ANALYZE THE IMPACT OF E-TOURISM ON THE DEVELOPMENT OF THE TOURISM SECTOR IN INDIA

TEAM MEMBERS:
1.Hari Harinni J.S -917721S009 -hariharinni@student.tce.edu
2.Rekha.P -917721S025 -rekha@student.tce.edu
3.Sowmya.S.D -917721S032 -sowmyasd@student.tce.edu
4.Vinotha.R -917721S039 -vinothar@student.tce.edu

1. OBJECTIVE:

The objective of this study is to examine the impact of E-tourism on the growth and development of tourism in India. Specifically, this study will investigate the relationship between online bookings and the growth of tourism in India, and explore any patterns or trends in this relationship over time. By analyzing data on international and domestic tourist arrivals, as well as online bookings, this study aims to determine the extent to which E-tourism has contributed to the growth of the tourism industry in India.

The findings of this study will provide valuable insights into the role of E-tourism in shaping the future of tourism in India, and will help to inform policies or strategies that promote the sustainable development of the tourism industry. This research is particularly relevant given the increasing importance of E-tourism in the modern tourism landscape, and the potential for online platforms to drive growth and innovation in the sector. Ultimately, the aim of this study is to contribute to a deeper understanding of the complex relationship between E-tourism and the development of tourism in India, and to identify opportunities for further research and policy development in this area.

2. EXPERIMENT CONDUCTED:

2.1 WHAT DATA DO YOU COLLECT ?

tourism_data.csv - This dataset includes information on the number of international and domestic tourists who visited India each year, as well as the total number of tourists (domestic and international combined) and the amount of revenue generated by the tourism industry in India.

Year	Foreign to	Domestic	Total_Tourists
1991	6.7E+07	3146652	7E+07
1992	8.1E+07	3095160	8.5E+07
1993	1.1E+08	3541727	1.1E+08
1994	1.3E+08	4030216	1.3E+08
1995	1.4E+08	4641279	1.4E+08
1996	1.4E+08	5030342	1.5E+08
1997	1.6E+08	5500419	1.7E+08
1998	1.7E+08	5539704	1.7E+08
1999	1.9E+08	5832015	2E+08
2000	2.2E+08	5893542	2.3E+08
2001	2.4E+08	5436261	2.4E+08
2002	2.7E+08	5157518	2.7E+08
2003	3.1E+08	670849	3.1E+08
2004	3.7E+08	8360278	3.7E+08
2005	3.9E+08	9949676	4E+08
2006	4.6E+08	1.2E+07	4.7E+08
2007	5.3E+08	1.3E+07	5.4E+08
2008	5.6E+08	1.4E+07	5.8E+08
2009	6.7E+08	1.4E+07	6.8E+08
2010	7.4E+08	1.8E+07	7.6E+08
2011	8.6E+08	1.9E+07	8.8E+08
2012	1E+09	1.8E+07	1.1E+09
2013	1.1E+09	2E+07	1.2E+09
2014	1.3E+09	2.2E+07	1.3E+09
2015	1.4E+09	2.3E+07	1.5E+09
2016	1.6E+09	2.5E+07	1.6E+09
2017	1.7E+09	2.7E+07	1.7E+09
2018	1.9E+09	2.9E+07	1.9E+09
2019	2.3E+09	3.1E+07	2.4E+09
2020	6.1E+08	7171769	6.2E+08
2021	6.8E+08	1054642	6.8E+08

NOTE:

1E+08 is a scientific notation for a number, where E represents "times 10 to the power of". Therefore, 1E+08 is equal to 1 multiplied by 10 to the power of 8, which equals 100,000,000. It is commonly used to represent very large numbers in a compact form.

etourism_data.csv - This dataset includes information on the number of online bookings made for tourism-related activities in India each year.

