Bank Customer Segmentation

Analysis of Customer Demographics & Transaction data

Project Brief

Source & wrangle Data → Derive Variables & Explore Relationship → Geospatial & Time series Analysis → Linear Regression & Kmeans clustering → Visualization & StoryTelling

Project Goals

- Identify distinct groups of customers based on their behavior or other traits. It enables efficient marketing resource allocation, the maximization of point-based approach to each customer group, as well as sales opportunities. It helps in enhancing customer service and retaining customers which are crucial for the banking industry.

Data

- The data was collected as a part of research project in collaboration with a bank. The dataset includes Customer demographics and transactions data from an Indian Bank.
- This data is an open-source data and downloaded from Kaggle.com.

Tools used

- Python, Anaconda, Jupyter Notebook
- Libraries (Pandas, Numpy, Seaborn, Matplotlib, Scipy, datetime, calendar, folium, sklearn, pylab, statsmodels.api)
- Tableau

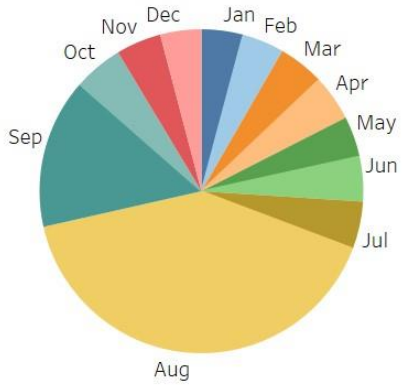
Skills

- Data Sourcing, Wrangling & Deriving Variables
- Exploratory Visual Analysis
- Geospatial Analysis & Time-Series Analysis
- Machine Learning –Linear regression & Kmeans Clustering
- Visualization & Storytelling

Exploratory Analysis

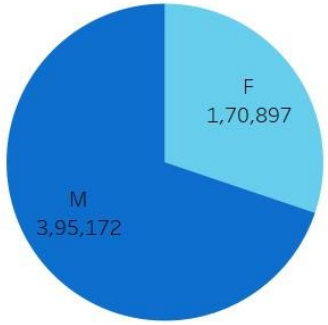
Month with the most and expensive transactions

The most transactions, as well as the most expensive transactions, occur in August.



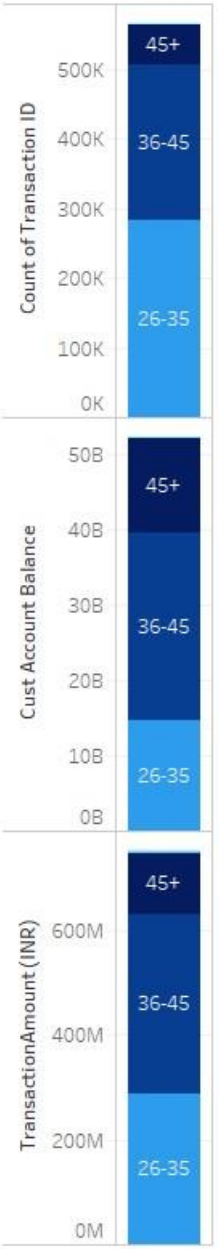
Gender Categorization

Comparing male and female customers, male customers are slightly more likely to spend money, have greater account balances, and make more transactions.



Age Categorization

Age groups 26 to 35 have more transactions than other age groups, while 36 to 45 have more expensive transactions and higher account balances than other age groups.

