Five Years Trends of Employment of Foreigners

Hang Zhao
Yimin Zhu

April 18, 2019

1 Background

The H-1B visa is a non-immigrant visa that allows foreign workers to be legally employed by employers in the United States. It is designed for workers in occupations that require the theoretical and practical application and a higher education in the specific specialty [1]. This is also the most common visa status applied for and held by international students once they graduate and begin to work in a full-time position. International students can be temporarily employed using Optional Practical Training (OPT) provided by their F-1 student visa, another non-immigrant visa for foreigners to study in the U.S. The period of OPT is three years maximum including one year for all majors and two addition years extension if the earned degree is in certain science, technology, engineering and math (STEM) fields. Hence, H-1B visa is mandatory for international students if they would like to stay employed in the U.S. longer.

Another way to obtain labor certification is to apply for Program Electronic Review Management (PERM). PERM and H-1B are both work visa. The major different is that PERM is the first step of the immigration visa process while H-1B is a non-immigrant visa. A prerequisite of PERM is to obtain the Prevailing Wage Determination (PWD) from the State Workforce Agency (SWA). PERM will only be granted if they pass the prevailing wage. The prevailing wage is the average wage in the requested occupation in area of intended employment [2].

2 Summary

2.1 Motivation

This project aims at investigating the last five fiscal years of work visa related program data to obtain better understanding of the recent employment status for foreigner workers. Compared to other projects [3, 4] that perform data analysis on H-1B data in the past, our project only note includes the most up-to-date data of fiscal year 2018, but also merges data set of PERM and PWD to achieve a more comprehensive understanding . Other than that, our project is unique in that it provides interactivity that is implemented utilizing D3 to make the visualization intuitive.

2.2 Hypothesis

Salary Factors

During job search, we will be asked about our salary expectation. Such kind of question is hard to answer due to the lack of knowledge about the salary range of that particular company. To make better estimation of salary, we would like to visualize the correlation between salary and other features to figure out the most significant factors affecting salary and how significant they are. Factors such as job categories, geological locations and size of company are expected to be major components.

Employment Summary

To make decision to accept an offer, it takes time to evaluate the employer. For instance, considering H-1B visa, the quality of a company can be measure by number of sponsorship, rate of of H-1B petition denial, and average salary. Visualization of the top companies with their indexes of quality is expected to give a brief overview of the employer.

Temporal Change

Another focus is the effect on job market caused by policy changes. Since the project utilizing datasets over five years, it can provide the chance of dynamic changes of work visas. We are interested in investigating changes on average salary, education, and top companies. The reason is that if you see a higher salary, higher education of the applicants, it implies more strict policies. We also expect that the more well-known the company is, the lower denial rate it will have.

3 Project Description

3.1 Tools

Language Python 3, HTML5, CSS, JavaScript

Library D3, Bootstrap, Pandas, NumPy

3.2 Data Preparation

We would like to investigate on of three work visa related programs: Labor Condition Application (LCA) Programs (H-1B, H-1B1, E-3), PWD, and PERM Program. For each program, we select five data sets of fiscal year from 2014 to 2018. Table 1 displays the dimensions for all data sets. The data sets are disclosed by the Office of Foreign Labor Certification (OFLC) [5].

3.3 Data Cleansing

Merging Datasets for each Program Rename the column to resolve inconsistency through different fiscal years. Preserve the intersection of columns to merge data sets of the same program. A new feature Year is generated after merging.

Table 1: Data Set Summary

	Fiscal Year	2014	2015	2016	2017	2018
PERM	Features	27	125	125	125	125
	Data Points	70876	89049	126143	97603	119776
LCA	Features	35	40	40	52	52
	Data Points	519504	615491	637714	624650	654360
PWD	Features	33	86	56	57	57
	Data Points	131999	138949	140940	182289	149409

Feature selection Consider the features that is salary is dependent on as significant. For example, the expected significant features for LCA program are case status, job title, wage rate, unit of wage, location of work site, name of employer, and number of works in the company.

