PROJECT REPORT ON

H – 1B VISA DATA ANALYSIS

"Impact of COVID – 19 on the VISA"

By

SHASHANK SINHA

DATA RANGE: **Q1 – 2019 : Q3 – 2024**

Tools Used

Python | Power BI | MS Office

ABSTRACT

The H-1B visa, pivotal in shaping the technological and innovation landscape of the United States, experienced profound disruptions during the COVID-19 pandemic. This project, driven by an acute awareness of the visa's significance and the business challenges heightened by the global health crisis, aimed to analyze and unravel the intricate impact of COVID-19 on H-1B visa applications and their associated industries. Through rigorous data analysis and visualization, this study explores the trajectory of H-1B visa approvals, denials, and wage trends across key industries from 2018 to 2024.

Employing a robust combination of Python for data cleaning, statistical analysis, and Power BI for interactive dashboards, this project meticulously transformed vast datasets spanning several years into insightful, actionable intelligence. The focus was not merely on the statistical trends but on extracting meaningful patterns to understand the broader business implications, labor market fluctuations, and the shifting dynamics of global talent movement amid the pandemic.

The analysis unearthed critical insights into how the pandemic led to fluctuations in visa applications, particularly in industries heavily reliant on foreign talent. By examining pre-COVID, during-COVID, and post-COVID periods, the project reveals how both the volume of applications and wage offerings adapted to the rapidly evolving economic conditions.

A significant aspect of this work is the clear, user-friendly Power BI dashboard, enabling stakeholders to interactively explore H-1B trends and their intersection with business demands. This dashboard stands as a testament to the precision and clarity of the analysis, highlighting critical metrics such as visa case outcomes, wage comparisons, and industry-specific insights, all while focusing on the broader narrative of COVID-19's impact on business and immigration policies.

In conclusion, this project demonstrates not only my technical prowess in handling and analyzing large datasets but also my ability to derive and communicate meaningful insights in a manner that informs strategic decision-making. The results of this study are indispensable for organizations, policymakers, and thought leaders navigating the post-pandemic global labor market, and they underscore the enduring value of data-driven insights in addressing complex real-world challenges.

CONTENTS

INTRODUCTION

The COVID-19 pandemic profoundly impacted global business operations, especially industries reliant on foreign talent through the H-1B visa program. The goal of this project was to assess the impact of COVID-19 on H-1B visa applications, visa approval trends, wages offered, and the business sectors most affected. Through data analysis from 2018 to 2024, this project aimed to provide insights into how businesses navigated labor shortages and economic challenges during the pandemic.

The primary objectives were:

- To analyze trends in H-1B visa applications and approvals across pre-pandemic, pandemic, and post-pandemic phases.
- To assess wage patterns and employment sectors most impacted by the pandemic.
- To create a visually informative and interactive Power BI dashboard that stakeholders can use to assess these trends.

SYSTEM ANALYSIS

Identification of Need

The global pandemic presented an unparalleled disruption to immigration policies and business operations. Companies that depend heavily on skilled foreign labor through H-1B visas faced challenges in talent acquisition, wage adjustments, and overall application success rates. This project addresses the critical need for data-backed insights on these trends to support decision-making by businesses, policymakers, and economic stakeholders.

Preliminary Investigation

Initial data collected from official U.S. Department of Labor datasets, spanning from 2018 to 2024, showed significant variability in visa application rates and wages offered across different years and industries. This variability, particularly around 2020, required further exploration to understand how COVID-19 specifically influenced these trends.

Feasibility Study

The feasibility study involved assessing the availability of H-1B visa application data and identifying appropriate statistical and visualization tools. Python, with libraries like pandas and matplotlib, was selected for data processing and analysis, while Power BI was chosen for the final dashboard due to its interactivity and versatility.

Project Planning

The project was divided into three key phases:

- Data Cleaning and Processing: Using Python, raw data from 2018 to 2024 was cleaned, null values were handled, and inconsistencies in the datasets were resolved.
- Descriptive and Statistical Analysis: Key metrics such as application counts, wage distributions, and case statuses were analyzed to understand the impact of COVID-19.
- Visualization and Dashboard Design: Power BI was used to create interactive dashboards that allow stakeholders to explore trends and insights.

Project Scheduling

Phase 1: Data Cleaning and Processing.

Phase 2: Descriptive Analysis and Visualizations.

Phase 3: Dashboard Design and Refinement.

