Battle of the Neighborhoods

In Mexico City

Carlos Villegas - June 2020

Introduction

In México the biggest city is Mexico City, and therefore has the largest number or service related locations.

Being a person with the intent of placing a service related establishment, you need to find a type of establishment that doesn’t have the same specialty of the most common in a certain Neighborhood.

Problem

- The objetive of this report is to find the highest profile Borough in Mexico City based in HDI (Human Development Index) that is a statistic composite of life expectancy, education and per capita income, among other indicators.

- Once that we profiled the highest borough, find the neighborhoods inside that borough and locate the food related locations on each neighborhood and find a correlation between the food establishments and obtain the top 10 on each cluster.

Data Acquisition

	ENTIDAD	Geo Point	Geo Shape
E CHAPULTEPEC	9	19.4228411174,-99.2157935754	{"type": "Polygon", "coordin
E REFORMA (LOMAS DE CHAPULTEPEC)	9	19.4106158914,-99.2262487268	{"type": "Polygon", "coordin
UE (POLANCO)	9	19.4342189235,-99.2094037513	{"type": "Polygon", "coordin
L DE SANTA URSULA I	9	19.314862237,-99.1477954505	{"type": "Polygon", "coordin
	9	19.324571116,-99.1561602234	{"type": "Polygon", "coordin
L MAUREL (U HAB)	9	19.3053003722,-99.1734305912	{"type": "Polygon", "coordin
ZARAGOZA I	9	19.4131119713,-99.0946992338	{"type": "Polygon", "coordin
	9	19.4257144309,-99.1226597674	{"type": "Polygon", "coordin
GOMEZ FARIAS	9	19.4161844072,-99.0935753088	{"type": "Polygon", "coordin
II	9	19.4415560237,-99.1183549378	{"type": "Polygon", "coordin
BRAVO	9	19.4538981425,-99.1231533534	{"type": "Polygon", "coordin
O	9	19.4484587894,-99.1999095691	{"type": "Polygon", "coordin
O ELIAS CALLES	9	19.4548396919,-99.1691894629	{"type": "Polygon", "coordin
RE COYOACAN (FRACC)	9	19.3088228782,-99.1170550984	{"type": "Polygon", "coordin
	9	19.4475072464,-99.1174322362	{"type": "Polygon", "coordin
GELES	9	19.4526551282,-99.1122459477	{"type": "Polygon", "coordin
	9	19.4384228454,-99.1079557082	{"type": "Polygon", "coordin
AFICAS	9	19.4113463978,-99.1258705526	{"type": "Polygon", "coordin
RBOLES	9	19.4164712493,-99.0844284558	{"type": "Polygon", "coordin
LHUACAN (U HAB)	9	19.307966041,-99.1102887862	{"type": "Polygon", "coordin
ZAPATA	9	19.326716273,-99.1380465169	{"type": "Polygon", "coordin
CISCO CULHUACAN (PBLO)	9	19.3355619401,-99.1182479778	{"type": "Polygon", "coordin
AN ANDRES	9	19.3457136742,-99.1483519404	{"type": "Polygon", "coordin
	9	19.3042837028,-99.1700844887	{"type": "Polygon", "coordin
SCO (U HAB)	9	19.5094365342,-99.0960012649	{"type": "Polygon", "coordin
LLO 2 (U HAB)	9	19.5341864173,-99.1443342522	{"type": "Polygon", "coordin
GON (U)	9	19.4526847398,-99.0560709204	{"type": "Polygon", "coordin

Data Processing

- Import CSV
- Import from JSON
- Convert to Pandas Dataframes
- Clean data
- One Hot Encoding
- Clean for KMeans Alrorithm

