Venues Analysis and Comparison of Two Fastest Developing Cities: Manisa&Gaziantep

Melik Haci Ahmet Ovali

March 14, 2021

1.   Introduction

1.a.   Discussion of the Background

Developing markets offer huge opportunities for entrepreneurs as there is flux of income for residents many of which ends in consumption. Therefore, developing markets always attract more investment than not promising ones. In a developing market where to locate a specific kind of business is always a challenging problem. As the location of a business is one of the most important factors for its success. Depending on the content of the business, location selection can be based on competition, proximity to other businesses or room for growth, etc. Decision is left to the entrepreneur, however, a good analysis and picture of the markets should be provided for him/her to launch the business in the right place.

1.b.   Description of the Problem

Location data can be used for the initial analysis of markets within a place or even in multiple places. Data may contain only names and coordinates of venues or may contain other attributes such as venue categories. Because, Google Maps has limited the number of free calls with its API, other location data ought to be used for analysis purposes. In this study Four-Square API will be adopted to analyze the number of venues, venue locations, venue categories and venue clusters for selected markets.

To enlarge the scope of the study, two market places, which are distinct cities in this study, are analyzed and compared to give initial insights on the venues and venue cluster locations. In order to determine the cities to analyze, highest city development ratios approach is adopted as follows. (Note: Year 2020 is excluded due to the Corona Virus Effects on the markets)

No	City	GDP Volume in 2018 (in Billion Turkish Liras)	GDP Volume in 2019 (in Billion Turkish Liras)	Change Rate
1	Manisa	29.67	31.17	5.1%
2	Osmaniye	7.31	7.60	4.0%
3	Gaziantep	31.95	33.12	3.7%

Table-1. First Three Cities of Turkey with Highest Gross Domestic Product (GDP) Growth Ratio in 2019[1]

From the table above, we see top three cities with highest development ratios of GDP, but one city, Osmaniye, has low GDP volume with respect to other two cities. By only looking at the development ratios of the cities may lead us to take into account small markets as promising ones. For this reason, one step further to determine cities to analyze, national standings of cities in terms of GDP Volumes is examined as below.

No	City	GDP Volume in 2018 (in Billion Turkish Liras)	GDP Volume in 2019 (in Billion Turkish Liras)	Change Rate
1	Istanbul	540.35	549.33	1.7%
2	Ankara	154.27	151.16	-2.0%
3	Izmir	109.22	111.12	1.7%
4	Bursa	72.65	73.58	1.3%
5	Kocaeli	70.24	71.23	1.4%
6	Antalya	53.21	53.88	1.3%
7	Konya	36.84	36.17	-1.8%
8	Adana	33.76	33.78	0.1%
9	**Gaziantep**	31.95	33.12	3.7%
10	**Manisa**	29.67	31.17	5.1%

Table-2. Top 10 Cities of Turkey with Highest Gross Domestic Product (GDP) in 2019[1]

From the table above, Manisa and Gaziantep are observed in both the Fastest Developing Cities and in Top-10 Cities with Highest GDP volumes. For that reason this study will focus on to analyze and compare how these two cities are similar or different from each other in terms of venues. As they are located in west and east part of the Turkey, respectively, analysis may reveal also interesting results due to the geographical and cultural differences as well.

1.c. Interest
Clearly, entrepreneurs who wants to decide where to launch their specific kind of business or who wants to see the picture of business clusters so as to detect the market need and to see whether there is a room for growth would benefit from this study. The power of this study stems from the fact that, it can be adjustable for anywhere and for any kind just by typing different city names at the beginning of codes.

2. Data

2.a. Data Source
In the study, location and venues data is acquired by using Four-Square API in Jupyter Notebook at IBM Cloud Watson Studio Service with Python Runtime Environment. Returned files from the API calls are JSON files with all the venues are within the 2000 meters radius from the coordinates of cities that are acquired via python's "geopy.geocoders" library.


```
In [7]: manisa_url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&oauth_token={}&v={}&radius={}&limit={}'.format(CLIENT_I
        man_results = requests.get(manisa_url).json()
        print(len(man_results['response']['venues']),'venues returned for Manisa')

        121 venues returned for Manisa

In [8]: gantep_url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&oauth_token={}&v={}&radius={}&limit={}'.format(CLIENT_I
        gzt_results = requests.get(gantep_url).json()
        print(len(gzt_results['response']['venues']),'venues returned for Gaziantep')

        129 venues returned for Gaziantep
```

Figure-1. Four-Square API Calls and Returned Values for Both Cities

In the study, there are 121 venues returned for Manisa City and 129 venues returned for the Gaziantep City. An important note here is that, number of venues returned from Four-Square API calls do not reflect the whole but only those are marked by users of Four-Square. However, these samples can still give us insights about the whole.