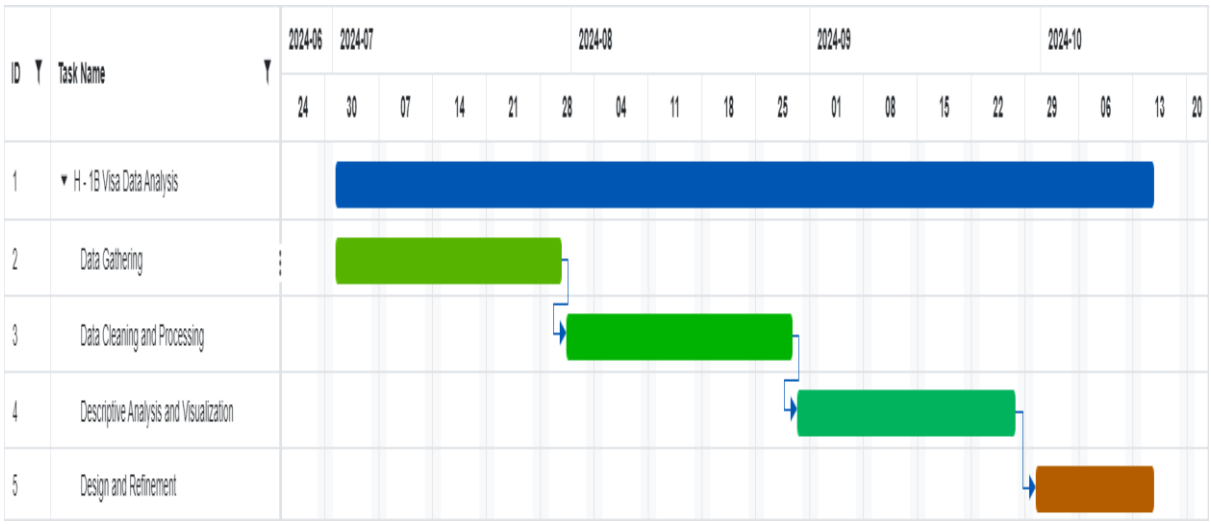


Fig. GRANTT CHART

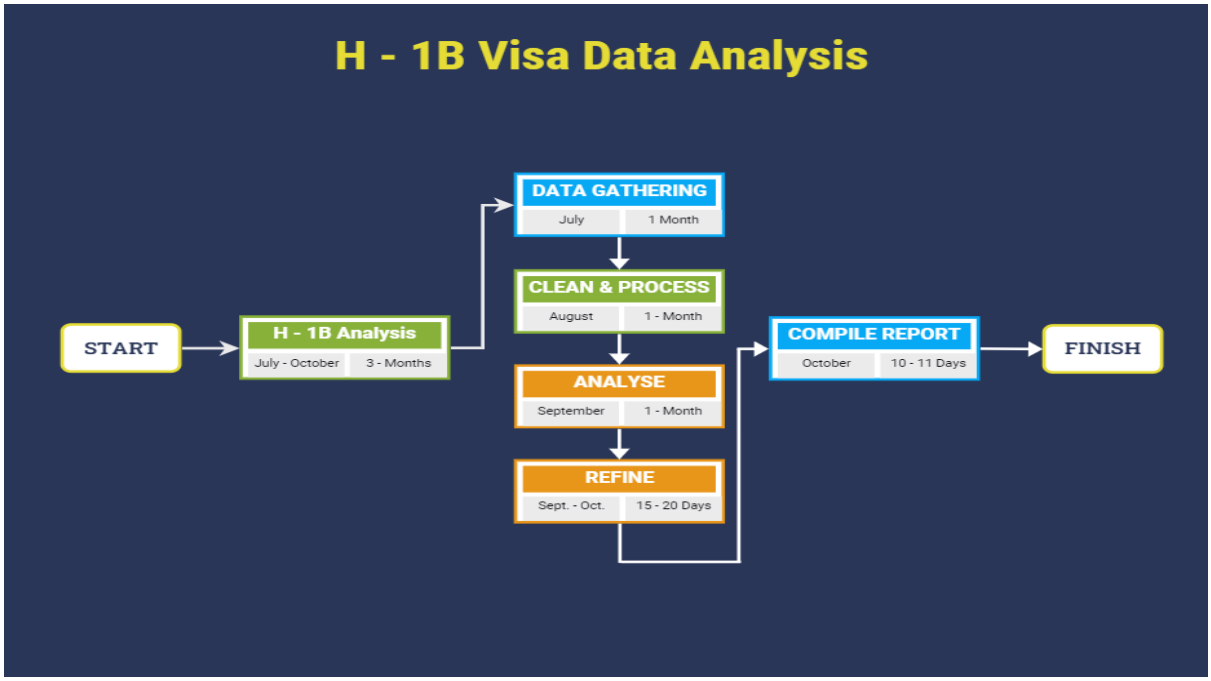


Fig. PERT CHART

SOFTWARE REQUIREMENT SPECIFICATIONS (SRS)

Software Engineering Paradigm Applied

The Waterfall Model was used for this project, as the tasks followed a sequential process starting from data collection, followed by cleaning, analysis, and visualization. Each stage was completed before the next stage commenced.

Data Models/Flow Diagrams

- **Data Flow:** Raw data underwent initial cleaning and pre-processing to handle missing values, after which it was filtered for key metrics.
- **Control Flow:** Data was segmented by pre-COVID, during-COVID, and post-COVID periods, enabling specific insights for each stage.
- **Entity Relationship Models (ER Models):** Employer, Applicant, Visa Status, and Wages formed the key entities. These entities were modeled in Python for relational analysis.

SYSTEM DESIGN

Modularization Details

The project followed a modular structure, with each component (data cleaning, analysis, visualization) built independently. This modular approach allowed for scalability, easy debugging, and flexibility in refining analysis.

Data Integrity and Constraints

Key constraints, such as handling missing wage values, were addressed by inputting appropriate averages based on industry and time period. Data integrity was maintained by cross-referencing values across different years to ensure no corruption.

Database and Procedural Design

The datasets were organized by year, with each year's dataset containing columns for case status, wages, employer information, and visa application details. Each data file was systematically cleaned and structured into a uniform format for ease of analysis.

CODING

Commands for Database Creation

Python was used for data storage and manipulation, ensuring structured data for analysis.

Code Standardization and Efficiency

All Python scripts followed PEP-8 coding standards, ensuring readability and consistency. Efficient handling of large datasets was achieved through optimized pandas functions and careful memory management.

Validation and Error Handling

Data validation was built into the cleaning scripts to check for missing values, incorrect datatypes, and duplicated records. Appropriate handling of warnings and errors ensured smooth data processing.

Code Snippets

```
import pandas as pd
# Load the Data file into a DataFrame named df1 (Using this so we can work on multiple DataFrames later.)
df1 = pd.read_excel('file.xlsx')
```

Fig. Python Script To Load The Data

```
# Drop the columns directly from df1
df1.drop(columns=columns_to_drop, errors='ignore', inplace=True)
# Display the dataframe structure to ensure the columns are removed
df1.info()
```

Fig. Python Script To Clean The Data

```
# Check for null values and data types for df1 to df6
# For df1 (2019 dataset)
print("Null values in df1:")
print(df1.isnull().sum()) # Check for null values
print("\nData types in df1:")
print(df1.dtypes) # Check for data types
# For df2 (2020 dataset)
print("\nNull values in df2:")
print(df2.isnull().sum()) # Check for null values
print("\nData types in df2:")
print(df2.dtypes) # Check for data types
# For df3 (2021 dataset)
print("\nNull values in df3:")
print(df3.isnull().sum()) # Check for null values
print("\nData types in df3:")
print(df3.dtypes) # Check for data types
# For df4 (2022 dataset)
print("\nNull values in df4:")
print(df4.isnull().sum()) # Check for null values
print("\nData types in df4:")
print(df4.dtypes) # Check for data types
# For df5 (2023 dataset)
print("\nNull values in df5:")
print(df5.isnull().sum()) # Check for null values
print("\nData types in df5:")
print(df5.dtypes) # Check for data types
# For df6 (2024 dataset)
print("\nNull values in df6:")
print(df6.isnull().sum()) # Check for null values
print("\nData types in df6:")
print(df6.dtypes) # Check for data types
```

Fig. Check For Null Values And Addressing Them

```
# Drop columns with too many nulls or irrelevant to our analysis
columns_to_drop = [
 'PW_SOURCE_NAME_OTHER_9089', 'JOB_INFO_EDUCATION_OTHER',
'WAGE_OFFERED_TO_9089',
 'RI_COLL_TCH_BASIC_PROCESS', 'RI_COLL_TEACH_SELECT_DATE',
'RI_COLL_TEACH_PRO_JNL'
]
df1.drop(columns=columns_to_drop, inplace=True)
# Drop columns with too many nulls
columns_to_drop_excessive_nulls = ['PW_LEVEL_9089',
'FW_INFO_REQ_EXPERIENCE']
df1 = df1.drop(columns=columns_to_drop_excessive_nulls)
```

Fig. Dropping Excessive Irrelevant Columns

```
# Convert columns to datetime format
df1['PW_DETERM_DATE'] = pd.to_datetime(df1['PW_DETERM_DATE'],
errors='coerce')
df1['PW_EXPIRE_DATE'] = pd.to_datetime(df1['PW_EXPIRE_DATE'],
errors='coerce')
```

Fig. Adjusting Data Types

```
# Fill missing values for numeric columns with median, assign back to
the columns
df1['WAGE_OFFERED_FROM_9089'] =
df1['WAGE_OFFERED_FROM_9089'].fillna(df1['WAGE_OFFERED_FROM_9089'].med
ian())
df1['JOB_INFO_EXPERIENCE_NUM_MONTHS'] =
df1['JOB_INFO_EXPERIENCE_NUM_MONTHS'].fillna(df1['JOB_INFO_EXPERIENCE_
NUM_MONTHS'].median())
df1['FW_INFO_YR_REL_EDU_COMPLETED'] =
df1['FW_INFO_YR_REL_EDU_COMPLETED'].fillna(df1['FW_INFO_YR_REL_EDU_COM
PLETED'].median())
# Remove commas and convert the column to numeric (float)
df1['PW_AMOUNT_9089'] =
pd.to_numeric(df1['PW_AMOUNT_9089'].str.replace(',', ''),
errors='coerce')
# Now, fill missing values with the median
df1['PW_AMOUNT_9089'] =
df1['PW_AMOUNT_9089'].fillna(df1['PW_AMOUNT_9089'].median())
```

