Tourism Dataset

INTRODUCTION

This tourism dataset offers a comprehensive view of tourism trends across multiple countries, showcasing data on visitor numbers, revenue, accommodation availability, and visitor ratings by tourism category. Key columns, such as Country, Category (e.g., Adventure, Cultural, Beach), Visitors, and Revenue, provide insights into travel preferences, high-demand locations, and economic impact by tourism type. Additional fields like Accommodation_Available and Rating_category enable analysis of how infrastructure and satisfaction levels vary across destinations.

This data serves as a valuable resource for understanding the factors that influence tourism success, revealing both high-performing locations and areas for improvement. By analyzing trends in visitor engagement and revenue, this dataset can help tourism boards and businesses make informed decisions to enhance visitor experiences, allocate resources effectively, and promote sustainable growth across the tourism sector.

AIM

This tourism dataset report aims to explore and interpret key data trends across various countries to provide actionable insights into the tourism industry. The report examines essential metrics—visitor counts, revenue generation, accommodation availability, and satisfaction ratings—to help understand how different types of tourism (e.g., cultural, adventure, and historical) perform globally. Through this analysis, the report seeks to highlight countries and tourism categories that attract the most visitors, generate significant revenue, or exhibit high satisfaction scores. It also pinpoints areas with growth potential or operational gaps, such as limited accommodation or low visitor satisfaction,

to recommend targeted improvements. Ultimately, the findings intend to assist tourism stakeholders—government agencies, private sector investors, and tourism boards—in making informed, data-driven decisions to enhance visitor experience, optimize resource allocation, and encourage sustainable tourism growth that benefits both tourists and local economies.

OBJECTIVES

Data Collection and Structuring:

- Load a structured dataset with fields such as Location, Country, Type, Visitors, Rating, Revenue, and Accommodation_Available using pandas for efficient handling and storage in DataFrames, making data organization and access straightforward.

Data Cleaning and Preprocessing:

- Use pandas to clean the dataset by addressing missing values, duplicates, and data inconsistencies. Prepare data types as needed (e.g., Visitors as integers, Revenue as floats) to ensure the dataset is ready for analysis.

Data Aggregation and Analysis:

- Perform data grouping and aggregation using pandas to calculate metrics like total visitors, average ratings, and revenue per tourism type or location. NumPy can be used to conduct calculations on arrays, enhancing efficiency for large datasets.

Exploratory Data Analysis (EDA):

- Conduct EDA with pandas, using functions to find patterns, outliers, and correlations. This includes calculating summary statistics, examining category distributions, and identifying popular locations or high-revenue types in the dataset.

Data Visualization:

- Use visualization libraries like Matplotlib and Seaborn to present trends in visitor counts, revenue distribution by type, and overall patterns, making complex information accessible for stakeholders.

DATA OVERVIEW

	COLUMN NAME	DATA TYPE	
	----------	----------	
	Location	object	
	Country	object	
	Category	object	
	Visitors	Integer	
	Rating	Float	
	Revenue	Float	
	Accommodation_Available	object	
	Rating_Category	object	
	High_Revenue_Location	object	

DATA ANALYSIS

1. **Descriptive Analysis**: The purpose of this analysis is to summarize tourism activity across different locations. Focus areas include visitor counts, average ratings, revenue leaders, and the availability of accommodations to highlight popular destinations and top-rated sites.
2. **Trend Analysis**: This analysis explores patterns in tourism metrics over time. The focus is on identifying seasonal trends in visitor counts and revenue to

understand fluctuations throughout the year.

3. **Visitor Demographics**: This analysis segments visitors by demographics, focusing on distribution across categories like country, category, and rating. The goal is to understand visitor preferences and demographics.
4. **Revenue Analysis**: The purpose of revenue analysis is to examine revenue distribution across locations and categories. This analysis focuses on identifying top-grossing destinations and categories for strategic insights.
5. **Accommodation Analysis**: This analysis assesses the impact of accommodation availability on tourism. By exploring the correlation between visitor numbers and accommodations, it reveals the role of accommodations in attracting tourists.
6. **Rating Analysis**: This analysis evaluates visitor satisfaction based on ratings across various destinations. By examining ratings, it provides insights into visitor satisfaction and experience quality.
7. **High-Revenue Identification**: The goal of this analysis is to identify high-revenue locations. This focus on "High Revenue" destinations helps pinpoint areas with potential for further development or investment.