2.b.   Data Cleansing

The returned JSON file has lots of key-value pairs, but not all of them are needed in the study. Thus, data cleansing is carried out to filter necessary attributes of each venue for both cities. Cleaned data ready for the analysis are transformed into distinct data frames with no lost-data. First five rows of each table is shown below.

Figure-2. Manisa City Venues Data Frame First Rows

Figure-3. Gaziantep City Venues Data Frame First Rows & Dimensions of Data Frames

In order not to lose any data before starting the analysis, venues without an address are also kept by assigning 'pass' markers to address values. We have still 121 venues for Manisa City and 129 venues for Gaziantep City after transforming JSON file into data frames. Now that we have our data ready and structured, we can begin to analyze and compare venues of two distinct cities.

3.   Methodology

3.a.   Exploratory Data Analysis

As this study aims to analyze and compare venues of two cities and present a picture of venue clusters so that an observer can see which venue categories are close to each other (accumulated) and which venue categories are located away from each other, we can start with finding the frequency of distinct venues for each city. Results are shown below.
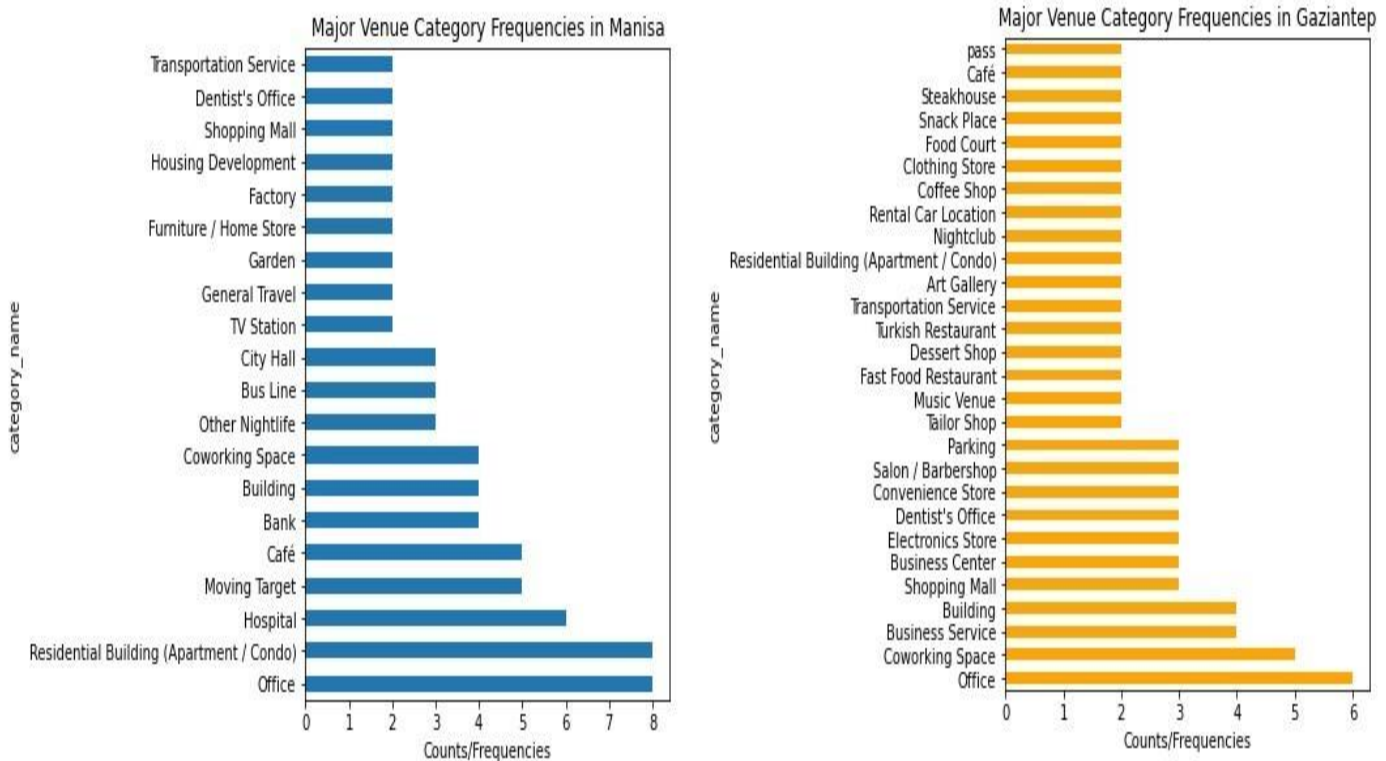