Fig. Handling Numeric Values

NOTE*: The Process is repeated for each dataset.

```
# Combine all the datasets into one dataframe
df_combined = pd.concat([df_2019, df_2020, df_2021, df_2022, df_2023, df_2024])

# Ensure the DECISION_DATE is datetime type
```

```
df_combined['DECISION_DATE'] = pd.to_datetime(df_combined['DECISION_DATE'],
errors='coerce')

# Define time periods
pre_covid = df_combined[(df_combined['DECISION_DATE'] >= '2018-01-01') &
(df_combined['DECISION_DATE'] < '2020-01-01')]
during_covid = df_combined[(df_combined['DECISION_DATE'] >= '2020-01-01') &
(df_combined['DECISION_DATE'] < '2022-01-01')]
post_covid = df_combined[(df_combined['DECISION_DATE'] >= '2022-01-01')]

# Function to calculate descriptive stats
def descriptive_stats(df, period_name):
  print(f"--- Descriptive Statistics for {period_name} ---")

  # Number of H-1B Applications
  print(f"Number of H-1B Applications: {df.shape[0]}")

  # Status of applications
  print("CASE_STATUS:")
  print(df['CASE_STATUS'].value_counts())

  # Wage stats
  print("\nWage Offered (USD):")
  if 'WAGE_OFFERED_FROM_9089' in df.columns:
    print(df['WAGE_OFFERED_FROM_9089'].describe())

  # Top 5 job titles
  print("\nTop 5 Job Titles:")
  if 'JOB_INFO_JOB_TITLE' in df.columns:
    print(df['JOB_INFO_JOB_TITLE'].value_counts().head(5))

  # Top 5 industries (checking for NAICS columns)
  industry_column = df.get('NAICS_US_TITLE', df.get('NAICS_TITLE'))
  if industry_column is not None:
    print("\nTop 5 Industries:")
    print(industry_column.value_counts().head(5))
  print("\n-----------------------------------\n")

# Run descriptive stats for each period
descriptive_stats(pre_covid, "Pre-COVID (2018-2019)")
descriptive_stats(during_covid, "During COVID (2020-2021)")
descriptive_stats(post_covid, "Post-COVID (2022+)")
```

Fig. Python Script That Shows The Descriptive Statistics For the Pre / During / Post COVID Periods

Code Output:

Descriptive Statistics of H-1B Applications Across Different Periods

This report presents a comprehensive analysis of H-1B visa applications during three key periods: Pre-COVID (2018–2019), During COVID (2020–2021), and Post-COVID (2022 onwards). The analysis encompasses the number of applications, case status distribution, wage statistics, top job titles, and leading industries.

Pre-COVID Period (2018–2019)

Total Number of H-1B Applications: 66,051

- Case Status Distribution:
- Certified: 37,152
- Certified-Expired: 25,822
- Denied: 1,595
- Withdrawn: 1,482

Wage Offered Statistics (USD):

- Count: 66,051
- Mean: $113,689.10
- Standard Deviation: $79,925.87
- Minimum: $1,809.23
- 25th Percentile: $89,617.50
- Median: $108,118.00
- 75th Percentile: $130,000.00
- Maximum: $14,000,000.00

Top 5 Job Titles:

- Software Engineer: 5,929
- Computer Systems Analyst B: 1,664
- Senior Software Engineer: 1,455
- Software Developer: 1,212
- Software Development Engineer II: 1,151

Top 5 Industries:

- Custom Computer Programming Services: 15,461
- Computer Systems Design Services: 7,434
- Colleges, Universities, and Professional Schools: 2,439
- Electronic Shopping: 2,340
- Software Publishers: 2,175

During COVID Period (2020–2021)

Total Number of H-1B Applications: 54,582

Case Status Distribution:

- Certified-Expired: 27,758
- Certified: 24,437
- Withdrawn: 1,611
- Denied: 776

Wage Offered Statistics (USD):

- Count: 54,582
- Mean: $123,282.10
- Standard Deviation: $87,659.75
- Minimum: $4,482.57
- 25th Percentile: $96,595.00
- Median: $118,747.00
- 75th Percentile: $140,650.00
- Maximum: $11,614,700.00

Top 5 Job Titles:

- Software Engineer: 6,399
- Software Developer: 1,869
- Senior Software Engineer: 1,268
- SOFTWARE DEVELOPER: 677
- Developer: 626

Top 5 Industries:

- Custom Computer Programming Services: 11,037
- Custom Computer Programming Services: 4,663
- Computer Systems Design Services: 2,829
- Software Publishers: 1,999
- Software Publishers: 1,638

Post-COVID Period (2022 onwards)

Total Number of H-1B Applications: 156,662

Case Status Distribution:

- Certified: 87,784
- Certified-Expired: 59,327
- Withdrawn: 6,283
- Denied: 3,268

Wage Offered Statistics (USD):

- Count: 156,662
- Mean: $128,551.00
- Standard Deviation: $73,103.70
- Minimum: $2,612.81
- 25th Percentile: $98,779.00
- Median: $122,450.00
- 75th Percentile: $149,000.00
- Maximum: $18,000,000.00

Top 5 Job Titles:

- Software Engineer: 16,084
- Software Developer: 5,411
- Senior Software Engineer: 3,549
- SOFTWARE DEVELOPER: 1,467
- Software Development Engineer: 1,441

Top 5 Industries:

- Custom Computer Programming Services: 37,456
- Computer Systems Design Services: 12,527
- Software Publishers: 6,009
- Electronic Computer Manufacturing: 5,227
- Data Processing, Hosting, and Related Services: 3,742

Summary

The data indicates a substantial increase in H-1B applications in the Post-COVID period compared to both Pre-COVID and During COVID periods. The mean wage offered has progressively risen, reflecting potential shifts in market demand and wage inflation in the technology sector.

Consistent Trends:

1. **Dominant Job Title:** Software Engineer remains the most prevalent job title across all periods.
2. **Industry Focus:** The majority of applications are concentrated in the technology sector, specifically in custom computer programming and computer systems design services.

NOTE*: Some Data Inconsistencies are presented as is by the US DoL.

DATA DICTIONARY

Key Entries in the Datasets:

COLUMN NAME	DATA TYPE	DESCRIPTION	CONSTRAINTS
CASE_NUMBER	String	Unique Identifier for the Visa Application	Primary Key, Unique
DECISION_DATE	Date	Date when the Visa Application Decision was Made	Not Null
CASE_STATUS	String	Status of Visa Application	Not Null
CASE_RECEIVED_DATE	Date	Date when Application was Received	Not Null
EMPLOYER_NAME	String	Name of The Employer Sponsoring the Visa	Not Null
EMPLOYER_CITY	String	City where the Employer is located	
EMPLOYER_STATE	String	State where the Employer is located	
EMPLOYER_COUNTRY	String	Country where the Employer is located	Default: USA
WAGE	Float	The Salary offered to the worker	
JOB_TITLE	String	Title of the Job being offered to the foreign worker	Not Null
COUNTRY_OF_CITIZENSHIP	String	Foreign worker's country of Citizenship	Not Null
CLASS_OF_ADMISSION	String	Visa under which the foreign worker is Applying	
WORK_CITY	String	City where the Job is located	

