H1-B petition Exploratory Data Analysis

June 18, 2018

Team Number : 12
Name : Andres Salcedo Noguera
Name : Sheldon Rodrigues

1 H1-B petition Exploratory Data Analysis

1.1 Background

The H-1B Visa in the United States is an employment-based, non-immigrant visa that allows US employers to employ foreign workers in specialty occupations.
For a foreign national to apply for H-1B visa, it is required that:
1) A US employer offers them a job
2) After this, the employer should submit a petition for a H-1B visa to the US immigration department.
This is the most common visa status applied for and held by international students once they complete college/higher education (Masters, PhD) and wish to work in a full-time position.

1.2 Motivation

Over past few months, President Trump has proposed several policy changes in the implementation of H1-B visa. Some of these include an increase in the minimum wage of employees who will qualify for the H-1B visa and setting a cap on certain companies that file petitions. As international students, we are concerned of our future post-graduation and hence through the exploratory analysis in this project we aim to gain further insight on the H1-B visas in aspects relating jobs, positions, and locations. In line with this, this analysis will point to interesting trends and association in the data that might help potential international students in deciding about their future in the U.S.

1.3 Data

The dataset we used comes the from instituttional records relating H1B visa petitions between 2014 and 2017 filed at the U.S. Department of Labor. The data specifically comes from the Labor Condition Application (LCA) filings which is a mandatory requirement for applying for an H1B visa. Each row of the dataset represents a visa petition. For our analysis we will be using the following fields (among the 40+ that are available):

The raw data was obtained from https://www.foreignlaborcert.doleta.gov/performancedata.cfm and processed in Python using the Pandas library. These steps consisted of selecting the desired

fields for the analysis (shown above) and the creation of some variables that will be used for merging complementary information. In particular, for this document we used complementary information from the U.S Census Bureau (state population), IRS (state taxes), and the Bureau of Labor Statistics (native-born unemployment). After running the code below, we imported the resulting data to Tableau and internally applied the required join operations for having the final data. For purposes of the analysis, we gruped application status in two categories: certified and not certified (the fist one referring to a successful application that then is subject to the lottery).

Following on, it should be noted that the interest variable in our dataset (case status being certified) does not actually translate into an H1B visa. Rather that this, having a certified status relates to the eligibility that an international worker has for obtaining the migratory status. As such, it is the first step to pass into getting an H1B visa and a more interesting in our opinion (i.e., as visa applications have historically been greater than the quota, final selection is done through a lottery allocation).

1.4 Data Cleaning and Processing

The code we used to perform our data cleaning steps is attached in the "Data Cleaning.ipnyb" jupyter Notebook.

1.5 Methods and Assumptions

For this analysis we mainly relayed in visualization techniques using Tableau. Prior to this, the required data was processed in Python and exported as a csv file. Once it was imported to Tableau, we performed four join operations with each of our interest datasets: state taxes, native-born unemployment by region, Fortune 500 company locations, and state population. With this completed, we focused in exploring different variables based on what we had stated in our project proposal. In line with this, the analysis presented below nurtures on our initial intent and exceeds by including other interesting policy implications. In doing this, however, caution should be taken since the data has important limitations.

