New Store Sales Forecasting

Team Presentation

Sampoornaa Ghosh

Christopher Porter

Samruddhi Supekar

Business Problem

➢ Optimizing New Store Performance

➢ Enhancing Operational Efficiency

➢ Measuring Success through Metrics

Purpose

➢ Correct Prediction of the First Year Annual Sales

➢ Correct Prediction of Daily Sales

Day 0 Performance Metrics

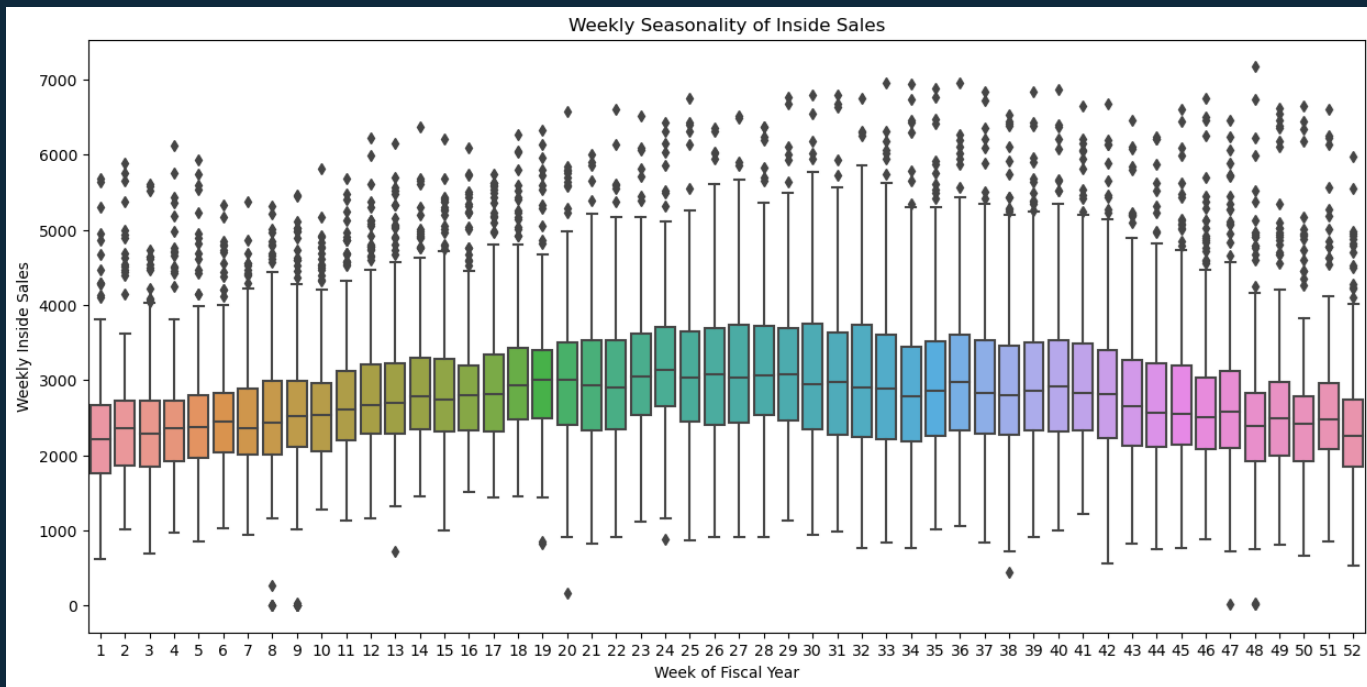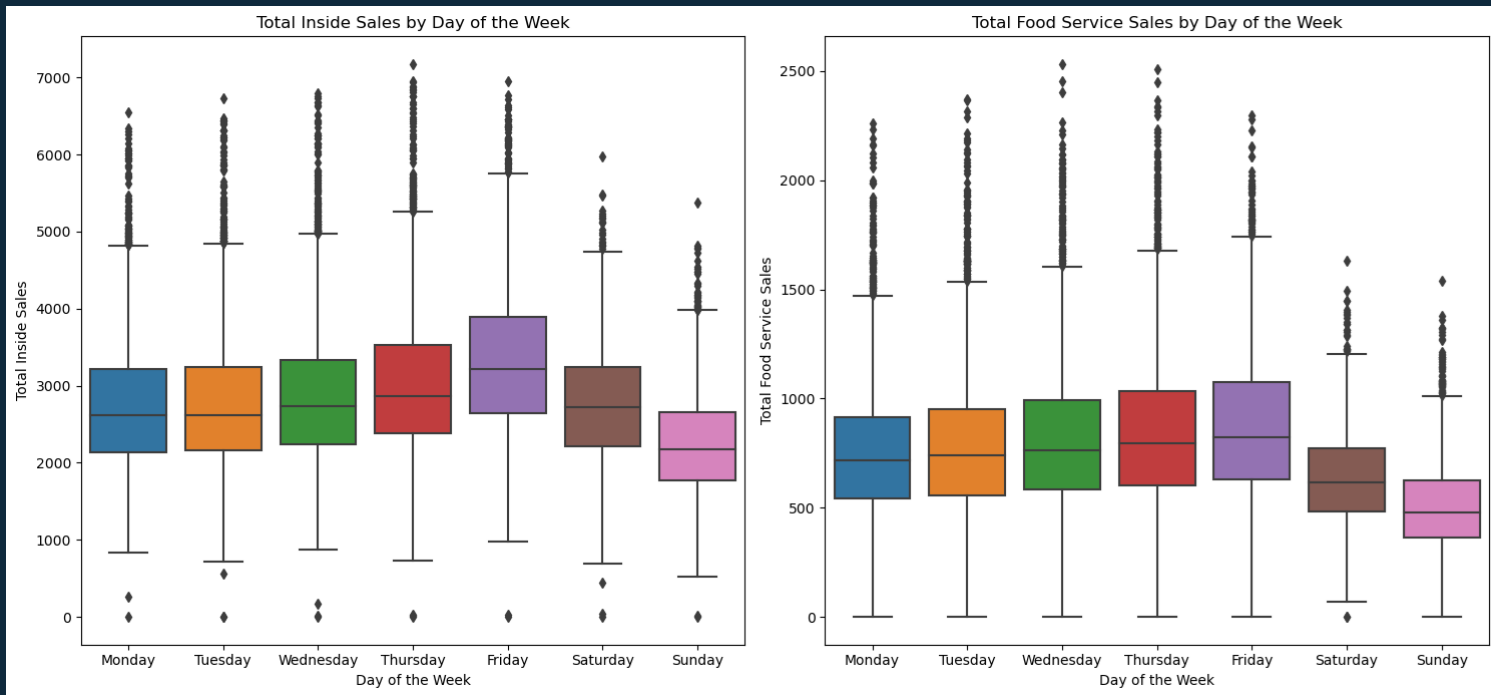| Metric	Inside Sales	Food Service	Diesel (gal)	Unleaded (gal)
Daily MAE	$688.27	$222.90	875	761
Daily MAPE	24.15%	27.96%	66.25%	54.71%
Annual Average Error per Store	-$140,468.76	-$69,233.33	251,607	109,111
Annual Cumulative Error %	-12.79%	-22.97%	45.22%	15.52%
Minimum Annual Cumulative Error % (Store)	1.54%	-15.84%	-3.61%	1.53%
Maximum Annual Cumulative Error % (Store)	-28.78%	-32.00%	68.54%	144.20%

