codebasics resume project 13

November 20, 2024

[2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df_city_target_passenger_rating = pd.read_csv(r"C:
 ↪\Users\Pinki\Downloads\RPC13_Input_For_Participants\RPC13_Input_For_Participants\datasets\c
 ↪csv")
df_dim_city = pd.read_csv(r"C:
 ↪\Users\Pinki\Downloads\RPC13_Input_For_Participants\RPC13_Input_For_Participants\datasets\c
 ↪csv")
df_dim_date = pd.read_csv(r"C:
 ↪\Users\Pinki\Downloads\RPC13_Input_For_Participants\RPC13_Input_For_Participants\datasets\c
 ↪csv")
df_reapeat_trip_distribution = pd.read_csv(r"C:
 ↪\Users\Pinki\Downloads\RPC13_Input_For_Participants\RPC13_Input_For_Participants\datasets\c
 ↪csv")
df_fact_passenger_summary = pd.read_csv(r"C:
 ↪\Users\Pinki\Downloads\RPC13_Input_For_Participants\RPC13_Input_For_Participants\datasets\c
 ↪csv")
df_fact_trips = pd.read_csv(r"C:
 ↪\Users\Pinki\Downloads\RPC13_Input_For_Participants\RPC13_Input_For_Participants\datasets\c
 ↪csv")
df_monthly_target_new_passengers = pd.read_csv(r"C:
 ↪\Users\Pinki\Downloads\RPC13_Input_For_Participants\RPC13_Input_For_Participants\datasets\c
 ↪csv")
df_monthly_target_trips = pd.read_csv(r"C:
 ↪\Users\Pinki\Downloads\RPC13_Input_For_Participants\RPC13_Input_For_Participants\datasets\c
 ↪csv")
```

[3]:
```
df_reapeat_trip_distribution.head()
```

[3]:

	month	city_id	trip_count	repeat_passenger_count
0	2024-01-01	AP01	10-Trips	7
1	2024-01-01	AP01	2-Trips	352
2	2024-01-01	AP01	3-Trips	158
3	2024-01-01	AP01	4-Trips	53
4	2024-01-01	AP01	5-Trips	38

[4]:
```
df_reapeat_trip_distribution.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 540 entries, 0 to 539
Data columns (total 4 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   month                   540 non-null    object
 1   city_id                 540 non-null    object
 2   trip_count              540 non-null    object
 3   repeat_passenger_count  540 non-null    int64
dtypes: int64(1), object(3)
memory usage: 17.0+ KB
```

[5]:
```
df_reapeat_trip_distribution['trip_count'] =␣
↪df_reapeat_trip_distribution['trip_count'].str.extract('(\d+)')
```

[6]:
```
df_reapeat_trip_distribution['trip_count'] = pd.
↪to_numeric(df_reapeat_trip_distribution['trip_count'], errors='coerce')
```

[7]:
```
df_reapeat_trip_distribution['TotalCount'] =␣
↪df_reapeat_trip_distribution['trip_count'] +␣
↪df_reapeat_trip_distribution['repeat_passenger_count']
```

[8]:
```
df_reapeat_trip_distribution.head()
```

[8]:

	month	city_id	trip_count	repeat_passenger_count	TotalCount
0	2024-01-01	AP01	10	7	17
1	2024-01-01	AP01	2	352	354
2	2024-01-01	AP01	3	158	161
3	2024-01-01	AP01	4	53	57
4	2024-01-01	AP01	5	38	43

[9]:
```
merged_df = pd.merge(df_reapeat_trip_distribution, df_dim_city, on='city_id')
```

[10]:
```
merged_df.head()
```

[10]:

	month	city_id	trip_count	repeat_passenger_count	TotalCount	city_name
0	2024-01-01	AP01	10	7	17	Visakhapatnam
1	2024-01-01	AP01	2	352	354	Visakhapatnam
2	2024-01-01	AP01	3	158	161	Visakhapatnam
3	2024-01-01	AP01	4	53	57	
4	2024-01-01	AP01	5	38	43	

```
3  Visakhapatnam
4  Visakhapatnam
```

1 Top 3 and bottom 3 cities by total trips

[11]:
```
top_3_cities = merged_df.nlargest(3, 'TotalCount')
```

[12]:
```
print("Top 3 Cities by Total Count:")
print(top_3_cities)
```

```
Top 3 Cities by Total Count:
          month city_id  trip_count  repeat_passenger_count  TotalCount  \
388  2024-02-01    RJ01           2                     999        1001
406  2024-04-01    RJ01           2                     991         993
307  2024-05-01    KL01           2                     971         973

    city_name
388    Jaipur
406    Jaipur
307     Kochi
```

[13]:
```
bottom_3_cities = merged_df.nsmallest(3, 'TotalCount')
```

[14]:
```
print("Bottom 3 Cities by Total Count:")
print(bottom_3_cities)
```

```
Bottom 3 Cities by Total Count:
          month city_id  trip_count  repeat_passenger_count  TotalCount  \
222  2024-01-01    KA01           7                       2           9
224  2024-01-01    KA01           9                       0           9
249  2024-04-01    KA01           7                       2           9

    city_name
222    Mysore
224    Mysore
249    Mysore
```

[15]:
```
df_fact_trips.head()
```

[15]:

	trip_id	date	city_id	passenger_type	distance_travelled(km)	fare_amount	passenger_rating	driver_rating
0	TRPLUC240113d55de2fb	2024-01-13	UP01	repeated				
1	TRPVAD240129a3b6dba8	2024-01-29	GJ02	repeated				
2	TRPCOI240107a42430fb	2024-01-07	TN01	repeated				
3	TRPKOC240325d7601389	2024-03-25	KL01	repeated				
4	TRPVIS2406027be97166	2024-06-02	AP01	new				

```
0                      11          158                5              5
1                       7           74                5              5
2                      11          155                8              8
3                      36          427                9             10
4                      17          265                8              8
```

[16]:
```
df_fact_trips.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 425903 entries, 0 to 425902
Data columns (total 8 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   trip_id                 425903 non-null  object
 1   date                    425903 non-null  object
 2   city_id                 425903 non-null  object
 3   passenger_type          425903 non-null  object
 4   distance_travelled(km)  425903 non-null  int64
 5   fare_amount             425903 non-null  int64
 6   passenger_rating        425903 non-null  int64
 7   driver_rating           425903 non-null  int64
dtypes: int64(4), object(4)
memory usage: 26.0+ MB
```

[17]:
```
merged_df1 = pd.merge(df_fact_trips, df_dim_city, on='city_id')
```

[18]:
```
merged_df1.head()
```

[18]:
```
                trip_id        date city_id passenger_type  \
0  TRPLUC240113d55de2fb  2024-01-13    UP01       repeated
1  TRPLUC240327a61cfe66  2024-03-27    UP01            new
2  TRPLUC240322c73f8439  2024-03-22    UP01       repeated
3  TRPLUC240420c114cb02  2024-04-20    UP01       repeated
4  TRPLUC2405094011dfa0  2024-05-09    UP01       repeated

   distance_travelled(km)  fare_amount  passenger_rating  driver_rating  \
0                      11          158                 5              5
1                       5           71                 7              8
2                      15          161                 6              7
3                      13          182                 7              5
4                      18          188                 6              5

  city_name
0   Lucknow
1   Lucknow
2   Lucknow
3   Lucknow
4   Lucknow
```

2 Average fare per trip by city

[19]:
```
city_metrics = merged_df1.groupby('city_name').agg(
    average_fare_per_trip=('fare_amount', 'mean'),
    average_distance=('distance_travelled(km)', 'mean'),
    total_trips=('fare_amount', 'count')  # Optional: Count trips for each city
).reset_index()
```

[20]:
```
print(city_metrics)
```

```
       city_name  average_fare_per_trip  average_distance  total_trips
0     Chandigarh             283.686950         23.518714        38981
1     Coimbatore             166.982183         14.979198        21104
2         Indore             179.838609         16.502473        42456
3         Jaipur             483.918128         30.023125        76888
4          Kochi             335.245079         24.065461        50702
5        Lucknow             147.180376         12.512963        64299
6         Mysore             249.707168         16.496921        16238
7          Surat             117.272925         10.997247        54843
8       Vadodara             118.566165         11.517736        32026
9  Visakhapatnam             282.672284         22.553938        28366
```

[21]:
```
highest_fare_city = city_metrics.nlargest(1, 'average_fare_per_trip')
```

[22]:
```
print(highest_fare_city)
```

```
  city_name  average_fare_per_trip  average_distance  total_trips
3    Jaipur             483.918128         30.023125        76888
```

[23]:
```
lowest_fare_city = city_metrics.nsmallest(1, 'average_fare_per_trip')
```

[24]:
```
print(lowest_fare_city)
```

```
  city_name  average_fare_per_trip  average_distance  total_trips
7     Surat             117.272925         10.997247        54843
```

3 Average ratings by city and passenger type

[25]:
```
rating_metrics = merged_df1.groupby(['city_name', 'city_id', 'passenger_type']).
  ↪agg(
    passenger_avg_rating=('passenger_rating', 'mean'),
    driver_avg_rating=('driver_rating', 'mean')
).reset_index()
```

[262]:
```
pd.set_option("display.max_columns", None)  # Show all columns
pd.set_option("display.width", 1000)
```

[263]:
```
print(rating_metrics)
```