INSIGHTS

Tourism Category Preferences by Country :

df . groupby ("Country")["Category"]. value_counts()

Egypt and **India**: Strong in Adventure, Beach, and Historical tourism.

France: Balanced across Cultural, Beach, and Nature tourism.

Australia: Notably high in Adventure and Nature tourism.

USA: Strong in Urban and Nature tourism.

These patterns help guide tourism strategies by showing where to invest resources and how to boost visitor interest. Each country can focus on promoting its top attractions to improve visitor experiences.

Accommodation Availability Insights Across Tourist Destinations :

df.groupby(['Country'])['Accommodation_Available']. value_counts()

High Accommodation Availability: Egypt and India have the most "Yes" responses, Ensuring a well-established accommodation network for visitors.

Mixed Availability: France, the USA, and Brazil display a balanced mix of available and unavailable accommodations, offering diverse tourism experiences.

Lower Availability: China and Australia show fewer accommodations, suggesting potential areas for infrastructure improvement in tourism.

Tourism Popularity by Category Based on Visitor Count :

df.groupby('Category')['Visitors'] .sum()

Top Categories: Adventure and Urban tourism have the highest visitor counts, signaling strong interest in these categories.

Broad Appeal: Beach, Cultural, and Historical categories also draw large numbers, showing their consistent popularity.

Potential Growth: Nature tourism, though slightly lower in visitor numbers, still has a substantial audience and potential for growth.

Tourism Trends in India :

df[df['Country'] == 'India']. value_counts().idxmax()

Popular Category: The most frequently visited category in India is **Cultural tourism**, with significant visitor numbers.

Visitor Engagement: Cultural tourism indicates a strong interest in India’s rich heritage and traditions.

Revenue Generation: The **revenue of ₹166,108.53** reflects the economic potential of promoting Cultural tourism further.

Accommodation Availability: The presence of accommodations indicates an ability to support an increase in tourists, enhancing the visitor experience.

Rating Insights: A rating of 1.74 indicates strong visitor engagement but highlights potential areas to improve visitor satisfaction.

Visitor Trends in Tourism Categories :

df.groupby(['Country', "Category"])['Visitors'].sum()

Top Adventure Destinations: Brazil leads in Adventure tourism with over **83 million visitors**, followed by Australia and India.

Balanced Interests: Countries like Egypt and France attract many visitors for Beach tourism, indicating diverse tourism preferences.

Cultural Draw: France and India show strong visitor numbers in Cultural tourism, reflecting interest in heritage experiences.

Urban Attraction: The USA has high numbers in Urban tourism, highlighting its appeal as a metropolitan destination.

Top-Rated Tourism Highlights :

df[df.Rating==df.Rating.max()]

All listed locations feature a perfect rating of 5.0, showcasing their high-quality tourism offerings and visitor satisfaction.

China and **Australia** have multiple entries with a 5.0 rating, indicating high satisfaction among visitors. Specifically, China has entries in Urban, Nature, Historical, and Cultural categories, while Australia has entries in Cultural and Adventure categories. This suggests that both countries are performing well in providing exceptional tourist experiences across various categories.

Visitor Satisfaction Insights :

df.groupby(['Country', 'Rating_category']).size()

“Excellent" Rating: France has the highest count, indicating strong visitor satisfaction in its tourism sector.

"Low" Rating: India shows a higher count in this category, suggesting areas for improvement in specific locations.

"Average" Rating: The USA has the most "Average" ratings, possibly showing a balance in tourist satisfaction but with potential to enhance visitor experiences.

High Revenue Locations by Country :

df[df['High_Revenue_Location'] == 'Yes'].groupby('Country').size()

Leading Country: India has the highest number of high-revenue locations, with a total of 467, indicating strong economic potential and tourism appeal.

Competitive Market: Egypt follows closely with 452 high-revenue locations, highlighting its status as a significant player in the tourism sector.