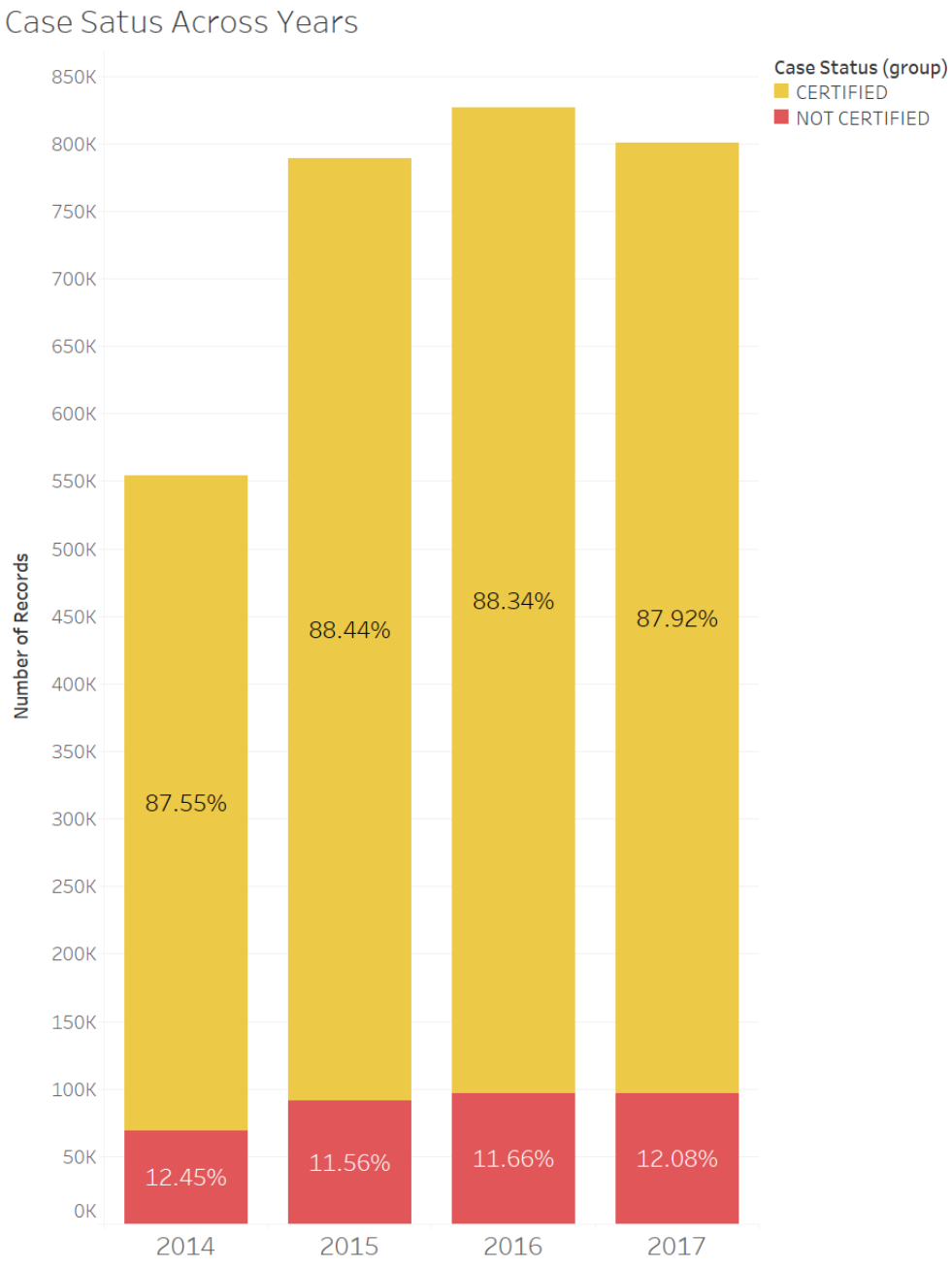
One the one hand, it refers to petitions (and not visa approvals), which might not necessarily correspond to the reality of the people that stay in the end. As such, our analysis is based on the assumption that the approval process, by functioning through a lottery assignment, will represent a similar distribution than the one we use in our data. On the other hand, we also relay in the quality of the data. Although it is clean in some fields others, like employer or job title, are messy and this difficults conducting a more thorough analysis. For example, when analyzed by organization Deloitte has more than 50 different references within the pooled dataset! Results would have probably been different by using standardized fields.

