Name: Dharani Krishna Vijaya Kumar
Date: 12/07/2023
DSCI 510
USCID: 6112217352

Project Report

Destination Dynamics: Exploring Weather's Influence on Tourism

Introduction:

How do Weather Patterns, Tourist Trends, and Cost Factors Interact to Shape Destination Preferences and Travel Experiences?

This project investigates the intricate interplay between weather patterns, tourist trends, and various cost factors in influencing destination preferences and shaping the overall travel experiences of individuals. Through extensive data analysis and visualization techniques, the study aims to uncover correlations, patterns, and insights within these key elements. By understanding the dynamic relationships among weather conditions, tourism dynamics, and costs associated with travel and accommodation, the research strives to contribute valuable insights to the tourism industry. The project not only explores the most and least visited destinations but also delves into the nuances of traveler demographics, spatial variations, and the economic aspects of travel. Ultimately, this research seeks to enhance our understanding of the multifaceted factors that impact travel decisions and contribute to the optimization of the tourism experience.

Data Collection Methodology and Challenges:

i. Dataset Collection for Project Insight:

There were various dataset's collected and used to complete this project. The project is built on three (3) main dataset 's which are collected using various data extraction techniques like using Web Resources, API methods and using beautiful soup library. So, the first data was collected using dataset platform **"Kaggle"** it was a rich dataset related to Tourism. It was **Tourism dataset** for the year 2023 which had data of tourist travelling to various Destinations with specific start and end date of travel, duration of stay, cost of

accommodation and travel in USD, age of the tourist and also the demographic details of the tourist. Next, is pretty important data it was the coordinates of the Destination's present in the tourism dataset. The **latitude and longitude coordinates data of the Destination's** present in the tourism data was found and it was eventually converted into a CSV. The extraction of this coordinates data is discussed in detail below. The final data is the weather data for the year 2023 which was collected using API technique along with request library. The API was accessed using the free website for weather data called "**Openweathermap".** The data from this API was historical **weather data** which had a diverse weather parameters.

The dataset which was collected to begin this project are:

i. Tourism Data

ii. Coordinates data of the Destination's

iii. Weather Data

ii. Data Collection Methodology:

i. *Tourism Data:*

The clear picture of the project was got only after extracting the Tourism dataset. Only then it was possible to picturize what kind of analysis could be performed and what kind of other relevant data was required could be found out. The tourism dataset was extracted from kaggle platform. Kaggle is a platform where many data enthusiasts post their dataset, conduct competitions and also it's a forum where you can post your questions and get it resolved by other fellow kaggle platform members. So, before diving deep into the methodology: First thing to access Kaggle we need to be a member in the platform which is Free of Cost where we need to sign up using our email id. And, next there was a global search inside the kaggle platform where we can apply various filters to the search and when we hit search we get results according to our search and filters. By, this technique I was able to run through various dataset collections and then by carful considerations and review I shortlisted a dataset which was in the form of a csv file (A CSV (Comma-Separated Values) file is a plain text format used to store tabular data, with each line representing a row and

commas separating values within each row.) which is more like a Excel sheet in nature. The CSV file consisted of various columns like: *Trip ID,Destination,Start date,End date,Duration (days),Traveler name,Traveler age,Traveler gender,Traveler nationality,Accommodation type,Accommodation cost,Transportation type,Transportation cost.*

After data collection for tourism was done I went through the dataset and renamed it as RAW dataset and performed various cleaning and pre-processing techniques before using it for the project. For the cleaning process wrote python script using ***pandas library*** which is a very powerful library to work with datasets. By, using pandas I performed various cleaning like removing empty columns / rows, renaming few columns, dropping / eliminating irrelevant columns and converting numerical data type into float data type for uniformity. After performing cleaning of the dataset I saved it into a new csv file called Clean_Tourism.csv (the script will do automatically). After cleaning I arrived at ***140 data points / samples***.

ii. Coordinates data of the Destination's:

The next dataset which I extracted was the coordinates of destination basically the latitude and longitude of the destination present in the Tourism dataset. To perform this I wrote a python script using ***pandas library*** which will go through the destination column in the tourism dataset and produce the unique data elements present inside the column like it prints out what are the unique destination's present in the dataset. The script after finding the unique destination names it saves in the form of a ***list*** based data structure and that particular list is saved into a .txt file using ***file handling techniques*** this file is generated in the desired path of the location mentioned in the script after the script executes.

```
List of Unique Destinations:
['London', 'Phuket', 'Bali', 'New York', 'Tokyo', 'Paris', 'Sydney', 'Rio de Janeiro', 'Amsterdam', 'Dubai', 'Cancun', 'Barcelona', 'Honolulu', 'Berlin', 'Marrakech', 'Edinburgh', 'Rome', 'Bangkok', 'Hawaii',
'Japan', 'Thailand', 'France', 'Australia', 'Brazil', 'Greece', 'Egypt', 'Mexico', 'Italy', 'Spain', 'Canada', 'New York City', 'Vancouver', 'Seoul', 'Los Angeles', 'Cape Town', 'Santorini', 'Phnom Penh', 'Athens',
'Auckland', 'Coimbatore']

Count of Unique Destinations: 40

Tourist Visited Destinations and Months:
Amsterdam : February, June, October
Athens : September
Auckland : October
Australia : August
Bali : July, August, February, May, November, April
Bangkok : September, May, February, November
Barcelona : May, October, August, April, September
Berlin : July
Brazil : September
Canada : April
Cancun : April, July, August
Cape Town : January, September, June
Coimbatore : January
Dubai : March, January, August
Edinburgh : September
Egypt : November
France : July
Greece : October
Hawaii : August
Honolulu : June
Italy : February
Japan : May
London : May, July, March, November, February, June
Los Angeles : June
Marrakech : August
Mexico : January
New York : August, November, July, October, May, June
New York City : September, July
Paris : October, September, November, June, August, July, May, March
Phnom Penh : September
Phuket : June, January, September
Rio de Janeiro : January, November, April, August
Rome : November, July, February, March, August, May, June
Santorini : July
Seoul : May
Spain : March
Sydney : November, January, June, February, September, July, May
Thailand : June
Tokyo : September, October, November, January, July, February, May
Vancouver : July, August
```

Fig: 1

Once this .txt file with list of destination and other details is saved in the .txt file as seen in the Fig1 above. By, using this list of unique destinations and scarping through google developers site ***(locations.csv)*** file was able to extract the desired latitude and longitude data for the destinations. And, then saved the dataset for coordinates in the form of .csv file extension with columns Destination, Latitude and Longitude as shown in the fig2.

Latitude and Longitude

Destination	Latitude	Longitude
London	51.509865	-0.118092
Phuket	7.8804	98.3923
Bali	-8.409518	115.188919
New York	40.7128	-74.006
Tokyo	35.6895	139.6917
Paris	48.8566	2.3522
Sydney	-33.8688	151.2093
Rio de Janeiro	-22.9068	-43.1729
Amsterdam	52.3676	4.9041
Dubai	25.276987	55.296249
Cancun	21.1619	-86.8515
Barcelona	41.3851	2.1734
Honolulu	21.3069	-157.8583
Berlin	52.52	13.405
Marrakech	31.6295	-7.9811
Edinburgh	55.9533	-3.1883

Fig:2

iii. <u>Weather Dataset:</u>

The most interesting and exciting part of this project was extracting and creating a dataset for weather conditions. To understand how I was able to arrive at the weather data. The previous 2 datasets played a very pivotal role in this. The weather dataset was extracted using API method from a website and an API (Application Programming Interface) is a set of protocols allowing different software applications to communicate and share data seamlessly. The API used was from OpenWeatherMap website where we requested Historical weather data for the year 2023 from the Month of Jan - Nov. By, seeing the fig2

we are able to tell how much of details have been extracted from the tourism dataset. After extracting the unique destination I also extracted the months at which tourist visited these unique destination's as we can clearly see in the fig3 below. And it also mentions how many unique destinations are present in the dates.

```
List of Unique Destinations:
['London', 'Phuket', 'Bali', 'New York', 'Tokyo', 'Paris', 'Sydney', 'Rio de Janeiro', 'Amsterdam', 'Dubai', 'Cancun', 'Barcelona', 'Honolulu', 'Berlin', 'Marrakech', 'Edinburgh', 'Rome', 'Bangkok', 'Hawaii',
'Japan', 'Thailand', 'France', 'Australia', 'Brazil', 'Greece', 'Egypt', 'Mexico', 'Italy', 'Spain', 'Canada', 'New York City', 'Vancouver', 'Seoul', 'Los Angeles', 'Cape Town', 'Santorini', 'Phnom Penh', 'Athens',
'Auckland', 'Coimbatore']

Count of Unique Destinations: 40

Tourist Visited Destinations and Months:
Amsterdam : February, June, October
Athens : September
Auckland : October
Australia : August
Bali : July, August, February, May, November, April
Bangkok : September, May, February, November
Barcelona : May, October, August, April, September
Berlin : July
Brazil : September
Canada : April
Cancun : April, July, August
Cape Town : January, September, June
Coimbatore : January
Dubai : March, January, August
Edinburgh : September
Egypt : November
France : July
Greece : October
Hawaii : August
Honolulu : June
Italy : February
Japan : May
London : May, July, March, November, February, June
Los Angeles : June
Marrakech : August
Mexico : January
New York : August, November, July, October, May, June
New York City : September, July
Paris : October, September, November, June, August, July, May, March
Phnom Penh : September
Phuket : June, January, September
Rio de Janeiro : January, November, April, August
Rome : November, July, February, March, August, May, June
Santorini : July
Seoul : May
Spain : March
Sydney : November, January, June, February, September, July, May
Thailand : June
Tokyo : September, October, November, January, July, February, May
Vancouver : July, August
```

