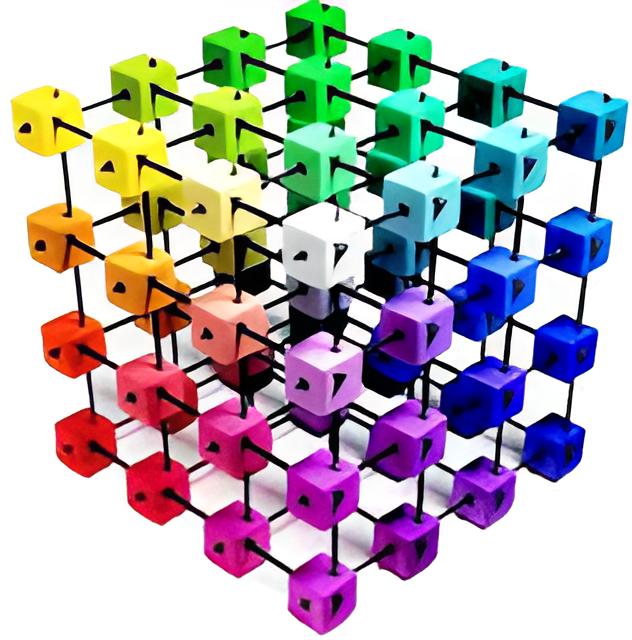
**Taller 1**

**Minería de Datos y Big Data**

**“Caso BikeStore”**



**Integrantes:**

* Jorge Gonzalez
* Roberto Neris

**11 may 2022**

# **Índice**

[**Índice**](#_m0febvclmnw9) **2**

[**Introducción**](#_eujegcxq92vw) **3**

[**Desarrollo**](#_jbf85qo7a0hj) **3**

[**KPI’s**](#_4nwzpqbw232p) **3**

[**Datawarehouse multidimensional**](#_tpn21wvhk7ei) **5**

[**Requerimientos para uso de la base**](#_91z5nputddq6) **5**

[**Carga de datos**](#_wn0890e1j6s5) **6**

[**Cuadro de mandos**](#_kkx2qhqdqyr0) **9**

[**Conclusión**](#_ofimsvhqq3rs) **10**

[**Referencias**](#_dcy5le9pq4bz) **10**

# 

# **Introducción**

En este informe se describen los pasos que fueron efectuados para realizar todas las actividades y exigencias dados en este Taller 1, partiendo desde la incorporación del backup dado por el profesor hasta el desarrollo de un cuadro de mando para las vistas de los KPIs. Dentro del desarrollo del informe veremos las actividades solicitadas a realizar tales como la definición detallada de los KPIs con sus respectivas características, el diagrama del modelo multidimensional realizado con sus componentes, el procedimiento de carga masiva que se realizó junto con la carga mensual, el cuadro de mando con las respectivas vistas de los KPIs y finalmente información adicional solicitada.

# **Desarrollo**

El taller consiste en realizar actividades relacionadas a un proyecto de ingeniería con el objetivo de ver la información de las ventas de bicicletas, esto a través de un cuadro de mando con las vistas asociadas a los KPIs que se definirían.

Como inicio, se nos facilitó un backup de un modelo relacional con el que debemos trabajar, este puede ser modificado y mejorado según se prefiera.

## **KPI’s**

Para establecer los KPIs necesarios, nos hacemos las siguientes preguntas:

* ¿Cuánto dinero ingresa por cada solicitud de compra?
* ¿Cuáles son los productos más vendidos?
* ¿Cuántos envíos fueron cancelados o fallaron?
* ¿Cuántos clientes acceden a registrar todos sus datos de contacto como su número de teléfono?

