Executive Summary   **Team:** The Pirates (Jacob Beaird, Moe Toyoda)

Overview:
As law enforcement regulations vary between different states, it is likely that clients' demands also differ depending on their state of residence. Thus, we chose to focus solely on data in Missouri. This data was obtained from the American Bar Association, which requested that we identify patterns or trends that could be of use to them and their state partners. During our study, our team had two main areas of focus. Firstly, we aimed to understand any relationships between clients' demographic data and question categories. By doing so, we could predict which category clients' questions are most likely to fall into. Secondly, we sought to gain insight into the attorney's perspective by identifying the most trending topics in each category based on question posts. This allowed us to determine which topics were currently the most relevant and in demand.

Analysis:
The first step of our analysis was to understand the association between clients' demographic data and the categories of questions they ask, using "questions.csv" and "clients.csv". To understand this relationship, we created box plots and histograms to visualize the distribution of each demographic factor within each question category. These descriptive data assisted us in selecting significant predictors (NumberInHousehold, Age, AnnualIncome, AllowedIncome, EthnicIdentity, MaritalStatus, Gender) for a random forest classifier that predicts clients' question categories based on their demographic data. New information can be entered into this model to anticipate each client's questions before they even ask.

Furthermore, we conducted an analysis of the top two most experienced attorneys in each category. To obtain their unique identifiers and the categories of questions they answered, we utilized data from the "question.csv" and "attorneys.csv". We then calculated the frequency of question post categories that each attorney answered.

Lastly, for text frequency analysis, we gathered data from "questionposts.csv" and "questions.csv" and combined them to create a dataset with variables for category and question. We converted the text to lowercase and removed any punctuation or stop words (such as "you", "I", "after", "but", "often", etc.). After the text mining process, we calculated the word frequency to create a word cloud plot for each category.

Conclusion:
With help from our client demographic analysis, random forest classifier, and best attorney investigation, the ABA has the power to match any given client with an attorney best suited to answer questions in the client's most likely category. In addition, our text frequency analysis and word cloud generation serve to prepare attorneys for what types of questions are in demand as well as understand the trending topic within each category. Implementation of this analysis will help the ABA increase the accuracy and efficiency of their services in Missouri, and comparable results can be obtained for other states by performing an identical analysis of the selected state's data.