



Proyecto de Inteligencia de Negocio

Producción de cine

Profesores

Joaquín Triñanes Fernández
David Mera Pérez

Inteligencia de Negocio

Master Universitario en Tecnologías de Análisis de Datos Masivos: Big Data

Escola Técnica Superior de Enxeñaría (ETSE)

Universidade de Santiago de Compostela

Tabla de contenido

1.- Introducción	5
2.- Diseño y construcción del almacén de datos.....	6
2.1 Diseño del almacén de datos	6
2.2 Creación de los cubos de datos.....	10
3.- Extracción, Transformación y Carga de Datos	12
4.- Análisis de datos con MDX y ROLAP	15
4.1 Consultas MDX	15
4.2 Consultas ROLAP	18
5.- Cuadros de mando e informes.....	22

1.- Introducción

Nuestra productora de cine se encarga de llevar a cabo proyectos de producción de películas, que incluyen la selección del equipo de producción y dirección, que se encarga de la creación de la película, la búsqueda de financiación para la ejecución, los procesos de publicidad y márketing, la distribución y venta y el análisis de los resultados obtenidos.

En este documento nos centraremos en la descripción de algunas herramientas de apoyo a la toma de decisión relacionadas con el proceso de selección del equipo de producción y del análisis de resultados.

Los objetivos generales perseguidos por la productora son los siguientes:

- Maximizar el margen de beneficio asociado a cada película.
- Maximizar la satisfacción de los espectadores con los productos generados.

Para poder alcanzar estos objetivos es fundamental tener éxito en la selección del director (directores) y productor (productores) más apropiados para llevar a cabo cada proyecto. Para realizar este proceso con solvencia es importante conocer:

- La satisfacción de los usuarios con las películas dirigidas por cada equipo de dirección y producidas por cada equipo de producción.
- Los resultados económicos alcanzados en las películas en las que participó cada equipo de dirección y producción.

En general, para conocer el progreso de la productora, es necesario conocer tanto la satisfacción de los usuarios con sus productos como los resultados económicos globales (costes, ingresos y beneficios).

En este documento se describen los procesos de diseño y construcción de algunos recursos de inteligencia de negocio asociados al problema anterior. En concreto, la Sección 2 se centra en el diseño y construcción del almacén de datos que gestiona la información necesaria. Los procesos de extracción, transformación y carga que permiten la creación y actualización del almacén de datos se describen en la Sección 3. La Sección 4 muestra algunas consultas de interés, e incluye tanto consultas MDX sobre los cubos de datos generados como consultas ROLAP sobre el modelo relacional del almacén de datos. Por último, la Sección 5 describe la creación de algunas interfaces de usuario, que incluyen cuadros de mando e informes.

2.- Diseño y construcción del almacén de datos

2.1 Diseño del almacén de datos

Teniendo en cuenta las fuentes de datos disponibles y los objetivos perseguidos, se seleccionan los siguientes procesos. El primer proceso tiene que ver con la producción y distribución de las películas. Cada mes, se recibe un conjunto de datos relacionado con las películas producidas por las productoras principales a nivel mundial. Este conjunto de datos tiene un registro por cada película producida, y de cada película se proporcionan los siguientes datos:

- Un identificador entero único para cada película generado por el productor de los datos.
- La fecha de emisión de la película.
- El título de la película.
- El presupuesto y los ingresos generados por la película.
- El listado de los directores de la película. De cada director se proporciona un identificador entero único y el nombre.
- El listado de los productores de la película. De cada productor se proporciona un identificador entero único y el nombre del productor.
- El listado de productoras involucradas. Este listado se restringe a las productoras principales a nivel mundial.

El formato en el que se proporcionan estos datos cada mes es XML y en la Figura 1 se puede ver un ejemplo de un par de películas del mes de junio de 2004.

```
<peliculas>
- <pelicula id="9890" emision="2004-06-10">
  <titulo>The Stepford Wives</titulo>
  <presupuesto>90000000</presupuesto>
  <ingresos>102000000</ingresos>
  - <directores>
    <director id="7908" nombre="Frank Oz"/>
  </directores>
  - <productores>
    <productor id="2997" nombre="Scott Rudin"/>
    <productor id="14929" nombre="Edgar J. Scherick"/>
    <productor id="58470" nombre="Donald De Line"/>
    <productor id="60010" nombre="Gabriel Grunfeld"/>
    <productor id="60013" nombre="Leslie J. Converse"/>
  </productores>
  - <productoras>
    <productora id="4" nombre="Paramount Pictures"/>
  </productoras>
</pelicula>
- <pelicula id="8193" emision="2004-06-11">
  <titulo>Napoleon Dynamite</titulo>
  <presupuesto>400000</presupuesto>
  <ingresos>46118097</ingresos>
  - <directores>
    <director id="53925" nombre="Jared Hess"/>
  </directores>
  - <productores>
    <productor id="53941" nombre="Sean Covey"/>
    <productor id="53942" nombre="Jeremy Coon"/>
    <productor id="1459908" nombre="Chris Wyatt"/>
  </productores>
  - <productoras>
    <productora id="4" nombre="Paramount Pictures"/>
  </productoras>
</pelicula>
```

Figura1: Ejemplo de archivo .XML de películas proporcionado cada mes.

El segundo proceso seleccionado tiene que ver con la monitorización de la satisfacción de los usuarios con las películas.

Cada mes también se proporciona un archivo .CSV con las votaciones que han realizado los usuarios sobre las películas de las principales productoras. Cada fila del archivo proporciona el voto realizado por un usuario (se proporciona un identificador numérico, con lo que los datos son anónimos), en un determinado instante temporal, para una película. En la Figura 2 se muestra un ejemplo con las primeras votaciones correspondientes al mes de septiembre de 2001.

```
1 pelicula,usuario,instante,voto
2 12103,1,2001-09-28 04:25:24,3.5
3 12103,2,2001-09-28 02:25:01,3.5
4 12103,6,2001-09-28 00:28:04,5.0
5 12103,16,2001-09-28 01:51:17,3.5
6 12103,47,2001-09-28 00:29:41,5.0
7 12103,50,2001-09-28 00:00:06,4.5
8 12103,110,2001-09-28 00:29:24,3.5
9 12103,111,2001-09-28 00:30:48,4.5
10 12103,147,2001-09-28 02:10:00,4.0
11 12103,150,2001-09-28 00:21:45,4.5
12 12103,160,2001-09-28 01:53:46,3.0
13 12103,165,2001-09-28 04:32:19,3.5
14 12103,196,2001-09-30 04:20:15,1.5
15 12103,215,2001-09-28 02:06:37,4.5
16 12103,223,2001-09-30 03:56:04,4.0
17 12103,253,2001-09-28 01:49:36,1.5
18 12103,260,2001-09-28 00:17:15,4.0
```

Figura2: Ejemplo de archivo .CSV de votaciones de usuarios.

La granularidad seleccionada para los datos almacenados de cada uno de los procesos anteriores es de un mes. Cada mes, con los datos obtenidos se agregarán datos de satisfacción y resultados.

Las dimensiones seleccionadas son las siguientes:

- **Tiempo:** Dada la granularidad seleccionada para los datos, la dimensión temporal de los mismos se representa mediante dos niveles, un nivel de año y un nivel de mes.
- **Director:** Cada elemento de esta dimensión representa al conjunto de directores de una película. Los identificadores y nombres de los directores se representan en atributos de tipo String.
- **Productor:** Cada elemento de esta dimensión representa al conjunto de productores involucrados en una película. Los identificadores y los nombres de los productores se representan en atributos de tipo String.
- **Productora:** Cada elemento de esta dimensión representa a una de las principales productoras existentes a nivel internacional.

