

ONLINE RETAIL CUSTOMER SEGMENTATION & SALES PREDICTION

BY AMIN NAZERZADEH



DATA

- The dataset has been taken from UCI Machine Learning Repository
- The client in this case is a UK-based and registered non-store online retail
- The data contains about 540,000 observations domestic and internationally
- The company mainly sells unique all-occasion gifts

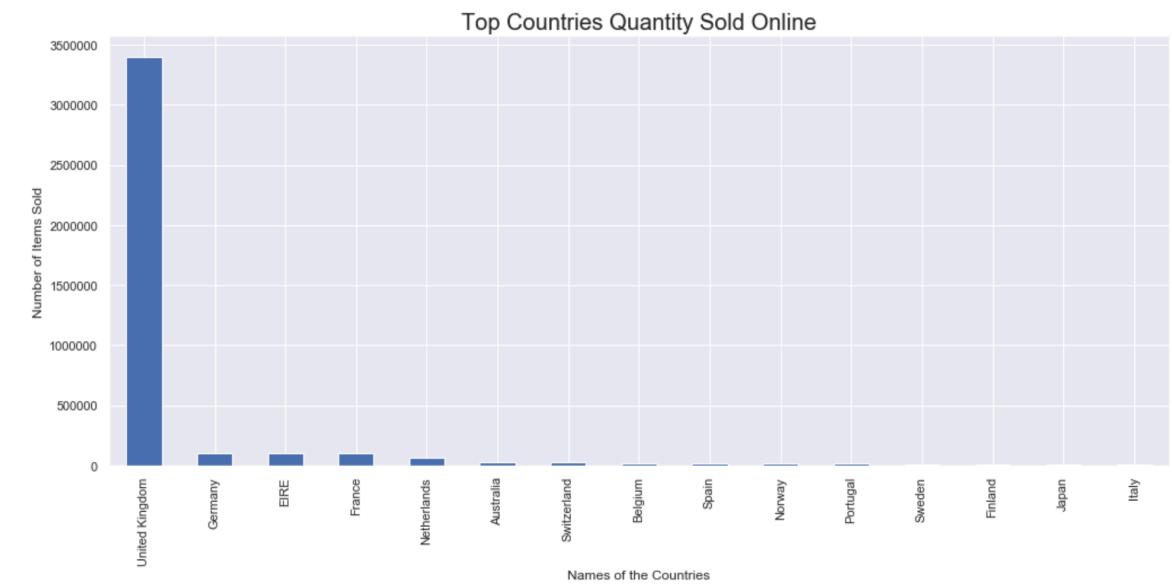
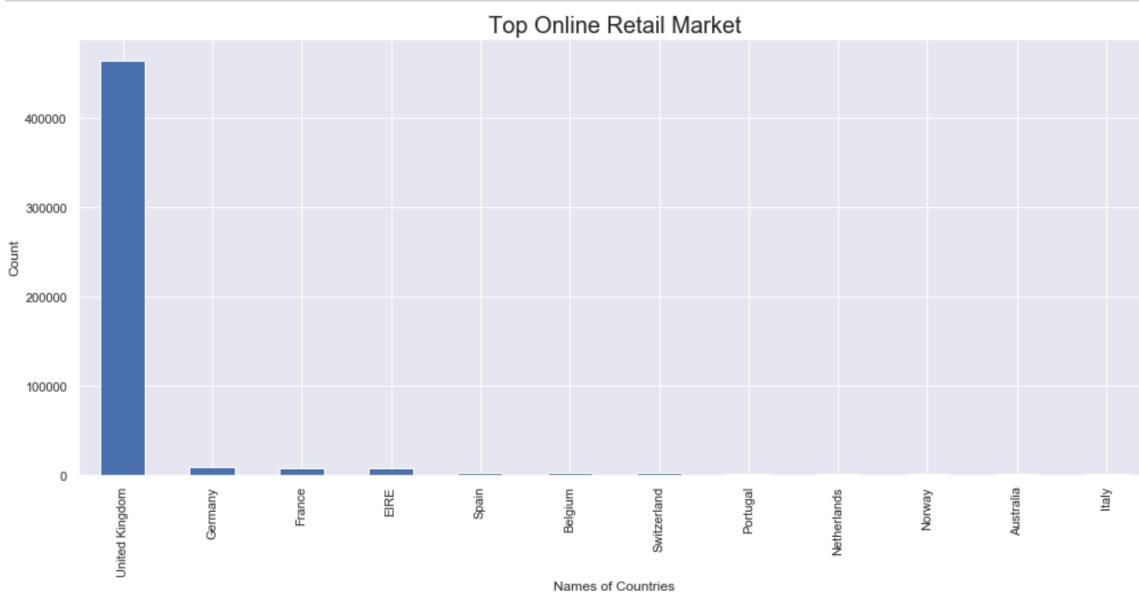
PROBLEM STATEMENT

- Due to the pandemic, online sales became very crucial for businesses. Predicting sales and RFM analysis is one of the most important business problems for any retail entity.
- An RFM analysis can help and be used in conjunction with sales prediction models to gain even further insight into customer behavior and sales predictions for online businesses.



