

Bank Customer Churn

Why I choose this data set?

I want to do a project in the financial field, so I spent a lot of time searching for this kind of data, but because of ethical issues and task needs, I finally found the Anonymous Multinational bank data set. In future interviews, I would like to simulate real-life scenarios to demonstrate my analytical reasoning.

Data

- **Source of Data:**

One of open source platform – kaggle.com and downloaded from following link
<https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn>

- **Data Collection:**

The data was collected from Governmental collection, because of Ethical issue, we wouldn't know which bank it is.

- **Data Content:**

This data include RowNumber—corresponds to the record (row) number and has no effect on the output. Customer ID, Surname, CreditScore, Geography, Gender, Age, Normally, Balance, NumOfProducts, HasCrCard(has a credit card or not). IsActiveMember, EstimatedSalary, Exited, Complain, Satisfaction Score, Card Type, Points Earned.

- **Data Profile & Cleaning data:**

1.The Data has 18 Columns and 10000 rows.

2.After I cleaned the data , I drop 2 Columns(Surname and RowNumber) because there is no impact on their decision to leave the bank. The Data has 16 Columns and 10000 rows.

3. There are no duplicates Values and missing values.

4.I rename the column 'CustomerId' to 'CustomerID'

5. Change data type: Card Type as a type of category, HasCrCard as a type of category, IsActiveMember as a type of category, Complain as a type of category, Exited as a type of category.

Columns	Description	Data type	Discrete / Continuous
RowNumber	corresponds to the record (row) number and has no effect on the output(I drop it)	Qualitative	Discrete
CustomerId	contains random values and has no effect on customer leaving the bank.	Qualitative	Discrete
Surname	The surname of a customer has no impact on their decision to leave the bank. (I drop it)	Qualitative	Discrete
CreditScore	A customer with a higher credit score is less likely to leave the bank.	Quantitative	Discrete
Geography	A customer's location	Qualitative	Discrete
Gender	It's interesting to explore whether gender plays a role in a customer leaving the bank.	Categorical	Discrete

Age	The customers age.	Quantitative	Discrete
Tenure	Refers to the number of years that the customer has been a client of the bank.	Quantitative	Discrete
Balance	How much money did they use.	Quantitative	Continuous
NumOfProducts	Refers to the number of products that a customer has purchased through the bank	Quantitative	Discrete
HasCrCard	Whether or not a customer has a credit card.	Categorical	Discrete
IsActiveMember	Active customers are less likely to leave the bank.	Categorical	Discrete
EstimatedSalary	Customers' salaries.	Quantitative	Continuous
Exited	Whether or not the customer left the bank.	Categorical	Discrete
Complain	Customer has complaint or not.	Categorical	Discrete
Satisfaction Score	Score provided by the customer for their complaint resolution.	Quantitative	Discrete
Card Type	Type of Debit Card hold by the customer.	Categorical	Discrete
Point Earned	The points earned by the customer for using credit card.	Quantitative	Discrete

4. limitations and ethics

Due to confidentiality concerns, the username needs to be removed first.

5. Questions to explore:

- What is the age group that predominantly utilizes our bank service?
- What is the Gender that predominantly utilizes our bank?
- Where is the customer's location that predominantly utilizes our bank service?
- Does the balance or salary of customers have an impact on the performance of our bank?
- How did customer complaints and satisfaction scores affect the performance of our bank?
- What kind of Debit Card in each age group like to use?
- How could we find more information from this data set to increase the number of customers in the future?