**Taller 1 Minería de Datos & Big Data**

Cristian Álvarez y Aníbal Páez

Facultad de Ingeniería y Tecnología, Universidad San Sebastián

Minería de Datos y Big Data

Profesor Mauricio Sepulveda

Universidad San Sebastián

Miércoles 11 de mayo de 2022

[**Introducción**](#_qr05gru5pchf) **4**

[**Proceso hacia Business Intelligence**](#_jbf85qo7a0hj) **4**

[Programas Requeridos](#_18c0eiwyamsn) 4

[KPI’s](#_4nwzpqbw232p) 4

[Modelo Multidimensional](#_e4mb1akqkgcr) 5

[Procesos ETL](#_uyg06tphczru) 6

[Flujo de Control](#_sxwtdlgq11x2) 6

[Generación tabla Calendario](#_t5h6ac84jgwm) 7

[Carga Masiva](#_zh4x3nofsl) 9

[Carga de Órdenes 2016](#_24ob30rxqk7l) 10

[Carga Incremental](#_xrb5lllexme6) 11

[Cuadros de mando](#_ena6sigpogix) 13

[Órdenes Mensuales](#_qixvk36a153l) 13

[Órdenes 2016. En este cuadro de mando se muestra las órdenes por mes del año 2016.](#_u52xtd3jox30) 13

[Órden Incremental. En este cuadro de mando se revelan las órdenes por mes del año 2017 hasta el 2018.](#_nla022569ufc) 14

[Año con más Venta](#_jjexxtct2j4y) 15

[Sucursal con más Venta](#_9pprsh8oqxw) 16

[Cantidad de Órdenes por Estado de Órden](#_2ts0lqmuyhuj) 18

[En esta vista se muestra la cantidad de órdenes que forman parte de un estado de órden en particular.](#_wfz8j5m5vuur) 18

[**Conclusión**](#_jnd8svgwyl64) **19**

[**Instructivo Visual Studio 2019**](#_24y8omytqp9x) **19**

[Paso 1](#_5uh0h1t1bgka) 19

[Paso 2](#_yxanriv3vms3) 19

[**Anexo**](#_ydctjzgtbx03) **21**

# 

# **Introducción**

En este informe se detallan las actividades que realizamos para ejecutar el proyecto de Inteligencia de Negocios relacionado al caso de **BikeStores**.

A partir de los datos que nos fueron entregados mencionaremos los KPI 's que implementamos, los cuales son los “Indicadores Claves de Rendimiento”. También se describe la ejecución de los Diagramas Multidimensionales, los procesos de ETL(Extract, Transform and Load) donde se ejecuta la carga de los datos de forma masiva, como también incremental, para finalmente realizar los cuadros de mando de los KPI 's correspondientes.

# **Proceso hacia Business Intelligence**

## **Programas Requeridos**

Para la solución de este Taller de Inteligencia de Negocios, requerimos de los siguientes Softwares:

* Microsoft SQL Server Management Studio (SSMS)
* Microsoft Visual Studio 2019
* MySQL Workbench
* Microsoft PowerBI

## **KPI’s**

Para establecer los KPIs necesarios, nos tenemos que hacer las siguientes preguntas con respecto a los datos que nos fueron entregados:

* ¿Qué mes fue donde la tienda recibió más órdenes?
* ¿Qué año fue el de más ventas en la tienda?
* ¿Qué Sucursal entre las 3 de la empresa tuvo más ventas?
* ¿En qué estado de pedido se encuentran las órdenes?

Luego, a partir de las preguntas anteriormente mencionadas podemos establecer los siguientes KPIs para trabajarlos en los cuadros de mandos correspondientes:

* Órdenes Mensuales: Este KPI nos permitirá visualizar sobre qué mes fue en donde se solicitaron más órdenes.
* Año con más Ventas: Año que contiene la mayor cantidad de ventas.
* Sucursal con más Ventas: Establecimiento que tendrá más ventas entre las 3 sucursales.
* Cantidad de Órdenes por Estado de pedido: Estado de pedido de las órdenes por cada orden.

## **Modelo Multidimensional**

Según lo mencionado anteriormente más la base de datos entregada desde un principio se conformó el siguiente Modelo Multidimensional. Para este diagrama usamos el software MySQL Workbench.

