Assessment cover sheet

Your name: Jeevan Kokkadan Johny

Mentem by UNSW

Instructions:

1. Include this cover sheet as the first page of your submission.

2. Convert your document(s) to PDF before uploading.

3. Submit this PDF along with any other deliverables specified in the assessment brief.

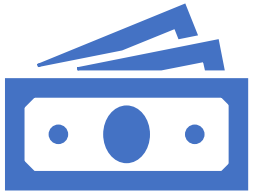
Addressing High Customer Churn Rates

Business Problem Statement:

The bank is facing a significant challenge with high customer churn rates, which is adversely affecting its revenue and growth. To tackle this issue, it is crucial to identify the key factors contributing to customer churn by analysing historical customer data. Understanding these factors will enable the bank to develop targeted strategies to retain customers and reduce churn rates, ultimately improving customer satisfaction and loyalty

Success Measures for Reducing Customer Churn

Success Measures:



Churn Rate Reduction

Explanation: The churn rate is the percentage of customers who leave the bank over a specific period. Reducing the churn rate is a primary indicator of the success of retention strategies.

Business Use Case:

- **Monitoring Customer Retention**: Track the effectiveness of retention strategies and identify areas for improvement.
- **Revenue Impact**: Retaining existing customers is more cost-effective than acquiring new ones, directly impacting revenue.

Improve Customer Satisfaction Score

Explanation: The Customer Satisfaction Score (CSAT) measures the level of satisfaction customers have with the bank's products and services.

Business Use Case:

- **Identifying Pain Points**: Analyse CSAT scores to identify areas of dissatisfaction and take corrective actions.
- **Improving Customer Experience**: Higher CSAT scores indicate happy customers, leading to increased loyalty and reduced churn.

Basic Statistical Measures

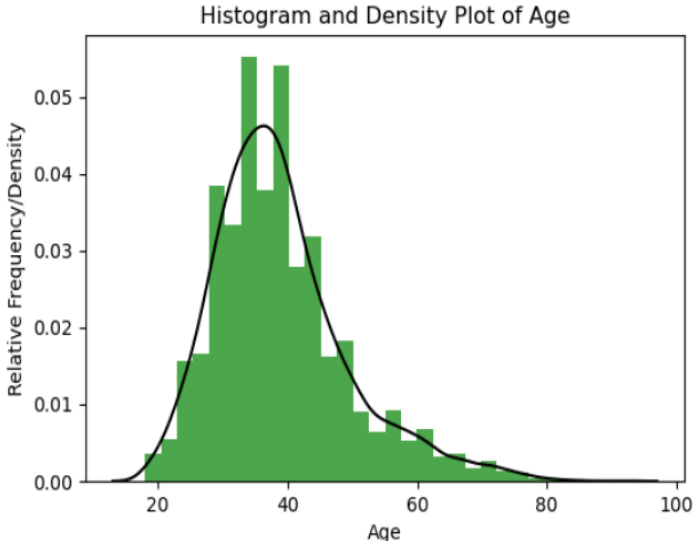
Attribute	Mean	Median	Mode	25th Percentile	50th Percentile (Median)	75th Percentile
CreditScore	650.53	652	850	584	652	718
Age	38.92	37	37	32	37	44
Tenure	5.01	5	2	3	5	7
Balance	$76,485.89	$97,198.54	$0.00	$0.00	$97,198.54	$127,644.20
NumOfProducts	1.53	1	1	1	1	2
HasCrCard	0.71	1	1	0	1	1
IsActiveMember	0.52	1	1	0	1	1
EstimatedSalary	$100,090.20	$100,193.94	$100,090.20	$39,620.41	$100,193.94	$149,388.24
Exited	0.2	0	0	0	0	0
Complain	0.2	0	0	0	0	0
Satisfaction Score	3.01	3	3	2	3	4
Point Earned	606.52	605	605	402	605	801

- **Mean**: The average value of each attribute, providing a central tendency of the data.
- By comparing the mean values of attributes like Credit Score, Age, Balance, and Satisfaction Score between churned and non-churned customers, the bank can identify which attributes are associated with higher churn rate
- **Median**: The middle value of each attribute when the data is sorted, indicating the central point of the data distribution.
- The median helps identify the typical value of each attribute, reducing the impact of outliers. By comparing the median values of attributes between churned and non-churned customers
- **Mode**: The most frequently occurring value in each attribute, showing the most common value in the dataset.
- The mode helps identify the most common characteristics of churned customers. For example, if the mode for the number of products is 1 for churned customers, it indicates that customers with only one product are more likely to churn
- **Percentiles**:
    - **25th Percentile**: The value below which 25% of the data falls, indicating the lower quartile.
    - **75th Percentile**: The value below which 75% of the data falls, indicating the upper quartile.