Data Explanation

37 Stores in total

➢ 1 Store (21980) is so different it was excluded

➢ 5 Stores (22925, 22680, 23415, 22785, 23380) were selected at random as test stores

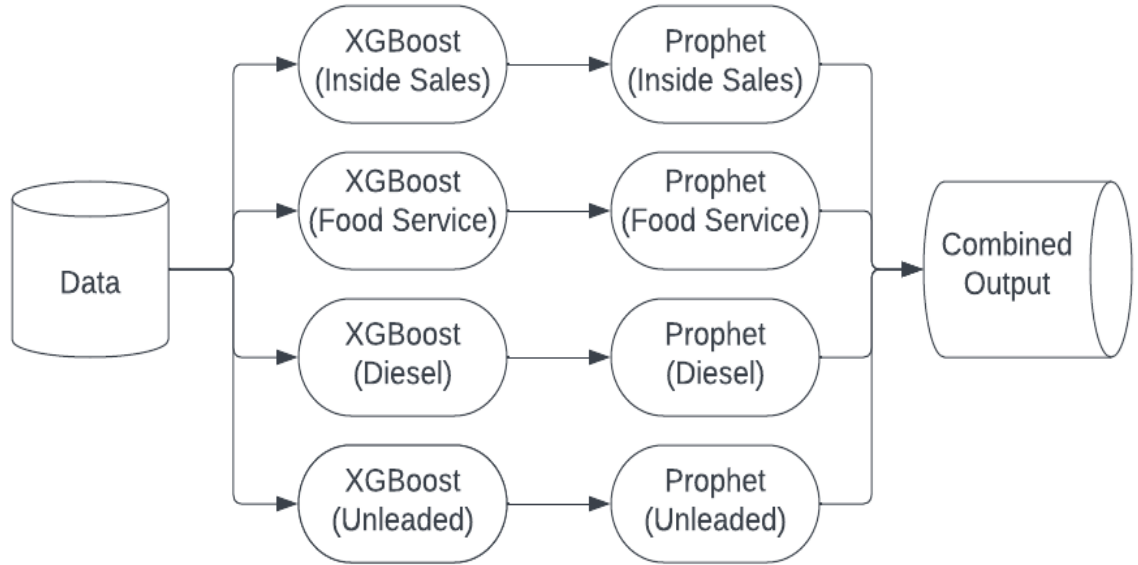➢ 31 Stores for Training Data

Exploratory Data Analysis

Weekly Seasonality of Inside Sales
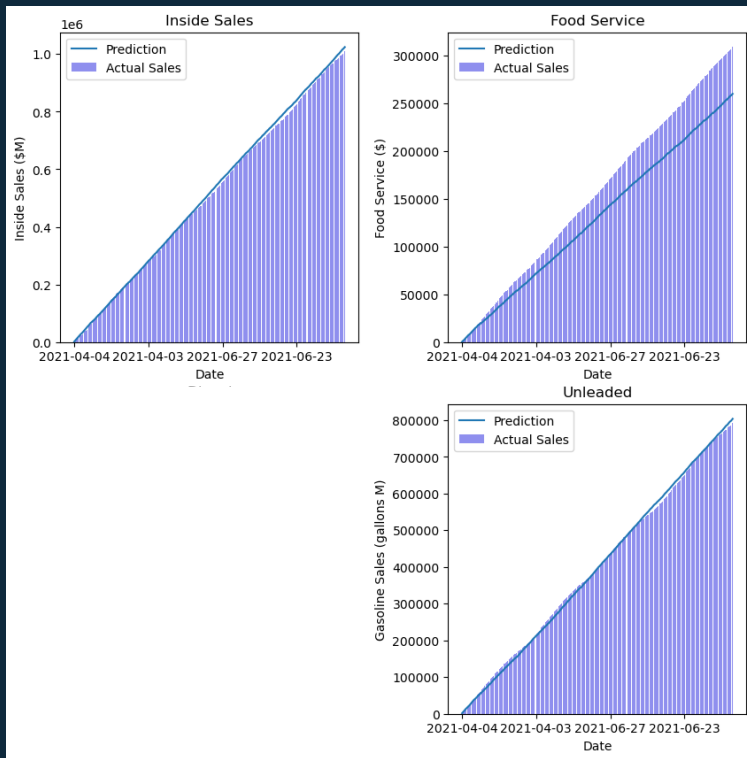
Exploratory Data Analysis

