



# Clustering and Comparing two neighborhoods in Mexico City to find the best place to relocate

---

GILLIAN ESCUDERO

MAY 31, 2020

# Part 1: Introduction and Business Problem

---

Before deciding on moving to a new area, families, and individuals, need to analyze a vast array of information to choose the location that is best suited to their needs. It is often the case that what prompts a move is a promotion, which forces families to investigate neighborhoods that are nearer the new office to avoid a long commute. In this scenario, finding a new place can be daunting, especially if we are not familiar with the area we are moving to. Some of the factors that could play a role in this decision are the proximity to bus stops, schools, parks, or other venues.

In this project we will explore, study, analyze, cluster, and compare the neighborhoods of these two boroughs. The goal is to provide valuable information for individuals looking to move to any of these two areas. We will compare the neighborhoods in each borough based on the quantity and category of the venues in the vicinity. By studying the neighborhoods, we will better understand what types of businesses thrive in each area, as well as finding how they are similar or how they are different.

# Part 2: Data acquisition and preparation

---

To carry out this project two types of data are needed:

1. A list of the neighborhoods in each borough, including geo location; and
2. The venues data for each neighborhood.

# Part 2: Data acquisition and preparation

---

## NEIGHBORHOOD DATA

The data for the neighborhoods for each borough can be extracted from Mexico City Data Portal (<https://datos.cdmx.gob.mx/explore/dataset/colonia-scdmx/table/>). The city has an open database that includes the list of neighborhoods by borough along with the geolocation. The data can be consumed through an API or directly downloading the data in any of the file formats they offer (csv, json, or excel). For ease of processing, the direct download in csv will be used.

## VENUES DATA

We will extract the venues data from Foursquare. The Places API offers real-time access to Foursquare's global database of rich venue data. The venue data is obtained by passing the required parameters for each neighborhood to the Places API. We will create a dataframe for each borough to contain the extracted venue data.

