

CUSTOMER SEGMENTATION IN E-COMMERCE

Ran Ma

INTRODUCTION



DATA SOURCE

1. Source: Online retail transaction from 2010 to 2011
2. Size: 8 columns, 460k rows

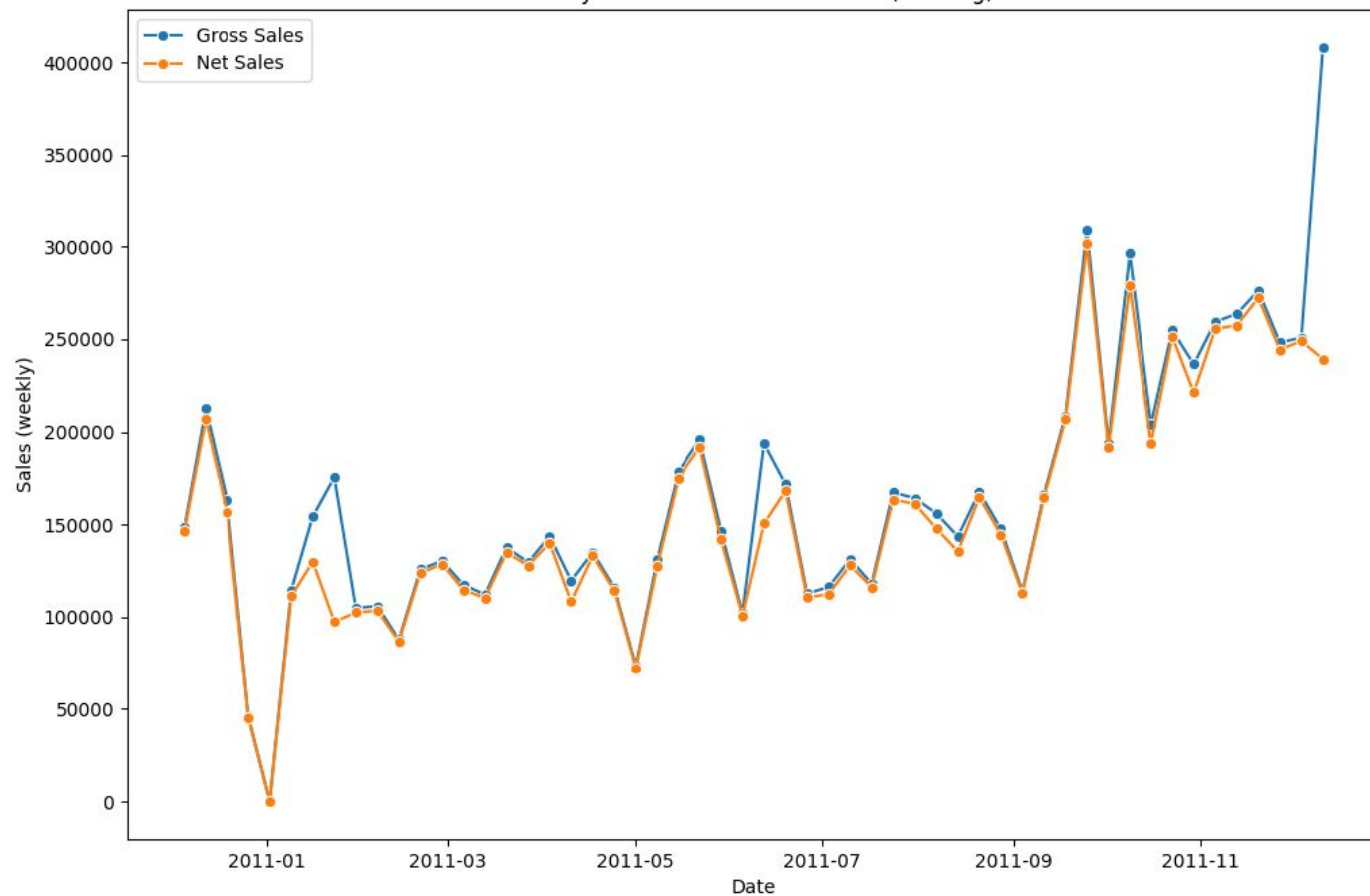
