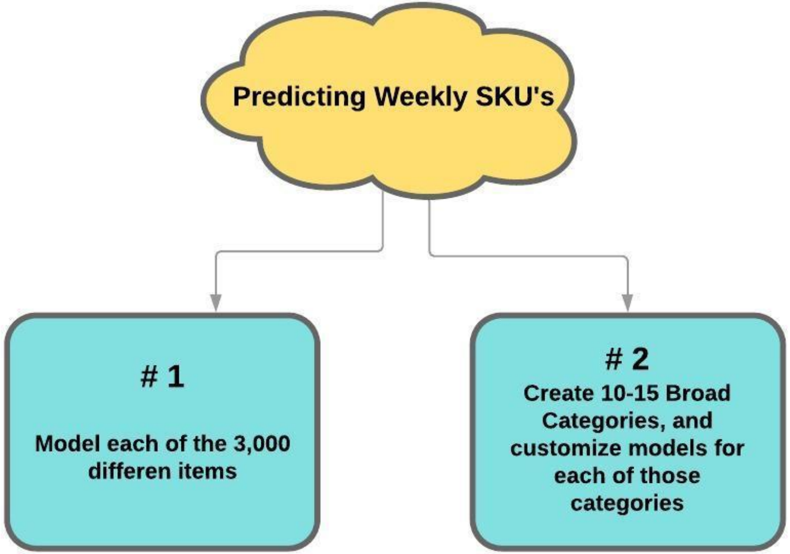
Forecasting Weekly Grocery Product Demand

Capstone 2 By: Kristen Colley

Why?

A local food distribution company buys produce from local farms and then distributes to wholesalers, upscale restaurants, grocery stores, and individuals. This company habitually over and under stocks their goods because they do not have an inventory demand forecast model. Buying decisions are mainly made by gut or innate industry knowledge. There is currently no systematic, data-driven stocking approach in place.

