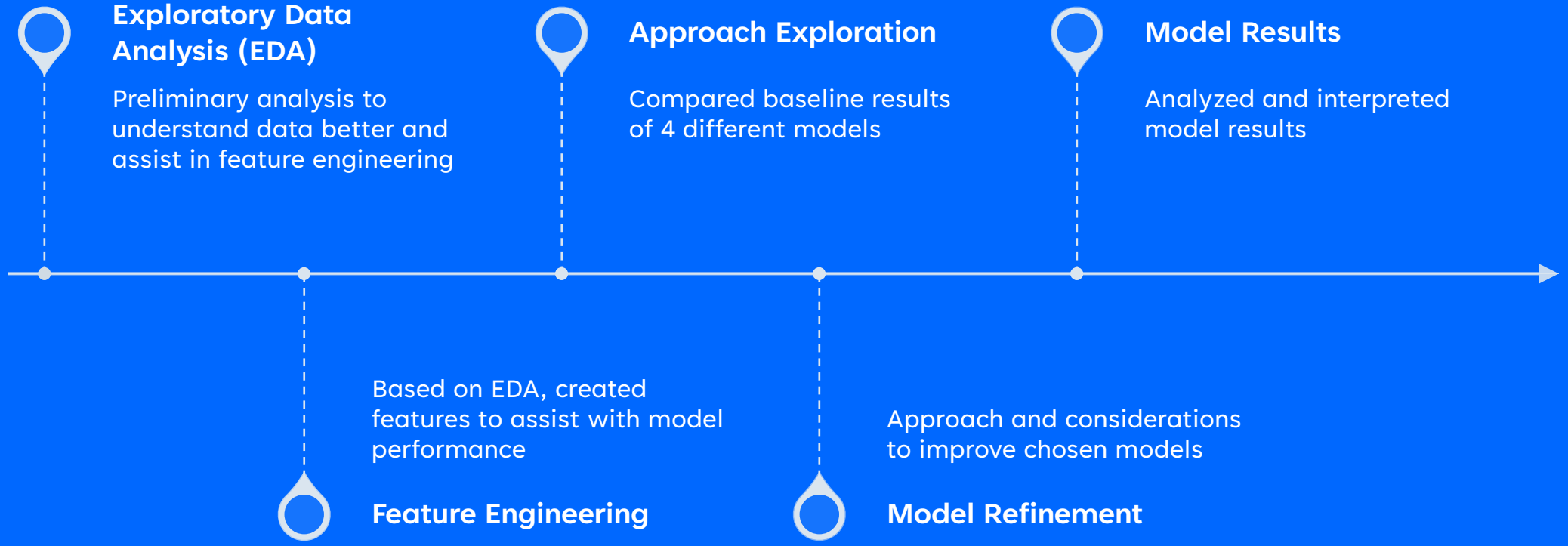
Best Buy Presentation

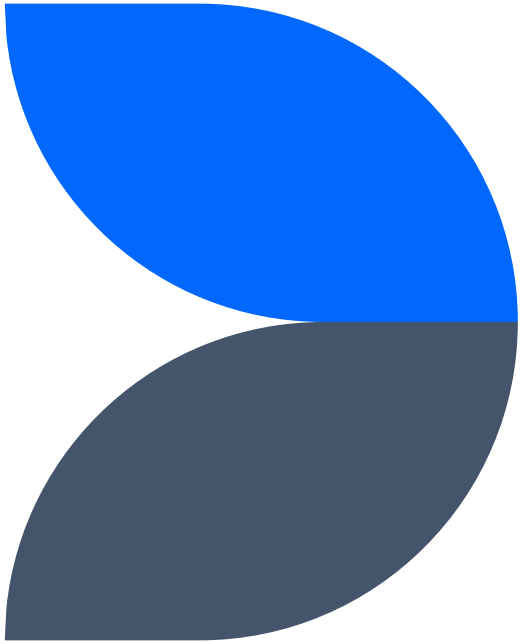
Introduction

- Slow-selling SKUs need forecasts
- Make up a large quantity of SKUs
- Forecasts need to be accurate and efficient