Figure-4. Distinct Venue Frequencies Chart for Both Cities

From the charts above, it is observed that most common venues have only 8 and 6 counts for each city respectively. In other words, most of the venues are unique (counts one) in their category. Although these charts give some insights about the characteristics of cities, with some data processing, venues can be clustered in terms of their types that will give better understanding about the market properties of the cities.

Before proceeding to cluster venues by their types, we can visualize the venues on city maps using the Python Folium Library. If the distinct venues were counted much, we would cluster them by their location as well.
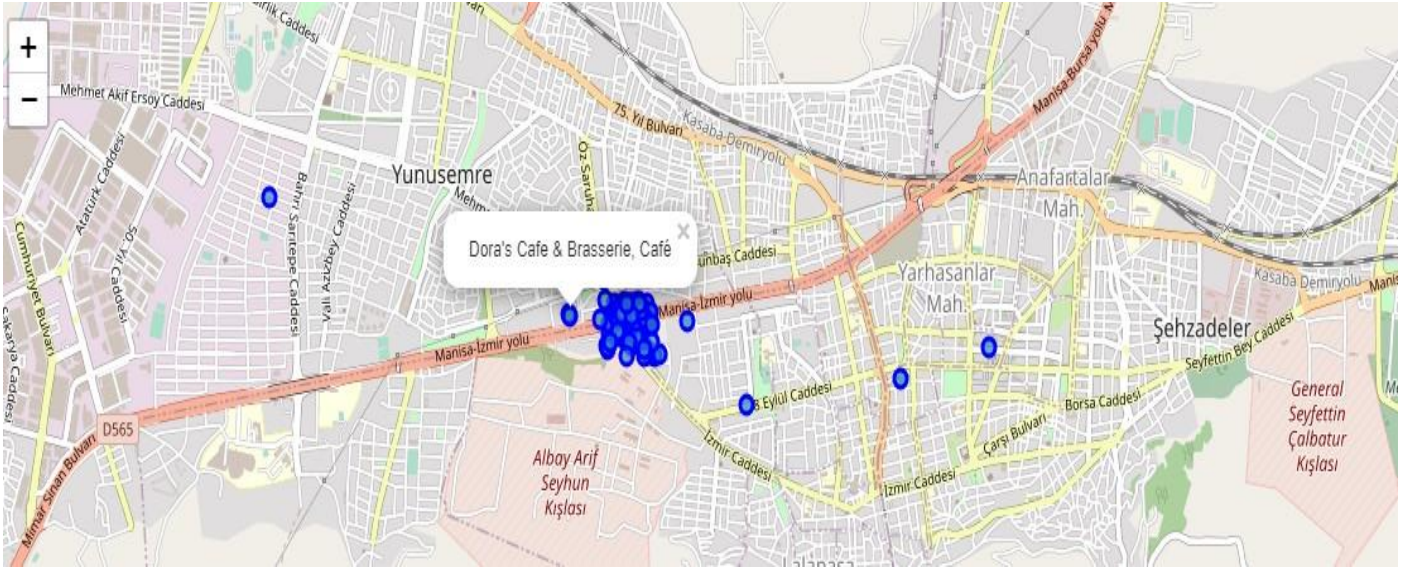
Figure-5. Venues with Name & Category in Manisa City

