Busiest Airports By Passenger Traffic Analysis (2022)

Overview

This analysis explores the busiest airports worldwide based on passenger traffic in 2022. Key tasks include data exploration, visualizations, and advanced insights into top airports and countries.

Tasks

1. Data Exploration
2. Data Visualizations
3. Advanced Insights

```
# Step 1: Import Libraries

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

# Step 2: Load Dataset

df = pd.read_csv(r"C:\Users\Nilay\Desktop\Busiest Airports Analysis\
busiest_airports_2022.csv")
df.head()
```

```
   Rank      lat      long
Airport  \
0   1.0  33.6324  -84.4277  Hartsfield–Jackson Atlanta International
Airport
1   2.0  32.8998  -97.0403           Dallas Fort Worth International
Airport
2   3.0  39.8561 -104.6737                       Denver International
Airport
3   4.0  41.9742  -87.9073                       O'Hare International
Airport
4   5.0  25.2532   55.3657                        Dubai International
Airport

                   Location               Country Code (IATA/ICAO)  \
0          Atlanta, Georgia         United States         ATL/KATL
1  Dallas–Fort Worth, Texas         United States         DFW/KDFW
2          Denver, Colorado         United States         DEN/KDEN
3         Chicago, Illinois         United States         ORD/KORD
4            Garhoud, Dubai  United Arab Emirates         DXB/OMDB
```

```
   Total passengers
0          93699630
1          73362946
2          69286461
3          68340619
4          66069981
```

```
# Step 3: Initial Exploration

# Basic Data Info
print('Dataset Shape:', df.shape)
print('\nDataset Info:')
df.info()
```

```
Dataset Shape: (50, 8)

Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 8 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   Rank              50 non-null     float64
 1   lat               50 non-null     float64
 2   long              50 non-null     float64
 3   Airport           50 non-null     object 
 4   Location          50 non-null     object 
 5   Country           50 non-null     object 
 6   Code (IATA/ICAO)  50 non-null     object 
 7   Total passengers  50 non-null     int64  
dtypes: float64(3), int64(1), object(4)
memory usage: 3.3+ KB
```

```
# Summary statistics
df.describe()
```

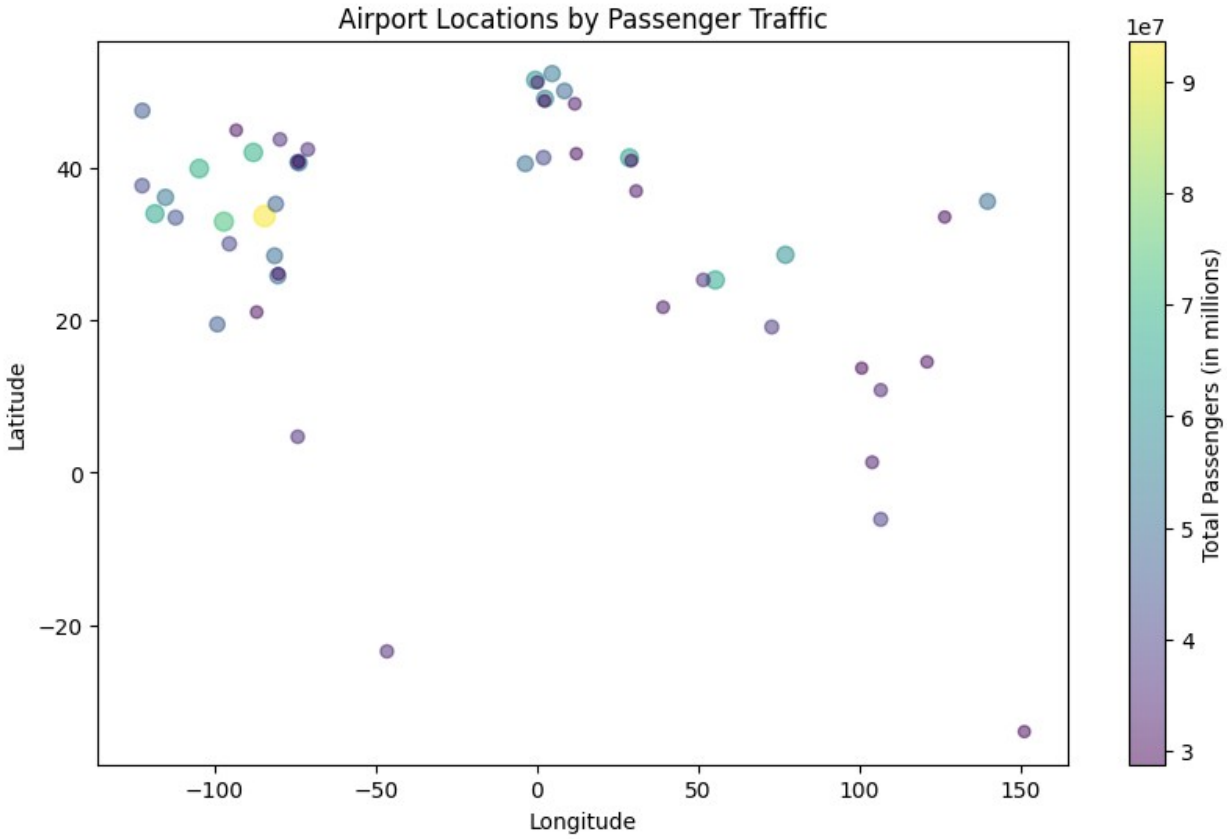
```
            Rank        lat         long  Total passengers
count  50.00000  50.000000   50.000000      5.000000e+01
mean   25.50000  30.802287  -15.451199      4.442828e+07
std    14.57738  18.186016   81.815943      1.471212e+07
min     1.00000 -33.950000 -122.381600      2.875431e+07
25%    13.25000  25.254925  -83.648463      3.165784e+07
50%    25.50000  35.381700   -2.012985      4.129557e+07
75%    37.75000  41.676370   37.075925      5.202524e+07
max    50.00000  52.308056  151.181700      9.369963e+07
```