Excluding Missing Values Drop out the data points with missing values or filled with nah for significant features. The reason we decide to drop out instead of filling in is that the data set is very large so that the drop proportion is small enough to be ignored.

Subsetting Datasets For example, considering the common possible visa type for international students. PERM subset should only include data points that the class of admission is F-1 and H-1B and LCA subset should only includes data points that the visa type is H-1B, and .

Fuse Datasets of Different Programs Datasets can be fused to find the data set of certified H-1B visas that is than approved for PERM. As the privacy issues are raised, we will decide whether to implement this data set later.

4 Project Description

4.1 Salary Factors

Method

Covariance, Correlation Matrix

Visualization

Scree Plot, Scatter Plot, Scatter Plot Matrix

Approach

Scatter plot is the most intuitive method to show relationship between two variables. There are more than ten features other than salary so that displaying scatter plot of salary and all other features will be redundant. Principal component analysis (PCA) can reduce dimensions to simplify the analysis while the principle components ordered by eigenvalues are hard to explain. As we would like to make the project intuitive, we decide to not apply PCA and keep the original dimension. We would like to apply scree plot to show the features in the order of variance and select some top features to be applied on scatter plot matrix to show the distribution of correlation.

4.2 Employment Summary

Method

K-means, MDS, PCA

Visualization

Scatter Plot, Bar Chart, Pie Chart, Grid Map, Biplot

Approach

Pie chart is a impressive way to show the proportion so it will be utilized to display the denial rate. For other measure of employer quality, it can be binned to display the summation of that factor on bar chart. Grid map is considered as well to display the average quality base on geography information. The recommended fields is visualized by K-means that clustering the relatively high qualities employers using scatter plot. Biplot is also a considered using top quality measure factors as basis to project the employment in different fields .

4.3 Temporal Change

Visualization

Line Chart

Approach

Using time as the horizontal axis and the feature we look into as vertical axis, it can show the data trends clearly through time. We can also figure out the peak and bottom easily. Another interesting point on the data set is that it contains the date January 20, 2017 when Trump become the president.A mainstream idea is that Trump raised more strict policies on immigration that increase the difficulty on obtaining a work visa. We would like to split the data set to two parts and visualize the difference between them to check if the idea is true.

5 Conclusion

Using the mentioned methods above to visualize work visa related data set, we would like to show a summary of foreigner employment situation. The trend of policy effect and the strategy people would like to deal with this situation based on the statistic analysis and thus to avoid potential risk ahead of us. By understanding the factors of salary and the relationship of salary with different locations and different occupations, we can have a proper estimation and get a better chance to win a salary negotiation and choose a suitable working environment for us.

References

[1] USCIS. *H-1B Specialty Occupations, DOD Cooperative Research and Development Project Workers, and Fashion Models*. Mar. 19, 2019. URL: https://www.uscis.gov/working-united-states/temporary-workers/h-1b-specialty-occupations-dod-cooperative-research-and-development-project-workers-and-fashion-models (visited on 04/16/2019).

[2] USCIS. *Foreign Labor Certification*. URL: https://www.foreignlaborcert.doleta.gov/pwscreens.cfm (visited on 04/16/2019).

[3] Aksanand. *H1B data analysis*. 2017. URL: https://www.kaggle.com/anandakshay44/h1b-data-analysis/data (visited on 04/16/2019).

[4] Sharan Narbole. *H-1B Visa Petitions Exploratory Data Analysis*. URL: https://nycdatascience.com/blog/student-works/h-1b-visa-petitions-exploratory-data-analysis/ (visited on 04/16/2019).

[5] OFLC. *OFLC Performance Data*. Feb. 11, 2019. URL: https://www.foreignlaborcert.doleta.gov/performancedata.cfm#dis (visited on 04/16/2019).