Summary of Insights on New Credit Card Product Opportunities

This is a summary of Insights on New Credit Card Product Opportunities, a report detailing analysis and recommendations for an anonymous bank based on 10,127 credit card user records.

Data source: https://www.kaggle.com/sakshigoyal7/credit-card-customers.

Data Source

The data used in this report are all sourced from this Kaggle dataset called Credit Card customers.[1] According to the Kaggle data description, this data was initially sourced from Leaps (https://leaps.analyttica.com/home). Leaps is an online learning platform and directory for data science. In this scenario, the data is from an anonymous bank containing anonymous credit card customer data points.

You can access the data from this link: https://www.kaggle.com/sakshigoyal7/credit-card-customers.

As businesses move to adopt data science into decision making, it is critical to leverage the data available to businesses and draw insights that can lead to real action and insights for business leaders.

It was our goal to select a dataset that reflected the type of information a Data Scientist working for a financial institution would be asked to use to draw insights or create a model. This dataset also matched our group's interest to work on some different types of models, including classifiers and clustering.

Analysis

The report broke down the analysis into three distinct sections: preliminary analysis, logistic regression, and clustering. The objective of the report was to create suggestions for the anonymous bank on new credit card product opportunities and insights on their current users.

---

[1] Goyal, S. (2020, November 19). Credit card customers. Retrieved April 11, 2021, from https://www.kaggle.com/sakshigoyal7/credit-card-customers

Each section approached the problem from slightly different ways. The preliminary analysis focused on the details of demographics on different variables, including attrition but also the distribution of the dataset across credit card products (Blue, Silver, Gold or Platinum). Interestingly, key demographics were identified, specifically as users in the age group 46-55 with Blue Cards (lowest credit card tier).

The logistic regression model served to predict churn of credit card customers (whether the attrition flag was true or not for a record). It was successful with an accuracy of 84% and a recall of 81%. This analysis also served to highlight the significant relationship between Platinum Cards (highest tier) and attrition.

The clustering model produced five clusters to group credit card users and provided useful insights onto each cluster, including detailed credit card distributions across the cluster, attrition, spending habits, and etc. The profiles were High Volume, Low Credit Limit, High Credit Limit, Pay It Off, and Low Volume.

The data was very detailed with a wide array of features in a space that some group members lacked some domain knowledge in. It was a challenge to determine relevant attributes and which ones to focus on, often causing project scope creep as the lens of our analysis kept growing. Through discussion and research, we were able to narrow down the scope of the project into a more reasonable focus.

Conclusions

- **Blue Card (lowest credit card tier) customers can be diversified into more targeted products.** These cards are prevalent in all user profiles, as well as across genders. These Blue Card users are generally high transaction users as well. The potential exists for this credit card user segment to be further targeted and more niche products used to better augment user experience and bank revenues. Niche targeting to certain demographics may help in designing these products, including elderly men and women with Blue Cards who have more transactions than other segments.
- **Opportunity lies in focusing on non-traditional user segments.** Users with current low credit limits are critical to DataBanks success as they hold low attrition and are mostly on the basic product (Blue Cards). This non-traditional user segment should be targeted for new products and marketing.
- **High-end segments and customers are exiting.** There is concerns on the attrition for high-end users with Platinum Cards, users with Doctorate degrees, and users with greater than $120,000 in annual income. Machine learning models, such as the logistic regression model used in this paper, should be developed and deployed to create early warning flags to address attrition and deploy solutions or support.