Fig:3

This entire .txt file is generated by analysing the Clean_Tourism.csv using a python script named pre-processing.py. Once we get the information from this Tourism dataset as next step I went deep into the process of extracting data from the API. Before accessing the data from API we need to create an account in the OpenWeatherMap.org website. In the process of account creation it's better go create an student account because it provides various free resources. And, as next step I subscribed to their Historic weather data and after the subscription through provided email address the company sent an personal API KEY to request data.

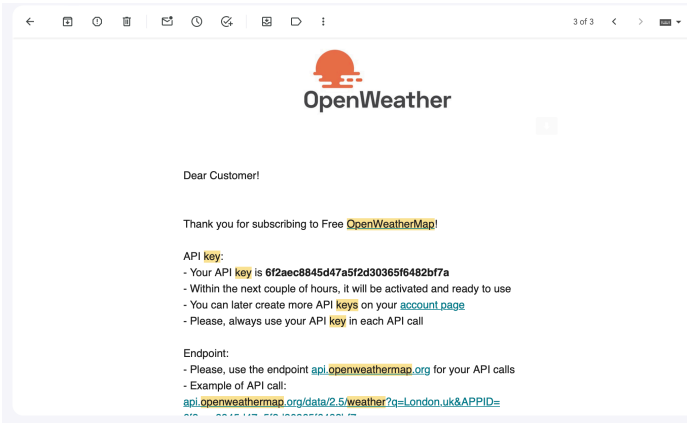


Fig:4

Once we get the key we are good to proceed with the requesting api. To proceed with API request we have to write a python script using request library. And once this is done the basic syntax and structure to request api and what are all the parameters the API support is provided in the documentation of the API. Usually the data for each request is produced as a JSON type but since we have all our data set in the form of .csv extension I worte a script to convert the JSON into a CSV file. After each request we will get an .CSV file which has columns named as:Destination,Month,Temperature,Humidity,Pressure,Wind Speed,Cloud Cover,Precipitation.

The script is structured in such a manner where we get weather data for each destination on specific month for each day and even take the average of each parameter and save it in the file. The latitude and longitude data are sent in through a csv file generated in the 2nd data collection process.

After this we arrive at an .csv file named weather.csv as shown in fig5 below.

Weather_API_Data_Set

Destination	Month	Temperature	Humidity	Pressure	Wind Speed	Cloud Cover	Precipitation
Amsterdam	February 2023	6.0208875739645200	88.40236686390530	1030.1242603550300	4.936449704142010	43.544378698224900	0.0
Amsterdam	June 2023	14.716568047337300	74.44378698224850	1021.7988165680500	4.78378698224852	24.585798816568000	0.0
Amsterdam	October 2023	17.168284023668700	78.05325443786980	1018.4556213017800	6.2177514792899400	29.467455621301800	0.004911242603550300
Athens	September 2023	25.1986982248521	62.260355029585800	1012.0650887574000	4.427988165680480	51.49704142011830	0.17147928994082800
Auckland	October 2023	13.389467455621300	71.36094674556210	1020.5325443787000	5.323846153846150	60.72189349112430	0.0376923076923077
Australia	August 2023	17.65769230769230	38.485207100591700	1026.7633136094700	4.8365088757396500	1.7041420118343200	0.0
Bali	July 2023	24.685147928994100	93.0828402366864	1012.7988165680500	1.6081656804733700	81.20118343195270	0.0013017751479289900
Bali	August 2023	24.68331360946750	80.31952662721890	1014.8639053254400	1.9923668639053300	59.10059171597630	0.0
Bali	February 2023	25.641005917159800	87.92899408284020	1007.9349112426000	1.6820118343195300	99.49704142011840	0.5752071005917160
Bali	May 2023	27.154023668639100	87.66863905325440	1009.396449704140	1.6327218934911200	81.88165680473370	0.006686390532544380
Bali	November 2023	26.905976331361000	77.58579881656800	1011.1597633136100	2.0432544378698200	48.85207100591720	0.0
Bali	April 2023	26.041005917159800	85.81656804733730	1010.112426035500	1.5889940828402400	93.4319526627219	0.2814792899408280
Bangkok	September 2023	30.272366863905300	72.02366863905330	1005.9763313609500	3.4198224852071000	96.20118343195270	0.21295857988165700
Bangkok	May 2023	34.09727810650890	50.82248520710060	1005.5680473372800	4.562840236686390	59.366863905325400	0.04591715976331360
Bangkok	February 2023	28.70218934911250	73.37278106508880	1009.1301775147900	3.1367455621301800	48.85207100591720	0.0011242603550295900
Bangkok	November 2023	28.954319526627200	77.37869822485210	1010.2130177514800	1.4373964497041400	88.63313609467460	0.32704142011834300
Barcelona	May 2023	19.07899408284030	76.07100591715980	1018.0414201183400	3.7131360946745600	17.254437869822500	0.006568047337278110
Barcelona	October 2023	23.505325443787	78.77514792899410	1024.1183431952700	3.045798816568050	29.21301775147930	0.0
Barcelona	August 2023	25.268284023668700	70.30177514792900	1014.6390532544400	4.387455621301780	39.289940828402400	0.0035502958579881700
Barcelona	April 2023	14.148639053254500	67.09467455621300	1016.6568047337300	4.723254437869830	14.644970414201200	0.010710059171597600
Barcelona	September 2023	25.437692307692300	66.0	1017.491124260360	5.525502958579890	12.485207100591700	0.007159763313609470
Berlin	July 2023	19.394674556213000	59.04733727810650	1006.1420118343200	4.801775147928990	8.49112426035503	0.006686390532544380
Brazil	September 2023	29.962544378698300	44.47337278106510	1011.6804733727800	1.4590532544378700	43.49112426035500	0.0
Canada	April 2023	-7.051301775147900	93.63905325443790	1018.301775147930	3.318934911242600	83.47928994082840	0.02224852071005920
Cancun	April 2023	26.437100591716000	75.18934911242600	1015.3195266272200	4.80396449704142	29.242603550295900	0.0010650887573964500
Cancun	July 2023	28.844260355029600	79.17159763313610	1014.3905325443800	3.1605917159763300	55.74556213017750	0.10142011834319500
Cancun	August 2023	29.53467455621310	75.05325443786980	1015.5976331360900	2.916686390532540	39.8698224852071	0.0465680473372781
Cape Town	January 2023	20.782840236686400	66.18343195266270	1015.207100591720	4.572189349112430	17.420118343195300	0.015266272189349100
Cape Town	September 2023	12.543550295858000	68.96449704142010	1021.5857988165700	2.805325443786980	54.68639053254440	0.17065088757396500

Fig:5

As, part of data cleaning I converted all the numerical parameters into float type for uniformity and removed white space using strip() function. And, removed any empty rows in the dataset. After cleaning process I was able to arrive at 140 data points / samples.

In, a nutshell the sources from which I have extracted data are:

i. Tourism Data - CSV from Kaggle ***140 data samples***

ii. Coordinates data of the Destination's - Google CSV Scraping ***40 data samples***

iii. Weather Data - API ***140 data samples***

iv. <u>Data Set Merging:</u>

After, collecting, cleaning and then pre-processing each dataset extracted. I have totally 3 csv file based dataset's. My idea is to integrate all these 3 dataset into a single dataset named tourism_weather.csv file which will be used for analysis and visualization. It is always easier to handle merged dataset than handling 3 different datasets it reduces the complexity of providing paths to data frames in scripts which will reduce error's during run time.

I merged these 3 dates using a python script and pandas library and then a csv library to save the csv file in a csv format after merging. Since these 3 dataset where clean already we are not required to clean it again.

