

## Exploratory Data Analysis of Visa Applications: Trends, Insights, and Key Findings

### Importing Libraries for Analysis

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

### Loading the Datasets

```
file_path = "Visadataset.csv" # Update with correct path
df = pd.read_csv(file_path)
```

### Seaborn theme

```
sns.set_theme(style="whitegrid")
```

### Statistical Summary of dataset

```
print("\nSummary Statistics (Numerical):")
print(df.describe())
print("\nSummary Statistics (Categorical):")
print(df.describe(include="object"))
```

#### Summary Statistics (Numerical):

	no_of_employees	yr_of_estab	prevailing_wage
count	25480.000000	25480.000000	25480.000000
mean	5667.043210	1979.409929	74455.814592
std	22877.928848	42.366929	52815.942327
min	-26.000000	1800.000000	2.136700
25%	1022.000000	1976.000000	34015.480000
50%	2109.000000	1997.000000	70308.210000
75%	3504.000000	2005.000000	107735.512500
max	602069.000000	2016.000000	319210.270000

#### Summary Statistics (Categorical):

	case_id	continent	education_of_employee	has_job_experience \
count	25480	25480	25480	25480
unique	25480	6	4	2
top	EZYV25480	Asia	Bachelor's	Y
freq	1	16861	10234	14802

	requires_job_training	region_of_employment	unit_of_wage \
--	-----------------------	----------------------	----------------

count	25480	25480	25480
unique	2	5	4
top	N	Northeast	Year
freq	22525	7195	22962

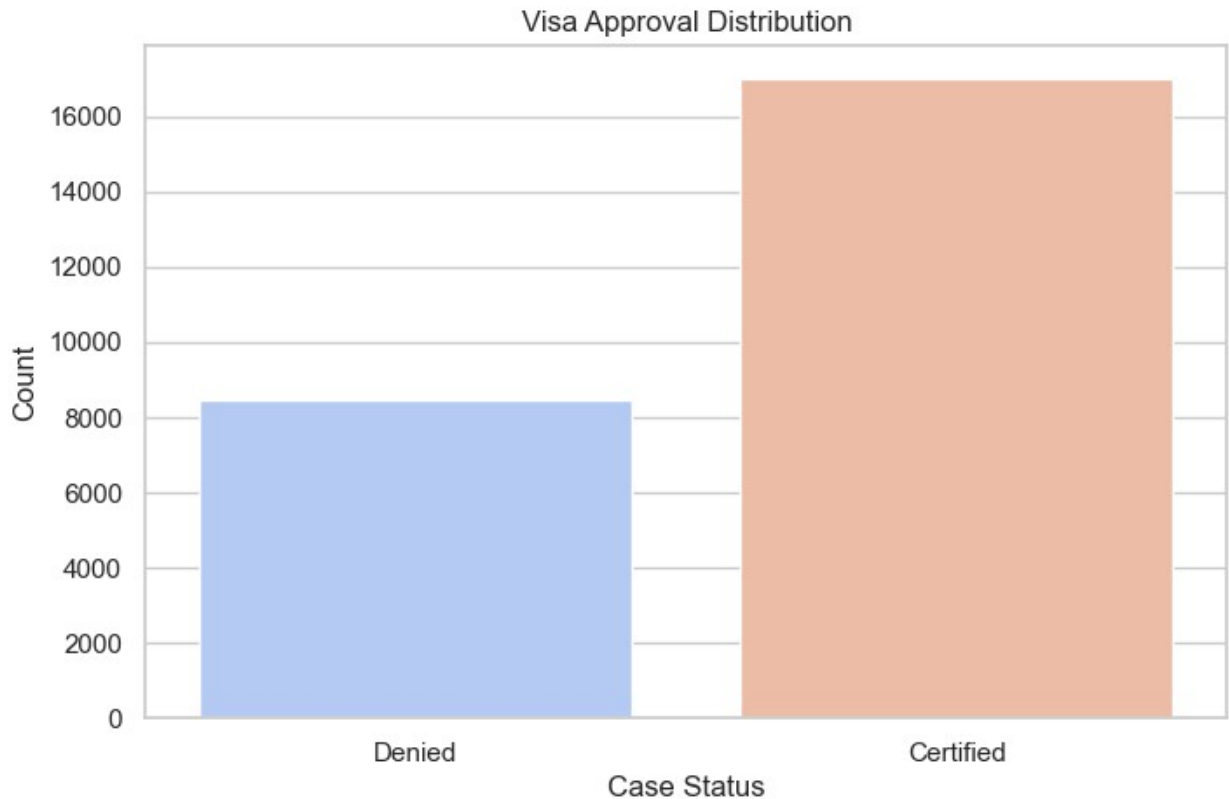
	full_time_position	case_status
count	25480	25480
unique	2	2
top	Y	Certified
freq	22773	17018

