Online Retail Customer Segmentation

Topics Covered

- Data import
- Data cleaning and wrangling
- EDA
- Preparing data attributes for modeling
- Building ML model
- Conclusion and future work

Goal

- ❑ Perform Cohort Analysis on Online Retail Data
- ❑ Conduct RFM Analysis Based on Online Retail Data
- ❑ Build Customer Segments Using K-Means Clustering Model
- ❑ Introduce the observations and results from this project to the Online Retail marketing department and collaborate in order to apply these insights to marketing strategies.
- ❑ GitHub link to the EDA and ML notebooks and utility functions can be found here.

Online Retail Data

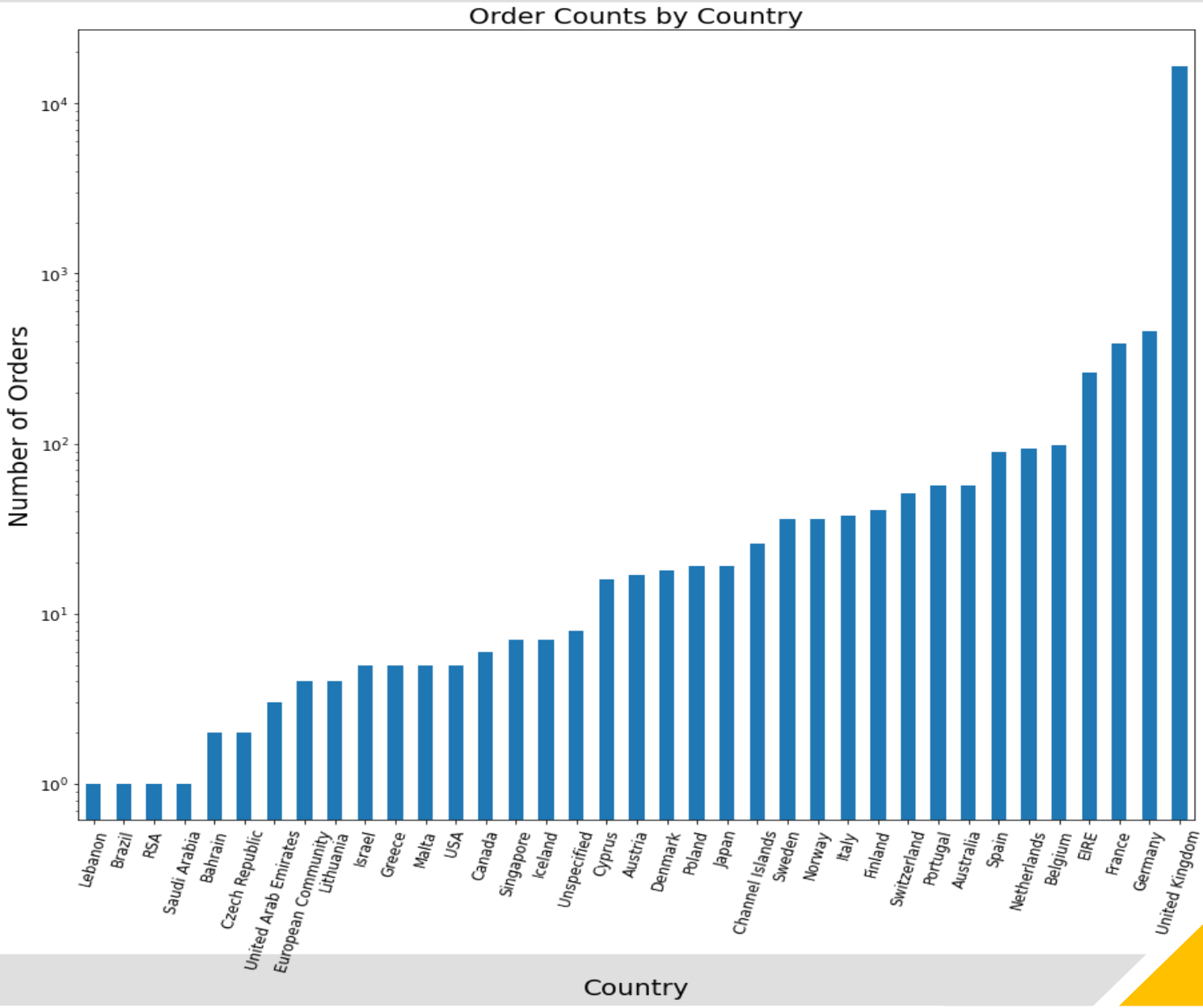
- Downloaded from Kaggel
- Contains all the transactions made from 2010-12-01 to 2011-12-09
- There were originally 541,909 rows and 8 columns in the data
- Each row was associated with one product in an order