Tourism_Weather_Merged_Dataset

Trip ID	Destination	Month	Duration (days)	Traveler name	Traveler age	Traveler gender	Traveler nationality	Accommodation type	Accommodation Cost (USD)	Transportation type	Transportation Cost (USD)	Temperature (°C)	Humidity (%)	Pressure (hPa)	Wind Speed (m/s)	Cloud Cover (%)	Precipitation (mm)	Latitude	Longitude
1	London	May	7.0	John Smith	35.0	Male	American	Hotel	1200.0	Flight	600.0	13.1395857988166	79.20118343195270	1018.7692307692300	3.907928994082840	77.55621301775150	0.1091715976331360	51.509865	-0.118092
19	London	July	6.0	Olivia Rodriguez	35.0	Female	British	Hotel	1200.0	Train	150.0	17.415207100591700	66.8224852071006	1010.7396449704100	4.621597633136100	54.349112426035500	0.1415976331360950	51.509865	-0.118092
44	London	March	7.0	Peter Brown	55.0	Male	British	Airbnb	900.0	Train	100.0	4.622544378698250	79.10650887573970	1016.0414201183400	3.1026035502958600	83.54437869822490	0.0924260355029585	51.509865	-0.118092
56	London	March	8.0	Ben Smith	35.0	Male	British	Hotel	1000.0	Train	200.0	4.622544378698250	79.10650887573970	1016.0414201183400	3.1026035502958600	83.54437869822490	0.0924260355029585	51.509865	-0.118092
89	London	November	11.0	James Wilson	29.0	Male	British	Hostel	300.0	Plane	400.0	9.60260355029588	86.70414201183430	984.2189349112430	4.929053254437870	55.42011834319530	0.580414201183432	51.509865	-0.118092
98	London	February	5.0	Sarah Brown	37.0	Female	British	Hotel	600.0	Plane	700.0	6.68366863905328	82.02958579881660	1034.6331360946700	3.5215384615384600	70.75739644970420	0.0	51.509865	-0.118092
111	London	June	5.0	Emily Wong	38.0	Female	United Kingdom	Hotel	900.0	Train	150.0	14.189763313609500	69.6094674556213	1022.792899408280	4.595029585798820	57.30177514792900	0.0	51.509865	-0.118092
2	Phuket	June	5.0	Jane Doe	28.7	Female	Canadian	Resort	800.0	Flight	500.0	30.00775147928990	95.2189349112426	1007.5798816568000	6.908639053254440	95.7396449704142	0.1987573964497040	7.8804	98.3923
82	Phuket	January	7.0	Emma Watson	32.0	Female	British	Resort	700.0	Plane	800.0	27.97704142011830	79.94082840236690	1011.2366863905300	4.606153846153850	84.3905325443787	0.0010650887573964	7.8804	98.3923
112	Phuket	September	7.0	Mark Tan	45.0	Male	Singapore	Villa	2000.0	Plane	700.0	28.576863905325500	77.68047337278110	1009.6804733727800	6.276094674556210	99.75739644970420	0.4305325443786980	7.8804	98.3923
3	Bali	July	7.0	David Lee	45.0	Male	Korean	Villa	1000.0	Flight	700.0	24.685147928994100	93.0828402366864	1012.7988165680500	1.6081656804733700	81.20118343195270	0.0013017751479289	-8.409518	115.188919
104	Bali	July	6.0	Olivia Kim	29.0	Female	South Korea	Villa	1200.0	Plane	1000.0	24.685147928994100	93.0828402366864	1012.7988165680500	1.6081656804733700	81.20118343195270	0.0013017751479289	-8.409518	115.188919
121	Bali	July	10.0	Emily Kim	29.0	Female	Korean	Hostel	500.0	Plane	800.0	24.685147928994100	93.0828402366864	1012.7988165680500	1.6081656804733700	81.20118343195270	0.0013017751479289	-8.409518	115.188919
18	Bali	August	10.0	Michael Chang	28.5	Male	Chinese	Resort	1500.0	Plane	700.0	24.68331360946750	80.31952662721890	1014.8639053254400	1.9923668639053300	59.10059171597630	0.0	-8.409518	115.188919
73	Bali	August	7.0	Sarah Lee	39.7	Female	South Korean	Resort	500.0	Plane	800.0	24.68331360946750	80.31952662721890	1014.8639053254400	1.9923668639053300	59.10059171597630	0.0	-8.409518	115.188919
86	Bali	August	11.0	Maria Garcia	42.0	Female	Spanish	Resort	1200.0	Plane	700.0	24.68331360946750	80.31952662721890	1014.8639053254400	1.9923668639053300	59.10059171597630	0.0	-8.409518	115.188919
109	Bali	August	8.0	Lisa Chen	30.0	Female	Taiwan	Resort	1500.0	Plane	1200.0	24.68331360946750	80.31952662721890	1014.8639053254400	1.9923668639053300	59.10059171597630	0.0	-8.409518	115.188919
66	Bali	February	7.0	Tom Wilson	27.0	Male	American	Resort	2200.0	Plane	1000.0	25.641005917159800	87.92899408284020	1007.9349112426000	1.6820118343195300	99.49704142011840	0.5752071005917160	-8.409518	115.188919
127	Bali	February	8.0	Katie Johnson	33.0	Female	Canadian	Hotel	800.0	Plane	800.0	25.641005917159800	87.92899408284020	1007.9349112426000	1.6820118343195300	99.49704142011840	0.5752071005917160	-8.409518	115.188919
71	Bali	May	8.0	Sarah Brown	37.0	Female	British	Resort	1800.0	Plane	1000.0	27.15402366863910	87.66863905325440	1009.396449704140	1.6327218934911200	81.88165680473370	0.0066863905325443	-8.409518	115.188919
78	Bali	November	7.0	Amanda Chen	25.0	Female	Taiwanese	Resort	600.0	Plane	700.0	26.905976331361000	77.58579881656800	1011.1597633136100	2.0432544378698200	48.85207100591720	0.0	-8.409518	115.188919
93	Bali	April	11.0	Putra Wijaya	33.0	Male	Indonesian	Villa	1500.0	Car rental	300.0	26.041005917159800	85.81656804733730	1010.112426035500	1.5889940828402400	93.4319526627219	0.2814792899408280	-8.409518	115.188919
4	New York	August	14.0	Sarah Johnson	29.0	Female	British	Hotel	2000.0	Flight	1000.0	23.39260355029590	62.84615384615390	1016.4674556213000	5.015029585798820	44.11242603550300	0.0939053254437869	40.7128	-74.006
81	New York	August	7.0	Tom Hanks	60.0	Male	American	Hotel	1500.0	Plane	1000.0	23.39260355029590	62.84615384615390	1016.4674556213000	5.015029585798820	44.11242603550300	0.0939053254437869	40.7128	-74.006
21	New York	November	5.0	Emily Lee	27.0	Female	American	Airbnb	600.0	Bus	100.0	9.96337278106512	61.85207100591720	1021.0769230769200	4.934319526627220	19.497041420118300	0.006094674556213	40.7128	-74.006
53	New York	November	7.0	Jessica Wong	28.0	Female	Canadian	Hotel	1400.0	Flight	800.0	9.96337278106512	61.85207100591720	1021.0769230769200	4.934319526627220	19.497041420118300	0.006094674556213	40.7128	-74.006
68	New York	July	6.0	James Chen	25.0	Male	American	Vacation rental	1000.0	Plane	800.0	25.91420118343200	74.88165680473370	1012.7278106508900	4.045798816568050	32.33727810650890	0.0575147928994082	40.7128	-74.006
88	New York	October	6.0	Emily Davis	31.0	Female	American	Airbnb	700.0	Car rental	200.0	19.963254437869900	76.31952662721890	1018.6804733727800	3.959171597633140	29.02366863905330	0.0532544378698224	40.7128	-74.006
99	New York	May	6.0	Michael Wong	50.0	Male	China	Airbnb	800.0	Car rental	300.0	13.150473372781100	63.56804733727810	1008.491124260360	4.984852071005920	55.29585798816570	0.0218934911242603	40.7128	-74.006
108	New York	June	5.0	Michael Wong	42.0	Male	Hong Kong	Hotel	1200.0	Car rental	200.0	19.114911242603600	56.994082840236700	1011.5443786982200	5.005917159763310	30.325443786982200	0.0214201183431952	40.7128	-74.006
5	Tokyo	September	7.0	Kim Nguyen	26.0	Female	Vietnamese	Airbnb	700.0	Train	200.0	28.116153846153900	78.4319526627219	1009.2899408284000	5.737869822485210	62.9585798816568	0.3031360946745560	35.6895	139.6917
62	Tokyo	September	7.0	Kenji Nakamura	28.0	Male	Japanese	Hostel	400.0	Train	300.0	28.116153846153900	78.4319526627219	1009.2899408284000	5.737869822485210	62.9585798816568	0.3031360946745560	35.6895	139.6917
96	Tokyo	September	9.0	Emily Johnson	50.8	Female	Canada	Airbnb	400.0	Train	200.0	28.116153846153900	78.4319526627219	1009.2899408284000	5.737869822485210	62.9585798816568	0.3031360946745560	35.6895	139.6917
107	Tokyo	September	9.0	Sarah Lee	28.0	Female	South Korea	Airbnb	800.0	Train	500.0	28.116153846153900	78.4319526627219	1009.2899408284000	5.737869822485210	62.9585798816568	0.3031360946745560	35.6895	139.6917
20	Tokyo	October	10.0	Kenji Nakamura	45.0	Male	Japanese	Hotel	1200.0	Plane	800.0	21.357751479290000	58.44378698224850	1009.9585798816600	5.0834911242603500	53.40236686390530	0.0510650887573964	35.6895	139.6917
51	Tokyo	October	10.0	David Kim	25.0	Male	American	Hostel	500.0	Bus	100.0	21.357751479290000	58.44378698224850	1009.9585798816600	5.0834911242603500	53.40236686390530	0.0510650887573964	35.6895	139.6917
124	Tokyo	October	8.0	Henry Kim	24.0	Male	Korean	Hotel	1200.0	Plane	700.0	21.357751479290000	58.44378698224850	1009.9585798816600	5.0834911242603500	53.40236686390530	0.0510650887573964	35.6895	139.6917
41	Tokyo	November	8.0	Sarah Wong	28.0	Female	Chinese	Hostel	500.0	Plane	900.0	20.68579881656810	69.9171597633136	1017.0710059171600	4.562840236686390	38.78698224852070	0.1537278106508880	35.6895	139.6917
74	Tokyo	January	8.0	Alex Kim	29.0	Male	American	Hotel	1000.0	Train	200.0	6.5368639053254700	45.70414201183430	1017.6568047337300	4.409644970414200	28.698224852071000	0.0	35.6895	139.6917
85	Tokyo	July	10.0	Sarah Lee	28.0	Female	Korean	Airbnb	500.0	Train	300.0	27.21727810650890	74.46745562130180	1004.6153846153800	5.65431952662722	60.43195266272190	0.0955621301775147	35.6895	139.6917
117	Tokyo	February	9.0	Alice Smith	32.0	Female	American	Hotel	1000.0	Plane	700.0	7.675562130177540	52.93491124260360	1016.5562130177500	4.939408284023670	52.3905325443787	0.009171597633136	35.6895	139.6917
130	Tokyo	May	7.0	Jane Smith	28.0	Female	British	Airbnb	7000.0	Train	1500.0	19.19165680473380	64.09467455621300	1012.5739644970400	8.277869822485220	42.66272189349110	0.2363313609467460	35.6895	139.6917
6	Paris	October	5.0	Michael Brown	42.0	Male	American	Hotel	1500.0	Flight	800.0	17.316568047337300	72.42011834319530	1024.1834319526600	2.5592307692307700	10.976331360946700	0.0063905325443786	48.8566	2.3522
17	Paris	September	9.0	Sarah Johnson	30.0	Female	American	Hotel	900.0	Plane	400.0	24.242426035503000	65.24260355029590	1017.4437869822500	3.430473372781060	8.804733727810650	0.0036094674556213	48.8566	2.3522
25	Paris	November	6.0	Lily Nguyen	24.0	Female	Vietnamese	Hotel	1400.0	Train	100.0	10.47479289940830	83.16568047337280	993.7337278106510	6.803431952662730	71.5680473372781	0.0762721893491124	48.8566	2.3522
126	Paris	November	8.0	Jack Smith	28.0	Male	American	Hostel	400.0	Plane	700.0	10.47479289940830	83.16568047337280	993.7337278106510	6.803431952662730	71.5680473372781	0.0762721893491124	48.8566	2.3522
39	Paris	June	7.0	Mia Johnson	25.0	Female	American	Hotel	1400.0	Plane	600.0	19.6385207100592	57.55621301775150	1017.1420118343200	5.198343195266270	4.940828402366860	0.0	48.8566	2.3522
76	Paris	June	7.0	John Smith	46.0	Male	British	Hotel	1200.0	Plane	700.0	19.6385207100592	57.55621301775150	1017.1420118343200	5.198343195266270	4.940828402366860	0.0	48.8566	2.3522
84	Paris	June	6.0	John Smith	35.0	Male	American	Hotel	800.0	Plane	500.0	19.6385207100592	57.55621301775150	1017.1420118343200	5.198343195266270	4.940828402366860	0.0	48.8566	2.3522
95	Paris	June	5.0	John Smith	35.0	Male	American	Hotel	500.0	Plane	800.0	19.6385207100592	57.55621301775150	1017.1420118343200	5.198343195266270	4.940828402366860	0.0	48.8566	2.3522
50	Paris	August	7.0	Jennifer Nguyen	31.0	Female	Canadian	Hotel	1200.0	Train	300.0	17.483431952662700	77.9585798816568	1011.0414201183400	5.747159763313610	66.59763313609470	0.1598816568047340	48.8566	2.3522
61	Paris	July	6.0	Rachel Sanders	35.0	Female	American	Hotel	1200.0	Plane	800.0	20.071183431952700	63.248520710059200	1014.2130177514800	4.4336094674556200	35.26627218934910	0.0709467455621301	48.8566	2.3522
67	Paris	May	6.0	Olivia Green	39.0	Female	French	Hotel	1100.0	Train	200.0	15.261360946745600	77.33727810650890	1018.5798816568000	3.8505325443787	39.54437869822490	0.2316568047337280	48.8566	2.3522
106	Paris	May	5.0	John Smith	35.0	Male	American	Hotel	1000.0	Plane	800.0	15.261360946745600	77.33727810650890	1018.5798816568000	3.8505325443787	39.54437869822490	0.2316568047337280	48.8566	2.3522
129	Paris	May	6.0	John Doe	35.0	Male	American	Hotel	5000.0	Airplane	2500.0	15.261360946745600	77.33727810650890	1018.5798816568000	3.8505325443787	39.54437869822490	0.2316568047337280	48.8566	2.3522

