

# Data Challenge and Proprietary Statement

## Proprietary Statement

Data from the American Bar Association are not public. These data are proprietary and are only to be used for the purposes of the American Statistical Association's DataFest.

By using this data, you agree to:

1. During participation in ASA's DataFest, store and manage the data securely and privately. This means you may NOT upload the data to websites that may provide access to the data to anyone other than yourselves. (For example, if using Tableau or Github, make sure privacy settings prevent others from seeing or accessing the data.) If you are not sure, then don't use that service! (Note: Rstudio Cloud does not have access to your data and so you can use it.)
2. Erase all data after your ASA's DataFest participation is complete.
3. Not identify or attempt to identify the information contained in the dataset, nor contact any of the individuals whose information is contained in the dataset.
4. Comply with all applicable U.S. federal and state laws and regulations relating to the maintenance of the dataset, the safeguarding of the confidentiality of the dataset, and the use and disclosure of the dataset.
5. Not publish results of your analysis of the data except that the final products of the competition (video, slide deck, one-page summary) may be displayed on team members' websites and on campus ASA DataFest websites.
6. Not share the data with anyone who is not a participant of ASA DataFest

Finally, please do NOT reveal the source of the data or any features of the data to ANYONE before April 29. This will ensure that all ASA DataFest participants around the globe have the same experience with the data as you. This means to not post anything to social media that might reveal clues, and to make sure that public github repositories are not discoverable (and if they are, that identifying work is removed).

## The Challenge

The American Bar Association provides pro bono (i.e. free of charge) legal services across the United States via an online platform available in some states and territories. The platform allows people who qualify (based on state-based low-income status) to post legal questions and receive legal advice from volunteer lawyers. The ABA would like to anticipate the sorts of legal questions that arise so that they can prepare volunteers to address those questions, better know how and when to recruit lawyers with specific expertise, and know how to advise state partners on general trends they're seeing. This information will help the ABA and their volunteer attorneys better adapt to new laws, societal changes, natural disasters, and other events that may cause an increase (and, in some cases, decrease) in the demand for pro bono legal services on particular legal topics.

Your data contains actual conversations between clients and lawyers (redacted to maintain confidentiality) and basic demographic information about the client. These exchanges have been previously classified into categories and sub-categories by the people posting questions; potentially, specific conversations might be sorted into different categories by platform administrators.

Your challenge is to provide advice to the ABA concerning any patterns or trends you see in these conversations that would help the ABA advise its state partners, create resources to respond to identified patterns, and devise outreach strategies to better reach potential user populations and volunteers. For one example, the ABA would like to help its lawyers better "connect" with clients, which means they should use similar language to describe issues as their clients, should understand the cultural, societal and emotional context of the clients' messages, and should understand how to best listen and engage with clients.

### Data notes

The data you are provided is a full history of all complete conversations from August 5, 2016 through January 24, 2022. There might be some data from earlier periods before the program was fully launched.

A conversation is marked as "closed" at the request of either the client, the attorney, or an administrator. Conversations are automatically marked as "closed" if there has been no communication for 10 days. Conversations may be closed because they have been completed,

or because an administrator has decided it is not appropriate for this web site, or for many other reasons.

If a potential client wishes to use the services, they must provide personal information. This information is used to determine eligibility for participation. Eligibility criteria vary by state (or territory) and depends on income. Not all states may have provided data on eligibility criteria.

In addition:

- Texts of posts in questionposts.csv will have commas in them. Watch out when reading!
- The data include a "state" labeled "US/ABA Federal", which is used for questions concerning immigration or coming from veterans.
- The data do not provide an obvious means for identifying which posts are from the attorney and which from the client.