DATA GLACIER - RETAIL FORECASTING PROJECT

WEEK 8

Group Name	Projection Wizard
Name	Hande Gul Atasagun
Email	hgatasagun@outlook.com
Country	Turkey
Specialization	Data Science

1. Problem description:

A large beverage company operating in the Australian market faces challenges in demand forecasting for its wide range of products. The demand is subject to fluctuations influenced by various factors, including holidays, seasonality, and promotional activities. The company is currently using an in-house forecasting software, which often produces unreliable and nonsensical forecasts. In pursuit of more accurate and efficient forecasting, the company aims to explore the potential of AI/ML-based solutions to replace their existing in-house system.

2. Dataset understanding:

The dataset provided is a collection of historical sales data for the company's products. The dataset comprises data recorded from February 5, 2017, to December 27, 2020. The data includes the following columns:

- Product: Product name.
- Date: Dates on which sales data were recorded weekly.
- Sales: The number of units sold weekly.
- Price Discount (%): The percentage of discount applied to the product's price during that week.
- In-Store Promo: Indicates the presence of in-store promotions (1 for yes, 0 for no) during the week.
- Catalogue Promo: ndicates the presence of catalogue promotions (1 for yes, 0 for no) during the week.
- Store End Promo: Indicates the presence of store end promotions (1 for yes, 0 for no) during the week.
- Google_Mobility: Data reflecting the impact of Google Mobility on sales.
- Covid_Flag: A flag representing the influence of COVID-19 on sales.
- V_DAY, EASTER, CHRISTMAS: Indicators of specific holiday or event occurrences and their impact on sales during the week.

3. Data types:

product	object
date	datetime64[ns]
sales	int64
price_discount	float64
in_store_promo	int64
catalogue_promo	int64
store_end_promo	int64
google_mobility	float64
covid_flag	int64
v_day	int64
easter	int64
christmas	int64

4. Problems in the data and approaches:

Observations for the last 6 weeks within the specified date range are missing for 5 products. To fill in these missing data, previous observations will be used. While there are outliers in the sales data, the approach to handling outliers will be determined based on the models to be used.