

# Capstone\_Presentation

Demand Forecast Project

Presented by - Swati Sharma

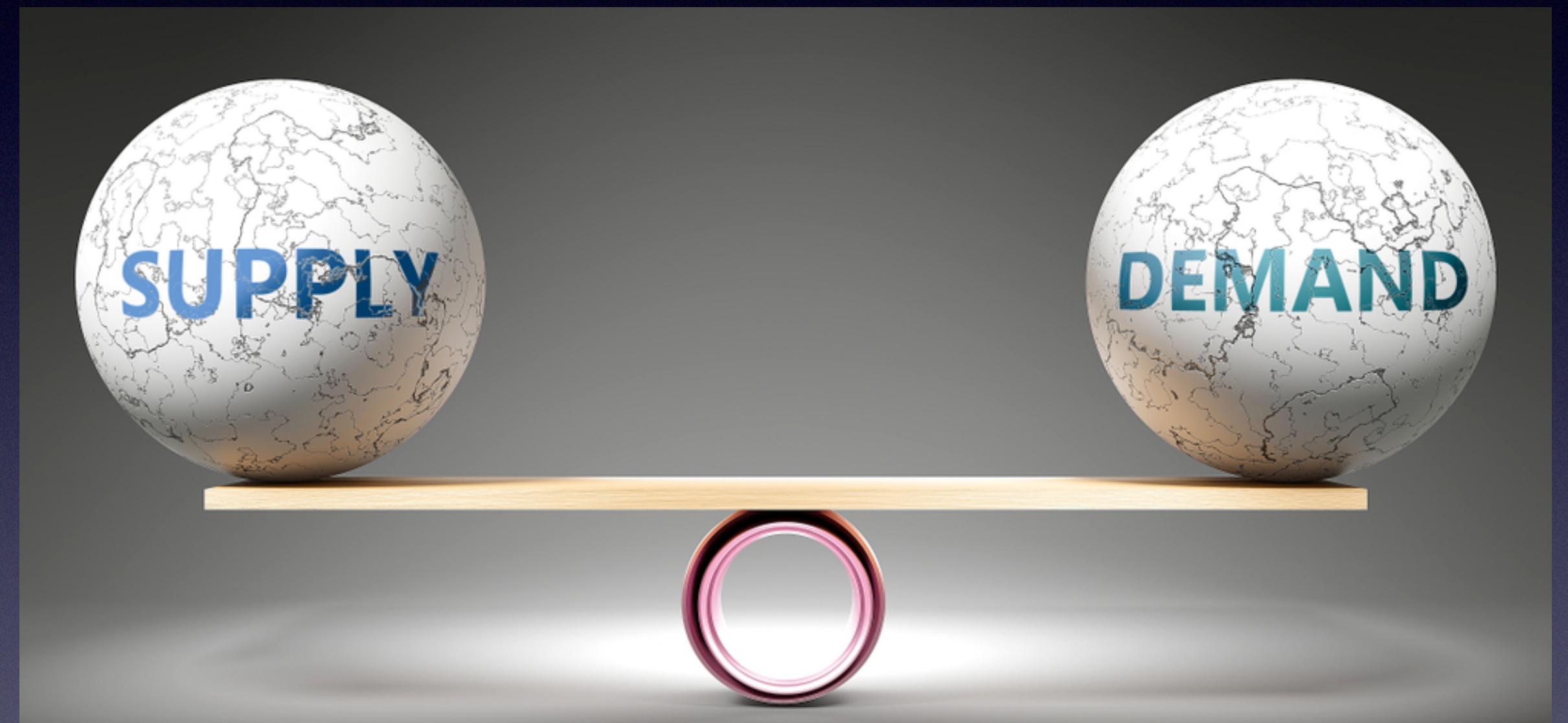
# Context

Fresh foods, a challenge to any grocer or retailer

- Short shelf life,
- Appearance,
- Season dependency,
- Specific transportation,
- Storage requirements

Led to Demand and Supply Imbalance.

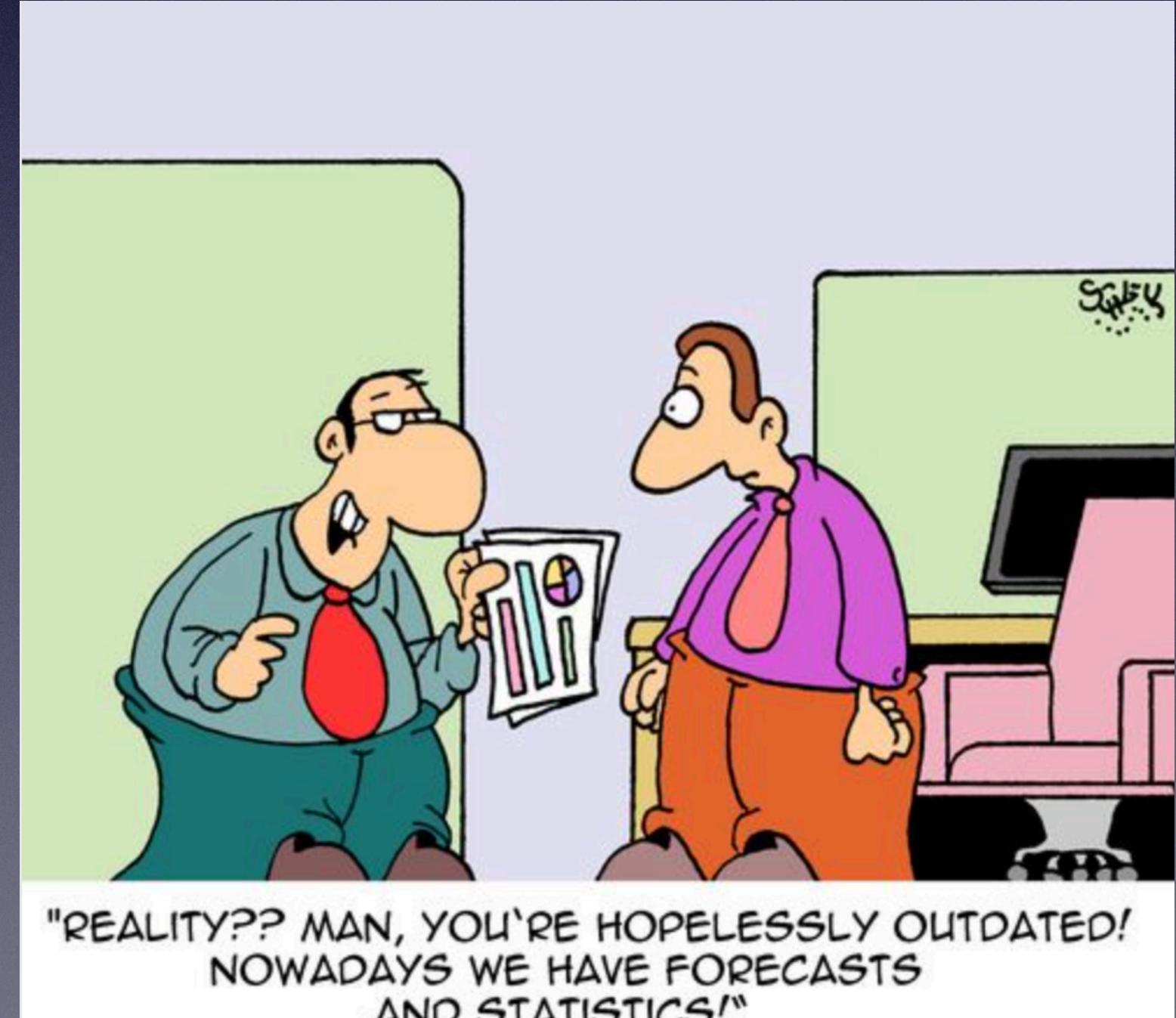
Food wastage, - 2.5 billion tons every year



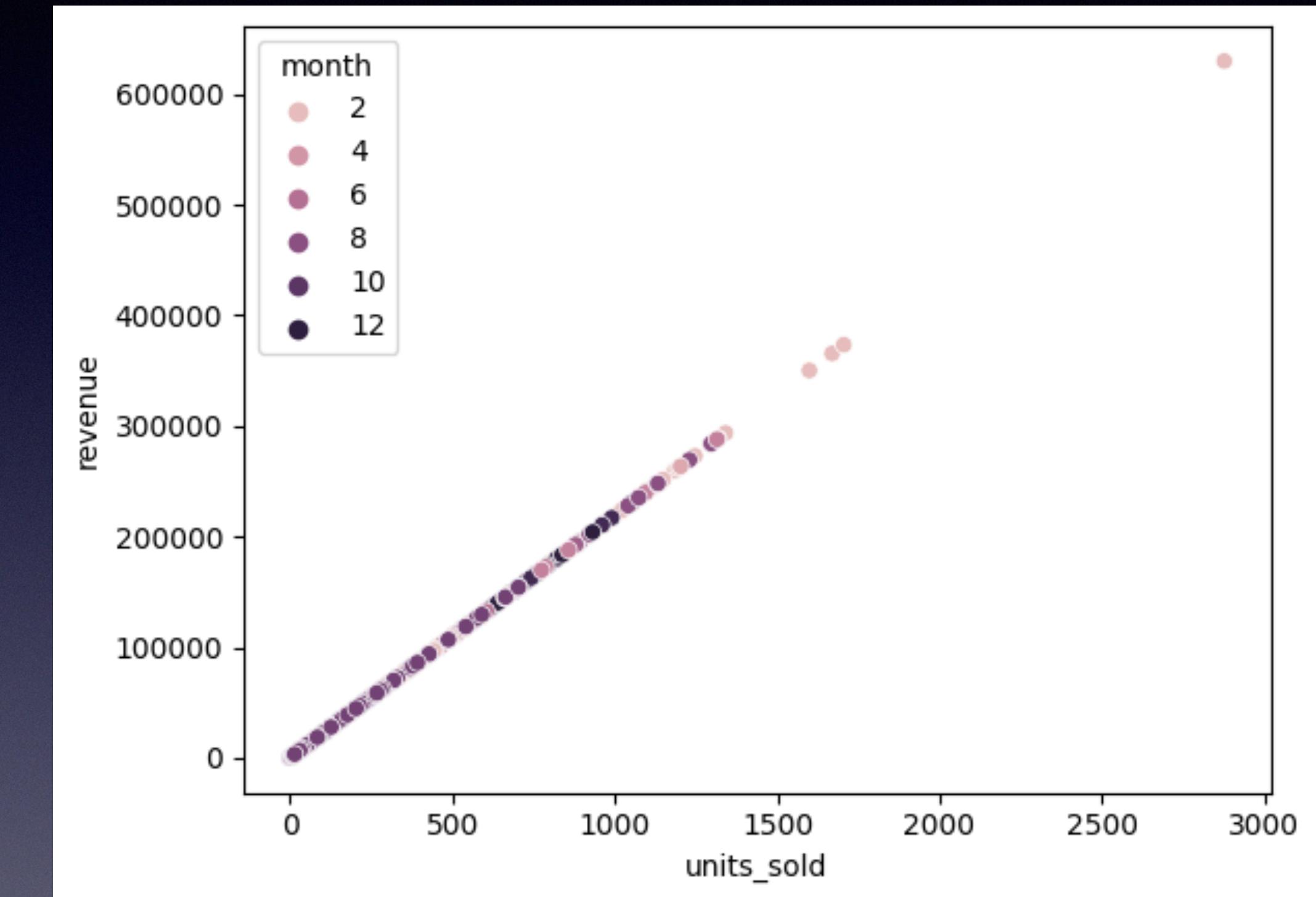
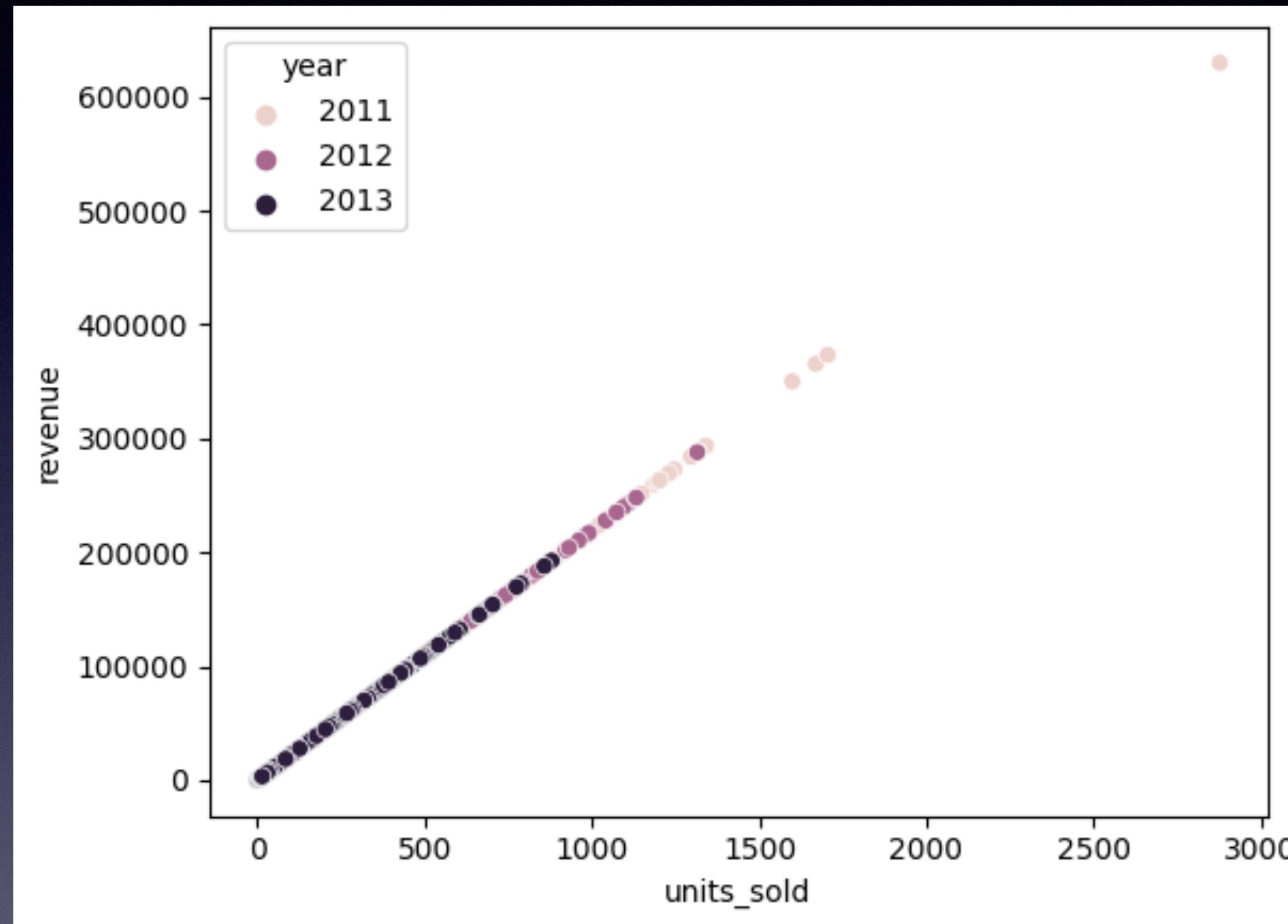
Need for a Demand Forecast system to resolve this issue at a global level.

# What is the Problem?

What opportunities exist for grocers or retailers to effectively develop and implement a new “Forecasting ML model” to better forecast the demand for perishable goods.



# Data Wrangling

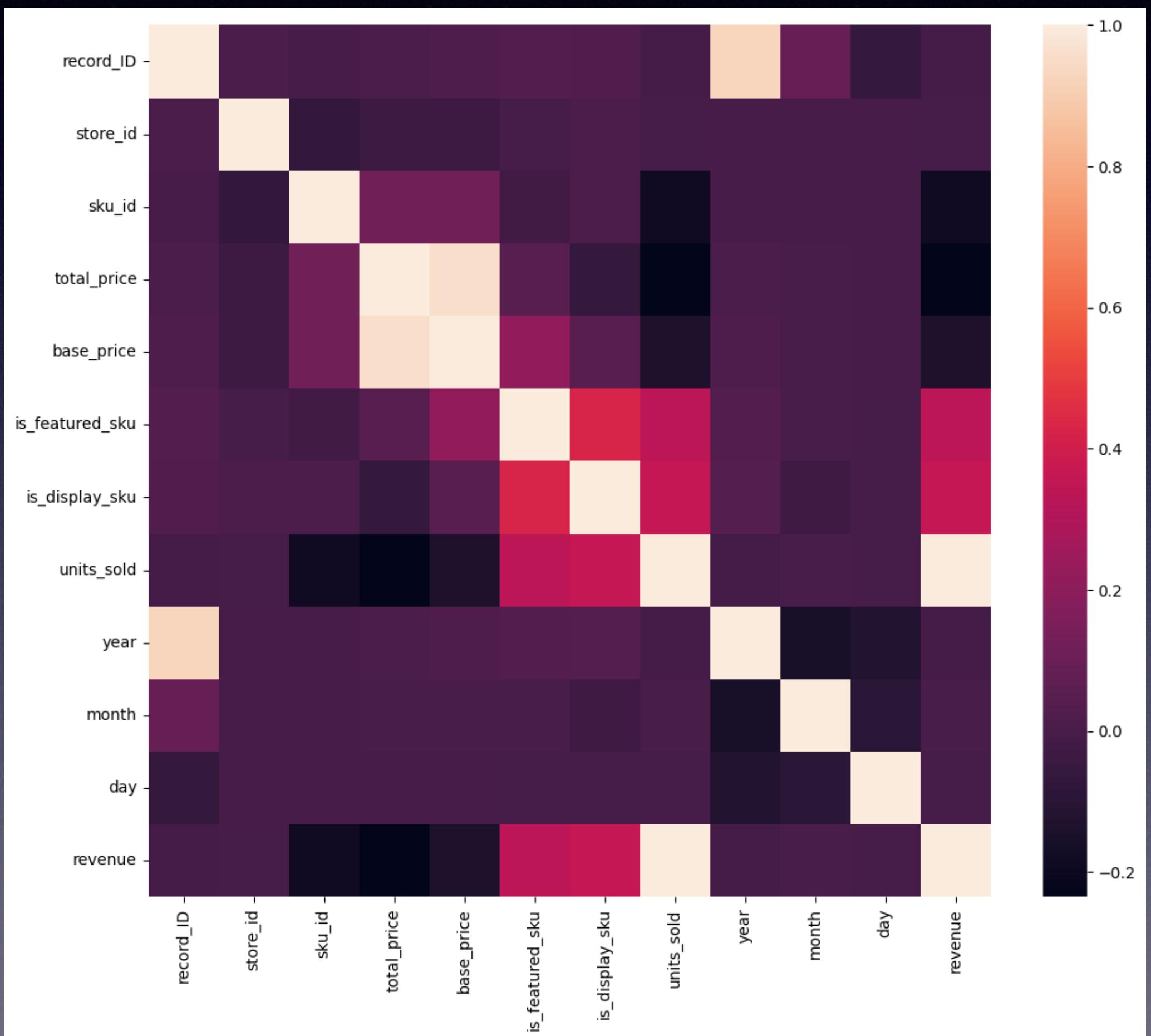


Decline in the sales per units sold between 2011 and 2013  
February (Month 2) saw the maximum purchases  
August (Month 8) saw the least amount of sales

# Exploratory Data Analysis

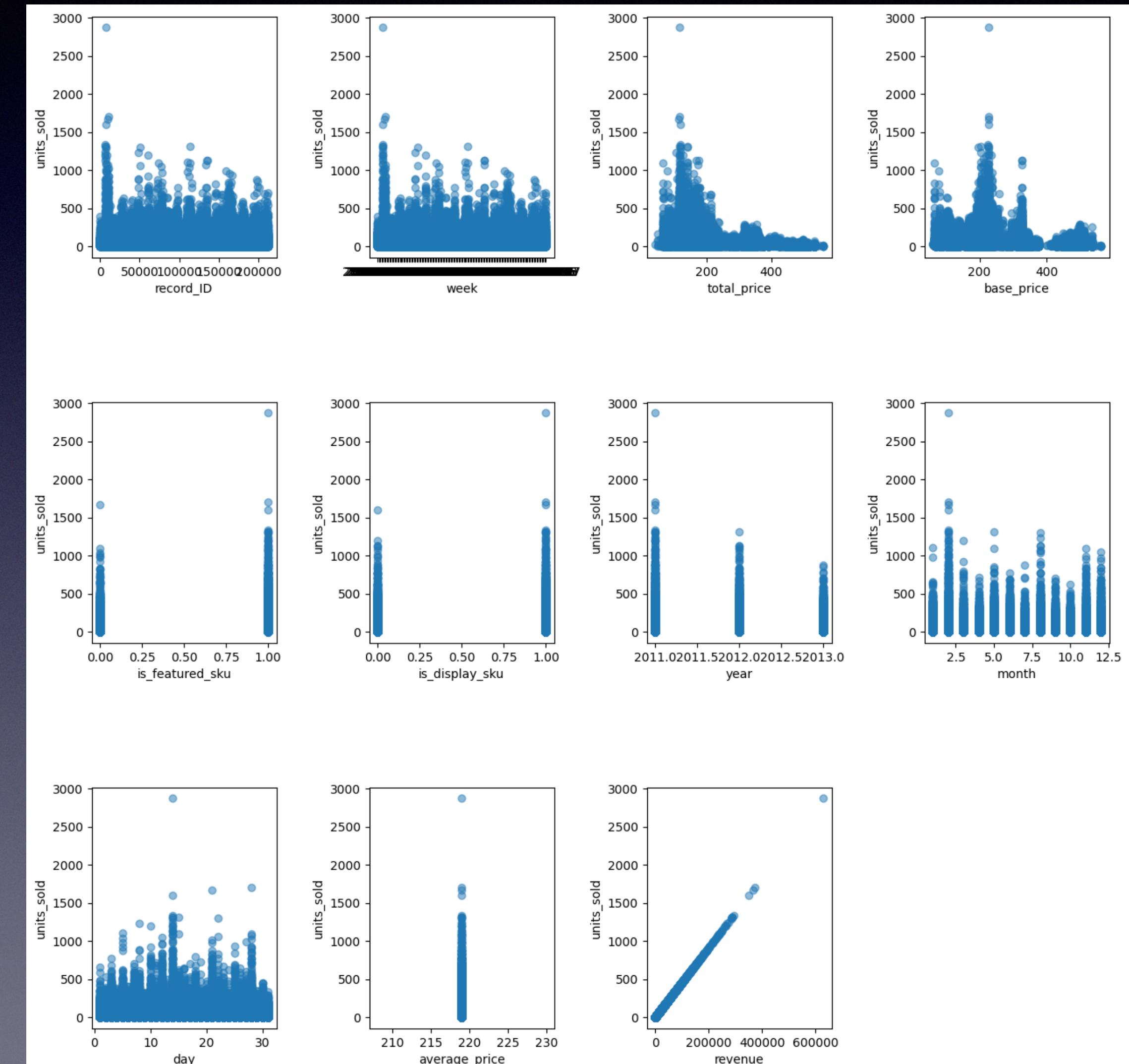
Gain a high level view of relationships amongst the features

- There's a strong positive correlation between units\_sold and revenue.
  - Year, Month and Day seems very useful to understand the seasonality and behavioral aspect.
  - There are some outliers present in almost all columns.



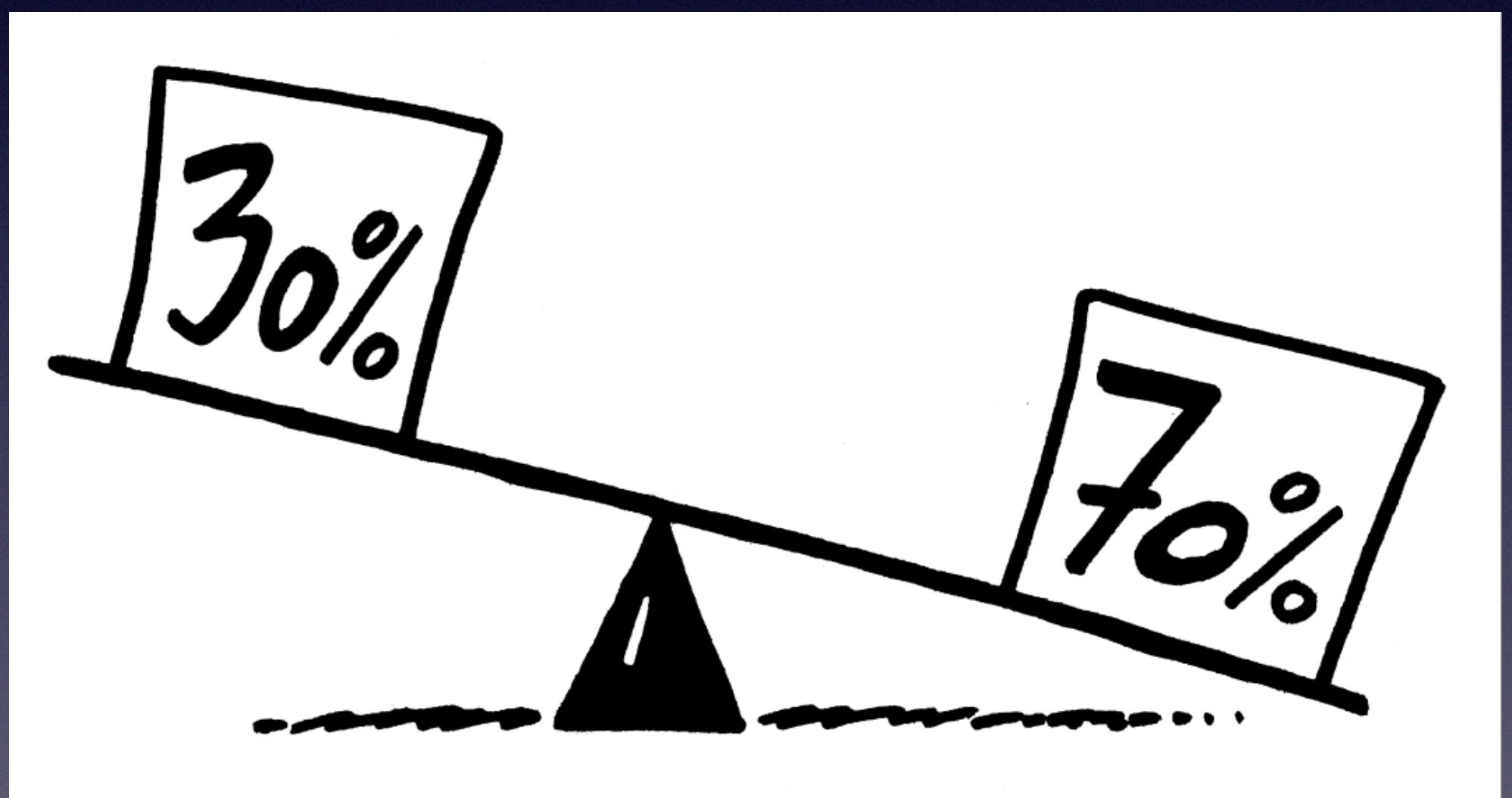
# Exploratory Data Analysis - Continue

Create Scatter plots for visualizing the relationship between a numeric feature against target variable, units sold.



# Further Analysis via Pre-processing and Training data

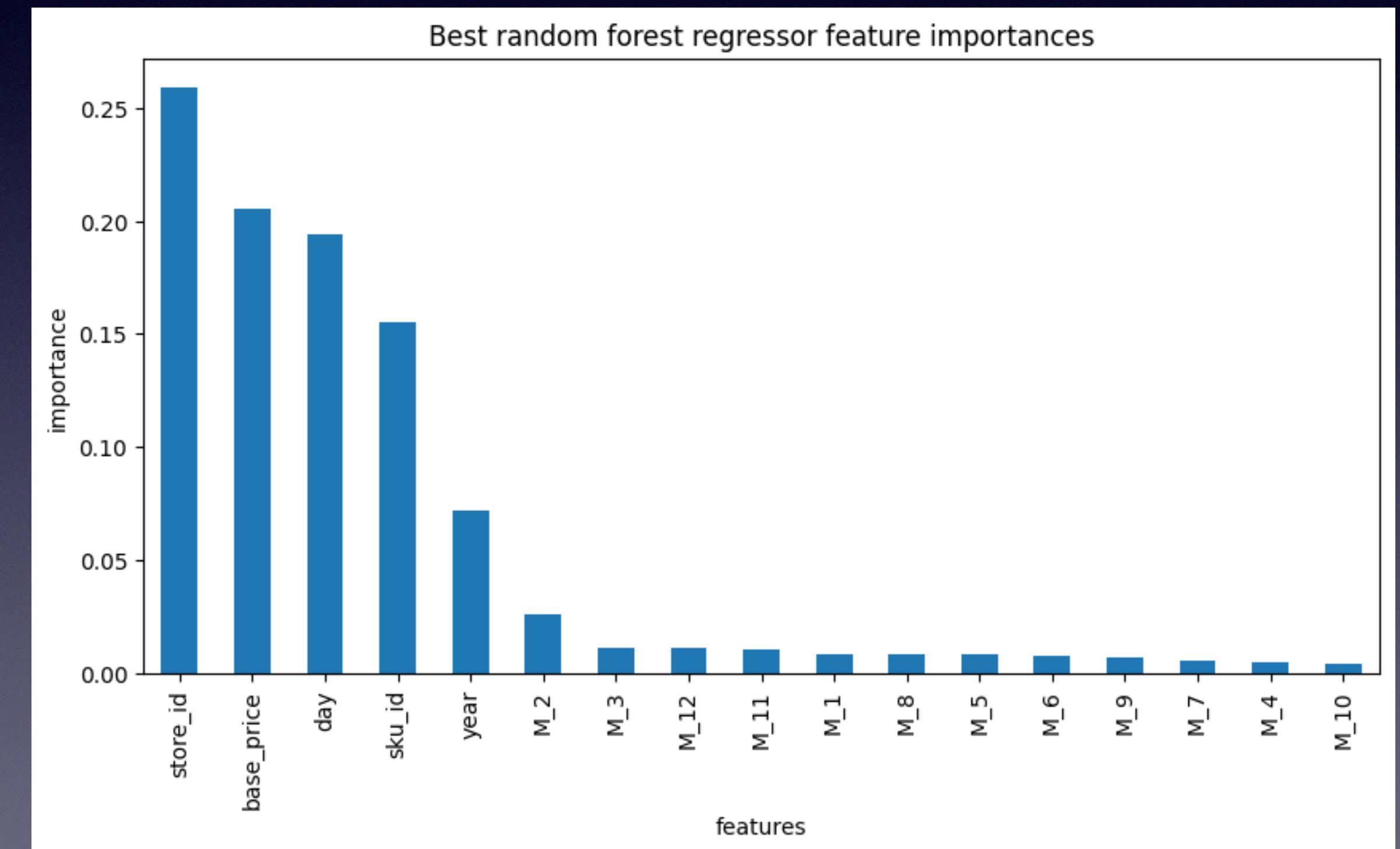
- Drop insignificant columns such as week, is\_featured\_sku, is\_display\_sku, average\_price.
- Impute missing values.
- Converted categorical column “month” into a format suitable for machine learning algorithms via “One-hot encoding”.
- Split the data into 70/30 train and test dataset.



# Modeling Results and Analysis

Identified dominant features using Random forest regression model

- store\_id
- base\_price
- day
- sku\_id
- M\_2
- year
- M\_3
- M\_12

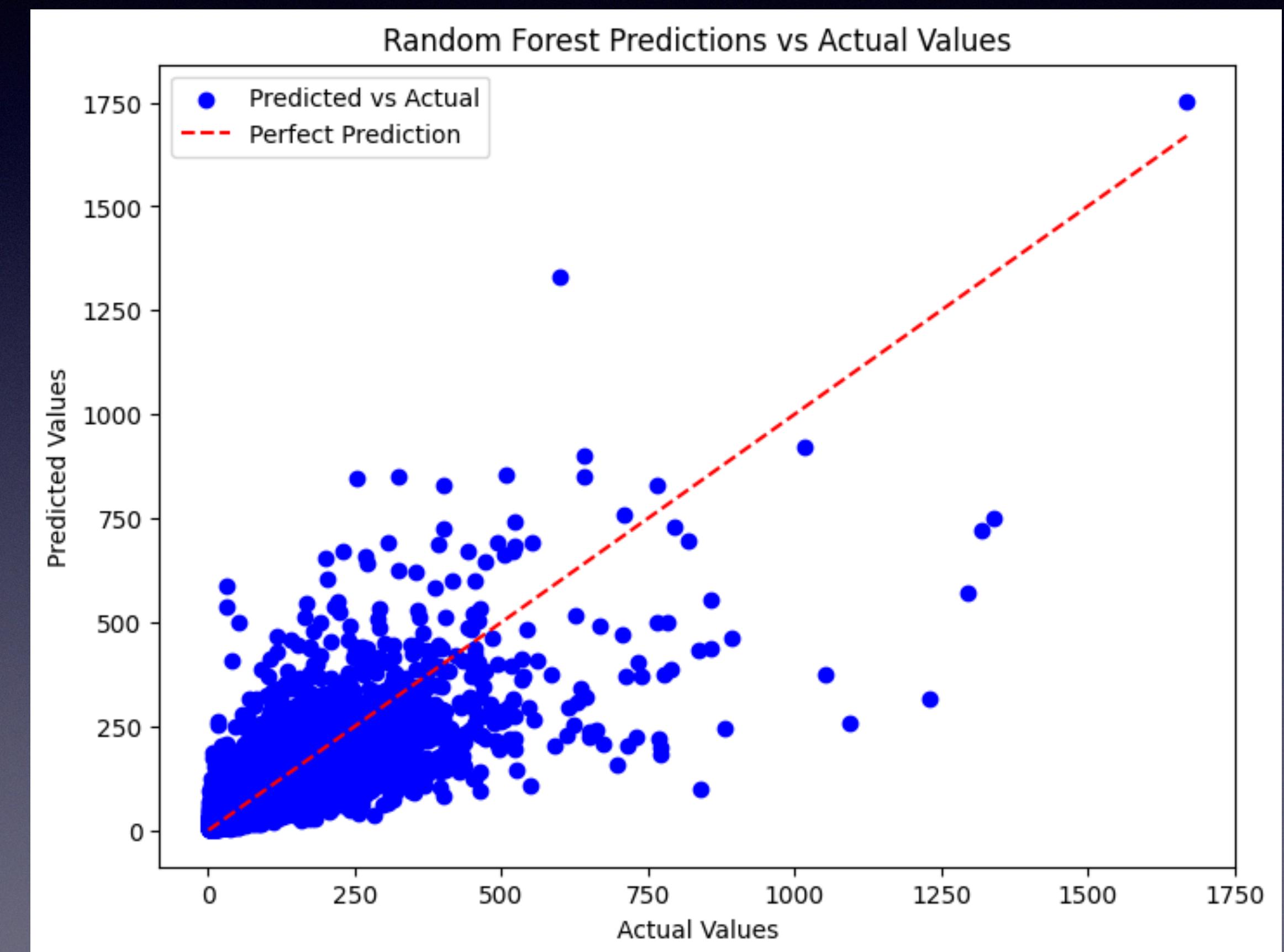


**Model Evaluation:** The random forest model has a lower cross-validation with least mean absolute error of 16.22. Exhibits less variability.

# Visualize the Predictions

Plotted the predicted values against the actual values ( $y_{\text{test}}$ ).

- Each point on the scatterplot represents a pair of **actual** and **predicted** values.
- The **red dashed line** represents perfect predictions ( $y = x$ ). If all the points lie on this line, the model has made perfect predictions.



# Recommendation and Key Findings

- Add more relevant features like external factors (economic indicators, holidays, etc.) to improve predictions.
- Include external data sources such as marketing campaigns, competitor actions to stay updated.
- Tune the model periodically with fresh data to keep it up to date.

