

PYTHON FOR

DATA SCIENCE

U S I N G



WITH EUROPE HOTEL BOOKING
SATISFACTION DATASET

By Muhammad Fikri Riyanto

TABLE OF CONTENT

DATASET

IMPORT, INFO, DESCRIBE

CODE & VISUALIZATION

Here is an quick overview of the dataset i will be working with

Europe Hotel Booking Satisfaction Score.CSV

id	Gender	Age	purpose_of	Type of Tra	Type Of Booki	Hotel wifi s	Departure/	Ease of Onl	Hotel locati	Food and d	Stay comfo	Common R	Checkin/Cl	Other servi	Cleanliness	satisfaction
70172	Male	13	aviation	Personal Travel	Not defined	3	4	3	1	5	5	5	4	5	5	neutral or dissatisfied
5047	Male	25	tourism	Group Travel	Group bookings	3	2	3	3	1	1	1	1	4	1	neutral or dissatisfied
110028	Female	26	tourism	Group Travel	Group bookings	2	2	2	2	5	5	5	4	4	5	satisfied
24026	Female	25	tourism	Group Travel	Group bookings	2	5	5	5	2	2	2	1	4	2	neutral or dissatisfied
119299	Male	61	aviation	Group Travel	Group bookings	3	3	3	3	4	5	3	3	3	3	satisfied
111157	Female	26	business	Personal Travel	Individual/Couple	3	4	2	1	1	1	1	4	4	1	neutral or dissatisfied
82113	Male	47	academic	Personal Travel	Individual/Couple	2	4	2	3	2	2	2	3	5	2	neutral or dissatisfied
96462	Female	52	aviation	Group Travel	Group bookings	4	3	4	4	5	5	5	4	5	4	satisfied
79485	Female	41	tourism	Group Travel	Group bookings	1	2	2	2	4	3	1	4	1	2	neutral or dissatisfied
65725	Male	20	academic	Group Travel	Individual/Couple	3	3	3	4	2	3	2	4	3	2	neutral or dissatisfied
34991	Female	24	academic	Group Travel	Individual/Couple	4	5	5	4	2	2	2	3	5	2	neutral or dissatisfied
51412	Female	12	tourism	Personal Travel	Not defined	2	4	2	2	1	1	1	5	5	1	neutral or dissatisfied
98628	Male	53	tourism	Group Travel	Individual/Couple	1	4	4	4	1	1	1	4	4	1	neutral or dissatisfied
83502	Male	33	academic	Personal Travel	Individual/Couple	4	2	4	3	4	4	4	2	2	4	satisfied
95789	Female	26	aviation	Personal Travel	Individual/Couple	3	2	3	2	2	2	2	2	1	2	neutral or dissatisfied
100580	Male	13	personal	Group Travel	Individual/Couple	2	1	2	3	4	1	4	1	3	4	neutral or dissatisfied
71142	Female	26	business	Group Travel	Group bookings	3	3	3	3	4	4	4	5	4	4	satisfied
127461	Male	41	tourism	Group Travel	Group bookings	4	4	2	4	4	4	5	3	5	5	satisfied
70354	Female	45	academic	Group Travel	Group bookings	4	4	4	4	3	5	5	3	5	4	satisfied
66246	Male	38	tourism	Personal Travel	Individual/Couple	2	3	3	2	5	5	5	3	2	5	neutral or dissatisfied
39076	Male	9	personal	Group Travel	Individual/Couple	2	4	2	4	2	1	2	4	3	2	neutral or dissatisfied
22434	Female	17	tourism	Personal Travel	Individual/Couple	3	1	3	3	5	5	5	3	4	5	neutral or dissatisfied
43510	Female	43	business	Personal Travel	Individual/Couple	3	5	3	5	5	5	3	3	3	4	neutral or dissatisfied
114090	Female	58	tourism	Personal Travel	Individual/Couple	4	5	4	5	4	4	4	2	4	2	neutral or dissatisfied
105420	Female	23	personal	Group Travel	Individual/Couple	5	0	5	1	1	1	1	3	5	1	satisfied
102956	Male	57	personal	Personal Travel	Individual/Couple	4	4	4	1	5	5	5	4	5	5	neutral or dissatisfied
18510	Female	33	personal	Group Travel	Group bookings	1	1	1	1	1	3	4	5	4	2	satisfied
14925	Female	49	business	Group Travel	Not defined	4	4	4	4	2	1	4	2	4	2	satisfied
118319	Female	36	tourism	Group Travel	Group bookings	3	1	1	1	1	1	3	2	3	2	neutral or dissatisfied
75460	Male	22	business	Personal Travel	Individual/Couple	3	2	3	3	3	1	3	4	2	3	neutral or dissatisfied
48492	Female	31	personal	Group Travel	Group bookings	4	4	4	4	5	5	5	1	5	5	satisfied
27809	Female	15	academic	Group Travel	Individual/Couple	2	2	2	3	5	5	5	2	4	5	neutral or dissatisfied
70594	Female	35	academic	Group Travel	Group bookings	4	5	4	4	4	4	3	4	3	4	satisfied
30089	Female	67	academic	Personal Travel	Individual/Couple	4	5	4	1	2	5	5	5	5	5	neutral or dissatisfied
58779	Male	37	tourism	Group Travel	Group bookings	3	3	3	4	1	1	1	1	4	1	neutral or dissatisfied
79659	Female	40	aviation	Group Travel	Individual/Couple	1	4	4	4	1	1	1	3	3	1	neutral or dissatisfied
110293	Female	34	academic	Group Travel	Group bookings	3	4	4	3	5	2	5	4	5	5	neutral or dissatisfied
18011	Male	40	personal	Personal Travel	Not defined	4	2	4	2	2	2	2	2	4	2	neutral or dissatisfied