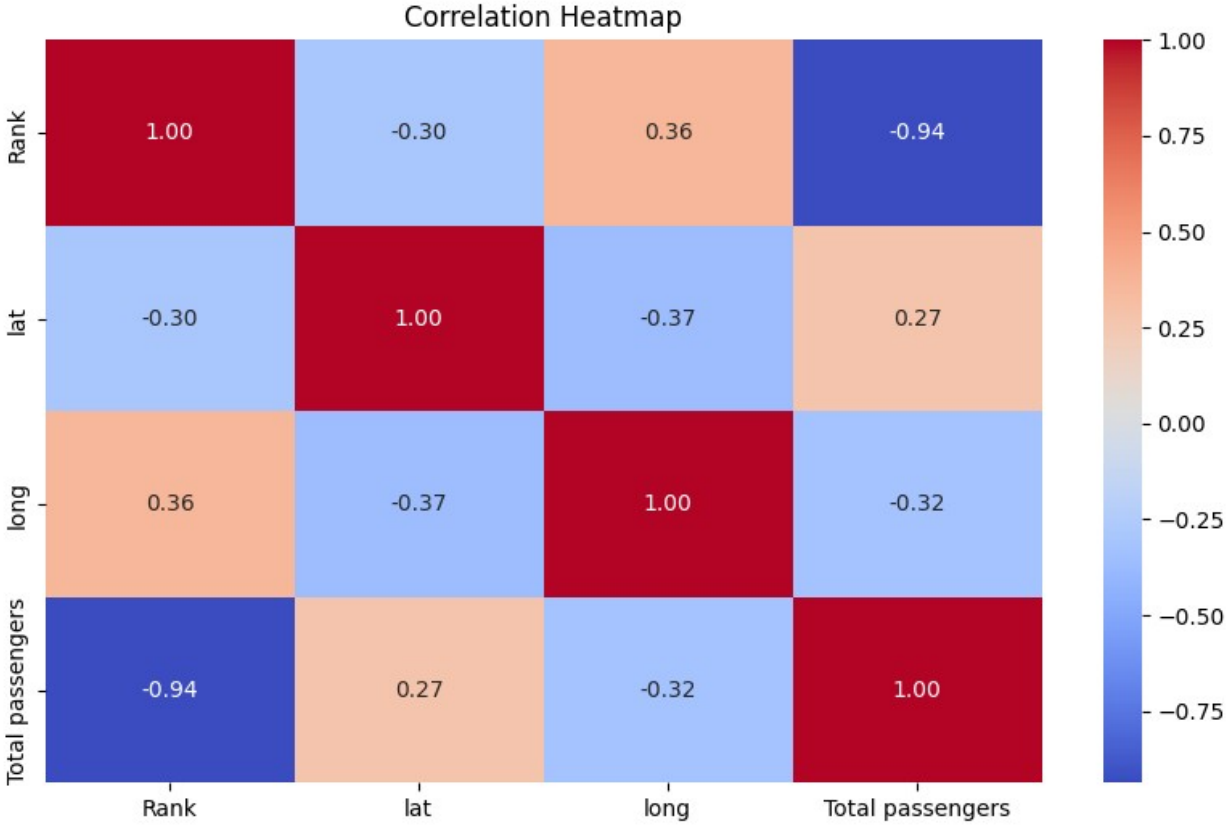
```
# Check duplicates and unique values
print('\nDuplicate Rows:', df.duplicated().sum())
print('\nUnique Values in Columns:')
print(df.nunique())
```

```
Duplicate Rows: 0

Unique Values in Columns:
Rank                50
lat                 50
long                50
Airport             50
Location            50
Country             24
Code (IATA/ICAO)    50
Total passengers    50
dtype: int64
```

