Sales Forecasting w/ Time Series

James Benasuli

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusions

Business Problem

- Covid has highlighted the importance of sales forecasting/demand planning capabilities for manufacturers, wholesalers, and retailers
    - Ex: Peloton Covid demand surge CapEx
- Without sales forecasting/demand planning tools, Walmart's marketplace sellers are unable to optimally plan their business activity
    - However implementation of ERPs/CRMs is too resource intensive for many (smaller) Marketplace sellers
- Our team, Walmart Store Sales DS, has been approached by members of the Marketplace Strategy and Product teams
- Can we leverage our expertise to develop an affordable forecasting/planning solution for Walmart Marketplace sellers

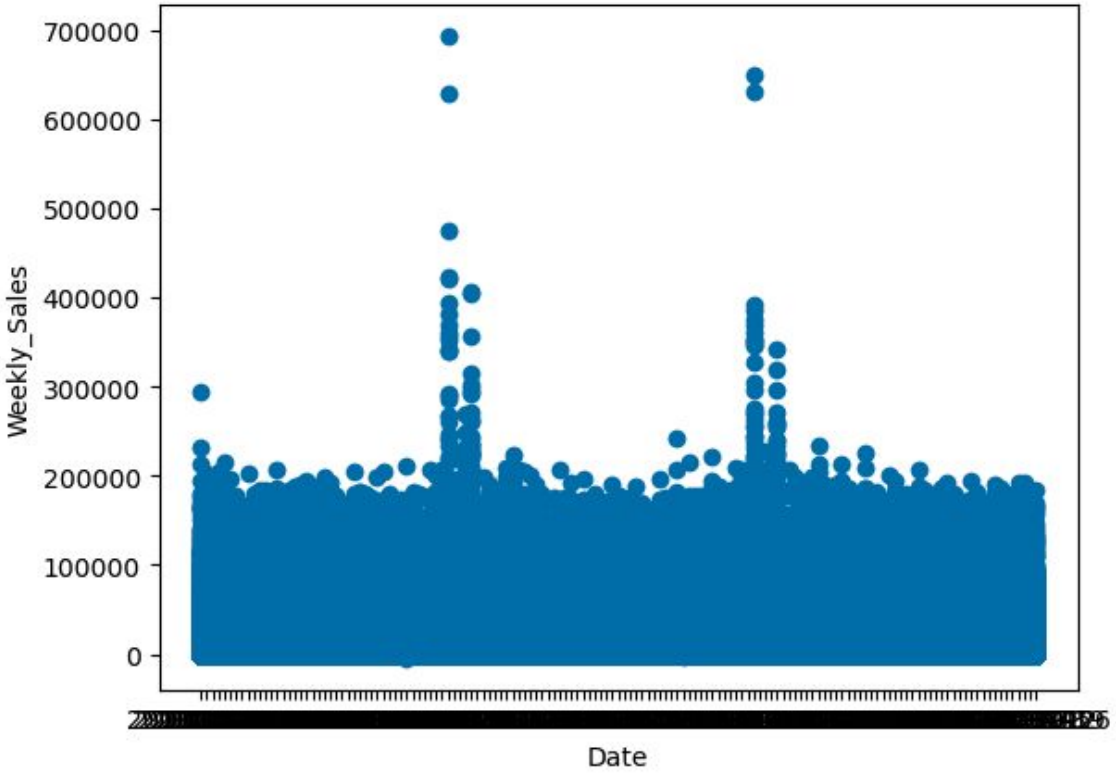
Data

- Historical Walmart store sales data
  - 2.75 yrs (2010-02-05 to 2012-10-26)
- Data was provided by colleagues from finance
  - Certain historical data was deleted prior to handoff to simulate the less than ideal data we expect our sellers to have at their disposal
- Target variable is Weekly Sales
- Features include various internal and external factors
  - Internal ex: store type, size, markdown levels, promotional periods
  - External ex: CPI, temperature, gas prices