Los hechos almacenados en el almacén de datos son los siguientes:

- **Finanzas:** Se almacena el coste e ingresos producidos por las películas de las que se tienen datos para cada combinación de las siguientes dimensiones: tiempo (mes de emisión de las películas), director, productor y productora. En el caso de las productoras, el coste e ingresos de cada película se divide a partes iguales entre todas las productoras involucradas.

- **Satisfacción de usuarios:** Se almacena el número de votos y la valoración media para cada combinación del siguiente conjunto de dimensiones: tiempo (mes de emisión de las películas), tiempo (mes en el que se han efectuado las votaciones), director, productor y productora. En el caso de las productoras, el número de votos y la media de las valoraciones se replica en el almacén para cada productora involucrada en cada película.

En la Figura 3, se proporciona el modelo conceptual con dos estrellas utilizado para almacenar los datos.

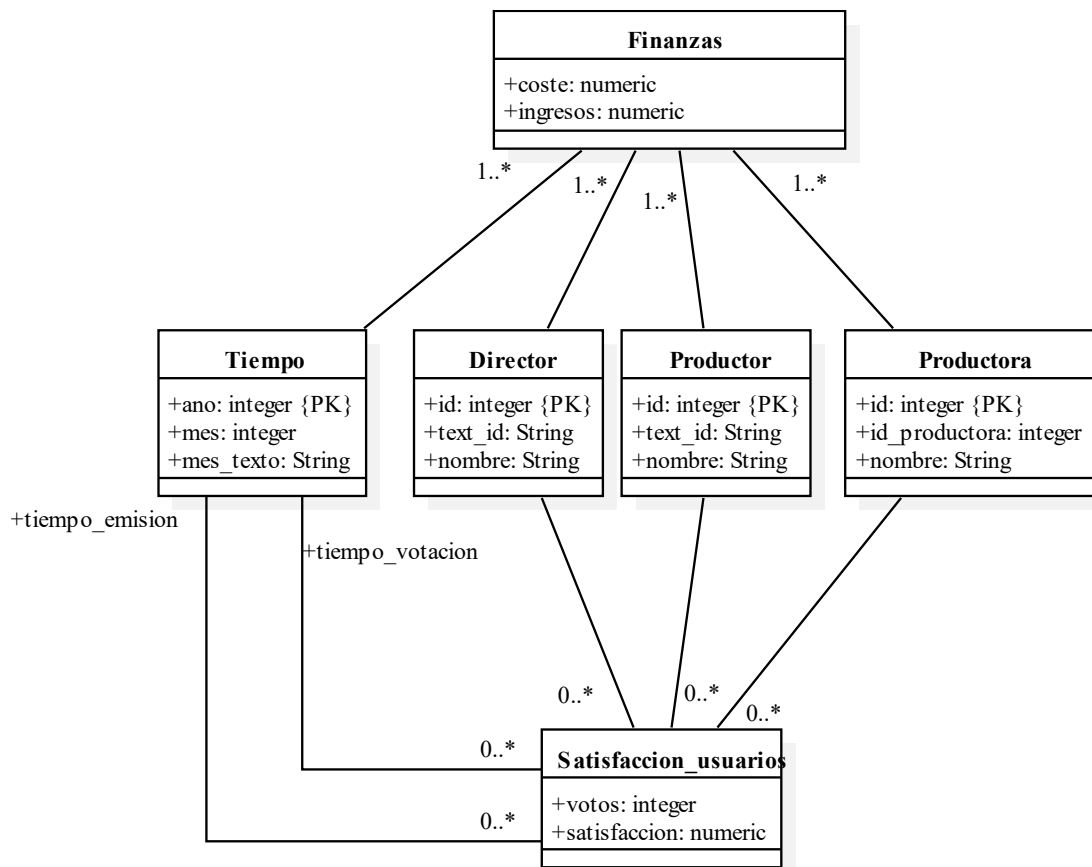


Figura 3: Modelo conceptual del almacén de datos.

Los detalles del diccionario de datos se proporcionan en la siguiente tabla.

Dimensiones		
-------------	--	--

Tiempo		
Atributo	Tipo	Descripción
id (cp)	Integer	Identificador autoincremental generado por el almacén cada vez que se inserta un nuevo mes.
ano	Integer	Número correspondiente al año
mes	Integer	Número correspondiente al mes dentro del año.
mes_texto	String	Texto con el nombre del mes (Enero, Febrero, etc.)

Director		
Atributo	Tipo	Descripción

id (cp)	Integer	Identificador autoincremental generado por el almacén cada vez que se inserta un nuevo equipo de dirección de una película.
text_id	String	Identificadores de los directores del equipo proporcionados por el proveedor de datos, codificados en un String separados por guiones (-).
nombre	String	Nombres de los directores del equipo codificados en un string separados por comas, con un espacio después de cada coma (,).

Productor		
Atributo	Tipo	Descripción
id (cp)	Integer	Identificador autoincremental generado por el almacén cada vez que se inserta un nuevo equipo de producción de una película.
text_id	String	Identificadores de los productores del equipo proporcionados por el proveedor de datos, codificados en un String separados por guiones (-).
nombre	String	Nombres de los productores del equipo codificados en un string separados por comas, con un espacio después de cada coma (,).

Productora		
Atributo	Tipo	Descripción
id (cp)	Integer	Identificador autoincremental generado por el almacén cada vez que se inserta una nueva productora de una película.
id_productora	Integer	Identificador de la productora proporcionado por el proveedor de datos.
nombre	String	Nombre de la productora.

Hechos		
--------	--	--

Finanzas		
Medida	Tipo	Descripción
coste	Numeric	Suma del coste de todas las películas vinculadas a cada combinación de las dimensiones. El coste de cada película se divide a partes iguales entre las productoras involucradas.
ingresos	Numeric	Suma de los ingresos de todas las películas vinculadas a cada combinación de las dimensiones. Los ingresos de cada película se dividen a partes iguales entre las productoras involucradas.

Satisfacción_usuarios		
Medida	Tipo	Descripción
votos	Integer	Número total de votos emitido para cada combinación de las dimensiones. Los votos de cada película se repiten para todas las productoras involucradas.
satisfacción	Numeric	Nivel de satisfacción (entre 0 y 5 con un decimal) medio expresado por los usuarios para cada combinación de los valores de las dimensiones. El valor de satisfacción de cada usuario con cada película se asocia también a cada productora involucrada.

2.2 Creación de los cubos de datos

Utilizando como base las estructuras de datos relacionales descritas en la subsección anterior se crea un esquema multidimensional con dos cubos. Dado que las dimensiones se comparten entre los dos cubos, se definen primero a nivel de esquema y después se reutilizan a nivel de cubo.

❖ Esquema: **ProduccionCine**

- Dimensión: **director**. Dimensión de tipo estándar.
 - Jerarquía: *jerarquiadirector*. Definida por el atributo *id*.
 - Nivel: *niveldirector*. Nivel de tipo regular definido por el atributo *nombre*
 - Tabla: *directores*
- Dimensión: **productor**. Dimensión de tipo estándar.
 - Jerarquía: *jerarquiaproductor*. Definida por el atributo *id*.
 - Nivel: *niveldirector*. Nivel de tipo regular definido por el atributo *nombre*.
 - Tabla: *productores*.
- Dimensión: **productora**. Dimensión de tipo estándar.
 - Jerarquía: *jerarquiaproductora*. Definida por el atributo *id*.
 - Nivel: *nivelproductora*. Nivel de tipo regular definido por el atributo *nombre*.
 - Tabla: *productores*.
- Dimensión: **tiempo**. Dimensión de tipo temporal
 - Jerarquía: *jerarquiat tiempo*. Definida por el atributo *id*.
 - Nivel: *anno*. Nivel de tipo *TimeYears* definido por el atributo *ano*.
 - Nivel: *mes*. Nivel de tipo *TimeMonths* definido por el atributo *mes*. Se utiliza el atributo *mes_texto* para nombrar a los elementos de este nivel.
 - Tabla: *tiempo*.
- Cubo: **satisfaccion**.
 - Tabla: *satisfaccion_usuarios*
 - Dimensiones
 - Dimensión usada: **director**. Se referencia la dimensión *director*. Se utiliza el atributo *director* como clave foránea.
 - Dimensión usada: **productor**. Se referencia la dimensión *productor*. Se utiliza el atributo *productor* como clave foránea.
 - Dimensión usada: **productora**. Se referencia la dimensión *productora*. Se utiliza el atributo *productora* como clave foránea.
 - Dimensión usada: **tiempoemision**. Se referencia la dimensión *tiempo*. Se utiliza el atributo *tiempo_emision* como clave foránea.
 - Dimensión usada: **tiempovotacion**. Se referencia la dimensión *tiempo*. Se utiliza el atributo *tiempo_votacion* como clave foránea.
 - Medidas
 - Medida: **votos**. Se agrega el atributo *votos* usando la función SUM.
 - Medida: **satisfaccion**. Se agrega el atributo *satisfacción* usando la función AVG.
- Cubo: **finanzas**
 - Tabla: *finanzas*
 - Dimensiones
 - Dimensión usada: **director**. Se referencia la dimensión *director*. Se utiliza el atributo *director* como clave foránea.