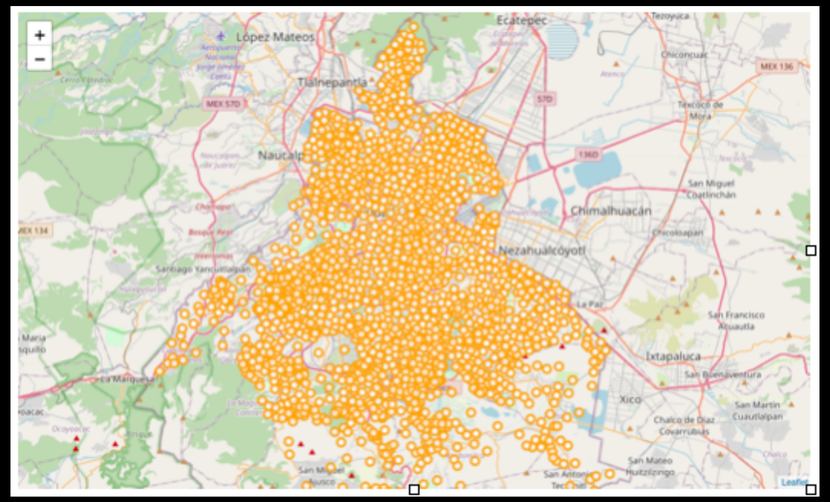
```
# Read CVS
df_colonias = pd.read_csv("2_colonias_cdmx_csv.csv")
df_colonias.head()
```

	COLONIA	ENTIDAD	Geo Point
0	LOMAS DE CHAPULTEPEC	9.0	19.4228411174,-99.215793575
1	LOMAS DE REFORMA (LOMAS DE CHAPULTEPEC)	9.0	19.4106158914,-99.226248726
2	DEL BOSQUE (POLANCO)	9.0	19.4342189235,-99.209403751
3	PEDREGAL DE SANTA URSULA I	9.0	19.314862237,-99.1477954505
4	AJUSCO I	9.0	19.324571116,-99.1561602234

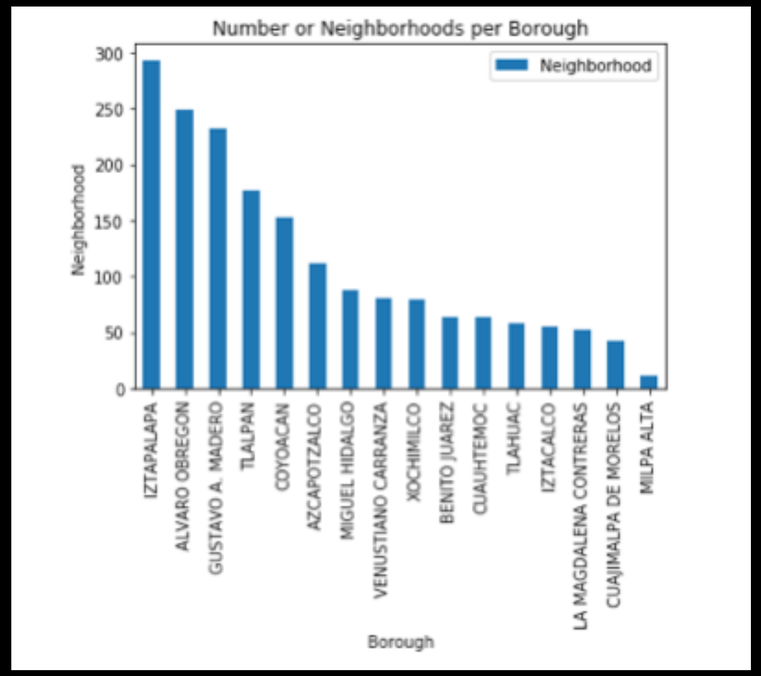
Methodology

Determine the Borough

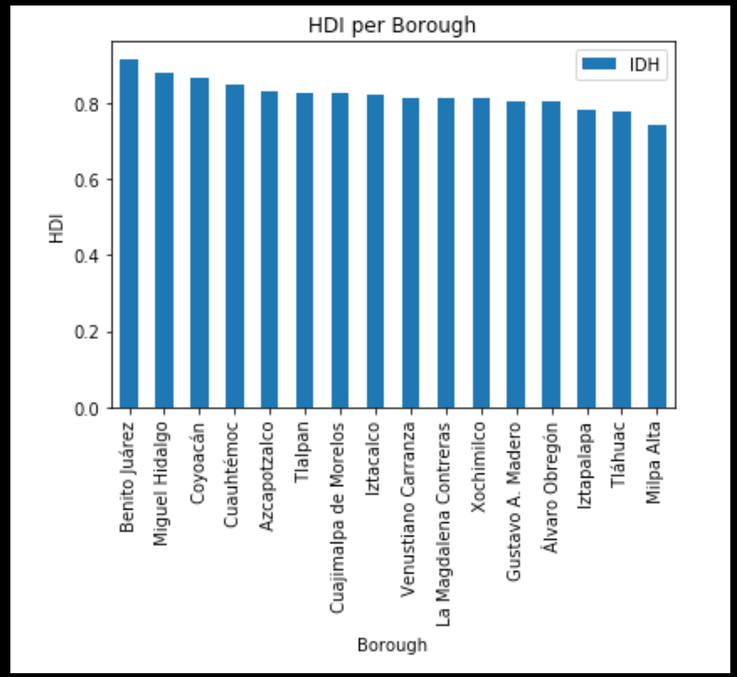
The amount of neighborhoods in Mexico City is considerable



