

Caso práctico: almacén de datos para el análisis del impacto conductual de la COVID-19 sobre la población

PRA3- Explotación de datos

Presentación

A partir de la solución oficial de la segunda práctica (PRA2) el estudiante debe implementar los cubos multidimensionales necesarios para la explotación de la información y el posterior análisis de datos. De este modo se facilitará la toma de decisiones de los usuarios potenciales.

Así pues, esta actividad tiene el objetivo de implementar un modelo *multidimensional online analytical processing* (OLAP) para el análisis multidimensional de la información con el fin de responder a las preguntas definidas en el análisis de requerimientos.

Adicionalmente, se facilitará junto a este enunciado el fichero «export_DW_COVID.sql», que contiene los *scripts* de generación y carga de todas las tablas planteadas en la solución, para que el estudiantado pueda partir de la misma base.

Descripción

Más concretamente esta tercera parte del caso práctico consiste en diseñar un modelo OLAP para el análisis multidimensional de la información disponible en el almacén de datos que permita dar respuesta a las siguientes cuestiones:

- 1) Análisis de las provincias con mayor porcentaje de movilidad según datos móviles.
- 2) Análisis del porcentaje de la población que evitaba las aglomeraciones según la comunidad autónoma.
- 3) Análisis del promedio de sanciones por habitante.
- 4) Evolución de las llamadas de urgencia al 112 en Cataluña por tipología de llamada.
- 5) Análisis de las llamadas de urgencia frente al porcentaje de la población que evitaba las aglomeraciones entre los meses comprendidos entre marzo y junio de 2020 en Cataluña, desglosado por provincia.
- 6) Determinación del día de la semana con menor número de denuncias.
- 7) Análisis de las diez fechas (*top ten*) con mayor número de llamadas de urgencia al 112 con tipología de tránsito registrada, tanto en época de COVID como antes.

Para un desarrollo y un diseño correctos del modelo OLAP el documento debe incluir como mínimo la descripción detallada de los siguientes pasos:

- 1) Creación del proyecto en Visual Studio.
- 2) Observación del origen de datos.
- 3) Creación e implementación de cubos.
- 4) Decisión sobre jerarquías, dimensiones y relación de atributos.
- 5) Explotación de la solución.

Además de entregar el documento con la solución de la PRA3, se considerará también la implementación de la máquina virtual proporcionada en el curso.

En resumen, el documento de la solución de la PRA3 debe incluir:

- La descripción de todas las acciones realizadas.
- Las capturas que demuestren la correcta definición de las vistas y cubos OLAP, así como las dimensiones, relaciones y jerarquías.
- Las capturas que muestren la correcta definición de las consultas, la visualización del resultado de las explotaciones y su correcta interpretación.

Guía de muestra

Esta guía de muestra se desarrolla con el fin de ayudar a alcanzar los objetivos planteados de la PRA3. La guía servirá de ejemplo para saber cómo realizar algunas de las tareas anteriormente descritas, es decir, el diseño de explotaciones de un modelo OLAP para el análisis del impacto conductual de la COVID-19 sobre la población.

Diseño del modelo OLAP

El diseño del cubo se realizará creando un proyecto multidimensional y de minería de datos en *visual studio*. Se deberán definir los orígenes de datos, las vistas de los orígenes de datos, los cubos, las relaciones de atributos, las dimensiones y las jerarquías necesarias para realizar la implementación de la solución y las explotaciones solicitadas en el enunciado de la PRA3.

A continuación, se muestra un ejemplo de explotación de datos que se puede realizar tras la implementación de la solución. El estudiantado debe incluir todos los apartados solicitados en este enunciado. Para acceder al visor OLAP, entrad en la pestaña *browser* de cada cubo creado.

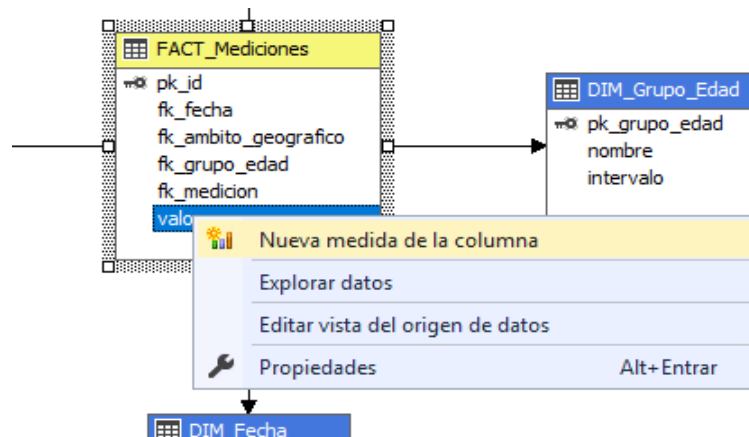
Análisis de las provincias con mayor porcentaje de movilidad según los datos móviles

Para analizar el porcentaje de movilidad dentro del cubo «Cubo_Mediciones» debéis seleccionar la dimensión «Medición» con el atributo «Movilidad de la población durante el estado de alarma». Después, seleccionad los diferentes atributos que queréis visualizar en la consulta y procesáis.