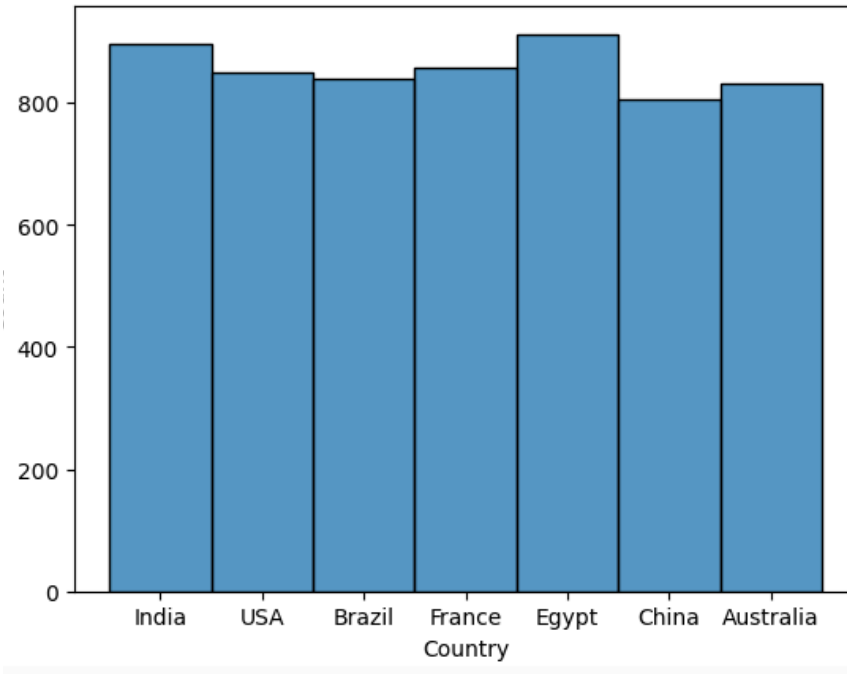
Balanced Tourism: The USA (425) and Brazil (421) also demonstrate a substantial presence of high-revenue locations, suggesting robust tourism infrastructure and diverse attractions.

Opportunities for Growth: Countries like Australia (417) and France (396) show healthy numbers, but there may be potential for increasing high-revenue locations through targeted investments in tourism initiatives.

Room for Improvement: China, with 385 high-revenue locations, presents opportunities to enhance its tourism offerings to attract more visitors.

Distribution of Visitors by Country :

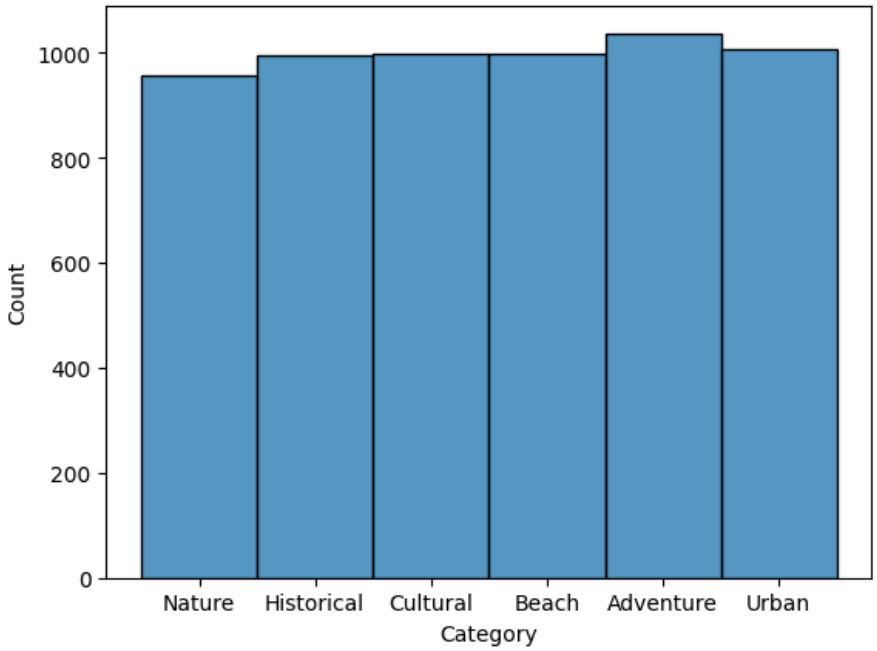
sns.histplot(data=df,x='Country')



The histogram provides a visual representation of the frequency of visitors from each country, indicating which countries contribute the most to tourism in the dataset.

