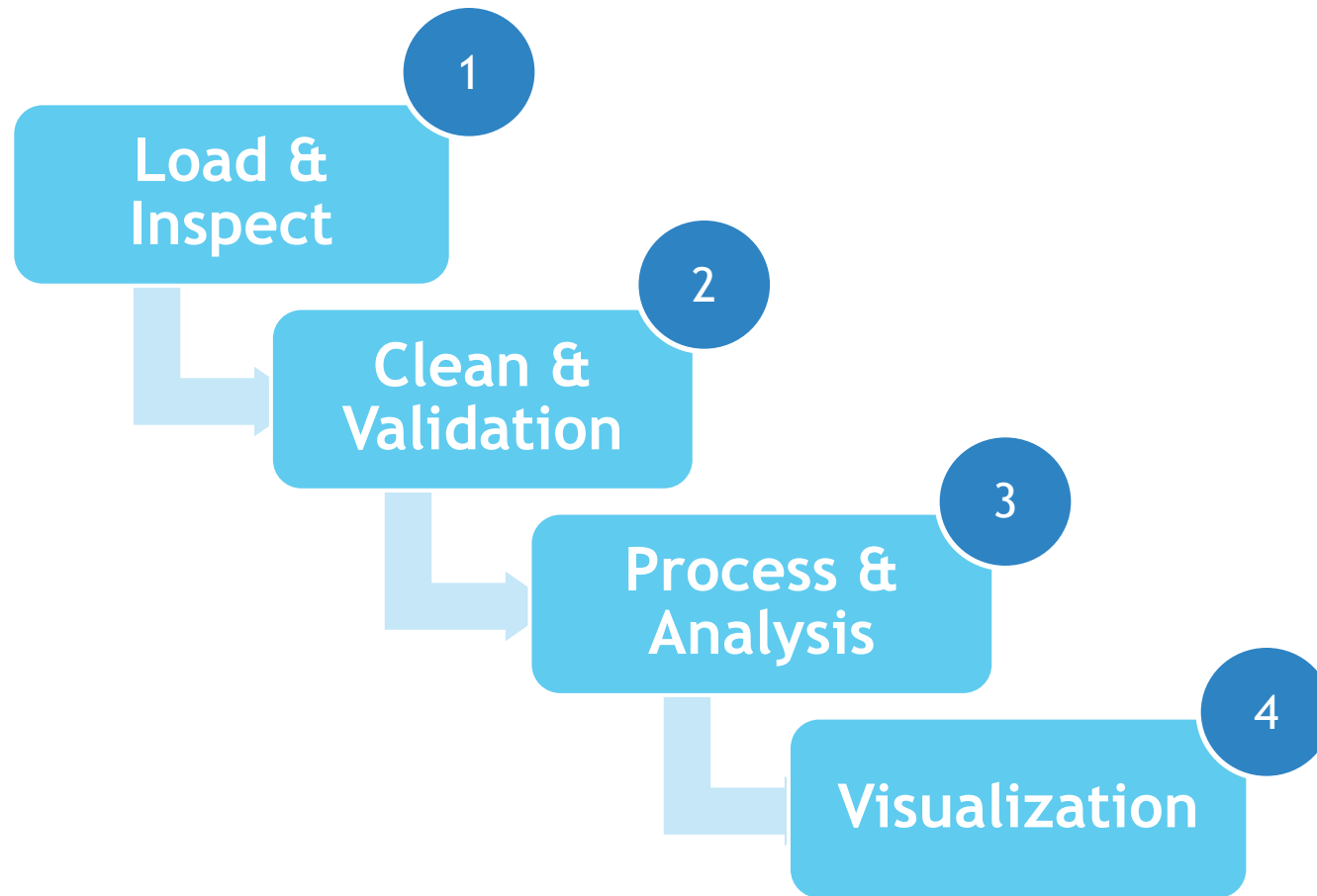


H1B Visa Applications Analysis



Steven Li

Table of Contents



Load & Inspect

- ▶ The H1B visa is an employment-based, non-immigrant visa for temporary workers. For this visa, an employer must offer a job in the US and apply for your H1B visa petition with the US Immigration Department.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

- ▶ This dataset contains five year's worth of H-1B petition data, with approximately 3 million records overall. The 11 columns in the dataset include case status, employer name, worksite coordinates, job title, prevailing wage, occupation code, and year filed.
- ▶ Data Source:
<https://www.kaggle.com/nsharan/h-1b-visa>

INSPECTING DATA

```
: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3002458 entries, 0 to 3002457
Data columns (total 11 columns):
Unnamed: 0                int64
CASE_STATUS                object
EMPLOYER_NAME              object
SOC_NAME                   object
JOB_TITLE                  object
FULL_TIME_POSITION         object
PREVAILING_WAGE            float64
YEAR                       float64
WORKSITE                   object
lon                        float64
lat                        float64
dtypes: float64(4), int64(1), object(6)
memory usage: 252.0+ MB
```

