H1B VISA DATA ANALYSIS: AN EXPLORATORY ANALYSIS ON THE FACTORS AFFECTING THE APPROVAL OF H1B VISAS

EAS 503 FINAL PROJECT

Group 24:
1. Amritha Subburayan
2. Kamalnath Sathyamurthy
3. Sriram Muralidharan

1. Abstract

H1-B visas are a type of non-immigrant visa for foreign national professionals who wish to work in the United States. As of FY 2022, there were around 308,613 H-1B registrations from 37,000 prospective petitioners. The visa requires the applicant to have an employment letter which indicates their legal employment in order to file an application to the US immigration service (USCIS). The number of applications approved by USCIS yearly should not exceed 85000 but the non-immigrants who wish to work in United States are outnumbered recently. The selection of the visa process is based on lottery; hence, it is crucial to understand the case status movements from the past years and to analyze which all factors thus affecting the approval of the visa. In this project, we have performed various exploratory data analysis to understand the trends and variation of H1-B visa application certification to help both the employee and the employer.

2. Introduction

This project intended to explain a clear trend, variation, and patterns about the H1-B visa application process over the period of time. We have collected data from Kaggle to analyze which factors influences the approval of H1-B visa for the non-immigrants to work legally in the United States. We have queried the dataset from the year 2016 till 2018 which contained around 40 lakhs entries of all the applicants. It is crucial for the US Department of Labor to figure out which companies are sending the most H1B Visa applications and to analyze which applications are legit to proceed further. This study will provide vital statistics about areas the U.S. government emphasizes to encourage domestic students to develop more local workforce in those job domains for the future. At the same time, it aims to provide guidance for employment in the US to talented international professionals.

3. Data Sourcing and Preprocessing

We have taken the following datasets for our analysis. They are mentioned below

Dataset Name	Source	Link	Description
H1B Visa Dataset - 2016, 2017, 2018	Kaggle	https://www.kaggle.com/jmpark746/h1b-visas?select=h1b16.csv https://www.kaggle.com/jmpark746/h1b-visas?select=h1b17.csv https://www.kaggle.com/jmpark746/h1b-visas?select=h1b18.csv	Dataset which contains information about Visa status, applicant information and employee information for the year 2016, 2017 and 2018.

3.1. Original Dataset Summary:

Variables	Description
CASE_STATUS	Status of H1B Visa
EMPLOYMENT_START_YEAR	Year in which employee started working with the employer
EMPLOYER_NAME	Name of the Employer
EMPLOYER_STATE	State in which Employer's company operates from
JOB_TITLE	Title of the employee job in the company
SOC_NAME	Job sector

3.2. Table Creation:

The datasets that are collected contains information about Visa status, Case information and employer information for the year 2016, 2017 and 2018 in which decisions for cases are made accordingly. After collecting the dataset from the Kaggle, we have imported the dataset to database and created below tables to store the variables.
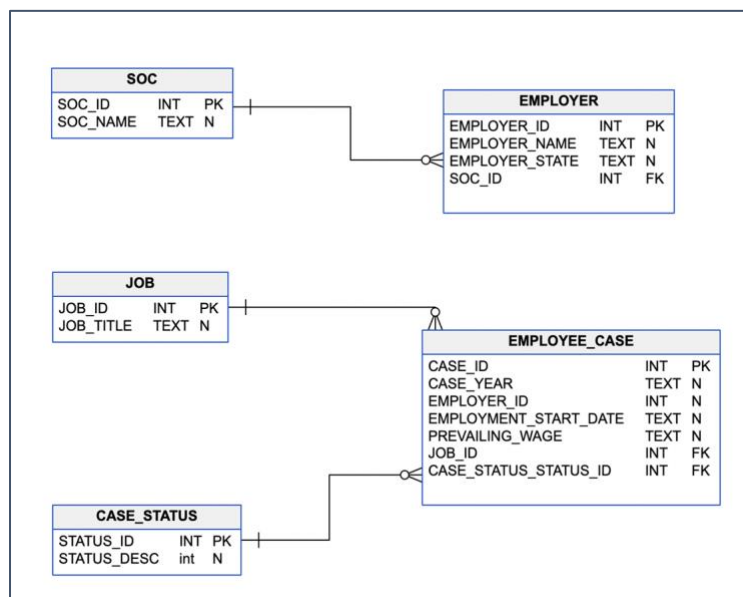
4. Proposed Model and Approach:

We have performed Exploratory Data Analysis to pursue below research questions of our project, which are listed below:

1. Which job sector possesses a higher percentage of visa approval?
2. Which employer files more petitions for their employees?
3. What is the Influence of Data science jobs on H1B visa and its growth trend?
4. How H1B applications filings changes over a period of time?
5. Classifying denial rate from historically collected data.

