**Project Report**

**“H1B Visa Data Analysis”**

CONTENTS

1. **Introduction**  
   1. H1B Visa ………………………………………………………………02
   2. Data Analysis ………………………………………………………….03
   3. Power BI……………………………………………………………….03

1. **Literature Survey**  
   1. Why is Data Analysis Important …………………………………….04
   2. Power BI used for ……………………………………….05
   3. Scope of Project………………………………………………………05

1. **System Analysis** 
   1. Import the Dataset ……………………………………………………06
   2. Dataset Cleaning………………………………………………………07
   3. Cleaning Steps ………………………………………………………...07
2. **Feasibility Study** 
   1. Data Transformation…………………………………………….08
   2. Power Query Editor…………………………………………….09

1. **Visualization Charts**
   1. Stacked Column Chart……………………………………….09
   2. Pie Chart……………………………………………………...11
   3. Line Chart ……………………………………………………12
   4. Table …………………………………………………………12
   5. Map ………………………………………………………….13
   6. Donut Chart ………………………………………………….15
   7. Stacked Area Chart ………………………………………….15
   8. Tree Map …………………………………………………….16
   9. Funnel Chart …………………………………………………16
2. **Implementation Phase**

6.1 Analysis …………………………………………………………….17

6.2 Screen Layouts ……………………………………………………...17

* 1. Dashboard ………………………………………………………….31

1. **Conclusion and discussion** …………………………………………….32

**1. Introduction**

**1.1 H1B Visa**

The H1B visa is a non-immigrant visa that allows US companies to employ foreign workers

in specialty occupations. The visa is intended for workers who have “a bachelor’s degree or

higher in a specialty occupation”. To qualify for an H1B visa, the foreign worker must have a

job offer from a US company in a specialty occupation.

**1.2 Characteristics of H1B Visa**

Here are few key aspects:

High Demand: The H-1B visa program typically experiences high demand from both

employers and foreign workers seeking employment in the United States.

Annual Cap: The program has an annual cap on the number of H-1B visas that can be issued.

This cap is set by the U.S. government and is often reached quickly, leading to a lottery

system for selection.

Specialized Occupations: The H-1B visa is designed for individuals with specialized

knowledge or skills in fields such as science, technology, engineering, and mathematics

(STEM), as well as other professional fields.

Education Requirements: Generally, H-1B applicants are required to have at least a bachelor's

degree or equivalent experience in a relevant field.

Employer Responsibility: Employers in the U.S. must sponsor H-1B visa applicants. They are

responsible for filing the necessary paperwork, complying with prevailing wage

requirements, and ensuring a suitable working environment.

Compliance Checks: The U.S. Citizenship and Immigration Services (USCIS) conducts

compliance checks to ensure that employers are meeting their obligations under the H-1B

program, such as paying the prevailing wage.

Green Card Transition: Many H-1B visa holders seek to transition to permanent residency

establish long-term residency in the United States.

Prevailing Wage: Employers sponsoring H-1B workers must pay them the prevailing wage

for the specific occupation and location. This is intended to prevent the undercutting of wages

for U.S. workers.

**1.3 Data Analysis**

Data analysis is the process of cleaning, changing, and processing raw data and extracting

actionable, relevant information that helps businesses make informed decisions. The

procedure helps reduce the risks inherent in decision-making by providing useful insights and

statistics, often presented in charts, images, tables, and graphs.

**1.4 Power BI**

Power BI is a powerful data visualization and analytics tool that can help you quickly make

sense of your data by extracting it from different data sources.

Power BI is a data visualization platform used primarily for business intelligence purposes.

Designed to be used by business professionals with varying levels of data knowledge, Power

BI’s dashboard is capable of reporting and visualizing data in a wide range of different styles,

including graphs, maps, charts, scatter plots, and more.

1. **LITRATURE SURVEY** 
   1. **Why is Data Analysis Important**

Better Customer Targeting: You don’t want to waste your business’s precious time,

resources, and money putting together advertising campaigns targeted at demographic groups that have little to no interest in the goods and services you offer. Data analysis helps you see

where you should be focusing your advertising and marketing efforts.

• You Will Know Your Target Customers Better: Data analysis tracks how well your

products and campaigns are performing within your target

demographic. Through data analysis, your business can get a better idea of your target

audience’s spending habits, disposable income, and most likely areas of interest. This data

helps businesses set prices, determine the length of ad campaigns, and even help project the

number of goods needed.

• Reduce Operational Costs: Data analysis shows you which areas in your business need

more resources and money, and which areas are not producing and thus should be scaled back

or eliminated outright.

• Better Problem-Solving Methods: Informed decisions are more likely to be successful

decisions. Data provides businesses with information. You can see where this progression is

leading. Data analysis helps businesses make the right choices and avoid costly pitfalls.

• You Get More Accurate Data: If you want to make informed decisions, you need data, but

there’s more to it. The data in question must be accurate. Data analysis helps businesses

acquire relevant, accurate information, suitable for developing future marketing strategies,

business plans, and realigning the company’s vision or mission

**2.2 Power BI used for**

Whether you’re a data pro or are just entering the business world, Power BI is designed to

empower you with data-driven insights. Some of the most common uses for the platform

include:

• Creating reports and dashboards that present data sets in multiple ways using visuals

• Connecting various data sources, such as Excel sheets, onsite data warehouses, and cloud-

based data storage, and then transforming them into business insights

• Turning data into a wide range of different visuals, including pie charts, decomposition

trees, gauge charts, KPIs, combo charts, bar and column charts, and ribbon charts – among

many other options

• Providing company-wide access to data, data visualization tools, and insights in order to

create a data-driven work culture

**2.3 Scope of Project**

Analyzing the H-1B visa program involves examining various aspects that have implications

for employers, foreign workers, policymakers, and the broader economy. The scope of H-1B

visa analysis encompasses several key areas:

Labor Market Dynamics:

Skills Gap: Analyzing the H-1B program involves understanding how it addresses skill

shortages in the U.S. job market, particularly in specialized fields like technology and STEM

(Science, Technology, Engineering, and Mathematics).

