

THE BATTLE OF NEIGHBORHOODS

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE CAPSTONE PROJECT REPORT

THE COFFEE PROJECT

- PHOENIX, ARIZONA



- PRAKASH SUDHAKAR

INTRODUCTION

Independent coffee shops pepper the light rail corridor in Midtown & Downtown Phoenix, creating a vibrant, interconnected community. Forget Portland or Seattle – Phoenix stands on its own as a unique coffee destination.

[Phoenix](#) —the fifth most populous city in the United States—has worked hard to forge a balance between urbanism and rurality. It is a metropolis, but one whose main attraction may be the great outdoors nearby. As the city continues to grow, Phoenix has seen generations of people committed to staying here joined by a new influx of young people with a mind to quality food and drink. These newcomers tend to see the city as ripe for innovation, and in synthesis with the city's appreciation for nature, have shaped a movement known as New Arizonan. Over the last few years, this emerging ethos has churned out its fair share of renowned brewers, chefs, and mixologists. In the midst of aggressively growing food and craft beer scenes, Phoenix coffee, too, has quietly begun to come into its own—feeling now more than ever like it's on the verge of a boom. Below are a few café's that have been at the forefront of the culture and which should be watched for new developments.

BUSINESS PROBLEM

The main idea behind the project is to analyze the Coffee chains in Phoenix and look out for the areas where opening a new coffee shop would be a grand success.

My client is a successful entrepreneur in California. It's only been 2 years since he started his business. But now, in 2020, He has 24 Coffee Shops in big cities around California like San Jose, San Francisco, Los Angeles and San Diego. Now, he wants to expand his business in other states and make his 25th Coffee Shop a success. He has a particular interest in Arizona. So, he wants to open a new restaurant in Phoenix.

TARGET AUDIENCE

The target market for **The Coffee Project** will be the working professionals downtown and the college students at the two local universities. The project is also beneficial to anyone who is looking to start a new Coffee Shop in Phoenix. The audience is quite large and includes the major brands like Starbucks Coffee, The Coffee Bean and Tea Leaf. The project can also benefit some locally-owned coffee shops to improve their business prospects.

Our target market of higher-educated people is interested in helping others in a tangible way through their affinity for specialty coffee drinks. According to the [Specialty Coffee Association](#), the U.S. market share of specialty coffee has grown 19% in just seven years and is expected to double by the year 2050. We intend to capture the attention of the roughly 18,000 millennials in Bartlesville who would look for more than just a caffeine fix.

DATA ACQUISITION AND PREPROCESSING:

In order to cluster & segment the neighborhoods and explore them, we will essentially need a dataset that contains the boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

- **Web Scrap Wikipedia Page:** [List of Neighborhood in Arizona](#) page gives us all the information about the neighborhoods present in Arizona. This page has the Neighborhood, borough, population, area and density of all the neighborhoods present in Phoenix. We can also directly download the information from this [link](#).
 - o **Name:** Name of the neighborhood
 - o **Type:** Categorizes into Town/City
 - o **County:** Name of the borough
 - o **Population:** Population of the borough recorded during 2010 census
 - o **Area:** Area of the borough in sq.km and sq.mi
 - o **Population Density:** Population per Area of the borough
 - o **Incorporated:** Year when the borough was incorporated

	Name	Type	County	Population	Area (2010) sq.mi	Area (2010) km2	Population density (2010)	Incorporated
0	Apache Junction	City	Pinal	35,840	35.00	90.6	1,024.0/sq mi (395.4/km	1978
1	Avondale	City	Maricopa	76,238	45.65	118.2	1,670.1/sq mi (644.8/km	1946
2	Benson	City	Cochise	5,105	41.46	107.4	123.1/sq mi (47.5/km	1924
3	Bisbee	City	Cochise	5,575	5.16	13.4	1,080.4/sq mi (417.2/km	1902
4	Buckeye	City	Maricopa	50,876	375.39	972.3	135.5/sq mi (52.3/km	1929

- **Geocoding Location Data:** The data scraped from Wikipedia only contains the names of the borough and its neighborhood. To further analyze the data, we would need their latitude and longitude. With the help of GeoPy library, we geocode the data into its respective geographic coordinates. We will use Nominatim Geocoding service, which is built on top of OpenStreetMap data.