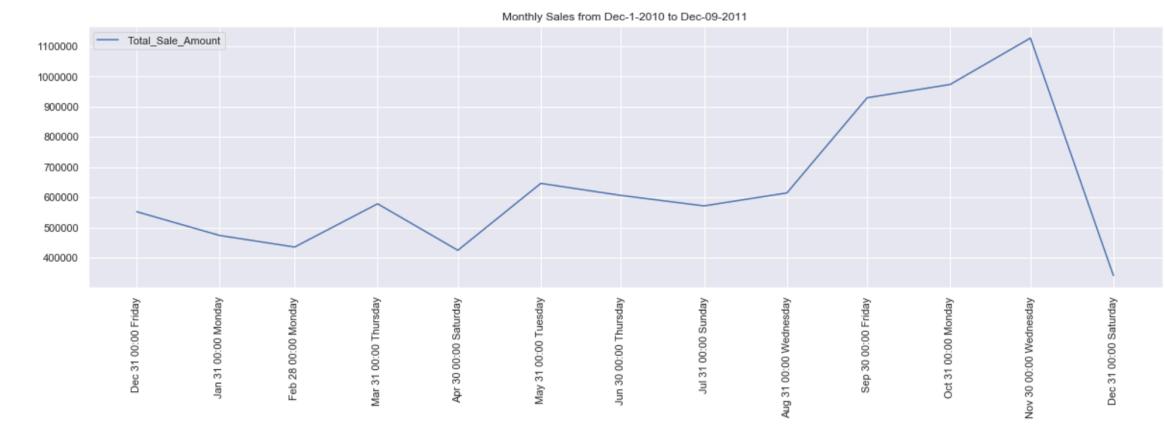
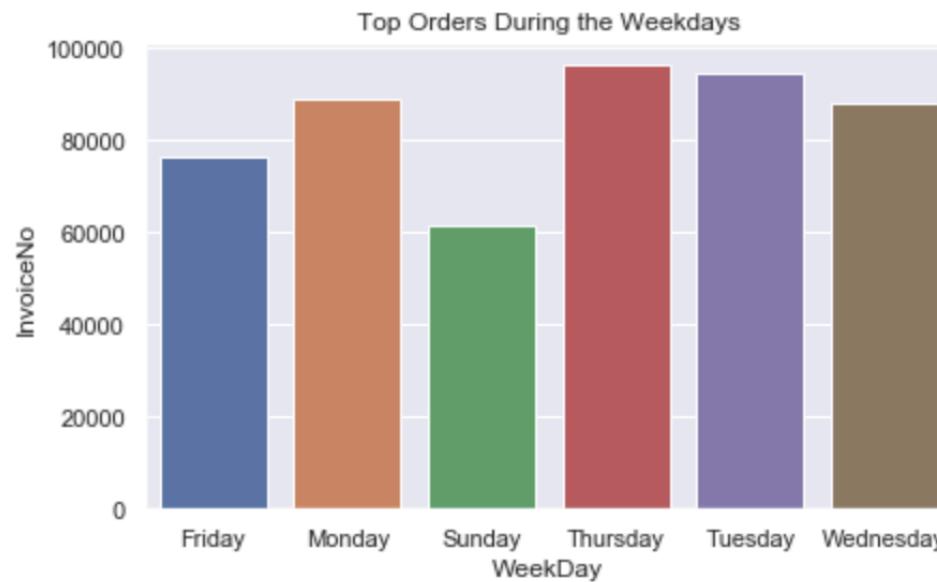
OBSERVATIONS

- Most sales comes from UK and about 90% customers also are from UK
- Outside UK, most of the sales is from Europe
- In international sales, Germany, France, and Ireland are among the highest
- Outside UK, most customers are from Germany and France
- Outside Europe, highest sales comes from Australia, while the highest number of cancellations come from United States.