DATA CLEANING AND PREPROCESSING

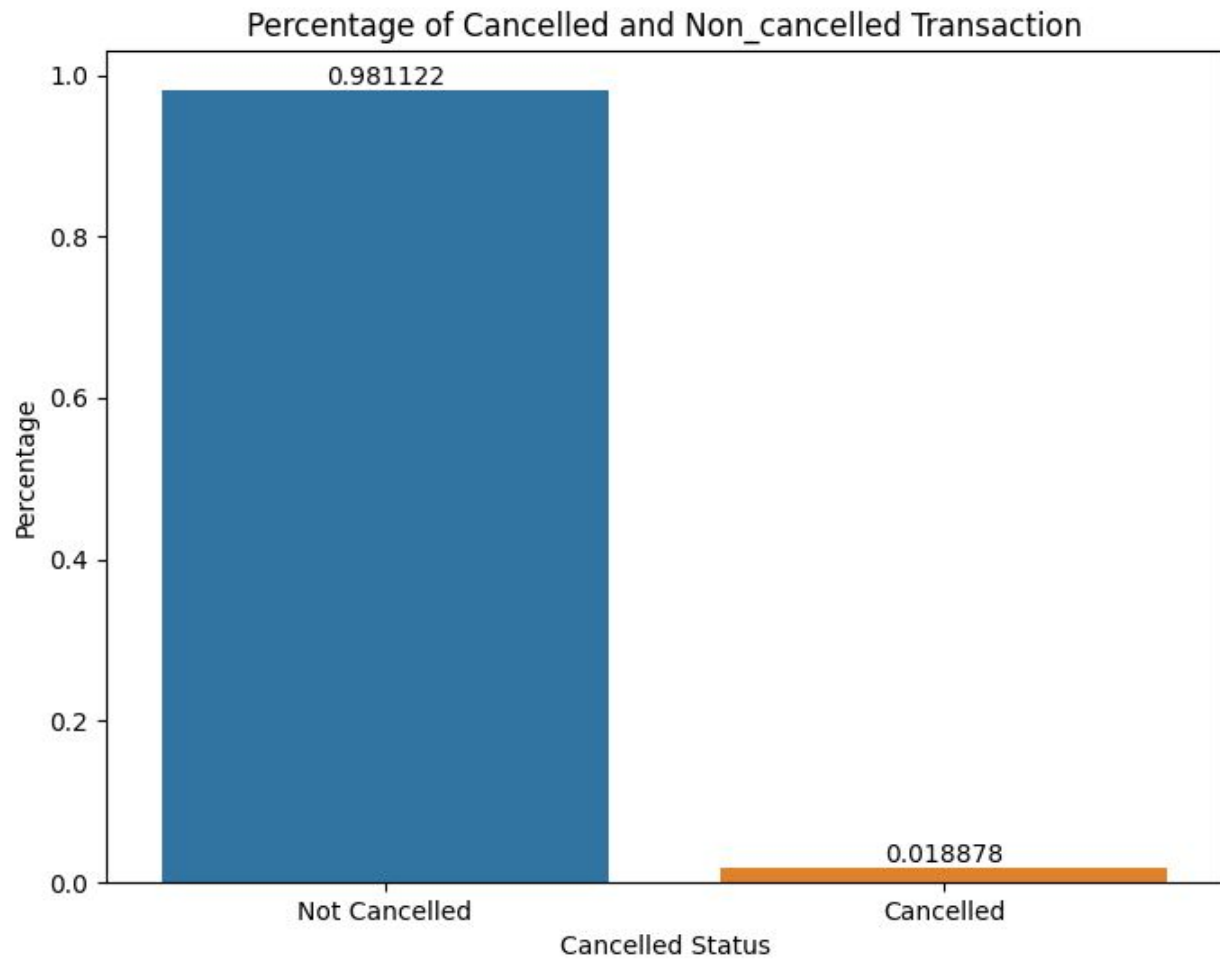
- Filtering Out Entries Without Customer ID
- Exclusion of Cancelled Transactions
- Identification and Marking of Corresponding Previous Order Transactions