	Neighborhood	Borough	Latitude	Longitude
0	Apache Junction	Pinal	33.415048	-111.549578
1	Avondale	Maricopa	33.435499	-112.349557
2	Benson	Cochise	31.967861	-110.294517
3	Bisbee	Cochise	31.448155	-109.928408
4	Buckeye	Maricopa	33.370275	-112.583867

- **Foursquare API:** Arizona geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venue information for each neighborhood. This link redirects you to [Foursquare API](#).

We need to define the corresponding URL and send the GET request to Foursquare API and examine the results. Next, get the relevant part of JSON and transform it into a pandas dataframe. I would like to explore all the venues in each neighborhood which are within 8 km radius. So, I define a query to search the top 100 venues in each neighborhood.

- **User ID:** Anonymous
- **Venue Name:** Returns the name of the Venue
- **Venue Category:** Returns the Category to which the venue belongs
- **Venue Latitude:** Returns the Latitude of the Venue
- **Venue Longitude:** Returns the Longitude of the Venue

	Borough	Neighborhood	Latitude	Longitude	Venue	Venue_Category	Venue_Latitude	Venue_Longitude
0	Pinal	Apache Junction	33.415048	-111.549578	Handlebar Pub And Grill	Bar	33.415463	-111.553158
1	Pinal	Apache Junction	33.415048	-111.549578	Walgreens	Pharmacy	33.412688	-111.547010
2	Pinal	Apache Junction	33.415048	-111.549578	CVS pharmacy	Pharmacy	33.413972	-111.549677
3	Pinal	Apache Junction	33.415048	-111.549578	Los Gringos Locos	Mexican Restaurant	33.412419	-111.551119
4	Pinal	Apache Junction	33.415048	-111.549578	Papa Murphy's	Pizza Place	33.415576	-111.568613

Methodology

Firstly, we need to get the list of neighborhoods in the city of Phoenix. Fortunately, the list is available in the following Wikipedia page 'https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_Arizona'. We will scrap using Python requests and beautifulsoup packages to extract the list of neighborhood data.

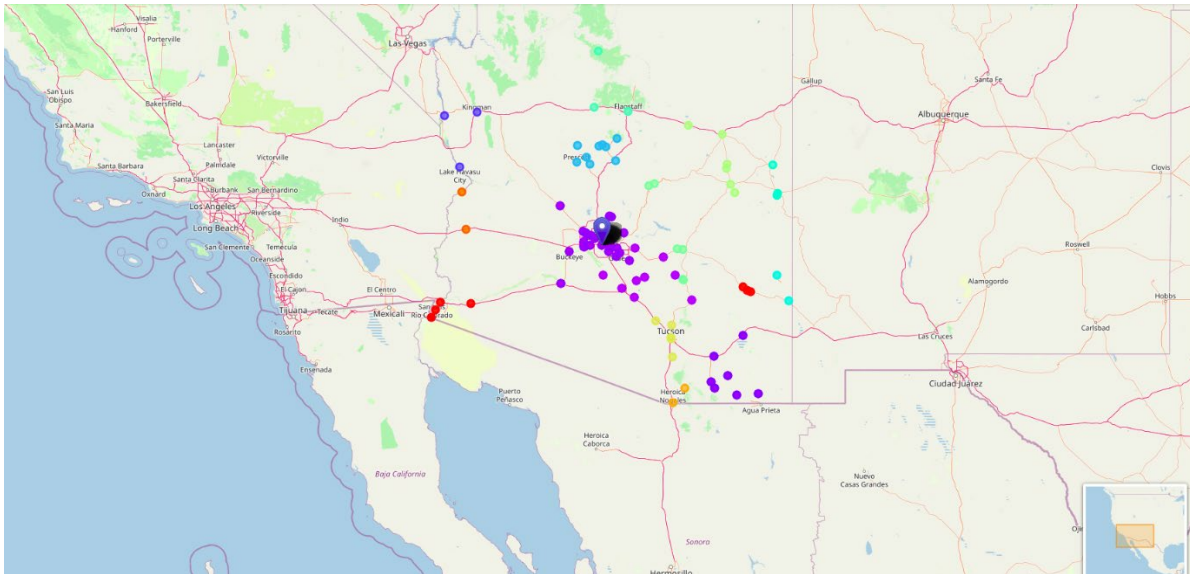
```
# specify which URL/web page we are going to be scraping
url = "https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_Arizona"

# open the url using urllib.request and put the HTML into the page variable
page = urllib.request.urlopen(url)

# parse the HTML from our URL into the BeautifulSoup parse tree format
soup = BeautifulSoup(page, "lxml")
print(soup.prettify())
```

