

# Cappadocia Venue Recommendation

Gulsum Elcin Sipahi

June 16, 2019



## 1. Introduction

*Cappadocia* is one of the most visited touristic attraction points in Turkey which is known for its fairy chimneys, valleys covered with natural rock formations, underground cities, historical monasteries, boutique hotels and hot air balloon tours.

Cappadocia region was formed 60 million years ago by erosion of soft layers of lava and ash from Mount Erciyes (Argeus), Mount Hasan and Mount Güllü composing with wind and rain over millions of years. It is a place where nature and history integrates. While geographical events created fairy chimneys – the people carved houses and churches inside these chimneys. Later on decorated them with frescoes in this historical process, bringing the traces of civilizations for thousands of years to present.



Due to its unique environment, the number of tourists visiting Cappadocia is increasing year by year. Cappadocia hosted over 2.2 million domestic and foreign tourists in 2017 and with a 33% increase rate it hosted over 2.9 million tourists in 2018.

The intention of this study is to recommend venues like hotels, museums, restaurants to visitors who are planning to visit the Cappadocia region according to FourSquare API venue recommendations with map visualizations. According to geographical coordinates of the venues, the most suitable cluster will be recommended for accommodation.

This kind of data analysis will be useful not only for tourists visiting a city for the first time but also for tourism agencies to inform & recommend their customers about the city and also for investors to invest in the city for new opportunities.



## 2. Data Section

The data used in study is provided by Foursquare location data. The data are grouped by landscape area. Data consists information about hotels, museums and restaurants which are visualized by maps.

The following information are retrieved:

Venue ID Venue Name Coordinates : Latitude and Longitude Category Name

Cappadocia Göreme Top Venue Recommendations from FourSquare API (FourSquare website: [www.foursquare.com](http://www.foursquare.com))

### Clean Data

I cleaned the data by using drop function to delete the multiple entries and unnecessary columns & rows.

To create a single table consisting all venue names I renamed the columns of the tables of hotels, museums and restaurants. I merged the new tables them and run K-means clustering algorithm in order to cluster the data to recommend the best place to accomodate.

## 3. Methodology

First of all I imported libraries which are essential for data analysis, handle requests and then I defined FourSquare Credentials. Since Göreme Region is one of the most visited regions of Cappadocia , I defined Göreme's latitude and longitude values .

Göreme is one of the most visited regions of Cappadocia so Göreme's latitude and longitude values are defined.

```
# define the city and get its latitude & longitude
city = 'Goreme'
geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(city)
latitude = location.latitude
longitude = location.longitude
print(latitude, longitude)
```

38.642089 34.8296234

### Search for Hotels

I searched the hotels within a range of 5 kms.