Luego, a partir de las preguntas anteriormente mencionadas podemos establecer los KPIs:

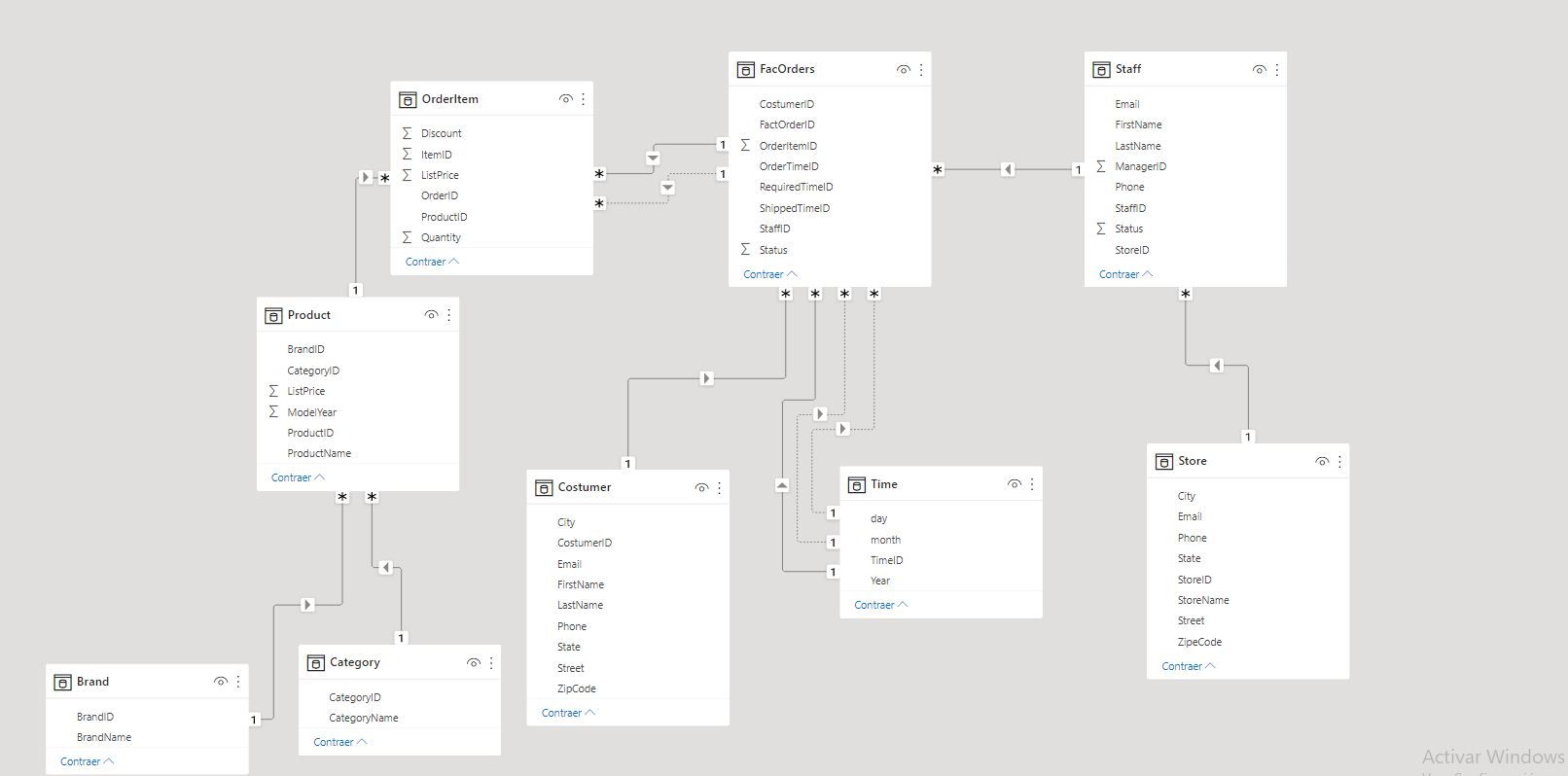
* **Producto más vendido:** Listado de los productos más vendidos dentro del catálogo.
* **Ingresos:** Los ingresos de cada orden solicitados con sus respectivos factores y cuál es el ingreso ideal.
* **Tasa de Fracaso:** Cantidad de pedidos o envíos fallidos en relación al total registrados
* **Confiabilidad de número de teléfono:** Cuántos clientes acceden a dar y registrar sus número de teléfono.

