**Micro Project Report**

On

**Tourism Under COVID Impact**

Submitted in Partial Fulfilment of

Award of

**BACHELOR OF TECHNOLOGY**

**In**

**Computer Science and Engineering**

**(Data Analytics)**

By

2023BCSE07AED059 – Harika.G

2023BCSE07AED054- Wayiz.K

2023BCSE07AED059 – Konda Reddy. Y

Under

The Supervision

of

**Dr. K Sasi Kala Rani**

**Professor, Department**

**OF CSE**



**ALLIANCE SCHOOL OF ADVANCED COMPUTING**

**ALLIANCE UNIVERSITY**

**BENGALURU**

APRIL 2025



**Department of computer science and Engineering**

**ALLIANCE SCHOOL OF ADVANCED COMPUTING**

**CERTIFICATE**

This is to certify that the Micro Project work entitled “Tourism Under Covid Impact” Harika.G[2023BCSE07AED059],Wayiz.K[2023BCSE07AED054],Konda Reddy.Y[2023BCSE07AED044] in partial fulfillment for the award of the degree of Bachelor of Technology DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (DATA ANALYTICS) of Alliance University, is a Bonafide work accomplished under my supervision and guidance during the academic year 2024-2025. This thesis report embodies the results of original work and studies conducted by students and the contents do not form the basis for the award of any other degree to the candidate or anybody else.

**Dr.K.Sasi Kala Rani**

**(Supervisor)**

**Table of Contents**

|  |  |  |
| --- | --- | --- |
|  | **Title** | **Page Number** |
| **1.** | Project Charter | 4 |
| 2. | Defining the Research Goal | 5 |
| 3. | Data Selection | 6 |
| 4. | Data Preparation | 7 |
| 5. | Exploratory Data Analysis | 8 |
| 6. | Visualization and Statistical Analysis | 10 |
| 7. | Pridiction | 15 |
| 8. | Evaluation using R² | 17 |
| 9. | GUI Graphical User Interface | 18 |
| 10. | Conclution | 20 |

1. **Project Charter:**

Purpose:

The purpose of this microproject is to analyze the impact of the COVID-19 pandemic on the tourism sector by examining visitor data from various tourist spots in India between the years 2019 and 2021. The project aims to understand how travel restrictions, lockdowns, and health concerns affected domestic and international tourism. By comparing visitor { Foreign Visitors and Domestic Visitors } numbers before, during, and after the peak of the pandemic, this study seeks to highlight the severity of the decline in tourism, identify emerging trends, and assess the gradual recovery of the sector. This analysis can help in formulating strategies for future crisis management and sustainable tourism development

**Introduction**:

* Tourism is a major global industry that contributes significantly to economies and cultural exchange.
* The COVID-19 pandemic caused severe disruptions with travel bans, lockdowns, and border closures.
* Flights were canceled, tourist spots closed, and international travel dropped sharply.
* The industry suffered massive economic losses, with many jobs and businesses impacted.
* Health concerns led to new safety protocols and a rise in virtual travel experiences.
* The pandemic reshaped tourism trends, emphasizing local travel and digital alternatives.

Tourism, a vital contributor to the global economy and a significant driver of cultural exchange, was one of the most severely affected sectors during the COVID-19 pandemic. With the rapid spread of the virus in early 2020, countries around the world implemented stringent measures such as lockdowns, travel restrictions, border closures, and quarantine protocols. While these steps were necessary to contain the virus, they brought international and domestic travel to an abrupt halt, leading to a profound impact on the tourism industry.

The economic repercussions rippled across various sectors, affecting not only large-scale businesses but also small and medium enterprises and local communities that heavily depended on tourism. Many workers in this industry faced unemployment or significant loss of income. Airlines grounded fleets, hotels reported historically low occupancy rates, and travel agencies were forced to shut down.Socially and culturally, the pandemic disrupted the way people travel and interact. Health safety became a primary concern, and terms like

“social distancing,” “contactless services,” and “quarantine tourism” entered the tourism lexicon.

**2.Defining the Research Goal** :

To interpret and examine the impact of COVID-19 on the tourism industry using publicly available data, travel statistics, and regional performance metrics, aiming to uncover trends in travel restrictions, economic losses, behavioral changes, and recovery patterns.

**Goals:**

* Gather and clean relevant tourism-related data (e.g., travel frequency, revenue loss, hotel occupancy rates, COVID case timelines, government regulations).
* Conduct exploratory data analysis (EDA) using statistical methods and key tourism indicators.
* Create new variables (e.g., % drop in tourist arrivals, revenue per traveler, travel recovery index).
* Visualize findings through line graphs, bar charts, heatmaps, and choropleth maps.
* Compare tourism performance across countries and regions before, during, and after the pandemic.
* Identify long-term changes in tourist behavior and preferences.
* Develop data-driven suggestions for tourism recovery, policy-making, and future crisis preparedness.

