H-1B Data Analyzer

[CIT 591] Final Project / User Manual

Adi Manjunath

Table of Contents:

1.0 Introduction:

Hi there! Thanks for using H-1B Data Analyzer.

This application reads in publicly available H-1B application data (converted into .csv files) from 2008-2017, and analyzes said data to reveal potential insight into the companies, attorneys, universities, cities, states, and employees/jobs (to name a few stakeholders) that are involved in the lottery process that takes place every April.

You, the user, pick a year, and then pick various questions you'd like to answer about your selected year's data set. You are also given the option to run a simulation of what the visa process would like if all 85,000 visas were handed to the highest bidders (on the basis of wage) as opposed to the luck of the lottery –– and how that affects which cities, states, and companies are more (or less) likely to receive H-1B visas.

First, let's get you set up.

2.0 Setup:

Step 1:

Download all files found on the H-1B Data Analyzer GitHub page (accessible here.)

Step 2:

Download the full datasets for 2008-2017 (accessible here) and replace each file in the test-data folder with the datasets available at the link above. Unfortunately with GitHub's restrictions on a repo's storage size, the test-data folder consists of incomplete data with only 1000 cases per year (as opposed to the 150,000-200,000+ cases that are found in the actual datasets.)

Step 3:

Add all the .jar files found in the lib folder to the libraries in your Java Build Path for whichever SDK you choose to use.
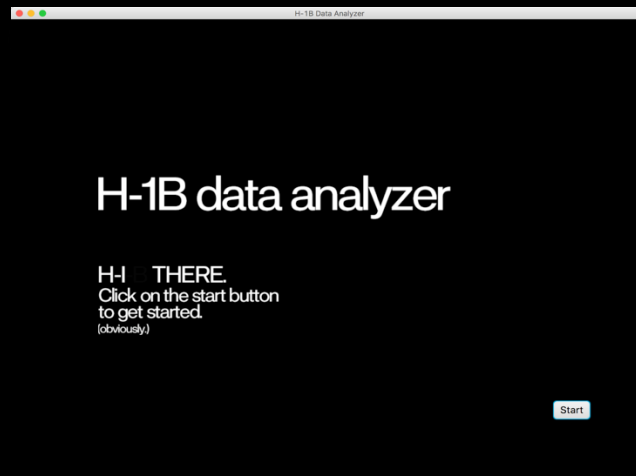
Step 4:

You're all set! Run the application via the H1BDataAnalyzer.java file.

3.0 How to use the application:

3.1 Welcome Page:

This is what you should see when you open the app. Click the start button to get started.

H-1B data analyzer

H-I THERE.
Click on the start button
to get started.
(obviously.)

Start

3.2 Pick Year Page:

Pick the year you'd like to analyze using the drop down box, and then click the "Confirm Year" button.

Pick the year you want to analyze:

2008

Go Back   Confirm Year
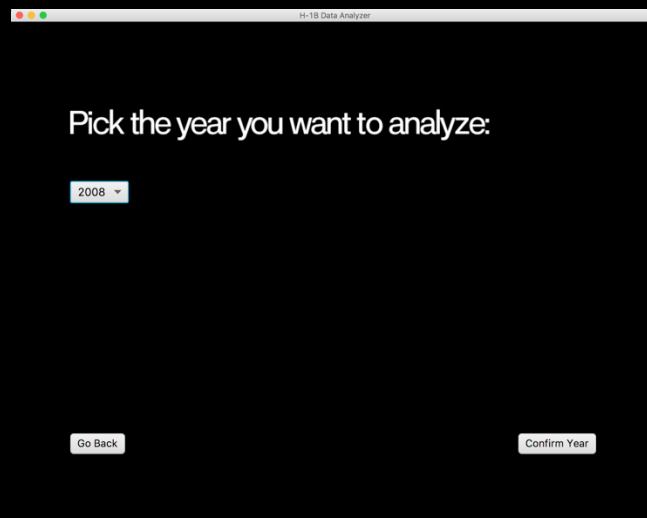
3.3 Pick Questions Page(s):

Done deciding which year you'd like to analyze? Great. One of two pages should pop up, depending on your selection. If you picked a year from 2008-2014, you should see a page that looks like this:

If you picked a year from 2015-2017, you should instead see a page that looks like this:

The different pages show up in response to which years' datasets contain information about the attorneys that filed each H-1B application (2015-2017 does) and which years do not (2008-2014 do not.)