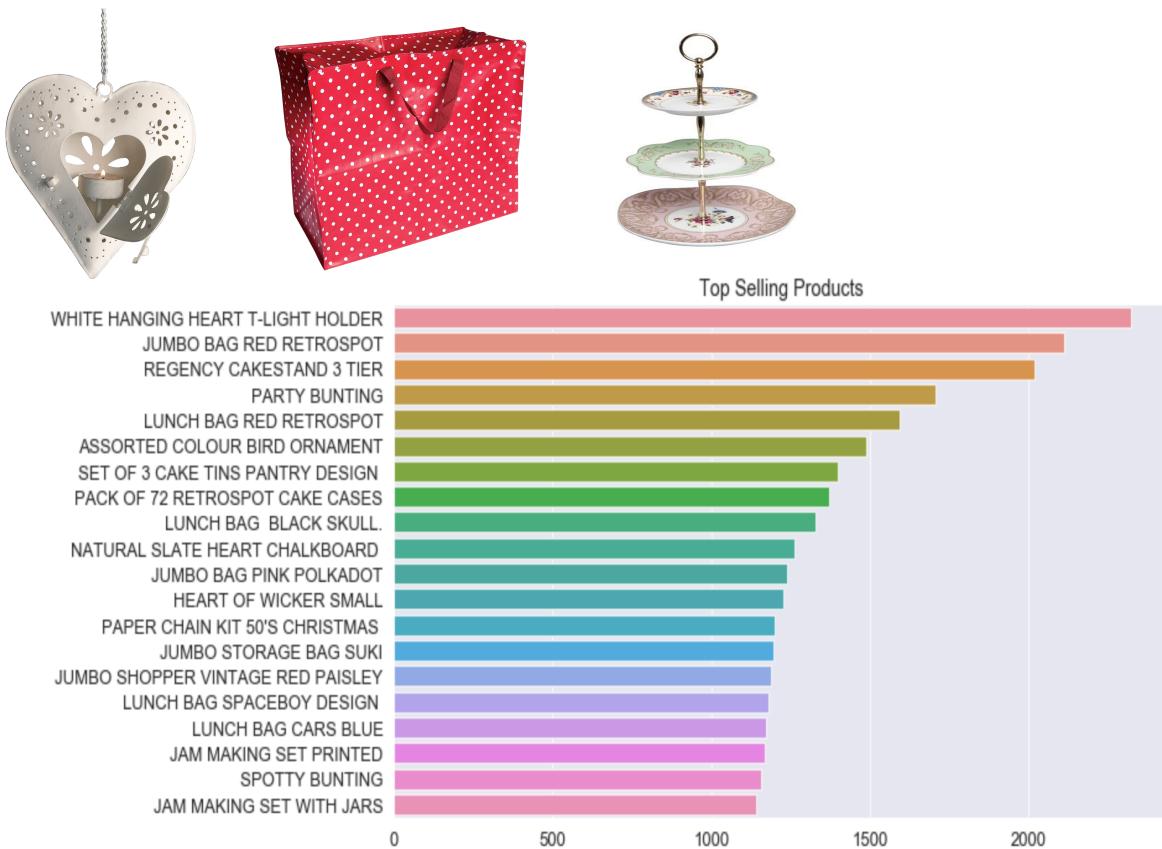
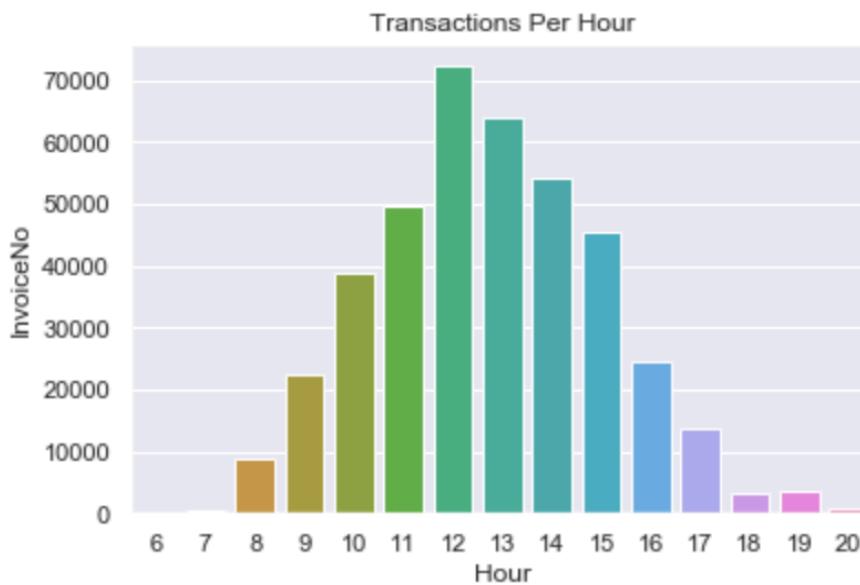
MONTHLY SALES & WEEKDAY SALES

