Analytics on LCA, PERM Applications And Prevailing Wages

Kanishka Sunnam, Nikeeta Akbari, Nikita Marathe, Rohit Tiwari, Vrunda Shah, Yuvasree Kalaiselvan
Department of Information Systems, California State University
Los Angeles

Abstract: This paper aims at performing data analysis on the LCA (Labor Condition Application) applications, Permanent Employment Certification applications, and prevailing wage determination. This data is available on the Department of Labor's Office of Foreign Labor Certificate. Every year the department avail a public access file, under disclosure data category for public use. We decided to work on the available data from the year 2008 to the year 2017 and tried to analyze the number of applications filed by employers, average wage, top job titles with the applications, etc. As the data size is enormous, we have used HIVE to perform data analysis and Tableau tool for data visualization. We find these data interesting as most of our class members are International students and hence deriving insights for job positions, top states for H-1B jobs and top employers that sponsor H-1B and permanent residence could help them to choose wisely.

1. Introduction

The dataset contains LCA application information for non-immigrant visa categories like H-1B, H-1B1, and E-3. The H-1B visa is a non-immigrant visa in the United States. It allows foreign workers to temporarily work in the United States in specialty occupations. The term 'specialty occupations' to the use of highly specialized technical and practical knowledge or occupational skills. Total stay in the United States on H-1B limited to 6 years. Initial approval is for three years, which can be extended for an increment of up to 3 years.

This dataset also contains the permanent labor certification applications which are filed by the employer to the Department Of Labor jurisdiction, when approved lets a foreign worker work permanently in the United States. We have also worked with the prevailing wages information, which is defined as the average wage paid to the employee under the same skills set, qualifications and the similar area of employment.

We chose to perform our analysis on these datasets because the data gives profound insights into the H1B Process, PERM process, and the prevailing wages information. We identified such research questions for our project so that answering these research questions will make it possible for foreign workers to target particular employers while looking for Jobs and will make it easier to decide whether the H1B process and PERM are getting simpler and more convenient.

2. System Specifications

To begin our data analysis, we have used Apache Hadoop as it popular to process massive data sets for analysis. We have also used Biginsights services of IBM cloud platform to connect to Hadoop cluster and hive. Below are the IBM Biginsights cluster details: -

- Cluster Type - Hadoop IBM Big Insights
- No. of Data Nodes – 1 node | vCPU = 4(24 GB RAM)
- No. of Management Nodes – 1 node | vCPU = 12 (48 GB RAM)
- Data Disk - 1 TB SATA | Data Storage- 244 GB
- CPU Speed - 2.30 GHz
- Version - IOP 4.2 [IBM Open Version Platform]
- Operating System - CentOS 6.6 [Linux]
- Data Centre - Washington DC

3. Workflow

We discovered data on Kaggle.com on H-1B applications and used the link provided to the data source, which is available on U.S. government Department of Labor website. We have downloaded Disclosure data on LCA, PERM, and PW available on OFLC performance data page. The dataset we got is in Microsoft Excel compatible (.xlxs format) format. We got 10 data sets for each year separate from 2008 to 2017. We have consolidated the data from all ten flat files into a single CSV file. There are few variables in our raw dataset which are not relevant to our analysis. Therefore, we filtered out those columns and some missing values from our raw dataset. After cleaning our dataset, we loaded our CSV files into the Hadoop distributed file system. For analyzing our dataset, we have used HIVE, which uses hive queries to perform advanced analysis. The hive queries require data to be loaded into the hive table and further analysis could be completed. After the hive queries are executed, we will load the result data into the Tableau, a visualization tool. The following diagram shows step by step workflow carried out for this analysis:-