**Deliverables:**

* Cleaned and structured tourism impact dataset.
* Python/R scripts for data cleaning, analysis, and visualization
* Comparative visual dashboards for cross-region analysis
* Final report with key trends, insights, and strategic recovery recommendations



**3.Dataset Selection:**

For this micro project, we have selected a dataset sourced from “Kaggle” that contains relevant data on visitor numbers to various tourist spots throughout in India during the period between 2019 and 2021. The dataset includes key metrics such as the number of visitors to popular tourist attractions. It helps analyze the impact of COVID-19 on tourism through metrics like visitor numbers and percentage growth/decline in both domestic and foreign tourism.

**Data:**

CSV dataset (All\_India\_Tourism\_Data\_2019\_2021.csv)

**Code Implementation** :

import pandas as pd

import matplotlib.pyplot as plt

df = pd.read\_csv('/content/archive (17).zip')

df

**Over View:**



**4.Data Preparation:**

Treated missing values with imputation methods and eliminated inconsistent or irrelevant records.

* **Missing Values**: Removed rows with missing data for foreign or total visitors
* **Duplicates**: Checked and removed any duplicate region entries
* **Outliers**: Excluded top 1% extreme values in foreign visitor counts to improve model accuracy

**Code Implementation :**

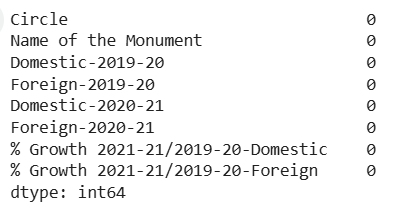
import pandas as pd

df = pd.read\_csv('/content/archive (17).zip')

df = df.dropna()

df = df.drop\_duplicates()

print(df.isnull().sum())



1.It also shows percentage growth for both categories.

2.All values are currently zero, indicating missing or placeholder data.

3.Dataset includes domestic and foreign visitor counts for 2019–20 and 2020–21.

**5.Exploratory Data Analysis (EDA):**

In the initial phase of data analysis, we aimed to understand the structure and key statistics of the dataset, particularly the tourism-related variables for the years 2019-2021. The dataset contains the following key columns:

* **Domestic-2019-20**
* **Foreign-2019-20**
* **Domestic-2020-21**
* **Foreign-2020-21**
* **% Growth 2021-21/2019-20-Domestic**
* **% Growth 2021-21/2019-20-Foreign**

To better understand the overall trends and central tendencies within these variables, we calculated the mean values for each of these columns. The mean values provide insight into the average number of visitors (both domestic and foreign) to the tourist spots during the pandemic years (2019-2021), as well as the average percentage growth or decline in tourism.

These calculated mean values helped us:

1. **Understand Trends**: By calculating the means, we could understand the general trend of domestic and foreign tourism across the years and how tourism behavior shifted between 2019-2020 and 2020-2021.
2. **Compare Pre- and Post-COVID Tourism Trends**: The means allowed us to quantify the impact of the pandemic on both domestic and international tourism, providing a basis for comparison before and after COVID-19.
3. **Prepare for Further Analysis**: With the central tendency of the data established, the dataset was then ready for further analysis, such as identifying trends, growth patterns, and assessing the recovery of tourism during 2021..

**Code Implementation** :

import pandas as pd

df = pd.read\_csv("/content/archive (17).zip")

df.columns = df.columns.str.strip().str.replace(" ", "\_")

df = df[~df['Name\_of\_the\_Monument'].isin(['Total', 'Grand Total'])]

cols = ['Domestic-2019-20', 'Foreign-2019-20', 'Domestic-2020-21', 'Foreign-2020-21']

df[cols] = df[cols].apply(pd.to\_numeric, errors='coerce')

avg\_dom\_2019 = df['Domestic-2019-20'].mean()

avg\_for\_2019 = df['Foreign-2019-20'].mean()

avg\_dom\_2020 = df['Domestic-2020-21'].mean()

avg\_for\_2020 = df['Foreign-2020-21'].mean()

print("Average Visitors in 2019-20:")

print(f"Domestic: {avg\_dom\_2019:,.0f}")

print(f"Foreign : {avg\_for\_2019:,.0f}")

print("\nAverage Visitors in 2020-21:")

print(f"Domestic: {avg\_dom\_2020:,.0f}")

print(f"Foreign : {avg\_for\_2020:,.0f}")