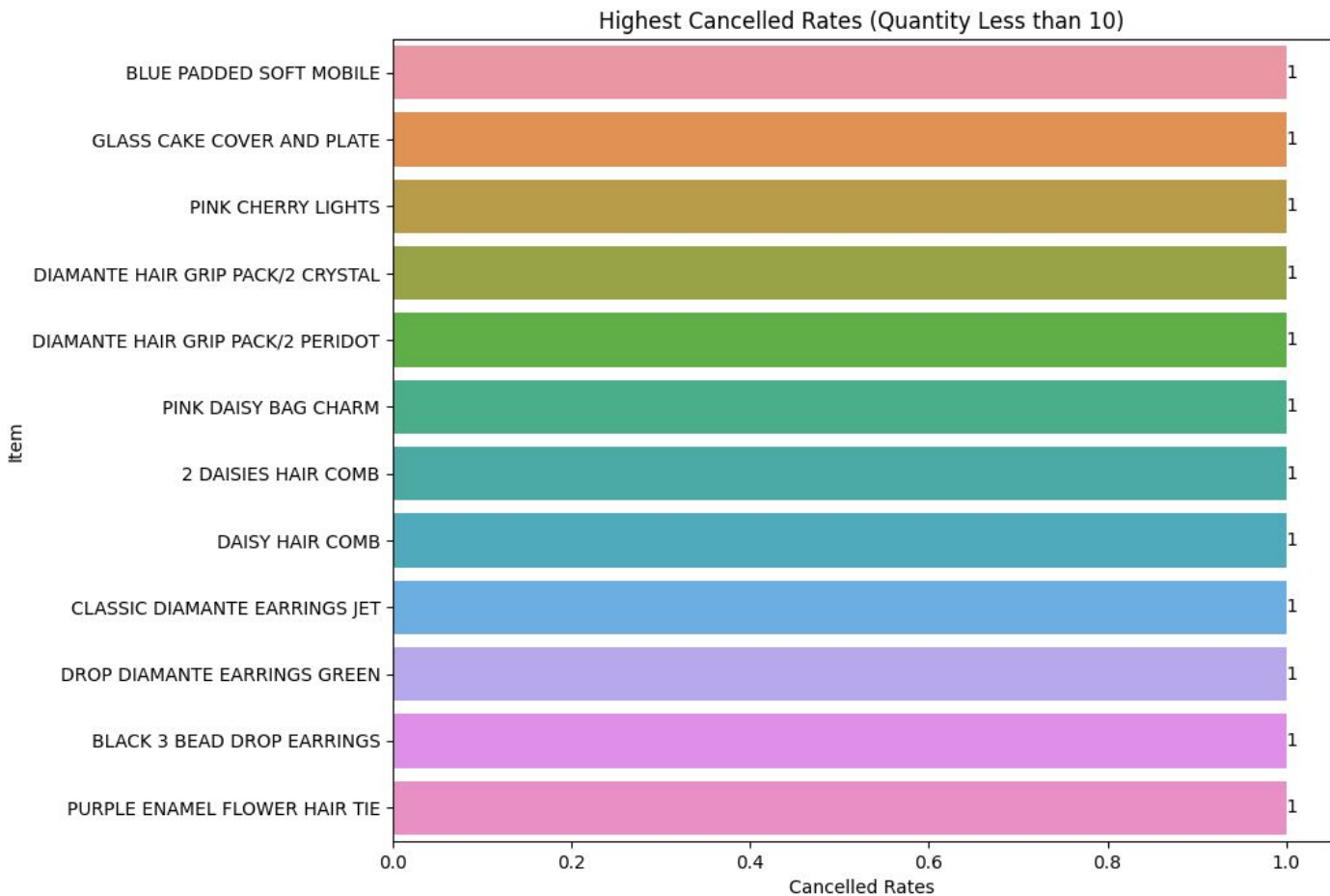
DATA EXPLORATION

- Approach to Sales Trend Analysis
- Analysis of Cancelled Transactions