8

Model Architecture

➢ XGBoost creates naive prediction of sales prior to store opening

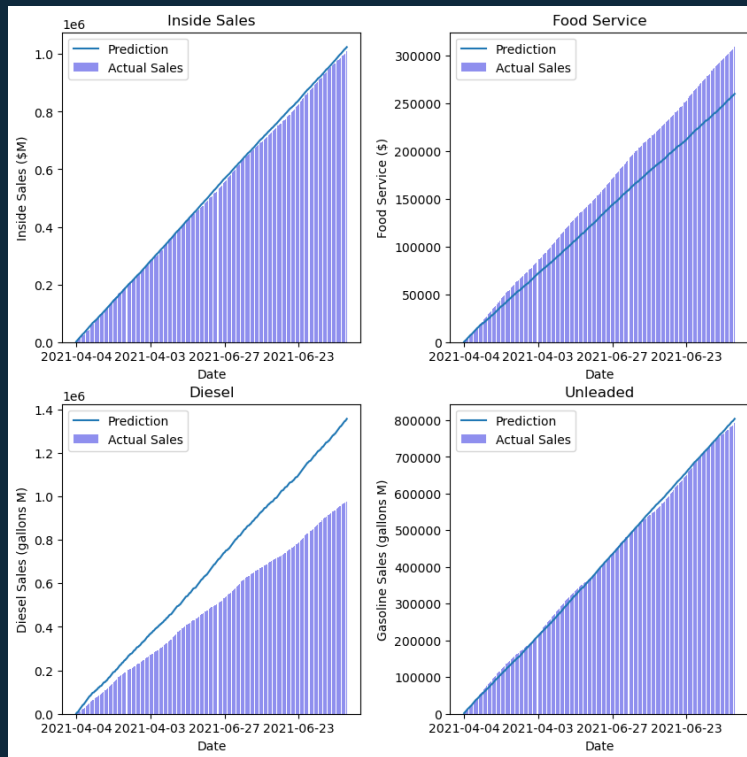➢ Prophet predicts future sales based on actual sales data

9

Model Performance

➢ XGBoost (Naive) Model can accurately predict Annual Sales in many instances

Model Performance

➢ XGBoost (Naive) Model can accurately predict Annual Sales in many instances.

But not in every case.

Two Models with different strengths

XGBoost

◇ Good prediction capability.
◇ No previous data needed
◇ Data Hungry
◇ No Temporal Awareness
◇ Stagnant Accuracy

Two Models with different strengths

XGBoost

- Good prediction capability.
- No previous data needed
- Data Hungry
- No Temporal Awareness
- Stagnant Accuracy

Prophet

- Temporal Awareness
- Needs prior data to be accurate
- Gets less accurate the longer the horizon
- Improving Accuracy

Two Models with different strengths

XGBoost

- ◇ Good prediction capability.
- ◇ No previous data needed
- ◇ Data Hungry
- ◇ No Temporal Awareness
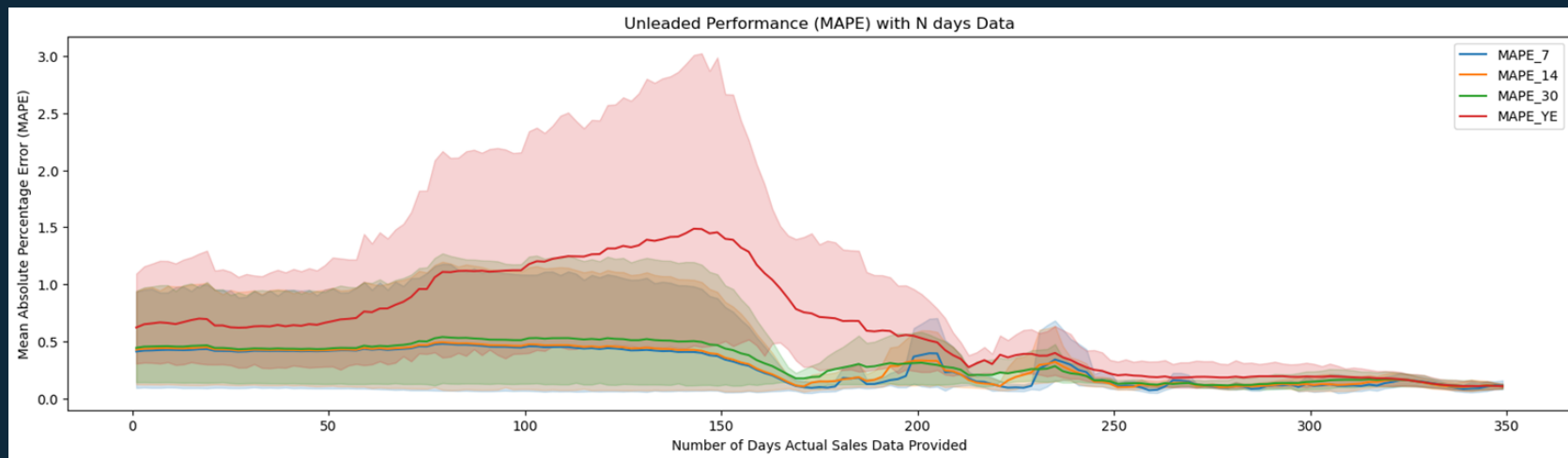- ◇ Stagnant Accuracy

Prophet

- ◇ Temporal Awareness
- ◇ Needs prior data to be accurate
- ◇ Gets less accurate the longer the horizon
- ◇ Improving Accuracy

Composite Model

- ◇ Can predict with no prior data
- ◇ Can use prior data to refine predictions
- ◇ Gets more accurate over time

A picture is worth a thousand words

Unleaded Performance (MAPE) with N days Data

Model Analysis

Strengths

◇ Can be very accurate
◇ Fulfills Use Case

Model Analysis

Strengths

- ◇ Can be very accurate
- ◇ Fulfills Use Case

Weaknesses

- ◇ Holidays
- ◇ Small Sample Size Issues
- ◇ 1 of 5 types of store not included

Model Analysis

Strengths

- ◇ Can be very accurate
- ◇ Fulfills Use Case

Weaknesses

- ◇ Holidays
- ◇ Small Sample Size Issues
- ◇ 1 of 5 types of store not included

Opportunities

- ◇ 400 Stores & Multiple Years
- ◇ Opportunities to improve sensitivity to Seasonality
- ◇ Better Forecasting Models

