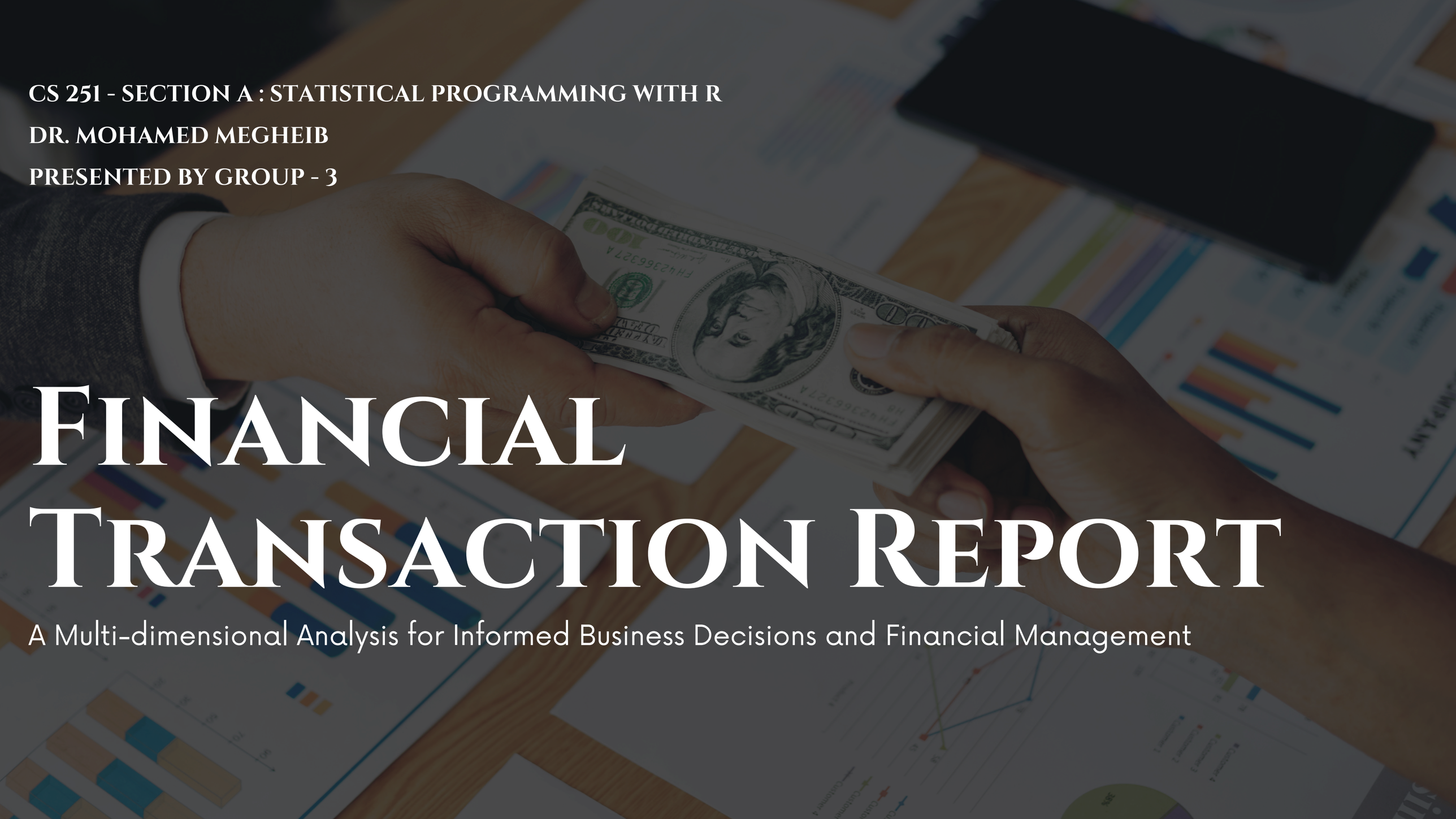


CS 251 - SECTION A : STATISTICAL PROGRAMMING WITH R

DR. MOHAMED MEGHEIB

PRESENTED BY GROUP - 3

Financial Transaction Report

A Multi-dimensional Analysis for Informed Business Decisions and Financial Management

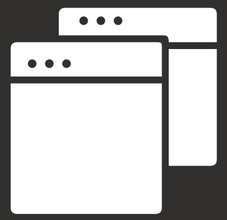


Table of Contents

Points to discuss:

INTRODUCTION

The rise of digital payment systems has helped us make purchases conveniently. But,

The problem is it cause impulsive Purchasing Behaviors (Boopathy & Kanagaraj, 2023).

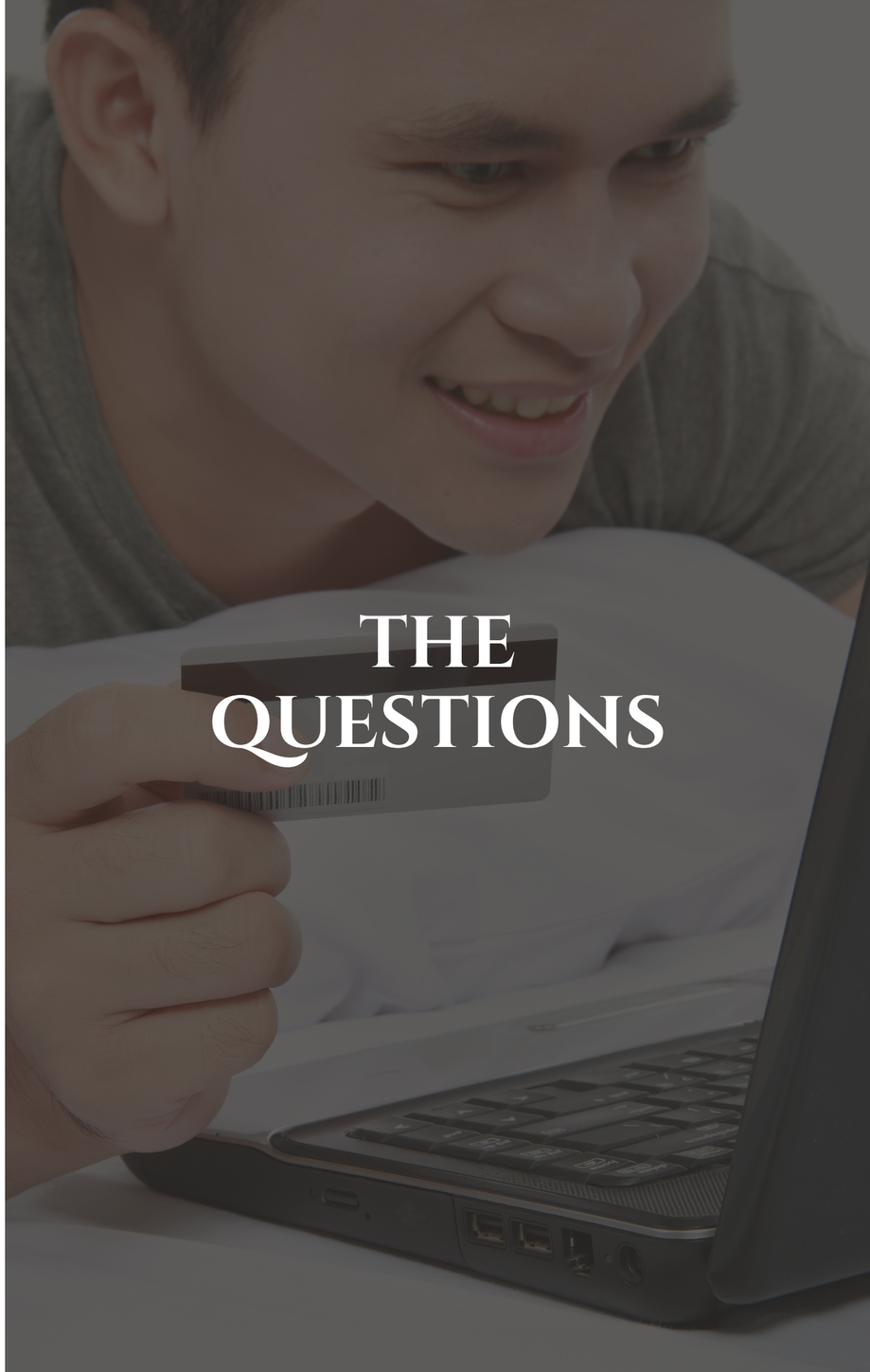
Digital payment market size - 4.7 trillions dollars in 2020 (Statista, 2020). It is expanding.