For each borough there are many neighborhoods, being the distribution as follows



Determine the Borough with the highest HDI in Mexico City. The selection for this study is Benito Juarez.

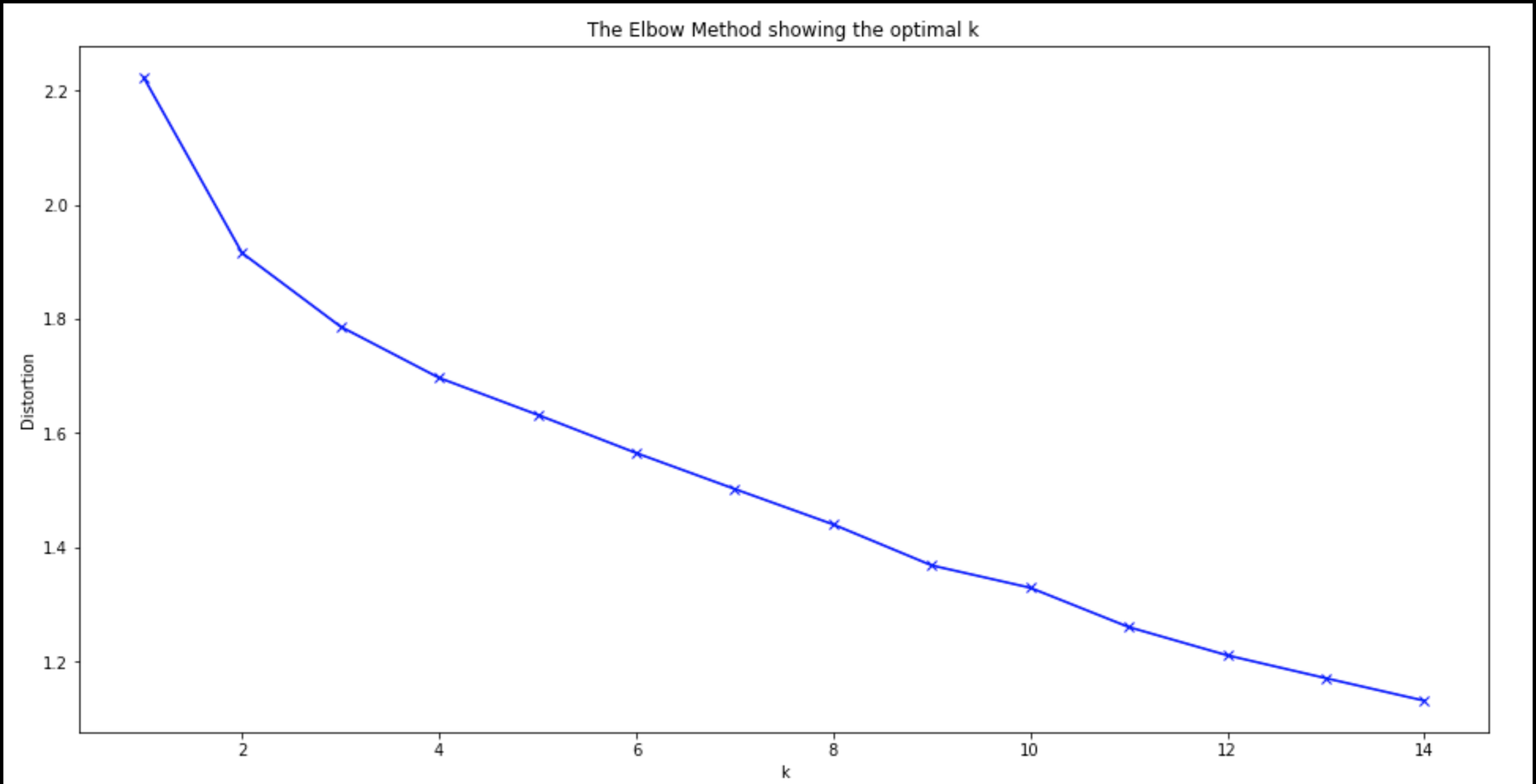


For this repot I decided to use Clustering because it is an unsupervised machine learning scenario, and using the “K-means” method in order to find relationships between the many venues of the borough.

