# AUTOMATING CATEGORY CLASSIFICATION USING ML:
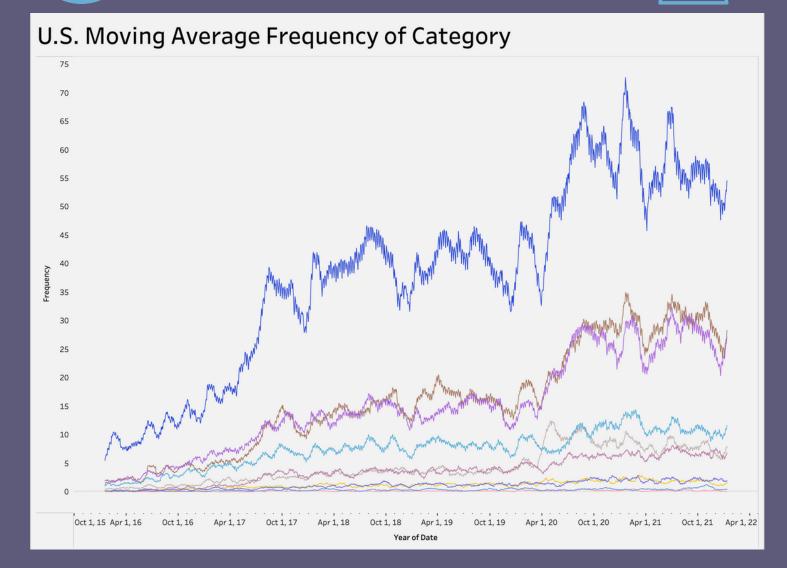## EXPLORATORY DATA ANALYSIS AND CLASSIFICATION METHODS OF ABA PRO-BONO ONLINE SERVICE

DUM
.PY

**Samarth Srinivasa**

**Aditi Behara**

**David Ryan**

**Surabhi Tadvalkar**

**Eliot Lee**

# Allocating the Right Lawyers

## U.S. Moving Average Frequency of Category



### Categories

- Consumer Financial Questions
- Education
- Family and Children
- Health and Disability
- Housing and Homelessness
- Income Maintenance
- Individual Rights
- Juvenile
- Other
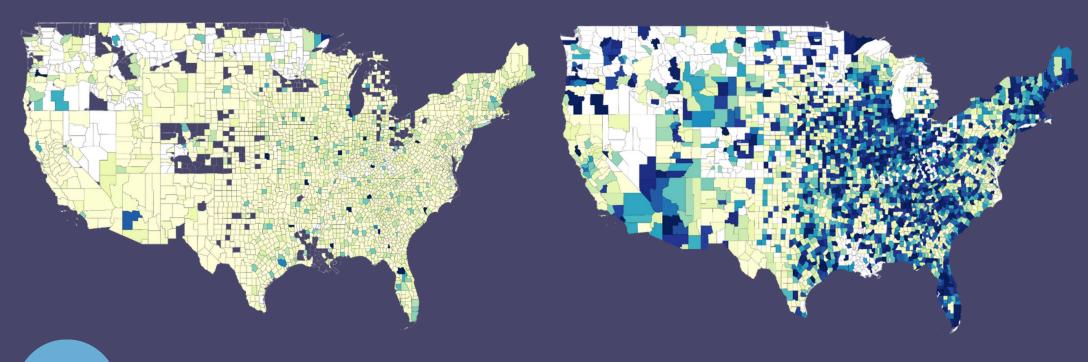- Work, Employment and Unemployment

It's important for the ABA to assign the correct specialty lawyers at the correct times to best assist clients.

We can predict the trends of when each category of lawyers is needed most by graphing the frequency of questions by category.
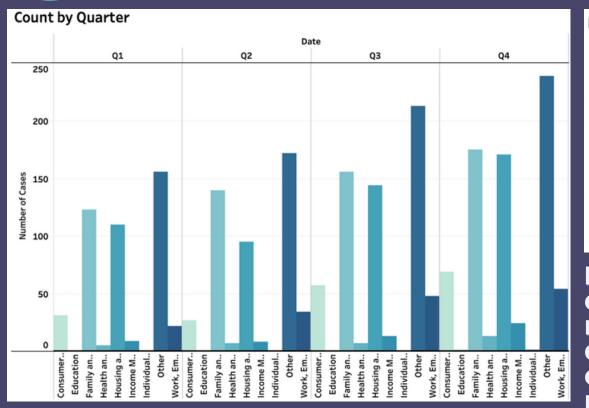
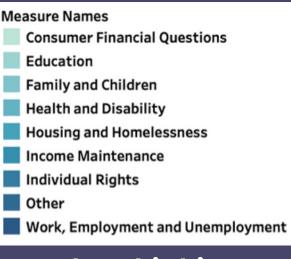# Frequency Of Attorneys & Clients in the Country - By County

## Attorney Cloropleth



## Clients Cloropleth



# Category Frequency Analysis

## Count by Quarter



### Measure Names

- Consumer Financial Questions
- Education
- Family and Children
- Health and Disability
- Housing and Homelessness
- Income Maintenance
- Individual Rights
- Other
- Work, Employment and Unemployment

Based on this histogram visualizing category trends, other miscellaneous cases aside, Family and Children is the most popular category within clients' question posts every quarter within an year.

# Category Predictor

**1** Merged datasets together in order to work with post questions and categories

**2** Removed weird symbols, stock words, and other unnecessary words

**3** Chose the parent question of every unique question identifier

**4** Ran a TDF-IF Vectorizer to give weights to words

**5** Ran a Logistic Regression Model

**Test Accuracy:** 72.17%

**Test Loss:** 27.83%

| | Precision | Recall | | Precision | Recall |
|---|---|---|---|---|---|
| Family and Children | 83% | 91% | Education | 45% | 27% |
| Housing and Homelessness | 79% | 81% | Income Maintenance | 45% | 18% |
| Work, Employment, or Underemployment | 71% | 71% | Individual Rights | 45% | 13% |
| Consumer Financial Questions | 64% | 55% | Health and Disability | 39% | 18% |
| Other | 48% | 52% | Juvenile | 0% | 0% |