In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
import os          # accessing directory structure
```

In [3]:
```
info_tourism = pd.read_csv("tourism_with_id.csv")
tourism_rating = pd.read_csv("tourism_rating.csv")
users = pd.read_csv("user.csv")
```

In [4]:
```
info_tourism
```

Out[4]:

	Place_Id	Place_Name	Description	Category	City	Price	Rating
0	1	Monumen Nasional	Monumen Nasional atau yang populer disingkat d...	Budaya	Jakarta	20000	4.6
1	2	Kota Tua	Kota tua di Jakarta, yang juga bernama Kota Tu...	Budaya	Jakarta	0	4.6
2	3	Dunia Fantasi	Dunia Fantasi atau disebut juga Dufan adalah t...	Taman Hiburan	Jakarta	270000	4.6
3	4	Taman Mini Indonesia Indah (TMII)	Taman Mini Indonesia Indah merupakan suatu kaw...	Taman Hiburan	Jakarta	10000	4.5
4	5	Atlantis Water Adventure	Atlantis Water Adventure atau dikenal dengan A...	Taman Hiburan	Jakarta	94000	4.5
...	...	...	...	...	...	...	...
432	433	Museum Mpu Tantular	Museum Negeri Mpu Tantular adalah sebuah museu...	Budaya	Surabaya	2000	4.4
433	434	Taman Bungkul	Taman Bungkul adalah taman wisata kota yang te...	Taman Hiburan	Surabaya	0	4.6
434	435	Taman Air Mancur Menari Kenjeran	Air mancur menari atau dancing fountain juga a...	Taman Hiburan	Surabaya	0	4.4
435	436	Taman Flora Bratang Surabaya	Taman Flora adalah salah satu taman kota di Su...	Taman Hiburan	Surabaya	0	4.6
436	437	Gereja Perawan Maria Tak	Gereja Katolik Kelahiran Santa	Tempat Ibadah	Surabaya	10000	4.8

	Place_Id	Place_Name	Description	Category	City	Price	Rating	T
		Berdosa Surabaya	Perawan Maria m...					

437 rows × 13 columns

In [5]:
```
info_tourism.head()
```

Out[5]:

	Place_Id	Place_Name	Description	Category	City	Price	Rating	Time
0	1	Monumen Nasional	Monumen Nasional atau yang populer disingkat d...	Budaya	Jakarta	20000	4.6	
1	2	Kota Tua	Kota tua di Jakarta, yang juga bernama Kota Tu...	Budaya	Jakarta	0	4.6	
2	3	Dunia Fantasi	Dunia Fantasi atau disebut juga Dufan adalah t...	Taman Hiburan	Jakarta	270000	4.6	
3	4	Taman Mini Indonesia Indah (TMII)	Taman Mini Indonesia Indah merupakan suatu kaw...	Taman Hiburan	Jakarta	10000	4.5	
4	5	Atlantis Water Adventure	Atlantis Water Adventure atau dikenal dengan A...	Taman Hiburan	Jakarta	94000	4.5	

In [6]:
```
info_tourism.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 437 entries, 0 to 436
Data columns (total 13 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Place_Id      437 non-null    int64  
 1   Place_Name    437 non-null    object 
 2   Description   437 non-null    object 
 3   Category      437 non-null    object 
 4   City          437 non-null    object 
 5   Price         437 non-null    int64  
 6   Rating        437 non-null    float64
 7   Time_Minutes  205 non-null    float64
 8   Coordinate    437 non-null    object 
 9   Lat           437 non-null    float64
 10  Long          437 non-null    float64
 11  Unnamed: 11   0 non-null      float64
 12  Unnamed: 12   437 non-null    int64  
dtypes: float64(5), int64(3), object(5)
memory usage: 44.5+ KB
```

In [7]:
```
info_tourism.describe
```

Out[7]:
```
<bound method NDFrame.describe of      Place_Id
Place_Name  \
0           1                             Monumen Nasional
1           2                                    Kota Tua
2           3                                Dunia Fantasi
3           4          Taman Mini Indonesia Indah (TMII)
4           5                    Atlantis Water Adventure
..        ...                                         ...
432       433                        Museum Mpu Tantular
433       434                             Taman Bungkul
434       435           Taman Air Mancur Menari Kenjeran
435       436               Taman Flora Bratang Surabaya
436       437  Gereja Perawan Maria Tak Berdosa Surabaya

                                           Description       Category  \
0    Monumen Nasional atau yang populer disingkat d...         Budaya
1    Kota tua di Jakarta, yang juga bernama Kota Tu...         Budaya
2    Dunia Fantasi atau disebut juga Dufan adalah t...  Taman Hiburan
3    Taman Mini Indonesia Indah merupakan suatu kaw...  Taman Hiburan
4    Atlantis Water Adventure atau dikenal dengan A...  Taman Hiburan
..                                                 ...            ...
432  Museum Negeri Mpu Tantular adalah sebuah museu...         Budaya
433  Taman Bungkul adalah taman wisata kota yang te...  Taman Hiburan
434  Air mancur menari atau dancing fountain juga a...  Taman Hiburan
435  Taman Flora adalah salah satu taman kota di Su...  Taman Hiburan
436  Gereja Katolik Kelahiran Santa Perawan Maria m...  Tempat Ibadah

         City   Price  Rating  Time_Minutes  \
0     Jakarta   20000     4.6          15.0
1     Jakarta       0     4.6          90.0
2     Jakarta  270000     4.6         360.0
3     Jakarta   10000     4.5           NaN
4     Jakarta   94000     4.5          60.0
..        ...     ...     ...           ...
432  Surabaya    2000     4.4          45.0
433  Surabaya       0     4.6           NaN
434  Surabaya       0     4.4          45.0
435  Surabaya       0     4.6           NaN
436  Surabaya   10000     4.8           NaN

                                        Coordinate       Lat        Long
\
0           {'lat': -6.1753924, 'lng': 106.8271528} -6.175392  106.827153
1    {'lat': -6.137644799999999, 'lng': 106.8171245} -6.137645  106.817125
2    {'lat': -6.125312399999999, 'lng': 106.8335377} -6.125312  106.833538
3    {'lat': -6.302445899999999, 'lng': 106.8951559} -6.302446  106.895156
4              {'lat': -6.12419, 'lng': 106.839134} -6.124190  106.839134
..                                              ...       ...         ...
432         {'lat': -7.4338593, 'lng': 112.7199058} -7.433859  112.719906
433  {'lat': -7.291346799999999, 'lng': 112.7398218} -7.291347  112.739822
434         {'lat': -7.2752955, 'lng': 112.7549381} -7.275296  112.754938
435  {'lat': -7.294330299999999, 'lng': 112.7617534} -7.294330  112.761753
436         {'lat': -7.2420758, 'lng': 112.7368158} -7.242076  112.736816

     Unnamed: 11  Unnamed: 12
0            NaN            1
```

```
1            NaN           2
2            NaN           3
3            NaN           4
4            NaN           5
..           ...         ...
432          NaN         433
433          NaN         434
434          NaN         435
435          NaN         436
436          NaN         437

[437 rows x 13 columns]>
```

In [8]:
```
info_tourism.columns
```

Out[8]:
```
Index(['Place_Id', 'Place_Name', 'Description', 'Category', 'City', 'Pric
e',
       'Rating', 'Time_Minutes', 'Coordinate', 'Lat', 'Long', 'Unnamed: 1
1',
       'Unnamed: 12'],
      dtype='object')
```

In [9]:
```
print(f"Column types in info_tourism:\n{info_tourism.dtypes}")
```

```
Column types in info_tourism:
Place_Id          int64
Place_Name       object
Description      object
Category         object
City             object
Price             int64
Rating          float64
Time_Minutes    float64
Coordinate       object
Lat             float64
Long            float64
Unnamed: 11     float64
Unnamed: 12       int64
dtype: object
```

In [10]:
```
info_tourism.count
```

Out[10]:
```
<bound method DataFrame.count of      Place_Id
Place_Name  \
0           1                            Monumen Nasional
1           2                                    Kota Tua
2           3                                Dunia Fantasi
3           4          Taman Mini Indonesia Indah (TMII)
4           5                    Atlantis Water Adventure
..        ...                                         ...
432       433                       Museum Mpu Tantular
433       434                            Taman Bungkul
434       435           Taman Air Mancur Menari Kenjeran
435       436               Taman Flora Bratang Surabaya
436       437  Gereja Perawan Maria Tak Berdosa Surabaya

                                           Description       Category  \
0    Monumen Nasional atau yang populer disingkat d...         Budaya
1    Kota tua di Jakarta, yang juga bernama Kota Tu...         Budaya
2    Dunia Fantasi atau disebut juga Dufan adalah t...  Taman Hiburan
3    Taman Mini Indonesia Indah merupakan suatu kaw...  Taman Hiburan
4    Atlantis Water Adventure atau dikenal dengan A...  Taman Hiburan
..                                                 ...            ...
432  Museum Negeri Mpu Tantular adalah sebuah museu...         Budaya
433  Taman Bungkul adalah taman wisata kota yang te...  Taman Hiburan
434  Air mancur menari atau dancing fountain juga a...  Taman Hiburan
435  Taman Flora adalah salah satu taman kota di Su...  Taman Hiburan
436  Gereja Katolik Kelahiran Santa Perawan Maria m...  Tempat Ibadah

         City   Price  Rating  Time_Minutes  \
0     Jakarta   20000     4.6          15.0
1     Jakarta       0     4.6          90.0
2     Jakarta  270000     4.6         360.0
3     Jakarta   10000     4.5           NaN
4     Jakarta   94000     4.5          60.0
..        ...     ...     ...           ...
432  Surabaya    2000     4.4          45.0
433  Surabaya       0     4.6           NaN
434  Surabaya       0     4.4          45.0
435  Surabaya       0     4.6           NaN
436  Surabaya   10000     4.8           NaN

                                        Coordinate       Lat        Long
\
0           {'lat': -6.1753924, 'lng': 106.8271528} -6.175392  106.827153
1    {'lat': -6.137644799999999, 'lng': 106.8171245} -6.137645  106.817125
2    {'lat': -6.125312399999999, 'lng': 106.8335377} -6.125312  106.833538
3    {'lat': -6.302445899999999, 'lng': 106.8951559} -6.302446  106.895156
4               {'lat': -6.12419, 'lng': 106.839134} -6.124190  106.839134
..                                               ...       ...         ...
432         {'lat': -7.4338593, 'lng': 112.7199058} -7.433859  112.719906
433  {'lat': -7.291346799999999, 'lng': 112.7398218} -7.291347  112.739822
434         {'lat': -7.2752955, 'lng': 112.7549381} -7.275296  112.754938
435  {'lat': -7.294330299999999, 'lng': 112.7617534} -7.294330  112.761753
436         {'lat': -7.2420758, 'lng': 112.7368158} -7.242076  112.736816