- Dimensión usada: **productor**. Se referencia la dimensión *productor*. Se utiliza el atributo *productor* como clave foránea.
- Dimensión usada: **productora**. Se referencia la dimensión *productora*. Se utiliza el atributo *productora* como clave foránea.
- Dimensión usada: **tiempoemision**. Se referencia la dimensión *tiempo*. Se utiliza el atributo *tiempo* como clave foránea.
- Medidas
 - Medida: **coste**. Se agrega el atributo *coste* usando la función SUM.
 - Medida: **ingresos**. Se agrega el atributo *ingresos* usando la función SUM.
 - Medida calculada: **beneficio**. Se genera un nuevo valor para la dimensión Measures usando la expresión MDX "[Measures].[ingresos]-[Measures].[coste]"

3.- Extracción, Transformación y Carga de Datos

En esta sección se describen los trabajos y transformaciones definidos en Pentaho Data Integration (PDI) para mantener el almacén de datos actualizado a partir de los archivos .xml y .csv proporcionados por las fuentes de datos cada mes.

Para definir la comunicación con los proveedores de datos, para simplificar el proceso, asumiremos que cada mes, en una determinada carpeta de la máquina en la que se ejecuta el ETL se colocarán los dos archivos de entrada, un archivo .xml con los datos de las películas, y un archivo .csv con los datos de las votaciones de los usuarios. Los nombres de estos archivos seguirán estos patrones (YYYY y MM representan el año y el mes):

- peliculas-YYYY-MM.xml
- votos-YYYY-MM.csv

A pesar de asumir que cada mes se colocará solo un archivo en la carpeta, el proceso ETL importará los datos de todos los archivos con los patrones anteriores presentes en la carpeta. Por lo tanto, después de haber realizado la carga de datos será necesario borrar los archivos de la carpeta para no volver a cargarlos en la siguiente ejecución. Este proceso de borrado no se realizará como parte del ETL.

El proceso ETL se define como un trabajo (job) en PDI. El flujo de transformaciones que forman parte del trabajo puede verse en la Figura 4.

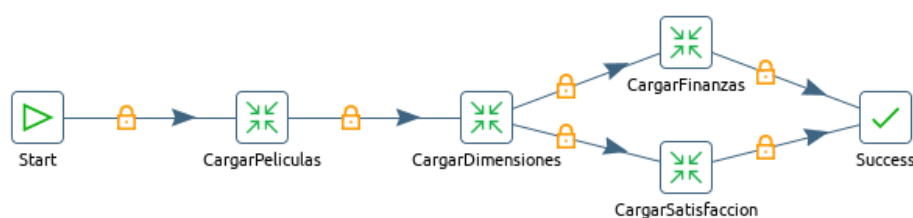


Figura 4: Flujo principal de transformaciones del proceso ETL.

En primer lugar, la transformación “CargarPelículas” carga en una tabla disponible en el área de ensayo del ETL información necesaria de cada película disponible en los archivos .xml. Esta tabla servirá de apoyo para el proceso de vinculación de los datos de cada película a cada registro de satisfacción de usuario. El esquema de esta tabla es el siguiente:

```
pelicula_productora(pelicula,  
                    fecha_emision,  
                    id_productora,  
                    text_id_director,  
                    text_id_productor)
```

Cada fila almacena el identificador de la película, la fecha de emisión, el identificador de la productora, los identificadores de los directores (separados por guiones), los identificadores de los productores (separados por guiones). El flujo de pasos de esta transformación puede verse en la Figura 5. Primero, se obtienen los datos de las películas, sus directores, sus productores y sus productoras en conjuntos de datos normalizados. Estos conjuntos de datos se ordenan, agrupan y combinan para generar un único conjunto de datos, utilizando pasos de ordenación, agrupación (Group By) y Join (merge join). Después, se utiliza el paso “if field value is null” para asignar directores y productores especiales a las películas que no tengan director o productor.

Se usarán identificadores 0 y nombres “Sin director” y “Sin productor” para estos casos. Para terminar el flujo, los datos se copian al trabajo para poder ser utilizados por transformaciones posteriores y se insertan en la tabla arriba mencionada.

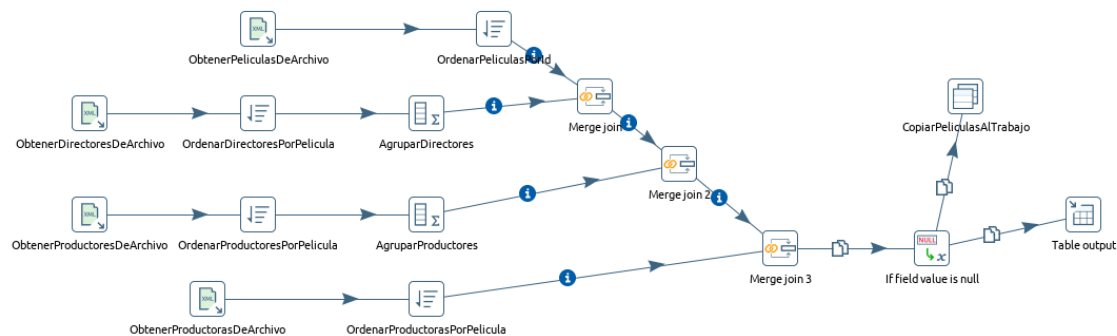


Figura 5: Flujo de pasos de la transformación “CargarPelículas”.

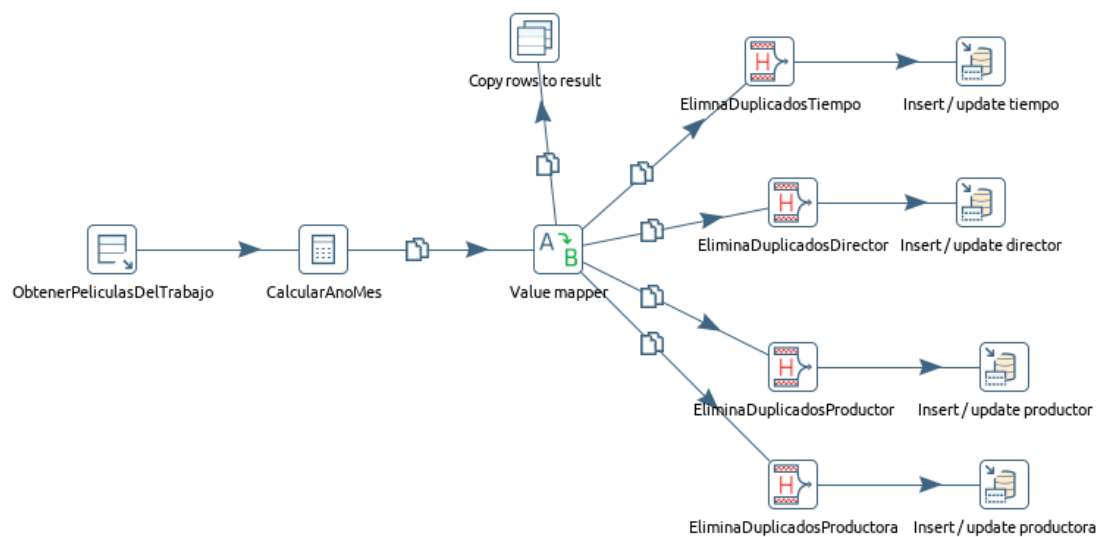


Figura 6: Flujo de pasos de la transformación “CargarDimensiones”.