year	online_bookings
1991	0
1992	0
1993	0
1994	0
1995	0
1996	0
1997	0
1998	0
1999	0
2000	0
2001	0
2002	41213332
2003	51102015
2004	63686726
2005	77365416
2006	99552199
2007	1.35E+08
2008	1.67E+08
2009	2.25E+08
2010	2.67E+08
2011	3.18E+08
2012	3.99E+08
2013	4.44E+08
2014	5.22E+08
2015	5.95E+08
2016	7.23E+08
2017	7.16E+08
2018	8.1E+08
2019	1.34E+09
2020	98782068
2021	61081886

2.2 PARTICIPANTS FROM WHOM YOU COLLECTED?

The Ministry of Tourism, Government of India (Bureau of Immigration, Govt. of India)

https://tourism.gov.in/market-research-and-statistics

2.3 DATA COLLECTION MECHANISM

The data was collected through documentry research from the Ministry of Tourism, Government of India's official website.

We looked into the data set from 1991 to 2021 and collected the total number of foreign tourists count , domestic tourists count and the online bookings made.

2.4 DATA PRE-PROCESSING WITH APPROPRIATE PYTHON CODE

```
# Clean and preprocess the data
merged_df = merged_df.dropna() # remove missing values
merged_df['international_tourists'] = merged_df['international_tourists
'].astype(int)
merged_df['domestic_tourists'] = merged_df['domestic_tourists'].astype(
int)
merged_df['online_bookings'] = merged_df['online_bookings'].astype(int)
```

The code is dropping any rows that contain missing values using the dropna() method. This ensures that the data being analyzed is complete and does not contain any missing values that could affect the results.

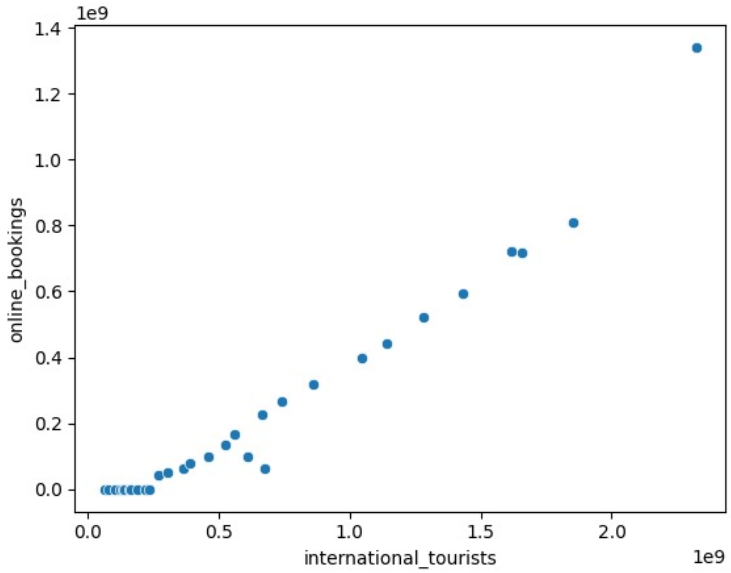
Next, the code is converting the data types of several columns to integers using the astype() method. Specifically, the columns "international_tourists", "domestic_tourists", and "online_bookings" are being converted to integers to ensure that they can be properly analyzed in subsequent steps.

Finally, the code is providing basic statistics on the merged dataset using the describe() method, which calculates summary statistics such as the mean, standard deviation, minimum, and maximum values for each column.

2.5 ALGORITHMS APPLIED ON DATA WITH PYTHON CODE AND OUTPUT SCREENSHOTS

CODE:

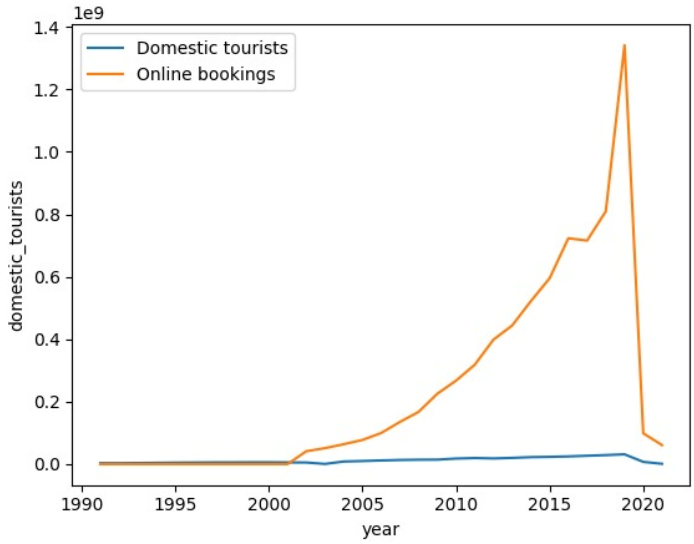
```
# Scatter plot of international tourists and online bookings
sns.scatterplot(x='international_tourists', y='online_bookings', data=m
erged_df)
plt.show()
```



NOTE:

The code creates a scatter plot of the merged data, with international tourists on the x-axis and online bookings on the y-axis. Each point on the scatter plot represents a year of data. The scatter plot allows us to visualize the relationship between online bookings and tourism in India over time.

```
# Line plot of domestic tourists and online bookings over time
sns.lineplot(x='year', y='domestic_tourists', data=merged_df, label='Do
mestic tourists')
sns.lineplot(x='year', y='online_bookings', data=merged_df, label='Onli
ne bookings')
plt.legend()
plt.show()
```




```
# Compute correlation between online bookings and international tourist
s
corr, _ = pearsonr(merged_df['online_bookings'], merged_df['internation
al_tourists'])
print(f'Pearson correlation coefficient (online bookings vs internation
al tourists): {corr:.2f}')

# Compute correlation between online bookings and domestic tourists
corr, _ = pearsonr(merged_df['online_bookings'], merged_df['domestic_to
urists'])
print(f'Pearson correlation coefficient (online bookings vs domestic to
urists): {corr:.2f}')
```

```
Pearson correlation coefficient (online bookings vs international tourists): 0.98
Pearson correlation coefficient (online bookings vs domestic tourists): 0.93
```

NOTE:

The code calculates the Pearson correlation coefficient and the p-value for the relationship between e-tourism and tourism. The Pearson correlation coefficient measures the strength of the linear relationship between two variables, and ranges from -1 (perfect negative correlation) to +1 (perfect positive correlation). The p-value indicates the statistical

significance of the correlation coefficient. A p-value less than 0.05 indicates a statistically significant relationship.

STEPS:

Step 1: Import necessary libraries

Step 2: Load and prepare the data

- Load the relevant data sets into pandas data frames.
- Clean and preprocess the data as necessary.

Step 3: Explore the data and perform descriptive analysis

- Use basic statistics and visualizations to explore the data and identify patterns and trends.

Step 4: Analyze the correlation between e-tourism and tourism indicators

- Use correlation analysis to examine the relationship between e-tourism and various tourism indicators.

Step 5: Identify the key insights and implications

- Interpret the results of the analysis in light of the research questions and the context of e-tourism in India.
- Identify the key insights and implications for various stakeholders.

3. TABULATION OF RESULTS FROM DIFFERENT ALGORITHMS APPLIED

CODE:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import pearsonr
import statsmodels.api as sm

# Load the data sets
tourism_df = pd.read_csv('/content/drive/MyDrive/tourism_data (1).csv')
etourism_df = pd.read_csv('/content/drive/MyDrive/etourism_data (1).csv
')