     Unnamed: 11  Unnamed: 12
0            NaN            1
```

```
1            NaN           2
2            NaN           3
3            NaN           4
4            NaN           5
..           ...         ...
432          NaN         433
433          NaN         434
434          NaN         435
435          NaN         436
436          NaN         437

[437 rows x 13 columns]>
```

In [11]:
```
info_tourism.shape
```

Out[11]:
```
(437, 13)
```

In [12]:
```
info_tourism.Category.unique()
```

Out[12]:
```
array(['Budaya', 'Taman Hiburan', 'Cagar Alam', 'Bahari',
       'Pusat Perbelanjaan', 'Tempat Ibadah'], dtype=object)
```

2nd table- tourism_rating

In [13]:
```
tourism_rating.head()
```

Out[13]:

	User_Id	Place_Id	Place_Ratings
0	1	179	3
1	1	344	2
2	1	5	5
3	1	373	3
4	1	101	4

In [14]:
```
tourism_rating.tail()
```

Out[14]:

	User_Id	Place_Id	Place_Ratings
9995	300	425	2
9996	300	64	4
9997	300	311	3
9998	300	279	4
9999	300	163	2

In [15]:
```
tourism_rating.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 3 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   User_Id        10000 non-null  int64
 1   Place_Id       10000 non-null  int64
 2   Place_Ratings  10000 non-null  int64
dtypes: int64(3)
memory usage: 234.5 KB
```

In [16]:
```
tourism_rating.describe
```

Out[16]:
```
<bound method NDFrame.describe of       User_Id  Place_Id  Place_Ratings
0           1       179              3
1           1       344              2
2           1         5              5
3           1       373              3
4           1       101              4
...       ...       ...            ...
9995      300       425              2
9996      300        64              4
9997      300       311              3
9998      300       279              4
9999      300       163              2

[10000 rows x 3 columns]>
```

In [17]:
```
tourism_rating.count()
```

Out[17]:
```
User_Id          10000
Place_Id         10000
Place_Ratings    10000
dtype: int64
```

In [18]:
```
tourism_rating.shape
```

Out[18]:
```
(10000, 3)
```

In [19]:
```
tourism_rating.columns
```

Out[19]:
```
Index(['User_Id', 'Place_Id', 'Place_Ratings'], dtype='object')
```

In [20]:
```
print(f"Column types in tourism_rating:\n{tourism_rating.dtypes}")
```

```
Column types in tourism_rating:
User_Id          int64
Place_Id         int64
Place_Ratings    int64
dtype: object
```

CHECK IF NULL VALUE PRESENT OR NOT

In [21]:
```
tourism_rating.isnull().sum()
```

Out[21]:
```
User_Id          0
Place_Id         0
Place_Ratings    0
dtype: int64
```

3rd table- users table

In [22]:
```
users.head()
```

Out[22]:

	User_Id	Location	Age
0	1	Semarang, Jawa Tengah	20
1	2	Bekasi, Jawa Barat	21
2	3	Cirebon, Jawa Barat	23
3	4	Bekasi, Jawa Barat	21
4	5	Lampung, Sumatera Selatan	20

In [23]:
```
users.tail()
```

Out[23]:

	User_Id	Location	Age
295	296	Lampung, Sumatera Selatan	31
296	297	Palembang, Sumatera Selatan	39
297	298	Bogor, Jawa Barat	38
298	299	Sragen, Jawa Tengah	27
299	300	Ponorogo, Jawa Timur	26

In [24]:
```
users.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   User_Id   300 non-null    int64 
 1   Location  300 non-null    object
 2   Age       300 non-null    int64 
dtypes: int64(2), object(1)
memory usage: 7.2+ KB
```

Statistical Summary

In [25]:
```
users.describe()
```

Out[25]:

	User_Id	Age
count	300.000000	300.000000
mean	150.500000	28.700000
std	86.746758	6.393716
min	1.000000	18.000000
25%	75.750000	24.000000
50%	150.500000	29.000000
75%	225.250000	34.000000
max	300.000000	40.000000

In [26]:
```
users.count()
```

Out[26]:
```
User_Id     300
Location    300
Age         300
dtype: int64
```

In [27]:
```
users.shape
```

Out[27]:
```
(300, 3)
```

In [28]:
```
users.columns
```

Out[28]:
```
Index(['User_Id', 'Location', 'Age'], dtype='object')
```

CHECK IF NULL VALUE PRESENT OR NOT

In [29]:
```
users.isnull().sum()
```

Out[29]:
```
User_Id     0
Location    0
Age         0
dtype: int64
```

Number of Places in the Dataset:

In [30]:
```
print(f"Number of places in the datasets : {len(info_tourism.Place_Id.unique
```

```
Number of places in the datasets : 437
```

Number of Users:

In [31]:
```
print(f"Number of users : {len(users.User_Id.unique())}")
```

```
Number of users : 300
```

The Number of Ratings Given by Users:

In [32]:
```
print(f"The number of ratings given by the user to the dataset : {len(touris
```

```
The number of ratings given by the user to the dataset : 10000
```

In [33]:
```
# Column data types

print(f"Column types in users:\n{users.dtypes}")
```

```
Column types in users:
User_Id      int64
Location    object
Age          int64
dtype: object
```

Check for duplicate rows

In [34]:
```
print("Duplicates in tourism_rating:", tourism_rating.duplicated().sum())
print("Duplicates in info_tourism:", info_tourism.duplicated().sum())
print("Duplicates in users:", users.duplicated().sum())
```

```
Duplicates in tourism_rating: 79
Duplicates in info_tourism: 0
Duplicates in users: 0
```

DATA PREPROCESSING

In [35]:
```
tourism_all= np.concatenate((
            info_tourism.Place_Id.unique(),
    tourism_rating.Place_Id.unique()
))

tourism_all= np.sort(np.unique(tourism_all))
tourism_all
```

Out[35]:
```
array([  1,   2,   3,   4,   5,   6,   7,   8,   9,  10,  11,  12,  13,
        14,  15,  16,  17,  18,  19,  20,  21,  22,  23,  24,  25,  26,
        27,  28,  29,  30,  31,  32,  33,  34,  35,  36,  37,  38,  39,
        40,  41,  42,  43,  44,  45,  46,  47,  48,  49,  50,  51,  52,
        53,  54,  55,  56,  57,  58,  59,  60,  61,  62,  63,  64,  65,
        66,  67,  68,  69,  70,  71,  72,  73,  74,  75,  76,  77,  78,
        79,  80,  81,  82,  83,  84,  85,  86,  87,  88,  89,  90,  91,
        92,  93,  94,  95,  96,  97,  98,  99, 100, 101, 102, 103, 104,
       105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117,
       118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130,
       131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143,
       144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156,
       157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169,
       170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182,
       183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195,
       196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208,
       209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221,
       222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234,
       235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247,
       248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260,
       261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273,
       274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286,
       287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299,
       300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312,
       313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325,
       326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338,
       339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351,
       352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364,
       365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377,
       378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390,
       391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403,
       404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416,
       417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429,
       430, 431, 432, 433, 434, 435, 436, 437], dtype=int64)
```

In [36]:
```
print(f"Total number of tourism: {len(tourism_all)}")
```

```
Total number of tourism: 437
```

In [37]:
```
all_tourism_rate = tourism_rating
all_tourism_rate
```

Out[37]:

	User_Id	Place_Id	Place_Ratings
0	1	179	3
1	1	344	2
2	1	5	5
3	1	373	3
4	1	101	4
...	...	...	...
9995	300	425	2
9996	300	64	4
9997	300	311	3
9998	300	279	4
9999	300	163	2

10000 rows × 3 columns

MERGE THE COLUMNS

In [38]:
```
all_tourism = pd.merge(all_tourism_rate,info_tourism[["Place_Id","Place_Name
all_tourism
```

Out[38]:

	User_Id	Place_Id	Place_Ratings	Place_Name	Description	City
0	1	179	3	Candi Ratu Boko	Situs Ratu Baka atau Candi Boko (Hanacaraka:ꦕꦤ...	Yogyakarta
1	1	344	2	Pantai Marina	Pantai Marina (bahasa Jawa: ꦥꦟ꧀ꦠꦲꦶꦩꦫꦶꦤ, trans...	Semarang
2	1	5	5	Atlantis Water Adventure	Atlantis Water Adventure atau dikenal dengan A...	Jakarta
3	1	373	3	Museum Kereta Ambarawa	Museum Kereta Api Ambarawa (bahasa Inggris: In...	Semarang
4	1	101	4	Kampung Wisata Sosro Menduran	Kampung wisata Sosromenduran merupakan kampung...	Yogyakarta
...	...	...	...	...	...	...
9995	300	425	2	Waterpark Kenjeran Surabaya	Waterpark Kenjeran Surabaya merupakan wisata k...	Surabaya
9996	300	64	4	Museum Sasmita Loka Ahmad Yani	Museum Sasmita Loka Ahmad Yani adalah salah sa...	Jakarta
9997	300	311	3	The Lodge Maribaya	The Lodge Maribaya adalah salah satu tempat wi...	Bandung
9998	300	279	4	Masjid Agung Trans Studio Bandung	Masjid Agung Trans Studio Bandung (TSB) berdir...	Bandung
9999	300	163	2	Watu Mabur Mangunan	Kawasan Tebing Watu Mabur ini terbilang belum ...	Yogyakarta

10000 rows × 7 columns

creating a new column called city_category

In [39]:
```
all_tourism['city_category'] = all_tourism[['City','Category']].agg(' '.joir
```

In [40]:
```
all_tourism
```

Out[40]:

	User_Id	Place_Id	Place_Ratings	Place_Name	Description	City
0	1	179	3	Candi Ratu Boko	Situs Ratu Baka atau Candi Boko (Hanacaraka:ꦕꦤ...	Yogyakarta
1	1	344	2	Pantai Marina	Pantai Marina (bahasa Jawa: ꦥꦟ꧀ꦠꦲꦶꦩꦫི, trans...	Semarang
2	1	5	5	Atlantis Water Adventure	Atlantis Water Adventure atau dikenal dengan A...	Jakarta
3	1	373	3	Museum Kereta Ambarawa	Museum Kereta Api Ambarawa (bahasa Inggris: In...	Semarang
4	1	101	4	Kampung Wisata Sosro Menduran	Kampung wisata Sosromenduran merupakan kampung...	Yogyakarta
...	...	...	...	...	...	...
9995	300	425	2	Waterpark Kenjeran Surabaya	Waterpark Kenjeran Surabaya merupakan wisata k...	Surabaya
9996	300	64	4	Museum Sasmita Loka Ahmad Yani	Museum Sasmita Loka Ahmad Yani adalah salah sa...	Jakarta
9997	300	311	3	The Lodge Maribaya	The Lodge Maribaya adalah salah satu tempat wi...	Bandung
9998	300	279	4	Masjid Agung Trans Studio Bandung	Masjid Agung Trans Studio Bandung (TSB) berdir...	Bandung
9999	300	163	2	Watu Mabur Mangunan	Kawasan Tebing Watu Mabur ini terbilang belum ...	Yogyakarta

10000 rows × 8 columns

Data Preparation:

check missing value for this column

In [41]:
```
all_tourism.isnull().sum()
```

Out[41]:
```
User_Id          0
Place_Id         0
Place_Ratings    0
Place_Name       0
Description      0
City             0
Category         0
city_category    0
dtype: int64
```

In [42]:
```
all_tourism.count()
```

Out[42]:
```
User_Id          10000
Place_Id         10000
Place_Ratings    10000
Place_Name       10000
Description      10000
City             10000
Category         10000
city_category    10000
dtype: int64
```

In [43]:
```
all_tourism.shape
```

Out[43]:
```
(10000, 8)
```

In [44]:
```
all_tourism.columns
```

Out[44]:
```
Index(['User_Id', 'Place_Id', 'Place_Ratings', 'Place_Name', 'Description',
       'City', 'Category', 'city_category'],
      dtype='object')
```

In [45]:
```
all_tourism.describe()
```

Out[45]:

	User_Id	Place_Id	Place_Ratings
count	10000.000000	10000.000000	10000.000000
mean	151.292700	219.416400	3.066500
std	86.137374	126.228335	1.379952
min	1.000000	1.000000	1.000000
25%	77.000000	108.750000	2.000000
50%	151.000000	220.000000	3.000000
75%	226.000000	329.000000	4.000000
max	300.000000	437.000000	5.000000

In [46]:

```
preparation= all_tourism.drop_duplicates("Place_Id")
preparation
```

Out[46]:

	User_Id	Place_Id	Place_Ratings	Place_Name	Description	City
0	1	179	3	Candi Ratu Boko	Situs Ratu Baka atau Candi Boko (Hanacaraka:ꦕꦤ...	Yogyakarta
1	1	344	2	Pantai Marina	Pantai Marina (bahasa Jawa: ꦥꦤ꧀ꦠꦲꦶꦩꦫི, trans...	Semarang
2	1	5	5	Atlantis Water Adventure	Atlantis Water Adventure atau dikenal dengan A...	Jakarta
3	1	373	3	Museum Kereta Ambarawa	Museum Kereta Api Ambarawa (bahasa Inggris: In...	Semarang
4	1	101	4	Kampung Wisata Sosro Menduran	Kampung wisata Sosromenduran merupakan kampung...	Yogyakarta
...	...	...	...	...	...	...
2008	62	370	1	Benteng Pendem	Benteng Pendem Cilacap (bahasa Belanda: Kustba...	Semarang
2399	74	350	4	Pantai Cipta	Pantai Cipta juga dikenal sebagai Pantai Petik...	Semarang
2448	75	10	2	Pulau Tidung	Pulau Tidung adalah salah satu kelurahan di ke...	Jakarta
2534	78	7	4	Kebun Binatang Ragunan	Kebun Binatang Ragunan adalah sebuah kebun bin...	Jakarta
2918	90	140	1	Bendung Lepen	Bendung Lepen sendiri dulunya merupakan salura...	Yogyakarta

437 rows × 8 columns

In [47]:
```
preparation.head()
```

Out[47]:

	User_Id	Place_Id	Place_Ratings	Place_Name	Description	City	Ca
0	1	179	3	Candi Ratu Boko	Situs Ratu Baka atau Candi Boko (Hanacaraka:ꦕꦤ...	Yogyakarta	
1	1	344	2	Pantai Marina	Pantai Marina (bahasa Jawa: ꦥꦤ꧀ꦠꦲꦶꦩꦫꦶꦤꦲ, trans...	Semarang	
2	1	5	5	Atlantis Water Adventure	Atlantis Water Adventure atau dikenal dengan A...	Jakarta	
3	1	373	3	Museum Kereta Ambarawa	Museum Kereta Api Ambarawa (bahasa Inggris: In...	Semarang	
4	1	101	4	Kampung Wisata Sosro Menduran	Kampung wisata Sosromenduran merupakan kampung...	Yogyakarta	

In [48]:
```
preparation.shape
```

Out[48]: (437, 8)

In [49]:
```
preparation.count
```

Out[49]:
```
<bound method DataFrame.count of       User_Id  Place_Id  Place_Ratings
Place_Name  \
0           1       179              3               Candi Ratu Boko
1           1       344              2                 Pantai Marina
2           1         5              5      Atlantis Water Adventure
3           1       373              3        Museum Kereta Ambarawa
4           1       101              4  Kampung Wisata Sosro Menduran
...       ...       ...            ...                           ...
2008       62       370              1                Benteng Pendem
2399       74       350              4                  Pantai Cipta
2448       75        10              2                  Pulau Tidung
2534       78         7              4        Kebun Binatang Ragunan
2918       90       140              1                 Bendung Lepen

                                            Description        City  \
0     Situs Ratu Baka atau Candi Boko (Hanacaraka:ꦕꦤ...  Yogyakarta
1     Pantai Marina (bahasa Jawa: ꦥꦤ꧀ꦠꦲꦶꦩꦫི, trans...    Semarang
2     Atlantis Water Adventure atau dikenal dengan A...     Jakarta
3     Museum Kereta Api Ambarawa (bahasa Inggris: In...    Semarang
4     Kampung wisata Sosromenduran merupakan kampung...  Yogyakarta
...                                                 ...         ...
2008  Benteng Pendem Cilacap (bahasa Belanda: Kustba...    Semarang
2399  Pantai Cipta juga dikenal sebagai Pantai Petik...    Semarang
2448  Pulau Tidung adalah salah satu kelurahan di ke...     Jakarta
2534  Kebun Binatang Ragunan adalah sebuah kebun bin...     Jakarta
2918  Bendung Lepen sendiri dulunya merupakan salura...  Yogyakarta

          Category             city_category
0           Budaya         Yogyakarta Budaya
1           Bahari           Semarang Bahari
2     Taman Hiburan     Jakarta Taman Hiburan
3           Budaya           Semarang Budaya
4           Budaya         Yogyakarta Budaya
...            ...                      ...
2008        Budaya           Semarang Budaya
2399        Bahari           Semarang Bahari
2448        Bahari            Jakarta Bahari
2534    Cagar Alam        Jakarta Cagar Alam
2918  Taman Hiburan  Yogyakarta Taman Hiburan

[437 rows x 8 columns]>
```

In [50]:
```
preparation.columns
```

Out[50]:
```
Index(['User_Id', 'Place_Id', 'Place_Ratings', 'Place_Name', 'Description',
       'City', 'Category', 'city_category'],
      dtype='object')
```

In [51]:
```
place_id = preparation.Place_Id.tolist()

place_name = preparation.Place_Name.tolist()

place_category = preparation.Category.tolist()

place_desc = preparation.Description.tolist()
```

```
place_city = preparation.City.tolist()

city_category = preparation.city_category.tolist()
```

In [52]:
```
tourism_new=pd.DataFrame({
          "id":place_id,
           "name": place_name,
           "category": place_category,
            "description": place_desc,
             "city":place_city,
            "city_category":city_category

