Capstone Project – The Battle of Neighborhoods in Madrid City for Food Chances

Introduction

In 2019, Spain was the second most visited country in the world, recording 83.7 million tourists which marked the seventh consecutive year of record-beating numbers.[1] The country has been a popular destination for summer holidays, especially with large numbers of tourists from the United Kingdom, Ireland, France, Germany, Italy, the Benelux, and the United States, among others. Accordingly, Spain's foreign tourist industry has grown into the second-biggest in the world.[2]

Spain's capital and its largest city, Madrid is a top destination for many visitors. And for good reason — it's a lively city that's both affordable by Western European standards and boasts outstanding culture, architecture, art, and excellent nightlife. But it's also a fairly modern city so you'll discover that much of the architecture is from the 20th century — which can sometimes be a surprise to many visitors looking for "old" Europe. That said, it is a city overflowing with energy and old-world charm that keeps pulling visitors back year after year.[3]

Something is happening in the world of gastronomy in Madrid. We are witnessing a phenomenon in which the cuisine permeates everything. New gastronomic spaces are born every little time in the city, specialty shops, cooking schools, culinary tours and, of course, establishments, make a living and highly attractive environment for visitors.[4]

The target of this paper is to discover a selection of that gastronomy using foursquare data from a list of the best tapas bar in Madrid to a most enjoyable culinary experience, from the peculiar character of Madrid`s cuisine alive to all kinds of influences such as the rest of the country, international even the most innovative movement.

Data Description

In order to carry out our research about the gastronomy in Madrid, it is necessary to obtain information about the restaurant and the districts.

To take the data about restaurant in Madrid, the website of Foursquare, a US tech company from New York focusing on location data. Their technology and data powers apps such as

---

[1] *Spain sets new foreign tourist arrival record for sixth consecutive year".* *4 February 2019. Retrieved 9 February 2019.*

[2] Archived 6 January 2011 at the Wayback Machine

[3] https://thesavvybackpacker.com/city-guide/madrid-travel-guide/

[4] https://www.turismomadrid.es/en/to-see-and-do/gastronomy/10175-gastronomy.html-.

Apple's Maps, Uber, Twitter and many other household names. Data such as the restaurant name, ID, location and category of food – Spanish, Argentina, Morocco etc.-.[5]To extract the data about the district of Madrid, Wikipedia is very useful in that sense. [6]
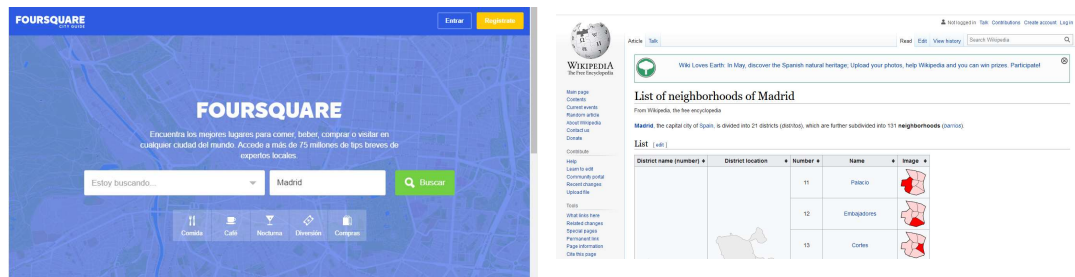
Methodology

Once it is explained what kind of data to be used, the next steps are to explain how to extract and to clean out the data with the unnecessary information utilizing the library in Python panda data frame. The results can be seen below.

|	Id	Neighborhood	District Location
0	1.0	Centro	Palacio (11)Embajadores (12)Cortes (13)Justici...
1	2.0	Arganzuela	Imperial (21)Acacias (22)Chopera (23)Legazpi (...
2	3.0	Retiro	Pacífico (31)Adelfas (32)Estrella (33)Ibiza (3...
3	4.0	Salamanca	Recoletos (41)Goya (42)Fuente del Berro (43)Gu...
4	5.0	Chamartín	El Viso (51)Prosperidad (52)Ciudad Jardín (53)...
5	6.0	Tetuán	Bellas Vistas (61)Cuatro Caminos (62)Castillej...
6	7.0	Chamberí	Gaztambide (71)Arapiles (72)Trafalgar (73)Alma...
7	8.0	Fuencarral-El Pardo	El Pardo (81)Fuentelarreina (82)Peñagrande (83...
8	9.0	Moncloa-Aravaca	Casa de Campo (91)Argüelles (92)Ciudad Univers...