**Output** :

Average Visitors in 2019-20:

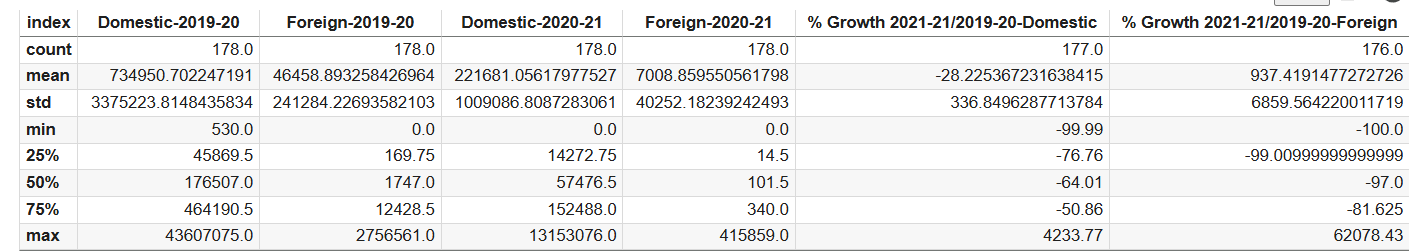
Domestic: 302,827

Foreign : 19,143

Average Visitors in 2020-21:

Domestic: 91,341

Foreign : 2,888



>Domestic tourist visits dropped by **~70%** from 2019–20 to 2020–21.

>Foreign tourist numbers declined sharply with some sites seeing **100% drop**

.

>Average domestic growth rate was **-28.2%**, showing widespread decline.

>High standard deviation indicates large variation across monuments

**6. Visualization** **and Statistical Analysis**:

To gain deeper insights into the impact of COVID-19 on tourism, we performed a comprehensive **Visualization and Statistical Analysis** of the dataset. Visualizations such as Heat Map, line graphs, and histograms were used to illustrate trends and comparisons across the years and Understanding Trends and Patterns, particularly focusing on the variations in **domestic** and **foreign** visitor numbers from 2019-2021.

**To Find Out Top 5 Most and Least Visited Cities:**

In this analysis, we identified the **Top 5 Most Visited** and **Top 5 Least Visited Cities** based on domestic and foreign visitor numbers from 2019-2021. Using **bar graphs**, we visually compared visitor totals for each city, highlighting tourism trends and how cities were impacted by the pandemic. The results reveal which cities saw the most significant declines and which ones showed recovery.

**Code Implementation** :

import matplotlib.pyplot as plt

import seaborn as sns

import pandas as pd

df['Total\_Visitors'] = df['Domestic-2019-20'] + df['Foreign-2019-20'] + df['Domestic-2020-21'] + df['Foreign-2020-21']

top5 = df.nlargest(5, 'Total\_Visitors')[['Name\_of\_the\_Monument', 'Total\_Visitors']]

bottom5 = df[df['Total\_Visitors'] > 0].nsmallest(5, 'Total\_Visitors')[['Name\_of\_the\_Monument', 'Total\_Visitors']]

fig, axes = plt.subplots(1, 2, figsize=(14, 6))

plt.subplots\_adjust(wspace=0.3)

sns.barplot(x='Total\_Visitors', y='Name\_of\_the\_Monument', data=top5, ax=axes[0], palette='viridis')

axes[0].set\_title("Top 5 Most Visited Monuments")

axes[0].set\_xlabel("Total Visitors (2019–21)")

axes[0].set\_ylabel("")

sns.barplot(x='Total\_Visitors', y='Name\_of\_the\_Monument', data=bottom5, ax=axes[1], palette='magma')

axes[1].set\_title("Top 5 Least Visited Monuments")