- Insights into Customer Geographic Distribution

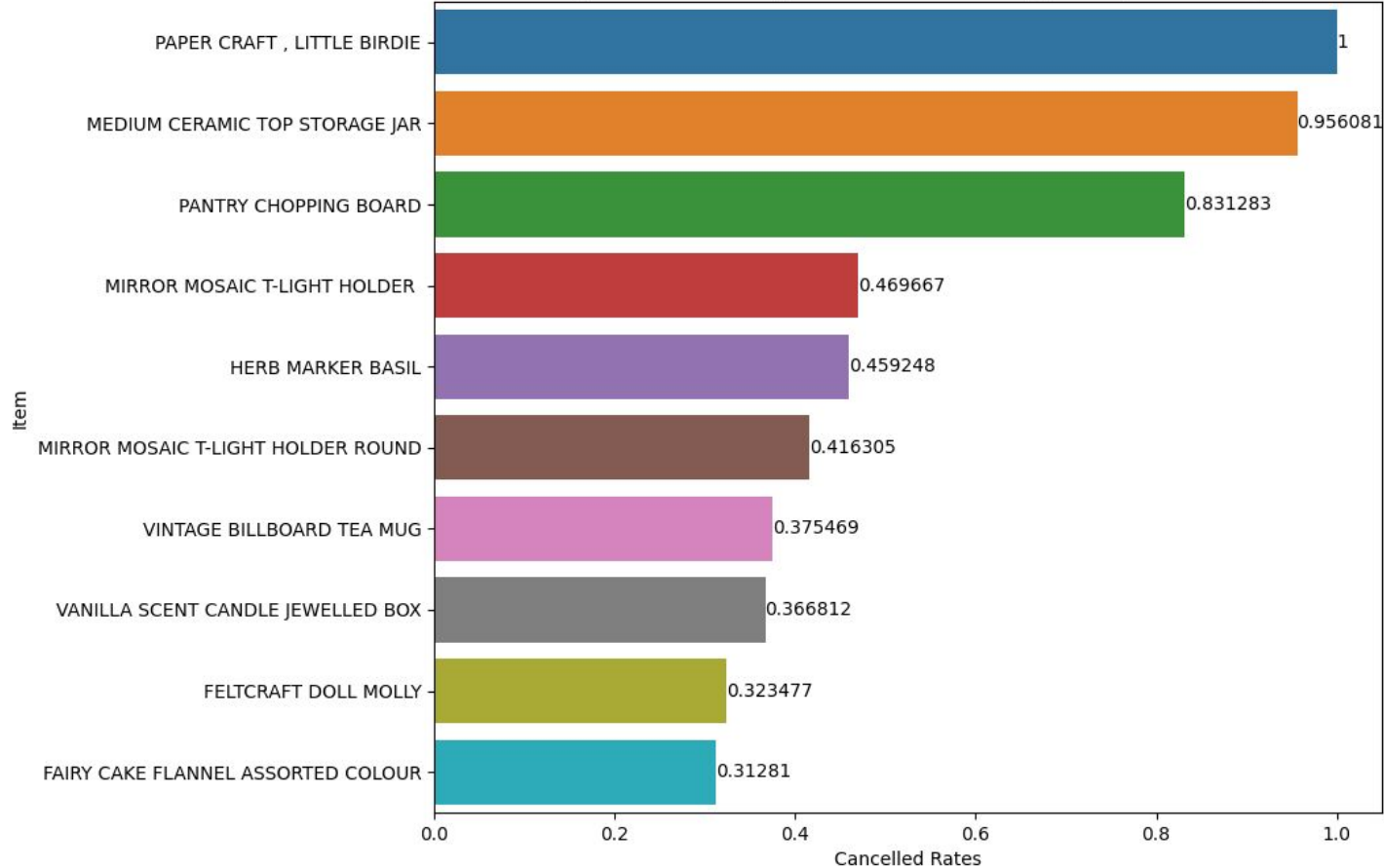
Weekly Gross Sales vs. Net Sales (Sterling)