En segundo lugar, se cargan los datos de las dimensiones en las tablas del almacén de datos (“CargarDimensiones”). Los pasos de esta transformación pueden verse en la Figura 6. En primer lugar se obtienen los datos del trabajo que han sido generados por la transformación anterior. Después se calculan los atributos “año” y “mes” a partir de la fecha de emisión de cada película. El paso “Value mapper” genera los nombres de cada mes a partir de su número (Enero, Febrero, etc.). El conjunto de datos se escribe en el trabajo para que pueda usarlo la transformación siguiente que genera los datos de finanzas. Para cada una de las dimensiones, se proyectan sus columnas del conjunto de datos generado, se eliminan duplicados y se ejecuta una operación de inserción / modificación. Esta operación inserta datos si la clave no existe en la tabla y los modifica en caso contrario.

Una vez cargados los datos de las dimensiones, los datos de las dos tablas de hechos pueden cargarse ya en paralelo. Para poder cargar los datos de finanzas, primero es necesario dividir los datos de coste e ingresos por el número de productoras de cada película. Después se debe obtener el identificador (atributo id) de cada dimensión consultando las tablas de dimensiones (recordar que estos identificadores los genera la base de datos cuando se insertan las filas de las

dimensiones). Se ordena el conjunto de datos por las dimensiones para poder agrupar los datos. Finalmente se insertan las filas generadas. Este flujo puede verse en la Figura 7.

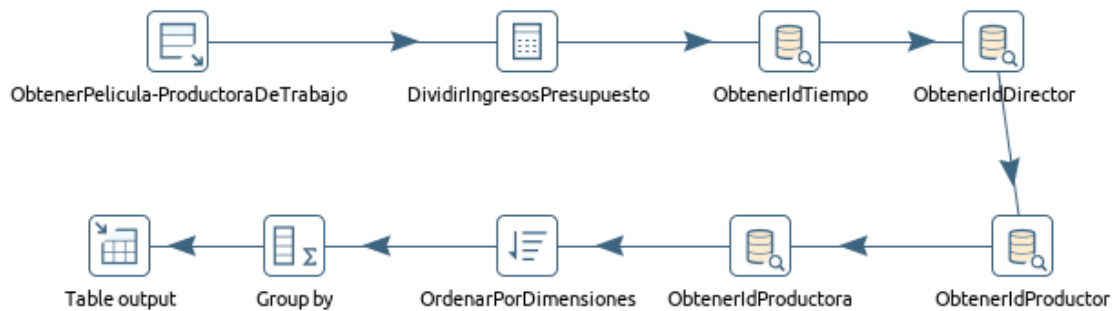


Figura 7: Flujo de pasos de la transformación "CargarFinanzas".

Para poder cargar los datos de satisfacción de usuarios, debemos combinar los datos de las películas con los datos obtenidos de los archivos .CSV. Para esto, primero cargamos todos los archivos .CSV, usando dos pasos, primero uno que lee todos los nombres de archivo, y después otro que lee los datos. Después de realizar el join con la tabla pelicula_productora, usando el paso "Database join", calculamos los años y meses de emisión y votación, y procedemos como con el proceso de carga de finanzas, obteniendo los identificadores de dimensión, ordenando, agrupando y almacenando en la tabla el resultado. Este flujo puede verse en la Figura 8.

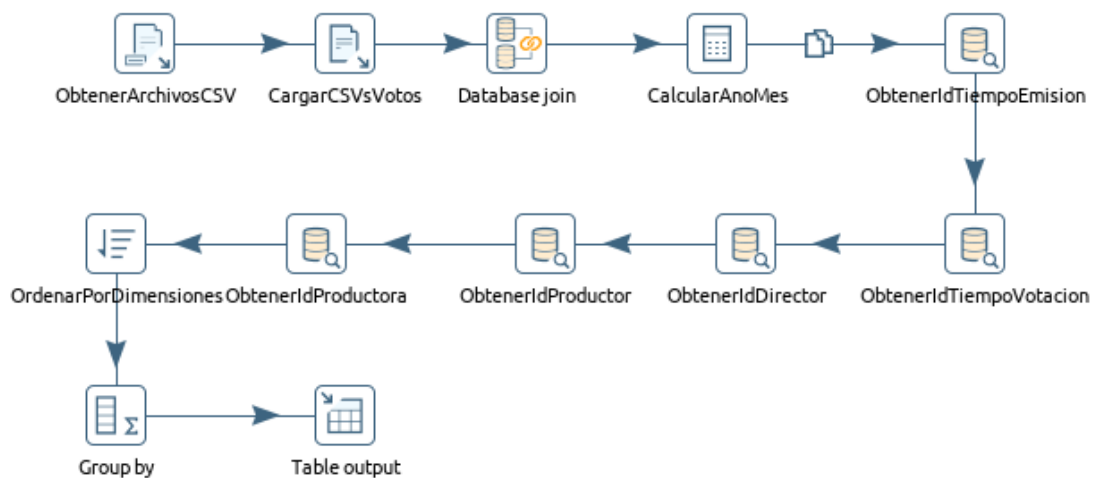


Figura 8: Flujo de pasos de la transformación "CargarSatisfaccion".

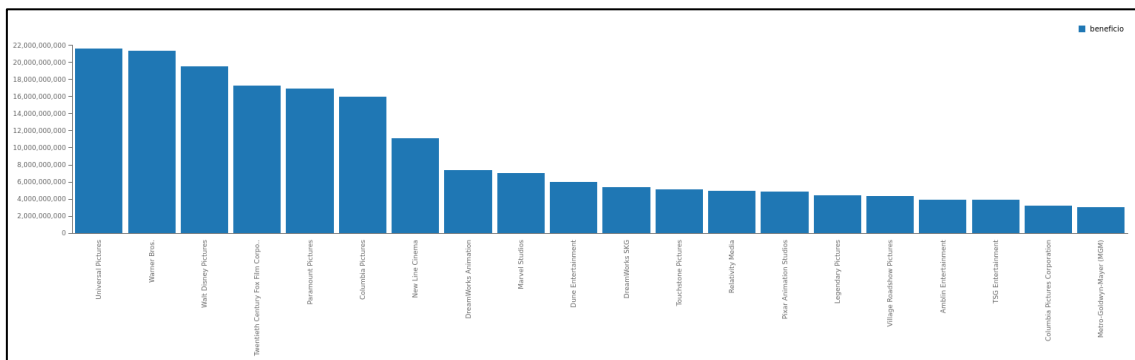
4.- Análisis de datos con MDX y ROLAP

4.1 Consultas MDX

Para ilustrar el uso de los cubos creados en la resolución de procesos de análisis se han resuelto las siguientes consultas usando el lenguaje MDX. Para cada consulta se proporciona su enunciado, el código MDX y una muestra del resultado.

