Coursera Capstone Project

İstanbul Venues

Savaş Türkoğlu 2020

Abstract

In this project, machine learning was used to estimate a good place for a tourist in center of Istanbul.

Contents :

1 Introduction

İstanbul is most popular city in Turkey and more than 20 million people choosing this beautiful city for their holiday every year from abroad. The best places to eat in Istanbul are deliciously diverse

In this favourite holiday city, there are many many restaurants and cafes of different ambiance from many different food cultures. also there are many different beaches resorts int this city.Traditional turkish cuisine, italian pizza restaurants, seafood restaurants, authentic middle east cafes, delicious kebab restaurants, turkish coffee, fast food, bistros, so many options for a tourist.

There is not enough time for a tourist who has been come on vacation for 10 days to get the experience up to the right place to choose the right restaurant for dinner.there is no time for a tourist to learn which restaurant is cooking better or which restaurant is cooking more delicious fish.

2 Data

A description of the data: the data used to solve this problem is geolocation data collected from FourSquare. Adequate explanation and discussion, with examples, of the data is the following. Data is a single dataframe, containing at least a location of the café and restaurants. Explanation of the location data is a standard tuple (lat, lng), where lat stands for latitude and lng for longitude. Some other metadata like name, postal code and so on is also collected, but let us discuss that they are

not absolutely necessary for the analysis. Example of the data used in analysis is shown in table

	venue.name	venue.id	venue.categories	venue.location.lat	venue.location.lng
0	İoki	50377a92e4b0fc42f6c6bb79	Sushi Restaurant	41.073893	29.057801
1	Kandilli Sahili	4d7f764d564b8cfa42840565	Beach	41.074378	29.059562
2	Adile Sultan Sarayı	4e1d862ac65b6cd0da9b2086	Historic Site	41.072159	29.058089
3	Mithat Alam Film Merkezi	4d959f46942ba0934c733e8c	Movie Theater	41.082960	29.050284
4	Bebek Parkı	4b811ccdf964a520379730e3	Park	41.076294	29.044012
...	...	...	...	...	...
95	Cup of Joy	5730bf4d498e9d5244e895a9	Coffee Shop	41.065425	29.018019
96	Kanlıca Sahili	4c14b748a1010f4740da4c18	Waterfront	41.099706	29.075694
97	Zorlu Center Park	527e37d511d299685ade1145	Garden	41.067336	29.015274
98	Abant Kahvecisi	58cedc897b88a71aacfedfb4	Coffee Shop	41.050100	29.053721
99	Maia Handmade Chocolate Atolye	5743104a498e6adb99b3ce0b	Café	41.051954	29.071567

Data will be used in the following way: by knowing the locations of already existing cafes and restaurants, it's possible to apply unsupervised learning technique like kernel density estimation to determine the area of influence of the existing cafes, and start up new café which is not in the area of influence.

3 Methodology

One of the key elements in the preferred neighborhood in this project.
We also need to know the coordinates and locations of these
neighborhoods, and therefore the geocode API has been used to
achieve this goal. This is important so that we can enter this information
into the location information provider like Foursquare.com to get venue
information in these neighborhoods, and that's exactly what we did in
this project. We will also use machine learning techniques such as
K-Means to segment and cluster them.
we can group them to understand their similarities. In other words, we
will not recommend management to enter a neighborhood where there is
already a high density of groceries and less demand in the
neighborhood.
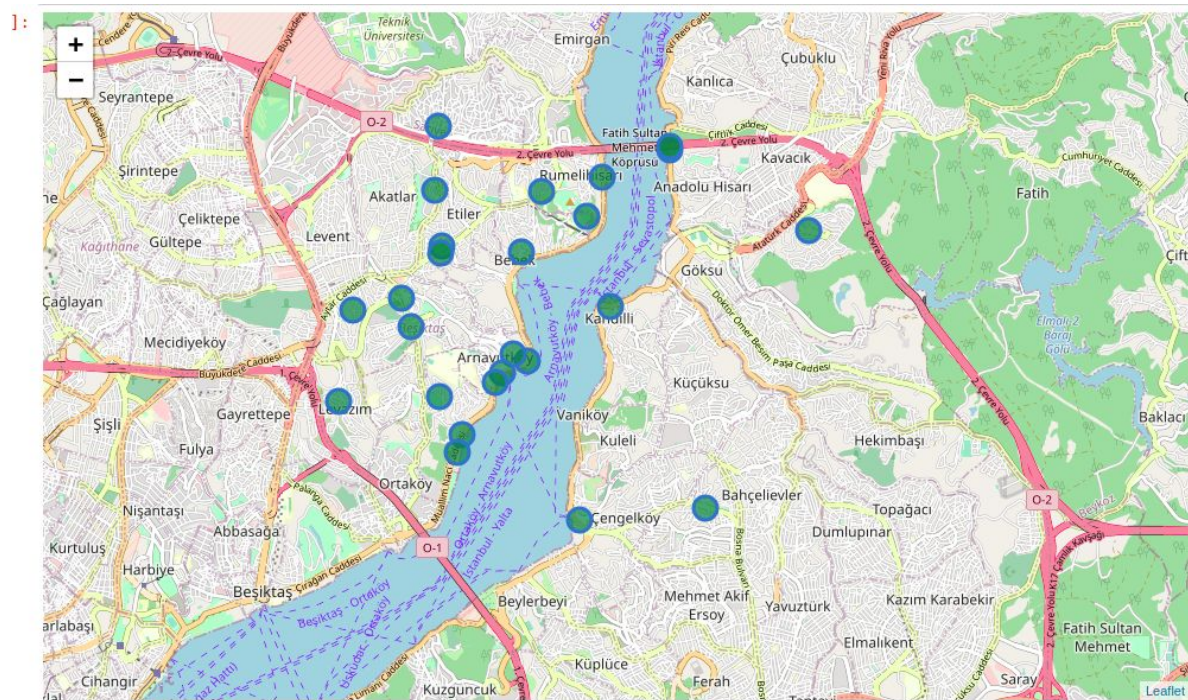
4- Results

With K-Means clustering technique, the top 5 clusters of similar neighborhoods have been apparent in the result, see below. These clusters are group together based on the similar nearby venues in each of the neighborhoods. This information is critical so that we can target on the cluster that offer the largest. their business model fairly quickly upon success in their first service offering in the selected neighborhood.

Discussion

From the above results we can state that the grouping of cuisines based on their rating is done very accurately but the results may have a limitation of accuracy as the whole results are based on the data we fetch from the Foursquare API calls hence further more advancements can be made by considering huge amounts of data.

Conclusion

In this project, we tried to categorize the venues in Istanbul according to the number of like and type.We can say that we have achieved a successful result even though the FourSquare data is very low for Istanbul. Also we can say it is very useful For tourists who want to travel around the city and discover new quality venues. Therefore this work

could be very useful for some company that creating advice programs and applications for tourist.