

Name: Felix Muriithi Nyagah

Email: fnyagah@gmail.com

Country: Kenya

Specialization: Data Science

[Github_repo](#)

Problem description

The large company who is into beverages business in Australia. They sell their products through various super-markets and also engage into heavy promotions throughout the year. Their demand is also influenced by various factors like holiday, seasonality. They needed forecast of each of products at item level every week in weekly buckets

Data understanding

The data has 12 columns and 1218 rows

The datatype for the columns include 3 objects, 8 integers and one float

The following are the columns and description:

Product: This column likely refers to the unique identifier or code for a particular product.

Date: This column indicates the date of sale for the product.

Sales: This column shows the total sales revenue generated by the product on the given date.

Price Discount (%): This column indicates the percentage discount that was applied to the product's price at the time of sale.

In-Store Promo: This column likely indicates whether there was a promotion or discount offered specifically for in-store purchases.

Catalogue Promo: This column likely indicates whether there was a promotion or discount offered through a catalogue or other print or digital media.

Store End Promo: This column likely indicates whether there was a promotion or discount offered specifically at the end of a store's promotional period.

Google_Mobility: This column may be a measure of the relative mobility or foot traffic in the area surrounding the store where the product was sold, based on Google's location data.

Covid_Flag: This column likely indicates whether there were any COVID-related restrictions or concerns in effect on the date of sale.

V_DAY: This column likely indicates whether the date of sale was close to or on Valentine's Day.

EASTER: This column likely indicates whether the date of sale was close to or on Easter.

CHRISTMAS: This column likely indicates whether the date of sale was close to or on Christmas.

Problems of the data

The data has no missing values

The data has no duplicates

The Sales column has outliers but they are genuine so they remain in data due to the seasonalities

The Sales column is skewed to the right due to seasonalities we will account for seasonality through seasonal decomposition

The date column is of object datatype it needs to be transformed to datetime

The Easter, Christmas, Covid_Flag, V_day, In-Store Promo, Catalogue Promo, Store End Promo need to be changed from integers to boolean

The V_Day column should be renamed to Valentines

The Price Discount column should be converted from object to integer after removing the percentage sign in the data