad	47
Figura 61: CM_ Bajas_País_Destino: Estados Unidos_Pestaña_Por Edad, Sexo y Estudios.	48

Figura 62: CM_Bajas_País_Destino: Estados Unidos_Pestaña_Por País Nac. y Nacionalidad	48
Figura 63: CM_Bajas_País_Destino: Perú_Pestaña_Por Edad, Sexo y Estudios	49
Figura 64: CM_Bajas_País_Destino: Perú_Pestaña_Por Nacionalidad y País Nac.....	49
Figura 65: CM_Bajas_Caducidad_ Años:2011-12, Distrito 5, Pestaña_Por País Nac y Nacionald	50
Figura 66: CM_Bajas_Caducidad_ Años:2011-12, Distrito 5, Pestaña_Bajas por Año y Mes	51
Figura 67: CM_Poblacion_ Nacionalidad Rumanía, Pestaña_Por Edad, Sexo y Estudios	52
Figura 68: CM_Población_ Nacionalidad Rumanía, Pestaña_Por País Nac y Procedencia	52
Figura 69: CM_Población_ Nacionalidad: Italia, Pestaña por Edad, Sexo y Estudios	53
Figura 70: CM_Población_ Nacionalidad: Italia, Pestaña por País Nac y Procedencia.....	53
Figura 71: CM_Población_Comparativa anual por Tramo de Edad	54
Figura 72: CM_Población_Comparativa anual por Nivel de Estudios	55
Figura 73: Cuadro Mando Flujo Migratorio, Año: 2013, Nacionalidad: España	57
Figura 74: CM_Bajas_button Bootstrap.....	58
Figura 75: Los 3 Cuadros de Mando entrelazados.....	59
Figura 76: Consola_usuario_pentaho_login.....	60

1. Introducción.

Durante las últimas dos décadas, con la expansión del uso del ordenador y del internet, la cantidad de datos disponibles se ha ido incrementando de forma exponencial. Hoy en día puedes acceder a un sinfín de datos y documentos de prácticamente cualquier tipo de información. Analizar estos conjuntos de datos no es siempre una tarea fácil o rápida, y puede ser muy tedioso volver a repetir las mismas tareas para otros conjuntos. Tal vez por eso las soluciones Business Intelligence han estado en aumento estos últimos años, facilitando la tarea del usuario.

Según el IT Glossary de Gartner, se define **Business Intelligence** como: *“La inteligencia empresarial (BI) es un término general que incluye las aplicaciones, la infraestructura, las herramientas y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento”*. (Gartner IT Glossary, 2017).

El BI se ha convertido en la actualidad en una de las técnicas de análisis de datos con la que las empresas pueden generar conocimiento y tomar decisiones estratégicas.

Por otro lado, los **Open data** o Datos Abiertos del sector público tienen como principal objetivo aprovechar al máximo los recursos públicos disponibles, permitiendo el acceso y reutilización de la información generada por las Administraciones públicas para beneficio de empresas, investigadores o ciudadanos en general. Estos proyectos van a impulsar una mejora de los servicios a la ciudadanía, generación de actividades de negocio y mayor transparencia en la gestión de la Administración.

Los datos pueden estar clasificados por distintos sectores o temas (economía, urbanismo, transporte, turismo, demografía...). Concretamente el conjunto de datos del sector demografía contiene información sobre la distribución de la población por distritos y secciones electorales, nivel de estudios, edad, nacionalidad así como las altas y bajas de población con su información asociada (motivos de alta/baja, país/provincia de procedencia/destino). Toda esta información es utilizada por las Administraciones públicas a la hora de gestionar recursos, infraestructuras en los ámbitos de la educación, sanidad (distribución de centros educativos o asistenciales), servicios de transporte público así como también para la adopción de políticas de acción social o cultural.

El Ayuntamiento de Santander ha puesto a disposición de la sociedad un Catálogo de Datos Abiertos sobre diversos temas, uno de los cuales es el sector **Demografía** que contiene un conjunto de datos históricos sobre Población y Bajas de Población. Por lo tanto, sería interesante llevar a cabo una solución BI para el sector de Demografía.

1.1. Motivación.

Los ficheros históricos de Población y Bajas de Población, son archivos sobre los que tradicionalmente siempre se han elaborado muchas estadísticas de población y demografía. Aparte de las estadísticas oficiales del INE, estas estadísticas muchas veces son objeto de noticia en medios de comunicación, se publican noticias referentes al “envejecimiento de la población”, “datos de inmigración/emigración durante la crisis...”, etc...

Se trata de aplicar a este conjunto de archivos históricos la potencia y el rendimiento que permite la tecnología de Business Intelligence, poder realizar análisis multidimensional y sobre todo, la creación de unos Cuadros de Mando dinámicos que permitan explorar y profundizar por diversos conceptos de los datos, consultas ágiles e intuitivas y además, que se puedan

deducir tendencias de evolución demográfica de la población, flujos migratorios, en fin, que se puedan corroborar todas esas estadísticas.

1.2. Objetivos.

El objetivo principal es la realización de una solución Business Intelligence completa, incluyendo unos Cuadros de Mando que optimicen el análisis de datos y las estadísticas demográficas por la aplicación de técnicas y herramientas de Business Intelligence para los archivos históricos de población del Ayuntamiento de Santander.

Para realizar la solución BI completa, se han de desarrollar los siguientes subobjetivos:

- Selección de una herramienta apropiada BI open source
- Estudio de la fuente de datos y modelo de datos
- Procesos ETL
- Creación de Cubos OLAP
- Realización de los Cuadros de Mando o Dashboards con las características anteriormente definidas.
- Pruebas y validación de los Dashboards
- Sugerencias de mejoras.

2. Herramientas y método.

En este apartado se describen las herramientas seleccionadas para el desarrollo de este proyecto, así como la metodología empleada. Durante el primer paso se tratará de seleccionar una herramienta de Business Intelligence adecuada para la realización de este proyecto y atendiendo a criterios como que sea open source, versión gratuita, de uso extendido y ampliamente documentada.

2.1. Selección de software BI.

Todos los años se publica la matriz de BI Gartner para Business Analytics donde se analizan las diferentes plataformas de BI. De manera gráfica y a través de una serie de criterios de evaluación, el cuadrante permite dar una perspectiva desde distintos ángulos sobre el posicionamiento de una empresa con respecto a sus competidores.

Figure 1. Magic Quadrant for Business Intelligence and Analytics Platforms



Figura 1: Gartner Magic Quadrant BI año 2017

Gartner define el eje X como el elemento que entrega “integridad de visión” y representa el conocimiento de los proveedores sobre cómo se puede aprovechar el momento actual del mercado para generar valor tanto para sus clientes como para ellos mismos. Por otro lado, el eje Y representa la “capacidad de ejecutar” y mide la habilidad de los proveedores para desarrollar con éxito su visión del mercado, además de analizar lo rápido que puede responder el proveedor a un cambio en las tendencias del mercado.

¿Cómo funciona un cuadrante mágico?

Un cuadrante mágico proporciona un posicionamiento gráfico competitivo de cuatro tipos de proveedores de tecnología, en mercados donde el crecimiento es alto y la diferenciación del proveedor es clara. Gartner distingue estas cuatro categorías:

Líderes: ejecutan bien en relación de su visión actual del negocio y están en buena posición para competir en el futuro.

Visionarios: entienden por dónde va el mercado o tienen una visión para el cambio de las reglas del mismo, pero todavía no lo manejan bien.

Jugadores de nicho: se centran con éxito en un pequeño segmento, o no se centran en superar o mejorar a su competencia.

Challengers: a día de hoy ofrecen una buena alternativa o pueden dominar un segmento grande, pero no demuestran una comprensión de la dirección del mercado. (Gartner Group, 2017, *Research Methodologies*).

Después de un breve estudio sobre algunas de las diferentes opciones, se selecciona **Pentaho** para este proyecto por constituir una completa plataforma BI, es modular, de implementación flexible, es open source y en su versión community es gratuita, con disponibilidad de muchos plugins y con extensa documentación y resolución de dudas en los foros de la comunidad.

Recientemente la plataforma Pentaho ha estado posicionada en el cuadrante “Jugadores de nicho” durante los años 2012 a 2015, mejoró posición durante el año 2016 como “Visionario” y este año 2017 vuelve a aparecer en el cuadrante de “Jugadores de nicho”.

Características de Pentaho BI.

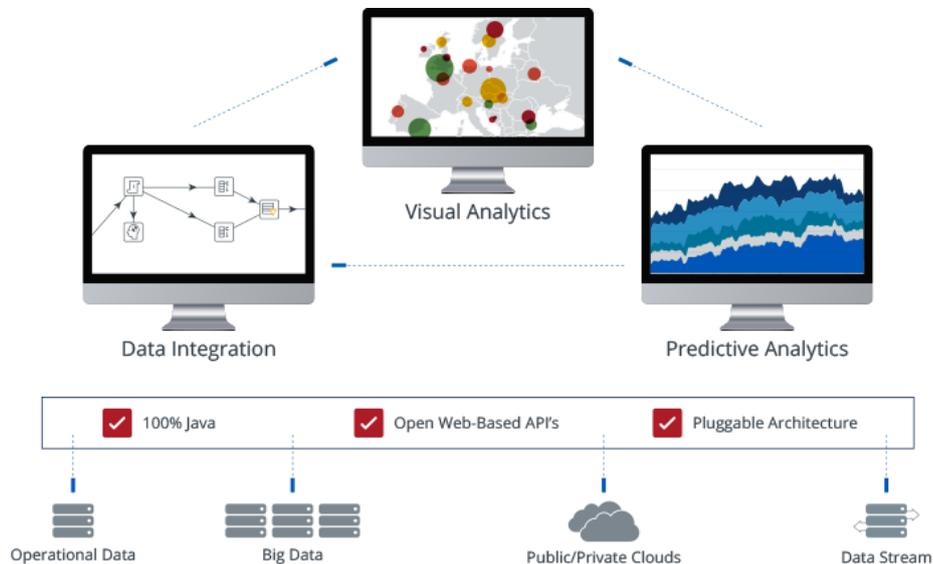


Figura 2: Características plataforma Pentaho

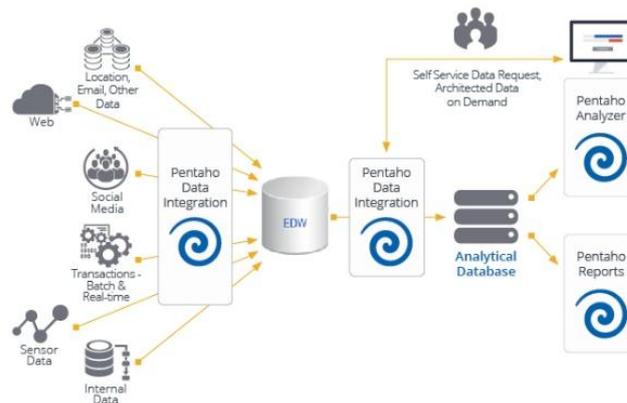
La plataforma Pentaho simplifica la preparación e integración de todos los datos e incluye un espectro de herramientas para analizar, visualizar, explorar, reportar y predecir fácilmente. Es open source, desarrollada en Java, integrable y extensible. (Pentaho Group,2017).

Los principales componentes son:

- Suite BI, es el componente base e incluye a Mondrian que forma parte del motor OLAP.
- Pentaho Reporting, para la generación de reportes.
- Pentaho Data Integration, para realizar procesos ETL.
- Schema Workbench, para creación de cubos Olap
- Community Dashboard Editor (CDE), para elaboración de Cuadros de mando.
- WEKA, para minería de datos.

■ Arquitectura

- A su vez, la arquitectura de Pentaho, consta de varias capas
- **Capa de datos:** utilizada para conectarse a cualquier origen de datos.



- **Capa del servidor:** capa intermedia dentro de la arquitectura, la aplicación "corre" en el servidor y, actúa como capa intermedia. Aquí, tenemos: **BA Server**, **Pentaho User Console** y **Pentaho Data Integration Server**.
 - **BA Server:** contenedor principal dónde se publican los dashboards e informes que se desarrollen, de modo que, posteriormente sean accesibles por los usuarios.
 - **Pentaho User Console:** nos permite establecer la seguridad a nivel de usuarios/roles y definir qué contenido pueden ver en función de los roles a los que pertenezcan.
 - **Pentaho Data Integration Server:** permite administrar y programar los trabajos ETL.
- **Capa del cliente:** compuesta por dos tipos principales:

- **Thin Client:** se ejecuta en el servidor y, entre otros, podríamos tener aplicaciones para la creación de dashboards con el **Community Dashboard Editor** o, herramientas la explotar la información subyacente al sistema a través de los visores web OLAP.
- **Thick Client:** representa las aplicaciones que no se ejecutan dentro del entorno de Pentaho pero, sí tienen una completa interactividad con éste, permitiendo publicar trabajos desarrollados con estas aplicaciones. Algunas de estas aplicaciones son: **Pentaho Data Integration**, **Pentaho Report Designer** o **Schema Workbench**.

Figura 3: Arquitectura plataforma Pentaho

(Stratebi , 2017 Comparativa de herramientas Business Intelligence).

2.2. Herramientas.

En esta sección se detallan los diferentes programas, plugins y herramientas seleccionados para la realización de este proyecto.

2.2.1. Pentaho BI.

La herramienta principal es la suite de Pentaho en su versión community 5.4, que incluye:

- biserver-ce-5.4.0.1-130 (Servidor BI de Pentaho)
- pdi-ce-5.4.0.1-130 (Procesos ETL)
- psw-ce-3.10.0.1-130 (Cubos OLAP)

2.2.2. Plugins para Pentaho BI.

Desde la opción Marketplace de Pentaho, que centraliza todos los plugins disponibles para la plataforma, se han instalado los siguientes Plugins:

- **Saiku Analytics** es un visor OLAP, open source, permite navegar cubos OLAP y mostrar esa información en forma de tablas y gráficos, soportando las funcionalidades típicas de los entornos OLAP como filtrar, desglosar y agrupar datos, exportar a formato CSV, Excel, generación de consultas en MDX.
Se ha utilizado esta herramienta para generar consultas MDX y también para crear widgets Saiku como un componente externo en el editor CDE de los Cuadros de Mando. Por ejemplo, en la visualización de la tabla de Años de Alta y Baja en el cuadro de mando de Flujo Migratorio.
- **IvyBC (Ivy IS Bootstrap Components)**, es un plugin para Pentaho que contiene una selección de componentes personalizados que permiten crear Dashboards (Cuadros de Mando), también para móviles. Dispone de una serie de componentes tales como: Bootstrap Button, Bootstrap Select, Bootstrap Panel. Se han utilizado estos componentes en todos los cuadros de mando, por ejemplo, el Bootstrap Button para enlazar los 3 cuadros de mando.
- **IvyDC (Ivy Dashboard Components)**, Ivy Dashboard Components es un plugin para Pentaho que permite utilizar una serie de componentes personalizados en los Dashboards tales como: Gauges, Funnel Chart, Vector Map, Video Player, Ion Range Slider. Por ejemplo, se ha utilizado el componente Ion Range Slider, en la selección del Período de años en el cuadro de mando de Bajadas de Población.

2.2.3. Otras herramientas.

- **Pencil Mockup** es una herramienta gratuita y de código abierto que permite diseñar prototipos para aplicaciones y sitios web. Permite crear bocetos exportables a PNG, PDF. Se ha utilizado para crear los mockups de los tres cuadros de mando.
- **Mockup Tiger**, en la opción wireframes on line, se ha utilizado para generar imágenes de los Pie Chart, Bar Chart y Line Chart utilizados en los mockups de los cuadros de mando.
- **SQLite Manager**, plugin para el navegador Firefox para gestionar la base de datos sqlite.

2.3. Metodología.

Teniendo en cuenta las fases que hay que realizar para el desarrollo de una aplicación con herramientas Business Intelligence, se determinan las siguientes fases en este proyecto:

- Análisis de requisitos.
- Diseño e implementación de la solución
 - Estudio de la fuente de datos y Modelo de datos
 - Creación del modelo Esquema en Estrella
 - Proceso ETL
 - Extracción, Transformación y Carga de datos
 - Creación del cubo OLAP
 - Realización de los cuadros de mando.
 - Evaluación y pruebas de los cuadros de mando
 - Integración en una página web.
- Conclusiones y Trabajos Futuros

3. Análisis y captura de requisitos.

Se determinan estos requisitos a lo largo del diseño y realización de las sucesivas etapas del proyecto.

- Se desea consultar la siguiente información del fichero de Bajas de Población (Motivo, fecha baja, estudios (nivel, titulación), nacionalidad, destino (país, provincia), nacimiento (país, provincia), edad, sexo, localización (distrito, sección, código postal).
- Se desea consultar la siguiente información del fichero de Población de Santander (Fecha de alta, estudios (nivel, titulación), nacionalidad, procedencia (país, provincia), nacimiento (país, provincia), edad, sexo, localización (distrito, sección, código postal).
- Asimismo se desea consultar los datos agregados por tiempo en distinto nivel de detalle:
 - Año, mes, día en fecha de baja
 - Año, mes en fecha de alta
- La visualización de los datos en la dimensión tiempo será como mínimo a nivel de años.
- Se definirán rangos de edades con intervalos de 5, 10 años.
- Los cuadros de mando permitirán como mínimo analizar los datos por 3 atributos seleccionables, tendrán un mínimo de 3 parámetros.
- El campo "distrito" será siempre uno de los seleccionables.
- Se definirá un parámetro adicional con un rango de valores (10,15,20,25) para que el usuario pueda elegir qué número de registros se visualizan en los gráficos o consultas del tipo "Top N Países".
- Los datos se visualizarán tanto en gráficos como en tablas.
- Las tablas podrán ser exportadas a Excel.
- En el acceso a los cuadros de mando estará implícito el usuario / password del Login, no lo tendrá que introducir el usuario.
- Desde un cuadro de mando se accederá directamente al resto de cuadros de mando, sin necesidad de volver a menú.

4. Diseño e implementación de la solución

En este capítulo se detallan todos los procesos realizados para el desarrollo de esta aplicación de Business Intelligence, empezando por el tratamiento de los datos (procesos ETL), el diseño de los cubos OLAP necesarios para realizar los análisis multidimensionales, el diseño y la creación de los propios Cuadros de Mando que permitan visualizar los datos, y ya por último varias pruebas realizadas con los mismos.

4.1. Estudio de fuentes de datos y modelo de datos.

En el Open Data del Ayuntamiento de Santander, en el sector Demografía se encuentran disponibles unos conjuntos de datos de Población Histórica y de Bajas de Población histórica que van a constituir nuestra fuente de datos externa, concretamente:

- Bajas: (Bajas Población 2010, 2011, 2012 y Motivos de baja).
- Población: (Población 2010, 2011, 2012, 2013, 2014).
- Población Actual: (Países y Niveles de Estudio).

Tras observar los conjuntos de datos de los que disponemos, pasamos a realizar un análisis más profundo sobre la información que contienen, con objeto de poder establecer el modelo de datos que dé respuesta a las preguntas de análisis que se han planteado en este trabajo.

En las Bajas:

- Se determinan las vistas o perspectivas de análisis que se pueden realizar a partir de esos datos (motivo baja, fecha baja, estudios, edad, sexo, nacionalidad, lugar nacimiento, lugar destino, distrito, código postal):
- Se seleccionan los atributos que conformarán las dimensiones que representaran estas vistas o perspectivas de análisis, definiendo jerarquía y niveles en las dimensiones:
 - o para la dimensión motivos de baja: motivo baja, motivo INE, descripción motivo baja
 - o para la dimensión de fechas de baja: fecha baja, año baja, mes baja, día baja
 - o para la dimensión nacionalidad: código nacionalidad, nombre país
 - o para la dimensión estudios: nivel estudios, titulación
 - o para la dimensión edad: tramo de edad de 10 años
 - o para la dimensión lugar de nacimiento: código país, código provincia, nombre país, nombre provincia.
 - o para la dimensión lugar de destino: código país, código provincia, nombre país, nombre provincia.
 - o para la dimensión sexo: código sexo, descripción sexo.
- La medida se calcula con la función "count" del número de bajas para las Bajas.

En las Población:

- Se determinan las vistas o perspectivas de análisis que se pueden realizar a partir de esos datos (año padrón, fecha alta, estudios, edad, sexo, nacionalidad, lugar nacimiento, lugar procedencia, distrito, código postal):
- Se seleccionan los atributos que conformarán las dimensiones que representaran estas vistas o perspectivas de análisis, definiendo jerarquía y niveles en las dimensiones:
 - o para la dimensión de fechas de alta: fecha alta, año alta, mes alta
 - o para la dimensión nacionalidad: código nacionalidad, nombre país
 - o para la dimensión estudios: nivel estudios, titulación
 - o para la dimensión edad: tramo de edad 5 años, tramo de edad 10 años

- para la dimensión lugar de nacimiento: código país, código provincia, nombre país, nombre provincia.
- para la dimensión lugar de procedencia: código país, código provincia, nombre país, nombre provincia.
- para la dimensión sexo: código sexo, descripción sexo.
- La medida se calcula con la función “count” del número de empadronados en Población.
- El nivel de detalle (grano) que se considera es:
 - para las Bajas (totalizan 26.000 registros) en fecha baja se puede llegar hasta el nivel del día (año-mes-día en fecha baja).
 - para Población (totalizan casi 900.000 registros) en fecha alta se considera llegar a nivel del mes (año-mes en fecha alta).

4.1.1. Esquemas en estrella

Se realizan los esquemas en estrella para las Bajas de Población Histórica y para los ficheros de Población Histórica.

En el esquema en estrella las tablas de dimensión se encuentran desnormalizadas, es decir, toda la información referente a una dimensión se almacena en la misma tabla. En el centro del esquema está la tabla de hechos con conexiones (Foreign Keys) hacia cada una de las dimensiones.

Esquema en estrella para Bajas Población

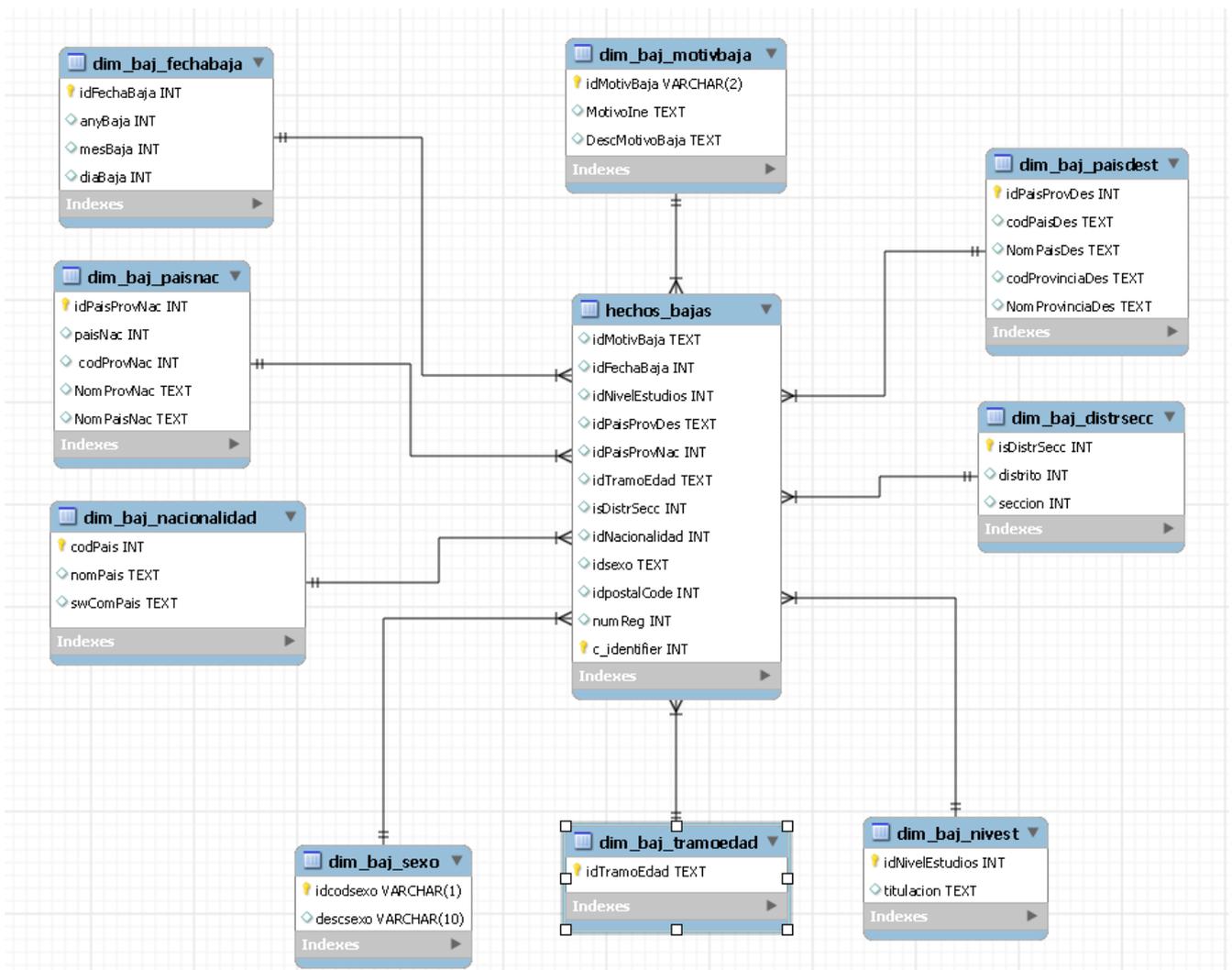


Figura 4: Esquema en estrella Bajas de Población

La figura 4 muestra el esquema en estrella utilizado para representar los datos relativos a las Bajas de Población. Se puede observar un hecho central (la tabla hechos_bajas) cuyos atributos direccionan a las Primary Keys de las tablas de dimensiones, con los atributos y agregación de datos indicados anteriormente para las Bajas.