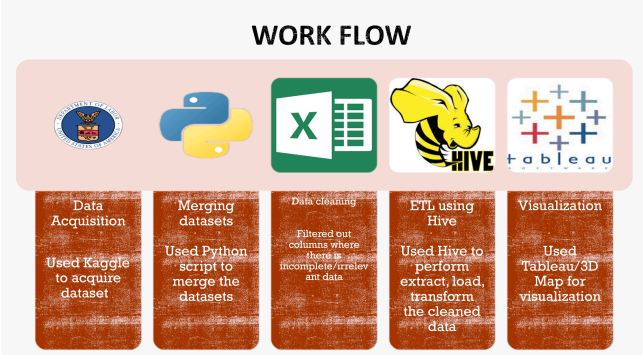

Figure 1: Workflow

4. Data Analysis

For all our Research question, we first transformed the data and loaded it into the hive tables over which we performed some hive queries. Later we used Tableau tool and its functionalities to visualize our data in different ways to reach our goal.

4.1 Number of H-1B Applications filed through the 2008 year to 2017 year

The trend for number of H-1B applications filed through 2008 to the 2017 year is shown in the below graph. The graph depicts a total number of H-1B applications submitted which includes both new and renewal applications. It was observed less number of applications were filed during the year 2009 to the 2011 year because the country faced recession during these years. Although the unfavorable political scenario is observed for employing foreign labor in the year 2017, there is not much decrease in the filing of the H-1B applications in this year.

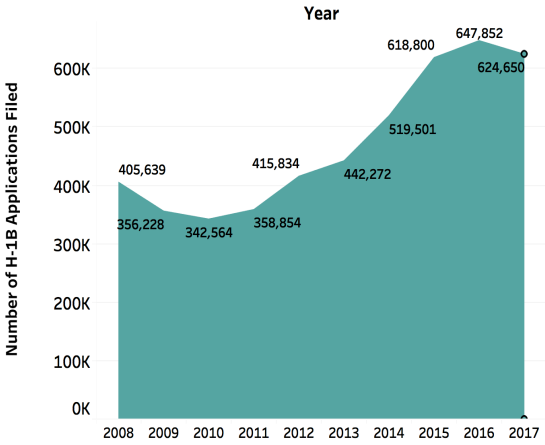

Figure 2: Number of H-1B Applications filed

4.2 The top 10 employers to file H1B and PERM applications and comparision between both the top 10 employers

The bar graph on top 10 employers of H1B and PERM gives the names of the employers to file a maximum number of applications from the year 2008 to the year 2017. We noticed that the top employers who filed H1B and PERM applications had only four employer names in common, which led us to conclude that the number of PERM applications filed are relatively less compared to by the same employers.

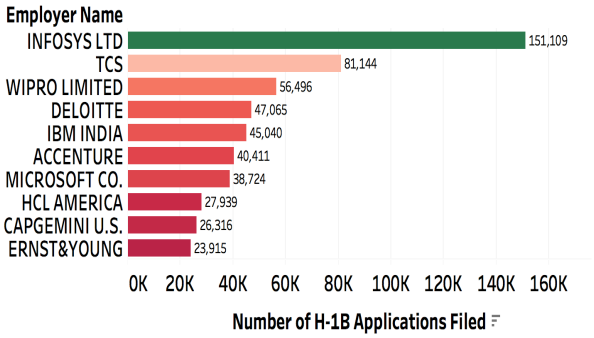

Figure 3: Top 10 employers to file H-1B Applications

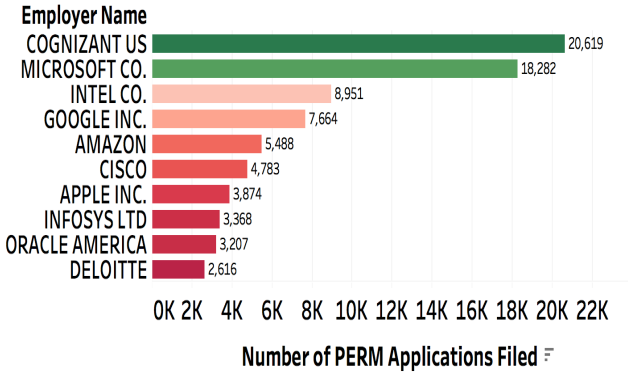

Figure 4: Top 10 employers to file PERM Applications

4.3 Top employers to file PERM applications for F-1 visa in the year 2017

Amazon, Antra Inc, Google, Intel Corporation, KFORCE Inc, Microsoft, NVIDIA Corporation, Wayne Farms LLC are top companies who directly filed PERM application for master's students, and an average number of submission is 150 for the year 2017. If international students want to work permanently in the country, they can approach these companies for jobs.