```
        city_name city_id passenger_type  passenger_avg_rating
driver_avg_rating
0      Chandigarh    CH01            new              8.489158
7.992120
1      Chandigarh    CH01       repeated              7.493798
7.472824
2      Coimbatore    TN01            new              8.485788
7.990604
3      Coimbatore    TN01       repeated              7.475457
7.480778
4          Indore    MP01            new              8.485837
7.970800
5          Indore    MP01       repeated              7.473961
7.477404
6          Jaipur    RJ01            new              8.985018
8.988246
7          Jaipur    RJ01       repeated              7.991042
8.984790
8           Kochi    KL01            new              8.987394
8.985350
9           Kochi    KL01       repeated              8.003665
8.989830
10        Lucknow    UP01            new              7.977429
6.990406
11        Lucknow    UP01       repeated              5.985741
6.491663
12         Mysore    KA01            new              8.982964
8.982878
13         Mysore    KA01       repeated              7.978495
8.965767
14          Surat    GJ01            new              7.984173
6.994925
15          Surat    GJ01       repeated              5.995511
6.479441
16       Vadodara    GJ02            new              7.979263
7.004147
17       Vadodara    GJ02       repeated              5.978629
6.481072
18  Visakhapatnam    AP01            new              8.976151
8.979995
19  Visakhapatnam    AP01       repeated              7.989628
8.992701
```

[27]:
```
highest_average_passenger_rating_city = rating_metrics.nlargest(1,␣
↪'passenger_avg_rating')
```

[28]:
```
highest_average_driver_rating_city = rating_metrics.nlargest(1,␣
↪'driver_avg_rating')
```

[29]:
```
print(highest_average_passenger_rating_city)
```

```
  city_name city_id passenger_type  passenger_avg_rating  driver_avg_rating
8     Kochi    KL01            new              8.987394            8.98535
```

[30]:
```
print(highest_average_driver_rating_city)
```

```
         city_name city_id passenger_type  passenger_avg_rating  \
19  Visakhapatnam    AP01       repeated              7.989628

    driver_avg_rating
19           8.992701
```

[31]:
```
lowest_average_passenger_rating_city = rating_metrics.nsmallest(1,␣
↪'passenger_avg_rating')
```

[32]:
```
lowest_average_driver_rating_city = rating_metrics.nsmallest(1,␣
↪'driver_avg_rating')
```

[33]:
```
print(lowest_average_passenger_rating_city)
```

```
   city_name city_id passenger_type  passenger_avg_rating  driver_avg_rating
17  Vadodara    GJ02       repeated              5.978629           6.481072
```

[34]:
```
print(lowest_average_driver_rating_city)
```

```
   city_name city_id passenger_type  passenger_avg_rating  driver_avg_rating
15     Surat    GJ01       repeated              5.995511           6.479441
```

4 Peak and low demand Months by city

[35]:
```
merged_df['month'] = pd.to_datetime(merged_df['month'])
```

[36]:
```
merged_df['month_number'] = merged_df['month'].dt.month
```

[37]:
```
merged_df['month_name'] = merged_df['month'].dt.strftime('%B')
```

[38]:
```
demand_metrics = merged_df.groupby(['city_name', 'city_id', 'month_name']).agg(
    month_highest_total_trip=('TotalCount', 'max'),
   month_lowest_total_trip=('TotalCount', 'min')
).reset_index()
```

[260]:
```
pd.set_option("display.max_columns", None)  # Show all columns
pd.set_option("display.width", 1000)
```

[261]:
```
print(demand_metrics)
```

	city_name	city_id	month_name	month_highest_total_trip	month_lowest_total_trip
0	Chandigarh	CH01	April	257	19
1	Chandigarh	CH01	February	298	23
2	Chandigarh	CH01	January	198	23
3	Chandigarh	CH01	June	222	28
4	Chandigarh	CH01	March	335	29
5	Chandigarh	CH01	May	340	25
6	Coimbatore	TN01	April	94	22
7	Coimbatore	TN01	February	71	13
8	Coimbatore	TN01	January	102	11
9	Coimbatore	TN01	June	91	15
10	Coimbatore	TN01	March	113	15
11	Coimbatore	TN01	May	119	13
12	Indore	MP01	April	423	27
13	Indore	MP01	February	334	28
14	Indore	MP01	January	401	23
15	Indore	MP01	June	378	24
16	Indore	MP01	March	339	35
17	Indore	MP01	May	615	32
18	Jaipur	RJ01	April	993	26
19	Jaipur	RJ01	February	1001	19
20	Jaipur	RJ01	January	707	23
21	Jaipur	RJ01	June	417	20
22	Jaipur	RJ01	March	845	27

23	Jaipur	RJ01	May	904
34				
24	Kochi	KL01	April	848
20				
25	Kochi	KL01	February	441
20				
26	Kochi	KL01	January	412
16				
27	Kochi	KL01	June	467
16				
28	Kochi	KL01	March	506
21				
29	Kochi	KL01	May	973
24				
30	Lucknow	UP01	April	383
28				
31	Lucknow	UP01	February	329
30				
32	Lucknow	UP01	January	357
24				
33	Lucknow	UP01	June	323
27				
34	Lucknow	UP01	March	329
27				
35	Lucknow	UP01	May	380
28				
36	Mysore	KA01	April	98
9				
37	Mysore	KA01	February	93
10				
38	Mysore	KA01	January	77
9				
39	Mysore	KA01	June	180
10				
40	Mysore	KA01	March	102
10				
41	Mysore	KA01	May	182
10				
42	Surat	GJ01	April	372
27				
43	Surat	GJ01	February	313
26				
44	Surat	GJ01	January	371
33				
45	Surat	GJ01	June	349
21				
46	Surat	GJ01	March	329
30				

```
47          Surat    GJ01        May                     374
29
48       Vadodara    GJ02      April                     183
20
49       Vadodara    GJ02   February                     143
18
50       Vadodara    GJ02    January                     114
18
51       Vadodara    GJ02       June                     139
24
52       Vadodara    GJ02      March                     156
22
53       Vadodara    GJ02        May                     166
27
54  Visakhapatnam    AP01      April                     429
17
55  Visakhapatnam    AP01   February                     435
14
56  Visakhapatnam    AP01    January                     354
16
57  Visakhapatnam    AP01       June                     403
15
58  Visakhapatnam    AP01      March                     507
16
59  Visakhapatnam    AP01        May                     502
15
```

5 Weekend vs weekday trip demand by city

[40]:
```
df_dim_date.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 182 entries, 0 to 181
Data columns (total 4 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   date            182 non-null    object
 1   start_of_month  182 non-null    object
 2   month_name      182 non-null    object
 3   day_type        182 non-null    object
dtypes: object(4)
memory usage: 5.8+ KB
```

[41]:
```
merged_df2 = pd.merge(merged_df , df_dim_date, on='month_name')
```

[42]:
```
merged_df2.head()
```

[42]:
```
        month city_id  trip_count  repeat_passenger_count  TotalCount  \
0 2024-01-01    AP01          10                       7          17
1 2024-01-01    AP01          10                       7          17
2 2024-01-01    AP01          10                       7          17
3 2024-01-01    AP01          10                       7          17
4 2024-01-01    AP01          10                       7          17

        city_name  month_number month_name        date start_of_month day_type
0  Visakhapatnam             1    January  2024-01-01     2024-01-01  Weekday
1  Visakhapatnam             1    January  2024-01-02     2024-01-01  Weekday
2  Visakhapatnam             1    January  2024-01-03     2024-01-01  Weekday
3  Visakhapatnam             1    January  2024-01-04     2024-01-01  Weekday
4  Visakhapatnam             1    January  2024-01-05     2024-01-01  Weekday
```

[43]:
```
print(merged_df2['day_type'].unique())
```