From the Manisa City map above, it is obvious that most venues are clustered closely independent of their venue category. Recalling that there are 121 venues for this city, only 7-8 of them are not in the cluster. For that reason, K-Means clustering which is one of the most common clustering algorithms can not be applied here.
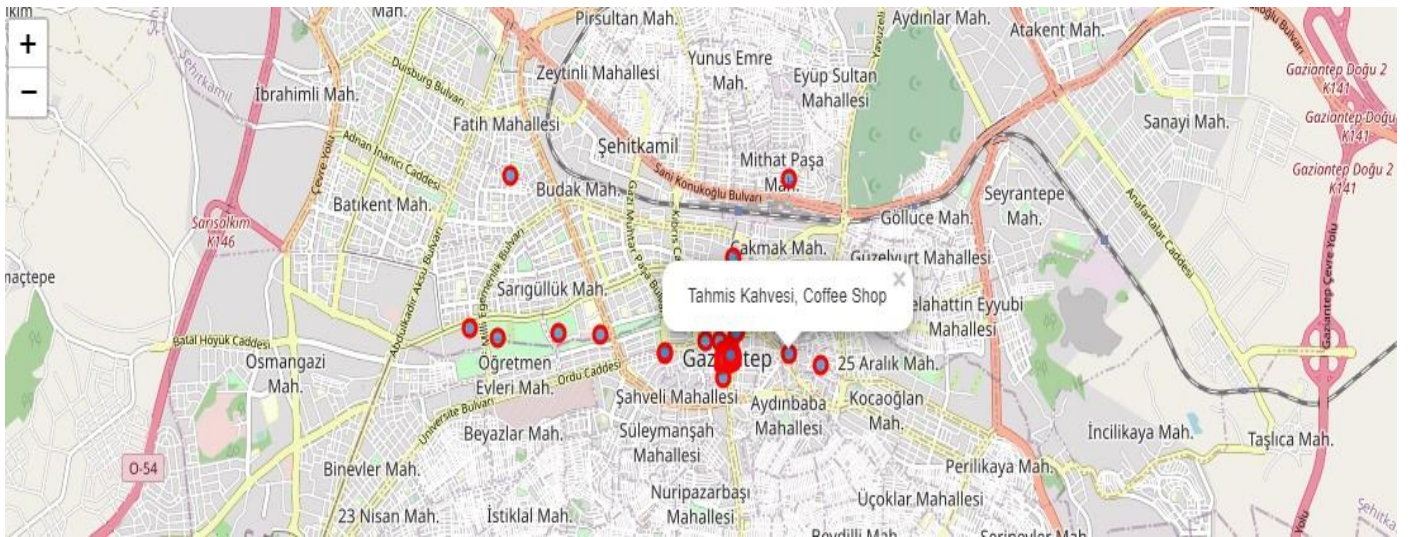

Figure-6. Venues with Name & Category in Gaziantep City

Now for the Gaziantep City map above, there seems to be only around 10 venues separated from the cluster. Recalling that there 129 venues in this city, again, K-Means clustering would not work as desired here, too. Grouping the venues by their category, however, may result in crowded venue category groups that would help us to carry out our analysis.

To form groups, venue categories with more than one count are cumulated under category group tags and all the venues are checked with a for-loop to append venues under these category groups. Category group tags and categories that fall under these tags are shown as below.

3.b. Data Wrangling

To draw more meaningful results, category names with more than one count (frequency) are gathered under tags as listed below. Venues with category names other than listed below would not be taken into consideration as there are venues with "default" marked category names and with unique category names as, well. Not only grouping categories under relative tags would produce crowded clusters, but also default marked or unique category names would be eliminated from the analysis. Therefore noise data would not be included in the analysis.

Group Tags	Housing 0	Medical 1	Business 2	Catering 3	Shopping 4	Dressing 5	Art&Music 6
Venue Categories	-Housing Development - Residential Building (Apartment /Condo)	- Dentist's Office - Hospital	- Coworking Space - Factory - General Travel - Office - Coworking Place - Business Service - Business Center - Rental Car Location	- Bakery - Fast Food Restaurant - Dessert Shop - Turkish Restaurant - Coffee Shop - Food Court - Snack Place - Steakhouse - Cafe - Café	- Shopping Mall Convenience Store - Clothing Store Electronics Store - Optical Shop - Furniture /Home Store	- Salon /Barbershop - Tailor Shop	- Music Venue - Art Gallery - Music Store