To add geospatial data to the data frame, that is the latitude and the longitude, installing conda-forge geopy package to take geopy functions.

	Id	Neighborhood	District Location	Latitude	Longitude
0	1.0	Centro	Palacio (11)Embajadores (12)Cortes (13)Justici...	47.549025	1.732406
1	2.0	Arganzuela	Imperial (21)Acacias (22)Chopera (23)Legazpi (...	40.396954	-3.697289
2	3.0	Retiro	Pacífico (31)Adelfas (32)Estrella (33)Ibiza (3...	6.061715	-75.510642
3	4.0	Salamanca	Recoletos (41)Goya (42)Fuente del Berro (43)Gu...	40.965157	-5.664018
4	5.0	Chamartín	El Viso (51)Prosperidad (52)Ciudad Jardín (53)...	40.701869	-4.957008

---

[5] https://es.foursquare.com/city-guide

[6] https://en.wikipedia.org/wiki/List_of_neighborhoods_of_Madrid

Once the data frame is created, the folium package is added to the code to draw a map and to show the 21 districts of Madrid.
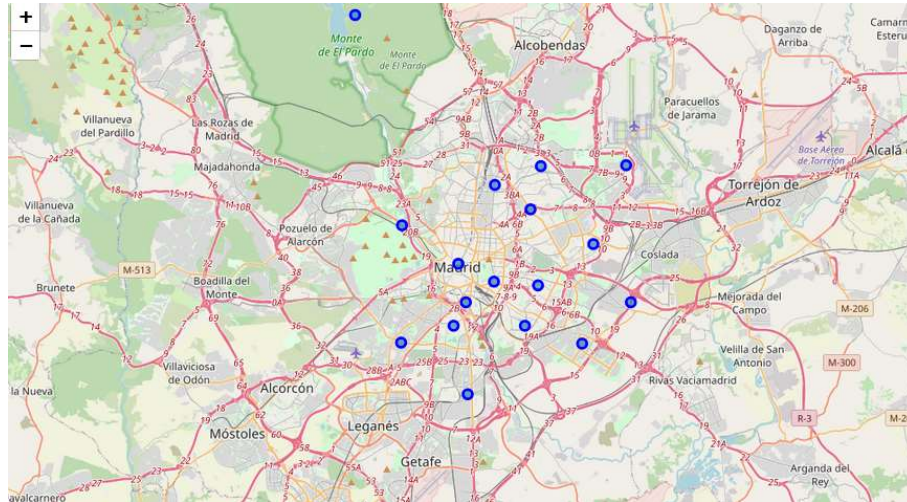

Now it is time for foursquare data. Firstly, a trial to find the top 15 venues in Madrid Center within a radius of 600 mts to check if foursquare works

```
6]: {'meta': {'code': 200, 'requestId': '60ae6bc6392465436085c5fd'},
     'response': {'suggestedFilters': {'header': 'Tap to show:',
       'filters': [{'name': 'Open now', 'key': 'openNow'}]},
      'headerLocation': 'Madrid',
      'headerFullLocation': 'Madrid',
      'headerLocationGranularity': 'city',
      'totalResults': 241,
      'suggestedBounds': {'ne': {'lat': 40.472310054000054,
        'lng': -3.631953961005949},
       'sw': {'lat': 40.36430994599994, 'lng': -3.773546038994051}},
      'groups': [{'type': 'Recommended Places',
        'name': 'recommended',
        'items': [{'reasons': {'count': 0,
           'items': [{'summary': 'This spot is popular',
             'type': 'general',
             'reasonName': 'globalInteractionReason'}]},
          'venue': {'id': '4adcda37f964a5201f3c21e3',
           'name': 'Puerta del Sol',
           'location': {'address': 'Pl. Puerta del Sol',
            'lat': 40.41679436967101,
            'lng': -3.703562021255493,
            'labeledLatLngs': [{'label': 'display',
              'lat': 40.41679436967101,
              'lng': -3.703562021255493}],
            'distance': 182,
            'postalCode': '28013',
            'cc': 'ES',
```

Secondly, to explore the neighborhoods in Madrid, getNearbyVenues function is used to retrieve foursquare data for all venues with a distance of less than 3000 meters from each city district center as shown with blue dots in the map above. It is obtained a list of 1541 venues of Madrid city where there are 475 restaurants with 43 unique categories such as Spanish, Italian, Mediterranean and so on.