LOAD DATA

```
import pandas as pd
import seaborn as sn
import matplotlib.pyplot as plt

df = pd.read_csv('Europe Hotel Booking Satisfaction Score.csv')
df
```

- 1.Importing 3 libraries : pandas, seaborn, and matplotlib.pyplot.
- 2.CSV file named “Europe Hotel Booking Satisfaction Score.CSV” is read and stored as a dataframe using pandas.
- 3.Finally, displaying the dataframe containing Europe Hotel Booking Satisfaction Score data.

	id	Gender	Age	purpose_of_travel	Type of Travel	Type Of Booking	Hotel wifi service	Departure/Arrival convenience	Ease of Online booking	Hotel location	Food and drink	Stay comfort	en
0	70172	Male	13	aviation	Personal Travel	Not defined	3	4	3	1	5	5	
1	5047	Male	25	tourism	Group Travel	Group bookings	3	2	3	3	1	1	
2	110028	Female	26	tourism	Group Travel	Group bookings	2	2	2	2	5	5	
3	24026	Female	25	tourism	Group Travel	Group bookings	2	5	5	5	2	2	
4	119299	Male	61	aviation	Group Travel	Group bookings	3	3	3	3	4	5	
...	
103899	94171	Female	23	business	Group Travel	Individual/Couple	2	1	2	3	2	2	
103900	73097	Male	49	tourism	Group Travel	Group bookings	4	4	4	4	2	5	
103901	68825	Male	30	tourism	Group Travel	Group bookings	1	1	1	3	4	5	

DISPLAY INFO DATA

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103904 entries, 0 to 103903
Data columns (total 17 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   id                                         103904 non-null  int64
1   Gender                                    103904 non-null  object
2   Age                                        103904 non-null  int64
3   purpose_of_travel                        103904 non-null  object
4   Type of Travel                           103904 non-null  object
5   Type Of Booking                          103904 non-null  object
6   Hotel wifi service                       103904 non-null  int64
7   Departure/Arrival convenience            103904 non-null  int64
8   Ease of Online booking                   103904 non-null  int64
9   Hotel location                           103904 non-null  int64
10  Food and drink                           103904 non-null  int64
11  Stay comfort                             103904 non-null  int64
12  Common Room entertainment                103904 non-null  int64
13  Checkin/Checkout service                 103904 non-null  int64
14  Other service                            103904 non-null  int64
15  Cleanliness                             103904 non-null  int64
16  satisfaction                             103904 non-null  object
dtypes: int64(12), object(5)
memory usage: 13.5+ MB
```

Based on information, there are 17 columns and 103904 rows. Most columns contain integers, except for 'Gender', 'purpose_of_travel', 'Type of Travel', 'Type Of Booking', and 'satisfaction', which are strings.

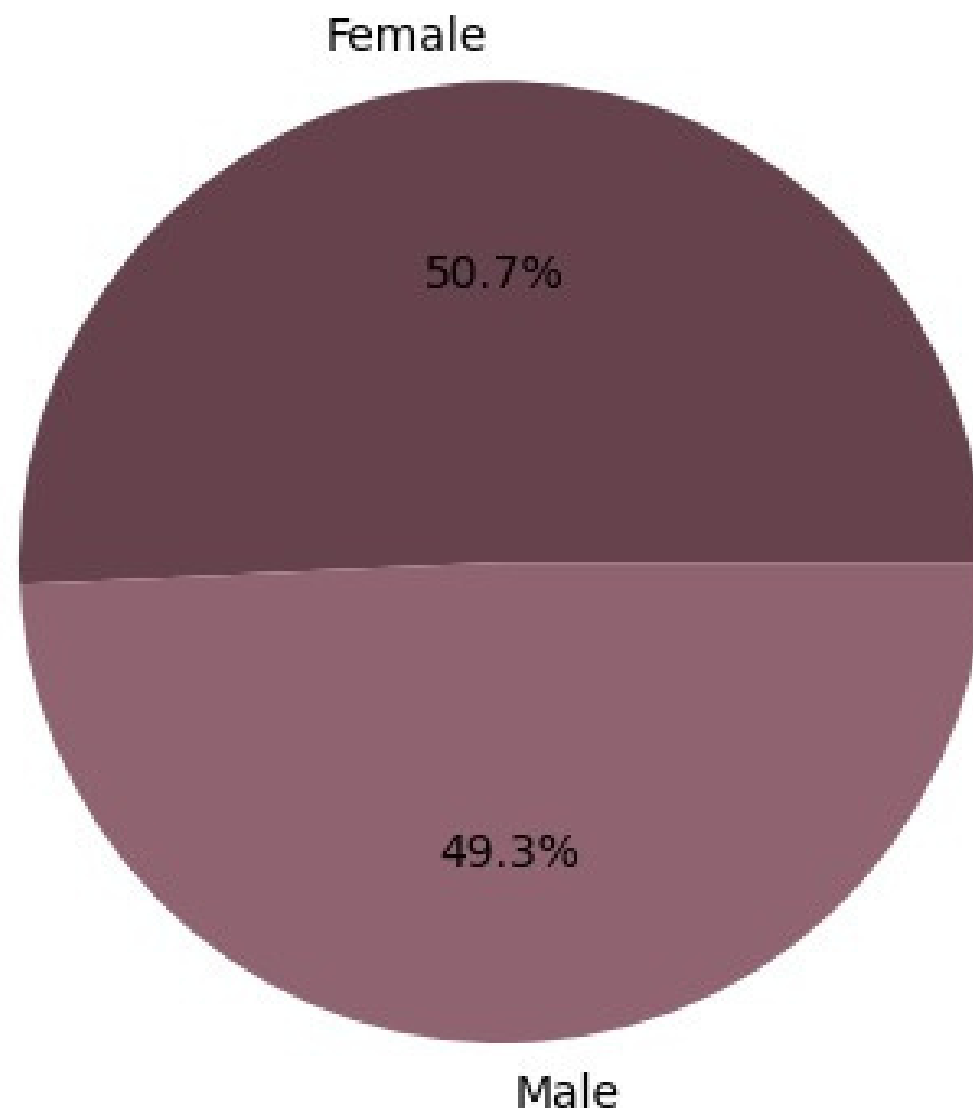
```
df.describe()
```

[illegible]

1

WHAT IS THE
PERCENTAGE DIFFERENCE
BETWEEN MALE AND FEMALE??

Customer gender comparison