We finalized to analyze our data with below EDA techniques to provide the trend analysis in visa statistics. Collected dataset were queried from table and modelled with use of below techniques.

Univariate non-graphical: Identifies the skewness and kurtosis in the dataset.
Univariate graphical: Analyze the frequency using histogram, stem and leaf plot, and box plots
Multivariate non-graphical: Analyze the relationship between multiple variables of data through cross-tabulation or statistics.
Multivariate graphical: Analyze relationship between multiple columns using Grouped bar plots, Bar chart, heatmap, scatterplot, parallel box plot.

5. Analysis and Results:

5.1.   Exploratory Data Analysis:

We performed various EDA techniques to understand the factors influencing the H1-B Visa approval.

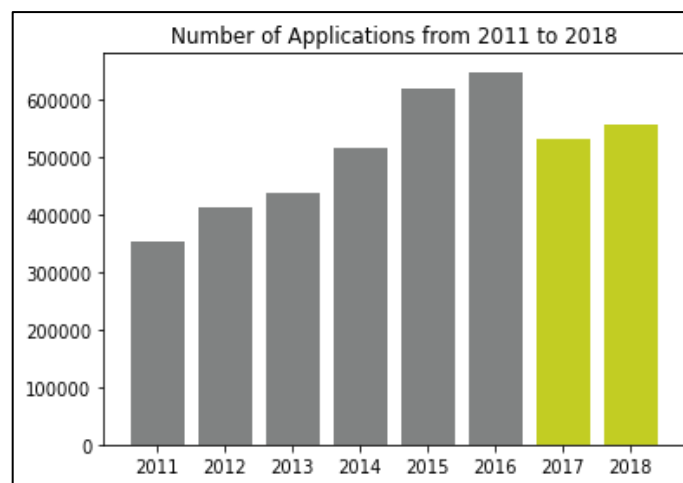### 5.1.1. Trend Analysis of H1-B Visa Applicants and Approval Status

Figure 1

Figure1 : The plot explains about the number of visa applications was submitted yearly. We can the increase in number of applications from 2011 till 2016 but from 2017 there was a drop in number of applications. This can also help to analyse which presidential period the H1-B visa's are offered more and not.
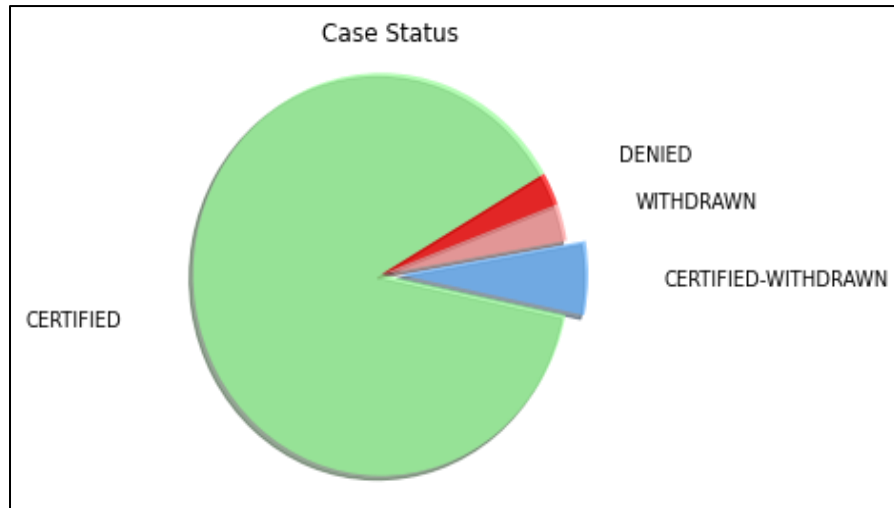

Figure 2

Figure 2: The pie chart shows the volume of case status, from which can interpret that many applications were approved whereas only few applications were denied of approving. Hence, this EDA will help us to focus more on the denied ones which can be further improved to get it approved in later stages.

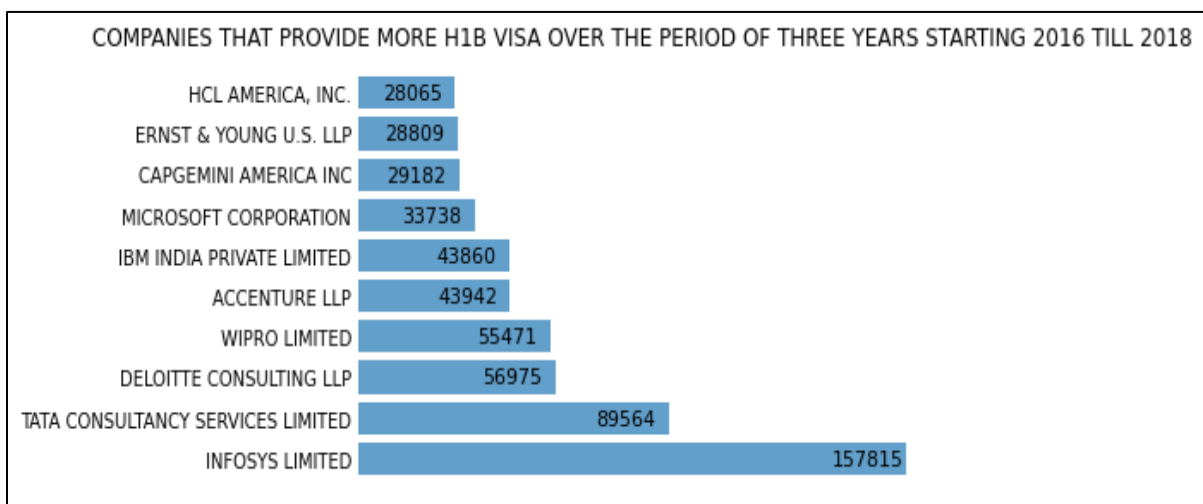**Trend Analysis of Companies H1B Sponsorship Over The Year**

Figure 3

Figure 3: The bar chart explains which companies has sponsored a greater number of H1-B visas to their employees between the year 2016 till 2018. We can interpret that, Infosys Limited has offered most of the H1-B visa to their employees. From this an individual can take wise decision to choose between the place to work considering these potential factors.
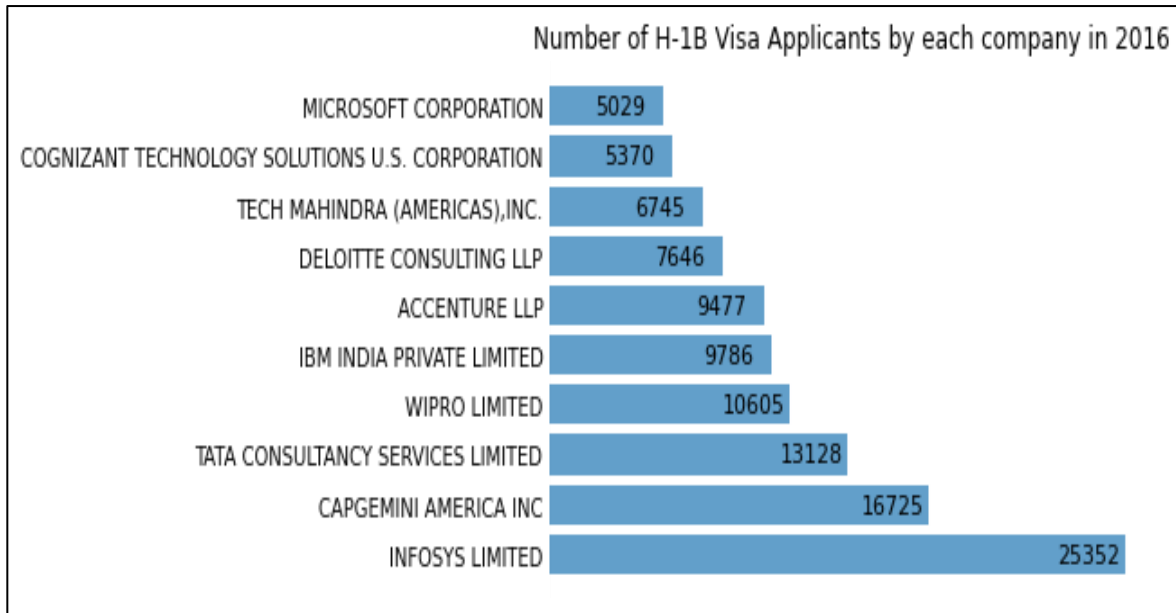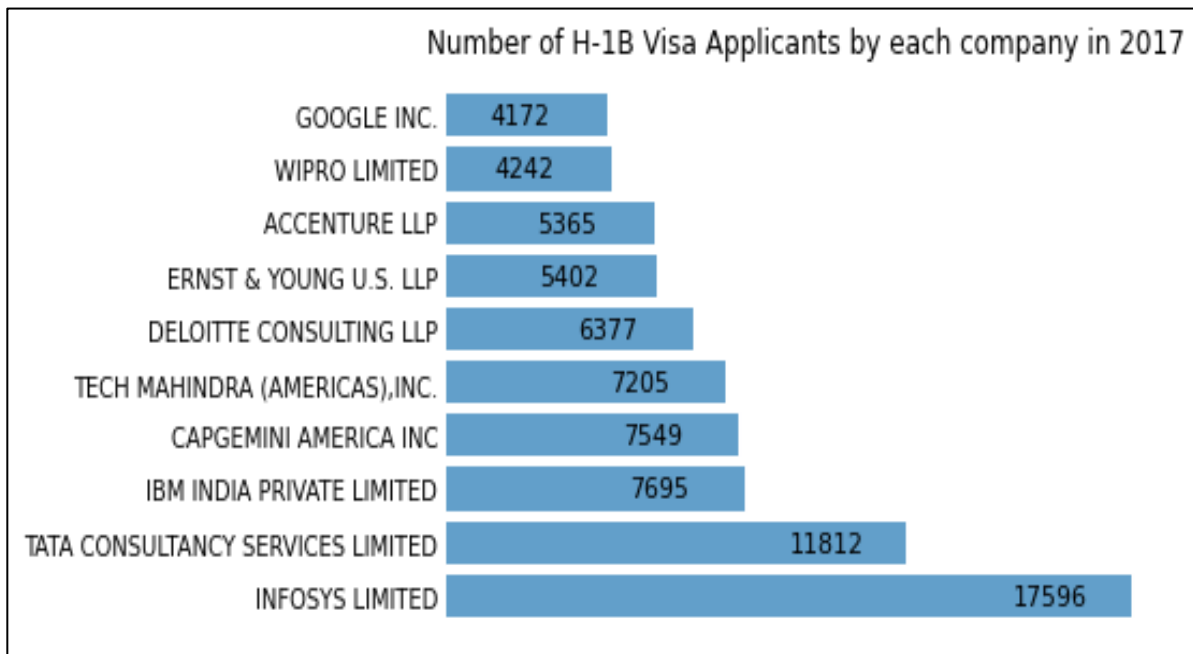

Number of H-1B Visa Applicants by each company in 2016

Company	Applicants
MICROSOFT CORPORATION	5029
COGNIZANT TECHNOLOGY SOLUTIONS U.S. CORPORATION	5370
TECH MAHINDRA (AMERICAS),INC.	6745
DELOITTE CONSULTING LLP	7646
ACCENTURE LLP	9477
IBM INDIA PRIVATE LIMITED	9786
WIPRO LIMITED	10605
TATA CONSULTANCY SERVICES LIMITED	13128
CAPGEMINI AMERICA INC	16725
INFOSYS LIMITED	25352

Figure 4: 2016 H1-B visa applicants

Number of H-1B Visa Applicants by each company in 2017

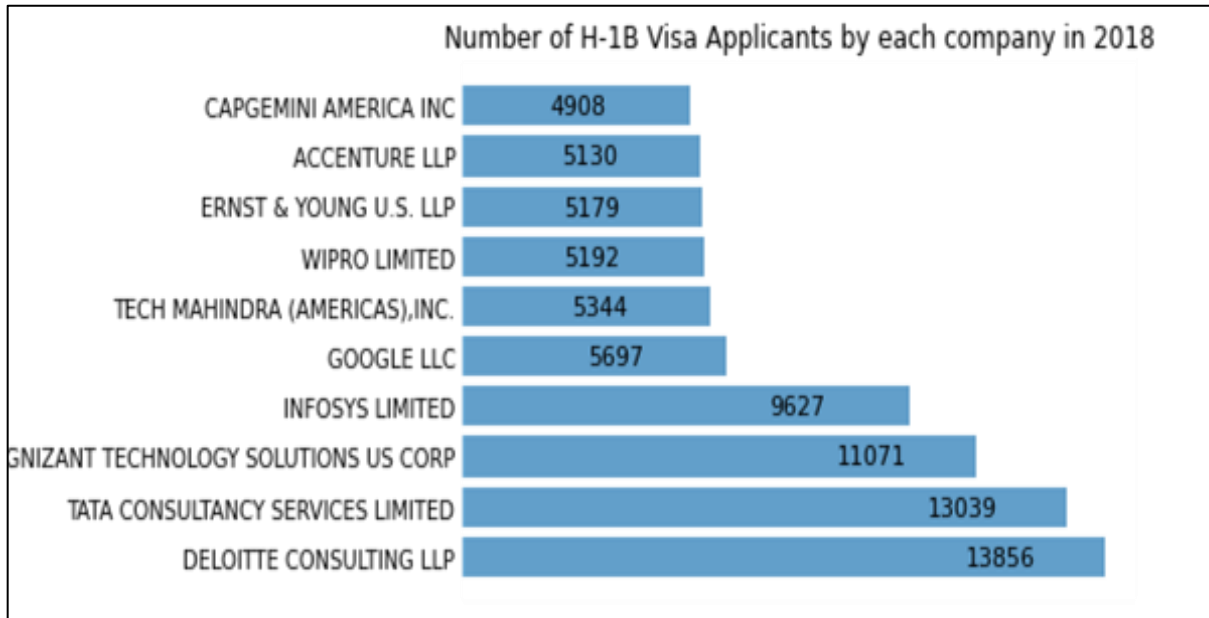| Company	Applicants
GOOGLE INC.	4172
WIPRO LIMITED	4242
ACCENTURE LLP	5365
ERNST & YOUNG U.S. LLP	5402
DELOITTE CONSULTING LLP	6377
TECH MAHINDRA (AMERICAS),INC.	7205
CAPGEMINI AMERICA INC	7549
IBM INDIA PRIVATE LIMITED	7695
TATA CONSULTANCY SERVICES LIMITED	11812
INFOSYS LIMITED	17596

Figure 5: 2017 H1-B visa applicants

Figure 6: 2018 H1-B visa applicants

Figure 4,5 and 6 shows visa sponsored by each company in the year 2016, 2017 and 2018 separately where Infosys Limited occupies more H1-B visa granted for its employees compared to others.

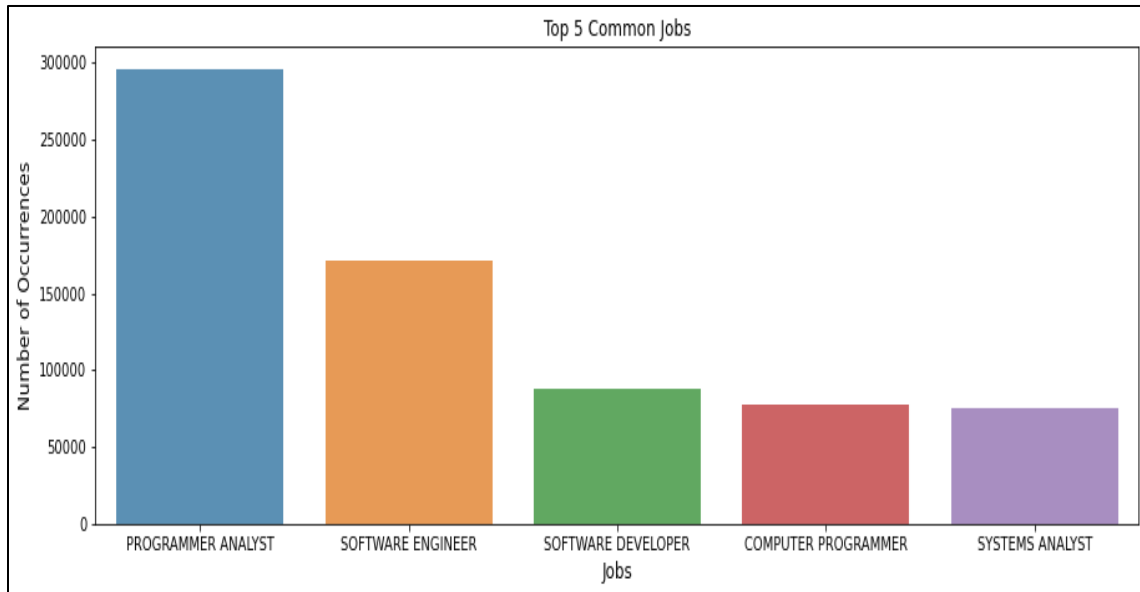**Job Sectors and Jobs with more H1-B Visa applicants**

Figure 5

Figure 5: Above bar plot shows the top 5 common jobs where most of the H1-B visa are certified. Program Analyst is the most popular job roles in late years and following are the software engineers.
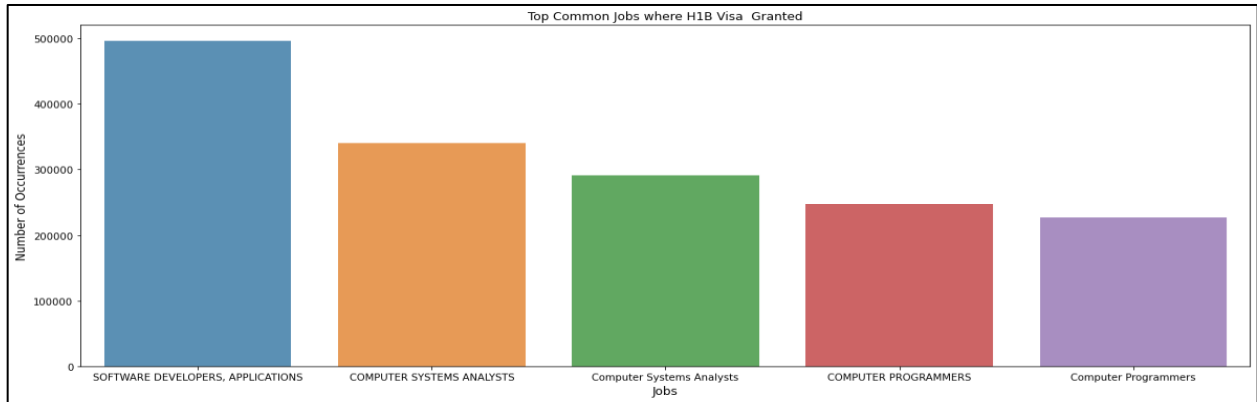

Figure 6

Figure 6: Above bar plot shows the top 5 common jobs sectors where most of the H1-B visa are granted from the employers. Software field is the only field where many applicants are granted for the H1-B visa.

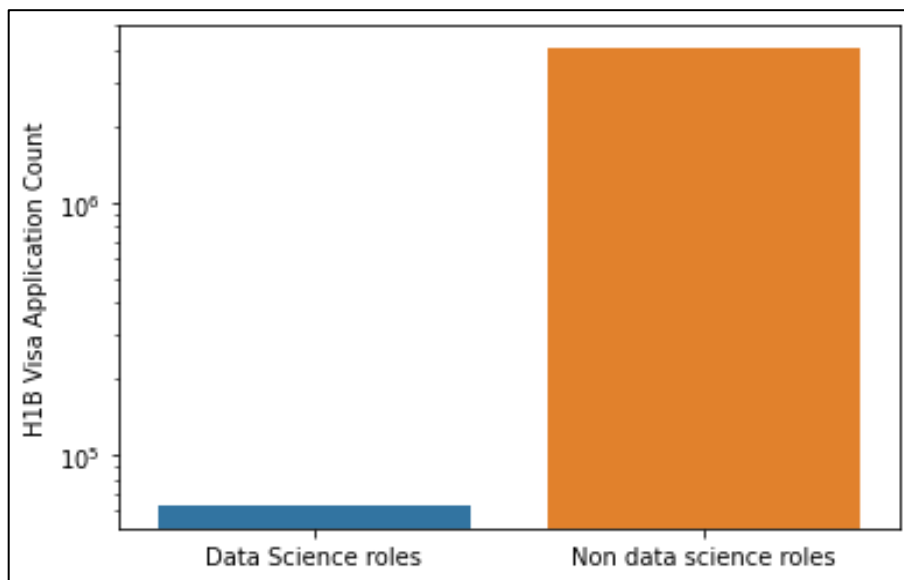## H1B VISA APPLICANTS: DATA SCIENCE VS NON-DATA SCIENCE

Figure 7

Figure 7: The bar plot shows the relationship between Data Science field and Non-Data Science field applications, from this we can conclude that Data Science jobs has started becoming popular whereas other field applicants are already huge in numbers.
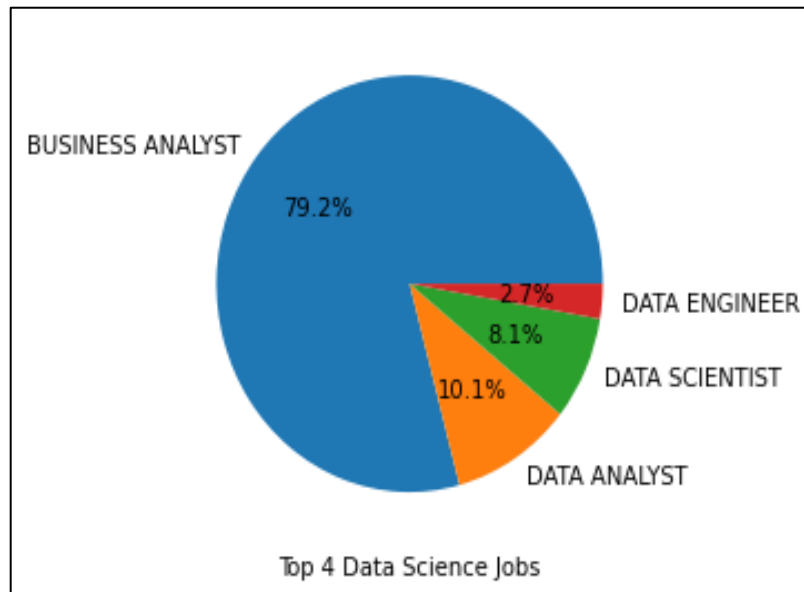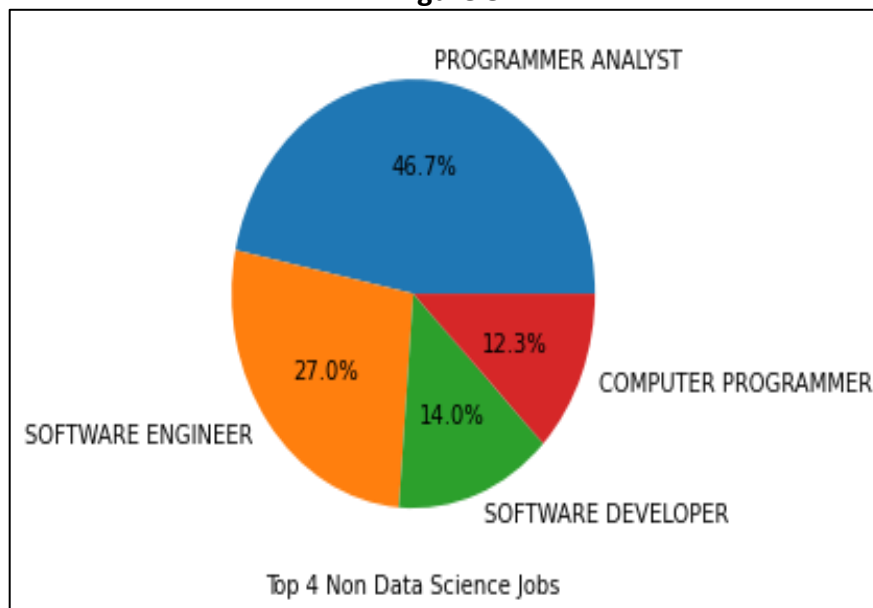
Figure 8

Figure 9

Figure 8 and Figure 9: The above pie chart shows the top four job roles that are currently in top of the market in both data science field and non-data science field.

6. Conclusion

With this analysis, we can be able to interpret that Computer sector plays major part in H1B Visa applicants, Data Science being with gradual growth from 2016 to 2018. From our research, we have understood which all factors influence in getting the H1-B visa approved from the US

government. Software field is the topmost sector to get the visa approved from the government without any hazzle, whereas data science is still in progress to achieve its place in the future.

7. Future Research Directions

Predicting the H1B Visa status based on various model and finding best model among them helps the employees and employers who are considering to sponsor them. Also, we need to collect the dataset from 2019 till 2021 to understand the growth of data science in recent years.

8. Reference

Debabrata Swain, Kushankur Chakraborty, Anay Dombe, Ashitosh Ashture, Nandakishor Valakunde: *"Prediction of H1B Visa Using Machine Learning Algorithms"*
Beliz Gunel, Onur Cezmi Mutlu: *"Predicting the Outcome of H-1B Visa Applications"*