Consumer Complaint Analysis

Michael Min
Cory Morales
Scott Young
Michael Gilroy

Question

Can we identify companies that may be committing malpractice?

There's plenty of complaints, but who is to blame? Is it the company's negligence, or is it consumers being naive?
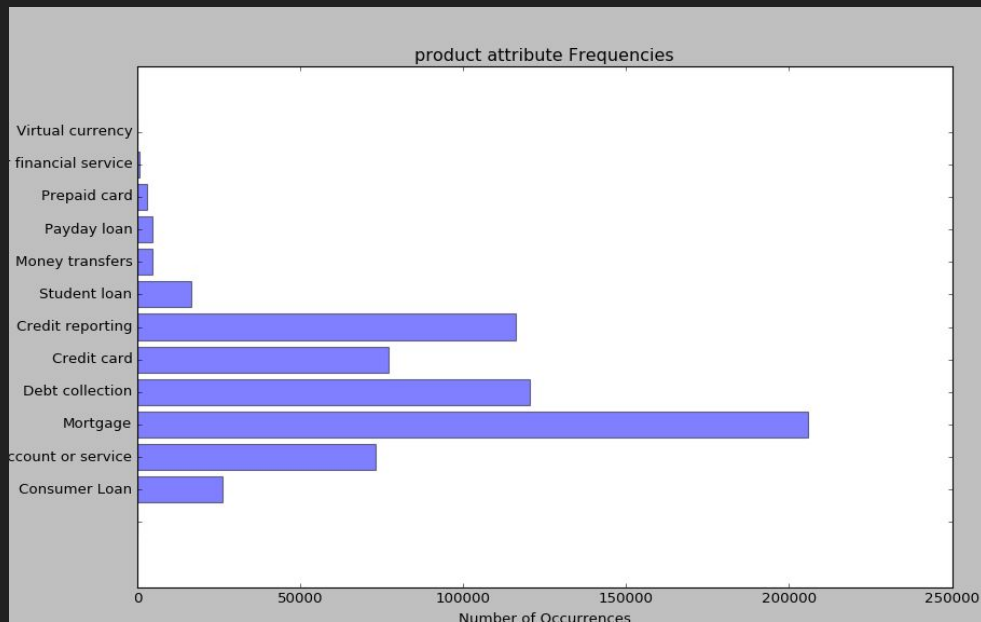
Data Preparation

- Minimal cleaning needed
- Filled field with ERROR and UNKNOWN as appropriate.
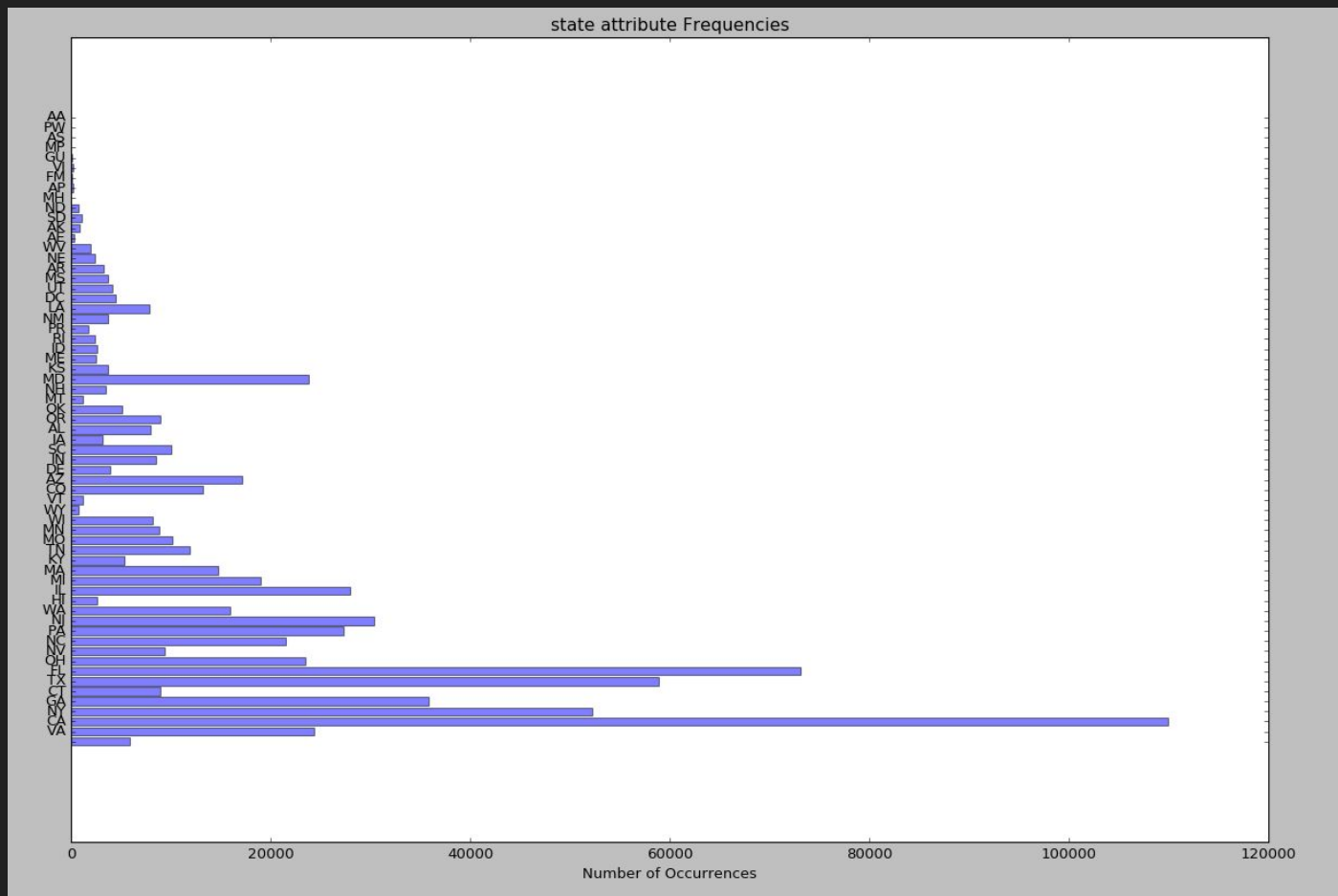- Ignored data that was problematic for a test