## 

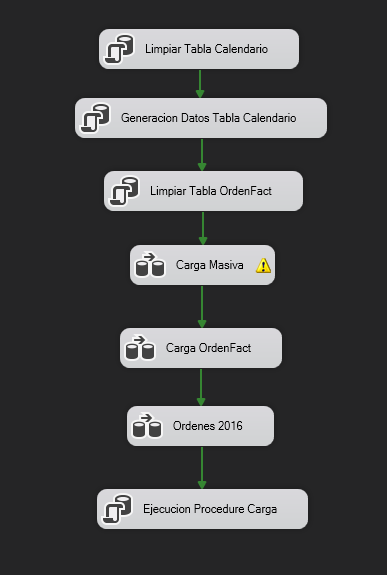
En caso de no poder visualizar la imagen le dejaremos adjunto un link en la sección de Anexo donde podrá visualizar mejor.

## **Procesos ETL**

Para comenzar con los procesos necesarios, primero se necesita restaurar la database en nuestro entorno de SSMS con el archivo llamado **BikeStores.bak**, que fue entregada por el profesor, para esto nos dirigimos a nuestra DataBase para así restaurar esta Base de Datos. Creamos nuestra base de datos que contiene nuestro modelo multidimensional, modelo que se pudo apreciar anteriormente y habiendo realizado estos pasos pasamos a realizar los flujos de cargas dentro de Visual Studio 2019.

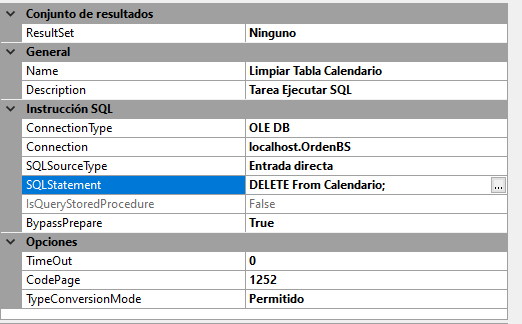
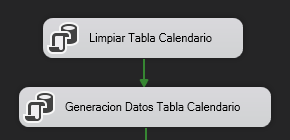
### **Flujo de Control**

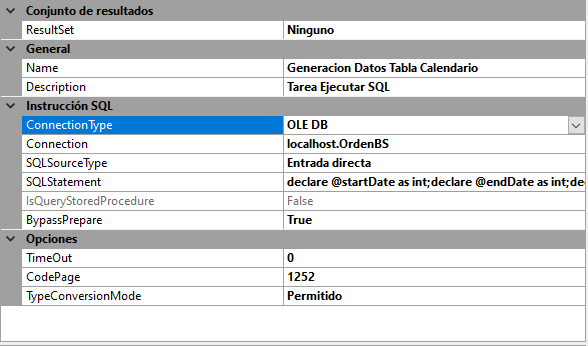
Antes que nada se debe generar una conexión con las bases de datos, la base BikeStores que se nos entregó y siendo la nuestra la base **OrdenBS**, dentro de Visual Studio 2019. Una vez realizado lo anterior comenzamos por generar el flujos control en el que armamos el orden de cómo se irán cargando los datos a la base destino utilizando la herramienta “Flow Task”.



### **Generación tabla Calendario**

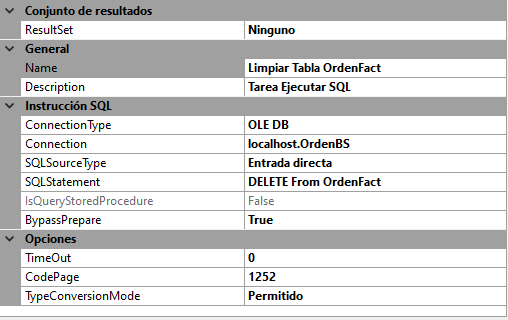
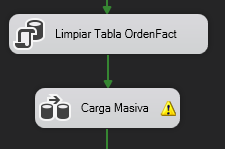
Al notar que la base de datos origen no contaba con una tabla de tiempo para poder extraer los datos por lo que hicimos uso de una herramienta que nos permite generar datos de tiempo solo con ingresar una Query dentro de esta misma. Primero limpiamos la tabla Calendario de la base OrdenBS y luego con la Query generamos los datos.