```
gender_counts = df['Gender'].value_counts()

#Defining the colors
warna = ['#66424D', '#8F6370']

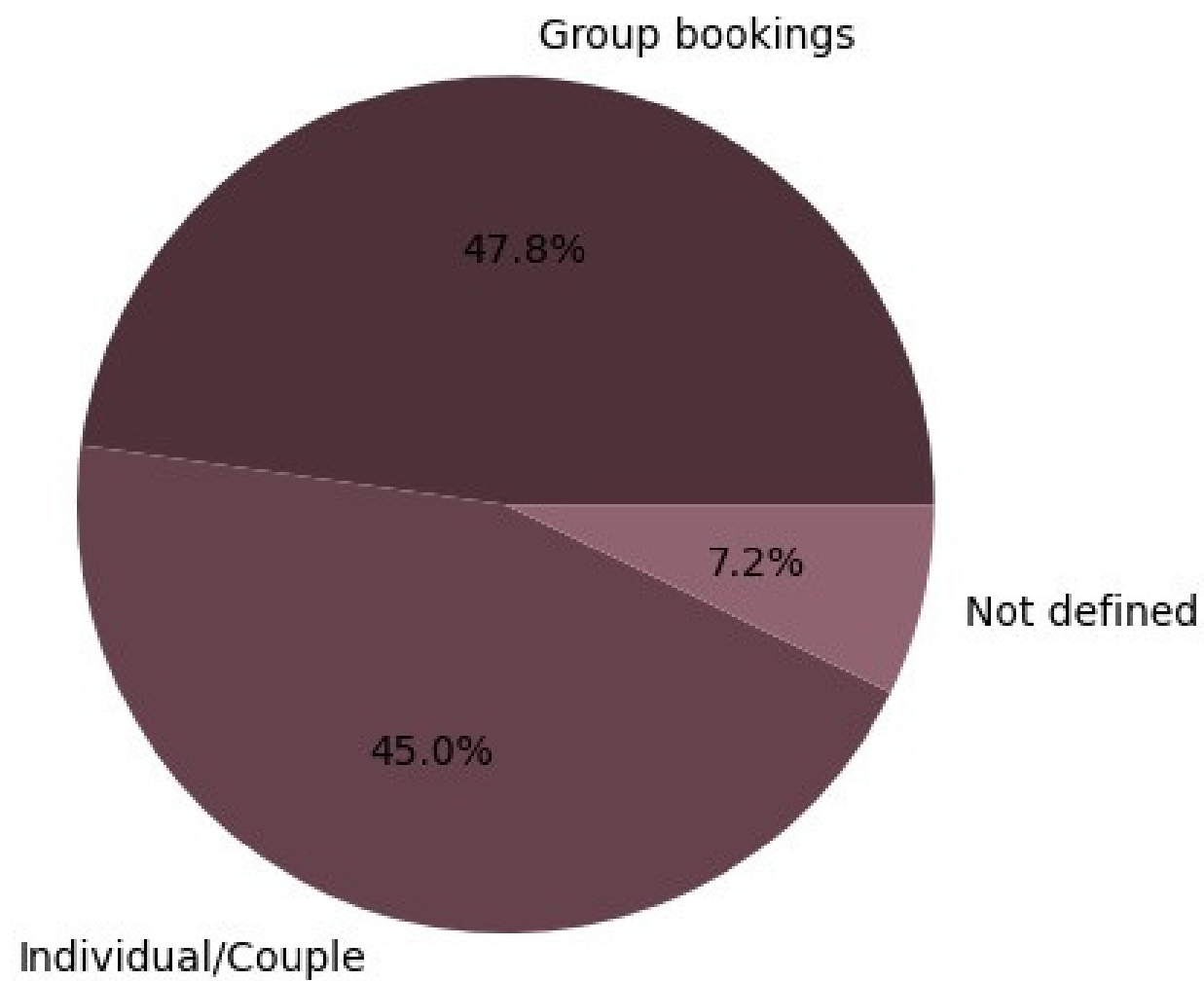
# Making chart pie
plt.pie(gender_counts, labels = gender_counts.index,
        colors = warna, autopct='%1.1f%%')
plt.title('Customer gender comparison')
plt.show()
```

50.7 % of customers are female while **49.3%** of customers are male.

2

WHAT IS THE
PERCENTAGE
DIFFERENCE TYPE OF
BOOKING?

Type Of Booking comparison