Visitor Counts Across Categories :

sns.histplot(data = df, x = 'Category')

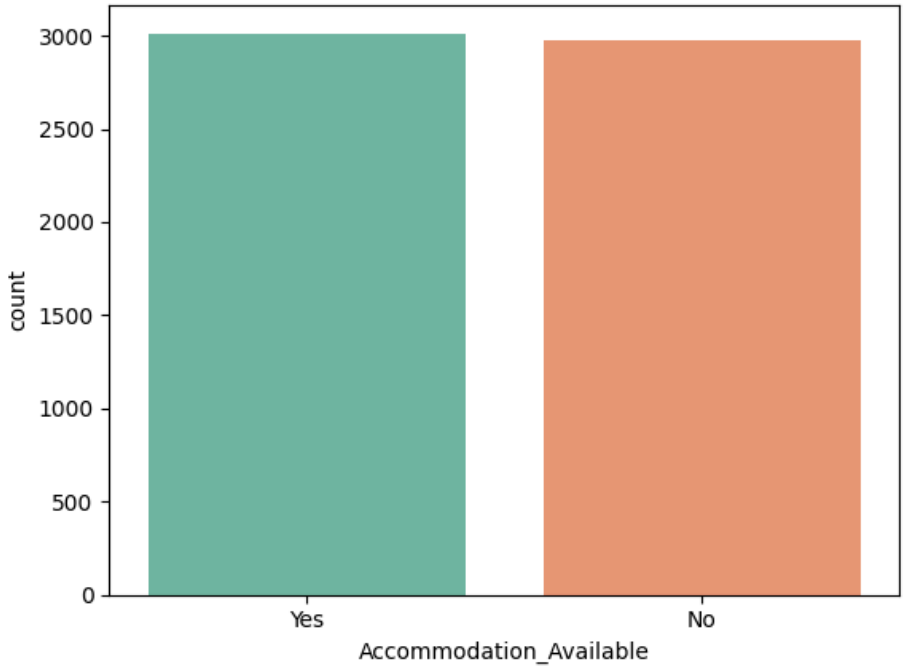


The histogram provides a visual representation of the frequency of occurrences across each category, highlighting Adventure categories have the highest visitor

counts and contribute most significantly to the dataset.

Availability of Accommodation for Visitors :

sns.countplot(x=df["Accommodation_Available"],palette="Set2")



This count plot shows the distribution of cases with and without available accommodations, making it easy to see the proportion of each category.

Country-wise Distribution of Accommodation Availability :

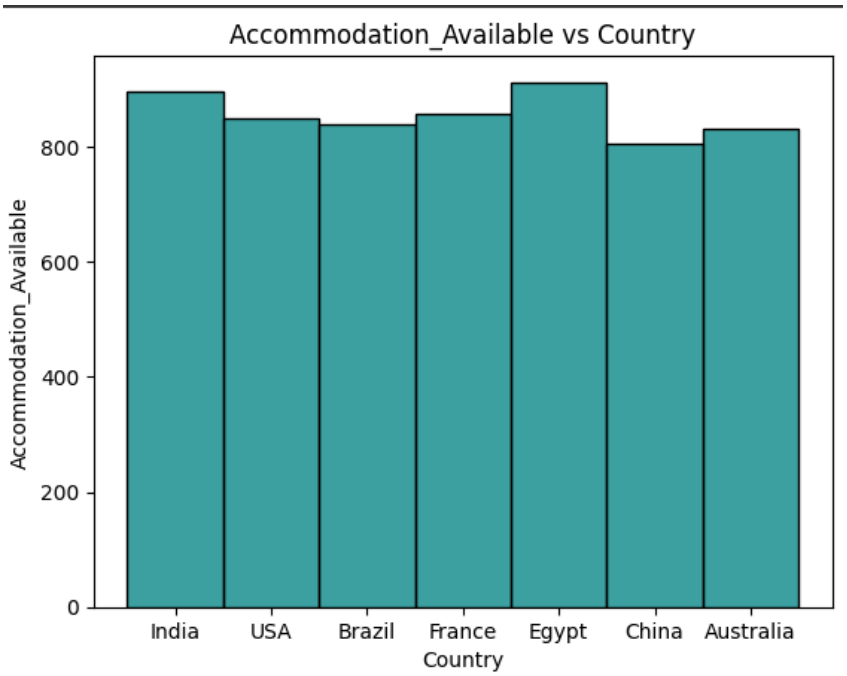
sns.histplot(df['Country'], color='teal')

plt.title('Accommodation_Available vs Country')

plt.xlabel('Country')

plt.ylabel('Accommodation_Available')