### Key Insights in Statements

- The number of employees in companies varies significantly, with an average of 5,667 and a median of 2,109. However, there is an anomaly with a negative value (-26), which is logically incorrect. Additionally, some companies have an exceptionally high employee count, with a maximum of 602,069, indicating large enterprises.
- The year of establishment for companies ranges from 1800 to 2016, with an average founding year of 1979 and a median of 1997. The year 1800 is likely an outlier or incorrect data entry, as most companies were established in the late 1900s or early 2000s.
- The prevailing wage shows a wide variation, with an average of 74,455 dollar and a median of 70,308 dollar. The minimum wage recorded is just 2.14 dollar, which is highly unrealistic and possibly an error. The majority of wages fall between 34,015 dollar and 107,735 dollar, suggesting a fair range for most employees.

### 1) Visualization: Visa Approval Distribution (Certified vs Denied)

```
plt.figure(figsize=(8,5))
sns.countplot(x="case_status", data=df, hue="case_status",
palette="coolwarm", legend=False)
plt.title("Visa Approval Distribution")
plt.xlabel("Case Status")
plt.ylabel("Count")
plt.show()
```



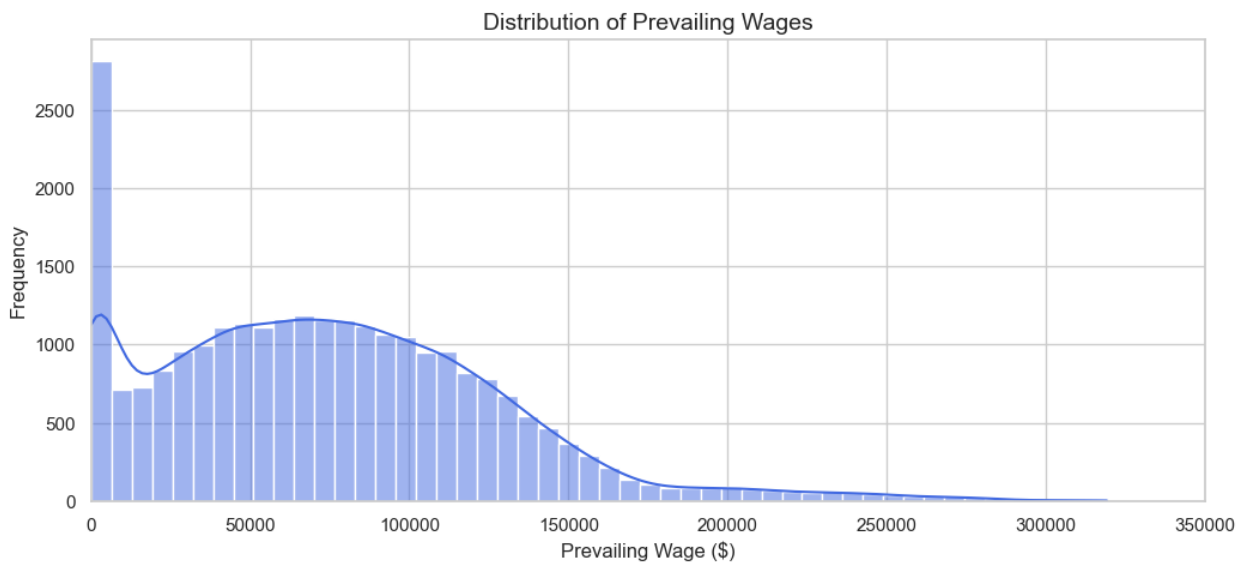
#### Insight from Visa Approval Distribution Chart:

- **Higher Approval Rate:**
- The majority of visa applications are Certified, significantly outnumbering Denied cases.
- This suggests a relatively high success rate in visa approvals.
- **Substantial Denial Rate:**
- Although more cases are certified, a considerable number of applications are denied (around half of the approved cases).
- This indicates that certain factors lead to rejection and may require further investigation.