1.6 Analysis

1.6.1 Descriptive Analysis

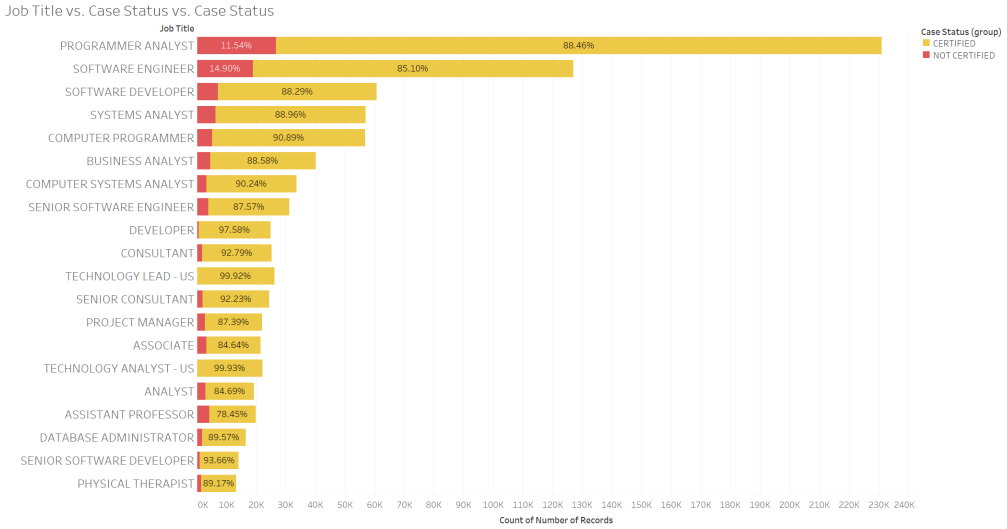
Our analysis started by analyzing the temporal pattern of the applications for the years anlayzed (2014-2017). From the graph below we can see that there was a considerable increase of applications in 2015 but, despite this increase, the proportion of certified petitions remained constant. In addition to this, we also analyzed the main job titles under which inmigrantes are applying for H1B visas (which is shown below the case status graph). In this case we see that a significant portion of the filed petitions came with IT related jobs. Notwithstanding this, we also see professors and health-related jobs among the most popular.

Satus Across Years.bb



image

Title vs. Case Status vs. Case Status.bb



image

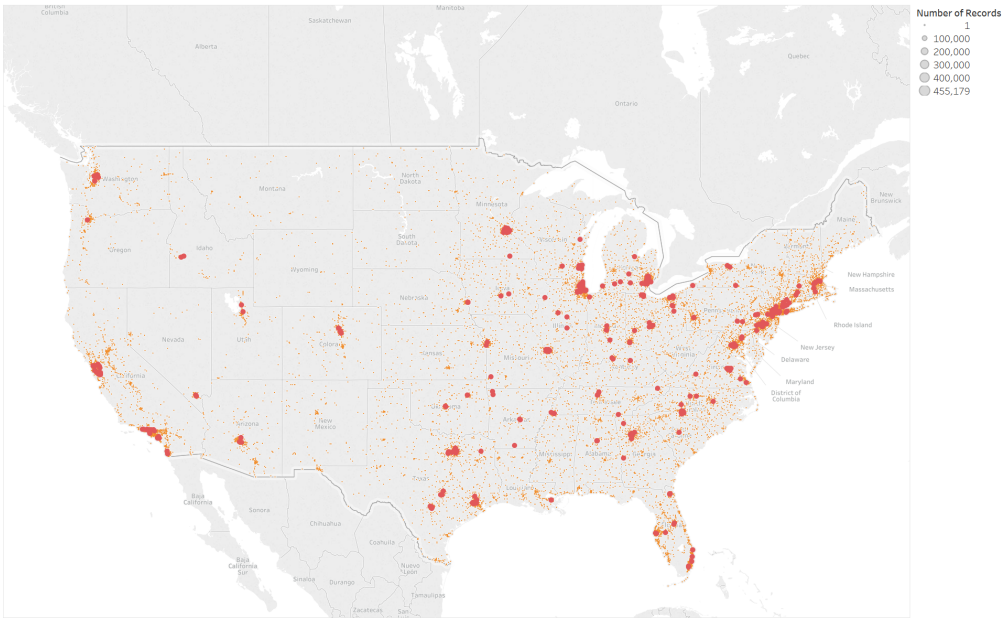
In relation to the geographical distribution of certified applications, we also looked if there was any relation between the worksite locations and those of the Fortune 500 corporations (showed in red points). The map produced below was obtained by mapping the centroid of the worksite's zip code. From it, we can see that the relationship of applications is not homogeneous for all organizations. One the one hand, we see that in places like Seattle, San Francisco, Chicago, and the Northeast coast the congrence of applications seems closely tied with the prescence of large organizations. However, there are also sparse relationships like in Indiana, Kansas, and Nebraska which prove this high association wrong.