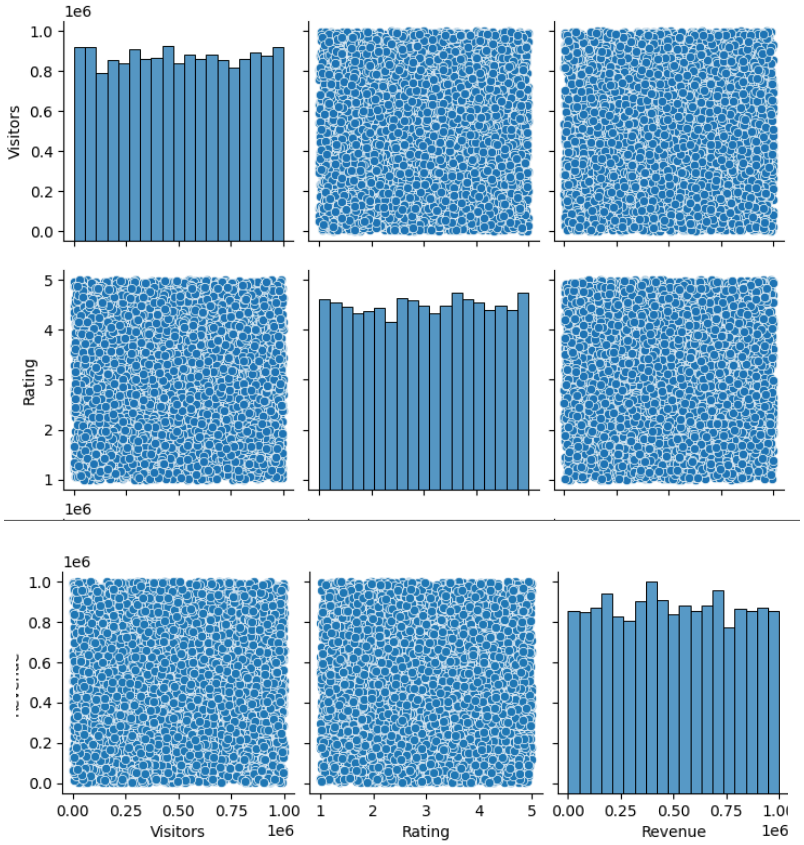
plt.show()



This histogram shows the frequency of records per country, highlighting India and Egypt have the most data entries for accommodation availability.

Pairwise Relationships Across Dataset Features :

sns.pairplot(df)

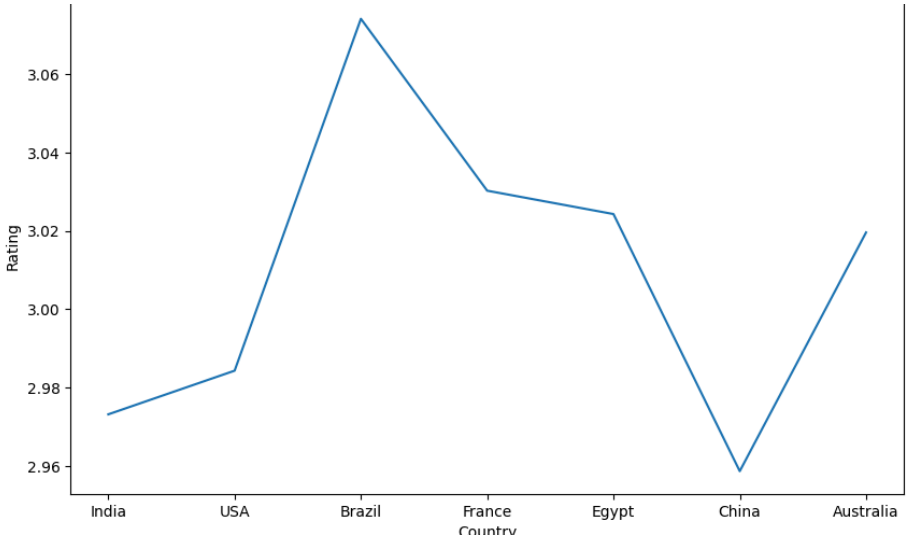


This pair plot visualizes relationships between all numerical features in the dataset, helping to identify correlations, patterns, and outliers across pairs of variables. Diagonal plots show individual feature distributions, while scatter plots reveal relationships between feature pairs

Average Rating by Country :

plt.figure(figsize = (10, 6))

sns.lineplot(data = df, x = 'Country', y = 'Rating', ci = None)