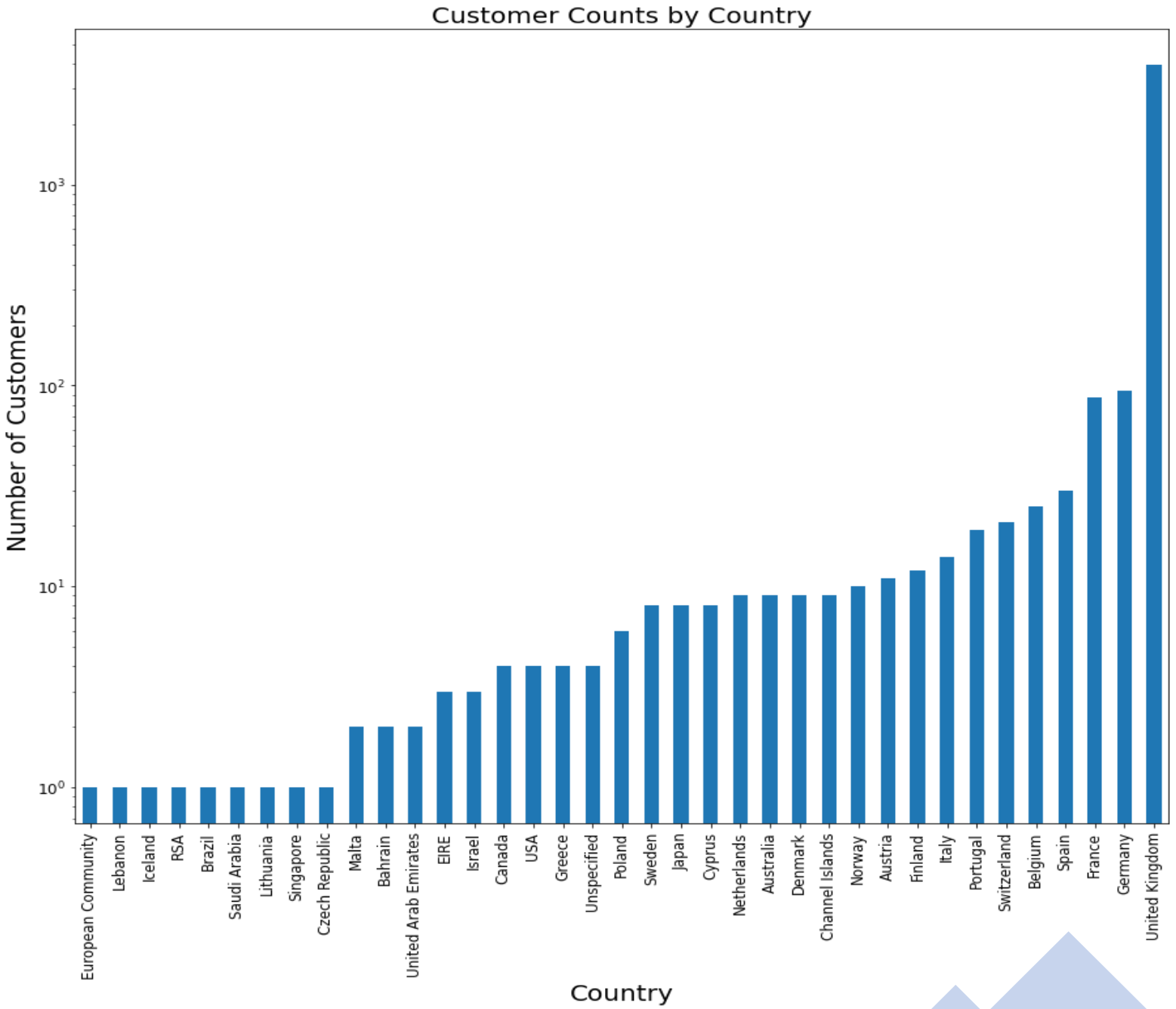
Data Field

- **InvoiceNo**: Invoice number. Nominal. A 6-digit integral number uniquely assigned to each transaction. If the InvoiceNo starts with the letter 'C', it indicates a cancellation.
- **StockCode**: Product code. Nominal. A 5-digit integral number uniquely assigned to each distinct product.
- **Description**: Product (item) name. Nominal.
- **Quantity**: The quantities of each product (item) per transaction. Numeric.
- **InvoiceDate**: Invoice date and time. Numeric. The day and time when a transaction was generated.
- **UnitPrice**: Unit price. Numeric.
- **CustomerID**: Customer number. Nominal. A 5-digit integral number uniquely assigned to each customer.
- **Country**: Country name. Nominal. The name of the country where a customer resides.

