**PROJECT DONE BY ZHIJIAN LIU AND ANGELA ERDIAW-KWASIE**

**DATA SCIENCE ANALYSIS OF HI-B VISA APPLICATION IN THE UNITED STATES OF AMERICA.**

**Abstract**

The H-1B is an employment-based, non-immigrant visa category for temporary foreign workers in the United States. Every year, the US immigration department receives over 200,000 petitions and selects 85,000 applications through a random process. The application data is available for public access to perform in-depth longitudinal research and analysis. This data provides key insights into the prevailing wages for job titles being sponsored by US employers under H1-B visa category. In particular, we utilize the 2011-2016 H-1B petition disclosure data to analyze the employers with the most applications, wage distribution and data science related job positions.

**H-1B Visa Data Introduction**

The H-1B is an employment-based, non-immigrant visa category for temporary foreign workers in the United States. For a foreign national to apply for H1-B visa, an US employer must offer a job and petition for H-1B visa with 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 work in a full-time position.

**Data Wrangling**

In this section, we discuss a few key data transformations performed on the raw dataset before data analysis could be performed. The exhaustive code for the data transformation can be found in the github link attached as well the overall codes used for this project.

**Data Set Description**

The dataset originally had 10 variables with 3002458 observations and we added 25 more variables that we needed in our analysis to make our computation easier.. The extra variables included :

|  |
| --- |
| apply\_city:total applications by city  applyrank\_city\_us:rank of applications  applyrank\_city\_state  apply\_employer\_us  applyrank\_employer\_us "applyrank\_employer\_state"  [22] "apply\_employer\_state" "apply\_employer\_city" "applyrank\_employer\_city"  [25] "wage\_state" "wagerank\_state" "wage\_city"  [28] "wagerank\_city\_us" "wagerank\_city\_state" "wage\_employer\_us"  [31] "wagerank\_employer\_us" "wage\_employer\_state" "wagerank\_employer\_state"  [34] "wage\_employer\_city" "wagerank\_employer\_city" |
|  |
| |  | | --- | | > | |

**DATA TRANSFORMATION**

How you cleaned the data.

Talk about the Geocoding and the problems you entered during that.

**DATA ANALYSIS**

The questions we will be answering through data include:

1) Which Employers submit the most number of H-1B visa applications?

2) What are the most common Job Titles applied for by the high applicant employers?

3) Distribution of the wages of the top 15 Job titles sponsored by the top employer applicants.

4)Which employer submit the most Data Science Job related applications**?**

5) Which state submit the highest data science related applications ?

6) Distribution of wages of among the top ten Data Science Related Jobs States.

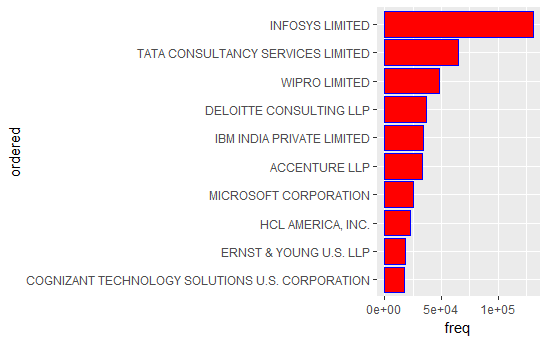
7)Distribution of applications within DC, VA and MD over the years?

8) Highest Data Science related job title h1b visa application state within DC, VA and MD.

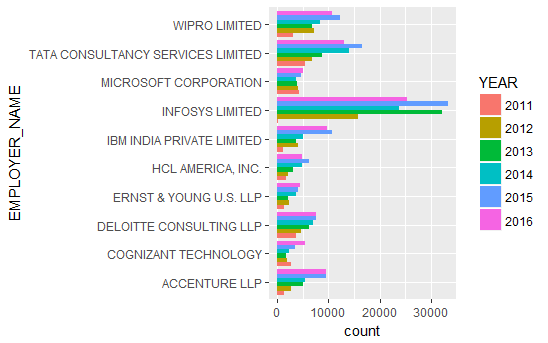
9) Among Virginia, DC and Maryland, which state is the highest data science related job title paying state?