Clean & Validation

```
# Deal with missing values
# Because most missing values are catagorical values
# and are not applicable for imputing with mean or median
# so we decide to delete rows that contain missing values.
# There is no outliers
data = data.dropna(axis=0)
```

```
# Check if the dataset has any missing values
nnull=data.isnull()
nnull.sum()
```

| | |
|--------------------|--------|
| Unnamed: 0 | 0 |
| CASE_STATUS | 13 |
| EMPLOYER_NAME | 59 |
| SOC_NAME | 17734 |
| JOB_TITLE | 43 |
| FULL_TIME_POSITION | 15 |
| PREVAILING_WAGE | 85 |
| YEAR | 13 |
| WORKSITE | 0 |
| lon | 107242 |
| lat | 107242 |
| dtype: int64 | |

Check
missing
values

Deal with
missing
values

Recheck

```
# Check missing value again
nnull=data.isnull()
nnull.sum()
```

| | |
|--------------------|---|
| Unnamed: 0 | 0 |
| CASE_STATUS | 0 |
| EMPLOYER_NAME | 0 |
| SOC_NAME | 0 |
| JOB_TITLE | 0 |
| FULL_TIME_POSITION | 0 |
| PREVAILING_WAGE | 0 |
| YEAR | 0 |
| WORKSITE | 0 |
| lon | 0 |
| lat | 0 |
| dtype: int64 | |

Processing & Analysis

1

What are the maximum, minimum, average and median wages for H1B application?

2

Which companies filled most of H1B visa applications?

3

Which employer with high applications offers the lowest median wages?

4

Which company has the most number of data-related positions?

5

Explore the salaries for data-related job

6

What is the trend of case status over years?

7

What is the fluctuation of H1B application amount over years?

8

Visualization of H1B application geographic distribution

9

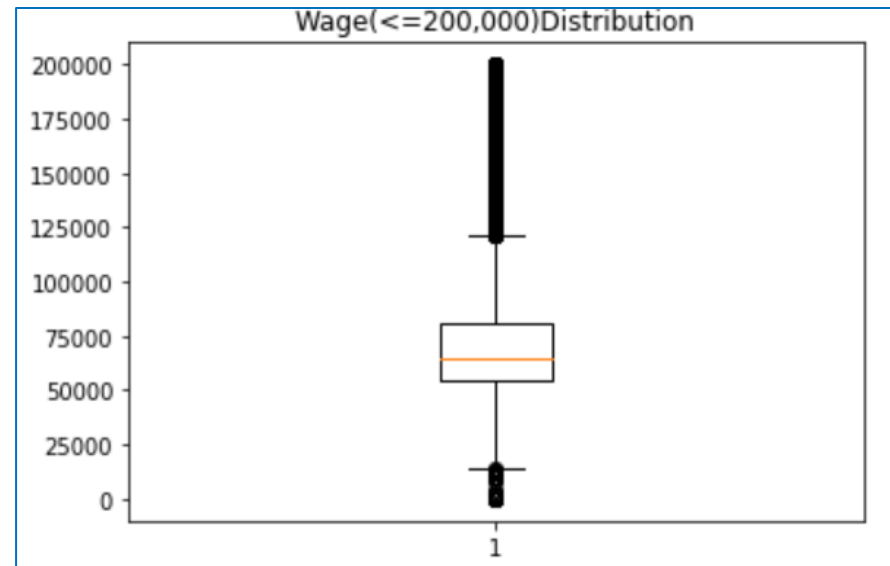
Explore the wage distribution, and what's the relationship between wage and case status?

