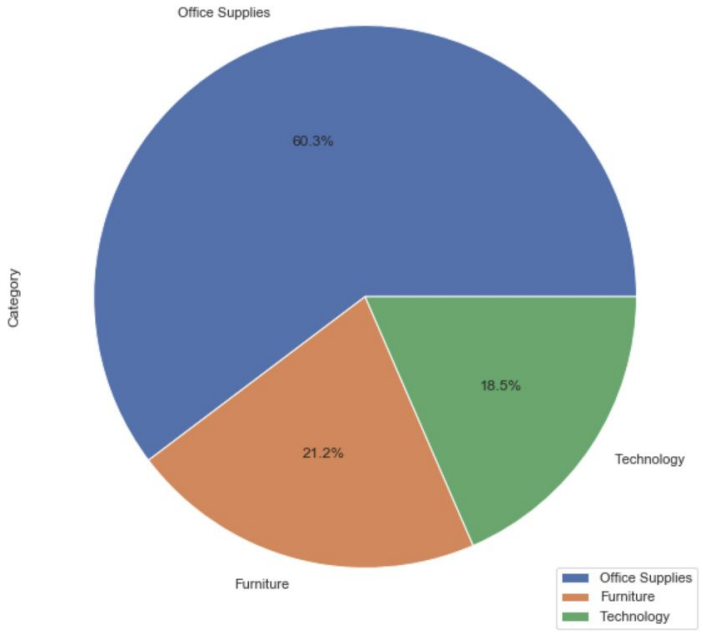
Superstore Analysis

Arnav Popat
08/01/2023
(Associate Operations Analyst Interview)

Background

1. “Superstore” is an e-commerce platform selling a variety of office products.
2. Dataset spans from the years 2014 - 2018.
3. 9994 data entries
4. Each row contains a single purchase along with relevant details.



- 60.3% of products belong to Office Supplies.
- 21.2% of products belong to Furniture.
- Only 18.5% of products belong to Technology.

Objectives

- Classifications
- Demographics
- Customer loss

- Seasonality
- Regional trends
- Discount effectiveness

- Predictive modelling
- Time series patterns

Customer Classification

- “Big spenders” and “Important developing customers” are low.
- Total percentage of “Lost” and “Almost lost” customers are is high **(28%). We need to investigate the reasons for losing customers.**
- There is a large amount of important recall customers.