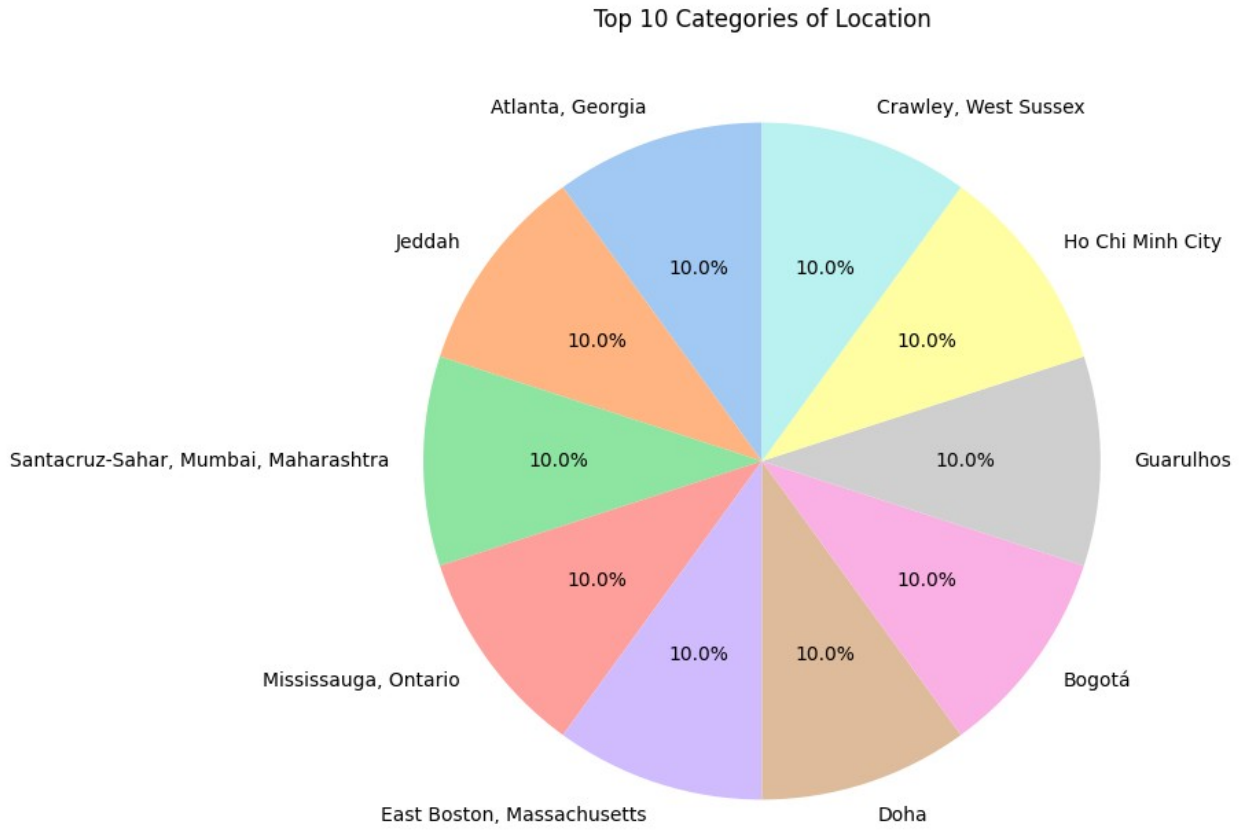
```python
# Pie chart for categorical columns
def plot_pie_chart(column, top_n=10):
    """Plots a pie chart for the top N categories of a column."""
    plt.figure(figsize=(8, 8))
    df[column].value_counts().head(top_n).plot.pie(
        autopct='%1.1f%%', startangle=90,
colors=sns.color_palette('pastel')
    )
    plt.title(f"Top {top_n} Categories of {column}")
    plt.ylabel('')  # Removes the y-axis label
    plt.show()

# Histogram for numerical columns
def plot_hist_chart(column):
    """Plots a histogram for a numerical column."""
    plt.figure(figsize=(8, 5))
    df[column].plot.hist(bins=20, color='skyblue', edgecolor='black')
    plt.title(f"Distribution of {column}")
    plt.xlabel(column)
    plt.ylabel('Frequency')
    plt.show()

# Correlation heatmap
def plot_correlation_heatmap():
    """Plots a heatmap showing correlations between numeric
columns."""
    # Select only numeric columns
    numeric_df = df.select_dtypes(include=['number'])
    # Compute the correlation matrix
    corr = numeric_df.corr()
    # Plot the heatmap
    plt.figure(figsize=(10, 6))
    sns.heatmap(corr, annot=True, cmap='coolwarm', fmt='.2f')
    plt.title("Correlation Heatmap")
    plt.show()
```

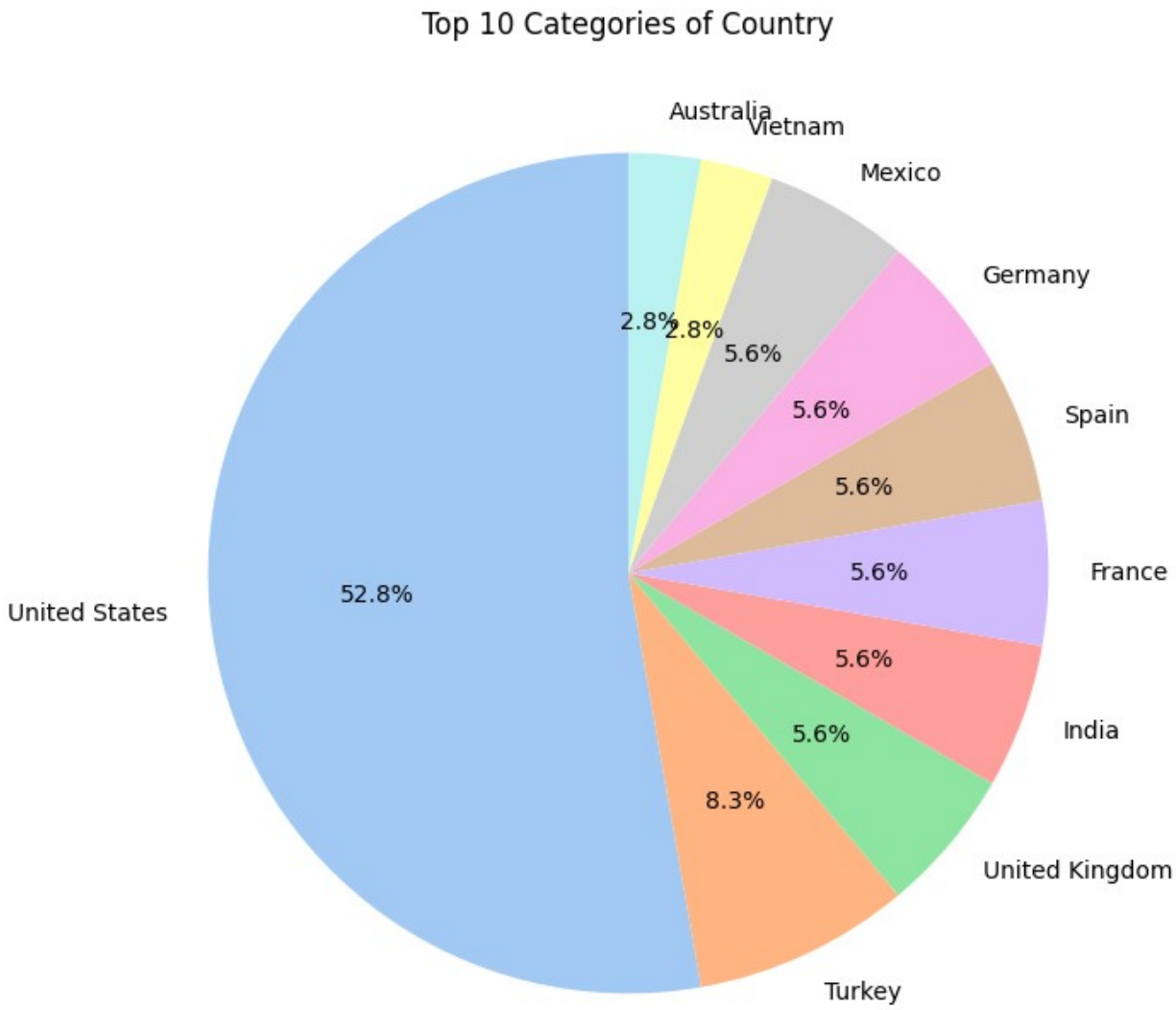
```
# Geospatial scatter plot
def plot_geospatial_scatter():
    """Plots a scatter plot of airport locations by longitude and
latitude."""
    plt.figure(figsize=(10, 6))
    scatter = plt.scatter(
        df['long'], df['lat'],
        s=df['Total passengers'] / 1000000, alpha=0.5,
        c=df['Total passengers'], cmap='viridis'
    )
    plt.colorbar(scatter, label="Total Passengers (in millions)")
    plt.title("Airport Locations by Passenger Traffic")
    plt.xlabel("Longitude")
    plt.ylabel("Latitude")
    plt.show()

### Step 5: Apply Visualizations

# Plot geospatial scatter
plot_geospatial_scatter()

# Plot correlation heatmap
plot_correlation_heatmap()

# Pie charts for categorical variables
for col in df.select_dtypes('object').columns:
    plot_pie_chart(col)

# Histograms for numerical variables
for col in df.select_dtypes(include=['float', 'int']).columns:
    plot_hist_chart(col)
```