# Part 3: Methodology

---

## Exploratory Data Analysis

- Most Common Venue Categories
- Most Widespread Venue Categories

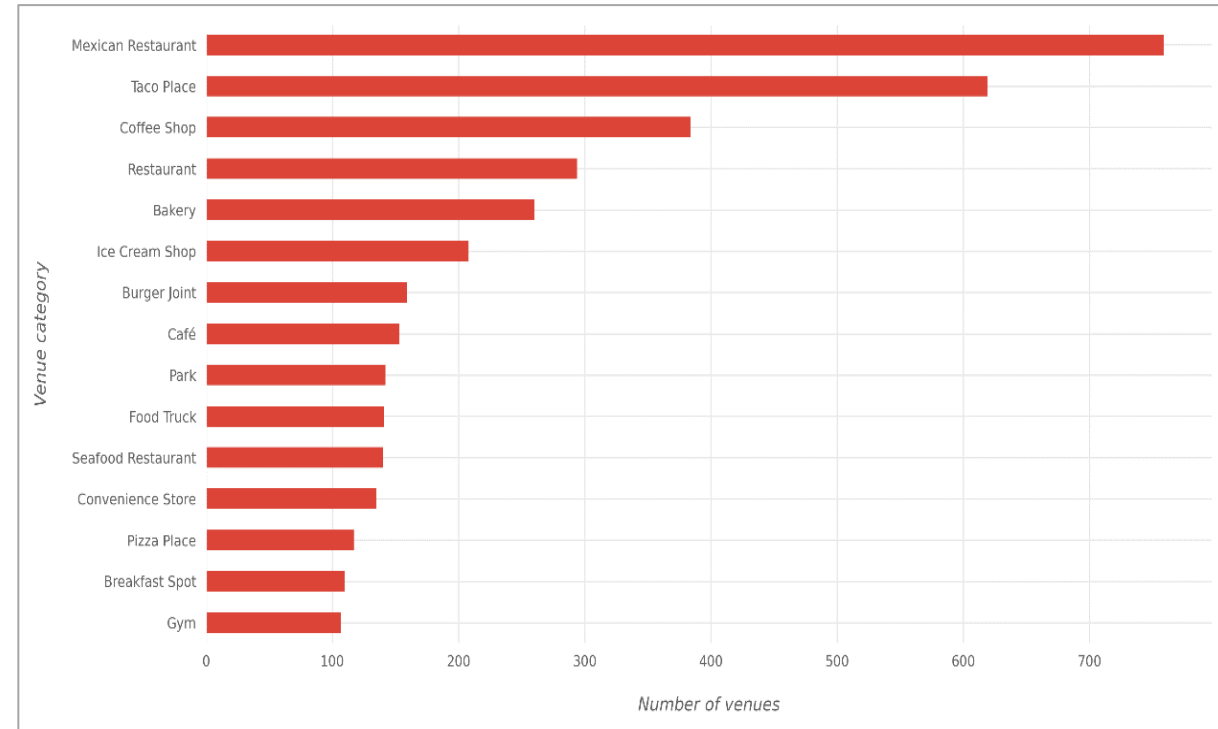
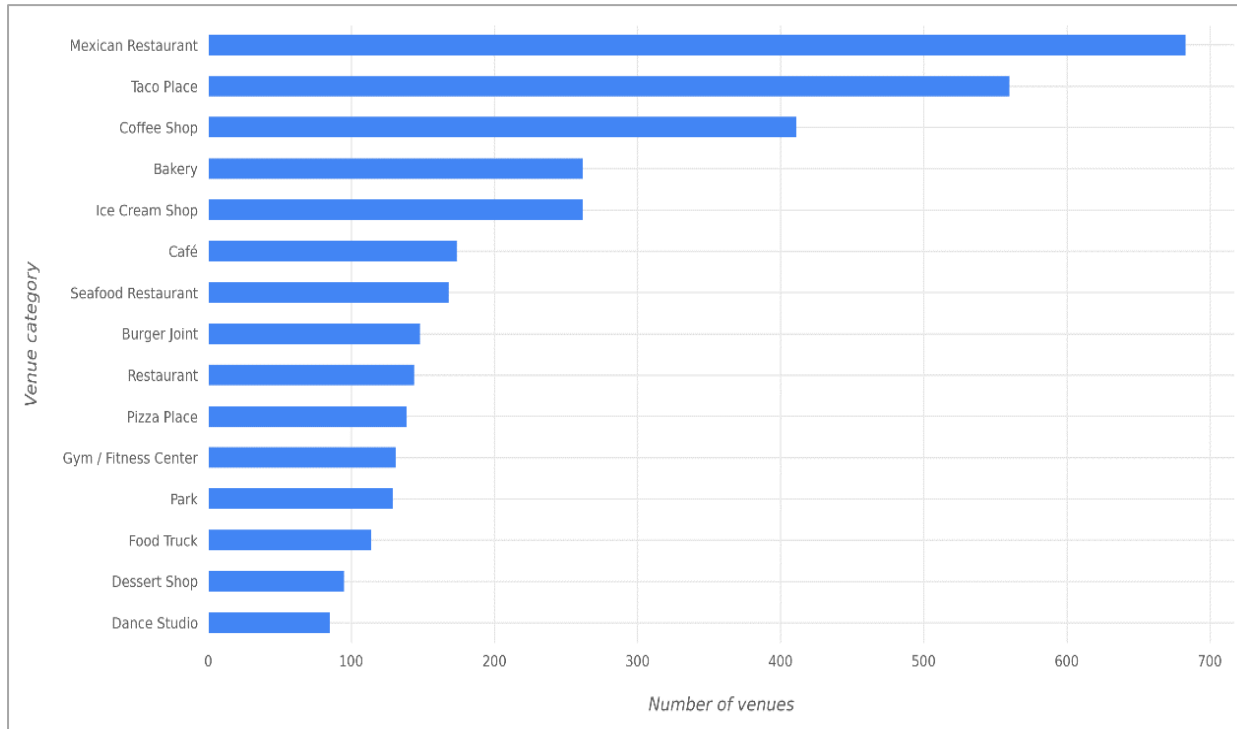
## Clustering of Neighborhoods

- One-hot encoding
- Combining the two dataframes
- The most common categories for each neighborhood
- K-Means clustering