La tabla hechos_bajas también contiene la medida, un campo “count” del número de bajas.

Esquema en estrella para Población Histórica

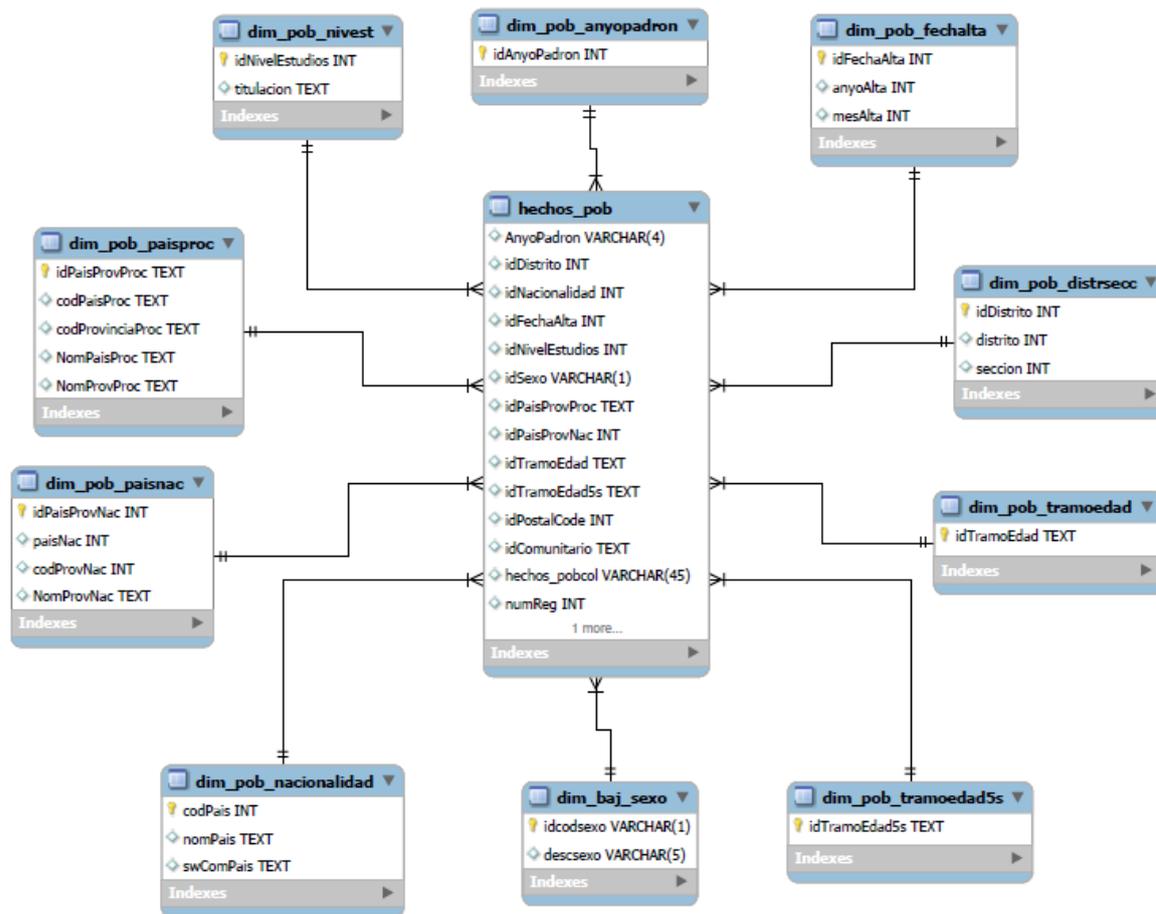


Figura 5: Esquema en estrella Población Histórica

La figura 5 muestra el esquema en estrella utilizado para representar los datos de Población Histórica. Se representa un hecho central (la tabla hechos_pob) cuyos atributos direccionan a las Primary Keys de las tablas de dimensiones, con los atributos y agregación de datos indicados anteriormente para Población.

Tiene una dimensión adicional según el año de padrón al que se refieren los datos y también se ha considerado un tramo de edad adicional de 5 años.

La tabla hechos_pob también contiene la medida, un campo “count” del número de empadronados.

4.2. Procesos ETL.

Un proceso ETL (Extraer, Transformar y Cargar) permite extraer los datos desde diversas fuentes, depurarlos, homogeneizarlos y cargarlos en otra base de datos o datawarehouse.

Para definir los procesos ETL necesarios para transformar los datos de entrada y almacenarlos en la base de datos, vamos a utilizar la herramienta Pentaho-Data Integration (PDI). Dicha herramienta utiliza una serie de términos necesarios para comprender la solución empleada y que por lo tanto, conviene presentar con antelación.

Una Transformación:

- . Representa una tarea ETL. Orientada a los Datos.
- . Es una colección de “pasos”. Cada paso es una operación particular sobre datos.
- . Los “pasos” se conectan entre sí a través de “saltos”, que indican el flujo de los datos.
- . Los “pasos” trabajan de manera **simultánea y asíncrona**

Un Trabajo:

- . Orientado a la Tarea y a los Datos.
- . Normalmente un Trabajo se compone de una o más Transformaciones/Trabajos, que serán ejecutadas **secuencialmente**.
- . La ejecución de cada “**Entrada de Trabajo**” presenta una salida de **status**, que puede ser analizada para la realización de diferentes acciones(p.e. tratamiento de errores). (Pierri,2011, Manipulación de datos con Kettle).

4.2.1. Procesos ETL para las Bajas de Población.

Se describe a continuación el proceso ETL para los datos de Bajas de Población, explicando los pasos más significativos para crear las tablas de dimensiones y la tabla de hechos.

Extracción de los datos de Entrada:

Se comienza con los siguientes pasos para “extraer” los ficheros de Entrada:

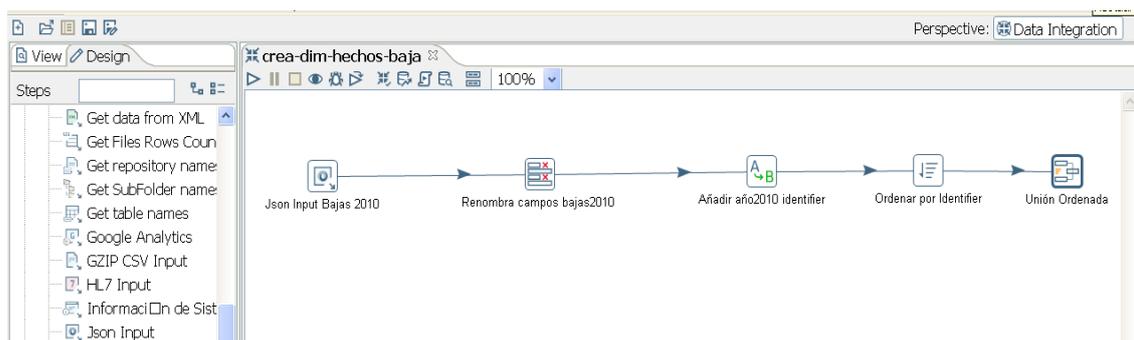


Figura 6: ETL_ Bajas_ pasos inicio fichero entrada json Bajas 2010

Se describen a continuación los pasos o steps más significativos de esta transformación:

Pasos:

- 1.- Paso JSON Input Bajas 2010 : se indica el path del fichero de entrada en formato json.
- 2.- Paso Renombra campos bajas2010 : se explica a continuación la necesidad de este paso.

DATOS

http://datos.santander.es/api/rest/datasets/poblacion_bajas2010.json

```
{
  "summary": {
    "items": 9101,
    "items_per_page": 50,
    "pages": 183,
    "current_page": 1,
    "resources": [
      {
        "ayto:codSexo": "39007",
        "ayto:paisNac": "CANTABRIA",
        "ayto:Motivo": "B3",
        "ayto:distrito": "BCR",
        "ayto:NomPaisDes": "108",
        "ayto:fechaAltaAnyo": "2006-02-10",
        "ayto:Edad": "V",
        "ayto:nacionalidad": "108",
        "ayto:codProvinciaProc": "ESPAÑA",
        "ayto:codPaisDes": "1",
        "ayto:NomProvNac": "39",
        "ayto:FechaBaja": "NAVARRA",
        "ayto:seccion": "02",
        "ayto:codNacional": "108",
        "ayto:fechaAltaMes": "2006",
        "ayto:codProvNac": "18",
        "gn:postalCode": "007",
        "dc:identifiier": "00000RHBT003469911",
        "ayto:NomProvinciaDes": "31",
        "ayto:NomPaisProc": "108",
        "ayto:comunitario": "02",
        "ayto:MotivoIne": "2010-07-22",
        "ayto:titulacion": "31",
        "ayto:nivelEstudios": "GUIPUZCOA",
        "ayto:NomProvProc": "20",
        "ayto:codProvinciaDes": "ESPAÑA",
        "ayto:codPaisProc": "ESPAÑA",
        "ayto:fechaAltaPad": "Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO",
        "dc:modified": "2017-04-19T22:30:10.481Z",
        "uri": "http://datos.santander.es/api/datos/poblacion_bajas2010/00000rhbt003469911.json"
      }
    ]
  }
}
```

Figura 7: ETL_ Bajas_ visualización fichero entrada json Bajas 2010

Problema con los ficheros de Baja Población: como se observa en la captura de pantalla anterior, al abrir el fichero JSON de Bajas 2010 la mayoría de los campos no se corresponden el nombre del campo con su contenido. Estas discrepancias deben de ser el resultado de algún tipo de fallo al recolectar o subir estos datos.

Por ejemplo, en ese registro hay campos con idéntico contenido como el valor “108” que aparece 3 veces y que puede corresponder al código de nacionalidad, país destino o país procedencia, algo similar ocurre con algunas fechas.

Para solucionarlo se estableció una consulta en una base de datos entre este fichero de bajas 2010 y el fichero de Población 2010 relacionados por la clave “dc:identifiier” que tienen ambos ficheros, ya que la mayoría de los registros contenidos en el fichero bajas durante el año 2010 están contenidos en el fichero Población 2010 (a 1-enero-2010), exceptuando los que han sido alta y baja durante el mismo año 2010, y así se pudo determinar el contenido correcto del fichero de Bajas ya que el fichero Poblacion presenta correctamente los campos.

En la captura de pantalla siguiente se detalla este paso Renombra campos bajas2010, se corrigen los nombres de los campos de entrada y se soluciona este error.

Ej: el campo “ayto:codProvNac (que en el fichero json tiene el valor 18) se renombra como “ayto:Edad” que es el nombre correcto.

Nombre paso Renombra campos bajas2010

Selecciona & Modifica Eliminar Meta-información

Campos :

#	Nombre campo	Renombrar a	Longitud	Precisión
1	dc:identifir			
2	ayto:Motivo			
3	ayto:codProvNac	ayto:Edad	3	
4	ayto:Motivolne	ayto:FechaBaja		
5	ayto:distrito	ayto:Motivolne		
6	ayto:codProvinciaDes	ayto:NomPaisDes		
7	ayto:paisNac	ayto:NomProvNac		
8	ayto:FechaBaja	ayto:NomProvinciaDes		
9	ayto:nacionalidad	ayto:codNacional		
10	ayto:NomPaisDes	ayto:codPaisDes		
11	ayto:NomProvNac	ayto:codProvNac		
12	ayto:NomProvinciaDes	ayto:codProvinciaDes		
13	ayto:Edad	ayto:codSexo		
14	ayto:seccion	ayto:distrito		
15	ayto:fechaAltaAnyo	ayto:fechaAlta		
16	ayto:fechaAltaMes	ayto:fechaAltaAnyo		
17	ayto:comunitario	ayto:fechaAltaMes		
18	ayto:codPaisProc	ayto:nacionalidad		
19	ayto:titulacion	ayto:nivelEstudios		
20	ayto:codNacional	ayto:paisNac		
21	gn:postalCode	ayto:seccion		
22	ayto:fechaAltaPad	ayto:titulacion		
23	ayto:codSexo	gn:postalCode		

Figura 8: ETL_ Bajas_paso_Renombrar campos bajas 2010

3.- Paso Añadir año2010 identifir:

En este paso se trata el campo “dc:identifir”, convirtiéndolo a formato numérico, se reemplaza la parte alfanumérica por “2010”, que es el año del fichero de bajas del que procede ya que posteriormente se unirán en uno solo todos los ficheros de Bajas.

Step name Añadir año2010 identifir

Fields string

	In stream field	Out stream f...	use RegEx	Search	Replace with	Set empty stri...	Replace with field
1	dc:identifir		N	00000RHBT	2010	N	

Figura 9: ETL_ Bajas_paso Añadir año2010 identifir

4.- Paso Unión Ordenada:

En este paso se unen todos los ficheros de Bajas previamente clasificados por el campo “dc:identifir”.

Nombre de paso Unión Ordenada

Campos :

	Nombre cá...	Ascendente
1	dc:identifir	S
2		

Figura 10: ETL_ Bajas_paso Unión Ordenada

La transformación con todos los ficheros de entrada quedaría así:

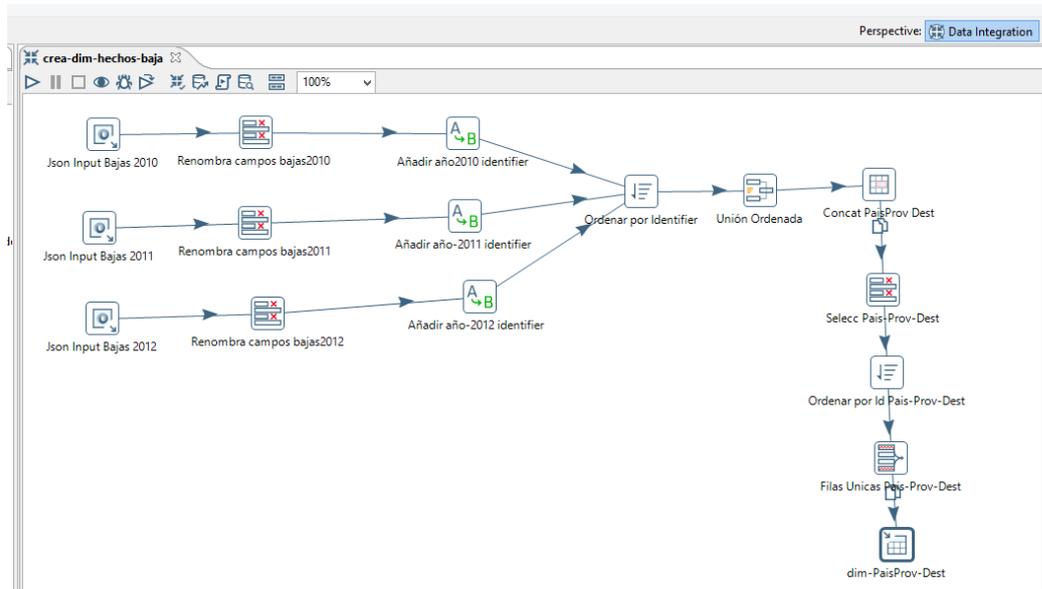


Figura 11: ETL_ Bajas_paso crear dimensión Pais_Prov_Destino

- La imagen contiene también la creación de la dimensión Pais_Prov_Destino, cuyos pasos son:
- 5.- Paso Concat PaisProv Dest: se concatenan codPais y codProv destino para formar la clave de la dimensión.
 - 6.- Paso Selecc Pais-Prov-Dest: Se seleccionan el resto de campos que forman la dimensión.
 - 7.- Paso Ordenar por Id Pais-Prov-Dest: Se ordenan por la clave de la dimensión.
 - 8.- Paso Filas Unicas Pais-Prov-Dest: Se forman filas únicas por clave de la dimensión.
 - 9.- Paso dim-PaisProv-Dest: Se graba la salida a tabla de la base de datos con la conexión definida anteriormente.

Transformación: bajas-crear-dim-hechos.ktr

La transformación completa “bajas-crear-dim-hechos” queda así:

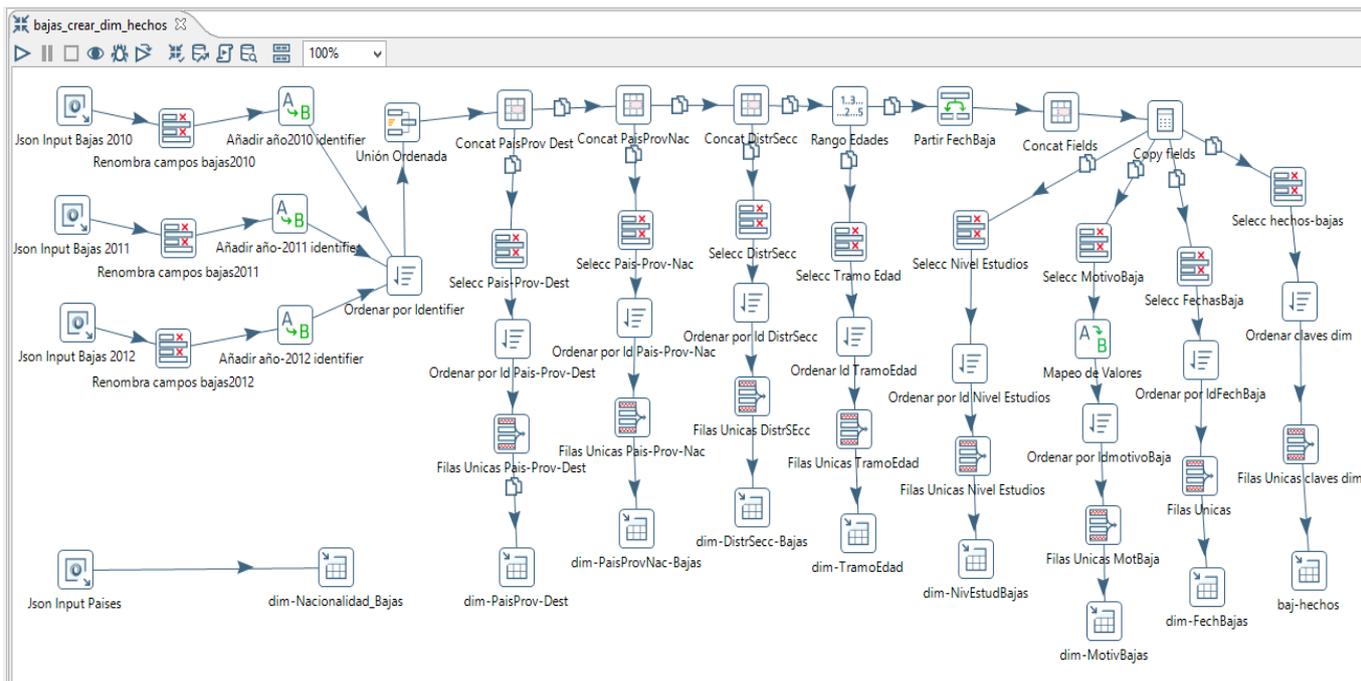


Figura 12: ETL_ Bajas_Transformación para crear todas dimensiones y hechos Bajas

Pasos para crear “hechos-Bajas:

1.- Paso Selecciona hechos-bajas:

Para crear la tabla de hechos, se seleccionarán el campo “identifier” y todos los campos que se han definido como claves de todas las dimensiones, serán las Foreign Key en la tabla de hechos que se relacionarán con las Primary Keys de las tablas dimensiones.

#	Nombre campo	Renombrar a	Longitud	Precisión
1	dc:identifier	identifier	13	
2	idMotivBaja			
3	idFechaBaja			
4	idNivelEstudios			
5	idPaisProvDes			
6	idPaisProvNac			
7	idTramoEdad			
8	isDistrSecc			
9	ayto:codNacional	idNacionalidad	3	
10	ayto:codSexo	idsexo	1	
11	gn:postalCode	idpostalCode	5	
12				

Figura 13: ETL_ Bajas_paso Selecciona campos hechos Bajas

2.- Paso Ordenar claves dim: ordenar por las claves de las dimensiones

3.- Paso Filas Unicas claves dim: agrupar por las claves de las dimensiones y se crea un campo contador “numReg”.

Los registros en la tabla de hechos no se graban individualmente sino agrupados por las claves de las dimensiones. Este campo numReg indica el nº de registros que hay por cada fila de la tabla de hechos.

Cuando se cree el cubo OLAP, la medida será una función suma de este campo numReg.

Nombre de paso: Filas Unicas claves dim

Settings

Añadir contador a la salida? Campo Contador: numReg

Redirect duplicate row Error description

Campos de comparación (dejar vacío para utilizar la fila entera)

#	Nombre de campo	Ignorar Mayúsculas/Minúsculas
1	idMotivBaja	N
2	idFechaBaja	N
3	idNivelEstudios	N
4	idPaisProvDes	N
5	idPaisProvNac	N
6	idTramoEdad	N
7	isDistrSecc	N
8	idNacionalidad	N
9	idsexo	N
10	idpostalCode	N

Figura 14: ETL_ Bajas_paso Filas Unicas claves dim

4.- Paso baj-hechos:

Se crea y carga la tabla baj_hechos, especificando la conexión anteriormente creada a la base de datos sqlite.

Nombre de paso: baj-hechos
Conexión: connect_lite [Editar...]
Esquema destino: []
Tabla destino: baj-hechos
Tamaño de transacción (commit): 1000
Vaciar tabla:
Ignorar errores de inserción:
Specify database fields:

Main options | Database fields

Fields to insert:

#	Table field	Stream field
1	idMotivBaja	idMotivBaja
2	idFechaBaja	idFechaBaja
3	idNivelEstudios	idNivelEstudios
4	idPaisProvDes	idPaisProvDes
5	idPaisProvNac	idPaisProvNac
6	idTramoEdad	idTramoEdad
7	isDistrSecc	isDistrSecc
8	idNacionalidad	idNacionalidad
9	idsexo	idsexo
10	idpostalCode	idpostalCode
11	numReg	numReg

Figura 15: ETL_ Bajas_paso Salida a Tabla baj_hechos

5.-Paso crear "dim-Nacionalidad-Bajas:

También se hace necesario incluir en esta transformación el fichero Poblacion_Pais, descargado de los ficheros Población del Open Data, para poder formar la dimensión dim_Nacionalidad_Bajas e incorporar los campos cod_Pais y nombre_Pais (o Nacionalidad).

Todo el proceso de creación de las tablas de dimensiones y hechos para las Bajas de Población se han ejecutado en 1 sola transformación, si fueran necesarias más transformaciones se articularían en un job, que ejecutaría secuencialmente esas transformaciones como se verá más adelante con los ficheros de Población.

4.2.2. Procesos ETL Población.

De forma análoga al proceso realizado con los datos de Bajas se describe a continuación el proceso ETL para los datos de Población, explicando los pasos más significativos para crear las tablas de dimensiones y la tabla de hechos.

Si visualizamos 1 registro vemos que la información se presenta correctamente en el sentido de que se corresponde el nombre del campo con su contenido, no existe el problema que había con los ficheros de Bajas de Población.

```
{
  "summary": {
    "items": 176031,
    "items_per_page": 50,
    "pages": 3521,
    "current_page": 1,
    "resources": [
      {
        "ayto:codSexo": "M",
        "ayto:paisNac": "108",
        "ayto:distrito": "03",
        "ayto:fechaAltaAnyo": "2006",
        "ayto:Edad": "32",
        "ayto:nacionalidad": "ESPAÑA",
        "ayto:codProvinciaProc": "39",
        "ayto:NomProvNac": "CANTABRIA",
        "ayto:seccion": "003",
        "ayto:codNacional": "108",
        "ayto:fechaAltaMes": "01",
        "ayto:codProvNac": "39",
        "gn:postalCode": "39001",
        "dc:identifier": "00000RHBT003469951",
        "ayto:NomPaisProc": "ESPAÑA",
        "ayto:comunitario": "S",
        "ayto:titulacion": "Enseñanza Primaria incompleta, cinco cursos de EGB o equivalente, Certificado de escolaridad o equivalente",
        "ayto:nivelEstudios": "22",
        "ayto:NomProvProc": "CANTABRIA",
        "ayto:codPaisProc": "108",
        "ayto:fechaAltaPad": "2006-01-03",
        "dc:modified": "2017-04-25T22:36:36.095z",
        "uri": "http://datos.santander.es/api/datos/poblacion_pob2014/00000rhbt003469951.json",
      }
    ]
  }
}
```

Figura 16: ETL_ Poblac_visualización fichero entrada json Pob 2014

Hay algunas diferencias respecto a las Bajas sobre todo en el nº de registros (26.000 para las Bajas y casi 900.000 registros para Población correspondientes a 5 años de Padrón.