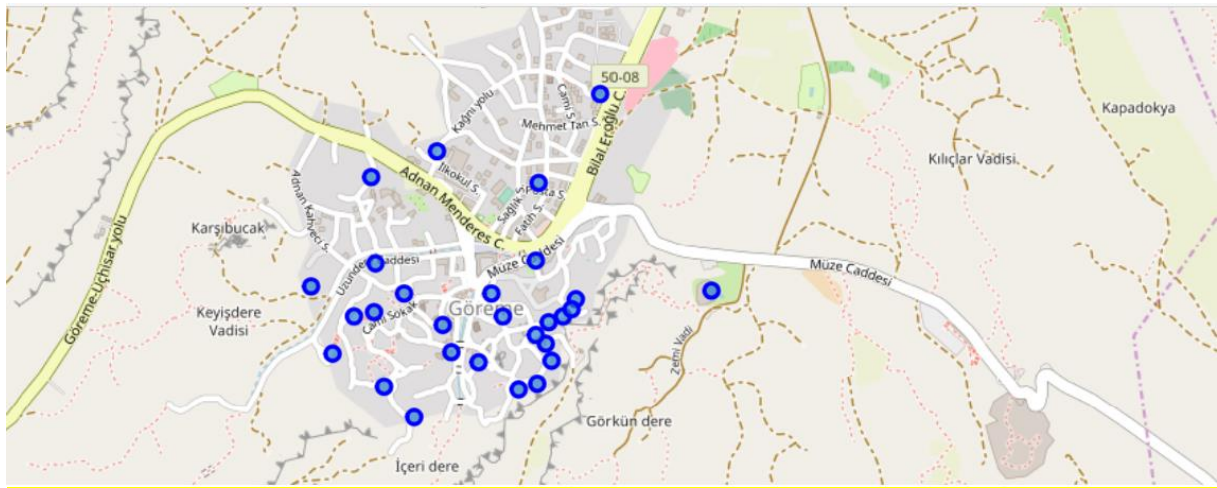
	categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	location.distance	location.formattedAddress	location.labeledLa
0	[[{"shortName": "Hotel", "icon": {"prefix": "htl..."}, {"label": "display 34.83234538336	False	4bd6491e75c0f47aa0ecb03	Gaferli Mah, Goreme	TR	Neveşehir	Türkiye	NaN	237	[Gaferli Mah, Göreme, 50180 Neveşehir, Türkiye]	[[{"label": "display 34.83234538336
1	[[{"shortName": "Hotel", "icon": {"prefix": "htl..."}, {"label": "display 34.82500880446	False	4bd845536a51ce728cbbde3f	Uzundere Cad No. 43	TR	Neveşehir	Türkiye	Göreme	419	[Uzundere Cad No. 43 (Göreme), Neveşehir, Türkiye]	[[{"label": "display 34.82500880446
2	[[{"shortName": "Hotel", "icon": {"prefix": "htl..."}, {"label": "display 34.82566624634	False	4cac1a4a36fa6dcbb77dfdb78	Goreme	TR	Neveşehir	Türkiye	Masat Sokak	459	[Göreme (Masat Sokak), 50180 Neveşehir, Türkiye]	[[{"label": "display 34.82437767703
3	[[{"shortName": "Hotel", "icon": {"prefix": "htl..."}, {"label": "display 34.82666624634	False	4c7813e693ef236a8107ab0f	Aydınlı Mah. Yavuz Sok No:1 Göreme Merkez	TR	Neveşehir	Türkiye	NaN	344	[Aydınlı Mah. Yavuz Sok. No:1 Göreme Merkez, 50180 Neveşehir, Türkiye]	[[{"label": "display 34.82666624634
4	[[{"shortName": "Hotel", "icon": {"prefix": "htl..."}, {"label": "display 34.82620066036	False	53fb94a3498ed0f02cd01a8	Orta Mahalle	TR	Göreme Tarihi Milli Parkı	Türkiye	Ayvaz Efendi Sokak	452	[Orta Mahalle (Ayvaz Efendi Sokak), 50180 Göreme, Neveşehir, Türkiye]	[[{"label": "display 34.82620066036

I cleaned the data by using drop function to keep only the columns including name, address and geographical coordinates.



	name	formattedAddress	lat	lng
0	Anatolian Houses Hotel	[Gaferli Mah, Göreme, 50180 Nevşehir, Türkiye]	38.642280	34.832345
1	Göreme Kaya Hotel	[Uzundere Cad No: 43 (Göreme), Nevşehir, Türkiye]	38.640995	34.825009
2	Flintstones Cave Hotel	[Göreme (Masat Sokak), 50180 Nevşehir, Türkiye]	38.642591	34.824378
3	Kelebek Special Cave Hotel	[Aydınlı Mah. Yavuz Sok. No:1 Göreme Merkez, 50180 Nevşehir, Türkiye]	38.641880	34.825666
4	Cappadocia Inn Hotel	[Orta Mahalle (Ayvaz Efendi Sokak), 50180 Göreme, Nevşehir, Türkiye]	38.645155	34.826201
5	Village Cave House Hotel	[Ünlü Sokak 18, Göreme Tarihi Milli Parkı, Türkiye]	38.641757	34.831525
6	Sarihan Cave Hotel	[Gaferli Mah. Cevizler Sok. No:20, Göreme-Cappadocia, Nevşehir, Türkiye]	38.641880	34.831929
7	Traveller's Cave Hotel	[Aydinli mahallesi gorceli sokak no:7, 50000 Nevşehir, Türkiye]	38.639513	34.827461

I used Python folium library to visualize the hotels in the Göreme Region. I used latitude and longitude values to get the visual as below:



## Search for Museums

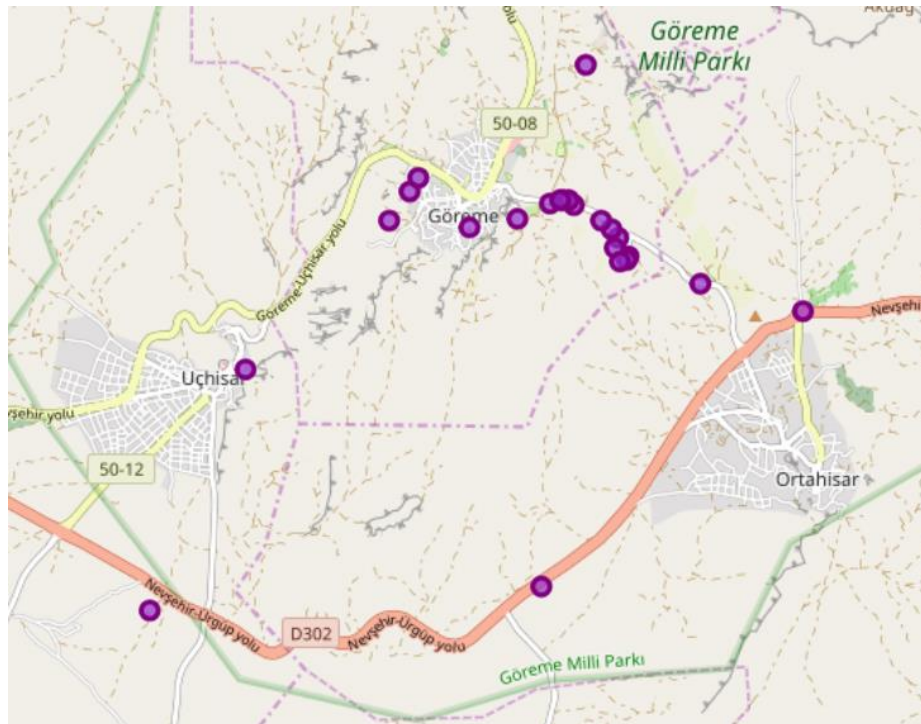
I searched the attraction points like museums and churches within a range of 5 kms and labeled the data as museums.

location.city	location.country	location.crossStreet	location.distance	location.formattedAddress	location.labeledLatLngs	location.lat	location.lng	location.postalCode	location.state	name	referrallid
Nevşehir	Türkiye	NaN	1361	[Göreme Milli Parkı, 50180 Nevşehir, Türkiye]	NaN	38.640302	34.845119	50180	Nevşehir	Göreme Açık Hava Müzesi	1560677039
Göreme Tarihi Milli Parkı	Türkiye	NaN	1303	[Göreme Tarihi Milli Parkı, Türkiye]	[[{"lat": 38.641041259414024, "label": "display..."}]]	38.641041	34.844558	NaN	Nevşehir	Tokalı Kilise	1560677039
NaN	Türkiye	NaN	971	[Türkiye]	[[{"lat": 38.64289261837788, "label": "display..."}]]	38.642893	34.840748	NaN	NaN	Saklı Kilise	1560677039
Göreme Tarihi Milli Parkı	Türkiye	NaN	1488	[Göreme Tarihi Milli Parkı, Türkiye]	[[{"lat": 38.638911628677945, "label": "display..."}]]	38.638912	34.846247	NaN	Nevşehir	Çankılı Kilise	1560677039
Aktepe	Türkiye	NaN	4234	[Zelve Yolu, Aktepe, Türkiye]	[[{"lat": 38.66939601696848, "label": "display..."}]]	38.669396	34.863531	NaN	Nevşehir	Zelve Açık Hava Müzesi	1560677039

I cleaned the data by using drop function to keep only the columns including name, address and geographical coordinates. There were multiple entries for the same churches so I used the drop function for rows to clean the data. Then I put the "name" column to the first column.