Fig:6

v. Adaptations and Challenges:

I would say there where lot's of learning than challenges. Right from project proposal till now the plan has not been changed but the method of achieving the goal has been modified due to difficulty in extracting datasets. Getting a robust and free dataset for historical weather data with adequate weather parameters was not an easy task because many websites do not provide free historic weather data. Along with it the number of API requests where also limited so getting the necessary data in a very constrained amount of time frame was difficult. It was difficult to find the common columns and correctly align all 3 datasets in correct columns since it was a very huge dataset.But, with a solid project planning I was able to attain correct results in a very limited time frame.

Note: When you try to request API for weather dataset using the python script you might find small fluctuations in the data it's due to floating point difference while requesting and also the websites constantly refine their database for quality.

iii. Analytical Insights and Visualizations:

I have performed various interesting and exciting Analysis and Visualizations which will provide a great detail about the tourism and weather trends and it will be helpful to gain insights and plan for tourism in a more efficient manner.

i. Analysis:

a. Analysis Techniques Used:

1. Descriptive Statistics: Leveraged measures of central tendency (mean, mode, median) and dispersion (range, variance, standard deviation) to gain a nuanced understanding of traveler ages.

2. Frequency Analysis: Conducted frequency analysis to identify patterns in visit counts, revealing the most and least visited destinations, nationalities, and transportation types.

3. Location-based Analysis: Employed location-based analysis to pinpoint the hottest and coldest destinations based on temperature data.Identified destinations with the highest and lowest precipitation levels, shedding light on weather conditions.

4. Duration and Accommodation Analysis: Analyzed the duration of stays to uncover maximum and minimum durations, providing insights into traveler preferences.Explored accommodation preferences, highlighting the most favored types.

5. Cost Analysis: Delved into transportation costs, revealing the costliest and least costly transportation methods. This analysis offers implications for budget-conscious travellers .

6. Correlation Analysis: Calculated the correlation matrix to discern relationships between numerical variables. This comprehensive analysis aids in understanding the interplay of different factors in the dataset.

Findings:

1. **Traveler Patterns:**Most and least visited destinations were identified, providing a snapshot of popular and less-explored locations.

2. **Weather Conditions:**Hottest and coldest destinations, along with precipitation levels, offer valuable information for travellers planning trips based on weather preferences.

3. Travel Duration and Preferences: Maximum and minimum durations of stays shed light on varying traveler preferences, informing the tourism industry about potential target segments.