To be more clear the barplot function, a bar chart with 10 most frequently restaurants in the whole city using seaborn and matplotlib packages. As it is shown below, Spanish restaurants including tapas restaurants which is as a kind of Spanish healthy fast food are the most frequently occurring restaurants in Madrid which is a pretty understandable taking into account Spaniards love eating their own gastronomy and even more the tourist who visits to Spain. In a second level, Italian and Mediterranean restaurants are predominant in Madrid City.

10 Most Frequently Occuring Venues in 9 City Districts of Madrid

<Figure size 1296x504 with 0 Axes>
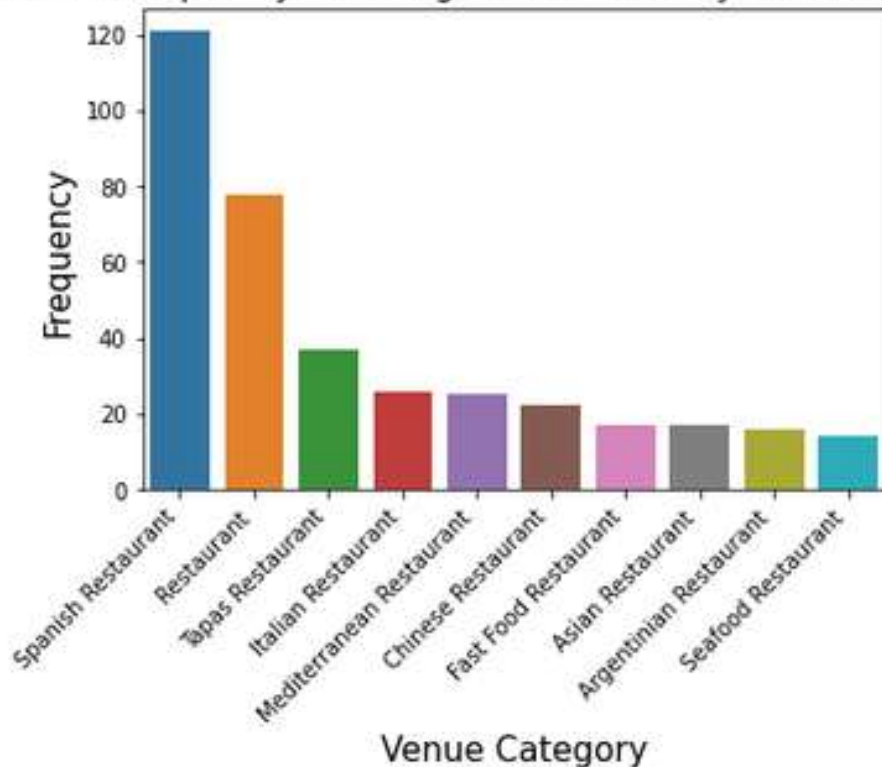
As the target is to find clusters of types of restaurant in different districts of Madrid, it is necessary to transform the data frame with the restaurant venues associated to city districts by on-hot encoding (0/1) as it can be seen below.

	Neighborhood	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	Dumpling Restaurant	Eastern European Restaurant	Falafel Restaurant	Fas Rest
1	Centro	0	0	0	0	0	0	0	0	0	0	0	
2	Centro	0	0	0	0	0	0	0	0	0	0	0	
3	Centro	0	0	0	0	0	0	0	0	0	0	0	
4	Centro	0	0	0	0	0	0	0	0	0	0	0	
5	Centro	0	0	0	0	0	0	0	0	0	0	0	

Now, neighborhood grouping by the mean of each restaurant category frequency in each city district.