Another interesting relationship that emerged from the previous map was whether some job categories had a higher application prevalence in remote areas. In particular, this analysis emerged considering that international physicians require to work for a number of years in small towns after completing their degrees (i.e., a waiver). In line with this hypothesis, we mapped medicine-related jobs to analyze if these were qualitatively more disperse than the entire set of applications. As can be seen (points maintain the same scale), these jobs seem to have a relatively high proportion of applications in some of the least popular states (e.g., New Mexico, Maine, and Iowa). Moreover, we also see that the applications to the popular cities of the west coast is vastly reduced.

Another important geographic feature to analyze is the average salary. Considering the inherent differences that exist between states, we decided to analyze how the wages behave across the range of continental states. The graph below depicts this relation, in addition to showing the total amount of certified petitions. From it we can see that in general higher wages are not necessarily related with higher applications (though it is the case for some of the top-paying states like California, Texas, and New York). In particular, we see cases like Illinois (high applications and lower relative salary) and Colorado (high relative salary but low applications) which seem to invalidate a possible association (meaning that people will apply to the places with higher wages).

Another important variable that was present in our dataset was related to the organizations that have a higher prevalence in sponsoring H1B and, in addition, what was the proportion of certified petitions. The graph below presents this in formation and from it we can get two main conclusions. The first is that the top two sponsorig organizations (Infosys and Tata) are Indian companies. Despite not having the data to corroborate this, it would be interesting to analyze