4. Cost Considerations:Insights into transportation costs help both travellers and the tourism industry make informed decisions based on budget considerations.

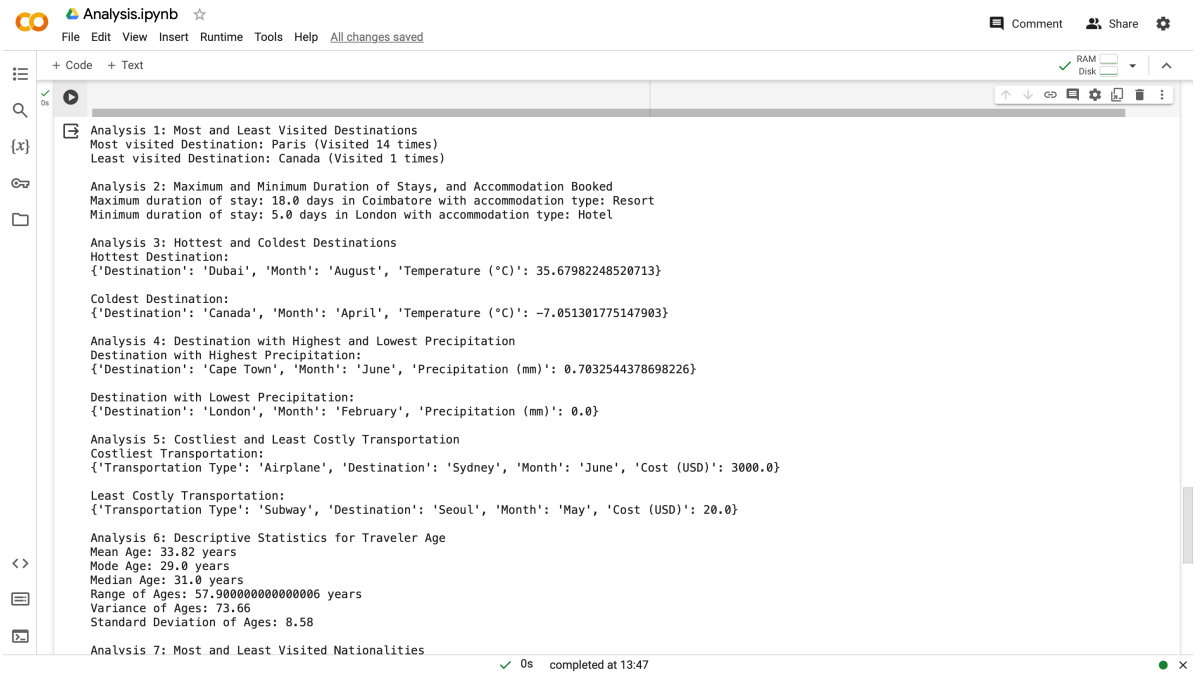
5. Demographic Insights: Descriptive statistics for traveler age provide a demographic overview, aiding in tailoring services to specific age groups.

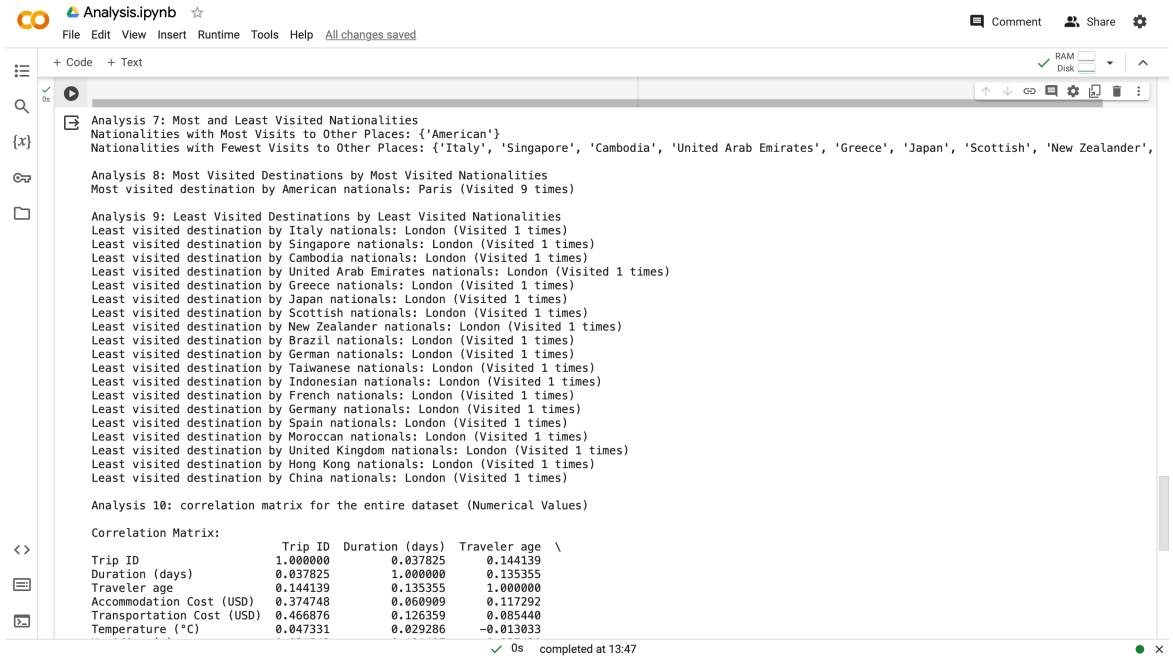
6. Nationality-based Insights: Most and least visited nationalities, along with their respective destinations, offer insights into global travel patterns.

7. Correlation Patterns: The correlation matrix unveils relationships between different variables, offering a holistic understanding of the dataset's dynamics.

In a nutshell, by using various methods to dig into the data, we've uncovered a whole bunch of useful information. This info isn't just interesting—it's gold for people involved in tourism, those making big decisions (like policymakers), and even for travellers planning their trips. These findings aren't just random facts; they're like puzzle pieces that, when put together, paint a clear picture of what's going on in the world of travel. This paints a roadmap for smarter choices and better plans for everyone involved in the travel game.

The Analysis output are displayed in the fig7 below.





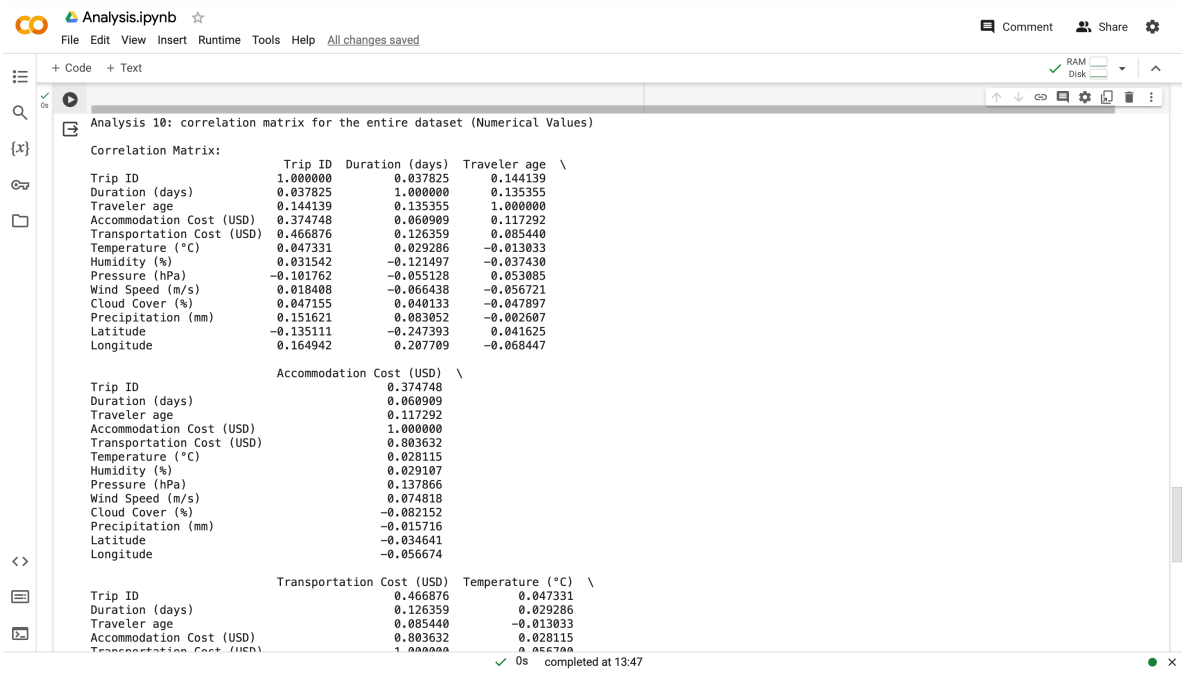


Fig:7

b. <u>Visualization:</u>

1. Best Months to Visit Most Popular Destination:

A bar plot depicting the number of visitors each month for the most visited destination.The plot allows you to identify the peak months of visitation for the most popular destination, aiding in travel planning. For instance, you might notice that more people visit during certain seasons or holidays.