Extracción de los datos de Entrada:

Vamos a ver los pasos correspondientes al tratamiento de todos los ficheros de entrada y creación de 1 dimensión:

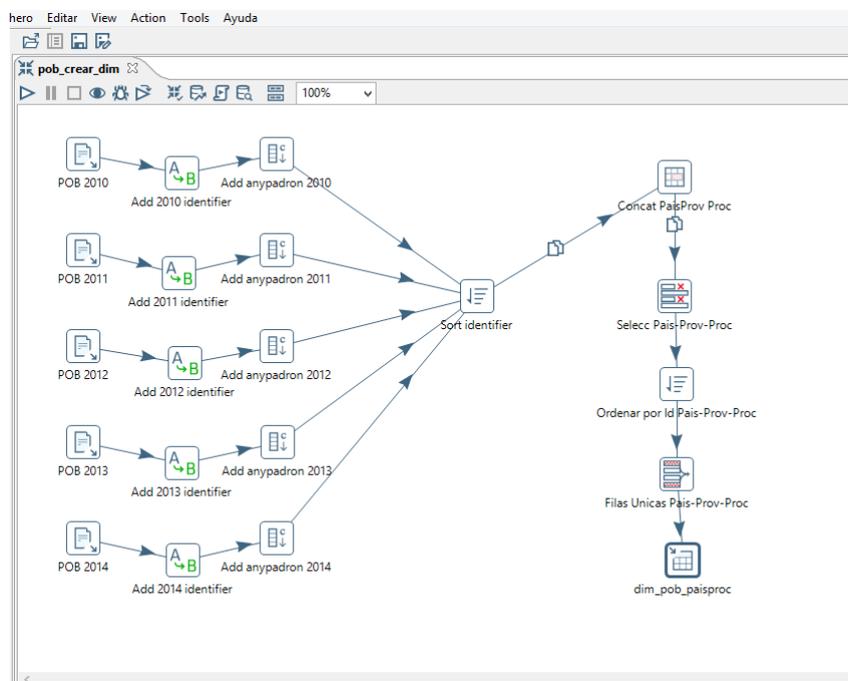


Figura 17: ETL_ Poblac_paso Unir fich. entrada y crear dimensión Pais_Prov_Proced.

Descripción de los pasos:

1.-Paso POB 2010 : se indica el path del fichero de entrada en formato csv, ya que para este nº elevado de registros era el formato en el que mejor se ejecutaba la transformación, (con los formatos json o xml daba error de insuficiente memoria o iba muy lento). También se seleccionan los campos del registro que pasarán a los siguientes pasos.

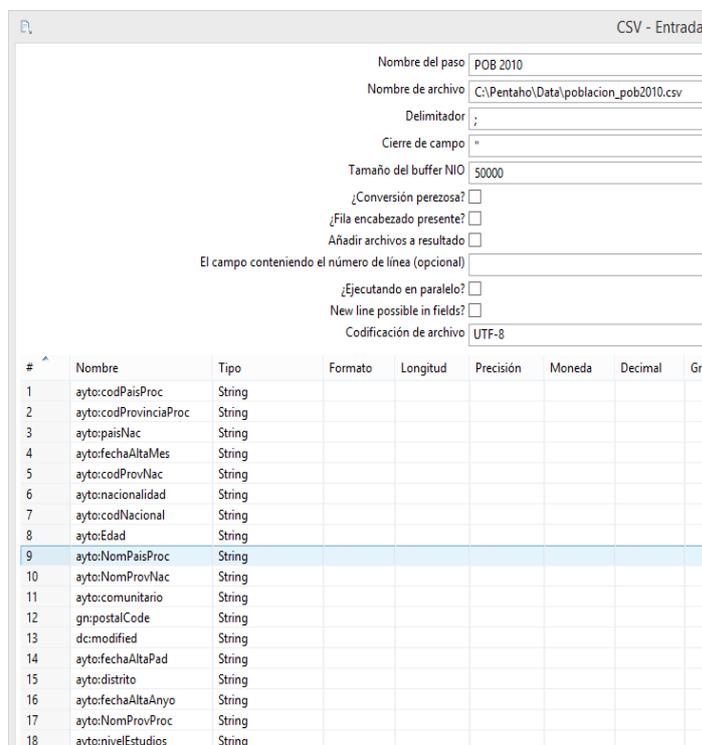


Figura 18: ETL_ Poblac_paso fichero entrada Pob2010

2.- Paso Add 2010 identifier : se convierte a numérico el campo “identifier”, campo identificador del registro, se reemplaza la parte alfanumérica por “2010”, el año padrón de dónde procede el registro, antes de unirse los ficheros de entrada y puede ser útil a la hora de visualizar registros y localizar errores. También se siguió este criterio en las Bajas.

3.- Paso Add anypadron 2010:



Figura 19: ETL_ Poblac_paso Add anypadron 2010

Es necesario añadir este campo “AnyoPadron” ya que como se observa en la figura 16 de visualización de un registro del fichero Población 2014 no aparece ningún campo con contenido 2014, este campo deberá incorporarse a la tabla de hechos y generará a su vez una dimensión “dim_pob_anyopadron”, ya que el tratamiento a la hora de realizar estadísticas del fichero Población difiere del tratamiento del fichero de Bajas por el siguiente motivo:

- los 3 ficheros de Bajas totalizan casi 27.000 registros y se pueden acumular esos registros durante un período: sería correcto decir que durante el período 2010-2012, en Santander se produjeron casi 27.000 bajas de personas empadronadas.
- los 5 ficheros de Población histórica totalizan casi 900.000 registros y no se pueden acumular esos registros durante un período, ya que el fichero de Población es un reflejo de la población existente (un recuento de personas) a una fecha determinada, y no se puede referir a la población de Santander durante el período 2010-2014 como de casi 900.000 personas empadronadas. Por ello, en las estadísticas siempre se considerará este campo, se tratará de comparativas seleccionando años de padrón.

Observaciones:

Hay un aspecto que podría ser acumulable desde los ficheros de Población anuales y se trata de las altas producidas durante el año anterior al de Padrón,(ya que el Padrón es a 1 de enero), es decir:

- en el fichero Población 2010 (a 01-01-2010), serían los registros con fecha-alta entre 01-01-2009 al 31-12-2009)
- en el fichero Población 2011 (a 01-01-2011), serían los registros con fecha-alta entre 01-01-2010 al 31-12-2010), y así con los otros años...

Se podría decir que las altas de población durante el período (2010-2014) serían la suma de estas altas anuales.

Así todo, sería una cifra aproximada de altas ya que no se están incluyendo las personas que han sido alta y baja durante el mismo año, es decir:

- una persona empadronada con fecha-alta 03-03-2010 y posteriormente se da de baja con fecha-baja 04-11-2010, esta persona constará como baja en el fichero de bajas 2010, pero en los ficheros de población, esta persona no existía a 01-01-2010 (fichero Población2010) y tampoco existe a 01-01-2011 (fichero Población 2011), este alta quedaría sin contabilizar.

Para contabilizar exactamente las altas de población hubiera sido necesario disponer de un fichero anual de Altas de Población, e incluyendo también el motivo del alta (alta por nacimiento, alta desde el extranjero, alta desde otro municipio de España, etc.), entonces se podría hacer para las altas un tratamiento similar a los ficheros de bajas.

Desde la perspectiva de diseño dimensional, se puede decir:

- la medida (num_empadronados) en Población histórica no es aditiva.
- la tabla de hechos (pob_hechos) en Población histórica es del tipo snapshot.
- la medida (num_bajas) en Bajas de Población es aditiva.
- la tabla de hechos (baj_hechos) en Bajas Población es del tipo transaccional.

4.- Paso Sort identifier:

Se van uniando los registros de todos los ficheros de población de entrada clasificados por el campo identifier.

De igual forma a la transformación que se realizó con las Bajas se van incorporando el resto de las dimensiones y la tabla de hechos.

La transformación completa “pob_crear_hechos_y_todas_dimensiones” quedaría así:

Transformación: pob crear hechos y todas dimensiones.ktr

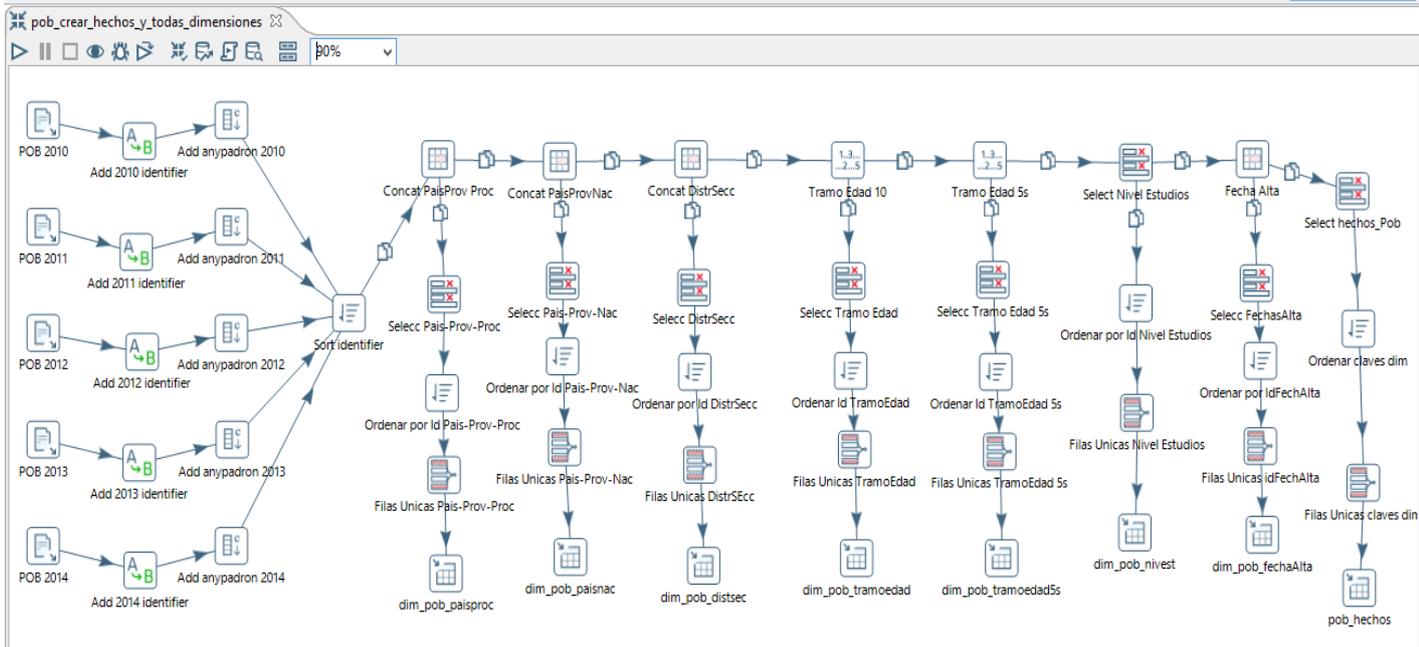


Figura 20: ETL_ Poblac_Transformación para crear todas dimensiones y hechos Población

A modo de ejemplo, se podría articular la ejecución secuencial de estas 2 transformaciones en un job (ficheros con extensión .kjb), como ilustra la figura siguiente:

job_crea_Bajas_Pob_Hechos_Dim.kjb

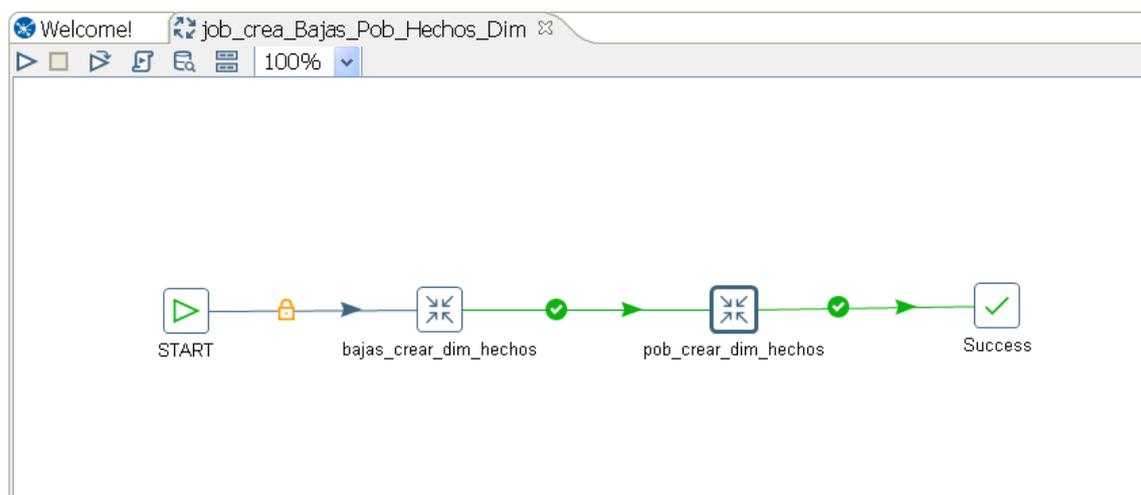


Figura 21: ETL_ Job_crear tablas y dimensiones Bajas y Población

Así se finaliza el proceso ETL para los ficheros de Bajas y Población quedando creadas y cargadas todas las tablas de dimensiones y hechos.

4.3. Creación Cubos OLAP.

Un cubo OLAP (Procesamiento Analítico en Línea) se define como una base de datos multidimensional, compuestas de medidas o dimensiones. Un cubo puede tener un número cualquiera de dimensiones, estas son las diferentes vistas de los datos; y una serie de medidas, representadas de forma numérica.

El almacenamiento físico de los datos se suele realizar sobre algún tipo de base de datos relacional, usando un esquema de estrella o copo de nieve. Las medidas se calculan a través de los registros de la tabla de hechos.

A continuación se describe la utilización de la herramienta gráfica de Pentaho SchemaWorkbench para la creación de los cubos OLAP para los datos de Bajas y Población.

Diseño Cubo OLAP Bajas de Población.

En la plataforma Open Source de PENTAHO se incluye la herramienta OLAP Mondrian, que permite crear cubos de información para análisis multidimensional.

Mondrian es un motor de procesamiento analítico en línea (OLAP) escrito en Java. Ejecuta consultas escritas en el lenguaje MDX, lee datos de una base de datos relacional (RDBMS), por tanto es un servidor ROLAP y presenta los resultados en un formato multidimensional a través de una API Java.

Esquema Mondrian

*Un esquema define una base de datos multi dimensional. Contiene un **modelo lógico**. Este modelo a su vez debe mapear a un **modelo físico**.*

- *El modelo lógico consiste en las estructuras usadas para escribir consultas en **lenguaje MDX**. Está formado por cubos, dimensiones, jerarquías, niveles y miembros.*
- *El modelo físico son las tablas de datos indicadas en el modelo lógico. Es típicamente un esquema en estrella, implementado en una base de datos relacional.*

*Los esquemas Mondrian son **archivos de tipo xml**. Se pueden editar manualmente, o bien por medio de la herramienta gráfica **Schema Workbench** de Pentaho. (Pentaho Group, 2017, Mondrian Documentation).*

Se siguen los siguientes pasos para crear el cubo OLAP:

- Se añade al Esquema el Cubo para las Bajas de Población y lo 1º que solicita es la tabla de hechos asociada:

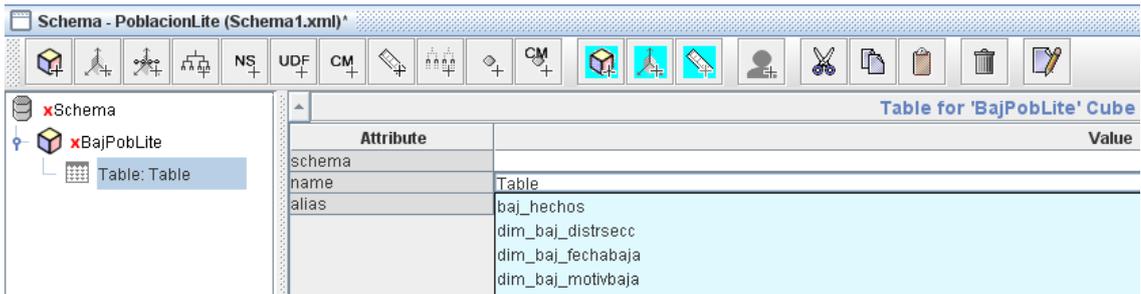


Figura 22: OlapCube_Bajas_Asociar Tabla hechos

- Se empieza a añadir las Dimensiones: dimensión MotivosBaja, se indica su Foreign Key en Tabla hechos, y el tipo de dimensión (Standard o Time).

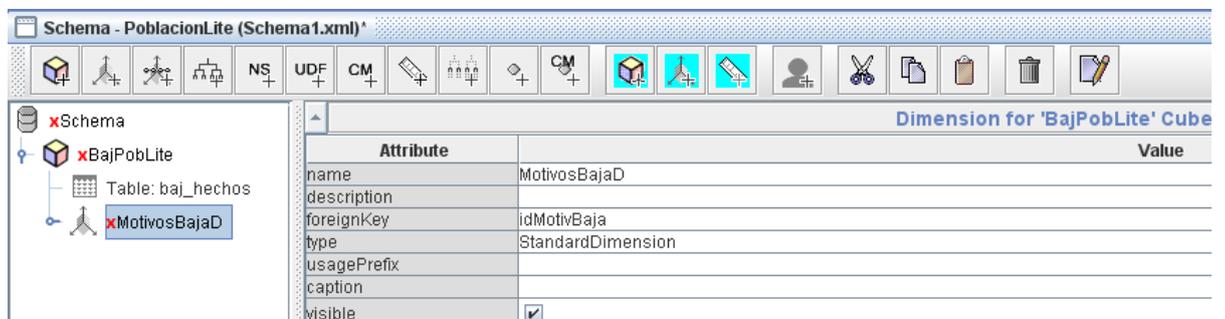


Figura 23: OlapCube_Bajas_Crear dimensión Motivos Baja

- Se añade la jerarquía y el nombre de la tabla dimensión correspondiente y se da un nombre al campo "AllMemberName" que se utilizará cuando se totalice por todos los registros de esa dimensión en las consultas Olap.

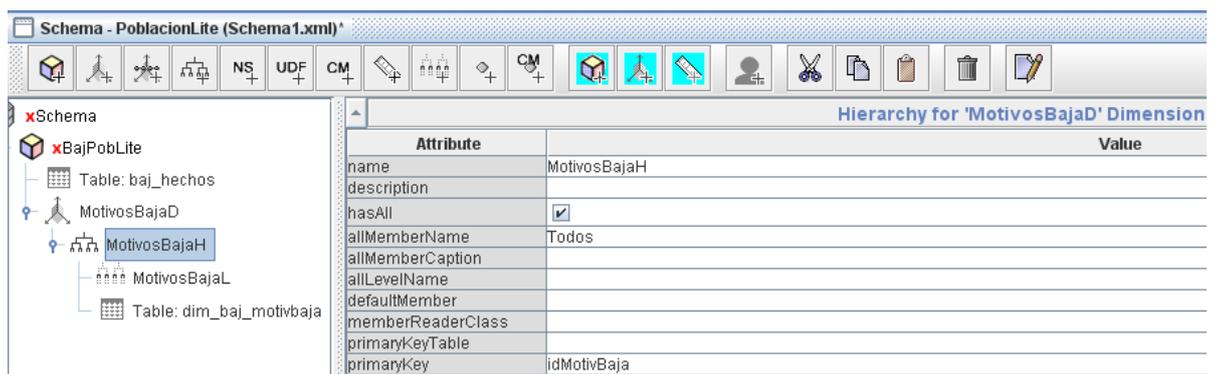


Figura 24: OlapCube_Bajas_Definir Jerarquía en Motivos Baja

- Se añade un nivel a la jerarquía, y se selecciona el campo de la tabla dimensión correspondiente a este nivel. En este caso sólo se añade un nivel, correspondiente la descripción del Motivo de Baja

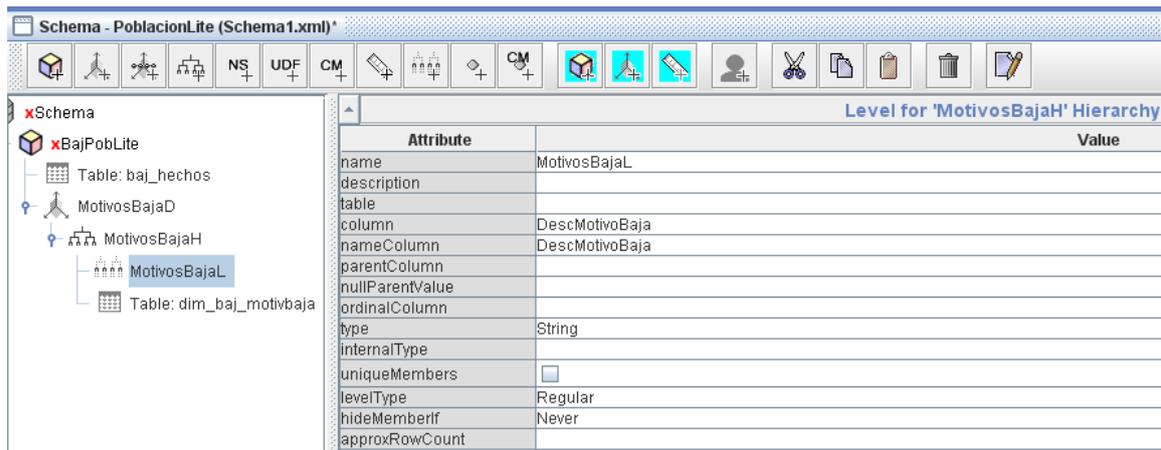


Figura 25: OlapCube_Bajas_Añadir nivel en Motivos Baja

- Después de añadir todas las dimensiones hay que definir la medida

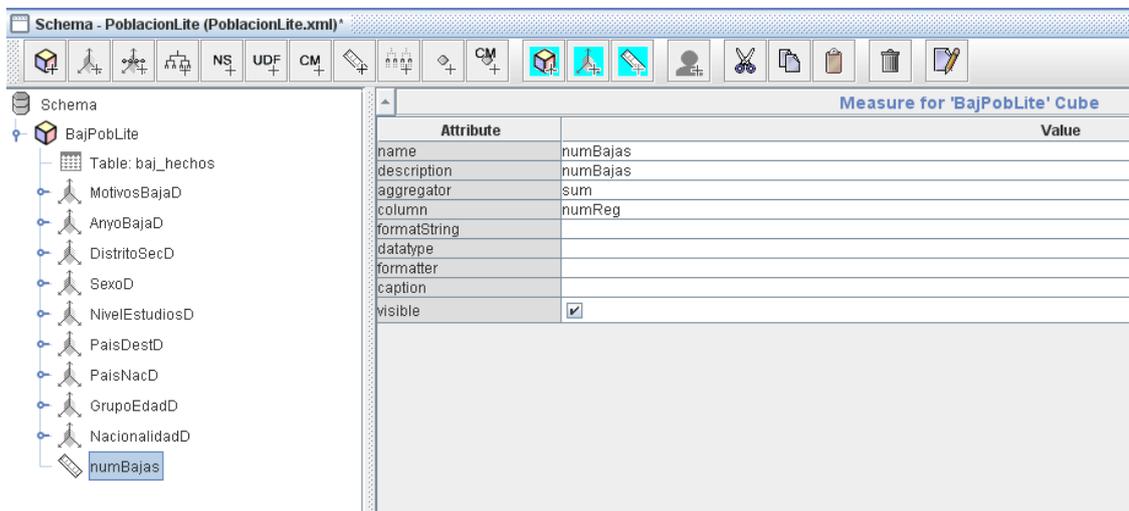


Figura 26: OlapCube_Bajas_Definir la medida

La medida en este cubo de Bajas de Población es en realidad un contador del número de bajas, pero como en el proceso ETL cuando se creó la Tabla de hechos no se grabaron todos los registros individuales de bajas, en el paso Filas Únicas por Claves Dimensión, se agruparon los registros por las claves de dimensiones y se creó el contador numReg que contabilizaba el número de registros de cada fila de la tabla de hechos. Por ello ahora la medida se define una función de suma para ese campo numReg.