IMPULSIVE PURCHASING BEHAVIORS LEAD TO **UNMANAGEABLE DEBT.**

The total credit card debt of Americans is around 1.166 trillions U.S Dollars in the third quarter of 2024. (Schulz & Shepard, 2024)

- **Are Americans having impulsive purchasing behaviors?**
- **What are their spending patterns and behaviors?**
- **Are they in debt?**

Downloaded a credit card financial transactions dataset from a U.S finance institution, available on Kaggle.



THE QUESTIONS

LITERATURE REVIEW

People abilities' to choose the right financial priorities are distracted by 'Impulsive purchasing behaviors'.

Young people have low credit scores. Credit card agencies indicate they do not use income as a parameter for calculating credit score.

Data Description

Source

From a financial institution, published on Kaggle

Usability Score

Perfect 10.0 on Kaggle, 100% in completeness, credibility, and compatibility.

Main Dataset Name

Financial Transactions Dataset

transactions_data,

Includes transaction amounts, timeline from 2010 to 2019.

13 Millions rows of data.

cards_data

Includes credit and debit card details.

merchant_category_data

Business Type explanation from numerical codes in transactions_data

users_data

Includes demographic information of card holders.

Data Formats

Sub-datasets in CSV file formats and a JSON format.

METHODOLOGY

Data Collection

- transactions_data, users_data, cards_data, and a JSON file of merchant categories

Exploratory Data Analysis/ Data Wrangling

- Overviews using summary statistics.
- Cleaned and converted data types for key variables

Data Cleaning Highlights

- transactions_data: Removed $/- signs from amount; changed transaction type (use_chip) to factor
- users_data: Converted yearly_income and total_debt columns to numeric
- cards_data: Converted credit_limit to numeric
- Merchant categories JSON: Converted mcc_code to integers

Merging Datasets

- Merged datasets using client_id and mcc_code keys into combined datasets
- Removed null values

LINEAR REGRESSION AND MODEL EVALUATION

Linear Regression and Feature Selection

- Relevant predictors based on domain knowledge

Visual Analysis

- Residual plots
- Predicted vs. Actual plot

Training and Testing

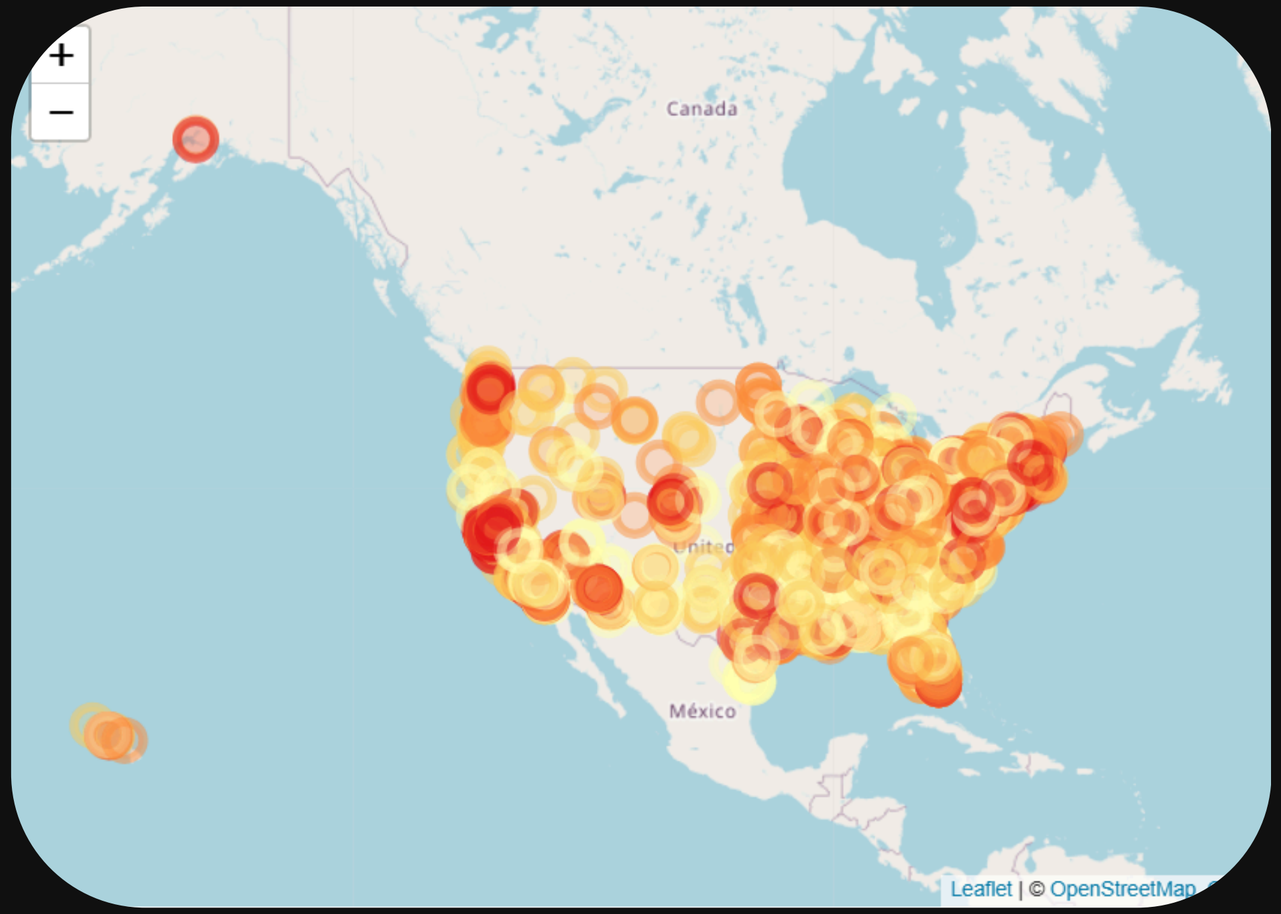
- Training (80%) and testing (20%)
- Random Forest Regression model

Evaluation Metrics

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2 Score

Research Findings

Finding 1.1: Geographical Locations of the Users using Interactive Maps



Maps based on user longitudes and latitudes from users_data using Leaflet library.