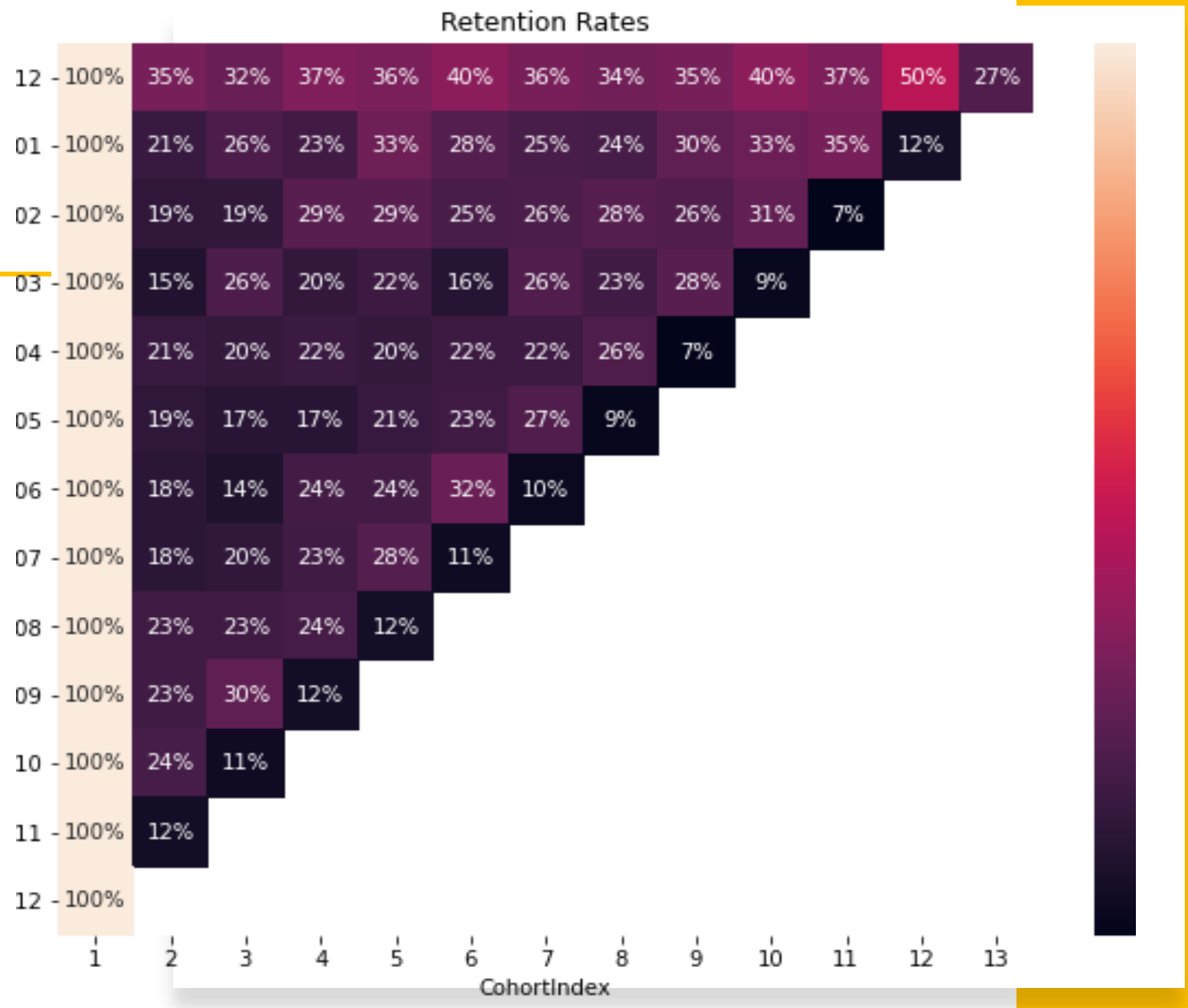
Data Cleaning

- ✓ *Removed 135,080 rows missing Customer ID*
- ✓ *Removed 8,905 rows with negative Quantity*
- ✓ *Removed 40 rows with UnitPrice 0*
- ✓ *Dropped Description column*