Basándonos en el modelo bsc se generó la siguiente tabla:

| **KPIs** | **Departamento** | **Objetivo** | **Categoría** | **Indicadores** | **Metas** |
| --- | --- | --- | --- | --- | --- |
| Most selled | Clientes | Definir que producto es el más cotizado por los clientes | Retail | Ventas de productos vs número | visualizar que producto es el más vendido por la tienda |
| Income | Resultados financieros | Obtener ingresos mayores a las planteadas | Financiero | Ingresos actuales vs planteados | Ingresos un 10% por sobre la planeación |
| Failure rate | Proceso interno del negocio | Mantener baja tasa de fallos | Financiero | valores NULL/total de datos en la columna shippedTime | Mantener una tasa menor al 5% |
| Cellphone confiability | Clientes | Mantener un buen número de personas que cedan su número telefónico | Clientes y mercado | Número de valores con datos no NULL/ número total de datos ingresados | Mantener por sobre un 20% la tasa de personas que ingresan su número telefónico a la base |

## **Datawarehouse multidimensional**

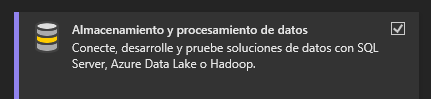
En base a los KPI's definidos, las relaciones y las variables de cada tabla presente en el modelo relacional se genero el siguiente modelo multidimensional:

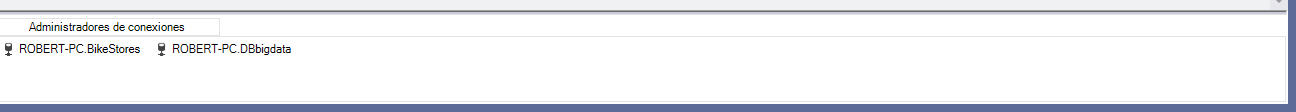


Este modelo pasó por diferentes diseños y cambios, los cuales concluyeron en el modelo anterior, constituido por 4 dimensiones y con estructura de copo de nieve.

## **Requerimientos para uso de la base**

Para utilizar la solución de la base para bikestore se requiere de utilizar **visual studio 2019** y de sql server manager studio. En el caso de visual studio se debe de tener el addon almacenamiento y procesamiento de datos perteneciente a las herramientas extras de visual studio.

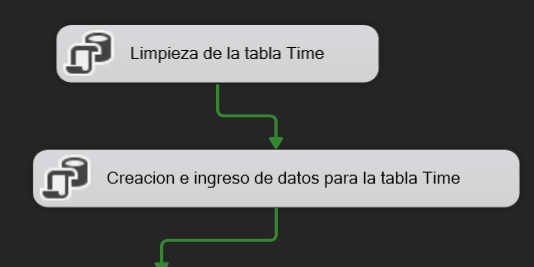


Además hay que cambiar las conecciones activas a que tienen como servidor **a localhost**  a la perteneciente al servidor que se esté utilizando

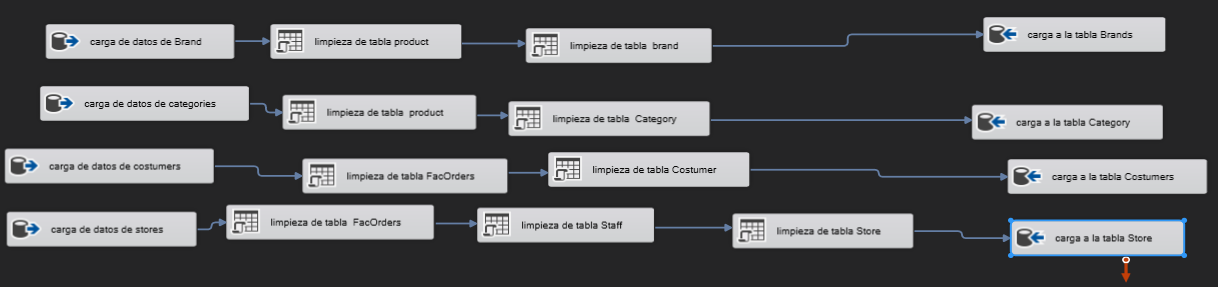
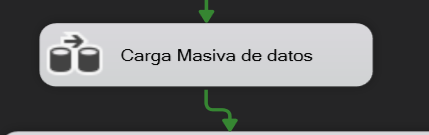
## **Carga de datos**