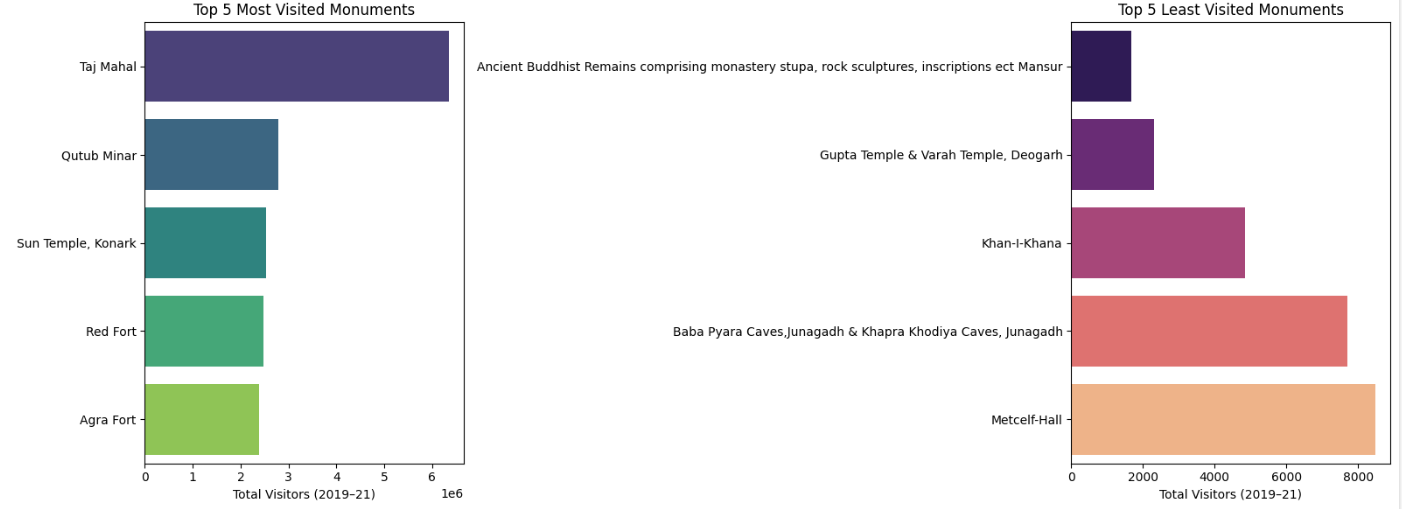
axes[1].set\_xlabel("Total Visitors (2019–21)")

axes[1].set\_ylabel("")

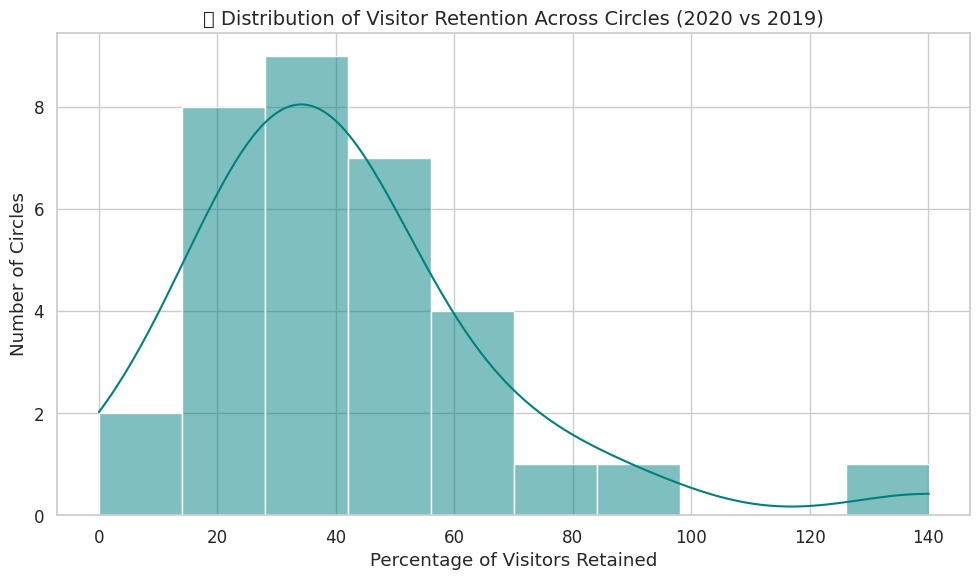
plt.tight\_layout()

plt.show()

**Output :**



**Histogram:**



**Visitor Retention Insights :**

>Histogram shows most regions retained **20–50%** of visitors.

>Few circles retained **over 100%**, showing rare resilience.

>Majority of regions faced **significant drop** in tourism.

>Distribution is **right-skewed**, with few high-performing outlier

**Regional Visitor Drop:**

In this analysis, we aimed to explore the **regional visitor drop** by comparing the number of **domestic** and **foreign** visitors to tourist spots from **2019** to **2021**. The goal was to highlight how the pandemic disproportionately affected both domestic and international tourism. By calculating the percentage drop in visitors for each region, we could assess which group (domestic or foreign) experienced a more significant decline.