El diseño del cubo de Bajas con todas las dimensiones quedaría así:

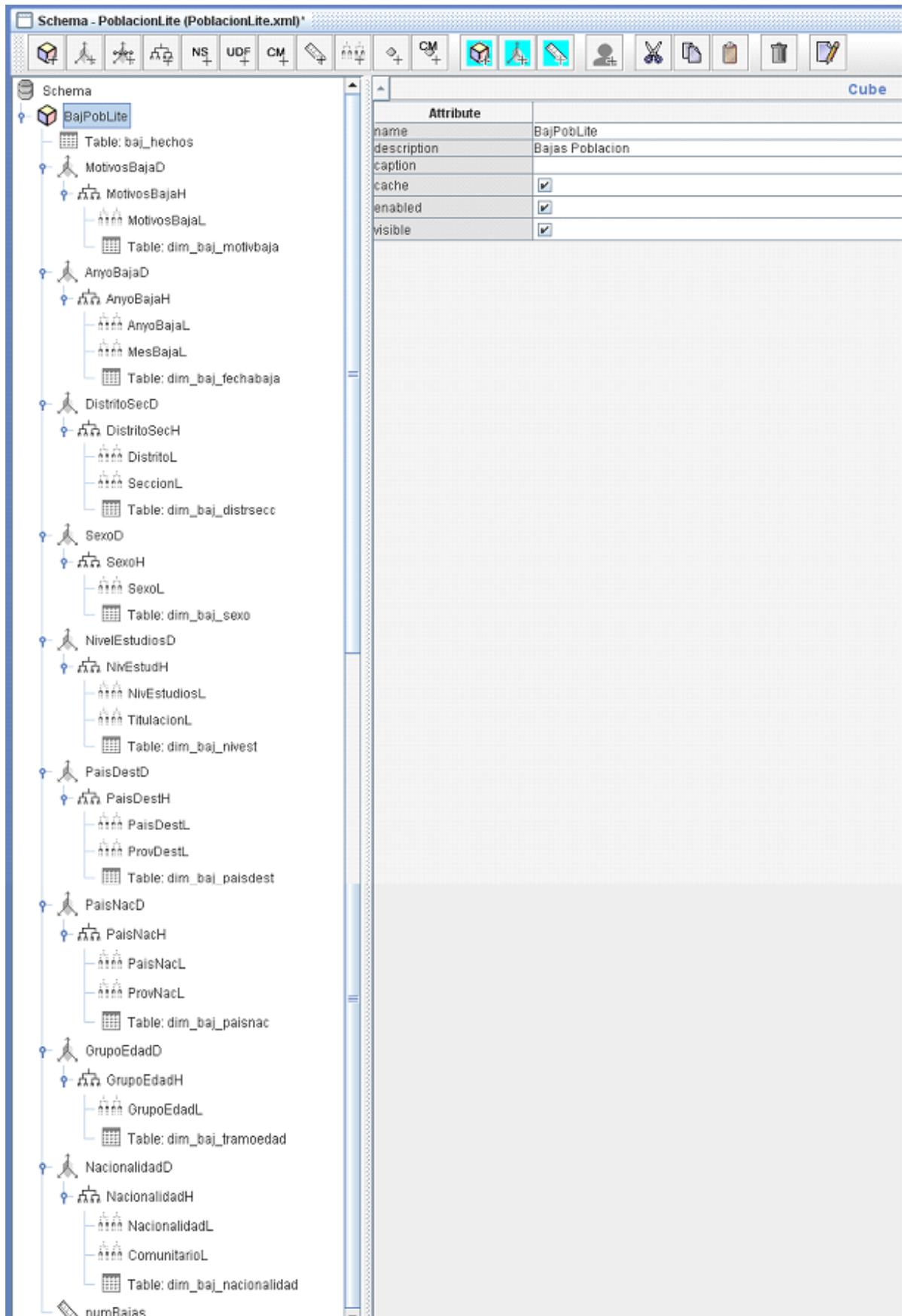


Figura 27: Olap_Cube Bajas de Población

Diseño cubo OLAP para Población Histórica:

Se realiza el mismo procedimiento para diseñar el cubo en Población Histórica

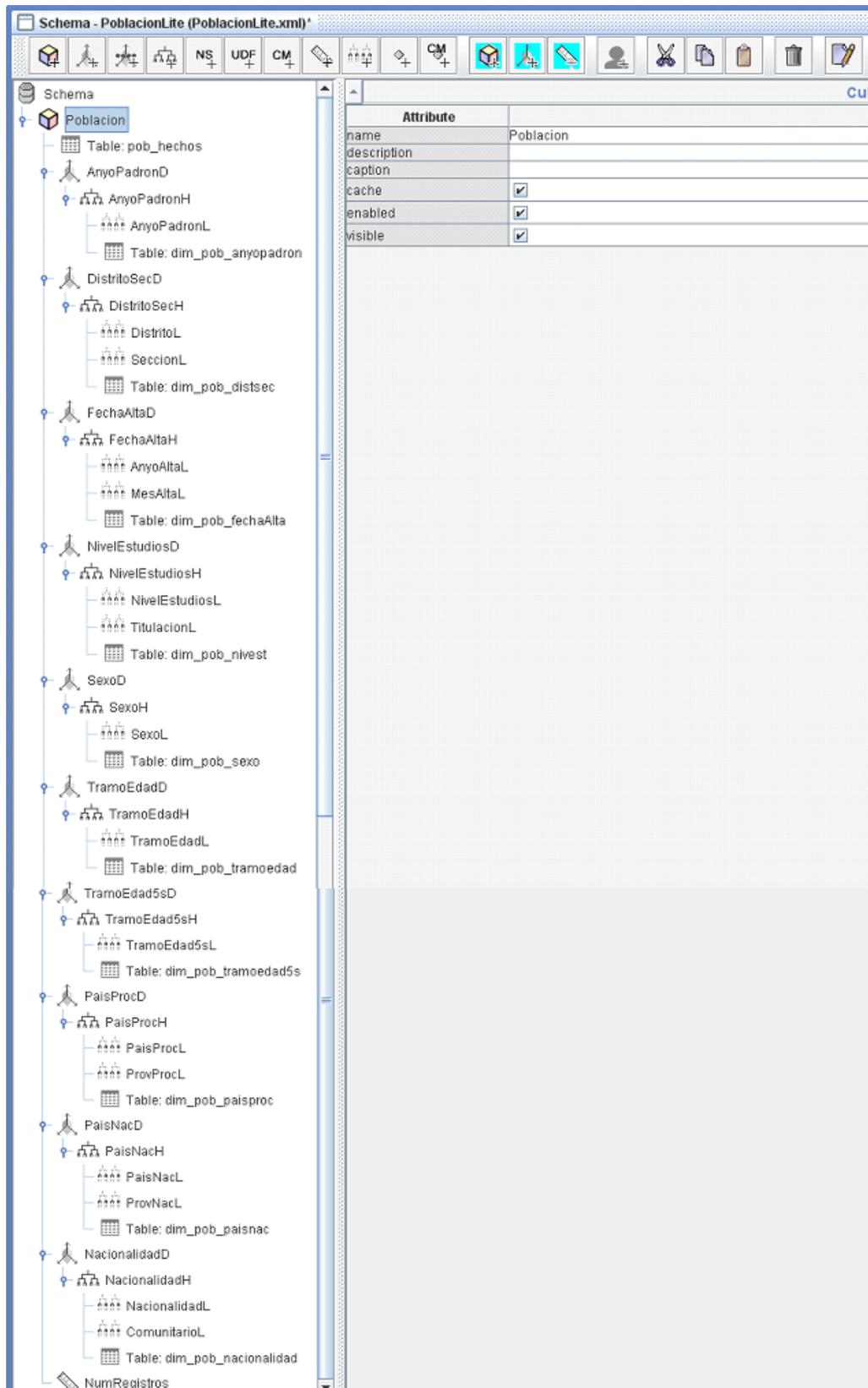


Figura 28: OlapCube_Población

Cubos Bajas y Población en formato XML

El archivo generado desde la herramienta Schema Workbench es un archivo XML correspondiente al Esquema llamado PoblacionLite.xml y contiene los cubos Poblacion y BajPobLite como se han denominado los cubos de Población Histórica y Bajas de Población.

4.4. Diseño y Desarrollo de Cuadros de Mando.

En esta sección se describen los pasos realizados para la creación de los 3 cuadros de mando diseñados para este proyecto: cuadros de mando de Bajas, de Población y de Flujo migratorio. Primero se explica el diseño de mockups y después la realización de los cuadros de mando con las herramientas de Pentaho.

Pero antes de todo, hay que definir que es un cuadro de mando (Dashboard):

- ***Un cuadro de mando es una representación simplificada de un conjunto de indicadores que aportan información de cómo se está comportando un área o un proceso de nuestra empresa.***
- ***Permite una visualización ágil e intuitiva de los datos, mediante el uso de tipo de gráficos, time line, mapas scorecards, alertas, ...***
- ***Es una herramienta de medición del rendimiento***
- ***Contribuye a reducir la incertidumbre***
- ***Es una herramienta de ayuda a la decisión.***
(Stratebi 2017, Cuadros-de-mando-y-scorecards).

4.4.1. Diseño de Mockups.

En este apartado se presentan los mockups de los 3 cuadros de mando, describiendo también el contenido de las pestañas en cada cuadro.

Mockup para Bajas de Población

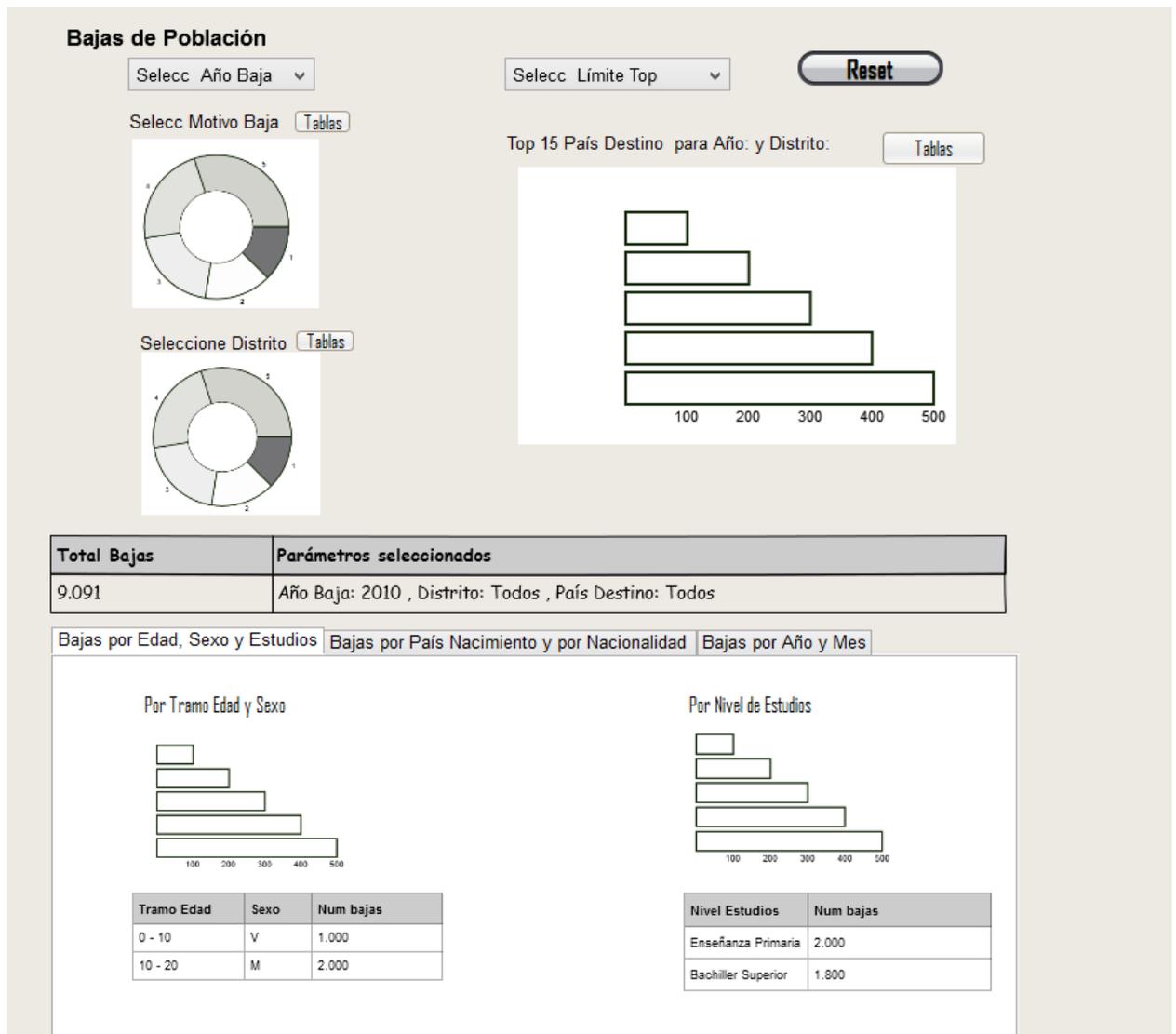


Figura 29: Mockup _Bajas Población _Pantalla inicio

Se distinguen en el mockup 3 partes:

1.-En la parte superior, un grupo de **selección de los siguientes parámetros**:

- Un selector de Año de Baja
- Un selector (Limit Top) que indica cuántos principales o primeros elementos (Top) se visualizan en el gráfico de barras horizontal clasificados por num Bajas. En la figura anterior donde se representa Top 15 País Destino, que se pueda seleccionar Top 20 o Top 25 País Destino.
- Un gráfico tipo "pie" clickable donde se puede seleccionar Motivo Baja.
- Un gráfico tipo "pie" clickable donde se puede seleccionar Distrito
- Un gráfico de barras donde se representan los principales Países de destino resultantes de la selección de parámetros anterior y que también es clickable (se seleccionará país de destino, por ejemplo, cuando el motivo Baja sea un Cambio de Residencia al Extranjero, sin embargo no tendrá sentido en otros motivos de Baja como son Cambio de Residencia dentro de España o Bajas por Defunción).

2.-En la parte central, una tabla que indica los **parámetros seleccionados** y el número total de bajas para esa selección.

3.- En la parte inferior, hay **tres pestañas** con una representación para esta selección en forma de gráficos y tablas, y donde se visualizan varias vistas del cubo de Bajas, por varias dimensiones que se detallan a continuación:

Pestaña “Bajas por Edad, Sexo y Estudios”:

- . una gráfica y tabla correspondiente de Bajas por Tramo Edad y Sexo
- . una gráfica y tabla correspondiente de Bajas por Nivel de Estudios.

Pestaña “Bajas por País de Nacimiento y por Nacionalidad”:

- . una gráfica y tabla correspondiente de Bajas por País de Nacimiento
- . una gráfica y tabla correspondiente de Bajas por Nacionalidad.

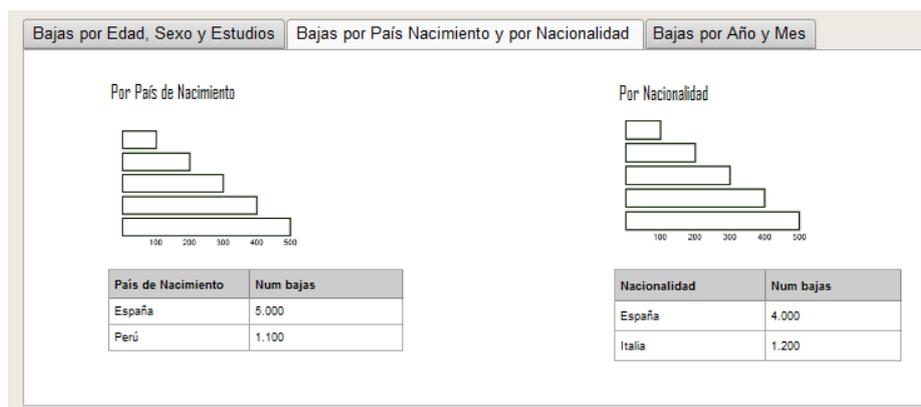


Figura 30: Mockup _Bajas Población _pestaña Bajas por País Nac. y Nacionalidad

Pestaña “Bajas por Año y Mes”:

- . una gráfica y tabla correspondiente de Bajas por Año y Mes



Figura 31: Mockup _Bajas Población _pestaña Bajas por Año y Mes

Mockup para Población Histórica

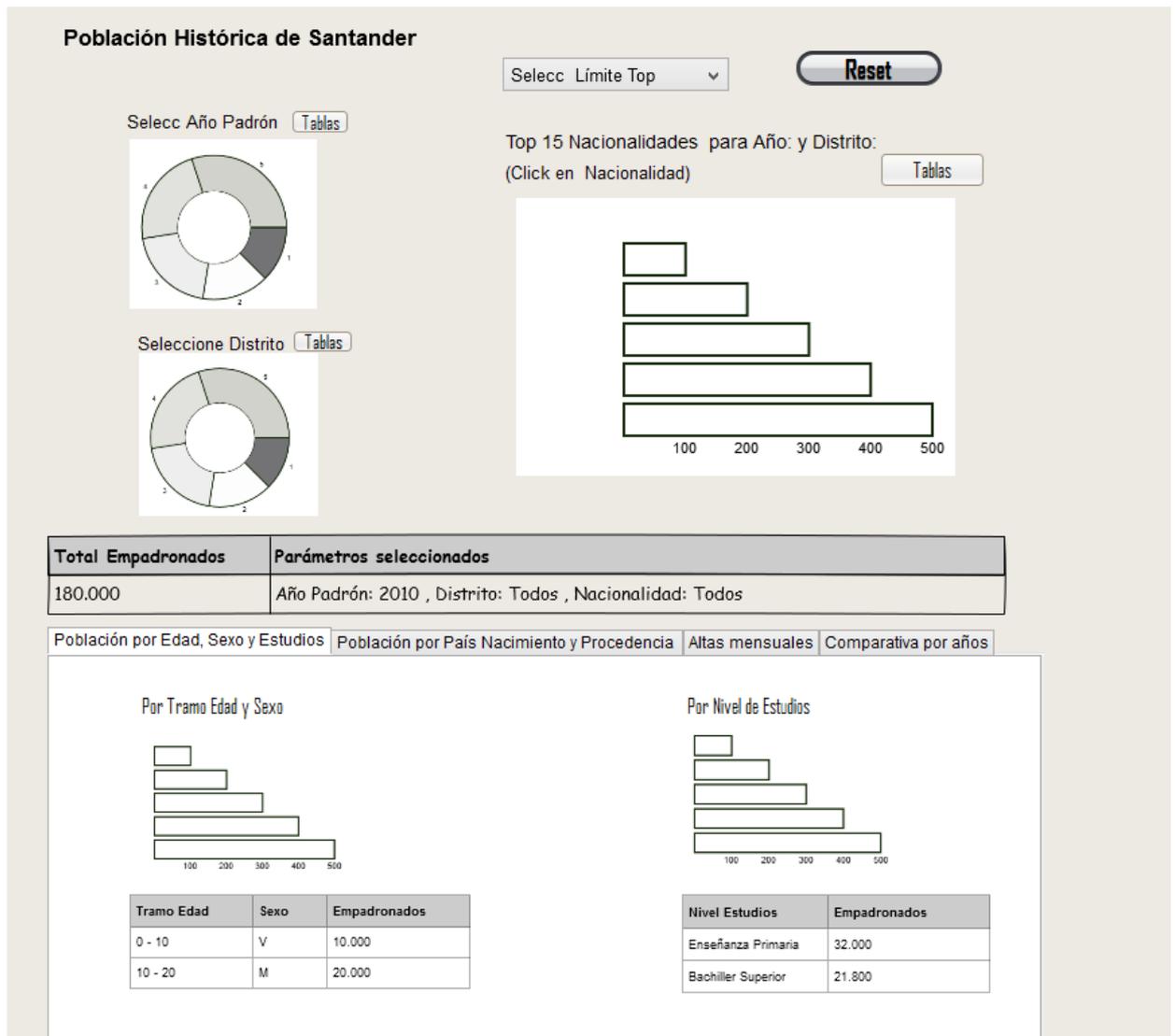


Figura 32: Mockup _Población Histórica _Pantalla inicio

Descripción del mockup:

Es similar al de Bajas, se distinguen en el mockup 3 partes:

1.-En la parte superior, **selección de los siguientes parámetros:**

- Un selector (**Limit Top**) que indica cuántos principales o primeros elementos (Top) se visualizan en el gráfico de barras horizontal clasificados por num Empadronados. En el caso de Población Histórica poder seleccionar Top 15, 20 o 25 Nacionalidades.
- Un gráfico tipo “pie” clickable donde se puede seleccionar **Año Padrón**.
- Un gráfico tipo “pie” clickable donde se puede seleccionar **Distrito**
- Un gráfico de barras donde se representan las principales Nacionalidades resultantes de la selección de parámetros anterior y que también es clickable (se seleccionará también Nacionalidad como parámetro).

2.-En la parte central, una tabla que indica los **parámetros seleccionados** y el número total de Empadronados para esa selección.

3.- En la parte inferior, hay **tres pestañas** con una representación para esta selección en forma de **gráficos y tablas**, y donde se visualizan varias vistas del cubo de Población Histórica, por varias dimensiones que se detallan a continuación:

Pestaña “Población por Edad, Sexo y Estudios”:

- . una gráfica y tabla correspondiente de Empadronados por Tramo Edad y Sexo
- . una gráfica y tabla correspondiente de Empadronados por Nivel de Estudios.

Pestaña “Población por País de Nacimiento y Procedencia”:

- . una gráfica y tabla correspondiente de Empadronados por País de Nacimiento
- . una gráfica y tabla correspondiente de Empadronados por País Procedencia

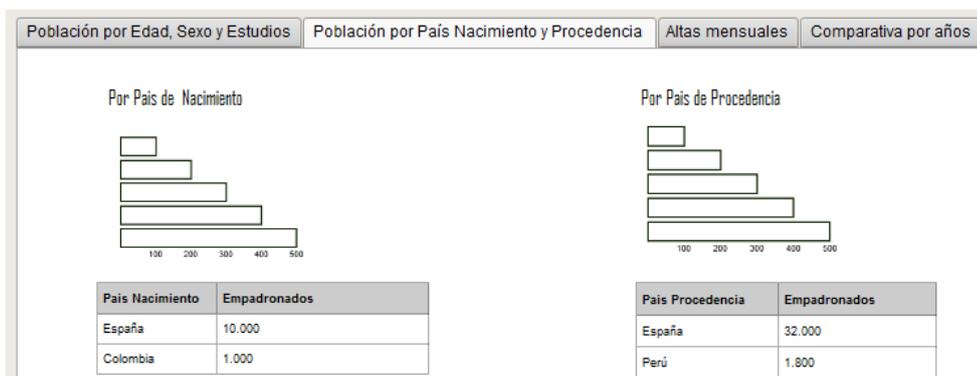


Figura 33: Mockup _ Población Histórica _pestaña Por País Nac. y Nacionalidad

Pestaña “Altas Mensuales”:

- . una gráfica y tabla correspondiente de Altas por Año y Mes

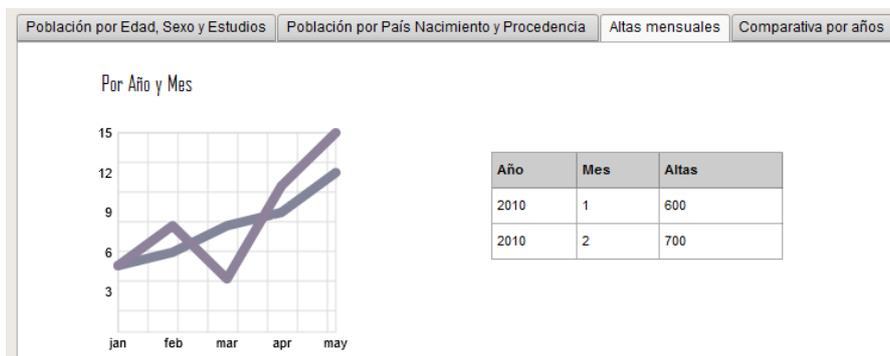


Figura 34: Mockup _ Población Histórica _pestaña Altas por Año y Mes

Pestaña “Comparativa por Años”:

- . Se hace una comparativa de todos los años de padrón por Tramos de Edad y por Nivel de Estudios (tablas y gráficos).

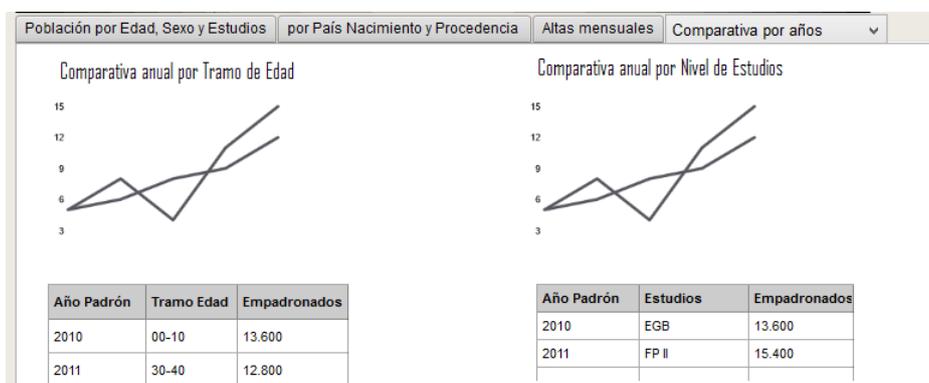


Figura 35. Mockup _ Población Histórica _pestaña Comparativa por Años

Mockup para Flujo Migratorio.

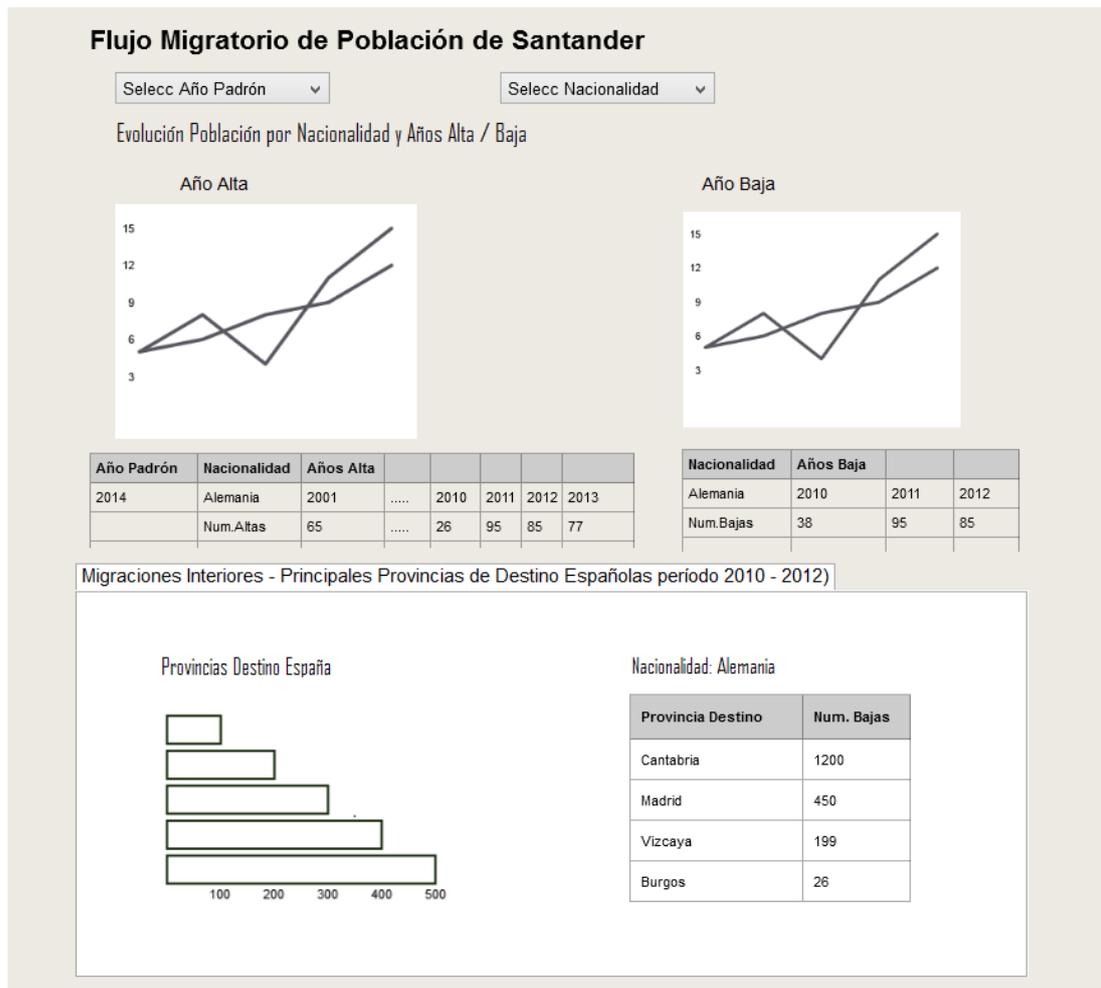


Figura 36: Mockup _ Flujo Migratorio

Se distinguen en el mockup Flujo Migratorio 2 partes:

1.-En la parte superior, selección de los siguientes parámetros:

- Un selector de Año Padrón.
- Un selector de Nacionalidad.

