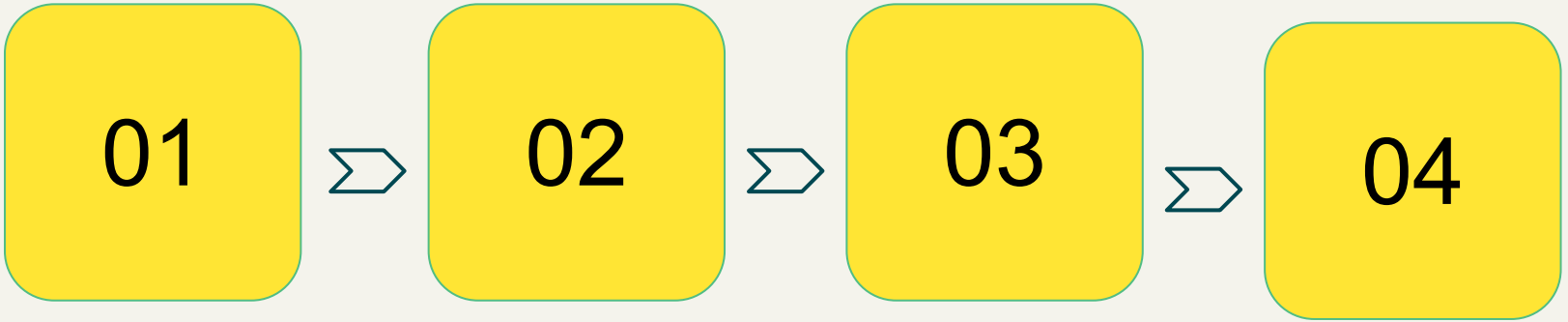
BY FADEL AHMAD FADHILAH

EXPLORING TRAVELER TRIP DATASET

Github: https://github.com/fadelaf/Portfolio---Traveler-Trip-Dataset

Outline


01
02
03
04
Data Introduction
Data Cleaning & Pre Processing
EDA and Data Visualization
Business Insight

01: Introduction

Traveler Trip Dataset

The travel dataset provides detailed information on various trips taken by travelers, including their destination, travel dates, duration of the trip in days, traveler demographics (name, age, gender, and nationality), as well as the type and cost of accommodation and transportation.
This dataset can be used to gain insights into travel patterns, preferences, and behaviors of different types of travelers.

Data consist of 139 data and 13 columns

Data Columns

- **Trip ID:** A unique identifier for each trip taken by a traveler.
- **Destination:** The name of the city or country visited by the traveler.
- **Start date:** The date the traveler started the trip.
- **End date:** The date the traveler ended the trip.

Duration (days): The number of days the traveler spent on the trip.

- **Traveler name:** The name of the traveler.
- **Traveler age:** The age of the traveler at the time of the trip.
- **Traveler gender:** The gender of the traveler.
- **Transportation cost:** The cost of transportation for the entire trip.

- **Traveler nationality:** The nationality of the traveler.
- **Accommodation type:** The type of accommodation the traveler stayed in, such as hotel, hostel, or Airbnb.
- **Accommodation cost:** The cost of the accommodation for the entire trip.
- **Transportation type:** The mode of transportation used by the traveler, such as plane, train, or car.

02: Data Cleaning and Preprocessing


```
### Columns that need to be converted ###

Traveler gender           object  -> object(string or mix)
Traveler nationality      object  -> object(string or mix)
Accommodation type        object  -> object(string or mix)
Accommodation cost        object  -> float
Transportation type       object  -> object(string or mix)
Transportation cost       object  -> float
Travel age                object  -> int
Duration (days)           float64 -> int
Start date                object  -> datetime
End date                  object  -> datetime
```

```
Columns name:  Traveler nationality

['American' 'Canadian' 'Korean' 'British' 'Vietnamese' 'Australian'
 'Brazilian' 'Dutch' 'Emirati' 'Mexican' 'Spanish' 'Chinese' 'German'
 'Moroccan' 'Scottish' 'Japanese' 'Italian' 'Indian' 'South Korean'
 'French' nan 'South African' 'Taiwanese' 'Indonesian' 'USA' 'Canada'
 'South Korea' 'UK' 'China' 'Taiwan' 'Japan' 'Spain' 'Brazil' 'Germany'
 'Hong Kong' 'United Kingdom' 'Singapore' 'Italy' 'Greece'
 'United Arab Emirates' 'Cambodia' 'New Zealander']
```

```
Counting Null Values

Trip ID                 0
Destination             2
Start date              2
End date                2
Duration (days)         2
Traveler name           2
Traveler age            2
Traveler gender         2
Traveler nationality    2
Accommodation type      2
Accommodation cost      2
Transportation type     3
Transportation cost     3
dtype: int64
```

