



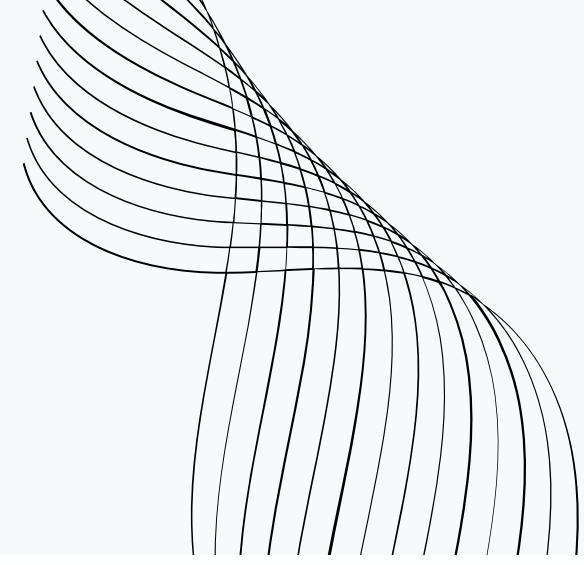
Katherine
Torian

EXPLORATORY DATA ANALYSIS U.S. CONSUMER COMPLAINTS

**UNDERSTANDING CONSUMER ISSUES TO
IMPROVE SERVICE QUALITY**

Overview

- 01** Why this problem?
- 02** Project Focus
- 03** Data Cleaning
- 04** Exploratory Data Analysis
- 05** Key Findings



Why this problem?

Every year, millions of consumers in the US file complaints about financial products and services. These complaints provide valuable insights into issues customers face, such as unfair practices, service delays, or product failures. Understanding these complaints helps businesses, regulators, and consumers identify systemic problems, improve customer experience, and promote transparency. By analyzing this data, we can detect trends, spot emerging issues early, and contribute to better decision-making across the industry.





1 every 20 seconds,
183 every hour,
4,383 every day,
133,333 every month...

1,600,000 complaints per year



Project Focus



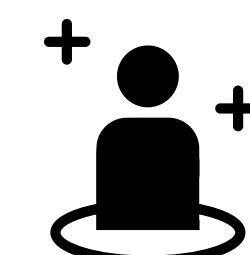
IDENTIFY TRENDS AND PATTERNS

Analyze complaint volumes and issue types over time, and compare how different companies respond to these complaints.



UNDERSTAND SUBMISSION AND RESPONSE:

Investigate the main channels consumers use to submit complaints and measure how quickly companies provide responses.



ANALYZE GEOGRAPHIC AND DEMOGRAPHIC IMPACT:

Examine how complaint volumes and types vary by region and demographic factors across the US.



EVALUATE COMPANY COMPLAINT RATES:

Identify companies with consistently high complaint numbers and assess their effectiveness in handling and resolving issues.



ASSESS RESOLUTION OUTCOMES:

Determine the most common complaint resolutions and analyze how often consumers dispute company responses.