Trata de reflejar las Migraciones Exteriores, representa la evolución de la población para una determinada Nacionalidad por Año de Llegada a la ciudad y por Año de Baja para las bajas. Se escoge la Nacionalidad como parámetro al no poder utilizar el País de Procedencia (muchos registros no tienen este dato, tienen valor null) y que habría sido más apropiado para el tema de migraciones.

Una vez seleccionados el Año Padrón y la Nacionalidad, se representan en gráfico y tablas el número de empadronados para esa nacionalidad según su año de llegada a la ciudad y por otra parte se representan las bajas indicando el número de bajas por año de baja.

2.-En la parte inferior, se representan las principales Provincias de Destino Españolas en los Cambios de Residencia dentro de España, durante el período 2010-2012, y por la Nacionalidad que se haya seleccionado en el parámetro.

Trata de reflejar las Migraciones Interiores, pero sólo en el sentido del destino (emigración interior), porque en la procedencia (inmigración interior) no se puede realizar por la falta de información en los campos de País y Provincia de Procedencia.

4.4.2. Desarrollo de Cuadros de Mando.

En este apartado se describen primero las herramientas utilizadas en Pentaho para la creación de Dashboards y después se detallan los principales componentes en la creación de los 3 cuadros de mando (Bajas de Población, Población Histórica y Flujo Migratorio).

Creación Dashboards en Pentaho.

Pentaho CTools es un conjunto de herramientas para crear y mantener dashboards en la versión Community de Pentaho. La plataforma Pentaho le permite utilizar componentes externos, o plugins, con el propósito de extender las funciones estándar.

Incluye:

1 . CDF: es el framework Open Source para la creación de cuadros altamente personalizables en el front end del servidor Pentaho Business Intelligence.

Características:

- *Basado en tecnología open source.*
- *Se basa en los estándares de desarrollo web como CSS, HTML5, Javascript, jQuery, Bootstrap*
- *Separa el diseño HTML de la definición del componente .*
- *Permite extensibilidad*

2. CCC: es el framework de librerías graficas open source, que se encuentra incluido dentro de las CTools y permite crear en los cuadros de mando componentes gráficos avanzados que permiten la interacción del usuario y con gran capacidad de personalización mediante las Extension Points. Es una biblioteca de visualización construida sobre Protovis.

3. CDA:

El acceso a datos de la comunidad (CDA) permite recopilar y combinar datos de varias fuentes de datos en una única estructura, que se puede utilizar en los dashboards. Creado por la necesidad de unificar el acceso a la capa de datos Pentaho, CDA fue desarrollado para crear una capa de abstracción entre las CTools dashboard y las conexiones físicas a diferentes bases de datos.

4. CDE:

El Community Dashboard Editor (CDE) y su tecnología subyacente (CDF, CDA y CCC) permiten el desarrollo y despliegue rápido y eficaz de Cuadros de Mando sobre Pentaho.

La herramienta CDE fue creada para simplificar los procesos de creación y diseño de los Dashboards, es una herramienta potente y completa, que integra perfectamente la interfaz de usuario con fuentes de datos y componentes personalizados.

(Pentaho Group 2016, - Documentation 6.1 CTools_Overview).

4.4.2.1. Creación Cuadro de Mando Bajas de Población.

En este apartado se presenta en la figura siguiente una captura de pantalla del cuadro de Bajas de Población desarrollado, describiendo los principales componentes de la herramienta CDE de Pentaho que se han utilizado, así como algunas consultas a la base de datos en formato MDX .

Bajas Población

Selecione Período

año_2_010 año_2_012

Limit 15 Reset

Tablas

Selecione Motivo Baja (click en gráfico)

- Baja Cambio Residencia España
- Baja por Defunción
- Baja por caducidad
- Baja Cambio de Residencia Extranj
- Baja por Inclusión Indevida
- Baja por Duplicado

41	(0.16%)
744	(2.85%)
957	(3.66%)
1,921	(7.35%)
5,582	(21.38%)
16,894	(64.63%)

Tablas

Top 15 PaisDestino --> Todos --> Distr: Todos

Selecione País de Destino (click en gráfico)

ESPAÑA	16,894
REINO UNIDO	1,921
FRANCIA	1,921
PERU	1,921
ESTADOS UNIDOS	1,921
ALEMANIA	1,921
COLOMBIA	1,921
ECUADOR	1,921
BRASIL	1,921
PARAGUAY	1,921
MOLDAVIA	1,921
ARGENTINA	1,921
CHILE	1,921
MEXICO	1,921
RUMANIA	1,921
ITALIA	1,921

Total Bajas

26.139

Parámetros Seleccionados

Años--> 2010 / 2012 Motivo--> Todos Distrito--> Todos País_Destino--> Todos

Bajas por Edad, Sexo y Estudios

Bajas por Nacionalidad y País de Nacimiento

Bajas por Año y Mes

Por Tramo-Edad y Sexo

Por nivel de Estudios

Tramo Edad y Sexo

Edad	Sexo	Num bajas
0-10	Mujer	1,025
0-10	Varon	1,089
10-20	Mujer	715
10-20	Varon	645
20-30	Mujer	2,724
20-30	Varon	2,476
30-40	Mujer	2,728
30-40	Varon	3,073
40-50	Mujer	1,270
40-50	Varon	1,595

Showing 1 to 10 of 20 entries

Nivel de Estudios

Nivel de Estudios	Num bajas
Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO	8,613
TITULACION INFERIOR AL GRADO DE ESCOLARIDAD	4,902
Bachiller superior. BUP. Bachiller LOGSE	3,105
Licenciado universitario, Arquitecto o Ingeniero Superior. Grado universitario	2,111
No aplicable por ser menor de 16 años	1,471
Diplomado de escuelas universitarias (Empresariales, Profesorado de EGB, ATS y similares)	1,091
Formación profesional primer grado. Formación profesional de Grado Medio. Oficialia industrial	1,025

Showing 1 to 7 of 18 entries

Figura 37: Cuadro_Mando_Bajas de Población.

Pasos que hay que dar para la creación de un cuadro de mando utilizando la herramienta CDE:

- Crear un Data source , importando el análisis en formato xml del cubo asociado a la conexión anterior.
- Empezar a definir layouts, componentes y data sources.
 - **El Layout** , indica donde se ubicarán los diferentes componentes visuales: gráficos, tablas, mapas ...
 - Es la presentación de la página.
 - Se pueden añadir recursos (archivos Css, Javascript), archivos Html.
 - **Los Componentes**, se definen los componentes específicos (gráficos, tablas, parámetros, mapas..). Conectarlos a los datos y ubicarlos en el layout...
 - Especifica su ubicación en el layout.
 - Especifica su conexión y consulta a los datos.
 - **Los DataSource (fuentes de datos)**, muestran las consultas a la base de datos (sql's, query's Mdx...).

Ejemplo: Seleccionar Motivo Baja

Se describen a continuación las 3 secciones del CDE (editor de Dashboard de Pentaho) para el Parámetro Seleccionar Motivo Baja:

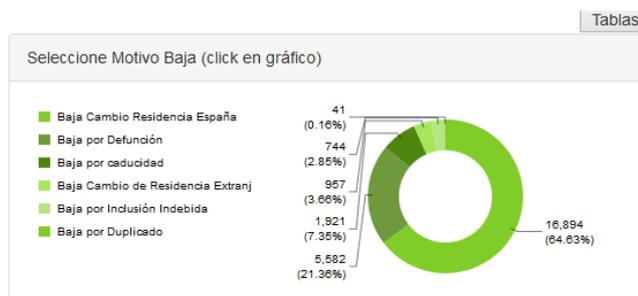


Figura 38: CM _Bajas_ Seleccionar Motivo Baja

- **El Layout,**

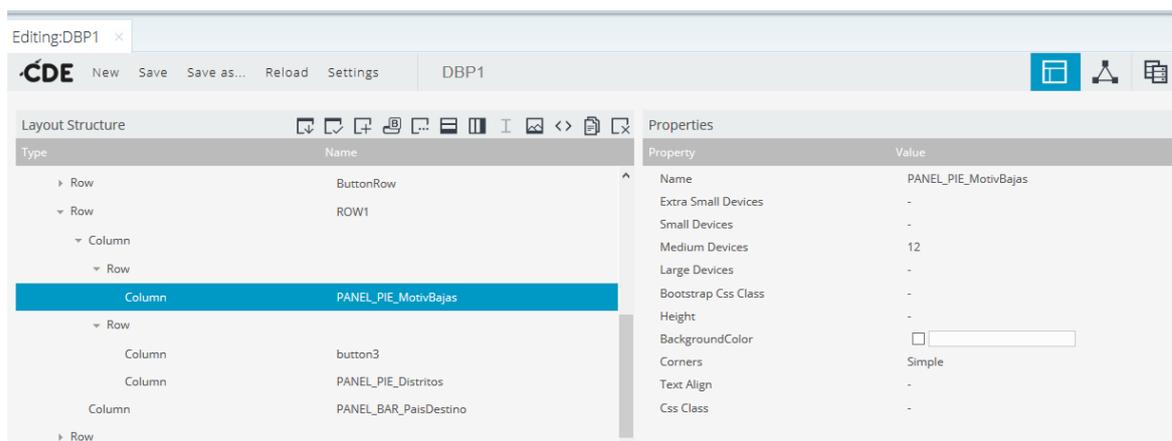


Figura 39: CM _Bajas_Layout_ Seleccionar Motivo Baja

Se le asigna una posición (Columna PANEL_PIE_MotivBajas).

- **Los Componentes,**

Se utilizan 2 componentes:

- Un componente Bootstrap Panel de Ivy DC, se le da un estilo Bootstrap , Header text, y en el HtmlObject se le indica el nombre que se le asignó en el Layout.

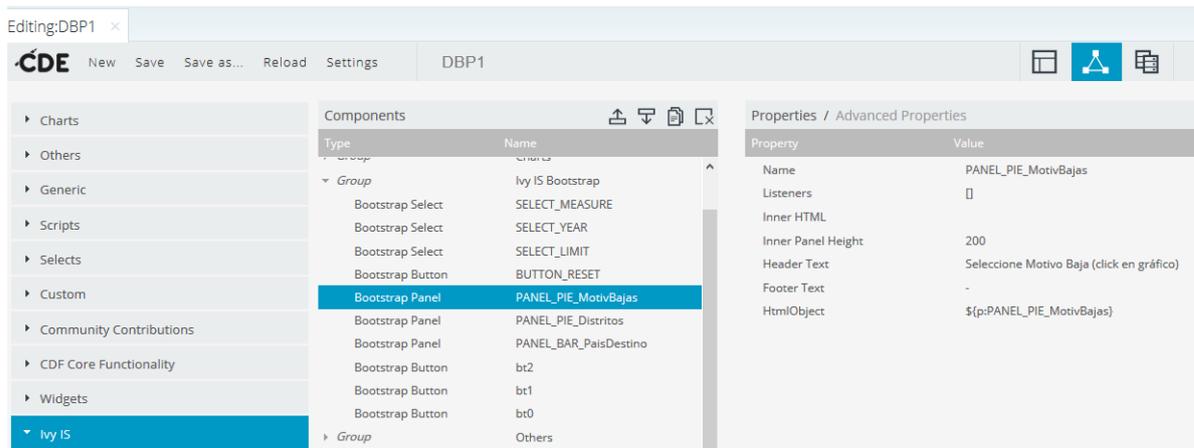


Figura 40: CM _Bajas_ Componente Bootstrap_Panel_ Seleccionar Motivo Baja

- El componente gráfico PIE_MotivBajas es del tipo Pie Chart, se le asocia el Datasource: DS_BY_MotivBajas, y los parámetros que utiliza: IR_YEAR, START_YEAR, END_YEAR, PARAM_DISTR.

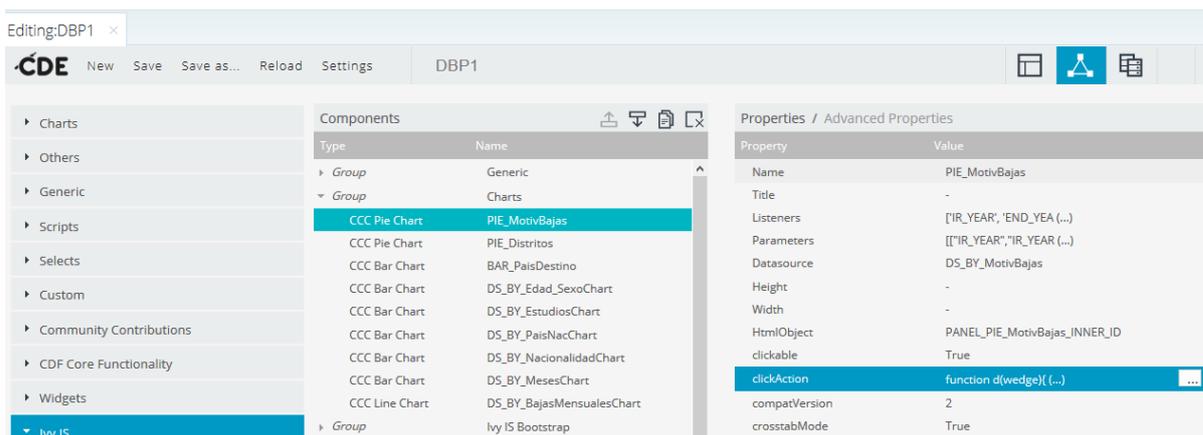


Figura 41: CM _Bajas_ Componente Pie_Chart_ Seleccionar Motivo Baja

Como este Chart tiene una función Select, se le pone clickable y en la propiedad clickAction se le define el siguiente script:

```
function d(wedge){
    Dashboards.fireChange("PARAM_MOTIV",wedge.vars.category.value); }
```

Figura 42: código_Script_clikAction_CM _Bajas_ Pie_Chart

En propiedades avanzadas, para cambiar como aparecen las etiquetas por defecto de las dimensiones en los tooltip, se ejecuta esta función en post fetch:

```
function f(data) {
    this.chartDefinition.dimensions = {
        series: {label: "series", isHidden:true},
        category: {label: "Motivo"},
        value: {label: "numBajas" }};
    return data; }
```

Figura 43: código_Script_modif_Tooltip_CM _Bajas_ Pie_Chart

- **El Data Source,**

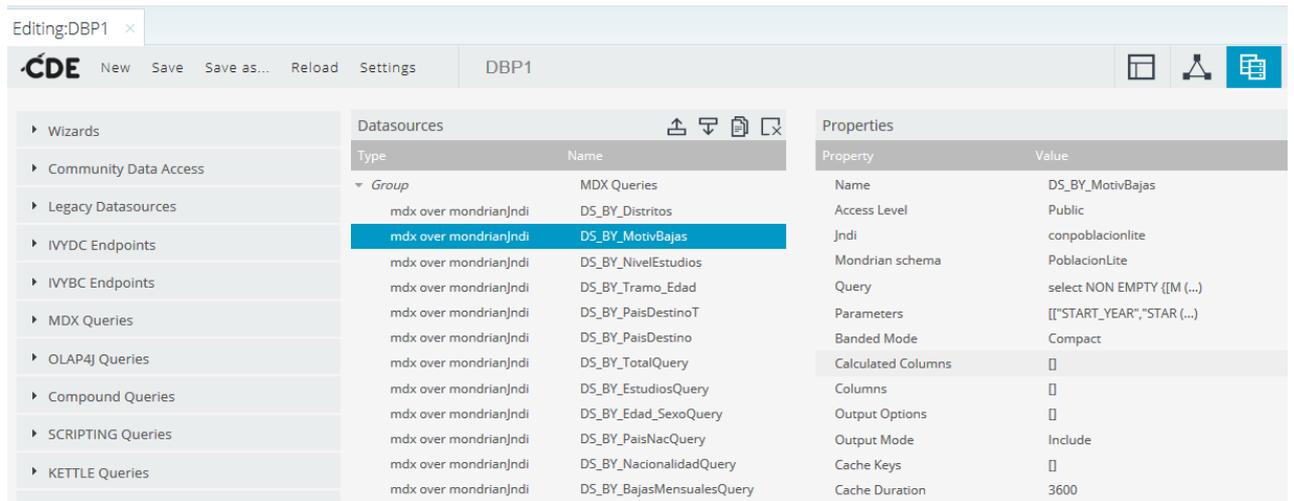


Figura 44: CM_Bajas_Datasource _ Seleccionar Motivo Baja

Como se observa en la figura están definidos el esquema Mondrian (PoblacionLite) y el Jndi o conexión a la Base de datos, como parámetros los mismos que se definieron en el componente y el Query es una consulta MDX:

```
select NON EMPTY {[Measures].[numBajas]} ON COLUMNS,
NON EMPTY Order([MotivosBajaD.MotivosBajaH].Children, [Measures].[numBajas], DESC) ON ROWS
from [BajPobLite]
where Crossjoin({[AnyoBajaD.AnyoBajaH].[${START_YEAR}]:[AnyoBajaD.AnyoBajaH].[${END_YEAR}]},
{[DistritoSecD.DistritoSecH].[${PARAM_DISTR}]})
```

Figura 45: código_ QueryMdx_CM_Bajas_ Seleccionar Motivo Baja

Otros Componentes utilizados: Componentes Popup y Button.

En la parte superior del Cuadro de Mando, donde están los gráficos de Motivos Bajas, Distrito y País Destino , hay encima de los gráficos unos botones “tablas”, que abren un popup para representar en forma de tabla los valores del gráfico, como se indica en la figura adjunta.

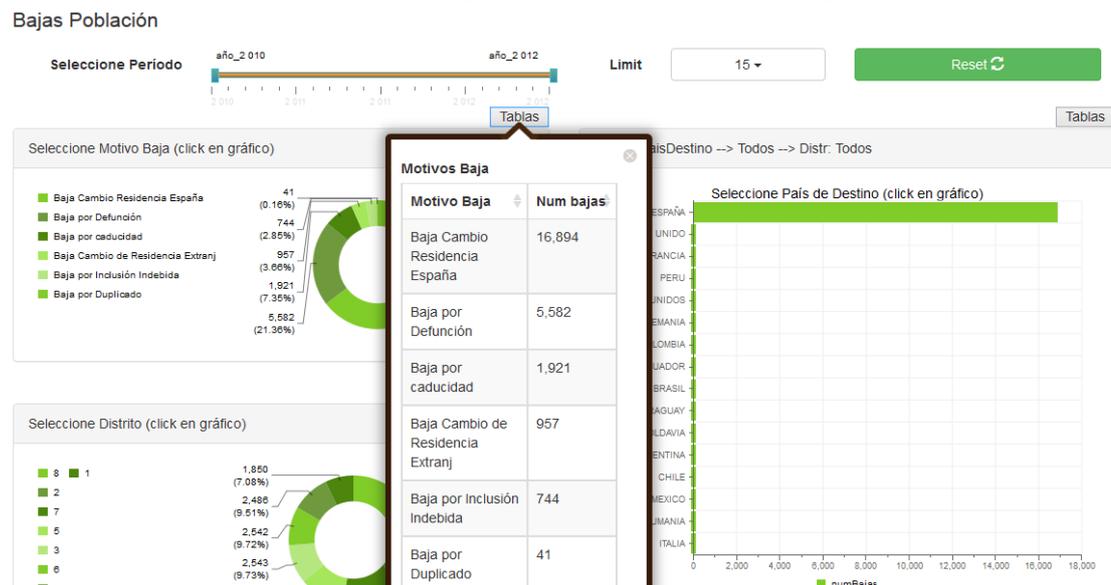


Figura 46: CM_Bajas_Popup Seleccionar Motivo Baja

4.4.2.2. Creación Cuadro de Mando de Población Histórica

En este apartado se presenta en la figura siguiente una captura de pantalla del cuadro de Población Histórica desarrollado, describiendo los principales componentes de la herramienta CDE de Pentaho que se han utilizado, así como algunas consultas a la base de datos en formato MDX .

Población Histórica de Santander

Limit 15 Reset

Tablas

Seleccione Año Padrón (click en gráfico)

Seleccione Distrito (click en gráfico)

Tablas

Top 15 Nacionalidades para año padrón --> 2010 --> Distr: Todos

(Click en Nacionalidad)

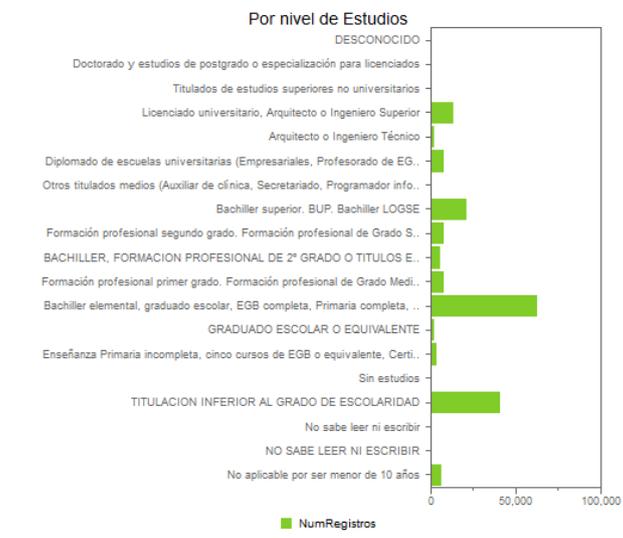
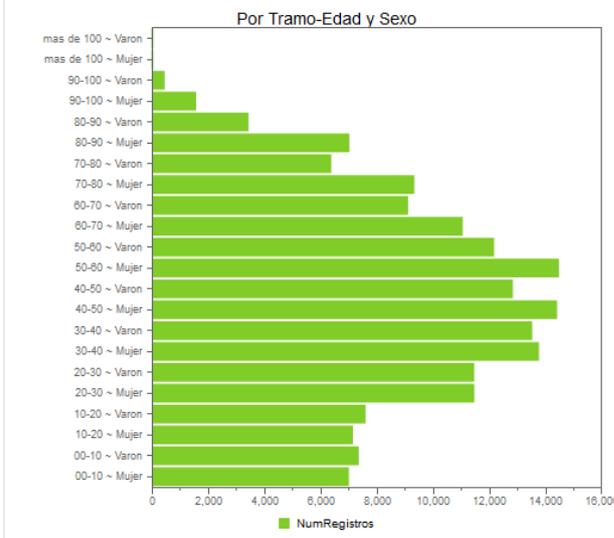
Total Empadronados

181,441

Parámetros Seleccionados

Año Padrón--> 2010 Distrito--> Todos Nacionalidad--> Todos

[Poblacion por Edad, Sexo y Estudios](#)
 [Poblacion por País de Nacimiento y Procedencia](#)
 [Altas Mensuales](#)
 [Comparativas por años: ▾](#)



Tramo Edad y Sexo

Edad	Sexo	Num bajas
00-10	Mujer	6,993
00-10	Varon	7,345
10-20	Mujer	7,137
10-20	Varon	7,588
20-30	Mujer	11,458

Nivel de Estudios

Nivel de Estudios	Empadronados
Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO	62,345
TITULACION INFERIOR AL GRADO DE ESCOLARIDAD	40,735
Bachiller superior. BUP. Bachiller LOGSE	20,623

Figura 47: Cuadro_Mando Población Histórica Año Padrón 2010

El Cuadro de Mando de Población es muy similar al de Bajas, se explican a continuación las siguientes diferencias:

- No se selecciona un período, un intervalo de años sino un año de padrón determinado por la razón de que Población, la tabla de hechos es del tipo Snapshot , representa una “foto” de la población a una fecha determinada, en este caso, a 1 de enero de cada año. Por defecto , la pantalla inicio corresponde al año 2010.
- Se selecciona Distrito y Límite Top como en las Bajas.
- Donde en las Bajas se seleccionaba País Destino, en Población se va a seleccionar Nacionalidad, atributo que poseen todos los registros. No se podía seleccionar País Procedencia porque como se verá en esa pestaña el valor es mayoritariamente null.
- Presenta una pestaña adicional “Comparativas por años” que se explicará más adelante.