# Most Common Venue Categories

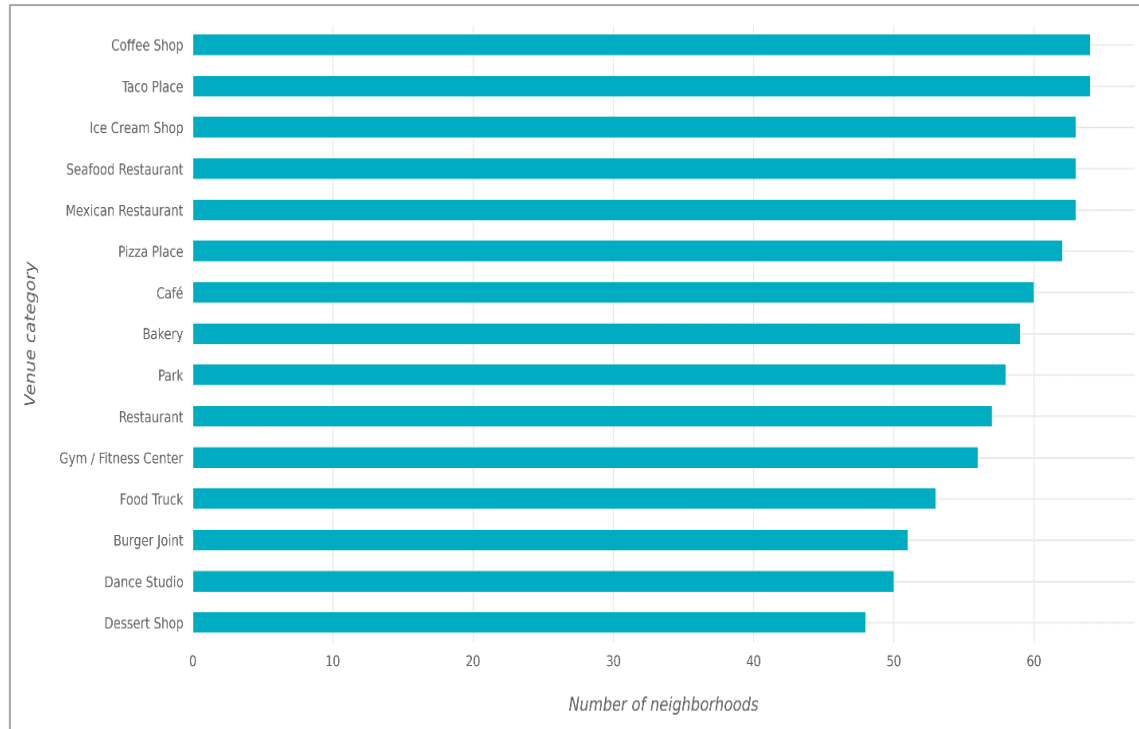
BENITO JUAREZ

MIGUEL HIDALGO

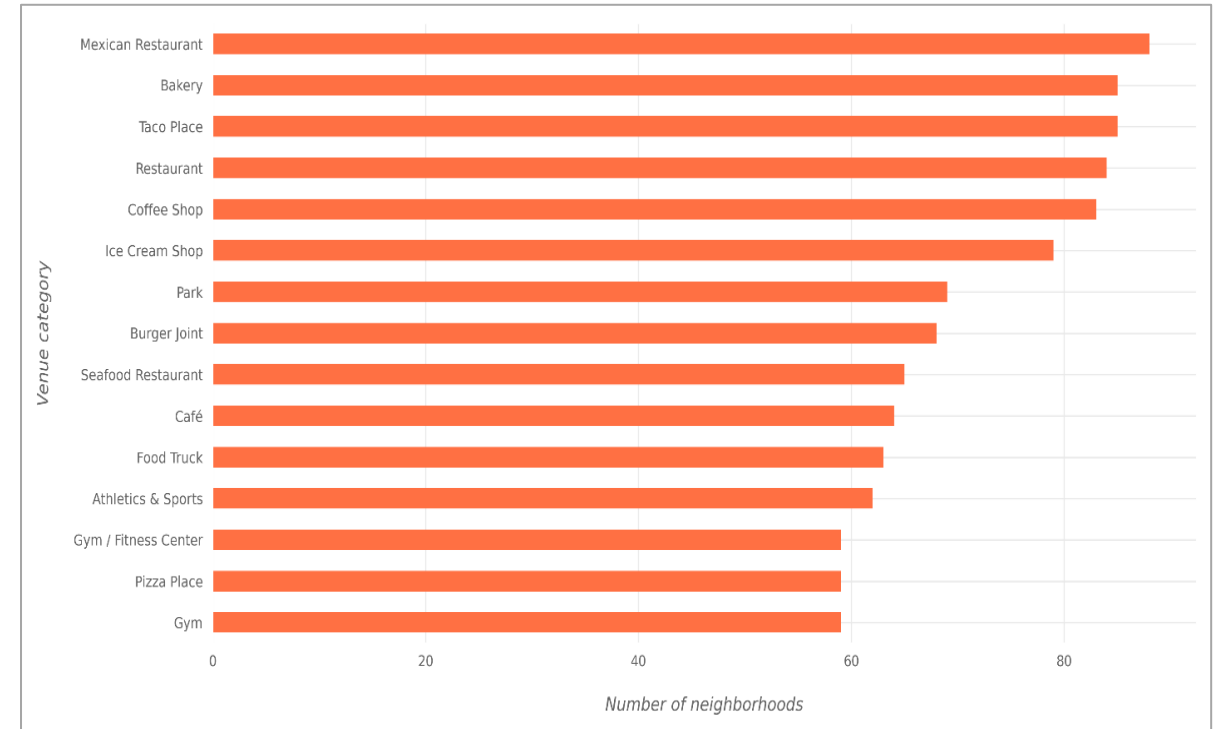


# Most Widespread Venue Categories

BENITO JUAREZ



MIGUEL HIDALGO



# One-hot encoding

---

	Neighborhood	Accessories Store	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	Automotive Shop
0	LOMAS DE CHAPULTEPEC	0	0	0	0	0	0	0	0	0	0	0	0
1	LOMAS DE CHAPULTEPEC	0	0	0	0	0	0	0	0	0	0	0	0
2	LOMAS DE CHAPULTEPEC	0	0	0	0	0	0	0	0	0	0	0	0
3	LOMAS DE CHAPULTEPEC	0	0	0	0	0	0	0	0	0	0	0	0
4	LOMAS DE CHAPULTEPEC	1	0	0	0	0	0	0	0	0	0	0	0

Before using the K-means clustering algorithm, we first need to prepare our data. For this purpose, one-hot encoding will be applied on the “Venue Category” feature and the result of the encoding will be used for the clustering.



# Combining the two dataframes

	Neighborhood	Accessories Store	Advertising Agency	African Restaurant	American Restaurant	Antique Shop	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Res
60	VERTIZ_NARVARTE_Benito Juarez	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.00	0.000000	0.0	
61	VILLA DE CORTES_Benito Juarez	0.0	0.0	0.00	0.0	0.0	0.00	0.01	0.01	0.00	0.000000	0.0	
62	XOCO_Benito Juarez	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.01	0.01	0.000000	0.0	
63	ZACAHUITZCO_Benito Juarez	0.0	0.0	0.00	0.0	0.0	0.00	0.02	0.00	0.00	0.000000	0.0	
64	10 DE ABRIL_Miguel Hidalgo	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.01	0.000000	0.0	
65	16 DE SEPTIEMBRE_Miguel Hidalgo	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.00	0.019608	0.0	