**Which Employers submit the most number of H-1B visa applications?**

|  |
| --- |
| Top10 EMPLOYERS  EMPLOYERS NUMBER OF APPLICATIONS  101628 INFOSYS LIMITED 130592  202574 TATA CONSULTANCY SERVICES LIMITED 64726  230799 WIPRO LIMITED 48117  57032 DELOITTE CONSULTING LLP 36742  98190 IBM INDIA PRIVATE LIMITED 34219  3522 ACCENTURE LLP 33447  135595 MICROSOFT CORPORATION 25576  91895 HCL AMERICA, INC. 22678  69986 ERNST & YOUNG U.S. LLP 18232  45941 COGNIZANT TECHNOLOGY SOLUTIONS U.S. CORPORATION 17528 |

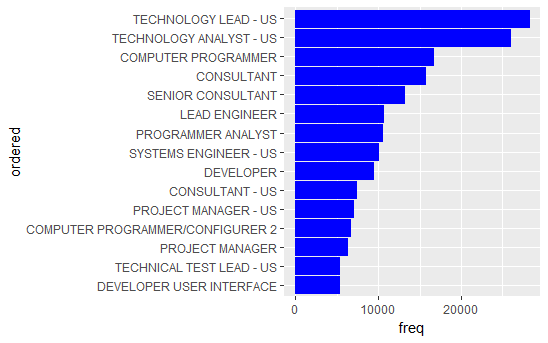


**Distribution of H1-B by top 10 Employers across the years(2011-2016)**



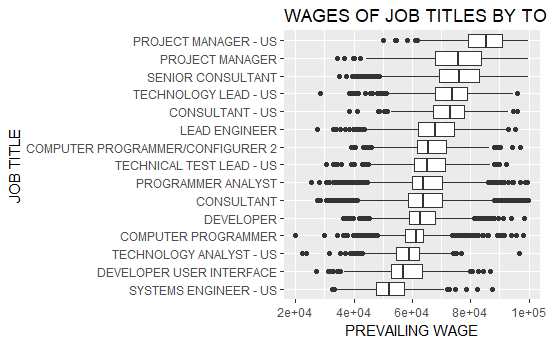
1. Infosys leads the pack by a huge margin with over 30000 applications in 2013 and 2015.
2. The Top 10 list is dominated by the Indian IT companies.(Wipro Limited, Microsoft Corporation, Tata Consultancy Limited, Cognizant Technology and Infosys)
3. In 2016, we observe a slight dip in the number of applications from Infosys, Wipro, Tata Consultancy, IBM India and HCL America. This might be because of increased incorporation of automation in the IT industry. According to an article(link to article will be attached) called the Quartz India, the Indian IT firms have been preparing for reduced number of H-1B visas for nearly a decade through increased focus on automation, cloud computing and artificial intelligence.

**What are the most common Job Titles applied for by the high applicant employers?**

**Observations:**

1. Technology related jobs fill up majority of the positions with the top 3 jobs being Technology Lead, Technology Analyst and Computer Programmer.
2. Consultant and Manager related jobs fill up the remaining spots.

**Distribution of the wages of the top 15 Job titles using the box plots**



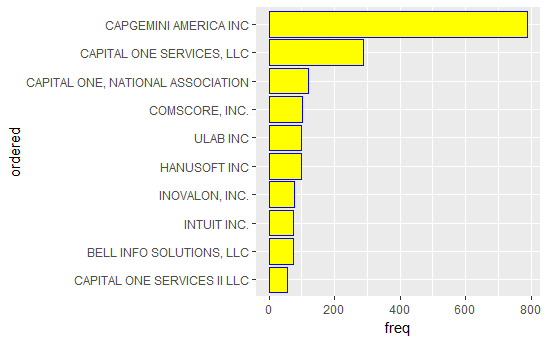
**observations**

Expectedly, the Manager level jobs and Lead Consultant job titles have the highest wages.

2. The Software Engineering jobs including Programmer analyst, Computer Programmer, Computer Systems Engineer have wages close to 60000 USD per annum.

3. Test Analyst and Systems Engineer have the lowest wages with the median slightly above 50000 USD.

**Which employer submit the most Data Science Job related applications? .**

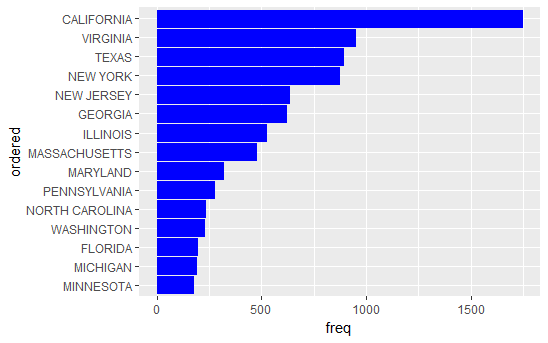


Observation:

1)CAPGEMIN AMERICA INC followed by Capital One Services LLC submit the most Data Science related applications.