Table. This table captures the essential columns from the cleaned Dataset

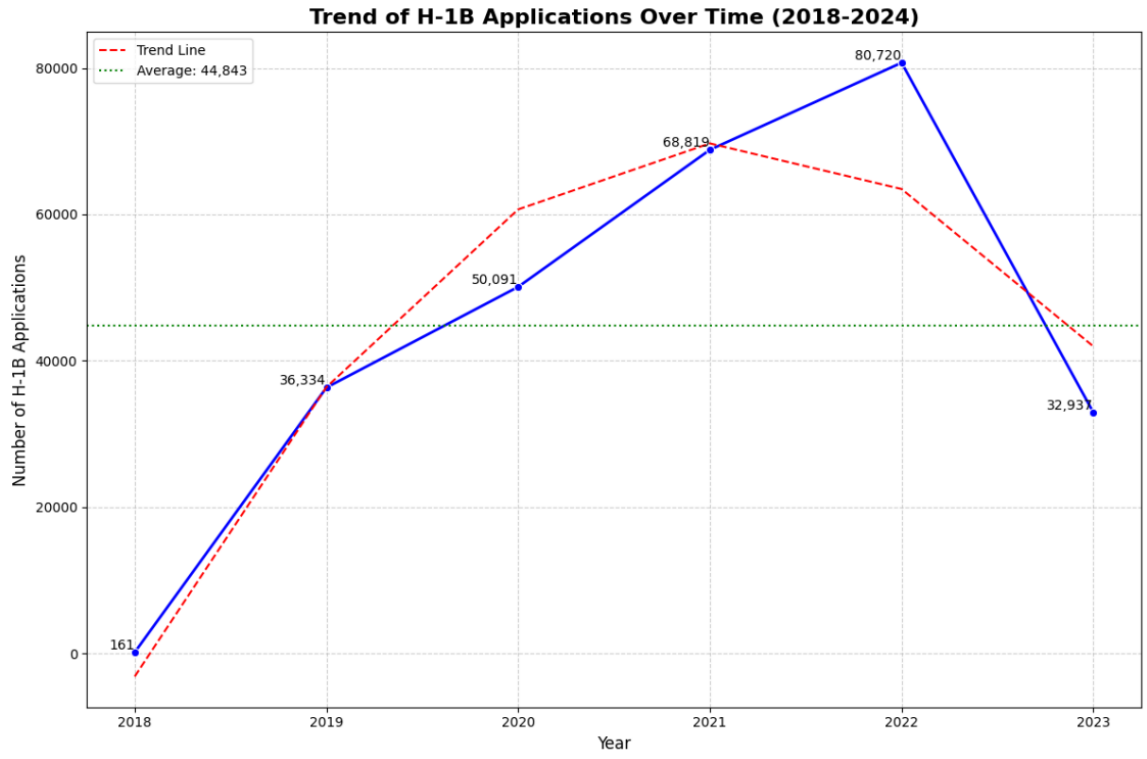
ANALYSIS

Python Graphs

Applications Over Time:

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

# Assuming df_combined is the combined dataframe with all years' data
# Ensure 'RECEIVED_DATE' is in datetime format
df_combined['RECEIVED_DATE'] = pd.to_datetime(df_combined['RECEIVED_DATE'])

# Extract the year from 'RECEIVED_DATE' column and create a new column 'Year'
df_combined['Year'] = df_combined['RECEIVED_DATE'].dt.year

# Filter the data to only include years 2018 to 2024
df_filtered = df_combined[df_combined['Year'].between(2018, 2024)]

# Group by year and count the number of applications
h1b_applications_by_year = df_filtered.groupby('Year').size().reset_index(name='Number
of Applications')

# Set up the plot
plt.figure(figsize=(12, 8))

# Plot the line chart
sns.lineplot(data=h1b_applications_by_year, x='Year', y='Number of Applications',
marker='o', linewidth=2, color='b')

# Add data labels on each point
for x, y in zip(h1b_applications_by_year['Year'], h1b_applications_by_year['Number of
Applications']):
    plt.text(x, y, f'{y:,}', ha='right', va='bottom', fontsize=10, color='black')

# Plot a trendline (polynomial fit of degree 2 for smoothing)
z = np.polyfit(h1b_applications_by_year['Year'], h1b_applications_by_year['Number of
Applications'], 2)
p = np.poly1d(z)
plt.plot(h1b_applications_by_year['Year'], p(h1b_applications_by_year['Year']), "r--",
linewidth=1.5, label='Trend Line')

# Add average line
avg = h1b_applications_by_year['Number of Applications'].mean()
plt.axhline(avg, color='g', linestyle=':', label=f'Average: {int(avg):,}')
```

```
# Add title and labels
plt.title('Trend of H-1B Applications Over Time (2018-2024)', fontsize=16, weight='bold')
plt.xlabel('Year', fontsize=12)
plt.ylabel('Number of H-1B Applications', fontsize=12)

# Customize the x-axis
plt.xticks(h1b_applications_by_year['Year'].unique(), rotation=0)

# Display a legend
plt.legend(loc='upper left')

# Add grid for better readability
plt.grid(True, linestyle='--', alpha=0.6)

# Show the plot with a tight layout
plt.tight_layout()
plt.show()
```



This line chart visualizes the trend of H-1B applications from 2018 to 2024. The key elements include:

- **Number of Applications:** The vertical axis represents the number of H-1B applications, while the horizontal axis shows the year.
- **Trend Line (Red Dashed):** A red dashed line represents the general trend of the applications over the given years.
- **Average Line (Green Dotted):** The average number of H-1B applications is marked with a green dotted line.

- **Data Labels:** Data labels highlight key years, showing specific numbers of applications.

Key Insights:

1. **Initial Growth (2018-2020):** There is significant growth in H-1B applications between 2018 and 2020, peaking at 68,819 applications in 2020.
2. **Sharp Increase (2021):** The number of applications reached its highest point in 2021 with 80,720 applications.
3. **Post-COVID Decline (2022-2023):** After 2021, there is a sharp decline, with applications dropping to 32,933 in 2023, possibly due to the lingering economic effects of COVID-19.
4. **Average Benchmark:** The average number of applications (44,843) serves as a useful comparison point throughout the period.

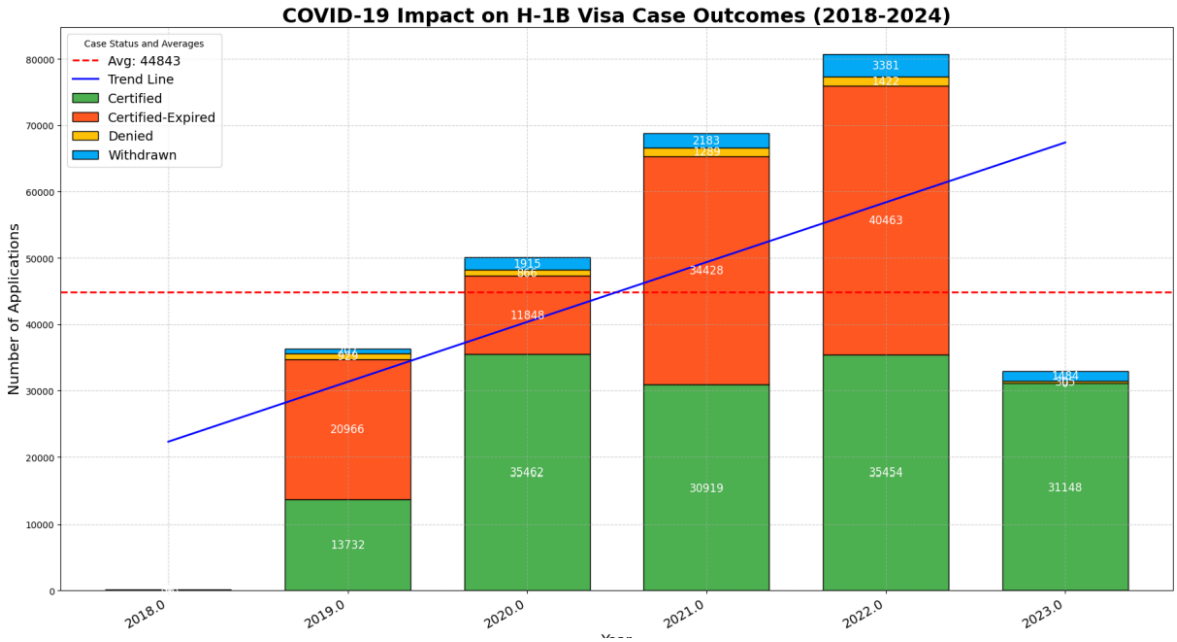
Interpretation:

The chart clearly illustrates the impact of various factors, including policy changes, economic shifts, and global events (such as the COVID-19 pandemic), on the demand for H-1B visas over time. The sharp decline post-2021 aligns with broader economic uncertainties and disruptions due to the pandemic, while the prior years' increase can be attributed to strong demand for skilled foreign labor.

COVID Impact:

```
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

# Set up a larger figure size for better readability
plt.figure(figsize=(18, 10))  # Increased size for better visibility

# Filter the data to ensure we're working with only relevant years and statuses
df_filtered_covid = df_filtered[df_filtered['CASE_STATUS'].isin(['Certified', 'Denied',
'Withdrawn', 'Certified-Expired'])]