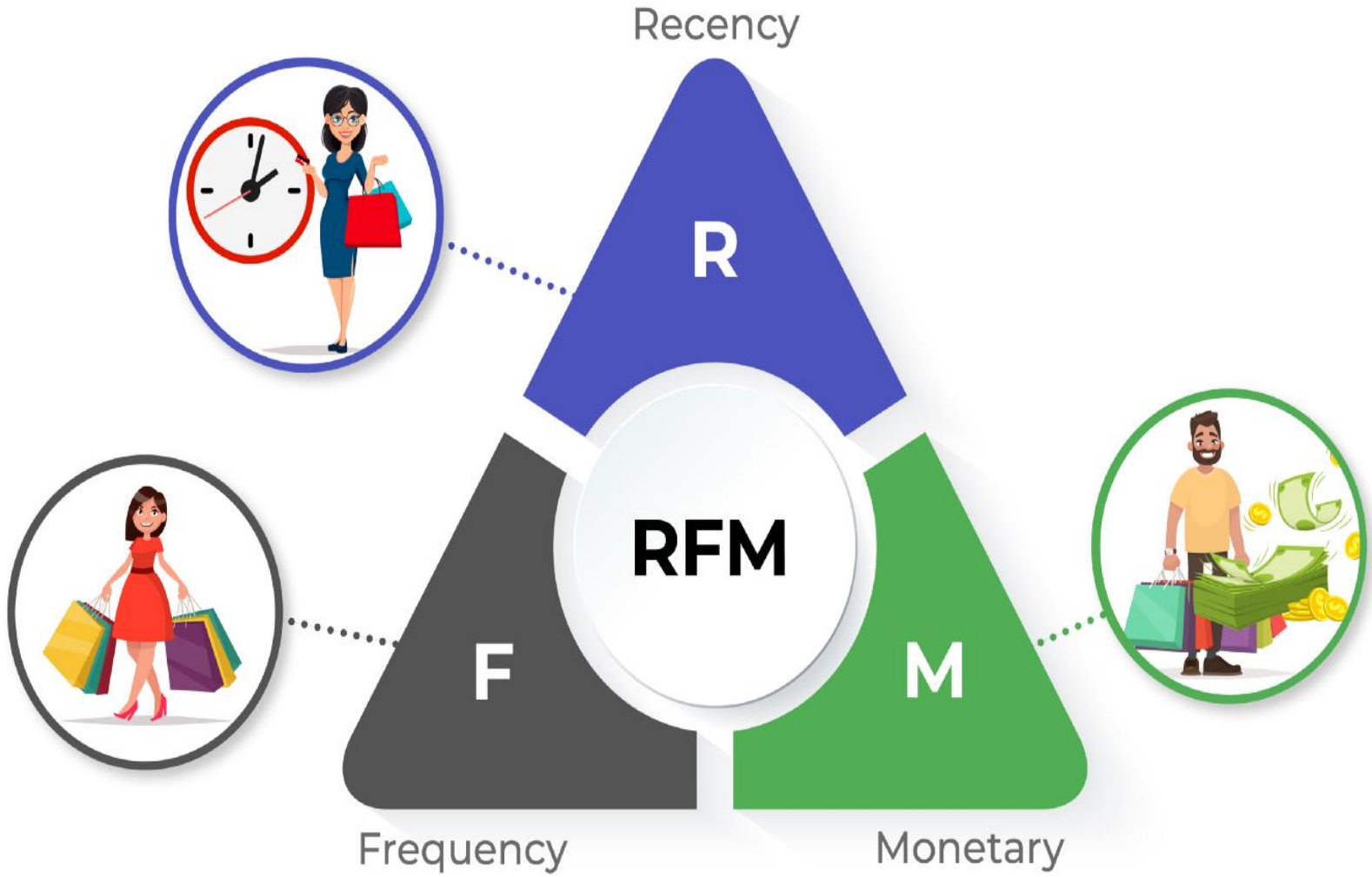

Order Counts by Country
Number of Orders
10^0
10^1
10^2
10^3
10^4
Lebanon
Brazil
RSA
Saudi Arabia
Bahrain
Czech Republic
United Arab Emirates
European Community
Lithuania
Israel
Greece
Malta
USA
Canada
Singapore
Iceland
Unspecified
Cyprus
Austria
Denmark
Poland
Japan
Channel Islands
Sweden
Norway
Italy
Finland
Switzerland
Portugal
Australia
Spain
Netherlands
Belgium
EIRE
France
Germany
United Kingdom
Country


Customer Counts by Country
Number of Customers
10^3
10^2
10^1
10^0
European Community
Lebanon
Iceland
RSA
Brazil
Saudi Arabia
Lithuania
Singapore
Czech Republic
Malta
Bahrain
United Arab Emirates
EIRE
Israel
Canada
USA
Greece
Unspecified
Poland
Sweden
Japan
Cyprus
Netherlands
Australia
Denmark
Channel Islands
Norway
Austria
Finland
Italy
Portugal
Switzerland
Belgium
Spain
France
Germany
United Kingdom
Country