Los parámetros que utiliza son :

- Año Padrón
- Límite Top (qué nº elementos se visualiza en los gráficos)
- Distrito
- Nacionalidad

Todos estos parámetros se utilizan en las pestañas de la parte inferior, por ejemplo, el query MDX de la consulta del gráfico por Tramo Edad y Sexo sería:

```
select NON EMPTY({Descendants([TramoEdadD.TramoEdadH].[Todos] ,[TramoEdadD.TramoEdadH].[TramoEdadL])
* Descendants([SexoD.SexoH].[Todos] ,[SexoD.SexoH].[SexoL])}) on ROWS,
NON EMPTY({[Measures].[NumRegistros]}) on Columns
from [Poblacion]
where Crossjoin(Crossjoin({[AnyoPadronD.AnyoPadronH].[${PARAM_YEAR}],
{[DistritoSecD.DistritoSecH].[${PARAM_DISTR}]}},
{[NacionalidadD.NacionalidadH].[${PARAM_NACION}]})
```

Figura 48: código_ QueryMdx_CM _Población_ Pestaña por Tramo Edad y Sexo.

En el caso de Población por País de Nacimiento, sería:

```
WITH
SET [~ROWS] AS
    Order(TopCount({[PaisNacD.PaisNacH].[PaisNacL].Members}, ${PARAM_LIMIT}, [Measures].[NumRegistros]),
[Measures].[NumRegistros], ASC)
SELECT
NON EMPTY {[Measures].[NumRegistros]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Poblacion]
where Crossjoin(Crossjoin({[AnyoPadronD.AnyoPadronH].[${PARAM_YEAR}],
{[DistritoSecD.DistritoSecH].[${PARAM_DISTR}]}},
{[NacionalidadD.NacionalidadH].[${PARAM_NACION}]})
```

Figura 49: código_ QueryMdx_CM _Población_ Pestaña por Provincia de Nacimiento.

Exportar a Excel tablas:

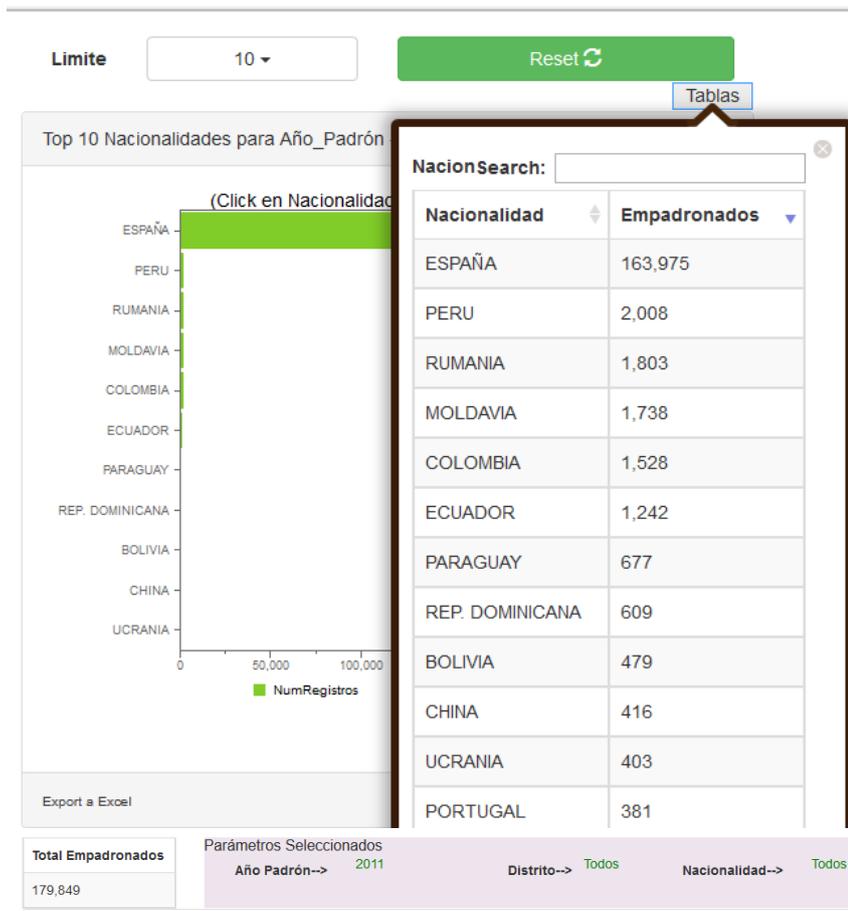


Figura 50: CM_Población año2011_Vista Tabla Nacionalidades

Como se observa en la figura anterior, aunque el gráfico muestre un Top 10 Nacionalidades, en la vista Tablas se relacionan todas y también se pueden exportar a Excel mediante un componente Export Popup Component (se observa en la figura en la parte izda de la tabla un button "Export a Excel").

Pestaña Altas Mensuales



Figura 51: CM_Población_Pestaña Altas Mensuales

Como ya se había dicho anteriormente, es una representación aproximada de altas porque no están incluidas las altas y bajas durante el mismo año, en la figura, si el año Padrón es 2011, se relacionan las altas durante el año 2010 ya que el Padrón es a fecha 01 de enero de 2011. Para ello, a partir del parámetro año Padrón hay que generar un nuevo parámetro que es el año anterior, se hace mediante este script en pre_Execution del Chart:

```
function f(){
    Dashboards.fireChange('PARAM_PREVIOUS_YEAR',PARAM_YEAR - 1);
    return PARAM_PREVIOUS_YEAR; }
```

Y el query de la consulta MDX sería:

```
WITH
SET [~COLUMNS] AS
    {[AnyoPadronD.AnyoPadronH].[${PARAM_YEAR}]}
SET [~FechaAltaD_FechaAltaD.FechaAltaH_AnyoAltaL] AS
    {[FechaAltaD.FechaAltaH].[${PARAM_PREVIOUS_YEAR}]}
SET [~FechaAltaD_FechaAltaD.FechaAltaH_MesAltaL] AS
    Exists({[FechaAltaD.FechaAltaH].[MesAltaL].Members}, [~FechaAltaD_FechaAltaD.FechaAltaH_AnyoAltaL])
SET [~ROWS] AS
    Hierarchize({[~FechaAltaD_FechaAltaD.FechaAltaH_MesAltaL]})
SELECT
NON EMPTY [~COLUMNS] ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Poblacion]

where Crossjoin({[DistritoSecD.DistritoSecH].[${PARAM_DISTR}]},
{[NacionalidadD.NacionalidadH].[${PARAM_PAISDEST}]})
```

Figura 52: código_CM_Población_queryMdx_Altas Mensuales.

4.4.2.3. Creación Cuadro de Mando de Flujo Migratorio

Al igual que en los dos apartados anteriores, la figura siguiente muestra una captura de pantalla del cuadro Flujo Migratorio desarrollado, describiendo los principales componentes de la herramienta CDE de Pentaho que se han utilizado, así como algunas consultas a la base de datos en formato MDX .

Flujo Migratorio de Población del Ayuntamiento de Santander

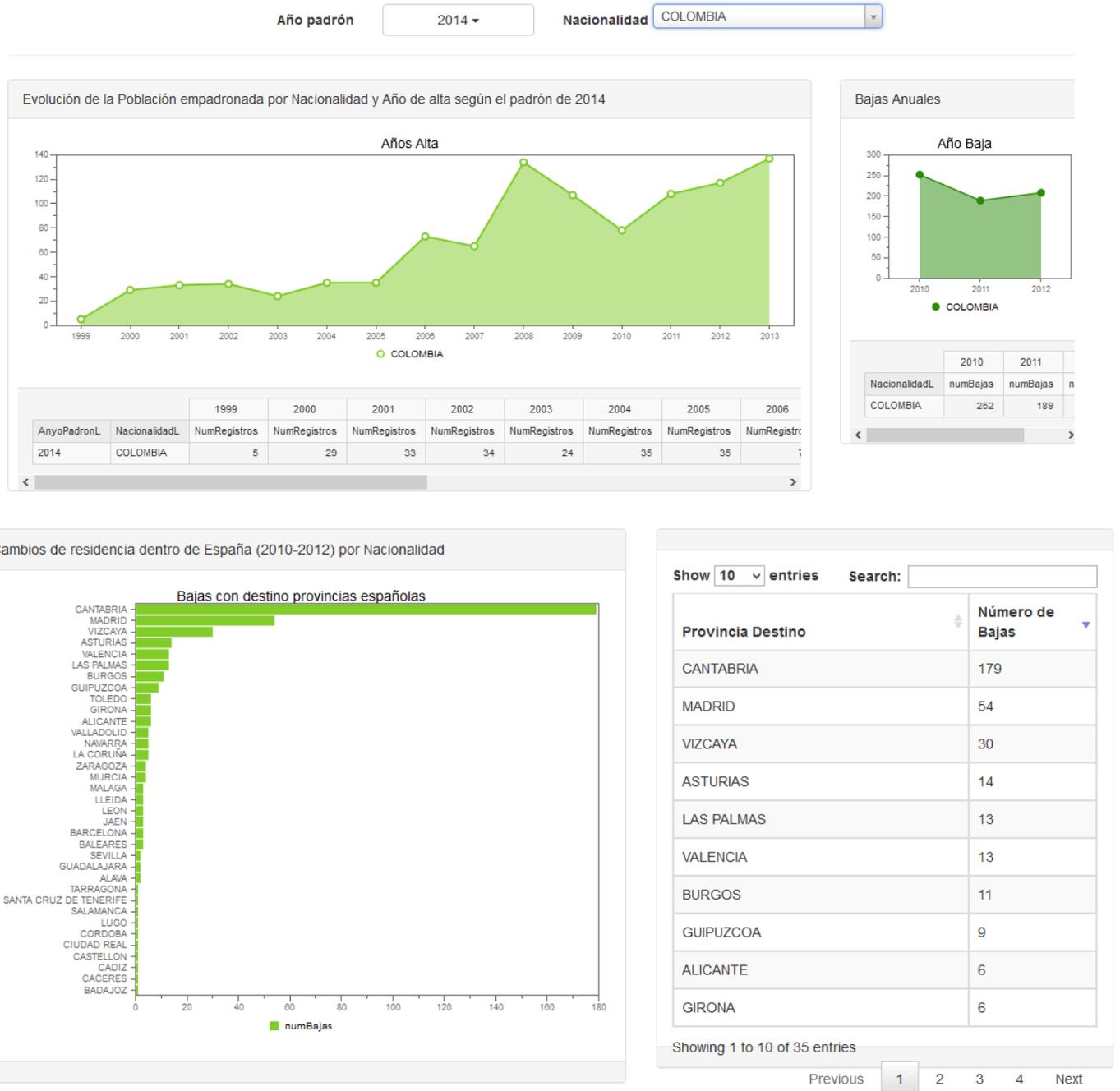


Figura 53: Cuadro Mando Flujo Migratorio, Año: 2014, Nacionalidad: Colombia

Características del Cuadro de Mando de Flujo Migratorio

Parámetros que utiliza:

- Año Padrón
- Nacionalidad.

Al no poder utilizar el País_Provincia de Procedencia por tener valor null en muchos de sus registros, se sustituye por la dimensión Nacionalidad.

Parte superior del Cuadro de Mando:

En la parte izquierda se representa mediante una gráfica una evolución de la población empadronada según el año de alta y para esa nacionalidad seleccionada, todo ello referido a un año Padrón determinado.

Concretamente, en cada Padrón se refleja la población de una nacionalidad, desglosada por el año de alta o de llegada a Santander, (año de alta a partir del año 1997). El padrón se informatiza en el año 1996, y a la población existente se le asigna una fecha genérica de alta: 1996-05-01. En el padrón del año 2010, hay casi 50.000 registros con esa fecha de alta.

Por ello, se excluye año 1996 en el siguiente query MDX de esta gráfica:

```
WITH
SET [~FILTER] AS
    {[AnyoPadronD.AnyoPadronH].[${PARAM_YEAR}]}
SET [~COLUMNS] AS
    {[${PARAM_NACION}]}
SET [~ROWS] AS
    Except({[FechaAltaD.FechaAltaH].[AnyoAltaL].Members}, {[FechaAltaD.FechaAltaH].[#null],
[FechaAltaD.FechaAltaH].[1996]})
SELECT
NON EMPTY [~COLUMNS] ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [Poblacion]
WHERE [~FILTER]
```

Figura 54: Código_queryMdx_AñosAlta_CM_Flujo Migratorio.

En la parte derecha se representa mediante una gráfica las bajas para esa nacionalidad en el período 2010-2012.

Tanto para las altas como las bajas, además de los gráficos se visualizan los datos en unas tablas obtenidas mediante widgets saiku, que son otros componentes del editor de cuadros de mando de Pentaho.

Parte Inferior del Cuadro de Mando:

Trata de ver en la **Emigración interior**, las principales provincias españolas como destino, seleccionando una determinada nacionalidad.

Para ello, se selecciona el motivo de baja: “Cambios de residencia dentro de España”, y se representan las principales 25 provincias españolas destino en un gráfico de barras y para la nacionalidad seleccionada.

En la vista de la tabla se representan todas las provincias españolas (no sólo el Top 25).

4.5. Pruebas de los Cuadros de Mando

En esta sección se incluyen una serie de pruebas con los 3 cuadros de mando definidos anteriormente y se hará un análisis de la información consultada.

- En el cuadro de mando de Bajas de Población:
 - o Motivos de Baja: por Defunción, por Caducidad, Cambio de Residencia en España.
 - o Motivo de Baja: Cambio de Residencia al Extranjero, se seleccionarán los países de destino: Reino Unido, Estados Unidos y Perú.
- En el cuadro de mando de Población Histórica:
 - o Se seleccionarán diferentes años de Padrón, Distritos y en Nacionalidad se seleccionará Rumania e Italia.
 - o Se probarán las pestañas de Comparativas Anuales por Edad, Sexo y por Nivel de Estudios.
- En el cuadro de mando de Flujo Migratorio:
 - o Se seleccionarán diferentes años de Padrón y Nacionalidad.

4.5.1. Prueba del Cuadro de Mando de Bajas de Población

Se irán probando todos los parámetros definidos: (Período, MotivoBaja, Distrito, Límite, PaísDestino)

1.- Prueba Selección Motivos de Baja

. Motivo Baja: Baja por Cambio de Residencia en España

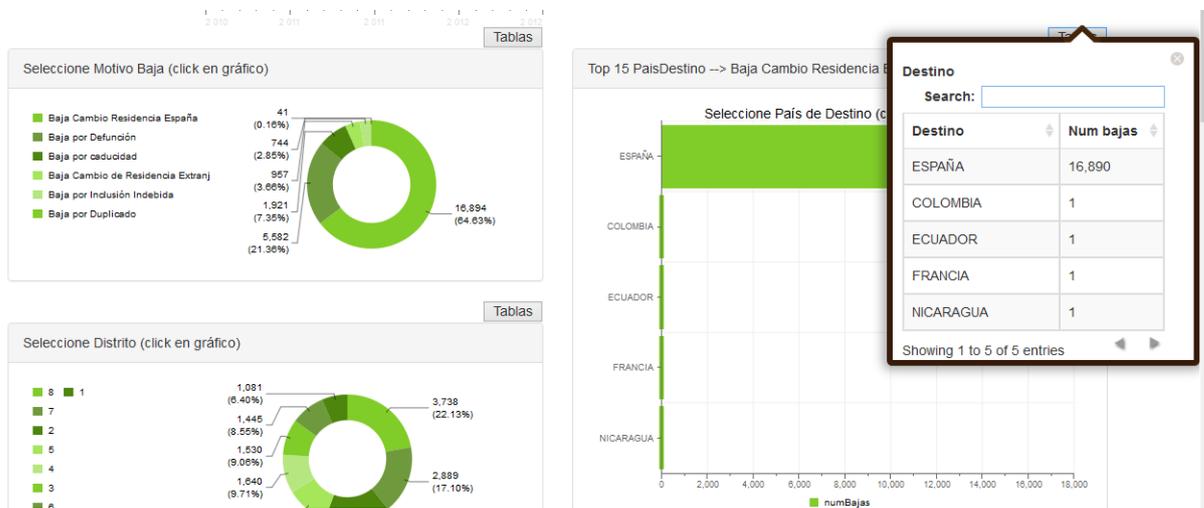


Figura 55: CM_Bajas_Error_Motivo Baja_Cambio Residencia España

Se detectan errores de inconsistencia de datos: hay 4 registros que tienen país de destino diferente de España cuando se trata de un movimiento de cambio residencia dentro de España. Se modificaría por comando sql en la base de datos.

. Motivo Baja: Baja por Defunción.

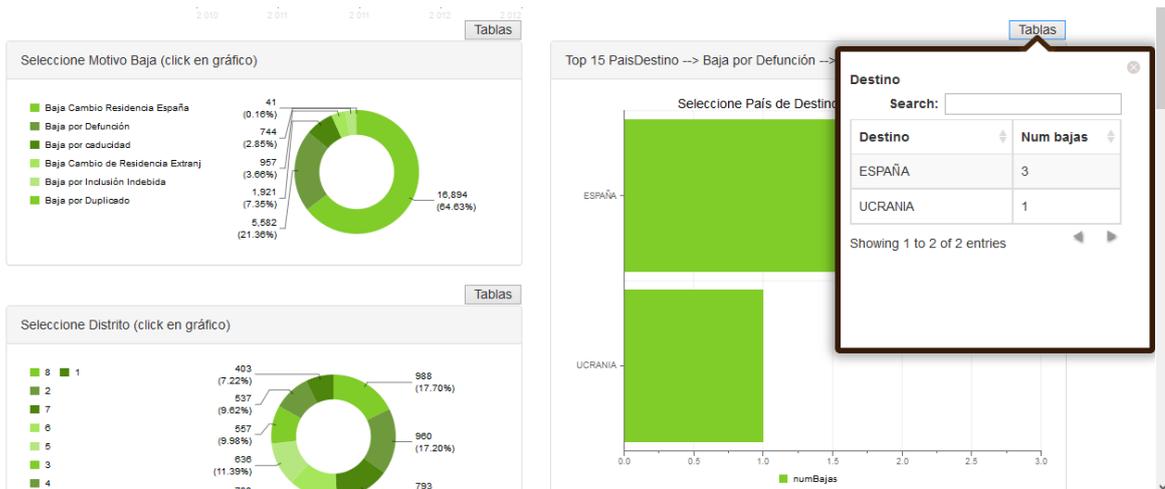


Figura 56: CM_Bajas_Error_Motivo Baja_Defunción

También hay 4 registros que tienen información errónea en país destino cuando debería ser null para las Bajas por Defunción. Se modificaría directamente en la base de datos.

En realidad, la selección por País Destino sólo tiene sentido para el motivo de baja de Cambio de Residencia al Extranjero como se verá a continuación.

. Motivo Baja: Baja por Cambio de Residencia al Extranjero.

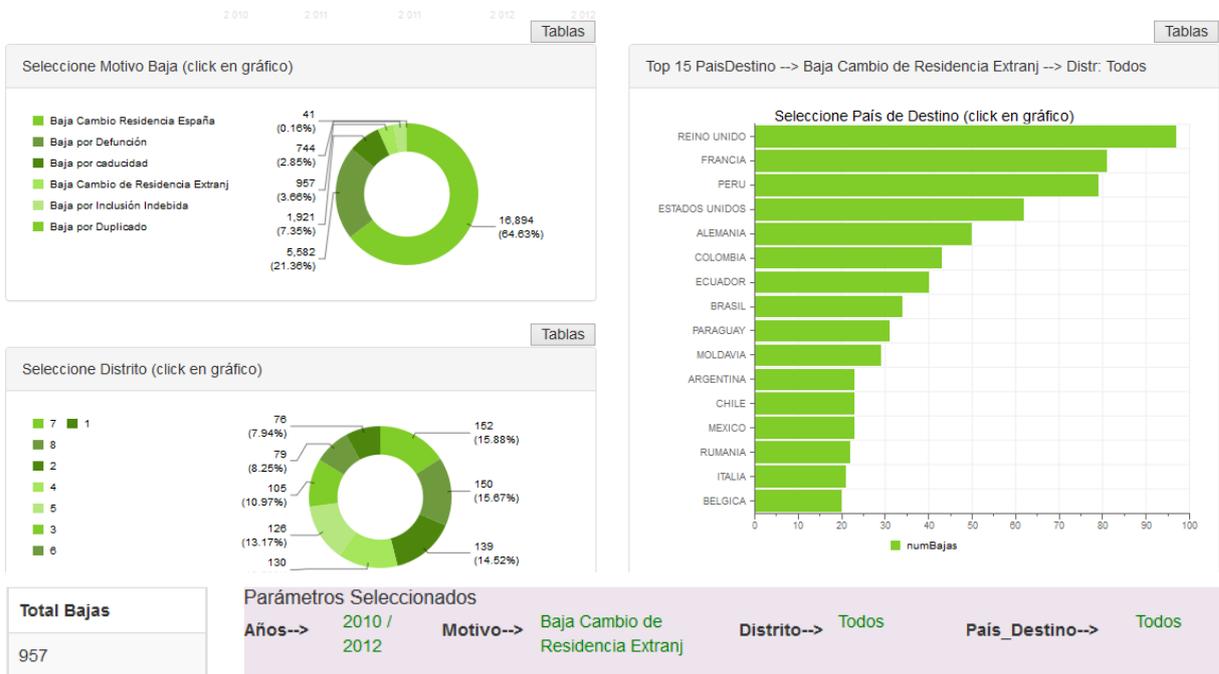


Figura 57: CM_Bajas_Top 15 Paises_Cambio Residencia al Extranjero

En la parte central del cuadro se indica el nº de bajas (957) para los parámetros seleccionados: Años (2010-2012), Motivo de baja: Cambio de residencia al Extranjero, considerando todos los distritos y todos los países de destino.

Emigración Exterior.

Se va a hacer un análisis de Emigración Exterior seleccionando algún país de destino.

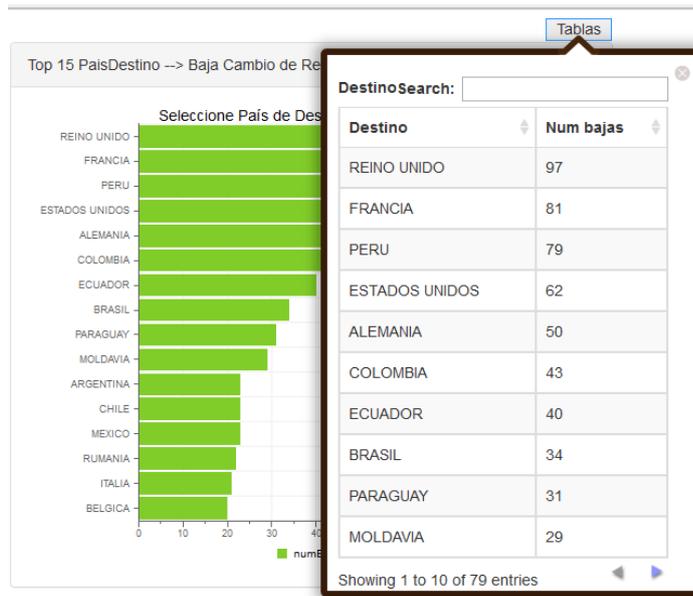


Figura 58: CM _Bajas_Top 15 Paises_Cambio Residencia Extranjero_Vista Tablas

1. País destino seleccionado: **Reino Unido**, (Pestaña): **Por Edad, Sexo y Estudios**

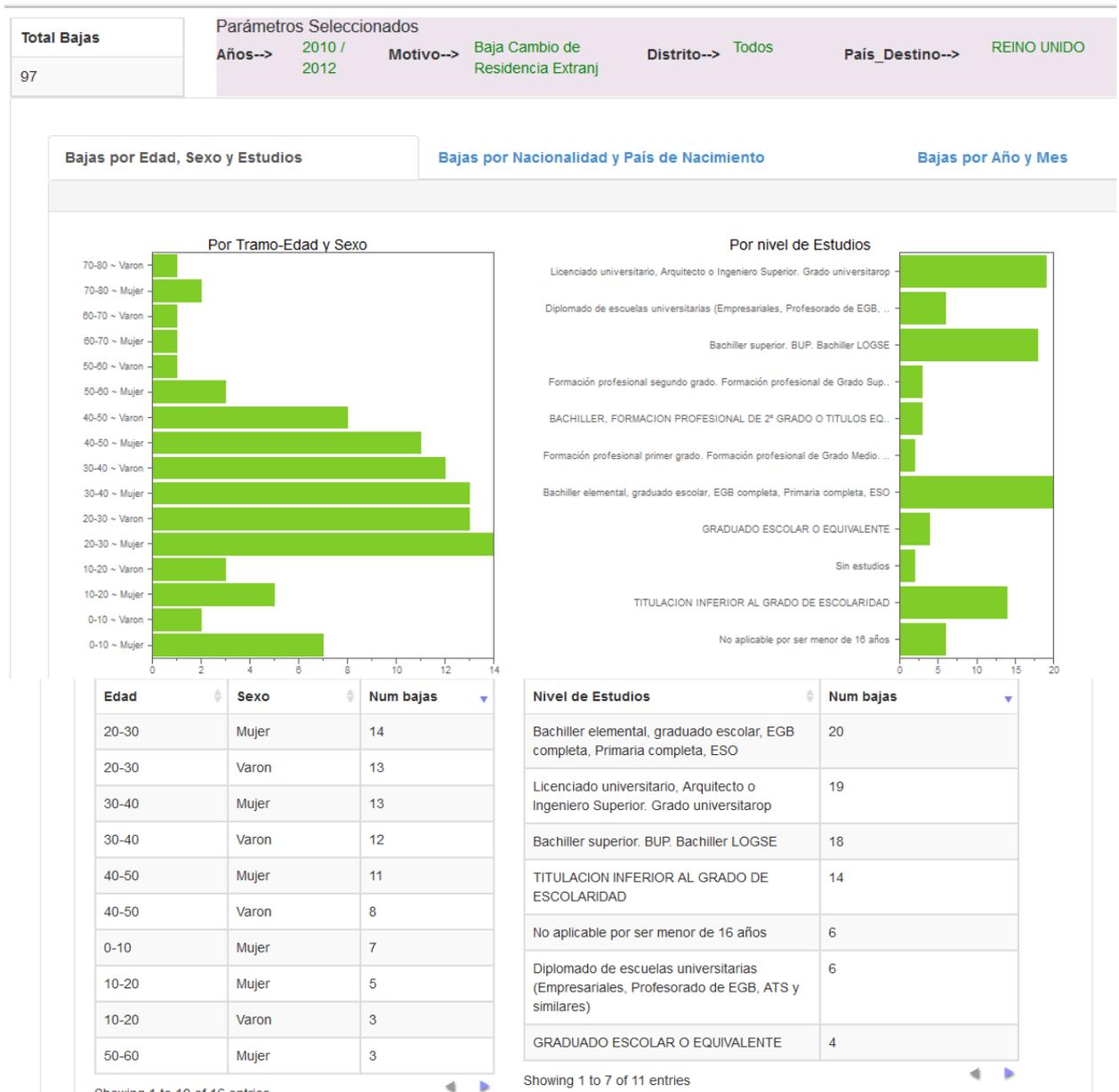


Figura 59: CM_Bajas_País_Destino: Reino Unido_Pestaña_Por Edad, Sexo y Estudios.

