



**Rajalakshmi Engineering College (An
Autonomous Institution) Rajalakshmi
Nagar, Thandalam- 602105**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND
MACHINE LEARNING**

AD23632 - Framework for Data Visualization and Analytics

Mini Project: Tourist Impact Analysis

Report submitted by

REGISTRATION NUMBER : 231501167

STUDENT NAME : S SURYA PRAJIN

YEAR : 2023-2027

SUBJECT CODE : AD23632



**Rajalakshmi Engineering College (An
Autonomous Institution) Rajalakshmi
Nagar, Thandalam- 602105**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND
MACHINE LEARNING**

AD23632 - Framework for Data Visualization and Analytics

Mini Project: Tourist Impact Analysis

Report submitted by

REGISTRATION NUMBER : 231501167

STUDENT NAME : S SURYA PRAJIN

YEAR : 2023-2027

SUBJECT CODE : AD23632

(Approved / Not Approved)
EXAMINER 1

(Approved / Not Approved)
EXAMINER 2

(Approved/Not Approved)
EXAMINER 3

(Approved / Not Approved)
HoD/AIML

Table of Contents

Chapter	Page No.
Abstract	3
Introduction	4
Dataset description	5
Objectives	6
Methodology	7
Python implementation	8
Power BI dashboard	9
Tableau dashboard	10
Analysis & findings	11
Conclusion	12
Future scope	13
Appendix (code)	15

Chapter 1: Abstract

The growing significance of tourism as a driver of national and regional economies has become a focal point for researchers and policymakers alike. This project aims to provide a comprehensive analysis of the relationship between tourism activities and economic performance by leveraging publicly available tourism and macroeconomic datasets. The dataset includes key indicators such as tourism revenue, contribution of tourism to GDP, employment in tourism-related sectors, and gross value added (GVA), along with complementary variables such as regional development levels and annual growth trends. Unlike time-series behavioral data, this dataset is aggregated and cross-sectional across multiple countries and years, offering a broad perspective on how tourism influences local and global economic structures.

The objectives of this study are threefold. First, it seeks to identify correlations between tourism growth and key economic indicators such as GDP contribution, employment generation, and revenue trends. Second, it examines how regional and national factors—such as development status, geographic location, and policy orientation—interact with tourism dynamics to influence economic performance. Third, it investigates whether post-crisis recovery patterns, particularly after global events such as the COVID-19 pandemic, show differential effects across regions. To achieve these objectives, the project employs a multi-tool analytical framework that integrates data exploration and visualization in Python, dashboard creation in Power BI, and interactive geo-visual storytelling in Tableau.

This study emphasizes the intricate relationship between tourism and economic development, highlighting the potential of the tourism sector as a catalyst for sustainable growth, job creation, and regional balance. It also demonstrates the value of combining multiple analytical tools—Python for quantitative exploration, Power BI for analytical dashboards, and Tableau for visual storytelling—to deliver both empirical rigor and intuitive insights. The findings derived from this analysis can assist policymakers, economists, and tourism planners in understanding the economic significance of tourism and in formulating strategies to maximize its positive impact on the local and global economy.

Chapter 2: Introduction

Tourism has emerged as one of the most dynamic sectors of the global economy, influencing not only economic growth but also employment generation, cultural exchange, and sustainable development. As countries increasingly recognize tourism as a strategic driver of prosperity, understanding its multifaceted impact on local economies has become essential. While tourism contributes directly through expenditure, job creation, and investment, it also generates indirect effects through infrastructure development, trade stimulation, and social innovation. This project seeks to analyze the complex relationship between tourism activity and economic performance using structured, cross-country datasets that include key indicators such as tourism's contribution to GDP, gross value added (GVA), and employment in tourism-related industries.

The study adopts a comprehensive analytical framework that integrates economic, social, and regional dimensions. It examines how tourism influences national and regional economies, identifies variations in performance across different countries, and highlights the role of policy and development level in shaping tourism's economic footprint. Unlike time-series forecasting models, this dataset provides a cross-sectional perspective that enables the exploration of comparative patterns and relationships among multiple nations and economic indicators.

By combining **Python** for exploratory and statistical analysis, **Power BI** for analytical dashboards, and **Tableau** for interactive visualization and storytelling, the project offers a multi-tool approach that merges data-driven rigor with intuitive communication. The ultimate goal is to uncover meaningful insights into how tourism drives economic outcomes, provide visual narratives that enhance policy understanding, and support evidence-based decision-making for sustainable tourism growth and economic resilience.

Chapter 3: Dataset Description

The dataset used in this project contains a wide range of tourism and economic variables that help understand how tourism affects a country's economy. It is a structured, cross-country dataset collected from reliable sources such as the OECD and World Bank. The data is not real-time but is organized by country and year, making it useful for comparing trends and relationships between tourism and economic growth.

Key variables include:

- Geographical and Temporal Attributes:

Country name, country code, and year — used to compare tourism and economic data across countries and over different time periods.