UK Customer Retention Rates



RFM Customer Segmentation

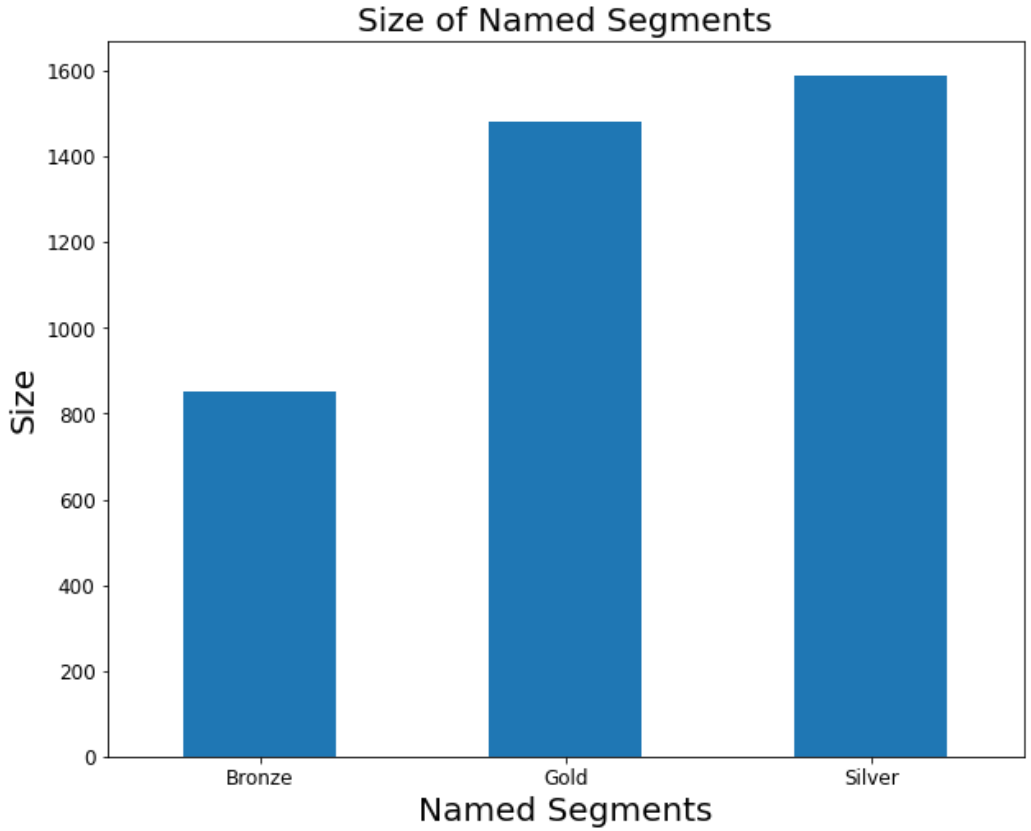


Summary Metrics of RFM Segmentation

RFM Score	Count	Recency Mean	Frequency Mean	Monetary Value Mean
3	396	265.5	1.0	158.6
4	454	184.7	1.1	280.7
5	443	110.5	1.3	363.6
6	407	90.3	1.7	704.1
7	374	76.6	2.3	698.8
8	366	58.7	3.0	1126.1