Converting Data Types:

Converting data types from object; (string or mix) to other type. Example:
Numerical to integer and float
Date string to datetime

Handling Inconsistent Data:

Handling to inconsistent data that has the same value but in different form: example:
There are “Japan” and “Japanese”

Handling Missing Value:

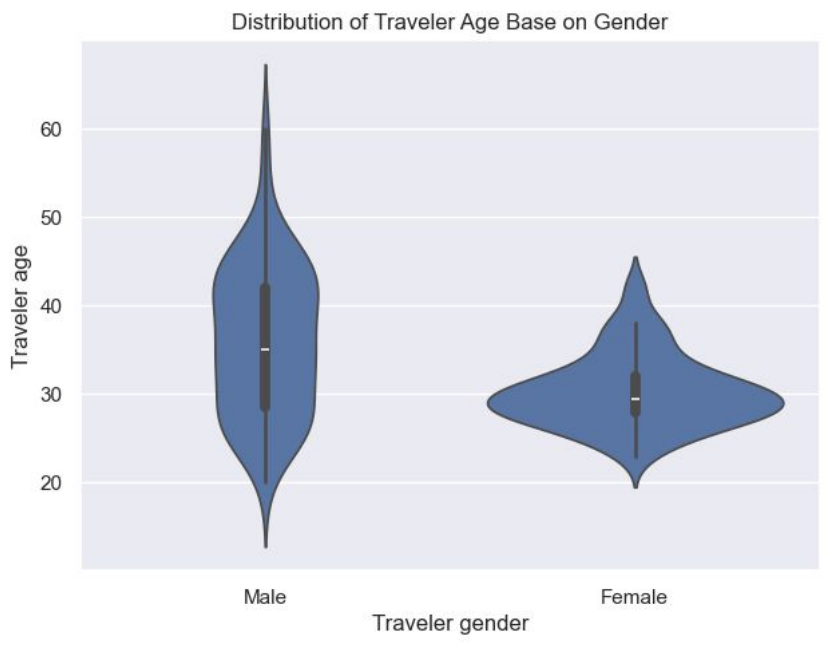
From 139 data, I remove 2 rows which poses nothing (all columns data are missing). While other, I fill it with the majority base on destination group

EDA and Data Visualization

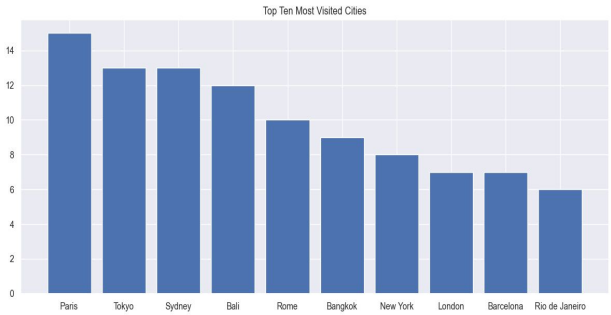
Distribution of Traveler Age Base on Gender

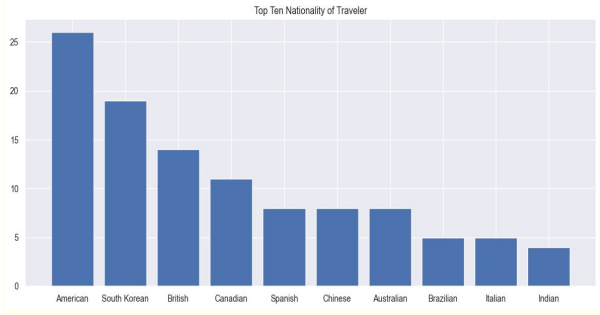
The violin chart gives information about the distribution of traveler age for Male and Female.