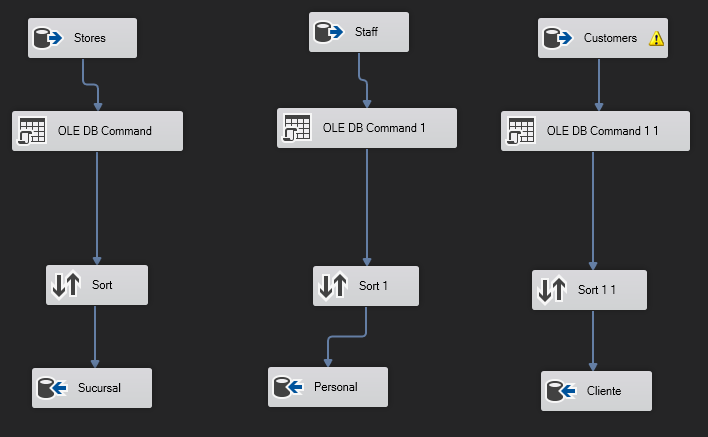




### **Carga Masiva**

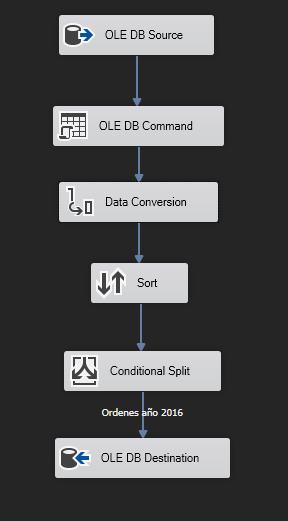
Comenzamos limpiando nuestra tabla de OrdenFact(Tabla de hechos) antes de entrar al Flow Task, luego dentro del Flow Task creamos tres flujos de datos en donde el origen serán las tablas sales.customer, sales.staff y sales.stores de la base de datos BikeStores y los datos de estas llegaran a las tablas de destino Cliente, Personal y Sucursal de la base OrdenBS, respectivamente.





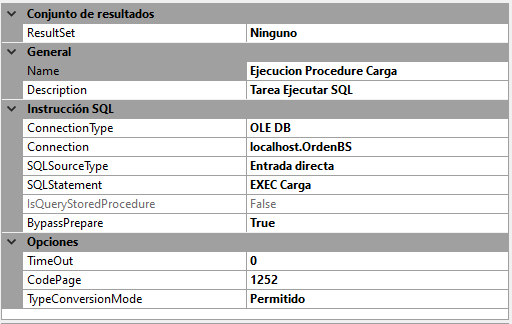
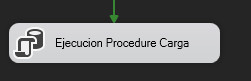
### **Carga de Órdenes 2016**

En esta sección realizamos una carga masiva hacia la tabla destino Orden2016 de la base OrdenBS, cargando las fechas de las órdenes de todo el año 2016, junto con las ID ‘s de estas órdenes y sus estados.



### **Carga Incremental**

En esta última carga nos enfocamos en realizar un incremento de datos de la tabla origen sales.orders hasta la tabla destino OrdenIncremental, cargando específicamente las fechas de las órdenes a partir del 2017 de manera mensual, es decir, se cargarán datos desde el primer mes del 2017 hasta el último mes del 2018. Para esta sección utilizamos un Store Procedure que ejecuta los comandos que van cargando incrementalmente los datos.