- Tourism Indicators:

Tourism receipts, international visitor arrivals, tourism exports, tourism departures, and tourism expenditures — showing both incoming and outgoing tourism activity and their financial impact.

- Economic Indicators:

Gross Domestic Product (GDP), inflation rate, and unemployment rate — showing the overall economic conditions that affect tourism and national growth.

- Temporal Attributes:

Year of observation — helps to study changes and trends in tourism and economic performance over time.

- Supplementary Contextual Variables:

Although not directly included, more details like tourism policies, sustainability efforts, and infrastructure data can be added for deeper analysis.

This dataset is valuable because it connects tourism performance with economic growth indicators. It provides a clear overview of tourism's role in the economy, helping identify patterns and trends that support sustainable development and job creation.

Chapter 4: Objective

The main objective of this project is to examine the impact of tourism on national and regional economic performance. To achieve this, the study defines specific research aims that provide clarity and direction:

1. Correlation Analysis:

Identify relationships between key tourism indicators (such as tourism receipts, international arrivals, and tourism share of GDP) and major economic outcomes (GDP growth, employment, and inflation rate).

2. Regional and Economic Moderators:

Explore how factors such as a country's development level, regional location, and economic structure interact with tourism activity to influence economic growth and stability.

3. Comparative Insights:

Compare tourism and economic performance across different countries and time periods to identify patterns, growth trends, and variations in tourism's contribution to national economies.

4. Policy and Development Implications:

Investigate how tourism-driven growth aligns with sustainable development goals, and how policy measures, infrastructure investment, and sustainability initiatives shape tourism's long-term economic impact.

5. Tool Demonstration:

Showcase the complementary strengths of analytical tools—**Python** for data cleaning, statistical analysis, and trend exploration; **Power BI** for interactive dashboards and comparative analysis; and **Tableau** for visual storytelling and geo-spatial insights.

By fulfilling these objectives, the project aims to deliver both analytical depth and practical relevance. For policymakers, it provides evidence-based insights for tourism development strategies. For economists, it offers data-backed evaluations of tourism's contribution to economic growth. Finally, for researchers, it demonstrates a comprehensive methodological framework that integrates multiple analytical and visualization tools to study the economic impact of tourism.

Chapter 5: Methodology

The methodology follows a multi-step approach to ensure systematic, accurate, and insightful analysis of tourism's impact on the economy:

1. Data Preprocessing:

Raw tourism and economic data from OECD and World Bank are cleaned in Python—handling missing values, standardizing columns, converting data types, and removing duplicates for consistency.

2. Exploratory Data Analysis (EDA):

Descriptive statistics, correlations, and visualizations (scatter plots, line charts) identify patterns and relationships between tourism and economic variables, comparing trends across developed and developing countries.

3. Feature Engineering:

Derived metrics like tourism contribution ratio, employment dependency ratio, and tourism growth rate are created to capture tourism's scale and economic impact.

4. Visualization Tools:

Python is used for statistical plots and EDA; Power BI builds interactive dashboards for filtering and comparison; Tableau develops geo-visual and storytelling dashboards to illustrate global and regional tourism impacts.

5. Interpretation and Analysis:

Results are contextualized to assess tourism's influence on GDP and employment, with regression or trend analyses exploring predictive relationships and regional policy implications.

Chapter 6: Python Implementation

Python serves as the primary environment for data preprocessing and exploratory analysis of tourism's economic impact. Libraries such as pandas, numpy, matplotlib, and seaborn are used for cleaning, summarizing, and visualizing data. The workflow begins by importing datasets from OECD and World Bank, standardizing column names, and converting key variables (GDP, tourism receipts, employment shares) into numeric types. Missing or inconsistent entries are addressed through imputation or removal.

Visualizations form the core of the Python workflow. Scatter plots and line charts examine relationships and trends between tourism receipts, GDP, and employment, while boxplots highlight differences across regions or economic classifications. Correlation heatmaps reveal interdependencies among economic and tourism indicators, showing, for example, how tourism growth relates to GDP changes. Comparative analyses across developed and developing countries help identify patterns in tourism's economic contribution.

Feature engineering is applied to create derived metrics such as tourism contribution ratio, employment dependency ratio, and tourism growth rate. These features deepen insights into both the scale and efficiency of tourism's economic role. Plots are exported for inclusion in reports or integration with Power BI and Tableau dashboards.

Python ensures a reproducible, transparent, and data-driven foundation, enabling systematic analysis of tourism's impact on the economy.

Chapter 7: Power BI Dashboard

Power BI is used to create interactive dashboards, offering stakeholders a business friendly way to explore the dataset. Data is imported from the cleaned CSV generated through Python preprocessing. Within Power BI, fields are classified appropriately numeric values for social media time and productivity scores, categorical values for job type and platform preference.