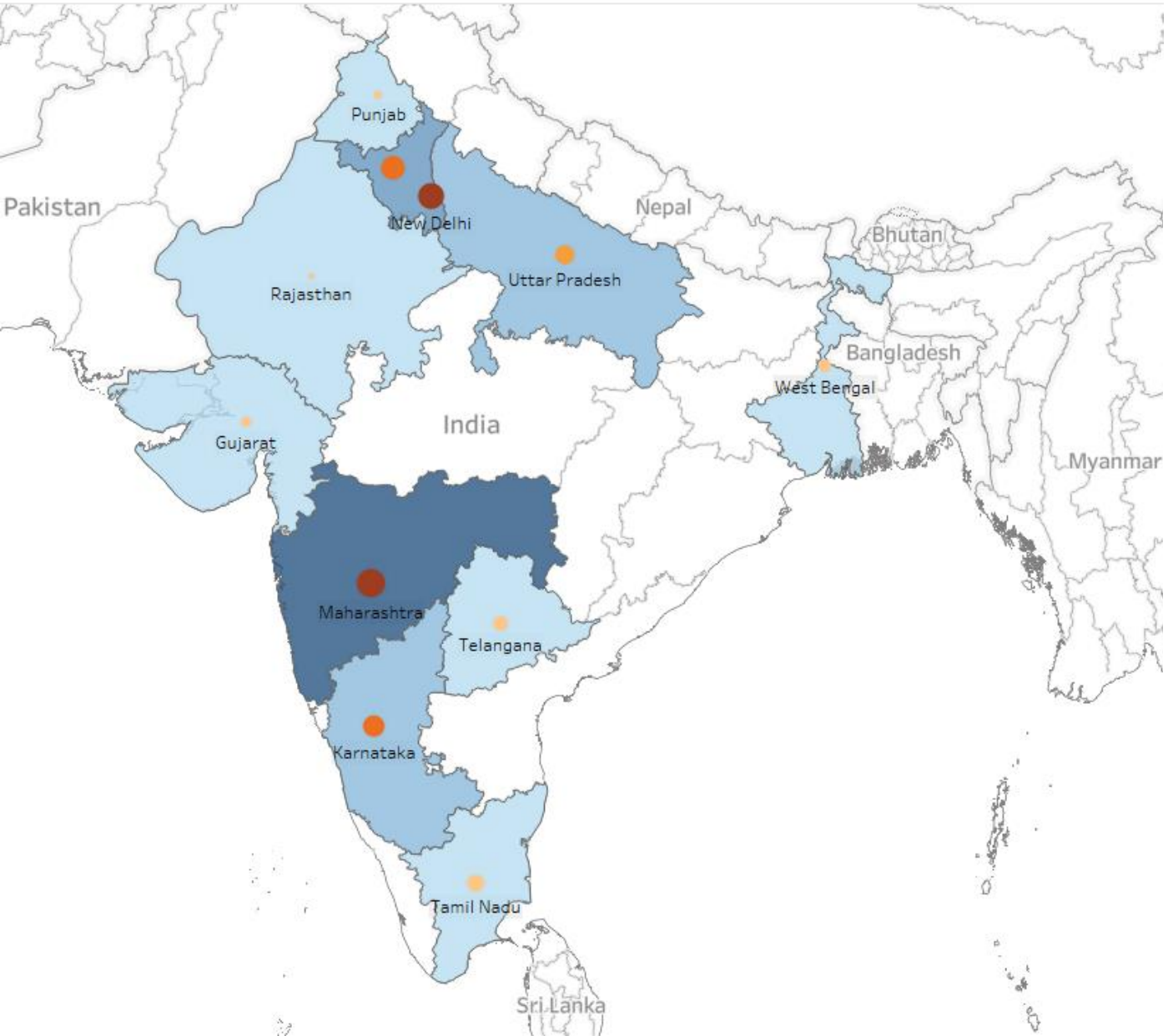


GeoSpatial Analysis

States with high customer balances, high customer spending, and high transaction volumes

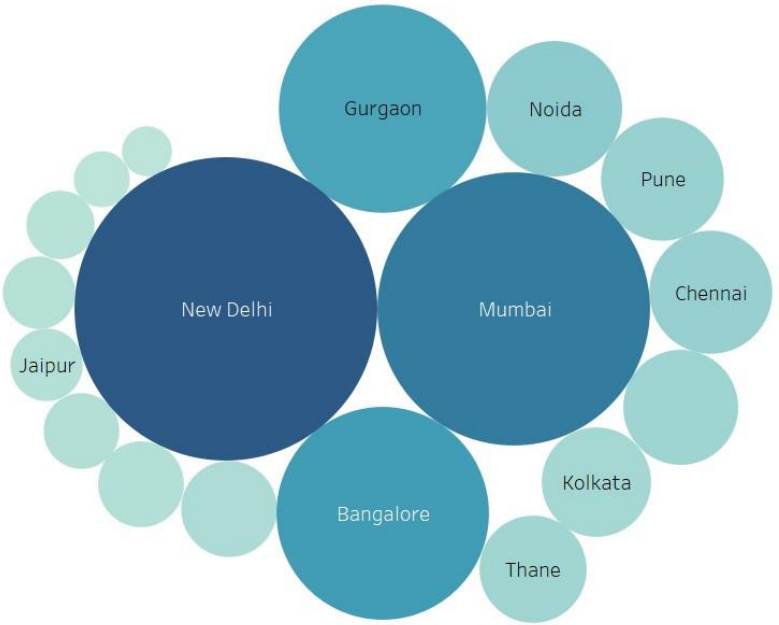
The darker the state,the higher the customer account balances.

The bigger the circle,higher the customer spending and darker the circle, higher the transaction volume.