})
```

In [53]:
```
tourism_new
```

Out[53]:

	id	name	category	description	city	city_category
0	179	Candi Ratu Boko	Budaya	Situs Ratu Baka atau Candi Boko (Hanacaraka:ꦕꦤ...	Yogyakarta	Yogyakarta Budaya
1	344	Pantai Marina	Bahari	Pantai Marina (bahasa Jawa: ꦥꦤ꧀ꦠꦲꦶꦩꦫꦶꦤꦲ, trans...	Semarang	Semarang Bahari
2	5	Atlantis Water Adventure	Taman Hiburan	Atlantis Water Adventure atau dikenal dengan A...	Jakarta	Jakarta Taman Hiburan
3	373	Museum Kereta Ambarawa	Budaya	Museum Kereta Api Ambarawa (bahasa Inggris: In...	Semarang	Semarang Budaya
4	101	Kampung Wisata Sosro Menduran	Budaya	Kampung wisata Sosromenduran merupakan kampung...	Yogyakarta	Yogyakarta Budaya
...	...	...	...	...	...	...
432	370	Benteng Pendem	Budaya	Benteng Pendem Cilacap (bahasa Belanda: Kustba...	Semarang	Semarang Budaya
433	350	Pantai Cipta	Bahari	Pantai Cipta juga dikenal sebagai Pantai Petik...	Semarang	Semarang Bahari
434	10	Pulau Tidung	Bahari	Pulau Tidung adalah salah satu kelurahan di ke...	Jakarta	Jakarta Bahari
435	7	Kebun Binatang Ragunan	Cagar Alam	Kebun Binatang Ragunan adalah sebuah kebun bin...	Jakarta	Jakarta Cagar Alam
436	140	Bendung Lepen	Taman Hiburan	Bendung Lepen sendiri dulunya merupakan salura...	Yogyakarta	Yogyakarta Taman Hiburan

437 rows × 6 columns

In [54]:
```
top_10 = tourism_new['id'].value_counts().reset_index()[0:10]
top_10 = pd.merge(top_10,preparation[['Place_Id','Place_Name']], how='left',
```

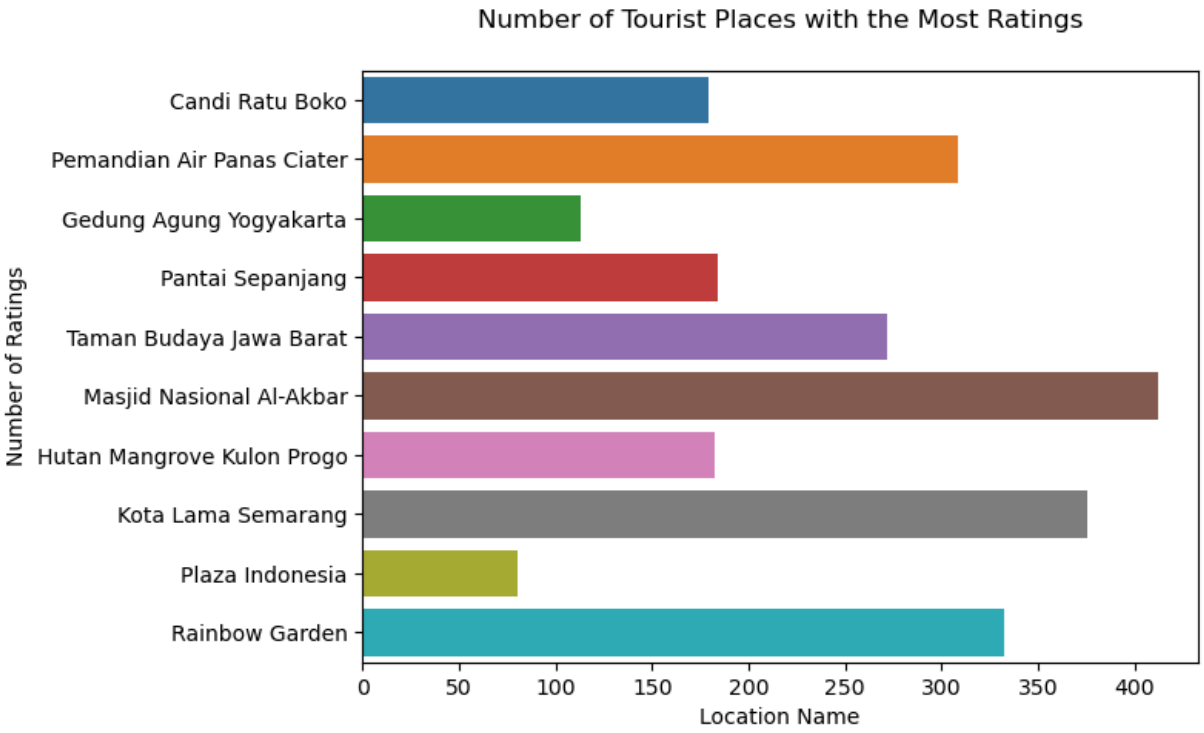
In [55]:
```
top_10
```

Out[55]:

	index	id	Place_Id	Place_Name
0	179	1	179	Candi Ratu Boko
1	308	1	308	Pemandian Air Panas Ciater
2	113	1	113	Gedung Agung Yogyakarta
3	184	1	184	Pantai Sepanjang
4	272	1	272	Taman Budaya Jawa Barat
5	412	1	412	Masjid Nasional Al-Akbar
6	182	1	182	Hutan Mangrove Kulon Progo
7	375	1	375	Kota Lama Semarang
8	80	1	80	Plaza Indonesia
9	332	1	332	Rainbow Garden

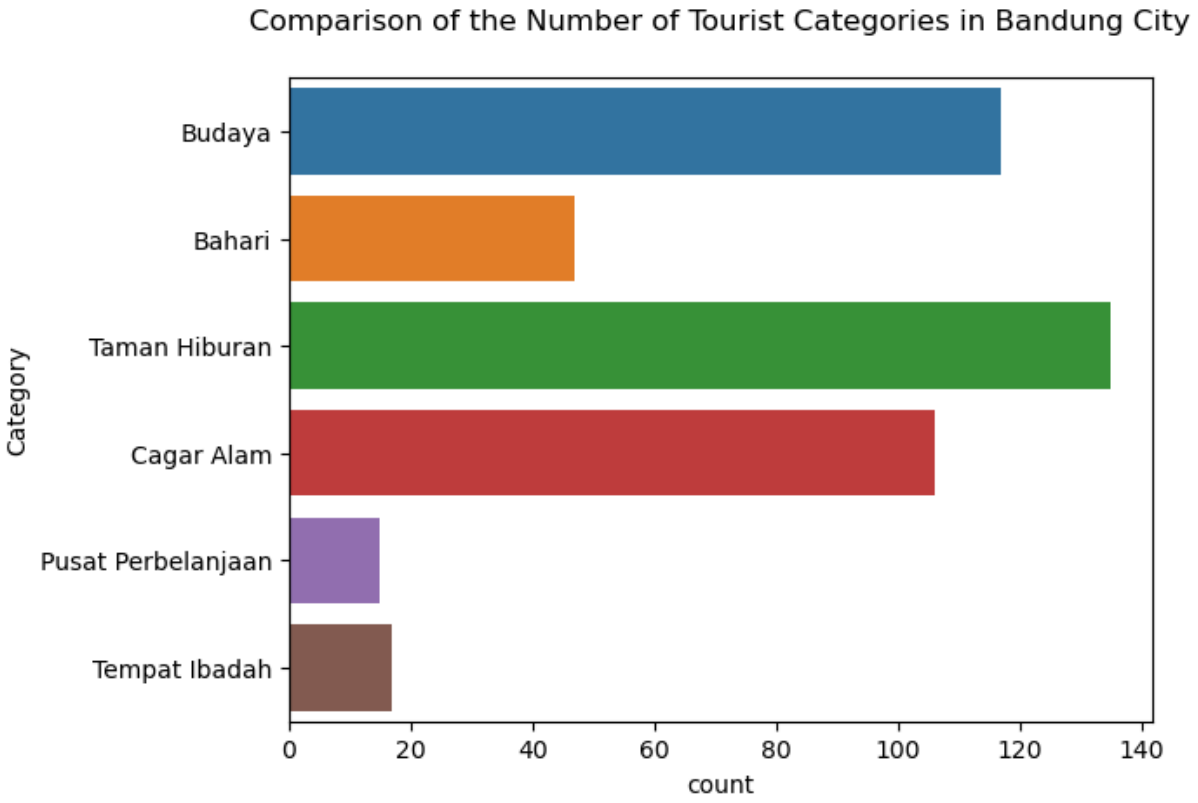
In [56]:
```
plt.figure(figsize=(7,5))
sns.barplot('Place_Id', 'Place_Name', data=top_10)
plt.title('Number of Tourist Places with the Most Ratings', pad=20)
plt.ylabel('Number of Ratings')
plt.xlabel('Location Name')
plt.show()
```