```
['Weekday' 'Weekend']
```

[44]:
```
start_date = '2024-01-01'
end_date = '2024-06-30'
filtered_df = merged_df2[(merged_df2['date'] >= start_date) &␣
 ↪(merged_df2['date'] <= end_date)]
```

[45]:
```
demand_metrics1 = filtered_df.groupby(['city_name', 'city_id',␣
 ↪'day_type','month_name']).agg(
    highest_total_trip=('TotalCount', 'max'),
   lowest_total_trip=('TotalCount', 'min')
).reset_index()
```

[269]:
```
pd.set_option("display.max_columns", None)  # Show all columns
pd.set_option("display.width", 1000)
```

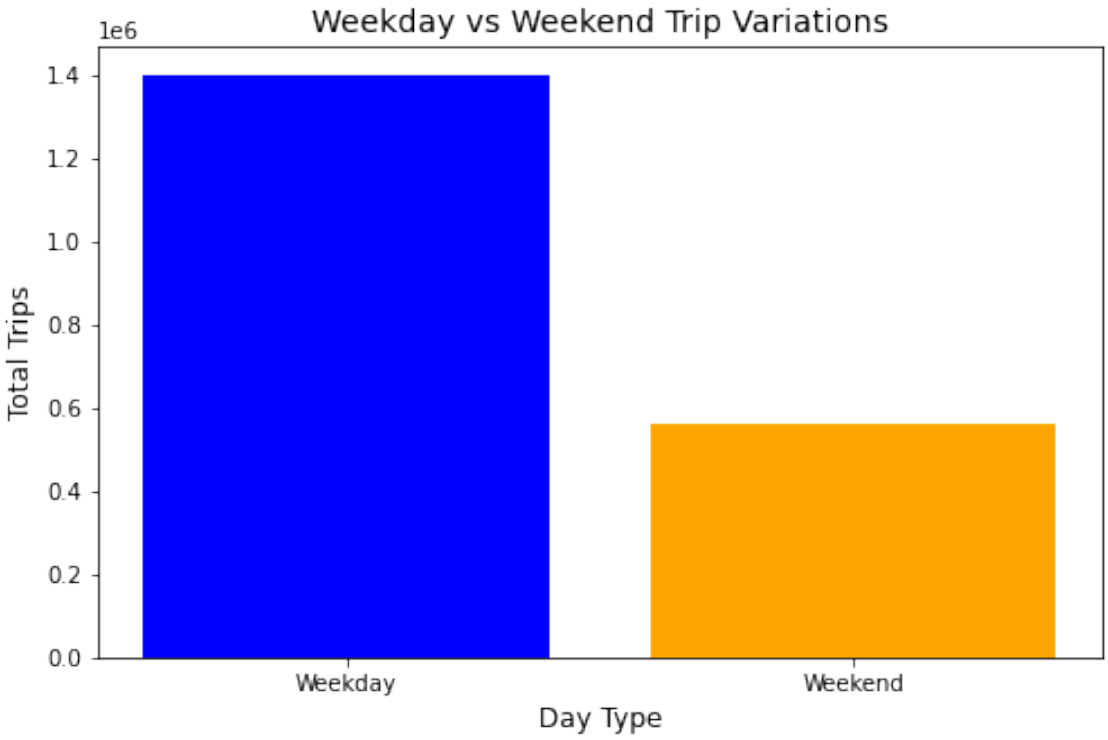
[270]:
```
print(demand_metrics1.head(10))
```

```
    city_name city_id day_type month_name  highest_total_trip  lowest_total_trip
0  Chandigarh    CH01  Weekday      April                 257                 19
1  Chandigarh    CH01  Weekday   February                 298                 23
2  Chandigarh    CH01  Weekday    January                 198                 23
3  Chandigarh    CH01  Weekday       June                 222                 28
4  Chandigarh    CH01  Weekday      March                 335                 29
5  Chandigarh    CH01  Weekday        May                 340                 25
6  Chandigarh    CH01  Weekend      April                 257                 19
7  Chandigarh    CH01  Weekend   February                 298                 23
8  Chandigarh    CH01  Weekend    January                 198                 23
9  Chandigarh    CH01  Weekend       June                 222                 28
```

[47]:
```
day_type_metrics = merged_df2.groupby('day_type').agg(
    total_trips=('TotalCount', 'sum')
```

```
).reset_index()
```

[48]:
```
plt.figure(figsize=(8, 5))
plt.bar(day_type_metrics['day_type'], day_type_metrics['total_trips'],␣
 ↪color=['blue', 'orange'])
plt.xlabel('Day Type', fontsize=12)
plt.ylabel('Total Trips', fontsize=12)
plt.title('Weekday vs Weekend Trip Variations', fontsize=14)
plt.xticks(fontsize=10)
plt.yticks(fontsize=10)
plt.show()
```



6 Repeat passenger frequency and city contribution analysis

[77]:
```
valuewise_analysis = merged_df.groupby(["city_name","city_id" ,"trip_count"]).
 ↪agg(
    total_trips=("trip_count", "sum"),
    repeat_passenger_trips=("repeat_passenger_count", "sum")
).reset_index()
```

[82]:
```
valuewise_analysis["repeat_passenger_trips_percentage"] = (
```

```
    valuewise_analysis["repeat_passenger_trips"] /␣
  ↪valuewise_analysis["total_trips"] * 100
)
```

[83]:
```
valuewise_analysis = valuewise_analysis.sort_values(
    by="repeat_passenger_trips_percentage", ascending=False
).reset_index(drop=True)
```

[85]:
```
valuewise_analysis["repeat_passenger_trips_percentage"] = valuewise_analysis[
    "repeat_passenger_trips_percentage"
].apply(lambda x: f"{x:.2f}%")
```

[267]:
```
pd.set_option("display.max_columns", None)  # Show all columns
pd.set_option("display.width", 1000)
```

[268]:
```
print(valuewise_analysis.head(20))
```

```
       city_name city_id  trip_count  total_trips  repeat_passenger_trips
repeat_passenger_trips_percentage
0          Jaipur    RJ01           2           12                    4855
40458.33%
1           Kochi    KL01           2           12                    3635
30291.67%
2   Visakhapatnam    AP01           2           12                    2618
21816.67%
3          Indore    MP01           2           12                    2478
20650.00%
4      Chandigarh    CH01           2           12                    1638
13650.00%
5          Jaipur    RJ01           3           18                    2007
11150.00%
6           Kochi    KL01           3           18                    1857
10316.67%
7          Indore    MP01           3           18                    1637
9094.44%
8         Lucknow    UP01           3           18                    1417
7872.22%
9         Lucknow    UP01           2           12                     927
7725.00%
10  Visakhapatnam    AP01           3           18                    1275
7083.33%
11          Surat    GJ01           2           12                     843
7025.00%
12          Surat    GJ01           3           18                    1232
6844.44%
13        Lucknow    UP01           4           24                    1555
6479.17%
14         Mysore    KA01           2           12                     720
```

```
6000.00%
15          Surat    GJ01           4           24                    1430
5958.33%
16        Lucknow    UP01           5           30                    1768
5893.33%
17          Surat    GJ01           5           30                    1706
5686.67%
18     Chandigarh    CH01           3           18                     976
5422.22%
19        Lucknow    UP01           6           36                    1937
5380.56%
```

[ ]:
```
# Jaipur has the maximum repeat passenger trips percentage for trip counts 2.
# Higher the repeat passenger trips and total trips higher the business.
```

7 Monthly target acheivement analysis

[89]:
```
df_monthly_target_trips.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60 entries, 0 to 59
Data columns (total 3 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   month               60 non-null     object
 1   city_id             60 non-null     object
 2   total_target_trips  60 non-null     int64 
dtypes: int64(1), object(2)
memory usage: 1.5+ KB
```

[90]:
```
df_monthly_target_trips.head()
```

[90]:

	month	city_id	total_target_trips
0	2024-03-01	MP01	7000
1	2024-05-01	KA01	2500
2	2024-04-01	UP01	11000
3	2024-02-01	GJ02	6000
4	2024-05-01	KL01	9000

[91]:
```
merged_df3 = pd.merge(df_monthly_target_trips, df_dim_city, on='city_id')
```

[92]:
```
merged_df3.head()
```

[92]:

	month	city_id	total_target_trips	city_name
0	2024-03-01	MP01	7000	Indore
1	2024-04-01	MP01	7500	Indore
2	2024-02-01	MP01	7000	Indore

```
3  2024-05-01    MP01                7500    Indore
4  2024-06-01    MP01                7500    Indore
```

[94]:
```
merged_df3['month'] = pd.to_datetime(merged_df3['month'])
```

[95]:
```
merged_df3['month_number'] = merged_df3['month'].dt.month
```

[96]:
```
merged_df3['month_name'] = merged_df3['month'].dt.strftime('%B')
```

[97]:
```
merged_df3.head()
```

[97]:
```
       month city_id  total_target_trips city_name  month_number month_name
0 2024-01-01    MP01                7000    Indore             1    January
1 2024-01-01    MP01                7500    Indore             1    January
2 2024-01-01    MP01                7000    Indore             1    January
3 2024-01-01    MP01                7500    Indore             1    January
4 2024-01-01    MP01                7500    Indore             1    January
```

[98]:
```
monthly_target_analysis = merged_df3.groupby(["city_name","city_id"␣
↪,"month_name"]).agg(
    total_target_trips=("total_target_trips", "sum"),
).reset_index()
```

[99]:
```
print(monthly_target_analysis)
```

```
        city_name city_id month_name  total_target_trips
0      Chandigarh    CH01        May               39000
1      Coimbatore    TN01       June               10500
2      Coimbatore    TN01        May               10500
3          Indore    MP01    January               43500
4          Jaipur    RJ01    January               67500
5           Kochi    KL01      April               24000
6           Kochi    KL01      March               25500
7         Lucknow    UP01   February               72000
8          Mysore    KA01   February                6500
9          Mysore    KA01    January                7000
10          Surat    GJ01       June               57000
11       Vadodara    GJ02      March               37500
12  Visakhapatnam    AP01      April               28500
```

[100]:
```
df_monthly_target_new_passengers.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60 entries, 0 to 59
Data columns (total 3 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   month                 60 non-null     object
 1   city_id               60 non-null     object
```