#### 2)Visualization: Distribution of Prevailing Wages

```
plt.figure(figsize=(12, 5))
sns.histplot(df["prevailing_wage"], bins=50, kde=True,
color="royalblue")
plt.title("Distribution of Prevailing Wages", fontsize=14)
plt.xlabel("Prevailing Wage ($)")
```

```
plt.ylabel("Frequency")
plt.xlim(0, 350000) # Limiting to avoid extreme outliers
plt.show()
```



### Insights from the Distribution of Prevailing Wages: Right-Skewed Distribution:

The wage distribution is skewed to the right, meaning most wages are concentrated at lower values (below \$100,000). A few high-paying jobs pull the tail of the distribution toward the right.

#### High Frequency at Low Wages:

The highest peak is at very low wages, which could indicate: Part-time jobs or entry-level positions. Potential data errors (since some wages are extremely low, like 2.14 dollar).

#### Majority of Wages Fall Between 34,015 dollar and 107,735 dollar:

Most wage values cluster within this range, aligning with the 25th and 75th percentile values from summary statistics. This suggests a reasonable salary range for most applicants.

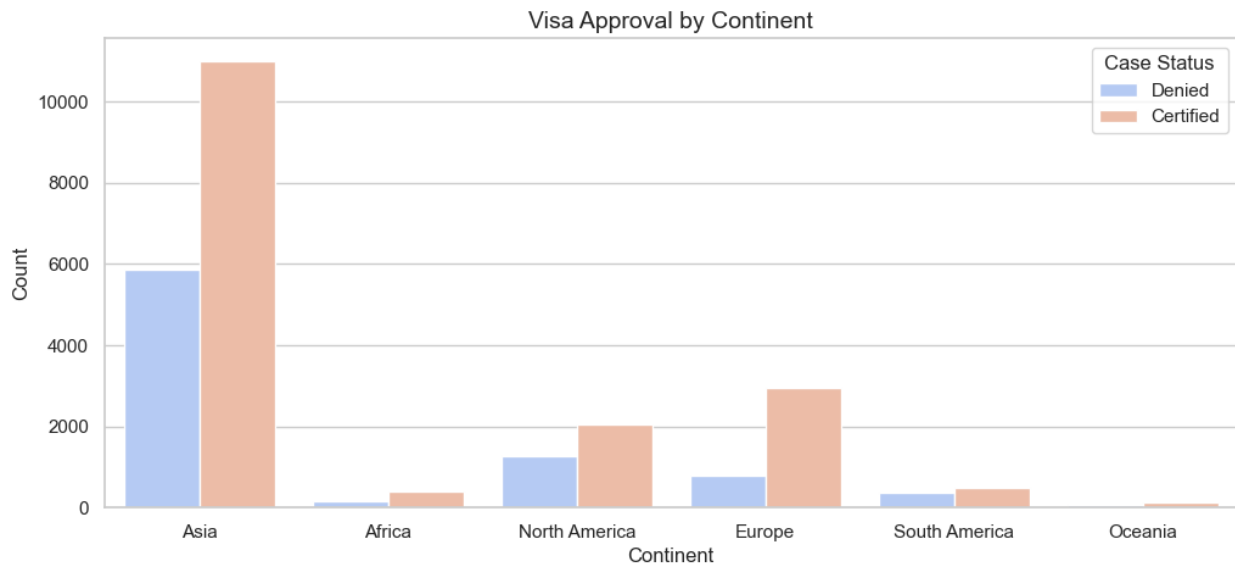
#### Outliers in High Salaries:

A small number of salaries exceed 200,000 dollar, suggesting high-paying roles in specialized industries. These could be executive or highly technical positions.

### 3)Visualization:Visa Approval by Continent

```
# Visa Approval by Continent
plt.figure(figsize=(12, 5))
sns.countplot(x="continent", hue="case_status", data=df,
palette="coolwarm")
plt.title("Visa Approval by Continent", fontsize=14)
plt.xlabel("Continent")
```

```
plt.ylabel("Count")
plt.legend(title="Case Status")
plt.show()
```



