Housing Sales Prices & Venues Data Analysis of Mexico City

Jesús Alfonso Juárez

February 2020

1.- Introduction

1a.-Description and discussion of the background

Greater Mexico City is the second largest metropolitan area of the western hemisphere and the largest spanish-speaking city in the world with **21.3** million of population. Mexico City has by itself **9 million** people gathered in just **1,485 square kilometers** turning into a high-density zone with **6,000 persons** by square kilometer. Mexico has the history embedded in their walls, originally named **Mexico Tenochitlan** by the aztecs has been witness of many stages from the pre-Hispanic to the modern era. Currently, the city is formed by **16 boroughs.[1]**

Mexico City is considered a megacity which means that is a high population density zone. Thus, there is a restricted supply of commercial and residential real estate. Moreover, the tendency to the vertical urbanization and the new structures of families demand a new approach in the housing sector. The city residents are seeking zones near to their jobs, with the venues that they attend and where the real estate values are lower, and Furthermore, investors are seeking to establish business in the neighborhoods with lower cost and less competition in the district.

Nowadays does not exist a tool that lead investors and city residents to make a data-based decision of the neighborhood to select. Consequently, we can create a map and information chart where the real estate index is placed on Mexico City and each district is clustered according to the venue density.

1b.-Data Description

To solve the problem, we can list the data as below:

- I found the Neighborhoods Coordinates of Mexico City in the Data Repository of the Mexico City government website [2]. The .geojson has the coordinates of all the districts and boroughs ('Delegaciones') of Mexico City
- I used Forsquare API to get the most common venues of given Borough ('Delegación') of Mexico City [3].
- The real estate as other markets has a widespread range of prices in similar housing, thus there is a myriad of information regarding the real estate costs. To overcome this issue, we are going to use the latest square meter Housing Sales Price (HSP)

Average for each Borough ('Delegación') of Mexico City retrieve from the real state retail web page [4].

1c.-Data Usage

The approach that we are going to take with the different sources of information is:

1.-The geojson file will be uploaded to the Jupyter notebook file.

Example in Excel of the Neighborhood's coordinates:

COLONIA	ENTIDAD	Geo Point	Geo Shape	CVE_ALC	ALCALDIA	CVE_COL	SECC_COM	SECC_PAR
IRRIGACION	9	19.442954925	{"type": "Poly	16	MIGUEL HIDALGO	16-035	5079, 5080, 5083, 5102	5068, 5082
MARINA NACIONAL (U HAB)	9	19.446631905	{"type": "Poly	16	MIGUEL HIDALGO	16-049	5137, 5182	
PEDREGAL DE STO DOMINGO VI	9	19.323402718	{"type": "Poly	3	COYOACAN	03-144	381, 394, 494, 416, 417, 439	
VILLA PANAMERICANA 7MA. SECCI	9	19.304604265	{"type": "Poly	3	COYOACAN	03-121		474, 475
VILLA PANAMERICANA 6TA. SECCIN	9	19.311223887	{"type": "Poly	3	COYOACAN	03-120		458
SANTA CRUZ AVIACION	9	19.422303952	{"type": "Poly	17	VENUSTIANO CARRANZA	17-056	5335	5325
MAGDALENA MIXHUCA	9	19.407130147	{"type": "Poly	17	VENUSTIANO CARRANZA	17-036	5384, 5386, 5387, 5396	
COPILCO UNIVERSIDAD	9	19.336117251	{"type": "Poly	3	COYOACAN	03-024	741, 740	359, 731
MORALES SECCION ALAMEDA (POL	9	19.433717401	{"type": "Poly	16	MIGUEL HIDALGO	16-054		4918
AJUSCO HUAYAMILPAS	9	19.323967921	{"type": "Poly	3	COYOACAN	03-003	403	373, 374
CHIMALISTAC	9	19.343796855	{"type": "Poly	3	COYOACAN	03-019		693, 727
EL ARENAL PTO AEREO (FRACC)	9	19.428691865	{"type": "Poly	17	VENUSTIANO CARRANZA	17-022	5432, 5433, 5434	
BAHIA (U HAB)	9	19.416474695	{"type": "Poly	17	VENUSTIANO CARRANZA	17-010		5462, 5463
CUADRANTE DE SAN FRANCISCO	9	19.342601525	{"type": "Poly	3	COYOACAN	03-035	720, 721, 745	
EL RELOJ	9	19.318273024	{"type": "Poly	3	COYOACAN	03-045	427	398, 399, 426, 428
EL MIRADOR	9	19.302360178	{"type": "Poly	3	COYOACAN	03-043		672, 671
LOS CIPRESES	9	19.31922947,	{"type": "Poly	3	COYOACAN	03-072	606	620
LOS CEDROS (FRACC)	9	19.313805224	{"type": "Poly	3	COYOACAN	03-071		688
CONSTITUCION DE LA REPUBLICA	9	19.483401334	{"type": "Poly	5	GUSTAVO A. MADERO	05-029	1548, 1549, 1550, 1551, 1552,	1553, 1554, 1555, 1556
CASTILLO GRANDE (AMPL)	9	19.537025598	{"type": "Poly	5	GUSTAVO A. MADERO	05-022		910, 912
7 DE NOVIEMBRE	9	19.460487752	{"type": "Poly	5	GUSTAVO A. MADERO	05-212	1139, 1140, 1212	1141, 1197, 1199
INDUSTRIAL I	9	19.473773255	{"type": "Poly	5	GUSTAVO A. MADERO	05-223	1089, 1090, 1095, 1105, 1106	1125
TEPETATAL	9	19.562424437	{"type": "Poly	5	GUSTAVO A. MADERO	05-180		838, 844, 852
SAN PEDRO EL CHICO	9	19.471173415	{"type": "Poly	5	GUSTAVO A. MADERO	05-167	1565, 1567, 1568	1155, 1566
PEMEX LINDAVISTA (U HAB)	9	19.506369491	{"type": "Poly	5	GUSTAVO A. MADERO	05-136	1054	1055
SAN JUAN III (U HAB)	9	19.480733627	{"type": "Poly	5	GUSTAVO A. MADERO	05-163		1546
GUADALUPE TEPEYAC	9	19.468067766	{"type": "Poly	5	GUSTAVO A. MADERO	05-074	1117, 1118, 1119, 1120, 1121	1102, 1125
GUADALUPE VICTORIA II	9	19.471907604	{"type": "Poly	5	GUSTAVO A. MADERO	05-076	1488, 1489	1491, 1492, 1493
LA PRADERA II (U HAB)	9	19.471701537	{"type": "Poly	5	GUSTAVO A. MADERO	05-110		1443, 1442

2.-The information retrieve of Metroscubicos website will be compiled in a csv file then uploaded to the Jupyter notebook file.

Example from Borough 'Coyoacan':

	Muestra (Unidades)	Tamaño Promedio (m2c)	Promedio por m2 (pesos)		
			Promedio	Máximo	Mínimo
Casa Sola	640	408.19	20,755.53	29,454.38	12,056.68
Departamento	874	95.23	24,808.50	32,954.91	16,662.09
Terreno	9	515.40	19,142.35	25,513.46	12,771.24
Casa en Condominio	142	321.80	23,481.23	28,664.52	18,297.94

3.- The gejson file and the csv files are going to pass through a cleansing stage in order to homologue the Boroughs names, and other fields.

4.- Afterwards, we are going to set a panda's data frame with the neighborhood name, coordinates, Borough, postcode.

5.- Leveraging the foursquare API we are going to retrieve the closest venues to each

district in a radio of 700 meters.

6.- Finally, we are going to transform this last panda's data frame, establishing the venues categories as columns with the get_dummies method and grouping by the neighborhood. The resultant data frame will be our input for the k-mean cluster method.

2.- References

[1] Mexico City recover from Wikipedia February 2020

[2] Coordinates of neighborhoods in Mexico City recover from CDMX government website February 2020

[3] Foursquare API

[4] Housing square meter average sales prices of each Borough recover from Metroscubicos February 2020