```
C:\Users\shaw3\anaconda3\lib\site-packages\seaborn\_decorators.py:36: Future
Warning: Pass the following variables as keyword args: x, y. From version 0.
12, the only valid positional argument will be `data`, and passing other arg
uments without an explicit keyword will result in an error or misinterpretat
ion.
  warnings.warn(
```

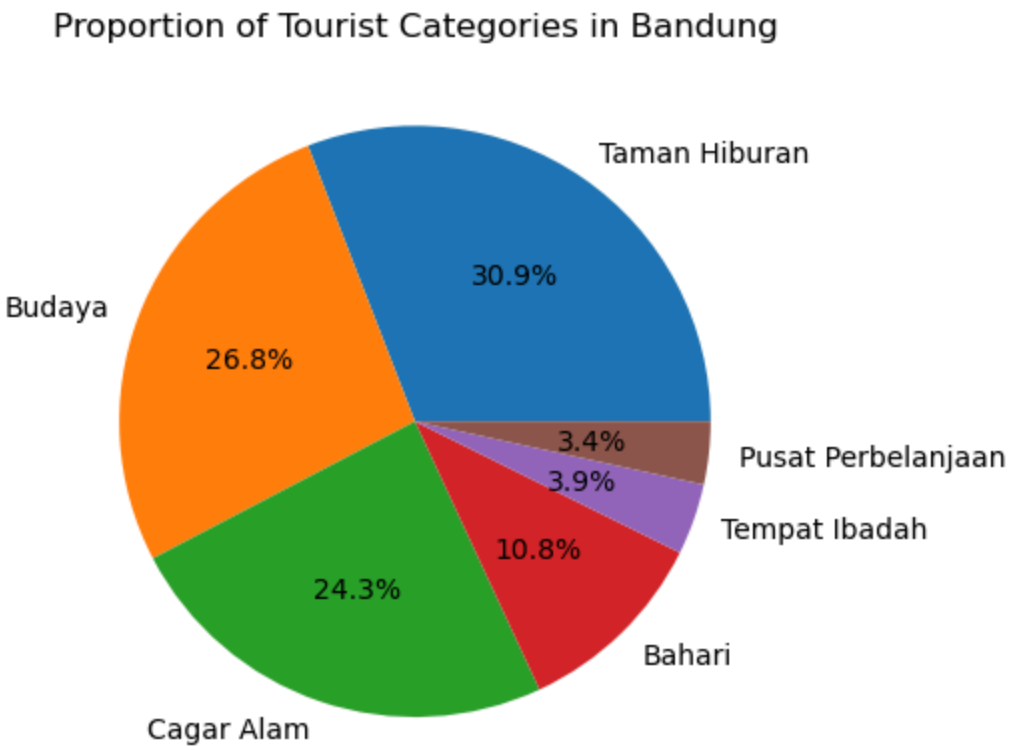


pad=20: Adds padding of 20 points between the title and the plot for better visual separation.

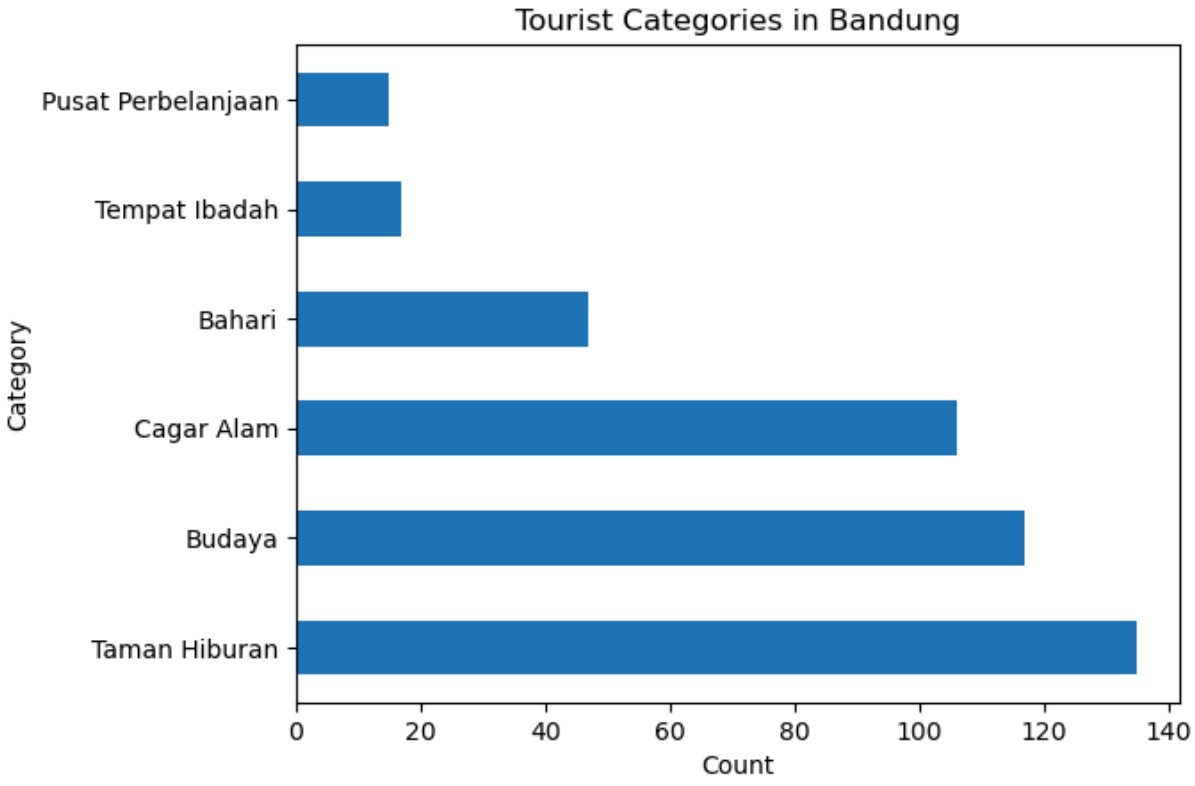
In [57]:
```
sns.countplot(y='Category', data=preparation)
plt.title('Comparison of the Number of Tourist Categories in Bandung City',
plt.show()
```



In [58]:
```
category_counts = preparation['Category'].value_counts()
plt.pie(category_counts, labels=category_counts.index, autopct='%1.1f%%')
plt.title('Proportion of Tourist Categories in Bandung')
plt.show()
```

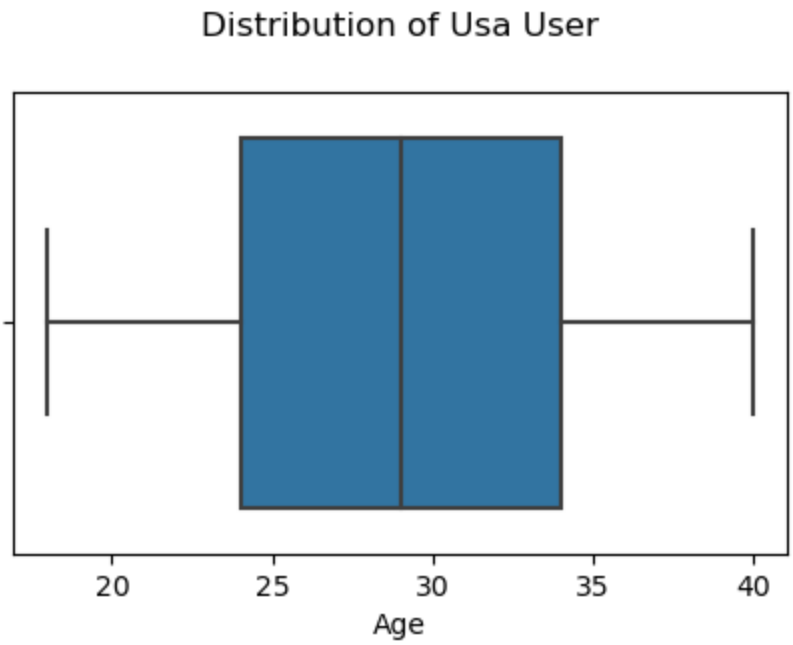


In [59]:
```
category_counts = preparation['Category'].value_counts()
category_counts.plot(kind='barh')
plt.title('Tourist Categories in Bandung')
plt.xlabel('Count')
plt.ylabel('Category')
plt.show()
```

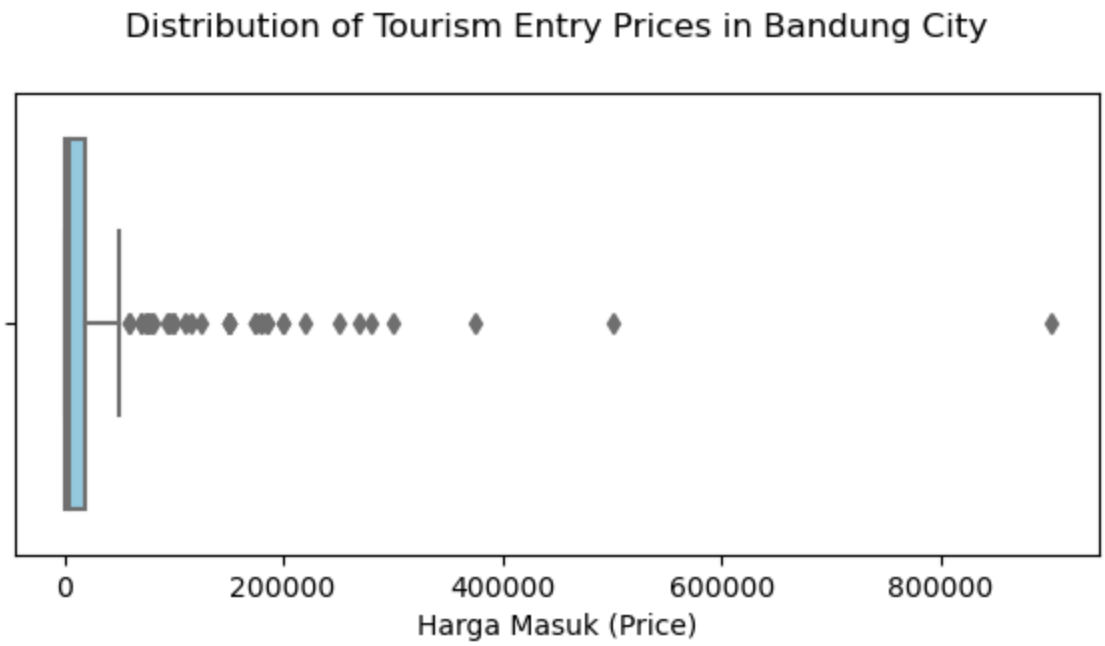


In [60]:
```
plt.figure(figsize=(5,3))
sns.boxplot(users['Age']);
plt.title('Distribution of Usa User', pad=20)
plt.show()
```

```
C:\Users\shaw3\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

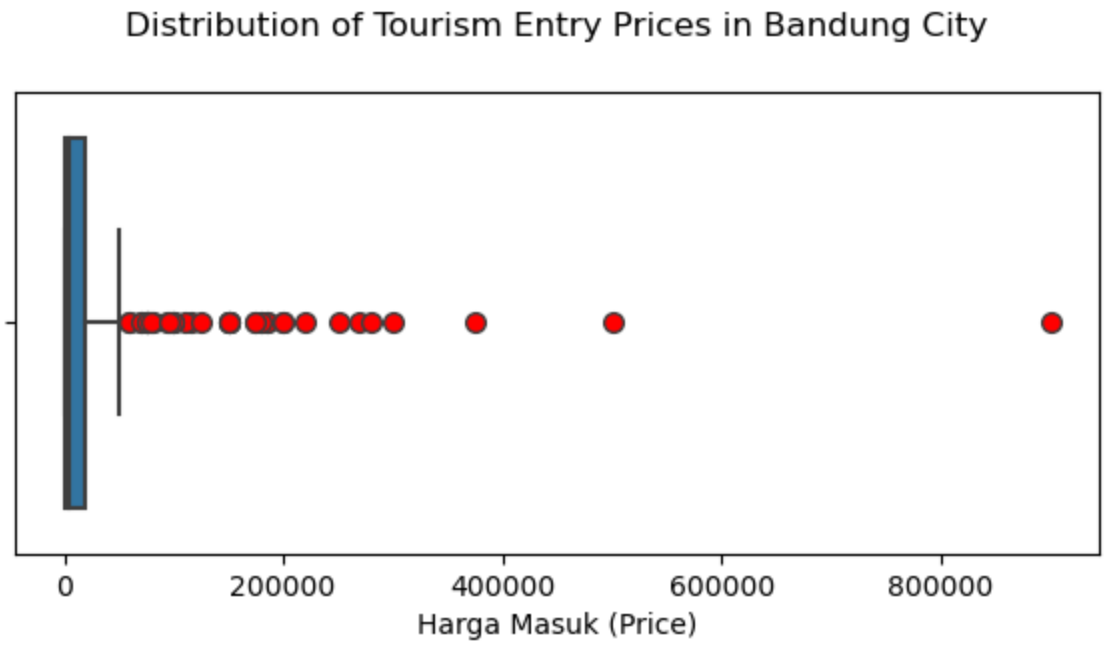


In [61]:
```
plt.figure(figsize=(7, 3))
sns.boxplot(x=info_tourism['Price'], color='skyblue')  # Adding a color
plt.title('Distribution of Tourism Entry Prices in Bandung City', pad=20)
plt.xlabel('Harga Masuk (Price)')
plt.show()
```