```
 2   target_new_passengers  60 non-null     int64
dtypes: int64(1), object(2)
memory usage: 1.5+ KB
```

[101]:
```
df_monthly_target_new_passengers.head()
```

[101]:

	month	city_id	target_new_passengers
0	2024-05-01	GJ01	1500
1	2024-05-01	GJ02	1500
2	2024-03-01	GJ01	2000
3	2024-05-01	UP01	2000
4	2024-05-01	MP01	2000

[103]:
```
merged_df4 = pd.merge(df_monthly_target_new_passengers, df_dim_city,␣
↪on='city_id')
```

[104]:
```
merged_df4.head()
```

[104]:

	month	city_id	target_new_passengers	city_name
0	2024-05-01	GJ01	1500	Surat
1	2024-03-01	GJ01	2000	Surat
2	2024-02-01	GJ01	2000	Surat
3	2024-04-01	GJ01	1500	Surat
4	2024-01-01	GJ01	2000	Surat

[105]:
```
merged_df4['month'] = pd.to_datetime(merged_df4['month'])
```

[106]:
```
merged_df4['month_number'] = merged_df4['month'].dt.month
```

[107]:
```
merged_df4['month_name'] = merged_df4['month'].dt.strftime('%B')
```

[108]:
```
merged_df4.head()
```

[108]:

	month	city_id	target_new_passengers	city_name	month_number	month_name
0	2024-05-01	GJ01	1500	Surat	5	May
1	2024-03-01	GJ01	2000	Surat	3	March
2	2024-02-01	GJ01	2000	Surat	2	February
3	2024-04-01	GJ01	1500	Surat	4	April
4	2024-01-01	GJ01	2000	Surat	1	January

[109]:
```
monthly_target_analysis = merged_df4.groupby(["city_name","city_id"␣
↪,"month_name"]).agg(
    target_new_passengers=("target_new_passengers", "sum"),
).reset_index()
```

[110]:
```
print(monthly_target_analysis)
```

```
       city_name city_id month_name  target_new_passengers
0      Chandigarh    CH01      April                   3000
```

1	Chandigarh	CH01	February	4000
2	Chandigarh	CH01	January	4000
3	Chandigarh	CH01	June	3000
4	Chandigarh	CH01	March	4000
5	Chandigarh	CH01	May	3000
6	Coimbatore	TN01	April	1000
7	Coimbatore	TN01	February	1500
8	Coimbatore	TN01	January	1500
9	Coimbatore	TN01	June	1000
10	Coimbatore	TN01	March	1500
11	Coimbatore	TN01	May	1000
12	Indore	MP01	April	2000
13	Indore	MP01	February	2700
14	Indore	MP01	January	2700
15	Indore	MP01	June	2000
16	Indore	MP01	March	2700
17	Indore	MP01	May	2000
18	Jaipur	RJ01	April	6000
19	Jaipur	RJ01	February	12000
20	Jaipur	RJ01	January	12000
21	Jaipur	RJ01	June	6000
22	Jaipur	RJ01	March	12000
23	Jaipur	RJ01	May	6000
24	Kochi	KL01	April	4000
25	Kochi	KL01	February	5000
26	Kochi	KL01	January	5000
27	Kochi	KL01	June	4000
28	Kochi	KL01	March	5000
29	Kochi	KL01	May	4000
30	Lucknow	UP01	April	2000
31	Lucknow	UP01	February	3200
32	Lucknow	UP01	January	3200
33	Lucknow	UP01	June	2000
34	Lucknow	UP01	March	3200
35	Lucknow	UP01	May	2000
36	Mysore	KA01	April	2000
37	Mysore	KA01	February	2000
38	Mysore	KA01	January	2000
39	Mysore	KA01	June	2000
40	Mysore	KA01	March	2000
41	Mysore	KA01	May	2000
42	Surat	GJ01	April	1500
43	Surat	GJ01	February	2000
44	Surat	GJ01	January	2000
45	Surat	GJ01	June	1500
46	Surat	GJ01	March	2000
47	Surat	GJ01	May	1500
48	Vadodara	GJ02	April	1500

```
49       Vadodara    GJ02   February                  1800
50       Vadodara    GJ02    January                  1800
51       Vadodara    GJ02       June                  1500
52       Vadodara    GJ02      March                  1800
53       Vadodara    GJ02        May                  1500
54  Visakhapatnam    AP01      April                  2000
55  Visakhapatnam    AP01   February                  2500
56  Visakhapatnam    AP01    January                  2500
57  Visakhapatnam    AP01       June                  2000
58  Visakhapatnam    AP01      March                  2500
59  Visakhapatnam    AP01        May                  2000
```

[114]:
```
merged_df5 = pd.merge(df_monthly_target_new_passengers,␣
↪df_fact_passenger_summary, on=["month","city_id"], how ="inner")
```

[265]:
```
pd.set_option("display.max_columns", None)  # Show all columns
pd.set_option("display.width", 1000)
```

[266]:
```
print(merged_df5)
```

```
        month city_id  target_new_passengers  new_passengers  repeat_passengers
total_passengers
0   2024-05-01    GJ01                   1500            1611                1606
3217
1   2024-05-01    GJ02                   1500            1388                 868
2256
2   2024-03-01    GJ01                   2000            1946                1494
3440
3   2024-05-01    UP01                   2000            1825                1662
3487
4   2024-05-01    MP01                   2000            2028                1563
3591
5   2024-01-01    AP01                   2500            2513                 650
3163
6   2024-03-01    MP01                   2700            2742                1091
3833
7   2024-06-01    CH01                   3000            2430                 867
3297
8   2024-04-01    AP01                   2000            1845                 992
2837
9   2024-01-01    KA01                   2000            1957                 172
2129
10  2024-03-01    UP01                   3200            3159                1622
4781
11  2024-02-01    AP01                   2500            2380                 790
3170
12  2024-04-01    GJ02                   1500            1637                 862
2499
```

13	2024-02-01	TN01	1500	1647	346	1993
14	2024-01-01	MP01	2700	2843	1033	3876
15	2024-03-01	TN01	1500	1538	427	1965
16	2024-05-01	KA01	2000	1921	349	2270
17	2024-05-01	TN01	1000	1039	504	1543
18	2024-02-01	GJ01	2000	2254	1313	3567
19	2024-04-01	CH01	3000	2496	789	3285
20	2024-01-01	GJ02	1800	2089	544	2633
21	2024-04-01	MP01	2000	2351	1295	3646
22	2024-06-01	TN01	1000	1226	402	1628
23	2024-04-01	KL01	4000	4939	1576	6515
24	2024-06-01	MP01	2000	2021	1131	3152
25	2024-03-01	CH01	4000	3228	872	4100
26	2024-01-01	RJ01	12000	10423	1422	11845
27	2024-05-01	RJ01	6000	5332	1842	7174
28	2024-02-01	MP01	2700	2878	1103	3981
29	2024-02-01	KA01	2000	2107	183	2290
30	2024-04-01	TN01	1000	1242	480	1722
31	2024-02-01	UP01	3200	3529	1659	5188
32	2024-04-01	KA01	2000	1836	236	2072
33	2024-03-01	AP01	2500	2170	923	3093
34	2024-02-01	GJ02	1800	2146	610	2756
35	2024-06-01	UP01	2000	1971	1727	3698
36	2024-01-01	TN01	1500	1822	392	2214

37	2024-03-01	KA01	2000	1986	208	2194
38	2024-06-01	KL01	4000	3011	1049	4060
39	2024-02-01	RJ01	12000	10789	1661	12450
40	2024-05-01	CH01	3000	2730	969	3699
41	2024-04-01	RJ01	6000	6120	1736	7856
42	2024-01-01	KL01	5000	4865	795	5660
43	2024-04-01	GJ01	1500	1843	1551	3394
44	2024-01-01	GJ01	2000	2432	1184	3616
45	2024-06-01	GJ01	1500	1540	1490	3030
46	2024-03-01	KL01	5000	4865	1348	6213
47	2024-06-01	RJ01	6000	5775	1181	6956
48	2024-04-01	UP01	2000	2311	1496	3807
49	2024-06-01	GJ02	1500	1104	703	1807
50	2024-05-01	KL01	4000	4369	1853	6222
51	2024-02-01	CH01	4000	4104	853	4957
52	2024-01-01	CH01	4000	3920	720	4640
53	2024-03-01	GJ02	1800	1763	759	2522
54	2024-01-01	UP01	3200	3465	1431	4896
55	2024-06-01	KA01	2000	1874	329	2203
56	2024-03-01	RJ01	12000	7417	1840	9257
57	2024-05-01	AP01	2000	1939	951	2890
58	2024-06-01	AP01	2000	1900	802	2702
59	2024-02-01	KL01	5000	4367	1005	5372

[244]:
```
merged_df6 = pd.merge(merged_df5, df_dim_city, on='city_id')
```

[245]:
```
merged_df6.head()
```

[245]:
```
        month city_id  target_new_passengers  new_passengers  repeat_passengers
total_passengers city_name
0  2024-05-01    GJ01                   1500            1611               1606
3217     Surat
1  2024-03-01    GJ01                   2000            1946               1494
3440     Surat
2  2024-02-01    GJ01                   2000            2254               1313
3567     Surat
3  2024-04-01    GJ01                   1500            1843               1551
3394     Surat
4  2024-01-01    GJ01                   2000            2432               1184
3616     Surat
```

[246]:
```
merged_df6.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 60 entries, 0 to 59
Data columns (total 7 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   month                  60 non-null     object
 1   city_id                60 non-null     object
 2   target_new_passengers  60 non-null     int64
 3   new_passengers         60 non-null     int64
 4   repeat_passengers      60 non-null     int64
 5   total_passengers       60 non-null     int64
 6   city_name              60 non-null     object
dtypes: int64(4), object(3)
memory usage: 3.8+ KB
```

[247]:
```
merged_df6['month'] = pd.to_datetime(merged_df6['month'])
```

[248]:
```
merged_df6['month_number'] = merged_df6['month'].dt.month
```

[249]:
```
merged_df6['month_name'] = merged_df6['month'].dt.strftime('%B')
```

[250]:
```
merged_df6.head()
```

[250]:
```
       month city_id  target_new_passengers  new_passengers  repeat_passengers
total_passengers city_name  month_number month_name
0 2024-05-01    GJ01                   1500            1611               1606
3217     Surat             5        May
1 2024-03-01    GJ01                   2000            1946               1494
3440     Surat             3      March
```