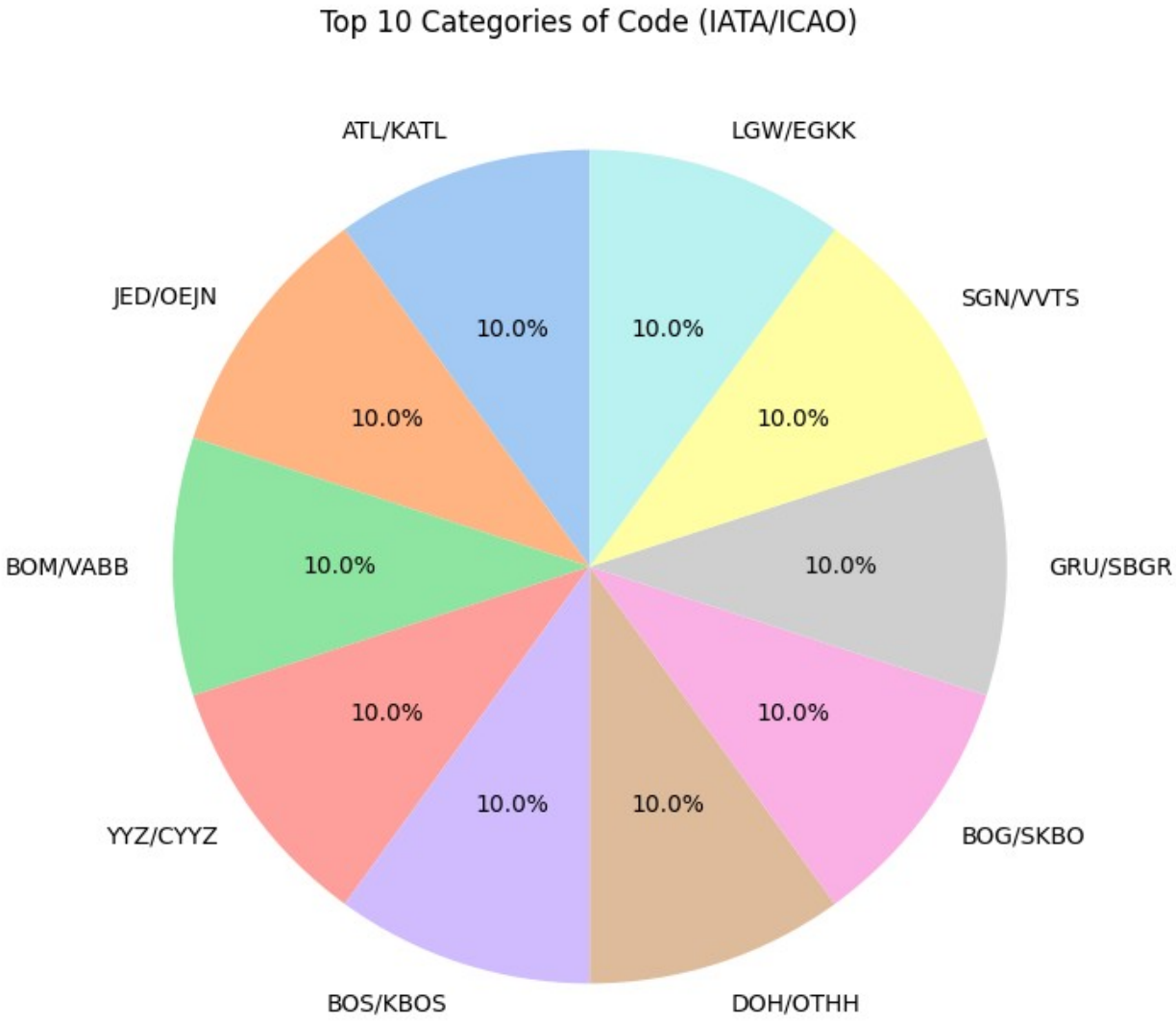

Airport Locations by Passenger Traffic
Latitude
Longitude
1e7
Total Passengers (in millions)
40
20
0
−20
−100
−50
0
50
100
150
9
8
7
6
5
4
3

