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: Popular Destination Analysis

Report submitted by

REGISTRATION NUMBER : 2315011076

STUDENT NAME : Y KAVYASHRI

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: Popular Destination Analysis

Report submitted by

REGISTRATION NUMBER : 2315011076

STUDENT NAME : Y KAVYASHRI

YEAR : 2023-2027

SUBJECT CODE : AD23632

(Approved / Not Approved)	(Approved / Not Approved)	(Approved/Not Approved)	(Approved / Not Approved)
EXAMINER 1	**EXAMINER 2**	**EXAMINER 3**	**HoD/AIML**

Table of Contents

Chapter 1: Abstract

Tourism is one of the fastest-growing industries worldwide, with destinations experiencing varying visitor patterns based on seasonality, geography, and cultural factors. This project, titled **"Popular Destination Analysis"**, focuses on identifying the **top tourist destinations** and examining their **seasonal trends** using data from state tourism boards and publicly available datasets such as Kaggle.

The methodology involves:

- **Data Preprocessing & Cleaning** (Python, Pandas, NumPy)
- **Exploratory Data Analysis (EDA)** to identify patterns in visitor inflows
- **Visualizations** in:
    - **Python (Matplotlib, Seaborn, Plotly)** → to build bar charts, heatmaps, and trend lines
    - **Power BI** → for interactive dashboards highlighting destination-wise traffic
    - **Tableau** → for advanced visual storytelling and seasonal trend analysis

The findings are expected to reveal:

1. **Top 10 tourist destinations by visitor count**
2. **Seasonal peaks and lows** in tourist arrivals
3. **Geographical comparison** of destinations

These insights will help tourism boards, policymakers, and travel companies in **strategic planning, targeted promotions, and infrastructure development** to enhance the tourism experience.

Chapter 2: Introduction

Tourism is a rapidly growing industry that significantly contributes to a country's economy by generating revenue, creating employment opportunities, and promoting cultural exchange. With the increasing accessibility of travel and the influence of social media, tourism patterns have evolved drastically over the years. Understanding which destinations attract the most visitors and identifying the factors that drive these trends have become essential for effective tourism management and planning.

The project **"Popular Destination Analysis"** aims to analyze tourism data to identify the most visited destinations and observe their seasonal variations. By studying patterns in visitor count, revenue, and tourist satisfaction ratings, this project seeks to uncover key insights into the popularity of different tourist locations. The findings from this analysis can help tourism boards, policymakers, and travel agencies make informed decisions about resource allocation, promotional campaigns, and infrastructure development.

For this analysis, datasets are obtained from **State Tourism Boards** and **Kaggle sources**, containing details such as visitor statistics, ratings, and revenue data for multiple destinations. Using data analytics tools such as **Python, Power BI, and Tableau**, the collected data is cleaned, processed, and visualized. Various visualization methods, including **bar charts, pie charts, line graphs, and heatmaps**, are used to represent the results in an interactive and easily understandable format.

Through this study, meaningful patterns are identified that reveal how tourism activities fluctuate across regions and seasons. The insights derived from this project can support the creation of data-driven strategies for improving the tourism experience and enhancing destination management. Ultimately, the project demonstrates how analytical tools can transform raw tourism data into valuable insights for sustainable tourism development.

Chapter 3: Dataset Description

The dataset for this project captures a wide range of variables that provide insights into global tourism trends, traveler preferences, and destination performance. It is a **structured and cross-sectional dataset**, not time series–based, making it particularly suitable for **comparative analysis across countries, categories, and tourism types**.

Key Variables Include:

● Destination Details:

- **Location** – Unique name or identifier for each tourist destination.
- **Country** – Specifies the country where the destination is located, enabling regional comparison and cross-country tourism analysis.
- **Category** – Type of attraction, such as *Nature, Historical, Adventure, Cultural, Religious,* or *Modern*, reflecting visitor interests and tourism diversity.

● Visitor and Engagement Metrics:

- **Visitors** – The total number of tourists visiting each destination, indicating popularity and footfall trends.
- **Rating** – Average visitor satisfaction rating (on a 1–5 scale), providing a measure of perceived quality and experience.

● Economic Indicators:

- **Revenue** – Total income generated from tourism activities at each location, reflecting economic impact and profitability.