2. Correlation Heat-map:

A heat-map displaying the correlation matrix of various numerical features such as duration, accommodation cost, transportation cost, and weather conditions.This visualization helps you understand the relationships between different factors. Strong positive or negative correlations indicate how changes in one variable might be associated with changes in another, providing insights into potential dependencies.

3. Temperature Distribution by Destination (Box Plot):

A box plot representing the distribution of temperatures across different destinations.This plot is valuable for comparing temperature variations between destinations. The box

indicates the interquartile range, and the whiskers show the range of temperatures. Outliers may also be visible, offering insights into potential climatic outliers.

4. Traveler Nationality Distribution (Treemap):

A treemap illustrating the distribution of travellers based on their nationality.The treemap provides a visual hierarchy of traveler nationalities, with larger sections representing a higher number of travellers from a specific country. This is useful for understanding the diversity of visitors and targeting marketing efforts.

5. Trip Duration vs. Total Cost (Line Chart):

A line chart depicting the relationship between trip duration and the total cost for various destinations.By tracking how total costs change over the duration of trips, this chart assists in budget planning. It may reveal trends, such as longer trips being associated with higher or lower overall costs.

6. Spatial Analysis and Exploration (Folium Map):

An interactive map with markers indicating destinations, displaying information like temperature, humidity, and month.This map facilitates spatial exploration, allowing you to visually inspect patterns across different locations. The pop-up information on each marker provides specific details for a more in-depth analysis of each destination.

7. Distribution of Traveler Gender (Pie Chart):

A pie chart showing the proportion of male and female travellers. This chart provides a quick overview of the gender distribution among travellers. It's a simple yet effective way to understand the demographic composition of the dataset.

<u>Visualization Figures:</u>

1.

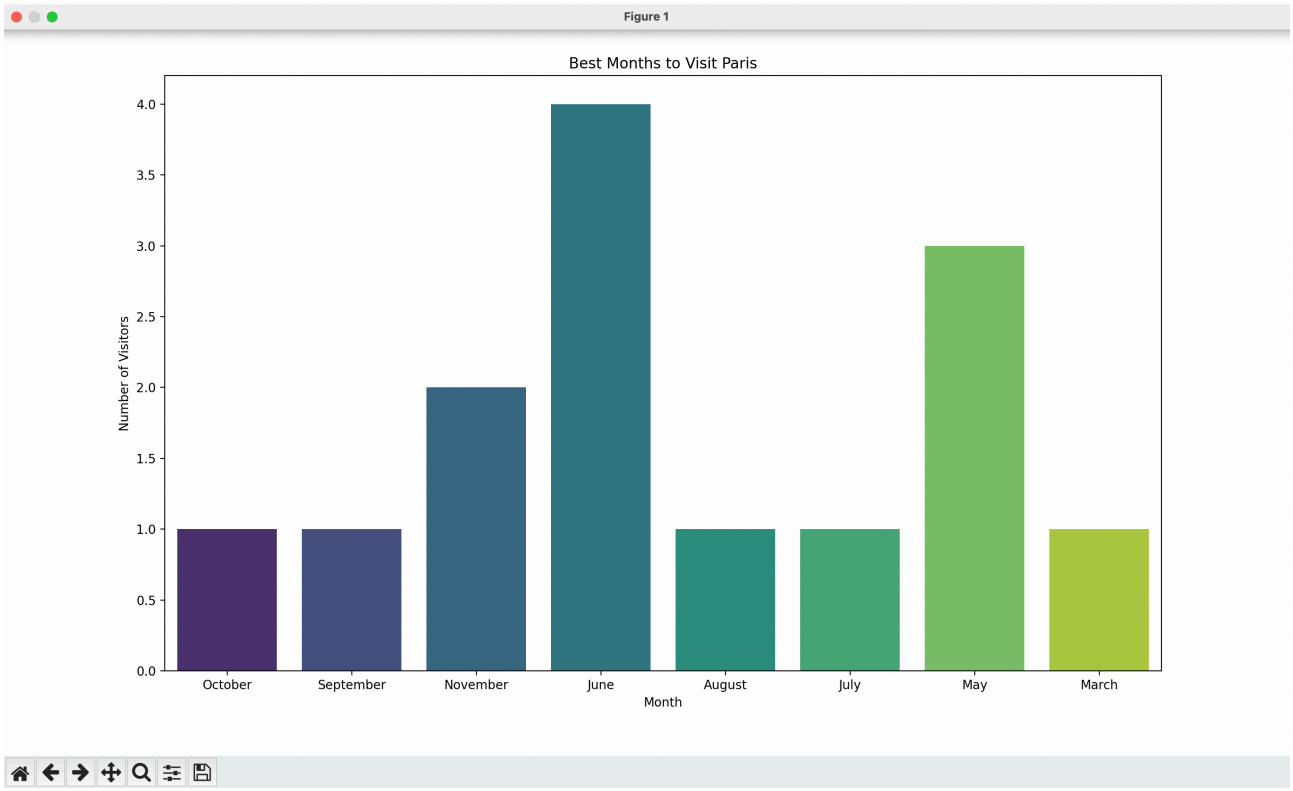


Figure 1
Best Months to Visit Paris
Number of Visitors
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
October
September
November
June
August
July
May
March
Month

2.