Cities with high spending customers and high transaction volume

The darker the state,the higher the transaction volume.The bigger the circle,higher the transaction value

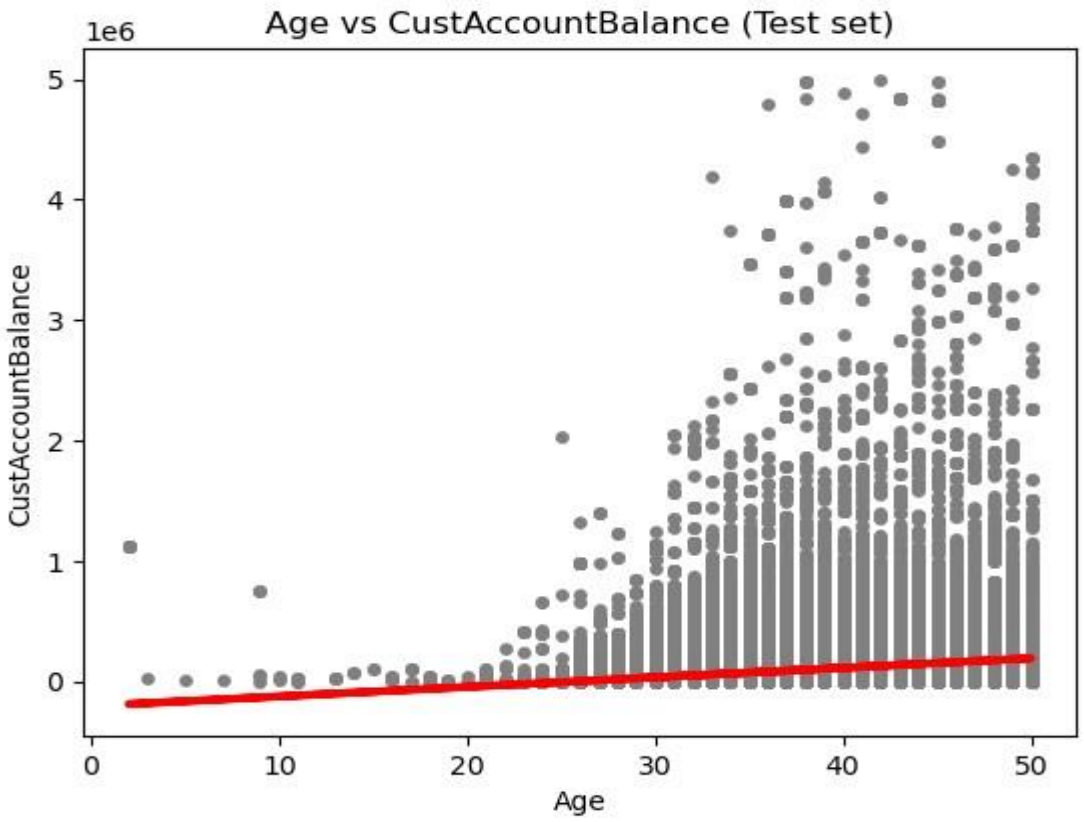


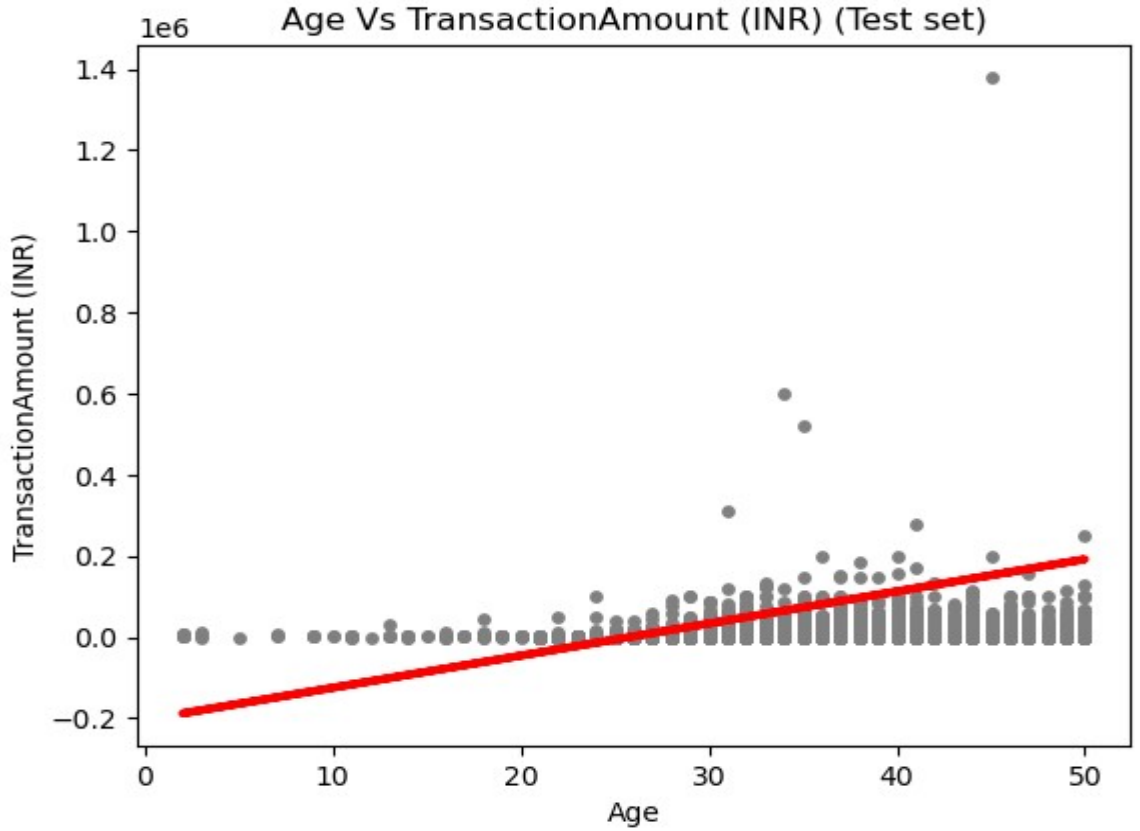
Time Series Analysis

2016 Monthly Transaction Count

The busiest month in 2016 was August, followed by September in terms of transactions.