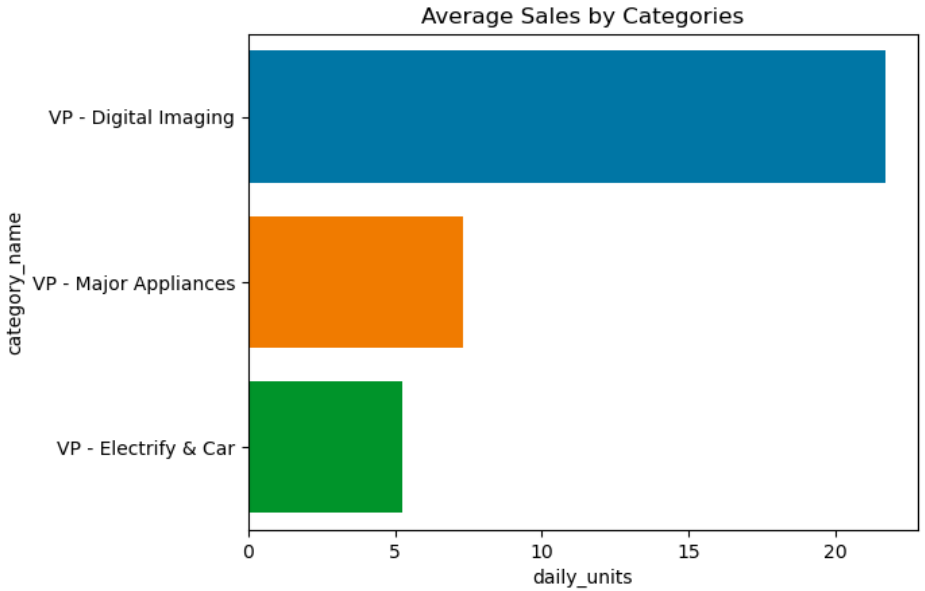
Timeline

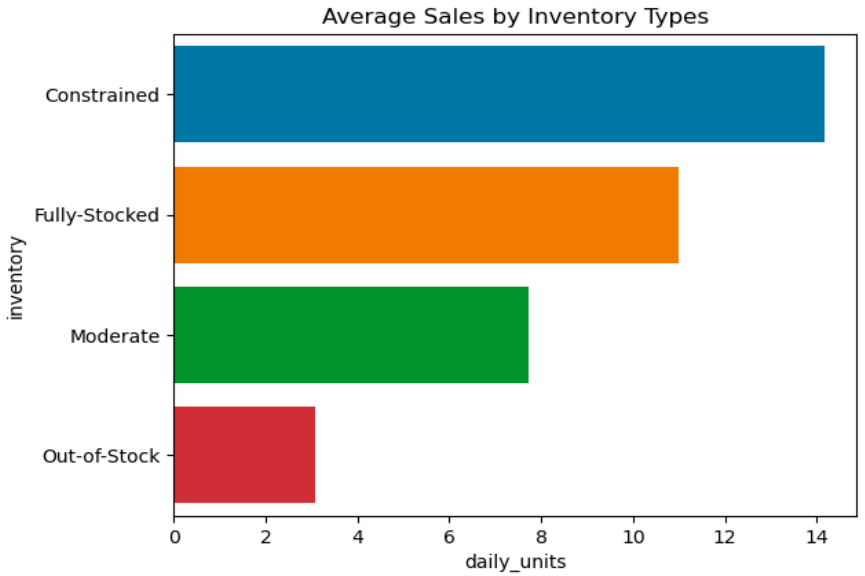

Exploratory Data Analysis (EDA)
Preliminary analysis to understand data better and assist in feature engineering
Approach Exploration
Compared baseline results of 4 different models
Model Results
Analyzed and interpreted model results
Based on EDA, created features to assist with model performance
Feature Engineering
Approach and considerations to improve chosen models
Model Refinement