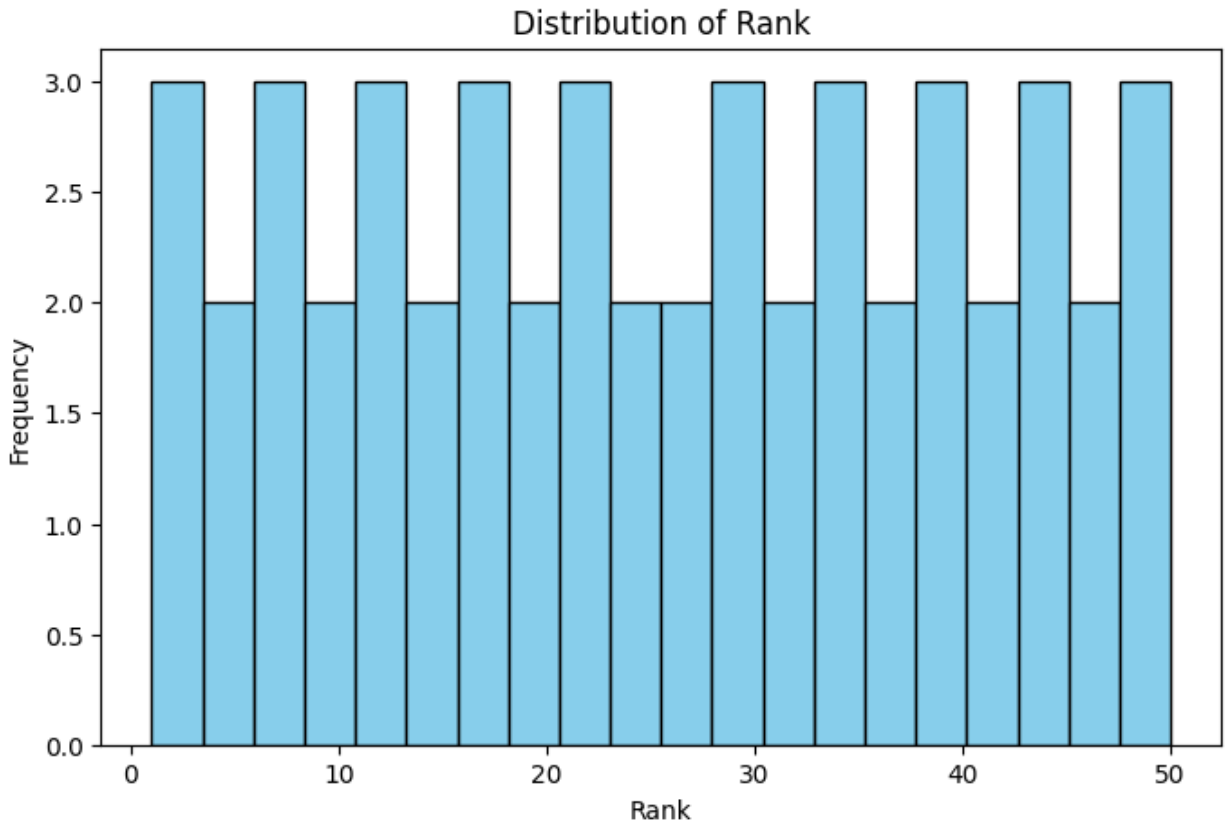

Correlation Heatmap
Rank
lat
long
Total passengers
1.00
-0.30
0.36
-0.94
-0.30
1.00
-0.37
0.27
0.36
-0.37
1.00
-0.32
-0.94
0.27
-0.32
1.00
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