Table-3. Venue Group Tags and Corresponding Venue Categories

Both city data frames are iterated through to check if there is any venue category matches to one of above. If yes, that venue is marked with a number and a grouping tag. Assigning each grouping tag with a number stems from the fact that, numerical analysis can not be performed with categorical variables. New city data frames with two additional columns as presented below. (Please check the last two columns.)

	name	address	latitude	longitude	category_name	venue_category	category_gr	category_gr_num
0	Merkez Efendi Devlet Hastanesi	Merkez Efendi Mh. 45020 Manisa, Türkiye	38.615879	27.408422	Hospital	medical	medical	1
1	Manisa Meteoroloji İl Müdürlüğü	pass	38.615385	27.404768	Spiritual Center	religious	None	NaN
2	Forum Magnesia	Laleli Mah. Ali Rıza Efendi Cad. No:4	38.616166	27.400331	Shopping Mall	mall	shopping	4
3	Lale Meydanı	pass	38.617454	27.402748	Plaza	plaza	None	NaN
4	Yunus Park	pass	38.615340	27.403179	Recreation Center	sportsoutdoors	None	NaN
5	Cafe Sera	Hakkı İplikçi Parkı	38.613658	27.422914	Café	cafe	catering	3
6	Yunusemre Belediyesi	Yunusemre	38.614899	27.402903	City Hall	cityhall	None	NaN
7	metoroloji müdürlüğü	Manisa	38.615083	27.405007	Farm	farm	None	NaN
8	Manisa Çarşısı	pass	38.615793	27.403437	Public Art	default	None	NaN
9	Manisa Çiftlik Parkı	Akmescit Mah. İzmir Cd. No:307	38.615225	27.404286	Park	park	None	NaN

Figure-7. Manisa City Data Frame with Grouping Names and Grouping Numbers

As seen from Figure-7&8, if a category is not commercial (not interesting for an entrepreneur), "None" is assigned to its category and "NaN" assigned (via Python's Numpy Library) as its group number. So that, non-commercial venues will be dropped later.

6

Figure-8. Gaziantep City Data Frame with Grouping Names and Grouping Numbers

Converting all categorical variables into numeric, we can drop the non-commercial venues and the two data frames take their final shape for the analysis. Checking two data frames we observe that **49** out of **121** venues for Manisa City and **62** out of **129** venues for Gaziantep City remain in the data frame which our analysis will be based on.

3.c. Initial Inferences

After dropping venues with "None" assigned categories, 49 venues category distribution of Manisa City and 62 venues distribution of Gaziantep City is shown below.
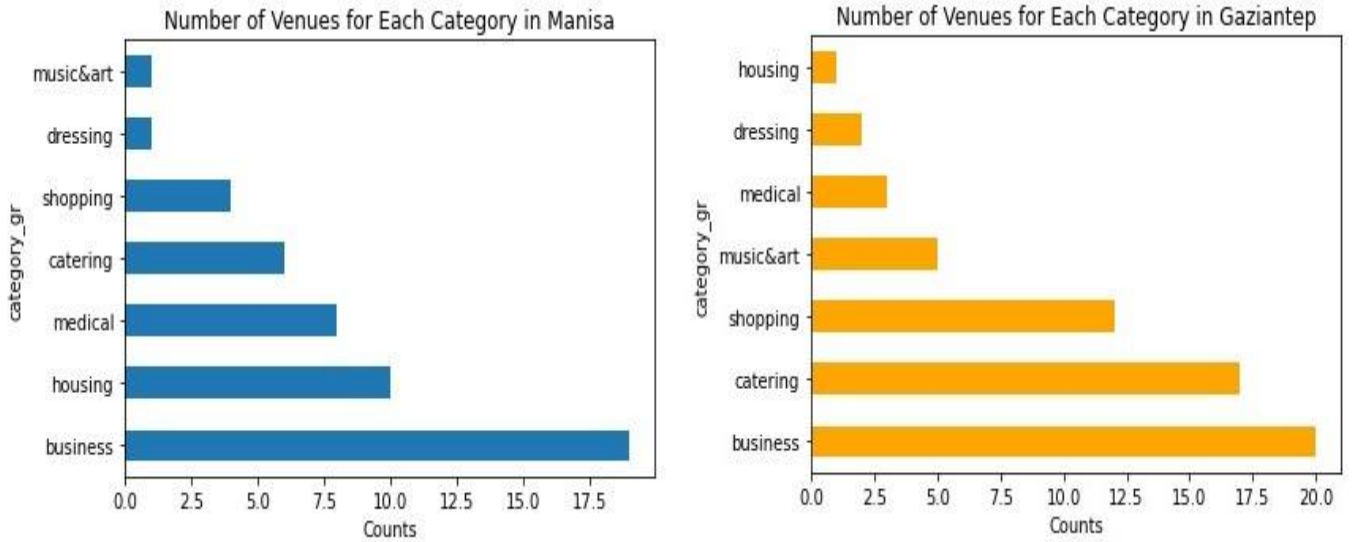