Highlighting Outliers

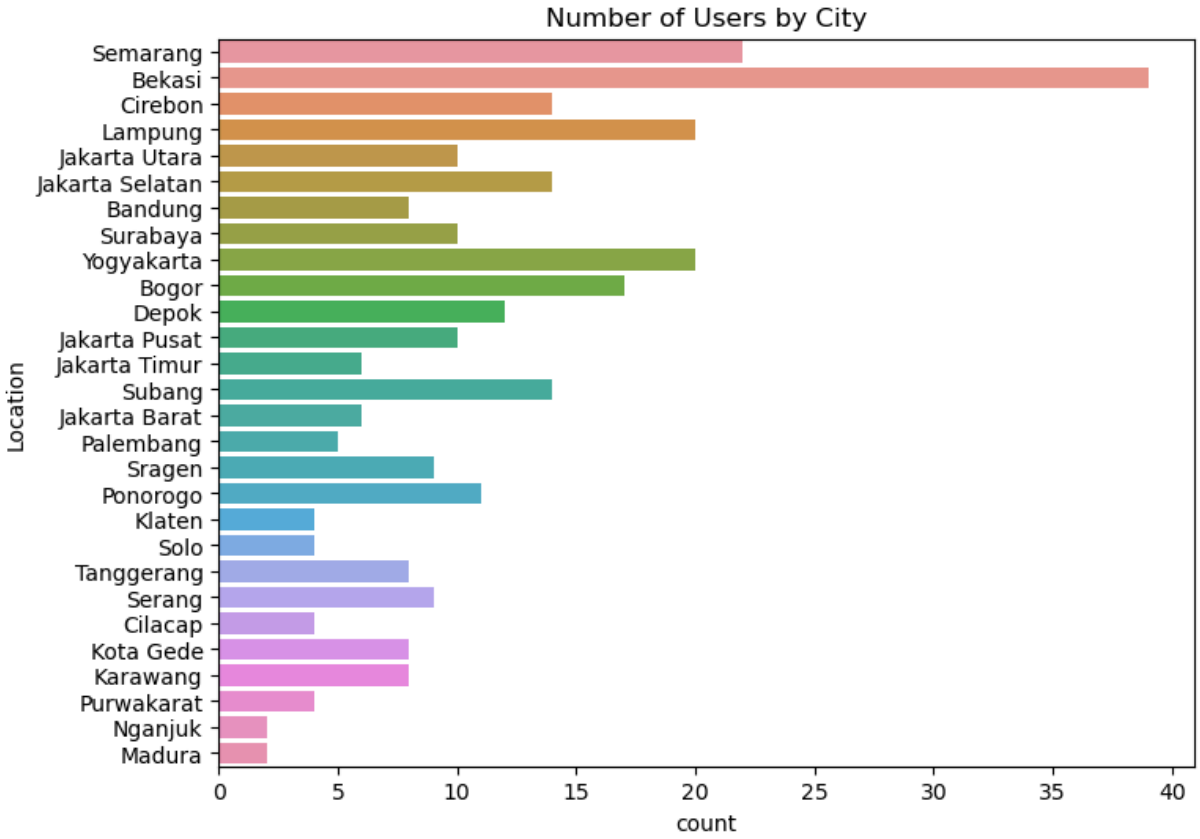
In [62]:
```
plt.figure(figsize=(7, 3))
sns.boxplot(x=info_tourism['Price'], flierprops={"marker": "o", "markerfaced
plt.title('Distribution of Tourism Entry Prices in Bandung City', pad=20)
plt.xlabel('Harga Masuk (Price)')
plt.show()
```



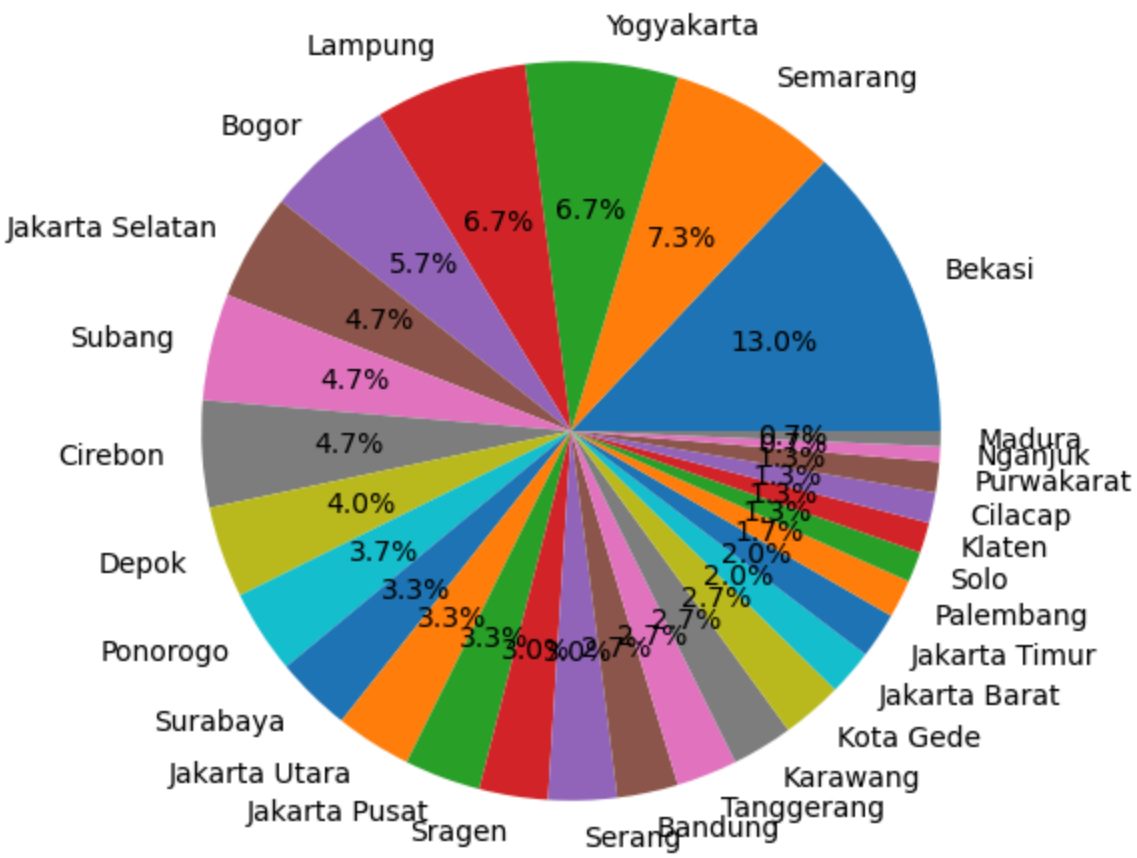
In [63]:
```
askot = users['Location'].apply(lambda x : x.split(',')[0])


plt.figure(figsize=(8,6))
sns.countplot(y=askot)
```

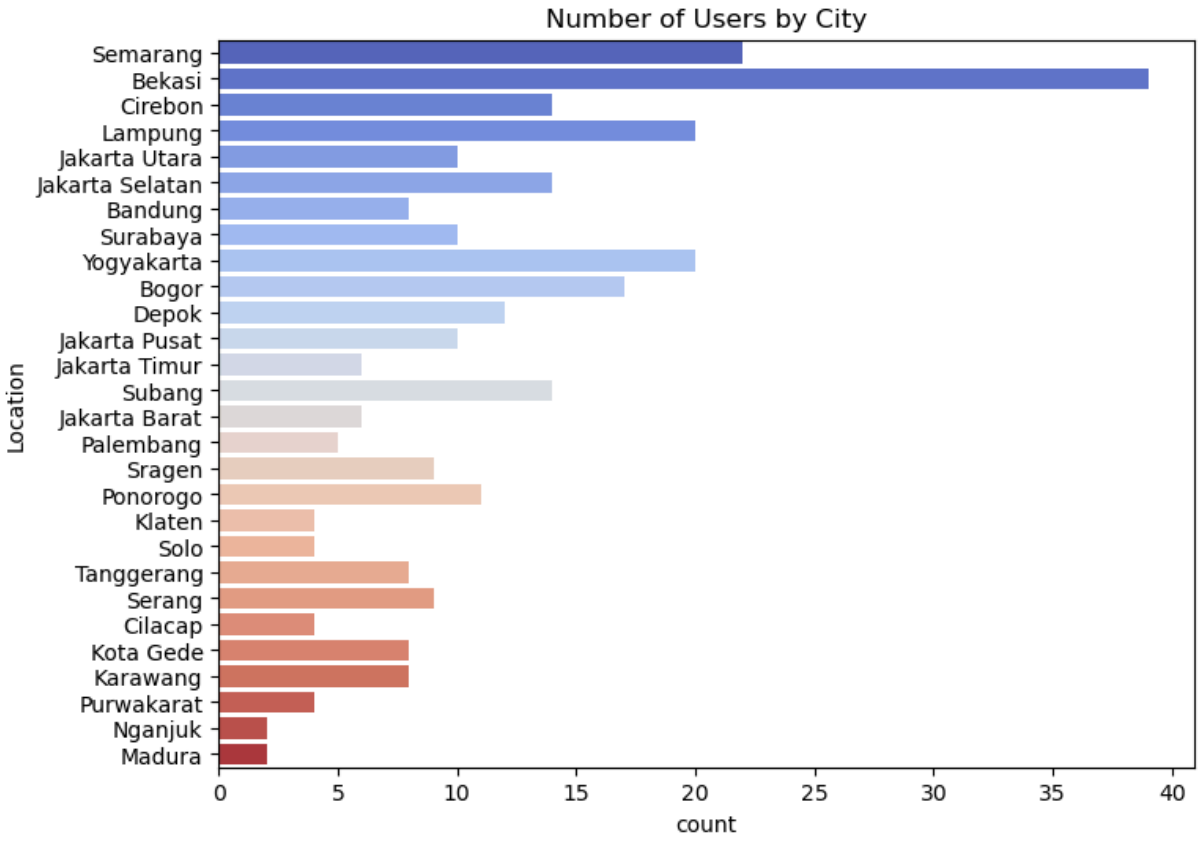
```
plt.title('Number of Users by City')
plt.show()
```



In [64]:
```
city_counts = askot.value_counts()
plt.figure(figsize=(8,6))
plt.pie(city_counts, labels=city_counts.index, autopct='%1.1f%%')
plt.title('Proportion of Users by City')
plt.show()
```



In [66]:
```
plt.figure(figsize=(8,6))
sns.countplot(y=askot, palette='coolwarm')
plt.title('Number of Users by City')
plt.show()
```



palette='coolwarm': which adds a color gradient to the bars from cool (blue) to warm (red).