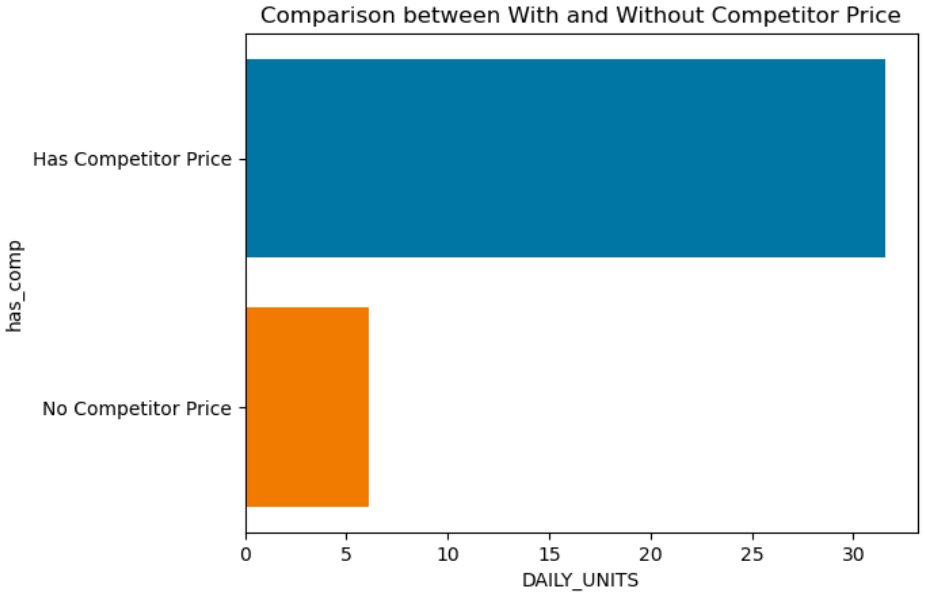
Exploratory Data Analysis

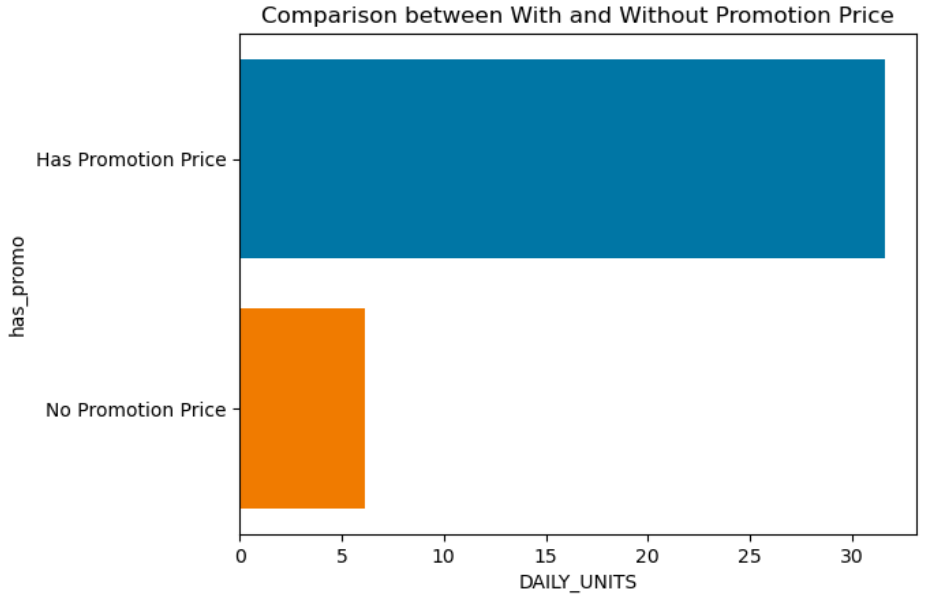


EDA

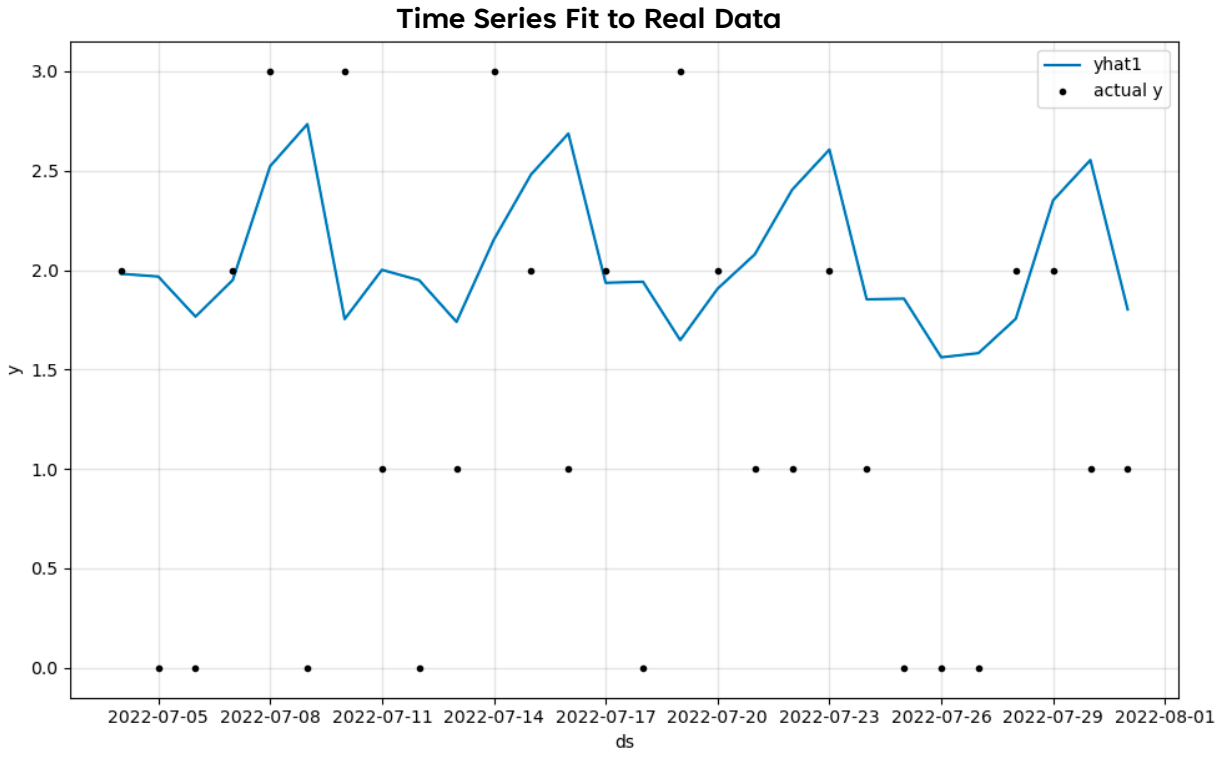

Average Sales by Categories
VP - Digital Imaging
VP - Major Appliances
VP - Electrify & Car
category_name
0
5
10
15
20
daily_units