Tools

- Python
  - numpy
  - matplotlib
- Jupyter

Narrowing Our Search

- Initial visualizations helped to know what to focus on
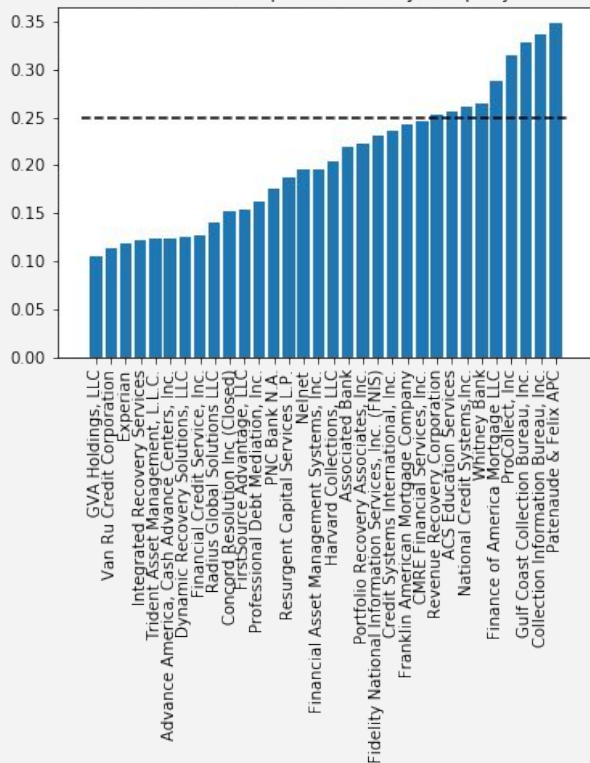- Complaints by location (state/zip code)