Processing & Analysis

Q1: Wage distribution for H1B applications



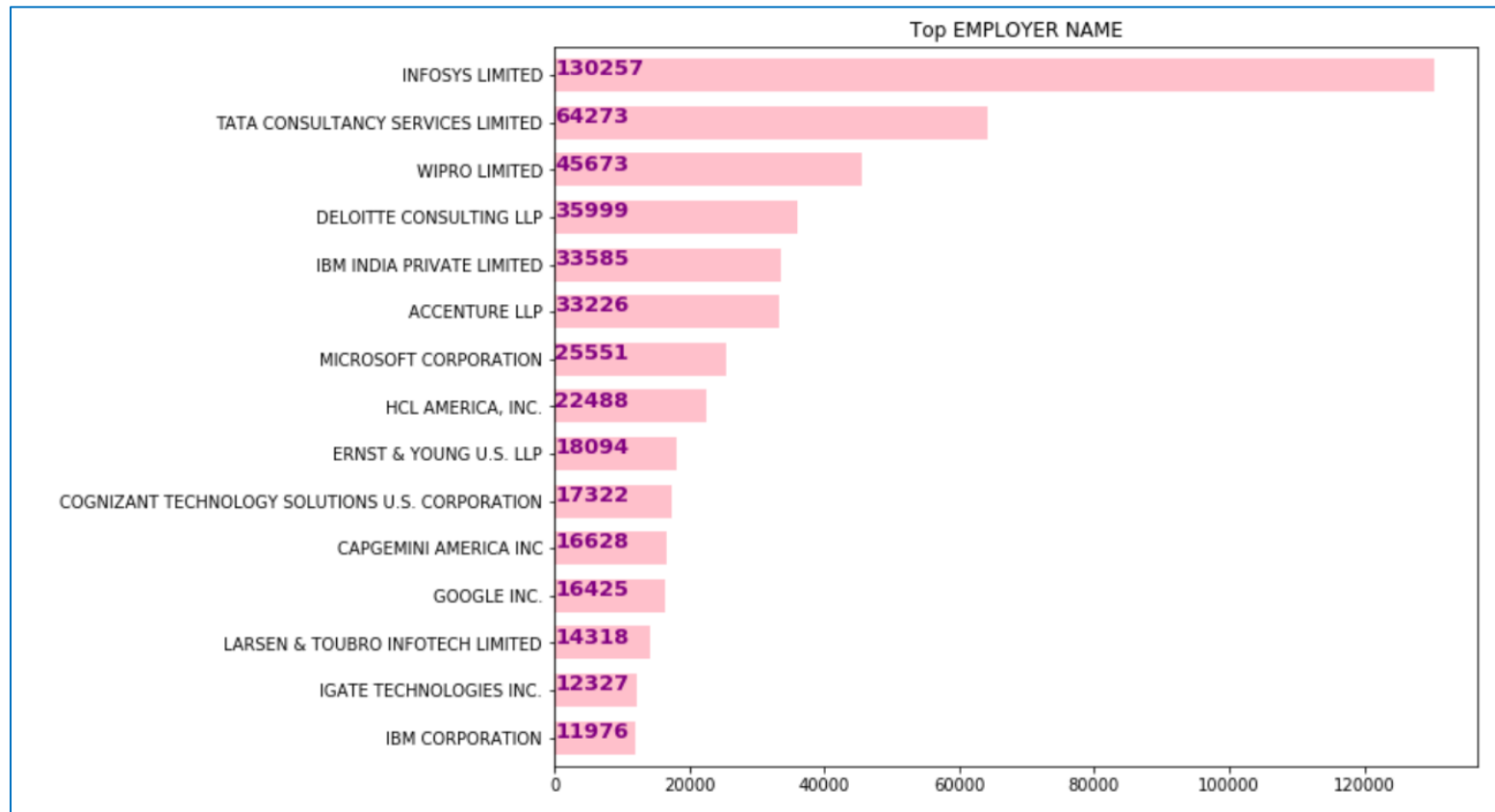
Too many outliers



Median \approx 62,000

Processing & Analysis

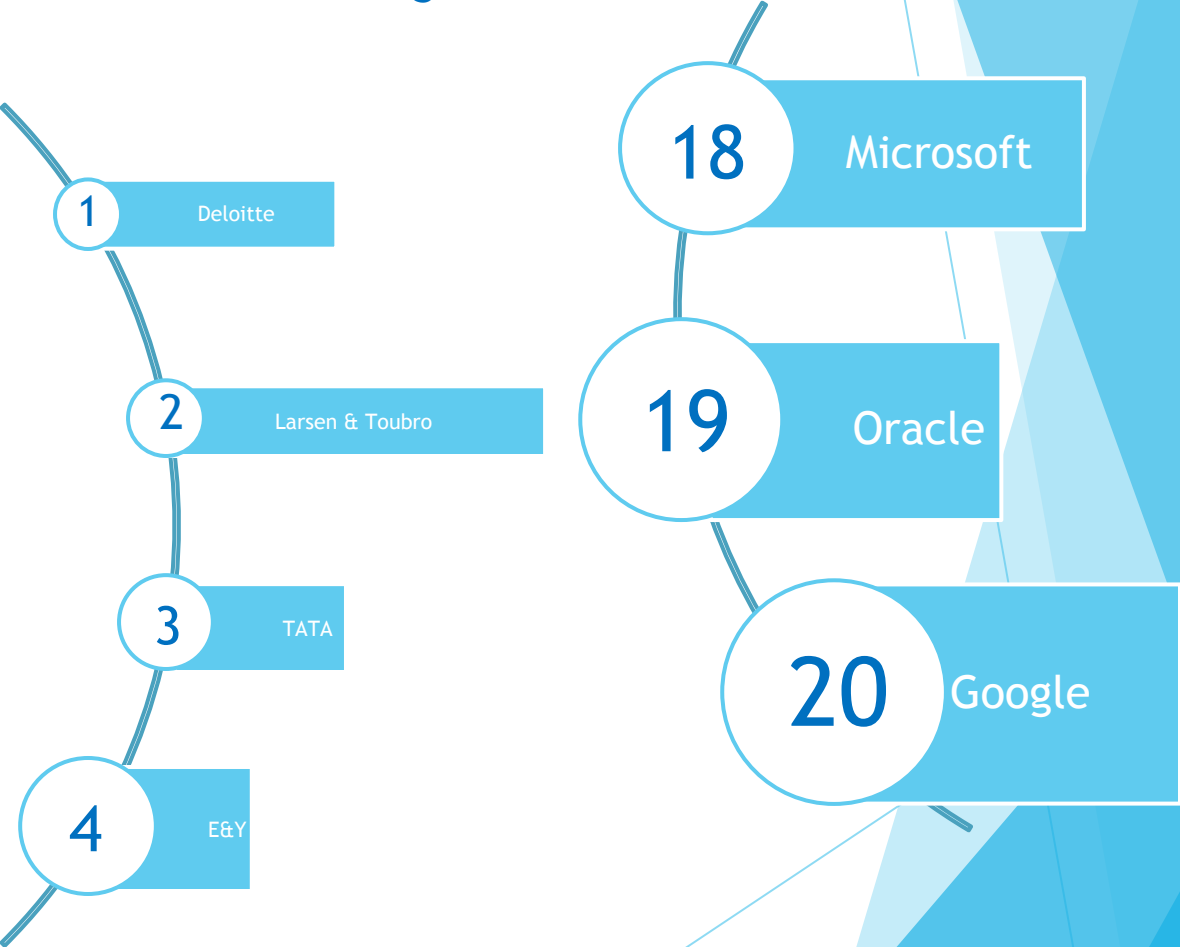
Q2. Which companies filled most of H1B visa applications?



