Project 2

IE6600 Computation and Visualization Fall Semester 2023

Group 11

Pramoth Guhan

guhan.p@northeastern.edu

Rohith Reddy Aleti

aleti.ro@northeastern.edu

Submission Date: 12/10/2023

Introduction:

In the ever-evolving realm of finance, the landscape of loaning serves as a transformative force, shaping economic activities and providing a pathway for individuals to realize their aspirations. This Tableau dashboard embarks on an insightful journey through an extensive dataset centered around loans, delving into the intricate web of factors that influence the borrowing and loaning processes. Approximately 19.4 million Americans currently hold personal loans, with an average size of $9,928.62. Despite the rise of online lending, traditional banks remain the preferred choice for a majority, with 56.2% of borrowers opting for bank loans, surpassing online lenders at 32.2%, while credit unions account for 26.5%. This exploration delves into the dynamics influencing the borrowing economy, offering insights into lender preferences and borrower behaviours, contributing to the ongoing dialogue on personal finance and economic empowerment.

Problem statement:

The contemporary financial landscape is heavily influenced by lending practices, making it imperative to comprehend the nuanced factors at play. This project addresses key questions, such as the determinants of loan approval, the impact of creditworthiness on interest rates, and the historical performance of loans. By articulating these questions, we aim to unravel the complexities of lending operations and glean actionable insights.

Data Description:

The dataset encompasses a comprehensive collection of over 5.7 million loan applications and borrower profiles sourced from LendingClub, spanning the years 2007 to 2018 in the Northern American region. With 115 columns, key attributes include loan amounts, annual income, interest rates, loan status, and various financial indicators. The total loan amount requested exceeds $750 million, distributed across 580,000 open accounts in the United States. The average loan amount is over $15,000, and the average interest rate for loans issued between 2007 and 2018 stands at 13.08%. The dataset offers insights into the magnitude of loan requests, associated risk and return, and the demographics of borrowers. Noteworthy columns such as Loan Amount, Interest Rate, Annual Income, Loan Status, Purpose, and Credit Score (FICO) provide critical information for analyzing the dataset. Additionally, the dataset categorizes borrowers into seven different grades and five subgrades based on their creditworthiness.

Methodology:

The project methodology is characterized by a systematic and thorough approach to ensure reliable and clear analysis. Initial efforts prioritize rigorous data cleaning to maintain data integrity, addressing inconsistencies and missing values. Subsequently, an exploratory data analysis (EDA) uncovers patterns, relationships, and correlations within the extensive dataset. The transition to the visualization phase involves utilizing Tableau to create insightful representations of the data, selecting charts and graphs thoughtfully to communicate key findings and trends. The overarching goal is to craft a coherent narrative through the development of three distinct dashboards, each capturing different facets of the loan analysis. By integrating data cleaning, exploratory analysis, and visualization with Tableau, the project aims to provide valuable insights into loan dynamics, borrower profiles, and associated financial indicators.

Design process:

The design process for this Tableau dashboard project is marked by a deliberate effort to distill intricate information about loans into visually appealing representations. With a dataset spanning over 5.7 million loan applications, the emphasis is on clarity and comprehension. Rigorous data cleaning ensures the integrity of the dataset, laying the foundation for insightful explorations. The choice of visualization types, such as charts and graphs, is carefully considered to effectively communicate key findings, including loan amounts, interest rates, and borrower demographics. A cohesive color scheme and layout are employed to enhance visual coherence and aid interpretation. Furthermore, the incorporation of interactive elements in the dashboards provides users with an immersive experience, enabling them to explore the data dynamically. The overarching goal is to create a visually engaging narrative that not only conveys complex financial information but also invites users to interact and gain deeper insights into the dynamics of loans and borrower behaviours.

Key insights from the data:

The key insights derived from the three dashboards—Loan Purpose Analysis, Loan Distribution Analysis, and Loan Grade Analysis are as follows:

Loan Purpose Analysis:

1. Average Interest Trend: The average interest rates exhibit a normal trend from 2006 to 2012, followed by a notable increase in 2013 and a subsequent decline from 2014 to 2016.
2. Square Area Visualization: A visual representation of loan purposes, with Debt Consolidation, Credit Card, and Home Improvement standing out over the years. The specific purposes vary annually.

3. Loan Purpose vs Amount: A comparative analysis of loan amounts for different purposes, highlighting trends and variations across various years.
4. Top 5 States: California, Florida, Illinois, New York, and Texas emerge as the top five states with the highest number of open accounts.

Loan Distribution Analysis:

1. Loan Amount vs States: Illustrates the distribution of loan amounts across different states in the USA between 2007 and 2018.
2. Loan Status Breakdown: A percentage breakdown of loan statuses, with Fully Paid at 47.35%, Current at 39.06%, and Charged Off at 11.87%, among others.
3. Loan Issued per Ownership: Analyses the number of loans issued based on home ownership status, showcasing a noticeable increase after 2013.
4. Loan Issued per Status Type: Examines the distribution of loans based on status types, including Fully Paid, Current, and Charged Off.

Loan Grade Analysis:

1. Loan Term Distribution: A comparison of loan distribution between 36 months and 60 months terms, with 60 months terms dominating at approximately 90 billion compared to 36 months terms at around 36 billion.
2. Number of Loans vs Loan Status: Demonstrates a post-2012 surge in the number of loans, enabling a comparison between Fully Paid and Charged Off categories to assess the proportion of good and bad loans.
3. Loan Distribution Across Grades for Application Type: Breaks down loan distribution across seven grades (A to G) with a filter for application types, revealing variations in amounts for individual and joint applications.
4. Trends of Annual Income: Explores trends in annual income over the years, providing insights into the changing dynamics of borrower income.

These insights collectively offer a comprehensive understanding of loan dynamics, borrower behaviours, and trends in interest rates and loan purposes over the years.

Conclusion:

In conclusion, the Tableau dashboards provide illuminating insights into loan dynamics, revealing trends in interest rates, loan purposes, and borrower behaviours. The visualizations offer a user-friendly exploration of the extensive dataset, facilitating a nuanced understanding of the financial landscape from 2007 to 2018.