**Code Implementation** :

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

circle\_visitors = df.groupby('Circle').agg({

'Domestic-2019-20': 'sum',

'Foreign-2019-20': 'sum',

'Domestic-2020-21': 'sum',

'Foreign-2020-21': 'sum'

})

circle\_visitors['Total\_2019'] = circle\_visitors['Domestic-2019-20'] + circle\_visitors['Foreign-2019-20']

circle\_visitors['Total\_2020'] = circle\_visitors['Domestic-2020-21'] + circle\_visitors['Foreign-2020-21']

circle\_visitors['% Drop'] = ((circle\_visitors['Total\_2019'] - circle\_visitors['Total\_2020']) / circle\_visitors['Total\_2019']) \* 100

heat\_data = circle\_visitors[['% Drop']].sort\_values('% Drop', ascending=False)

plt.figure(figsize=(10, 8))

sns.heatmap(heat\_data, annot=True, cmap='coolwarm', fmt='.1f', linewidths=0.5, cbar\_kws={'label': '% Drop'})

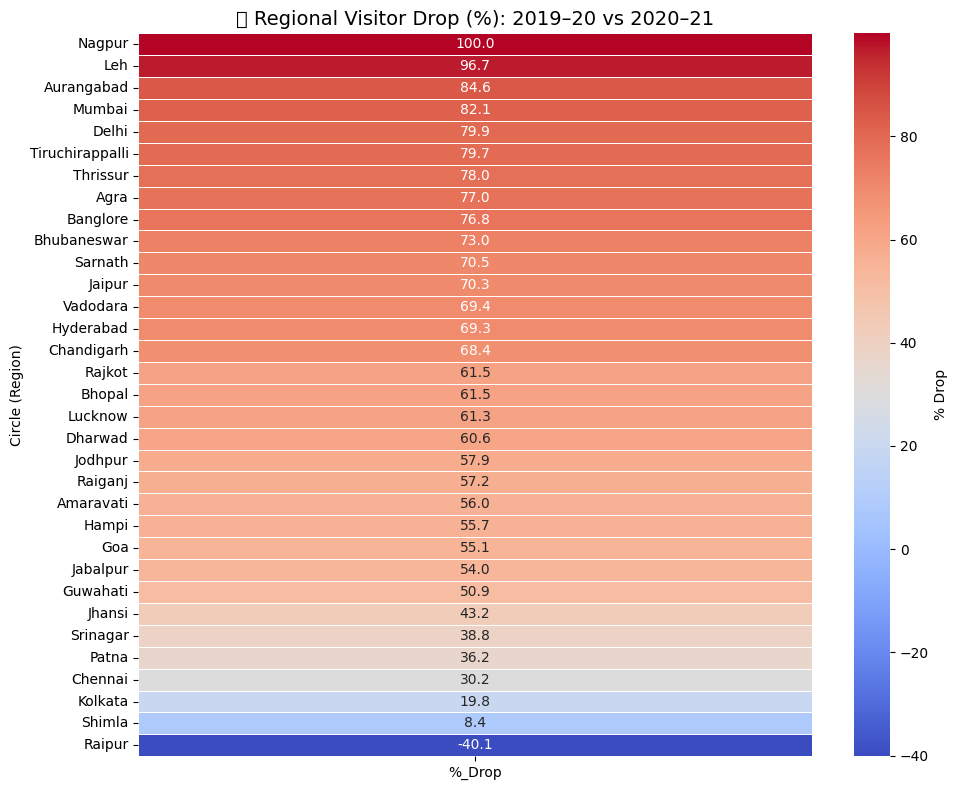
plt.title(" Regional Visitor Drop (%): 2019-20 vs 2020-21", fontsize=14)

plt.ylabel("Circle (Region)")

plt.xlabel("")

plt.tight\_layout()

plt.show()



**Correlation between Foreign Visitors and Overall Drop between 2019 to 2021:**

This analysis investigates the relationship between the **drop in foreign visitors** and the **overall visitor drop** during the years 2019-2021.

**Code Implementation :**

import matplotlib.pyplot as plt

import seaborn as sns

plt.figure(figsize=(8, 6))

sns.regplot(

x='Foreign\_Dependency\_Rate',

y='Total\_Drop',

data=df,

scatter\_kws={'alpha': 0.7, 's': 60, 'color': 'royalblue'},

line\_kws={'color': 'red'}

)

plt.title("Foreign Dependency vs. Total Visitor Drop (2019-21)", fontsize=14)

plt.xlabel("Foreign Dependency Rate (2019-20)")

plt.ylabel("Total Visitor Drop (COVID Impact)")

plt.grid(True)

corr = df[['Foreign\_Dependency\_Rate', 'Total\_Drop']].corr().iloc[0, 1]