Impact on U.S. Workers: Assessing the impact of H-1B visa holders on the wages and

employment opportunities of U.S. workers is an important aspect of analysis.

Economic Impact:

Innovation and Productivity: Researching how the H-1B program contributes to innovation

and productivity in the U.S. economy, especially in technology and research sectors.

Entrepreneurship: Examining whether H-1B visa holders contribute to entrepreneurship by

starting new businesses and creating jobs in the United States.

Employer Practices and Compliance:

Employer Behavior: Analyzing how employers utilize the H-1B program, including their

hiring practices, adherence to prevailing wage requirements, and overall compliance with program regulations.

Workplace Conditions: Evaluating the working conditions and treatment of H-1B workers to

ensure they are not subjected to exploitation or unfair labor practices.

Policy Considerations:

Regulatory Changes: Staying abreast of changes in H-1B visa policies, regulations, and

legislation at both the federal and state levels.

Program Reforms: Analyzing proposals for reforms to the H-1B program and their potential

impact on immigration policy and the economy

1. **SYSTEM ANALYSIS** 
   1. **Import the Dataset**

Initially, we identify the type of data we are working with, such as Excel data, SQL Server

data, web data, or Power BI datasets. Once determined, we proceed to load and transform the

data, initiating the cleaning process as the first step.

**3.2 Dataset Cleaning**

There are some steps we use to clean data in Power BI, which are helpful in creating meaningful insights.

1. Use the first row as a header.

2. Remove duplicate rows.

3. Remove duplicate and null columns.

4. Remove blank rows.

5. Check for null and blank values; if they exist, delete them.

6. Check data types; always convert numerical values into decimal or fixed decimal values. 7. Check date data types and convert them into the US and UK time zones using their respective locales

**3.3 Cleaning Steps**

Here, I commence the cleaning process for H1B visa data. The initial steps include:

1. Using the first row as the header.

2. Renaming the default column index as 'Unnamed.'

3. Changing 'Employer Name' to 'Company Name.'

4. Removing duplicate and blank rows.

5. Checking for null and blank values and deleting them if they exist.

6. Converting the data types of each column; numerical values default to decimals, and dates

are adjusted to their local zones, such as US and UK.

7. For this H1B dataset, I created a duplicate column for 'Worksite' and subsequently split the states and cities.

Following these steps, I close and apply the transformed dataset, where I proceed to conduct insightful analysis.

1. **Feasibility**

**4.1 Data Transformation:**

**Power BI Desktop has three views**:

**Report view** – You can use queries that you create to build compelling visualizations, arranged as you want them to appear, and with multiple pages, that you can share with others.

**Data view** – See the data in your report in data model format, where you can add measures, create new columns, and manage relationships.

**Model view** – Get a graphical representation of the relationships that are established in your data model, and manage or modify them as needed.

Access these views by selecting one of the three icons along the left side of Power BI Desktop. In the following image, Report view is selected, indicated by the yellow band beside the icon.

Power BI Desktop also comes with Power Query Editor. Use Power Query Editor to connect to one or many data sources, shape and transform the data to meet your needs, then load that model into Power BI Desktop.

**4.2 Power Query Editor**

To get to Power Query Editor, select Transform data from the Home tab of Power BI Desktop.

With no data connections, Power Query Editor appears as a blank pane, ready for data.

After a query is loaded, Power Query Editor view becomes more interesting. If you connect to a Web data source using the New Source button in the top left, Power Query Editor loads information about the data, which you can then begin to shape.

To connect to data and begin the query building process, select New Source. A menu appears, providing the most common data sources.

**Advanced Editor**

The Advanced Editor lets you see the code that Power Query Editor is creating with each step. It also lets you create your own code in the Power Query M formula language. To launch the advanced editor, select View from the ribbon, then select Advanced Editor. A window appears, showing the code generated for the selected query.

1. **Visualizations**

**5.1 Stacked Column Chart & Stacked Bar Chart**

Both are most usable visuals in Power BI. Stacked Column Chart is useful to compare multiple dimensions against a single measure. In a stacked column chart, the vertical axis represents the numerical values of the data, while the horizontal axis displays the categories or time periods.

Stacked column charts are useful for comparing the total values of different groups, as well as for identifying the contribution of each group to the overall total.

**How to formatting Stacked bar Chart?**

Visual Tab:

X-axis : Mange the X-axis Value font size, color, Title and Display Unit etc.

Y-axis: Mange the Y-axis Value font size, color and Title.

Legend: Specified the Legend Text color, font size, Position & Title.

Bars: Change the color of chart & maintain the spacing between bars.

Data Labels: Enable the data labels on chart, manage the display unit of labels.

Total Labels: Enable Total labels on chart.

General Tab:

Properties: In this section you can manage the chart height, width, horizontal & vertical position.

Title: Under general tab you can see the title section> Here you can set the below properties-

Title- Specified the title for chart, & manage the font size, color, background for chart.

Subtitle- Specified the subtitle for chart, & manage the font size, color, background for chart.

Divider- Enable the line between Title & chart.

Spacing- Manage the space between title, Subtitle & chart area.