Visualizations include:

- **Line and bar charts** comparing average productivity against social media time.
- **Scatter plots** displaying correlations between notifications and productivity.
- **Stacked bar charts** showing differences in platform preference by gender or job type.
- **Slicers and filters** for users to interactively explore the dataset by age, gender, or profession.



Fig 7.1: Power BI Dashboard

Chapter 8: Tableau Dashboard

Tableau complements Power BI by creating visually engaging dashboards that emphasize storytelling and presentation. The cleaned tourism and economic datasets are imported, and calculated fields are created, such as tourism contribution ratio, employment dependency ratio, and tourism growth rate.

Dashboards combine multiple sheets—maps, bar charts, and line charts—into a cohesive narrative that illustrates global and regional tourism trends, economic contributions, and employment impacts. The visual storytelling capabilities of Tableau make it effective for communicating insights to policymakers, stakeholders, or the public, highlighting patterns, comparisons, and the overall economic influence of tourism.

Tourist Impact Analysis Dashboard

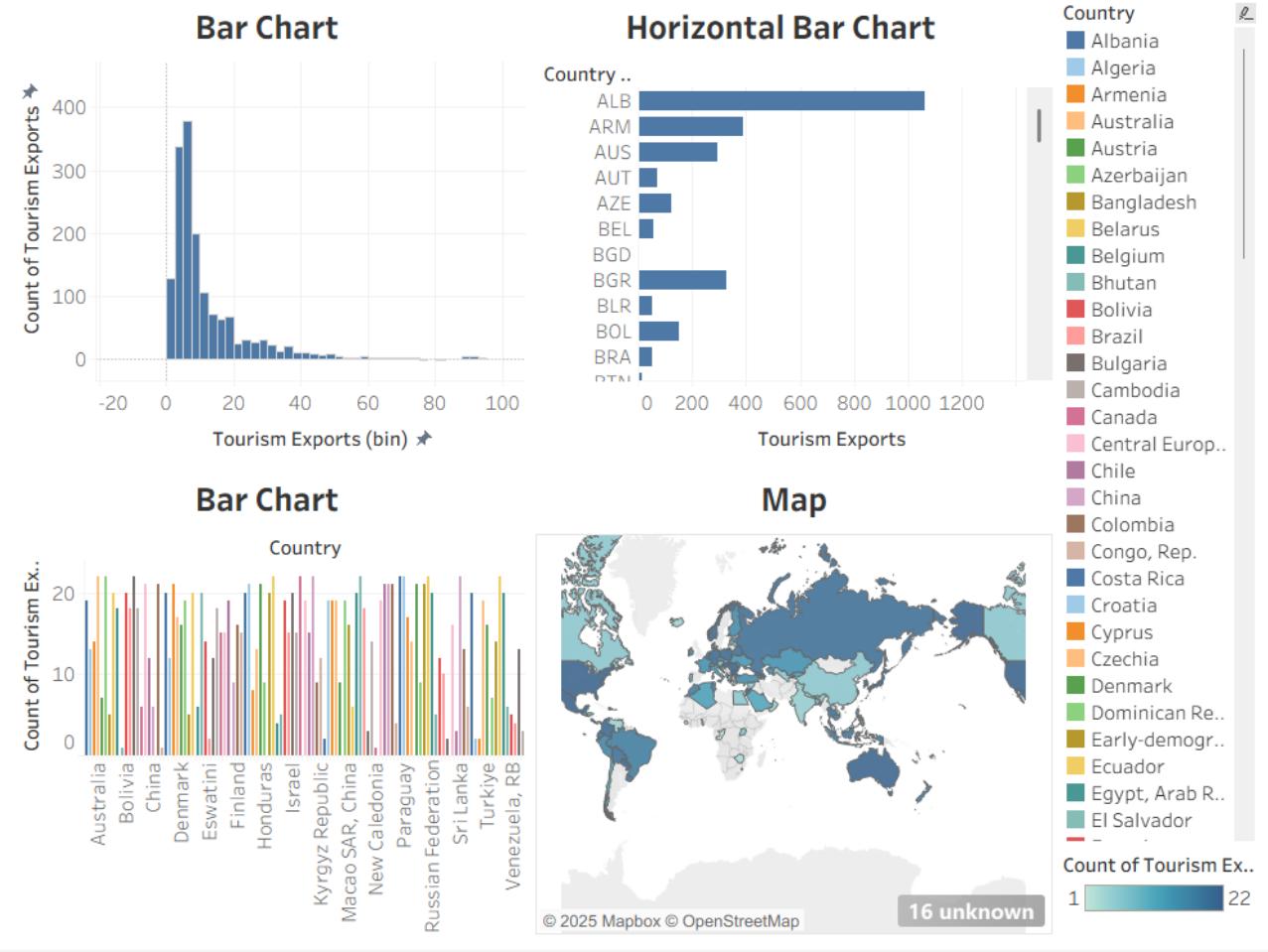


Fig 8.1: Tableau Dashboard

Chapter 9: Analysis

The analysis highlights several key patterns in tourism's economic impact. First, higher tourism receipts generally correlate with stronger GDP growth and greater employment shares, though the strength of this relationship varies across countries and regions. Developed economies with robust tourism infrastructure tend to experience steady economic gains, while developing economies show more variability, reflecting sensitivity to fluctuations in tourist arrivals.