Model Analysis

Strengths

◇ Can be very accurate
◇ Fulfills Use Case

Weaknesses

◇ Holidays
◇ Small Sample Size Issues
◇ 1 of 5 types of store not included

Opportunities

◇ 400 Stores & Multiple Years
◇ Opportunities to improve sensitivity to Seasonality

Threats

◇ Kum 'n' Go
◇ Regionality

19

Business Value

➢ Accurately Predict Annual and Daily Sales prior to store opening within 25% of actual performance in key sales segments

➢ Improve Predictions throughout the year as actual sales data becomes available

➢ Provide very accurate 7 day, 2 week, and 1 month forecasts to benchmark store performance
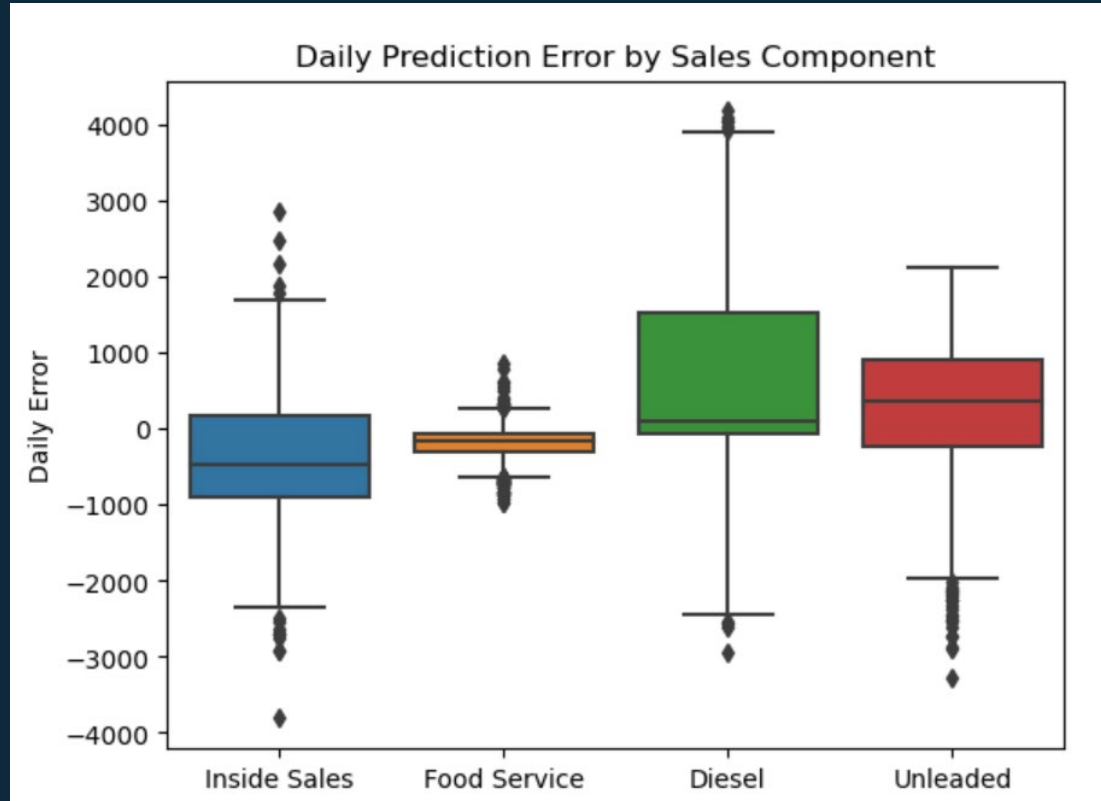
Thank You!

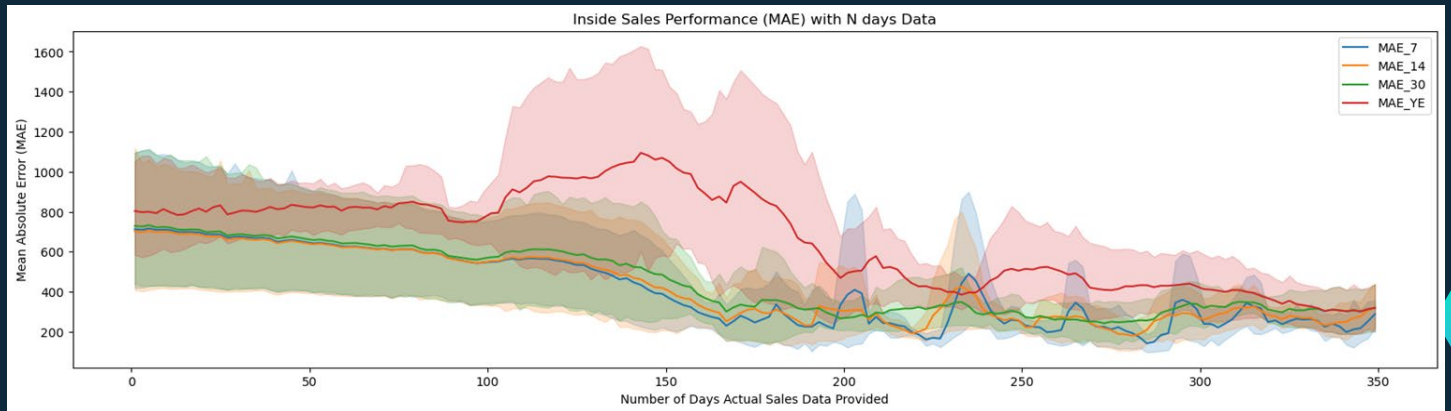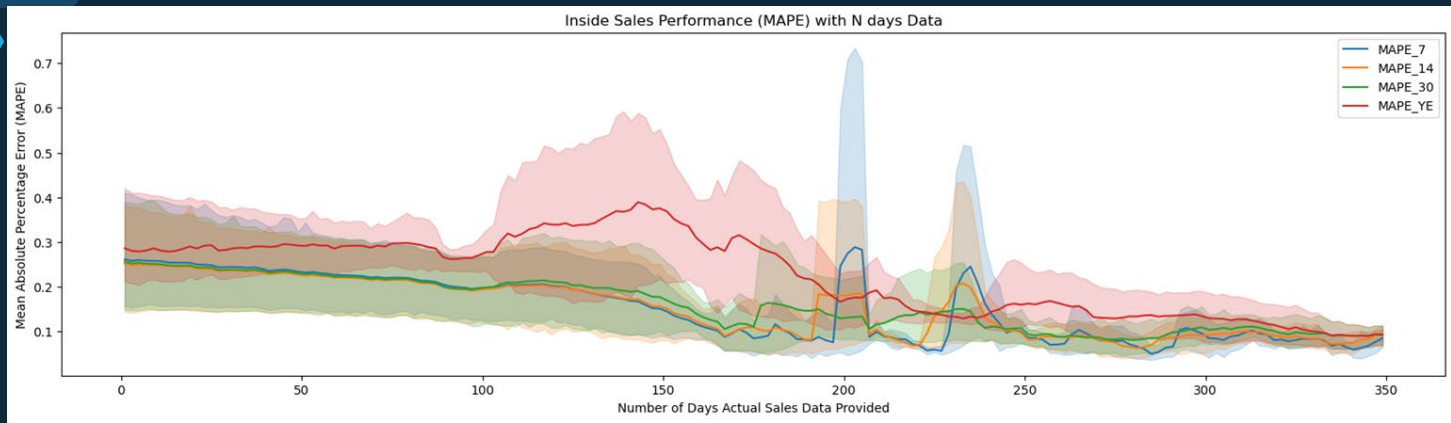# Extra Resources