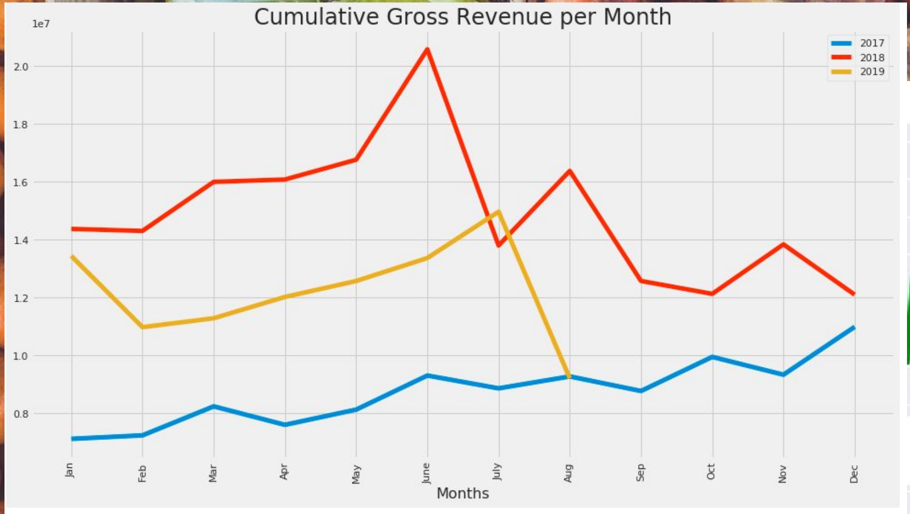
Method

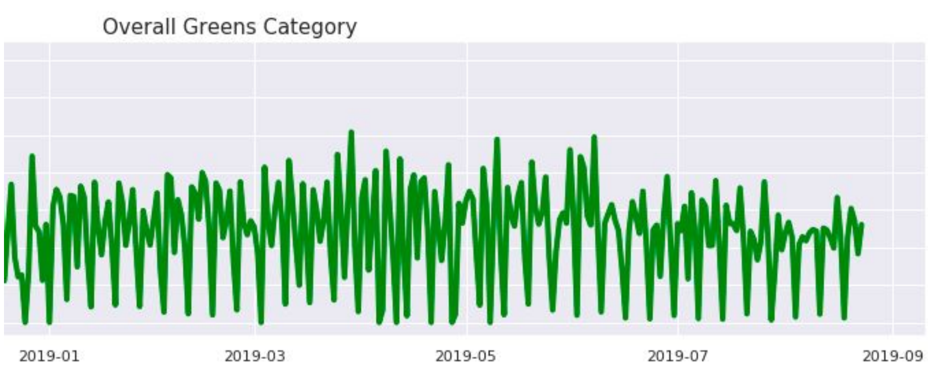


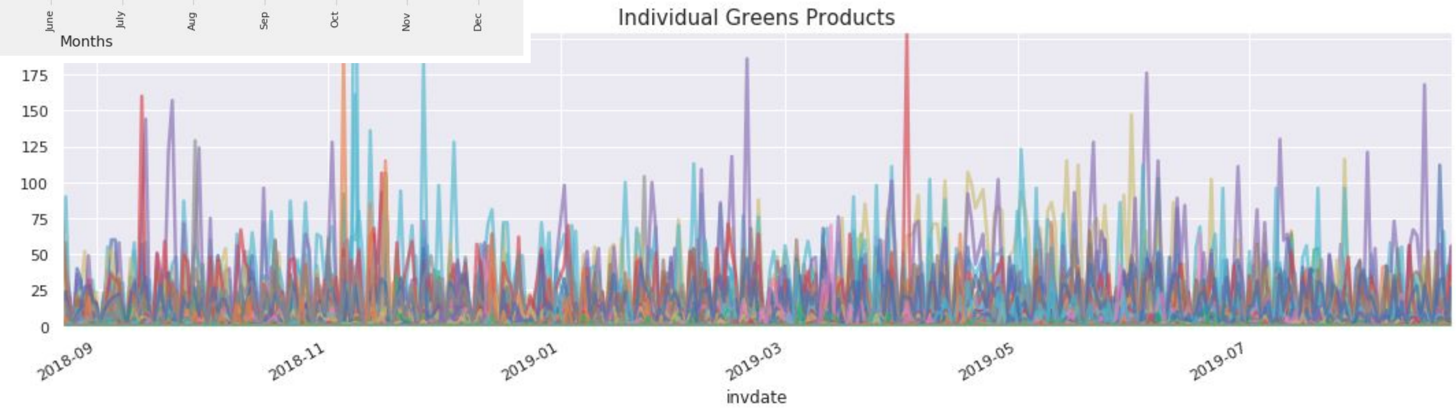
Models

1. Naive Model
2. Classical Models (SMA, SES, HLT, HWM)
3. Prophet (Facebook)
4. Random Forest Regression
5. Light Gradient Boost model