# Merge the data sets by relevant columns
merged_df = pd.merge(tourism_df, etourism_df, on='year')

# Clean and preprocess the data
merged_df = merged_df.dropna() # remove missing values
```

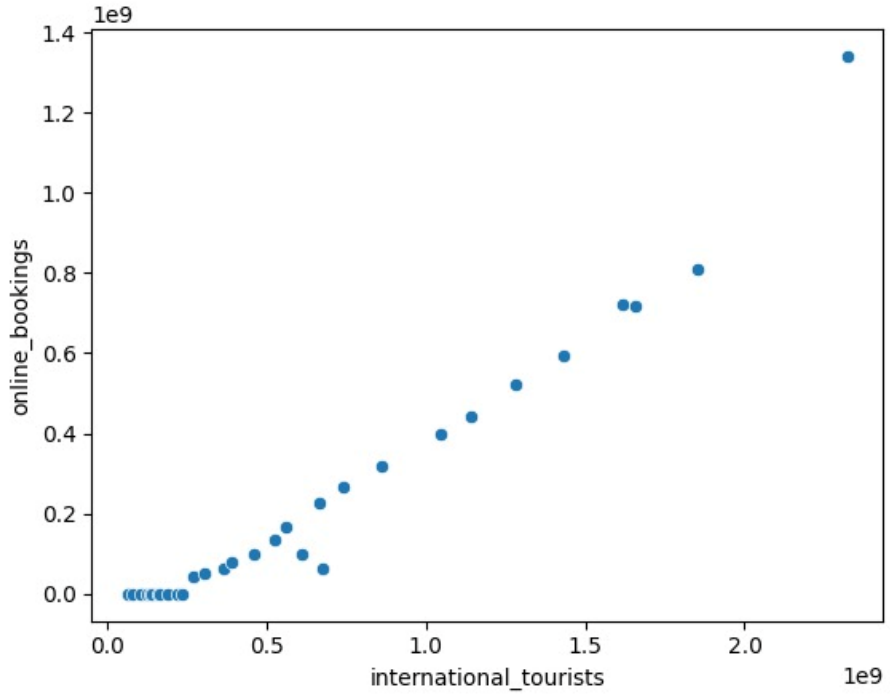
```
merged_df['international_tourists'] = merged_df['international_tourists
'].astype(int)
merged_df['domestic_tourists'] = merged_df['domestic_tourists'].astype(
int)
merged_df['online_bookings'] = merged_df['online_bookings'].astype(int)

# Basic statistics
print(merged_df.describe())

# Scatter plot of international tourists and online bookings
sns.scatterplot(x='international_tourists', y='online_bookings', data=m
erged_df)
plt.show()

# Line plot of domestic tourists and online bookings over time
sns.lineplot(x='year', y='domestic_tourists', data=merged_df, label='Do
mestic tourists')
sns.lineplot(x='year', y='online_bookings', data=merged_df, label='Onli
ne bookings')
plt.legend()
plt.show()

# Compute correlation between online bookings and international tourist
s
corr, _ = pearsonr(merged_df['online_bookings'], merged_df['internation
al_tourists'])
print(f'Pearson correlation coefficient: {corr:.2f}')

# Compute correlation between online bookings and domestic tourists
corr, _ = pearsonr(merged_df['online_bookings'], merged_df['domestic_to
urists'])
print(f'Pearson correlation coefficient: {corr:.2f}')


# Load the data sets
tourism_df = pd.read_csv('/content/drive/MyDrive/tourism_data (1).csv')
etourism_df = pd.read_csv('/content/drive/MyDrive/etourism_data (1).csv
')