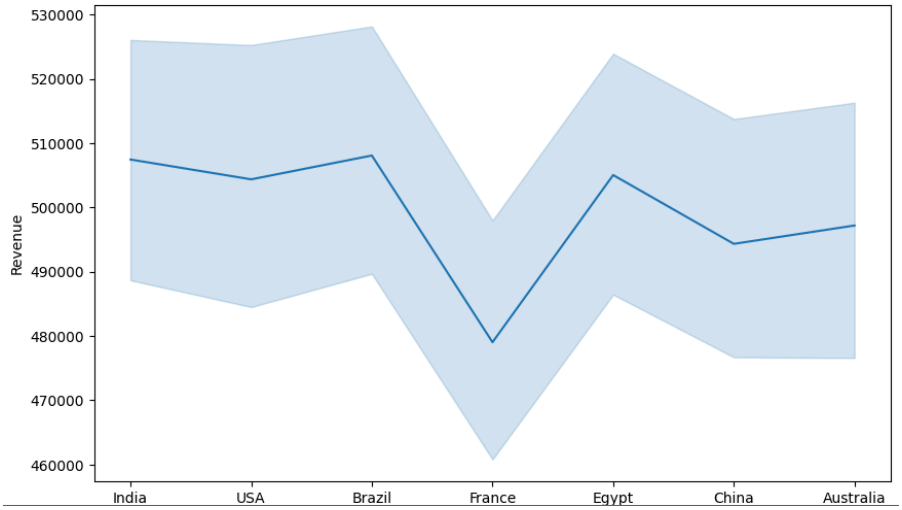


This line plot displays the average rating for each country, highlighting country-specific trends in ratings

Revenue Trends by Country :

plt.figure(figsize = (10, 6))

sns.lineplot(data = df, x = 'Country', y = 'Revenue')



This line plot illustrates the revenue generated from each country, showcasing variations in revenue levels

Distribution of Ratings :

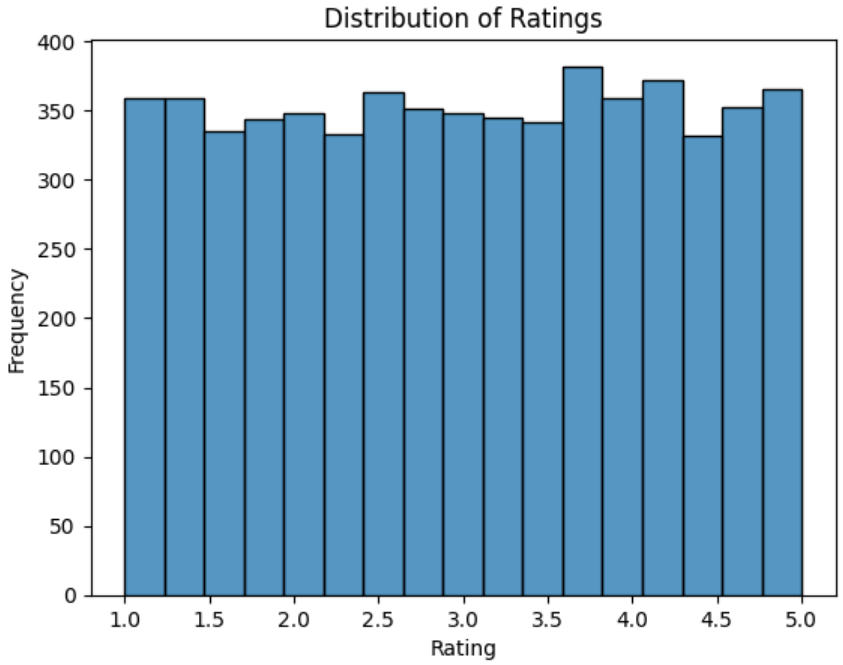
sns.histplot(df['Rating'], bins=17)

plt.title('Distribution of Ratings')

plt.xlabel('Rating')

plt.ylabel('Frequency')

plt.show()



It helps identify trends, such as the most common rating values and the overall rating distribution pattern

Revenue Trends by Category :

plt.figure(figsize=(7,4))

sns.lineplot(x='Category', y='Revenue', data=df, marker='o',color="red")

plt.title('Revenue by Category')