Why are certain factors key to accurate forecasts

Certain factors can have huge effects on consumer demand and store sales

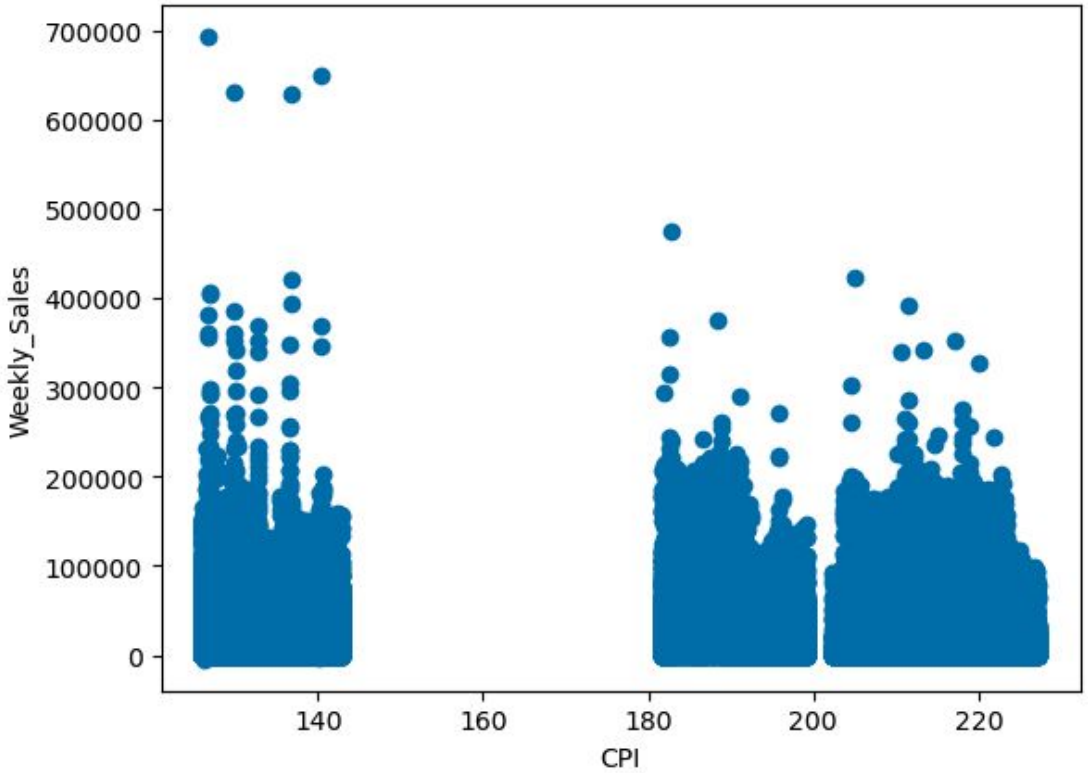
- As evidenced here by the distribution of sales over a year