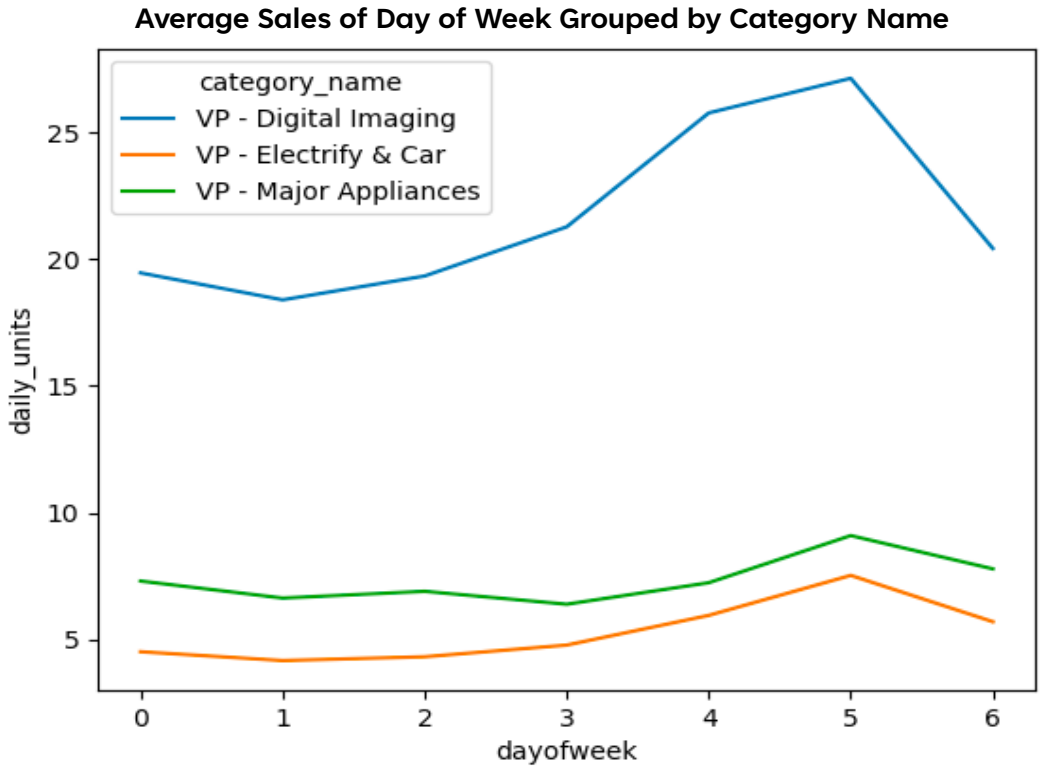

Average Sales by Inventory Types
Constrained
Fully-Stocked
Moderate
Out-of-Stock
inventory
0
2
4
6
8
10
12
14
daily_units


Comparison between With and Without Competitor Price
Has Competitor Price
No Competitor Price
has_comp
0
5
10
15
20
25
30
DAILY_UNITS