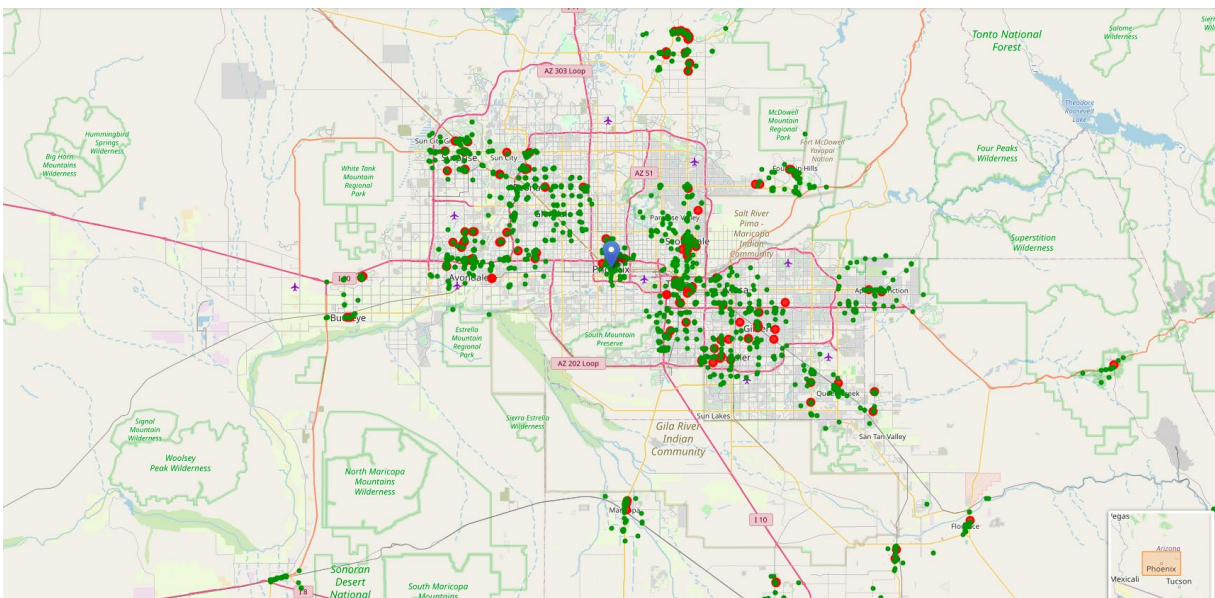
After gathering the data from Wikipedia, we got a list of 90 neighborhoods that were in turn fed into Geopy (Geocoder Package) to retrieve the geographical coordinates of Arizona. We can use the location's latitude and longitude values to visualize the neighborhood in the map using the Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data

returned by Geocoder are correctly plotted. Each Borough is being clustered and shown in the form of rainbow colors for better visualization.



