

DEFINICIÓN DEL PROYECTO

- **Objetivo:** Analizar las ventas del supermercado para verificar si factores de los clientes tales como ubicación geográfica, sexo y situación marital influyen en sus decisiones de compras.
- **Alcance:** Normalización de una base de datos de ventas de supermercados para crear un tablero en Power BI de análisis de información e indicadores.
- **Nivel de aplicación del análisis:** Estratégico.
- **Usuario final:** Directorio de la cadena de supermercado.

DEFINICIÓN FUENTE DE DATOS

Temática de los datos: En el siguiente trabajo, se realizará un análisis de las ventas de una cadena de supermercados de Estados Unidos, en donde los registros van desde diciembre 2011 hasta diciembre 2022.

Hipótesis: En este trabajo se evaluará si tanto ubicación geográfica, sexo y situación civil afectan la compra valorizada de las familias de productos comida, no-consumible y bebestibles.

Data set: El data set a trabajar corresponde a registros de ventas de una cadena de supermercados de Estados Unidos con sucursales en distintos estados, poseyendo 14.060 registros que entregan información tanto de compras como información de los clientes entre los periodos diciembre 2011 hasta diciembre 2013.

- A continuación, se entrega el data set desde un archivo Excel.

DEFINICIÓN TABLAS

- **STORE_CITY:** En esta tabla se encuentra información sobre las ciudades en los cuales están las tiendas.
 - Posee los campos: ID_CITY (PK INDEX) Y CITY.
- **STORE_STATE:** En esta tabla se encuentra información sobre los estados en los cuales están las tiendas.
 - Posee los campos: ID_STATE(PK INDEX) y STATE
- **STORE_COUNTRY:** En esta tabla se encuentra información sobre los países en los cuales están las tiendas.
 - Posee los campos: ID_COUNTRY(PK INDEX) Y COUNTRY.
- **PRODUCT_FAMILY:** En esta tabla se encuentra información sobre las familias de productos que se encuentran en las tiendas.
 - Posee los campos: ID_PFAMILY (PK INDEX) y PRODUCT_FAMILY.
- **PRODUCT_DEPARTMENT:** En esta tabla se encuentra información sobre los departamentos a los cuales pertenecen los productos.
 - Posee los campos: ID_PDEPARTMENT (PK INDEX) y PRODUCT_DEPARTMENT.
- **PRODUCT_CATEGORY:** En esta tabla se encuentra información sobre las categorías a las cuales pertenecen los productos.
 - Posee los campos: ID_PCATEGORY(PK INDEX) y PRODUCT_CATEGORY.
- **CUSTOMERS:** En esta tabla se encuentra información sobre los clientes de las tiendas.
 - Posee los campos: ID_CUSTOMER(PK INDEX), GENDER, MARITAL_STATE, HOMEOWNER, CHILDREN, ANNUAL_INCOME_INF_LIMIT y ANNUAL_INCOME_SUP_LIMIT.
- **SALES:** En esta tabla se encuentra información sobre la venta de productos en las tiendas.
 - Posee los campos: ID_TRANSACTION(PK INDEX), PURCHASE_DATE, ID_CUSTOMER (FK), ID_CITY(FK), ID_STATE (FK), ID_COUNTRY (FK), ID_PFAMILY(FK), ID_PDEPARTMENT(FK), ID_PCATEGORY(FK), UNITS_SOLD Y REVENUE_USD.

DIAGRAMA ENTIDAD-RELACIÓN

