

Coursera Capstone Project -

The Battle of Neighborhoods

Mejdi Souki

02 Nov 2020

1. Introduction

The combination of breathtaking natural beauty, historic and cultural sites, along with seaside resorts along its Aegean and the Mediterranean Sea coasts makes Turkey one of the top tourist destinations in the world. As in previous years, the number of incoming tourists to Turkey increased by 13.7 percent in 2019, reaching 51.8 million people, setting an all-time record.

In fact, Turkey becomes the sixth most visited destination by international tourist arrivals in 2020.

Otherwise, a person wanting to visit Turkey will have some confusion to choose the city to visit due to the variety and the richness of venues in different cities of Turkey.

That is why we have to help anyone who wants to visit Turkey having some confusion about what cities he must go to be totally satisfied.

This solution will help Travel agencies in other countries because they can help them make their choices easily when they ask for help based on the type of tourism (Natural, Historic, seaside tourism) they tend to and provide the most trending venues they should go to for each city chosen.

2. Data collection and preparation

- i. At first, we extract the list of the largest cities of Turkey with their location (latitude, longitude) taking from a site using pandas library (by using read_html method to a html component).

```
Cities.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   City        36 non-null    object
1   Latitude    36 non-null    object
2   Longitude   36 non-null    object
dtypes: object(3)
memory usage: 992.0+ bytes
```

- ii. At second, we extract venues for each city by using Foursquare API with the coordinates taken from our first dataframe and given the Id

of categories related to the three important types of tourism in Turkey as mentioned in the introduction.

[12]:

	City Name	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Manisa	38.630554	27.422222	Şehit Ömer Halisdemir Parkı	38.625438	27.429206	Park
1	Manisa	38.630554	27.422222	Tarihi Bedesten Nikah Salonu	38.611251	27.432323	Historic Site
2	Manisa	38.630554	27.422222	Spil Dağı	38.556317	27.444257	Mountain
3	Manisa	38.630554	27.422222	Mesir Tabiat Parkı	38.603193	27.385859	Park
4	Manisa	38.630554	27.422222	Akpınar piknik alanı	38.599578	27.500354	Park
...
1793	Afyonkarahisar	38.756886	30.538704	Erkmen Şelale	38.756282	30.483831	Waterfall
1794	Afyonkarahisar	38.756886	30.538704	Sarı köyü	38.669280	30.556475	Field
1795	Afyonkarahisar	38.756886	30.538704	Otogar Camii	38.790348	30.510718	Mosque
1796	Afyonkarahisar	38.756886	30.538704	Harbiş	38.735048	30.568914	Field
1797	Afyonkarahisar	38.756886	30.538704	Ataköy Üçgen Park	38.730381	30.552597	Park

3. Methodology

- i. Next, we group the venues by city and apply the sum operation to the venues related to categories corresponding to a type of tourism.

[13]:

	Venue Category_Art Museum	Venue Category_Beach	Venue Category_Botanical Garden	Venue Category_Campground	Venue Category_Castle	Venue Category_Cave	Venue Category_Church	Venue Category_Coliseum
City Name								
Adana	6	1	1	0	0	0	0	0
Afyonkarahisar	7	0	1	2	1	0	0	0
Amasya	19	0	2	2	1	0	0	0
Ankara	9	0	0	0	1	0	0	0
Antalya	3	27	0	0	0	0	0	0
Artvin	4	1	2	6	0	0	0	1
Bolu	11	0	1	3	0	0	0	0
Bursa	12	0	2	0	1	0	0	0
Denizli	5	0	0	2	0	0	0	0
Diyarbakir	16	0	2	1	2	0	0	0
Edirne	18	0	3	1	0	0	0	1
Elazığ	13	1	2	3	0	1	0	0
Fakicehir	12	0	1	1	0	0	0	0

	City Name	Historic_venues	Seaside_venues	Natural_venues
0	Adana	6	1	34
1	Afyonkarahisar	7	0	39
2	Amasya	19	0	30
3	Ankara	9	0	35
4	Antalya	3	27	13
5	Artvin	4	1	42
6	Bolu	11	0	33
7	Bursa	12	0	32
8	Denizli	5	0	41
9	Diyarbakir	16	0	27
10	Edirne	18	0	25
11	Elazığ	13	1	31
12	Eskişehir	13	0	35
13	Gaziantep	10	0	35

- ii. Finally, for each city we extract the most trending venues tourists should visit using Foursquare API for trending venues.