	name	formattedAddress	lat	lng
0	Göreme Açık Hava Müzesi	[Göreme Milli Parkı, 50180 Nevşehir, Türkiye]	38.640302	34.845119
1	Tokalı Kilise	[Göreme Tarihi Milli Parkı, Türkiye]	38.641041	34.844558
2	Saklı Kilise	[Türkiye]	38.642893	34.840748
3	Çarıklı Kilise	[Göreme Tarihi Milli Parkı, Türkiye]	38.638912	34.846247
4	Zelve Açık Hava Müzesi	[Zelve Yolu, Aktepe, Türkiye]	38.669396	34.863531
5	Elmalı Kilise	[50180 Göreme Belediyesi, Türkiye]	38.639540	34.844933
6	Karanlık Kilise	[Göreme Açık Hava Müzesi, Göreme Tarihi Milli ...]	38.638551	34.846021
7	Yılanlı Kilise	[Göreme Açık Hava Müzesi, Ürgüp, Türkiye]	38.638528	34.845307

I used Python folium library to visualize the museums in the Göreme Region. I used latitude and longitude values to get the visual as below:



## Search for Restaurants

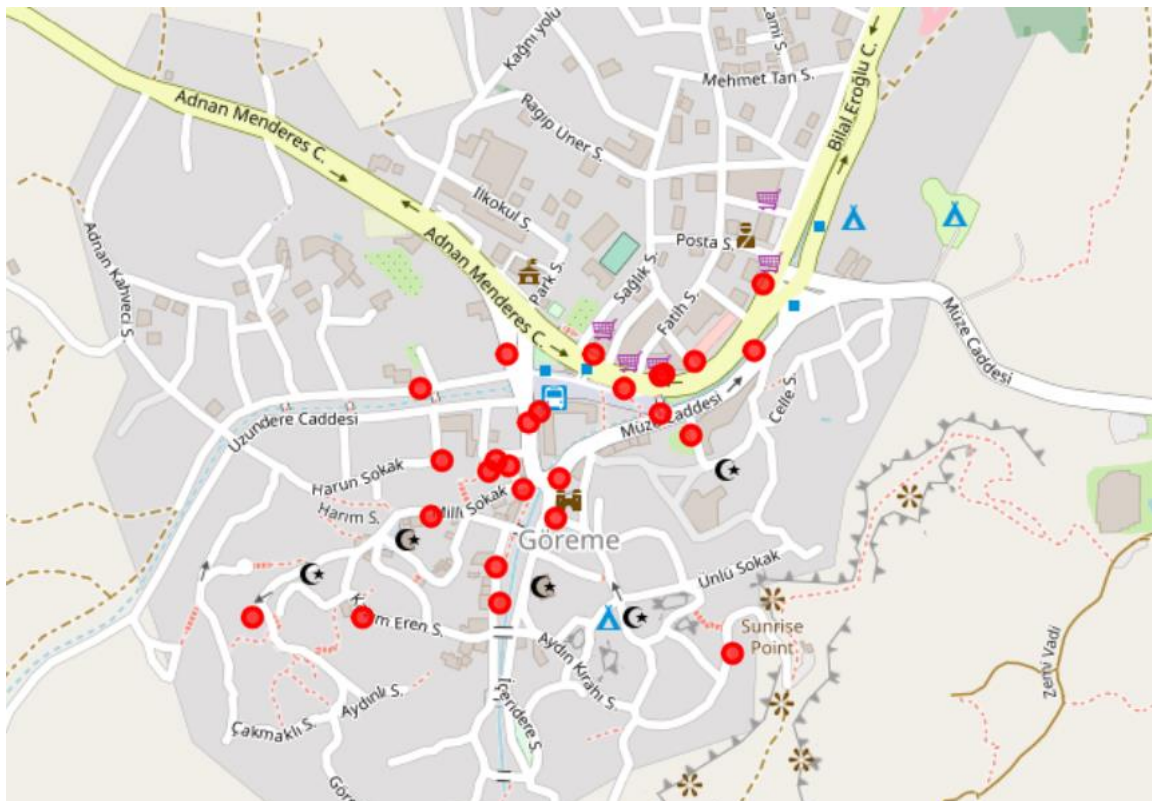
I searched the restaurants within a range of 5 kms.

	categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	location.distance	location.formattedAddress	locati
0	[{"primary": True, "id": "4bf58dd8d4898d1c494..."}]	False	4c22570413c00f47635587de	Belediye Cd.	TR	Göreme Tarihi Milli Parkı	Türkiye	NaN	127	[Belediye Cd., 50180 Göreme Tarihi Milli Parkı...]	3
1	[{"primary": True, "id": "4bf58dd8d4898d1c494..."}]	False	56e3f6b1498e8a4e8acbe205	Orta Aydınli Mah. Milli Sk. Göreme	TR	Ürgüp	Türkiye	NaN	107	[Orta Aydınli Mah. Milli Sk. Göreme, Ürgüp, Tu...]	3
2	[{"primary": True, "id": "4bf58dd8d4898d1c494..."}]	False	54f457f498ea30870927516	Orta Mah.Uzun Dere Cad. No:9	TR	Göreme Tarihi Milli Parkı	Türkiye	NaN	201	[Orta Mah.Uzun Dere Cad. No:9, 50180 Göreme Ta...]	3
3	[{"primary": True, "id": "4bf58dd8d4898d1c494..."}]	False	51eae36c498e9c7da2723ee8	NaN	TR	Nevşehir	Türkiye	NaN	97	[50000 Nevşehir, Türkiye]	3
4	[{"primary": True, "id": "4bf58dd8d4898d1c494..."}]	False	510fe1ede4b0aaab99fce8c1	Orta Mah. Milli Sok. No:6 Göreme	TR	Nevşehir	Türkiye	NaN	110	[Orta Mah. Milli Sok. No:6 Göreme, Nevşehir, ...]	3

I cleaned the data by using drop function to keep only the columns including name, address and geographical coordinates. Then I put the "name" column to the first column.

	name	formattedAddress	lat	lng
0	One Way Cafe Restaurant	[Belediye Cd., 50180 Göreme Tarihi Milli Parkı...	38.643175	34.829141
1	Viewpoint Cafe & Restaurant	[Orta Aydınllı Mah. Milli Sk. Göreme, Ürgüp, Tü...	38.642711	34.828678
2	Şaziye Restaurant	[Orta Mah.Uzun Dere Cad. No:9, 50180 Göreme Ta...	38.643803	34.828888
3	Cappadocia Pide House Restaurant	[50000 Nevşehir, Türkiye]	38.641496	34.828805
4	My Mother's Cafe & Restaurant	[Orta Mah. Milli Sok. No: 6 Göreme, Nevşehir, ...	38.642821	34.828761
5	Alaturka Restaurant	[Goreme, Cappadocia, Türkiye]	38.643253	34.830709
6	Dibek Cafe & Restaurant	[Hakkı Paşa Meydanı No: 1 Göreme (Kapadokya), ...	38.642551	34.829095
7	Köy Evi Restaurant	[Göreme Tarihi Milli Parkı, Türkiye]	38.641024	34.831573
8	Mozaik Restaurant	[Belediye Caddesi, El Sanatları Çarşısı, No 31...	38.643264	34.829287
9	Urizip Korean Restaurant	[Türkiye]	38.643494	34.830290
10	Millocal Restaurant Kapadokya	[Tekeli Mah. Divanlı Sok. No:20 (Göreme), Üçhi...	38.631171	34.808063
11	Old Cappadocia Cafe & Restaurant	[Göreme Tarihi Milli Parkı, Türkiye]	38.641825	34.828764

I used Python folium library to visualize the restaurants in the Göreme Region. I used latitude and longitude values to get the visual as below:



## Machine Learning: K- Means Algorithm

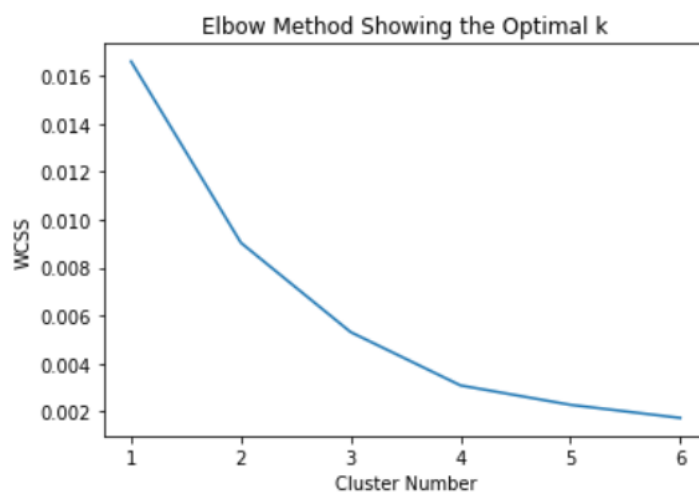
I will try to find the best region to accommodate which is near to venues as much as possible. To do this K- means algorithm is used.