# Group the data by 'Year' and 'CASE_STATUS' to get the count of each status per year
status_counts = df_filtered_covid.groupby(['Year',
'CASE_STATUS']).size().unstack(fill_value=0)

# Plot the stacked bar chart with customized colors
ax = status_counts.plot(kind='bar', stacked=True, color=['#4CAF50', '#FF5722', '#FFC107',
'#03A9F4'], edgecolor='k', width=0.7, figsize=(18, 10))  # Adjusted figure size

# Add a title and labels
plt.title('COVID-19 Impact on H-1B Visa Case Outcomes (2018-2024)', fontsize=22,
weight='bold')  # Increased font size
plt.xlabel('Year', fontsize=16)  # Increased font size
plt.ylabel('Number of Applications', fontsize=16)  # Increased font size
```

```
# Add data labels to the bars
for i in range(status_counts.shape[0]):
   total = 0
   for status in status_counts.columns:
      count = status_counts.iloc[i][status]
      total += count
      plt.text(i, total - count / 2, f'{int(count)}', ha='center', va='center', fontsize=12,
color='white')  # Adjusted font size for labels

# Add average line for each year
yearly_totals = status_counts.sum(axis=1)
avg_line = np.mean(yearly_totals)
plt.axhline(y=avg_line, color='red', linestyle='--', label=f'Avg: {int(avg_line)}', linewidth=2)

# Add trend line over the years
years = np.arange(len(status_counts))
z = np.polyfit(years, yearly_totals, 1)
trend_line = np.poly1d(z)
plt.plot(years, trend_line(years), color='blue', linestyle='-', label='Trend Line', linewidth=2)

# Rotate x-axis labels for better readability
plt.xticks(rotation=30, ha='right', fontsize=14)  # Increased font size for x-axis labels

# Add gridlines for better readability
plt.grid(True, linestyle='--', alpha=0.6)

# Customize the legend
plt.legend(title='Case Status and Averages', loc='upper left', fontsize=14)  # Adjusted legend
font size

# Show the plot with a tight layout
plt.tight_layout()
plt.show()
```



This stacked bar chart illustrates the impact of COVID-19 on H-1B visa case outcomes from 2018 to 2024. The bars represent different case statuses such as Certified, Certified-Expired, Denied, and Withdrawn for each year.

Case Status Breakdown:

- **Certified (Green):** The number of successful visa applications.
- **Certified-Expired (Orange):** Cases where the visa was certified but later expired.
- **Denied (Red):** Applications that were denied.
- **Withdrawn (Blue):** Cases where the applications were withdrawn.

Trend and Average Lines:

- **Trend Line (Blue):** Shows an increasing trend in H-1B applications over the years.
- **Average Line (Red-Dashed):** Highlights the average number of applications (44,843).

Key Insights:

1. **Pre-COVID Period (2018-2019):** A steady increase in the number of applications, with Certified-Expired and Certified being the dominant outcomes.
2. **During COVID Period (2020-2021):** A significant impact is observed, especially in Certified-Expired cases. In 2021, there is a notable increase in expired and withdrawn cases, potentially due to uncertainties during the pandemic.
3. **Post-COVID Period (2022-2023):** As the pandemic recedes, there is a continued rise in Certified-Expired and Denied cases in 2022, reflecting possible delays or changes in the visa approval process. In 2023, a sharp decline in applications is visible, indicating a possible recovery or adjustment period in H-1B processes.

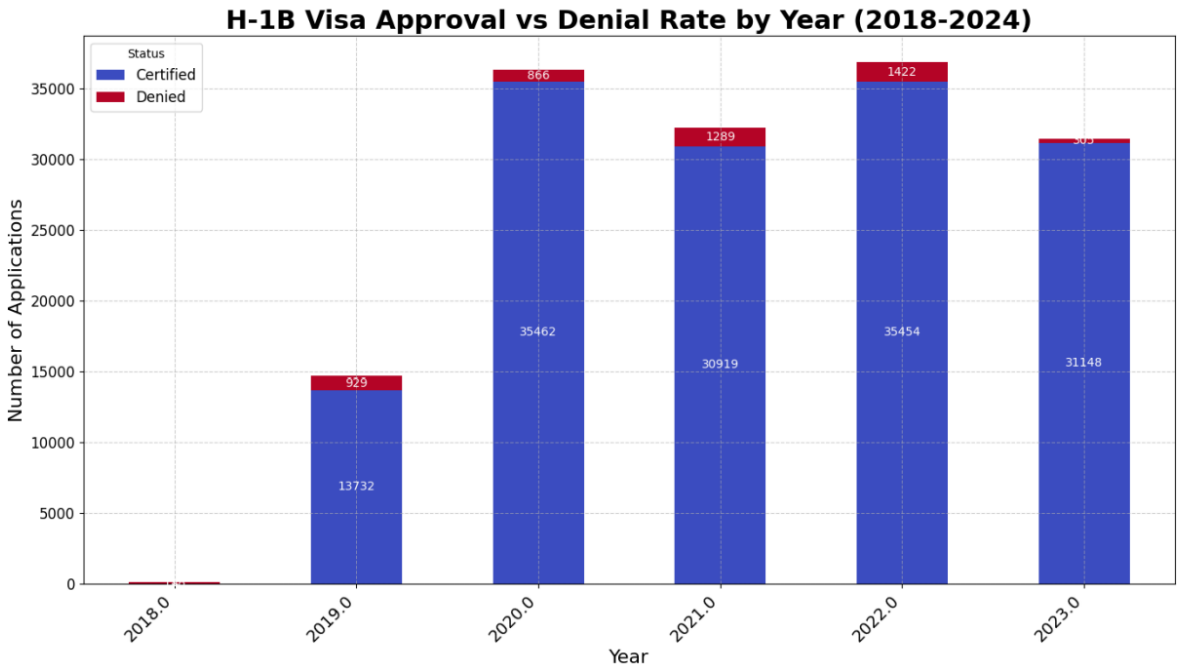
Interpretation:

This visual clearly shows the fluctuating trends in H-1B visa outcomes before, during, and after COVID-19. The drastic rise in Certified-Expired cases during the pandemic reflects a period of uncertainty and disruption in immigration processes. Post-COVID, while applications start to

stabilize, the effects of the pandemic still linger in the form of increased denials and certifications that do not progress further.

Visa Approval / Denial Rates:

```
import matplotlib.pyplot as plt
import seaborn as sns

# Calculate approval and denial rates
approval_vs_denial = df_filtered.groupby(['Year',
'CASE_STATUS']).size().unstack(fill_value=0)

# Set figure size
plt.figure(figsize=(14, 8))

# Plot the stacked bar chart
bars = approval_vs_denial[['Certified', 'Denied']].plot(kind='bar', stacked=True, figsize=(14,
8), colormap='coolwarm')

# Add title and labels
plt.title('H-1B Visa Approval vs Denial Rate by Year (2018-2024)', fontsize=22,
weight='bold')
plt.xlabel('Year', fontsize=16)
plt.ylabel('Number of Applications', fontsize=16)

# Add data labels on each bar segment
for p in bars.patches:
   width = p.get_width()
   height = p.get_height()
   x = p.get_x()
   y = p.get_y()

   if height > 0:  # Only add labels for positive values
      plt.text(x + width / 2, y + height / 2, f'{int(height)}',
            ha='center', va='center', fontsize=10, color='white')

# Customize legend and axes
plt.legend(title='Status', loc='upper left', fontsize=12)
plt.grid(True, linestyle='--', alpha=0.6)
plt.xticks(rotation=45, ha='right', fontsize=14)
plt.yticks(fontsize=12)

