

Study customer churn analytics for the banking industry

Customer Churn:-

Customer churn, also known as customer attrition or customer turnover, refers to the rate at which customers stop doing business with a company during a certain period. It is a critical metric for businesses, particularly in subscription-based or recurring revenue models, as it indicates the loss of customers who were previously generating revenue for the company. Churn can result from various factors, such as dissatisfaction with the product or service, competitive alternatives, or changes in the customer's circumstances.

As per the Requirement of the Banking Company Summary Statistics of the Credit Score and Balance of Customers:-

Report on Credit Score Analysis:-

Introduction:

Credit Score:- A credit score in the banking industry is a numerical representation of a person's creditworthiness. It is a crucial factor that lenders, such as banks and financial institutions, use to assess the risk associated with lending money to an individual or business.

In this report, we analyze the credit score data of customers in the "Churn Modelling" dataset. Our primary focus is on key statistical measures, including the average credit score, median credit score, first quartile (Q1), third quartile (Q3), and the range between the minimum and maximum credit scores. These statistics provide valuable insights into the distribution of credit scores within the dataset, helping us understand potential outliers and the overall spread of credit scores.

	average_creditscore numeric	median_creditscore double precision	q1_creditscore double precision	q3_creditscore double precision	max_creditscore integer	min_creditscore integer
1	650.5288000000000000	652	584	718	850	350

Key Findings:

Average Credit Score: The average credit score in the dataset is approximately 650.53. This metric represents the central tendency of credit scores and serves as a baseline for understanding the typical creditworthiness of customers in the dataset.

Median Credit Score: The median credit score, which is 652, indicates the middle point of the credit score distribution when the scores are sorted in ascending order. It's noteworthy that the median is quite close to the average, suggesting a relatively symmetric distribution.

Q1 and Q3 Credit Scores: The first quartile (Q1) is at 584, and the third quartile (Q3) is at 718. These quartiles divide the data into four equal parts. Q1 represents the 25th percentile of credit scores, and Q3 represents the 75th percentile. The spread between Q1 and Q3 is an essential indicator of data dispersion. In this case, it spans from 584 to 718, demonstrating a relatively wide interquartile range.

Range of Credit Scores: The maximum credit score in the dataset is 850, and the minimum credit score is 350. This range reflects the full spectrum of credit scores within the dataset. The substantial range suggests the presence of both high and low credit scores, contributing to the overall variability in the data.

Interpretation:

The analysis of credit scores in the "Churn Modelling" dataset reveals several important insights:

Symmetric Distribution: The close proximity of the mean and median (650.53 and 652, respectively) suggests that the distribution of credit scores is reasonably symmetric. This indicates that there may be a balanced number of customers with higher and lower credit scores, leading to a relatively even spread.

Outliers: The presence of extremely low and high credit scores, as evidenced by the minimum (350) and maximum (850) values, indicates the existence of outliers. Outliers can have a significant impact on the mean and may lead to its divergence from the median. In this case, the outlier on the lower end (350) likely contributes to a slightly lower mean compared to the median.

Dispersion: The interquartile range, spanning from 584 to 718, is relatively wide. This suggests that a significant portion of customers falls within this range, but there are also outliers with credit scores well below and above this range. The wide dispersion indicates that credit scores in the dataset vary significantly.

Conclusion:

The analysis of credit scores in the "Churn Modelling" dataset reveals a relatively symmetric distribution with some noticeable outliers on both ends. The presence of these outliers contributes to a wider interquartile range and a slightly lower mean compared to the median. These findings emphasize the importance of understanding and addressing outliers in credit score analysis. Additionally, the wide dispersion of credit scores highlights the diversity of creditworthiness among customers in the dataset.

Understanding these credit score metrics is crucial for making informed decisions, particularly in the banking industry, as it helps in risk assessment, customer profiling, and credit evaluation. It is recommended to further investigate the outliers and their potential impact on the data analysis to gain a comprehensive understanding of the credit score distribution.