RFM Analysis

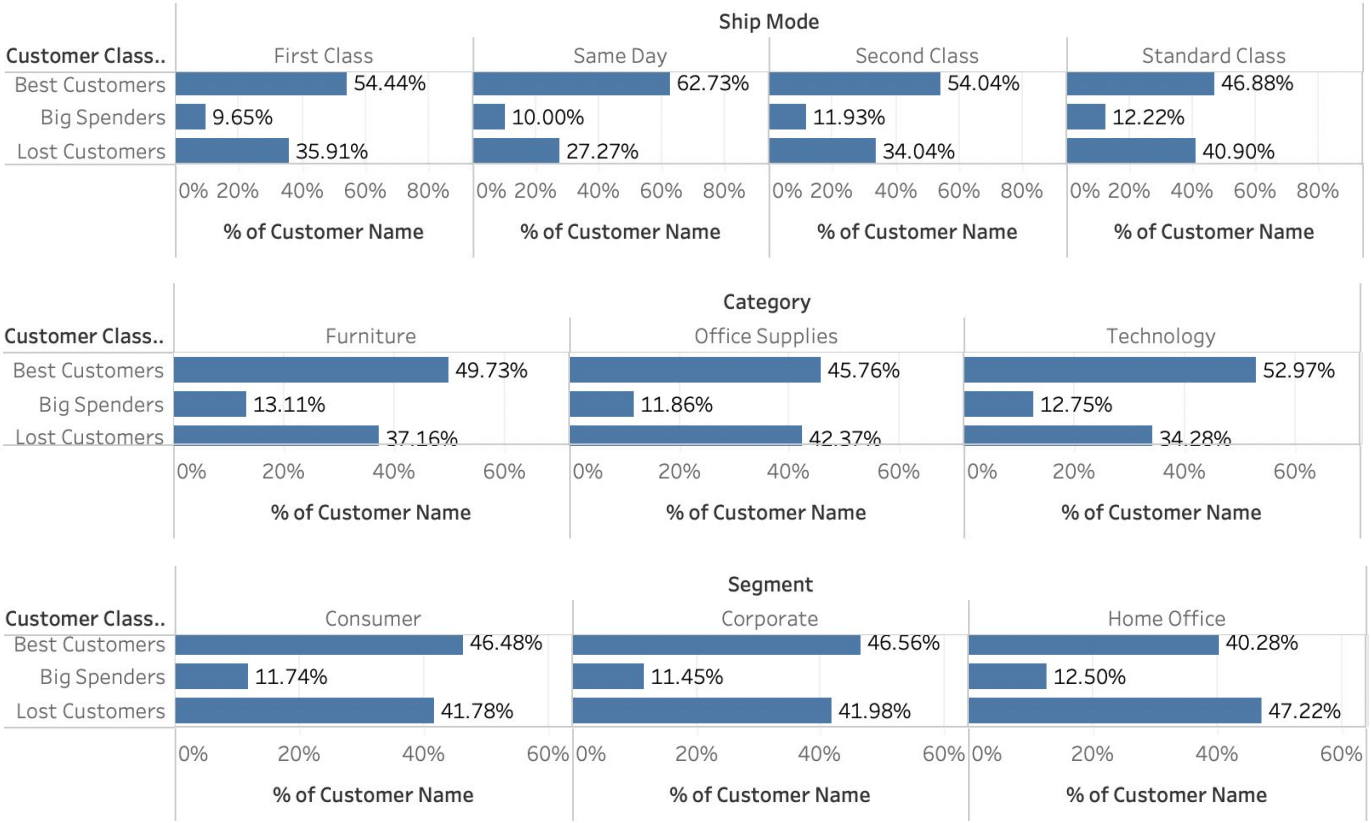
Customer Classification	Distinct count of..	% of Total Distin..
Almost Lost Customers	51.0	6.43%
Best Customers	189.0	23.83%
Big Spenders	49.0	6.18%
Important Developing Cus..	41.0	5.17%
Important recall Custome..	118.0	14.88%
Lost Customers	178.0	22.45%

RFM: Recency, Frequency, Monetary

Customer Classification

- Recency represents the customer's loyalty.
- We can see that "Big Spenders" have relatively low customer loyalty and it may be an area we need to improve.

Why Are Customers Leaving?

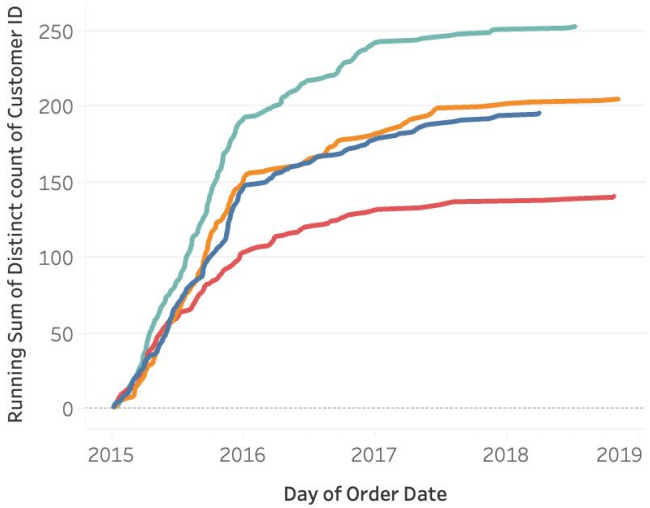


Insights and Recommendation

- Most lost customers choose Standard Shipment. Investigating this shipping mode and the waiting time may help us find the reason for losing customers.
- Using a membership account that provides an upgrade of shipping mode may be a strategy to turn other customers into “Best Customers”
- A large number of lost customers bought office supplies. This may be due to discounts and promotions on these products.
- Discounts increase sales but may not be a good strategy to build a healthy customer structure.