Figure-9. Venues Category Distribution of Both Cities

From the charts above, it is observed that:

1. Business venues are at most for both cities, as city centers are usually good places for accumulation of business.

2. For the second rows, Manisa City has housing venues with a number of 10, Gaziantep City has catering venues. Although it may seem reasonable for the Manisa City to have housing (apartments or condos) in the second row, result for the Gaziantep City, 17 catering venues for second row, is due to the fact that Gaziantep City is listed by the UNESCO as one of the cities of gastronomy.[2]

3. For the third rows, 8 medical (mostly dentists) venues for Manisa City catches some attention, 12 shopping venues for Gaziantep City, however, can be inferred as high shopping habits.

3.d.   Visualization of Cluster Groups

Data frames for both cities still have latitude and longitude columns, which enables us to examine the venue category groups map distribution throughout the cities. These maps can also present any clusters of category groups that is also referred in the location problems as "competitive". Let's start our analysis with Manisa City. Group colors represent venue categories for both cities as follows.,

Purple: Medical, **Turquoise**: Catering, **Pistachio Green**: Shopping, **Blue**: Business,
Red: Housing, **Orange**: Music&Art, **Warm Ivory**: Dressing

Figure-10. Venues Category Group Clusters for Manisa City

From the high view map of Manisa City, we observe that:

1. **Medical** (Purple) venues cluster together, except for the state hospital seen alone in the right, that can be inferred as they locate competitively. From the data frame we see that almost all of them are dentist's offices.
2. **Catering** (Turquoise) venues scatter around the other venues, which can be inferred as they aim to catch the need of customers just in place.
3. **Shopping** (Pistachio Green) venues are both stands close to other venue clusters to meet the needs as well.
4. **Business** (Blue) venues mostly cluster closely to each other, not because they are competitive, but because customers want to find them in close range to have their all needs met once they get to the point. Another inference can be that business venues mostly related to each other even if they have different categories. For example, if there is a customer service center or notary, there should be also an photocopy center or a photographer.
5. **Housing** (Red) venues are also clustered together, as there are a lot of advantages such as reaching utilities and creating an lively environment.
6. **Music&Art** (Orange) venue is single in Manisa City data frame, but it is remarkable that it is located at the center point of the housing venues that hosts most of its customers.
7. **Dressing** (Warm Ivory) venue is observed in Manisa City clusters, as it counts only one.