Regression Analysis- Linear Regression

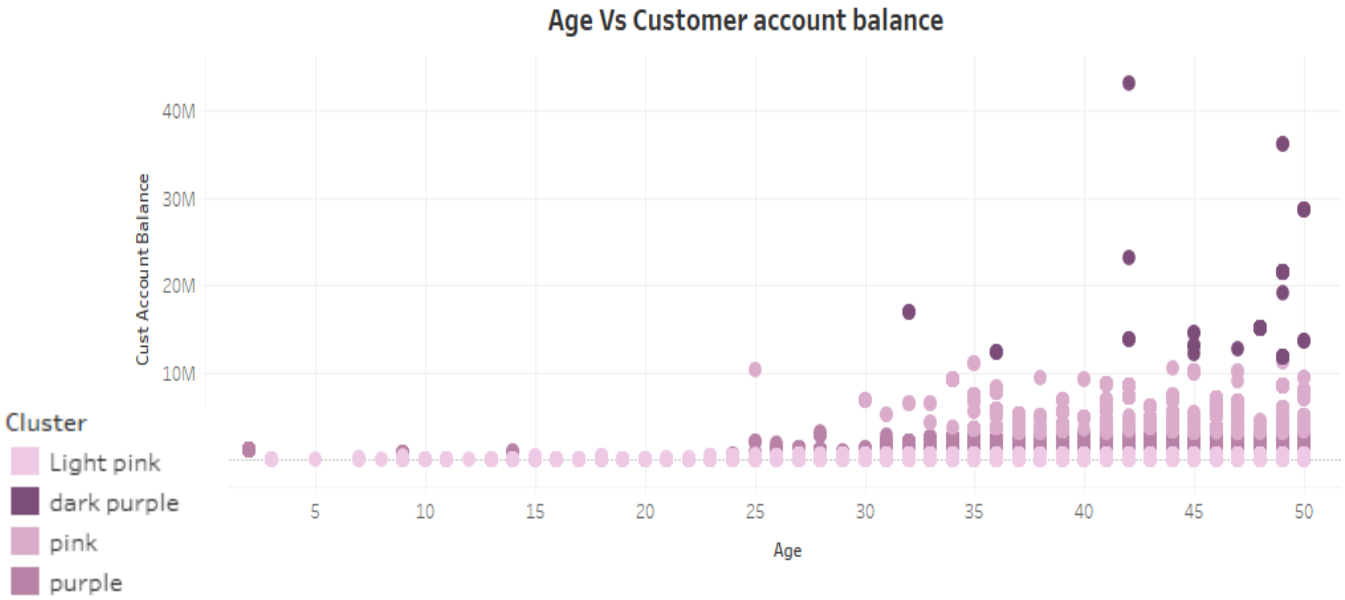


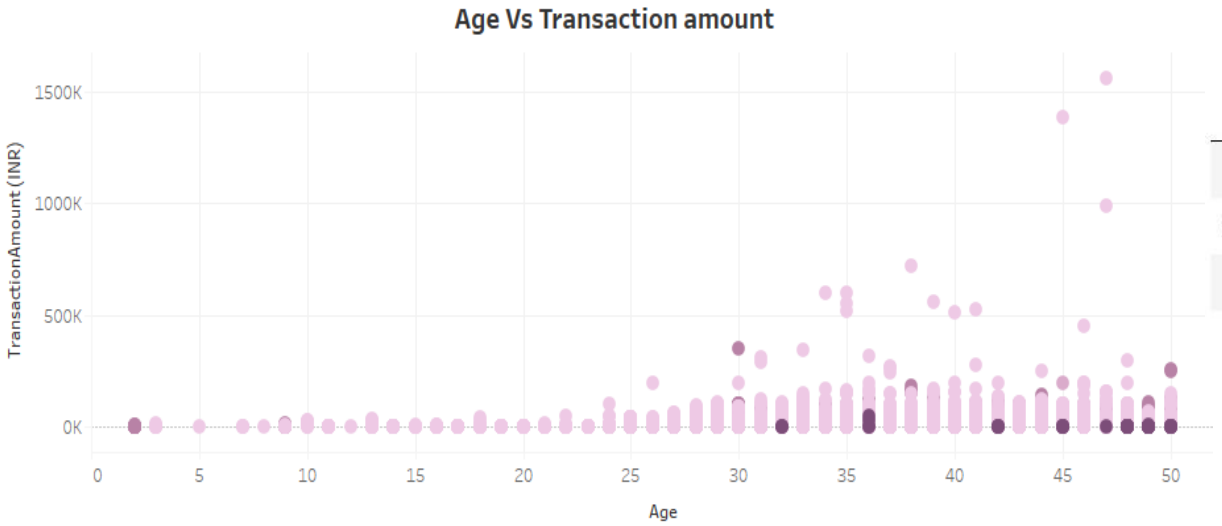


- There is no linear relationship between the variables because most of the data points lie outside of the trend line.
- The R-squared value is always in the range between 0 and 1, with values closer to 1 denoting a regression model that fits the data more closely. Therefore, the extremely low R-squared value of almost zero and the huge Mean Squared Error demonstrate that the Linear regression is not the best model for the dataset.
- Even though the positive slope value indicates a solid association between the variables in this case, the MSE & R2 score value prevents us from utilizing this model to generate predictions.

Clustering Analysis – Kmeans Clustering

Since Linear regression model didn't seem to fit well to the data, the analysis moved on to machine learning using k-means clustering. The optimal number of clusters has been identified to be four using the Elbow technique. After that, a model was created using kmeans clustering and the resulting clusters produced insightful results.





Cluster Analysis Results

- In terms of statistics across all categories, the dark purple cluster outperforms pink, purple, and light pink.
- In comparison to other groups, the dark purple cluster's customers are older, have larger account balances, and do more transactions.
- As a result, it is evident that a customer's account balance and transactional or spending amounts increase with age.

	Age		CustAccountBalance		TransactionAmount (INR)	
	mean	median	mean	median	mean	median
cluster						