Pick the questions you'd like to get the answer to by checking the corresponding checkbox. Hover over any question to get a brief explanation of what each question refers to. You are free to choose as many as you'd like before clicking the "Confirm Questions" button.

If you pick a question that requires you to pick a city, state, or job, a pop-up window will appear prompting you to make a choice. Make your selection, and then click the "Confirm" button to exit out of the pop-up window.

Click the "Confirm Questions" button when you're ready to see your answers. If you have not selected any questions, a pop-up window will appear prompting you to select a question.

3.4 Answers page:

The answers to your selected questions should show up on a window that looks like this:
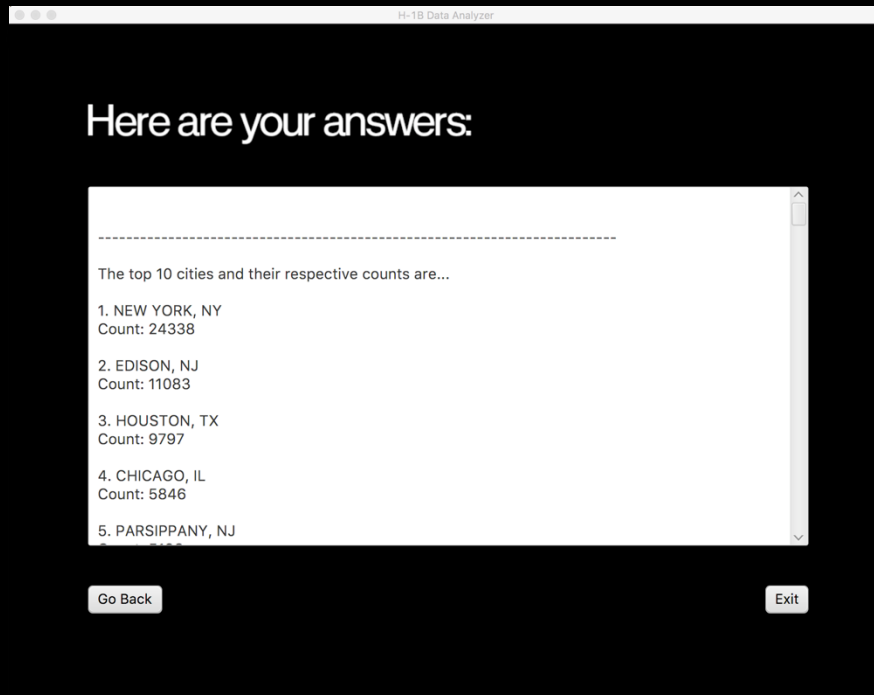
Scroll down within the text area to view all your answers.

At this point, you are free to either go back and choose different questions (or go back further to choose a different year), or to exit the program using the "Exit" button.

4.0 Questions: Explained!

Having a hard time understanding what each question means? Try hovering over the question to get a brief explanation of what each question deals with.

Still having trouble? Worry not; here's a (slightly more in-depth) explanation to each question.

4.1   Top Cities (by # of applications):

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 H-1B employer cities. The list contains each city's name + state, and their respective count of applications for the selected year's data.

4.2   Top States (by # of applications):

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 H-1B employer states. The list contains each state, and their respective count of applications for the selected year's data.

4.3   Household vs H-1B Income (by state):

This question parses through census data to load the average household income of a state (that you, the user, selects), and then parses through the selected year's data to find all H-1B cases from the same state to calculate the average wage offered to H-1B candidates in your selected state. The final answer should display the difference between the average household income, and the average H-1B wage.