  Percentiles help identify the spread and variability of each attribute. By analyzing the percentiles of attributes like Credit Score, Age, Balance, and Satisfaction Score, the bank can identify patterns and trends that contribute to churn. For example, if the 25th percentile of satisfaction scores is significantly lower for churned customers, it indicates that a substantial portion of churned customers are dissatisfied

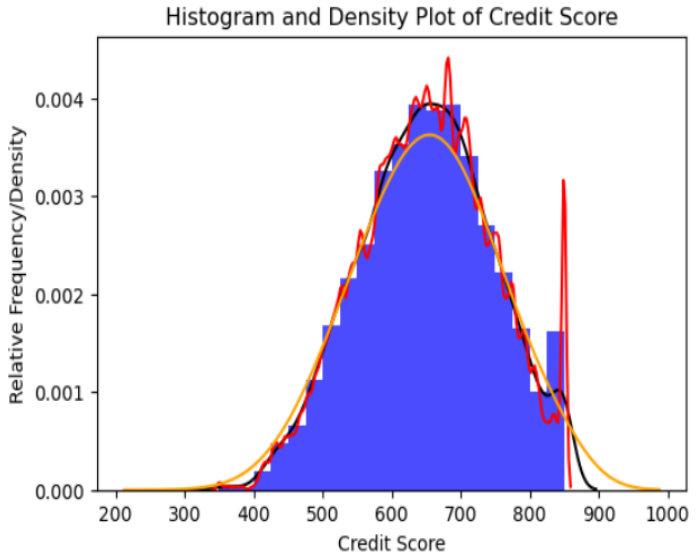
These statistical measures provide a comprehensive overview of the customer churn dataset, highlighting key attributes and their distributions. This information can be used to identify patterns and trends that contribute to customer churn, helping the bank develop targeted strategies to retain customers and reduce churn rates.

Key Data Visualizations for Customer Churn Analysis

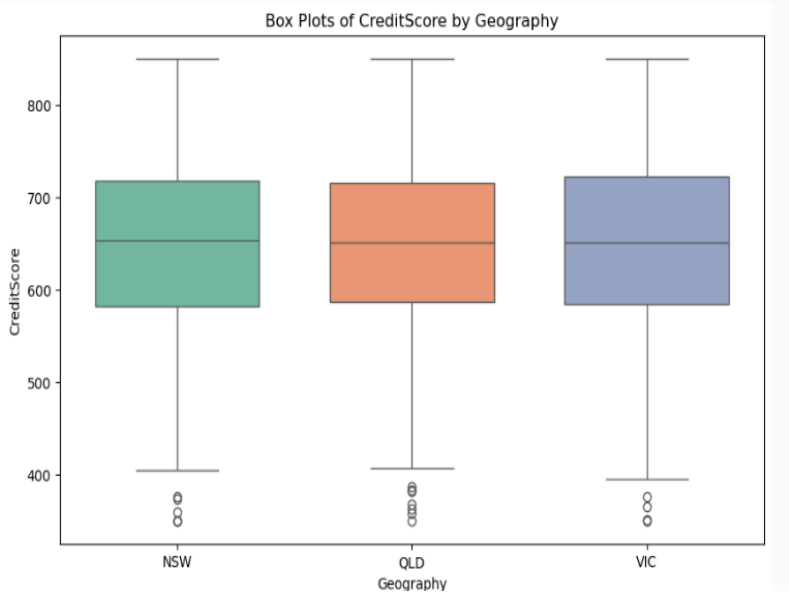


- **Central Tendency**: The histogram and density plot indicate that the majority of customers are concentrated around the mean age. The peak of the density plot represents the most common age range among customers.
- **Spread**: The spread of the histogram and the width of the density plot show the variability in customers' ages. The distribution appears to be relatively symmetric, with a slight skew towards older ages

Understanding the age distribution of customers helps the bank identify age groups that may require targeted retention strategies



- **Central Tendency**: The histogram and density plots indicate that the majority of customers have credit scores concentrated around the mean. The peak of the density plots represents the most common credit score range among customers.
- **Spread**: The spread of the histogram and the width of the density plots show the variability in customers' credit scores. The distribution appears to be relatively symmetric, with a slight skew towards lower credit scores.
- **Bandwidth Adjustments**: The different bandwidth adjustments (black, red, and orange lines) provide varying levels of smoothing for the density plots



This box plot visualizes the distribution of customers' credit scores across different geographical regions (NSW, QLD, VIC). Each box represents the interquartile range (IQR) of credit scores for a specific region, with the line inside the box indicating the median credit score. The whiskers extend to the minimum and maximum values within 1.5 times the IQR, and any points outside this range are considered outliers.

By analysing the credit score distribution in conjunction with churn rates, the bank can determine if certain credit score ranges are more likely to churn and develop targeted interventions to improve retention

Correlation Matrix Heatmap - Key Points

- This heatmap visualizes the correlation matrix for key attributes such as Credit Score, Age, Balance, Estimated Salary, Point Earned, Complain, Satisfaction Score, and Exited (churn). The correlation values range from -1 to 1, where values close to 1 indicate a strong positive correlation, values close to -1 indicate a strong negative correlation, and values close to 0 indicate no correlation. Key findings from the heatmap include

- **Credit Score and Exited**: There is a negative correlation between Credit Score and Exited, indicating that customers with lower credit scores are more likely to churn.