● Infrastructure Availability:

- **Accommodation_Available** – Indicates whether lodging facilities (Yes/No) are available near the destination, serving as a key factor in accessibility and visitor convenience.

Chapter 4: Objective

The main objective of this project is to **analyze and visualize global tourism patterns** to identify the most popular and high-performing destinations based on multiple factors such as visitor count, revenue generation, ratings, and accessibility. By leveraging data analytics and visualization tools, the project aims to uncover meaningful insights that can support **data-driven decision-making in the tourism industry**.

Specific Objectives:

1. **Identify Popular Tourist Destinations:**
   Determine which countries and locations attract the highest number of visitors and generate the most tourism revenue.
2. **Analyze Tourism Categories:**
   Compare different categories of destinations — such as *Nature, Historical, Cultural, Adventure, Religious,* and *Modern* — to understand their relative popularity and economic contribution.
3. **Understand Visitor Satisfaction:**
   Study how visitor ratings relate to factors like accommodation availability, country, and destination type, to measure overall tourist satisfaction.
4. **Assess Economic Impact:**
   Evaluate how tourism contributes to revenue generation across destinations and examine the correlation between visitor volume and economic performance.
5. **Explore Infrastructure Influence:**
   Investigate whether the presence of accommodation facilities influences the number of visitors, ratings, and overall profitability of a location.
6. **Develop Interactive Dashboards:**
   Use **Power BI** and **Tableau** to create engaging visual dashboards that display trends, comparisons, and summaries of the tourism data.
7. **Apply Data Analytics with Python:**
   Perform **data cleaning, preprocessing, and exploratory data analysis (EDA)** using Python to ensure the dataset is accurate, consistent, and ready for visualization.

Chapter 5: Methodology

The methodology for analyzing popular tourist destinations involves **data collection, cleaning, analysis, and visualization** using Python, Power BI, and Tableau. The steps are as follows:

1. **Data Collection:**
    - Gather data from reliable sources such as travel websites, tourism boards, or open datasets on tourist visits, ratings, reviews, and demographic information.
    - Ensure the dataset contains attributes like destination name, location, number of visitors, tourist ratings, seasonality, and travel type.
2. **Data Preprocessing (Python):**
    - Handle missing values through imputation or removal.
    - Remove duplicates and irrelevant columns.
3. **Exploratory Data Analysis (Python):**
    - Analyze trends and patterns using descriptive statistics and visualizations (histograms, bar charts, and scatter plots).
    - Identify top destinations based on visitor count, ratings, or other popularity metrics.
    - Explore correlations between variables like tourist ratings, seasonality, and location.
4. **Data Visualization:**
    - **Power BI:** Create interactive dashboards showing top destinations, visitor trends over time, and regional comparisons.
    - **Tableau:** Visualize geographic distributions using maps, heatmaps, and trend charts to highlight destination popularity.
5. **Insights and Reporting:**
    - Identify patterns such as most visited destinations, peak travel seasons, and visitor preferences.
    - Provide actionable insights for tourism planning, marketing strategies, or travel recommendations.
6. **Conclusion:**
    - Summarize findings with visual dashboards and descriptive insights.
    - Highlight popular destinations and factors influencing their popularity.

Chapter 6: Python Implementation

Python is used as the primary tool for **data preprocessing and exploratory analysis** of tourist destinations. Libraries such as **pandas** and **numpy** manage data cleaning, type conversion, and handling of missing or inconsistent entries, ensuring the dataset is accurate and ready for analysis.

Exploratory analysis is performed using **visualizations**. **Bar charts** and **histograms** identify the most visited destinations and seasonal trends. **Scatter plots** explore relationships between visitor counts, ratings, and travel seasons, while **boxplots** compare popularity across regions or travel types. **Correlation heatmaps** reveal patterns and interdependencies among variables, helping to understand factors influencing destination popularity.

Feature engineering is applied to create metrics like a **popularity score** (combining ratings and visitor counts) and **peak travel periods** for each destination. These processed datasets and visualizations are exported for further **interactive analysis in Power BI and Tableau**, enabling map-based and dynamic dashboards.

Overall, Python provides a **reproducible and transparent foundation** for analysis, allowing insights into the most popular destinations, visitor behavior, and seasonal trends in a data-driven manner.

Chapter 7: Power BI Dashboard

This Power BI dashboard analyzes global tourism trends using cleaned CSV data processed in Python. It highlights **top tourist countries**, **visitor counts**, **ratings**, **revenues**, and **tourism categories** through visuals such as bar, line, pie, gauge, and card charts.