Data Cleaning

From +7GB to under 2GB

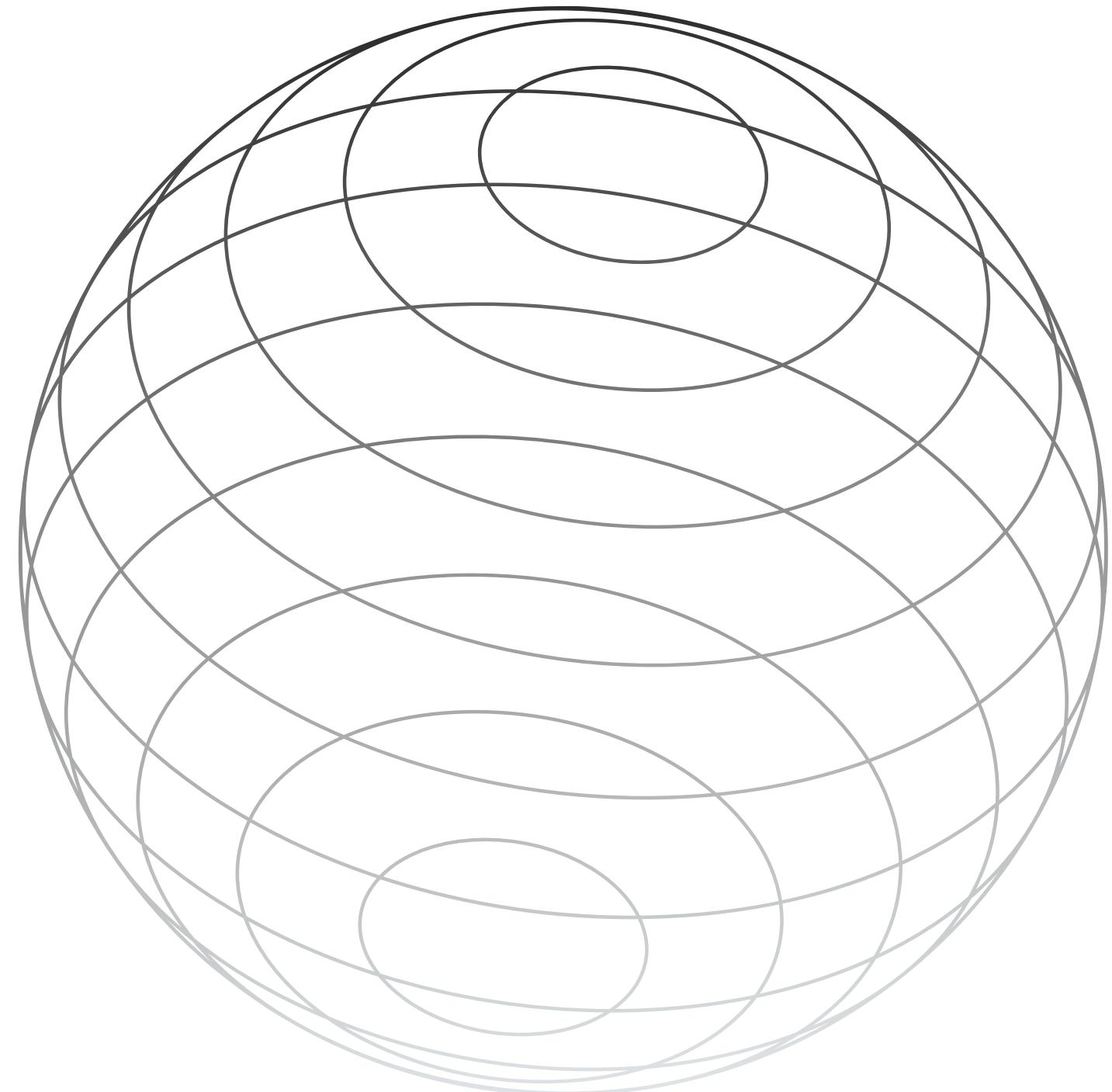
The original Consumer Complaint database was very large, around 7 GB, which made it difficult to process efficiently. To manage this, I first filtered the dataset by year, selecting only the most recent and relevant years. This reduced the dataset size to under 2 GB, making it more manageable for analysis.

Choosing columns wisely

I selected specific columns that were most relevant to the project goals. Removing the redundant columns was a good way to focus on the analysis and improve processing speed.

The outcome?