Second, tourism's contribution interacts with broader economic and policy factors. Regions with targeted tourism strategies, sustainable initiatives, or diversified tourism offerings often see higher efficiency in employment and GDP contributions. Conversely, areas reliant on seasonal or unregulated tourism experience uneven benefits, sometimes leading to economic instability.

Third, demographic and regional patterns emerge. Service-oriented and knowledge-based employment sectors benefit differently, with some regions showing higher dependency ratios on tourism employment. Spatial analyses reveal that certain countries or regions disproportionately benefit from tourism, highlighting the importance of strategic planning and investment to maximize economic returns.

Overall, tourism plays a multifaceted role in the economy: it drives growth, generates employment, and can support regional development, but the magnitude and stability of its impact depend on infrastructure, policy frameworks, and the balance between sustainable and high-volume tourism.

Chapter 9: Conclusion

The study concludes that tourism has a measurable and multifaceted impact on economic growth and employment. While higher tourism receipts and employment shares are generally associated with stronger GDP growth, the magnitude and stability of these effects vary across regions and countries. Developed economies with robust infrastructure tend to benefit steadily, whereas developing regions show greater variability and higher dependency on tourism activity.

For policymakers and stakeholders, this highlights the importance of strategic planning, investment in tourism infrastructure, and sustainable tourism initiatives to maximize economic benefits. Diversifying tourism offerings and managing seasonal fluctuations can improve stability and efficiency, while targeted policies can enhance employment outcomes and regional development.

Future work could expand this study by incorporating longitudinal data to track tourism's impact over time, integrating more granular datasets such as regional or sector-specific statistics, and applying advanced econometric or machine learning models to better predict the relationship between tourism and economic indicators. Comparative studies across cultural and economic contexts, as well as controlled interventions (e.g., policy changes in tourism management), would further strengthen causal insights and guide effective economic planning.

Chapter 10: Appendix

10.1 Python Code

Import required libraries:-

```
import pandas as pd  
import numpy as np  
import seaborn as sns  
import matplotlib.pyplot as plt  
import plotly.graph_objects as go  
import plotly.express as px
```

Check data link by loading head:-

```
file_path = "cleaned_dataset_1.csv" # <-- change path if needed  
df = pd.read_csv(file_path)
```

```
df.columns = df.columns.str.strip().str.lower()
```

Data Cleaning:-

```
num_cols = ['tourism_receipts', 'tourism_arrivals', 'tourism_exports',  
           'tourism_departures', 'tourism_expenditures', 'gdp',  
           'inflation', 'unemployment']  
for col in num_cols:  
    df[col] = pd.to_numeric(df[col], errors='coerce')  
  
df = df.dropna(subset=['country', 'year', 'gdp'])
```

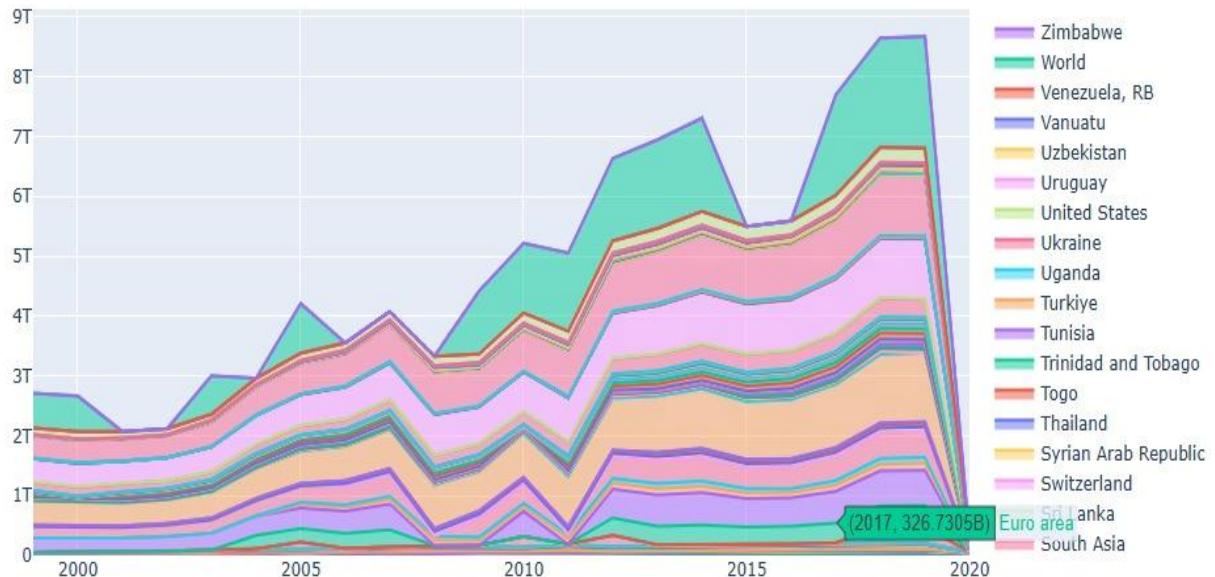
Derived indicators:-

```
df['tourism_gdp_share'] = (df['tourism_receipts'] / df['gdp']) * 100  
df['receipts_per_arrival'] = df['tourism_receipts'] / df['tourism_arrivals']  
df['export_dependency'] = (df['tourism_exports'] / df['gdp']) * 100  
  
print(f" ✅ Dataset ready. Shape: {df.shape}")
```