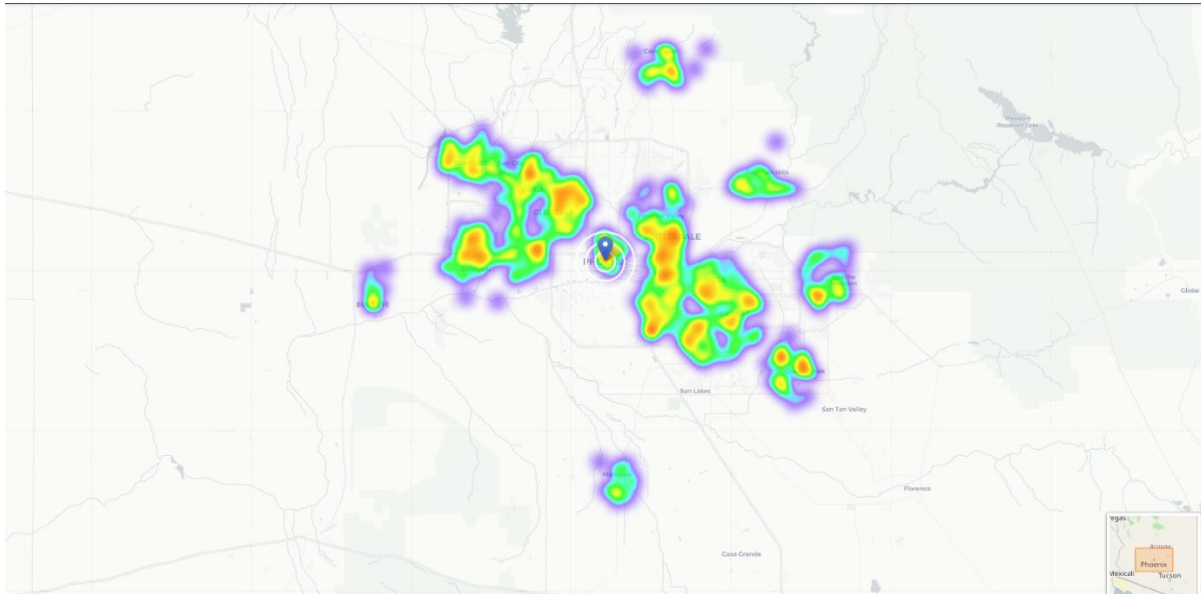
Next, we can find the top 100 venues with 8 km in each neighborhood by connecting to the Foursquare API. I have created a Foursquare developer account to obtain an account ID and API key to pull the data. From Foursquare, I can pull the names, categories, latitude and longitude of the venues. With this data, I can also check how many unique categories that I can get from these venues. It returns a JSON file containing all the venues in each neighborhood which is converted into a data frame. In short, we have 303 unique categories that were returned from 2653 distinct venues by Foursquare API.

Exploring the Data and Major Neighborhoods of Arizona:

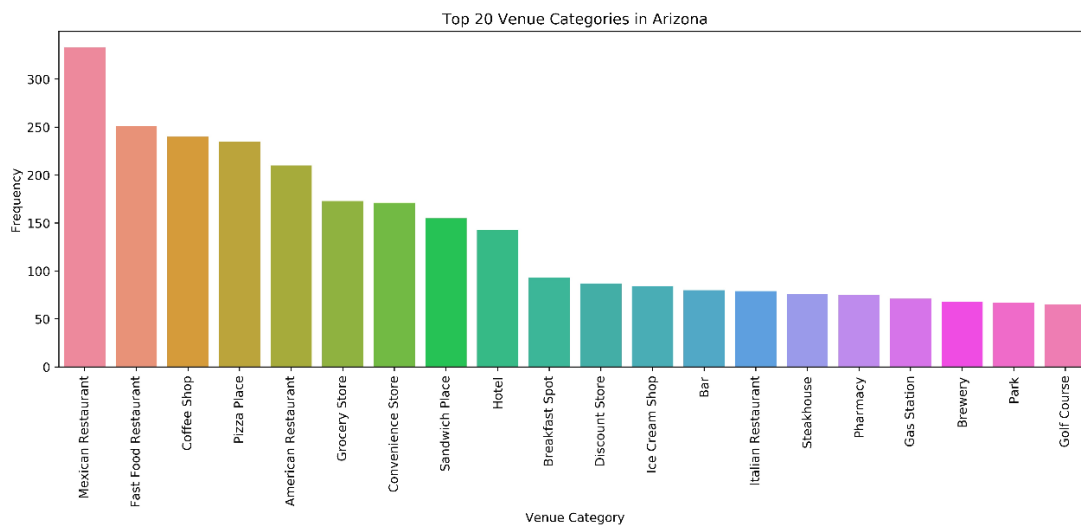


There are 24 neighborhoods present near the Phoenix Area, Arizona. They are visualized using the Folium package in Python. The green circles represent all venues other than coffee shops, while the red-colored circles represent coffee shops.