Narrowing Our Search
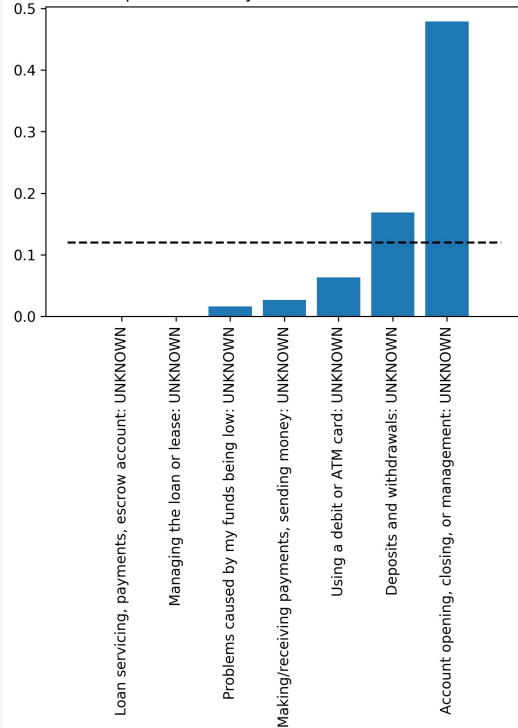

state attribute Frequencies

Analysis

- Identify Companies by complaint dispute rate
- Identify Companies by fraction of disputes for a type of complaint
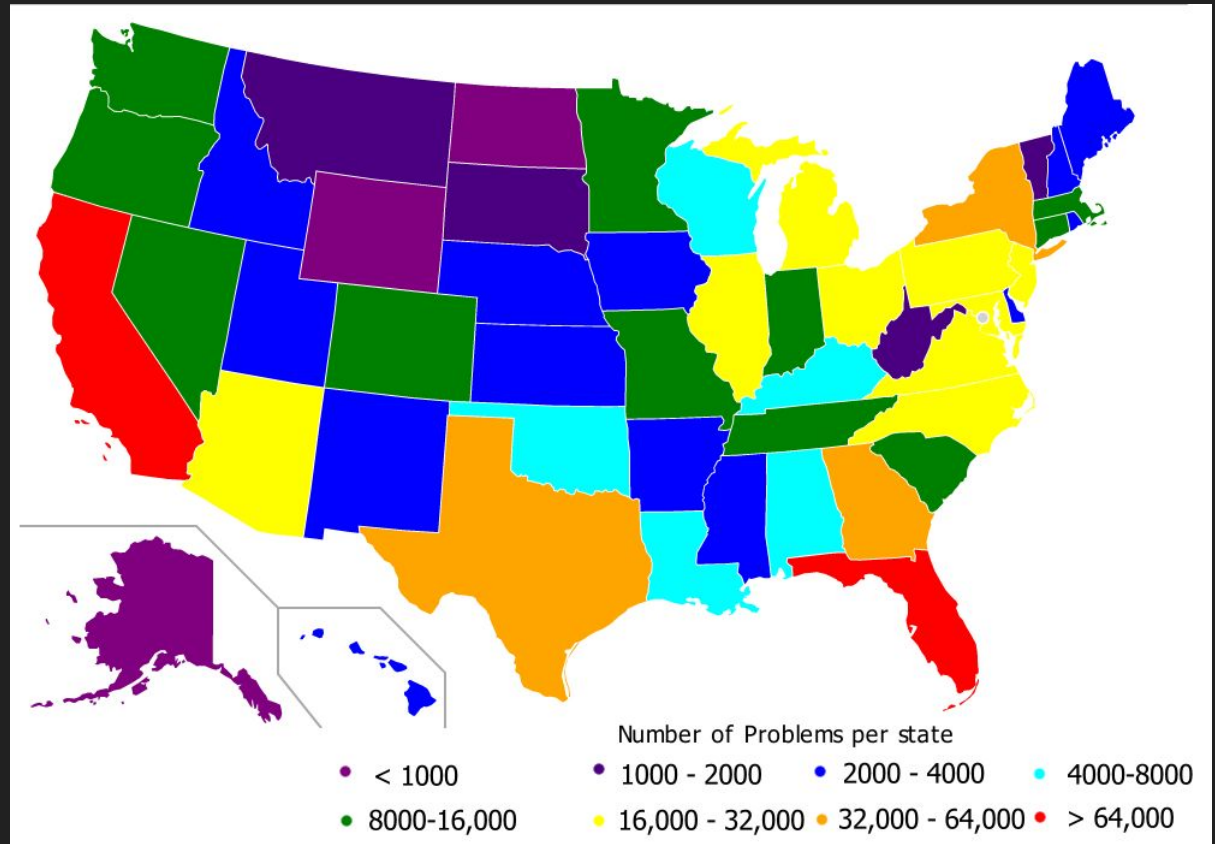

