**Exploratory Data Analysis Using Python**

Excited to continue my project series! Today, I performed an in-depth exploratory data analysis (EDA) and data visualization on a tourism dataset sourced from Kaggle.

**Project Objectives:**

1. Analyze trends in visitor count and tourism revenue generated on a country and category basis.
2. Examine whether visitor count impacts tourism revenue by country and tourism category.
3. Assess the influence of tourism category ratings on visitor count.

**Dataset Overview:** This dataset includes information on the country, tourism category, revenue generated, rating, and visitor count.

**Data Import & Cleaning:**

* Imported essential libraries such as Pandas and Matplotlib.
* Loaded the CSV file and removed irrelevant columns, such as "location."
* Checked for duplicates and adjusted column data types as needed.
* Created pivot tables to calculate revenue and visitor counts by country, and average ratings and visitor counts by category.

**Data Visualisation**

* **Bar Plot (Country vs. Revenue):**

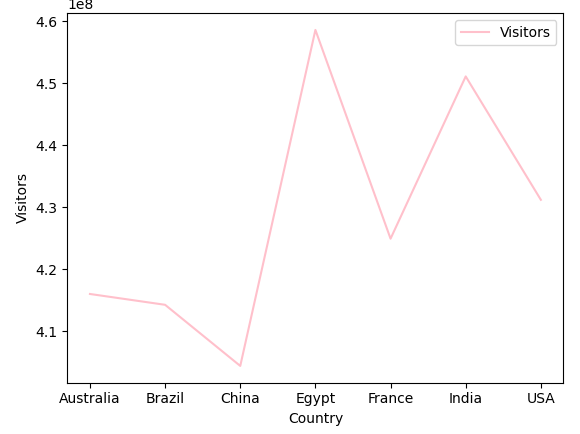
Egypt leads in tourism revenue, followed by India, while China records the lowest.

A graph of green bars with names

Description automatically generated

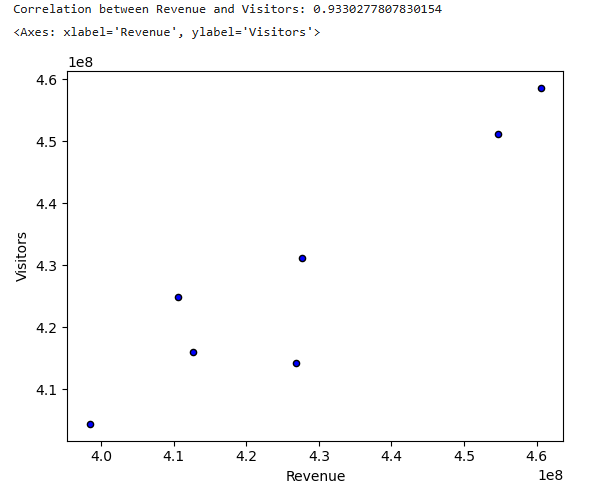
* **Line Plot (Country vs. Visitor Count):**

Consistent with the bar plot, Egypt has the highest visitor count, with China trailing.



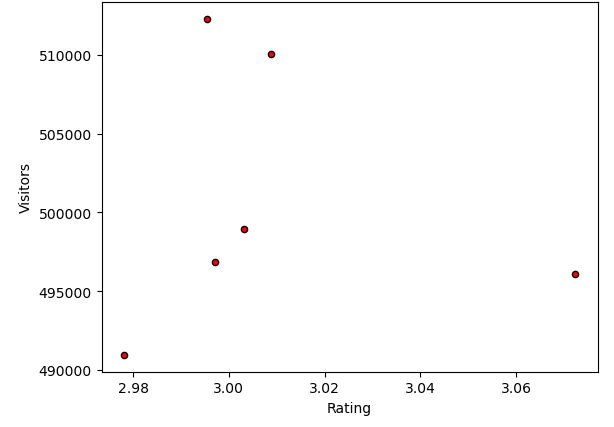
* **Scatter Plot (Revenue vs. Visitor Count):**

A high positive correlation (0.9330) indicates that an increase in visitor count is strongly associated with higher tourism revenue.



* **Scatter Plot (Category Rating vs. Visitor Count):**

Initially, a slight negative correlation (-0.0670) was observed, likely influenced by outliers. After removing outliers based on the Interquartile Range (IQR), the recalculated correlation (0.1308) suggests a positive relationship, indicating that higher-rated tourism categories attract more visitors.



A graph with red dots

Description automatically generated