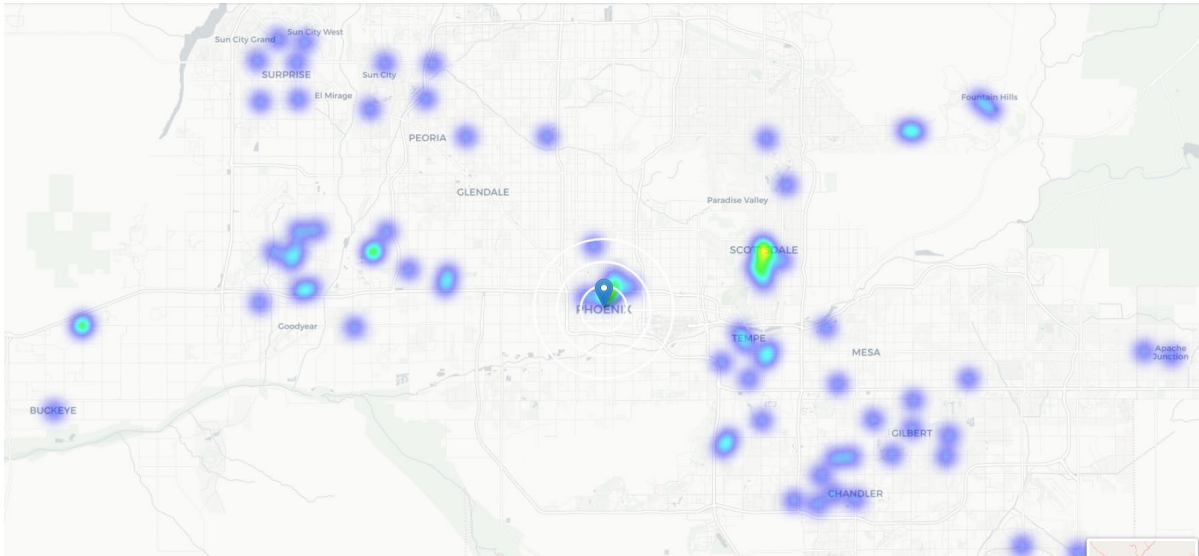
Venues other than Coffee Shops - Neighborhood in Phoenix



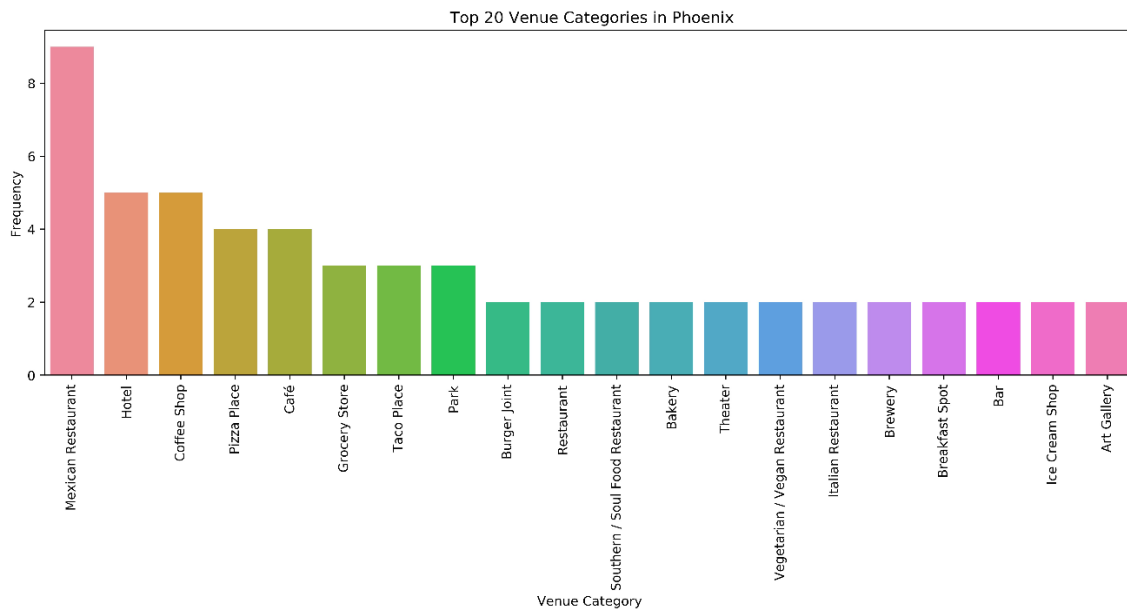
The above heatmap shows us the presence of venues in and around Phoenix Area, Arizona. We can also plot a bar chart with the most populated venues in a neighborhood.