Why are certain factors key to accurate forecasts

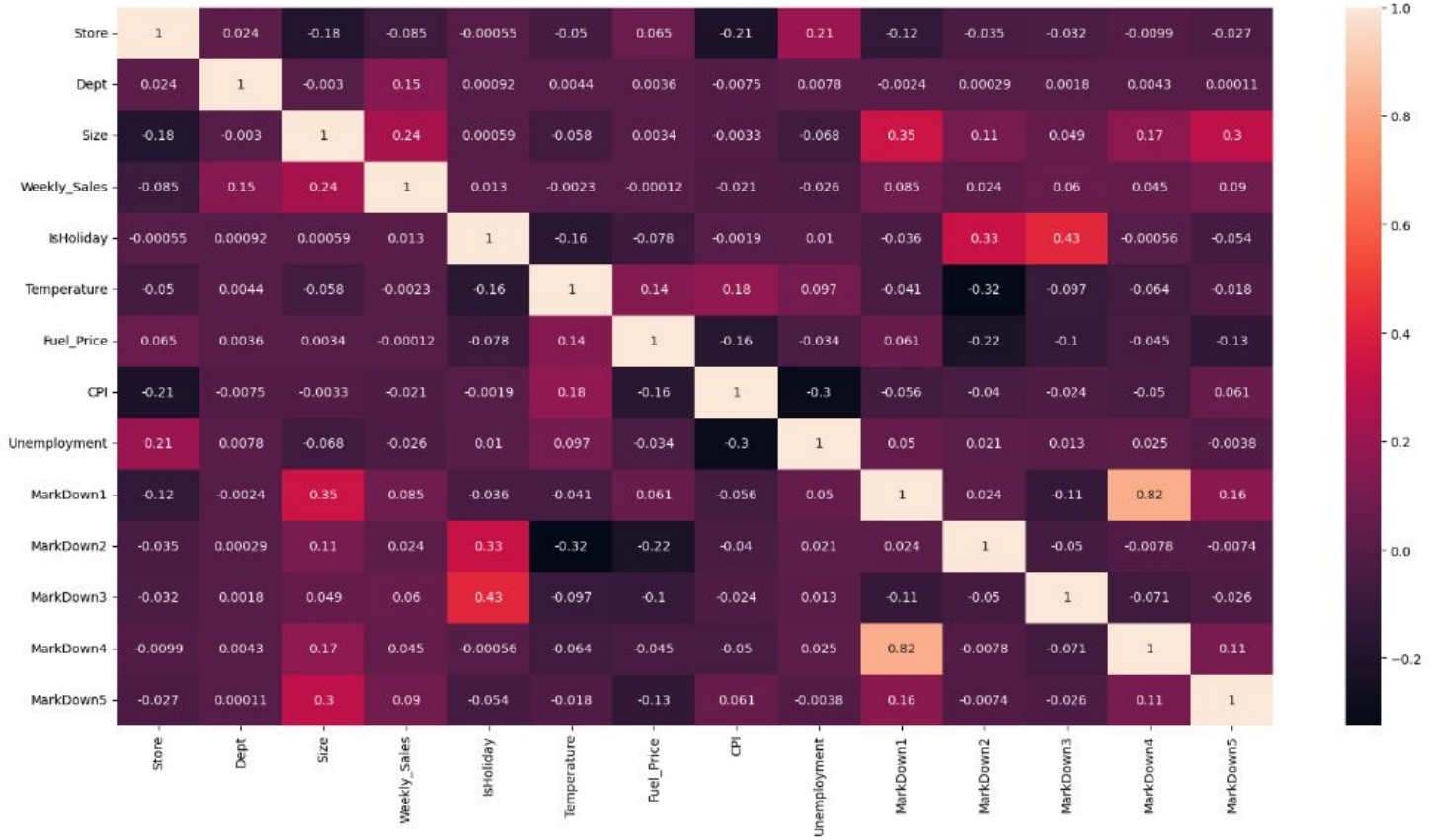
Other factors one might assume would have a large effect actually might not

- As evidenced here by the distribution of sales over different values of CPI (consumer price index aka inflation)



So what features should we focus on?