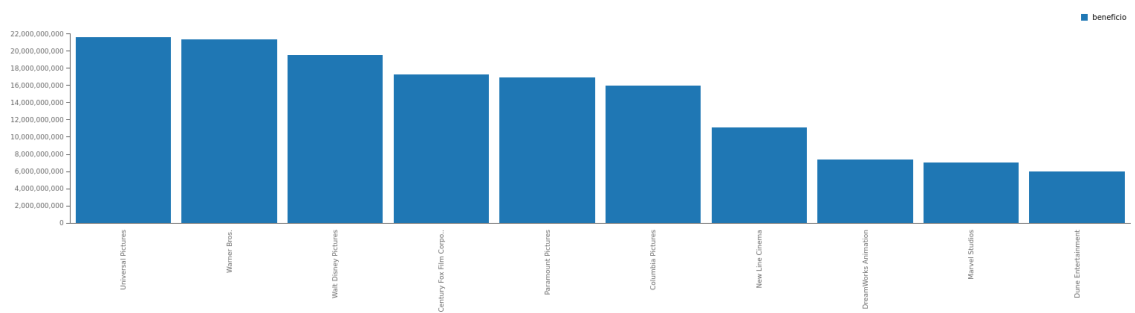
1.- Obtener el beneficio generado por cada productora. Ordena el resultado por beneficio de forma descendente.

```
WITH
SET [~ROWS] AS
    Order([productora.jerarquiaproductora].[nivelproductora].Members),
[Measures].[beneficio], DESC)
SELECT
NON EMPTY {[Measures].[beneficio]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [finanzas]
```



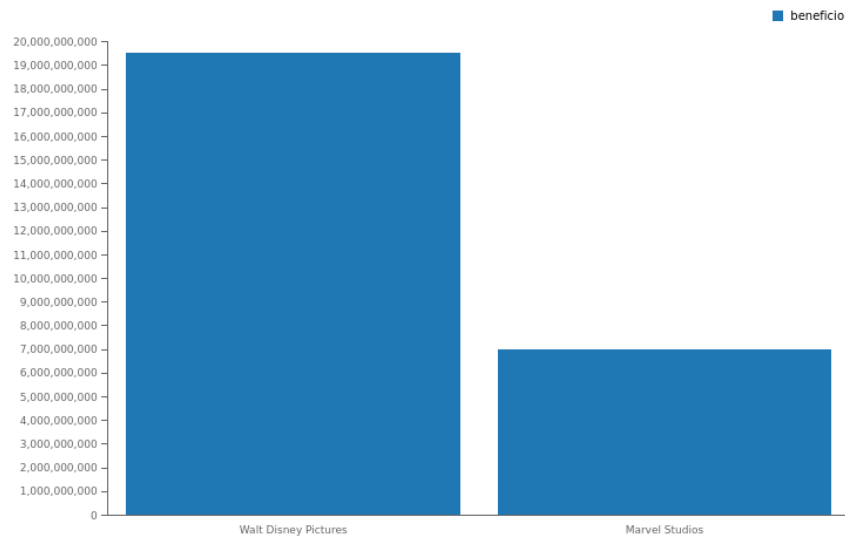
2.- Obtener el beneficio generado por cada productora. Ordena el resultado por beneficio de forma descendente. Limita el resultado a solo las 10 productoras que generan más beneficios

```
WITH
SET [~ROWS] AS
    Order(TopCount([productora.jerarquiaproductora].[nivelproductora].Members),
10, [Measures].[beneficio]), [Measures].[beneficio], DESC)
SELECT
NON EMPTY {[Measures].[beneficio]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [finanzas]
```



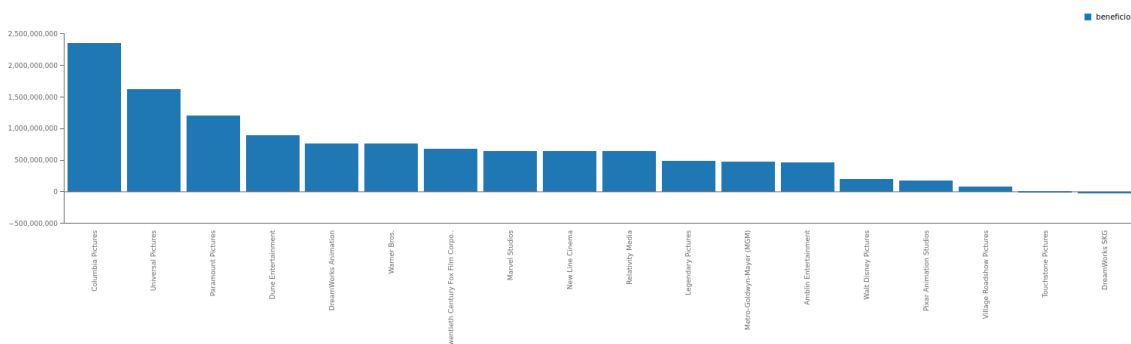
3.- Obtener el beneficio generado por las productoras 'Walt Disney Pictures', 'Columbia Pictures'. Ordena el resultado por beneficio de forma descendente.

```
WITH
SET [~ROWS] AS
    Order({[productora.jerarquiaproductora].[Walt Disney Pictures],
[productora.jerarquiaproductora].[Marvel Studios]}, [Measures].[beneficio],
DESC)
SELECT
NON EMPTY {[Measures].[beneficio]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [finanzas]
```



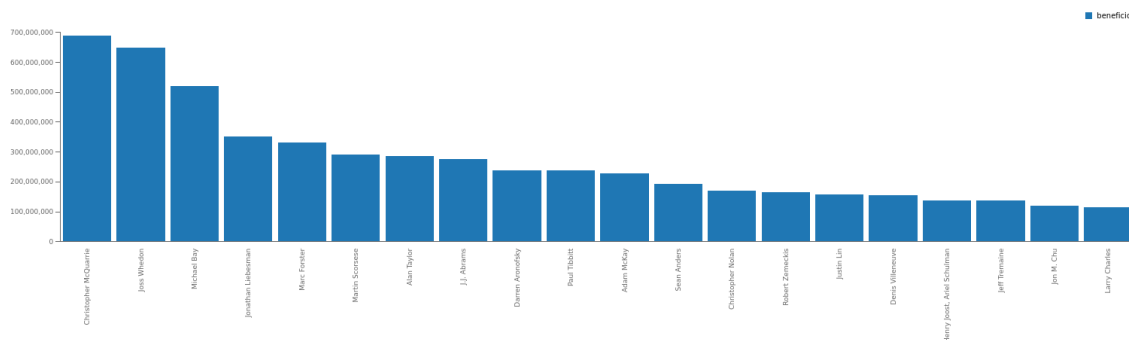
4.- Mostrar los datos de beneficio de cada productora, pero solo para el año 2012

```
WITH
SET [~FILTER] AS
    {[tiempoemision.jerarquiatiempo].[2012]}
SET [~ROWS] AS
    Order({[productora.jerarquiaproductora].[nivelproductora].Members},
[Measures].[beneficio], DESC)
SELECT
NON EMPTY {[Measures].[beneficio]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [finanzas]
WHERE [~FILTER]
```