```
2 2024-02-01    GJ01                    2000            2254                1313
3567     Surat             2   February
3 2024-04-01    GJ01                    1500            1843                1551
3394     Surat             4      April
4 2024-01-01    GJ01                    2000            2432                1184
3616     Surat             1    January
```

[251]:
```
merged_df6 = merged_df6.sort_values(
    by="month_number", ascending=True
).reset_index(drop=True)
```

[253]:
```
merged_df6.head()
```

[253]:
```
       month city_id  target_new_passengers  new_passengers  repeat_passengers
total_passengers city_name  month_number month_name
0 2024-01-01    MP01                    2700            2843                1033
3876    Indore             1    January
1 2024-01-01    UP01                    3200            3465                1431
4896   Lucknow             1    January
2 2024-01-01    GJ01                    2000            2432                1184
3616     Surat             1    January
3 2024-01-01    KA01                    2000            1957                 172
2129    Mysore             1    January
4 2024-01-01    KL01                    5000            4865                 795
5660     Kochi             1    January
```

[254]:
```
conditions = [
    merged_df6["new_passengers"] > merged_df6["target_new_passengers"],
    merged_df6["new_passengers"] == merged_df6["target_new_passengers"],
    merged_df6["new_passengers"] < merged_df6["target_new_passengers"]
]
```

[255]:
```
choices = ["Exceeded", "Met", "Missed"]
```

[256]:
```
merged_df6["TargetStatus"] = np.select(conditions, choices, default="Unknown")
```

[257]:
```
columns_to_display = [col for col in merged_df6 if col not in ["month",␣
↪"month_number","repeat_passengers","total_passengers","target_achieved"]]
```

[258]:
```
print(merged_df6[columns_to_display])
```

```
   city_id  target_new_passengers  new_passengers      city_name month_name
TargetStatus
0     MP01                   2700            2843         Indore    January
Exceeded
1     UP01                   3200            3465        Lucknow    January
Exceeded
2     GJ01                   2000            2432          Surat    January
```

```
Exceeded
3     KA01                2000            1957         Mysore    January
Missed
4     KL01                5000            4865          Kochi    January
Missed
5     CH01                4000            3920     Chandigarh    January
Missed
6     GJ02                1800            2089       Vadodara    January
Exceeded
7     AP01                2500            2513  Visakhapatnam    January
Exceeded
8     RJ01               12000           10423         Jaipur    January
Missed
9     TN01                1500            1822     Coimbatore    January
Exceeded
10    TN01                1500            1647     Coimbatore   February
Exceeded
11    AP01                2500            2380  Visakhapatnam   February
Missed
12    KL01                5000            4367          Kochi   February
Missed
13    UP01                3200            3529        Lucknow   February
Exceeded
14    MP01                2700            2878         Indore   February
Exceeded
15    GJ02                1800            2146       Vadodara   February
Exceeded
16    RJ01               12000           10789         Jaipur   February
Missed
17    KA01                2000            2107         Mysore   February
Exceeded
18    GJ01                2000            2254          Surat   February
Exceeded
19    CH01                4000            4104     Chandigarh   February
Exceeded
20    AP01                2500            2170  Visakhapatnam      March
Missed
21    KA01                2000            1986         Mysore      March
Missed
22    CH01                4000            3228     Chandigarh      March
Missed
23    RJ01               12000            7417         Jaipur      March
Missed
24    TN01                1500            1538     Coimbatore      March
Exceeded
25    GJ01                2000            1946          Surat      March
Missed
26    MP01                2700            2742         Indore      March
```

Exceeded					
27	UP01	3200	3159	Lucknow	March
Missed					
28	GJ02	1800	1763	Vadodara	March
Missed					
29	KL01	5000	4865	Kochi	March
Missed					
30	KA01	2000	1836	Mysore	April
Missed					
31	GJ01	1500	1843	Surat	April
Exceeded					
32	RJ01	6000	6120	Jaipur	April
Exceeded					
33	GJ02	1500	1637	Vadodara	April
Exceeded					
34	TN01	1000	1242	Coimbatore	April
Exceeded					
35	CH01	3000	2496	Chandigarh	April
Missed					
36	KL01	4000	4939	Kochi	April
Exceeded					
37	UP01	2000	2311	Lucknow	April
Exceeded					
38	AP01	2000	1845	Visakhapatnam	April
Missed					
39	MP01	2000	2351	Indore	April
Exceeded					
40	TN01	1000	1039	Coimbatore	May
Exceeded					
41	KL01	4000	4369	Kochi	May
Exceeded					
42	RJ01	6000	5332	Jaipur	May
Missed					
43	GJ01	1500	1611	Surat	May
Exceeded					
44	KA01	2000	1921	Mysore	May
Missed					
45	CH01	3000	2730	Chandigarh	May
Missed					
46	AP01	2000	1939	Visakhapatnam	May
Missed					
47	MP01	2000	2028	Indore	May
Exceeded					
48	UP01	2000	1825	Lucknow	May
Missed					
49	GJ02	1500	1388	Vadodara	May
Missed					
50	TN01	1000	1226	Coimbatore	June

```
Exceeded
51    CH01                    3000             2430     Chandigarh       June
Missed
52    RJ01                    6000             5775        Jaipur       June
Missed
53    KL01                    4000             3011         Kochi       June
Missed
54    MP01                    2000             2021        Indore       June
Exceeded
55    UP01                    2000             1971       Lucknow       June
Missed
56    GJ02                    1500             1104      Vadodara       June
Missed
57    GJ01                    1500             1540         Surat       June
Exceeded
58    KA01                    2000             1874        Mysore       June
Missed
59    AP01                    2000             1900  Visakhapatnam       June
Missed
```

[259]:
```
# There are mixed consequence of target achievements.
```

[264]:
```
merged_df7 = pd.merge(df_city_target_passenger_rating, df_dim_city,␣
↪on='city_id')
```

[271]:
```
print(merged_df7)
```

```
  city_id  target_avg_passenger_rating      city_name
0    CH01                         8.00     Chandigarh
1    UP01                         7.25        Lucknow
2    AP01                         8.50  Visakhapatnam
3    MP01                         8.00         Indore
4    RJ01                         8.25         Jaipur
5    KA01                         8.50         Mysore
6    GJ01                         7.00          Surat
7    TN01                         8.25     Coimbatore
8    KL01                         8.50          Kochi
9    GJ02                         7.50       Vadodara
```

[273]:
```
merged_df6["percentage_difference"] = (
   ( merged_df6["target_new_passengers"]- merged_df6["new_passengers"])/␣
↪merged_df6["target_new_passengers"] * 100
)
```

[274]:
```
merged_df6.head()
```

[274]:
```
        month city_id  target_new_passengers  new_passengers  repeat_passengers
total_passengers city_name  month_number month_name TargetStatus
```

```
percentage_difference
0 2024-01-01    MP01                    2700            2843                1033
3876    Indore             1    January     Exceeded              -5.296296
1 2024-01-01    UP01                    3200            3465                1431
4896   Lucknow             1    January     Exceeded              -8.281250
2 2024-01-01    GJ01                    2000            2432                1184
3616     Surat             1    January     Exceeded             -21.600000
3 2024-01-01    KA01                    2000            1957                 172
2129    Mysore             1    January       Missed               2.150000
4 2024-01-01    KL01                    5000            4865                 795
5660     Kochi             1    January       Missed               2.700000
```

[279]:
```
merged_df6["percentage_difference"] = merged_df6[
    "percentage_difference"
].apply(lambda x: f"{x:.2f}%")
```

[280]:
```
print(merged_df6.drop(columns=["month",␣
 ↪"month_number","repeat_passengers","total_passengers"]))
```