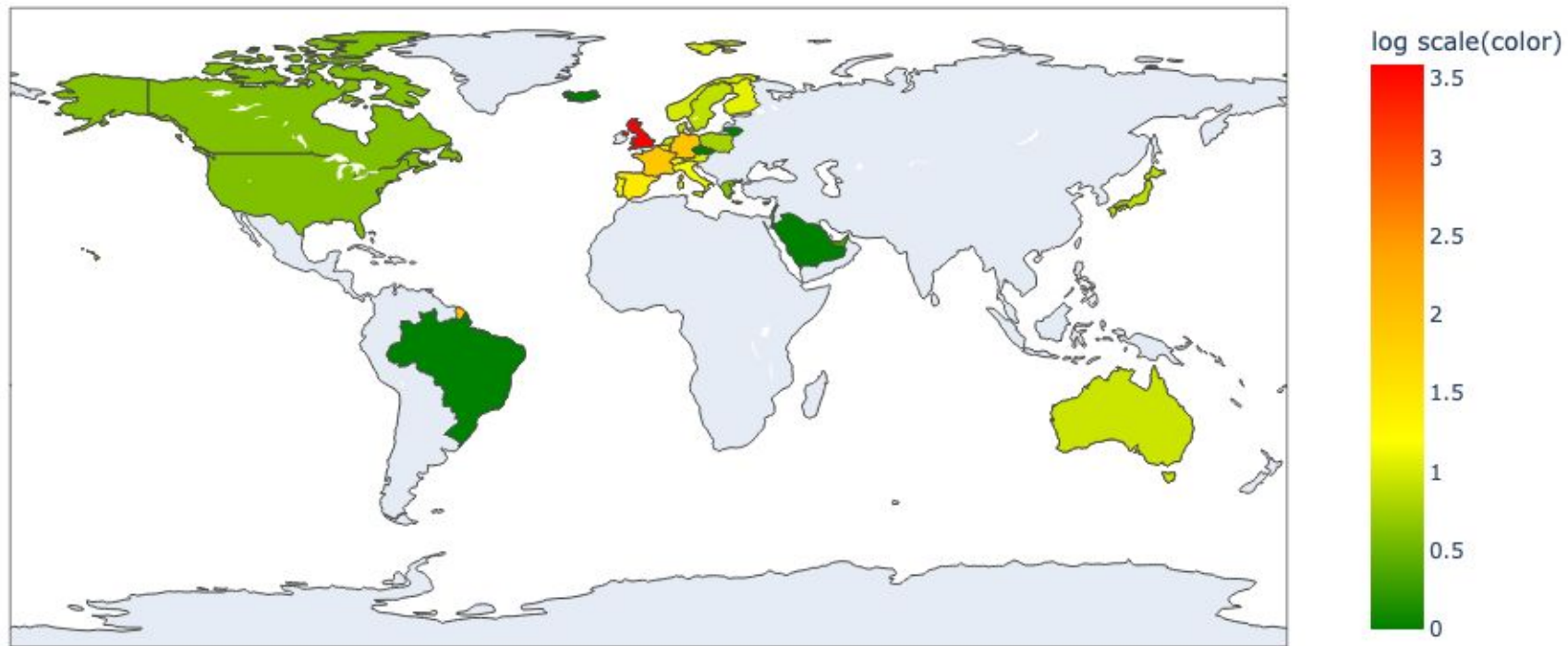




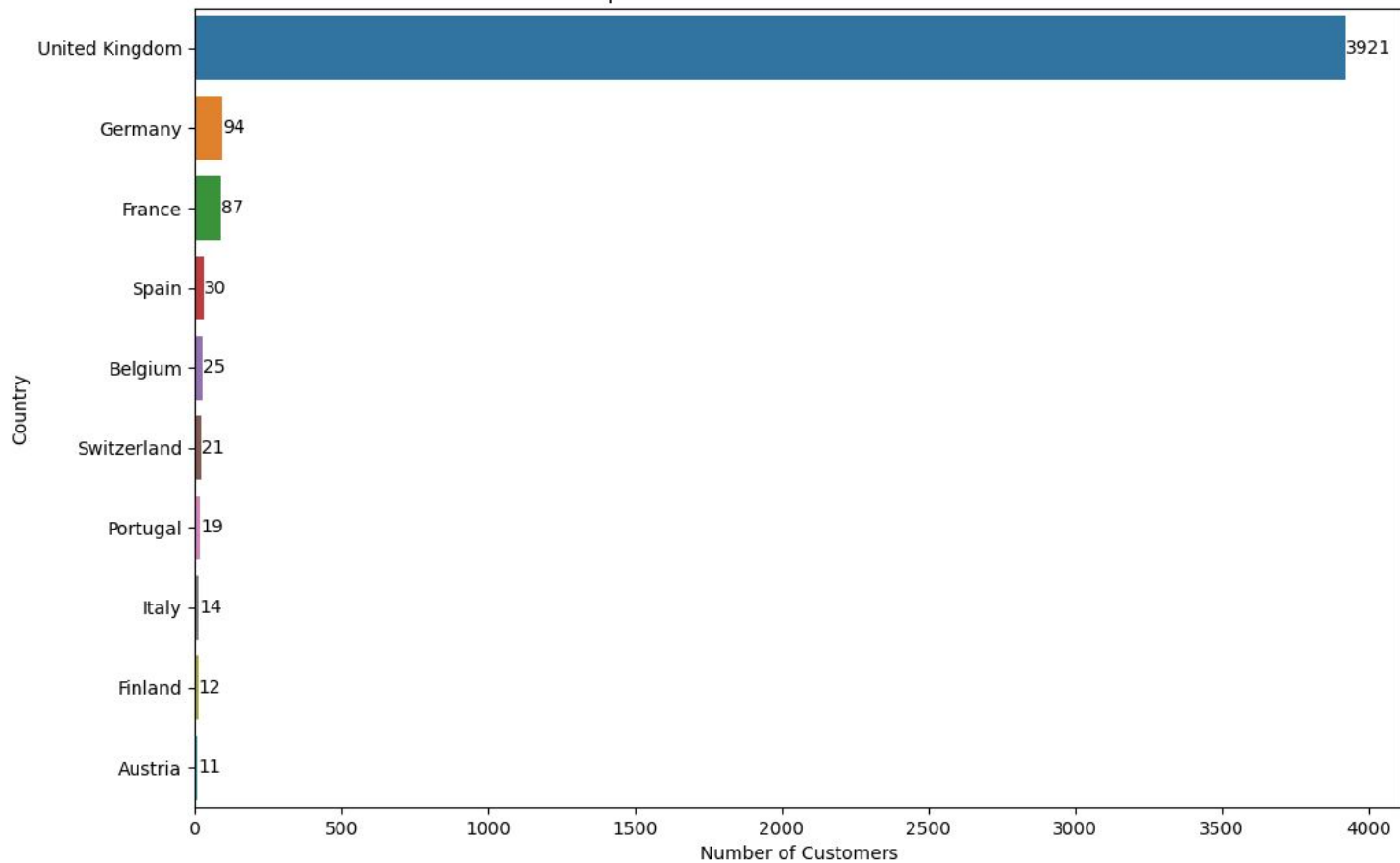
Top 10 Items with Highest Cancelled Rates (Quantity Over 1000)