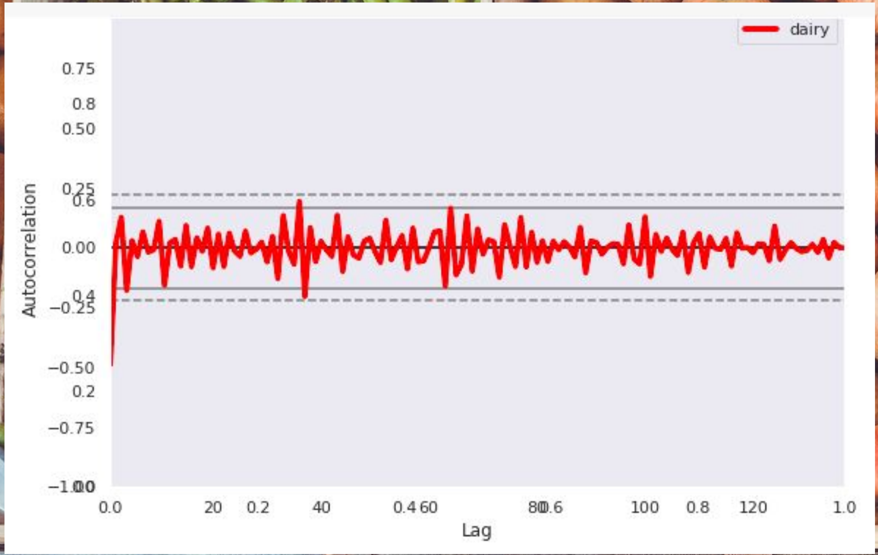
EDA

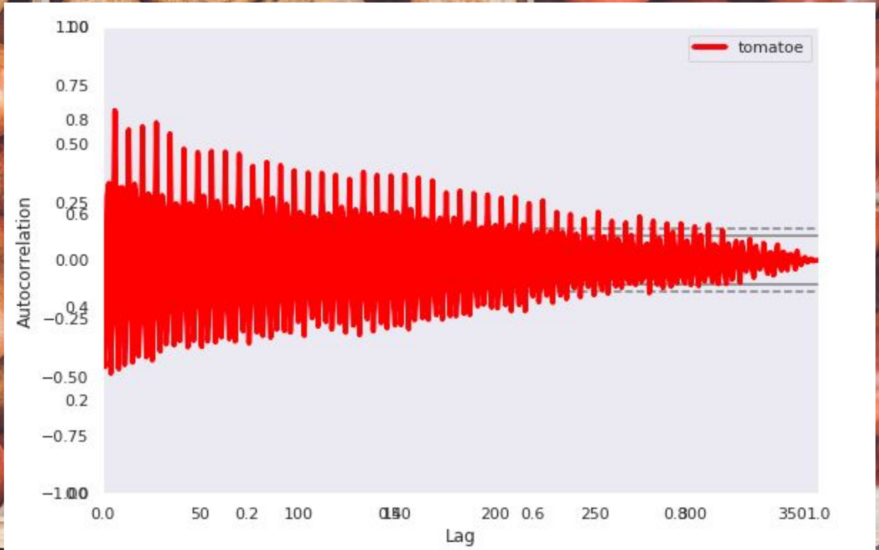

Cumulative Gross Revenue per Month
1e7
2.0
1.8
1.6
1.4
1.2
1.0
0.8
2017
2018
2019
Jan
Feb
Mar
Apr
May
June
July
Aug
Sep
Oct
Nov
Dec
Months


Overall Greens Category
2019-01
2019-03
2019-05
2019-07
2019-09


Individual Greens Products
175
150
125
100
75
50
25
0
2018-09
2018-11
2019-01
2019-03
2019-05
2019-07
invdate