```
   city_id  target_new_passengers  new_passengers      city_name month_name
TargetStatus percentage_difference
0     MP01                   2700            2843         Indore    January
Exceeded               -5.30%
1     UP01                   3200            3465        Lucknow    January
Exceeded               -8.28%
2     GJ01                   2000            2432          Surat    January
Exceeded              -21.60%
3     KA01                   2000            1957         Mysore    January
Missed                 2.15%
4     KL01                   5000            4865          Kochi    January
Missed                 2.70%
5     CH01                   4000            3920     Chandigarh    January
Missed                 2.00%
6     GJ02                   1800            2089       Vadodara    January
Exceeded              -16.06%
7     AP01                   2500            2513  Visakhapatnam    January
Exceeded               -0.52%
8     RJ01                  12000           10423         Jaipur    January
Missed                13.14%
9     TN01                   1500            1822     Coimbatore    January
Exceeded              -21.47%
10    TN01                   1500            1647     Coimbatore   February
Exceeded               -9.80%
11    AP01                   2500            2380  Visakhapatnam   February
Missed                 4.80%
12    KL01                   5000            4367          Kochi   February
Missed                12.66%
13    UP01                   3200            3529        Lucknow   February
```

Exceeded	-10.28%					
14	MP01	2700	2878	Indore	February	
Exceeded	-6.59%					
15	GJ02	1800	2146	Vadodara	February	
Exceeded	-19.22%					
16	RJ01	12000	10789	Jaipur	February	
Missed	10.09%					
17	KA01	2000	2107	Mysore	February	
Exceeded	-5.35%					
18	GJ01	2000	2254	Surat	February	
Exceeded	-12.70%					
19	CH01	4000	4104	Chandigarh	February	
Exceeded	-2.60%					
20	AP01	2500	2170	Visakhapatnam	March	
Missed	13.20%					
21	KA01	2000	1986	Mysore	March	
Missed	0.70%					
22	CH01	4000	3228	Chandigarh	March	
Missed	19.30%					
23	RJ01	12000	7417	Jaipur	March	
Missed	38.19%					
24	TN01	1500	1538	Coimbatore	March	
Exceeded	-2.53%					
25	GJ01	2000	1946	Surat	March	
Missed	2.70%					
26	MP01	2700	2742	Indore	March	
Exceeded	-1.56%					
27	UP01	3200	3159	Lucknow	March	
Missed	1.28%					
28	GJ02	1800	1763	Vadodara	March	
Missed	2.06%					
29	KL01	5000	4865	Kochi	March	
Missed	2.70%					
30	KA01	2000	1836	Mysore	April	
Missed	8.20%					
31	GJ01	1500	1843	Surat	April	
Exceeded	-22.87%					
32	RJ01	6000	6120	Jaipur	April	
Exceeded	-2.00%					
33	GJ02	1500	1637	Vadodara	April	
Exceeded	-9.13%					
34	TN01	1000	1242	Coimbatore	April	
Exceeded	-24.20%					
35	CH01	3000	2496	Chandigarh	April	
Missed	16.80%					
36	KL01	4000	4939	Kochi	April	
Exceeded	-23.47%					
37	UP01	2000	2311	Lucknow	April	

Exceeded	-15.55%					
38	AP01	2000	1845	Visakhapatnam	April	
Missed	7.75%					
39	MP01	2000	2351	Indore	April	
Exceeded	-17.55%					
40	TN01	1000	1039	Coimbatore	May	
Exceeded	-3.90%					
41	KL01	4000	4369	Kochi	May	
Exceeded	-9.22%					
42	RJ01	6000	5332	Jaipur	May	
Missed	11.13%					
43	GJ01	1500	1611	Surat	May	
Exceeded	-7.40%					
44	KA01	2000	1921	Mysore	May	
Missed	3.95%					
45	CH01	3000	2730	Chandigarh	May	
Missed	9.00%					
46	AP01	2000	1939	Visakhapatnam	May	
Missed	3.05%					
47	MP01	2000	2028	Indore	May	
Exceeded	-1.40%					
48	UP01	2000	1825	Lucknow	May	
Missed	8.75%					
49	GJ02	1500	1388	Vadodara	May	
Missed	7.47%					
50	TN01	1000	1226	Coimbatore	June	
Exceeded	-22.60%					
51	CH01	3000	2430	Chandigarh	June	
Missed	19.00%					
52	RJ01	6000	5775	Jaipur	June	
Missed	3.75%					
53	KL01	4000	3011	Kochi	June	
Missed	24.73%					
54	MP01	2000	2021	Indore	June	
Exceeded	-1.05%					
55	UP01	2000	1971	Lucknow	June	
Missed	1.45%					
56	GJ02	1500	1104	Vadodara	June	
Missed	26.40%					
57	GJ01	1500	1540	Surat	June	
Exceeded	-2.67%					
58	KA01	2000	1874	Mysore	June	
Missed	6.30%					
59	AP01	2000	1900	Visakhapatnam	June	
Missed	5.00%					

8 Highest and lowest repeat passenger rate by city and month

[281]:
```
merged_df["repeat_passenger_rate"] = (
    merged_df["repeat_passenger_count"]/ merged_df["TotalCount"] * 100
)
```

[283]:
```
merged_df.head()
```

[283]:
```
        month city_id  trip_count  repeat_passenger_count  TotalCount
city_name  month_number month_name  repeat_passenger_rate
0 2024-01-01    AP01          10                       7          17
Visakhapatnam             1    January              41.176471
1 2024-01-01    AP01           2                     352         354
Visakhapatnam             1    January              99.435028
2 2024-01-01    AP01           3                     158         161
Visakhapatnam             1    January              98.136646
3 2024-01-01    AP01           4                      53          57
Visakhapatnam             1    January              92.982456
4 2024-01-01    AP01           5                      38          43
Visakhapatnam             1    January              88.372093
```

[291]:
```
merged_df["repeat_passenger_rate"] = merged_df[
    "repeat_passenger_rate"
].apply(lambda x: f"{x:.2f}%")
```

[292]:
```
print(merged_df.drop(columns=["month", "month_number"]))
```

```
    city_id  trip_count  repeat_passenger_count  TotalCount      city_name
month_name repeat_passenger_rate
0      AP01          10                       7          17  Visakhapatnam
January                41.18%
1      AP01           2                     352         354  Visakhapatnam
January                99.44%
2      AP01           3                     158         161  Visakhapatnam
January                98.14%
3      AP01           4                      53          57  Visakhapatnam
January                92.98%
4      AP01           5                      38          43  Visakhapatnam
January                88.37%
..      …         …                       …         …             …
…                    …
535    UP01           5                     272         277        Lucknow
June                98.19%
536    UP01           6                     272         278        Lucknow
June                97.84%
537    UP01           7                     246         253        Lucknow
June                97.23%
538    UP01           8                      83          91        Lucknow
```

```
June                91.21%
539    UP01           9                      19          28        Lucknow
June                67.86%

[540 rows x 7 columns]
```

[287]:
```
merged_df["month_name"].unique()
```

[287]:
```
array(['January', 'February', 'March', 'April', 'May', 'June'],
      dtype=object)
```

[289]:
```
start_date = '2024-01-01'
end_date = '2024-06-30'
filtered_df1 = merged_df[(merged_df['month'] >= start_date) &␣
 ↪(merged_df['month'] <= end_date)]
```

[290]:
```
filtered_df1
```

[290]:
```
          month city_id  trip_count  repeat_passenger_count  TotalCount
city_name  month_number month_name  repeat_passenger_rate
0    2024-01-01    AP01          10                       7          17
Visakhapatnam             1    January              41.176471
1    2024-01-01    AP01           2                     352         354
Visakhapatnam             1    January              99.435028
2    2024-01-01    AP01           3                     158         161
Visakhapatnam             1    January              98.136646
3    2024-01-01    AP01           4                      53          57
Visakhapatnam             1    January              92.982456
4    2024-01-01    AP01           5                      38          43
Visakhapatnam             1    January              88.372093
..          …     …          …                       …          …
…            …        …                      …
535 2024-06-01    UP01           5                     272         277
Lucknow             6       June              98.194946
536 2024-06-01    UP01           6                     272         278
Lucknow             6       June              97.841727
537 2024-06-01    UP01           7                     246         253
Lucknow             6       June              97.233202
538 2024-06-01    UP01           8                      83          91
Lucknow             6       June              91.208791
539 2024-06-01    UP01           9                      19          28
Lucknow             6       June              67.857143