**Key features of a Stacked Bar Chart in Power BI include:**

* Multiple Categories: Each bar on the chart represents a specific category, and the individual segments within the bar represent the subcategories or components.
* Stacked Segments: Subcategories are stacked on top of each other within a single bar, illustrating the contribution of each subcategory to the total of that category. o
* Comparison and Composition: Stacked bar charts allow you to compare both the overall distribution of categories and the distribution of subcategories within each category.
* Color Coding: Different subcategories are typically assigned distinct colors to aid in differentiation.
* Legend: The legend helps identify the subcategories and their colors.
* Tooltip: Hovering over a segment provides detailed information about the value it represents

**5.2 Pie Chart in Power BI**

Pie Chart in Power BI is a built-in visualization chart available with all versions of Power BI. The pie chart is a round-shaped circle chart where each category data set is shown in a pie shape based on the value of each data label. In addition, the pie chart of each category is adjusted against the overall portion of the data labels.

Legend: This is nothing based on what column values we need to see the pie chart. In our example, we need to see the “Region-wise” pie chart. Our legend will be the “Region” column from “Pie Table.”

Details: If you want to show any further partition of the data you can add here, we will come back to this in the second example.

Values: This is nothing but what numerical value we need to show in the pie circle.

Tooltips: This is nothing but adding an extra element to the pie chart when we hover on any of the slices of the pie.

**5.3 Line Chart in Power BI**

A line chart is a sequence of data points defined by dots and joined by straight lines. A line chart may include one or multiple lines. It is one of the visuals to display the data in the Power BI report in a visually immersive and interactive manner.

It is used to define continuous data sets and the line charts maintain an X and a Y-axis in Power BI.

**Multiple lines on a Line chart in Power BI**

Let us see how to display multiple lines plotted on a Line Chart visual in Power BI,

Yes, it is possible to display the line chart with multiple lines plotted on it.

**To achieve this, follow the below steps:**

* Initially, make sure that you have loaded the data source to the Power BI desktop.
* Once the data is loaded you can confirm the data under the Fields section and the fields pane contains the column data presented in the source data.
* Now under the Visualizations pane, select the Line Chart option, you can see that the line chart is added to the report canvas.

**5.4 Tables in Power BI reports**

A table is a grid that contains related data in a logical series of rows and columns. A table can also contain headers and a row for totals. Tables work well with quantitative comparisons where you're looking at many values for a single category. In the following example, the table displays five different measures for the Category items, including average prices, year over year sales, and sales goals.

Power BI helps you create tables in reports and cross-highlight elements within the table with other visuals on the same report page. You can select rows, columns, and even individual cells, and then cross-highlight the values. You can also copy and paste individual cells and multiple cell selections into other applications.

When to use a table

Tables are a great choice for several scenarios:

Represent numerical data by category with multiple measures.

Display data as a matrix or in a tabular format with rows and columns.

Review and compare detailed data and exact values rather than visual representations.

**5.5 Map visualization**

The map visualization in Power BI is a useful tool for analyzing spatial data. It allows you to create a map that represents your data in a way that is easy to understand and visually compelling. The map visualization works by integrating with Bing Maps, which provides a rich set of map styles and geographic data.

The map visualization gives you the ability to add multiple layers and data points to your map, customize the map's appearance, and create drill-down hierarchies for deeper analysis.

One of the key features of the map visualization in Power BI is the ability to use custom map layers.

This means that you can import your own geographic data, Additionally, the map visualization supports a wide range of data types, including latitude and longitude coordinates, addresses, and even postal codes, making it a versatile tool for spatial analysis.

1. Categorize geographic fields

In Power BI Desktop, you can ensure fields are correctly geo-coded by setting the Data Category on the data fields. In Data view, select the desired column. From the ribbon, select the Column tools tab and then set the Data Category to Address, City, Continent, Country, County, Postal Code, State, or Province.

2. Use more than one location column

Sometimes, even setting the data categories for mapping isn't enough for Bing to correctly guess your intent. Some designations are ambiguous because the location exists in multiple countries/regions. For example, there's a Southampton in England, Pennsylvania, and New York.

3. Use specific Latitude and Longitude

Add latitude and longitude values to your semantic model. This data removes any ambiguity and returns results more quickly. Latitude and Longitude fields must be in Decimal Number format, which you can set in the data model.

4. Use Place category for columns with full location information

While we encourage you to use geo-hierarchies in your maps, if you must use a single location column with full geographical information, you can set the data categorization to Place. For example, if the data in your column is full addresses, such as 1 Microsoft Way, Redmond Washington 98052, this generalized data category works best with Bing.

**In Power BI: tips to get better results when using map visualizations**

1. Use latitude and longitude fields (if they exist)

In Power BI, if the semantic model you are using has fields for longitude and latitude--use them! Power BI has special buckets to help make the map data unambiguous. Just drag the field that contains your latitude data into the Visualizations > Latitude area. And do the same for your longitude data. When you add this data, you also need to fill the Location field when creating your visualizations. Otherwise, the data is aggregated by default, so for example, the latitude and longitude would be paired at the state level, not the city level.

**5.6 Donut Chart**

A donut chart is a circular chart, which could present values of a dataset in the form of slices of a donut. The donut chart is exactly the same as a pie chart, the only difference is pie chart has a circle, but a donut chart has a hole in the circle. We have various options to format donut charts, we can change the value of the legends, rotation, detail labels, etc. In this article, we will learn how to format a donut chart in Power BI and explore its various options.

**Formatting a Donut Chart In Power BI**

After the successful, creation of a donut chart in Power BI, we have multiple options to format it. For example, adding the title to the chart, changing the color, and position of the chart, and adding tooltips, slicer colors, and detail labels to the chart.