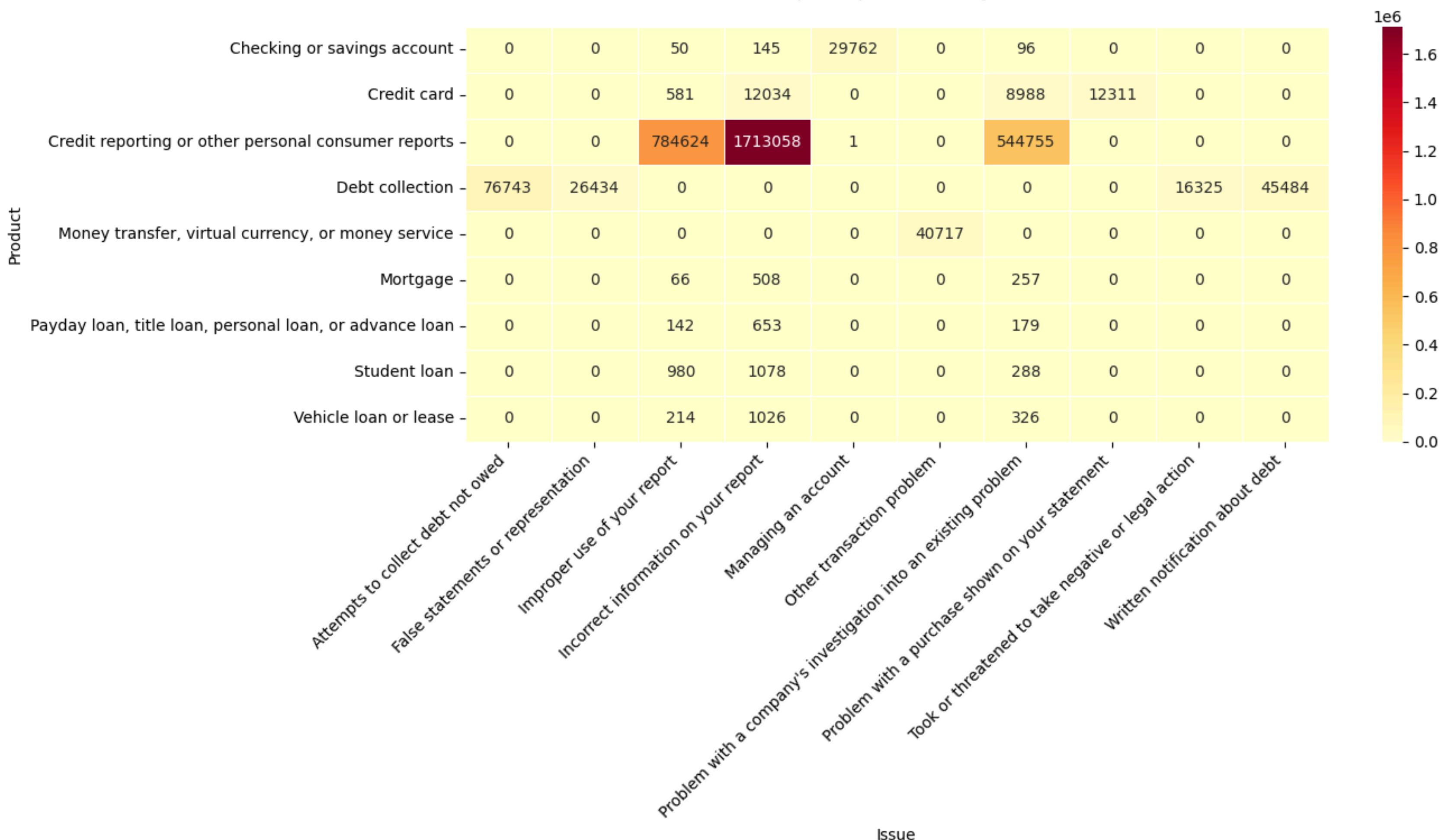
I ensured that the dataset was clean, relevant, and optimized for exploratory data analysis.



Question #1

WHAT ARE THE TRENDS OR PATTERNS THAT CAN BE FOUND IN CONSUMER COMPLAINT VOLUMES, ISSUES AND IN THE RESPONSE BEHAVIOUR ACROSS DIFFERENT FINANCIAL PRODUCTS AND THROUGHOUT DIFFERENT COMPANIES?