[540 rows x 9 columns]
```

[295]:
```
filtered_df2 = filtered_df1.drop(columns=["month", "month_number"])
```

[304]:
```
filtered_df2 = filtered_df2.sort_values(
    by="repeat_passenger_rate", ascending=False
).reset_index(drop=True)
```

[306]:
```
filtered_df2.head(20)
```

[306]:

	city_id	trip_count	repeat_passenger_count	TotalCount	city_name	month_name	repeat_passenger_rate
0	RJ01	2	999	1001	Jaipur	February	99.800200
1	RJ01	2	991	993	Jaipur	April	99.798590
2	KL01	2	971	973	Kochi	May	99.794450
3	RJ01	2	902	904	Jaipur	May	99.778761
4	KL01	2	846	848	Kochi	April	99.764151
5	RJ01	2	843	845	Jaipur	March	99.763314
6	RJ01	2	705	707	Jaipur	January	99.717115
7	MP01	2	613	615	Indore	May	99.674797
8	AP01	2	505	507	Visakhapatnam	March	99.605523
9	KL01	2	504	506	Kochi	March	99.604743
10	AP01	2	500	502	Visakhapatnam	May	99.601594
11	KL01	2	465	467	Kochi	June	99.571734
12	KL01	2	439	441	Kochi	February	99.546485
13	AP01	2	433	435	Visakhapatnam	February	99.540230
14	AP01	2	427	429	Visakhapatnam	April	99.533800
15	MP01	2	421	423	Indore	April	99.527187
16	RJ01	2	415	417	Jaipur	June	99.520384
17	KL01	2	410	412	Kochi	January	99.514563
18	AP01	2	401	403	Visakhapatnam	June	99.503722
19	MP01	2	399	401	Indore		

```
January               99.501247
```

[296]:
```
highest_repeat_passenger_rate = filtered_df2.nlargest(2,␣
↪'repeat_passenger_rate')
```

[297]:
```
print(highest_repeat_passenger_rate)
```

```
    city_id  trip_count  repeat_passenger_count  TotalCount city_name month_name
repeat_passenger_rate
388    RJ01           2                     999        1001    Jaipur   February
99.80020
406    RJ01           2                     991         993    Jaipur      April
99.79859
```

[298]:
```
# Jaipur has highest repeat passenger rate in the month of February.
```

[299]:
```
lowest_repeat_passenger_rate = filtered_df2.nsmallest(2,␣
↪'repeat_passenger_rate')
```

[300]:
```
print(lowest_repeat_passenger_rate)
```

```
    city_id  trip_count  repeat_passenger_count  TotalCount city_name month_name
repeat_passenger_rate
224    KA01           9                       0           9    Mysore    January
0.0
234    KA01          10                       0          10    Mysore      March
0.0
```

[301]:
```
# Mysore has lowest repeat passenger rate in the month of January.
```

[307]:
```
# City demographics might contribute to higher and lower repeat passenger rate␣
↪in different cities.
```

[308]:
```
merged_df = merged_df.sort_values(
    by="trip_count", ascending=False
).reset_index(drop=True)
```

[311]:
```
merged_df.head(50)
```

[311]:
```
        month city_id  trip_count  repeat_passenger_count  TotalCount
city_name  month_number month_name repeat_passenger_rate
0  2024-01-01    AP01          10                       7          17
Visakhapatnam             1    January                41.18%
1  2024-03-01    GJ02          10                      12          22
Vadodara             3      March               54.55%
2  2024-03-01    KL01          10                      16          26
Kochi              3      March               61.54%
3  2024-02-01    KL01          10                      10          20
Kochi              2   February               50.00%
```

```
4  2024-06-01    KA01          10                    1          11
Mysore             6       June                9.09%
5  2024-05-01    KA01          10                    3          13
Mysore             5        May               23.08%
6  2024-04-01    KA01          10                    1          11
Mysore             4      April                9.09%
7  2024-03-01    KA01          10                    0          10
Mysore             3      March                0.00%
8  2024-02-01    KA01          10                    1          11
Mysore             2   February                9.09%
9  2024-01-01    KA01          10                    1          11
Mysore             1    January                9.09%
10 2024-06-01    GJ02          10                   15          25
Vadodara           6       June               60.00%
11 2024-05-01    GJ02          10                   17          27
Vadodara           5        May               62.96%
12 2024-04-01    GJ02          10                   10          20
Vadodara           4      April               50.00%
13 2024-02-01    GJ02          10                    8          18
Vadodara           2   February               44.44%
14 2024-01-01    CH01          10                   14          24
Chandigarh         1    January               58.33%
15 2024-01-01    GJ02          10                    8          18
Vadodara           1    January               44.44%
16 2024-06-01    GJ01          10                   11          21
Surat              6       June               52.38%
17 2024-05-01    GJ01          10                   19          29
Surat              5        May               65.52%
18 2024-04-01    GJ01          10                   24          34
Surat              4      April               70.59%
19 2024-03-01    GJ01          10                   20          30
Surat              3      March               66.67%
20 2024-02-01    GJ01          10                   16          26
Surat              2   February               61.54%
21 2024-01-01    GJ01          10                   27          37
Surat              1    January               72.97%
22 2024-06-01    CH01          10                   18          28
Chandigarh         6       June               64.29%
23 2024-05-01    CH01          10                   15          25
Chandigarh         5        May               60.00%
24 2024-04-01    CH01          10                   12          22
Chandigarh         4      April               54.55%
25 2024-03-01    CH01          10                   19          29
Chandigarh         3      March               65.52%
26 2024-04-01    KL01          10                   10          20
Kochi              4      April               50.00%
27 2024-05-01    KL01          10                   14          24
```

```
Kochi             5        May               58.33%
28 2024-06-01    KL01          10                   6          16
Kochi             6       June               37.50%
29 2024-01-01    MP01          10                  13          23
Indore            1    January               56.52%
30 2024-06-01    UP01          10                  17          27
Lucknow           6       June               62.96%
31 2024-05-01    UP01          10                  18          28
Lucknow           5        May               64.29%
32 2024-04-01    UP01          10                  20          30
Lucknow           4      April               66.67%
33 2024-03-01    UP01          10                  17          27
Lucknow           3      March               62.96%
34 2024-02-01    UP01          10                  20          30
Lucknow           2   February               66.67%
35 2024-01-01    UP01          10                  14          24
Lucknow           1    January               58.33%
36 2024-06-01    TN01          10                   5          15
Coimbatore        6       June                33.33%
37 2024-05-01    TN01          10                   3          13
Coimbatore        5        May                23.08%
38 2024-04-01    TN01          10                  12          22
Coimbatore        4      April                54.55%
39 2024-03-01    TN01          10                   7          17
Coimbatore        3      March                41.18%
40 2024-02-01    TN01          10                   3          13
Coimbatore        2   February                23.08%
41 2024-01-01    TN01          10                   1          11
Coimbatore        1    January                 9.09%
42 2024-06-01    RJ01          10                  10          20
Jaipur            6       June               50.00%
43 2024-05-01    RJ01          10                  24          34
Jaipur            5        May               70.59%
44 2024-04-01    RJ01          10                  16          26
Jaipur            4      April               61.54%
45 2024-03-01    RJ01          10                  17          27
Jaipur            3      March               62.96%
46 2024-02-01    RJ01          10                  14          24
Jaipur            2   February               58.33%
47 2024-01-01    RJ01          10                  13          23
Jaipur            1    January               56.52%
48 2024-06-01    MP01          10                  14          24
Indore            6       June               58.33%
49 2024-05-01    MP01          10                  22          32
Indore            5        May               68.75%
```

[312]:
```
# Higher trip count is observed in all over the different cities for␣
 ↪differentmonths,
#but repeat passenger rate is not showing much higher values.
```

[313]:
```
merged_df7 = pd.merge(df_fact_passenger_summary, df_dim_city, on='city_id')
```

[314]:
```
merged_df7['month'] = pd.to_datetime(merged_df7['month'])
```

[315]:
```
merged_df7['month_number'] = merged_df7['month'].dt.month
```

[316]:
```
merged_df7['month_name'] = merged_df7['month'].dt.strftime('%B')
```

[324]:
```
merged_df7 = merged_df7.sort_values(
    by="month_number", ascending=True
).reset_index(drop=True)
```

[325]:
```
merged_df7
```

[325]:

	month	city_id	new_passengers	repeat_passengers	total_passengers	city_name	month_number	month_name
0	2024-01-01	UP01	3465	1431	4896	Lucknow	1	January
1	2024-01-01	MP01	2843	1033	3876	Indore	1	January
2	2024-01-01	GJ01	2432	1184	3616	Surat	1	January
3	2024-01-01	TN01	1822	392	2214	Coimbatore	1	January
4	2024-01-01	GJ02	2089	544	2633	Vadodara	1	January
5	2024-01-01	RJ01	10423	1422	11845	Jaipur	1	January
6	2024-01-01	AP01	2513	650	3163	Visakhapatnam	1	January
7	2024-01-01	KA01	1957	172	2129	Mysore	1	January
8	2024-01-01	KL01	4865	795	5660	Kochi	1	January
9	2024-01-01	CH01	3920	720	4640	Chandigarh	1	January
10	2024-02-01	KA01	2107	183	2290	Mysore	2	February
11	2024-02-01	UP01	3529	1659	5188	Lucknow	2	February
12	2024-02-01	CH01	4104	853	4957	Chandigarh	2	February
13	2024-02-01	GJ02	2146	610	2756			