**5.7 Stacked Area Chart**

A stacked area chart is formed by combining the line chart, with the shaded area under that line. This chart is generally, used when we want to see the trends, that which field is performing better, in a particular time frame. For example, considering the stock prices of different companies, in the past 5 months, then a stacked area chart can be very useful, to see this trend. In this article, we will learn how to create a stacked area chart in Power BI.

**Creating a Stacked Area Chart**

A Stacked Area chart has multiple options while creating, and customizing it. We will take a look at each of the options. For example, we are given a data set of Employees, and we want to make a Stack Area chart, consisting of legends, and small multiples, segregated by year. We will explore each option while creating this stacked area chart.

**5.8 Tree Map**

TreeMap in Power BI is the hierarchical chart, which is used to show the parent and child data distribution. TreeMap is shown by a group of rectangles, these rectangles are segregated on the basis of the category. The larger numeric values are present at the top, and lower numeric values are present at the bottom. In this article, we will learn how to create a TreeMap in Power BI using steps.

**Loading Dataset for a TreeMap**

When generating and designing a treemap, there are several options available. We’ll examine all of the possibilities. We will use a given dataset of employees, for instance, we could wish to create a treemap with the following categories: Department, Details, Employee Name, Values, and Bonus. While building this treemap, we will investigate each possibility.

**5.9 Funnel Chart**

A Funnel Chart is a type of chart that is used to represent how the data moves through a process or system. It shows how data flows throughout all the stages of some process which is easy to read and understand. It represents a Linear process with sequential and connected stages.

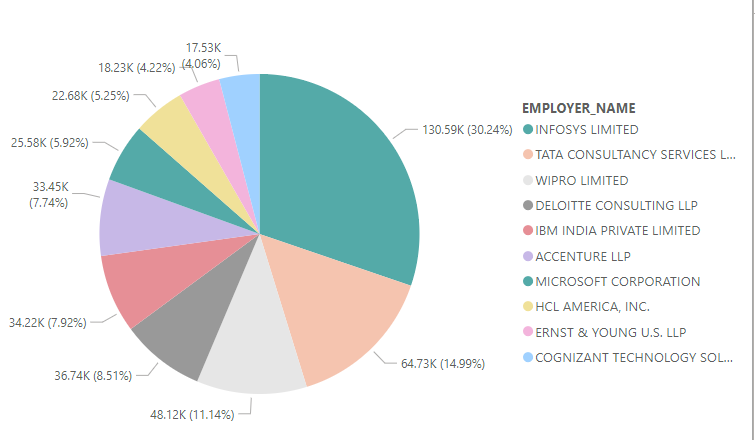
A Funnel chart looks like a broad head and narrow neck at the bottom showing some data flow in a chart. Funnel charts are widely used to represent the sales funnels, recruitment process, and item order fulfillment process which means multiple stages of a whole long process.

**6. Implementation Phase**:

**6.1 Analysis:**

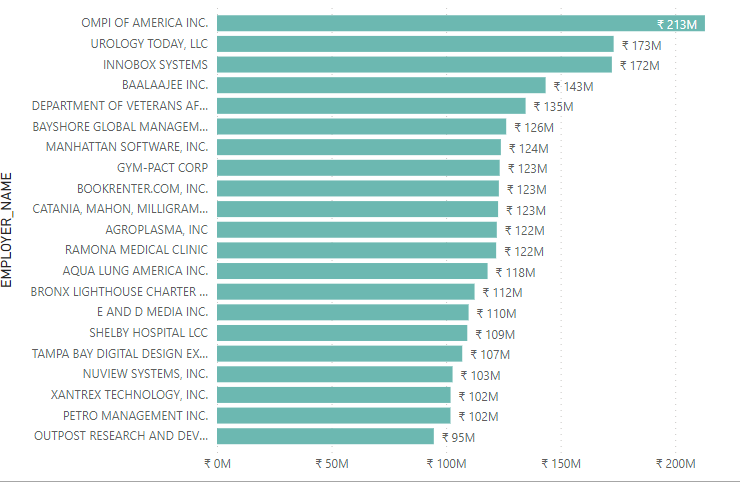
**1. Top 10 Hiring Companies:**

Create visualizations to showcase the top 10 hiring companies.



**Analysis:** Firstly, input the company names and their respective counts. Next, apply basic filtering by selecting the "Top N" option and entering the value 10. Then, drag the company names into the designated area. Finally, observe the results to identify the top 10 hiring companies. Who grants the visa for our employees to work in other branches in another country

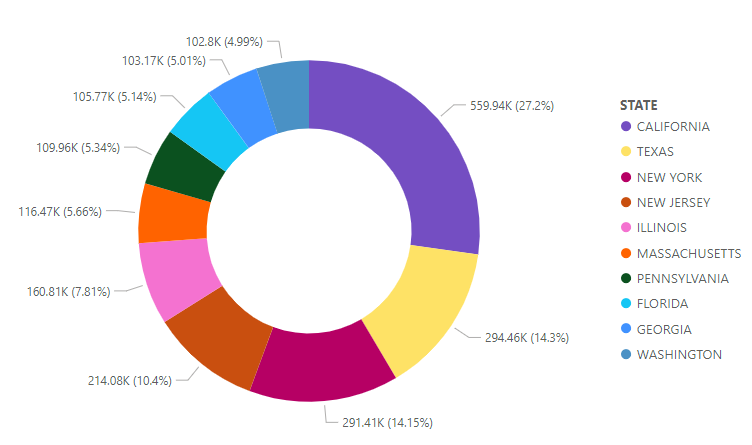
**2. Average Prevailing Wage by Employer:**



**Analysis:** Here, input the company names and Average of Prevailing Wage their respective. Then, drag the company names into the designated area. Finally, observe the results to identify the Average Prevailing wage by Company name. I found out the prevailing wage of employees who work in branches of the company in other countries.

