

# Data Science : Retail Forecasting Group Project

**Group Name: Data Solver**

Member 1:

Name: Yiu Hung Leong

Email: ernest9773@gmail.com

Country: Malaysia

College/Company: Monash University

Specialization: Data Scientist

Member 2:

Name: Oladimeji Adaramoye

Email: adaramoyeoladimeji@gmail.com

Country: United Kingdom

College/Company: Roehampton University

Specialization: Data Science

Github repo link:

<https://github.com/ErnestCodeHub/Data-Glacier-Internship/tree/main/Group%20Project>

## Problem description: Forecasting Beverage Demand

A large company in the beverage industry in Australia sells their products through various supermarkets and heavily invests in promotions throughout the year. They face the challenge of accurately forecasting the demand for each product at the item level on a weekly basis. This forecast is crucial for effective inventory management and production planning.

The company's demand patterns exhibit various complexities, including trends, seasonality, and non-linear patterns. As a result, we will build 4-5 multivariate forecasting models, including ML or deep learning-based models, using PySpark and parallel computing techniques and demonstrate best-in-class forecast accuracy by achieving a low weighted Mean Absolute Percentage Error (MAPE). Lastly, we will provide explainability of the models by showcasing the contribution of each variable in the forecasting process.

## Business Understanding

Accurate demand forecasting is crucial for the success of any business, especially in the fast-paced consumer goods industry. The ability to predict future demand patterns helps companies optimize their inventory, production, and supply chain processes, leading to reduced costs and improved customer satisfaction.

In the case of the beverage industry, demand forecasting becomes even more critical due to the perishable nature of the products and the highly competitive market. The company in question heavily relies on promotions and marketing campaigns to drive sales, making it challenging to accurately forecast demand.

By implementing advanced AI/ML-based forecasting models, we aim to provide the company with reliable predictions of beverage demand at the item level. This will help them make informed decisions regarding production planning, inventory management, and marketing strategies.

## Project Life Cycle and Deadline

The project will follow an iterative and collaborative approach, consisting of the following phases:

### **Week 8 ( Due date 26 May 2024)**

Data Collection and Exploration: In this phase, we will collect historical demand data for the company's beverage products, along with relevant external factors such as promotional activities, holidays, and weather data. We will explore the data to understand its characteristics, identify any data quality issues, and perform initial visualizations to gain insights.

### **Week 9 ( Due date 2 June 2024)**

Conduct further data cleaning

### **Week 10 ( Due date 9 June 2024)**

Feature Engineering: In this phase, we will leverage feature engineering techniques to derive additional variables that capture the underlying factors affecting beverage demand. This may include lagged variables, seasonality indicators, promotional indicators, and other relevant features. We will also perform any necessary data transformations and scaling

### **Week 11 ( Due date 16 June 2024)**

EDA Presentation and present proposed modeling technique.

### **Week 12 ( Due date 23 June 2024)**

In this phase, we will evaluate the performance of the developed models using metrics such as weighted MAPE and forecast accuracy. We will compare the models' results and select the best-performing model(s) for further refinement and analysis. Lastly, we will refine the selected model(s) by incorporating feedback from stakeholders and addressing any identified limitations or issues. We will also analyze the contribution of each variable in the forecasting process to provide explainability and insights to stakeholders.

### **Week 13 ( Due date 01 July 2024)**

Submit Final Project Report and Code

# Data Intake Report

**Name:** Data Science : Retail Forecasting Group Project

**Report date:** 19/05/2024

**Internship Batch:** LISUM32: 01 Apr 24 - 01 July 24

**Version:** 1.0

**Data intake by:** Yiu Hung Leong

**Data intake reviewer:** Oladimeji Adaramoye

**Data storage location:**

<https://github.com/ErnestCodeHub/Data-Glacier-Internship/tree/main/Group%20Project>

**Tabular data details:**

**Data Dataset:**

<b>Total number of observations</b>	<b>14616</b>
<b>Total number of files</b>	<b>1</b>
<b>Total number of features</b>	<b>12</b>
<b>Base format of the file</b>	<b>.csv</b>
<b>Size of the data</b>	<b>74 KB</b>