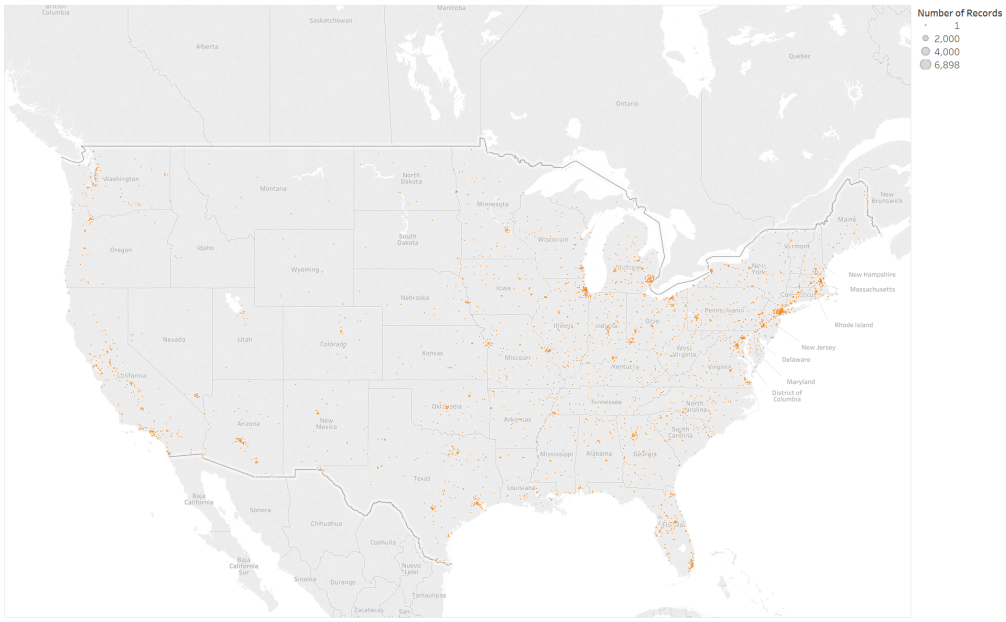
Jobs vs. Fortune 500



vs. Fortune 500.bb

image

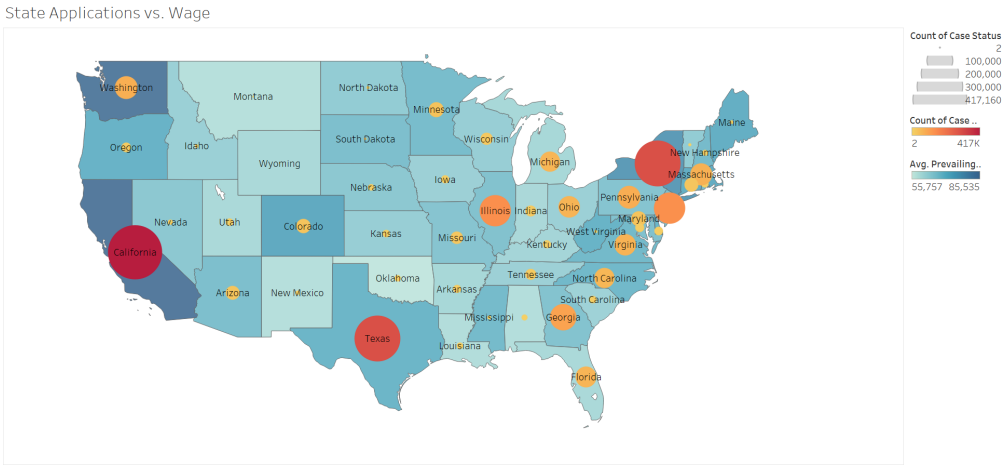
Remote Jobs



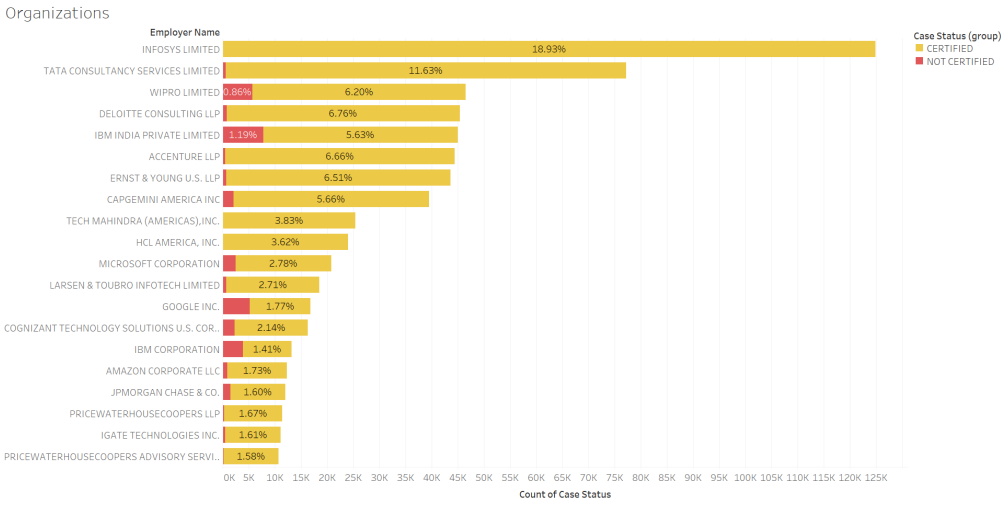
Jobs.bb

image

Applications vs. Wage.bb



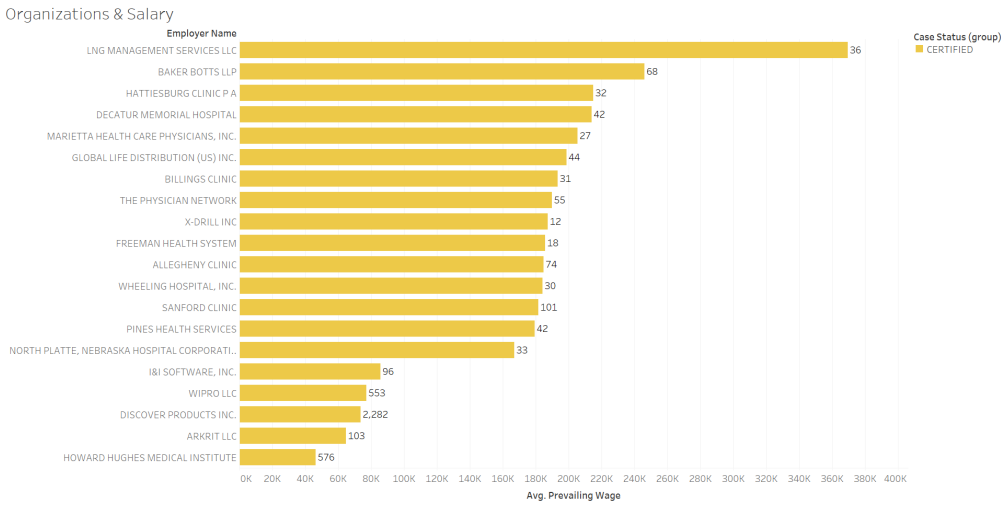
image



image

what proportion of sponsorees are Indians vs. other nationalities. In addition to this, we can see that the prevalence of not certified petitions changes depending on the organization. For example, Infosys has a very low proportion of not certified petitions vs. organizations like IBM India or Wipro, which exhibit a higher prevalence of these cases. In relationship to this information, it would have been interesting to see if the law firms that these organizations use is different and might explain the differences observed in the graph.

When we compare the previous graph with that of the top paying organizations we observe that there is little congruency between both (although there are exceptions like Wipro). For obtaining the graph below, we filtered for