**Modeling Approach**

Our modeling approach had different stages and steps to get our team in a place where we felt happy with the overall model performance. Our main goal was to create a final deliverable with a model that produces accurate forecast, considering seasonality for each of the sales metrics, in which the model would update the forecasts as new data comes in for the site ensuring accuracy.

**Step 1: Variable Selection Using Linear Regression:**

We first brainstormed and researched different possible models strategies like LSTM, Weighed moving averages, linear regressions and more complex time series models like VAR and ARIMA. We initially used more basic linear regression models and Variance Inflation Factor in order to perform variable selection and identify the predictors with better predictive power for each of the sales metrics.

**Step 2: Penalized Regression Using LASSO:**After concluding linear regression for variable selection, we decided to create our first model by using LASSO penalized regression. Our main idea behind this choice was because of how computationally effective LASSO is, and the fact of how LASSO chooses the best predictors variables for each of the sales metrics and reduces the coefficients of the less important predictors. In order to implement that we created a loop in which the sales metrics would be added as a lag from previous historical data into the model as new columns, being used as predictors for the next day forecast.

**Step 3: Vector Autoregressive (VAR) and ARIMA models:**

After creation of the LASSO model, our team decided together with the Professor’s advising to try to develop a more comprehensive time series model in order to use lag values from previous days for the different sales metrics and being able to increase accuracy and decrease RMSE values. With this in mind, we tried two different time series models that per our research have been shown to be effective in time series forecasting analysis: VAR and ARIMA. We chose Vector Autoregressive because of its capacity of capturing the dynamic interactions and feedback effects among multiple variables. On the other hand, we decided to create an ARIMA model because of how it accounts for various patterns, such as linear or nonlinear trends, constant or varying volatility, and seasonal or non-seasonal fluctuations. We believed both models could provide us good insights, and ARIMA could be really helpful considering the nature and fluctuations of our data.

**Step 4: Discussion of accuracy metrics for each model:**

Our team thinks that the creation of these different models was really helpful in weaknesses and strengths, and using different approaches to see which one would best suit our data and goal for this Capstone Project.