

Project 1 Proposal – Group 12

1. Introduction:

- a. Our selected dataset, Financial Transactions Dataset: Analytics, contains financial payment card and card holder demographic information from 2010 through 2019. We chose this dataset to explore how financial health, measured by debt-to-income (DTI) ratio, varies based on demographics, individual financial attributes, and geographic regions, and to determine the correlations between these characteristics and overall financial health.
- b. Selected Dataset:
<https://www.kaggle.com/datasets/computingvictor/transactions-fraud-datasets/code>

2. High level questions:

- a. How do customer demographics (e.g., age, gender, income) correlate with customer financial health?
- b. How do financial attributes (e.g., card type, loan amount, credit score, total debt) correlate with customer financial health?
- c. How does customer geographic information (e.g. state, city, per capita income) correlate with customer financial health?

3. Inspiration:

In recent years, rising inflation resulted in broader societal focus on consumer spending habits and individual financial health. From government bodies to the media, the conversation largely focuses on the resilience of consumers in the face of inflation, and on the sustainability of consumer spending habits as inflation impacts overall financial health. Assessing the pre-pandemic financial health of consumers will provide insight regarding the “normal” financial habits and financial health of consumers prior to the impacts of the COVID-19 pandemic and subsequent inflation. This may facilitate future comparative analysis of financial health characteristics from these two distinct periods.

Comparable research in this area also serves as conceptual inspiration for this effort. Examples include:

- Center for Economic and Policy Research study, “Before and After the Pandemic: Income Volatility, Health Care Affordability, and Debt”

(<https://cepr.net/report/before-and-after-the-pandemic-income-volatility-health-care-affordability-and-debt/>)

- Financial Health Network’s study, “Financial Health Pulse 2023 U.S. Trends” (<https://finhealthnetwork.org/wp-content/uploads/2023/09/2023-Pulse-U.S.-Trends-Report-Final.pdf>).

4. Possible Visualizations

- A minimum of two bar or hbar charts showing the following:
 - Average or median debt-to-income levels by age group
 - Average per-capita income by state
- A split violin chart showing the debt-to-income distribution by gender
 - Potentially perform a t-test on male vs. female debt-to-income
- A scatterplot showing the average count of cards held by credit score ranges
- A donut chart displaying the distribution of total debt by credit score range
- A map plotting the highest and lowest debt-to-income ratios by geographic locations.
 - Potential heat map/gradient color scheme embedded in each plotted location to show the highest debt-to-income locations (red) and lowest debt-to-income locations (green)
 - Potential box plot for locations
 - Potential ANOVA test across several locations
- Liner regressions measuring the following relationships:
 - Debt-to-income ratio vs. income
 - Debt-to-income ratio vs. credit score

5. Color Theme

- <https://coolors.co/palette/386641-6a994e-a7c957-f2e8cf-bc4749>

6. Roles & Responsibilities

- Question 1 (and slides) – Charles
 - A minimum of two bar or hbar charts showing the following:
 - Average or median debt-to-income levels by age group
 - Average per-capita income by state
 - A split violin chart showing the debt-to-income distribution by gender
 - Potentially perform a t-test on male vs. female debt-to-income
- Question 2 (and slides) – Gavin
 - A scatterplot showing the average count of cards held by credit score ranges

- ii. A donut chart displaying the distribution of total debt by credit score range
- c. Question 3 (and slides) – Caleb
 - i. A map plotting the highest and lowest debt-to-income ratios by geographic locations.
 - 1. Potential heat map/gradient color scheme embedded in each plotted location to show the highest debt-to-income locations (red) and lowest debt-to-income locations (green)
 - 2. Potential box plot for locations
 - 3. Potential ANOVA test across several locations
- d. Regression (and slides) – Luisa
 - i. Debt-to-income ratio vs. income
 - ii. Debt-to-income ratio vs. credit score
- e. Other Presentation Requirements
 - i. Introduction and Inspiration – Caleb
 - ii. Goals/Objectives – Charles
 - iii. Bias and Limitations – Gavin
 - iv. Conclusion/Final thoughts – Luisa
- f. Write-up: Collaborative Effort