9	409	45.8	4.0	1401.8
10	347	30.0	5.2	2337.6
11	301	21.1	8.0	3476.3
12	423	7.7	15.8	8469.8

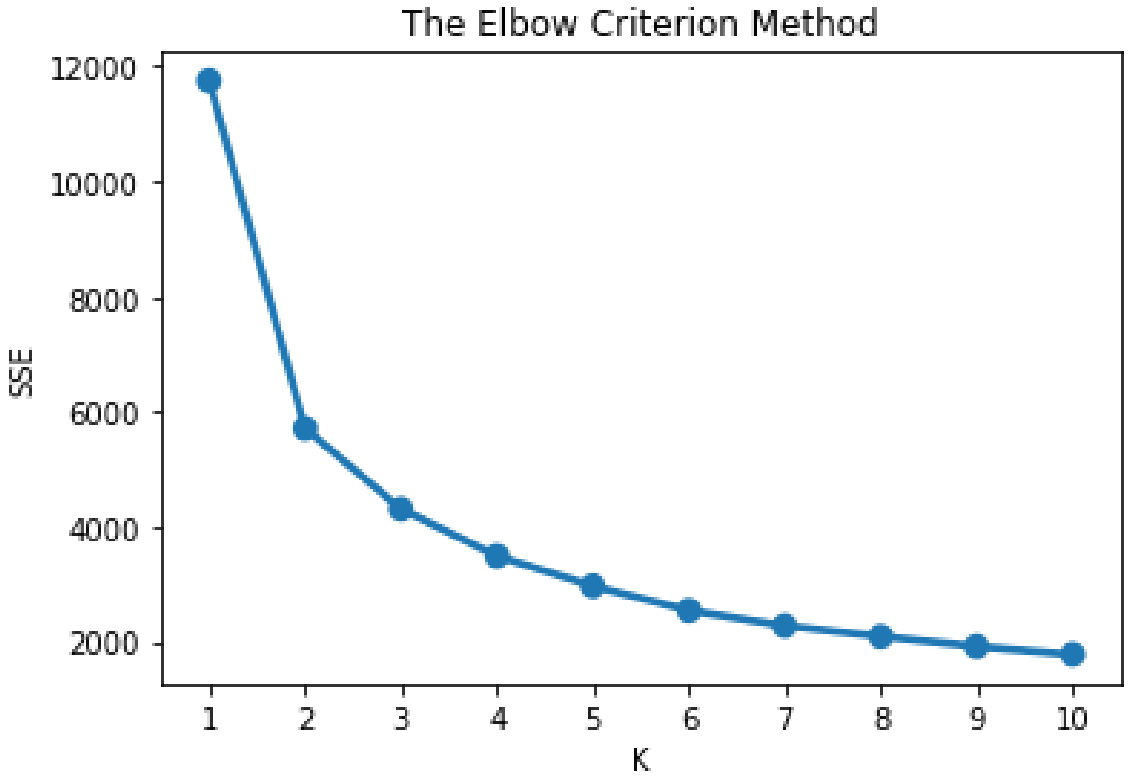
Named Customer Segments

Based on customers' RFM Scores, we also group customers into named segments:
Gold (RFM Score > = 9), Silver (9>= RFM Score >= 5), and Bronze (RFM Score <= 4).



