

Data Warehouse

Adultos condenados anualmente en España según su género, nacionalidad, edad, año de condena, cantidad de delitos cometidos y número de condenas

Fecha: 3 de Marzo de 2022

Asignatura: Bases de Datos Avanzadas

Profesor: Julián Ruiz Fernández

Componentes: Jesús Baltasar Fernández

Carlos Gómez Fernández

Domenico Fasciano

Tabla de contenido

1.	Resumen.....	3
2.	Introducción	3
3.	Desarrollo	4
	a. Procesamiento de datos	4
	b. Explicación tecnológica.....	4
	c. Diseño del Data Warehouse	4
4.	Resultados	5
5.	Conclusiones.....	6

1. Resumen

El desarrollo del trabajo se basa en a partir de datos disponibles públicamente, de iniciativas *OpenData*, diseñar e implementar un *Data Warehouse*, teniendo en cuenta tanto los procedimientos para su creación como las preguntas a las que se podrá responder a partir de su análisis.

2. Introducción

Para facilitar la comunicación entre los componentes del equipo en los avances en el desarrollo del trabajo se ha optado por la utilización de la plataforma *GitHub*, para controlar y mantener las versiones del trabajo en su desarrollo incremental.

En cuanto al desarrollo del cuaderno de laboratorio se optó por la utilización de una de las plantillas proporcionadas en el enunciado y el cual se ha realizado haciendo uso de la herramienta *Overleaf*, la cual proporciona un editor del lenguaje *LaTeX*.

Uno de los primeros hitos importantes fue la selección de las herramientas y aplicaciones que se emplearán en el desarrollo del trabajo, entre las cuales cabe destacar *Excel*, para el tratamiento preliminar de los datos extraídos, *Microsoft Access*, para el desarrollo de la base de datos la cual de soporte al *Data Warehouse*, y *Python*, como lenguaje de programación mediante el cual acceder a la base de datos y realizar consultas sobre la misma.

Posteriormente se comenzó con el proceso de extracción del *dataset* a utilizar para la creación del *Data Warehouse*, el primero seleccionado contenía el precio medio del metro cuadrado de suelo urbano en municipios de menos de 1.000 habitantes, obtenido, en formato *Excel*, de la web del Ministerio de Transporte, Movilidad y Agenda Urbana.

Tras analizar en profundidad los datos que contenía, reparamos en que eran insuficientes para la construcción del *Data Warehouse*, por lo que se optó por modificar la fuente de datos por una con una cantidad mayor tanto de atributos como de registros, la cual corresponde con el índice de precios de propiedades por ciudad desde el año 2009 hasta el año 2021, la cual constaba de un total de 13 archivos *Excel* y se extrajo de la web *Kaggle*.

Tras analizar en profundidad los datos que contenía, reparamos, de nuevo, en que eran insuficientes para la construcción del *Data Warehouse*, por lo que se optó por modificar la fuente de datos por una con una cantidad mayor tanto de atributos como de registros, la cual corresponde con los condenados mayores de edad según sexo, edad, nacionalidad y número de delitos en España desde el año 2013 hasta el año 2020, en formato *Excel* y extraída del Instituto Nacional de Estadística (INE).

3. Desarrollo

a. Procesamiento de datos

Para comenzar con el desarrollo e implementación del *Data Warehouse*, y tras realizar el proceso de extracción de los datos, se iniciarán los procesos de transformación del *dataset* para adaptarlo tanto al diseño como al objetivo en el análisis de los datos, es por ello por lo que se prescindió de algunos registros, ya que aportaban redundancia innecesaria en los datos, que posteriormente se vería paliada con los resultados de las consultas contra el *Data Warehouse*, tras ello se generaron tanto las tablas de dimensiones como la tabla de hechos que componen el diseño del *Data Warehouse*. Todas las acciones correspondientes con las operaciones de transformación se realizaron haciendo uso de la herramienta *Excel*.

Los procesos de carga se basaron en, haciendo uso de funcionalidades de la herramienta *Microsoft Access* y teniendo divididas en archivos *Excel* independientes cada una de las tablas que conforman el *Data Warehouse*, almacenar y relacionar estas en la herramienta utilizada como motor de la base de datos relacional que subyace en el diseño del *Data Warehouse*.

b. Explicación tecnológica

Tras finalizar todos los procesos *ETL* con el *dataset* seleccionado, se optó por comenzar con la implementación de un *script* en *Python*, el cual diese soporte a la conexión con la base de datos, la creación y realización de consultas *SQL* contra la base de datos y la generación de gráficos que mostrasen la información contenida en el *dataset*.