Visualizations - STREAMGRAPH - Tourism Receipts Over Time

```
stream_data = df.pivot(index='year', columns='country',
values='tourism_receipts').fillna(0)
fig = go.Figure()
for country in stream_data.columns:
    fig.add_trace(go.Scatter(
        x=stream_data.index, y=stream_data[country],
        mode='lines', stackgroup='one', name=country
    ))
fig.update_layout(title='Streamgraph: Tourism Receipts by Country Over Time')
fig.show()
```

Streamgraph: Tourism Receipts by Country Over Time



Visualizations - LOLLIPOP CHART - Latest Year Tourism GDP Share

```
latest_year = df['year'].max()  
latest_data = df[df['year'] == latest_year].sort_values(by='tourism_gdp_share',  
ascending=True)
```

```
fig = go.Figure()
```

```
fig.add_trace(go.Scatter(
```

```
    x=latest_data['tourism_gdp_share'],  
    y=latest_data['country'],  
    mode='markers+lines',  
    marker=dict(size=10, color='crimson'),  
    line=dict(color='gray', width=2)
```

```
))
```

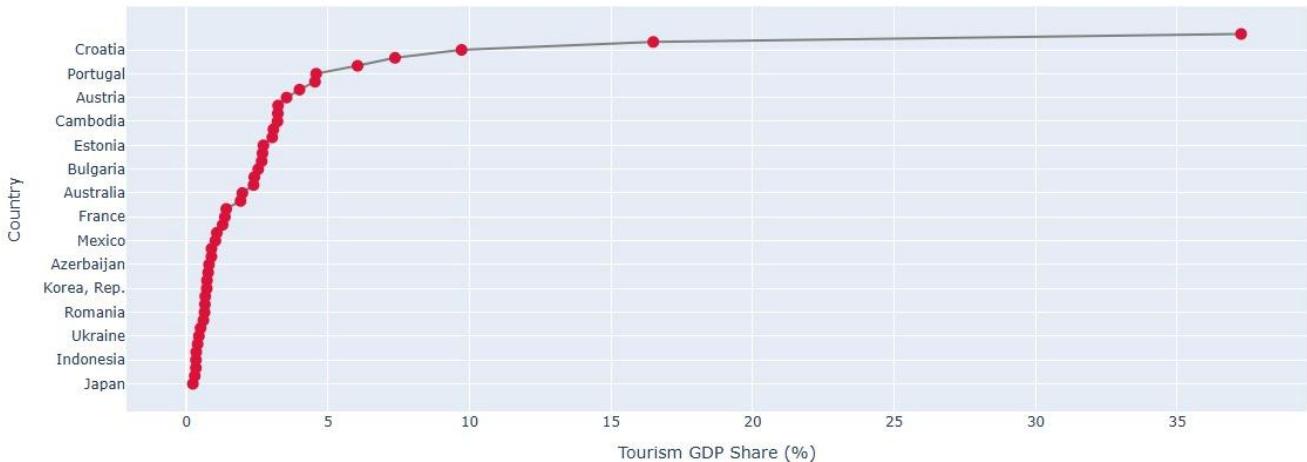
```
fig.update_layout(
```

```
    title=f'Lollipop Chart: Tourism GDP Share by Country ({latest_year})',  
    xaxis_title='Tourism GDP Share (%)', yaxis_title='Country'
```

```
)
```

```
fig.show()
```

Lollipop Chart: Tourism GDP Share by Country (2020)