Benchmark Comparison

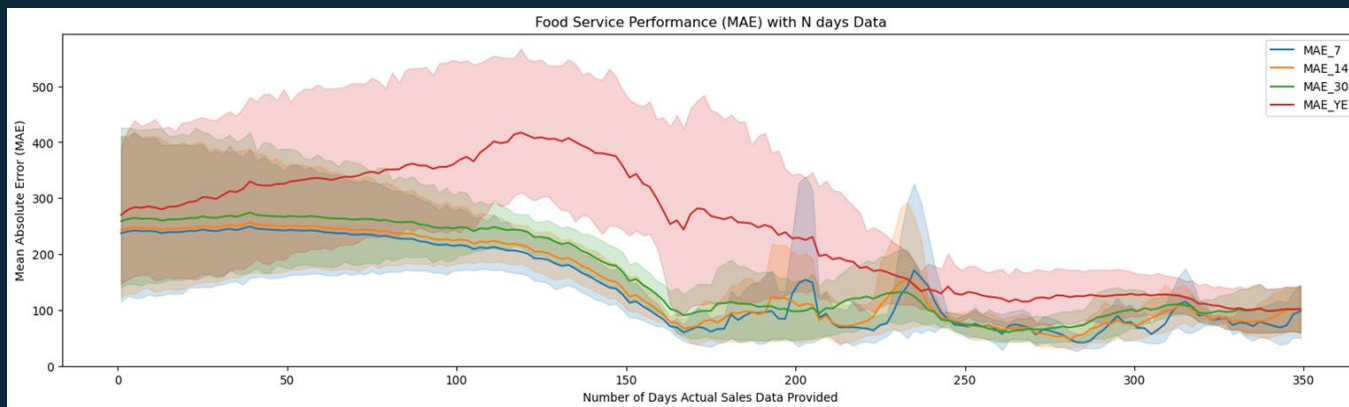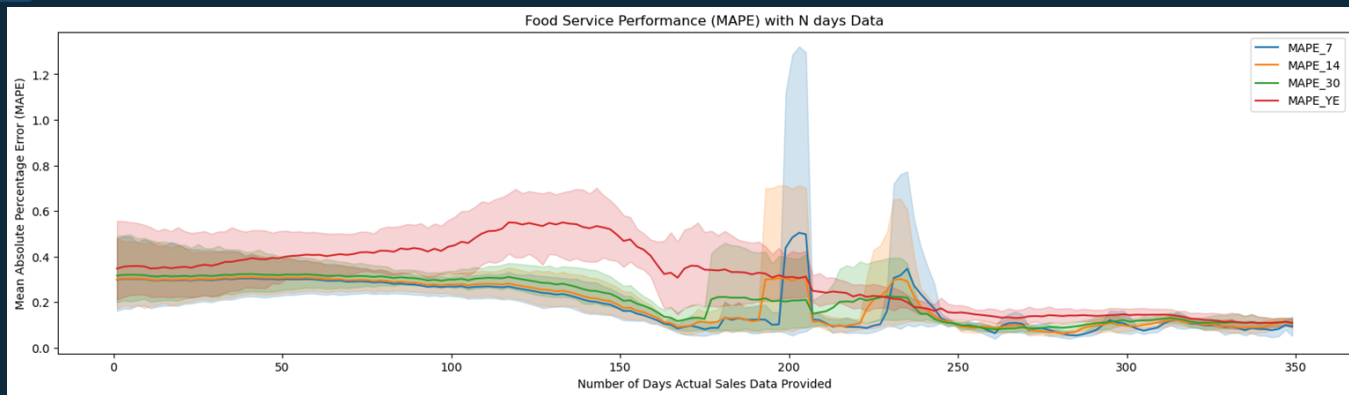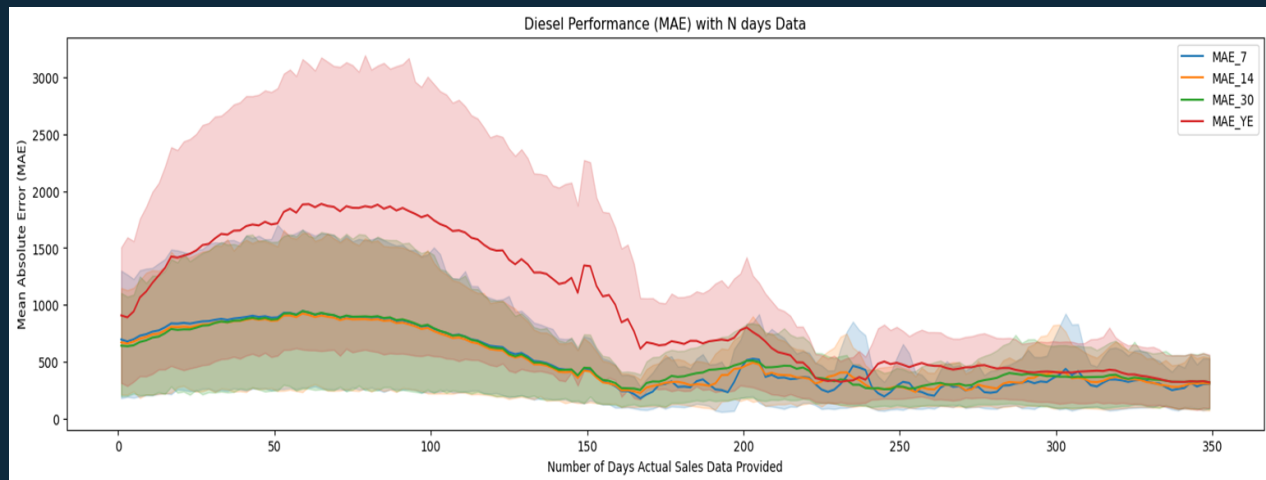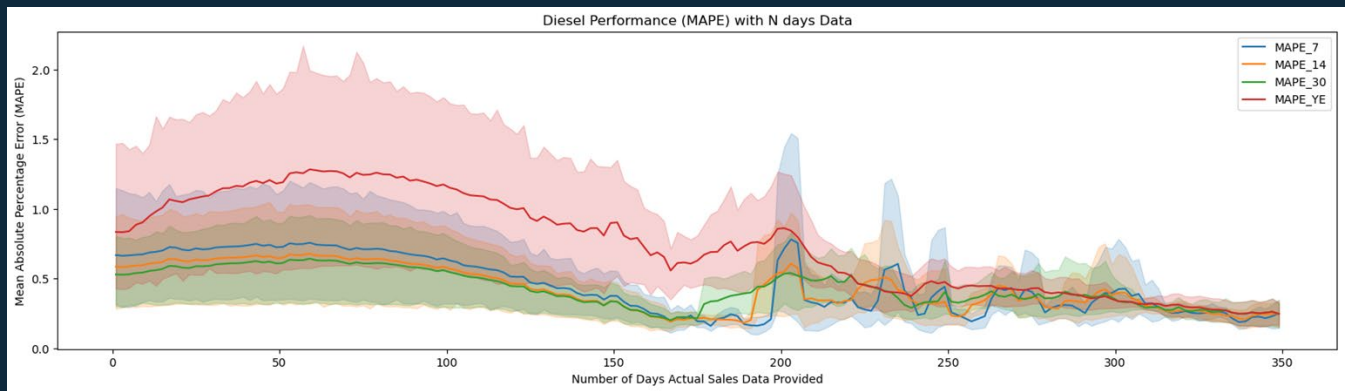|	Inside Sales	Food Service	Diesel	Unleaded
14-day MSE	661003	80331	1600325	831857
14-day RMSE	793	265	996	835
21-day MSE	646277	79586	1450307	824445
21-day RMSE	784	263	951	826
Minimum MSE	241722	40325	5316	150815