Heatmap of Top 10 Issues by Product

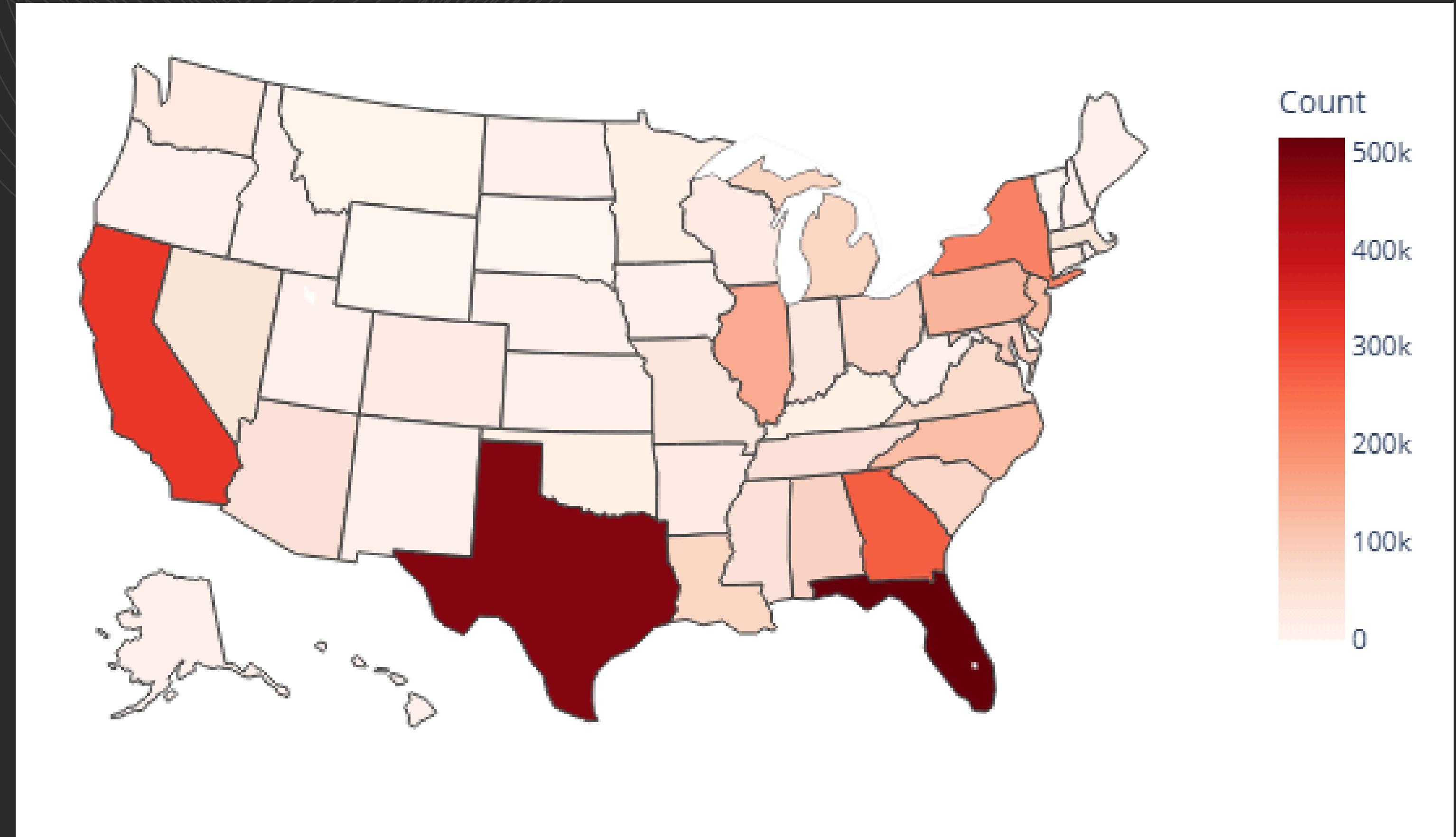


This means...