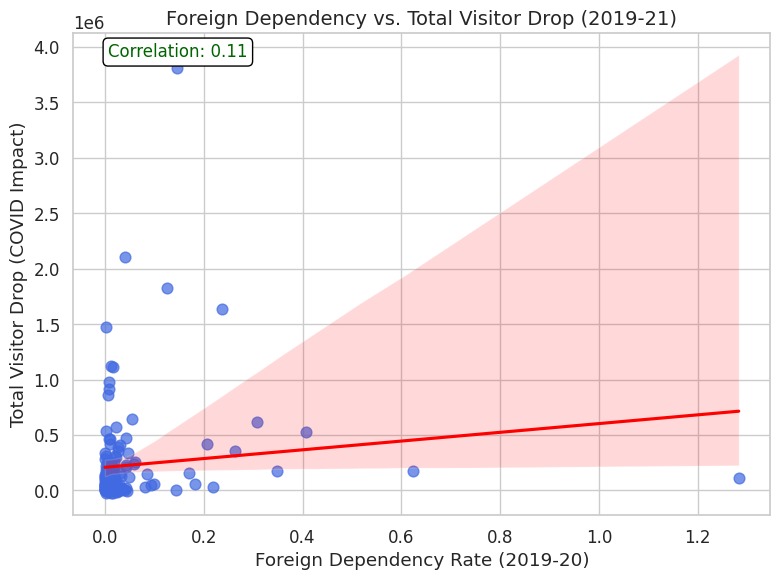
plt.text(0.05, 0.95, f"Correlation: {corr:.2f}", transform=plt.gca().transAxes,

fontsize=12, color="darkgreen", bbox=dict(facecolor='white', edgecolor='black', boxstyle='round'))

plt.tight\_layout()

plt.show()

**Output**

****

**Foreign Dependency vs Total Visitor Drop :**

>Foreign visitors dropped drastically.

>This caused a steep % drop in overall tourism in highly foreign-visited monuments.

>So the effect of foreign tourists is site-specific

**7. Prediction:**

**Simple Linear Regression:**

In this analysis, **Simple Linear Regression** was used to explore the relationship between two continuous variables: the **foreign visitor drop** and the **total visitor drop** from 2019 to 2021. The goal of this regression analysis is to predict the total visitor drop based on the change in foreign visitors. By fitting a linear model to the data, we aim to identify if there is a linear trend or pattern that explains how the decline in foreign tourism correlates with the overall decline in tourist numbers.

The regression line represents the best fit for the data, showing whether increases or decreases in foreign visitors are associated with proportional changes in the total number of visitors. The **slope** of the line indicates the strength and direction of the relationship between these two variables, while the **intercept** represents the expected value when the foreign visitor drop is zero.

**Code Implementation** :

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import r2\_score

import numpy as np

df['Total\_Visitors\_2019'] = df['Domestic\_2019-20'] + df['Foreign\_2019\_20']

df['Total\_Visitors\_2020'] = df['Domestic\_2020\_21'] + df['Foreign\_2020\_21']

df\_clean = df[['Foreign\_2019\_20', 'Total\_Visitors\_2019', 'Total\_Visitors\_2020']].dropna()

X = df\_clean[['Foreign\_2019\_20']]

y\_2019 = df\_clean['Total\_Visitors\_2019']

model\_2019 = LinearRegression().fit(X, y\_2019)

r2\_2019 = r2\_score(y\_2019, model\_2019.predict(X))

y\_2020 = df\_clean['Total\_Visitors\_2020']

model\_2020 = LinearRegression().fit(X, y\_2020)

r2\_2020 = r2\_score(y\_2020, model\_2020.predict(X))

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

sns.scatterplot(x=X['Foreign\_2019\_20'], y=y\_2019, label='2019-20 Total Visitors', color='blue')

sns.lineplot(x=X['Foreign\_2019\_20'], y=model\_2019.predict(X), color='blue', label=f'2019-20 Regression (R²={r2\_2019:.2f})')

sns.scatterplot(x=X['Foreign\_2019\_20'], y=y\_2020, label='2020-21 Total Visitors', color='green')

sns.lineplot(x=X['Foreign\_2019\_20'], y=model\_2020.predict(X), color='green', label=f'2020-21 Regression (R²={r2\_2020:.2f})')

plt.xlabel("Foreign Visitors (2019-20)")

plt.ylabel("Total Visitors")

plt.title(" Impact of Foreign Visitors on Total Tourism: Pre vs. Post COVID")

plt.legend()

plt.grid(True)