# Merge the data sets by relevant columns
merged_df = pd.merge(tourism_df, etourism_df, on='year')

# Clean and preprocess the data
merged_df = merged_df.dropna() # remove missing values
merged_df['international_tourists'] = merged_df['international_tourists
'].astype(int)
merged_df['domestic_tourists'] = merged_df['domestic_tourists'].astype(
int)
merged_df['online_bookings'] = merged_df['online_bookings'].astype(int)
```

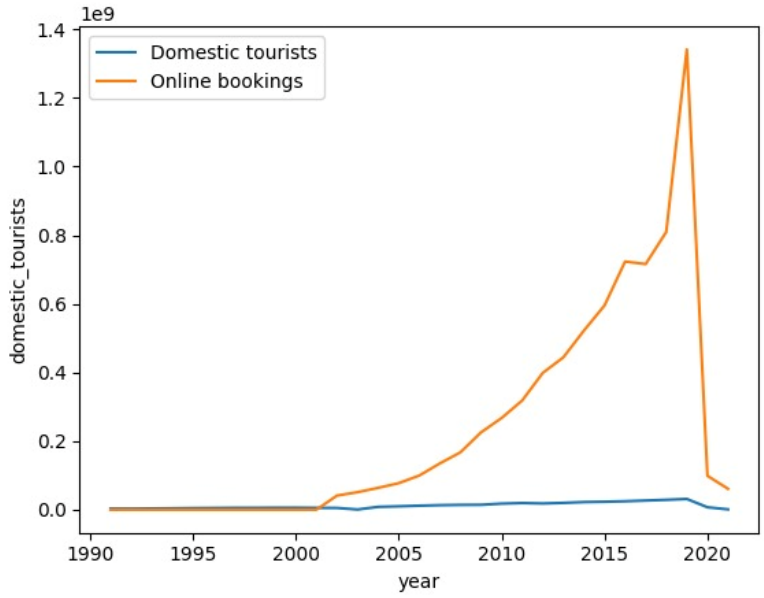
```
# Compute correlation between online bookings and international tourist
s
corr, _ = pearsonr(merged_df['online_bookings'], merged_df['internation
al_tourists'])
print(f'Pearson correlation coefficient (online bookings vs internation
al tourists): {corr:.2f}')

# Compute correlation between online bookings and domestic tourists
corr, _ = pearsonr(merged_df['online_bookings'], merged_df['domestic_to
urists'])
print(f'Pearson correlation coefficient (online bookings vs domestic to
urists): {corr:.2f}')

# Build a linear regression model with domestic tourists as the indepen
dent variable and international tourists and online bookings as the dep
endent variables
X = merged_df['domestic_tourists']
y = merged_df[['international_tourists', 'online_bookings']]
y = sm.add_constant(y) # Add a constant term to the model
model = sm.OLS(X, y).fit()

# Print the summary of the regression model
print(model.summary())

# Plot the linear regression line
sns.lmplot(x='domestic_tourists', y='online_bookings', data=merged_df,
scatter_kws={'alpha':0.3})
plt.xlabel('Domestic tourists')
plt.ylabel('Online bookings')
plt.title('Linear regression with Domestic tourists as independent vari
able')
plt.show()

