Answer the questions

1. What is the total number of trips?

Answer: total number of trips : 10906

In [19]:
```
####Qustion 1
import pandas as pd
summary_transaction = pd.read_csv('summary_transaction.csv')
print ('total number of trips is :',summary_transaction[summary_transaction.columns[0]].count())
```

```
total number of trips is : 10906
```

2. How many provinces are there in the trip with the most number of provinces?

Answer: the most number of provinces: 12

In [43]:
```
####Qustion 2
import pandas as pd
summary_transaction = pd.read_csv('summary_transaction.csv')
summary_transaction['count_province'] = summary_transaction['province_list'].str.count(',') + 1
#count_province['count'].max()
print("the most number of provinces :", summary_transaction['count_province'].max())
```

```
the most number of provinces : 12
```

3. What are the most common province pairs that people travel to in the same trip?

Answer: the most common province : Pathum Thani-Bangkok

In [89]:
```
####Qustion 3
import pandas as pd
summary_transaction = pd.read_csv('summary_transaction.csv')
#summary_transaction = summary_transaction[summary_transaction['user_id'].isin([1, 5])]
summary_transaction_2 = summary_transaction[['Trip_id','province_list']]
summary_transaction_2['province'] = summary_transaction_2['province_list'].str.split(',')
summary_transaction_3 = summary_transaction_2.explode('province')s
summary_transaction_3.drop(columns=['province_list'], inplace=True)
summary_transaction_4 = pd.merge(summary_transaction_3, summary_transaction_3, on='Trip_id', how='left')
summary_transaction_5 = summary_transaction_4.loc[summary_transaction_4['province_x'] != summary_transaction_4['province_y']]
summary_transaction_6 = summary_transaction_5.groupby(['province_x', 'province_y']).size().reset_index(name='Count')
max_count = summary_transaction_6['Count'].max()
summary_transaction_7 = summary_transaction_6[summary_transaction_6['Count'] == max_count]
summary_transaction_7['most_popular'] = summary_transaction_7['province_x'] + '-' + summary_transaction_7['province_y'].astype(st
summary_transaction_7[['most_popular','Count']].reset_index(drop=True)
```

```
C:\Users\var_m\AppData\Local\Temp\ipykernel_14184\30978007.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy
  summary_transaction_2['province'] = summary_transaction_2['province_list'].str.split(',')
C:\Users\var_m\AppData\Local\Temp\ipykernel_14184\30978007.py:13: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-ve
rsus-a-copy
  summary_transaction_7['most_popular'] = summary_transaction_7['province_x'] + '-' + summary_transaction_7['province_y'].astyp
e(str)
```

Out[89]:

	most_popular	Count
0	Pathum Thani-Bangkok	382
1	Bangkok- Pathum Thani	382

4. How can we use the data to support our client's business?

Answer: We can use the output of the data to plan the best route or make decisions on promotions by analyzing the periods and locations where people tend to travel the most