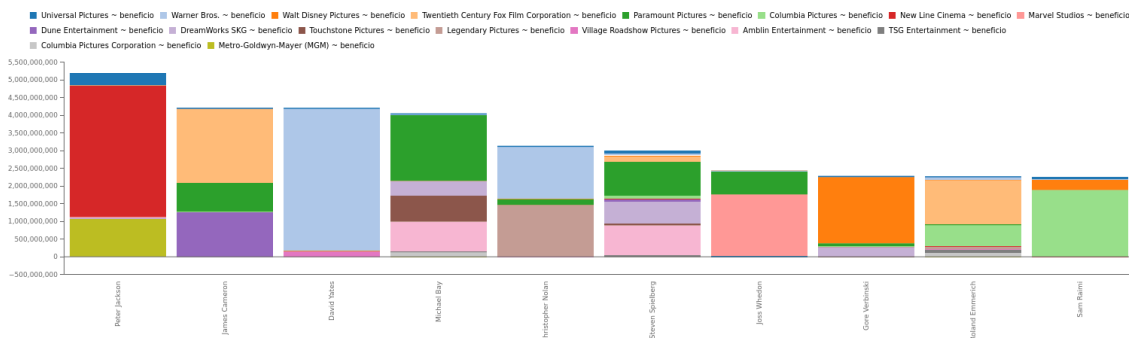
5.- Mostrar los datos de beneficio de los 20 directores con más beneficio, pero solo para los últimos 5 años y para la productora 'Paramount Pictures'

```
WITH
SET [~ROWS] AS
    Order(TopCount([director.jerarquiadirector].[niveldirector].Members),
    20, [Measures].[beneficio]), [Measures].[beneficio], DESC)
SELECT
NON EMPTY {[Measures].[beneficio]} ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [finanzas]
WHERE CrossJoin([productora.jerarquiaproductora].[Paramount Pictures]),
{[tiempoemision.jerarquiatiempo].[2012],
[tiempoemision.jerarquiatiempo].[2013],
[tiempoemision.jerarquiatiempo].[2014],
[tiempoemision.jerarquiatiempo].[2015],
[tiempoemision.jerarquiatiempo].[2016]}}
```



6.- Mostrar datos de beneficio para combinaciones de productoras y directores (los 10 con más beneficio)

```
WITH
SET [~COLUMNS] AS
    Order([productora.jerarquiaproductora].[nivelproductora].Members),
    [Measures].[beneficio], DESC)
SET [~ROWS] AS
    Order(TopCount([director.jerarquiadirector].[niveldirector].Members), 10,
    [Measures].[beneficio]), [Measures].[beneficio], DESC)
SELECT
NON EMPTY CrossJoin([~COLUMNS], {[Measures].[beneficio]}) ON COLUMNS,
NON EMPTY [~ROWS] ON ROWS
FROM [finanzas]
```



4.2 Consultas ROLAP

Con las siguientes consultas ilustramos el uso de la extensión del SQL para OLAP para solucionar problemas de análisis.

1.- Obtener para las productoras 'New Line Cinema' y 'Paramount Pictures' la evolución mensual de los beneficios entre principios de 2014 y finales de 2016

```
select p.nombre as productora, t.ano as ano, t.mes as mes,
       sum(ingresos-coste) as beneficio
from dw.finanzas f, dw.productoras p, dw.tiempo t
where f.productora = p.id
      and f.tiempo = t.id
      and p.nombre in ('Paramount Pictures', 'New Line Cinema')
      and t.ano between 2014 and 2016
group by p.id, t.ano, t.mes
order by productora, ano, mes
```

productora	123 ano	123 mes	123 beneficio
New Line Cinema	2.014	7	80.525.432
New Line Cinema	2.014	8	89.238.518
New Line Cinema	2.014	10	248.773.813
New Line Cinema	2.014	11	130.588.862
New Line Cinema	2.014	12	353.009.894
New Line Cinema	2.015	5	125.723.678
New Line Cinema	2.015	7	115.948.598
New Line Cinema	2.015	8	57.878.330
New Line Cinema	2.015	11	45.522.527
New Line Cinema	2.016	1	74.343.513
New Line Cinema	2.016	4	40.596.378
New Line Cinema	2.016	5	140.085.004
New Line Cinema	2.016	6	177.458.808
New Line Cinema	2.016	12	25.708.010
Paramount Pictures	2.014	1	-9.450.893
Paramount Pictures	2.014	3	237.637.473
Paramount Pictures	2.014	6	440.702.548

2.- Obtener para las productoras 'New Line Cinema' y 'Paramount Pictures' la evolución mensual de los beneficios entre principios de 2014 y finales de 2016. Añade también los datos agregados por productora y año, por productora y el total

```
select p.nombre as productora, t.ano as ano, t.mes as mes,
       sum(ingresos-coste) as beneficio
from dw.finanzas f, dw.productoras p, dw.tiempo t
where f.productora = p.id
      and f.tiempo = t.id
      and p.nombre in ('Paramount Pictures', 'New Line Cinema')
      and t.ano between 2014 and 2016
group by rollup (p.nombre, t.ano, t.mes)
order by productora, ano, mes
```

productora	123 ano	123 mes	123 beneficio
New Line Cinema	2.014	7	80.525.432
New Line Cinema	2.014	8	89.238.518
New Line Cinema	2.014	10	248.773.813
New Line Cinema	2.014	11	130.588.862
New Line Cinema	2.014	12	353.009.894
New Line Cinema	2.014	[NULL]	902.136.519
New Line Cinema	2.015	5	125.723.678
New Line Cinema	2.015	7	115.948.598
New Line Cinema	2.015	8	57.878.330
New Line Cinema	2.015	11	45.522.527
New Line Cinema	2.015	[NULL]	345.073.133
New Line Cinema	2.016	1	74.343.513
New Line Cinema	2.016	4	40.596.378
New Line Cinema	2.016	5	140.085.004
New Line Cinema	2.016	6	177.458.808
New Line Cinema	2.016	12	25.708.010
New Line Cinema	2.016	[NULL]	458.191.713
New Line Cinema	[NULL]	[NULL]	1.705.401.365
Paramount Pictures	2.014	1	-9.450.893
Paramount Pictures	2.014	3	237.637.473
Paramount Pictures	2.014	6	440.702.548

3.- Generar un cubo de datos para las productoras 'Paramount Pictures' y 'New Line Cinema' para la medida del beneficio de las películas, con las dimensiones de productora, año y mes. Usa los datos solo de los años 2014, 2015 y 2016

```
select p.nombre as productora, t.ano as ano, t.mes as mes,
       sum(ingresos-coste) as beneficio
from dw.finanzas f, dw.productoras p, dw.tiempo t
where f.productora = p.id
      and f.tiempo = t.id
      and p.nombre in ('Paramount Pictures', 'New Line Cinema')
      and t.ano between 2014 and 2016
group by cube (p.nombre, t.ano, t.mes)
order by productora, ano, mes
```

productora	123 ano	123 mes	123 beneficio
New Line Cinema	2.014	7	80.525.432
New Line Cinema	2.014	8	89.238.518
New Line Cinema	2.014	10	248.773.813
New Line Cinema	2.014	11	130.588.862
New Line Cinema	2.014	12	353.009.894
New Line Cinema	2.014	[NULL]	902.136.519
New Line Cinema	2.015	5	125.723.678
New Line Cinema	2.015	7	115.948.598
New Line Cinema	2.015	8	57.878.330
New Line Cinema	2.015	11	45.522.527
New Line Cinema	2.015	[NULL]	345.073.133
New Line Cinema	2.016	1	74.343.513
New Line Cinema	2.016	4	40.596.378
New Line Cinema	2.016	5	140.085.004
New Line Cinema	2.016	6	177.458.808
New Line Cinema	2.016	12	25.708.010
New Line Cinema	2.016	[NULL]	458.191.713
New Line Cinema	[NULL]	1	74.343.513
New Line Cinema	[NULL]	4	40.596.378
New Line Cinema	[NULL]	5	265.808.682
New Line Cinema	[NULL]	6	177.458.808
New Line Cinema	[NULL]	7	196.474.030
New Line Cinema	[NULL]	8	147.116.848
New Line Cinema	[NULL]	10	248.773.813
New Line Cinema	[NULL]	11	176.111.389
New Line Cinema	[NULL]	12	378.717.904
New Line Cinema	[NULL]	[NULL]	1.705.401.365
Paramount Pictures	2.014	1	-9.450.893
Paramount Pictures	2.014	3	237.637.473