Now, we can examine Gaziantep City venues cluster map in the same way. Colors represent the same venue categories as stated above.
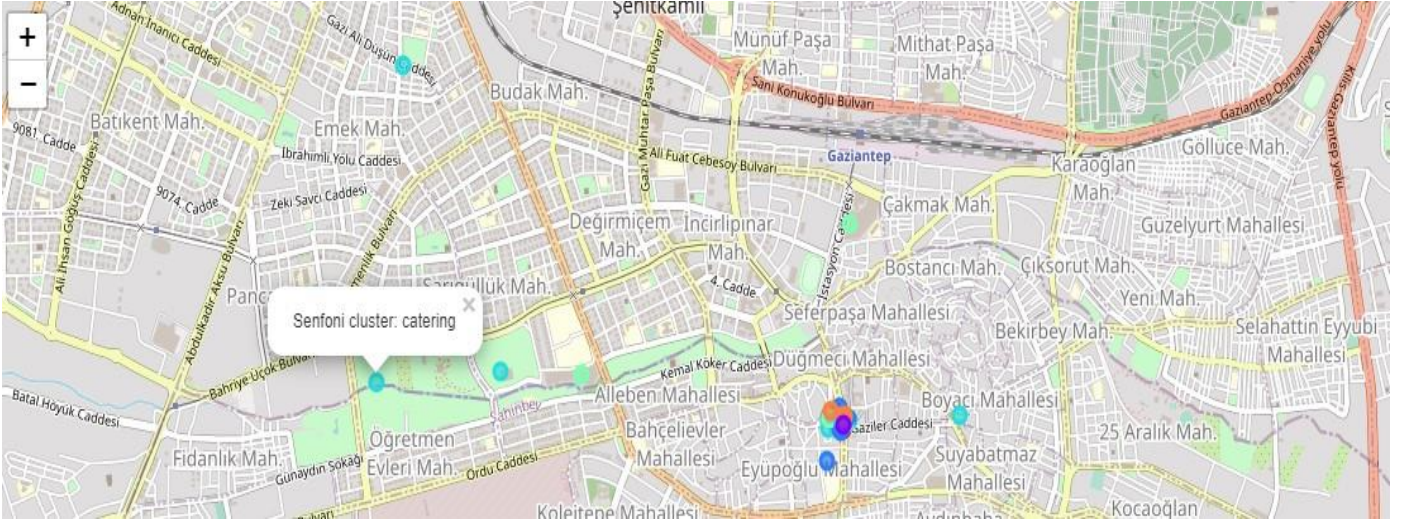
Figure-11. Venues Category Group Clusters for Gaziantep City

From the map above, it is obvious that catering venues are spread throughout the city. Shopping venues (which is quite common in this city) also spread on wide ranges. Other than there two category group clusters are accumulated in the city center. Thus, we need to examine center cluster closely.

Figure-12. Close Examination of Venue Category Clusters for Gaziantep City

From the map in Figure-12 above, we observe that:

1. **Medical** (Purple) venues cluster together again. From the data set we see these medical venues are dentist's offices, which makes sense to locate them closely. They share the market competitively.

2. **Catering** (Turquoise) venues scatter around the other venues again for the same reason stated earlier.

3. **Shopping** (Pistachio Green) venues which are common in this city, clustered in small groups and also scattered around the other venues. This can be stated as, they target to meet demand in place, and within the place, they are competitive to receive customers.

4. **Business** (Blue) venues are, again, clustered together for the same reasons stated above.

5. **Housing** (Red) venue is sole in the map above, for that reason, no inference can be drawn from it.

6. **Music&Art** (Orange) venues are two small clusters with two venues. Just like the medical venues, they share the market in a competitive manner.

7. **Dressing** (Warm Ivory) venues are spread within the cluster, which can be interpreted as they meet the need just in place without acting competitively.

4. Results

4.a. Venue Characteristics of Both Cities

By examining both city venues closely , we see both similarities and differences in terms of venue categories and venue clusters.

To start with similarities:

1. We see that both city center have "business" as most common venue category. This observation is quite reasonable and holds true for the most city centers.

2. We see also that dentist's offices are common as "medical" venues in both cities and they are located closely, which points out they are competitive.

3. When it comes to "catering", it is observed that this type of venues; such as restaurants, snack place or cafes, try to exist where the demand is obvious. In other words, they are spread within the other types of clusters which have both employees and customers. However, it is also remarkable that catering venues avoid competition as possible as they can, but they try to meet the need where ever it exists.

To continue with differences:

1. It is observed from the maps that Gaziantep City has its venues more spread than Manisa City.

2. Within the shopping group of Gaziantep City, due to target more customer, we see that shopping venues are located in a competitive manner.

3. Although both cities have close GDP's, it is observed that Gaziantep City has more commercial venues. That may be an indicator of high tendency of consumptions.