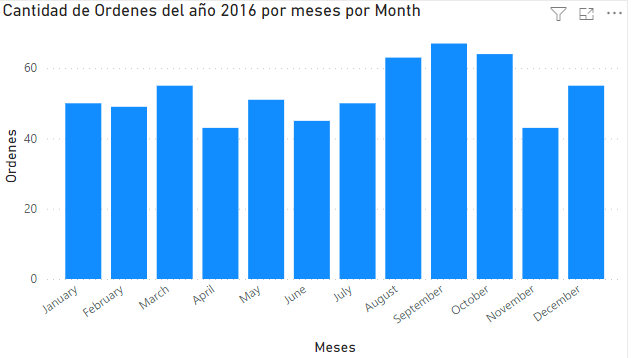
## 

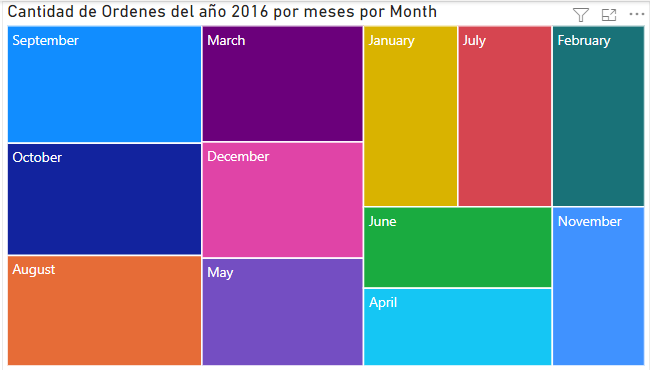
## **Cuadros de mando**

A continuación se presentan los cuadros de mandos correspondientes para poder visualizar los KPI ’s escogidos a través del programa PowerBI.

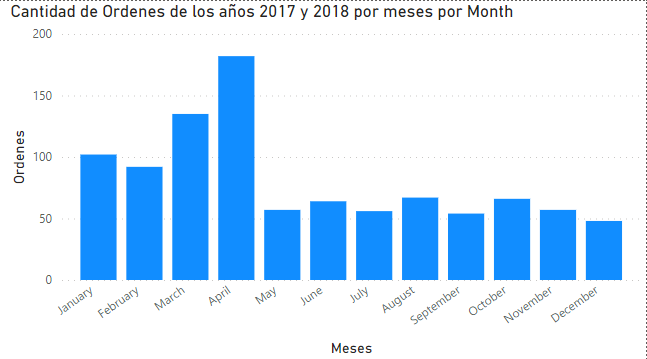
### **Órdenes Mensuales**

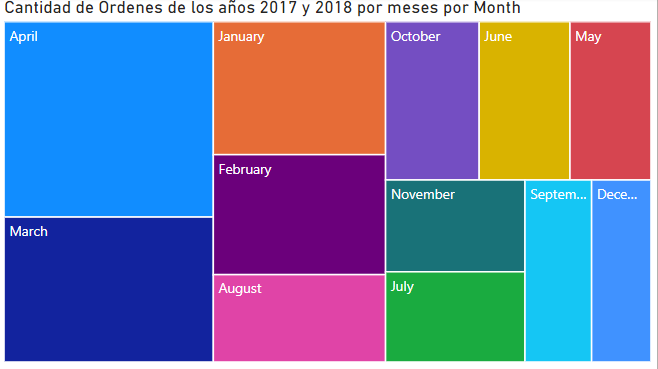
#### **Órdenes 2016.** En este cuadro de mando se muestra las órdenes por mes del año 2016.





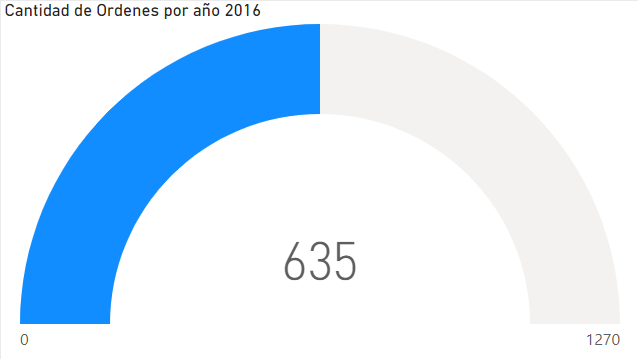
#### **Órden Incremental.** En este cuadro de mando se revelan las órdenes por mes del año 2017 hasta el 2018.