1. **Top 10 States with Highest Application Counts:**

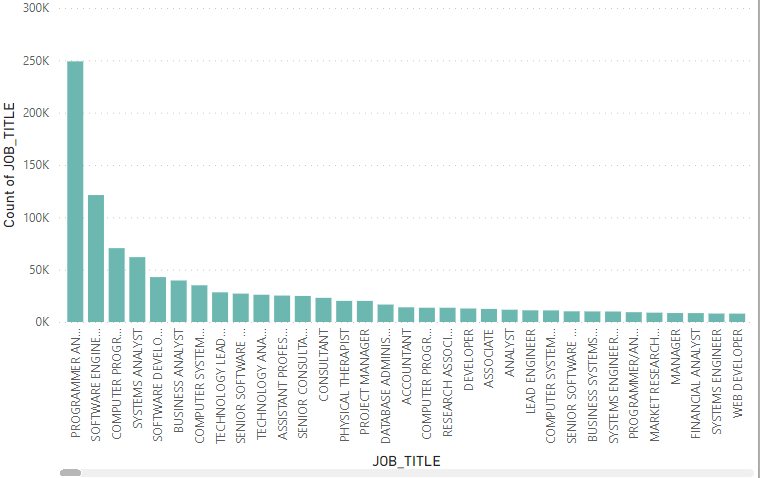
If state information is available, create visualizations to display the top 10 states with the highest number of visa applications.



**Analysis:** I used a Donut chart with a column for the state and its corresponding count. I identified the top 10 states with the highest application counts.

1. **Distribution of Applications by Job Titles:**

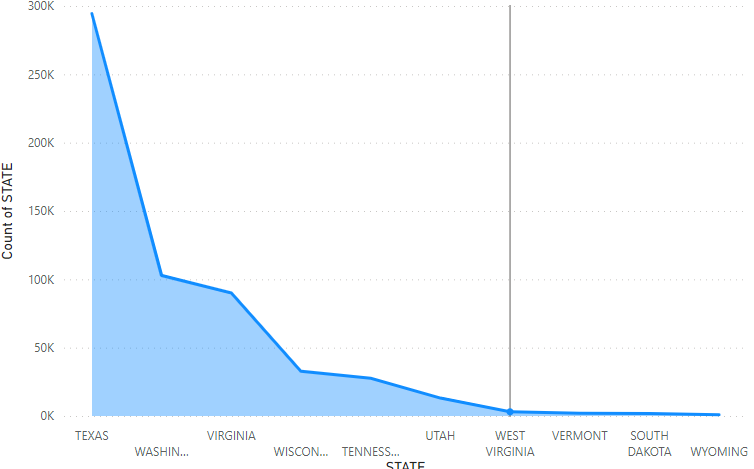
Develop visualizations to explore the distribution of applications based on job titles



Analysis: I used a stacked bar chart with a column for job\_title and its corresponding count. I analyzed the distribution of applications based on their job titles.

1. **Top 10 States with Highest Denied Petitions:**

If state information is available, create visualizations to display the top 10 states with the highest number of denied visa petitions.

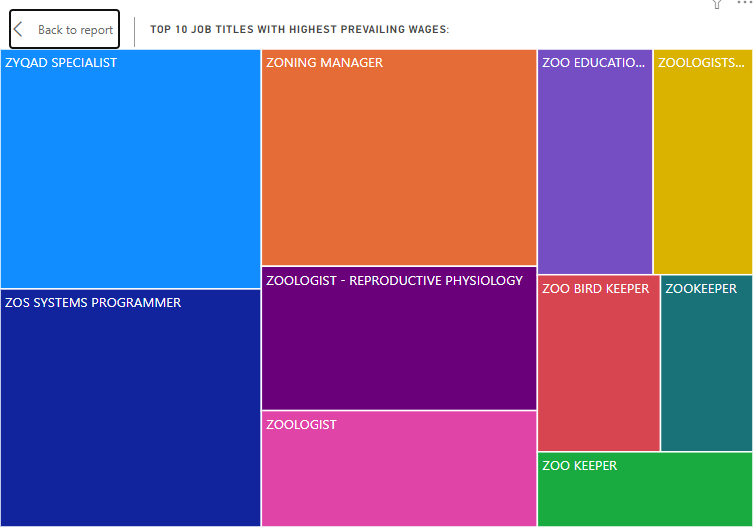


**Analysis:** Here, we have a column for state and the count of states. We identified the top 10 states with the highest number of denied petitions.

**What is Denied Petitions?**

"Denied petitions" typically refers to applications or requests that have been formally rejected or refused. In the context you provided, if you're dealing with immigration or legal matters, denied petitions would likely refer to applications or requests that have not been approved. For example, if someone applied for a visa or a work permit, a denied petition means that their application was not accepted, and they did not receive the permission they were seeking.