> KMeans definition: k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster.
>
> - Wikipedia

Using the optimal K to find the number of clusters

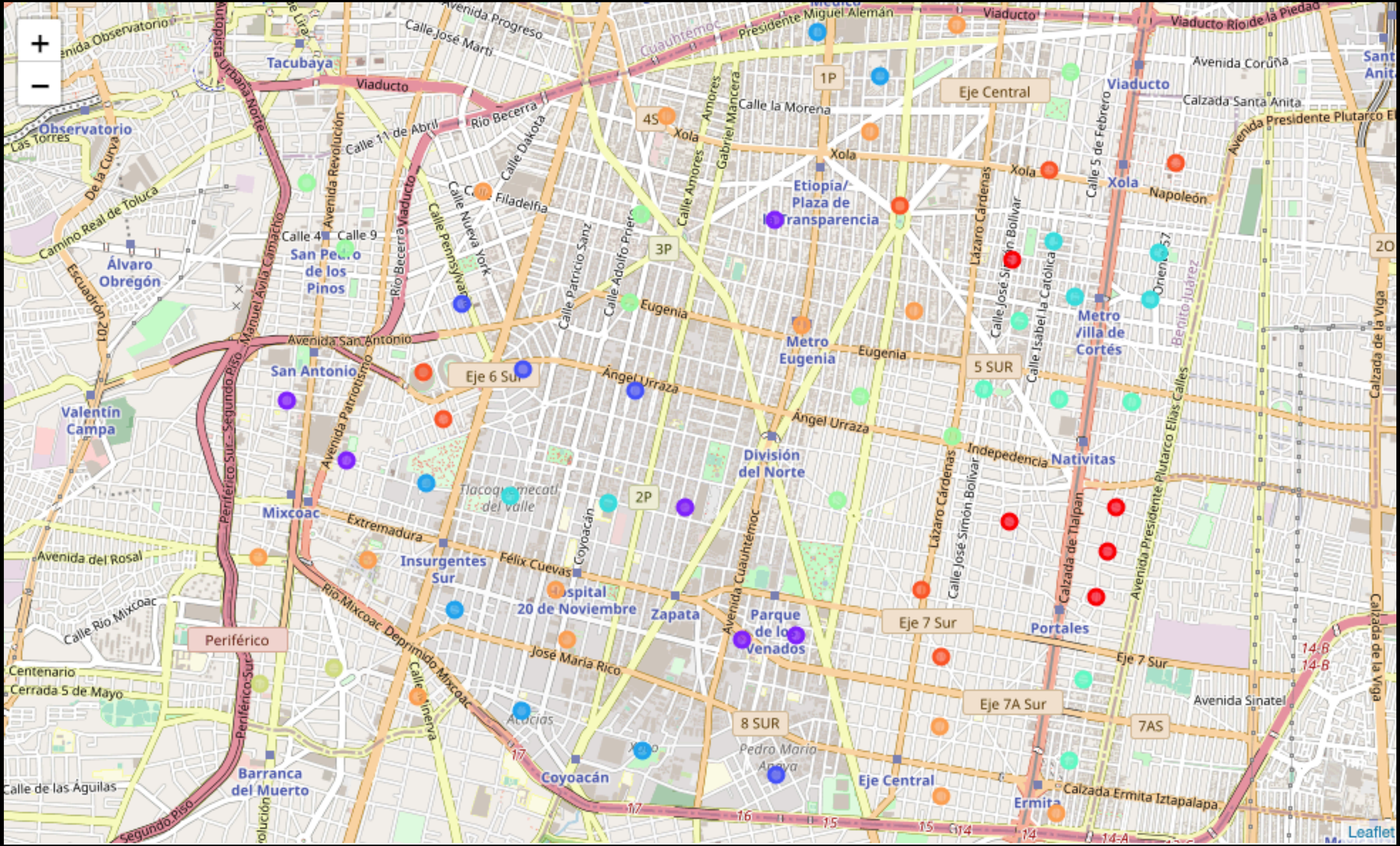
K = 10



Explanation

- With the data found all the boroughs in Mexico City and then all the neighborhoods.
- Based on HDI, determine the best suitable borough to analyze.
- Retrieve information from Foursquare regarding service venues for each neighborhood in the borough.
- Transform the information retrieved into a DataFrame that can be manipulated trough one hot encoding, merge and join with the neighborhood information.
- The previous step was to prepare the data in order to apply K-means algorithm in order to find patterns in the venues on each neighborhood.