Para realizar la conexión con la base de datos se ha hecho uso de la librería *pyodbc*, para la realización de las consultas *SQL* contra la base de datos se empleó la librería *pandas*, la cual muestra el resultado de la misma por la terminal, por lo que se optó por la generación de gráficos haciendo uso de la librería *plotly* y finalmente, se automatizó la generación de gráficos, desarrollando un sencillo e interactivo *dashboard* empleando la librería *dash*.

c. Diseño del Data Warehouse

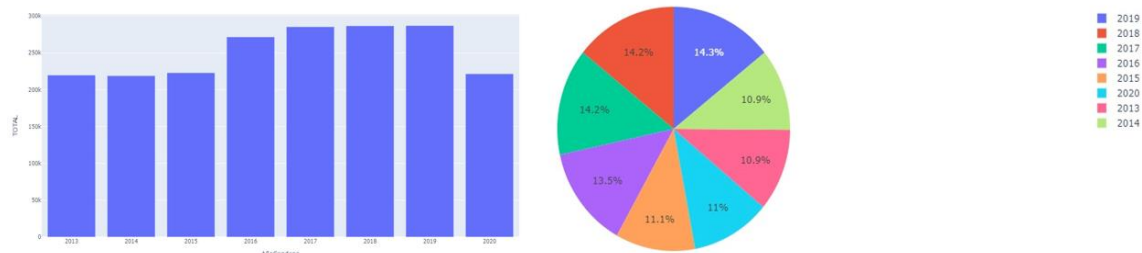
Tras consultar diversas fuentes de información se seleccionó un diseño en estrella, ya que presenta una estructura más sencilla; al incluir muchas menos tablas, pero más grandes y con más información, a diferencia del modelo en copo de nieve, en el cual se tienen muchas más tablas, lo que obligaría a hacer muchos más *Join* y cada uno de estos ralentizaría la consulta.

Entre las tablas de dimensiones del *Data Warehouse* se encuentra la tabla con el género del delincuente, la nacionalidad del delincuente, la edad del delincuente, la fecha de comisión del delito y la cantidad de delitos cometidos por el delincuente.

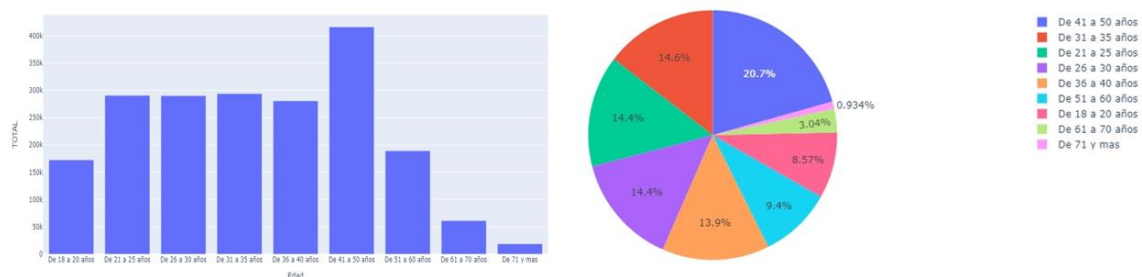
La tabla de hechos del *Data Warehouse* incluye como atributos el género del delincuente, la nacionalidad del delincuente, la edad del delincuente, la fecha de comisión del delito, la cantidad de delitos cometidos por el delincuente y la cantidad de delincuentes que tienen todas las características.

4. Resultados

- Total de condenas por año



- Total de condenas por rango de edad



- Total de condenas por género



- Total de condenas por nacionalidad



- Total de condenas por número de delitos cometidos



5. Conclusiones

- Total de condenados por año

Como muestran los gráficos en los últimos años la cantidad de condenas ha crecido de manera constante, lo cual se puede analizar y/o entender desde dos puntos de vista, o bien la delincuencia crece cada año, o bien la justicia y las fuerzas del orden cada vez son más eficientes y estrictas. Por otra parte, se puede observar como en el año 2020 se ha roto esta tendencia al alza seguramente influenciado por las restricciones de movilidad y la situación social vivida en los peores meses de la pandemia.

- Total de condenas por rango de edad

Como muestran los gráficos, en la edad adulta la cantidad de condenas es más o menos constantes en todos los rangos de edad, a excepción de los más jóvenes y los cuarentones, los cuales son condenados en mucha mayor medida, lo cual se puede deber a que delinchan por necesidad, sin muchos recursos ni planificación previa, lo que los lleve a ser condenados por las autoridades. En los últimos años de edad adulta se aprecia un descenso muy significativo en la cantidad de condenas, debido sin duda por cuestiones físicas de los delincuentes.

- Total de condenas por género

Como muestran los gráficos, los hombres son condenados en mucha mayor proporción que las mujeres, lo cual se puede deber a que, o bien delinquen más o son menos metódicos, organizados y perfeccionistas que las mujeres.

- Total de condenas por nacionalidad

Como muestran los gráficos, los españoles son condenados en mayor proporción que los extranjeros, datos que se deberían analizar con más detenimiento y profundidad, ya que, la proporción de extranjeros en España no se aproxima a la de españoles, representando la población extranjera un 13% de la población de España.

- Total de condenas por número de delitos cometidos

Como muestran los gráficos, cuanto mayor es la cantidad de delitos cometidos menor cantidad de condenas hay, lo cual se debe a varias razones, la principal es que si son condenados, les es mucho más complicado seguir cometiendo delitos, otra podría ser que muchos de los delitos cometidos en España son por necesidad, y si los delincuentes son condenados, aprecian que pierden más de lo que consiguen delinquiendo.