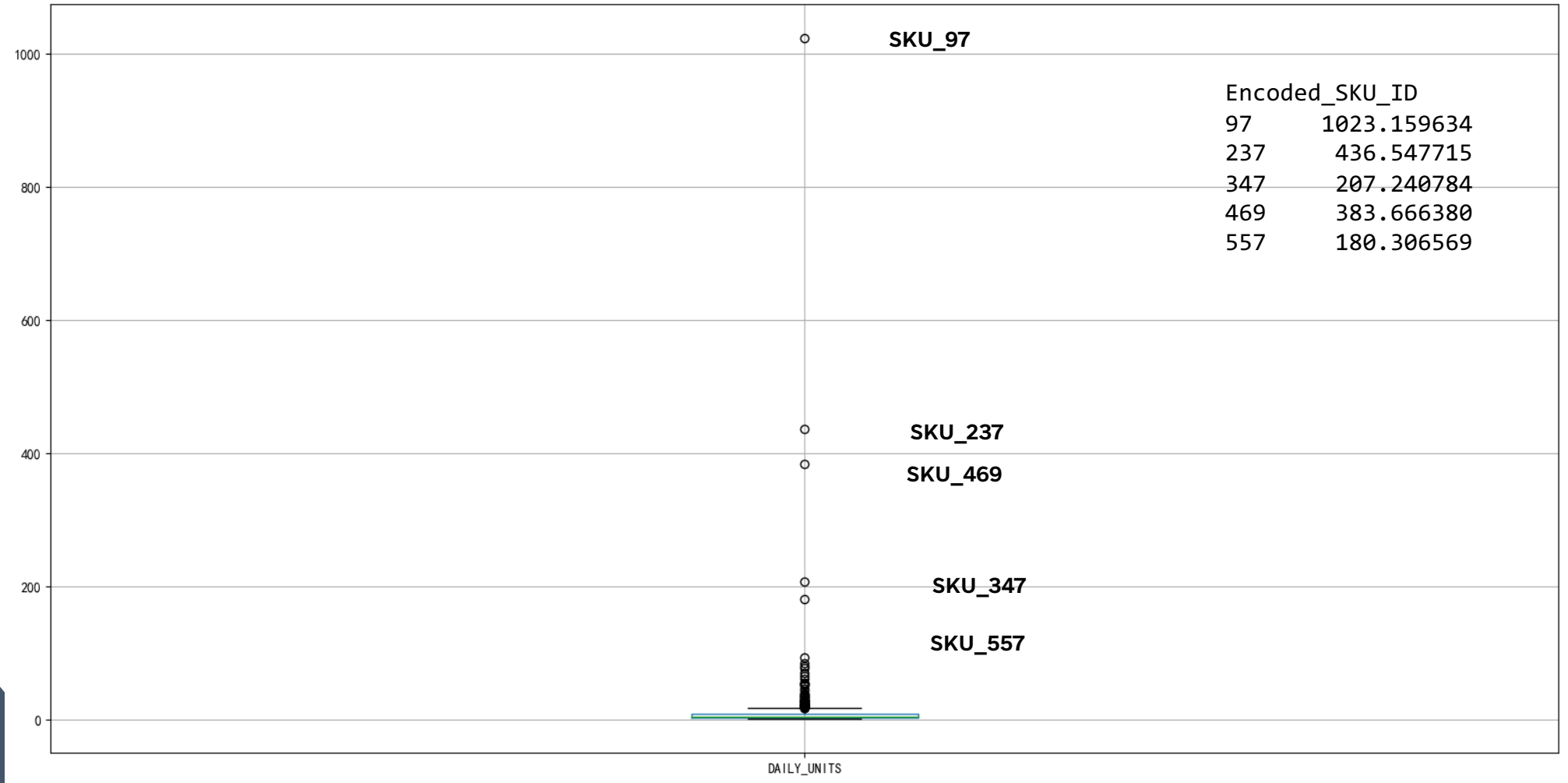

Comparison between With and Without Promotion Price
Has Promotion Price
No Promotion Price
has_promo
0
5
10
15
20
25
30
DAILY_UNITS

EDA continued



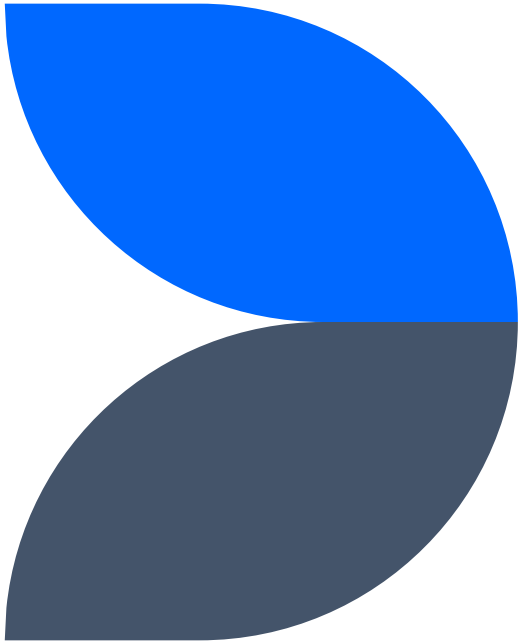


EDA Outlier Analysis



Encoded_SKU_ID	
97	1023.159634
237	436.547715
347	207.240784
469	383.666380
557	180.306569

