Individual Report (Group 5)

Project Title: Demographic Impact on Credit Card Usage
Name: Sparshika Ajmaan Dinesh Kumar (2180247)
Date: 11/14/2024

---

1. Responsibilities:

In this project, my primary responsibilities included:

- **Multiple Linear Regression and Lasso Regression Models**: I developed and implemented both models to analyze the influence of demographic, behavioral, and transaction factors on credit card usage and retention. These models aimed to predict credit limits and identify variables with strong predictive power.
- **Exploratory Data Analysis (EDA)**: I worked on visualizing key data trends and distributions, analyzing variables related to transaction amount, transaction count, credit utilization, and demographics. Through EDA, we uncovered insights that shaped our modeling approach.

These tasks required a thorough understanding of data preprocessing, model tuning, and regression techniques.

2. Detailed Analysis:

- **Multiple Linear Regression and Lasso Regression**: Multiple linear regression was chosen to understand the linear relationship between demographic variables and credit card usage patterns. The model highlighted key demographic variables affecting credit card retention. Following this, I implemented Lasso regression to regularize and refine our predictor set by penalizing less important variables, helping improve model accuracy by minimizing overfitting.
- **Exploratory Data Analysis (EDA) Plot**: The EDA plot provided a comprehensive view of the data distribution, helping us identify patterns and trends among key variables. For instance, variables like transaction amount and transaction count showed skewed distributions, prompting a logarithmic transformation to normalize the data. This EDA plot also revealed significant outliers in certain variables, such as transaction frequency and credit utilization, which we addressed before building models. By visualizing correlations and distributions, the EDA plot served as a foundational tool for selecting meaningful predictors for our models.

3. Citations:

- Source for dataset: Internal bank database (cited as "Bank Churners" dataset).
- Various data visualization and analysis packages in R, including corrplot for correlation analysis and glmnet for Lasso regression.

4. Personal Reflection:

Reflecting on this project, I gained a practical understanding of how demographic and behavioral data drive customer retention strategies in financial institutions. Working with multiple regression models revealed the impact of factors like income level, transaction history, and credit utilization. Implementing Lasso regression, in particular, was a valuable experience, as it clarified the role of regularization in managing variable importance and overfitting, which are crucial for predictive modeling.

The group's collaborative approach made this project insightful. Each team member brought a unique perspective, allowing us to analyze the dataset thoroughly. One challenge we faced was balancing model complexity with interpretability, especially as we moved toward interaction and polynomial models. This experience highlighted the importance of choosing models that not only perform well but are also understandable for stakeholders.

If improvements were to be made, I would suggest enhancing the model's interpretability by exploring interaction terms further, as this could capture more nuanced relationships between demographic and behavioral factors. Additionally, feedback on the report and presentation format could improve clarity. A structured rubric could help guide students on report expectations, ensuring consistent standards across individual submissions.

Conclusion
Overall, this project has been a rewarding experience, both in terms of applying technical skills and understanding the practical implications of demographic analysis in credit card usage. By analyzing customer retention drivers, I've gained a deeper appreciation for how financial institutions can leverage data to make informed, customer-centric decisions. This project underscored the importance of thorough data analysis in real-world applications, and I look forward to applying these learnings to future endeavors.