Random Walk Process



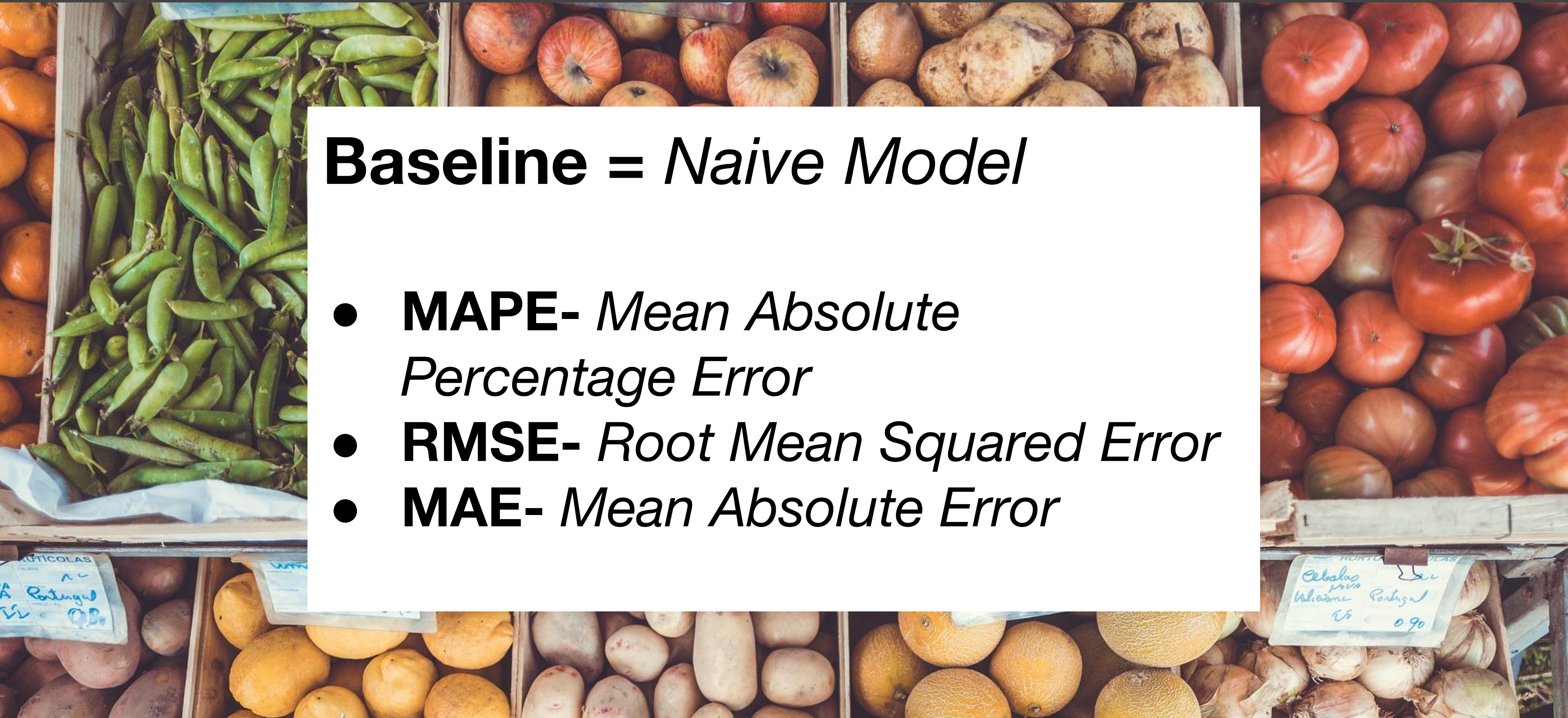


WINNER: ONE YEAR WEEKLY DATA

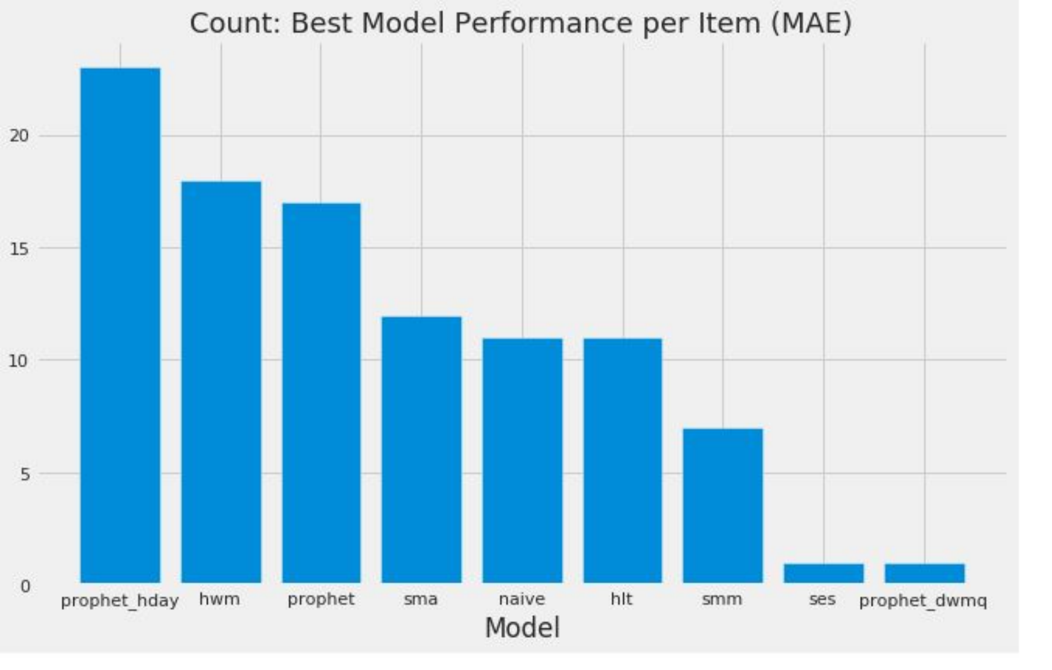
How to Measure Success

Baseline = *Naive Model*

- **MAPE-** *Mean Absolute Percentage Error*
- **RMSE-** *Root Mean Squared Error*
- **MAE-** *Mean Absolute Error*