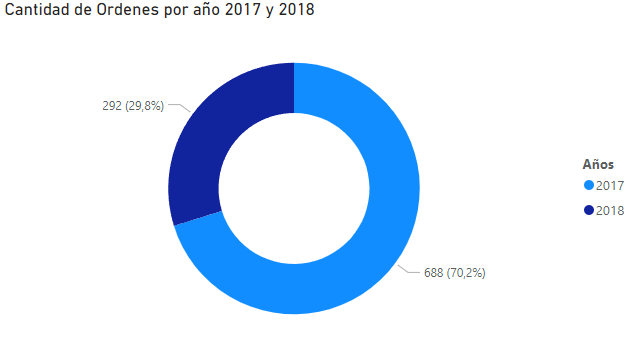




### **Año con más Venta**

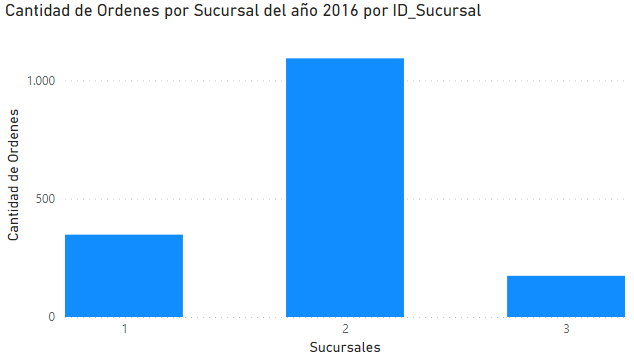
En esta vista se aprecia el año que tuvo más ventas o bien más órdenes.

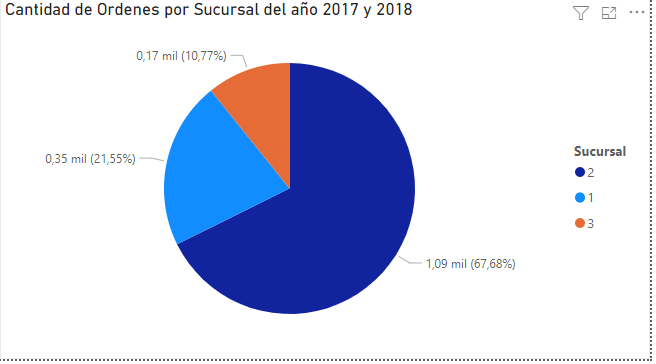




### **Sucursal con más Venta**

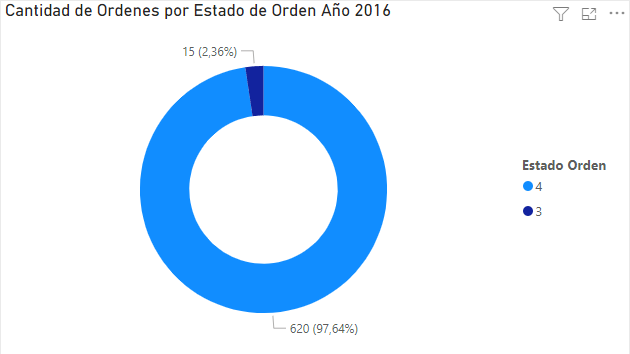
En este cuadro se muestra la comparación entre las sucursales sobre cuál tuvo la mayor cantidad de ventas u órdenes.

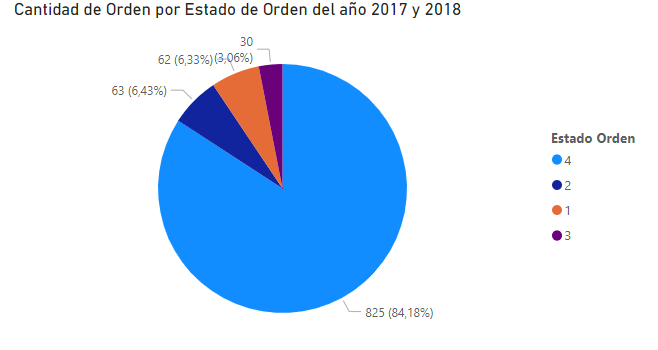




### **Cantidad de Órdenes por Estado de Órden**

#### En esta vista se muestra la cantidad de órdenes que forman parte de un estado de órden en particular.





# **Conclusión**

Este trabajo ha demostrado ser un desafío sobre todo en todo en el proceso de ETL, pues un ligero error en el modelo multidimensional podría causar el rediseño de este mismo y tener que comenzar de nuevo con el procesos de cargas, por lo tanto, antes de realizar los procesos de ETL es necesario analizar y evaluar la base de datos de origen para luego proponer de manera adecuada las KPI ‘s y finalmente diseñar el modelo multidimensional.