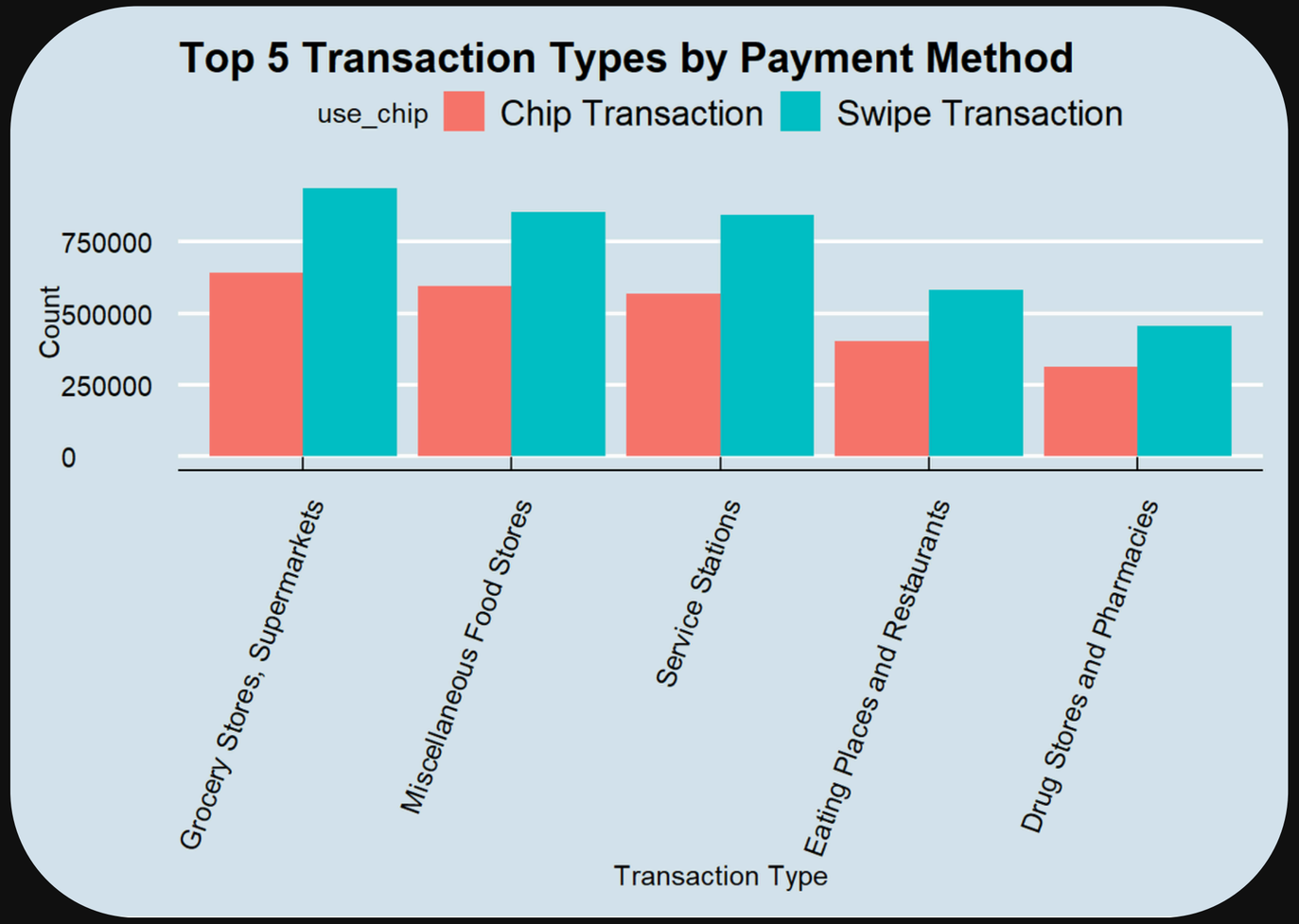
Users span diverse states within the United States.

Finding 1.2: Users with Highest Income Location

Highest incomes concentrated in major financial states: New York and California.

Suggest better financial opportunities in regions with strong economies

Finding 2.1: Transaction Types by Payment Method

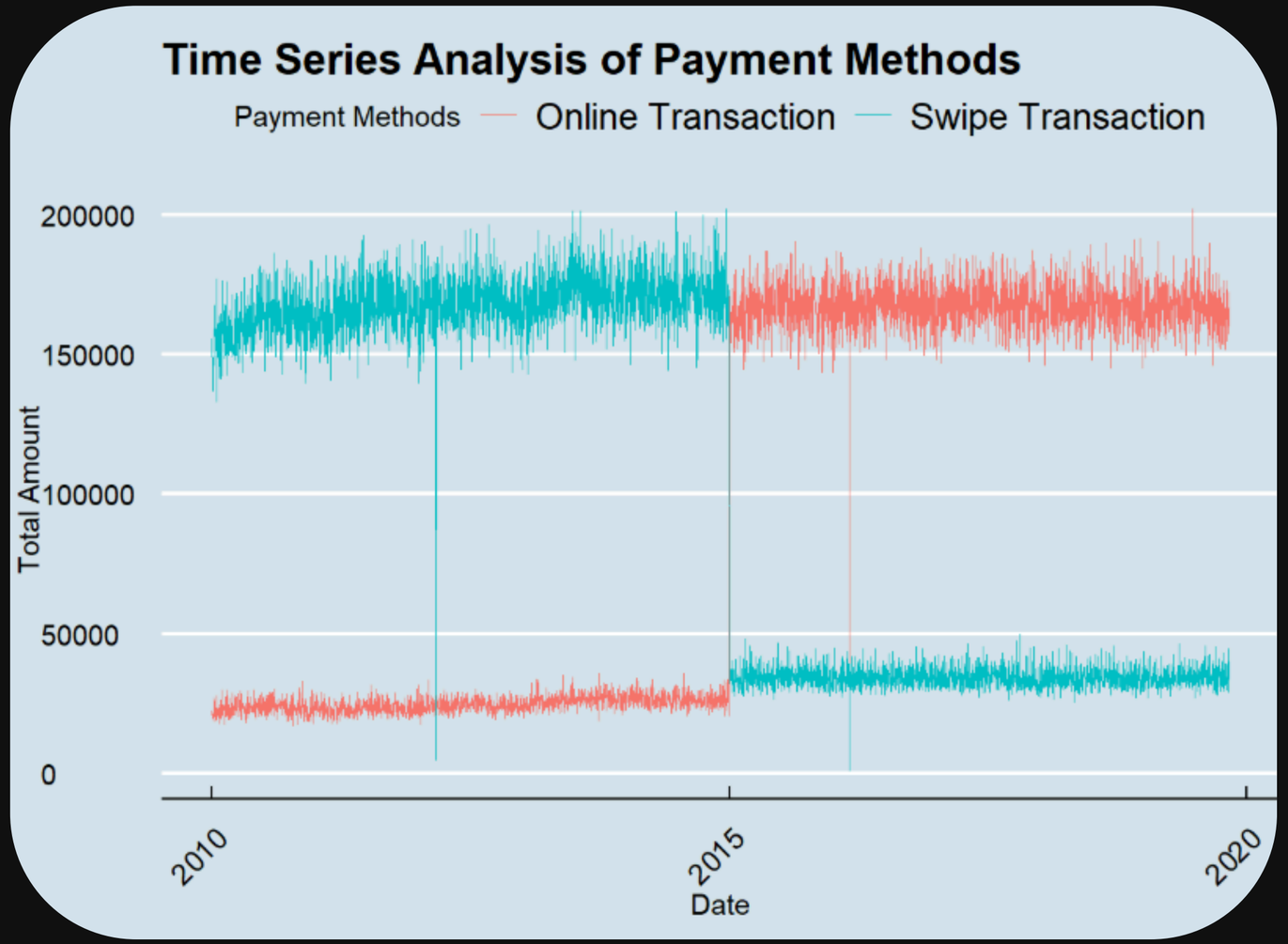


Preference for In-Store Purchases: Individuals prefer purchasing items in stores.