4.b. Venues Comparison

To compare venue category groups closely, let's render a quick look at the venues comparison chart below:

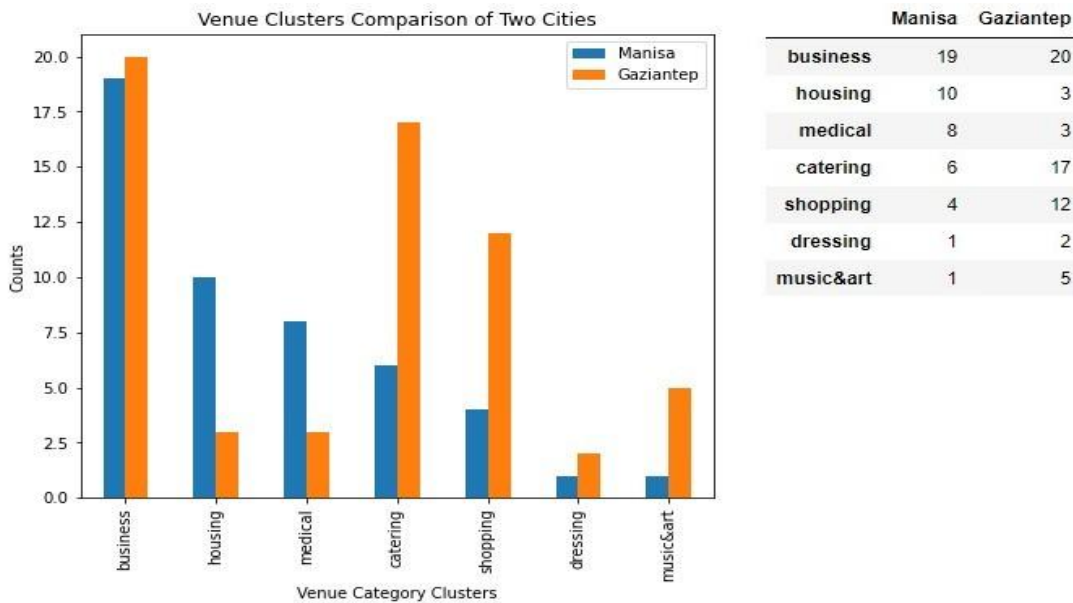|	Manisa	Gaziantep
business	19	20
housing	10	3
medical	8	3
catering	6	17
shopping	4	12
dressing	1	2
music&art	1	5

Figure-13. Venues Category Groups Comparison of Two Cities

From the chart above:

1. It is clear that no city has superior to the other in terms of business venues category.

2. For the housing category, Manisa City has outnumbered the Gaziantep City.

3. For the catering, shopping and music-art venues category groups, Gaziantep City has an obvious superiority. This finding is important to indicate that consumption inclination seems to be quite higher in this city.

5. Discussion

In the study, **121** venues for Manisa City and **129** venues for Gaziantep City were returned by Four-Square API calls, which composes the data set. Although, our analysis depends on total of 250 data elements, it is still capable of giving initial insights on how the venues of both cities located and how two cities are different or similar.

Even with a relatively small data set, visualization and comparison managed to provide basic indicators of market and customer tendencies as well as similarities and differences between two places. Finding catering venues common in Gaziantep City was also verified with a quick browser research that lists the city as one of gastronomy places of world by the UNESCO.[2] "1 "

This study can only be one part of the market research, as more research should support the entrepreneurs decision to launch a business at any place. However, the attracting point in this study is that, market places can be analyzed and compared without any on-site visit research.

6. Conclusion

In the study, firstly, venues data of two cities is acquired via the Four-Square API calls, then, these venues are exploratory analyzed which showed that a clustering method should be developed to form relational and crowded groups. Next, to see if the venues would be clustered with K-Means clustering algorithm which depends on location proximity, Python Folium Mapping tool used and maps are generated. Examining the maps for both cities, venues are observed to cluster closely no matter their category type. Thus, clustering approach moved from location proximity to categorical relations.

To form categorical groups, venue categories are listed under relative grouping tags and all venues are iterated to check if there is any match. By this way, new data frames are formed with **49** venues for Manisa City and **62** venues for Gaziantep City. Plotting these venues on maps of both cities, we drew important inferences on venues and their distribution as well as compared two cities to see how similar or different they are.

7. References

[1]   https://www.tepav.org.tr/upload/files/1602216836-7.Gece_Isiklariyla_Il_Bazinda_GSYH_Tahmini _2019___da_81_Ilin_Kisi_Basi_Geliri.pdf

[2]   https://jotags.org/2018/vol6_issue4_article48.pdf