organizations that had at least thirty petitions for H1B visas. Otherwise, we obtained a significant number of outliers that were related to particular cases within small organizations (that did not sponsor many candidates). Under these considerations, the graph below shows that, in general, healthcare organizations pay higher salaries which is expected. However, they are also highly selective for pettitioning H1B visas which probably is explained by taylored requirements related to a particular field of medicine. Despite the concen-



image

tration of this type of organizations, we also see the prescense of law firms (Baker Botts), oil and gas (X-Drill), consulting (Arkrit), and IT (Wipro).

1.6.2 What-if Analysis

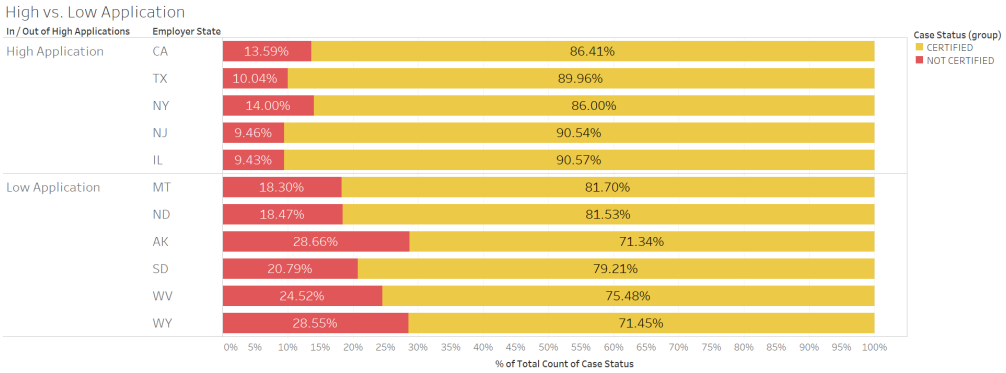
To some extent, our dataset allows us to phrase questions in terms of the naive likelihood of having a certified petition. In line with this, we used the observed outcome in the data to make interesting questions regarding the feasibility of being certified as different factors are considered.