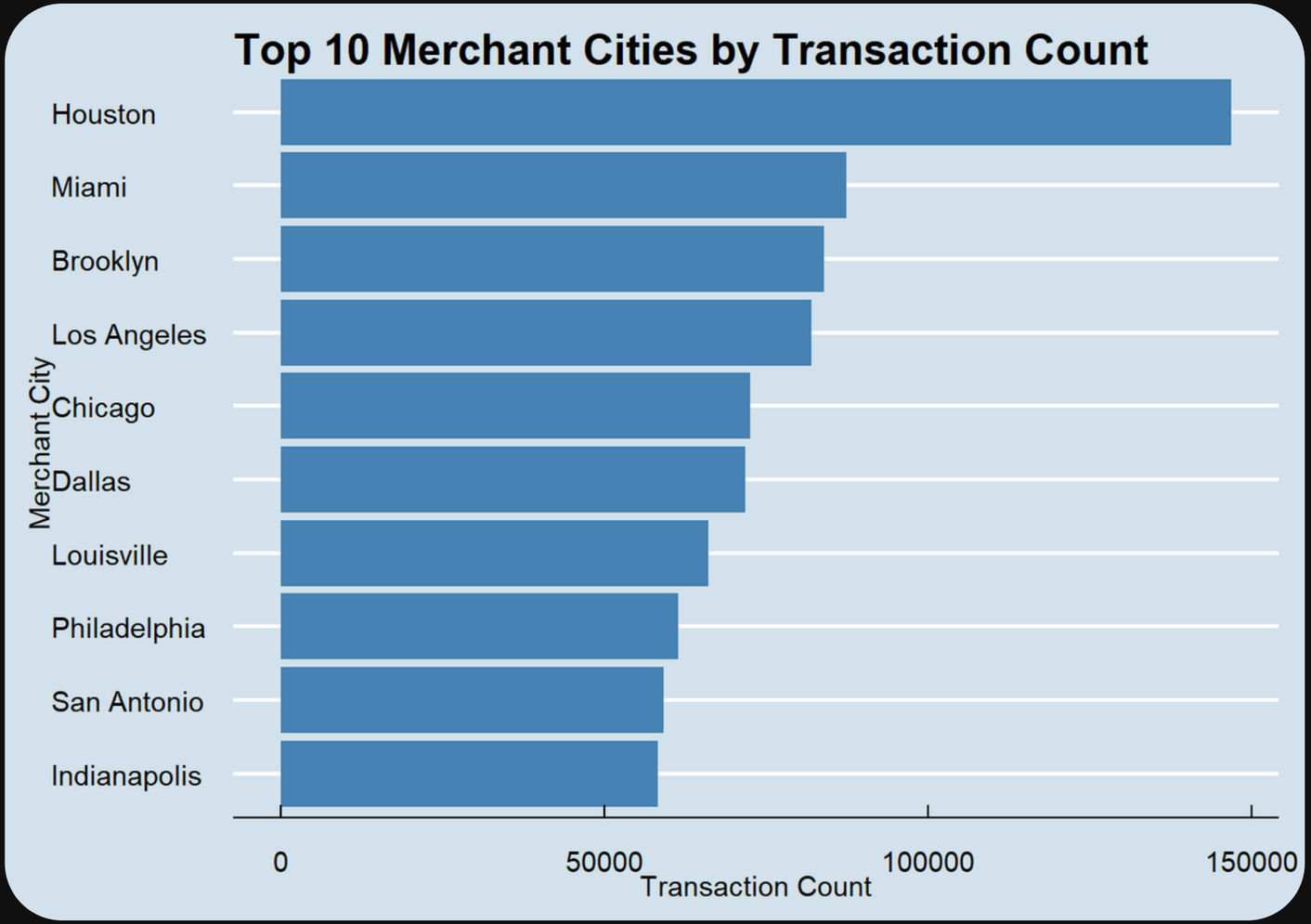
Financial Management Insights: Different payment methods can help businesses make informed decisions and optimize their financial management.

Finding 2.2: Time Series Analysis of Payment Methods

- The shift indicates broader adoption of e-commerce and digital wallets.
- 2016 Anomaly: A sudden drop to zero for both swipe and online transactions suggests a possible technical error.
- Business Insight: Businesses can understand changing customer preferences in payment methods and make informed decisions.



Finding 2.3: Top Merchant Cities by Transaction Count



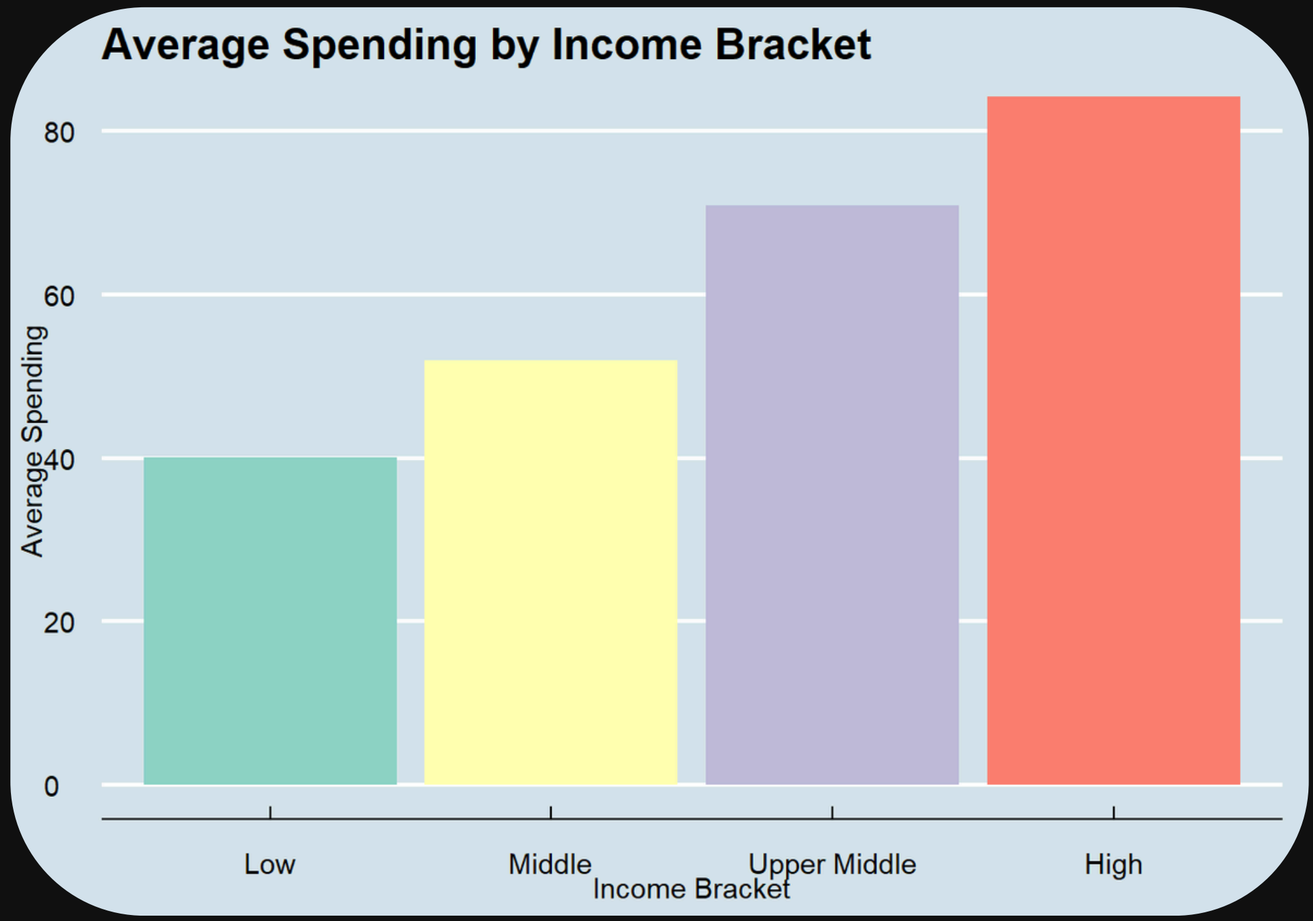
Analysis of transaction data highlights key cities excluding online transactions.