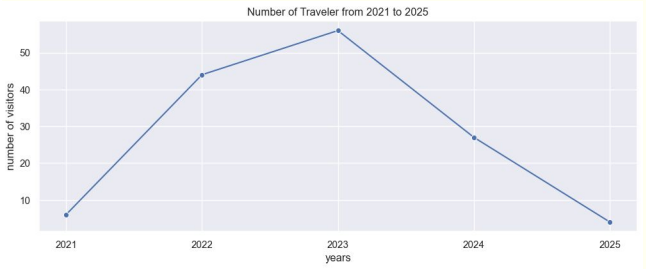
It is obvious that males poses more variance age than females, from under 20 until more than 60, with mean 33 years old. Meanwhile, females centered most in around 30, with minimum and maximum age approximately 23 and 42 respectively.



EDA and Data Visualization







Top Ten Most Visited Cities

The bar chart displays 10 most visited cities by traveler. Paris ranked 1 with 15 visitors, followed by Tokyo and Sydney, both have 13. Bali as the third, 12 visitors

Top Ten Nationality of Traveler

The graph shows that 25 traveler are American, which is the highest. South Korean ranked second with nearly 20. Followed by British with 11 traveler.

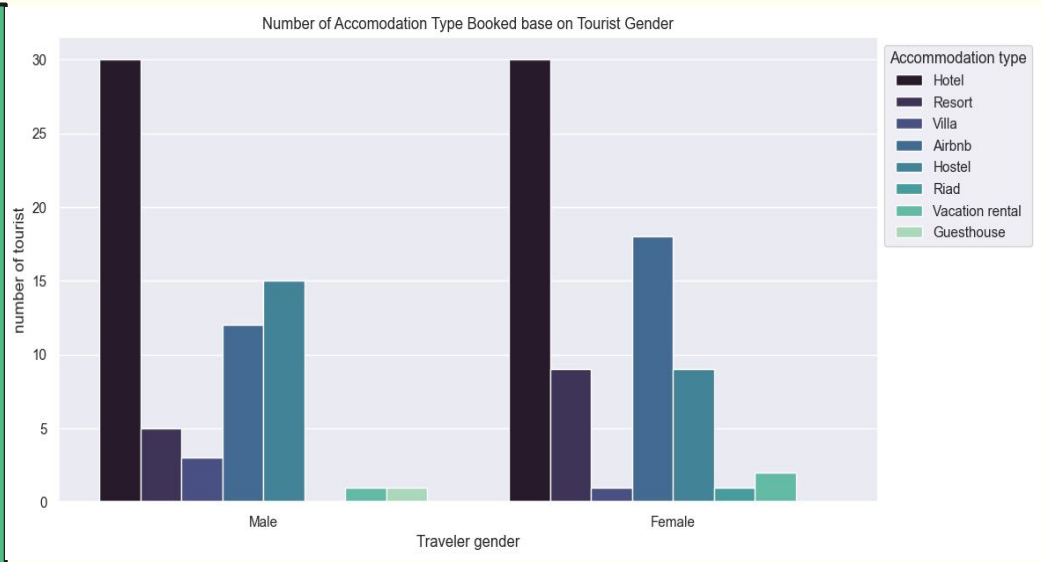
Number of Traveler from 2021 to 2025

It is evident, displayed by the line chart, number of traveler trend are fluctuated. Starting from 2021 which had less than 10 visitors climbed up to reach its peak in 2023 with more than 50. However, it extremely dropped until the lowest in 2025.