Firstly, we analyzed whether applying to low-popularity states might increase the chances of having a certified status. In terms of migratory policy, this represents an interesting analysis that might allow policymakers to attract human capital in the most needed areas (assuming of course that there is demand for skilled labor). The graph below shows the prevalence of certified petition in the top and bottom five states among the continental states (i.e., excluding autonomous territories). The results are surprising into some extent; least popular states have in general a higher prevalence of not certified petitions! Following this, if the U.S. government is interested in attracting more people to these states then an approach could be to save a portion of the visa quotas to them (or at least detecting why the high prevalence is occuring and if there is a way to reduce it).

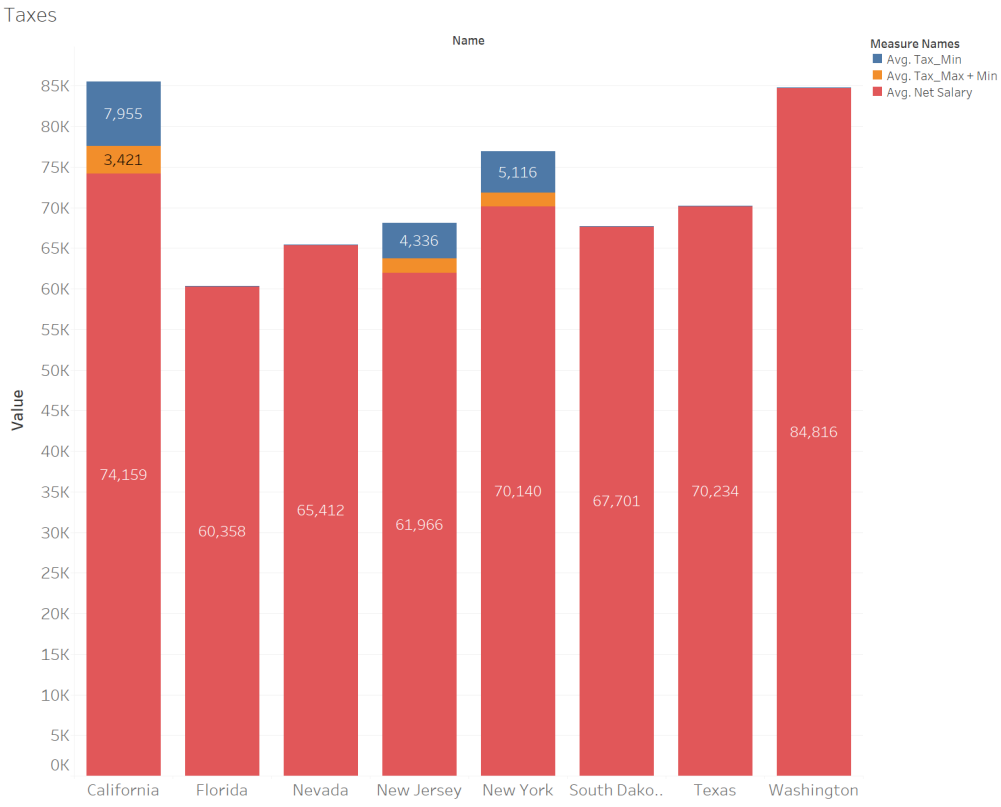
Secondly, we analyzed in term of after-tax economic benefit what will be the 'best' state to consider. For doing this, we obtained the minimum and maximum state tax rates applied to the income and compared if some locations have relative advantages as could be perceived by international inmigrants. The graph below shows the result of the excercise conducted, which looked at both at states with higher salaries and those states that currently do not charge a local tax on income. From the graph an interesting pattern results: altough California is the state with the highest average salary, once the taxes are accounted for this advantage is lost. In particular, if the minimum tax rate for California is applied (which will represent $7,995) then this advantage is lost against Washington (which does not charges any state tax). (Therefore if the maximum rate is applied, which will be $7,995 + $3,421, the difference becomes wider). As this case, there are other presented in the graph (New Jersey vs. Nevada, South Dakota, and Texas).