	Neighborhood	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Brazilian Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	Dumpling Restaurant	Eastern European Restaurant	Falafel Restaurant
0	Arganzuela	0.000000	0.00	0.095238	0.000000	0.000000	0.047619	0.000000	0.047619	0.000000	0.000000	0.000000
1	Barajas	0.000000	0.00	0.074074	0.000000	0.000000	0.074074	0.000000	0.000000	0.000000	0.000000	0.000000
2	Carabanchel	0.000000	0.00	0.000000	0.071429	0.035714	0.035714	0.000000	0.000000	0.000000	0.000000	0.000000
3	Centro	0.000000	0.00	0.043478	0.000000	0.000000	0.043478	0.000000	0.000000	0.000000	0.000000	0.000000
4	Chamartín	0.000000	0.00	0.017857	0.035714	0.017857	0.000000	0.017857	0.000000	0.000000	0.017857	0.000000
5	Chamberí	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Ciudad Lineal	0.000000	0.00	0.041667	0.062500	0.000000	0.062500	0.000000	0.000000	0.000000	0.020833	0.000000
7	Fuencarral-El Pardo	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	Hortaleza	0.022727	0.00	0.045455	0.068182	0.000000	0.090909	0.000000	0.000000	0.000000	0.000000	0.000000
9	Moncloa-Aravaca	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.142857	0.000000	0.000000
10	Moratalaz	0.000000	0.00	0.000000	0.028571	0.000000	0.028571	0.000000	0.000000	0.000000	0.028571	0.000000
11	Puente de Vallecas	0.000000	0.00	0.035714	0.000000	0.000000	0.035714	0.000000	0.000000	0.000000	0.035714	0.000000
12	Retiro	0.000000	0.00	0.055556	0.000000	0.000000	0.027778	0.000000	0.027778	0.000000	0.000000	0.000000
13	Salamanca	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
14	San Blas-Canillejas	0.000000	0.00	0.000000	0.115385	0.000000	0.115385	0.000000	0.000000	0.000000	0.000000	0.000000

With the above information, it is already possible to create a data frame to see the most common restaurant venue types for each city district.

Results

With the data frame with the most common restaurant venues categories for each city district, an unsupervised machine learning algorithm could be run. In this case, a k-means clustering algorithm from the scikit-learn package with the elbow method to systematically define k value for this paper of 5.

Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
1	Arganzuela	Restaurant	Spanish Restaurant	Italian Restaurant	Argentinian Restaurant	Tapas Restaurant	Mediterranean Restaurant	Thai Restaurant	Sushi Restaurant
2	Barajas	Spanish Restaurant	Restaurant	Japanese Restaurant	Argentinian Restaurant	Tapas Restaurant	Chinese Restaurant	Himalayan Restaurant	Mexican Restaurant
1	Carabanchel	Spanish Restaurant	Italian Restaurant	Tapas Restaurant	Restaurant	Seafood Restaurant	Asian Restaurant	Paella Restaurant	Mediterranean Restaurant
1	Centro	Restaurant	Tapas Restaurant	Sushi Restaurant	Mediterranean Restaurant	Argentinian Restaurant	Spanish Restaurant	Seafood Restaurant	Chinese Restaurant
1	Chamartín	Spanish Restaurant	Restaurant	Japanese Restaurant	Thai Restaurant	Tapas Restaurant	Mediterranean Restaurant	Seafood Restaurant	Asian Restaurant
3	Chamberí	French Restaurant	Fast Food Restaurant	Venezuelan Restaurant	Falafel Restaurant	Japanese Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant
1	Ciudad Lineal	Spanish Restaurant	Restaurant	Tapas Restaurant	Chinese Restaurant	Indian Restaurant	Asian Restaurant	Thai Restaurant	Sushi Restaurant
4	Fuencarral-El Pardo	Spanish Restaurant	Venezuelan Restaurant	Falafel Restaurant	Japanese Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant	Greek Restaurant