Business Opportunities: Cities where businesses can target strategically

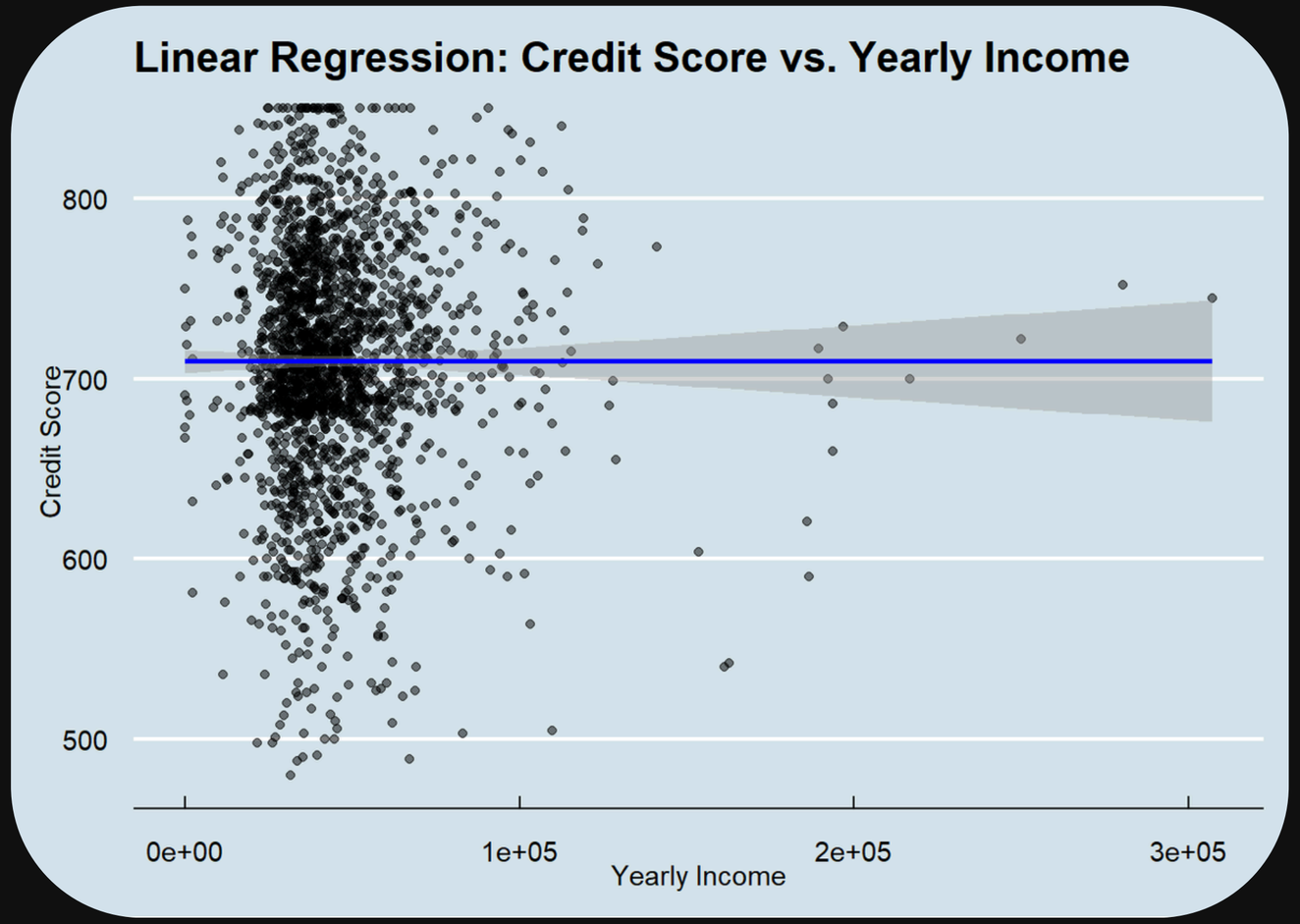
Finding 2.4: Average Spending by Income Bracket

- Grouped by yearly income:
  - Low ($0–$30k)
  - Middle ($30k–$70k)
  - Upper middle ($70k–$150k)
  - High (above $150k).

- Spending increases with income.



Finding 3.1: Relationship Between Credit Score and Yearly Income



- Yearly Income Coefficient: 0.0000004872
- Other Influential Factors: Credit history, debt payment, etc.