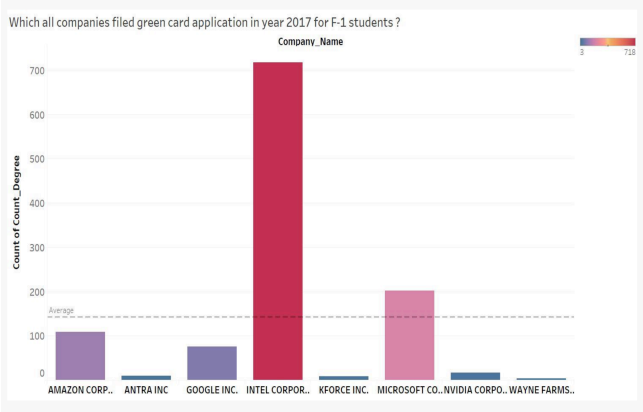

Figure 5: Top employers to file PERM applications for F-1 visa

4.4 Top job titles with highest applications over the year 2008 to the year 2017

As per the data from 2008 to 2017 of top 2 job titles with the highest application, we can see the job which is in higher demand. Also, we can see that in last eight years the top 2 job profiles were "ANALYST" and "SOFTWARE DEVELOPER." Based on that we can assume that after graduating from CSULA and getting a job, the company will apply for our H-1B.

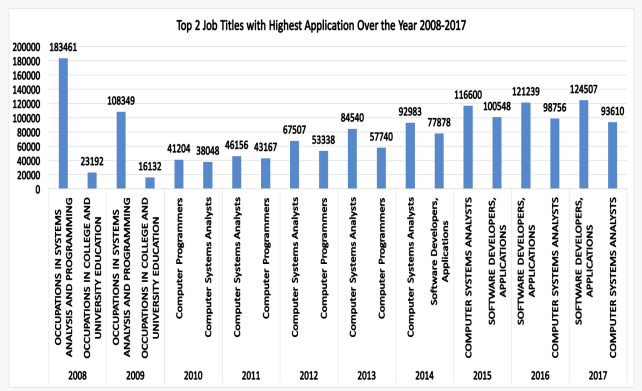

Figure 6: Top job titles with highest applications over the year 2008-2017

4.5 Average salary for an analysts position having a green card from the year 2011 to the year 2017.

As we all know, Green Card holders have some benefits over other foreign workers, i.e., Higher Salary, allow you to apply for US citizenship, can work for your sponsor company or yourself, Sponsor your spouses or relatives, etc. The figure shows average salary trend from the year 2011 to the year 2017 for analysts who hold green cards. This position can be anything such as budget analyst, computer system analyst, data analysts, Market research analysts, etc. Salary goes from 76K to 91K. In Recent time, everything connected to internet transfers data to it and Volume of data increase with time. To analyze this data and get some helpful insights analysts are needed for every field, and thus salary scale goes up consistently. For the year 2017, it is approximately 91K. Besides political changes and other affecting factors, this trend doesn't change.