4.- Obtener para las productoras 'New Line Cinema' y 'Paramount Pictures' la evolución mensual de los beneficios entre principios de 2014 y finales de 2016. Crea una vista "beneficio_mensual1" con el resultado.

```
create view beneficio_mensual1 as
select p.nombre as productora, t.ano as ano, t.mes as mes,
       sum(ingresos-coste) as beneficio
from dw.finanzas f, dw.productoras p, dw.tiempo t
where f.productora = p.id
      and f.tiempo = t.id
      and p.nombre in ('Paramount Pictures', 'New Line Cinema')
      and t.ano between 2014 and 2016
group by p.id, t.ano, t.mes
order by productora, ano, mes
```

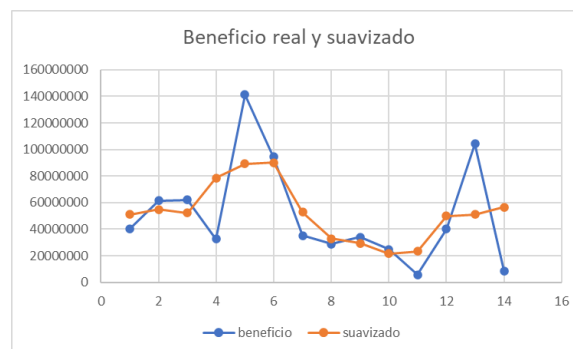
5.- Para cada una de las productoras obtener el porcentaje que representa el beneficio de cada mes en el beneficio anual. Obtener también el porcentaje que representa cada mes en el beneficio total de cada productora

```
select *,
       sum(beneficio) over (partition by productora, ano) as beneficio_anual,
       (beneficio / sum(beneficio)
        over (partition by productora, ano))*100 as porcentaje_beneficio_anual,
       (beneficio / sum(beneficio)
        over (partition by productora))*100 as porcentaje_beneficio_productora
from beneficio_mensual1
```

ABC productora	123 ano	123 mes	123 beneficio	123 beneficio_anual	123 porcentaje_beneficio_anual	123 porcentaje_beneficio_productora
New Line Cinema	2.014	7	40.262.716	337.707.347	11,9223689854	5,6428587884
New Line Cinema	2.014	8	61.587.969	337.707.347	18,2370829498	8,6316137275
New Line Cinema	2.014	10	62.193.453	337.707.347	18,4163754661	8,7164728987
New Line Cinema	2.014	11	32.459.252	337.707.347	9,6116511199	4,5491957228
New Line Cinema	2.014	12	141.203.957	337.707.347	41,8125214788	19,7898718437
New Line Cinema	2.015	5	94.292.758	192.476.316	48,9892782445	13,2152216995
New Line Cinema	2.015	7	35.102.498	192.476.316	18,2373077008	4,9196492193
New Line Cinema	2.015	8	28.939.165	192.476.316	15,0351823026	4,0558521077
New Line Cinema	2.015	11	34.141.895	192.476.316	17,7382317521	4,785019775
New Line Cinema	2.016	1	24.781.171	183.332.615	13,5170553259	3,4731052064
New Line Cinema	2.016	4	5.566.327	183.332.615	3,03619026	0,7801261403
New Line Cinema	2.016	5	40.024.287	183.332.615	21,8315148126	5,6094427323
New Line Cinema	2.016	6	104.391.494	183.332.615	56,9410380144	14,6305693673
New Line Cinema	2.016	12	8.569.336	183.332.615	4,6742015871	1,201000771
Paramount Pictures	2.014	1	-1.575.149	426.826.641	-0,3690371801	-0,1734299183
Paramount Pictures	2.014	3	59.409.368	426.826.641	13,9188518928	6,5411982209
Paramount Pictures	2.014	6	110.175.637	426.826.641	25,8127366984	12,1307582457

6.- Suaviza los valores del beneficio mensual realizando la media de cada valor con el valor anterior y siguiente

```
select *,
       avg(beneficio)
       over (partition by productora
            order by ano, mes
            rows between 1 preceding and 1 following) as beneficio_suavizado
from beneficio_mensual1
```



7.- Obtener para cada productora y mes, la posición que ocupa ese mes (ordenando por beneficio) en todos los meses de ambas productoras y en la productora actual, el percentil de ese mes en la distribución de beneficios de la productora y la ratio de incremento de ese mes respecto al mes anterior.

```
select *,
    rank() over (order by beneficio desc) as ranking_total,
    rank()
        over (partition by productora order by beneficio desc) as ranking_productora,
    cume_dist() over (partition by productora order by ano, mes) as percentil,
    beneficio / lag (beneficio, 1, null)
        over (partition by productora order by ano, mes) as ratio_incremento_beneficio
from beneficio_mensual
order by productora, ano, mes
```

ABC productora	123 ano	123 mes	123 beneficio	123 ranking_total	123 ranking_productora	123 percentil	123 ratio_incremento_beneficio
New Line Cinema	2.014	7	40.262.716	14	6	0,0714285714	[NULL]
New Line Cinema	2.014	8	61.587.969	12	5	0,1428571429	1,5296526196
New Line Cinema	2.014	10	62.193.453	11	4	0,2142857143	1,0098312058
New Line Cinema	2.014	11	32.459.252	21	10	0,2857142857	0,5219078606
New Line Cinema	2.014	12	141.203.957	1	1	0,3571428571	4,3501913414
New Line Cinema	2.015	5	94.292.758	8	3	0,4285714286	0,6677770227
New Line Cinema	2.015	7	35.102.498	18	8	0,5	0,3722714103
New Line Cinema	2.015	8	28.939.165	22	11	0,5714285714	0,824418963
New Line Cinema	2.015	11	34.141.895	19	9	0,6428571429	1,1797816212
New Line Cinema	2.016	1	24.781.171	24	12	0,7142857143	0,7258288094
New Line Cinema	2.016	4	5.566.327	28	14	0,7857142857	0,2246192079
New Line Cinema	2.016	5	40.024.287	15	7	0,8571428571	7,1904304221
New Line Cinema	2.016	6	104.391.494	4	2	0,9285714286	2,6082037139
New Line Cinema	2.016	12	8.569.336	27	13	1	0,0820884506
Paramount Pictures	2.014	1	-1.575.149	33	19	0,05	[NULL]
Paramount Pictures	2.014	3	59.409.368	13	8	0,1	-37,7166655345

5.- Cuadros de mando e informes

En esta sección se muestran dos cuadros de mando y un informe generados con las herramientas de Pentaho. En la Figura 9 se muestra una captura de pantalla del primer cuadro de mando, orientado a la selección de directores y productores para un nuevo proyecto. En la parte superior se proporcionan dos buscadores, para directores y productores, en los que aparecen los mismos ordenados por valoración. Se muestran solo los equipos (directores y productores) que han obtenido más de 100 votos y se muestran resultados para los 10 últimos años. Junto con el valor de cada equipo se muestra el valor del percentil en la distribución de valoraciones.