- There is no sales happening on Saturdays.
- Sales falls during the holidays
- The overall sales has increased through the months
- Most orders happened on Thursdays and Tuesdays



TOP SELLING PRODUCTS AND TRANSACTION HOURS

- Sales happens only during the working hours - between 8am and 8pm
- Most sales happening between 12 – 2 pm
- The No. 1 best selling product sells almost double the quantity of the 10th best seller

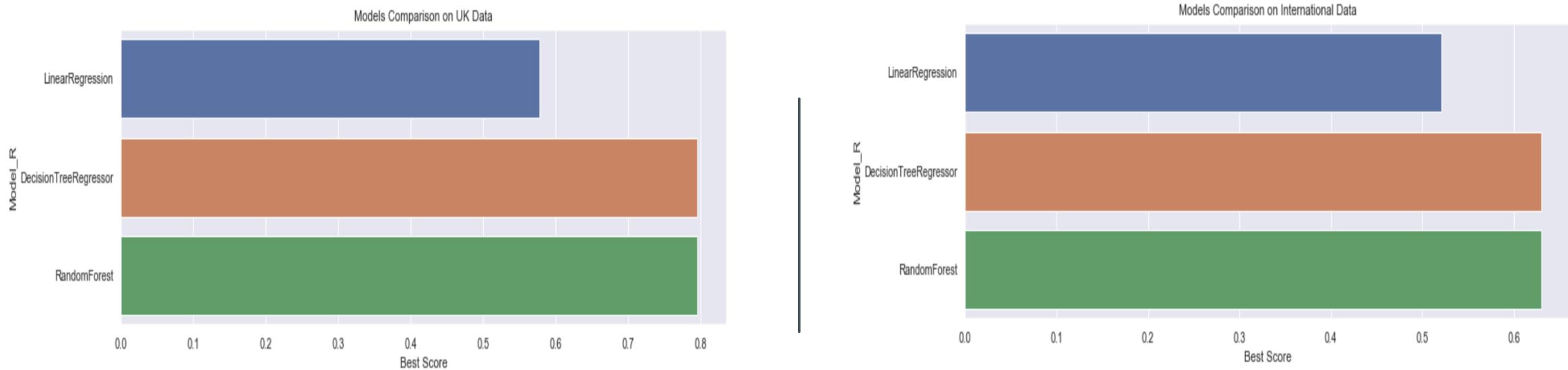


RFM (RECENCY, FREQUENCY, MONETARY) ANALYSIS

Best Customers: 618
Loyal Customers: 1080
Big Spenders: 1087
Almost Lost: 58
Recent Customers: 154
Lost Customers: 14
Lost Cheap Customers: 361

- The RFM analysis is a marketing framework that seeks to understand and analyze customer behavior based on three factors: recency, frequency, and monetary
- Recency: When was the last time the customer bought something from your store?
- Frequency: How often do customers come to the store?
- Monetary value: How much has the customer spent in the past?
- RFM score is from 111 thru 444 to each customer.
- 4 is best and highest value and 1 is the lowest and worst value
- RFM Score can be very helpful for businesses and marketers how to target or deal with each segment to improve their sales.

		Recency	Frequency	Monetary	Rank	R_Quartile	F_Quartile	M_Quartile	RFM_Score	Customers Segments
CustomerID	Country									
17504	United Kingdom	3516	127	2997.03	556.0	1	4	4	144	Lost Customers
13588	United Kingdom	3332	174	2478.95	710.0	4	4	4	444	Best Customers
17289	United Kingdom	3332	104	931.02	1741.0	4	4	3	443	Recent Customers
15549	United Kingdom	3393	114	1681.46	1074.0	2	4	4	244	Almost Lost
12386	Australia	3636	10	401.90	2848.0	1	1	2	112	Others
15863	United Kingdom	3332	18	68.84	4279.0	4	2	1	421	Others
12758	Portugal	3424	97	2791.59	603.0	2	3	4	234	Others
12479	Germany	3332	31	527.20	2512.0	4	2	2	422	Others
13267	United Kingdom	3302	211	4521.49	313.0	4	4	4	444	Best Customers
16031	United Kingdom	3393	98	547.90	2459.0	2	3	2	232	Others



MODELING TO PREDICT THE SALES

- We ran three models to test our model: Linear regression, Decision Tree Regressor, and Random Forest Regressor.
- We created two datasets to test our model with it: UK and International datasets
- Random Forest Regressor appears to be the best model for our prediction on our datasets



Why there is no sales on Saturdays



How to improve sales during the holidays



There is still room for model improvement



Building a reward system



Our model can be replicated to any similar online or physical store selling any kind of product

FURTHER ANALYSIS AND CONCLUSION

ANY QUESTIONS