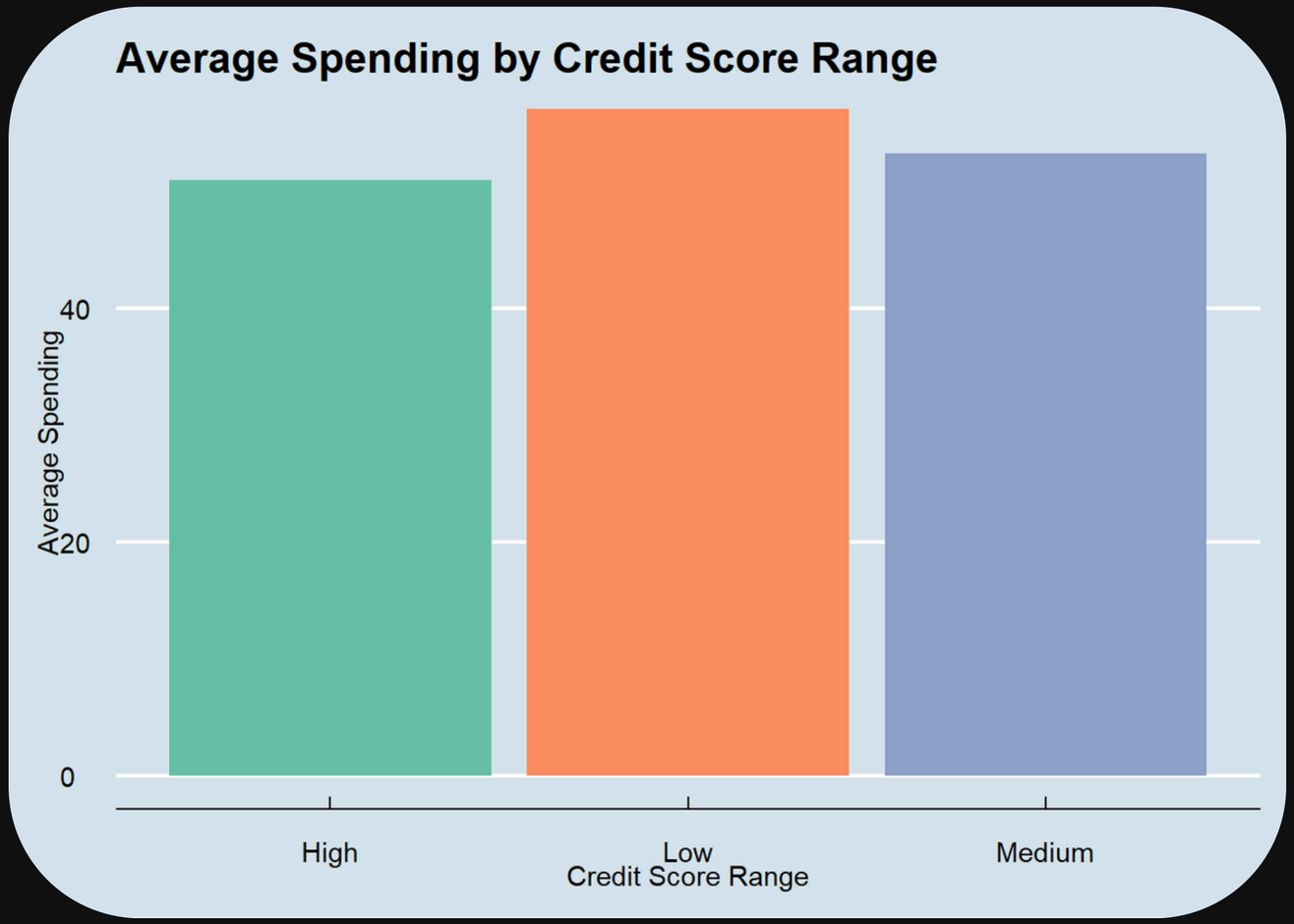
Finding 3.2: Average Spending by Credit Score Range

Credit Score Categories:

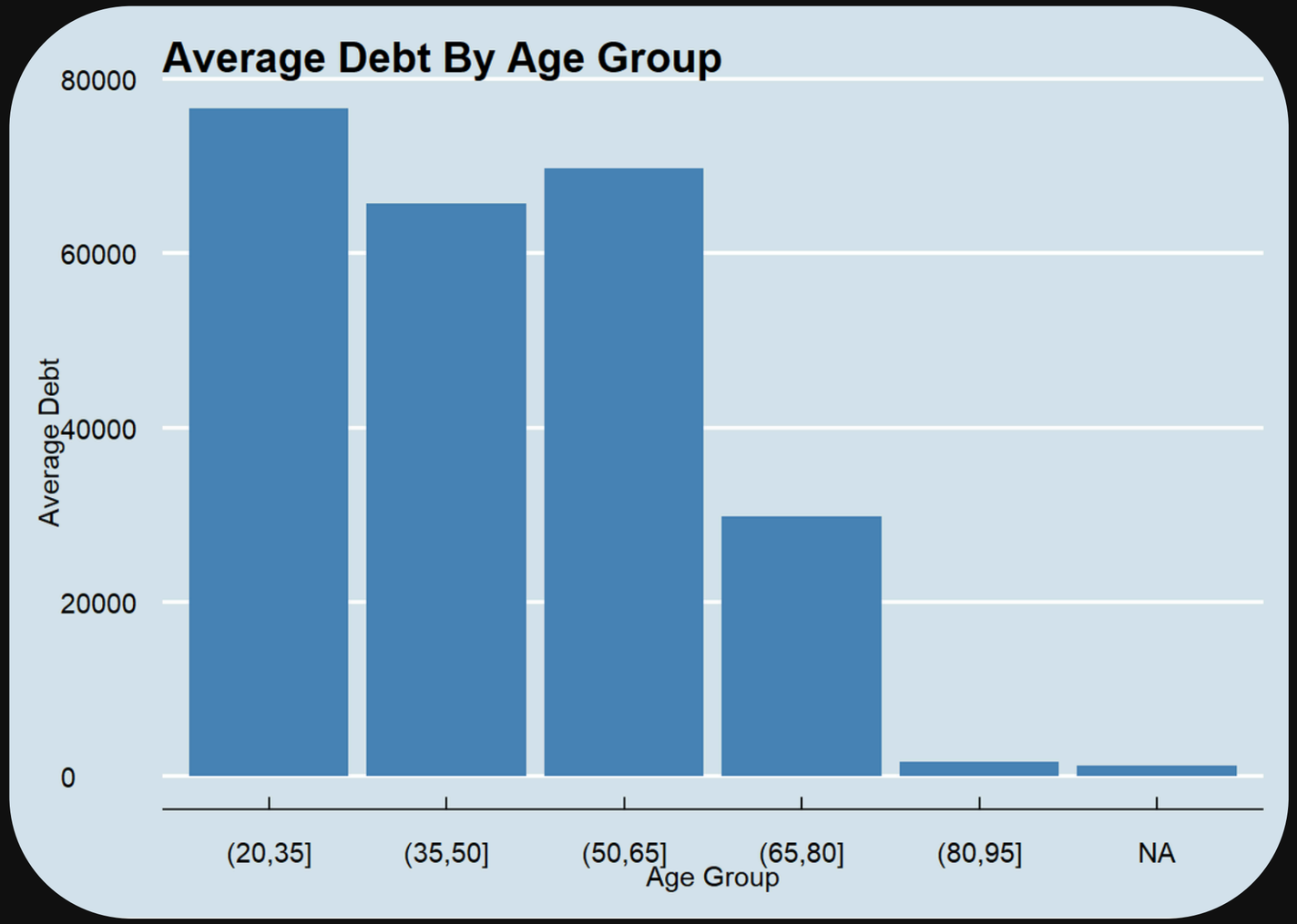
- Low (below 600),
- medium (600–699),
- high (700 and above)