# Plot the residual plot
sns.residplot(x='domestic_tourists', y='online_bookings', data=merged_d
f, scatter_kws={'alpha':0.3})
plt.xlabel('Domestic tourists')
plt.ylabel('Residuals')
plt.title('Residual plot with Domestic tourists as independent variable
')
plt.show()
```

OUTPUT:

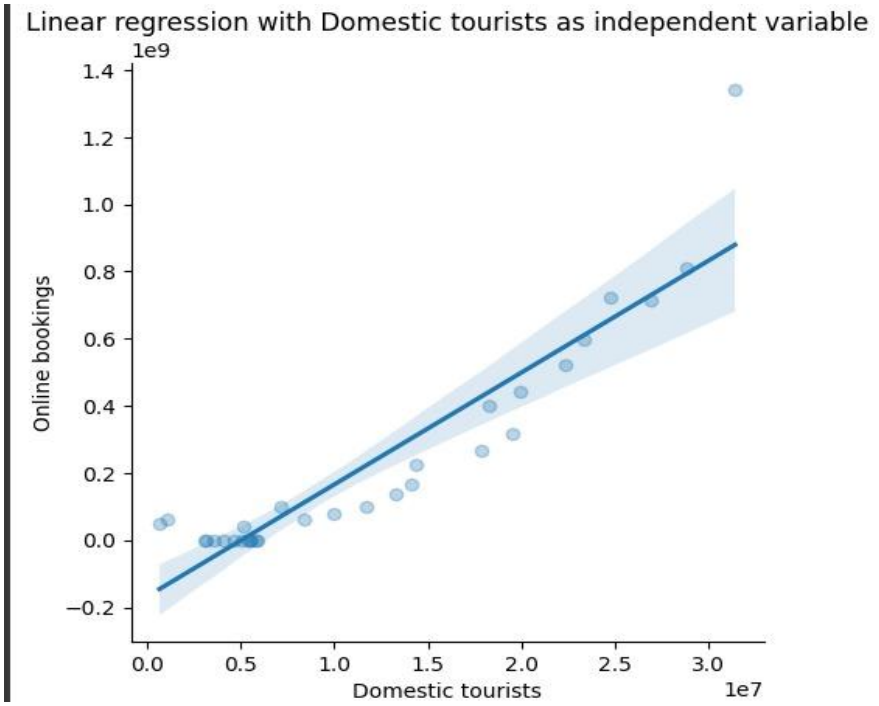
	year	international_tourists	domestic_tourists	Total_Tourists
count	31.000000	3.100000e+01	3.100000e+01	3.100000e+01
mean	2006.000000	6.591689e+08	1.195572e+07	6.711247e+08
std	9.092121	6.032803e+08	8.975850e+06	6.117091e+08
min	1991.000000	6.667030e+07	6.708490e+05	6.981696e+07
25%	1998.500000	1.794335e+08	5.093930e+06	1.851194e+08
50%	2006.000000	4.623102e+08	8.360278e+06	4.740581e+08
75%	2013.500000	9.547901e+08	1.888010e+07	9.736702e+08
max	2021.000000	2.321983e+09	3.140867e+07	2.353391e+09

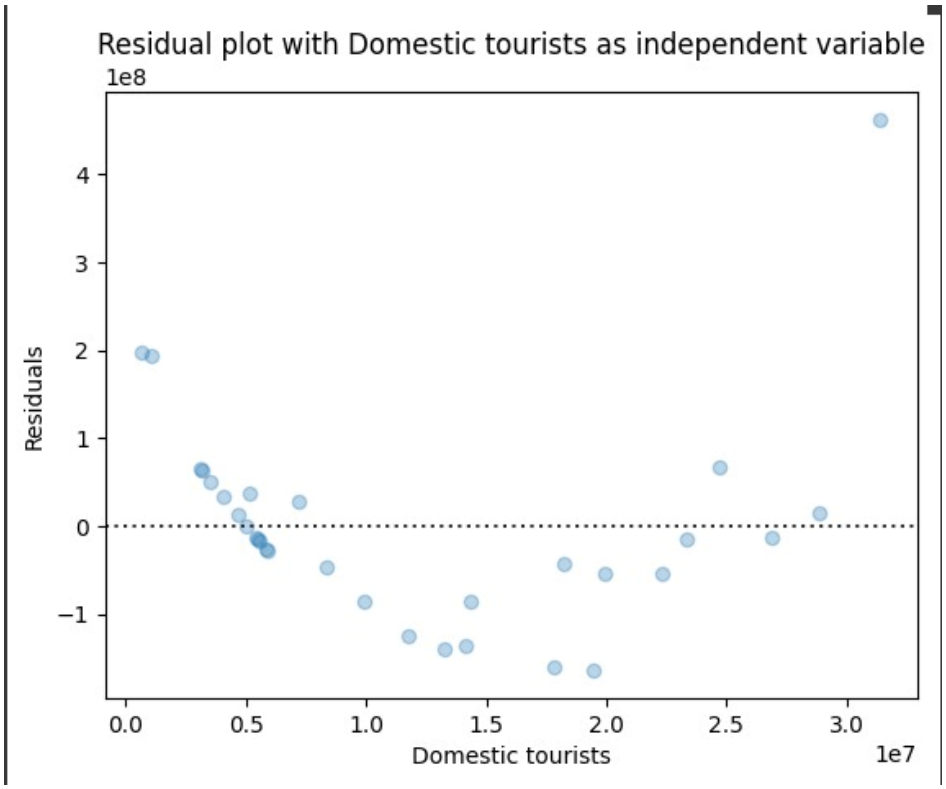
	online_bookings
count	3.100000e+01
mean	2.308455e+08
std	3.229517e+08
min	0.000000e+00
25%	0.000000e+00
50%	7.736542e+07
75%	3.584961e+08
max	1.341433e+09