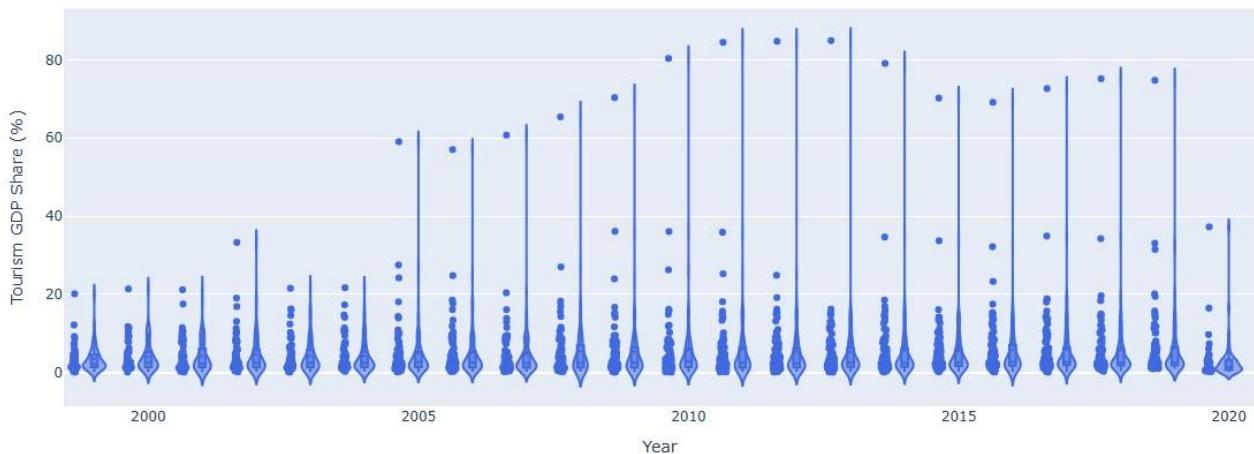
Visualizations - RAINCLOUD (VIOLIN) PLOT - Distribution of Tourism GDP Share Over Years

```

fig = go.Figure()
fig.add_trace(go.Violin(
    x=df['year'], y=df['tourism_gdp_share'],
    points='all', box_visible=True, line_color='royalblue',
    meanline_visible=True
))
fig.update_layout(
    title='Raincloud Plot: Distribution of Tourism GDP Share Over Years',
    yaxis_title='Tourism GDP Share (%)', xaxis_title='Year'
)
fig.show()

```

Raincloud Plot: Distribution of Tourism GDP Share Over Years



Visualizations - CORRELOGRAM - Economic Indicators Correlation

```

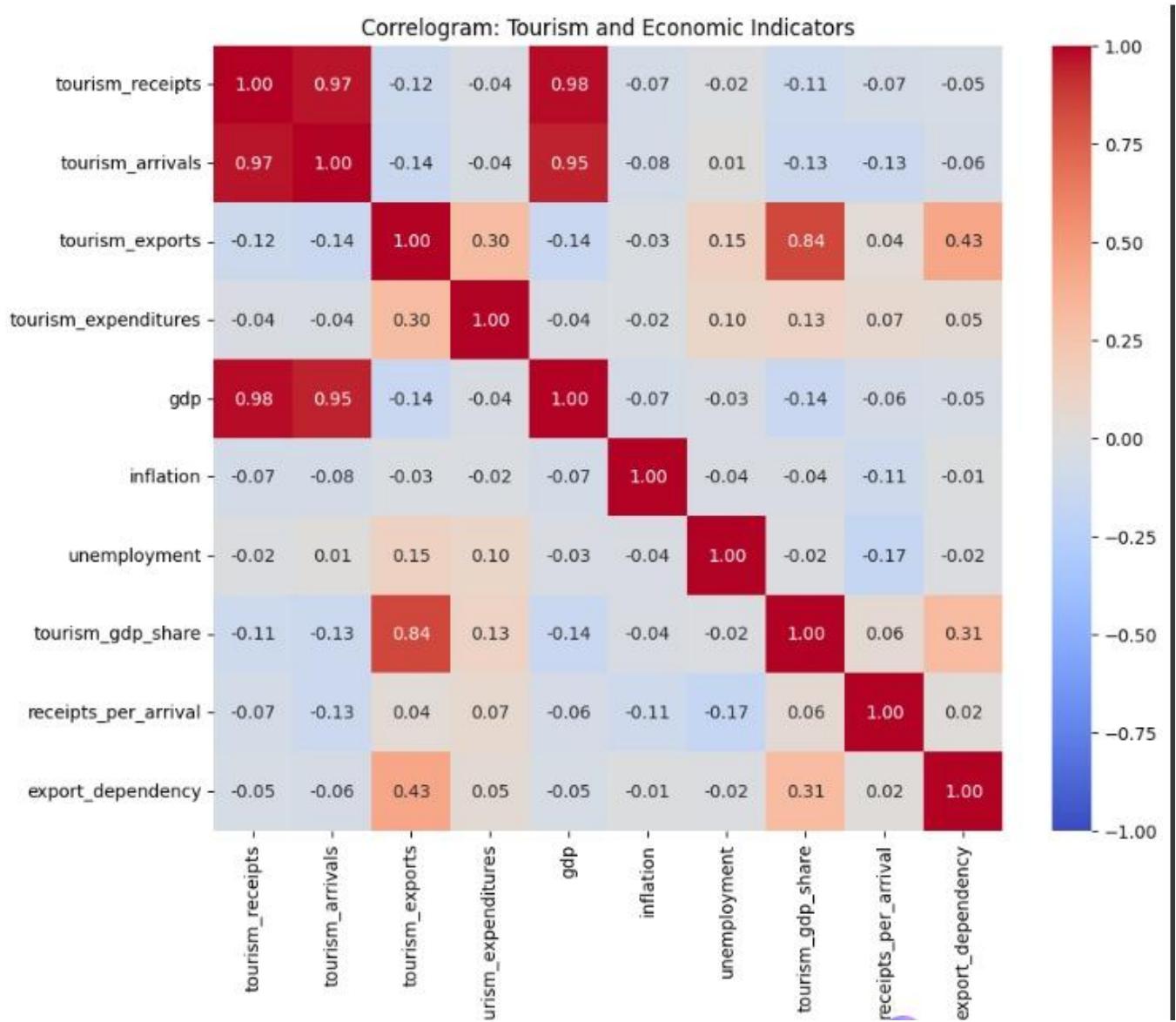
corr_cols = ['tourism_receipts', 'tourism_arrivals', 'tourism_exports',
            'tourism_expenditures', 'gdp', 'inflation', 'unemployment',
            'tourism_gdp_share', 'receipts_per_arrival', 'export_dependency']
corr = df[corr_cols].corr()

