Milestone 5

Initial Dataset:

From the beginning of the project, I took part in finding an appropriate dataset that aligned with my and group members interests. After considering a couple of datasets, our group decided to work on an e-commerce dataset, which we later replaced. Initially, I took the initiative to create exploratory graphs like bar graphs and scatter plots, and I also assessed the normality of the data through QQ plots to analyze the distribution. Additionally, I constructed the main weekly sales for the top-selling product in that dataset. After receiving valuable feedback, our group switched to a different sales dataset because the initial one lacked sufficient data for comprehensive analysis.

Store Sales Forecast Dataset:

I mainly participated in model construction of the store sales dataset, starting from ACF, PACF, EACF, and Box plots to KPSS and ADF tests to analyze and construct different ARIMA models of the product with the highest sales that is 'GROCERY I'. This required a deep understanding of how to handle non-stationary series and how to deal with series that are not white noise. I also tried different transformation techniques such as taking logs, differencing, and log returns. However, when differencing, the series was over-differenced. Although, I also tried looking at the log transformed series and analyzed its ACF, PACF, EACF etc because I visualized a little bit of transformation in the series after appyling log.

After accounting for the non-stationary series, I attempted to build the model manually. During this process, I examined the residuals and performed backtesting. This was very challenging, as it requires significant time to try different models, especially with seasonality, to achieve residuals that are white noise. I also did forecasting after constructing the best-fit model. In addition, I analyzed the auto ARIMA models to compare and evaluate their performance with the manaully constructed model.

This entire process of building an ARIMA model has greatly helped me understand the concepts of stationarity, non-stationarity, backtesting, and forecasting. Although I am still in the process of coming up with the best-suited model due to an initial dataset change, the series still seems quite challenging.

This project significantly enhanced my understanding and application of time series techniques and concepts. I learned how to analyze and build model for non-stationary series, perform various transformations, and implement ARIMA models effectively. Additionally, I gained valuable experience in diagnosing model performance through residual analysis and backtesting, which are crucial for accurate forecasting.

In addition to this, I took the initiative to arrange online meetings for the group to discuss the analysis, assign various tasks, and plan the next steps. Furthermore, on behalf of my group, I presented the initial series in class and received valuable feedback from the professor. I have contributed to completing every milestone so far.