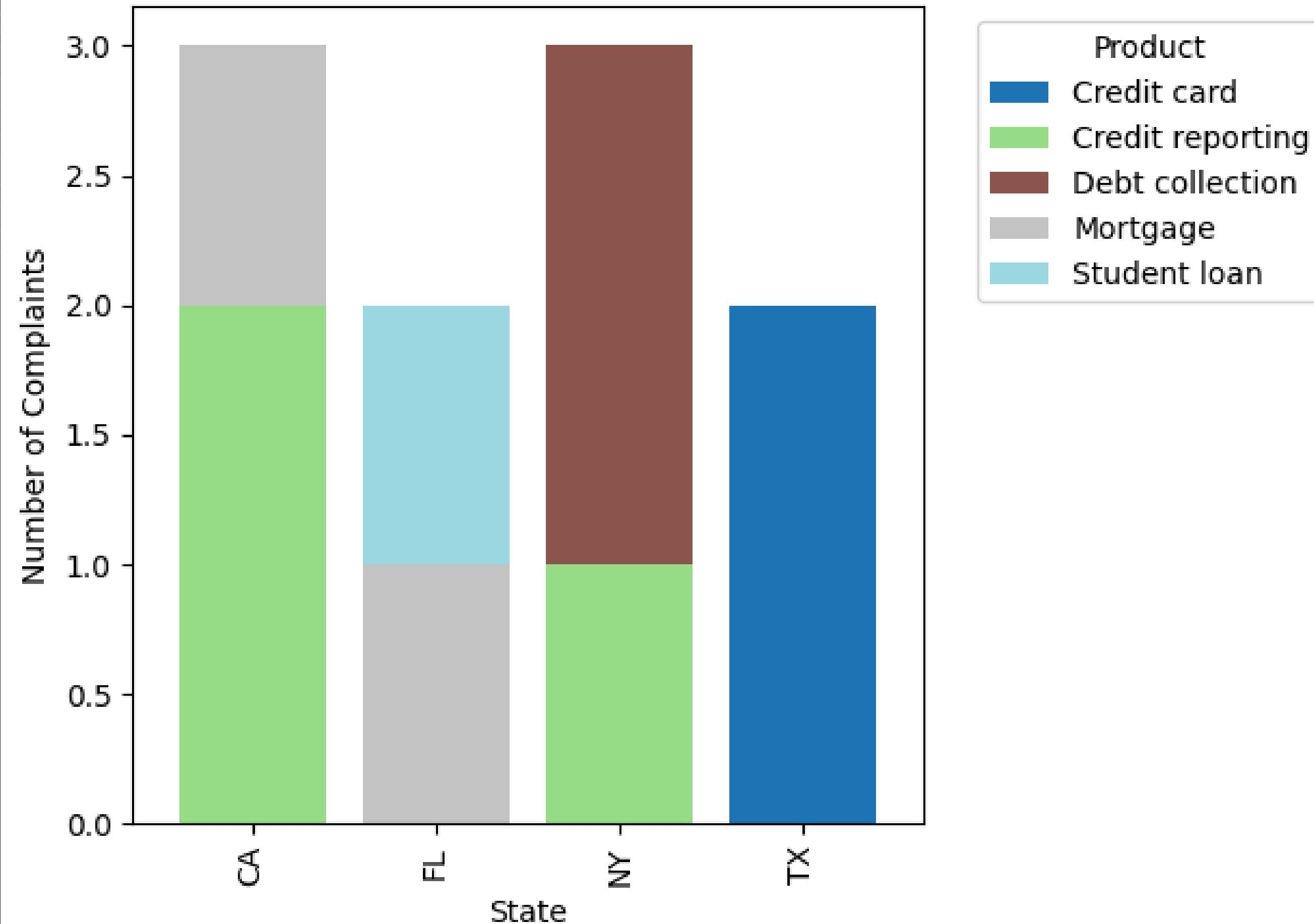
- DOMINANT COMPLAINT CATEGORIES: CREDIT REPORTING AND DEBT COLLECTION DRIVE MOST COMPLAINTS.
- KEY ISSUES WITHIN DOMINANT CATEGORIES: INCORRECT DATA AND FALSE DEBT CLAIMS ARE PRIMARY CONCERNS.

Question #2

WHICH DEMOGRAPHIC OR
GEOGRAPHIC FACTORS INFLUENCE
COMPLAINT VOLUMES THE MOST, AND
ARE THERE REGIONAL DIFFERENCES IN
THE TYPES OF FINANCIAL PRODUCTS
OR ISSUES REPORTED?



Complaint Volumes by Product Across Top 5 States



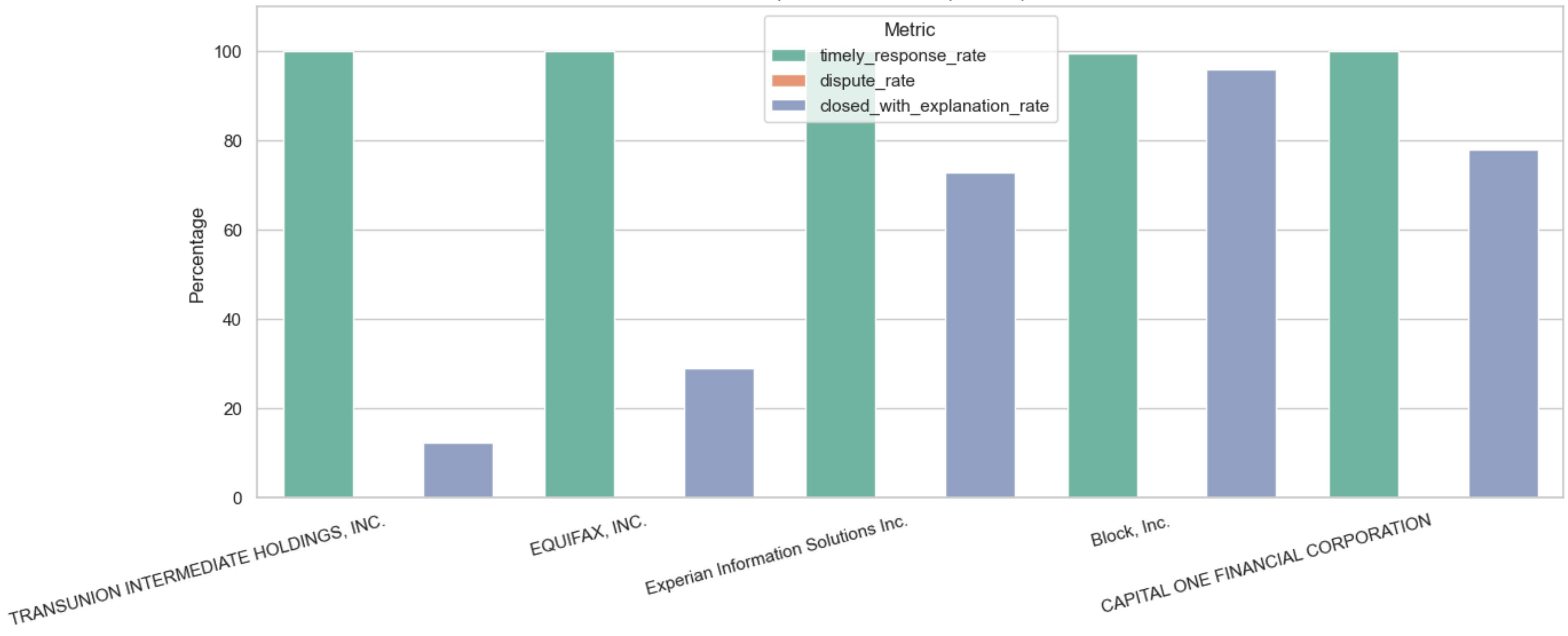
This means...