4.4   Top Wage Difference Cities:

Quick explainer for the terminology used below: "prevailing wage" is the submitted wage amount in each case for what the average pay for the H-1B candidate's job is in their worksite city. H-1B applications generally require the wage offered to exceed the "prevailing wage", and the difference between the two can vary depending on the candidate's worksite location.

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 cities where the average difference between the "prevailing wage" and the wage offered to H-1B candidates are the highest. The list contains each job, and their respective wage difference for the selected year's data. This potentially gives insight into how much more certain companies / businesses need to pay to get top talent based on where they're based.

4.5   State vs H-1B Population %:

This question parses through census data to load the percentage of your selected state (compared to the national population), then parses through the selected year's data to load the H-1B application percentage by employer state, and then compare the two. This potentially gives insight into how much location factors in where a majority of H-1B jobs are, and which states over/under-index in terms of H-1B jobs.

4.6   Top H-1B Jobs (by city):

This question asks you to pick a city from a predefined list. The question then parses through the selected year's data and returns a list of (based on the number of applications) the top 10 H-1B jobs in your selected city. The list contains each job, and their respective count of applications for your selected city in the selected year's data. This potentially gives insight into what cities are hubs for what industries.

4.7   Avg. Wage Difference (by job):

Quick explainer for the terminology used below: "prevailing wage" is the submitted wage amount in each case for what the average pay for the H-1B candidate's job is in their worksite city. H-1B applications generally require the wage offered to exceed the "prevailing wage", and the difference between the two can vary depending on the candidate's worksite location.

This question asks you to pick a job from a predefined list. The question then parses through the selected year's data, keeping track of the average wage offered and the average prevailing wage, and then returns the difference between the two (along with both averages.) This potentially gives insight into what H-1B jobs on average require a higher wage offered to H-1B candidates compared to regular candidates.

4.8   Top H-1B Jobs:

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 jobs for your selected year's cohort of H-1B candidates. The list contains each job, and their respective count of applications for the selected year's data.

4.9   Top C-Suite Cities:

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 cities for all "C-Suite" (CFO, CEO, CMO, COO, CTO, to name

a few) jobs. The list contains each city's name + state, and their respective count of applications for the selected year's data.

4.10   Average C-Suite Pay:

This question parses through the selected year's data and returns the average wage for specific C-Suite jobs, and the number of applications filed for each of those jobs for the selected year's data.

4.11   Top Universities (by # of applications):

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 universities for H-1B candidates. The list contains each university, and their respective count of applications for the selected year's data.

4.12   Top University H-1B Jobs:

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 jobs for H-1B candidates at universities. The list contains each job, and their respective count of applications for the selected year's data.

4.13   Top Attorneys (by # of applications):

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 attorneys for H-1B candidates. The list contains each attorney's name, and their respective count of applications for the selected year's data.

4.14   Top Attorney Cities (by # of applications):

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 cities that attorneys are based in for H-1B candidates. The list contains each city, and their respective count of applications that were submitted by attorneys in said city for the selected year's data.

4.15   Top Attorney States (by # of applications):

This question parses through the selected year's data and returns a list of (based on the number of applications) the top 10 states that attorneys are based in for H-1B candidates. The list contains each state, and their respective count of applications that were submitted by attorneys in said state for the selected year's data.

4.16   Run H-1B Allocation Simulation:

This question runs a simulation and then answers multiple questions pertinent to what would happen if the yearly quota of 85,000 visas was not allocated via lottery (as it is currently, but instead to the 85,000 highest bidders based on the wage offered to each candidate. This analyzes information about the 85,000 that get their visas, and information about the remaining who do not –– for instance, where are the 85,000 winning cases based out of, and where are the rest of the losing cases based out of? What is the average, highest and lowest wage offered amongst the winning cases? Which cases make up the top 25 cases that make it (given they are extremely likely to be high-profile seven-figure wage jobs.) Answers to all of these questions (and a few more) are returned when you run this simulation.

Adi Manjunath

MCIT Candidate @
University of Pennsylvania
@adi.txt / work@adithya.co