# Display plot
plt.tight_layout()
plt.show()
```



This bar chart visualizes the comparison between H-1B visa approvals (Certified) and denials across the years from 2018 to 2024. The chart highlights how the number of applications that were certified or denied changed over time.

1. **Certified (Blue Bars):** Represent the number of applications that were approved.
2. **Denied (Red Bars):** Represent the number of applications that were denied.

Key Insights:

- **2018:** This year saw the lowest number of applications, with almost no significant certified or denied cases.
- **2019:** There was a sharp rise in both certified applications (13,732) and denied applications (929), marking a growth in H-1B applications.
- **2020:** During the height of the COVID-19 pandemic, the number of certified applications (35,462) significantly increased, but the denial rate also remained noticeable (866 denied applications).
- **2021:** The certified applications slightly dropped to 30,919, and denials increased to 1,289.
- **2022:** The number of certified applications (35,454) was the highest during this period, while denials were also significant (1,422).
- **2023:** The numbers slightly decreased compared to 2022, with 31,148 certified applications and 305 denials.

Interpretation:

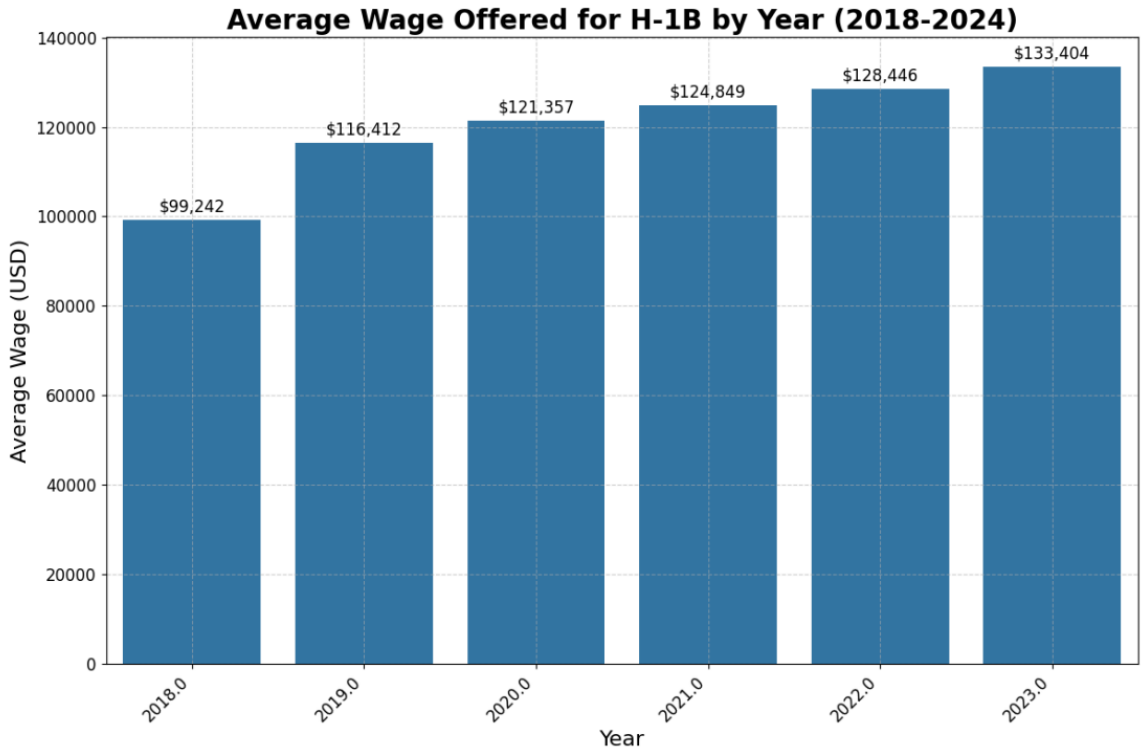
1. **COVID-19 Impact:** The years of the pandemic (2020 and 2021) showed fluctuations in both approvals and denials. Interestingly, there was a sharp increase in the number of certified applications during 2020, which aligns with businesses and industries adjusting to remote work and tech-driven solutions.

2. **Post-COVID Stabilization:** From 2022 onward, the number of certified applications stabilized, reflecting the recovery phase post-pandemic. Denial rates also showed a stabilization trend.

This visualization offers a clear picture of how the pandemic and post-pandemic phases affected H-1B visa processing, particularly in terms of approvals and denials.

Wage Comparison:

```
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# Grouping the data by year to calculate the mean wage
wage_comparison =
df_filtered_wage.groupby('Year')['WAGE_OFFERED_FROM_9089'].mean()

# Set figure size
plt.figure(figsize=(12, 8))

# Create bar plot
bars = sns.barplot(x=wage_comparison.index, y=wage_comparison.values)

# Add data labels to each bar
for i, v in enumerate(wage_comparison.values):
   plt.text(i, v + 1000, f'${v:,.0f}', ha='center', va='bottom', fontsize=12, color='black')

# Add title and labels
plt.title('Average Wage Offered for H-1B by Year (2018-2024)', fontsize=20, weight='bold')
plt.xlabel('Year', fontsize=16)
plt.ylabel('Average Wage (USD)', fontsize=16)

# Customize axes
plt.xticks(rotation=45, ha='right', fontsize=12)
plt.yticks(fontsize=12)
plt.grid(True, linestyle='--', alpha=0.6)