```
Type_Of_Booking_counts = df['Type Of Booking'].value_counts()

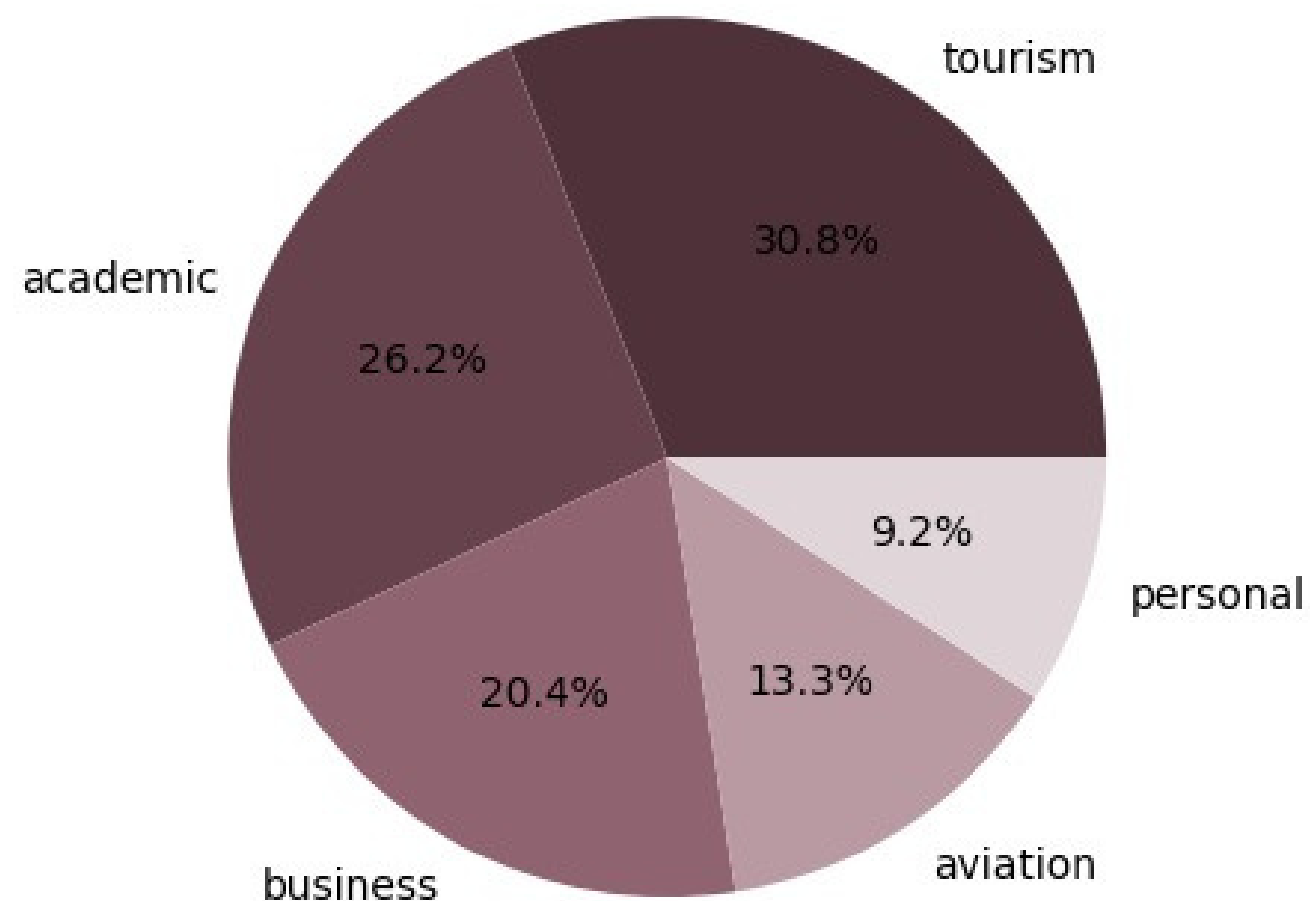
# Making chart pie
plt.pie(Type_Of_Booking_counts, labels = Type_Of_Booking_counts.index,
        colors = warna2, autopct='%1.1f%%')
plt.title('Type Of Booking comparison')
plt.show()
```

47.8% of the type of bookings are group bookings and **45.0%** of the type of bookings are individual/couple while **7.2%** are not defined

3

WHAT IS THE
PERCENTAGE
DIFFERENCE PURPOSE
OF TRAVEL?

purpose of travel comparison