Daily Prediction Error by Sales Component

Inside Sales Performance (MAPE) with N days Data

Inside Sales Performance (MAE) with N days Data

Food Service Performance (MAPE) with N days Data

Food Service Performance (MAE) with N days Data

Diesel Performance (MAPE) with N days Data

Diesel Performance (MAE) with N days Data
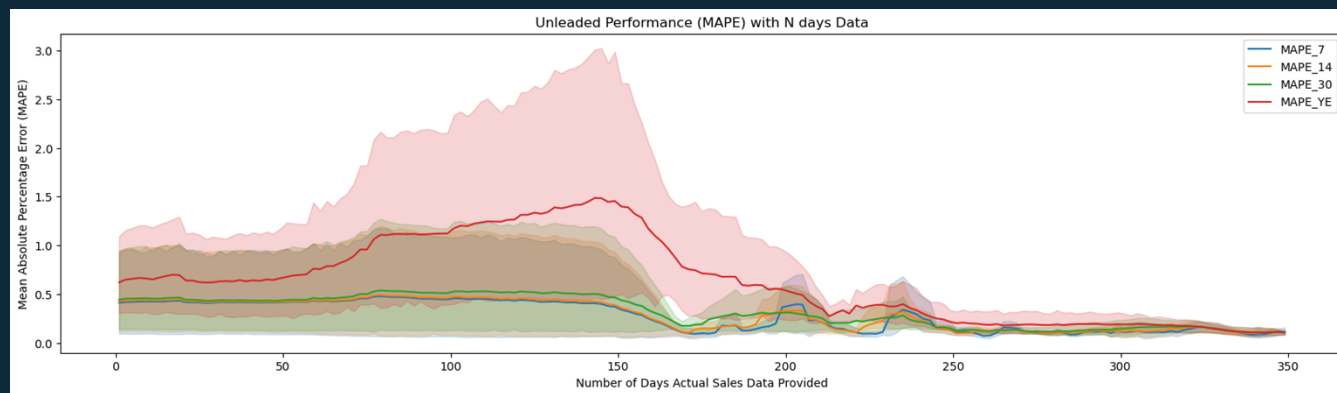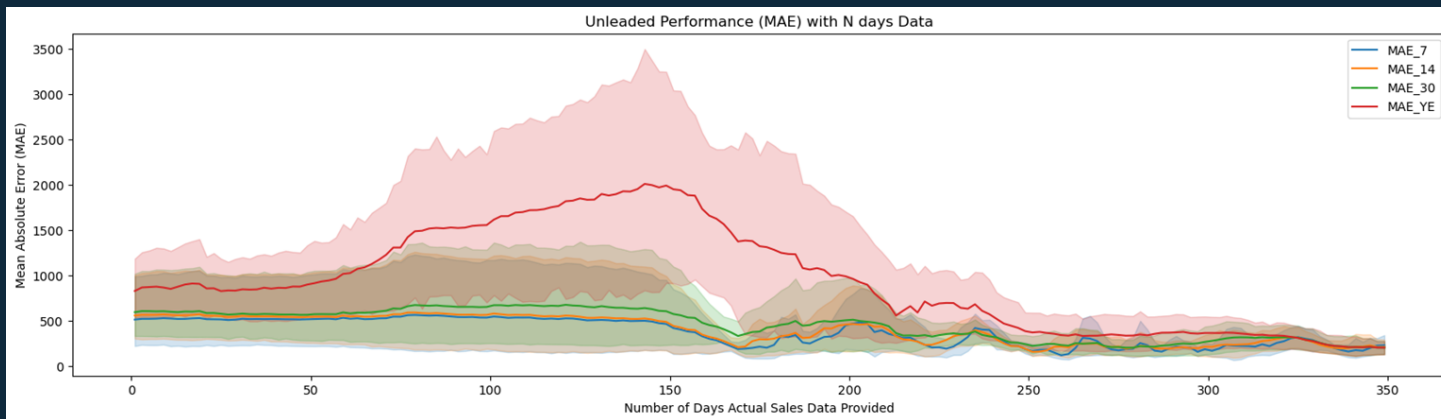
Unleaded Performance (MAE) with N days Data