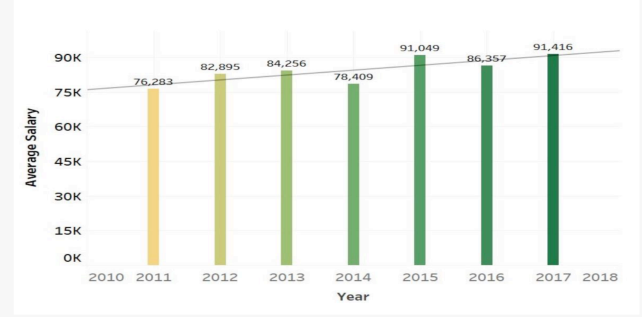

Figure 7: Average salary for an analyst position having a green card

4.6 Highest average salaries based on job title for the year 2017

The bar graph lists the top four most sought job titles which get the maximum average salary as submitted by the US employer in the Fiscal Year 2017. Data Engineer tops the chart with an average salary of 92000, Data scientist with 87000, Consultant with 70000 and Analysts with 64000. Analyst Job positions include Big Data Analyst, Business Analyst, Computer Analyst & Data Science Analyst. The Average Salary is the average offered salary on the LCA.

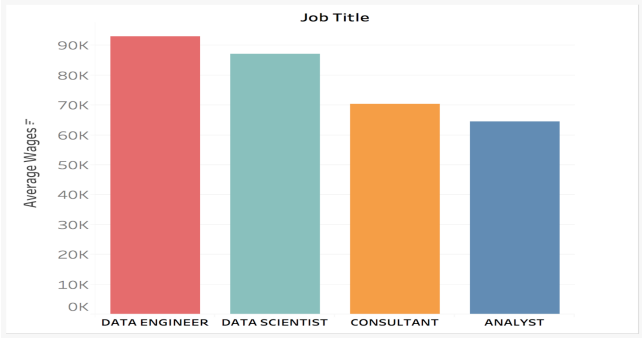

Figure 8: Highest average salaries based on job title for the year 2017

4.7 Highest average salaries based on worksite state

The graph shows the top 10 States in the United States whose average salary ranges from 94000 to 77000. We have also calculated the average resulting salary after applying the state tax to each respective state. As we can infer from the data, it is better to choose the states that balance both the average wage & state tax, rather than choosing a state which gives maximum wage.

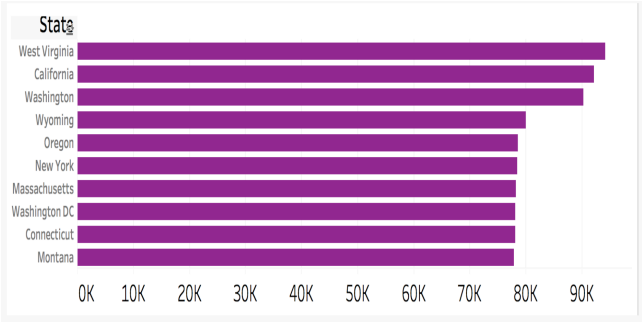

Figure 9: Highest average salaries based on worksite state

5. Summary

- The number of the H-1B applications are relatively higher than the number of the PERM applications for the top 10 employers.
- Microsoft, Infosys and Deloitte Consulting are the companies that are in the top 10 employer list for both H1B as well as PERM so if you are planning to stay and work in USA for longer duration then these three companies are your best bet.
- We also found Amazon, Antra Inc, Google, Intel Corporation, KFORCE Inc, Microsoft, NVIDIA Corporation, Wayne Farms LLC are top companies who directly filed green card application for International students on F-1 Visa.
- Average wages of Data Engineer, Scientist, Consultant and Analyst are between 62k to 92k.
- Data Engineer is the highest paid job and Data Analyst lower paid.
- Analysis also provided yearly salary increment for analyst's position.
- Any F1-B (Foreign International student) should be able to wisely chose the job title, Employer and State from the data analysis in this presentation based on their criteria like job stability, High Wages, VISA requirement, etc.

6. GitHub Link

https://github.com/Kan1shka9/H1B-Data-Analysis-using-Hive-on-Hadoop-Cluster.git

7. References

[1]https://console.ng.bluemix.net/catalog/services/biginsights-for-apache-hadoop
[2]https://www.tableau.com
[3]https://www.foreignlaborcert.doleta.gov/performancedata.cfm