Data Glacier Group Project: Retail Forecasting

Data Cleansing and Transformation

Team Member Details:

Kierra Dangerfield
kierradachelle@yahoo.com
United States of America
Freelance
Specialization: Data Science

Problem Description:

The large company which 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 a forecast of each of the products at item level every week in weekly buckets.

Github Repo Link:

https://github.com/KierraDangerfield/Data-Glacier/tree/main/Week_9

Data Cleansing and Transformation

The dataset only has 1218 entries. There are no null values or duplicate values.

- I transformed the date column from object to datetime.
- I changed the “Price Discount (%)” column from string to float and renamed the column to “Discount”

I added 3 columns:
DayOfWeek: which day the date was on
Quarter: Quarter of the year the date of sales is in

WeekNum: The week number of the year

There are outliers in the sales column. Because I am doing a forecasting model, I did not remove the outliers. Most of the outliers come from quarter 3 and 4. If the outliers are affecting the model, I might have to remove them from the dataset or replace the values with mean, median, or mode values.

I am looking at experimenting with Linear Regression, tree-based models, and other models as well as some type of regularization/standardization.