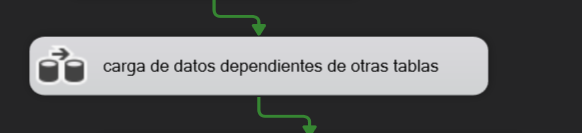
Para este procedimiento realizamos una serie de pasos dentro de un flujo de control que se irá ejecutando en orden y para cargar los datos se utilizó el data Staging para cargar los datos de la siguiente manera:

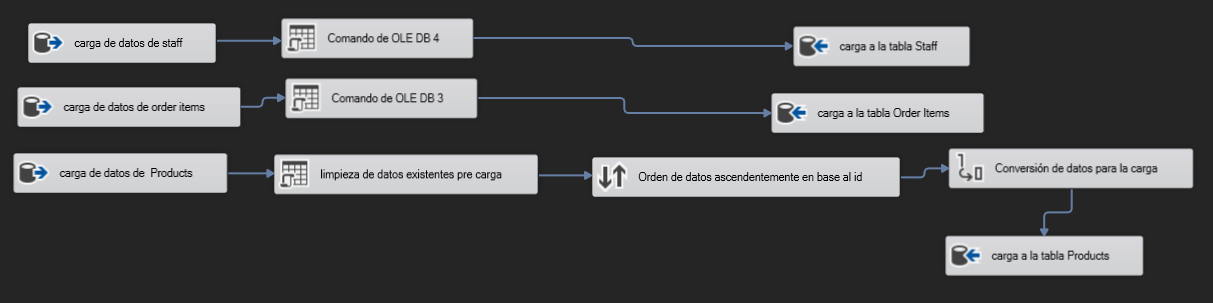
* Primero se borra cualquier dato existente en la tabla Time para evitar errores, luego se genera la tabla “Time” utilizando una query sql.



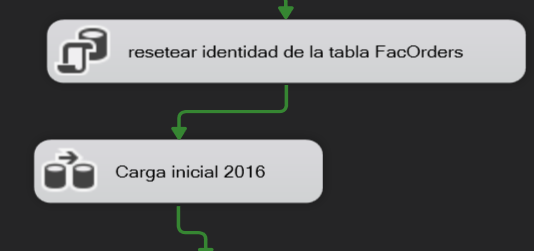
* Se cargan los datos iniciales de 2016 mediante Carga Masiva considerando primero los datos que no dependan de una foreign key y luego quienes sí dependen. Estos procesos componen un gran flujo de datos el cual se ve a continuación.







* Luego se resetea la identidad de la tabla FacOrder y pasamos directamente a la carga inicial del 2016 donde se extraen los datos de la tabla relacional, luego se limpian las tablas de datos desde las más grandes a las más pequeñas y luego se cargan los datos a sus tablas correspondientes. Después de cargar los datos iniciales se cargan los datos dependientes de otra tabla con el mismo método para finalmente ingresar los datos correspondientes a 2016.





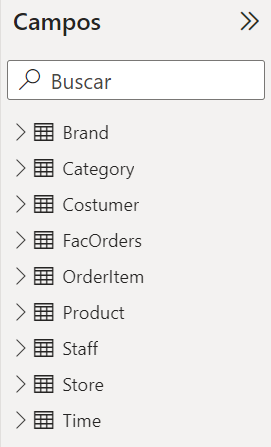
* Después de cargar los datos iniciales se cargan los datos dependientes de otra tabla con el mismo método para finalmente ingresar los datos correspondientes a 2016,
* Se cargan los datos mensuales de 2017 en adelante mediante