66	5 DE MAYO_Miguel Hidalgo	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.00	0.000000	0.0	
67	AGRICULTURA_Miguel Hidalgo	0.0	0.0	0.01	0.0	0.0	0.00	0.00	0.01	0.00	0.010000	0.0	
68	AHUEHUETES ANAHUAC_Miguel Hidalgo	0.0	0.0	0.00	0.0	0.0	0.01	0.00	0.00	0.00	0.010000	0.0	
69	AMERICA_Miguel Hidalgo	0.0	0.0	0.00	0.0	0.0	0.00	0.00	0.00	0.00	0.013514	0.0	

# The most common categories for each neighborhood

	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	9th Most Common Venue	10th Most Common Venue
0	ACACIAS_Benito Juarez	Coffee Shop	Mexican Restaurant	Bakery	Ice Cream Shop	Boutique	Cosmetics Shop	Steakhouse	Shopping Mall	Seafood Restaurant	Burger Joint
1	ACTIPAN_Benito Juarez	Ice Cream Shop	Café	Coffee Shop	Cosmetics Shop	Gym / Fitness Center	Argentinian Restaurant	Taco Place	Supermarket	Diner	Dessert Shop
2	ALAMOS I_Benito Juarez	Mexican Restaurant	Taco Place	Bakery	Burger Joint	Seafood Restaurant	Coffee Shop	Dessert Shop	Café	Ice Cream Shop	Restaurant
3	ALAMOS II_Benito Juarez	Taco Place	Mexican Restaurant	Burger Joint	Café	Coffee Shop	Ice Cream Shop	Bakery	Breakfast Spot	Dance Studio	Bar
4	ALBERT_Benito Juarez	Mexican Restaurant	Taco Place	Breakfast Spot	Food Truck	Coffee Shop	Bakery	Gym	Soccer Field	Pool	Flea Market

Due to the variety of venues, only the top 10 common venues are selected for each neighborhood as the features to train the K-means clustering algorithm. This dataframe is created by retrieving the 10 categories with the largest values for each neighborhood.