Next we analyzed if having H1B dependants is associated to a higher proportion of certified petitions. From our analysis, presented in the graph below, it is clear that within the % of cases

vs. Low Application.bb



image



.bb

image

who applied with dependants across 2014-2017 we observed that there is a higher chance of obtaining a sucessfull petition if a person has a dependant (i.e., the proportion of certified petitions is higher for people with dependants).

Regarding the proportion of certified petitions, we also looked at whether having a full-time position is positively associated to our itnerest outcome. Similar to our belief, but contrary to our previous expectations, the proportions ended being up fairly similar with a higher value for full-time offers (presented in the graph below). In this case the analysis is more nuanced because we have information pertaining to full-time positions only for 2016-2017 (the field was not collected before). Although in 2016 we see having a full-time position increases your chances of obtaining a certified petition, in 2017 both proportions become fairly similar. In line with this, our surprise regarding the graph is related with the relative success that part-time positions have in the data.

1.7 Policy Analysis

In 2017, President Donald Trump spoke about major changes in the H1-B visa application process and based on this we decided to further dive into these aspects to see how implications on salary cap H1-B visa would affect. In relationship to this, we also analyzed if, as President Trump claims, international inmigration has a conexion with the higher unemployment of native borns. Finally, we also validate if state level fluctuations in the population might be explained by the pattern of certified applications.

1.7.1 1) Will a Floor on H1-B prevailing wage impact the number of petitions ?

Analysis
In 2016, it was seen that 75 percentile of the H1-B application petitions had a prevailing wage of 84,781 dollars, while the 90 percentile had a prevailing wage of 106,101 dollars.In 2017, it was seen that 75 percentile of the H1-B application petitions had a prevailing wage of 90,750 dollars, while the 90 percentile had a prevailing wage of 112,549 dollars.So if Donald trump really wants to prevent people from filing H1-B applications we believe that he should raise the 90,000 dollars cap even higher to allow only skilled immigrants(who earn more) to enter the American workforce

1.7.2 2) Does State Population and H1-B have a relation ?

Analysis
In general we tried to see if increase in Average State Population had a relation with the number of H1-B petitions applied. We evaluated the top-4 states(California, Illinois,New York and Texas) where maximum number of H1-B petitions came from during 2015-2016 and its relation to the average State Population. We observed that California had the highest average population followed by Texas, New York and Illinois. In general in terms of H1-B petitions we observed positive trends i.e an increase in average state population the H1-B petitions increased.However, in Texas we observed the Average State Population to be much higher than number of case status this could be possible that Texas is the 2nd largest state in the US and the number of application are lesser than the average population.

1.7.3 3) Are immigrants taking jobs ?

Analysis
A major reason behind Trump's policy change in H1-B visa applications was that immigrants

are taking American jobs. Based on unemployment information status obtained from 2015-2016 within different regions across America. We choose to narrow our analysis based on the top three regions where most H1-B petitions come from. We observed that as the number of H1-B case status increased the unemployment decreased for Middle Atlantic and Pacific regions, but remained constant for West South Central. Our further analysis seemed to match with our previous analysis where West South Central (containing Texas) the average population increase was higher than the number of case status applied.In general, the trend does seem to tell us as number of case status increased the uemployment decreased so it could be that immigrants are taking jobs.

1.7.4 Conclusion

This project has taken us to explore a relevant dataset as international students. Through an exploratory analysis driven by visualizations, we were able to convey an interesting message that pertains the population of international students. Despite this results, we are open to suggest caution when interpreting them since these were not obtained under a causality or prediction setting. Moreover, as was discussed before, there are important assumptions behind the analysis provided. Nevertheless, we think the visualizations proved good for generating additional insights and maybe for exploring further topics in the future.

1.7.5 References

1) Kaggle,H-1B Visa Petitions 2011-2016

2) US Department of Labor US Department of Labor

3) What is H1B LCA? Why file it? Salary, Processing times – DOL

4) H1B Application Process: Step by Step Guide