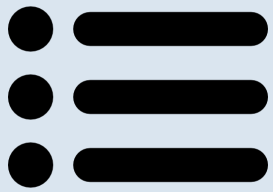
Feature Engineering



Feature Engineering

Category Features

- Sales vary by category
- Ex: Flash memory has much higher sales



Price Features

- Promo prices increase sales
- Having lower price than competitors = higher sales

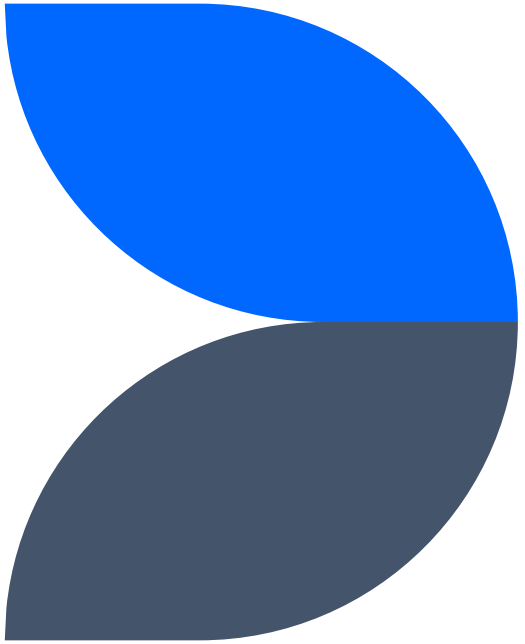
Inventory Features

- Sales slow when items are out-of-stock
- Sales increase when supply is limited

Time Features

- Sales increase during holidays
- Sales increase during weekends

Approaches Exploration



Approaches

Time Series Approach:

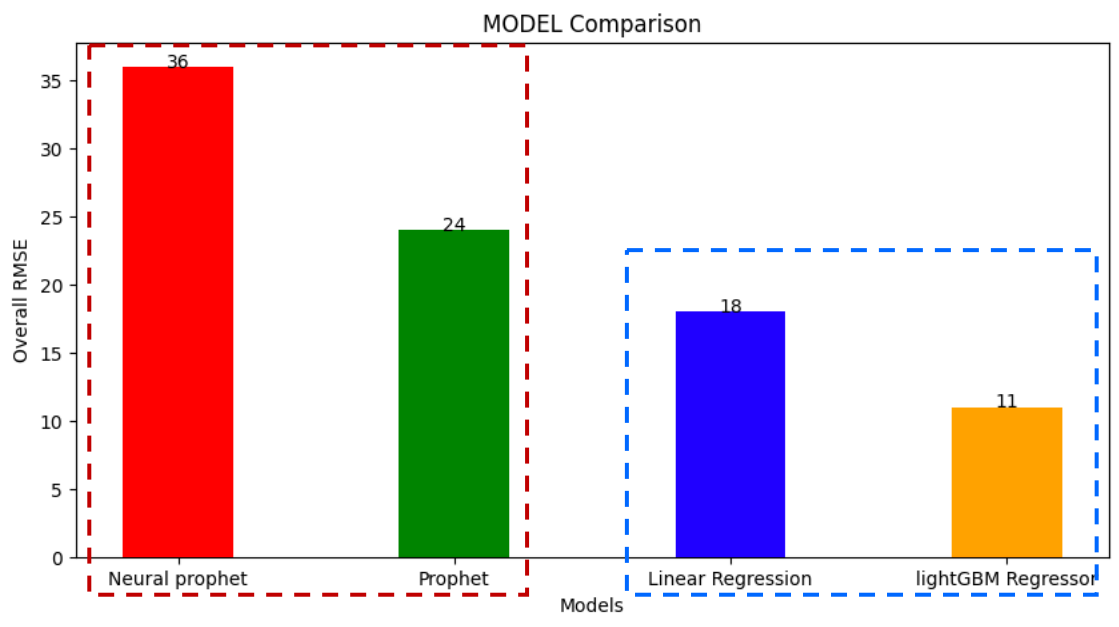
Time series is traditional, but analysis shows that the sales data is too sparse and might not be appropriate.