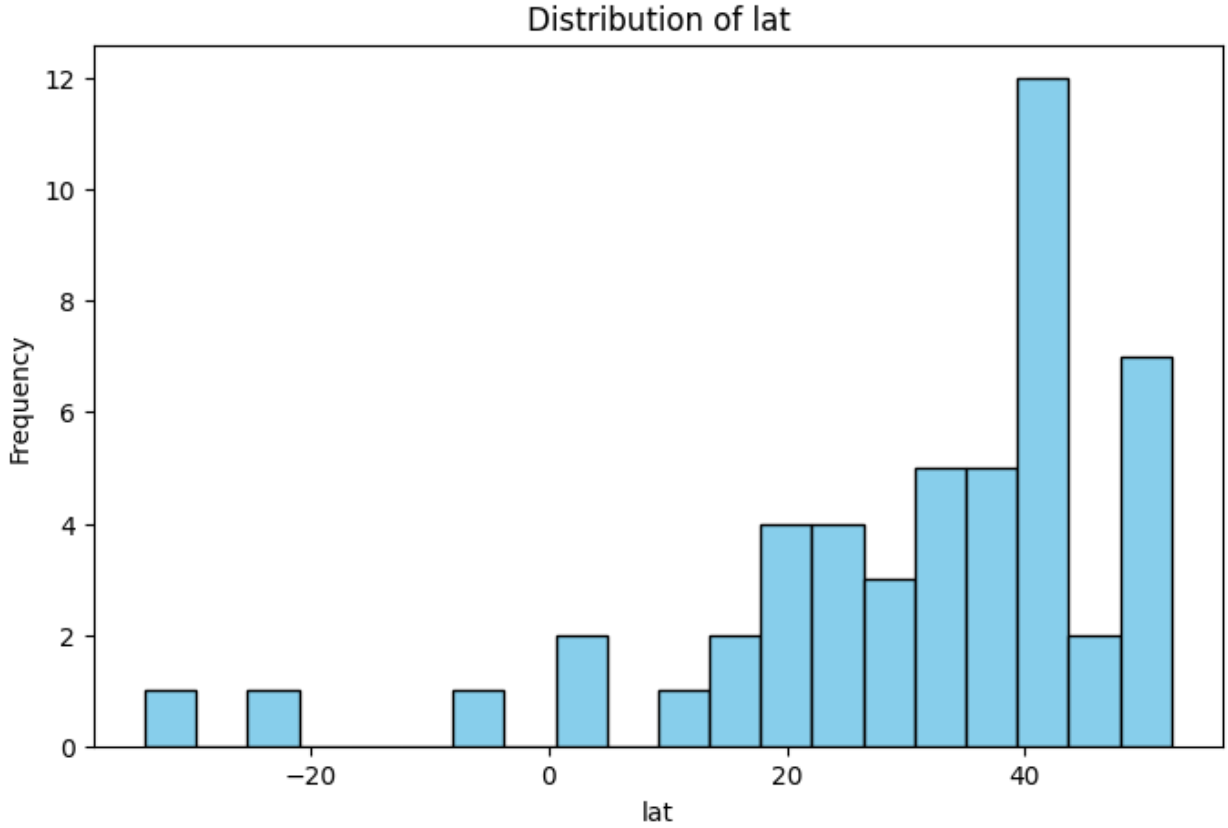

Top 10 Categories of Airport
Hartsfield–Jackson Atlanta International Airport
London Gatwick Airport
King Abdulaziz International Airport
Tan Son Nhat International Airport
Chhatrapati Shivaji Maharaj International Airport
São Paulo/Guarulhos International Airport
Toronto Pearson International Airport
El Dorado International Airport
Logan International Airport
Hamad International Airport
10.0%

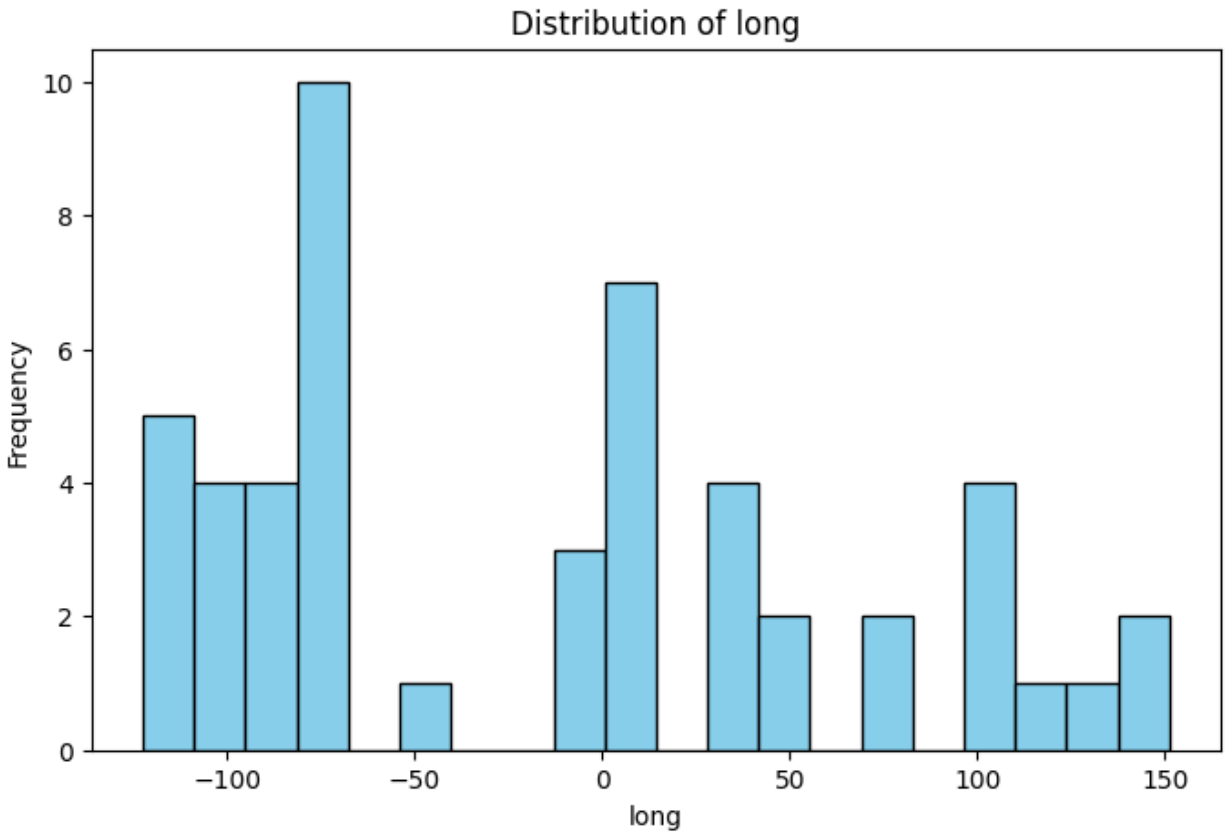

Top 10 Categories of Location
Atlanta, Georgia
Crawley, West Sussex
Jeddah
Ho Chi Minh City
Santacruz-Sahar, Mumbai, Maharashtra
Guarulhos
Mississauga, Ontario
Bogotá
East Boston, Massachusetts
Doha
10.0%