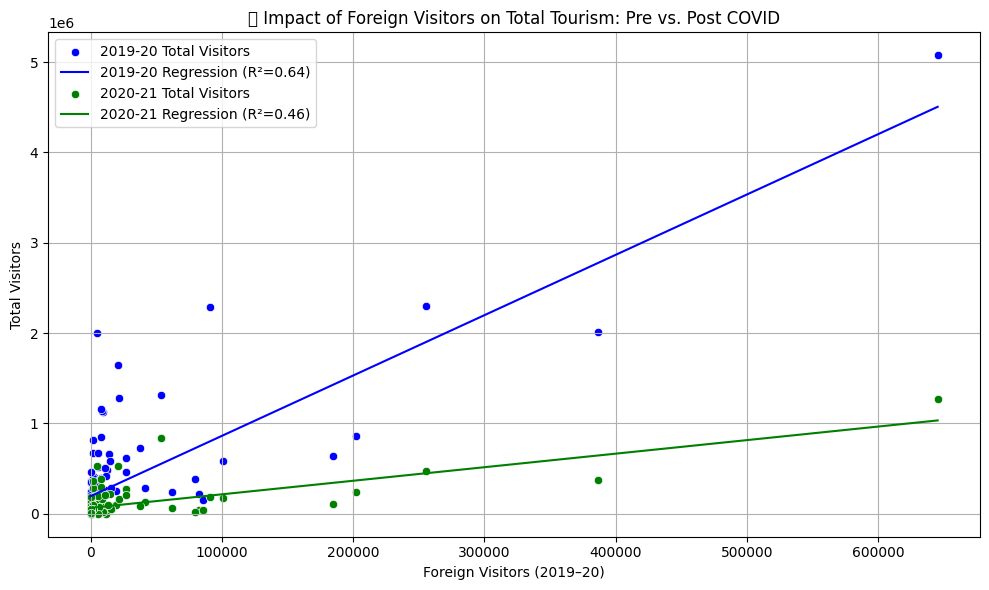
plt.tight\_layout()

plt.show()

print(f"R² in 2019-20: {r2\_2019:.2f} (foreign visitors were strongly related to total tourism)")

print(f"R² in 2020-21: {r2\_2020:.2f} (relationship weakened due to COVID impact)")

**Output :**



>The chart shows a linear regression analysis of the impact of foreign visitors on total tourism.

>2019–20 data (blue) shows a stronger correlation (R² = 0.64) between foreign and total visitors.

>2020–21 data (green) shows a weaker correlation (R² = 0.46), reflecting the effects of COVID-19.

>Each dot represents a region's visitor data, with both years plotted against foreign visitors from 2019–20.

>The regression lines reveal that foreign tourism influenced total tourism more before the pandemic

**8.Evaluation using R² :**

**Pre-COVID (2019–20 Total Visitors):**

* + **R² Score = 0.64**
  + This means **64%** of the variation in total visitors for 2019–20 can be explained by foreign visitor numbers.

**Post-COVID (2020–21 Total Visitors):**

* + **R² Score = 0.46**
  + This means **46%** of the variation in total visitors for 2020–21 can be explained by 2019–20 foreign visitor numbers.

>R² Score

Indicates the quality of the model in explaining variation in Infosys metrics. The closer to 1 = better prediction.

>Standard Error (SEE)

Indicates average prediction error. Lower SEE = better financial forecasts.

>Regression Toward the Mean

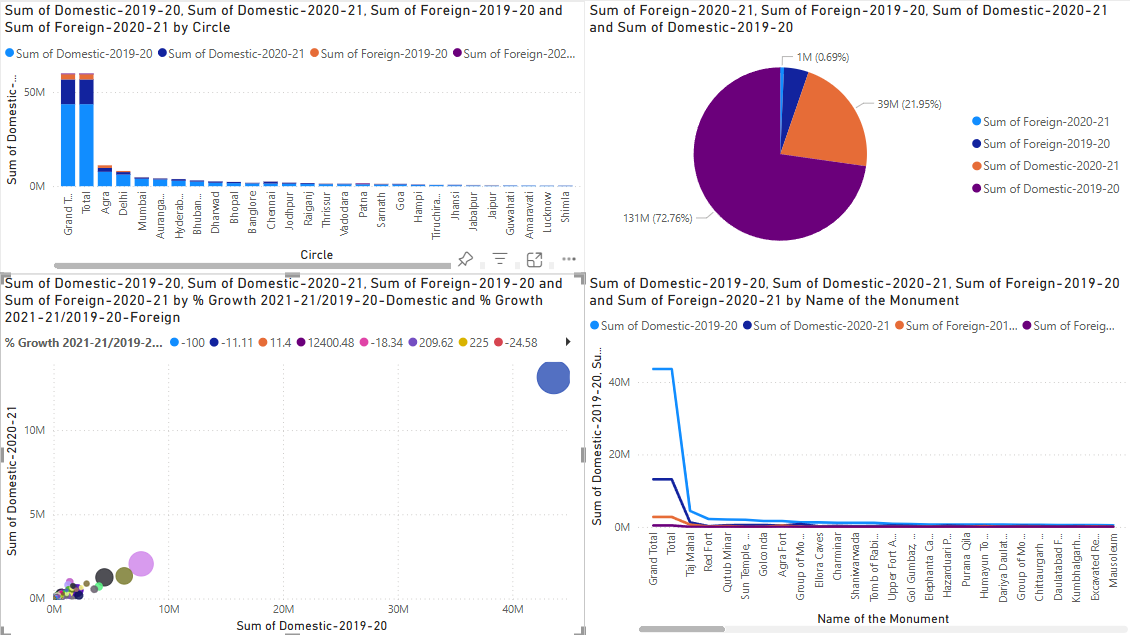
Determines whether extreme revenue/profit values stabilize over time, helpful in identifying long-term trends.