Cluster Map


+
−
Tacubaya
Viaducto
Observatorio
Álvaro
Obregón
San Pedro
de los
Pinos
San Antonio
Valentín
Campa
Mixcoac
Insurgentes
Sur
Periférico
Barranca
del Muerto
Coyoacán
Zapata
Parque
de los
Venados
División
del Norte
Metro
Eugenia
Etiopía/
Plaza de
la Transparencia
Xola
Metro
Villa de
Cortés
Nativitas
Portales
Ermita
Eje Central
Eje 6 Sur
Eje 7 Sur
Eje 7A Sur
5 SUR
8 SUR
7AS
1P
2P
3P
4S
Eugenia
Ángel Urraza
Independencia
Extremadura
Félix Cuevas
Hospital
20 de Noviembre
José María Rico
Avenida San Antonio
Calle 11 de Abril
Río Becerra
Filadelfia
Calle la Morena
Napoleón
Avenida Coruña
Calzada Santa Anita
Viaducto Río de la Piedad
Presidente Miguel Alemán
Avenida del Rosal
Calle Río Mixcoac
Centenario
Cerrada 5 de Mayo
Calle de las Águilas
Avenida Sinatel
Calzada Ermita Iztapalapa
Calzada de Tlalpan
Calzada de la Viga
Avenida Presidente Plutarco Elías Calles
Tlacoquemecatl
del Valle
Pedro María
Anaya
Leaflet

Example of Cluster

	COLONIA	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
4	LETRAN VALLE	Taco Place	Bakery	Mexican Restaurant	Ice Cream Shop	Pizza Place	Pet Store	IT Services	Liquor Store	Dessert Shop	Park
10	INDEPENDENCIA	Mexican Restaurant	Coffee Shop	Food Truck	Bakery	Steakhouse	Taco Place	Dessert Shop	Dance Studio	Pool	Farmers Market
32	SAN PEDRO DE LOS PINOS	Mexican Restaurant	Bakery	Ice Cream Shop	Park	Food Truck	Yoga Studio	Café	Bistro	Pharmacy	Burger Joint
34	DEL VALLE II	Mexican Restaurant	Health & Beauty Service	Bakery	Pet Store	Ice Cream Shop	Deli / Bodega	Park	Sushi Restaurant	Taco Place	Tapas Restaurant
45	DEL VALLE III	Mexican Restaurant	Dance Studio	Coffee Shop	Spa	Taco Place	Bakery	Liquor Store	Drugstore	Park	Camera Store
50	OCHO DE AGOSTO	Mexican Restaurant	Bakery	Taco Place	Food Truck	Performing Arts Venue	Hotel	Burger Joint	Motorcycle Shop	Seafood Restaurant	Bistro
58	VERTIZ NARVARTE	Mexican Restaurant	Bakery	Taco Place	Flower Shop	Steakhouse	Russian Restaurant	Coffee Shop	Spa	Southern / Soul Food Restaurant	German Restaurant
60	ALAMOS I	Mexican Restaurant	Bakery	Coffee Shop	Taco Place	Gym / Fitness Center	Pie Shop	Market	Steakhouse	Martial Arts Dojo	Food Truck

Discussion

Given the problem, I found that there is an over saturation of venues in the Borough of Benito Juarez in México City.

This over saturation made difficult to establish a pattern where you could place a new business inside the best borough in terms of HDI.

The k-indicator didn’t have a clear elbow point because of this.

I think that a market segmentation data base will provide better results into a specialization business analysis.

Conclusion

- By far the most common venue is a Mexican Restaurant
- There is demand for food services in Benito Juarez
- If a business most be opened I would recommend the following:
- Mexican Restaurant in Cluster 1 or Cluster 2
- Do not place anything in Cluster 9
- Taco Place in Cluster 10
- In all other clusters a Coffee Shop or Ice Cream Shop is a good option.