Low credit scores = Highest Average Spending

Individual interests, loans, and debts



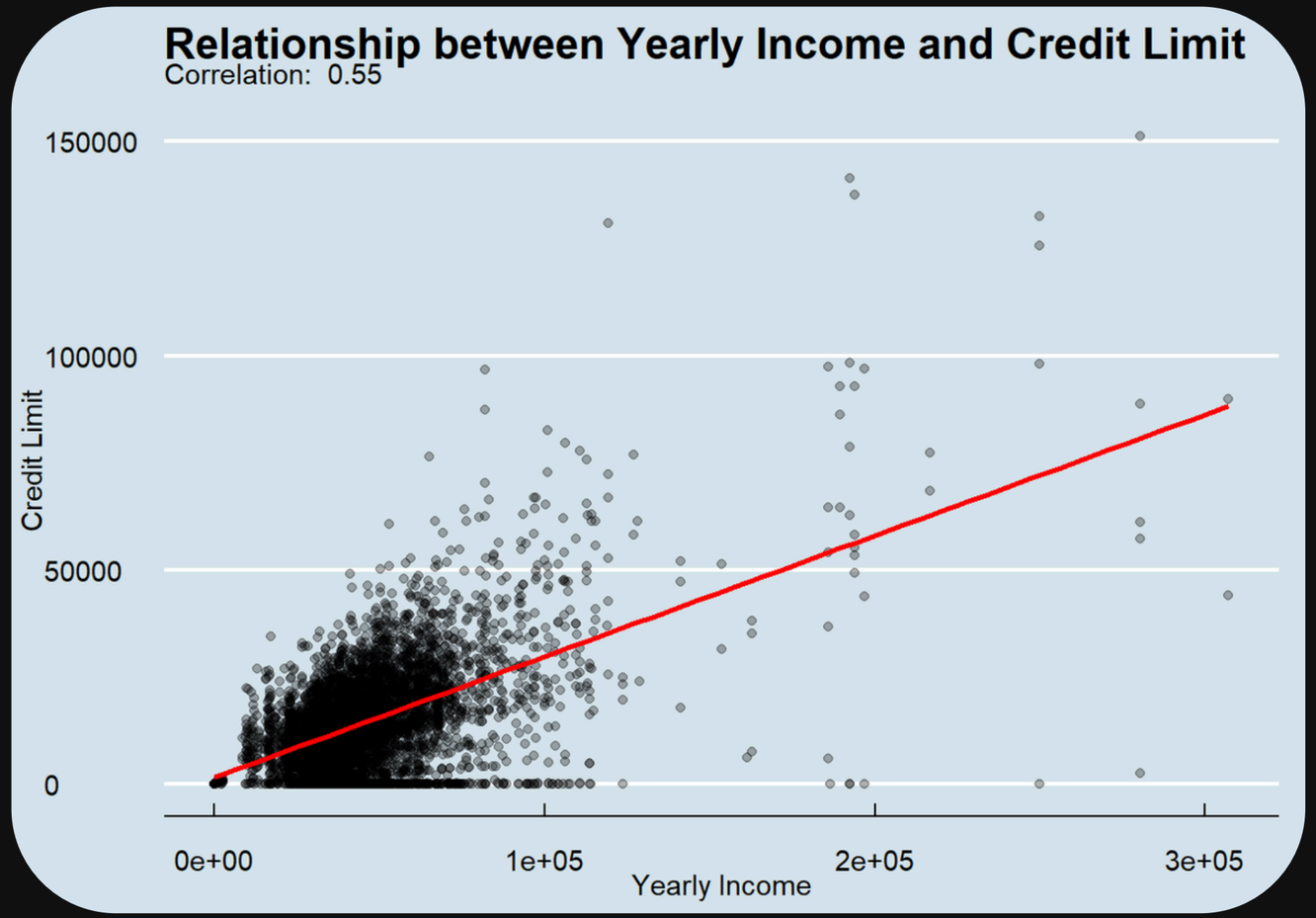
Finding 5: Relationship between Age Groups and Average Debt



Highest Debt Levels: 20–35 age group -> almost $80,000

Life-Cycle Effect: Trends suggest a potential life-cycle effect on debt accumulation and repayment.

Finding 7: Relationship between Yearly Income and Credit Limit



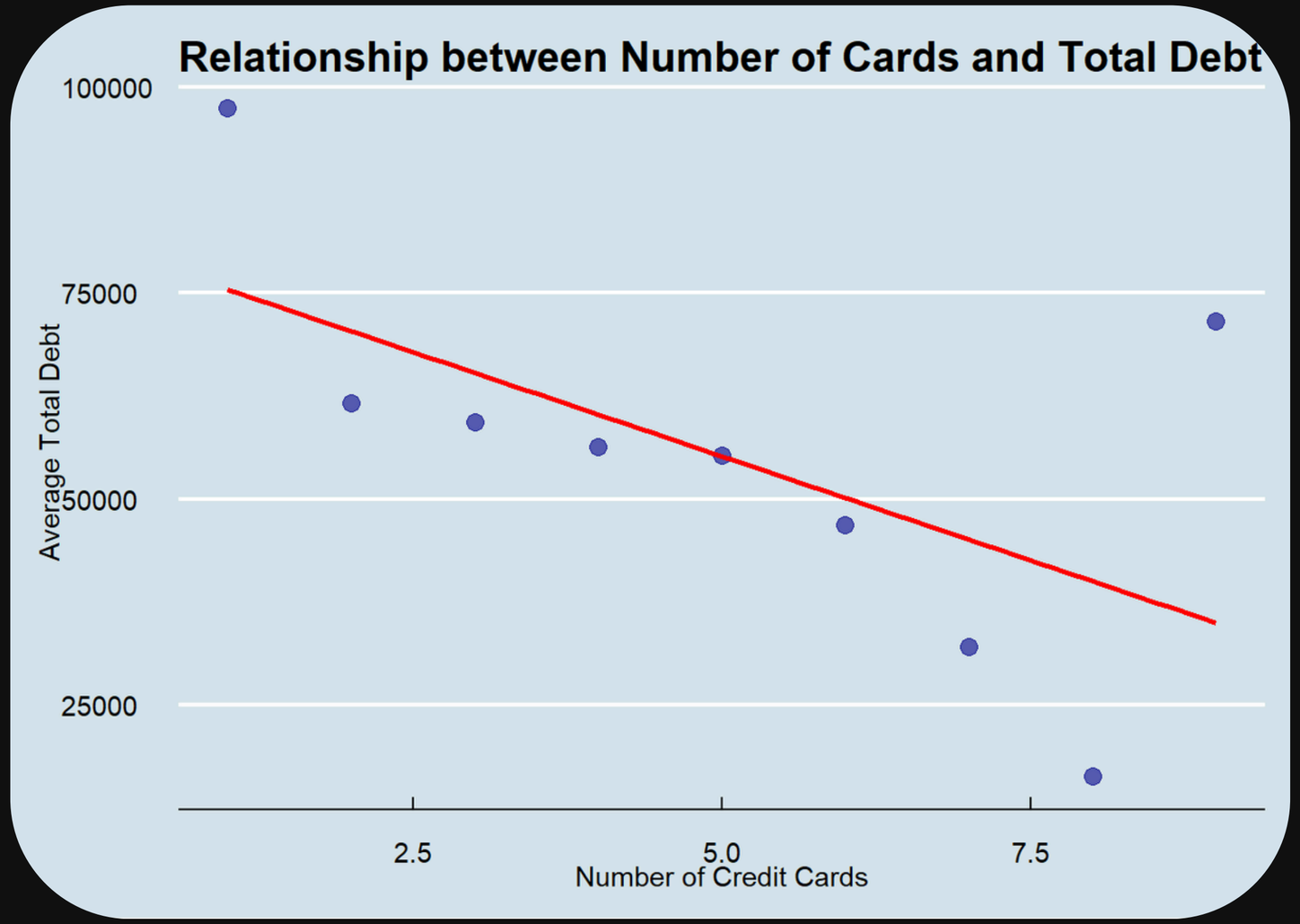
55% positive correlation

Higher incomes = increased credit limits.

Lenders' Confidence: Reflects lenders' confidence in individuals' ability to manage debt effectively.