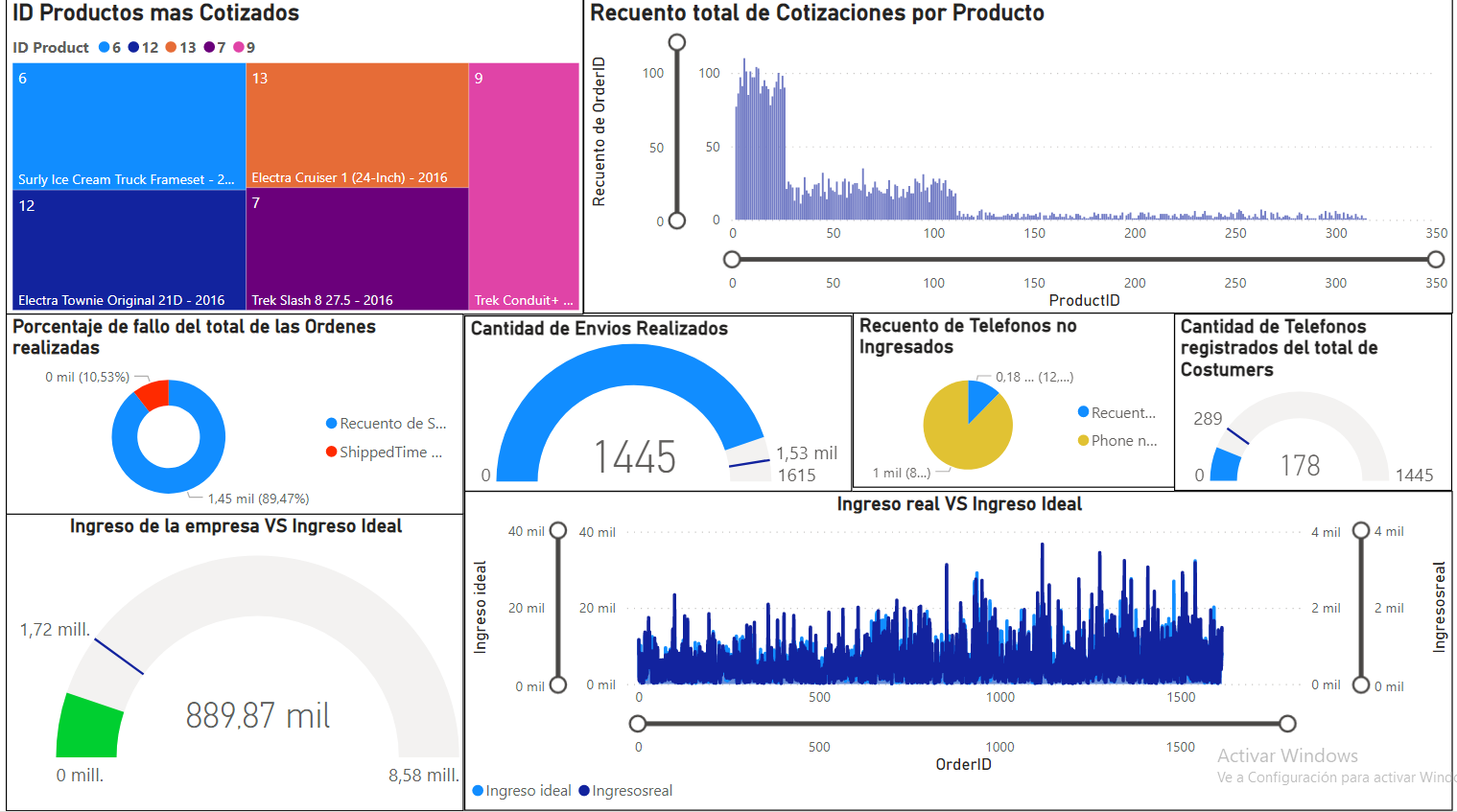
## **Cuadro de mandos**

Para visualizar los KPIs se utilizo PowerBI para crear el cuadro de mandos quedando el siguiente resultado:

**Campos resultantes**



**Vistas del cuadro de mando**



# **Conclusión**

En base a la producción de este DW multidimensional se puede concluir que el hecho de hacer un DW no es fácil y requiere de un buen conocimiento de la empresa para poder realizar un buen DW, los cuadros de mandos facilitan enormemente el poder visualizar las medidas de la empresa y poder tomar decisiones respecto a la información almacenada que posee una empresa.

# **Referencias**

link de powerBI : <https://app.powerbi.com/links/d2N9cemlWK?ctid=7a599002-001c-432c-846e-1ddca9f6b299&pbi_source=linkShare>