**Project Review: Customer Churn in Banking – Exploratory Data Analysis**

Ramesh Talapaneni

Bellevue University

DSC 530 – Data Exploration and Analysis

March 01, 2024

**Introduction**

Based on the Exploratory Data Analysis (EDA) we conducted on the customer churn dataset for a bank, we can summarize the findings, reflect on potential omissions, and discuss the challenges faced during the analysis.

**Outcome of EDA**

The EDA revealed several key insights into factors influencing a customer's decision to leave the bank. Notably, age and balance showed a significant relationship with customer churn, with younger customers and those with lower balances more likely to leave. Gender also appeared to influence churn, albeit to a lesser extent. Active membership and credit card ownership were inversely related to churn, suggesting that engaged customers with more ties to the bank are less likely to churn. The number of products held by a customer and their tenure with the bank were also explored, but their relationship with churn was less clear from initial observations. Hypothesis testing confirmed some of these relationships, while the impact of credit score and geography required further investigation due to mixed results.

**Missed Aspects**

One limitation of the EDA was the lack of deeper temporal analysis, which could have provided insights into churn trends over time and potentially identified seasonal patterns or the impact of specific events. Additionally, the interaction effects between variables were not fully explored, which might mask important insights, for instance, the combined effect of age and balance on churn.

**Potential Additional Variables**

One limitation of the EDA was the lack of deeper temporal analysis, which could have provided insights into churn trends over time and potentially identified seasonal patterns or the impact of specific events. Additionally, the interaction effects between variables were not fully explored, which might mask important insights, for instance, the combined effect of age and balance on churn.

**Assumptions and Incorrect References**

Some assumptions made during the analysis may have been oversimplifications. For example, the assumption that higher credit scores always correlate with lower churn may not hold for all customer segments. Also, categorizing age into broad groups might have obscured more specific age-related patterns in churn.

**Challenges and Understanding**

Challenges encountered during the analysis included handling missing data and ensuring that data distribution did not bias the results. The assumption that the dataset was representative of the broader customer base might not be accurate, which could affect the generalizability of the findings. Additionally, the causal relationships could not be definitively established due to the nature of observational data; further analysis, possibly through a longitudinal study or randomized controlled trials, would be required to draw firmer conclusions.

**Conclusion**

In conclusion, while the EDA provided valuable insights, it also highlighted the complexity of customer churn as a phenomenon influenced by multiple, often interrelated factors. Further analysis, enriched by additional data and refined methodologies, would be necessary to develop a more comprehensive understanding and to inform the creation of targeted retention strategies.

**References**

Kollipara, R. ([n.d.]). Bank customer churn. [Dataset]. Kaggle. <https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn/data>

Marconato, R. (2023, October 26). Customer churn EDA, balancing, and ML (83.63) [Computer code]. Kaggle. <https://www.kaggle.com/code/raphaelmarconato/customer-churn-eda-balancing-and-ml-83-63>