

## Predicting the Likelihood of Acceptance for an H1B LCA application

An H1B visa is a non-immigrant visa that allows foreign born workers to enter the United States and work temporarily for up to three years with a possibility of extension for up to 6 years. In order for a worker to receive an H1B visa, an employer must offer them a position and then submit an H1B visa application with the Department of Immigration. H1B visas are commonly applied to by international students who are looking to work after completing their education in the United States.

A preliminary step before an H1B application can be filed is the submission of the LCA (Labor Condition Application) to the Department of Labor. The LCA contains information about the job title being offered, duration of the job, whether job is full time, rate of pay, location of the job, and the prevailing wage in the area. The purpose of the LCA is to bind the employer to agree to pay a fair wage and provide benefits to a foreign born worker that are equal to the prevailing wage and benefits at that occupation's location. In addition, the employer should provide the same working conditions to H1B applicants as other workers under the LCA application.

The dataset contains two years worth of data (2015-2016) and approximately 1 million records of H1B LCA application results. I want to prevent employers from having their LCA application denied by identifying and communicating to the employer what the employee/employment characteristics are for a high acceptance rate and low acceptance rate. In addition, I will be able to predict whether an employer's prospective employee will have the LCA approved so that the employer can focus their time on hiring people that will have a certified LCA.

My client is any company that wishes to hire a foreign born worker. The client should care about H1B LCA outcomes because they do not want to spend time filing and looking for candidates that will eventually not be certified with a successful H1B LCA application. Based on my analysis, the employer will be able to find workers that are likely to be H1B LCA certified by looking for candidates with characteristics that have a high H1B LCA certification rate.

In order to find out which characteristics are prevalent in successful H1B LCA applications, I will find out which employers have the most certified applications and the highest rate of applications certified. In addition, I will find out which occupations, job titles, and cities have the best certification rates. Finally, I will determine whether full time or part time employment affects applications as well as if the prevailing wage for the position and year of filing affects application outcome. Once the probability of

receiving an approved H1B LCA application is found for every relevant characteristic, a model can be made that weights characteristics depending on their correlation to application approval. If characteristic has a strong correlation with with application approval, then it will receive a large weight and therefore be more influential in the model predicting whether the application will be approved. For example, the relationship between prevailing wage and application approval can be found by graphing prevailing wage as the independent variable and the application approval as a dependent variable represented by 0 or 1 (rejected or approved). Then a correlation coefficient is obtained to determine how well the independent variable can predict the application approval. This coefficient will be used in a machine learning model.

In addition to application approval prediction, I want to do some EDA by asking various questions about data science positions in relation to H1B LCA applications. I want know which industries file for the most H1B LCA applications for a data scientist position. I also want to know which employers file the most petitions. Also, I want to know how does the wage for data science positions vary by location and employer. Finally, figuring out where data scientist positions are concentrated by the H1B LCA application data is useful to knowing where to look for open data related jobs.

I am using a dataset found here: <https://www.kaggle.com/nsharan/h-1b-visa>. The dataset is a csv file with six fields and approximately 1 million rows. The csv file can simply be downloaded and opened in Excel.

I am planning to include code and a paper as part of the deliverables for this project.

The proposal should address the following questions:

- What is the problem you want to solve?
- Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?
- What data are you using? How will you acquire the data?
- Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.
- What are your deliverables? Typically, this includes code, a paper, or a slide deck.