First I merged all the hotels, museums and restaurants tables into a single table using concat function. Now, our new table(df\_new) consists **87 rows** and **5 columns**. I get the row and column numbers using shape function.

```
df_new = pd.concat([museum_dataframe4,hotel_dataframe3,resto_dataframe3])
df_new.shape
```

```
]: (87, 5)
```

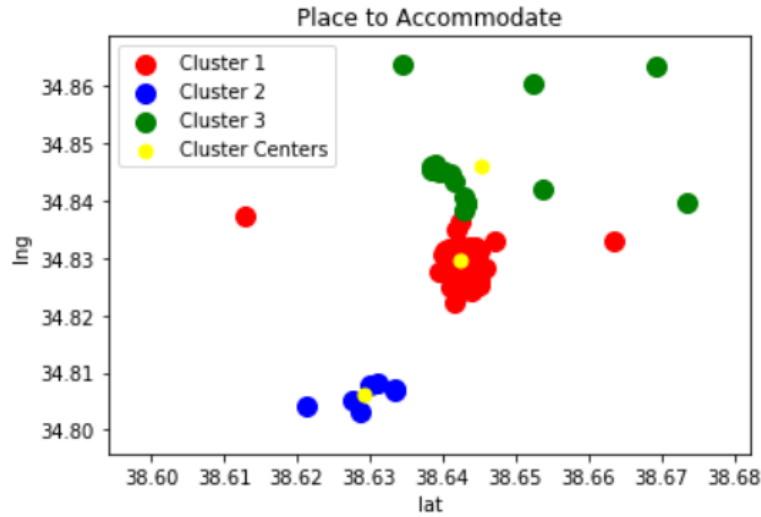
Then I looked for the elbow method which suggests the optimum k of the K-Means.



I took k as 3 degrees and cluster the venues into 3 subsets. I then applied K-means algorithm to cluster venues according to geographical coordinates and find the optimum cluster for accommodation.

#### 4. Results

I used matplotlib.pyplot to visualize the results. All the venues are shown in the graphs according to their geographical coordinates. The K-Means algorithm clustered the venues in the 3 areas shown in green, red and blue colors. The cluster centers of each cluster are shown with yellow points.



## 5. Discussion

Cappadocia is a region spreading an area covering the provinces particularly to Nevsehir, Kirsehir, Nigde, Aksaray and Kayseri. The region nowadays has a great importance in terms of Tourism.

The main worthseeing places are Ürgüp, Avanos, Göreme, Akvadi, Uçhisar and Ortahisar Castles, Güvercinlik Valley, Derinkuyu, Kaymaklı, Özkonak underground cities, Ihlara Valley, Çavuşin, Güllüdere Valley.

In this study I mainly focused on the Göreme region but the study can be expanded including the above regions and attraction points. As there is such variety, different approaches can be tried in clustering and classification studies.

I used the Kmeans algorithm as a machine learning method. When I tested the Elbow method, I set the optimum k value to 3. However, if the study is expanded to other regions and attraction points, other clustering methods can be preferred.

In this study to recommend the best cluster to accommodate I cluster the areas according to geographical proximity. Other data like prices of the hotels and ratings can also be analyzed together to gain a better insight.

## 6. Conclusion

*Cappadocia* is one of the most visited touristic attraction points in Turkey which is known for its fairy chimneys, natural rock formations, historical monasteries, boutique hotels, hot air balloon tours and now it is very popular destination with a growing amount of visitors each year.

In this study, according to FourSquare API venue recommendations I listed the tables of hotels, museums, restaurants in the Göreme region and visualized the map of these venues. Then I merged all the tables into a single table and used K-means clustering algorithm as a machine learning method to cluster the venues according the distances between the venues. I defined centroids of the clusters and found that Cluster 1 will be the best choice to accommodate.



This kind of data analysis will be useful not only for tourists visiting a city for the first time but also for tourism agencies to inform & recommend their customers about the city and also for investors to invest in the city for new opportunities.

