# INFORME DE VENTAS

## CALIDAD DE DATOS

A manera general la calidad de datos es buena.

Sin embargo, se encontraron datos faltantes en el peso del producto y el tamaño del local. En este sentido los puntos a mejorar:

* En **grocery store** falta más de la mitad de las observaciones del peso del producto y el tamaño de la tienda.
* En **Supermarket Type1** faltan observaciones del tamaño de la tienda.
* En **Supermarket Type3** no hay datos sobre el peso del producto.
* El etiquetado de los datos no tiene el mismo formato y puede llevar a conclusiones erroneas (ej: Low Fat -LF, Regular-reg) Esto podria generar que se categoricen de forma incorrecta los productos.

Sería importante que la data de estos locales se llene de manera adecuada para evitar inferencias que pueden no ser exactas a los datos reales. Tal vez el analizar el como se alimentan las bases puede ayudar a que se realice un control de calidad periodico y evitar estos inconvenientes.

A pesar de esto, existen metodos que se aplicaron para llenar los valores faltantes.

## PRINCIPALES HALLAZGOS

**Variedad de productos**

* **Tipo de local:** El Supermercado tipo 1 es el que tiene más variedad de productos.
* **Tamaño de local:** Los locales medianos son los que tienen más variedad.
* **Locación de local:** Existe la misma variedad en todas las zonas (tier1 tier2 y tier3).

**Ventas**

### **Tipo de local:** El Supermercado tipo 1 es el que tiene más ventas. Sin embargo, el promedio de ventas del Supermercado tipo 3 es mayor (menos frecuencia, mayor cantidad). Igualmente en este supermercado es donde se dan las ventas por producto más altas.

* **Tamaño de local:** Los locales medianos son los que venden más. Se ve que generan más ventas pero el valor de la venta máxima y minima es similar que en las otras locaciones.
* **Locación de local:** Los locales ubicados en el tier 3 venden más. Se ve que generan más ventas pero el valor de la venta máxima y minima es similar que en las otras locaciones.
* El local que genera mas ventas es el 27. En general la venta promedio es superior que en los otros locales. El valor maximo de venta es superior que los demás. Los locales 13, 45, 17, 46, 35 y 49 tienen valores maximos y minimos de venta y promedio de venta similares.
* En todos los locales la participación en la venta de los productos regular fat y low fat es igual. Sin embargo el abastecimiento de productos low fat es mayor que el regular fat.
* Los productos que generan más ventas son las frutas y vegetales seguidas de los snacks.

## PREDICCION DE VENTAS

Se corrieron tres modelos para predecir las ventas pero la que dio mejores resultados fue el modelo Random Forest Regression ya que el score fue el mayor.