- NY'S DEBT COLLECTION PROBLEM: NEW YORK HAS THE HIGHEST COMPLAINTS, DOMINATED BY DEBT COLLECTION.
- CA'S CREDIT REPORTING FOCUS: CALIFORNIA'S COMPLAINTS ARE PRIMARILY DRIVEN BY CREDIT REPORTING.
- FL'S DIVERSE COMPLAINTS: FLORIDA SHOWS A MIX OF COMPLAINTS ACROSS CREDIT REPORTING, MORTGAGE, AND STUDENT LOAN.
- TX'S CREDIT CARD CONCENTRATION: TEXAS COMPLAINTS ARE EXCLUSIVELY CREDIT CARD RELATED.
- CLEAR REGIONAL PRODUCT DIFFERENCES: THE CHART CLEARLY SHOWS DIFFERENT FINANCIAL PRODUCT ISSUES ARE DOMINANT IN DIFFERENT STATES.

Question #3

DO CERTAIN COMPANIES HAVE
CONSISTENTLY HIGHER COMPLAINT
RATES OR PATTERNS OF UNRESOLVED
ISSUES, AND HOW DOES THIS CORRELATE
WITH THEIR RESPONSE STRATEGIES (E.G.
TIMELY RESPONSE, DISPUTE HANDLING,
CLOSED WITH EXPLANATION)?

Response Metrics for Top 5 Companies



This means...

- EXCELLENT TIMELY RESPONSES: ALL TOP COMPANIES CONSISTENTLY PROVIDE TIMELY RESPONSES.
- VARIED EXPLANATIONS: MOST COMPANIES HIGHLY UTILIZE "CLOSED WITH EXPLANATION," BUT EXPERIAN DOES SO LESS FREQUENTLY.
- MISSING DISPUTE DATA: THE CHART LACKS VISIBLE DISPUTE RATES, MAKING IT HARD TO ASSESS UNRESOLVED ISSUES.
- TIMELINESS ≠ SATISFACTION: HIGH RESPONSE TIMELINESS DOESN'T GUARANTEE FULL RESOLUTION OR SATISFACTION.

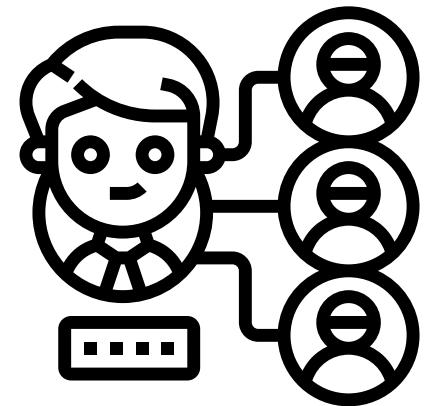
KEY FINDINGS



Top Issues: Credit Reporting & Debt Collection
(incorrect data, false claims).



Regional Differences:
NY leads in debt collection, CA in credit reporting, and TX in credit card issues.



Company Responses:
Companies are timely, but "closed with explanation" rates vary. Dispute data is missing, limiting full satisfaction assessment.



THANK YOU

Any questions?