Customer Demographics

New Customers

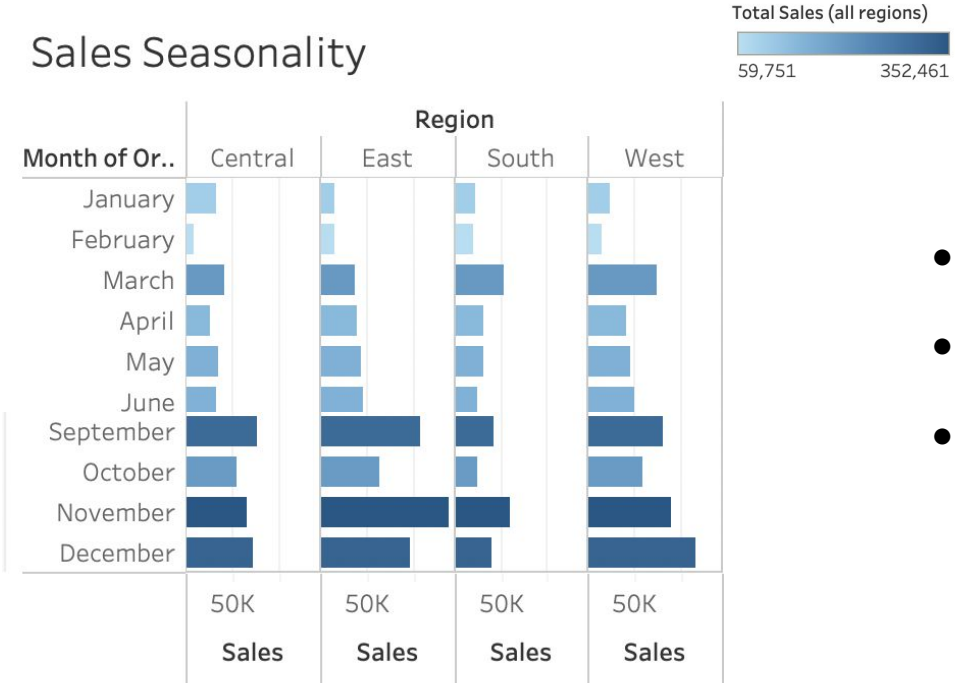


Increase of new customers slows after 2016 for all regions.

Loss Customers Analysis

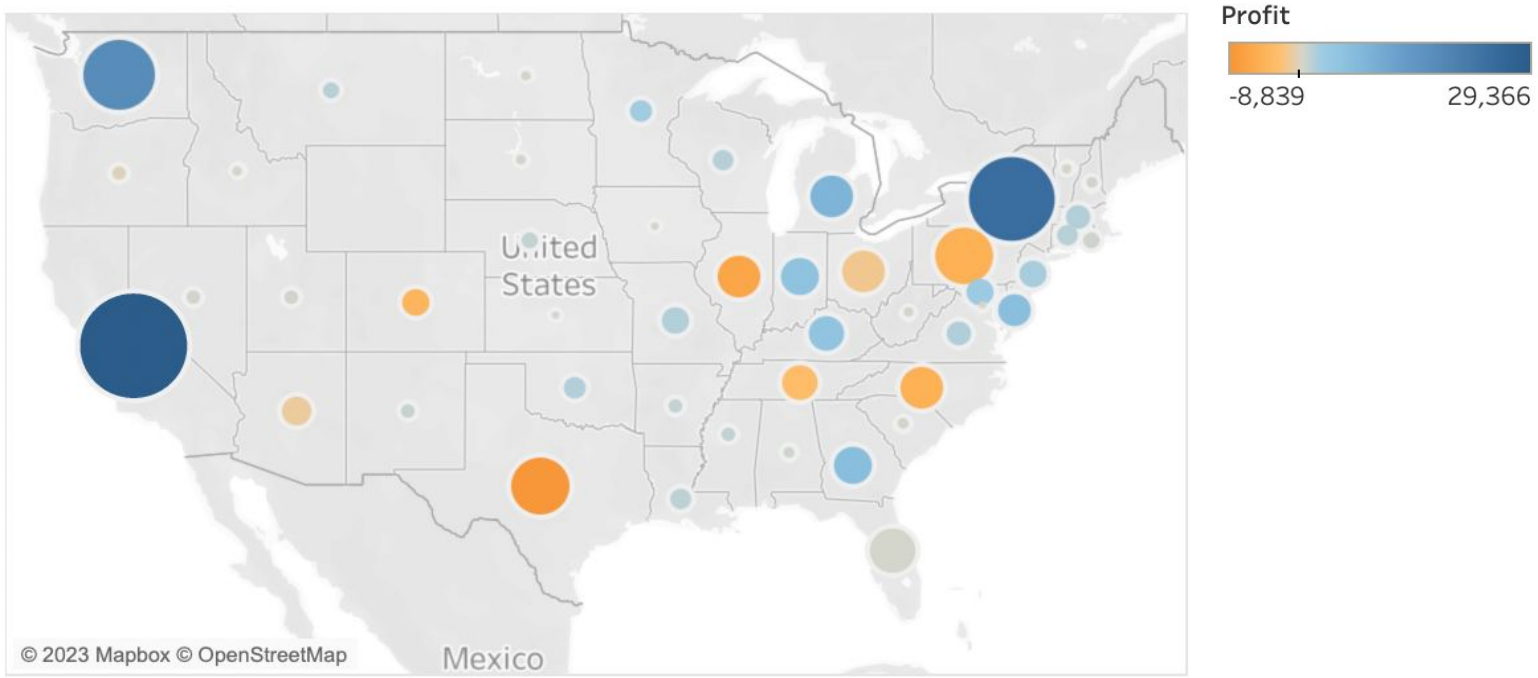
- A small part of lost customers are unprofitable. They buy discounted products at that time and don't return without discounts.
- For those losing customers who generate profits, it is important to attract them back