Unleaded Performance (MAPE) with N days Data

Test Store
XGBoost
Performance

Test Store
XGBoost
Performance

Test Store
XGBoost
Performance
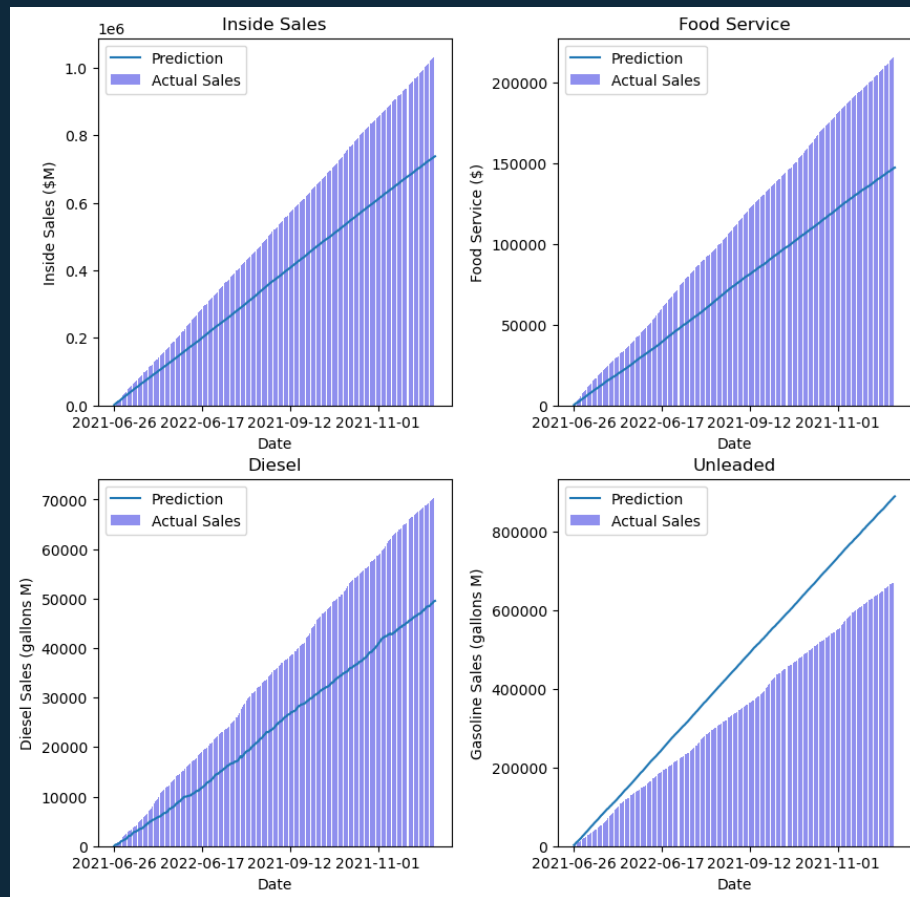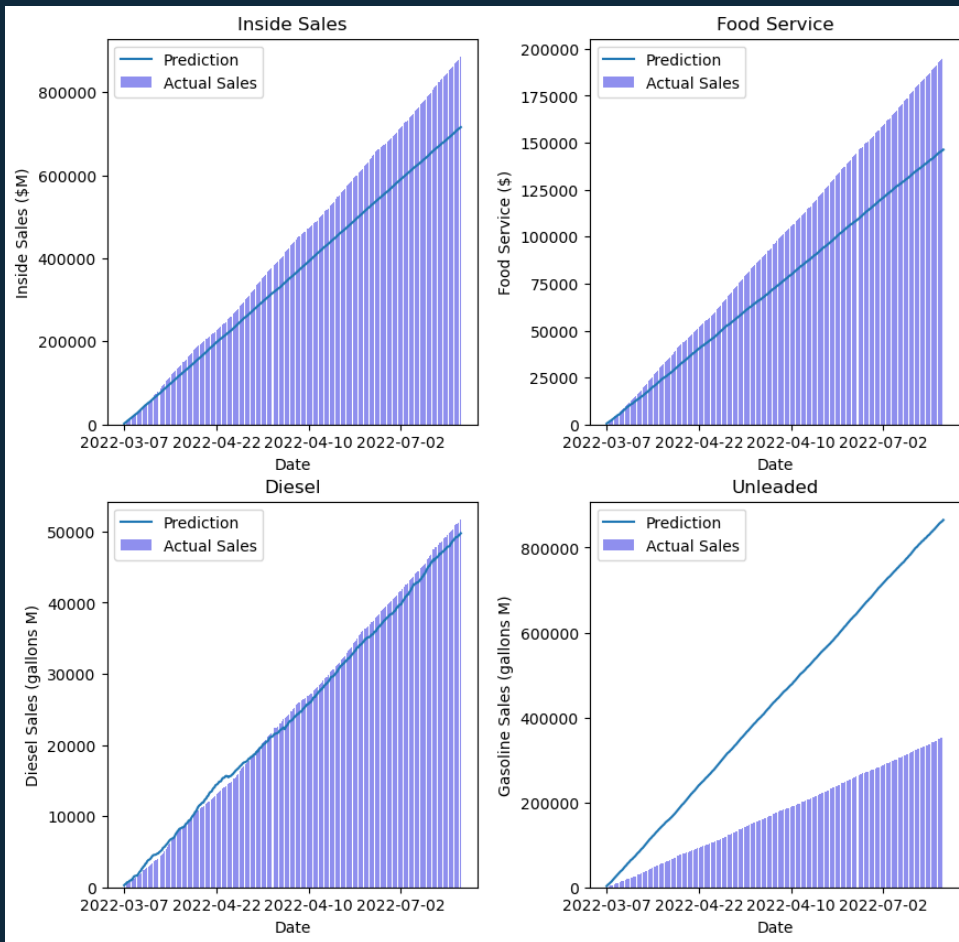