```
purpose_of_travel_counts = df['purpose_of_travel'].value_counts()

# Defining colors
warna2 = ['#4F3139', '#66424D', '#8F6370', '#B99AA3', '#E0D6D9']

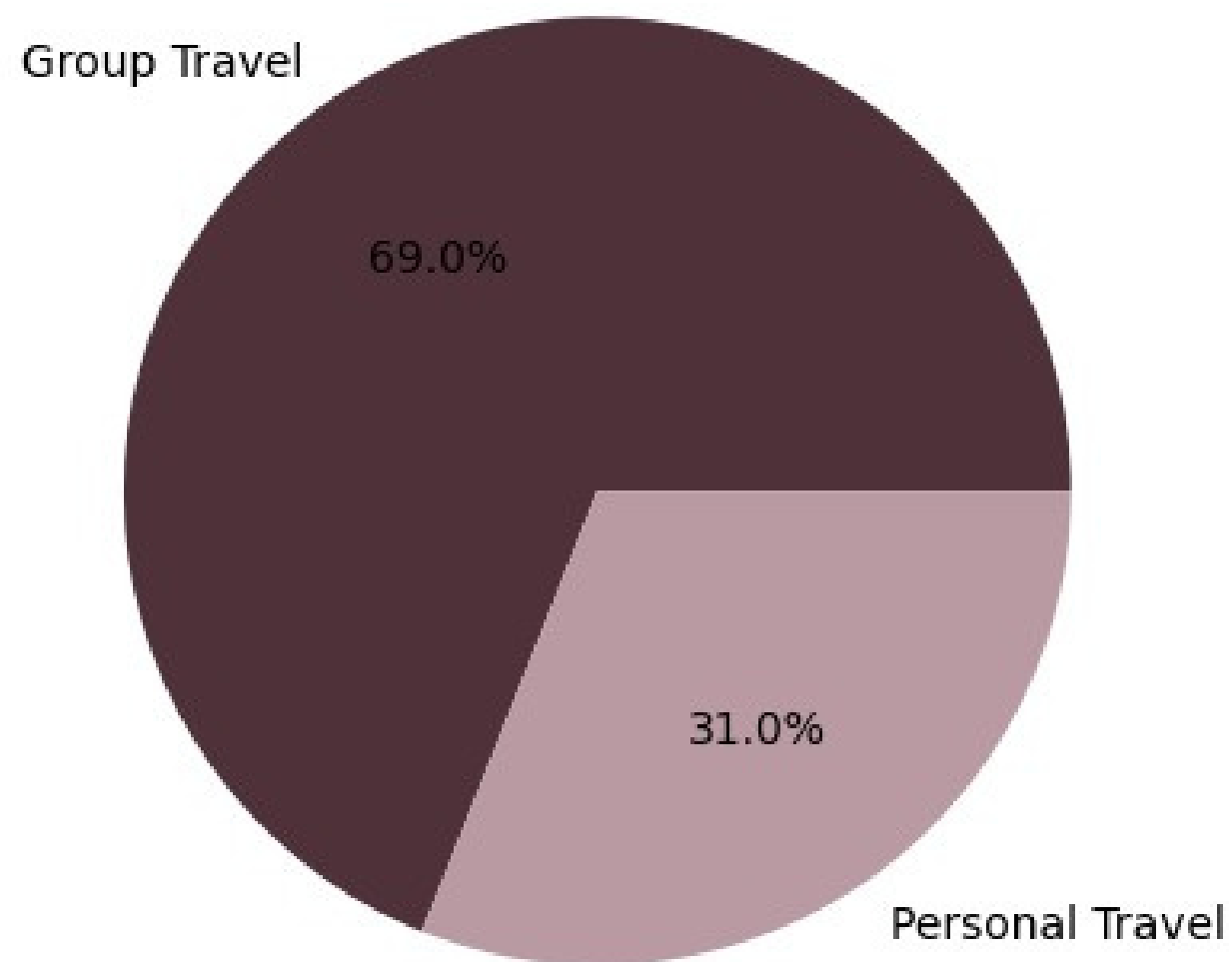
# Making chart pie
plt.pie(purpose_of_travel_counts, labels = purpose_of_travel_counts.index,
        colors = warna2, autopct='%1.1f%%')
plt.title('purpose of travel comparison')
plt.show()
```

30.8 of the purpose of the travel is tourism. While for academic it is **26.2%**, for business it is **20.4%**, for aviation it is **13.3%** and for personal it is only **9.2%**.

4

WHAT IS THE
PERCENTAGE
DIFFERENCE TYPE OF
TRAVEL?

Type of Travel comparison