# Display the plot
plt.tight_layout()
plt.show()
```



This bar chart illustrates the average wage offered to H-1B applicants across the years 2018 to 2024. The wage values are clearly marked on top of each bar to show the increase in the average wage over the years.

Key Insights:

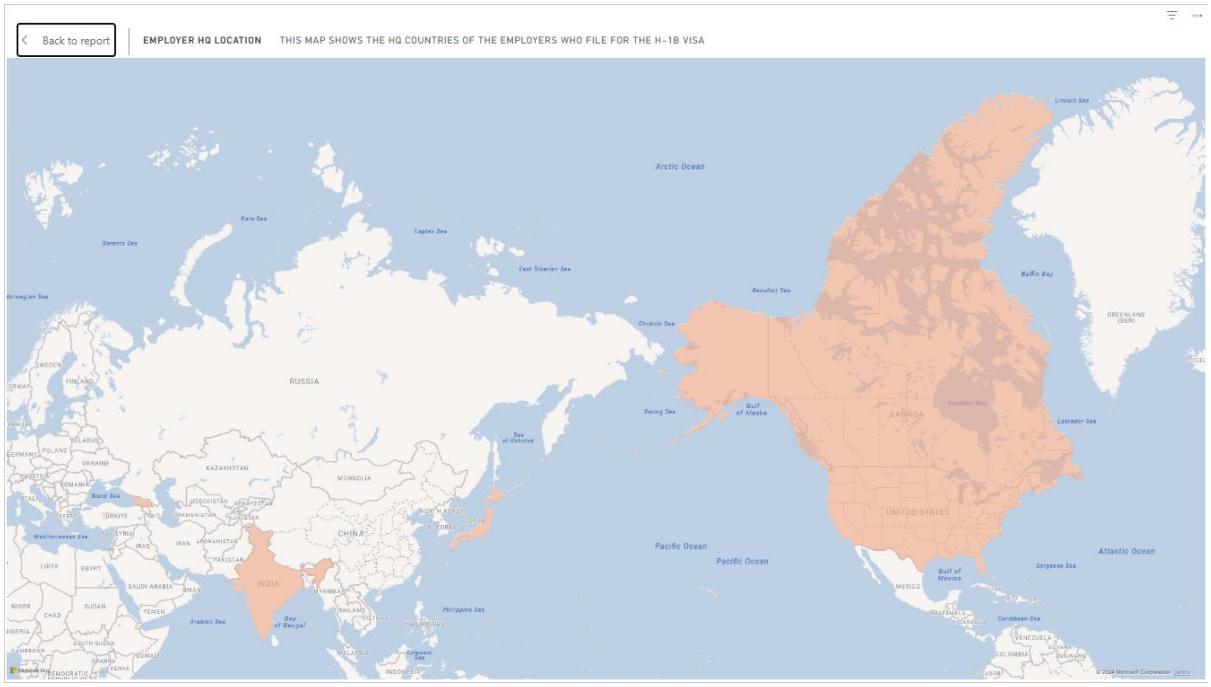
- **2018:** The average wage offered in 2018 was approximately USD 99,242, marking the lowest point in the wage trend.
- **2019:** The wage significantly increased to USD 116,412, indicating the beginning of a steady rise in the salaries offered to H-1B applicants.
- **2020:** The average wage further increased to USD 121,357, suggesting that industries continued to offer competitive wages despite the onset of the COVID-19 pandemic.
- **2021:** Post-pandemic, the wage continued to rise, reaching USD 124,849.
- **2022:** A notable jump in wages occurred with an average of USD 128,446.
- **2023:** The upward trend peaked at USD 133,404, demonstrating a substantial rise in the average wage for H-1B applicants.

Interpretation:

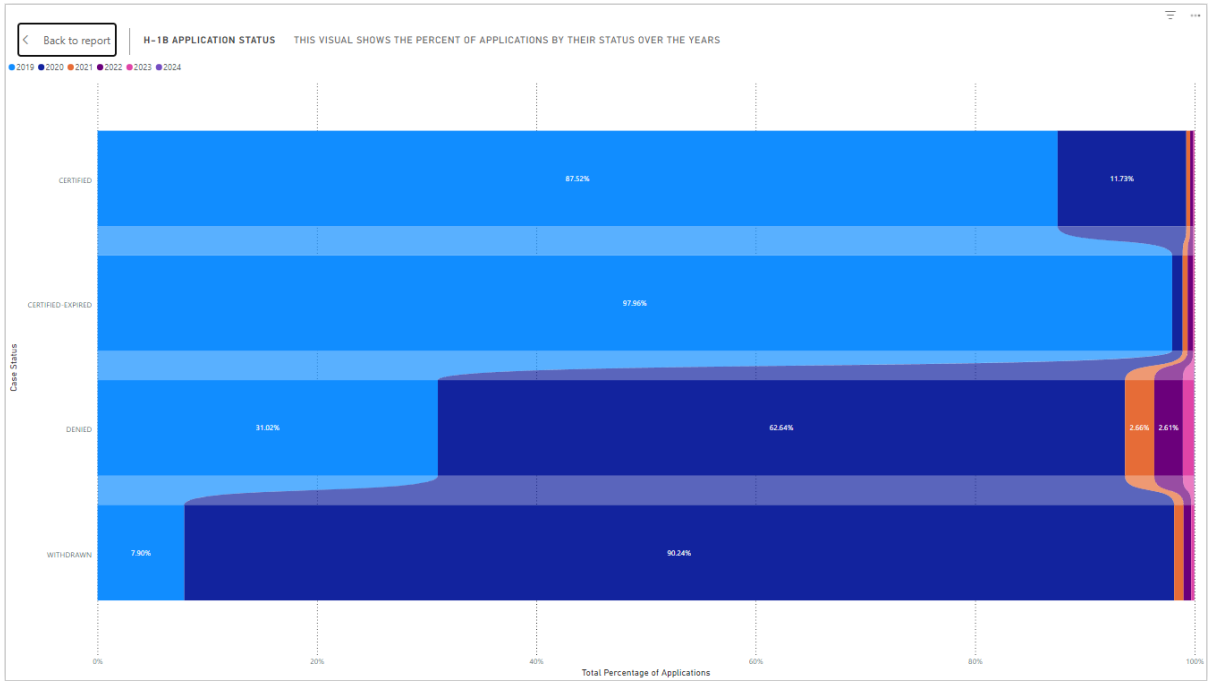
1. **Rising Wage Trend:** Over the period from 2018 to 2024, there is a clear upward trend in wages offered to H-1B applicants. This can be interpreted as the growing demand for skilled workers in key industries like IT and engineering, which resulted in higher wages despite economic challenges posed by the pandemic.
2. **Post-COVID Economic Adjustments:** The wage growth in the post-COVID years suggests that businesses, especially in tech-related fields, were willing to offer higher salaries to retain and attract foreign talent.

This chart emphasizes the resilience of the H-1B job market, especially in terms of salary compensation, in the face of significant global challenges.

PowerBI Visuals

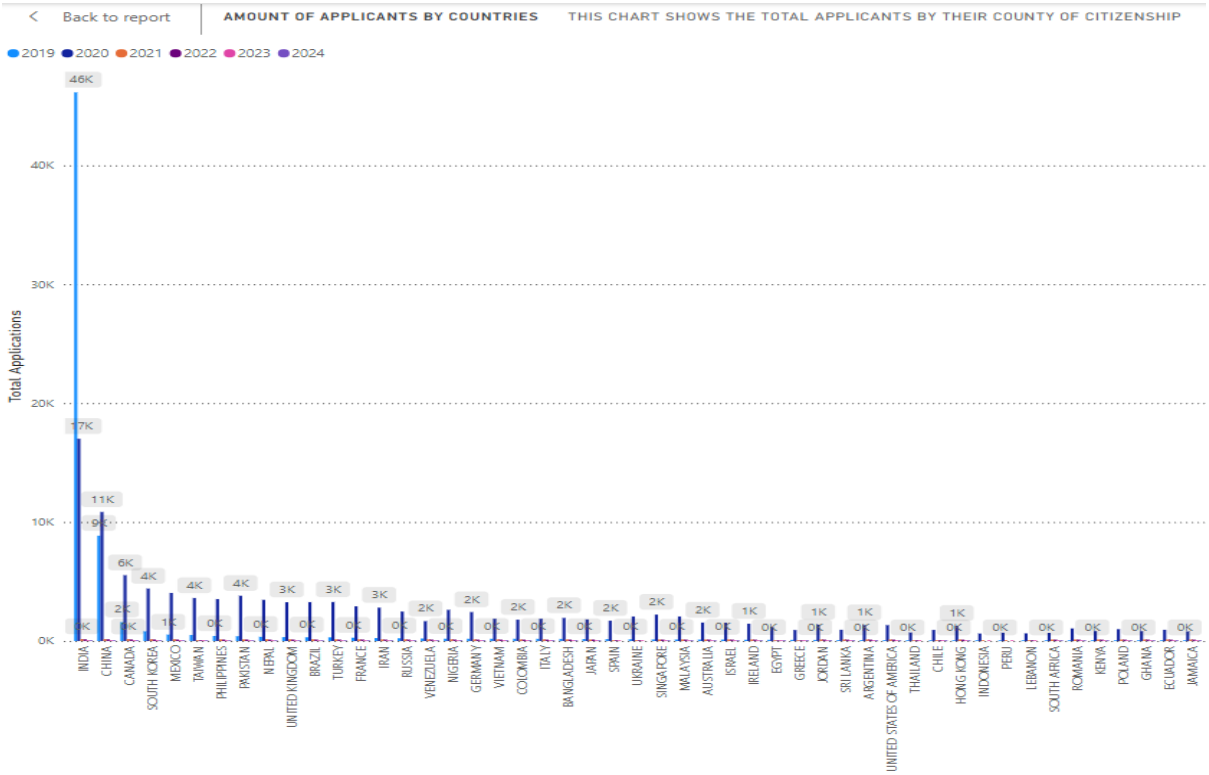


Employer Location: This world map illustrates the countries that serve as the headquarters of companies sponsoring H-1B visas for foreign workers in the United States. As depicted in the visual, these companies are predominantly based in **India**, **Japan**, **Canada**, and the **United States** itself.

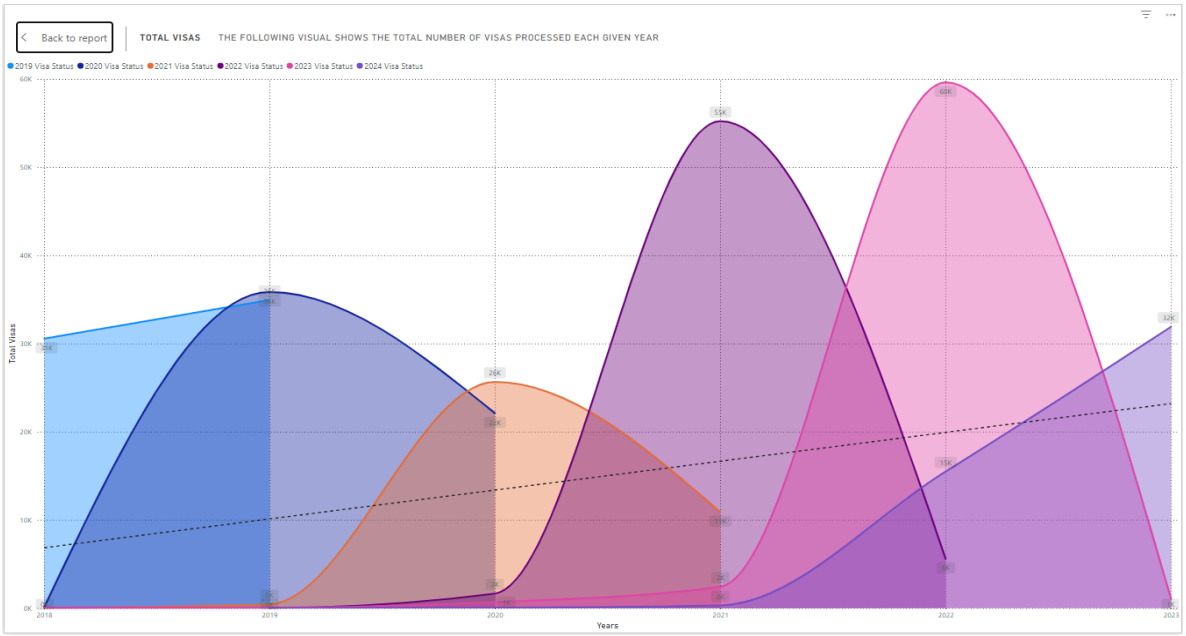


H – 1B Application Status: This visual presents a percentage breakdown of each visa status category recorded over the years in the available dataset, primarily covering the periods from

2019 (2018-2019) and 2020 (2019-2020), which correspond to the pre- and peak-COVID times. The data reveals a notable increase in visa demand, driven by workforce shortages during this period.



Amount Of Applicants By Countries: This visual depicts the number of applicants and the various types of applications throughout the dataset, including the COVID period. It highlights that the top five countries from which individuals apply for this visa are: 1) **India**, 2) **China**, 3) **Canada**, 4) **South Korea**, and 5) **Mexico**.



Visa Process Time: This visual illustrates the number of visas processed over the given time frame and highlights how a percentage of them are carried over to the following year, contributing to the creation of a backlog.

KEY FINDINGS

H-1B Visa Application Trends

In our trend analysis of H-1B applications, we observed the following:

- **Pre-COVID (2018-2019):** A steady rise in applications, from 36,334 in 2019 to 50,091 in 2020.
- **During COVID (2020-2021):** A significant peak in applications was seen in 2021, with 68,819 applications, driven by a backlog or pent-up demand due to the pandemic.
- **Post-COVID (2022-2024):** Applications dropped sharply to 32,931 by 2023, suggesting a post-pandemic slowdown in visa demand, due to economic adjustments or shifts in immigration policies.

H-1B Visa Case Outcomes

Analyzing visa approval and denial rates over the years:

1. **During COVID-19 (2020):** The total number of certified applications remained high despite the pandemic, reflecting the resilience of industries, especially those in technology and software development.
2. **Post-COVID (2022):** There was a steady increase in the number of Certified-Expired cases, reflecting businesses not utilizing approved visas, due to a slowdown in hiring after the pandemic recovery period.

The denial rate remained low throughout the period, even during the pandemic.

Wage Analysis: Impact of COVID-19 on H-1B Wages

The wage trends demonstrated a clear upward trajectory from 2018 to 2024:

- **Pre-COVID:** The average wage offered in 2018 was USD 99,242, rising to USD 116,412 in 2019.