Light pink	36.197638	35.0	4.937166e+04	16385.54	1273.485022	435.56
dark purple	46.563380	48.0	1.794515e+07	15158474.07	2997.812676	2000.00
pink	43.042343	44.0	5.184852e+06	4831162.46	4444.372308	1334.00
purple	40.223630	40.0	1.096728e+06	906665.85	3082.458258	1000.00

Insights from Data

The statistics supported my hypothesis, "Customers' account balances and transactional or spending quantities increase as they get older."

Insights

- Male customers are more likely than female customers to spend money, have higher account balances, and make more transactions.
- August is the month with the most transactions and the most expensive transactions, followed by September.
- The age groups 26 to 35 have the most transactions, whereas the age groups 36 to 45 have the most expensive transactions and the highest account balances.
- The state of Maharashtra has the highest customer balances, customer spending, and transaction volumes, followed by New Delhi and Haryana.
- The cities with the highest spending and transaction volume are New Delhi, Mumbai, and Bangalore.

Limitations

- Since the data was provided by a bank as part of a research project in 2016, the analysis won't take into account recent trends.

Next steps

- The most recent bank transaction data will be updated, and further research will be conducted by adding other variables including the type of transaction.

Click below to view detailed Report, Scripts & Visualizations

Let's connect and explore how I can help you achieve your business objectives

sruthysreekanth@gmail.com

https://www.linkedin.com/in/sruthy-sreekanth/

https://github.com/sruthysreekanth

https://public.tableau.com/app/profile/sruthy.sreekanth