1. **Top 10 Job Titles with Highest Prevailing Wages**



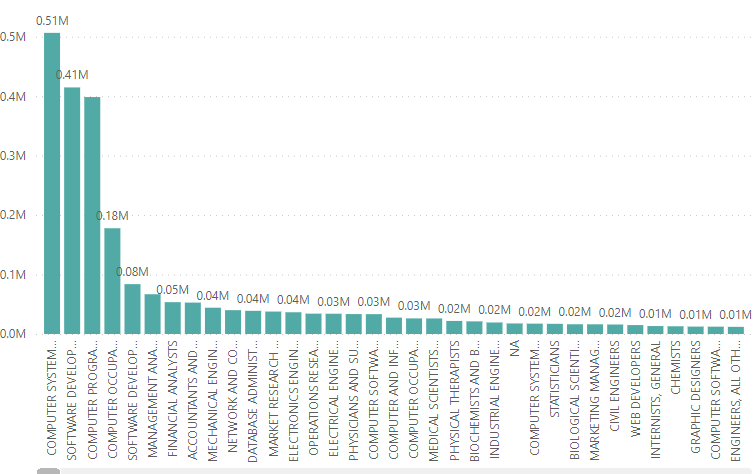
**Analysis:** Here, I have columns for job\_title and prevailing\_wage, and I identified the top 10 job titles with the highest prevailing wages

**What is meaning of Prevailing Wages?**

"Prevailing wages" refer to the average wages paid to workers in a specific occupation, typically within a particular geographic area. These wages are determined by government agencies or regulatory bodies and are meant to reflect the standard pay rates for various jobs in a given location. The idea behind prevailing wages is to ensure that workers are fairly compensated in a particular region and to prevent the undercutting of local labor markets by employers paying significantly lower wages. Prevailing wages are often associated with government contracts and projects, and they aim to maintain a level playing field for workers and contractors.

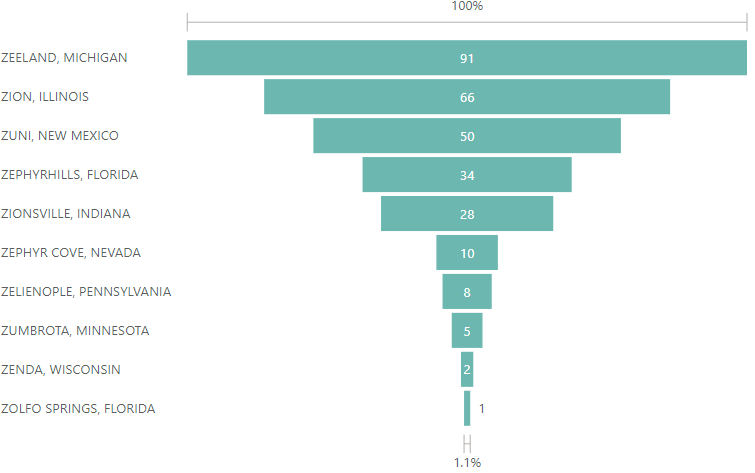
1. **Occupation Analysis:**

Explore the distribution of applications based on different job occupations



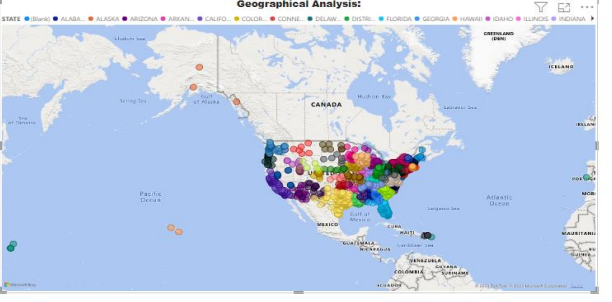
**Analysis:** I used a stacked bar chart with a column for Soc\_name and its corresponding count. I analyzed the occupation analysis

1. **Top 10 Worksite Locations**



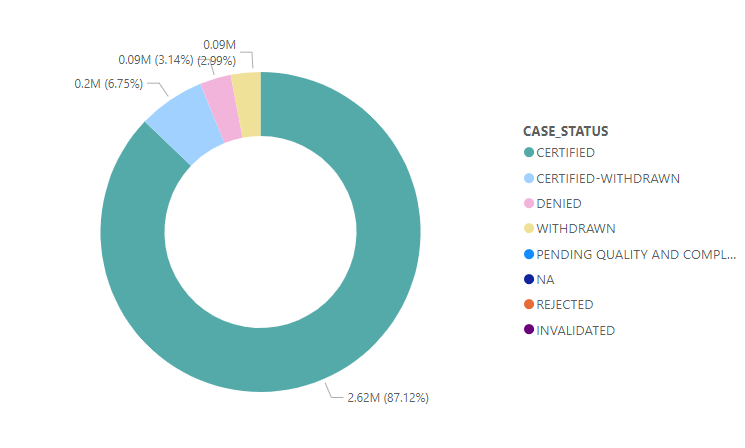
**Analysis:** Firstly, input the worksite location and their respective counts. Next, apply basic filtering by selecting the "Top N" option and entering the value 10. Then, drag the worksite into the designated area. Finally, observe the results to identify the top 10 worksite location.

1. **Map visualization**



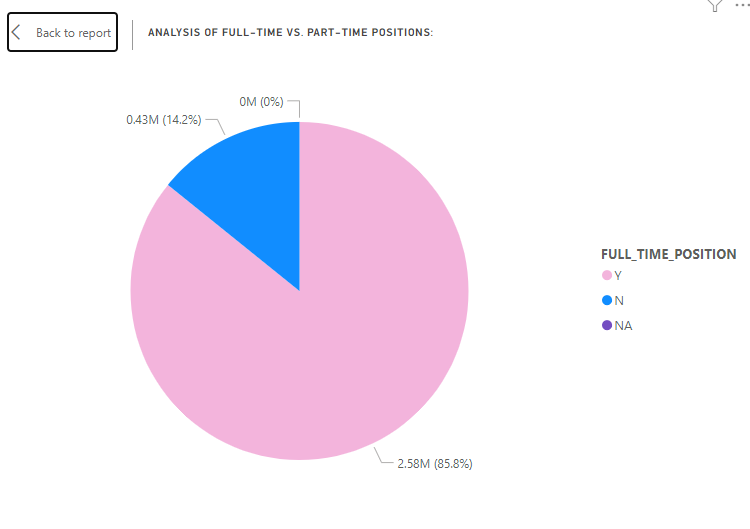
**Analysis:** Here, I utilized a map for geographical analysis, initially capturing data on states, latitudes, and longitudes. Within the available columns, I sought to identify the state where the majority of applicants are employed.