Content Based Filtering

In [67]:
```
data = tourism_new
data.sample(5)
```

Out[67]:

	id	name	category	description	city	city_category
136	116	Jurang Tembelan Kanigoro	Taman Hiburan	Jurang Tembelan Kanigoro berada di Desa Wisata...	Yogyakarta	Yogyakarta Taman Hiburan
205	234	Amazing Art World	Budaya	Amazing Art World Bandung, sebuah objek wisata...	Bandung	Bandung Budaya
426	217	Kebun Binatang Bandung	Cagar Alam	Kebun Binatang Bandung merupakan salah satu ob...	Bandung	Bandung Cagar Alam
165	45	Jakarta Aquarium dan Safari	Taman Hiburan	Jika telah mengunjungi Seaworld Ancol, mungkin...	Jakarta	Jakarta Taman Hiburan
208	197	Pantai Jungwok	Bahari	Pantai Jungwok adalah pantai yang terletak di ...	Yogyakarta	Yogyakarta Bahari

In [68]:
```
data.shape
```

Out[68]:
```
(437, 6)
```

In [69]:
```
data.count
```

Out[69]:
```
<bound method DataFrame.count of       id                          name
category  \
0    179               Candi Ratu Boko         Budaya
1    344                 Pantai Marina         Bahari
2      5      Atlantis Water Adventure  Taman Hiburan
3    373        Museum Kereta Ambarawa         Budaya
4    101  Kampung Wisata Sosro Menduran         Budaya
..   ...                           ...            ...
432  370                Benteng Pendem         Budaya
433  350                  Pantai Cipta         Bahari
434   10                  Pulau Tidung         Bahari
435    7        Kebun Binatang Ragunan     Cagar Alam
436  140                  Bendung Lepen  Taman Hiburan

                                           description        city  \
0    Situs Ratu Baka atau Candi Boko (Hanacaraka:ꦕꦤ...  Yogyakarta
1    Pantai Marina (bahasa Jawa: ꦥꦤ꧀ꦠꦲꦶꦩꦫ꧀ꦶꦤ, trans...    Semarang
2    Atlantis Water Adventure atau dikenal dengan A...     Jakarta
3    Museum Kereta Api Ambarawa (bahasa Inggris: In...    Semarang
4    Kampung wisata Sosromenduran merupakan kampung...  Yogyakarta
..                                                 ...         ...
432  Benteng Pendem Cilacap (bahasa Belanda: Kustba...    Semarang
433  Pantai Cipta juga dikenal sebagai Pantai Petik...    Semarang
434  Pulau Tidung adalah salah satu kelurahan di ke...     Jakarta
435  Kebun Binatang Ragunan adalah sebuah kebun bin...     Jakarta
436  Bendung Lepen sendiri dulunya merupakan salura...  Yogyakarta

                city_category
0           Yogyakarta Budaya
1             Semarang Bahari
2       Jakarta Taman Hiburan
3             Semarang Budaya
4           Yogyakarta Budaya
..                        ...
432           Semarang Budaya
433           Semarang Bahari
434            Jakarta Bahari
435        Jakarta Cagar Alam
436  Yogyakarta Taman Hiburan

[437 rows x 6 columns]>
```

In [70]:
```
data.columns
```

Out[70]:
```
Index(['id', 'name', 'category', 'description', 'city', 'city_category'], d
type='object')
```

TF-IDF(Term Frequency-Inverse Document Frequency) Vectorizer:

In [71]:
```
from sklearn.feature_extraction.text import CountVectorizer

cv = CountVectorizer()
```

```
cv.fit(data['city_category'])

print("Features Name: ", list(cv.vocabulary_.keys()))
```

```
Features Name:  ['yogyakarta', 'budaya', 'semarang', 'bahari', 'jakarta', 't
aman', 'hiburan', 'bandung', 'cagar', 'alam', 'surabaya', 'pusat', 'perbelan
jaan', 'tempat', 'ibadah']
```

In [72]:
```
cv_matrix = cv.transform(data['city_category'])

cv_matrix.shape
```

Out[72]:
```
(437, 15)
```

In [73]:
```
cv_matrix
```

Out[73]:
```
<437x15 sparse matrix of type '<class 'numpy.int64'>'
        with 1147 stored elements in Compressed Sparse Row format>
```

The expression cv_matrix.todense() is used to convert a sparse matrix into a dense matrix

In [74]:
```
cv_matrix.todense()
```

Out[74]:
```
matrix([[0, 0, 0, ..., 0, 0, 1],
        [0, 1, 0, ..., 0, 0, 0],
        [0, 0, 0, ..., 1, 0, 0],
        ...,
        [0, 1, 0, ..., 0, 0, 0],
        [1, 0, 0, ..., 0, 0, 0],
        [0, 0, 0, ..., 1, 0, 1]], dtype=int64)
```

In [75]:
```
pd.DataFrame(
    cv_matrix.todense(),
    columns=list(cv.vocabulary_.keys()),
    index = data.name
).sample(5)
```

Out[75]:

name	yogyakarta	budaya	semarang	bahari	jakarta	taman	hiburan	k
Museum Mandala Wangsit Siliwangi	0	0	1	1	0	0	0	
Curug Anom	1	0	1	0	1	0	0	
Air Mancur Menari	0	0	0	0	0	1	0	
Monumen Sanapati	0	0	0	1	0	0	0	
Benteng Pendem	0	0	0	1	0	0	0	

In [76]:
```
pd.DataFrame(
    cv_matrix.todense(),
    columns=list(cv.vocabulary_.keys()),
    index = data.name
).sample(5)
```

Out[76]:

name	yogyakarta	budaya	semarang	bahari	jakarta	taman	hiburan
Dago Dreampark	0	0	1	0	0	1	0
Museum Benteng Vredeburg Yogyakarta	0	0	0	1	0	0	0
Waduk Jatibarang	1	0	0	0	1	0	0
Monumen Bambu Runcing Surabaya	0	0	0	1	0	0	0
Masjid Agung Trans Studio Bandung	0	0	1	0	0	0	1

Cosine Similarity¶

In [77]:
```
from sklearn.metrics.pairwise import cosine_similarity
```

```
cosine_sim1 = cosine_similarity(cv_matrix)
cosine_sim1
```

Out[77]:
```
array([[1.        , 0.        , 0.        , ..., 0.        , 0.        ,
        0.40824829],
       [0.        , 1.        , 0.        , ..., 0.5       , 0.        ,
        0.        ],
       [0.        , 0.        , 1.        , ..., 0.40824829, 0.33333333,
        0.66666667],
       ...,
       [0.        , 0.5       , 0.40824829, ..., 1.        , 0.40824829,
        0.        ],
       [0.        , 0.        , 0.33333333, ..., 0.40824829, 1.        ,
        0.        ],
       [0.40824829, 0.        , 0.66666667, ..., 0.        , 0.        ,
        1.        ]])
```

In [78]:
```
cosine_sim_df = pd.DataFrame(cosine_sim1,index=data['name'],columns=data['na
cosine_sim_df.sample(5,axis=1).sample(10,axis=0)
```

Out[78]:

name	Indonesia Kaya Park	Geoforest Watu Payung Turunan	Kidzania	Selasar Sunaryo Art Space	Taman Mini Indonesia Indah (TMII)
name					
Museum Macan (Modern and Contemporary Art in Nusantara)	0.000000	0.000000	0.408248	0.000000	0.408248
Atlantis Water Adventure	0.666667	0.000000	1.000000	0.666667	1.000000
Monumen Perjuangan Rakyat Jawa Barat	0.000000	0.000000	0.000000	0.408248	0.000000
Perpustakaan Nasional	0.000000	0.000000	0.408248	0.000000	0.408248
Jalan Braga	0.000000	0.000000	0.000000	0.408248	0.000000
Monumen Tugu Pahlawan	0.000000	0.000000	0.000000	0.000000	0.000000
Masjid Raya Bandung	0.000000	0.000000	0.000000	0.333333	0.000000
Gedung Sate	0.000000	0.000000	0.000000	0.408248	0.000000
Pemandian Air Panas Ciater	0.000000	0.666667	0.000000	0.333333	0.000000
Bukit Bintang Yogyakarta	0.666667	0.333333	0.666667	0.666667	0.666667

Recommendation

In [79]:
```
def tourism_recommendations(place_name,similarity_data=cosine_sim_df,items=c
    index = similarity_data.loc[:,place_name].to_numpy().argpartition(range(

    closest = similarity_data.columns[index[-1:-(k+2):-1]]

    closest = closest.drop(place_name,errors='ignore')

    return pd.DataFrame(closest).merge(items).head(k)
```

In [80]:
```
tourism_recommendations("Air Mancur Menari")
```

Out[80]:

	name	category	description	city
0	Taman Prestasi	Taman Hiburan	Taman Prestasi Surabaya merupakan salah satu t...	Surabaya
1	Taman Pelangi	Taman Hiburan	Kalau pelangi biasanya ada di siang hari pasca...	Surabaya
2	Atlantis Land Surabaya	Taman Hiburan	Sejak diresmikan pada bulan Desember 2017, Atl...	Surabaya
3	Taman Buah Surabaya	Taman Hiburan	Wisata Taman Buah Undaan di Surabaya adalah sa...	Surabaya
4	Taman Keputran	Taman Hiburan	Ntah, mengapa nama taman ini disebut dengan ta...	Surabaya

In [103…
```
tourism_recommendations('Ocean Ecopark')
```

Out[103…

	name	category	description	city
0	Taman Impian Jaya Ancol	Taman Hiburan	Taman Impian Jaya Ancol merupakan sebuah objek...	Jakarta
1	Waterboom PIK (Pantai Indah Kapuk)	Taman Hiburan	Waterbom Jakarta merupakan sebuah wahana perma...	Jakarta
2	Sea World	Taman Hiburan	Seaworld Indonesia adalah sebuah miniatur peso...	Jakarta
3	The Escape Hunt	Taman Hiburan	Escape Hunt adalah salah satu tempat rekreasi ...	Jakarta
4	Dunia Fantasi	Taman Hiburan	Dunia Fantasi atau disebut juga Dufan adalah t...	Jakarta

In [ ]:
```

```

Create a Recommendation Model using another approach

SIMILARITY_SCORE METHOD

In [81]:
```
# Merge ratings with tourism data
merged_data = tourism_rating.merge(info_tourism[['Place_Id', 'Place_Name']],
print(merged_data.head())
```

```
   User_Id  Place_Id  Place_Ratings       Place_Name
0        1       179              3  Candi Ratu Boko
1       22       179              4  Candi Ratu Boko
2       40       179              3  Candi Ratu Boko
3       49       179              5  Candi Ratu Boko
4       74       179              3  Candi Ratu Boko
```