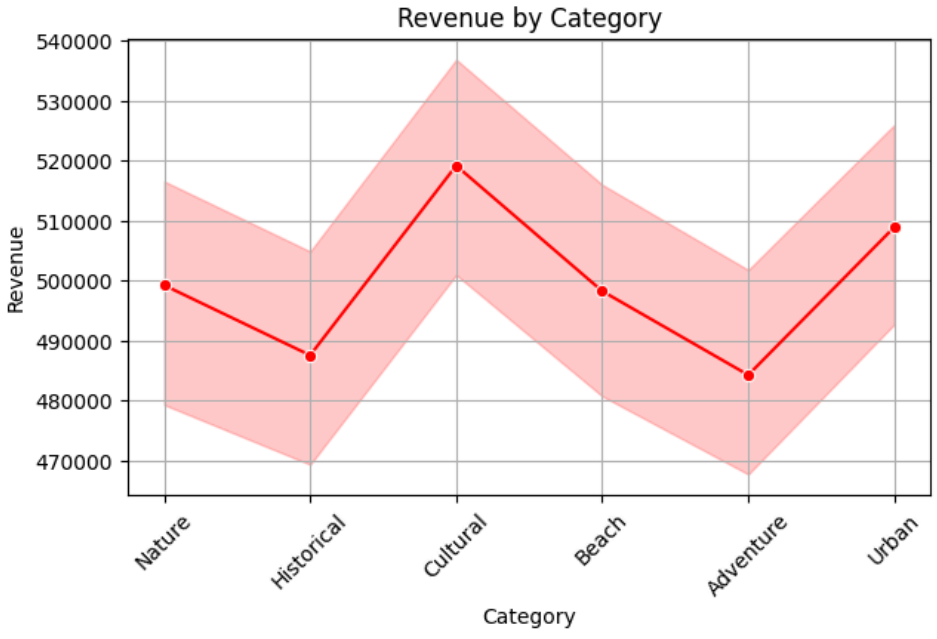
plt.xlabel('Category')

plt.ylabel('Revenue')

plt.xticks(rotation=45)

plt.grid(True)

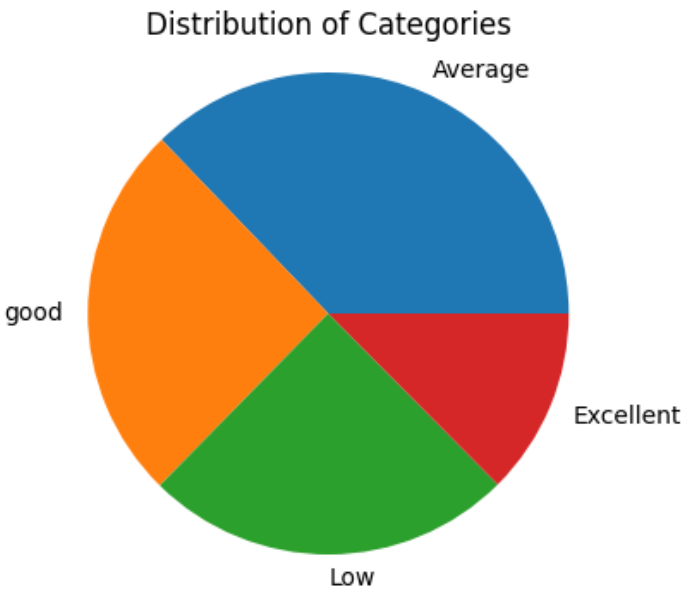
plt.show()



This line plot illustrates the revenue generated by each category, with data points marked for clarity. The red line highlights trends in revenue across categories, making it easy to identify which categories contribute the most to overall revenue

Rating Category Distribution :

```
counts = df['Rating_category'].value_counts()

plt.figure(figsize=(7, 4))

plt.pie(counts, labels=counts.index)

plt.title('Rating Category Distribution')

plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.

plt.show()
```



It helps identify which categories are most prevalent, enabling better understanding and analysis of visitor satisfaction levels within the dataset

CONCLUSION

The tourism dataset analysis offers important insights into what attracts visitors and the economic opportunities available in different countries. It shows that cultural and adventure tourism are particularly popular, with high visitor numbers in countries like India and Egypt, which also feature many high-revenue locations. The availability of accommodations is sufficient to handle

the influx of tourists, but some areas show potential for improvement in visitor satisfaction.

This analysis highlights the need for a well-rounded approach to tourism development. It suggests promoting a variety of tourism experiences while also focusing on improving areas that are underperforming. By strategically investing in the most promising tourism categories, countries can achieve sustainable growth in their tourism sectors. Overall, addressing visitor needs and enhancing their experiences can lead to increased engagement and better economic outcomes for the tourism industry.