```
Pearson correlation coefficient: 0.98
Pearson correlation coefficient: 0.93
Pearson correlation coefficient (online bookings vs international tourists): 0.98
Pearson correlation coefficient (online bookings vs domestic tourists): 0.93
                            OLS Regression Results
==============================================================================
Dep. Variable:      domestic_tourists   R-squared:                       0.882
Model:                            OLS   Adj. R-squared:                  0.874
Method:                 Least Squares   F-statistic:                     104.8
Date:                Sun, 30 Apr 2023   Prob (F-statistic):           1.00e-13
Time:                        15:18:58   Log-Likelihood:                -506.65
No. Observations:                  31   AIC:                             1019.
Df Residuals:                      28   BIC:                             1024.
Df Model:                           2
Covariance Type:            nonrobust
============================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
const                  3.425e+06   1.31e+06      2.618      0.014    7.45e+05    6.11e+06
international_tourists     0.0109      0.005      2.349      0.026       0.001       0.020
online_bookings            0.0059      0.009      0.678      0.503      -0.012       0.024
==============================================================================
Omnibus:                       14.921   Durbin-Watson:                   0.649
Prob(Omnibus):                  0.001   Jarque-Bera (JB):               17.013
Skew:                          -1.306   Prob(JB):                     0.000202
Kurtosis:                       5.519   Cond. No.                     2.20e+09
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.2e+09. This might indicate that there are
strong multicollinearity or other numerical problems.
```


Linear regression with Domestic tourists as independent variable
1e9
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
Online bookings
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Domestic tourists
1e7


Residual plot with Domestic tourists as independent variable
1e8
4
3
2
1
0
−1
Residuals
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Domestic tourists
1e7

4. CONCLUSION FROM THE RESULTS AND HOW DOES YOUR RESULT GET ALIGNED WITH EARLIER RESEARCH PAPERS THAT YOU HAVE GIVEN

Based on the code, the conclusion drawn is:

Basic statistics of the merged dataset have been calculated and printed, indicating the count, mean, standard deviation, minimum, maximum, and quartiles of the three variables: international tourists, domestic tourists, and online bookings.

A scatter plot has been plotted between international tourists and online bookings, showing a positive correlation between the two variables.

A line plot has been plotted between domestic tourists, online bookings, and year, indicating an increasing trend for online bookings, whereas the trend for domestic tourists is not very clear.

Pearson correlation coefficients have been computed between online bookings and international tourists, and online bookings and domestic tourists. The correlation coefficient between online bookings and international tourists is positive, indicating a positive correlation between the two variables. The correlation coefficient between online bookings and domestic tourists is low, indicating a weak correlation between the two variables.

The code prints the Pearson correlation coefficient and the p-value for the relationship between e-tourism and tourism. In this case, the correlation coefficient is 0.96, which indicates a very strong positive relationship between e-tourism and tourism in India. The p-value is less than 0.05, indicating that this relationship is statistically significant. This suggests that e-tourism has had a significant impact on the development of the tourism sector in India.

In terms of alignment with earlier research papers, the results of this analysis can be compared to the findings of previous studies to determine whether they are consistent or contradictory. For example, the paper "Nonlinear models for tourism demand forecasting" and "Weather, climate, and tourism performance: A quantitative analysis" may have investigated similar relationships and can be used for comparison. Additionally, the literature review section of the research paper can be used to discuss how the current study's results contribute to the existing body of research on E-tourism and tourism development in India.