As it can be seen in the table, the city districts, their most common venues and the assigned 5 different cluster labels from 0 to 4. Now using the folium function, it is possible to show the city districts marked by a cluster – specific color on a map.
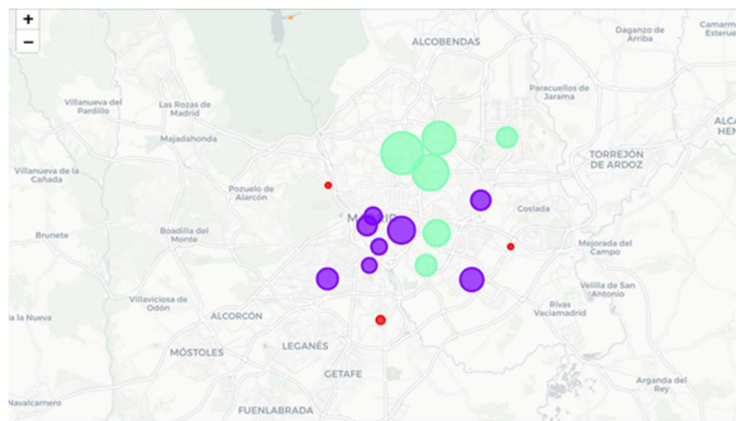
Finally, the five clusters of restaurant categoris for Madrid City can be seen below.

Cluster 1

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
Moncloa-Aravaca	0	Spanish Restaurant	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Mediterranean Restaurant	Dumpling Restaurant	Venezuelan Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant
Villaverde	0	Spanish Restaurant	Fast Food Restaurant	Latin American Restaurant	Chinese Restaurant	Mediterranean Restaurant	Restaurant	Brazilian Restaurant	French Restaurant	Italian Restaurant
Vicálvaro	0	Spanish Restaurant	Restaurant	Mediterranean Restaurant	Fast Food Restaurant	Falafel Restaurant	Eastern European Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant

Cluster 2

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
Centro	1	Restaurant	Tapas Restaurant	Sushi Restaurant	Mediterranean Restaurant	Argentinian Restaurant	Spanish Restaurant	Seafood Restaurant	Chinese Restaurant	Italian Restaurant
Arganzuela	1	Restaurant	Spanish Restaurant	Italian Restaurant	Argentinian Restaurant	Tapas Restaurant	Mediterranean Restaurant	Thai Restaurant	Sushi Restaurant	Cuban Restaurant
Retiro	1	Restaurant	Spanish Restaurant	Mediterranean Restaurant	Tapas Restaurant	Argentinian Restaurant	Cuban Restaurant	Chinese Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant
Salamanca	1	Tapas Restaurant	Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Spanish Restaurant	Japanese Restaurant	Paella Restaurant	Falafel Restaurant	Indian Restaurant
Latina	1	Restaurant	Tapas Restaurant	Spanish Restaurant	Argentinian Restaurant	Mediterranean Restaurant	Sushi Restaurant	Seafood Restaurant	Paella Restaurant	Chinese Restaurant
Carabanchel	1	Spanish Restaurant	Italian Restaurant	Tapas Restaurant	Restaurant	Seafood Restaurant	Asian Restaurant	Paella Restaurant	Mediterranean Restaurant	Latin American Restaurant
Usera	1	Argentinian Restaurant	Spanish Restaurant	Tapas Restaurant	Italian Restaurant	Chinese Restaurant	Seafood Restaurant	Cuban Restaurant	Arepa Restaurant	Theme Restaurant
Villa de Vallecas	1	Restaurant	Fast Food Restaurant	Spanish Restaurant	Asian Restaurant	Tapas Restaurant	Chinese Restaurant	Mediterranean Restaurant	Mexican Restaurant	Falafel Restaurant
San Blas-Canillejas	1	Restaurant	Spanish Restaurant	Asian Restaurant	Chinese Restaurant	Fast Food Restaurant	Italian Restaurant	Mediterranean Restaurant	Tapas Restaurant	Sushi Restaurant

Cluster 3

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Chamberí	2	French Restaurant	Fast Food Restaurant	Venezuelan Restaurant	Falafel Restaurant	Japanese Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant	Greek Restaurant	German Restaurant