Interactive filters and slicers allow users to explore data by **country** and **category**.
Key insights: Egypt and India attract the most visitors and revenue, while cultural and adventure tourism remain the most popular types.

Overall, the dashboard provides a clear and interactive overview for understanding and improving global tourism performance.

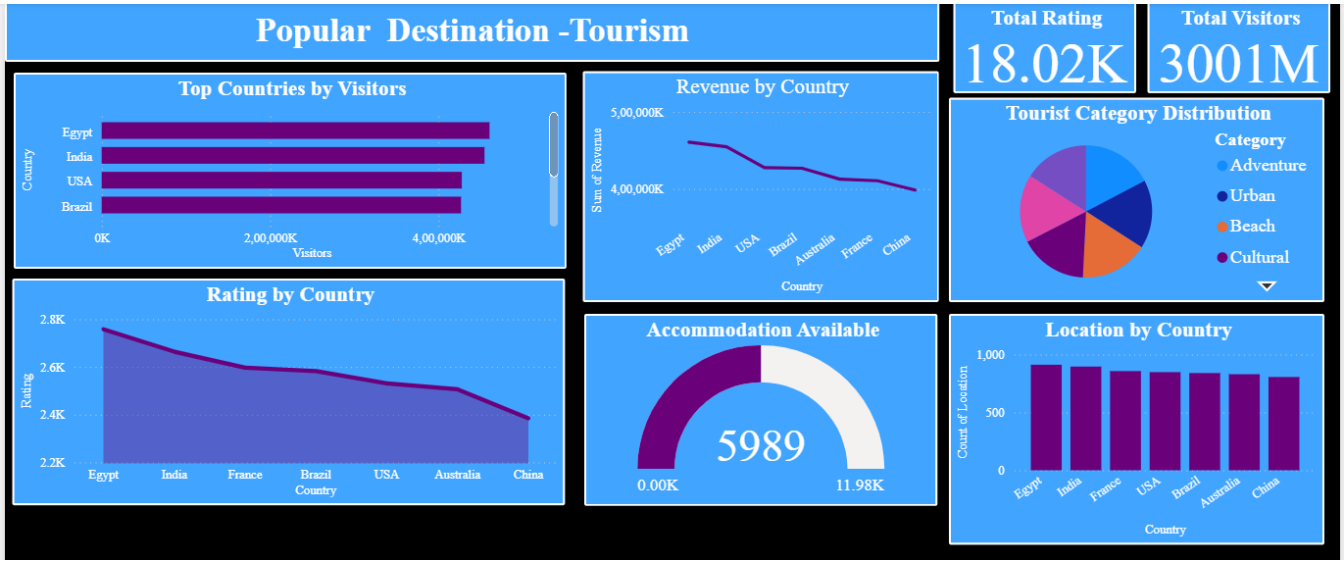


Fig 7.1: Power BI Dashboard

Chapter 8: Tableau Dashboard

Tableau complements Power BI by focusing on visually engaging dashboards suitable for presentations and storytelling. The cleaned dataset is imported, and calculated fields are created, such as the ratio of social media time to work hours and the perceived productivity gap.

This tableau dashboard focuses on analysing the usage of social media sites. Dashboards combine multiple sheets into a cohesive story.The visual storytelling aspect of Tableau makes it highly effective for communicating results to a broad audience. Tables, bar charts, and line charts have been used to understand how much media is being used in current times.