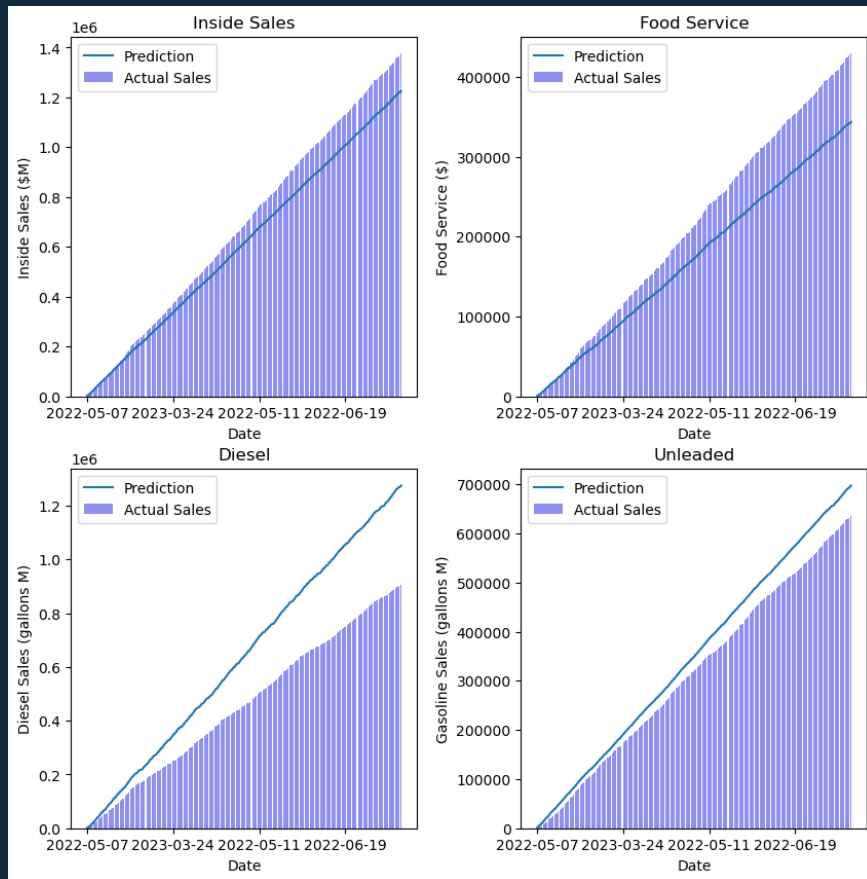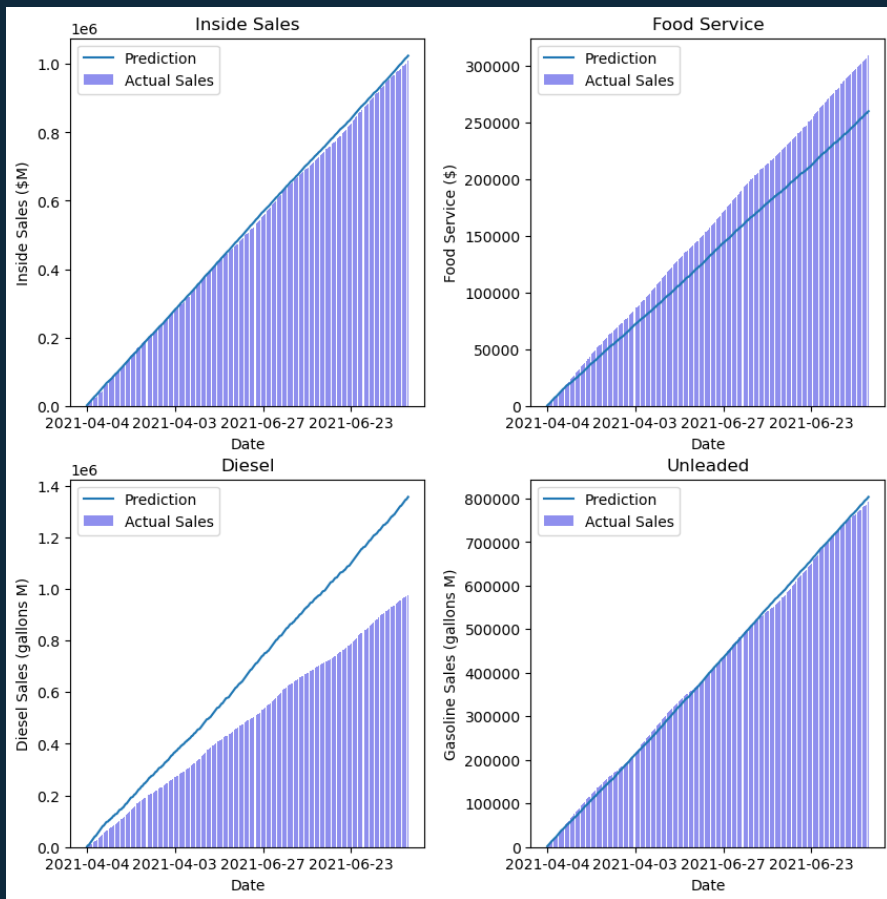
Test Store
XGBoost
Performance

Test Store
XGBoost
Performance

Feature Importance
XGBoost

Feature Importance
XGBoost

Feature Importance
XGBoost

Feature Importance
XGBoost