```
Type_of_Travel_counts = df['Type of Travel'].value_counts()

# Making chart pie
plt.pie(Type_of_Travel_counts, labels = Type_of_Travel_counts.index,
        colors = warna1, autopct='%1.1f%%')
plt.title('Type of Travel comparison')
plt.show()
```

50.7 % of customers are female while **49.3%** of customers are male.

5

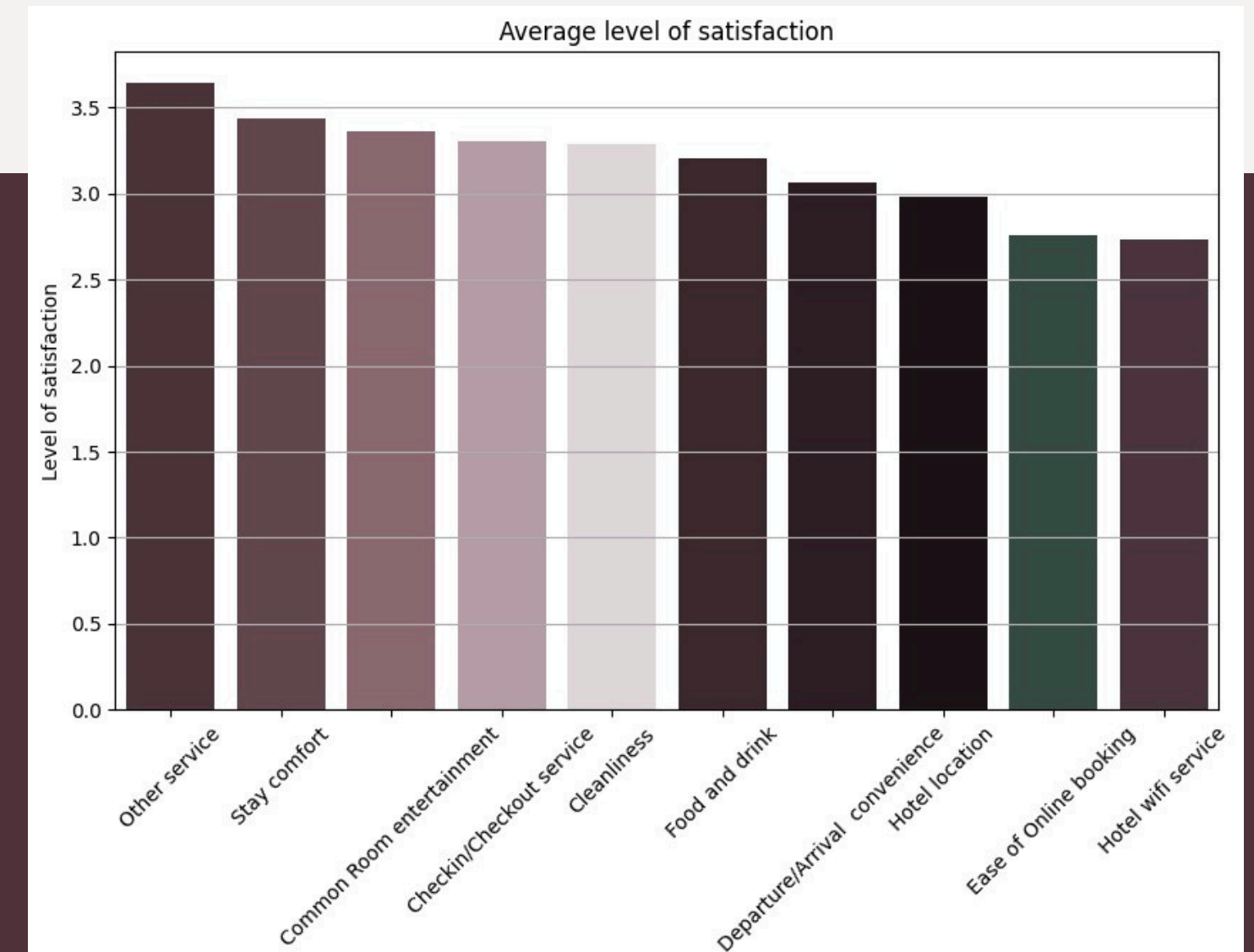
WHAT IS THE HIGHEST
SATISFACTION
FACTOR BASED ON
DATA FORM
CUSTOMERS?