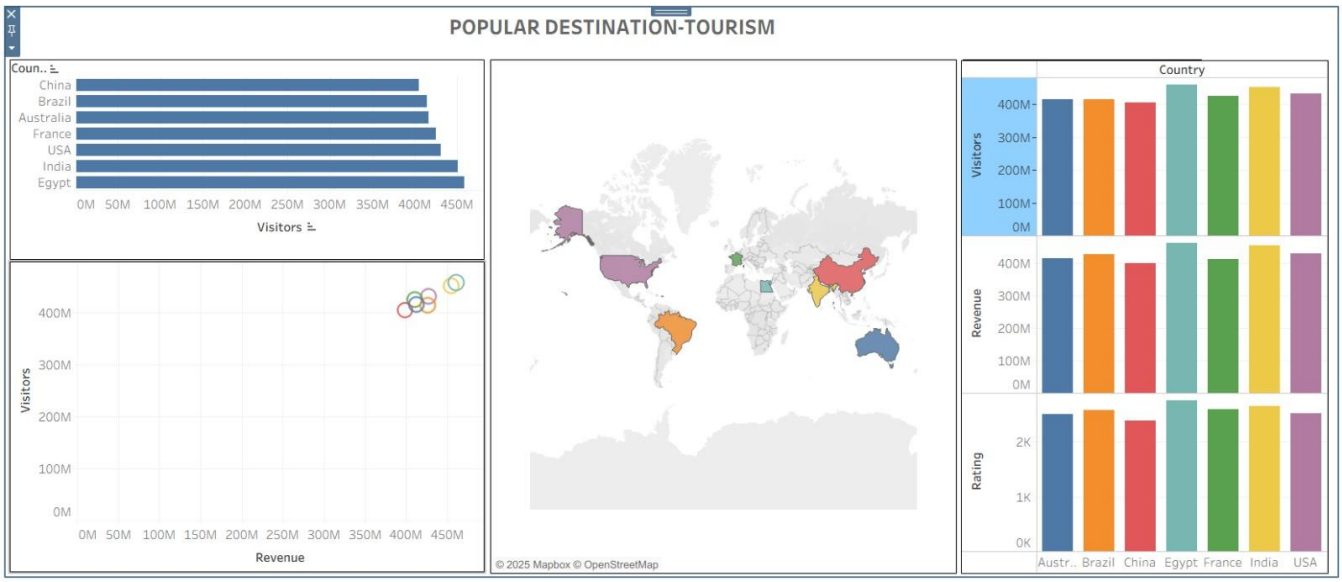


Fig 8.1: Tableau Dashboard

Chapter 9: Analysis & Finding

The project **"Popular Destination – Tourism Analysis"** aims to study global tourism trends and identify the most visited countries using data analytics and visualization tools. It focuses on understanding patterns in visitors, revenue, ratings, and accommodation availability across various nations.

Data preprocessing was performed using **Python**, where missing values were handled, duplicates were removed, and data types were standardized. The cleaned dataset containing key attributes such as *Country*, *Visitors*, *Revenue*, and *Rating* was then exported as a CSV file for further visualization.

An interactive dashboard was built in **Power BI** to present tourism insights visually. It includes bar, line, pie, and gauge charts showing top tourist countries, revenue trends, ratings, and accommodation statistics. Dynamic filters and slicers enable easy exploration by country and category, with Egypt and India emerging as top destinations.

Finally, **Tableau** was used to create a geographical visualization through maps, bar charts, and scatter plots, displaying visitor distribution, revenue, and rating comparisons globally. Together, Python, Power BI, and Tableau provide a comprehensive, data-driven analysis that supports better decision-making in the tourism industry.

Chapter 9: Conclusion

The **Popular Destination – Tourism Analysis** project effectively demonstrates how data analytics and visualization tools can provide deep insights into global tourism trends. By using **Python** for data preprocessing, the dataset was cleaned, refined, and structured to ensure accuracy and reliability. This foundation allowed for seamless integration into visualization platforms.

Through **Power BI**, interactive dashboards were developed to analyze visitors, ratings, revenue, and accommodation availability across different countries. The dashboard enabled dynamic exploration using filters and slicers, helping identify key tourism leaders like **Egypt and India**, which consistently perform well in terms of both visitors and revenue.

Meanwhile, **Tableau** offered a powerful visual representation through maps, bar charts, and scatter plots. This helped in understanding geographical tourism patterns and comparing country-wise performance more effectively.

Overall, the combination of Python, Power BI, and Tableau provided a complete end-to-end analytical workflow — from data cleaning to visualization — offering valuable insights for policymakers, travel organizations, and investors to make informed decisions and enhance global tourism growth.