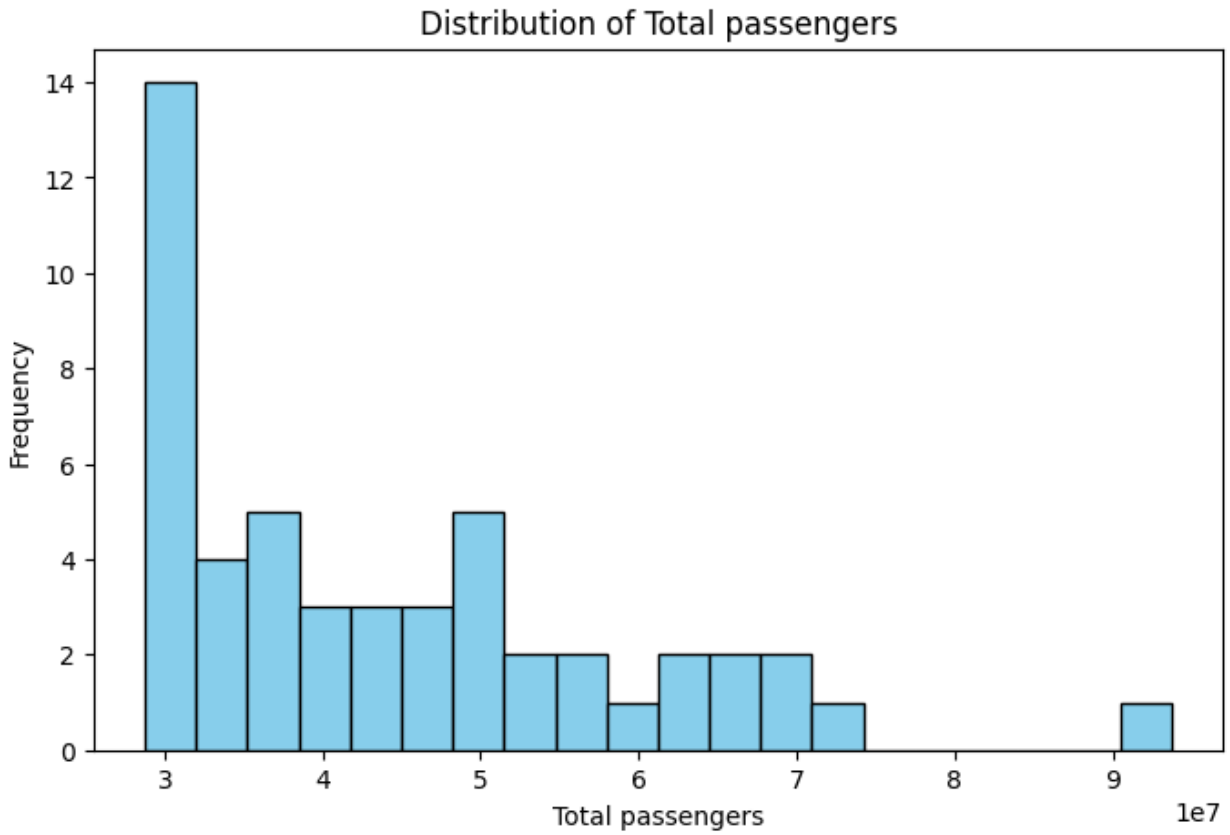

Top 10 Categories of Country
Australia
Vietnam
Mexico
Germany
Spain
France
India
United Kingdom
Turkey
United States
2.8%
2.8%
5.6%
5.6%
5.6%
5.6%
5.6%
5.6%
8.3%
52.8%


Top 10 Categories of Code (IATA/ICAO)
ATL/KATL
LGW/EGKK
JED/OEJN
SGN/VVTS
BOM/VABB
GRU/SBGR
YYZ/CYYZ
BOG/SKBO
BOS/KBOS
DOH/OTHH
10.0%
10.0%
10.0%
10.0%
10.0%
10.0%
10.0%
10.0%
10.0%
10.0%


Distribution of Rank
Frequency
Rank
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
10
20
30
40
50