```
# Analysis 1: Factors that affect cutomers satisfaction
factors = ['Hotel wifi service', 'Departure/Arrival convenience',
           'Ease of Online booking', 'Hotel location', 'Food and drink',
           'Stay comfort', 'Common Room entertainment',
           'Checkin/Checkout service', 'Other service', 'Cleanliness']
factor_means = df[factors].mean().sort_values(ascending=False)

#Defining the colors of the bar chart
warna = ['#4F3139', '#66424D', '#8F6370', '#B99AA3', '#E0D6D9', '#3E262D',
         '#2E1C21', '#1E1115', '#314F45', '#4F313F']

# Visualisasi Factors that affect cutomers satisfaction
plt.figure(figsize=(10, 6))
sn.barplot(x=factor_means.index, y=factor_means.values, palette=warna)
plt.title('Average level of satisfaction')
plt.xticks(rotation=45)
plt.ylabel('Level of satisfaction')
plt.xlabel(" ")
plt.grid(axis='y')
plt.show()
```

The level of satisfaction with other servuces has the highest satisfaction compared to other factors, which is inversely proportional to the hotel wifi service wich has the lowest level so that it needs to be improved.



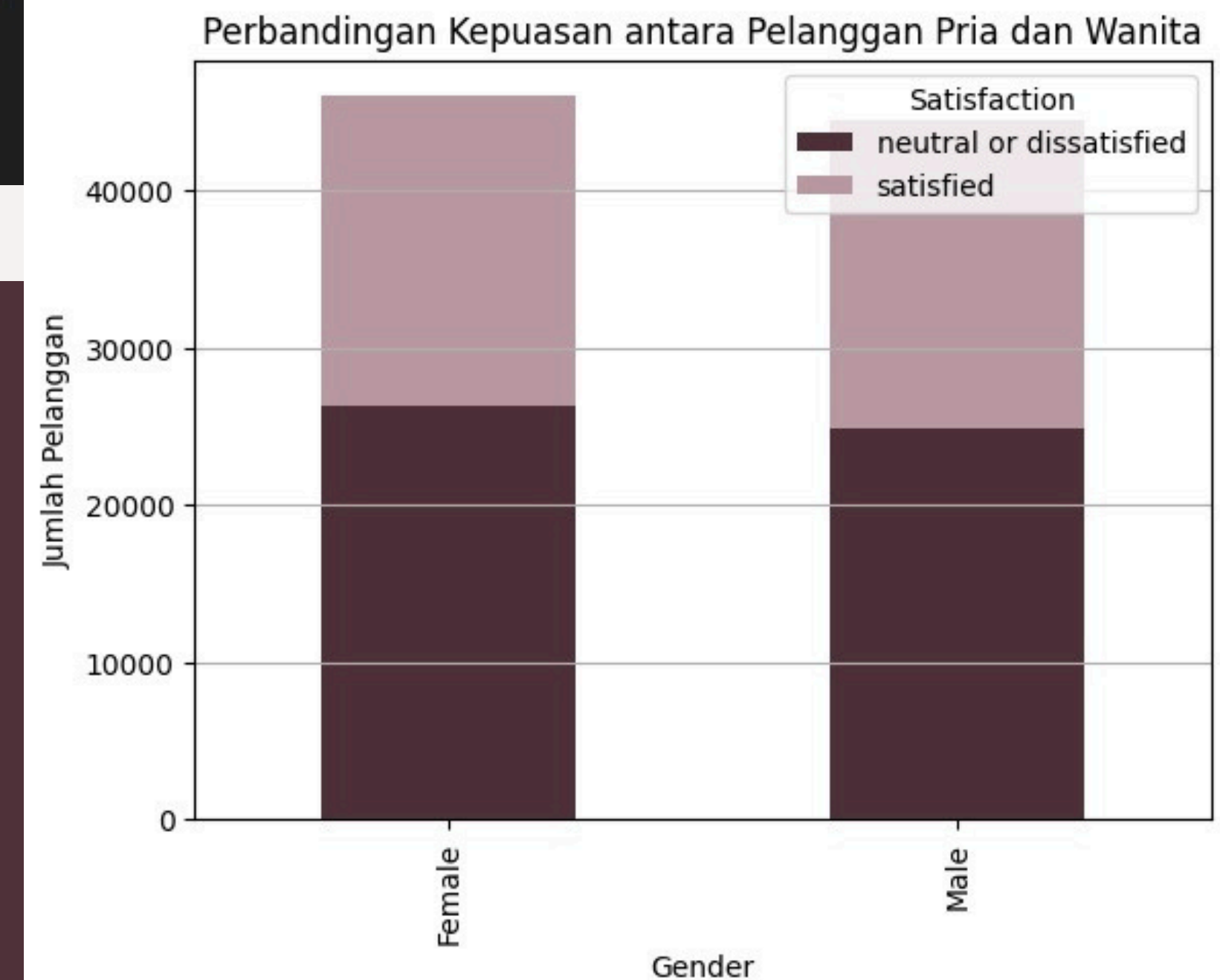
6

WHICH GENDER FEELS
THE MOST SATISFIED?