Chapter 9: Future Scope

The Tourism Dashboard provides insights into visitor trends, revenue, and accommodation patterns. Its future enhancements can make it more powerful and actionable:

1. **Real-Time Data Integration:** Incorporate live data from travel portals, social media, and booking platforms for dynamic tracking of tourist behavior.
2. **Predictive Analytics:** Use machine learning to forecast visitor inflow, revenue trends, and popular destinations, aiding in better planning and resource allocation.
3. **Geospatial Visualization:** Add interactive maps and heatmaps to identify tourism hotspots, under-visited areas, and optimize itineraries.
4. **Personalized Recommendations:** Suggest destinations, accommodations, and packages based on visitor preferences and ratings.
5. **Web & Mobile Deployment:** Make the dashboard accessible online or via mobile apps for tourists, agencies, and policymakers.
6. **Enhanced Interactivity:** Introduce filters, drill-downs, and scenario simulations to explore "what-if" analyses for tourism planning.
7. **Sustainability Insights:** Include environmental impact metrics to promote eco-friendly tourism and monitor tourist density and resource usage.

Conclusion: With these enhancements, the dashboard can evolve into a comprehensive decision-support system for tourists, businesses, and policymakers, enabling data-driven tourism strategies.

Chapter 10: Appendix

10.1 Python Code

Import libraries

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

LOAD DATASET

```
df = pd.read_csv("tourism_dataset.csv")
```

Basic setup

```
sns.set(style="whitegrid", palette="muted")
df['Accommodation_Available'] = df['Accommodation_Available'].astype('category')
```

STATIC DASHBOARD (MATPLOTLIB + SEABORN)

```
fig, axes = plt.subplots(3, 3, figsize=(20, 16))
plt.suptitle("Popular Destination Dashboard", fontsize=18, fontweight='bold')
```

Bubble Chart: Visitors vs Revenue by Country

```
sns.scatterplot(
    data=df,
    x="Visitors",
    y="Revenue",
    size="Rating",
    hue="Country",
    alpha=0.7,
    sizes=(50, 800),
    palette="Set2",
    ax=axes[0, 0]
```

```
)
axes[0, 0].set_title("Visitors vs Revenue (Bubble Chart)")
axes[0, 0].set_xlabel("Visitors")
axes[0, 0].set_ylabel("Revenue")
```

Top 5 Countries by Visitors

```
country_visitors = df.groupby('Country')['Visitors'].sum().sort_values(ascending=False).head(5)
sns.barplot(
    x=country_visitors.index,
    y=country_visitors.values,
    palette='viridis',
    ax=axes[0, 1]
)
axes[0, 1].set_title("Top 5 Countries by Total Visitors")
axes[0, 1].set_ylabel("Total Visitors")
axes[0, 1].set_xticklabels(axes[0, 1].get_xticklabels(), rotation=45)
```

Accommodation Availability Pie Chart

```
df['Accommodation_Available'].value_counts().plot.pie(
    autopct='%1.1f%%',
    startangle=90,
    colors=['#66b3ff', '#99ff99'],
    ax=axes[0, 2]
)
axes[0, 2].set_title("Accommodation Availability")
axes[0, 2].set_ylabel('')

# Average Rating by Category
sns.barplot(
    x='Category',
    y='Rating',
    data=df,
```

```
    estimator=np.mean,
    palette='Set2',
    ax=axes[1, 0]
)
axes[1, 0].set_title("Average Rating by Tourism Category")
axes[1, 0].set_xlabel("Category")
axes[1, 0].set_ylabel("Average Rating")
axes[1, 0].set_xticklabels(axes[1, 0].get_xticklabels(), rotation=45)

Correlation Heatmap
sns.heatmap(
    df[["Visitors", "Revenue", "Rating"]].corr(),
    annot=True,
    cmap="YlGnBu",
    linewidths=0.5,
    ax=axes[1, 1]
)
axes[1, 1].set_title("Correlation Between Visitors, Revenue, and Ratings")

Revenue Distribution
sns.histplot(df['Revenue'], bins=25, kde=True, color='purple', ax=axes[1, 2])
axes[1, 2].set_title("Revenue Distribution")
axes[1, 2].grid(True)

Rating Distribution
sns.histplot(df['Rating'], bins=15, kde=True, color='orange', ax=axes[2, 0])
axes[2, 0].set_title("Rating Distribution")
axes[2, 0].grid(True)

Turn off empty axes
axes[2, 1].axis('off')
axes[2, 2].axis('off')
```

Adjust layout

```
plt.tight_layout(rect=[0, 0, 1, 0.97])
plt.show()
```

INTERACTIVE DASHBOARD (PLOTLY)