Cluster 4

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
Chamartín	3	Spanish Restaurant	Restaurant	Japanese Restaurant	Thai Restaurant	Tapas Restaurant	Mediterranean Restaurant	Seafood Restaurant	Asian Restaurant	Comfort Food Restaurant
Tetuán	3	Spanish Restaurant	Italian Restaurant	Moroccan Restaurant	Venezuelan Restaurant	Falafel Restaurant	Indian Restaurant	Himalayan Restaurant	Greek Restaurant	German Restaurant
Puente de Vallecas	3	Spanish Restaurant	Restaurant	Tapas Restaurant	Italian Restaurant	Mediterranean Restaurant	Eastern European Restaurant	Argentinian Restaurant	Korean Restaurant	Greek Restaurant
Moratalaz	3	Spanish Restaurant	Seafood Restaurant	Tapas Restaurant	Restaurant	Mediterranean Restaurant	Mexican Restaurant	Italian Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant
Ciudad Lineal	3	Spanish Restaurant	Restaurant	Tapas Restaurant	Chinese Restaurant	Indian Restaurant	Asian Restaurant	Thai Restaurant	Sushi Restaurant	Seafood Restaurant
Hortaleza	3	Spanish Restaurant	Restaurant	Chinese Restaurant	Asian Restaurant	Indian Restaurant	Argentinian Restaurant	Tapas Restaurant	Italian Restaurant	Mediterranean Restaurant
Barajas	3	Spanish Restaurant	Restaurant	Japanese Restaurant	Argentinian Restaurant	Tapas Restaurant	Chinese Restaurant	Himalayan Restaurant	Mexican Restaurant	Fast Food Restaurant

Cluster 5

Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Fuencarral-El Pardo	4	Spanish Restaurant	Venezuelan Restaurant	Falafel Restaurant	Japanese Restaurant	Italian Restaurant	Indian Restaurant	Himalayan Restaurant	Greek Restaurant	German Restaurant	French Restaurant

Discussion

This research about the gastronomy in Madrid, firstly, started taking data about restaurant in Madrid like name, ID, location and category food from the website of Foursquare and about the districts in Madrid. To extract and clean out those data with the unnecessary information, panda data frame library was utilized. Moreover, geospatial data was added to the data frame as latitude and longitude installing conda-forge geopy package to show a map of Madrid with the districts using folium package to show.

Secondly, the getNearbyVenues function to explore the neighborhoods in Madrid and to retrieve foursquare data for all venues and a list of Madrid restaurant with unique categories. To be more clear, a barplot function with bar chart with the 10 most frequently restaurants in the whole city using seaborn and matplotlib packages.

Thirdly, it was necessary to transform the data frame with restaurant venues associated to city districts by on-hot encoding to find clusters of types of restaurant in different districts of Madrid. After this, neighborhood grouping by the mean of each restaurant category frequency in each district to create a data frame in which it is possible to see the most common restaurant venue types for each city district.

Finally, an unsupervised machine learning algorithm could be running, in this case, with a k-means clustering algorithm from the scikit-learn package with the elbow method to systematically define k value for this paper of 5.

Conclusion

The target of this paper was to discover a selection of the gastronomy in Madrid for each district of the city using foursquare data. Spanish restaurants including tapas restaurants which is as a kind of Spanish healthy fast food are the most frequently occurring restaurants in Madrid which is a pretty understandable taking into account Spaniards love eating their own gastronomy and even more the tourist who visits Spain. In a second level, Italian and Mediterranean restaurants are predominant in Madrid City. Another important point is the average income for each district, the higher average income is more possible to find international cuisine restaurants, the less average income is more probably to see in those districts Spanish restaurants specially tapas restaurants and fast food restaurants too.

Acknowledgment & sources

- https://www.coursera.org/professional-certificates/ibm-data-science
- https://github.com/
- https://stackoverflow.com/
- https://simpleanalytical.com/
- https://es.foursquare.com/city-guide
- https://en.wikipedia.org/wiki/List_of_neighborhoods_of_Madrid
- *Spain sets new foreign tourist arrival record for sixth consecutive year"*. 4 February 2019. Retrieved 9 February 2019.

- [Archived](#) 6 January 2011 at the [Wayback Machine](#)
- https://thesavvybackpacker.com/city-guide/madrid-travel-guide/
- https://www.turismomadrid.es/en/to-see-and-do/gastronomy/10175-gastronomy.html-