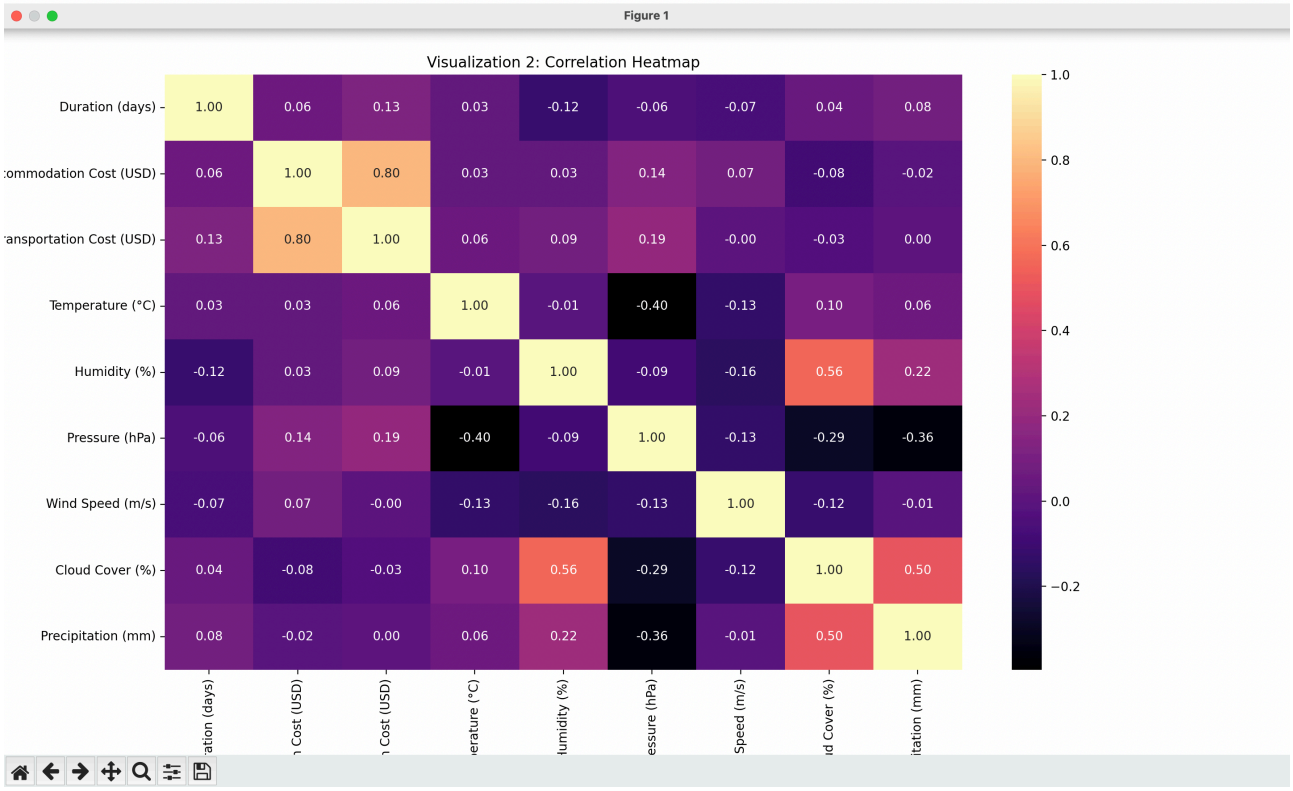
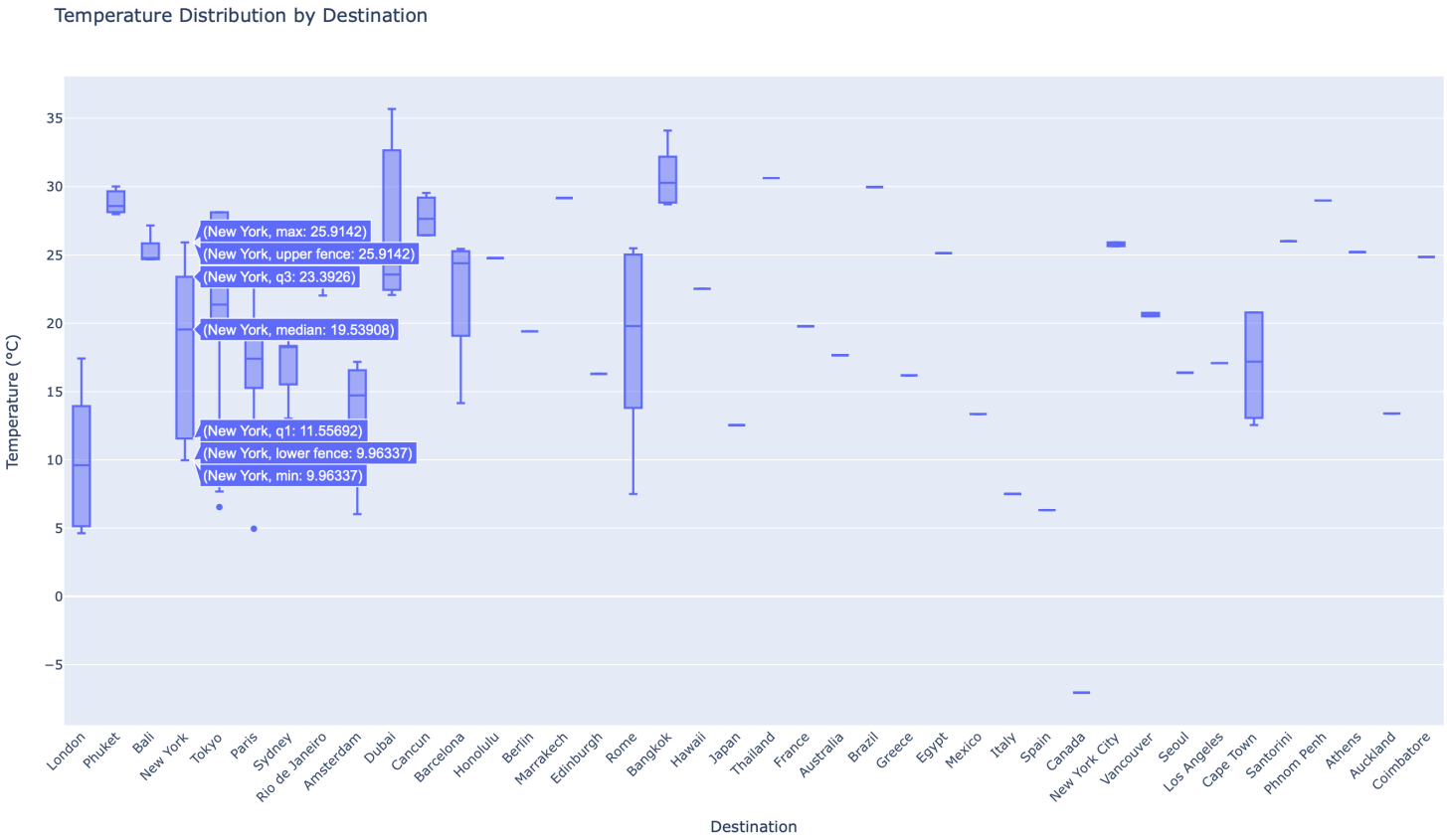


Figure 1
Visualization 2: Correlation Heatmap
Duration (days) 1.00 0.06 0.13 0.03 -0.12 -0.06 -0.07 0.04 0.08
ommodation Cost (USD) 0.06 1.00 0.80 0.03 0.03 0.14 0.07 -0.08 -0.02
ansportation Cost (USD) 0.13 0.80 1.00 0.06 0.09 0.19 -0.00 -0.03 0.00
Temperature (°C) 0.03 0.03 0.06 1.00 -0.01 -0.40 -0.13 0.10 0.06
Humidity (%) -0.12 0.03 0.09 -0.01 1.00 -0.09 -0.16 0.56 0.22
Pressure (hPa) -0.06 0.14 0.19 -0.40 -0.09 1.00 -0.13 -0.29 -0.36
Wind Speed (m/s) -0.07 0.07 -0.00 -0.13 -0.16 -0.13 1.00 -0.12 -0.01
Cloud Cover (%) 0.04 -0.08 -0.03 0.10 0.56 -0.29 -0.12 1.00 0.50
Precipitation (mm) 0.08 -0.02 0.00 0.06 0.22 -0.36 -0.01 0.50 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

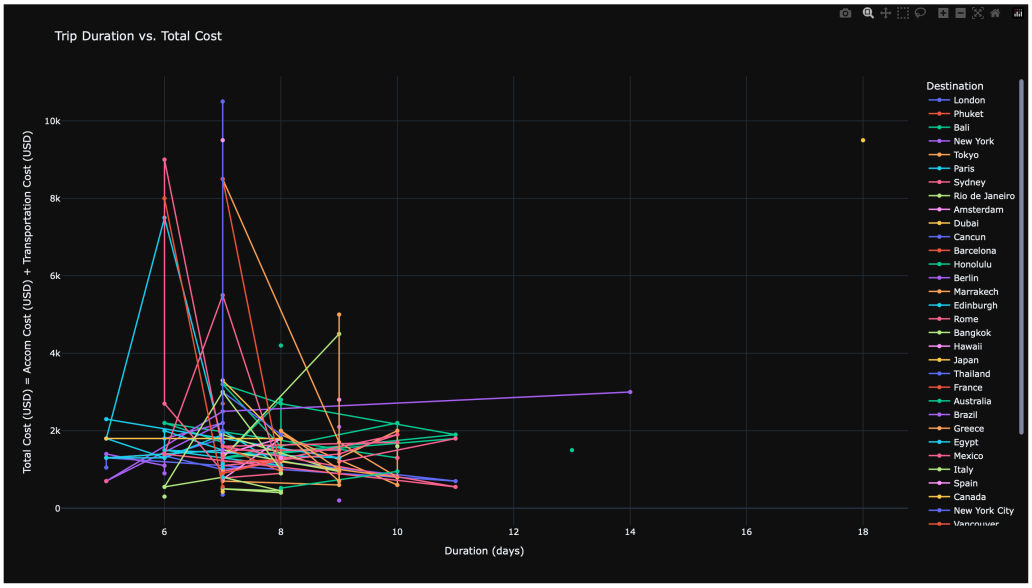
3.


Temperature Distribution by Destination
Temperature (°C)
35
30
25
20
15
10
5
0
−5
(New York, max: 25.9142)
(New York, upper fence: 25.9142)
(New York, q3: 23.3926)
(New York, median: 19.53908)
(New York, q1: 11.55692)
(New York, lower fence: 9.96337)
(New York, min: 9.96337)
London
Phuket
Bali
New York
Tokyo
Paris
Sydney
Rio de Janeiro
Amsterdam
Dubai
Cancun
Barcelona
Honolulu
Berlin
Marrakech
Edinburgh
Rome
Bangkok
Hawaii
Japan
Thailand
France
Australia
Brazil
Greece
Egypt
Mexico
Italy
Spain
Canada
New York City
Vancouver
Seoul
Los Angeles
Cape Town
Santorini
Phnom Penh
Athens
Auckland
Coimbatore
Destination

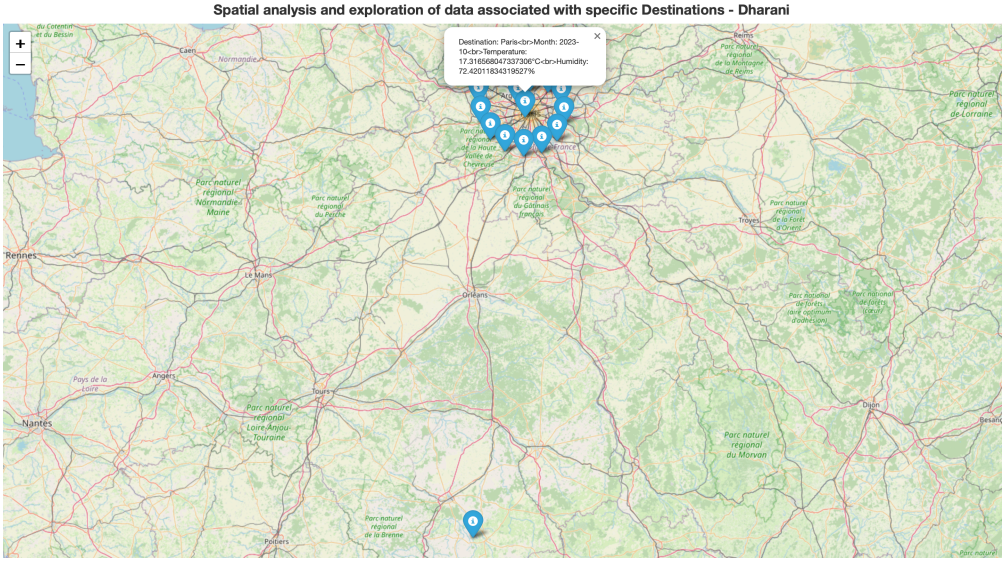
4.