Regression Approach:

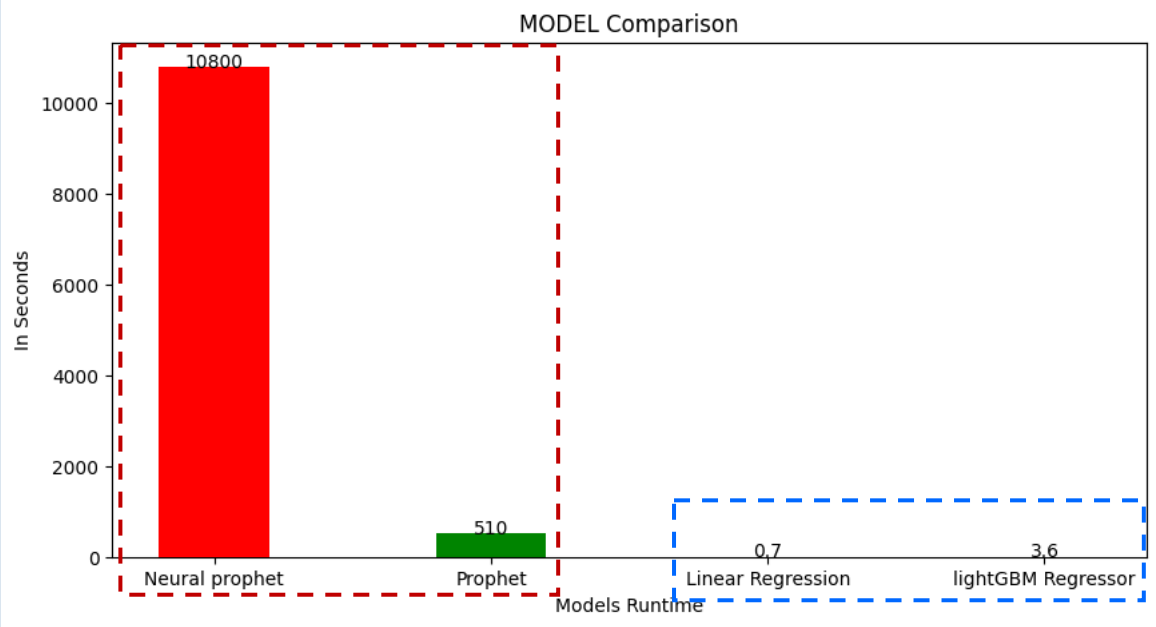
Formulate 7 regression problems for 7 days of weeks due to the clear weekly trend.

We separated data into train (before 07-04-2022) and validation (07-04-2022 to 07-31-2022).

Approach Comparison



RMSE Comparison

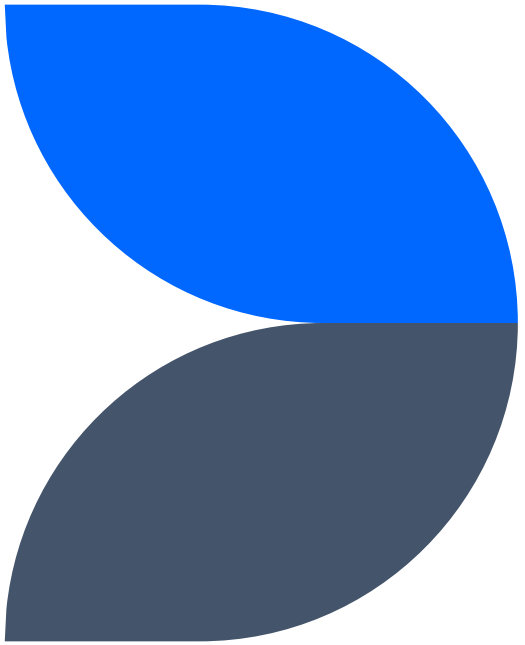


Runtime Comparison (seconds)

Time-series approaches

Regression approaches

Model Refinement



Model Refinement

1

New Features

Added additional features such as: holidays, median sales same category, etc.

2

Outlier Removal

Analyzed models' errors and iteratively improved performance with new features and tuning