```
# Menghitung jumlah pelanggan berdasarkan gender dan kepuasan
gender_satisfaction_counts = df.groupby(['Gender', 'satisfaction']).size().unstack()

# Membuat bar chart
gender_satisfaction_counts.plot(kind='bar', stacked=True, color = warna1)
plt.xlabel('Gender')
plt.ylabel('Jumlah Pelanggan')
plt.title('Perbandingan Kepuasan antara Pelanggan Pria dan Wanita')
plt.legend(title='Satisfaction', loc='upper right')
plt.grid(axis='y')
plt.show()
```

It can be seen that male are more satisfied than female by **0.25%.**



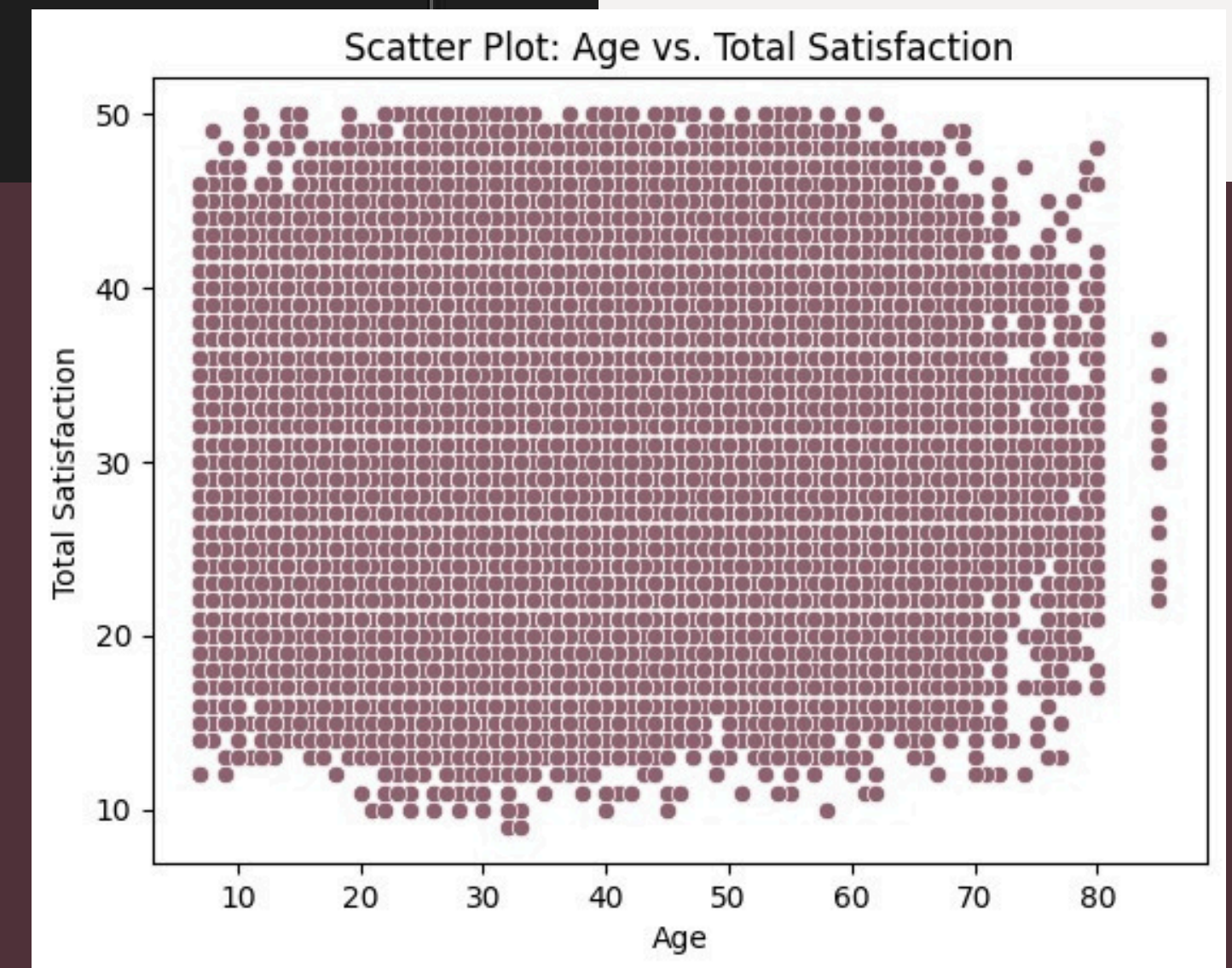
7

IS THERE ANY
RELATION BETWEEN
AGE VS TOTAL SATISFACTION??

```
# Calculating the total satisfaction
df['Total Satisfaction'] = df[['Hotel wifi service',
                              'Departure/Arrival convenience',
                              'Ease of Online booking', 'Hotel location',
                              'Food and drink', 'Stay comfort',
                              'Common Room entertainment', 'Checkin/Checkout service',
                              'Other service', 'Cleanliness']].sum(axis=1)

# Making scatter plot
sns.scatterplot(x='Age', y='Total Satisfaction', data=df, color = '#8F6370')
plt.xlabel('Age')
plt.ylabel('Total Satisfaction')
plt.title('Scatter Plot: Age vs. Total Satisfaction')
plt.show()
```

Based on this data, it can be seen that the data is spread evenly across various age ranges. There was no clear pattern indicating that certain ages had higher or lower levels of total satisfaction. And also the density of dots on the plot shows that most respondents have varying levels of total satisfaction in almost every age range. This suggests that age does not significantly influence the level of total satisfaction reported.



— —

Thank You