Processing & Analysis

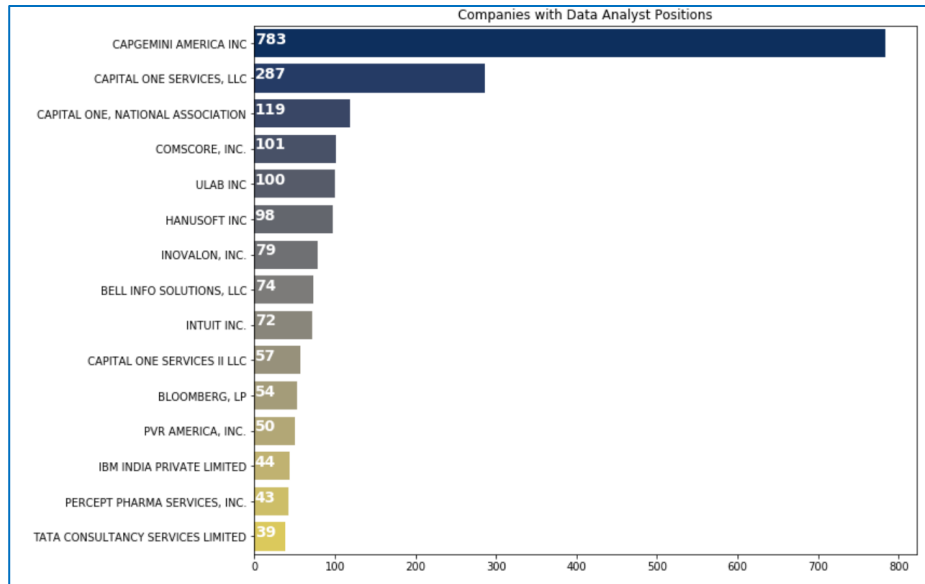
Q3: Which employer with high applications offers the lowest median wages?

| EMPLOYER_NAME | PREVAILING_WAGE |
|---|-----------------|
| DELOITTE & TOUCHE LLP | 55827.0 |
| LARSEN & TOUBRO INFOTECH LIMITED | 56534.0 |
| TATA CONSULTANCY SERVICES LIMITED | 61381.0 |
| ERNST & YOUNG U.S. LLP | 66539.0 |
| COGNIZANT TECHNOLOGY SOLUTIONS U.S. CORPORATION | 67101.0 |
| IGATE TECHNOLOGIES INC. | 67350.0 |
| INFOSYS LIMITED | 67371.0 |
| IBM INDIA PRIVATE LIMITED | 67621.0 |
| WIPRO LIMITED | 68141.0 |
| HCL AMERICA, INC. | 68557.0 |
| TECH MAHINDRA (AMERICAS),INC. | 70782.0 |
| ACCENTURE LLP | 71074.0 |
| DELOITTE CONSULTING LLP | 71510.0 |
| IBM CORPORATION | 75483.0 |
| CAPGEMINI AMERICA INC | 78083.0 |
| INTEL CORPORATION | 85779.0 |
| AMAZON CORPORATE LLC | 86278.0 |
| MICROSOFT CORPORATION | 87547.0 |
| ORACLE AMERICA, INC. | 92539.0 |
| GOOGLE INC. | 99653.0 |