Cubo_Mediciones		Dimensión	Jerarquía	Operador	Expresión de filtro
Metadatos		DIM Medicion	Nombre	Igual	{ Movilidad de la poblaci
Ruscar modelo		<Seleccionar dimensión>			
Grupo de medida:					
<Todos>					
Cubo_Mediciones		Provincia	Unidad Medida	Valor	
Measures		CEUTA	%	1003,42	
FACT Mediciones		MELILLA	%	1253,86	
Recuento FACT Mediciones		ALMERÍA	%	1031,66	
Valor		CÁDIZ	%	1128,26	
KPI		CÓRDOBA	%	1113,93	
DIM Ambito Geografico		GRANADA	%	1114,59	
DIM Fecha		HUELVA	%	1098,52	
DIM Grupo Edad		JAÉN	%	860,53	
DIM Medicion		MÁLAGA	%	1076,6	
Nombre		SEVILLA	%	1197,2	
Unidad Medida					

Sin embargo, la explotación ofrece datos incoherentes, dado que los porcentajes no deberían superar el 100 %. Un análisis más detallado os revela que esto es debido a que el valor se está agregando como un sumatorio. Así, sería más interesante realizar una media de los porcentajes en función de la provincia. Para ello, vais a añadir un nuevo valor agregado.

En el apartado «Estructura de cubo» seleccionad la opción «Nueva medida de la columna» del menú contextual sobre el campo «valor» de la tabla de hechos de mediciones.



Esto crea una nueva medida disponible. Haciendo doble clic sobre ella se abre el menú para configurarla y debéis asignarle un nuevo tipo de agregación. Para cambiarle el nombre, editadlo en «Propiedades».

The **Editar medida** dialog box shows the configuration for a new measure. It includes the following fields:

- Uso:** Promedio a lo largo de un período de tiempo
- Tabla de origen:** FACT_Mediciones
- Columna de origen:** valor

The **Propiedades** window displays the properties for the measure **Valor - FACT Mediciones**. It is organized into two sections:

- Avanzada (Advanced):**
 - AggregateFunction:** AverageOfChildren
 - DataType:** Double
 - DisplayFolder:**
 - MeasureExpression:**
 - Visible:** True
- Básica (Basic):**
 - Description:**
 - FormatString:**
 - ID:** Valor - FACT Mediciones
 - Name:** Media valor
 - Source:** FACT_Mediciones.valor (Double)

Ahora solo hay que seleccionar los atributos adecuados para el análisis. Dado que solo os interesa la movilidad durante el estado de alarma, se seleccionan los meses adecuados para el análisis.

Dimensión	Jerarquía	Operador	Expresión de filtro
DIM Medicion	Nombre	Igual	{ Movilidad de la población durante el estado de alarma }
DIM Fecha	Tiempo	Igual	{ 2020, 3, 4, 5, 6 }
<Seleccionar dimensión>			

Provincia	Media valor
CEUTA	10,794888888889
MELILLA	13,542333333333
ALMERÍA	11,192333333333
CÁDIZ	12,264444444444
CÓRDOBA	12,128333333333
GRANADA	12,114888888889
HUELVA	11,963666666667
JAÉN	9,353555555556
MÁLAGA	11,698777777778
SEVILLA	13,013888888889
HUESCA	10,802333333333
TERUEL	6,519777777778
ZARAGOZA	14,380555555556
PALMAS,	12,796333333333
SANTA	12,735
CANTABRIA	13,434

Formato y fecha de entrega

La entrega final de esta actividad debe realizarse enviando un único mensaje al buzón de registro de AC del apartado «Evaluación del aula». Adjunto al mensaje se enviará un único archivo en formato Word o PDF con la solución de la PRA3. El nombre del archivo debe ser la composición del nombre de usuario y «_BDA_PRA3». Por ejemplo, si el nombre de usuario es «bantich», el nombre del archivo debe ser «bantich_BDA_PRA3.pdf».

La fecha máxima de entrega es el **02/06/2021** a las 23:59 horas.