Summary Metrics for Named Segments

Named Segment	Count	Recency Mean	Frequency Mean	Monetary Value Mean
Gold	1480	26.2	8.5	$4,063
Silver	1590	85.4	2.0	$705
Bronze	850	222.4	1.1	$223



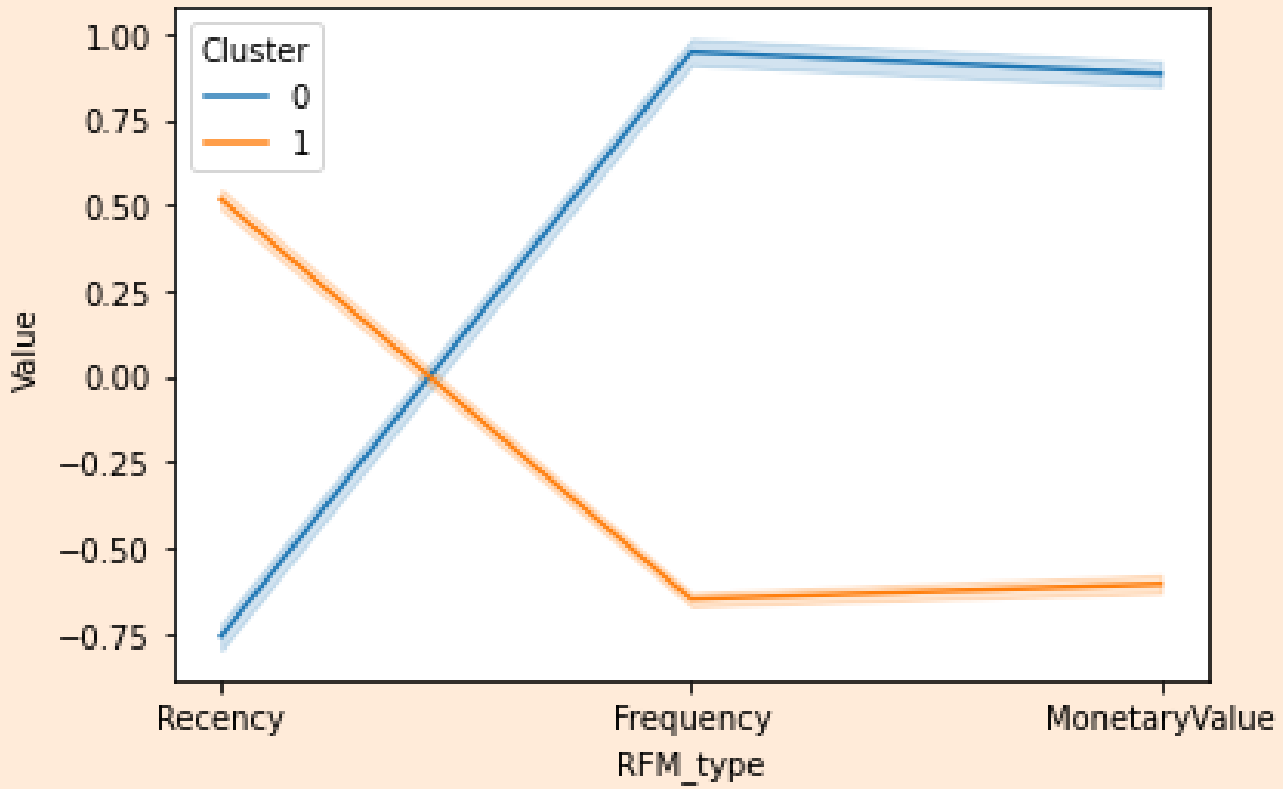
Machine Learning Modeling

- K-Means Clustering Model
- Used the elbow criterion method to choose number of clusters
- The best: K=2

The Summary Metrics of K-Means Clusters

Cluster	Count	Recency Mean	Frequency Mean	Monetary Value Mean
0	1592	28.7	8.1	$3,957
1	2328	136.5	1.6	$440

The Plot of Standardized RFM



The average Recency, Frequency and Monetary Value of customers in cluster 0 are much better than those of customers in cluster 1.

Conclusion

- Target Silver segment
  - Largest group + most potential
  - Ex: offer deals on next purchase before a certain date, bundle discounts
- Maintain Gold segment
  - Best customers + most profitable
  - Ex: implement a point system with exclusive benefits to reward continued customer involvement

Future Work

- Distinguish between recency, frequency, and monetary value in terms of impact on overall RFM score (to better understand customer behavior within each segment)
- Pinpoint the products that have been purchased most frequently and most recently