Sales Seasonality



- The period from September to December is the peak season of sales.
- Quarter 1 is the offseason of sales for all regions.
- It may be related to the holidays, promotions, and market events.

Regional Profits



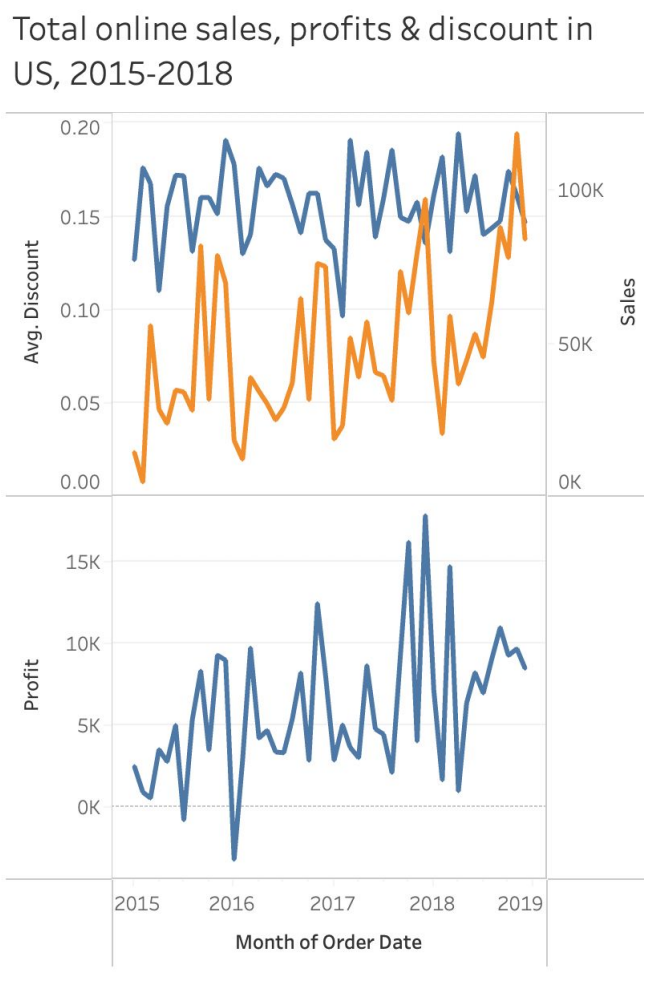
We find that there are large negative profits in states such as **Texas, Illinois and Pennsylvania**, and large profitable markets in **Washington, California and New York**.