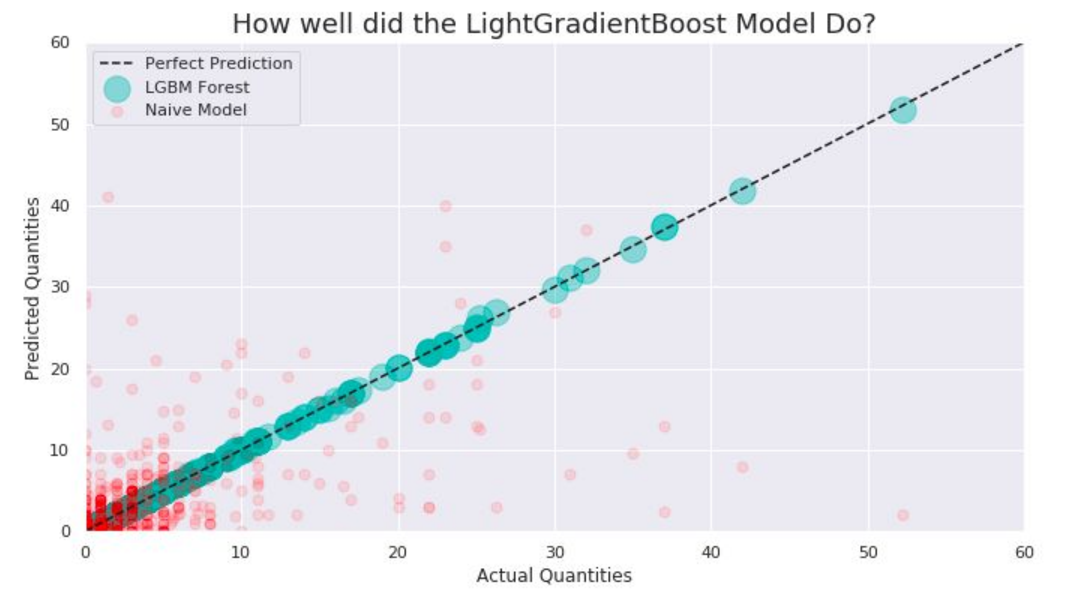
Classical Models + Prophet



IMPROVEMENT OVER NAIVE MODEL=
43%-50%

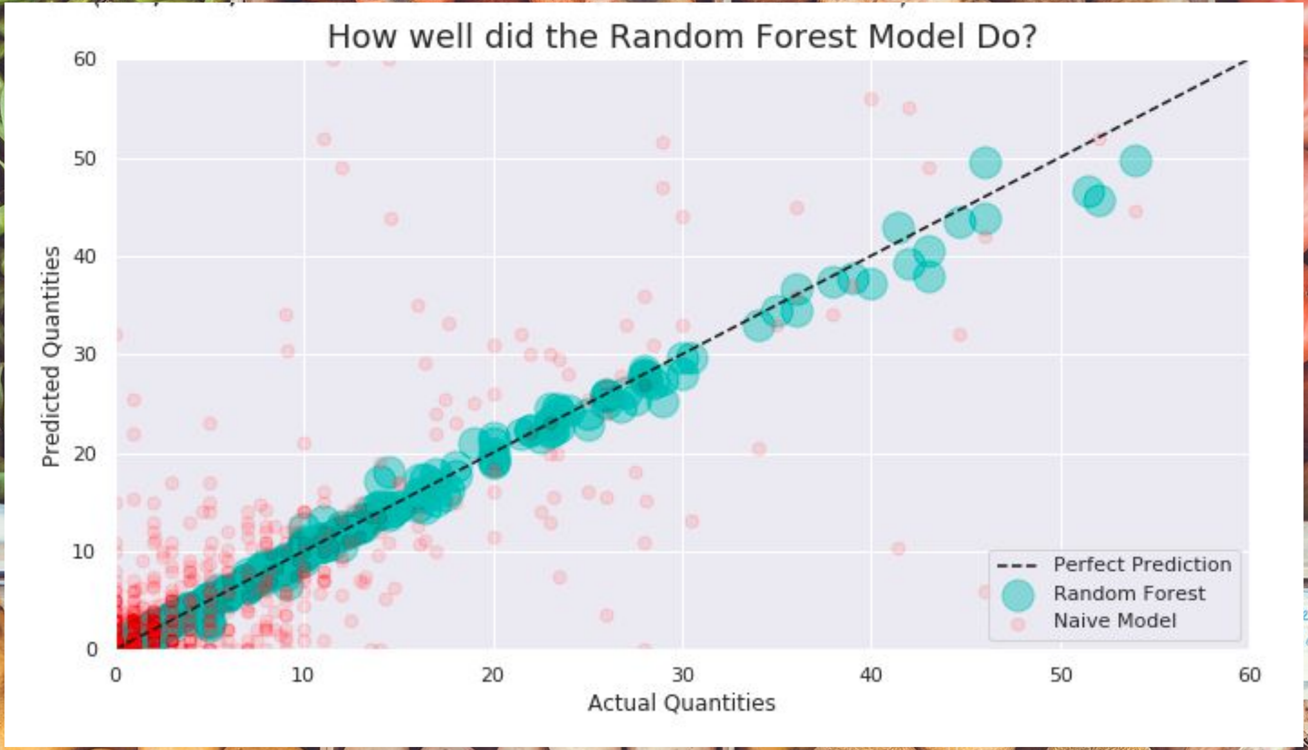
11% STILL NOT BEATING NAIVE MODEL

Trees- LGBM, Random Forest

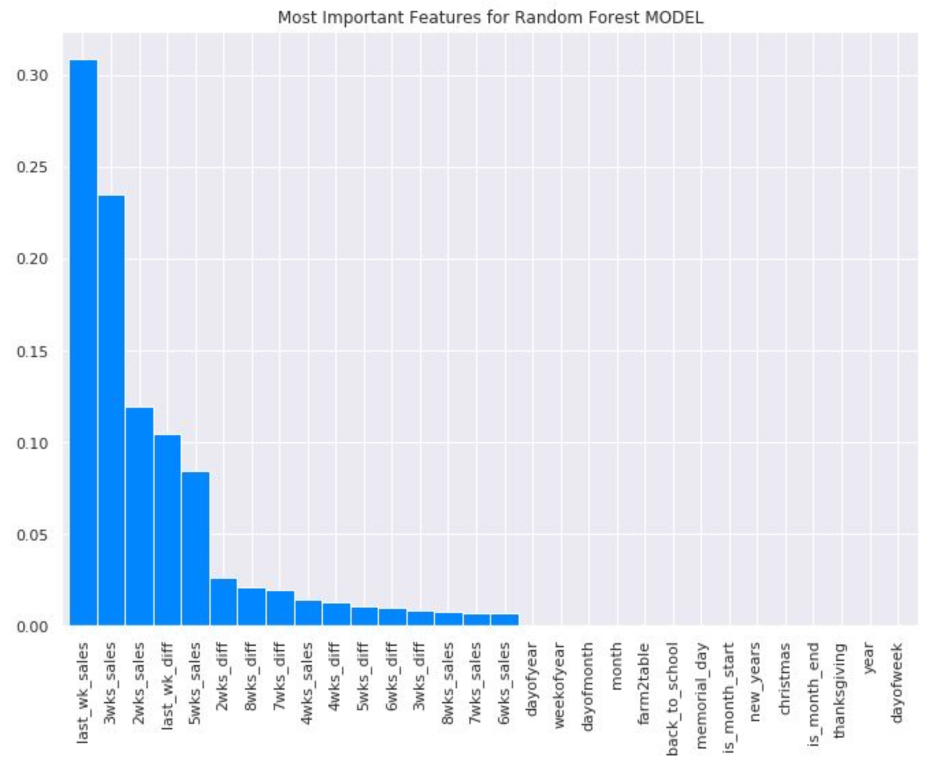


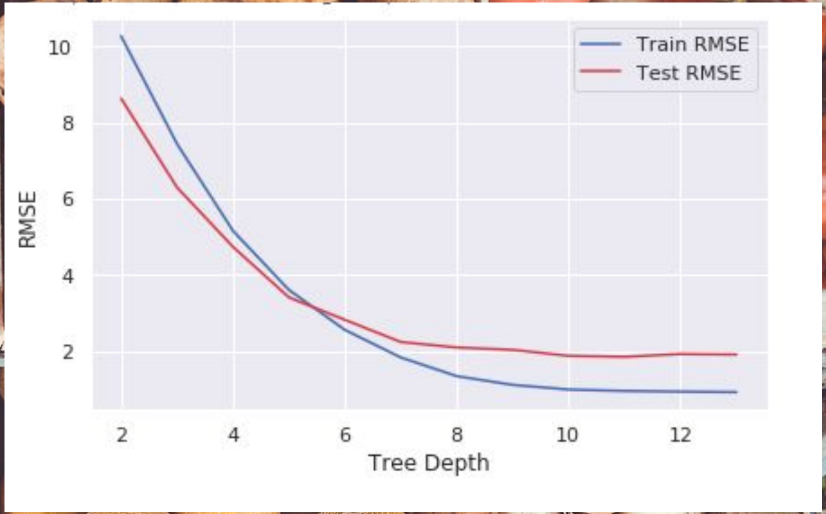
- **Best Models**
- **Beating Naive**
    - 60-88%
- **Can Handle Lots of zeros**
- **LGBM overfitting**

WINNER = Random Forest Regression



Feature Engineering & Parameter Tuning





Conclusion

The Random Forest Regression model was able to beat the customers current naive model by 60-88% (depending on the week). This is a great result for the forecasting modeling space where you don’t really see incredibly accurate results. Also, the most recent sales were the biggest factor influencing the prediction in the RFR model.