Distribution of lat
Frequency
lat
0
2
4
6
8
10
12
−20
0
20
40

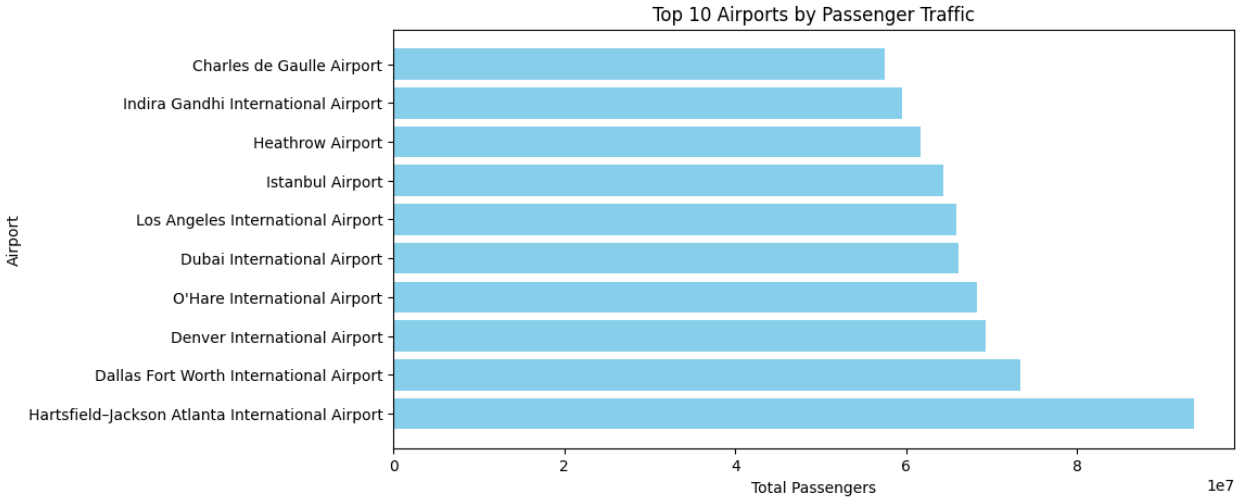

Distribution of long
Frequency
long
0
2
4
6
8
10
−100
−50
0
50
100
150

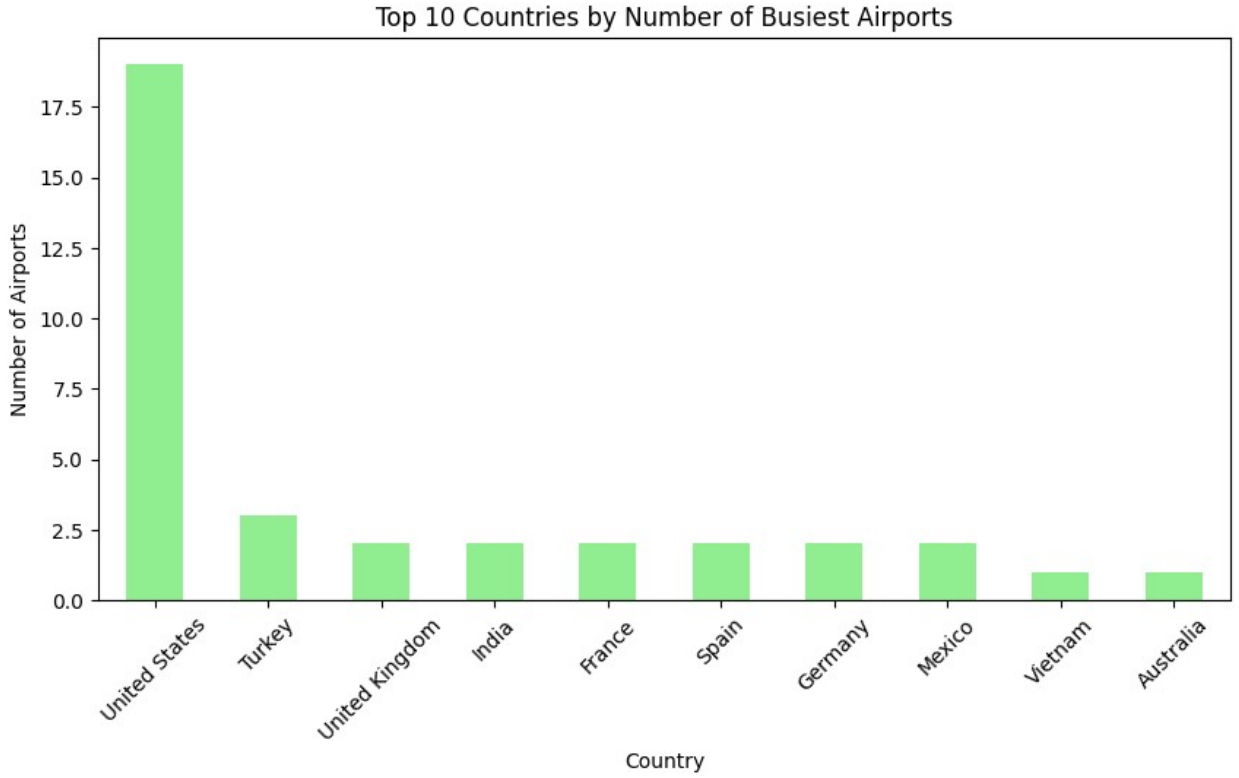



```
# Step 6: Advanced Insights

# Top 10 airports by passenger traffic
top_10_airports = df.nlargest(10, 'Total passengers')[['Airport',
'Location', 'Country', 'Total passengers']]
print('Top 10 Airports by Total Passengers:')
print(top_10_airports)

# Visualize top 10 airports
plt.figure(figsize=(10, 5))
plt.barh(top_10_airports['Airport'], top_10_airports['Total
passengers'], color='skyblue')
plt.title('Top 10 Airports by Passenger Traffic')
plt.xlabel('Total Passengers')
plt.ylabel('Airport')
plt.show()

# Distribution of airports across countries
plt.figure(figsize=(10, 5))
df['Country'].value_counts().head(10).plot(kind='bar',
color='lightgreen')
plt.title('Top 10 Countries by Number of Busiest Airports')
plt.xlabel('Country')
plt.ylabel('Number of Airports')
```

```
plt.xticks(rotation=45)
plt.show()
```

```
Top 10 Airports by Total Passengers:
                                        Airport  \
0  Hartsfield–Jackson Atlanta International Airport
1           Dallas Fort Worth International Airport
2                       Denver International Airport
3                       O'Hare International Airport
4                        Dubai International Airport
5                 Los Angeles International Airport
6                                 Istanbul Airport
7                                 Heathrow Airport
8               Indira Gandhi International Airport
9                          Charles de Gaulle Airport

                          Location               Country  Total
passengers
0                 Atlanta, Georgia         United States
93699630
1         Dallas–Fort Worth, Texas         United States
73362946
2                 Denver, Colorado         United States
69286461
3                Chicago, Illinois         United States
68340619
4                   Garhoud, Dubai  United Arab Emirates
66069981
5          Los Angeles, California         United States
65924298
6             Arnavutköy, Istanbul                Turkey
64289107
7               Hillingdon, London        United Kingdom
61614508
8                     Palam, Delhi                 India
59490074
9  Roissy-en-France, Île-de-France                France
57474033
```





Conclusion

This analysis provided a detail overview of the busiest airports by passenger traffic in 2022. Key findings include:

- Identification of top-performing airports.
- Geospatial analysis of airport locations.
- Insights into the distribution of airports across countries.
- Correlation analysis highlighting key relationships between numerical variables.