3

Feature Selection

Used LASSO for feature selection

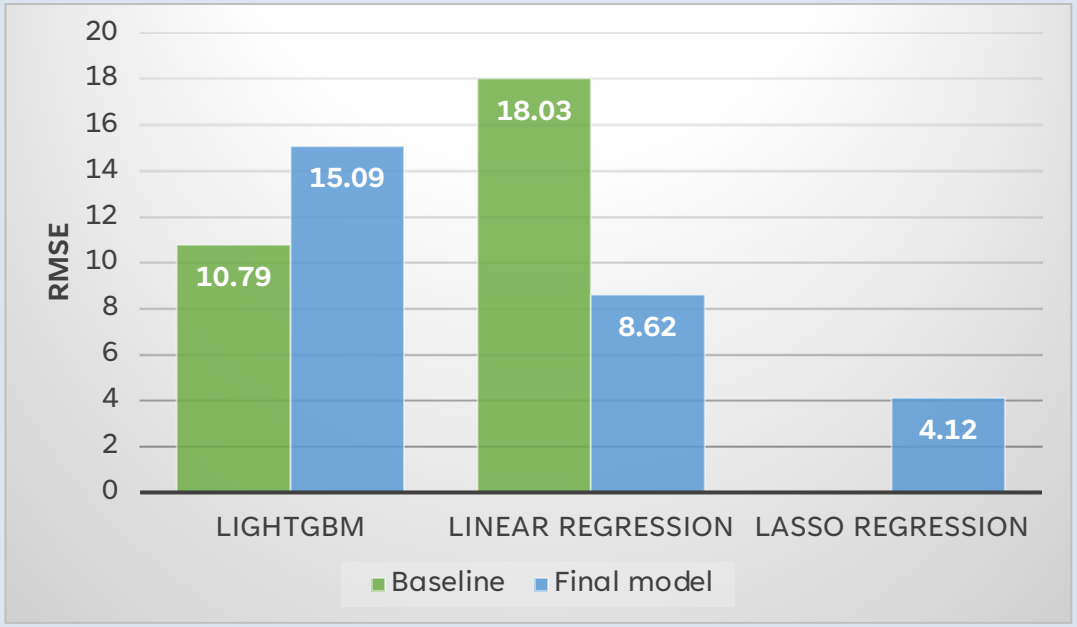
4

Model Tuning

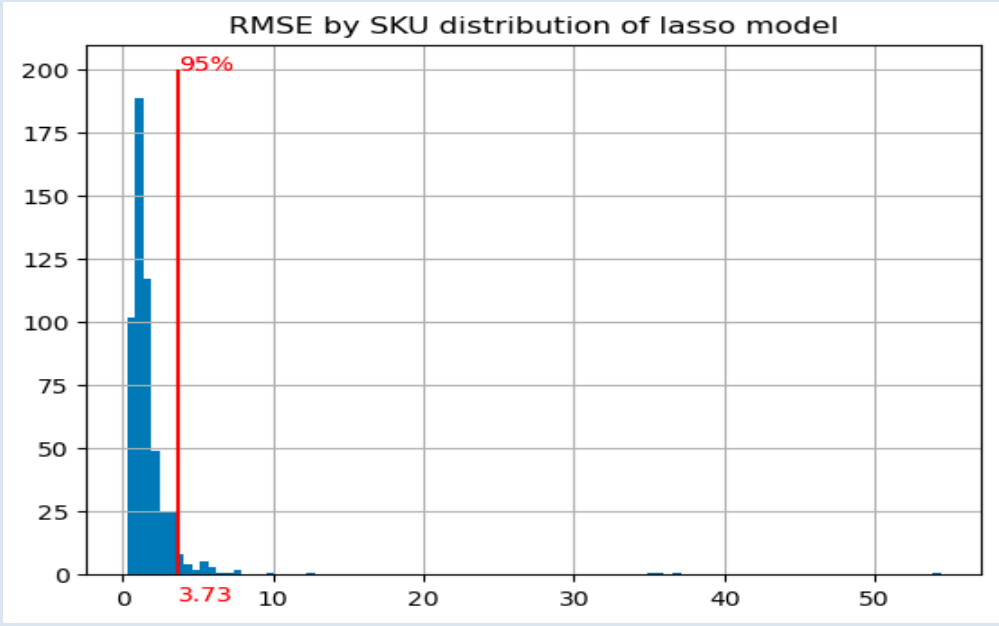
Tuned models' hyper parameters

Model Results

Model Results



RMSE baseline vs Final model

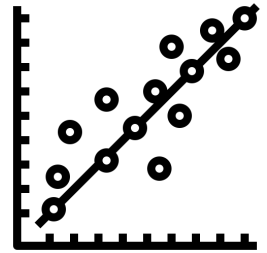


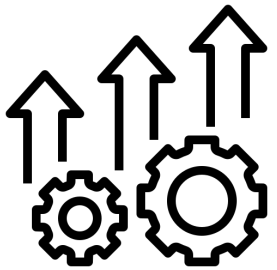
Distribution of RMSE by SKU for Lasso Regression

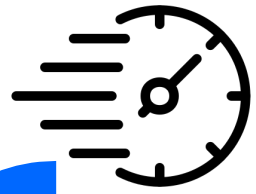
Summary

Summary

- Regression approach outperformed time series approach
- Improvement to 4.12 RMSE & 0.7 second run time
- Can run forecast as frequently as needed







Questions?

> The goal of forecasting is not to predict the future but to tell you what you need to know to take meaningful action in the present.
>
> -Paul Saffo