**9. GUI Graphical User Interface** :

Graphical User Interfaces like **Power BI, Tableau, and Streamlit** were employed to visually analyze the impact of COVID-19 on Indian tourism. These tools offered interactive dashboards, insightful charts, and simplified reporting capabilities, transforming complex tourism data into intuitive visual stories for stakeholders.

Power BI/Tableau provided dynamic and visually compelling dashboards to explore trends in **domestic and foreign visitors**, **monument-wise footfall**, and **regional variations** across two key years (2019–20 and 2020–21).

The dashboard delivered a comprehensive analysis of tourism metrics before and after the pandemic. It featured **four key interactive charts**:



**1.Regional Visitor Distribution by Year (Bar Chart):**

A clustered bar chart showing the sum of domestic and foreign tourists across different **tourism circles** (regions) for both 2019–20 and 2020–21. This visual highlights how tourism dropped across states due to the pandemic.

**2.Overall Visitor Composition (Pie Chart):**

A pie chart representing the **share of total visitors** by type (domestic/foreign) and year. This visual quantifies the dominance of domestic tourism and illustrates the steep decline in foreign arrivals post-COVID.

**3.Growth Analysis: Domestic vs. Foreign Visitors (Scatter Chart):**

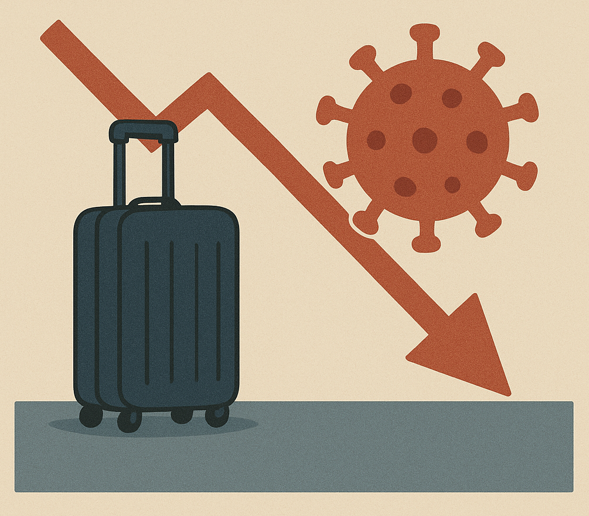
A scatter/bubble chart displaying **2019–20 vs. 2020–21 tourism** for each region or monument, color-coded by the percentage growth or decline. It helps to visually assess which areas were hit the hardest or showed resilience during the pandemic.

**4.Monument-wise Visitor Trend (Line Chart):**

A multi-series line chart comparing **monument-level footfall** across years and visitor types. This enables a deeper understanding of how landmark sites like the Taj Mahal, Red Fort, and Qutub Minar were impacted.

**10. Conclusion:**

The analysis reveals that the **COVID-19 pandemic** had a significant impact on tourism, with both **domestic** and **foreign visitor numbers** experiencing a sharp decline from 2019 to 2021, although foreign tourism was affected more severely due to global travel restrictions. The correlation analysis indicated that in **2019-20**, foreign visitors were strongly correlated with total tourism, as evidenced by a high R² value, but this relationship weakened in **2020-21** as international travel plummeted. The **regional visitor drop** analysis highlighted that tourist spots relying heavily on foreign visitors saw the most substantial declines. Additionally, **simple linear regression** showed that the decline in foreign visitors was closely linked to the overall visitor drop, further emphasizing the pandemic’s disproportionate effect on international tourism. The findings suggest that regions dependent on foreign visitors need to consider diversifying their tourism sources and focus more on domestic markets to mitigate future risks, while investing in **digital tourism** and **localized marketing** could aid in recovery efforts.

****