Sunburst Chart: Visitors by Country, Category, Accommodation

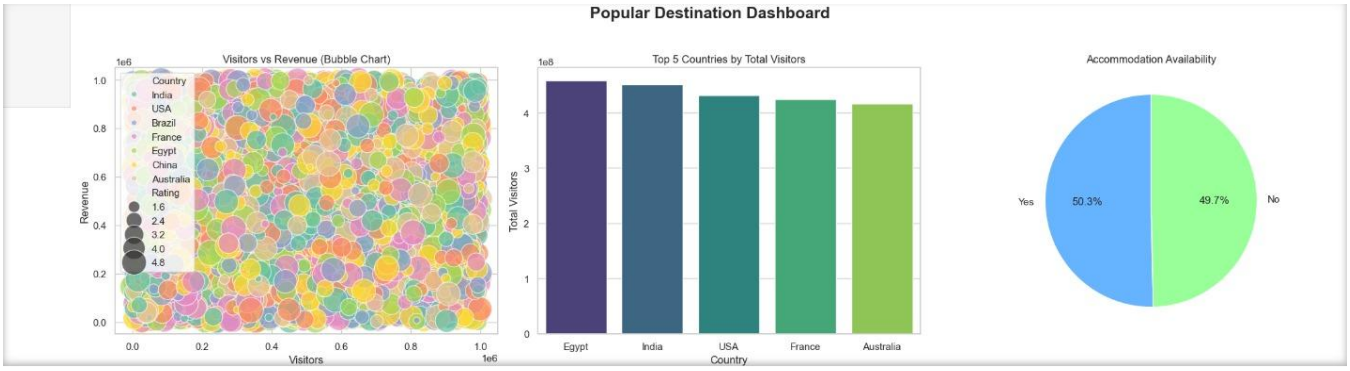
```
fig_sunburst = px.sunburst(
    df,
    path=["Country", "Category", "Accommodation_Available"],
    values="Visitors",
    color="Country",
    hover_data=["Revenue", "Rating"],
    title="Tourism Distribution by Country, Category, and Accommodation",
    color_discrete_sequence=px.colors.qualitative.Pastel
)
fig_sunburst.show()
```

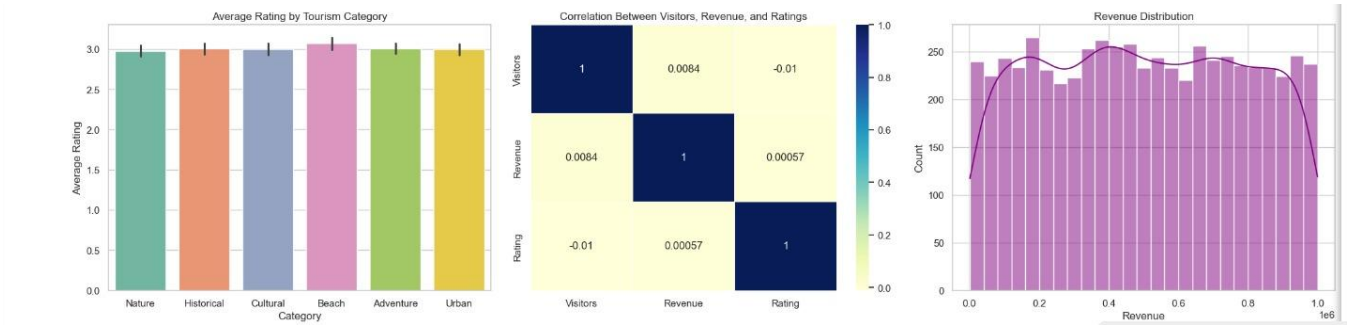
Treemap Chart: Revenue across Countries and Categories

```
fig_treemap = px.treemap(
    df,
    path=["Country", "Category"],
    values="Revenue",
    color="Rating",
    hover_data=["Visitors"],
    color_continuous_scale="Viridis",
    title="Revenue Distribution Across Countries and Categories"
)
fig_treemap.show()
```

Interactive Bubble Chart

```
fig_scatter = px.scatter(
    df,
    x="Visitors",
    y="Revenue",
    size="Rating",
    color="Country",
    hover_name="Category",
    hover_data=["Accommodation_Available"],
    size_max=60,
    title="Visitors vs Revenue (Interactive Bubble Chart)"
)
fig_scatter.show()
```





Tourism Distribution by Country, Category, and Accommodation

Revenue Distribution Across Countries and Categories