Customer Analytics dashboard:

Understanding customers is of paramount importance for any business. This Customer Analytics dashboard aims to analyse customers and their purchasing habits of a superstore. For this project, I used the Sample superstore dataset from Kaggle.

Methodology:

I began with data exploration to review the data. I cleaned the data next by removing the rows with missing values and also those sales row with negative or zero pricing. I also removed the duplicates and updated the date column to date type. It was earlier in an object type.

I then proceeded calculate the RFM metrics by defining the table, then creating the table and also reset the index. I looked up additional customer features such as looking up number of unique products purchased per customer , total transaction per customer and merged them in a table. Then calculated product diversity.

After visualizations, I next worked on Customer segmentation with K-Means Clustering and the Elbow method to find the optimal k. I applied the K-Means clustering and also created a pac for 2d visualization. This was followed by building cluster profiles and comparing them and then building customer personas based on cluster characteristics.

Market Basket Analysis: The data was then prepared for market basket analysis and applied the Apriori algorithm and generated the Association rule. Finally, worked on the product recommendation function and for the Business insights and recommendations.

Summary: Most customers haven't purchased in a long time, indicating possible churn or inactivity. Most customers have made between 5 and 25 purchases with the majority of customers spending less than $5000 total. There is only a small amount of high spenders which may be valuable for targeted marketing. The plot was heavily skewed with most values at or near 1, indicating single-production transactions are common.

The biggest challenge for this project was finding good data source. Even with this dataset, I had to lower my numbers to get the association rules. There simply weren't that many customers buying many products. This dataset had a large based of moderately engaged, low to mid spend customers.

References

- Sample_superstore_ Kaggle
- Professor Avinash Jairam Customer_segmenation collab notebook