Insights and Recommendations

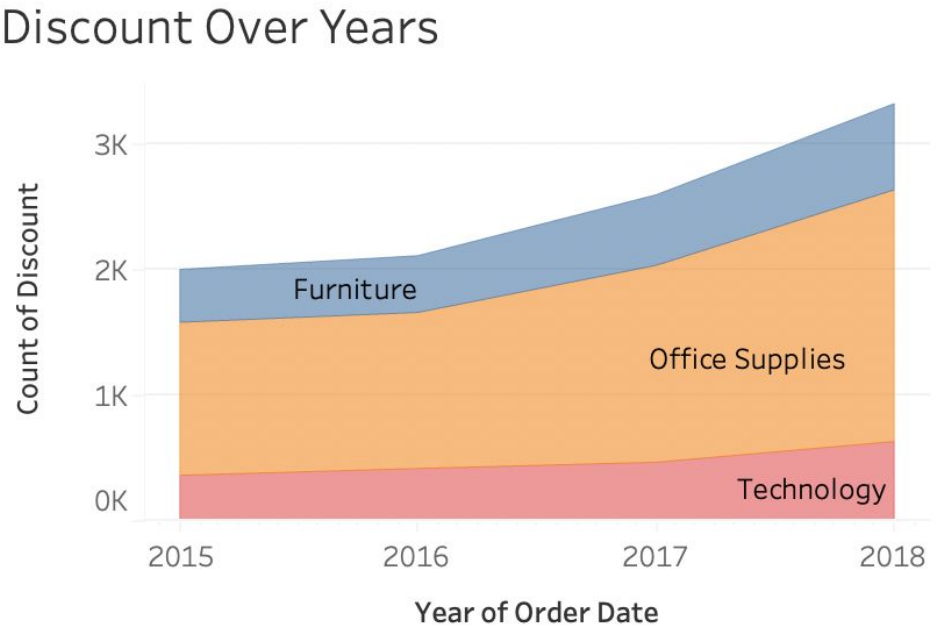
- We find that location and months of the year both affect profits and sales. Many economic and social factors such as capital market performance, national holidays, and regional culture, need to be considered and further researches could be conducted.
- It is advised to investigate the regional difference first, and try to improve the performance of sales in unprofitable states.

Effectiveness of Discounts

- The trends of Avg. Discount is consistent with the trend of Sales and Profit for Month of Order Date. It means that the promotion events are efficient and the discount could be transformed into sales and profits.
- However, in the year 2017, the promotion did not increase sales and profit.



Discount Breakdown



The times of promotion is increasing over the years, especially for **Office Supplier**. It may be one of the reasons why the sales of Office Supplier boost in the year 2018.