Coffee Shops – Neighborhood in Phoenix



Here's a heatmap that represents the places which are populated with coffee shops. It gives us an understanding of where to expect competition from when opened in these regions. We can see that Phoenix and Scottsdale are highly populated with coffee shops, thereby increasing our chances of opening considering the competition level and popularity of our coffee service.



Coffee shops are everywhere! Not only in the overall state, but it's also in the top 3 most visited venue category in Phoenix.

As the first step of cluster analysis, one hot encoding was conducted to give binary values to each venue categories.

```
# one hot encoding
phoenix_vc_onehot = pd.get_dummies(phoenix_vicinity_venues[['Venue_Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
phoenix_vc_onehot['Neighborhood'] = phoenix_vicinity_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [phoenix_vc_onehot.columns[-1]] + list(phoenix_vc_onehot.columns[:-1])
phoenix_vc_onehot = phoenix_vc_onehot[fixed_columns]
phoenix_vc_onehot.set_index(['Neighborhood'], inplace = True)

print(phoenix_vc_onehot.shape)
phoenix_vc_onehot.head()
```

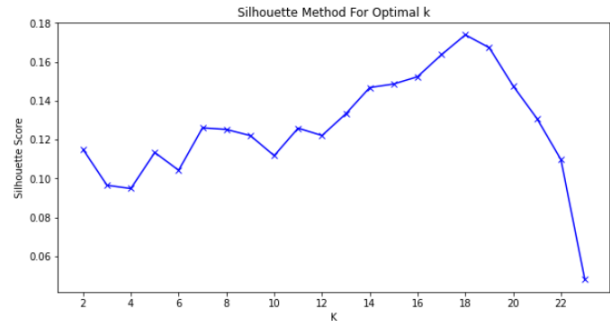
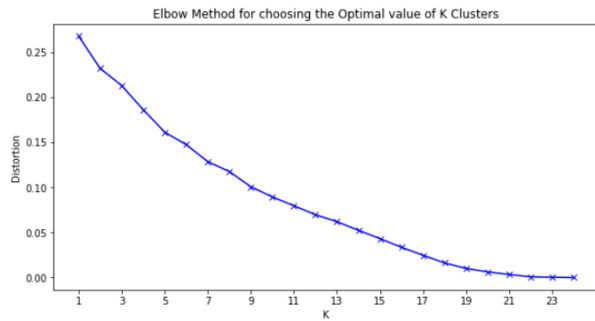
(2256, 222)

	Zoo	ATM	Adult Boutique	Advertising Agency	American Restaurant	Antique Shop	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Bagel Shop	Bakery	Bank	Bar	Baseball Field	Baseball Stadium	Basketball Stadium	Beer Bar	Beer Store	Big Box Store
Neighborhood																							
Apache Junction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Apache Junction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apache Junction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apache Junction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Apache Junction	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Then, the data were grouped by borough names to find out how many venues of each category exist in the boroughs within the top 100 venues. However, as some boroughs display less than 100 venues due to lack of Foursquare data, the category counts were altered to the frequency of how often the category appears among others. Based on the frequency, we could attain a list of most common venue categories in each borough as follows.

	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	Apache Junction	Fast Food Restaurant	Sandwich Place	Trail	Pharmacy	Grocery Store	Pizza Place	Convenience Store	Mexican Restaurant	Breakfast Spot	Spa
1	Avondale	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Gym	Pizza Place	Convenience Store	Frozen Yogurt Shop	Seafood Restaurant	Donut Shop
2	Buckeye	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Gas Station	Discount Store	Coffee Shop	Bank	Pharmacy	Sandwich Place	Wings Joint
3	Carefree	Coffee Shop	Golf Course	American Restaurant	Grocery Store	Mexican Restaurant	Convenience Store	BBQ Joint	Sandwich Place	Pizza Place	Bar
4	Cave Creek	Coffee Shop	American Restaurant	Golf Course	Mexican Restaurant	BBQ Joint	Diner	Sandwich Place	Convenience Store	Grocery Store	Pizza Place

Once we get quite a broad overview of the different types of venues, mainly coffee shops around the major neighborhoods in Phoenix, it is time to use clustering the districts using K-Means.