# K-means clustering

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
VILLA DE CORTES_Benito Juarez	2	Mexican Restaurant	Taco Place	Ice Cream Shop	Coffee Shop	Bakery	Park	Burger Joint	Dessert Shop	Pizza Place	Gym
XOCO_Benito Juarez	0	Coffee Shop	Ice Cream Shop	Bakery	Taco Place	Boutique	Sushi Restaurant	Shopping Mall	Lingerie Store	Cosmetics Shop	Pizza Place
ZACAHUITZCO_Benito Juarez	2	Mexican Restaurant	Taco Place	Food Truck	Dessert Shop	Café	Coffee Shop	Gym	Sushi Restaurant	Bakery	Sandwich Place
10 DE ABRIL_Miguel Hidalgo	4	Mexican Restaurant	Coffee Shop	Restaurant	Taco Place	Gym / Fitness Center	Ice Cream Shop	Bakery	Sandwich Place	Shopping Mall	Park
16 DE SEPTIEMBRE_Miguel Hidalgo	1	Taco Place	Mexican Restaurant	Convenience Store	Food Truck	Coffee Shop	Restaurant	Pharmacy	Seafood Restaurant	Bus Station	Bakery

The clustering algorithm assigns a cluster label from 0 to 4 to each neighborhood, these labels denote the cluster assigned to each record.