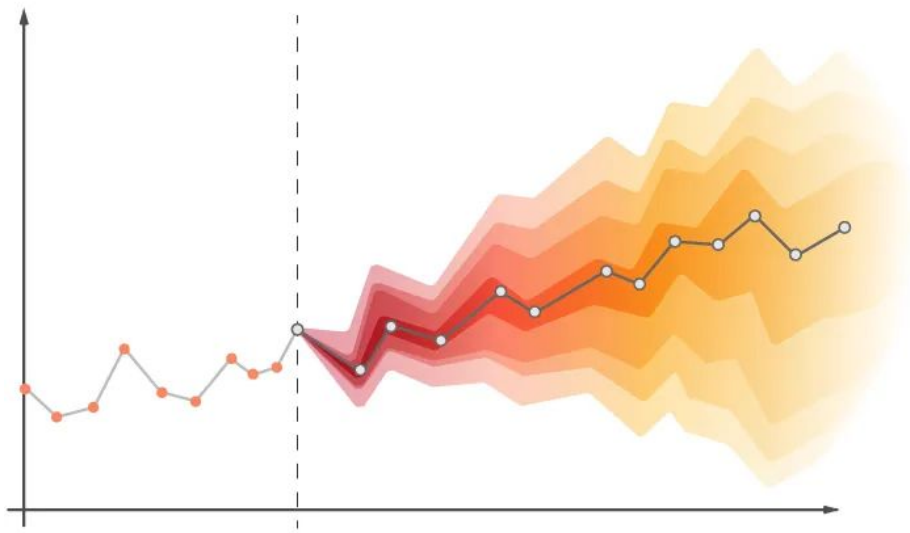
Insights and Recommendations

- Discounts and promotional levers are critical for retailers who want to reach new shoppers to drive incremental sales.
- A location based strategy could be designed based on how each product responds to the promotion.
- Analysis shows that Chairs, Phones and Copiers respond to promotions effectively.

Forecasting

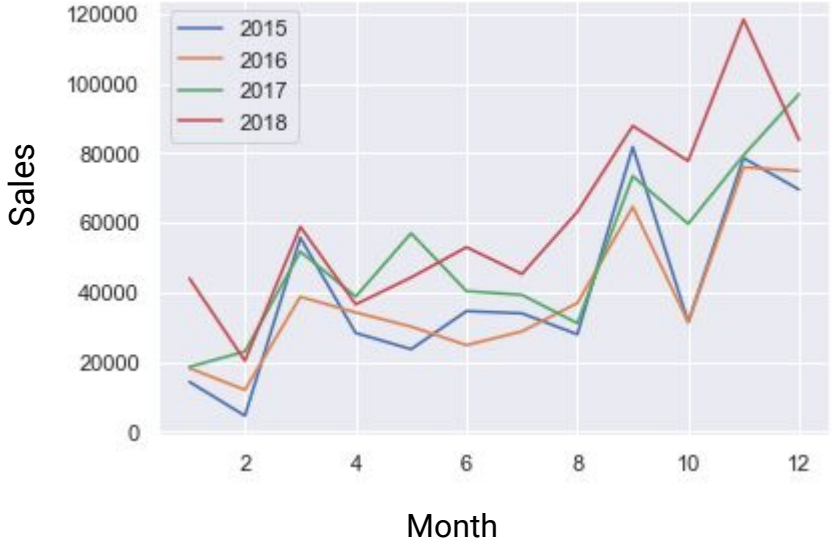
Enables businesses to plan and make informed decisions about future operations, marketing, and resource allocation.

Accurate sales forecasting can help businesses anticipate future demand, identify potential problems or opportunities, and adjust their strategies accordingly.



ARIMA Model

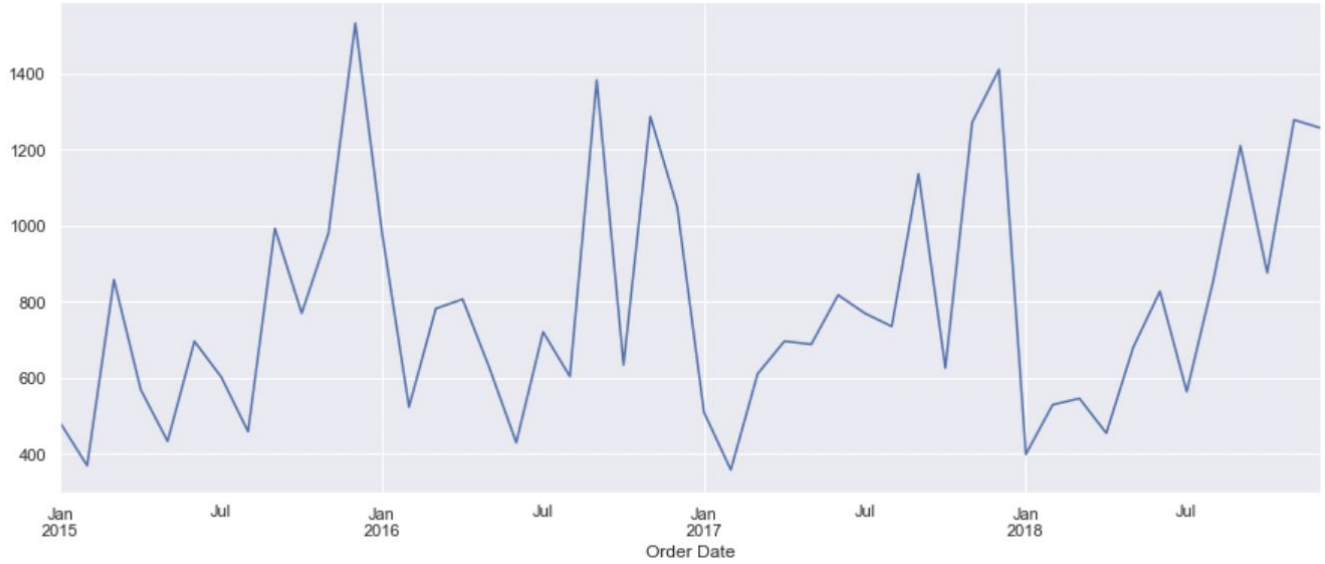
- Prediction model used for time series analysis and forecasting.
- Forecasts are usually more accurate and reliable.
- Better understanding of time series patterns.
- Captures only linear relationships.
- Our dataset could use the ARIMA model since it looks like there is a linear relationship between sales and years.
- However, we need to exclude the trend and seasonal effects to get a stationary dataset.