#### Insights from Visa Approval by Continent Chart:

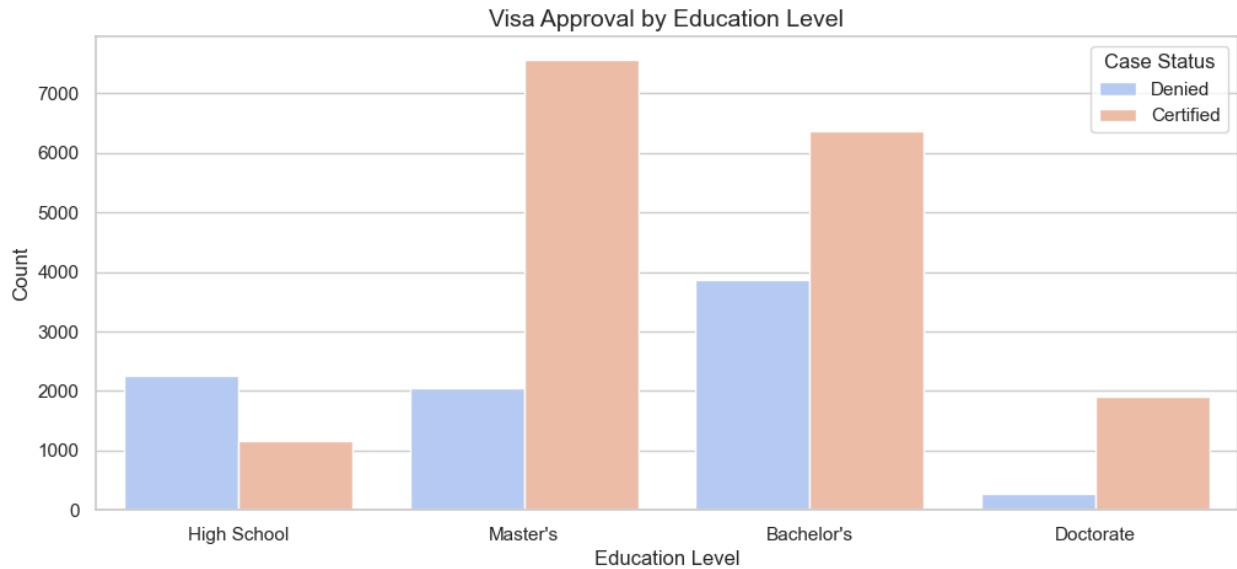
**Asia Dominates Visa Applications:** The highest number of applications come from Asia, with a significant number getting approved. However, Asia also has the highest denial count, suggesting a high application volume but also stricter scrutiny.

**Europe and North America Show Balanced Trends:** Both continents have a moderate number of applications, with approvals outpacing denials. This suggests a relatively higher acceptance rate compared to Asia.

**Low Application Rates from Africa, South America, and Oceania:** These continents have minimal representation in the dataset. The approval and denial rates are relatively small, possibly due to fewer applicants or specific job market demands.

#### 4)Visualization:Visa Approval by Education Level

```
plt.figure(figsize=(12, 5))
sns.countplot(x="education_of_employee", hue="case_status", data=df,
palette="coolwarm")
plt.title("Visa Approval by Education Level", fontsize=14)
plt.xlabel("Education Level")
plt.ylabel("Count")
plt.legend(title="Case Status")
plt.show()
```



**Insights from Visa Approval by Education Level Chart: Higher Education Increases Visa Approval Chances:** Master's degree holders have the highest approval rate, followed by Bachelor's degree holders. This suggests that employers and visa authorities prefer candidates with advanced qualifications.

**High Denial Rate for Bachelor's Degree Holders:** While many Bachelor's degree holders are approved, they also have a high number of denials. This could indicate a competitive job market where a Bachelor's degree alone may not always meet employer or visa requirements.

**High School Graduates Face High Rejections:** More high school graduates are denied visas than those approved. This suggests that lower education levels struggle with meeting job and visa requirements.

**Doctorate Holders Have Low Denial Rates but Fewer Applications:** The number of doctorate holders applying is relatively low, but they have a high approval rate. This suggests specialized positions that favor highly educated candidates.

---

## Conclusion from Visa Dataset Analysis

**Visa Approval Rate:** Around 67% of visa applications are approved, while 33% are denied. A significant number of applications get rejected, indicating possible strict criteria or common issues in applications.

**Wage Distribution:** Most applicants have a prevailing wage below 150,000 dollar. There are outliers with extremely high wages, but these are rare cases. The wage distribution is right-skewed, meaning a small percentage of jobs offer exceptionally high salaries.

**Approval by Continent:** Asia has the highest number of applications, but also a high number of denials. Europe and North America have higher approval rates, suggesting regional variations in approval likelihood.

**Education and Approval Rate:** Higher education increases the chances of approval. Master's degree holders have the highest approval rate. High school graduates face the most denials, showing that higher qualifications may be preferred for visa approval.

**Key Takeaways** Improving education qualifications and job experience could enhance visa approval chances. Applicants from Asia might face stricter scrutiny or higher competition. Employers should ensure wage offers are competitive to align with approval trends.

-----end of  
analysis-----