**11.Case Status Distribution:**



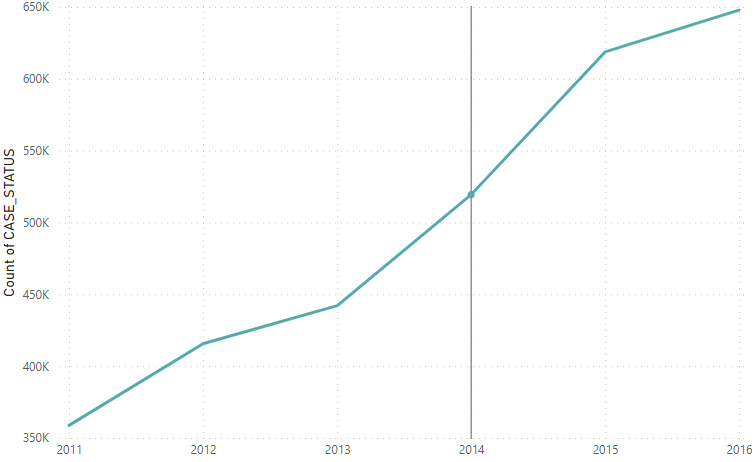
**Analysis:** I employed a pie chart, utilizing the basic columns of "case\_status" and the corresponding count of each status. This analysis aimed to identify the distribution of employment certifications, withdrawals, denials, and rejections within the company, as illustrated.

**12.Analysis of Full-Time vs. Part-Time Positions:**



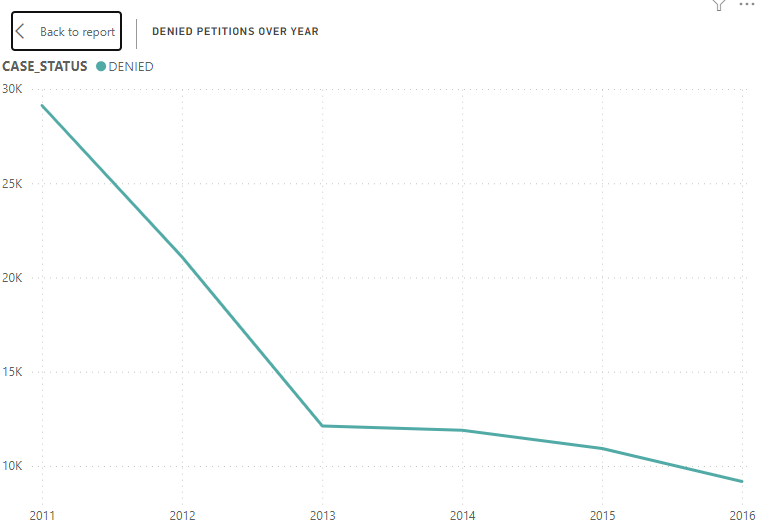
**Analysis:** I employed a pie chart, utilizing the basic columns of "Full\_Time\_Position" and the corresponding count of each status. This analysis aimed to identify the analysis of full-time vs. analysis of part-time position, as illustrated.

**13.Visa Application Trends Over the Years:**



**Analysis:** Here, I conducted an analysis of visa applications over the years, examining the data based on the "year" column and its corresponding case statuses. This analysis includes visa applications over year, quarter and month.

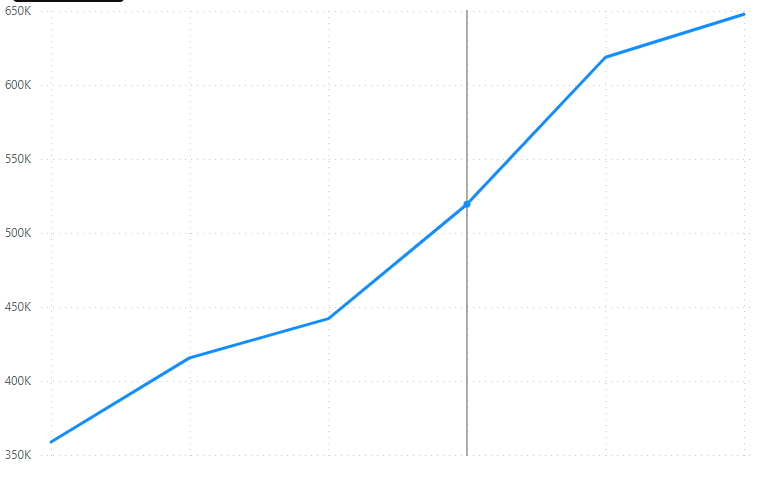
**14.Denied Petitions Over the Years:**



**Analysis:** Here, I conducted an analysis of visa applications denied petition over the years, examining the data based on the "year" column and its corresponding case statuses. This analysis includes a denied petition over year.