País destino seleccionado: **Reino Unido**, (Pestaña): **Por País de Nacimiento y Nacionalidad**

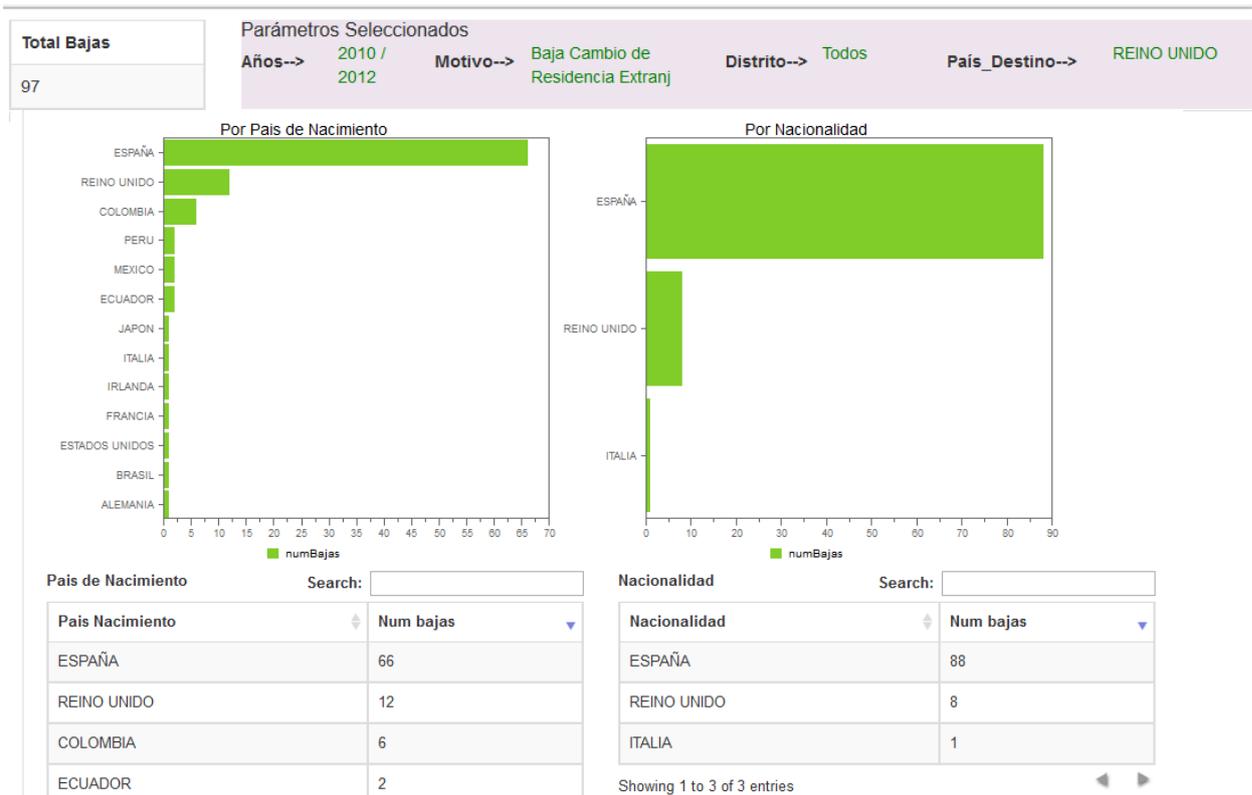


Figura 60: CM _Bajas_País_Destino: Reino Unido_Pestaña_Por País Nac. y Nacionalidad

Se deduce que el perfil de personas que ha emigrado(2010-2012) a Reino Unido es:

- Intervalos de edad (20–30, 30-40 años)
- Estudios: Bachiller elemental, Licenciado, Bachiller superior
- Nacionalidad: mayoritariamente española

A continuación se ve algún país de destino más, pero sin representar los gráficos, sólo la vista de las tablas.

2. País destino seleccionado: **Estados Unidos**, (Pestaña): **Por Edad, Sexo y Estudios**

Total Bajas	Parámetros Seleccionados			
62	Años--> 2010 / 2012	Motivo--> Baja Cambio de Residencia Extranj	Distrito--> Todos	País_Destino--> ESTADOS UNIDOS

Tramo Edad y Sexo			Nivel de Estudios	
Edad	Sexo	Num bajas	Nivel de Estudios	Num bajas
30-40	Varon	13	Licenciado universitario, Arquitecto o Ingeniero Superior. Grado universitarop	17
30-40	Mujer	9	Bachiller superior. BUP. Bachiller LOGSE	13
20-30	Mujer	7	TITULACION INFERIOR AL GRADO DE ESCOLARIDAD	10
20-30	Varon	7	Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO	8
40-50	Varon	6	Doctorado y estudios de postgrado o especialización para licenciados	4
40-50	Mujer	5	No aplicable por ser menor de 16 años	3
0-10	Mujer	3	GRADUADO ESCOLAR O EQUIVALENTE	2
0-10	Varon	2		
10-20	Mujer	2		
50-60	Mujer	2		

Showing 1 to 10 of 14 entries

Showing 1 to 7 of 12 entries

Figura 61: CM _Bajas_País_Destino: Estados Unidos_Pestaña_Por Edad, Sexo y Estudios.

País destino seleccionado: **Estados Unidos**, (Pestaña): **Por País de Nacimiento y Nacionalidad**

Total Bajas	Parámetros Seleccionados			
62	Años--> 2010 / 2012	Motivo--> Baja Cambio de Residencia Extranj	Distrito--> Todos	País_Destino--> ESTADOS UNIDOS

País de Nacimiento		Search:	Nacionalidad		Search:
País Nacimiento	Num bajas		Nacionalidad	Num bajas	
ESPAÑA	47		ESPAÑA	59	
ESTADOS UNIDOS	5		COLOMBIA	2	
COLOMBIA	3		ESTADOS UNIDOS	1	
CUBA	2				
MEXICO	1				
PERU	1				
REP. DOMINICANA	1				
SUIZA	1				
VENEZUELA	1				

Showing 1 to 9 of 9 entries

Showing 1 to 3 of 3 entries

Figura 62: CM _Bajas_País_Destino: Estados Unidos_Pestaña_Por País Nac. y Nacionalidad

Se **deduce** que el perfil de personas que ha emigrado (2010-2012) a Estados Unidos es:

- Intervalos de edad (30–40, 20-30 años)
- Estudios: Licenciado, Bachiller superior
- Nacionalidad: mayoritariamente española

3. País destino seleccionado: **Perú**, (Pestaña): **Por Edad, Sexo y Estudios**

Total Bajas	Parámetros Seleccionados			
79	Años--> 2010 / 2012	Motivo--> Baja Cambio de Residencia Extranj	Distrito--> Todos	País_Destino--> PERU

Tramo Edad y Sexo			Nivel de Estudios	
Edad	Sexo	Num bajas	Nivel de Estudios	Num bajas
20-30	Varon	14	Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO	23
30-40	Varon	14	Bachiller superior. BUP. Bachiller LOGSE	17
40-50	Varon	9	No aplicable por ser menor de 16 años	6
20-30	Mujer	7	Licenciado universitario, Arquitecto o Ingeniero Superior. Grado universitarop	6
30-40	Mujer	7	TITULACION INFERIOR AL GRADO DE ESCOLARIDAD	5
0-10	Mujer	5	Enseñanza Primaria incompleta, cinco cursos de EGB o equivalente, Certificado de escolaridad o equivalente	4
10-20	Mujer	5	Formación profesional segundo grado. Formación profesional de Grado Superior. Maestría industrial	4
40-50	Mujer	5		
50-60	Mujer	4		
50-60	Varon	4		

Showing 1 to 10 of 13 entries

Showing 1 to 7 of 12 entries

Figura 63: CM _Bajas_País_Destino: Perú_Pestaña_Por Edad, Sexo y Estudios.

País destino seleccionado: **Perú**, (Pestaña): **Por País de Nacimiento y Nacionalidad**

Total Bajas	Parámetros Seleccionados			
79	Años--> 2010 / 2012	Motivo--> Baja Cambio de Residencia Extranj	Distrito--> Todos	País_Destino--> PERU

País de Nacimiento		Search:	Nacionalidad		Search:
País Nacimiento	Num bajas		Nacionalidad	Num bajas	
PERU	67		PERU	59	
ESPAÑA	11		ESPAÑA	18	
ESTADOS UNIDOS	1		ESTADOS UNIDOS	2	

Showing 1 to 3 of 3 entries

Showing 1 to 3 of 3 entries

Figura 64: CM _Bajas_País_Destino: Perú_Pestaña_Por Nacionalidad y País Nac.

Se deduce que el perfil de personas que ha emigrado (2010-2012) a Perú es:

- Intervalo de edad (20–30, 30-40 años)
- Estudios: Bachiller elemental, Bachiller superior
- Nacionalidad: mayoritariamente peruana.

Se observa el diferente perfil de población según el país de destino seleccionado, en el caso de Reino Unido y Estados Unidos es población española, muchos con estudios superiores que habrán emigrado buscando trabajo, mientras que en el caso de Perú, es población con estudios más inferiores, con nacionalidad del país destino, parece que se trata de emigrantes que retornan a su país.

4. Se hace una última prueba cambiando todos los parámetros:
 - Período: 2011-2012
 - Limit Top: 20
 - Motivo Baja: Caducidad
 - Distrito: 5

Bajas Población

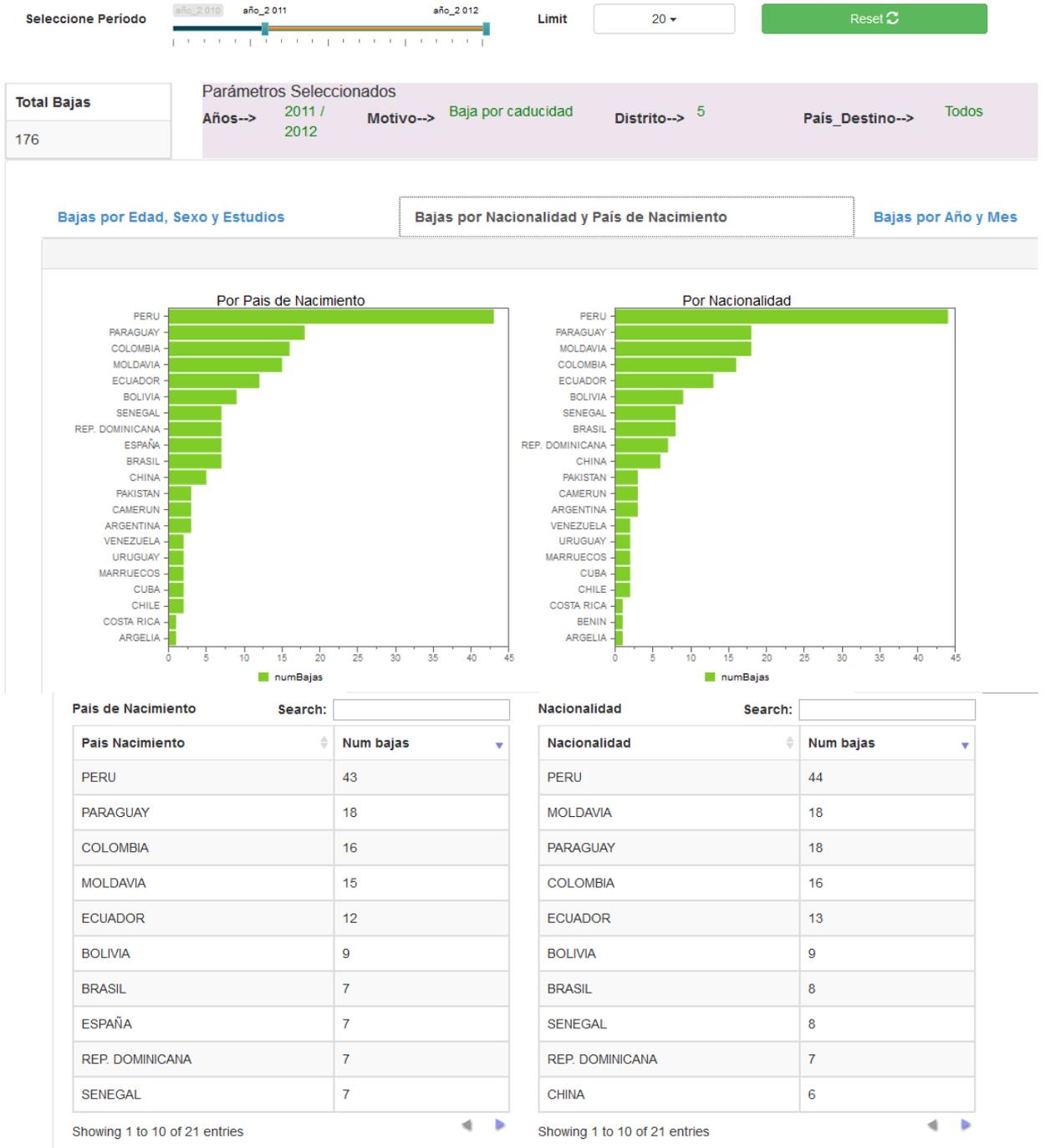


Figura 65: CM _Bajas_Caducidad_ Años:2011-12, Distrito 5, Pestaña_Por País Nac y Nacionald

Las bajas por Caducidad no tienen información en País Destino. En la figura anterior, se observa la pestaña País Nacimiento y Nacionalidad donde aparecen representados 20 países que fue el valor que se puso al límite Top, aunque en las vistas de las tablas siempre se representan todos los registros).

. Pestaña Bajas por Año y Mes para estos mismos parámetros

Bajas Población

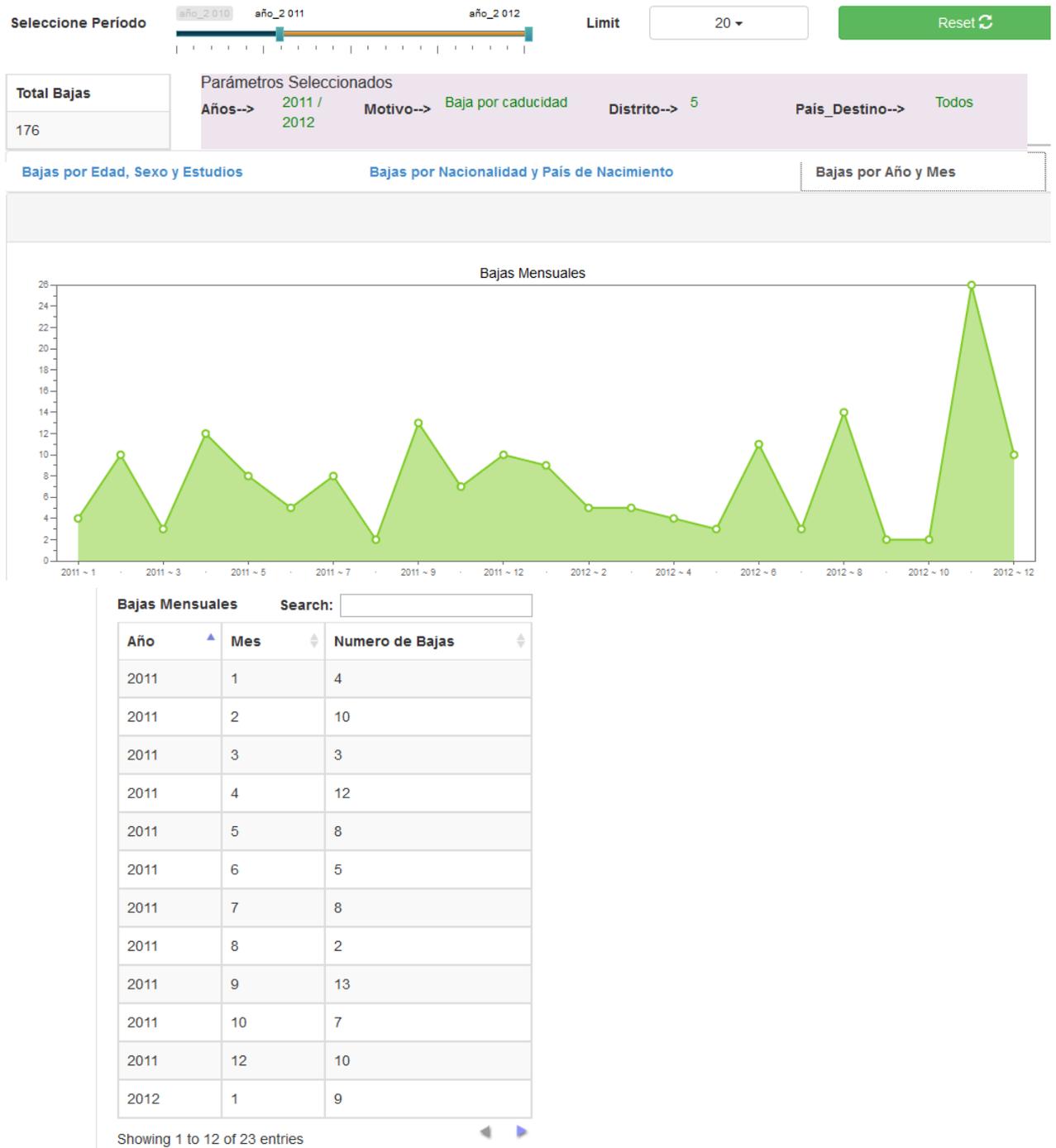


Figura 66: CM _Bajas_Caducidad_ Años:2011-12, Distrito 5, Pestaña_Bajas por Año y Mes

4.5.2. Prueba del Cuadro de Mando de Población Histórica.

Se prueba cambiando el Año Padrón a **2011** y seleccionando alguna nacionalidad.

1. Nacionalidad seleccionada: **Rumanía**, Pestaña : **Población por Edad, Sexo y Estudios**

Total Empadronados	Parámetros Seleccionados		
1,803	Año Padrón--> 2011	Distrito--> Todos	Nacionalidad--> RUMANIA

Tramo Edad y Sexo			Search:	Nivel de Estudios		Search:
Edad	Sexo	Empadronados		Nivel de Estudios	Empadronados	
20-30	Mujer	283		Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO	612	
30-40	Varon	270		TITULACION INFERIOR AL GRADO DE ESCOLARIDAD	346	
20-30	Varon	242		Bachiller superior. BUP. Bachiller LOGSE	222	
30-40	Mujer	216		No aplicable por ser menor de 10 años	145	
40-50	Varon	150		BACHILLER, FORMACION PROFESIONAL DE 2º GRADO O TITULOS EQUIVALENTES O SUPERIORES	91	
40-50	Mujer	144				
00-10	Mujer	96				
00-10	Varon	89				
10-20	Varon	81				

Figura 67: CM _Poblacion_ Nacionalidad Rumanía, Pestaña_Por Edad, Sexo y Estudios

Nacionalidad seleccionada: **Rumanía**, Pestaña : **Población por País Nac y Procedencia**

Total Empadronados	Parámetros Seleccionados		
1,803	Año Padrón--> 2011	Distrito--> Todos	Nacionalidad--> RUMANIA

Pais Nacimiento	Empadronados	Pais de Procedencia	Empadronados
RUMANIA	1,627	RUMANIA	1,088
ESPAÑA	104	ESPAÑA	497
MOLDAVIA	72	#null	184
		MOLDAVIA	33
		ITALIA	1

Showing 1 to 3 of 3 entries

Showing 1 to 5 of 5 entries

Figura 68: CM _Poblacion_ Nacionalidad Rumanía, Pestaña_Por País Nac y Procedencia

Se **deduce** que el perfil de personas empadronadas en 2.011 con nacionalidad rumana es:

- Intervalo de edad (20–30, 30-40 años)
- Estudios: mayoritariamente estudios inferiores a Bachiller superior
- Países Nacimiento: Rumanía, España
- Países Procedencia: Rumanía, España

2. Nacionalidad seleccionada: **Italia**, Pestaña : **Población por Edad, Sexo y Estudios**

Total Empadronados		Parámetros Seleccionados		
323		Año Padrón--> 2011	Distrito--> Todos	Nacionalidad--> ITALIA
Tramo Edad y Sexo		Search: <input type="text"/>		
Edad	Sexo	Empadronados		
30-40	Varon	63		
20-30	Varon	48		
40-50	Varon	35		
20-30	Mujer	34		
40-50	Mujer	28		
30-40	Mujer	25		
50-60	Varon	13		
10-20	Varon	12		
10-20	Mujer	11		
60-70	Varon	10		
Nivel de Estudios		Search: <input type="text"/>		
Nivel de Estudios	Empadronados			
Licenciado universitario, Arquitecto o Ingeniero Superior	62			
Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO	58			
Bachiller superior. BUP. Bachiller LOGSE	53			
Diplomado de escuelas universitarias (Empresariales, Profesorado de EGB, ATS y similares)	31			
BACHILLER, FORMACION PROFESIONAL DE 2º GRADO O TITULOS EQUIVALENTES O SUPERIORES	28			

Figura 69: CM_Población_Nacionalidad: Italia, Pestaña por Edad, Sexo y Estudios

Nacionalidad seleccionada: **Italia**, Pestaña **Población por País Nac y Procedencia**

Total Empadronados		Parámetros Seleccionados		
323		Año Padrón--> 2011	Distrito--> Todos	Nacionalidad--> ITALIA
País Nacimiento		Empadronados		
ITALIA		171		
ARGENTINA		78		
ESPAÑA		17		
BRASIL		10		
COLOMBIA		8		
VENEZUELA		6		
PERU		5		
ALEMANIA		4		
PARAGUAY		4		
País de Procedencia		Empadronados		
ITALIA		110		
#null		96		
ESPAÑA		80		
ARGENTINA		19		
COLOMBIA		6		
VENEZUELA		4		
PARAGUAY		2		
BRASIL		1		
GUATEMALA		1		

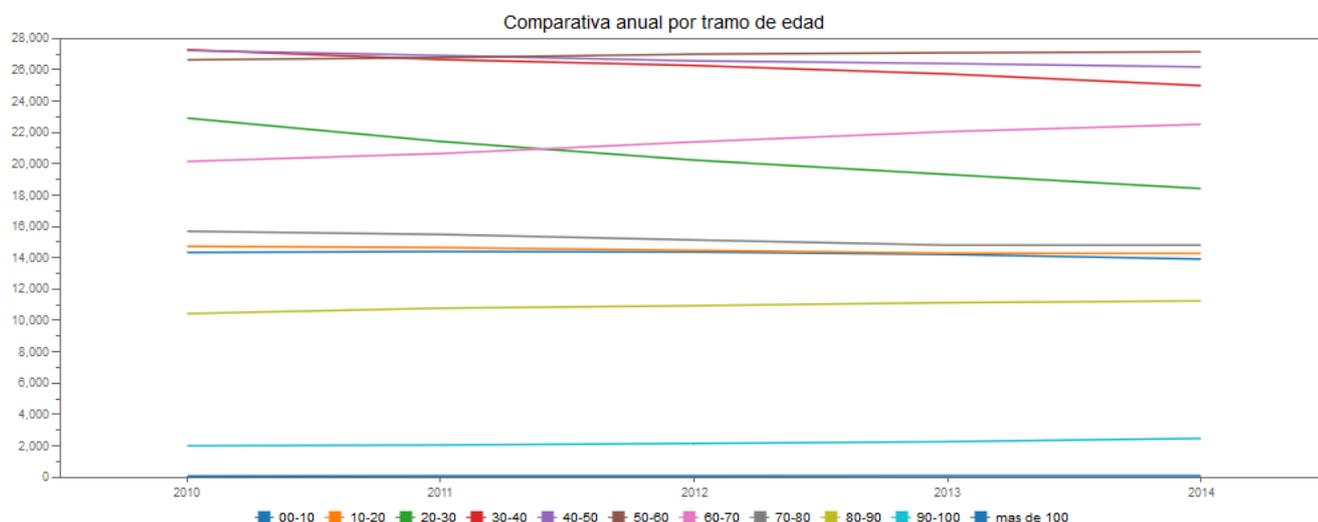
Figura 70: CM_Población_Nacionalidad: Italia, Pestaña por País Nac y Procedencia.