EDA and Data Visualization

Number of Accommodation Type Booked base on Gender

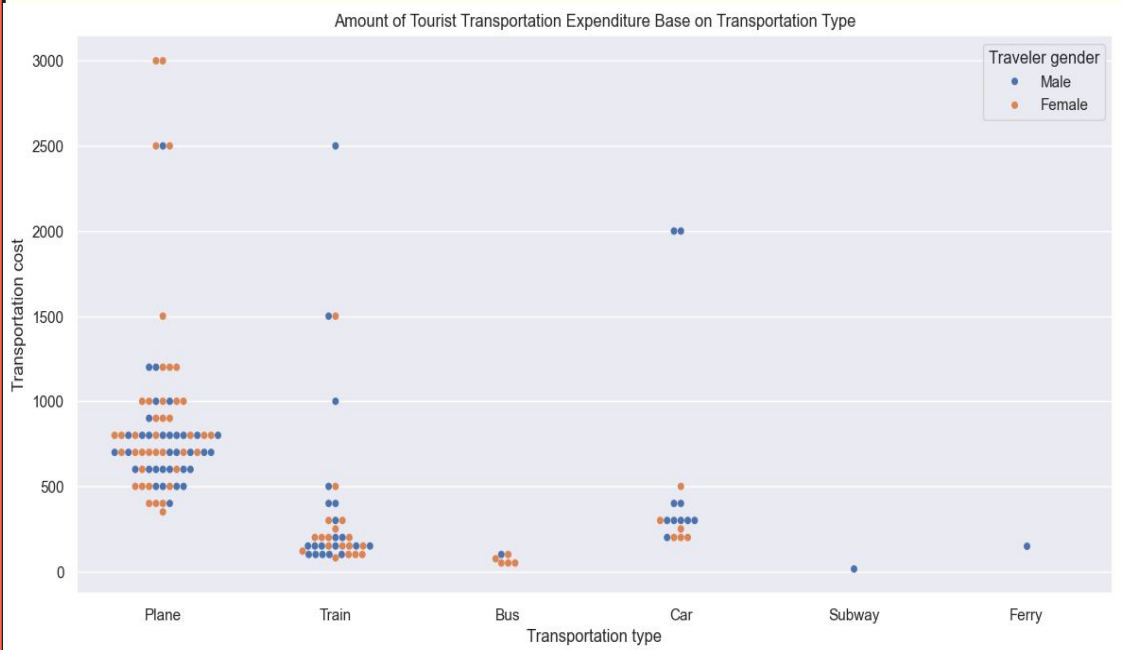
The bar chart gives information concerning preferable accommodation type by gender. It is apparent that male and female prefer hotel the most. Furthermore, male are likely to choose hostel, instead of Airbnb. In contrast with females, they tend to choose Airbnb over Hostel.



EDA and Data Visualization

Amount of Tourist Transportation Expenditure Base on Transportation Type

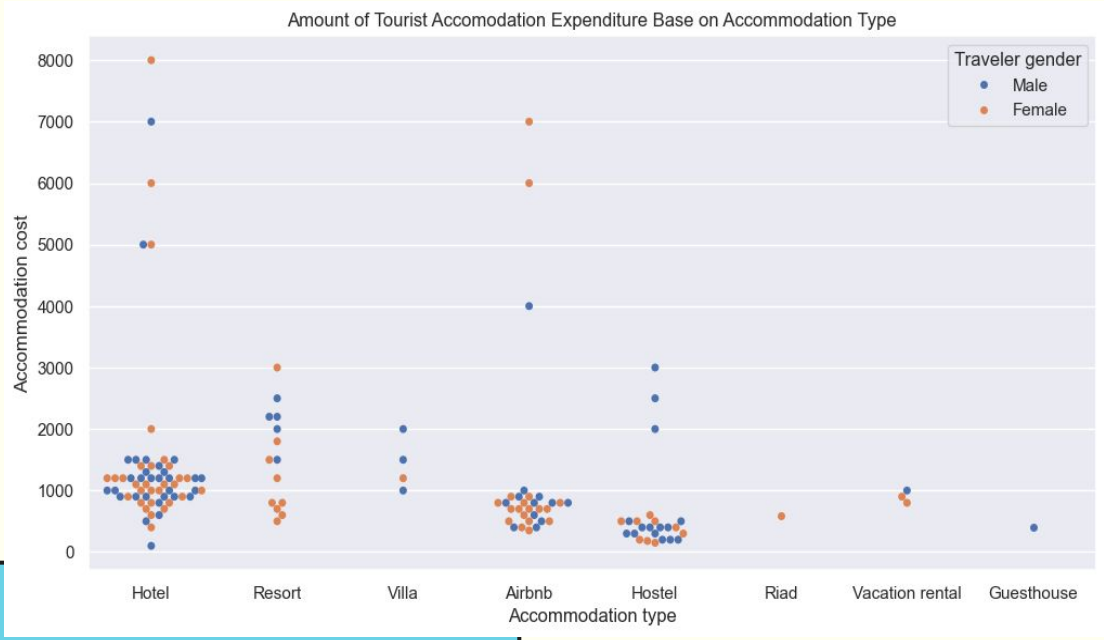
It is evident from the scatter plot given that majority of the tourist are likely to take plane to travel, although it higher has higher cost among all. The price level of train and bus almost in the same level, but train is more preferred. Meanwhile, there are also travelers riding car.