Limitations

It only works on a single variable. Therefore it cannot exploit the leading indicators or explanatory variables such as the effects of external factors, such as marketing, competition, or events, on the time series.

Stationarity



- We will use the average daily sales value to represent that month, and we are using the start of each month as the timestamp
- The time-series has seasonality pattern, such as sales are always low at the beginning of the year and high at the end of the year. There is always an upward trend within any single year with a couple of low months in the mid of the year.
- The dataset is unstationary.

Selecting Parameters

ARIMA models contain 3 distinct parameters:

1. P: autoregressive
2. D: integrated
3. Q: moving average

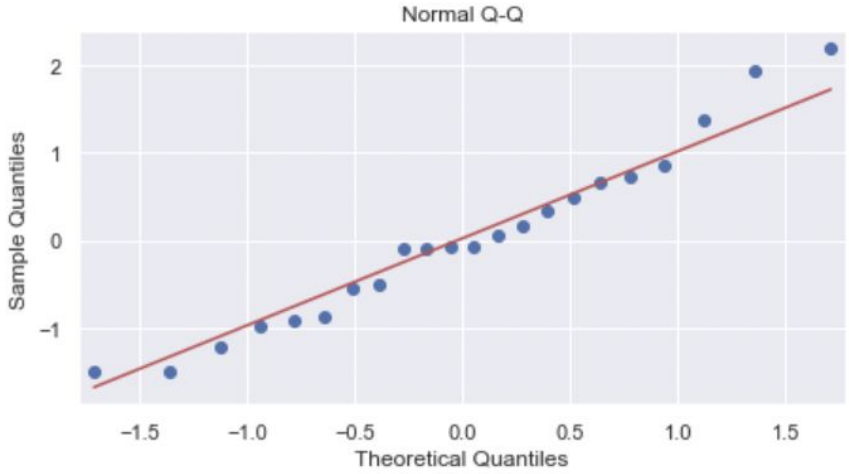
My approach was to brute force every possible combination of the 3 parameters and use the combination with the lowest AIC score (lower the better).