Feature correlation heatmap

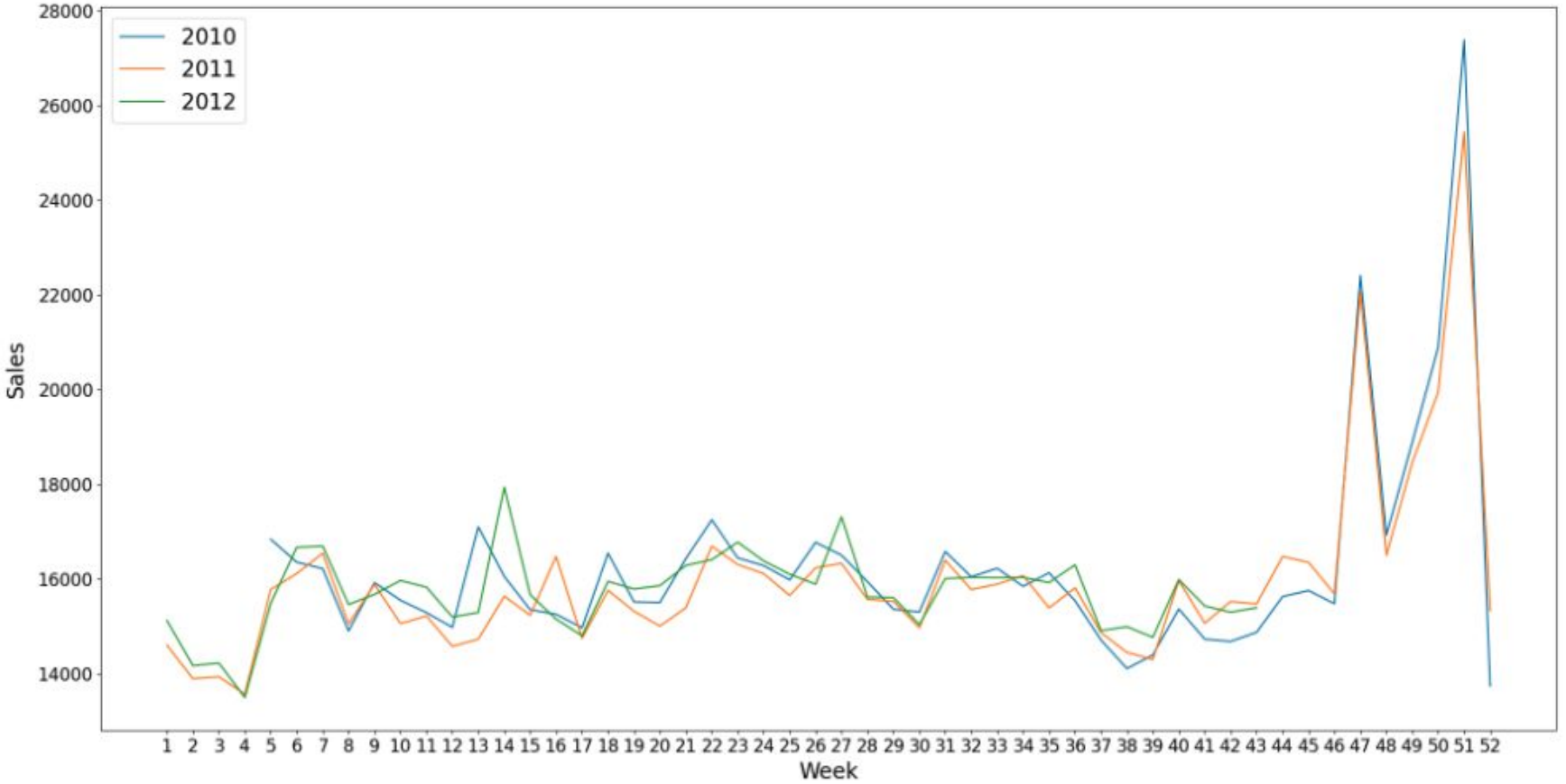


Methods

As nothing popped out on our heatmap we employed the following methods:

- General feature engineering
  - Ordinally encoded the store type variable
  - Extracted the year, month, and week from the date associated with each record
- Holiday/promotional period feature engineering
  - Our experience in store sales has emphasized the importance of considering holidays/promotional periods when planning business activity
- Basic pipelines and column transformers were built to iterate through different models and scaling techniques
  - This architecture will allow us to scale more complicated transformation in the future
- Pipelines were passed to random forest regressors for modeling