2)The top ten companies with the highest overall applications do not submit more applications for Data Science Related Job.

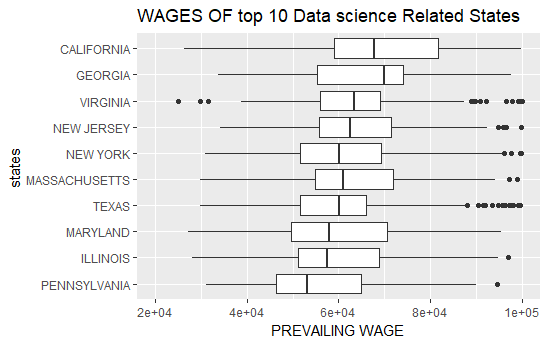
**Which state submit the highest data science related applications .**



**Observations:**

1. California leads the pack by a huge margin with over 1500 applications.
2. This result is expected as these states are hub of technology innovation with California housing the Silicon Valley, NY housing the Finance and media corporations.

**Distribution of wages of among the top ten Data Science Related Jobs States**

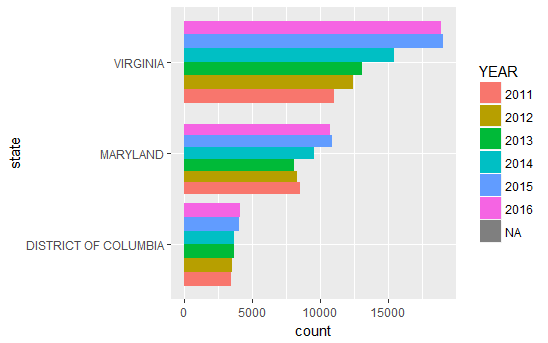


**Observations**

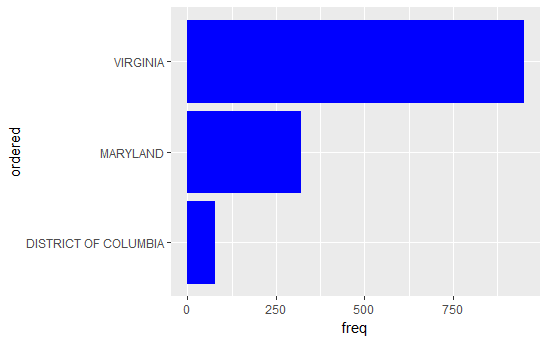
1. California has not only got the most number of jobs but also the highest wages. This might be due to the higher cost of living as I will analyze later.

2. Significant variation in the mean wage across the states.

**Distribution of applications within DC, VA and MD over the years**

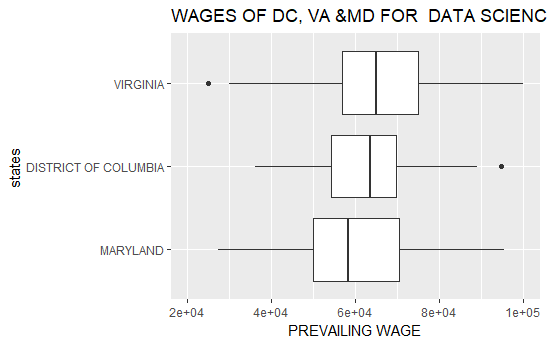


8) Highest Data Science related job title h1b visa application state within DC, VA and MD



Virginia leads with over 949 applications followed by Maryland with 318 applications and then District of Columbia which has 74 applications.

Among Virginia, DC and Maryland, which state is the highest data science related job title paying state?



**Conclusion**

To conclude, in this project, we performed exploratory data analysis on the H-1B visa petition disclosure data for the period 2011-201. Also, I found that the Data Scientist position has experiences an exponential growth in terms of H-1B visa applications. Interestingly, the Data Scientist jobs are clustered in a few hostpots with San Francisco region having the highest number. The top ten companies with the highest overall applications do not submit more applications for Data Science Related Job.

We expanded this project to build a shiny app to ………………(talk about whatever you involved in your shiny app.