EDA and Data Visualization

Amount of Tourist Accommodation Expenditure Base on Accommodation Type

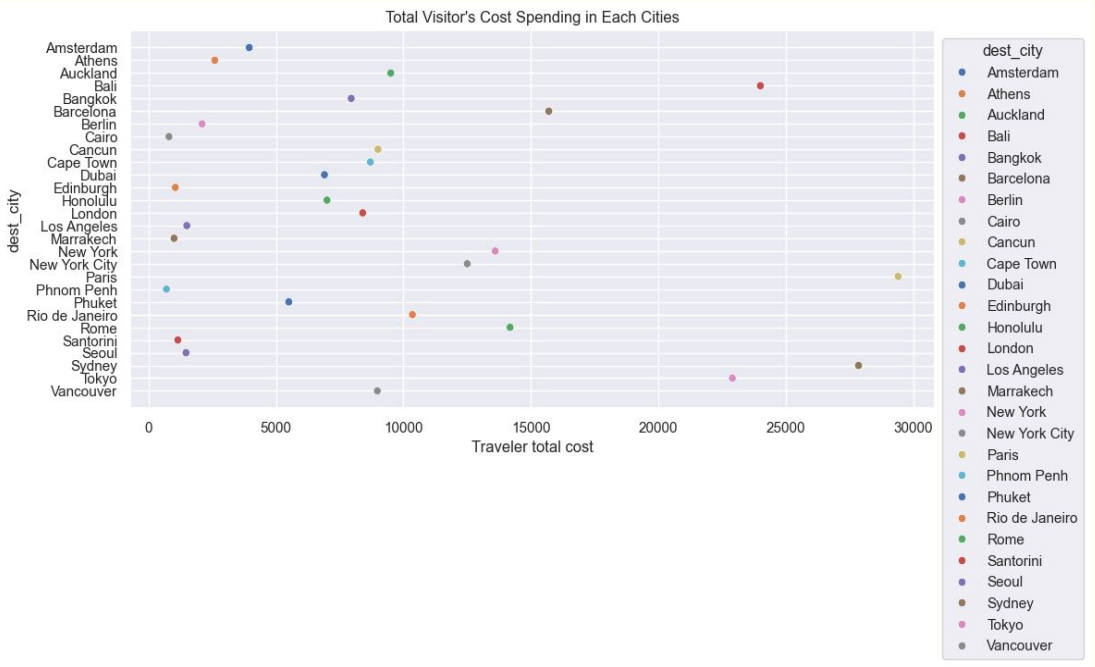
The graph shows hotel are the most popular place to stay, with median cost in 1100. Airbnb's average in 750 and hostel in 400. Another highlight, cost for resort has more variances.



EDA and Data Visualization

Total Traveler's Cost Spending in Each CIties

We can highlight that traveler are willing to spend money to nearly 30,000 while in Paris. Followed by Sydney and Bali. In contrast, Pnom Penh and Cairo are city those traveler spend the least.



EDA and Data Visualization

Relationship between Accommodation Cost and Transportation Cost

The graph shown there is a positive relationship between the level of accommodation cost and transportation cost. Higher accommodation cost tend to have higher transportation cost.


```
# Uji Korelasi Pearson

test_pearson = np.corrcoef(traveler_data['Accommodation cost'],traveler_data['Transportation cost'])
print(test_pearson)
```

```
[[1.         0.78732028]
 [0.78732028 1.        ]]
```

Business Insight: Recommendations

- Since Paris known as the famous destination for traveler, travel agent could promote or make various trip packages to Paris.
- Travel agent could target some niche market, for example, offering trip package for South Korean with tour guide in their language.
- Property owner could promote their rooms targeting female by using Airbnb or build some hostel for male niche.