Holidays Drive Variation in Sales Performance



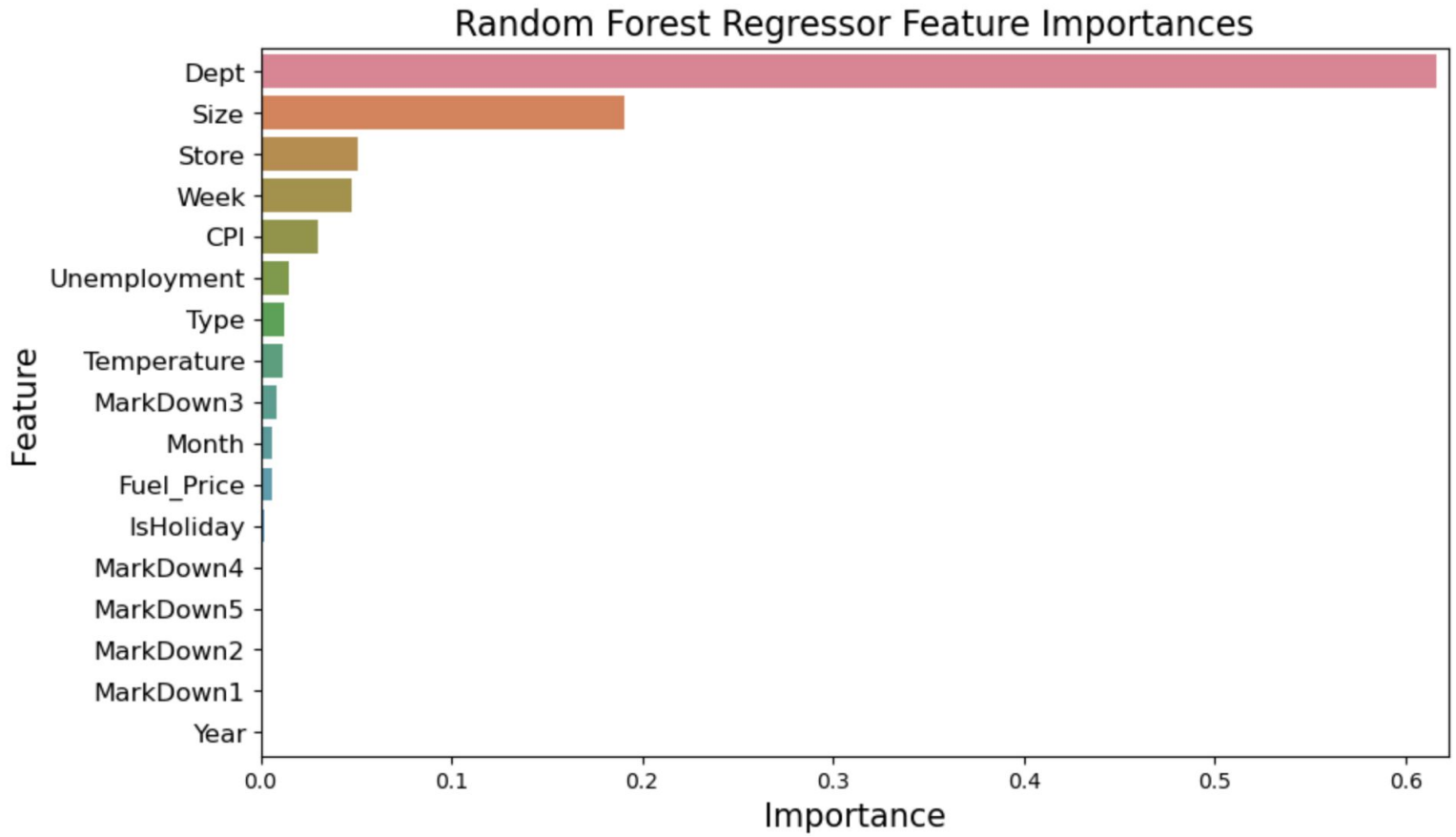
Modeling Evaluation Metrics

- RMSE
  - Root mean squared error
  - Standard deviation of the residuals (prediction errors)
  - Gives more weight to large errors
- MAE
  - Mean absolute error
  - Calculated as the sum of absolute errors divided by the sample size
  - Easier to interpret
- Features of both metrics
  - In the unit of the target variable
  - Non negative and disregard error direction
  - Lower the better

Model Results

- Baseline RandomForestRegressor (all features)
  - RMSE: 5,030.87
  - MAE: 2,296.4
- RFR w/ scaled (MinMax) features
  - RMSE: 5,037.66
  - MAE: 2,295.92
- Only date and holiday features (clearly we are benefitting from other features)
  - RMSE: 22,113.89
  - MAE: 14,879.50

Random Forest Regressor Feature Importances



Model Results

- Using top 10 features from Random Forest feature importance attribute
  - RMSE: 5,063.34
  - MAE: 2,294.50
- Manually selecting only store feature and date features
  - RMSE: 3,545.91
  - MAE: 1,672.03

Conclusions

- Even with incomplete historical sales data, we can cheaply and reliably forecast future sales
- This tool can provide material benefit for Marketplace sellers
  - Differentiator for Walmart compared to competitor such as Amazon
  - Equipping sellers with forecasting tool ultimately benefits our online customers - Walmart.com's most important stakeholder
- With continued work on our models we are confident in our ability to further reduce RMSE/MAE
- Next steps:
  - Additional feature engineering
    - Identify additional date related factors potentially driving sales fluctuations
    - Are there external factors, which might be affecting sales which he haven't considered/ have data for (e.g. GDP, local cost of living metrics, equity market values)?
  - Construct separate models for different Marketplace Seller Segments
    - Not all sellers will have a product mix which tracks the traditional holiday calendar
    - Ex: pool/beach supply sellers


+6.31%
THANK YOU!
+5.27%
+6.31%
+7.12%
-3.54%