Percent of disputed cases by company

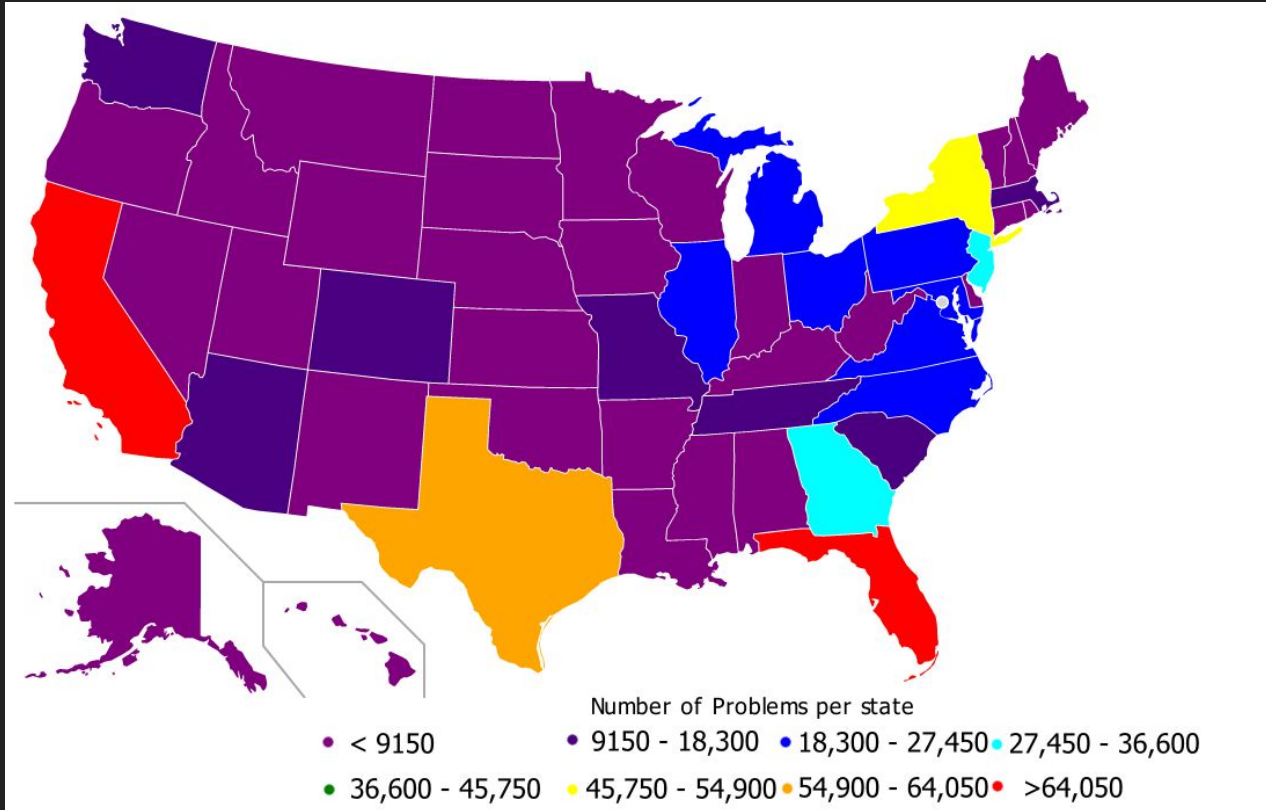Disputed cases by issue for Scottrade Bank

Analysis

- Number of reported problems ranged heavily by state
- Florida and California led the pack
- Alaska, Wyoming, and North Dakota had the fewest complaints

Number of Problems per state

- • < 1000
- • 1000 - 2000
- • 2000 - 4000
- • 4000-8000
- • 8000-16,000
- • 16,000 - 32,000
- • 32,000 - 64,000
- • > 64,000