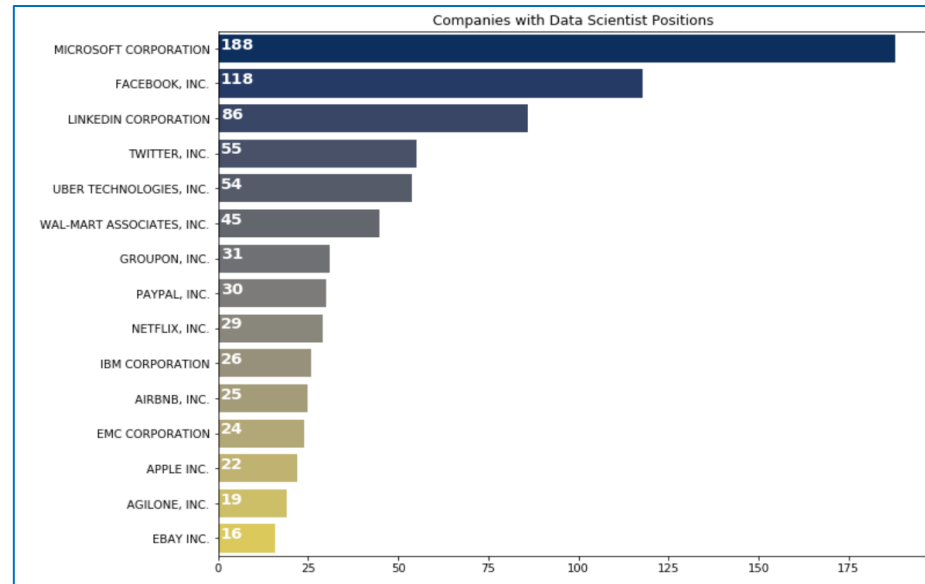


Processing & Analysis

Q4: Which company has the most number of Data related positions?



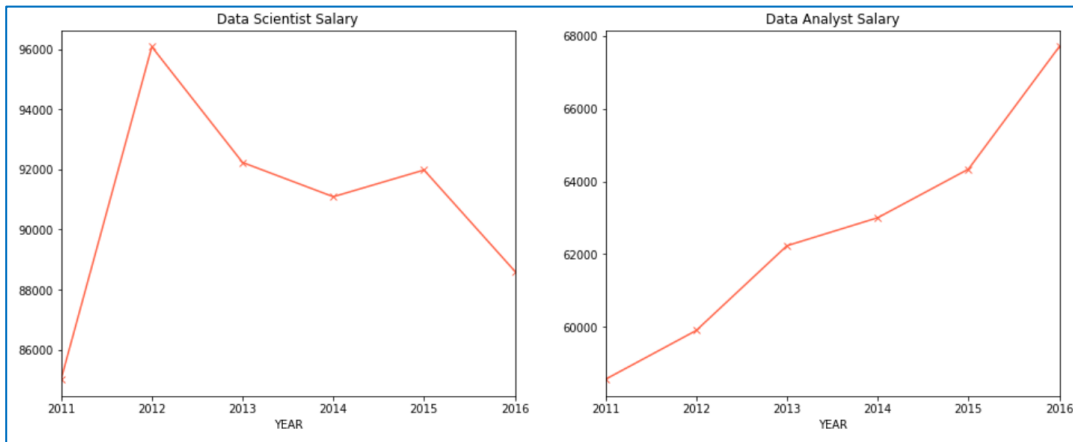
The techniques that data analysts require are not complicated and difficult than those of data scientists, so it is obvious that most data analyst positions are filled by business and financial institutions.



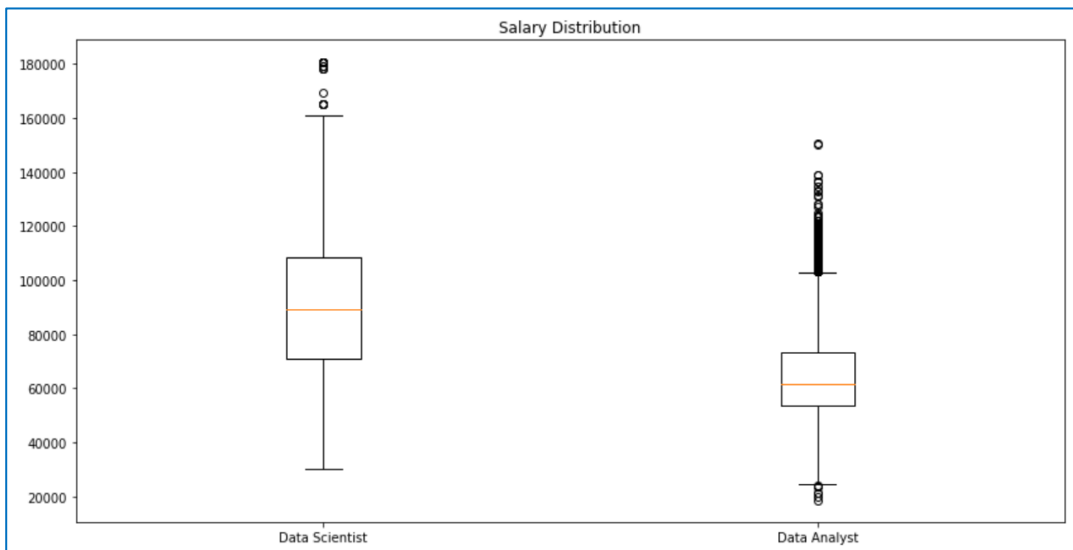
Data scientists are expected to produce more sophisticated outcome, so they are more common in technology industry.