In [82]:
```
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.model_selection import train_test_split

user_item_matrix = merged_data.pivot_table(index='User_Id', columns='Place_I

# Calculate cosine similarity
user_similarity = cosine_similarity(user_item_matrix)

# Create a DataFrame for user similarities
user_sim_df = pd.DataFrame(user_similarity, index=user_item_matrix.index, co
```

In [83]:
```
user_sim_df
```

Out[83]:

User_Id	1	2	3	4	5	6	7
User_Id							
1	1.000000	0.058921	0.010902	0.120602	0.041520	0.027104	0.000000
2	0.058921	1.000000	0.048176	0.000000	0.086006	0.029943	0.011765
3	0.010902	0.048176	1.000000	0.028665	0.063653	0.011081	0.065302
4	0.120602	0.000000	0.028665	1.000000	0.032752	0.116877	0.131601
5	0.041520	0.086006	0.063653	0.032752	1.000000	0.166165	0.000000
...	...	...	...	...	...	...	...
296	0.017975	0.142319	0.006124	0.050415	0.026238	0.091350	0.107673
297	0.020515	0.127489	0.068144	0.026971	0.000000	0.026065	0.076805
298	0.057127	0.069972	0.058387	0.073146	0.149857	0.000000	0.074390
299	0.108022	0.163305	0.043580	0.068766	0.041494	0.106903	0.051083
300	0.044273	0.013043	0.030166	0.077607	0.074677	0.119996	0.070719

300 rows × 300 columns

In [84]:

```
def get_user_recommendations(user_id, user_item_matrix, user_sim_df, top_n=5
    # Get similar users
    similar_users = user_sim_df[user_id].sort_values(ascending=False).index[

    # Calculate weighted ratings from similar users
    weighted_ratings = {}
    for similar_user in similar_users:
        similar_user_ratings = user_item_matrix.loc[similar_user]
        for place_id, rating in similar_user_ratings.items():
            if rating > 0:  # Consider only rated places
                if place_id not in weighted_ratings:
                    weighted_ratings[place_id] = 0
                weighted_ratings[place_id] += rating * user_sim_df[user_id][

    # Sort and return top N recommendations
    recommended_places = sorted(weighted_ratings.items(), key=lambda x: x[1]
    return recommended_places[:top_n]

# Example usage
user_id = 1
recommendations = get_user_recommendations(user_id, user_item_matrix, user_s
print("Recommended Places:")
for place_id, score in recommendations:
    place_name = info_tourism[info_tourism['Place_Id'] == place_id]['Place_N
    print(f"{place_name} (Score: {score})")
```

```
Recommended Places:
Taman Sungai Mudal (Score: 10.352040605708352)
Museum Mandala Wangsit Siliwangi (Score: 9.931266605765721)
Grand Maerakaca (Score: 9.915958218381787)
Masjid Agung Trans Studio Bandung (Score: 8.848300426709493)
Museum Gedung Sate (Score: 8.45126876562424)
```

In [ ]:

In [ ]:

In [85]:
```python
import pandas as pd

# Merge tourism_rating with place information
merged_df = tourism_rating.merge(preparation, on="Place_Id", how="left")


print(merged_df.head())
```

```
   User_Id_x  Place_Id  Place_Ratings_x  User_Id_y  Place_Ratings_y  \
0          1       179                3          1                3
1          1       344                2          1                2
2          1         5                5          1                5
3          1       373                3          1                3
4          1       101                4          1                4

                     Place_Name  \
0                Candi Ratu Boko
1                  Pantai Marina
2       Atlantis Water Adventure
3         Museum Kereta Ambarawa
4  Kampung Wisata Sosro Menduran

                                          Description        City  \
0  Situs Ratu Baka atau Candi Boko (Hanacaraka:ꦕꦤ...  Yogyakarta
1  Pantai Marina (bahasa Jawa: ꦥꦤ꧀ꦠꦲꦶꦩꦫ꧀ꦶꦤ, trans...    Semarang
2  Atlantis Water Adventure atau dikenal dengan A...     Jakarta
3  Museum Kereta Api Ambarawa (bahasa Inggris: In...    Semarang
4  Kampung wisata Sosromenduran merupakan kampung...  Yogyakarta

        Category          city_category
0        Budaya      Yogyakarta Budaya
1        Bahari        Semarang Bahari
2  Taman Hiburan  Jakarta Taman Hiburan
3        Budaya        Semarang Budaya
4        Budaya      Yogyakarta Budaya
```

In [86]:
```python
merged_df
```

Out[86]:

	User_Id_x	Place_Id	Place_Ratings_x	User_Id_y	Place_Ratings_y	Place_
0	1	179	3	1	3	Canc
1	1	344	2	1	2	I
2	1	5	5	1	5	A Adv
3	1	373	3	1	3	Mu Amb
4	1	101	4	1	4	Kar Wisata Mer
...	...	...	...	...	...	
9995	300	425	2	20	4	Wat Ke Sur
9996	300	64	4	6	5	Mu Sasmit Ahma
9997	300	311	3	24	5	The Ma
9998	300	279	4	17	4	Masjid Trans Ba
9999	300	163	2	13	4	Watu Man

10000 rows × 10 columns

In [87]:
```
merged_df.columns
```

Out[87]:
```
Index(['User_Id_x', 'Place_Id', 'Place_Ratings_x', 'User_Id_y',
       'Place_Ratings_y', 'Place_Name', 'Description', 'City', 'Category',
       'city_category'],
      dtype='object')
```

In [88]:
```
user_place_matrix = merged_df.pivot_table(index='User_Id_x', columns='Place_
```

In [89]:
```
user_place_matrix
```

Out[89]:

Place_Id	1	2	3	4	5	6	7	8	9	10	...	428	429
User_Id_x													
1	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	NaN	...	3.0	NaN
2	NaN	5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
4	NaN	NaN	NaN	4.0	5.0	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
5	NaN	NaN	NaN	4.0	NaN	NaN	NaN	NaN	5.0	NaN	...	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...
296	NaN	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	...	NaN	NaN
297	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	NaN	...	NaN	3.0
298	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN
299	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	...	NaN	NaN
300	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0	NaN	NaN	...	NaN	NaN

300 rows × 437 columns

In [90]:
```
pip install implicit
```

```
Requirement already satisfied: implicit in c:\users\shaw3\anaconda3\lib\site
-packages (0.7.2)
Requirement already satisfied: tqdm>=4.27 in c:\users\shaw3\anaconda3\lib\si
te-packages (from implicit) (4.64.1)
Requirement already satisfied: numpy>=1.17.0 in c:\users\shaw3\anaconda3\lib
\site-packages (from implicit) (1.21.6)
Requirement already satisfied: scipy>=0.16 in c:\users\shaw3\anaconda3\lib\s
ite-packages (from implicit) (1.9.1)
Requirement already satisfied: threadpoolctl in c:\users\shaw3\anaconda3\lib
\site-packages (from implicit) (2.2.0)
Requirement already satisfied: colorama in c:\users\shaw3\anaconda3\lib\site
-packages (from tqdm>=4.27->implicit) (0.4.5)
Note: you may need to restart the kernel to use updated packages.
```

In [93]:
```
import pandas as pd
import numpy as np
import implicit
```

In [94]:
```
pip install scipy
```

```
Requirement already satisfied: scipy in c:\users\shaw3\anaconda3\lib\site-pa
ckages (1.9.1)
Requirement already satisfied: numpy<1.25.0,>=1.18.5 in c:\users\shaw3\anaco
nda3\lib\site-packages (from scipy) (1.21.6)
Note: you may need to restart the kernel to use updated packages.
```

In [95]:
```
import scipy.sparse
```

In [96]:
```
# Convert the DataFrame to a sparse matrix format for implicit
user_place_sparse = scipy.sparse.csr_matrix(user_place_matrix.values)
```

In [97]:
```
user_place_sparse
```

Out[97]:
```
<300x437 sparse matrix of type '<class 'numpy.float64'>'
        with 9597 stored elements in Compressed Sparse Row format>
```

In [98]:
```
# Check for NaN values in the matrix
print(user_place_matrix.isna().sum().sum())
```

```
0
```

In [99]:
```
# Fill NaN values with 0 (It indicats no interaction)
user_place_matrix = user_place_matrix.fillna(0)
```

In [100…
```
user_place_matrix
```

Out[100…

Place_Id	1	2	3	4	5	6	7	8	9	10	...	428	429	430	431
User_Id_x															
1	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	0.0	...	3.0	0.0	0.0	0.0
2	0.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	4.0	5.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	5.0	0.0	...	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
296	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
297	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	...	0.0	4.0	5.0	0.0
298	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0
299	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	...	0.0	0.0	2.0	0.0
300	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0

300 rows × 437 columns

In [101…
```
# Convert the DataFrame to a sparse matrix format for implicit
import scipy.sparse
```

```
user_place_sparse = scipy.sparse.csr_matrix(user_place_matrix.values)
```

In [102…]
```
# Train the ALS (Alternating Least Squares) model from implicit
model = implicit.als.AlternatingLeastSquares(factors=50, regularization=0.01
model.fit(user_place_sparse.T)  # Transpose the matrix
```

```
C:\Users\shaw3\anaconda3\lib\site-packages\implicit\cpu\als.py:95: RuntimeWa
rning: OpenBLAS is configured to use 12 threads. It is highly recommended to
disable its internal threadpool by setting the environment variable 'OPENBLA
S_NUM_THREADS=1' or by calling 'threadpoolctl.threadpool_limits(1, "blas")'.
Having OpenBLAS use a threadpool can lead to severe performance issues here.
  check_blas_config()
C:\Users\shaw3\anaconda3\lib\site-packages\implicit\cpu\als.py:95: RuntimeWa
rning: Intel MKL BLAS is configured to use 6 threads. It is highly recommend
ed to disable its internal threadpool by setting the environment variable 'M
KL_NUM_THREADS=1' or by callng 'threadpoolctl.threadpool_limits(1, "blas")'.
Having MKL use a threadpool can lead to severe performance issues
  check_blas_config()
C:\Users\shaw3\anaconda3\lib\site-packages\implicit\utils.py:164: ParameterW
arning: Method expects CSR input, and was passed csc_matrix instead. Convert
ing to CSR took 0.0010142326354980469 seconds
  warnings.warn(
  0%|          | 0/10 [00:00<?, ?it/s]
```

In [ ]: