

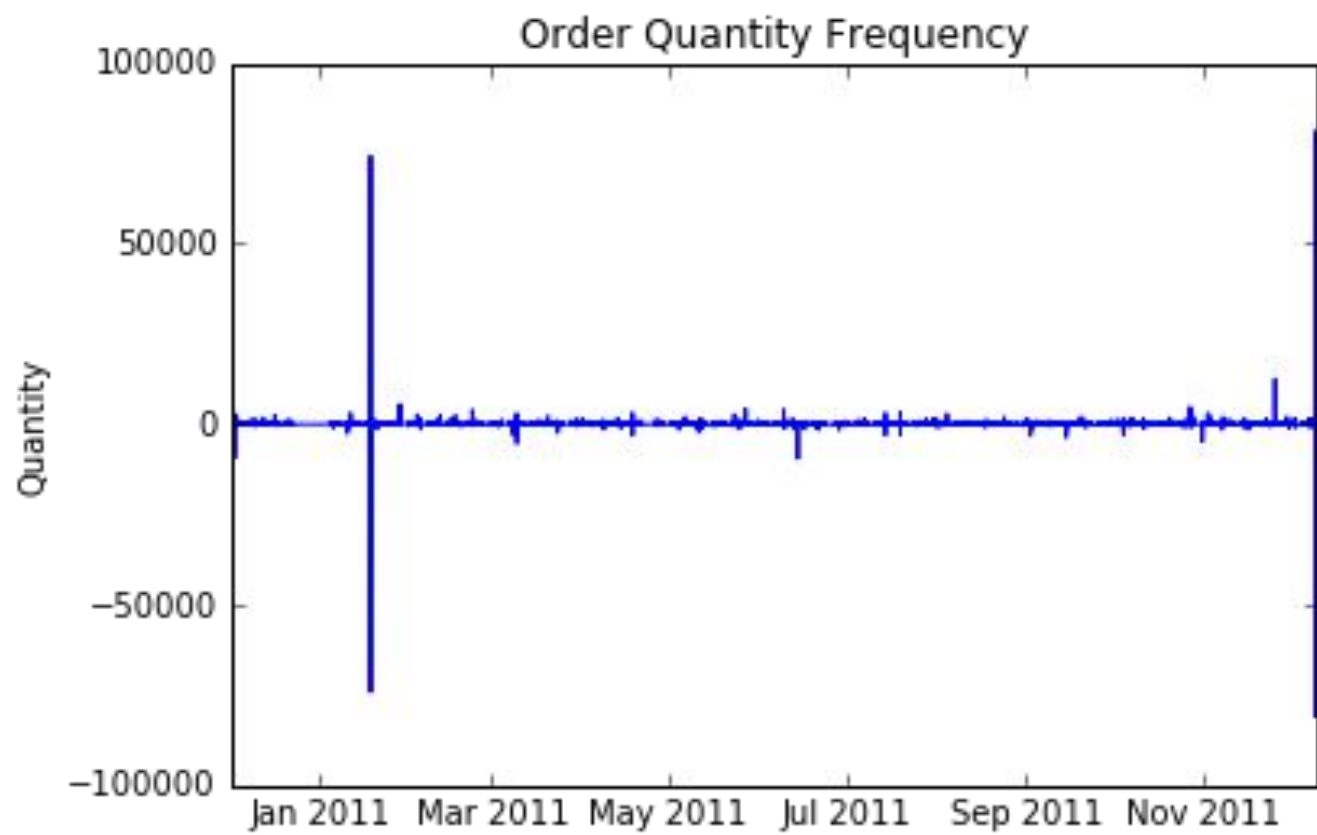
Capstone Project Presentation

To Build a Recommendation Systems for an Online
Retail Dataset

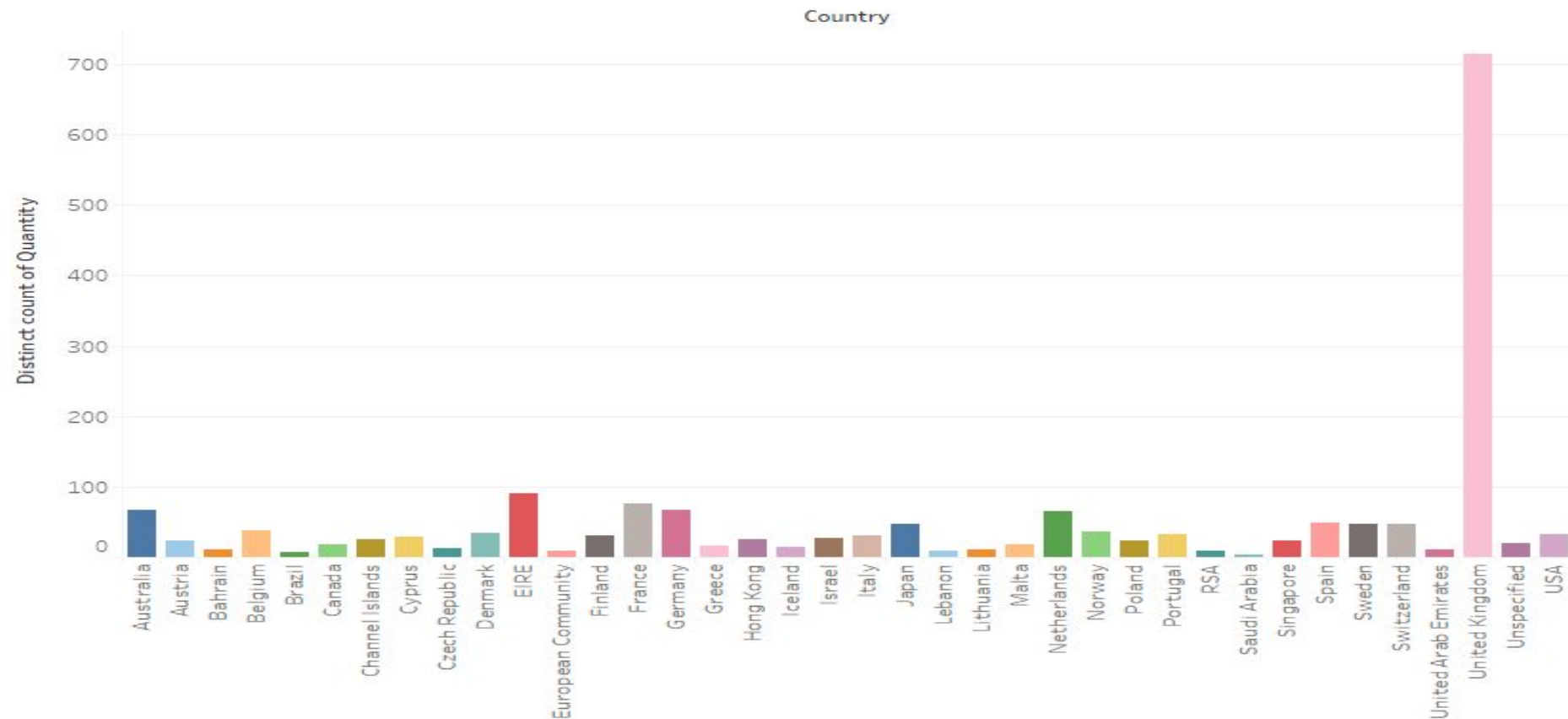
Sumit Patel

Objective of Social Network Ad

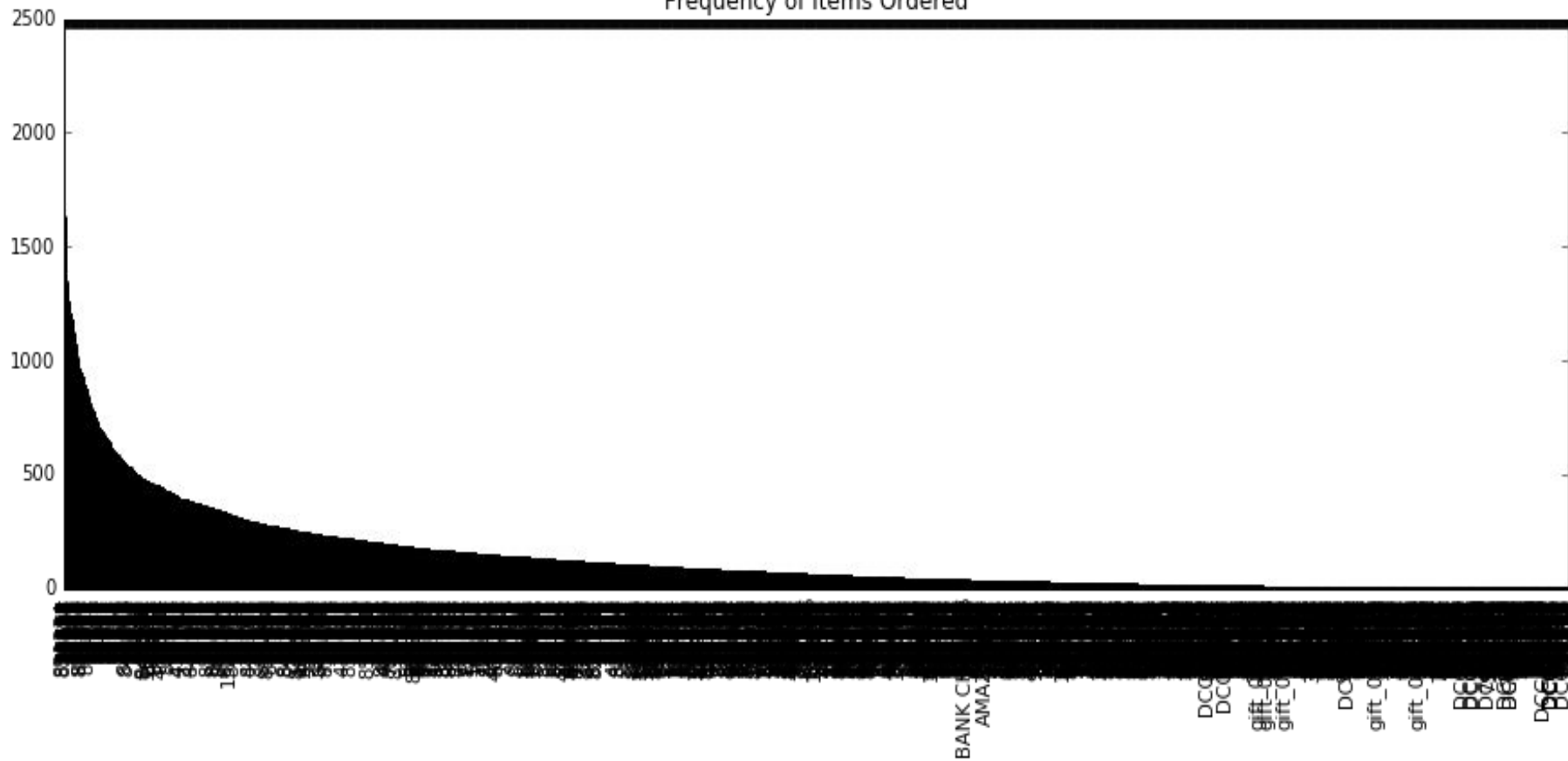
- The objective of a recommender system is to use specific features or parameters to help provide recommendations to customers based on other customer data, similar items, popular items, etc.
- The initial step is to identify the features of interest that will help to build the recommendation system. The features that will be necessary are `user_id`, `item_id` and `target`. We will be using graphlab to build our recommender system.



Quantity Ordered by Country



Frequency of Items Ordered



Purpose of the Model

- To provide a recommendation with very little error-rate by items to customers to possibly make an impact on purchasing behavior and increase sales through suggestive recommendations.
 - User_id = CustomerID
 - Item_id = StockCode (we merged this with Description so that you can tell what the item is)
 - Target = Quantity

Data Description and Processing

- The dataset is comprised of 8 attributes consisting of InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID and Country. The dataset is composed of 541,909 rows
- After selecting the features of interest we then convert the dataframe into an SFrame to use in graphlab. We will be using the features of CustomerID, Items (which we derived by concatenating StockCode with Description) and Quantity. Grouping CustomerID and Items we reduce our dataset to 406,829 rows of unique CustomerID instances.
- We then aggregate values for mean and standard deviation for quantity and filter out quantities with standard deviation=0 which reduces our dataset to 269,270 rows

Preprocessing and Model Evaluation

The following was performed:

- Mean and standard deviation was calculated for each instance of Quantity
- Z-score was calculated for scaling the data with since the feature of Quantity does not have a normal distribution
- Using z-score of quantity as target variable we retrieved our normalized RMSE values

RMSE of Models on Original Dataset

Recommender System	RMSE Overall
Recommender	127.61870150218108
Factorization	174.9675095375155
Ranking Factorization	145.38689486980064
Popularity	196.00268472641525
Item Similarity	212.4734752720519

RMSE of Models on Normalized Data using z-score of Quantity

Recommender System	RMSE Overall
Recommender	7.832946415412624
Factorization	5.038368894044742
Ranking Factorization	6.427522664597214
Popularity	5.984642479064843
Item Similarity	8.672835866566814

Conclusion

- Recommender systems require normally distributed values to perform ideally as we have seen when comparing unscaled measure of quantity against the z-score of the quantity which vastly improved our RMSE values
- Without normalizing our data we would have been likely to choose the wrong recommender for our data, this would have resulted in many incorrectly matched recommendations which would not have improved or benefitted the business operations.