***Trabajo final bases***

1. Un grupo de inversionistas desea empezar desde cero con un nuevo proyecto, buscan tener una base sólida de información que les permita tomar decisiones a la hora de saber en qué tipo de negocio y en qué ubicación geográfica invertir para que dicho proyecto sea realmente rentable y exitoso para todos los involucrados.
2. preguntas de negocio:

* ¿En qué tipo de tiendas desearía invertir mi dinero como inversor?
* con el fin de maximizar el flujo de clientes y de dinero ¿en que barrio seria bueno iniciar con el proyecto?

1. preguntas analiticas:

* ¿Cuál es la cantidad de clientes dependiendo del tipo de documento?
* ¿Cuáles son los principales sectores(barrios) en los cuales hay un mayor flujo de dinero?
* ¿Cuál es la cantidad de tiendas según el tipo y qué tipo de tiendas son las que reciben más dinero por sus ventas?
* según el tipo de documento como se distribuyen los clientes?
* ¿Cuál fue la tienda más visitada por los clientes y en qué barrio se encuentra?
* ¿Cuál es la distribución de tiendas según el barrio? ¿Cuáles son los barrios con más tiendas y cuáles los barrios con menos?

***descripción del proceso de limpieza de los datos***

**Etapa bronce**: Se recibieron los datos originales en los que se pudieron encontrar algunos errores, en la columna de número de documento varios valores estaban en negativo lo cuál no tiene mucho sentido teniendo en cuenta el contexto del atributo por lo tanto se pusieron todos los valores en positivo, en la columna tipo de tienda y nombre del barrio los campos estaban mezclados entre letras mayúsculas y minúsculas, se toma la decisión de convertirlos todos a mayúscula para evitar errores a la hora de hacer las diferentes consultas, además de esto tomamos la decisión de retirar la columna de latitud y longitud. En esta etapa el nombre del archivo es: [**Sales\_Dataset\_Input.xlsx**](https://github.com/MIGUEVARVAL/FinalDataBase/blob/master/bronze_stage/Sales_Dataset_Input.xlsx)

**Etapa plata:** Al revisar los datos nos dimos cuenta que todas las preguntas importantes que nos hicimos acerca del negocio se podían responder con los datos ya suministrados y por eso el archivo de la etapa plata pasa a ser el de la etapa oro. en esta etapa el nombre del archivo es: [**Sales\_Dataset\_Input\_Clean.xlsx**](https://github.com/MIGUEVARVAL/FinalDataBase/blob/master/Components/silver_stage/Sales_Dataset_Input_Clean.xlsx)

**Etapa oro:** ya con los datos en esta etapa pudimos empezar con el proceso de pensar y crear el modelo relacional para la posterior ingesta de los datos a la base. El archivo en esta etapa se llamó: [**Sales\_Dataset\_Input\_Info.xlsx**](https://github.com/MIGUEVARVAL/FinalDataBase/blob/master/Components/golden_stage/Sales_Dataset_Input_Info.xlsx)

***descripción del modelo relacional***

tomamos la decisión de dividir los datos en cinco tablas:

* La tabla barrio posee dos atributos: el identificador del barrio y el nombre del barrio.
* Después tenemos la tabla cliente que contiene el número de documento del cliente y el tipo de documento de este.
* La tabla llamada “tbTienda\_has\_tbCliente” combina dos atributos: el número de documento del cliente y el código de la tienda que esté visitó.
* la tabla tienda que contiene el código de la tienda, el total de las ventas que se hicieron en la tienda, el tipo de tienda y el barrio en el que esta se encuentra.
* por último tenemos la tabla tipo de tienda que tiene el atributo identificador y en texto el tipo de tienda que es cada identificador.