Se **deduce** que el perfil de personas empadronadas en 2.011 con nacionalidad italiana es:

- Intervalo de edad (30–40, 20-30 años) y mayor nº de varones
- Estudios: alto porcentaje de estudios superiores
- Países de Nacimiento: Italia, Argentina, España
- Países de Procedencia: Italia, España (se observa el elevado nº de nulls que hay en procedencia).

A continuación se ve la última pestaña de Comparativa por Años
Pestaña comparativa por años

1. Comparativa por Tramo de Edad y Años



Tramo Edad

Tramo Edad	2010	2011	2012	2013	2014
00-10	14,338	14,393	14,366	14,215	13,914
10-20	14,725	14,650	14,468	14,288	14,277
20-30	22,912	21,423	20,233	19,316	18,414
30-40	27,276	26,650	26,270	25,732	24,987
40-50	27,232	26,914	26,566	26,397	26,179
50-60	26,630	26,777	26,993	27,086	27,143
60-70	20,147	20,651	21,391	22,048	22,516
70-80	15,688	15,488	15,133	14,802	14,804
80-90	10,431	10,781	10,938	11,133	11,244
90-100	1,997	2,045	2,144	2,259	2,467

Showing 1 to 10 of 11 entries

Figura 71: CM_Población_Comparativa anual por Tramo de Edad

En la figura anterior se observa que en el período 2010-2014, los tramos de Edad que más umentan son:

60-70 años, 50-60 años, 80-90 años, 90-100 años

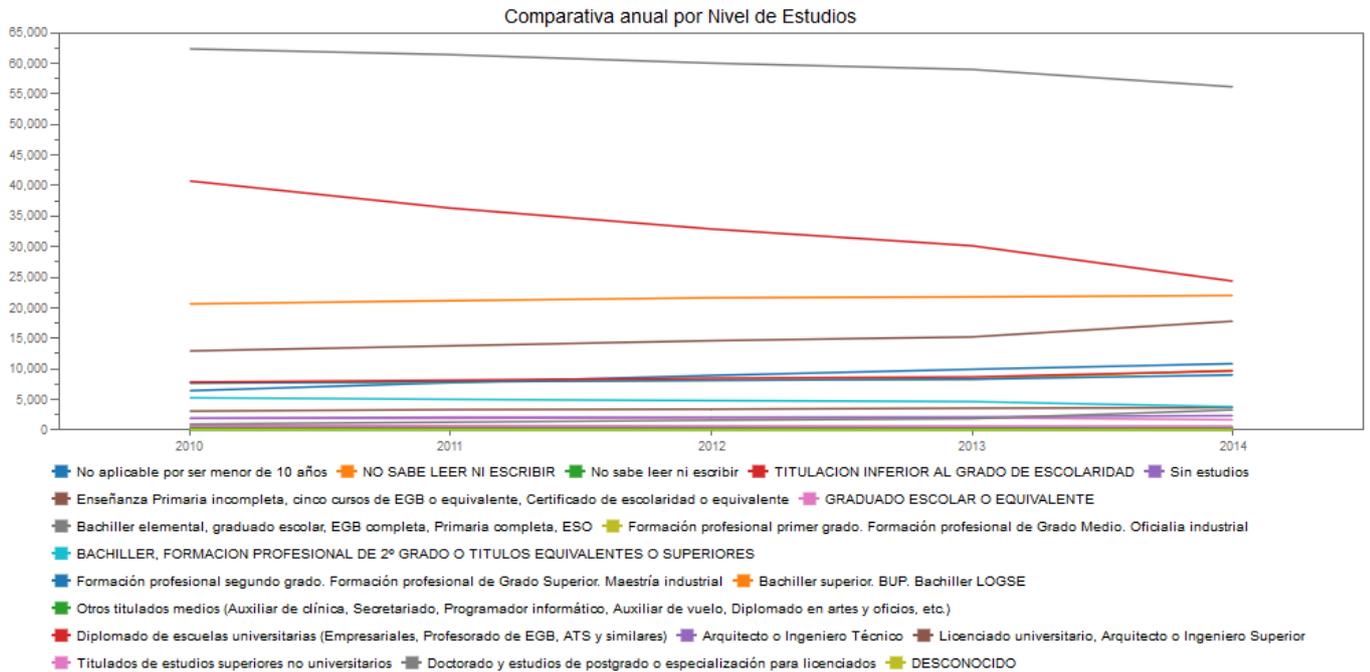
Y disminuyen:

20-30 años, 30-40 años, 40-50 años, 70-80 años, 10-20 años, 00-10 años.

Se **deduce** que:

La población hasta los 50 años disminuye y aumenta a partir de los 50 años (exceptuando el tramo de 70-80 años), sólo comparando estos 5 años de Padrón. Es evidente la **tendencia al envejecimiento de la población** en este período.

2. Comparativa por Nivel de Estudios y Años



Nivel Estudios	2010	2011	2012	2013	2014
Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO	62,345	61,400	60,011	58,976	56,134
TITULACION INFERIOR AL GRADO DE ESCOLARIDAD	40,735	36,298	32,880	30,133	24,344
Bachiller superior. BUP. Bachiller LOGSE	20,623	21,151	21,622	21,773	22,005
Licenciado universitario, Arquitecto o Ingeniero Superior	12,929	13,764	14,606	15,223	17,797
No aplicable por ser menor de 10 años	6,422	7,737	8,920	9,924	10,843
Formación profesional primer grado. Formación profesional de Grado Medio. Oficialía industrial	7,667	7,870	8,113	8,298	9,762
Diplomado de escuelas universitarias (Empresariales, Profesorado de EGB, ATS y similares)	7,821	8,143	8,471	8,673	9,662
Formación profesional segundo grado. Formación profesional de Grado Superior. Maestría industrial	7,691	7,886	8,125	8,312	8,993
BACHILLER, FORMACION PROFESIONAL DE 2º GRADO O TITULOS EQUIVALENTES O SUPERIORES	5,264	5,012	4,807	4,639	3,793
Doctorado y estudios de postgrado o especialización para licenciados	950	1,272	1,593	1,850	3,281
Arquitecto o Ingeniero Técnico	1,920	2,086	2,094	2,124	2,343
GRADUADO ESCOLAR O EQUIVALENTE	1,897	1,902	1,985	1,991	1,676
Titulados de estudios superiores no universitarios	706	683	656	621	584

Figura 72: CM_Población_Comparativa anual por Nivel de Estudios

Se observa que los niveles de estudio que disminuyen son:

- Titulación inferior al grado de Escolaridad
- Bachiller elemental, graduado escolar
- FP II

Y aumentan:

- Licenciado universitario
- Bachiller superior, bachiller LOGSE
- Diplomado universitario
- Doctorado, Postgrado

Se **deduce** que:

En este período (2010-2014) se observa también que está **aumentando el nivel de Estudios** de la Población de Santander

4.5.3. Prueba del Cuadro de Mando de Flujo Migratorio

A continuación se prueba este cuadro de mando con la nacionalidad española.

Año Padrón **2013**, Nacionalidad: **España**

Flujo Migratorio de Población del Ayuntamiento de Santander

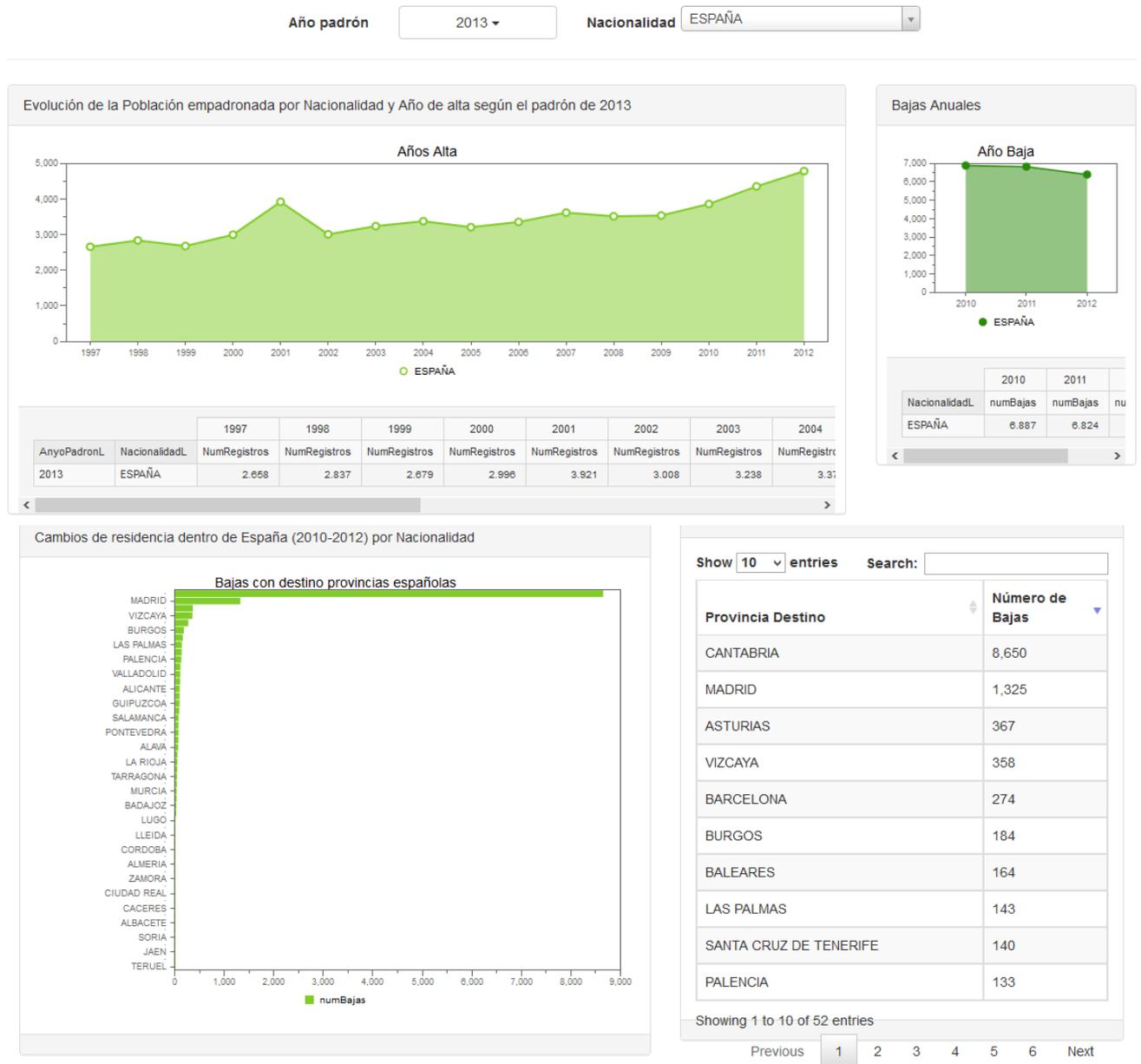


Figura 73: Cuadro Mando Flujo Migratorio, Año: 2013, Nacionalidad: España

Se deduce que:

En el período (2010-2012), en los Cambios de Residencia dentro de España, la población con **nacionalidad española**, las **principales provincias de destino** fueron:

- Mayoritariamente en Cantabria, movimientos hacia otros municipios dentro de la región
- Fuera de Cantabria, las provincias de destino elegidas fueron: Madrid, Asturias, Vizcaya, Barcelona... (Según se indica en la tabla adjunta).

4.6. Integración de Cuadros de Mando.

Entrelazar los 3 Cuadros de Mando

Los 3 Cuadros de Mando anteriormente generados se quieren entrelazar para que desde un cuadro de mando se pueda acceder directamente a los otros dos, sin necesidad de utilizar un menú específico.

Para ello, se van a utilizar unos botones bootstrap, obteniendo la url donde se encuentra almacenado cada Cuadro de Mando y enlazando los dos botones pertinentes en cada uno de los 3 Cuadros de Mando. Los botones utilizados son:

- bt0 para enlazar con el cuadro de mando Bajas de Población
- bt1 para enlazar con el cuadro de mando Población Histórica
- bt2 para enlazar con el cuadro de mando Flujo Migratorio

Y se incluyen en cada cuadro de mando, por ejemplo, en el de Bajas de Población

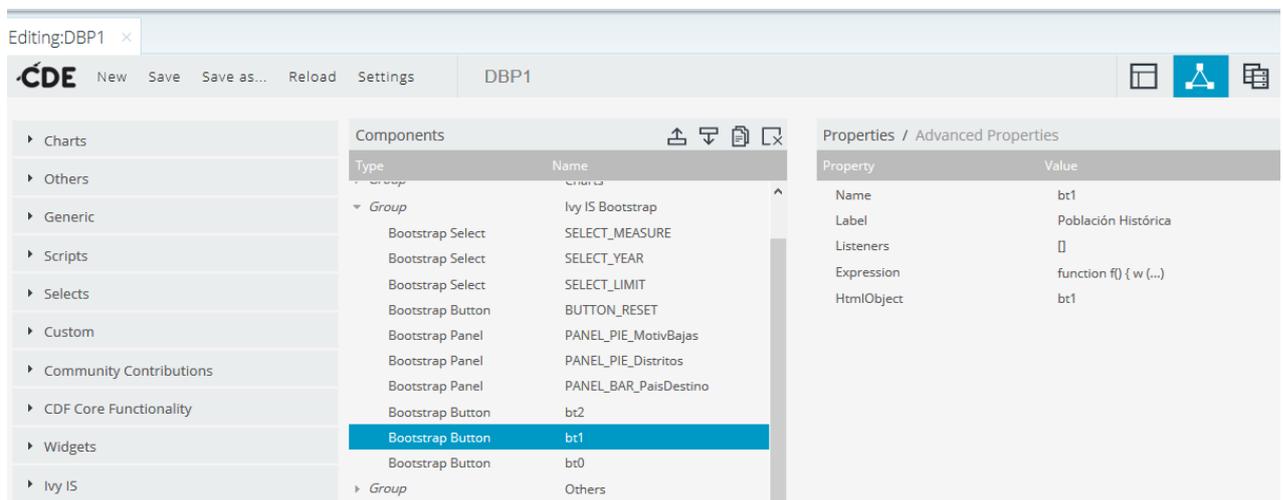


Figura 74: CM_Bajas_button Bootstrap

El button bt1 tiene la etiqueta de Población histórica y en expresión la siguiente función:

```
function f()
{ window.location =
"http://localhost:8080/pentaho/api/repos/%3Apublic%3APob%3ADP1.wcdf/generatedContent"; }
```

Está direccionando al cuadro de mando de Población Histórica.

Así se haría en los 3 cuadros de mando y se obtiene el resultado de la siguiente figura:

Bajas Población

Seleccione Período: añ_2_010 añ_2_012 Limit: 15 Reset

Selecciones: Selecciones Motivo Baja (click en gráfico) Top 15 PaisDestino --> Todos --> Distr: Todos

Selecciones Motivo Baja (click en gráfico)

Baja Cambio Residencia España	41 (0.16%)
Baja por Defunción	744 (2.85%)
Baja por caducidad	957 (3.60%)
Baja Cambio de Residencia Extranj	1,921 (7.35%)
Baja por Inclusión Indevida	5,582 (21.36%)
Baja por Duplicado	16,894 (64.63%)

Selecciones Distrito (click en gráfico)

8	1	1,850 (7.08%)	5,232 (20.02%)
2		2,488 (9.51%)	
7		2,542 (9.72%)	
5		2,543 (9.73%)	4,271 (16.34%)
3		2,976 (11.39%)	4,239 (16.22%)
6			
4			

Selección País de Destino (click en gráfico)

ESPAÑA	16,894 (64.63%)
REINO UNIDO	
FRANCIA	
PERU	
ESTADOS UNIDOS	
ALEMANIA	
COLOMBIA	
ECUADOR	
BRASIL	
PARAGUAY	
MOLDAVIA	
ARGENTINA	
CHILE	
MEXICO	
RUMANIA	
ITALIA	

Total Bajas: 26,139

Parámetros Seleccionados: Años--> 2010 / 2012; Motivo--> Todos; Distrito--> Todos; Pais_Destino--> Todos

Bajas por Edad, Sexo y Estudios | **Bajas por Nacionalidad y País de Nacimiento** | **Bajas por Año y Mes**



Tramo Edad y Sexo

Edad	Sexo	Num bajas
0-10	Mujer	1,025
0-10	Varon	1,089
10-20	Mujer	715
10-20	Varon	645
20-30	Mujer	2,724
20-30	Varon	2,476
30-40	Mujer	2,728

Nivel de Estudios

Nivel de Estudios	Num bajas
Bachiller elemental, graduado escolar, EGB completa, Primaria completa, ESO	8,613
TITULACION INFERIOR AL GRADO DE ESCOLARIDAD	4,902
Bachiller superior. BUP. Bachiller LOGSE	3,105
Licenciado universitario, Arquitecto o Ingeniero Superior. Grado universitarop	2,111
No aplicable por ser menor de 16 años	1,471

Figura 75: Los 3 Cuadros de Mando entrelazados.

Así, desde un cuadro de mando se accede directamente a los otros dos, mediante estos botones bootstrap.

Integrar, incrustar Cuadros de Mando en Pentaho en una página web

Hasta ahora, para visualizar el cuadro de mando, una vez iniciado el servidor pentaho (start-pentaho.bat), se hacía login en la consola de usuario de Pentaho (<http://localhost:8080/pentaho>) y se introducía Usuario = Admin y Password = password, según la figura adjunta:

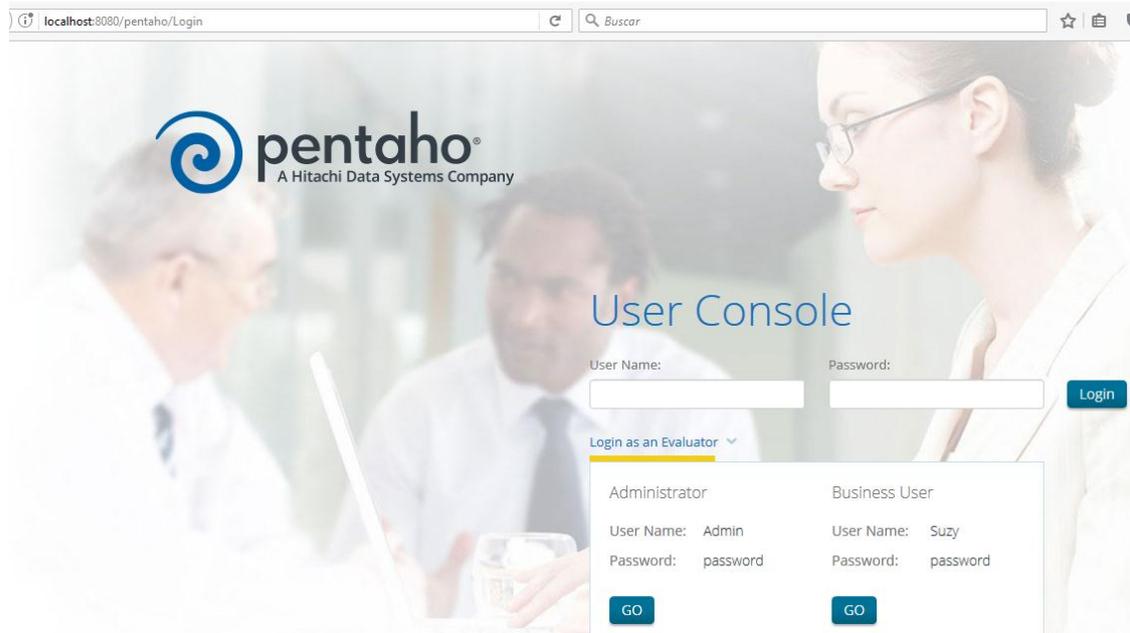


Figura 76: Consola_usuario_pentaho_login

Se puede incluir en un archivo html un enlace a uno de los Cuadros de Mando anteriores y de forma implícita hacer este login asignando usuario y password.

En un archivo html, se incluye el siguiente enlace, que contiene la url anterior del cuadro de mando de Bajas, al que se ha añadido el usuario y password:

```
<!DOCTYPE HTML>
<html>
  <a
    href="http://localhost:8080/pentaho/api/repos/%3Apublic%3APob%3ADBP1.wcdf/generated
    Content?userid=Admin&password=password">Cuadros Mando</a>
</html>
```

Al abrir este archivo en el navegador, el enlace Cuadros Mando lleva directamente al cuadro de mando de Bajas y desde él, a los otros Cuadros de Mando.

También hay que modificar el archivo:

```
\pentaho\biserve-ce\pentaho-solutions\system\security\properties
```

Y poner a true este parámetro: requestParameterAuthenticationEnabled=true

(Pedro Alves, 2015-02-24, Easy authentication in Pentaho 5.3)

5. Conclusiones y trabajos futuros.

Se ha desarrollado un proyecto de Business Intelligence sobre un conjunto de archivos históricos de demografía del Ayuntamiento de Santander logrando realizar todos los pasos de un proyecto BI (modelo de datos, proceso ETL, cubos OLAP) hasta llegar a elaborar unos cuadros de mando que permiten consultar y analizar la información profundizando en algunos datos demográficos de la población así como también sobre los flujos migratorios que se han producido en estos años objeto de estudio.

Se han cumplido objetivos en el sentido de que los cuadros de mando nos han permitido confirmar y además obtener información detallada sobre aspectos tales como el envejecimiento de la población, la búsqueda de empleo en el extranjero por parte de muchos jóvenes durante estos años de crisis, el retorno de inmigrantes a su país...

Un resumen de la información obtenida de las pruebas de los Cuadros de Mando es:

- De las comparativas anuales de Población durante el período 2010-2014, se deduce:
 - o El envejecimiento de la población
 - o El aumento del nivel de estudios (mayor nº de personas con niveles más altos de estudios).
 - o Visualizar perfil de población empadronada (edad, sexo, estudios...) por nacionalidad.
- En Emigraciones exteriores durante el período 2010-2012, y según el País de Destino, se deduce:
 - o La salida al extranjero de gente joven, muchos con estudios superiores a países como Reino Unido, Estados Unidos.
 - o El retorno de inmigrantes a su país de origen como Perú.
- En Emigraciones interiores se obtiene para cada nacionalidad las provincias españolas más solicitadas como destino.
 - o En el caso de la nacionalidad española, mayoritariamente han tenido como destino otros municipios de Cantabria, y de las otras provincias españolas, las más frecuentes han sido: Madrid, Asturias, Vizcaya y Barcelona.

También se han reflejado algunas limitaciones por la falta de información de algunos datos en las altas de población (sin información en el país, provincia de procedencia..)

Como Trabajo futuro se consideran las siguientes sugerencias que mejorarían estos Cuadros de Mando realizados sobre la población histórica del Ayuntamiento de Santander.

- Disponer de más años de Población histórica y Bajas de Población, ya que en el Open Data de Santander, sólo están disponibles Población histórica (años 2010-2014) y Bajas de Población (años 2010-2012).
- Si es posible, generar también un fichero de Altas anuales, similar al de Bajas, que incluyera los motivos del Alta (Alta por nacimiento, Alta desde el Extranjero o desde otro municipio de España...) y con los datos de País, Provincia de Procedencia, es decir, un fichero análogo al de Bajas, que permitiría dar una información más completa en las Inmigraciones (tanto exterior como interior) y en las estadísticas de Altas mensuales.
- También mejoraría este proyecto si se programa un proceso ETL con el PDI (Pentaho Data Integration) que incorpore los nuevos archivos históricos que se fueran depositando en el Open Data, programando la actualización de las tablas de hechos y dimensiones tanto de Bajas como de Población Histórica.

6. Bibliografía

Gartner Group.(2017). *Research Methodologies* .Obtenido de:
http://www.gartner.com/technology/research/methodologies/research_mq.jsp

Gartner IT Glossary.(2017). Obtenido de: <http://www.gartner.com/it-glossary/business-intelligence-bi/>

Pedro Alves, 2015-02-24, Easy authentication in Pentaho 5.3. Obtenido de:
<http://pedroalves-bi.blogspot.com.es/2015/02/useful-tips-easy-authentication-in.html>

Pentaho BI Suite Tutorials (2014-06). Obtenido de:
<http://pentaho-bi-suite.blogspot.com.es/2014/06/high-performance-dashboard-sales.html>

Pentaho Group 2016, - Documentation 6.1 CTools_Overview. Obtenido de:
https://help.pentaho.com/Documentation/6.1/0R0/CTools/CTools_Overview

Pentaho Group .(2017). <http://www.pentahocom/>. Obtenido de:
<http://www.pentaho.com/product/product-overview>

Pentaho Group .(2017). Mondrian Documentation. Obtenido de:
<http://mondrian.pentaho.com/documentation/schema.php>

(Pierri,2011, Manipulacion de datos con Kettle. Obtenido de:
<https://es.slideshare.net/mpierri/manipulacion-de-datos-con-kettle>

Stratebi .(2017, Comparativa de herramientas Business Intelligence).
<http://www.stratebi.com/>. Obtenido de: <http://todobi.blogspot.com.es/2017/04/comparativa-de-herramientas-business.html>

(Stratebi 2017, Cuadros de mando y Scorecards). Obtenido de:
<http://www.stratebi.com/-/cuadros-de-mando-y-scorecards>