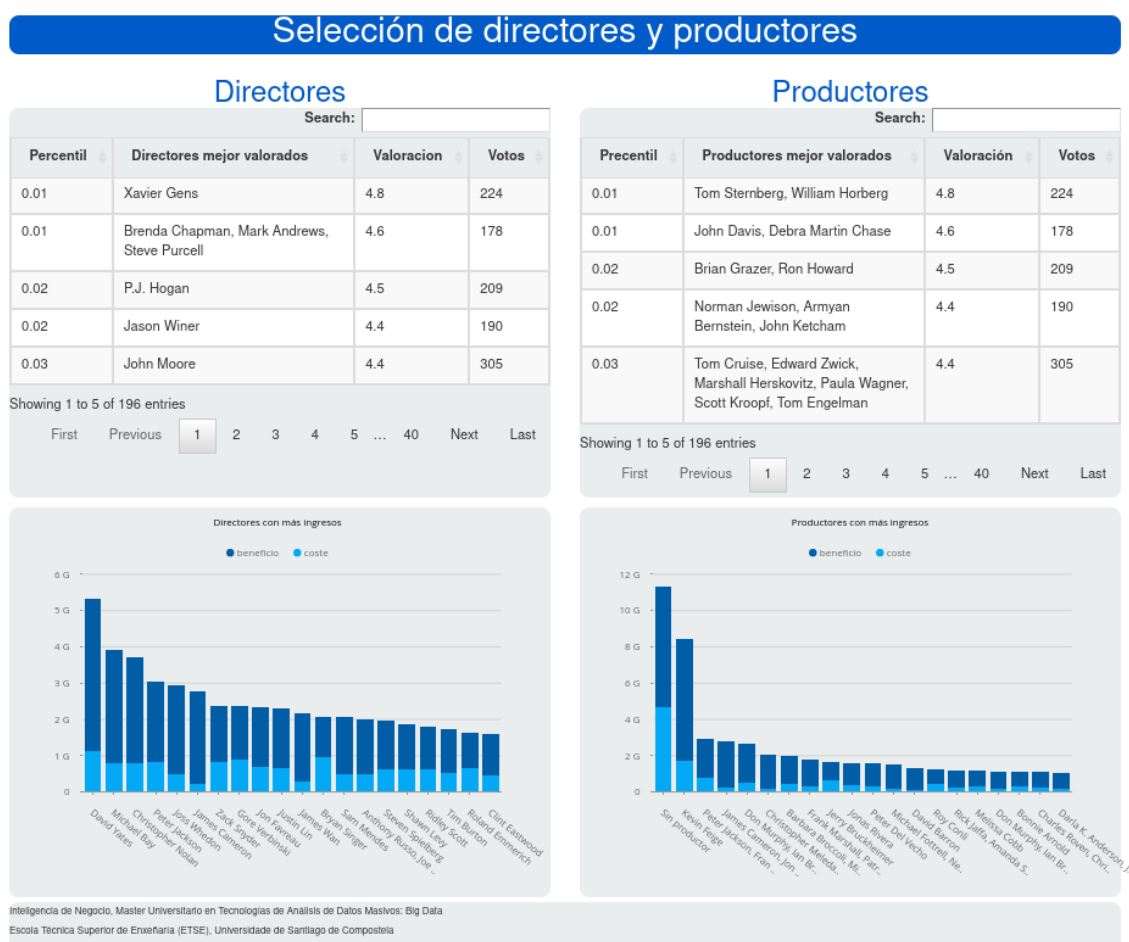


Figura 9: Cuadro de mando de selección de directores y productores.

Debajo de los buscadores tenemos una gráfica de barras que muestra los 20 equipos que han generado más ingresos en los últimos 10 años. El cuadro de mando permite así tener en cuenta tanto la satisfacción del usuario como los resultados financieros para seleccionar a los mejores directores y productores. Los nombres de las personas mejor situadas en las gráficas de resultados pueden utilizarse para buscar en las tablas de la fila superior y ver así el percentil que tienen en la distribución de valoraciones.

El segundo de los cuadros de mando permite visualizar y comparar los resultados financieros de las productoras principales durante los últimos 10 años. La Figura 10 muestra una captura de pantalla de este cuadro de mando. En la parte superior se muestran los resultados acumulados de los últimos 10 años para cada productora, ordenados por productora. En la parte inferior se pueden comparar los resultados de dos productoras elegidas por el usuario (por defecto se

muestran los datos de “Warner Bros.” en ambas gráficas). Para facilitar la comparativa, ambas gráficas utilizan la misma escala para mostrar los resultados.

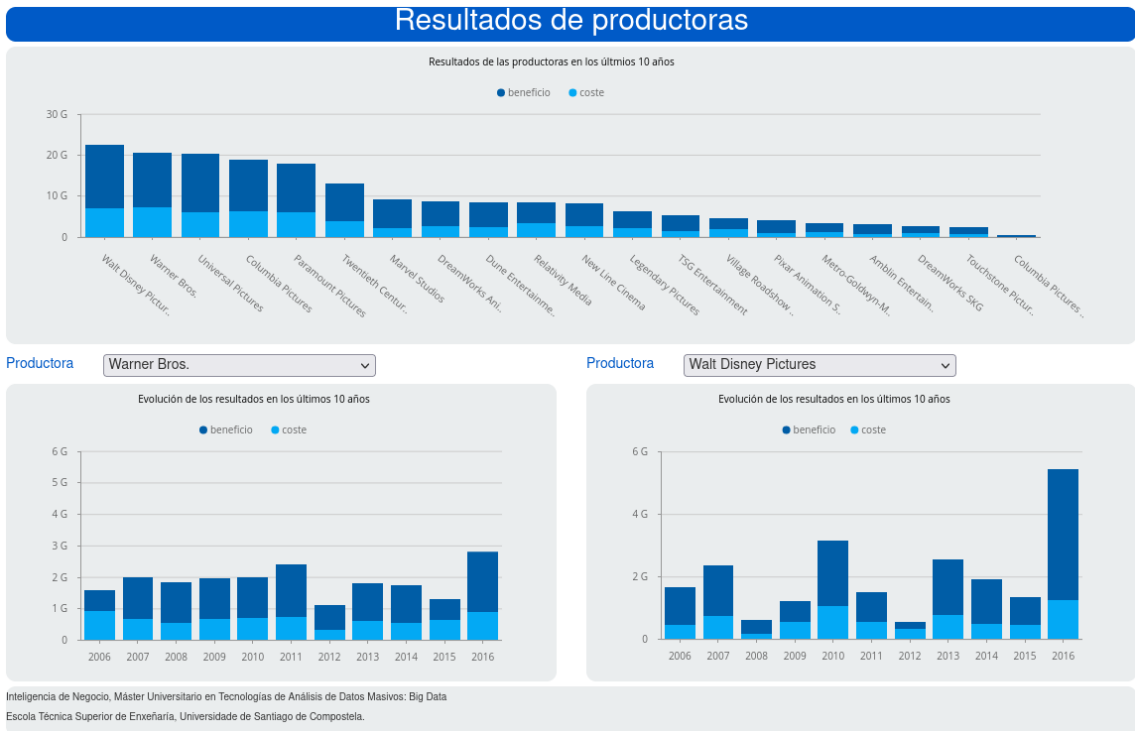


Figura 10: Cuadro de mando de resultados de productoras.

Además de los dos cuadros de mando, se ha implementado un informe que muestra la evolución mensual de los resultados de todas las productoras durante el año 2016. En la Figura 11 se muestra parte de la primera página de este informe.

septiembre 12, 2021 @ 10:25

Resultados mensuales por productora

Productora: Amblin Entertainment			
Mes	Ingresos	Coste	Beneficios
Junio	91.672.794	70.000.000	21.672.794
Total	91.672.794	70.000.000	21.672.794

Productora: Columbia Pictures			
Mes	Ingresos	Coste	Beneficios
Enero	109.906.372	38.000.000	71.906.372
Febrero	58.661.032	37.500.000	21.161.032
Mayo	349.779.543	73.000.000	276.779.543
Junio	119.100.758	17.000.000	102.100.758
Julio	255.279.076	91.000.000	164.279.076
Septiembre	54.120.212	30.000.000	24.120.212
Octubre	220.021.259	75.000.000	145.021.259
Diciembre	151.572.076	55.000.000	96.572.076
Total	1.318.440.328	416.500.000	901.940.328

Productora: DreamWorks Animation			
Mes	Ingresos	Coste	Beneficios
Enero	260.585.412	72.500.000	188.085.412
Octubre	173.432.231	62.500.000	110.932.231
Total	434.017.643	135.000.000	299.017.643

Productora: DreamWorks SKG			
Mes	Ingresos	Coste	Beneficios
Septiembre	25.956.113	20.000.000	5.956.113
Noviembre	57.250.649	22.500.000	34.750.649
Total	83.206.762	42.500.000	40.706.762

Productora: Dune Entertainment			
Mes	Ingresos	Coste	Beneficios

Figura 11: Informe de resultados mensuales durante el año 2016.