Finding 8: Relationship between Number of Cards and Total Debt

- Average total debt < the number of credit cards increases.
- Negative Correlation
- More Cards, more likely to manage their debts
- Bank Strategy: Encourage more cards, subscribe to premium products



Credit Limit Allocation Prediction Model

Objective: Predict Random Forest Regression (RFR) model.

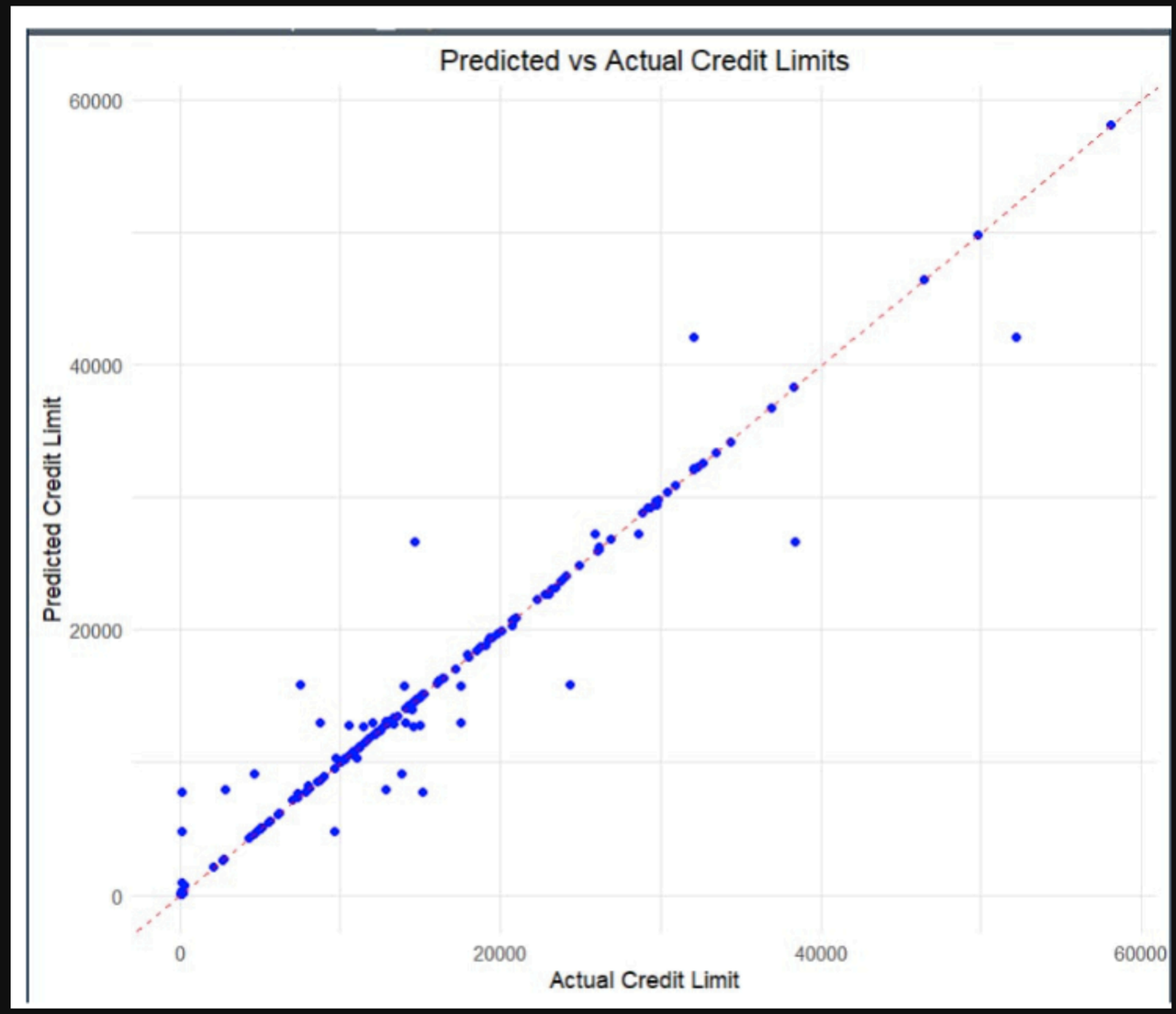
Features: Yearly income, credit score, current age, total debt, and average spending

Training Process:

- 80% training and 20% testing data
- 100 decision trees.

Evaluation Metrics:

- Mean Absolute Error (MAE): 780.64
- Root Mean Squared Error (RMSE): 2264.05
- R^2 Score: 0.95



Discussion and Results

LIMITATIONS

- Short Analysis Period
- Analytical Tools Limitation
- Data Size Challenge
- Data Complexity
- Historical Insights
- Dataset Representation
- Potential Bias

- Better targeting and customer segmentation
- Adapting to the new trends
- Effective financial management system
- Suggestion for more data exploration
- Help banks, businesses, and individuals

Discussion

Conclusion

Analysis includes financial status, spending, transaction trends, cash flows, and financial management across the US.

- Geographical Influence
- Payment Method Shift
- Spending Patterns
- Income and Credit Score

- Debt Levels
- Credit Card Management
- Credit Limits and Year Income

References

Beer, R., Ionescu, F., & Li, G. (2018, August 13). Are Income and Credit Scores Highly Correlated? FEDS Notes. https://www.federalreserve.gov/econres/notes/feds-notes/are-income-and-credit-scores-highly-correlated-20180813.html

Boopathy, G., & Kanagaraj, P. (2023). The Impact of Digital Payments on Consumer Spending Habits. Tuijin Jishu/ Journal of Propulsion Technology, ISSN: 1001 - 4055, 44(4), 5373. DOI: https://doi.org/10.52783/tjjpt.v44.i4.1896

Best, R. D. (2024, May 22). Credit card penetration in 161 countries worldwide 2011, 2014, 2017, 2021. https://www.statista.com/statistics/675371/ownership-of-credit-cards-globally-by-country/

Clearly Payments (2023). The Growth of the Credit Card Industry in 2023. https://www.clearlypayments.com/blog/growth-of-credit-card-industry-in-2023/

Chowdhury, M. H. (2023). The Financial Management Practices of Young Adults in Helsinki. A Comparative Study. Arcada University of Applied Sciences: International Business, 2023. https://www.theseus.fi/bitstream/handle/10024/817389/Manzar_Chowdhury.pdf?sequence=2&isAllowed=y

Nyrhinen, J., Sirola, A., Koskelainen, T., Munnukka, J., & Wilska, T. (2024, April). Online antecedents for young consumers' impulse buying behavior. Computers in Human Behavior. Vol.153. https://doi.org/10.1016/j.chb.2023.108129

Putrevu, J., & Mertzanis, C. (2023). The adoption of digital payments in emerging economies: challenges and policy responses. Digital Policy Regulation and Governance. DOI: 10.1108/DPRG-06-2023-0077

Schulz, M., & Shepard, D. (2024, December 6). 2024 Credit Card Debt Statistics. https://www.lendingtree.com/credit-cards/study/credit-card-debt-statistics/

Yee, A. (2024, February 10). Credit card debt among young people rising, Fed report shows. Chicago Sun Times. https://chicago.suntimes.com/2024/2/9/24067906/credit-card-debt-auto-loans-mortgage-interest-rates-millennials-genz-nyfed-report

Do you have any questions?

Please let us know!

With Kindness, from Group - 3!