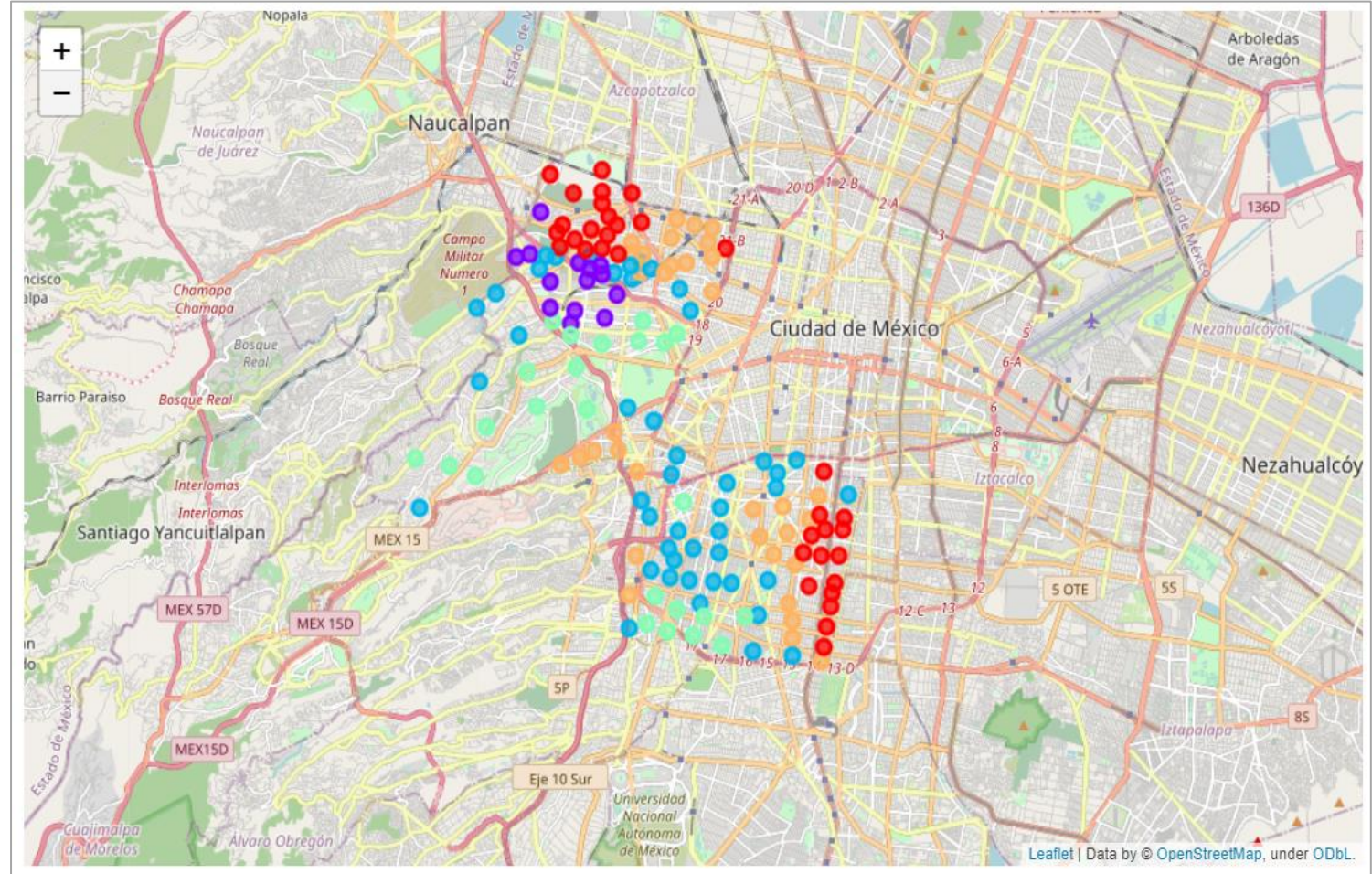
# Part 4: Results

---

The output of the clustering operation is 5 clusters with labels 0, 1, 2, 3, and 4. Each cluster is comprised of a group of neighborhoods that are similar based on the most common venue categories in each neighborhood. The clustering algorithm was run on the 64 neighborhoods in Benito Juarez and the 88 neighborhoods of Miguel Hidalgo.

Table 1: Cluster distribution		
Cluster Label	Number of Neighborhoods	Color
0	24	Pale green
1	34	Orange
2	34	Red
3	14	Purple
4	46	Sky blue

# Map of Mexico City with Benito Juárez and Miguel Hidalgo neighborhood clusters



# Part 5:

## Discussion

From this figure we can distinguish some of the differences in each cluster:

- In the first cluster, Coffee shops are the most common venues.
- In the second cluster, the most common venues are Taco Places, with 12.34% of the venue's categories.
- Mexican Restaurants are the most common venue categories in Clusters 3, 4 and 5.
- The fourth cluster is the only one with Shopping Malls and Boutiques in the top 5 venue categories.

Cluster 1:

Category	% of venues
Coffee Shop	6.595448
Bakery	4.133767
Mexican Restaurant	4.040873
Ice Cream Shop	3.901533
Seafood Restaurant	3.204830

Cluster 2:

Category	% of venues
Taco Place	12.346814
Mexican Restaurant	10.753676
Coffee Shop	5.422794
Bakery	3.890931
Restaurant	3.523284

Cluster 3:

Category	% of venues
Mexican Restaurant	17.504964
Taco Place	13.666446
Coffee Shop	3.970880
Bakery	3.507611
Restaurant	3.375248

Cluster 4:

Category	% of venues
Mexican Restaurant	7.988381
Coffee Shop	6.535948
Boutique	3.994190
Ice Cream Shop	3.195352
Shopping Mall	2.832244

Cluster 5:

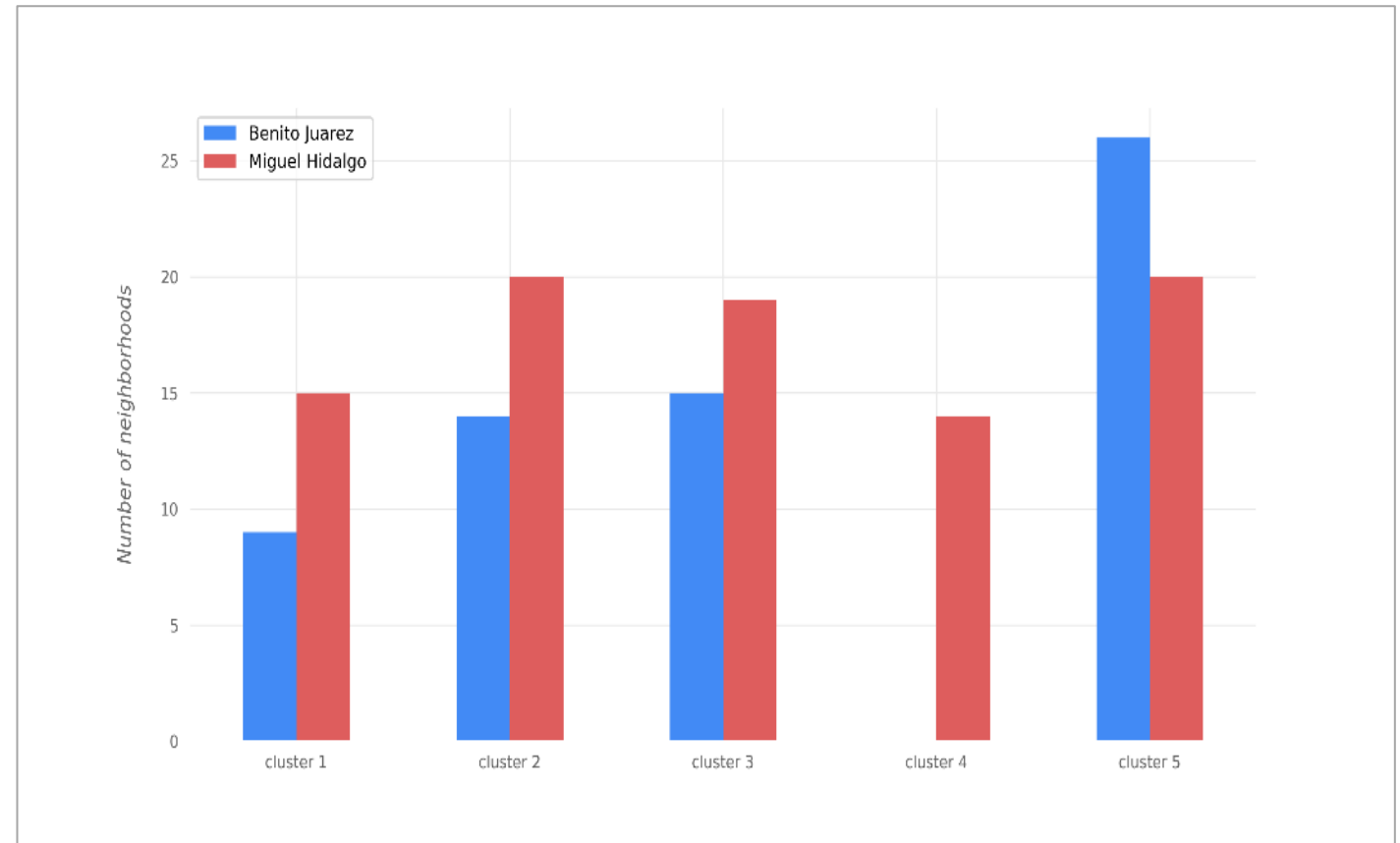
Category	% of venues
Mexican Restaurant	8.293570
Taco Place	6.271302
Coffee Shop	6.044081
Bakery	3.794592
Ice Cream Shop	3.726426

# Number of neighborhoods from Benito Juarez and Miguel Hidalgo in each cluster

From this bar chart, we can examine how the neighborhoods from each borough are distributed in each cluster.

Considering Miguel Hidalgo had ~20 more neighborhoods than Benito Juarez, it is natural to see more neighborhoods from this borough in each cluster. This is the case for most clusters, except for Cluster 5, where there are more neighborhoods from Benito Juarez.

Cluster 4 does not have any neighborhoods from Benito Juarez. Let's remember this is the only cluster where there were shopping malls and boutiques in the top 5 venue category.



# Part 6: Conclusions

---

The goal of this project was to help people decide the best neighborhoods to relocate based on the proximity of venues in Mexico City. By using public data, the neighborhoods of the two of the hottest boroughs from Mexico City were analyzed. The results show that both boroughs, Benito Juárez and Miguel Hidalgo, are remarkably similar. There are clusters of neighborhoods that share common characteristics in the two boroughs, as well as some that do not.

This analysis could be expanded by adding crime rate data, which is also available in Mexico City's Data portal. Another feature to add would be a seismic activity feature, considering Mexico is prone to earthquakes.