Number of Customers by Country



Top Contries with Number of Customers



CUSTOMER SEGMENTATION - RFM ANALYSIS

Recency: The period since a customer's last purchase

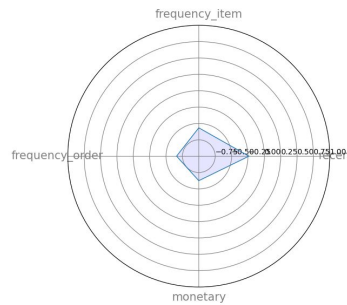
Frequency: A measurement of a customer's purchasing regularity.

Monetary: An estimate of a customer's total spending.

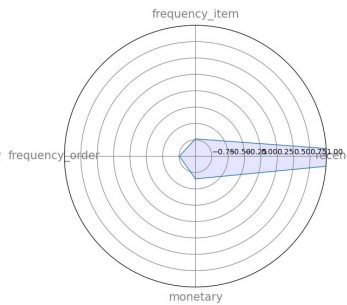
Derived Columns: recency, frequency_item, frequency_order,
monetary

Clustering Approach: KMean (k = 8)

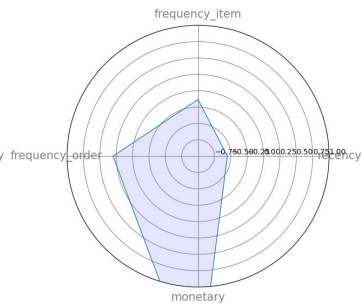
Cluster 0



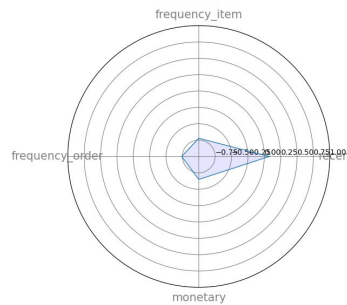
Cluster 1



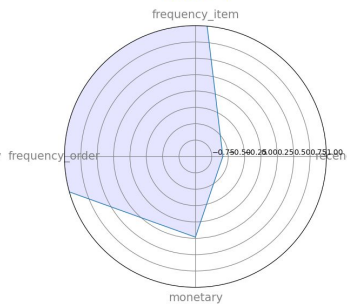
Cluster 2



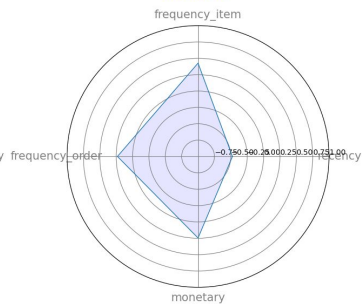
Cluster 3



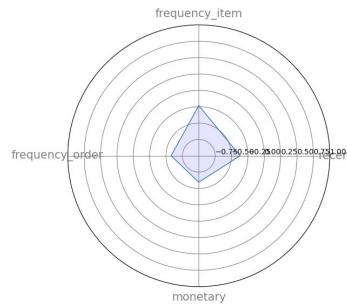
Cluster 4



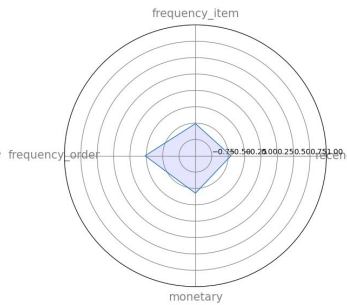
Cluster 5



Cluster 6



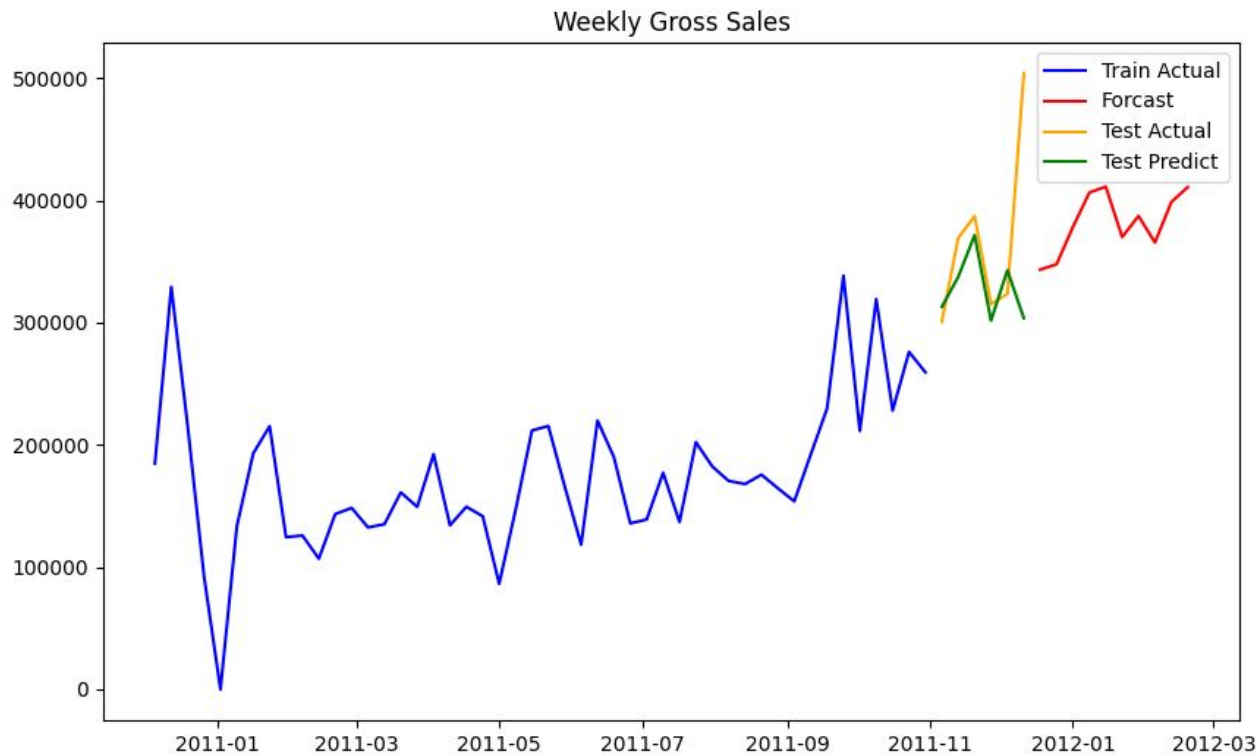
Cluster 7



CUSTOMER SEGMENTATION - CLASSIFICATION

	balanced_accuracy	recall	precision	f1_score
Logistic Regression	0.930339	0.979724	0.980729	0.979770
Multinomial Naive Bayes	0.546568	0.595392	0.672699	0.606465
K-Nearest Neighbor	0.989813	0.979724	0.979863	0.979742
Decision Tree	0.989813	0.979724	0.979863	0.979742
Random Forest	0.989813	0.979724	0.979863	0.979742
XGBoost	0.989813	0.979724	0.979863	0.979742
Average Weighting	0.989813	0.979724	0.979863	0.979742
Stacking	0.859300	0.982488	0.981628	0.982029

SALES FORECASTING



CONCLUSION AND RECOMMENDATION

- Cancelled Items: Scrutinize items with high cancellation rates.
- Inactive Customers: Encourage engagement through incentives.
- High-value Segments: Boost loyalty strategies for these customers.
- Sales Forecasting: Leverage model for efficient operational Planning.

THANK YOU