Report on Balance Analysis:-

Introduction:

In this report, we analyze the balance data of customers in the "Churn Modelling" dataset. Our primary focus is on key statistical measures, including the average

balance, median balance, first quartile (Q1), third quartile (Q3), and the range between the minimum and maximum balances. These statistics provide valuable insights into the distribution of balances within the dataset, helping us understand potential outliers and the overall spread of account balances.

	average_balance numeric	median_balance double precision	q1_balance double precision	q3_balance double precision	max_balance numeric	min_balance numeric
1	76485.889288000000	97198.540000000001	0	127644.24	250898.09	0

Key Findings:

Average Balance: The average balance in the dataset is approximately \$76,485.89. This metric represents the central tendency of account balances and serves as a baseline for understanding the typical account balance of customers in the dataset.

Median Balance: The median balance, which is approximately \$97,198.54, indicates the middle point of the balance distribution when the balances are sorted in ascending order. Notably, the median is significantly higher than the average, suggesting a potential asymmetry in the distribution.

Q1 and Q3 Balances: The first quartile (Q1) balance is \$0, and the third quartile (Q3) balance is approximately \$127,644.24. These quartiles divide the data into four equal parts. Q1 represents the 25th percentile of balances, and Q3 represents the 75th percentile. The wide spread between Q1 and Q3 is indicative of a substantial variability in account balances.

Range of Balances: The maximum account balance in the dataset is \$250,898.09, while the minimum balance is \$0. The range from \$0 to \$250,898.09 encompasses the entire spectrum of account balances within the dataset. The presence of both zero balances and high balances highlights the diversity of financial positions among customers.

Interpretation:

The analysis of account balances in the "Churn Modelling" dataset reveals several important insights:

Asymmetric Distribution: The substantial difference between the mean and median (average balance and median balance, respectively) suggests that the distribution of account balances is likely skewed. In this case, the median is significantly higher than the mean, indicating that there may be a relatively small number of customers with very high balances that pull the median up.

Outliers: The presence of customers with zero account balances and high balances, as indicated by the minimum (\$0) and maximum (\$250,898.09) values, points to the existence of outliers. Outliers can significantly affect the mean and contribute to the observed asymmetry in the distribution.

Dispersion: The wide spread between Q1 and Q3 balances, spanning from \$0 to \$127,644.24, suggests a substantial variation in account balances. This dispersion underscores the diversity of financial situations within the dataset.

Conclusion:

The analysis of account balances in the "Churn Modelling" dataset indicates an asymmetric distribution with the potential presence of outliers. The median balance is notably higher than the average balance, suggesting that a minority of customers may have significantly higher balances, contributing to the observed skewness in the distribution.

Understanding the distribution of account balances is crucial for making informed financial decisions and risk assessments, particularly in the banking industry. Further investigation into the outliers and their impact on data analysis is recommended to gain a comprehensive understanding of the balance distribution. Additionally, strategies for managing and addressing accounts with low or zero balances may be necessary to optimize financial services and customer relationships.

Best Parameters to Understand and Analyse Customer Churn:-

To better understand and analyze customer churn in the "Churn Modelling" dataset, several key metrics should be considered. These metrics are crucial for understanding and analyzing customer churn in this dataset for the following reasons:

Churn Rate: This is the central metric to monitor and understand the extent of churn. It directly quantifies the percentage of customers who are leaving, which is the primary focus of churn analysis.

Customer Tenure: Understanding the tenure of customers can help distinguish between new and long-standing customers. New customers may have different churn patterns compared to loyal, long-term customers.

Product Usage: The number of products used by a customer indicates their engagement level. Analyzing this metric can reveal the relationship between product diversity and churn.

Active Membership: Engaged customers are less likely to churn. Monitoring active membership helps in assessing the importance of customer engagement for retention.

Credit Score and Balance: Financial stability is often associated with churn behavior. A lower credit score or a certain balance level may be correlated with higher churn rates, making these metrics crucial for financial institutions.

Geography and Gender: Demographic factors can provide valuable segmentation for churn analysis. Differences in churn rates across regions or between genders may reveal opportunities for targeted retention strategies.

Incorporating these metrics into the analysis will provide a comprehensive view of churn patterns and enable the development of data-driven strategies to reduce churn and improve customer retention.