Traveler Nationality Distribution
labels=American
count=25
parent=
id=American
Nationality=American
American
Korean
Canadian
Australian
Chinese
Spanish
Indian
Vietnamese
South Korean
South Korea
Brazilian
Italian
Mexican
Japanese
South African
Taiwan
Dutch
Emirati
Canada
Taiwanese
Scottish
New Zealander
China
Brazil
Indonesian
Singapore
United Kingdom
Cambodia
Spain
Moroccan
German
Japan
French
United Arab Emirates
Germany
Greece
Hong Kong
Italy

5.


Trip Duration vs. Total Cost
Total Cost (USD) = Accom Cost (USD) + Transportation Cost (USD)
Duration (days)
Destination
London
Phuket
Bali
New York
Tokyo
Paris
Sydney
Rio de Janeiro
Amsterdam
Dubai
Cancun
Barcelona
Honolulu
Berlin
Marrakech
Edinburgh
Rome
Bangkok
Hawaii
Japan
Thailand
France
Australia
Brazil
Greece
Egypt
Mexico
Italy
Spain
Canada
New York City

6.


Spatial analysis and exploration of data associated with specific Destinations - Dharani
Destination: Paris
Month: 2023-10
Temperature: 17.31656804737306°C
Humidity: 72.42011834319527%

7.

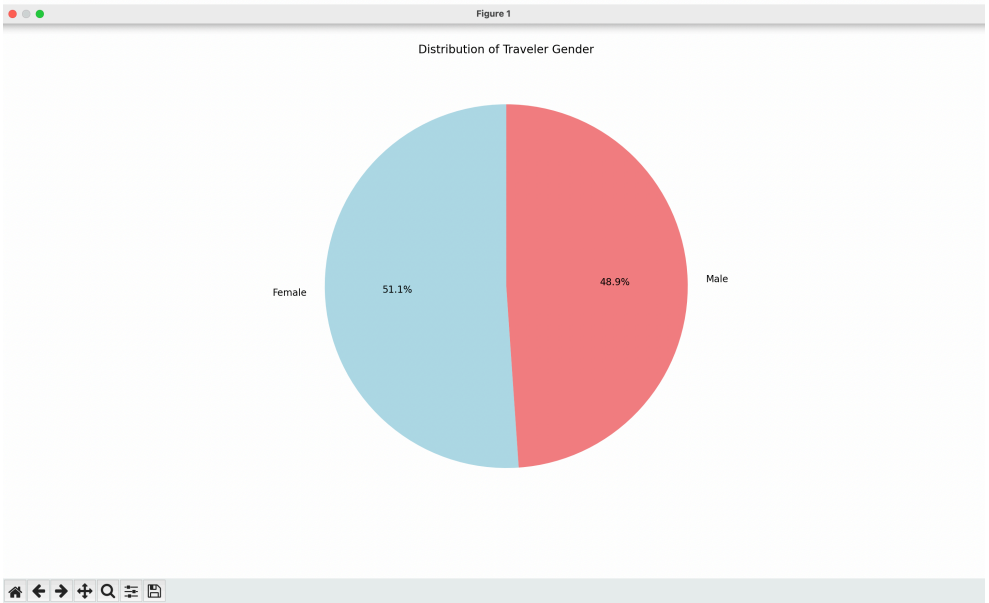


Figure 1
Distribution of Traveler Gender
Female
51.1%
48.9%
Male

Fig:8

This dynamic visualization ensemble transforms raw data into captivating insights, allowing users to seamlessly navigate through travel patterns, climate nuances, and demographic snapshots. From interactive maps that bring destinations to life to correlation heatmaps unraveling hidden associations, this empowers users with a visually immersive and informative journey through the world of tourism data. These interactive visual features can used by accessing the Visiulization.py. file in the package submitted. Please give the dates path in the data set path field in the code.

c. Observations and Conclusion:

Visualization:

The visualization part for the tourism and weather dataset is impressively comprehensive, offering a diverse array of insights through a range of visualizations, including bar plots, heat-maps, geographical maps, and treemaps. This approach ensures a holistic understanding of the data, addressing various aspects such as temporal trends, geographical context, and multivariate relationships. The Folium map with marker clusters provides a valuable spatial analysis tool, while the monthly variations in visitor count and temperature contribute to a thorough temporal analysis. The incorporation of interactive features through Plotly and Folium enhances the user-friendly representation of the data. Notably, the inclusion of a pie chart depicting traveler gender distribution adds a demographic perspective, enriching the overall comprehension of the dataset.

Analysis:

The analysis part delves deep into the numerical and statistical aspects of the dataset, offering a holistic understanding through a diverse range of analyses. From pinpointing popular destinations to examining extreme values in duration, temperature, precipitation, and transportation costs, the code provides a comprehensive overview. Notably, insights into temporal and weather patterns emerge, highlighting the hottest and coldest destinations, as well as those experiencing extreme precipitation levels. Consideration of transportation costs, distinguishing the costliest from the least costly, contributes to a nuanced understanding of budget considerations for travelers. Moreover, the code delves into traveler demographics and preferences, utilizing descriptive statistics to unravel age distribution among travelers and identifying the most and least visited nationalities. For a more tailored

understanding, Analysis 8 and Analysis 9 focus on destinations favored by the most and least visited nationalities. Lastly, the correlation matrix adds a numerical perspective, uncovering relationships and dependencies among various features in the dataset.

Conclusion:

In conclusion, the amalgamation of compelling visualizations and comprehensive statistical analyses equips users with a powerful toolkit to decipher the complexities of tourism and weather interactions. The insights derived from this approach pave the way for informed decision-making, be it in the place of travel planning. This holistic exploration lays the foundation for deeper investigations into the dynamic interplay between tourism dynamics and weather conditions.

d. Impact of Findings:

The impact of these analyses and visualizations extends beyond mere data exploration; it serves as a catalyst for informed decision-making in the realms of travel management and research.By discerning temporal trends, geographical intricacies, and traveler demographics, stakeholders are empowered to tailor offerings and services to meet the dynamic preferences of their audience. The insights derived from cost considerations and extreme values in various parameters offer a pragmatic understanding of budget constraints and exceptional scenarios. Furthermore, the correlations uncovered through the analyses provide a nuanced comprehension of interdependencies within the dataset. This holistic approach not only aids in optimizing travel-related strategies but also lays the groundwork for more targeted and effective research initiatives, fostering a continuous cycle of improvement and innovation in the tourism industry.

iv. Future Work:

With additional time, I would focus on refining and expanding the project in key areas. Firstly, I would delve into predictive modeling, implementing machine learning algorithms to forecast future travel patterns based on historical data. This could provide invaluable insights for businesses and tourism management in strategic planning. Additionally, enhancing user interaction within visualizations by incorporating features like dropdowns or sliders would offer a more personalized exploration of the dataset. I would also invest time

in advanced geospatial analysis to unveil deeper insights into the geographical distribution of tourism activities. Conducting time-series analysis on crucial variables and performing a more in-depth exploration of traveler demographics would contribute to a comprehensive understanding of temporal dynamics and visitor characteristics. Moreover, integrating user feedback mechanisms, optimizing code for efficiency, and exploring collaboration opportunities with industry stakeholders would ensure the project's relevance and real-world impact. I aim to craft an interactive travel dashboard. Users can effortlessly select their preferred destination, gaining instant access to tailored analyses. The dashboard will offer insights on optimal travel months, weather patterns, and personalized details, revolutionizing the travel planning experience.