```

```

plt.figure(figsize=(10, 8))
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f", vmin=-1, vmax=1)
plt.title('Correlogram: Tourism and Economic Indicators')
plt.show()

```

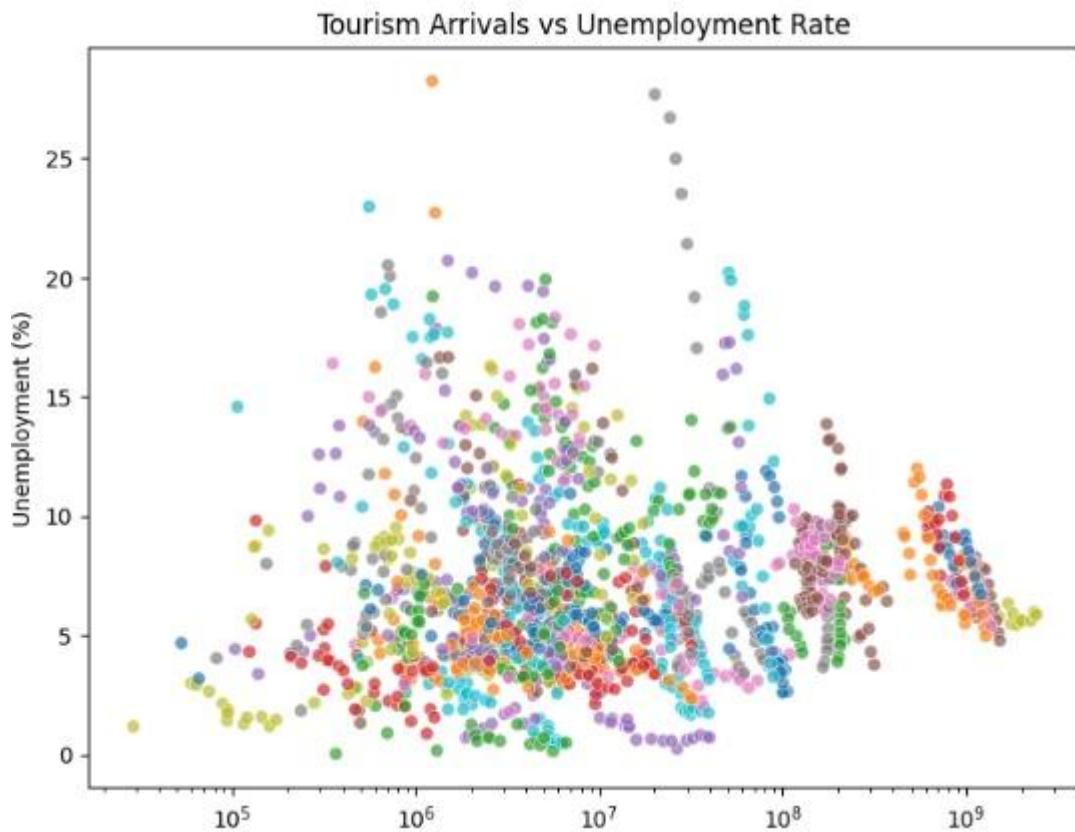


Visualizations - SCATTER - Tourism Arrivals vs Unemployment

```
plt.figure(figsize=(8, 6))
sns.scatterplot(
    data=df, x='tourism_arrivals', y='unemployment',
    hue='country', palette='tab10', alpha=0.7
)
```