- **Balance and Exited**: There is a positive correlation between Balance and Exited, suggesting that customers with higher balances are more likely to churn.

- **Satisfaction Score and Exited**: There is a strong negative correlation between Satisfaction Score and Exited, indicating that customers with lower satisfaction scores are more likely to churn.

- **Complain and Exited**: There is a positive correlation between Complain and Exited, suggesting that customers who have raised complaints are more likely to churn.

Identifying and Addressing Data Quality Issues

1. Missing Values:

- **Identification**: No missing values were found in the dataset. This means that every cell in the Data Frame has a valid entry, and there are no NaN (Not a Number) or null values present in any of the columns.

2. Duplicate Records:

- **Identification**: There are no duplicate rows in the dataset. This means that each row in the Data Frame is unique, and there are no repeated entries.

3. Inconsistent Data Formats:

- **Identification**: The data looks clean and consistent for the "Geography," "Card Type," and "Gender" columns.

4. Outliers:

- **Identification**: Used box plots to identify outliers in key numerical columns. Outliers were found for the "Age" and "Credit Score" columns.
- **Action Taken**: Handled outliers by capping them at the 1st and 99th percentiles to reduce their impact on the analysis.
- **Result**: Outliers were effectively managed, ensuring a more accurate analysis.

Methodology for Handling Data Quality Issues

1. **Handling Missing/Duplicate Data**:
    1. **Method**: Dropped rows with missing values.
    2. **Justification**: Ensures data integrity and prevents potential biases in the analysis.
2. **Handling Outliers**:
    1. **Method**: Capped outliers at the 1st and 99th percentiles.
    2. **Justification**: Reduces the impact of extreme values on the analysis, leading to more accurate and reliable results.

Formulating and Evaluating Hypotheses

- **Null Hypothesis (H0)**: There is no significant difference in churn rates between customers with higher balances and those with lower balances.
- **Alternate Hypothesis (H1)**: Customers with higher balances are more likely to churn compared to those with lower balances.
- **Business Objective Alignment**: Understanding the relationship between account balances and churn rates, the bank can develop targeted strategies to retain high-balance customers and reduce churn rates.
- **Methodology:**

1. **Define Balance Threshold**: The median balance is used as the threshold to separate customers with lower and higher balances.
2. **Separate Data**: The data is divided into two groups: customers with lower balances and customers with higher balances.
3. **Perform T-Test**: A t-test is conducted on the 'Exited' column to test the hypothesis.

Results:

- **T-statistic**: -11.49
- **P-value**: 2.21e-30
- **Interpretation**: The p-value is less than 0.05, indicating that we reject the null hypothesis. This means that customers with higher balances are significantly more likely to churn.
- **Churn Rates:**
- **Lower Balance Group**: 15.78%
- **Higher Balance Group**: 24.98%

- **Null Hypothesis (H0)**: There is no significant difference in churn rates between customers who have raised complaints and those who have not.
- **Alternate Hypothesis (H1)**: Customers who have raised complaints are more likely to churn compared to those who have not raised complaints.
- **Business Objective Alignment :** understanding the relationship between complaints and churn rates, the bank can develop targeted strategies to address customer issues and improve retention.
- **Methodology:**

1. **Separate Data**: The data is divided into two groups based on complaint status: customers who have raised complaints and customers who have not.
2. **Perform T-Test**: An independent two-sample t-test is conducted on the 'Exited' column to test the hypothesis.

Results:

- **T-statistic**: 1073.80
- **P-value**: 0.0
- **Interpretation**: The p-value is less than 0.05, indicating that we reject the null hypothesis. This means that customers who have raised complaints are significantly more likely to churn.
- **Churn Rates:**
- **Complained Customers**: 99.51%
- **Not Complained Customers**: 0.05%

Designing a Controlled Experiment to Reduce Customer Churn

- **Controlled Experiment: Customers Who Have Raised Complaints are More Likely to Churn**
- **Objective:** To test the hypothesis that customers who have raised complaints are more likely to churn and to develop strategies to reduce churn rates among these customers.
- **Hypothesis:** Customers who have raised complaints are more likely to churn.
- **Experiment Design:**

1. **Control Group:**
    1. **Description**: Customers who have not raised complaints.
    2. **Size**: 7,956 customers.
2. **Treatment Group:**
    1. **Description**: Customers who have raised complaints.
    2. **Size**: 2,044 customers.
3. **Duration:**
    1. **Length**: 6 months.
4. **Relevant Metrics:**
    1. **Churn Rate**: The percentage of customers who leave the bank.
    2. **Customer Satisfaction Score**: Measured through surveys to assess customer satisfaction levels.
    3. **Number of Resolved Complaints**: Tracked through customer service records to measure the effectiveness of complaint resolution.

By comparing the churn rates, satisfaction scores, and resolved complaints between the control group and the treatment group, the bank can evaluate the effectiveness of strategies implemented to address customer complaints and reduce churn rates