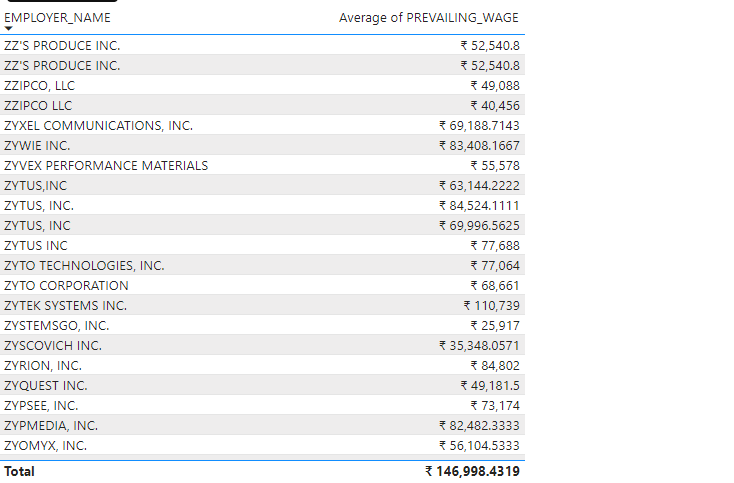
**15. Petitions by Year:**

Generate visualizations to showcase the number of petitions filed each year.



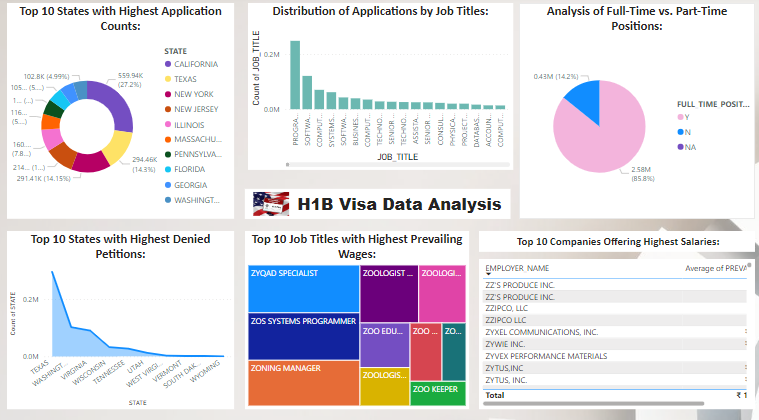
**Analysis:** Here, I conducted an analysis of visa applications petition over the years, examining the data based on the "year" column and its corresponding case statuses. This analysis includes a breakdown petition over year.

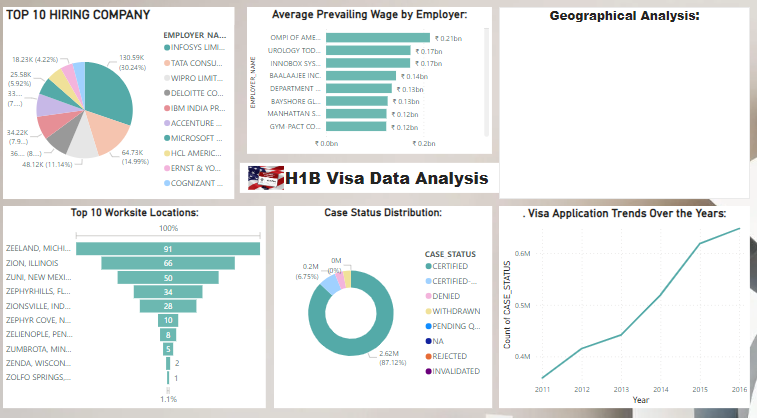
**17. Top 10 Job Titles with Highest Prevailing Wages:**

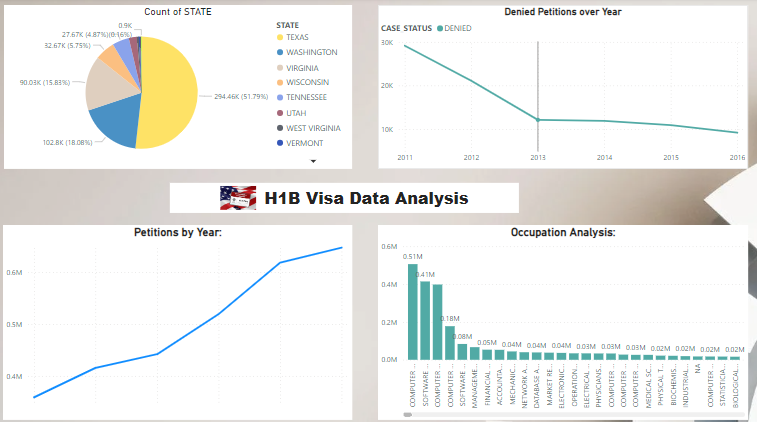


**Analysis:** Here, we display records in the form of tables with columns for company names, and we conduct an analysis to identify the top 10 companies with the highest salaries.

**6.3 Dashboards:**







1. **Conclusion**

The project taught us many innovative things to work on. In conclusion, the H1B visa analysis project involved an in-depth exploration of various facets within the dataset. The investigation focused on identifying the top 10 companies facilitating visa applications, understanding the distribution of approved petitions versus denied ones, analysing job titles associated with these applications, and pinpointing the prime worksite locations where employees engage in both part-time and full-time roles across America.

Additionally, scrutiny was applied to discern the certification status of visas, tracking the number of applications over the years and quarters. This comprehensive analysis unearthed meaningful insights that not only contribute to understanding the dynamics of H1B visa applications but also provide valuable information for companies aiming to enhance their application processes.

The findings serve as a guide for improving visa application forms and expanding opportunities for employees to realize their career aspirations in dream companies. The project underscores the importance of continuous analysis and improvement in facilitating the visa application process, thereby fostering a more seamless experience for both employers and applicants