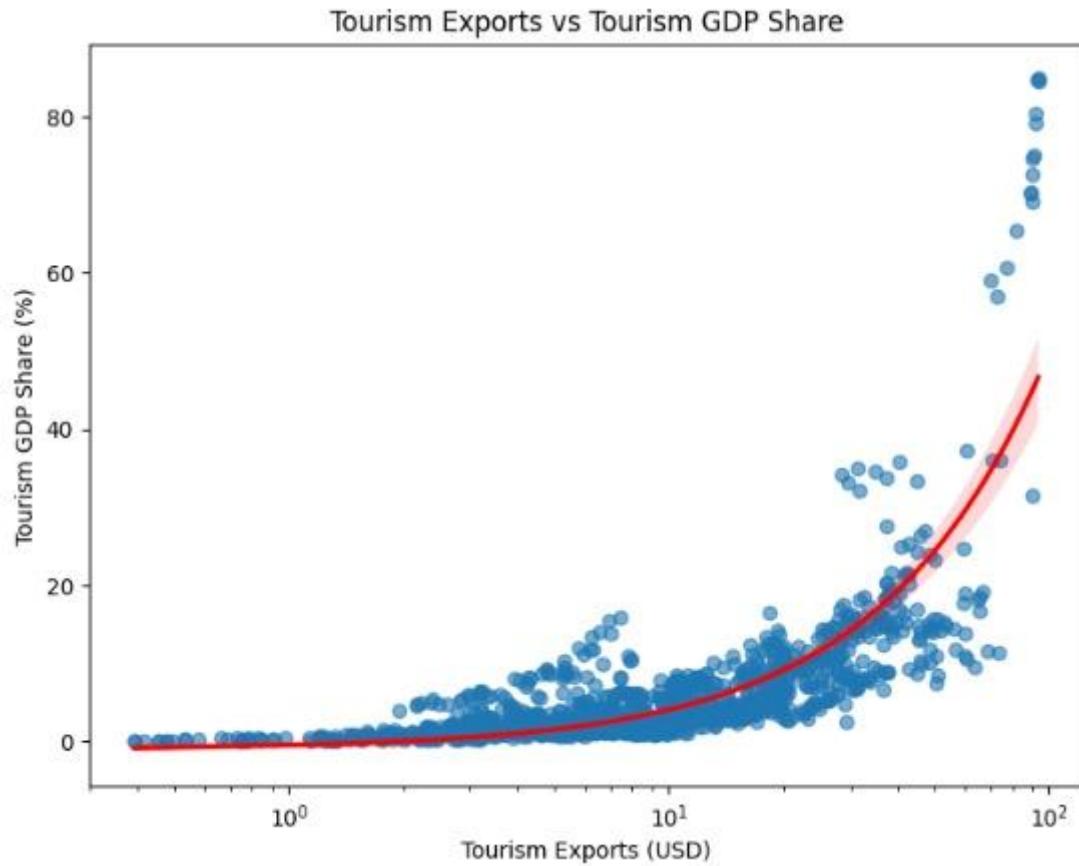
```
plt.title('Tourism Arrivals vs Unemployment Rate')
plt.xlabel('Tourism Arrivals')
plt.ylabel('Unemployment (%)')
plt.xscale('log')
plt.legend([],[], frameon=False)
```

```
plt.show()
```



Visualizations - SCATTER + REGRESSION - Tourism Exports vs Tourism GDP Share

```
plt.figure(figsize=(8, 6))
sns.regplot(
    data=df, x='tourism_exports', y='tourism_gdp_share',
    scatter_kws={'alpha':0.6}, line_kws={'color':'red'}
)
plt.title('Tourism Exports vs Tourism GDP Share')
plt.xlabel('Tourism Exports (USD)')
plt.ylabel('Tourism GDP Share (%)')
plt.xscale('log')
plt.show()
```



Visualizations - LINE CHART - Tourism GDP Share vs Arrivals Over Time

```
country_focus = 'Japan' # <-- change this to any country name
country_df = df[df['country'] == country_focus]
```

```
fig = go.Figure()
fig.add_trace(go.Scatter(
    x=country_df['year'], y=country_df['tourism_gdp_share'],
    mode='lines+markers', name='Tourism GDP Share (%)',
    line=dict(color='royalblue', width=3)
))
fig.add_trace(go.Scatter(
    x=country_df['year'], y=country_df['tourism_arrivals'],
    mode='lines+markers', name='Tourism Arrivals',
    line=dict(color='green', width=3, dash='dot'),
    yaxis='y2'
))
fig.update_layout(
```

```

title=f" {country_focus}: Tourism GDP Share vs Tourism Arrivals Over Time",
xaxis=dict(title='Year'),
yaxis=dict(title='Tourism GDP Share (%)',
           titlefont=dict(color='royalblue'),
           tickfont=dict(color='royalblue')),
yaxis2=dict(title='Tourism Arrivals',
            titlefont=dict(color='green'),
            tickfont=dict(color='green'),
            overlaying='y', side='right'),
legend=dict(x=0.02, y=0.98, bgcolor='rgba(255,255,255,0.7)'),
template='plotly_white'
)
fig.show()

```

Japan: Tourism GDP Share vs Tourism Arrivals Over Time