Analysis

Number of Problems per state

- < 9150
- 9150 - 18,300
- 18,300 - 27,450
- 27,450 - 36,600
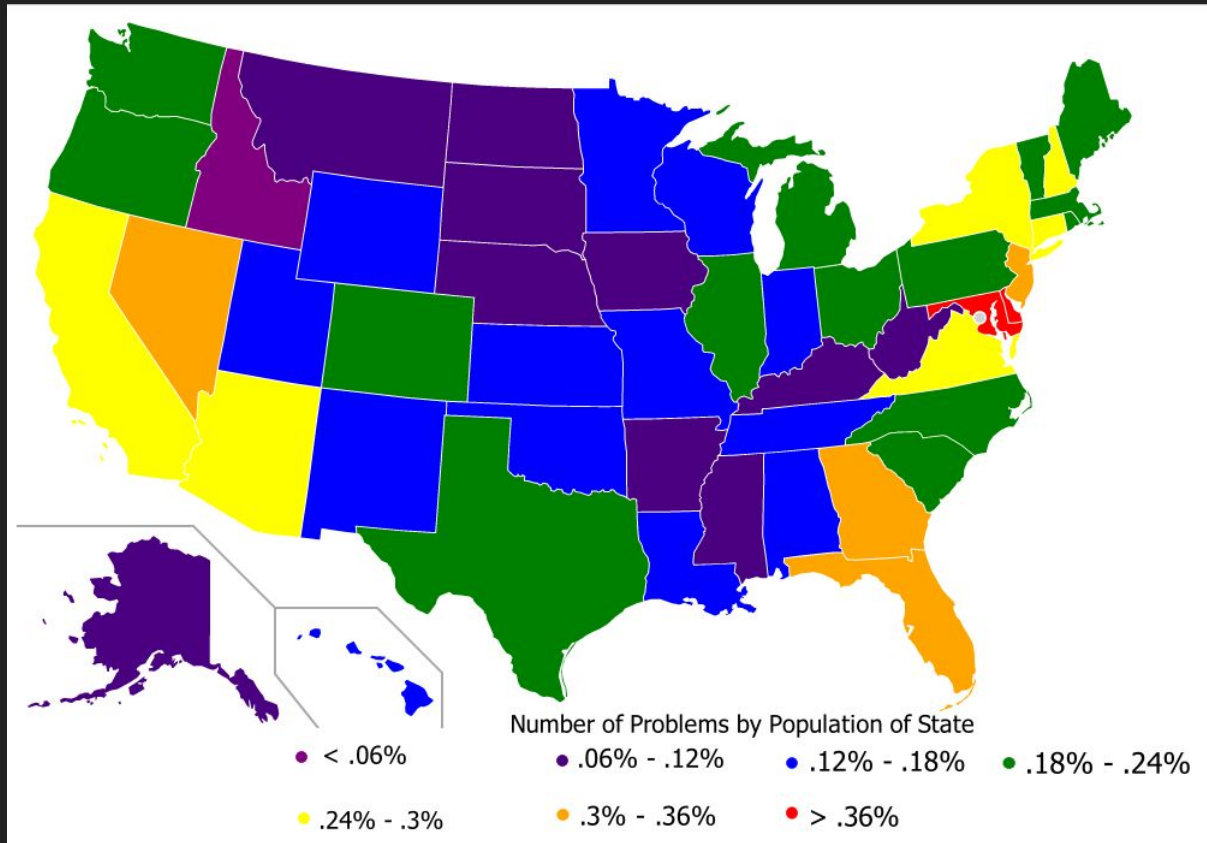- 36,600 - 45,750
- 45,750 - 54,900
- 54,900 - 64,050
- >64,050

Analysis

- When compared to population, the results are noticeably different
- Idaho trails behind the rest of the country
- Maryland and Delaware lead the country

Number of Problems by Population of State

- < .06%
- .06% - .12%
- .12% - .18%
- .18% - .24%
- .24% - .3%
- .3% - .36%
- > .36%

Analysis

Lift Correlation:

- The big correlation type that we used to analyze the data set's issue-company pairs
- Checks to see how commonly a pair appears in relation to the amount of times the individual members appear at all

$$lift(A, B) = \frac{P(A \cup B)}{P(A)P(B)}$$

Issue	Company	Lift
Late Fee	FDIC	13.962
Credit Card Protection/Debt Protection	Square One Financial, LLC	92.879
Credit Reporting	Capital One	8.983
Advertising, marketing, or disclosures	NetSpend Corporation, a TSYS Company	423.112
Arbitration	GAMACHE & MYERS, PC	246.555
Cash Advance Fee	Commerce Bank	17.310

Analysis

Chi-Square:

- Another kind of correlation, equally impactful as lift
- Checks to see the probability of an issue-company combination being mere randomness in the data set

Issue: Late Fee
Average Chi-Square Correlation: 532.135

Company	Chi-Square Correlation
Bank of America	6094.632
Wells Fargo & Company	3775.741
FDIC	3262.525
Alliance Data Card Services	2935.158
Experian	2886.604

Application

- Data can be used for the following:
  - Finding companies to avoid
  - Finding states that hold companies more accountable
  - Finding if more should be done to protect consumers