AIC: Akaike Information Criterion

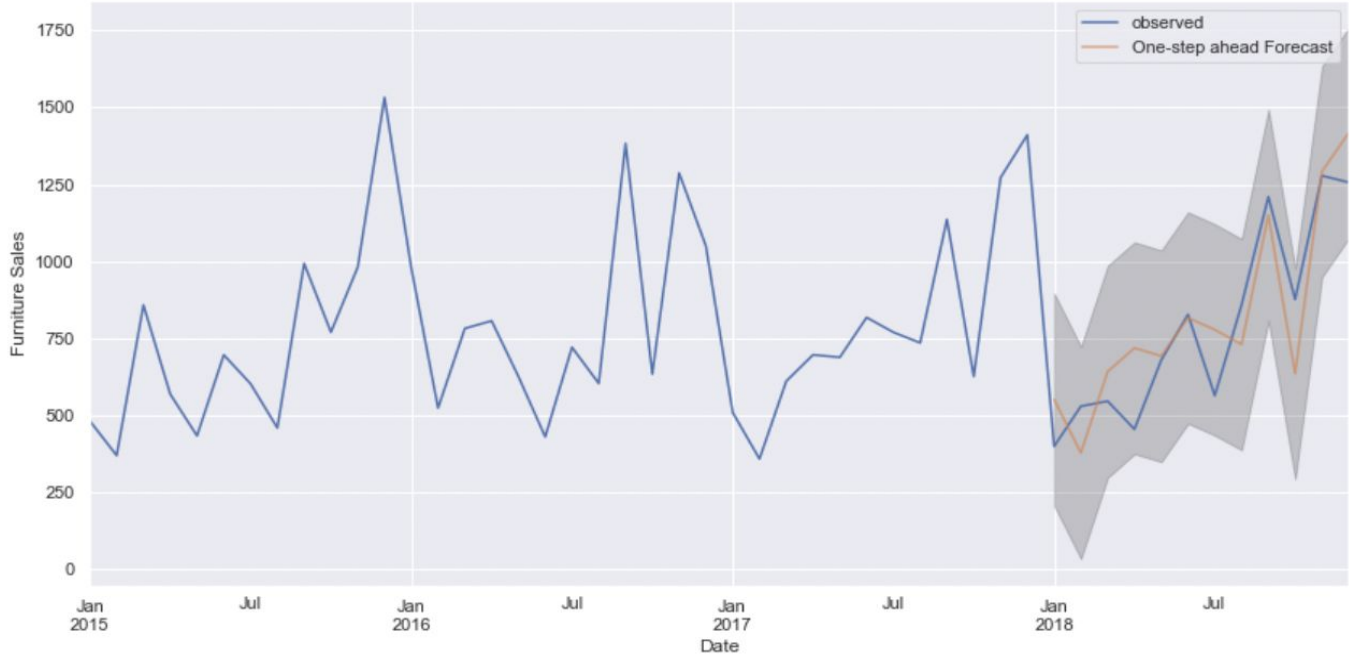
Fitting and Validating Our Model

After fitting our model and analyzing diagnostics, I had good indications that the residuals are normally distributed

Those observations lead me to conclude that our model produces a satisfactory fit that could help us understand our time series data and forecast future values.



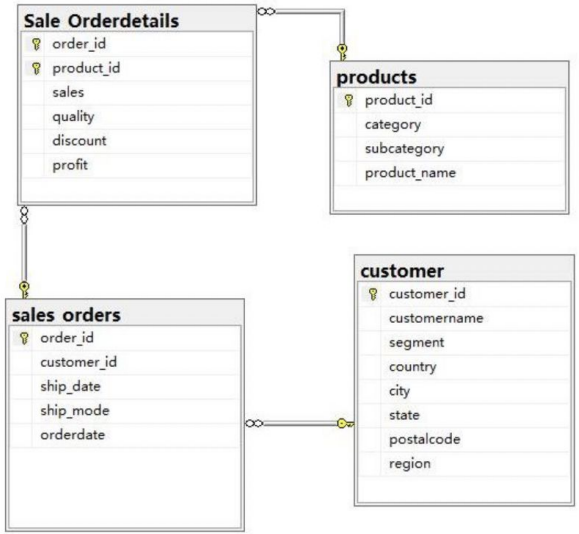
Validating Forecasts



After generating our model, we can plot the predicted and observed values to get an idea of how accurate the forecast it.

Database Design

Normalization of relational database helps us avoid redundancy and anomaly. It helps us insert, update, and delete records easily, and it performs better than excel files when we want to share the dataset among a group or understand the relationship between variables quickly.



Summary

- Customers buying office supplies for a home office using standard shipping have the lowest retention rates.
- A membership strategy which upgrades shipping can help to increase retention rates.
- While discounts and promotions are effective for short term sales growth, they are unable to retain customers.
- Southern region is the most unprofitable and has the least number of new customers.
- Location based advertising campaigns could lead to increased sales.
- Forecasting shows growth for the next year.

Thank You!