```
Vadodara             2   February
14 2024-02-01    KL01            4367              1005              5372
Kochi             2   February
15 2024-02-01    TN01            1647               346              1993
Coimbatore             2   February
16 2024-02-01    AP01            2380               790              3170
Visakhapatnam             2   February
17 2024-02-01    GJ01            2254              1313              3567
Surat             2   February
18 2024-02-01    RJ01           10789              1661             12450
Jaipur            2   February
19 2024-02-01    MP01            2878              1103              3981
Indore            2   February
20 2024-03-01    GJ01            1946              1494              3440
Surat             3      March
21 2024-03-01    MP01            2742              1091              3833
Indore            3      March
22 2024-03-01    RJ01            7417              1840              9257
Jaipur            3      March
23 2024-03-01    KA01            1986               208              2194
Mysore            3      March
24 2024-03-01    KL01            4865              1348              6213
Kochi             3      March
25 2024-03-01    GJ02            1763               759              2522
Vadodara             3      March
26 2024-03-01    TN01            1538               427              1965
Coimbatore             3      March
27 2024-03-01    CH01            3228               872              4100
Chandigarh             3      March
28 2024-03-01    UP01            3159              1622              4781
Lucknow           3      March
29 2024-03-01    AP01            2170               923              3093
Visakhapatnam             3      March
30 2024-04-01    GJ02            1637               862              2499
Vadodara             4      April
31 2024-04-01    UP01            2311              1496              3807
Lucknow           4      April
32 2024-04-01    GJ01            1843              1551              3394
Surat             4      April
33 2024-04-01    RJ01            6120              1736              7856
Jaipur            4      April
34 2024-04-01    AP01            1845               992              2837
Visakhapatnam             4      April
35 2024-04-01    KA01            1836               236              2072
Mysore            4      April
36 2024-04-01    MP01            2351              1295              3646
Indore            4      April
```

```
37 2024-04-01    KL01            4939             1576             6515
Kochi             4      April
38 2024-04-01    TN01            1242              480             1722
Coimbatore             4      April
39 2024-04-01    CH01            2496              789             3285
Chandigarh             4      April
40 2024-05-01    AP01            1939              951             2890
Visakhapatnam            5        May
41 2024-05-01    GJ01            1611             1606             3217
Surat             5        May
42 2024-05-01    KL01            4369             1853             6222
Kochi             5        May
43 2024-05-01    KA01            1921              349             2270
Mysore            5        May
44 2024-05-01    UP01            1825             1662             3487
Lucknow           5        May
45 2024-05-01    TN01            1039              504             1543
Coimbatore             5        May
46 2024-05-01    MP01            2028             1563             3591
Indore            5        May
47 2024-05-01    GJ02            1388              868             2256
Vadodara           5        May
48 2024-05-01    RJ01            5332             1842             7174
Jaipur            5        May
49 2024-05-01    CH01            2730              969             3699
Chandigarh             5        May
50 2024-06-01    RJ01            5775             1181             6956
Jaipur            6       June
51 2024-06-01    TN01            1226              402             1628
Coimbatore             6       June
52 2024-06-01    GJ01            1540             1490             3030
Surat             6       June
53 2024-06-01    CH01            2430              867             3297
Chandigarh             6       June
54 2024-06-01    GJ02            1104              703             1807
Vadodara           6       June
55 2024-06-01    AP01            1900              802             2702
Visakhapatnam            6       June
56 2024-06-01    UP01            1971             1727             3698
Lucknow           6       June
57 2024-06-01    KA01            1874              329             2203
Mysore            6       June
58 2024-06-01    MP01            2021             1131             3152
Indore            6       June
59 2024-06-01    KL01            3011             1049             4060
Kochi             6       June
```

[326]:
```
# In the tourism season January and February Jaipur has 10420 and 10789 new␣
 ↪passengers,
# which shows that demand of Goodcabs increases seasonally.
# Obviously tailoring marketing efforts to these events increase trip volume in␣
 ↪tourism oriented cities.
```

[330]:
```
merged_df7 = merged_df7.sort_values(
    by="new_passengers", ascending=False
).reset_index(drop=True)
```

[331]:
```
merged_df7
```

[331]:

	month	city_id	new_passengers	repeat_passengers	total_passengers	city_name	month_number	month_name
0	2024-02-01	RJ01	10789	1661	12450	Jaipur	2	February
1	2024-01-01	RJ01	10423	1422	11845	Jaipur	1	January
2	2024-03-01	RJ01	7417	1840	9257	Jaipur	3	March
3	2024-04-01	RJ01	6120	1736	7856	Jaipur	4	April
4	2024-06-01	RJ01	5775	1181	6956	Jaipur	6	June
5	2024-05-01	RJ01	5332	1842	7174	Jaipur	5	May
6	2024-04-01	KL01	4939	1576	6515	Kochi	4	April
7	2024-03-01	KL01	4865	1348	6213	Kochi	3	March
8	2024-01-01	KL01	4865	795	5660	Kochi	1	January
9	2024-05-01	KL01	4369	1853	6222	Kochi	5	May
10	2024-02-01	KL01	4367	1005	5372	Kochi	2	February
11	2024-02-01	CH01	4104	853	4957	Chandigarh	2	February
12	2024-01-01	CH01	3920	720	4640	Chandigarh	1	January
13	2024-02-01	UP01	3529	1659	5188	Lucknow	2	February
14	2024-01-01	UP01	3465	1431	4896	Lucknow	1	January
15	2024-03-01	CH01	3228	872	4100	Chandigarh	3	March
16	2024-03-01	UP01	3159	1622	4781			

```
Lucknow             3      March
17 2024-06-01    KL01            3011              1049              4060
Kochi               6       June
18 2024-02-01    MP01            2878              1103              3981
Indore              2   February
19 2024-01-01    MP01            2843              1033              3876
Indore              1    January
20 2024-03-01    MP01            2742              1091              3833
Indore              3      March
21 2024-05-01    CH01            2730               969              3699
Chandigarh          5        May
22 2024-01-01    AP01            2513               650              3163
Visakhapatnam           1    January
23 2024-04-01    CH01            2496               789              3285
Chandigarh          4      April
24 2024-01-01    GJ01            2432              1184              3616
Surat               1    January
25 2024-06-01    CH01            2430               867              3297
Chandigarh          6       June
26 2024-02-01    AP01            2380               790              3170
Visakhapatnam           2   February
27 2024-04-01    MP01            2351              1295              3646
Indore              4      April
28 2024-04-01    UP01            2311              1496              3807
Lucknow             4      April
29 2024-02-01    GJ01            2254              1313              3567
Surat               2   February
30 2024-03-01    AP01            2170               923              3093
Visakhapatnam           3      March
31 2024-02-01    GJ02            2146               610              2756
Vadodara            2   February
32 2024-02-01    KA01            2107               183              2290
Mysore              2   February
33 2024-01-01    GJ02            2089               544              2633
Vadodara            1    January
34 2024-05-01    MP01            2028              1563              3591
Indore              5        May
35 2024-06-01    MP01            2021              1131              3152
Indore              6       June
36 2024-03-01    KA01            1986               208              2194
Mysore              3      March
37 2024-06-01    UP01            1971              1727              3698
Lucknow             6       June
38 2024-01-01    KA01            1957               172              2129
Mysore              1    January
39 2024-03-01    GJ01            1946              1494              3440
Surat               3      March
```

```
40 2024-05-01    AP01            1939              951              2890
Visakhapatnam           5        May
41 2024-05-01    KA01            1921              349              2270
Mysore            5        May
42 2024-06-01    AP01            1900              802              2702
Visakhapatnam           6       June
43 2024-06-01    KA01            1874              329              2203
Mysore            6       June
44 2024-04-01    AP01            1845              992              2837
Visakhapatnam           4      April
45 2024-04-01    GJ01            1843             1551              3394
Surat             4      April
46 2024-04-01    KA01            1836              236              2072
Mysore            4      April
47 2024-05-01    UP01            1825             1662              3487
Lucknow           5        May
48 2024-01-01    TN01            1822              392              2214
Coimbatore            1    January
49 2024-03-01    GJ02            1763              759              2522
Vadodara              3      March
50 2024-02-01    TN01            1647              346              1993
Coimbatore            2   February
51 2024-04-01    GJ02            1637              862              2499
Vadodara              4      April
52 2024-05-01    GJ01            1611             1606              3217
Surat             5        May
53 2024-06-01    GJ01            1540             1490              3030
Surat             6       June
54 2024-03-01    TN01            1538              427              1965
Coimbatore            3      March
55 2024-05-01    GJ02            1388              868              2256
Vadodara              5        May
56 2024-04-01    TN01            1242              480              1722
Coimbatore            4      April
57 2024-06-01    TN01            1226              402              1628
Coimbatore            6       June
58 2024-06-01    GJ02            1104              703              1807
Vadodara              6       June
59 2024-05-01    TN01            1039              504              1543
Coimbatore            5        May
```

[332]:
```
# Jaipur is getting higher new passengers all over the month from January to␣
 ↪June.
# Goodcabs can consider integrating electric vehicles or eco friendly␣
 ↪initiatives to stay competitive.
```

[ ]:
```
# As in tourism season deamnd increases there are oppurtunities for Goodcabs to␣
↪partener with local business.
# To make Goodcabs more Data driven Wait time of passenger can be collected to␣
↪give more insights.
```