Processing & Analysis

Q5: Explore the salaries for data related job



- The salary for **Data Scientists** has dropped in the past few years, but still above median wage.
- The wage for **Data Analysts** has increased tremendously over the years.

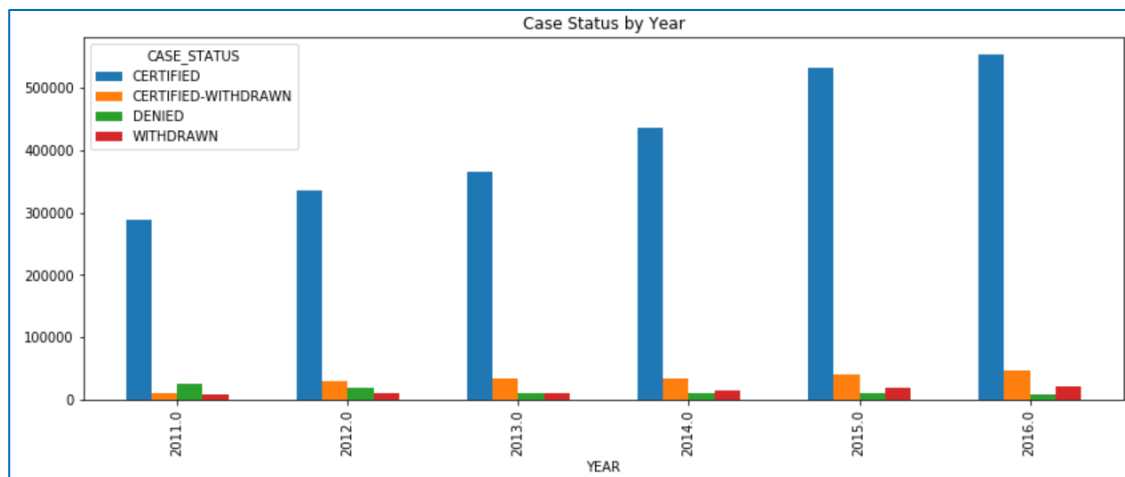
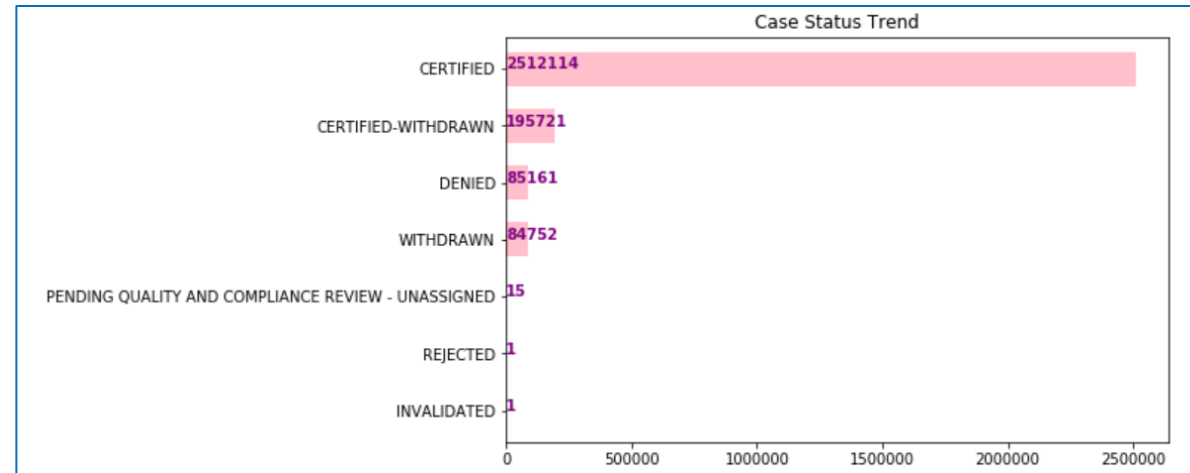


- The median wage of data scientists is about 90K, and the median wage of data analysts is 62k.
- There are more outliers in data analysts.

Processing & Analysis

Q6: What is the trend of case status over years?

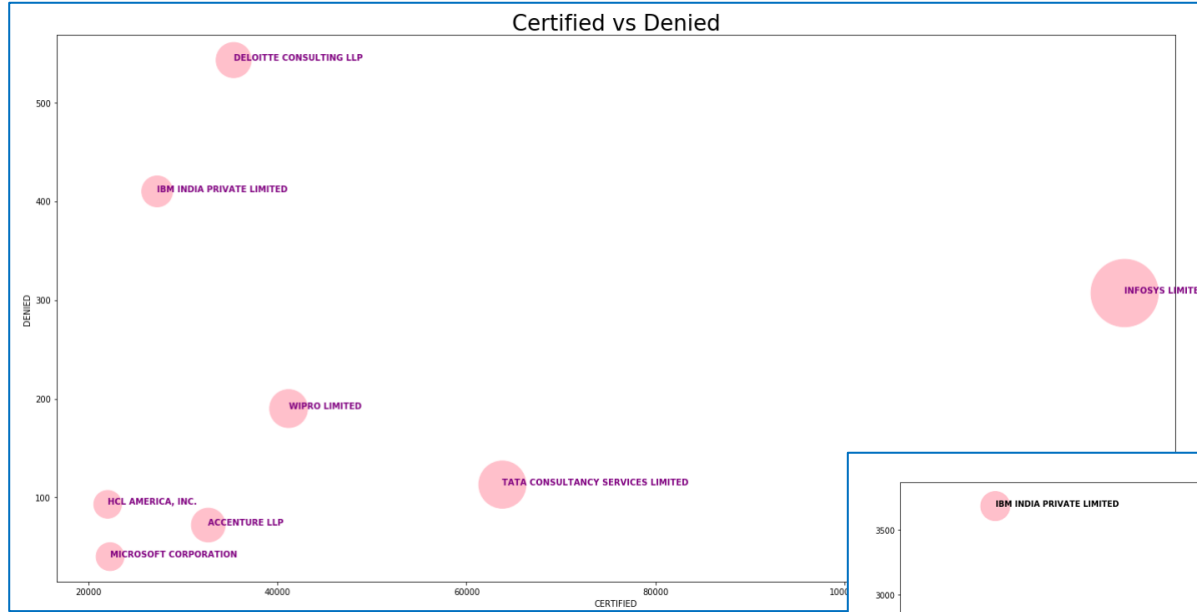
- Certified cases: 2.5 million.
- Denied or withdrawn cases: < 90,000.



- The percentage of certified cases is increasing with a slower speed
- The percentage of denied cases is decreasing with a slower trend.

Processing & Analysis

Q6: What is the trend of case status over years?

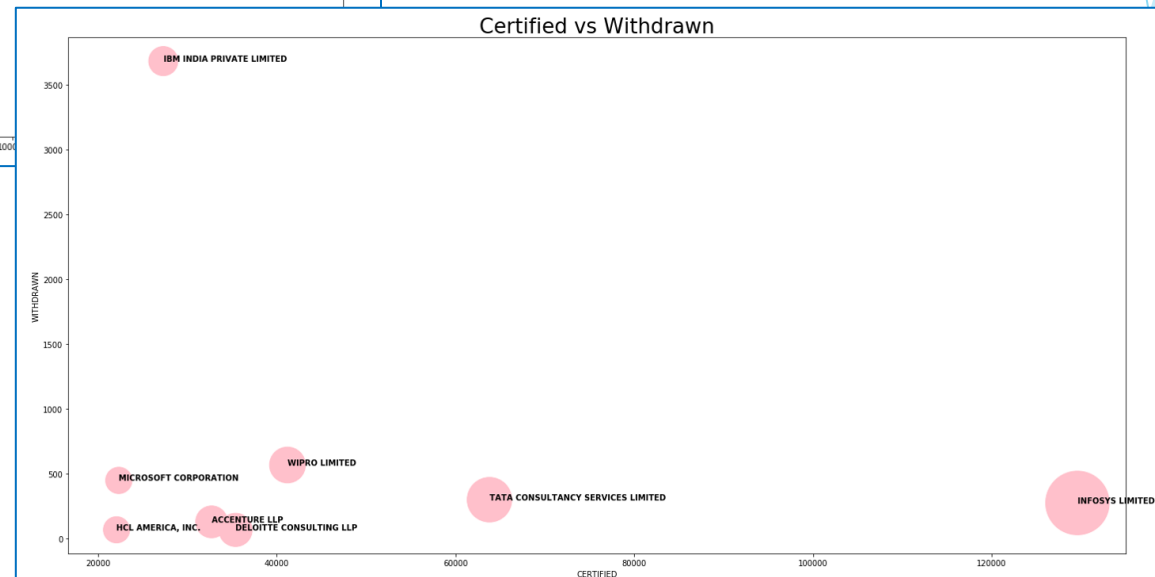


The bubble size represents the amount of corresponding cases.

- INFOSYS LIMITED has the largest amount of certified cases.
- DELOITTE has the largest amount of denied cases.

The scatter map also shows a relative trend.

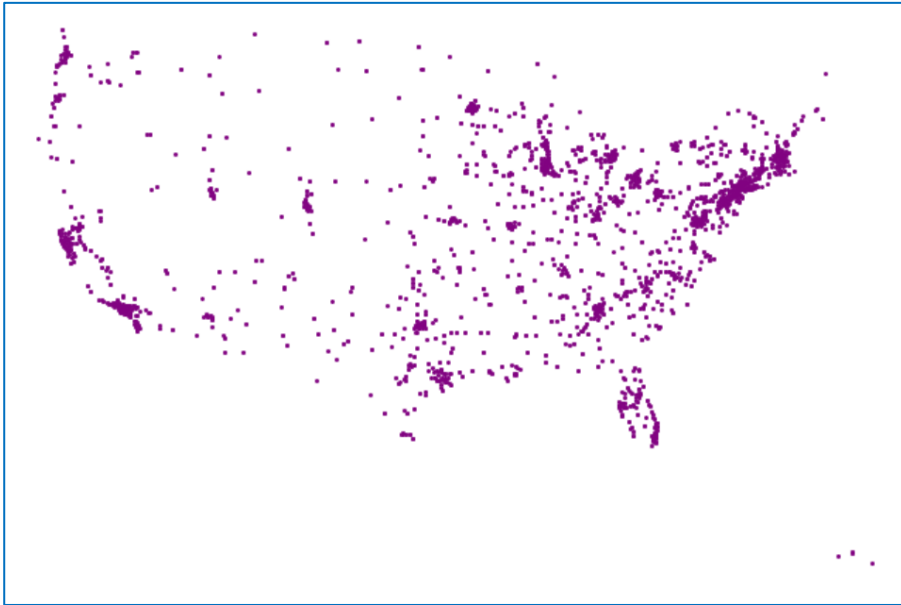
- For example, TATA and INFOSYS have around the same number of withdrawn cases but INFOSYS has over 2 times amount of certified cases.



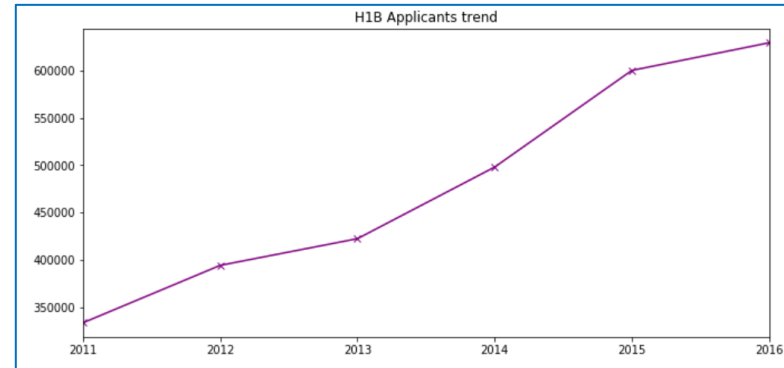
Processing & Analysis

Q7: What is the fluctuation of H1B application amount over years?

Q8: Visualization of H1B application geographic



Density of applicant location

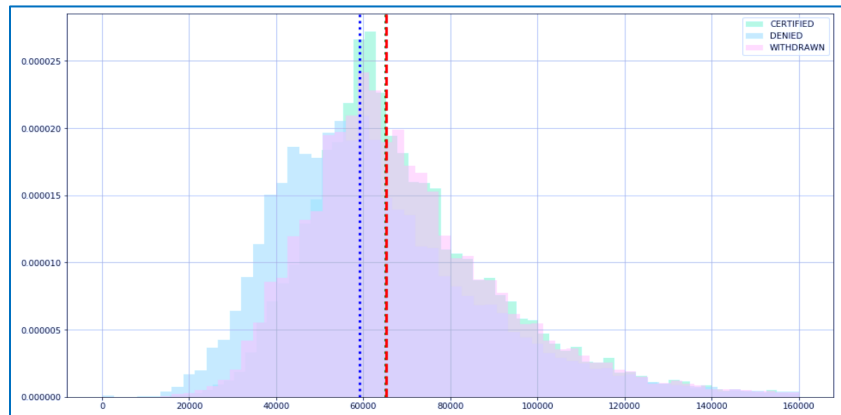
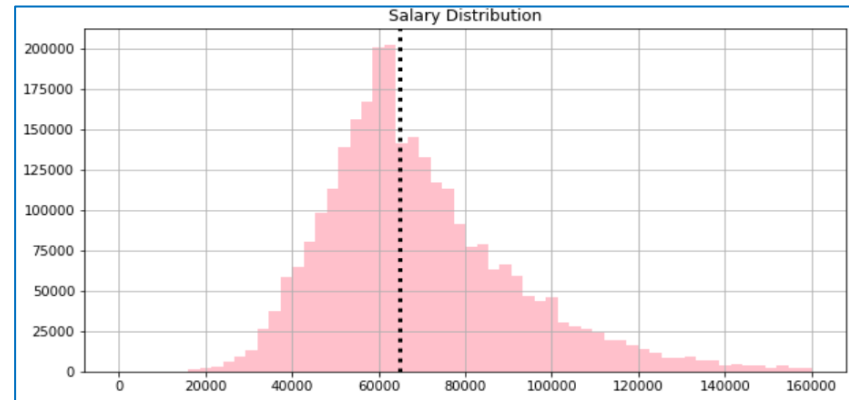


The application amount of H1B is stably increasing from 2011 to 2016, the number nearly doubled through these years.

Processing & Analysis

Q9: Explore the wage distribution, and what's the relationship between wage and case status?

- The distribution is skewed towards the right, which is likely to bias the mean estimate.
- Thus the **median** wage would give a better estimate compared to the **mean** wage.



- The median wage for Denied Cases is lower than that of Certified Cases.
- This might be a reason for application denial-those applicants might have a lower wage than required.

The background features abstract, overlapping geometric shapes in various shades of blue, ranging from light sky blue to deep navy blue. These shapes are primarily located on the left and right sides of the frame, creating a modern, dynamic border around the central text.

Thank You