Problem Statement Formulation

Can UGR (UK Grocery Retail) increase unit sales of their brick-and-mortar stores (excludes ecommerce sales) by changing pricing and distribution of top selling items?

Data source: SalesSample Exploratory.cvs
Target: SalesQuantity (explore both item level and total grocery sales quantity)
Features to Explore: NewBasePrice, SellingStoreRatio (distribution), WeekIdentifier (for seasonality) and Territory

Context

UGR is under new leadership. The new CEO expects a top line overview how the company is doing: specifically, which sales channels, territories and sub-segments are driving performance.

He strongly believes the company can fuel future sales growth through changing pricing and distribution* decisions. He wants his Data Science team to prove or disprove this hypothesis, focusing on the brick and mortar stores (excluding eCommerce) and the top selling items that account for 70-80% of the sales. (Note: assume that bottom 20% will be replaced by more distribution of strong selling items or new products)

He also wants the team to develop a model to predict unit sales of the top selling items, by changing price and/or distribution.

He will welcome any additional insight that will increase sales.

Scope of Solution Space

- Sales Channel: Brick and Mortar Stores Stores
- Top 20-30% of Items that control 80% of Store Revenues (Dollars)

Success Criteria:

- Deliver predictive model
- Grow Sales Quantity and show how this impacts revenue (dollars)
- Deliver the final recommendation by Dec. 15, 2025

Constraints

- Data on retail margins, how much USG actually earns from the sale of each good, is not provided.
- The stores have fixed amount of space, so increasing distribution of an item means you need to decrease the distribution of another item.

Stakeholders: New CEO, Springboard (Mentor and Evaluation Team)

Data Sources:

- SampleSales.csv - contains detailed weekly sales data covering the years 2018 and 2019.
- Items.csv - provides essential information about the product hierarchy, grouping products into logical categories for easier analysis and categorization.
- PriceAdjustments.csv - offers insights into the historic price changes for each product, aiding in understanding pricing strategies over time

The link to the files is provided here:
https://www.kaggle.com/datasets/ziedzen/uk-grocery-retailer-sales-and-pricing-analysis

Initial thoughts on how to solve the problems:

- Wrangle data and confirm scope Mm5PoR88 is a good sub-category to model on (ie. meaningful in size and has enough variability in its pricing and assortment histories)
- Narrow and clean the data sets to the items in scope
- Get a preliminary view of the factors impacting sub-category sales (ie. Mm5PoR88 in AXX stores).
- Join relevant information into the modelling dataset.
- Evaluate different models to predict sub-category sales.
- Explore multiple regression and other more complex techniques (like random forests).
- Determine the model the best predict sub-category sales
- Use the best model to predict impact of changing price and assortment, and make recommendations based on it.
- Throughout the process, gather insights for potential growth and future exploration

Deliverables:

A GitHub repo containing the work you complete for each step of the project

Timely completion of each step.

By December 15

- Full Project Documentation (doc)
- Executive Level Power point presentation