[95]:

	City	Latitude	Longitude	1st Most Trending Venue	2nd Most Trending Venue	3rd Most Trending Venue	4th Most Trending Venue	5th Most Trending Venue	6th Most Trending Venue	7th Most Trending Venue	8th Most Trending Venue	9th Most Trending Venue
0	Manisa	38.6306	27.4222	Manisa Çarşı	Cafe Sera	nan	nan	nan	nan	nan	nan	nan
1	Diyarbakir	37.91	40.24	Sülüklü Han	Firuze Kahve Lab	Diyarbakir	Dağkapı Meydanı	Diyarbakir Ceylan Avm	nan	nan	nan	nan
2	Izmit, Kocaeli	40.7667	29.9167	Izmit Çarşı	Yürüyüş Yolu	Seka Park	Marina Cafe	Fethiye Caddesi	nan	nan	nan	nan
3	Gaziantep	37.0667	37.3833	Sanko Park	Gaziantep Çarşı	Forum Gaziantep	Gaziantep	Tahmis Kahvesi	Starbucks	nan	nan	nan
4	Sanliurfa	37.1583	38.7917	Gönül Kahvesi	Bahçelievler	Piazza	Karaköprü	Balıkli Göl	Cafe Nobel	Şanlıurfa	Starbucks	Gümüş H
5	Tekirdag	40.9778	27.5153	Tekirdağ Sahil	Tekirdağ Çarşı	Tekira	nan	nan	nan	nan	nan	nan
6	Adana	37	35.3213	Adana Çarşı	Adana Optimum	Adana	Ziyapaşa Bulvarı	Kazım Büfe	Tucco Gastropub	01 Burda	Storie Store	nan
					Luna		Isiklan	Adana AVM		Yat	Denize	Gül

As unsupervised machine learning algorithm, we used the K-Means algorithm as a clustering algorithm because in our case we need an unsupervised method.

4. Results

After applying K-Means algorithm with number of clusters, we can add the number of cluster to each city.

[27]:

	Cluster Labels	City Name	Historic_venues	Seaside_venues	Natural_venues
0	0	Adana	6	1	34
1	0	Afyonkarahisar	7	0	39
2	1	Amasya	19	0	30
3	0	Ankara	9	0	35
4	2	Antalya	3	27	13
5	0	Artvin	4	1	42
6	1	Bolu	11	0	33
7	1	Bursa	12	0	32
8	0	Denizli	5	0	41
9	1	Diyarbakir	16	0	27
10	1	Edirne	18	0	25
11	1	Elazığ	13	1	31
12	1	Eskişehir	13	0	35
13	0	Gaziantep	10	0	35

After clustering, we can print each cluster with the most trending venues for each city in the cluster.

Cluster 0:

```
[95]: print ('cluster 0')
      CT.loc[CT['Cluster Labels']==0]
```

cluster 0

[95]:	City_x	1st Most Trending Venue	2nd Most Trending Venue	3rd Most Trending Venue	4th Most Trending Venue	5th Most Trending Venue	6th Most Trending Venue	7th Most Trending Venue	8th Most Trending Venue	9th Most Trending Venue	10th Most Trending Venue	Cluster Labels
0	Manisa	Manisa Çarşısı	nan	nan	nan	nan	nan	nan	nan	nan	nan	0
2	Izmit, Kocaeli	İzmit Çarşısı	Yürüyüş Yolu	Fethiye Caddesi	Seka Park	nan	nan	nan	nan	nan	nan	0
3	Gaziantep	Gaziantep Çarşısı	Sanko Park	Forum Gaziantep	Gazi Muhtar Paşa Bulvarı	Gaziantep	nan	nan	nan	nan	nan	0
6	Adana	Adana Çarşısı	Adana Optimum	Adana	Ziypaşa Bulvarı	01 Burda	nan	nan	nan	nan	nan	0
9	Yozgat	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	0
10	Bolu	Bolu Çarşısı	İzzet Baysal Caddesi	nan	nan	nan	nan	nan	nan	nan	nan	0
12	Van	Van AVM	Albino Cafe	Maraş Caddesi	Kahve Deryası	Star 2000 Cafe	Van	Balkon Etiler	nan	nan	nan	0
13	Denizli	Çınar Meydanı	Denizli	Denizli Şehirler Arası Otobüs Terminali	nan	nan	nan	nan	nan	nan	nan	0

Cluster 1:

```
: print ('cluster 1')
CT.loc[CT['Cluster Labels']==1]
```

cluster 1

[illegible]

Cluster 2:

```
print ('cluster 2')
CT.loc[CT['Cluster Labels']==2]
```

cluster 2

[illegible]

5. Discussion

We can observe that:

- Cluster 0 corresponds to natural destinations due to the huge number of natural venues in this cluster.
- Cluster 1 corresponds to seaside destinations due to the important number of seaside venues in this cluster.
- Cluster 2 corresponds to historic destinations due to the important number of historic venues in this cluster.

The most of largest cities of Turkey are historic and natural destinations.

This is related basically to the legacy of ottoman civilization and the variety of natural places in Turkey.

6. Conclusion

In this study, I used Foursquare API and some of python libraries (pandas, scikit-learn) to build my solution through many steps. I built clustering model to predict to segment the most common destinations of Turkey into three categories. This model can be very useful in helping travel agencies and tourists having ambition to visit Turkey. For example, it could help them choose the right destination according to their propensities and have an idea about the most trending venues they can visit there.