Capstone project

Problem:
Online sales are growing at an unprecedented rate, making it challenging for the distribution centers to keep ample stock of top selling products, as well as ensure that the centers are staffed appropriately to handle demand.

Objective:
Drive speed and efficiency in our online business through better forecasting for our distribution centers.

Strategy:
To determine whether a forecasting model built by our team would bring value to the supply chain team, we decided to test out a model on the most challenging category. So I built a proof of concept forecasting model that forecasts weekly sales 4 weeks in advance for the most difficult use-case: low-volume, highly seasonal products

Scope:
- Snow products only
- Orders from Ontario
- Predicting demand for Black Friday

Feature Engineering
In order to forecast demand, we amalgamated numerous resources and data sources to capture as much information as possible. These resources included:
- ➢ Historic sales
  - o Units Sold
  - o Order Numbers
- ➢ Price Information
  - o Current Price
  - o Discount Price
- ➢ Website Behavior
  - o Product Clicks
  - o Product Views
  - o Product Orders
  - o Product Visits
- ➢ Weather Data
  - o Temperature
  - o Precipitation
  - o Snow Depth
  - o Min Temp
  - o Max Temp

Model Features

Sales

Units Sold
Orders

Website Behavior

Product Clicks
Product Views
Product Orders
Product Visits

Price

Price
Promo

Weather

Temperature
Precipitation
Snow Depth
Min Temp
Max Temp
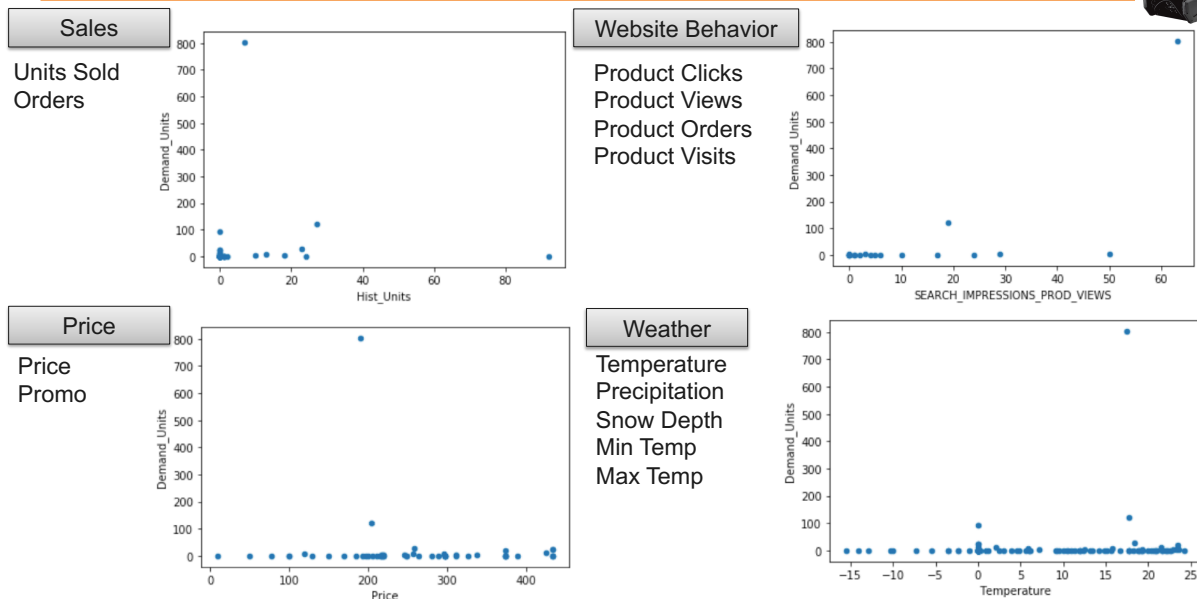
Each feature individually did not show a strong correlation to demand sales.

Model Strategy
Since snow blowers are only bought in the winter, and in low volume, we had to use some creative techniques to get the best model.

Step 1: Binary Classifier of Sales
We created a classifier random forest model that predicted whether a product would sell or not. This model helped us predict weeks where a store would likely not sell any snow blowers.

Step 2: Random Forest Regression
We used all the features mentioned above to create a random forest model that predicts how many units of each product will sell.
We forecasted zero sales if the binary classifier gave an over 40% chance that nothing would be sold that week and used the forecast of the regression model in all other cases.

Results
The application of this modelling strategy improved the current forecast by 5%. This is a large improvement as every percentage increase in forecast accuracy translates to millions in sales.