Así mismo, se deben tener los datos claros que se van a ocupar y cargar en los distintos flujos, para la simplificación del uso de estos datos en el software de Power BI y escoger los cuadros de mando con la mejor visualización posible para su entendimiento.

# **Instructivo Visual Studio 2019**

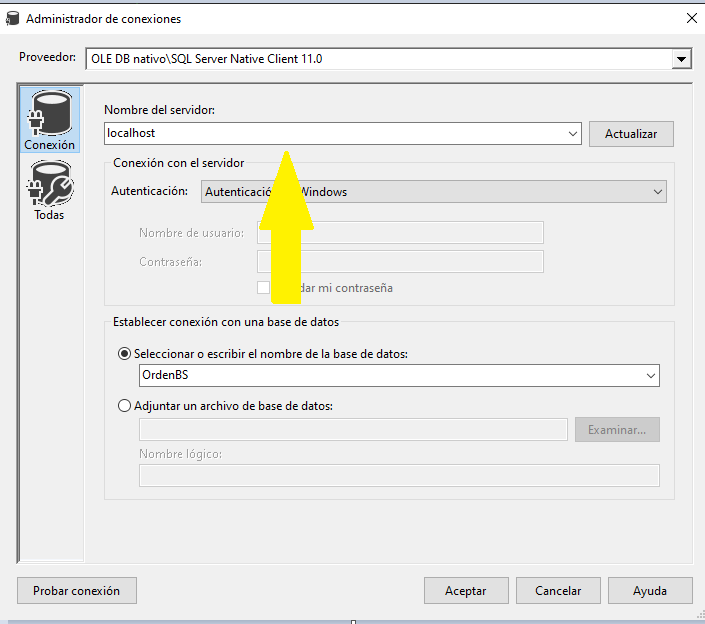
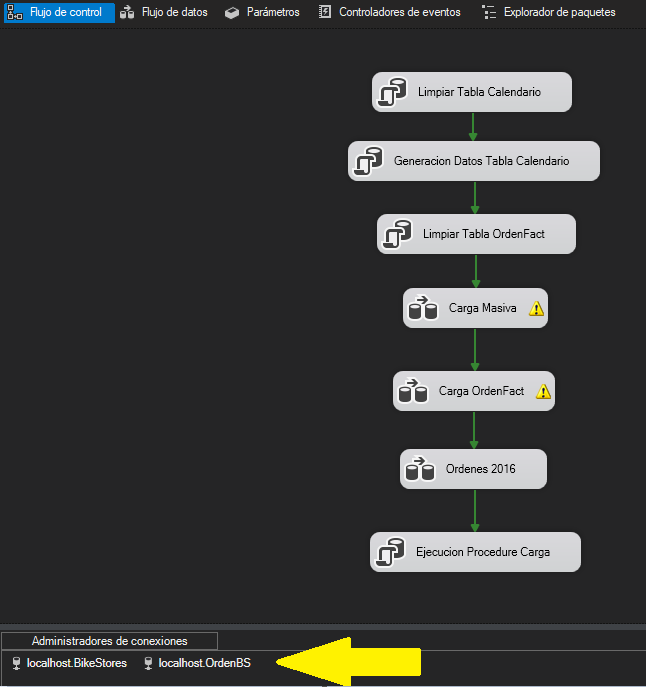
En caso de querer hacer uso de este programa deberá realizar los siguientes pasos.

## **Paso 1**

Dirigirse al localhost.OrdenBS y hacer click como se muestra en la imagen.

## **Paso 2**

Ir a la sección de Nombre de servidor y seleccionar su servidor local.



# **Anexo**

**Imagen del modelo multidimensional**

[ModeloMultidimensional.png](https://drive.google.com/file/d/17DNPyyb6ZM4vwsnQGkPXWEHzLi_Yw99r/view?usp=sharing)

**Power BI**

<https://app.powerbi.com/groups/d13ab270-c044-4bd8-b74c-fae8e592bdac/reports/f5e52d82-ba25-4774-b9f4-2311830e8ab8?ctid=7a599002-001c-432c-846e-1ddca9f6b299&pbi_source=linkShare>