



Visualization of Complex Data

DATS 6401

Final Term Project- Proposal

Name : Aman Jaglan
GWID: G45030269

Introduction:

For my final term project, I propose utilizing the American Express dataset, which comprises over 5 million rows and 190 columns. American Express stands as one of the world's leading payment card issuers. This challenge entails employing time-series behavioral data and anonymized customer profiles to construct a model that outperforms current risk management strategies. The objective is to refine lending decisions, elevate customer experiences, and bolster business economics. Successful models have the potential to transform credit default prediction, offering significant rewards and possibly leading to career opportunities with American Express. This contribution aims at making credit card approvals both safer and more accessible.

Objective:

The objective of this project is to predict the probability that a customer does not pay back their credit card balance amount in the future based on their monthly customer profile. The target binary variable is calculated by observing 18 months performance window after the latest credit card statement, and if the customer does not pay due amount in 120 days after their latest statement date it is considered a default event.

The dataset contains aggregated profile features for each customer at each statement date. Features are anonymized and normalized, and fall into the following general categories:

D_* = Delinquency variables
S_* = Spend variables
P_* = Payment variables
B_* = Balance variables
R_* = Risk variables

Static Plots and Features:

For static plots, the project will leverage:

- **Bar Charts:** To compare delinquency rates, average spend, and payment habits across different customer segments.
- **Histograms:** For visualizing the distribution of balance amounts and risk scores among cardholders.
- **Box Plots:** To examine variability in spend and payment amounts, segmented by risk levels.
- **Scatter Plots:** Exploring correlations between spend behaviors and delinquency variables, as well as balance amounts and payment habits.
- **Line Graphs:** To observe trends over time in payment regularity, spend patterns, and risk score evolution.
- **Area Charts:** Visualizing cumulative metrics like total spend or balance amounts over time, segmented by risk categories.
- **Pie and Donut Charts:** Displaying the proportion of delinquency incidents across different risk levels.

- **Violin Plots:** For a deeper analysis of the distribution patterns in spend and balance variables across various customer demographics.
- **Heat Maps:** Identifying strong correlations among the delinquency, spend, payment, balance, and risk variables.
- **Stacked Bar Charts:** Comparing total and segmented behaviors in spending and payment patterns across different risk categories.
- **Treemap:** Offering a hierarchical view of consumer behaviors, segmented by delinquency levels, risk categories, and spend types.
- **Radar Charts:** To contrast the profiles of different consumer segments based on their financial behaviors and risk scores.

Interactive Dashboard Features and plan:

For the interactive plots on the dashboard, we will focus on the following key features:

- **Dynamic Selection Filters:** Allowing users to filter the dataset by risk categories, spend types, balance brackets, and more, to tailor the analysis to specific interests.
- **Interactive Scatter Plots:** With dynamic filtering capabilities for deeper exploration of the relationships between spend behaviors, payment regularity, and risk assessment.
- **Customizable Time Series Analysis:** Users can trace the evolution of key metrics over time, adjusting for factors like risk level or customer segment.
- **Comparison Tool:** Facilitating side-by-side comparisons of consumer behaviors and risk profiles across different segments.

To ensure the dashboard is practical and user-friendly, the plan includes:

To maximize utility and user engagement, the project's dashboard will be designed with a focus on intuitiveness, responsiveness, and performance optimization. This includes a user-friendly layout, cross-device compatibility, comprehensive help resources, and efficient data handling. By combining rigorous data analysis with accessible visualization tools, this project aims to advance the field of credit default prediction, offering actionable insights for risk management and consumer lending strategies.

Link : https://www.kaggle.com/competitions/amex-default-prediction/data?select=train_data.csv

Please note: The dataset for this project exceeds 15GB in size, making it impractical to upload directly to Blackboard. You can download the dataset or view its overview through the above link. This ensures you have access to the necessary data for analysis without encountering file size limitations on the educational platform.