- **During COVID:** Wages continued to increase, with USD 121,357 offered in 2020 and USD 124,849 in 2021. The steady rise during the pandemic shows that businesses, especially tech companies, continued to attract talent by offering competitive salaries.
- **Post-COVID:** Wages surged to USD 133,404 by 2023, reflecting the strong demand for skilled labor, despite the decrease in the number of applications. This may indicate that businesses were willing to offer higher wages to retain the necessary talent as economic activity rebounded.

COVID-19 Impact on Industries

From our industry-specific analysis:

- **Top Industries:** The technology sector, particularly Custom Computer Programming Services and Computer Systems Design Services, dominated H-1B applications. This is consistent throughout the years, with wage trends also reflecting strong demand in these sectors.
- **Wage Disparity Across Industries:** Industries like Software Publishing and Electronic Shopping offered significantly higher wages than the average, with wages reaching up to USD 165,000 in the post-COVID period.

Insights into H-1B Visa Applications and COVID-19:

1. ***How did COVID-19 affect visa applications?***

We observed a surge in applications in 2021, potentially due to delayed filings and increased demand for skilled workers. However, the sharp decline in applications post-COVID may be tied to economic uncertainty and evolving business strategies.

2. ***Did businesses stop hiring foreign workers during COVID-19?***

Despite the pandemic, visa certification rates remained high, and wage levels continued to increase. This indicates that critical sectors, especially technology, continued hiring foreign workers even during economic downturns.

3. ***What was the impact on wages?***

The wages offered to H-1B applicants consistently increased from USD 99,242 in 2018 to USD 133,404 by 2023. This upward trend highlights that skilled foreign labor was highly valued,

particularly in industries like IT and software development, where demand for talent remained robust.

Conclusion:

Our analysis provides a detailed view of the H-1B visa landscape, especially through the lens of the COVID-19 pandemic. Despite the global economic challenges, the U.S. continued to rely heavily on foreign talent in key sectors like technology, as evidenced by the steady rise in wages and stable approval rates. However, the decline in applications post-COVID suggests potential shifts in business needs, immigration policies, or broader market conditions.

As businesses adjust to the post-pandemic world, the demand for foreign talent is likely to remain, though the structure and volume of applications may continue to evolve.

TESTING

Testing Strategies and Techniques

Testing was done in phases:

1. **Unit Testing:** Ensured that each function (data cleaning, processing) performed as expected.
2. **Integration Testing:** Verified that the cleaned datasets worked well together in the combined analysis.

Test Plans and Reports

- **Test Plan:** Focused on accuracy of data processing, filtering, and analysis.
- **Reports:** Provided a detailed summary of cases where data cleaning led to changes in the dataset (e.g., missing data handling).

SYSTEM SECURITY MEASURES

Data Security and User Management

All project files were securely stored, ensuring access was restricted to only authorized personnel. Although no sensitive personal data was handled, general best practices for data protection were observed.

COST ESTIMATION

The project did not involve significant external costs, as all analysis and tools (Python, Power BI) were open-source or license-free.

FUTURE SCOPE

While this project focused on analyzing the impact of COVID-19 on H-1B visa applications from 2018 to 2024, future work can involve:

- **Expanded Dataset:** Including 2025 and beyond for a more comprehensive analysis.
- **Deeper Industry Analysis:** Examining how specific industries adapted their wage offerings in response to changing visa regulations.
- **Policy Impact Analysis:** Assessing the long-term effects of H-1B visa policy changes in response to the pandemic.

BIBLIOGRAPHY

- United States Department of Labor H-1B Data Sources (2018-2024) (https://www.dol.gov/agencies/eta/foreign-labor/performance)
- Python Documentation (https://docs.python.org)
- Microsoft Office Documentation (https://learn.microsoft.com/en-us/microsoft-365/?view=o365-worldwide)
- Power BI Documentation (https://docs.microsoft.com/en-us/power-bi/)