Nflverse: Kaiden Kaiser-Barton, Ryan Trachsel, Chris Blake, Sam Fink

Background

We analyzed data provided by the American Bar Association (ABA). The data included information regarding clients, attorneys, meeting times, online conversations and locations. The goal of our analysis is to analyze the data and identify patterns or trends that could be used by the ABA to advise state partners, allocate resources and create outreach strategies for potential users and volunteers. One particular area of interest is helping lawyers connect better with their clients. This can be done in a number of ways, including by using similar language, understanding the context of messages and providing clients with lawyers with an area of expertise that best suits their needs.

Analysis

Our group started by formulating a research question: Are there any trends, gaps or correlations that can be utilized to improve client-attorney relationships at the state level? We decided that focusing on visualization would be a good way to communicate patterns to law professionals and clients. After compiling the data in R, we used Tableau to generate a map showing the client-to-attorney ratio in each state we had data for (38 states).

We decided the best way to explore improvements for client-attorney relationships at the state level was by focusing on statewide demographics. Since we sought to merely find patterns, our analysis was centered around unsupervised learning. For the purpose of our analysis, we created a "prime" data set in which each row was a single state. One variable we included was *Hours* from the "AttorneyTimeEntries" data file. This variable represented the hours spent responding to clients. We computed the average of the *Hours* variable for each state and included it in the data set. We also created our own variable, *client_to_attorney_ratio*, for each state. We thought this would be interesting, as it provides a metric to evaluate each state's resources. Additionally, we included categorical demographic variables derived from the "Clients" file. These included *EthnicIdentity*, *Gender*, *Veteran*, *Imprisoned,* and *AnnualIncome*. We also included *Category* from the "questions.csv" file. For each of these, we organized each entry into subcategories, created a column for each created subcategory, then displayed the count for each subcategory for each state.

We created various maps and plots to visualize the relationships between variables, computing ratios between categories as needed. The state of Texas had the largest number of entries, so we decided to create a more specific series of graphics for the state. We implemented clustering in order to make better sense of the relationship between variables. We constructed a scatter plot matrix for the quantitative (non-count) variables. Focusing on one plot, we made clusters to sort the states into different three income levels: low, medium and high income.

Conclusion

We found that certain categories had statewide trends. There is great value in individual state analysis. Our findings demonstrate that different states have different needs, demographics and resources. There are an infinite number of conclusions that can be drawn depending on specific interests, locations and demand. Using our code and "prime" data set, the ABA can recreate almost any statewide analysis it would like.