We iterate over loops to find the optimal value of K for the K-Means Clustering and Silhouette approach. I have decided to segment these neighborhoods into 5 clusters based on the above graphs and use K-Means clustering to segment by venue categories. So, our expectation would be based on the similarities of venue categories, these neighborhoods will be clustered.

RESULTS

Based on our analysis above, we can draw several conclusions that will be useful to aid any visitor visiting the city of Phoenix, Arizona. The results from k-means clustering show that we can categorize Phoenix neighborhoods into 5 clusters based on how many Coffee Shops are in each neighborhood:

Cluster 0: Neighborhoods with little Coffee Shops

	Borough	Neighborhood	Latitude	Longitude	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
2	Maricopa	Buckeye	33.370275	-112.583867	0	Pizza Place	Fast Food Restaurant	Mexican Restaurant	Gas Station	Discount Store	Coffee Shop	Bank	Pharmacy	Sandwich Place	Wings Joint
13	Pinal	Maricopa	33.058129	-112.047755	0	Pizza Place	Fast Food Restaurant	American Restaurant	Donut Shop	Gym / Fitness Center	Mexican Restaurant	Wings Joint	Sandwich Place	Pharmacy	Coffee Shop

Cluster 1: Neighborhoods with a high number of Coffee Shops

	Borough	Neighborhood	Latitude	Longitude	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
1	Maricopa	Avondale	33.435499	-112.349557	1	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Gym	Pizza Place	Convenience Store	Frozen Yogurt Shop	Seafood Restaurant	Donut Shop
10	Maricopa	Goodyear	33.435367	-112.357601	1	Coffee Shop	Mexican Restaurant	American Restaurant	Fast Food Restaurant	Gym	Pizza Place	Golf Course	Frozen Yogurt Shop	Grocery Store	Convenience Store
12	Maricopa	Litchfield Park	33.493380	-112.358124	1	Coffee Shop	Mexican Restaurant	Pizza Place	Gym	American Restaurant	Grocery Store	Fast Food Restaurant	Golf Course	Japanese Restaurant	Sushi Restaurant
15	Maricopa	Paradise Valley	33.542801	-111.955600	1	American Restaurant	Pizza Place	Breakfast Spot	Hotel	Grocery Store	Trail	Mexican Restaurant	Spa	Steakhouse	Restaurant
19	Maricopa	Scottsdale	33.494219	-111.926018	1	Coffee Shop	Bar	Mexican Restaurant	Hotel	Pizza Place	Park	Restaurant	American Restaurant	New American Restaurant	Sushi Restaurant
21	Maricopa	Tempe	33.425506	-111.940012	1	Coffee Shop	Brewery	Sandwich Place	Park	Pizza Place	American Restaurant	Convenience Store	Bar	Middle Eastern Restaurant	Mexican Restaurant

Cluster 2: Neighborhoods with a high number of Coffee Shops

	Borough	Neighborhood	Latitude	Longitude	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
3	Maricopa	Carefree	33.822261	-111.918203	2	Coffee Shop	Golf Course	American Restaurant	Grocery Store	Mexican Restaurant	Convenience Store	BBQ Joint	Sandwich Place	Pizza Place	Bar
4	Maricopa	Cave Creek	33.833333	-111.950833	2	Coffee Shop	American Restaurant	Golf Course	Mexican Restaurant	BBQ Joint	Diner	Sandwich Place	Convenience Store	Grocery Store	Pizza Place
7	Maricopa	Fountain Hills	33.611711	-111.717361	2	Pizza Place	Golf Course	American Restaurant	Mexican Restaurant	Coffee Shop	Grocery Store	Italian Restaurant	Bank	Convenience Store	Casino

Cluster 3: Neighborhoods with a high number of Coffee Shops

	Borough	Neighborhood	Latitude	Longitude	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
5	Maricopa	Chandler	33.306222	-111.841332	3	Mexican Restaurant	Pizza Place	Coffee Shop	Grocery Store	Café	Beer Bar	Fast Food Restaurant	Sandwich Place	Sushi Restaurant	Chinese Restaurant
6	Maricopa	El Mirage	33.613034	-112.324487	3	Mexican Restaurant	Grocery Store	Convenience Store	Coffee Shop	Sandwich Place	Pizza Place	Fast Food Restaurant	Pharmacy	Park	Ice Cream Shop
8	Maricopa	Gilbert	33.352763	-111.789037	3	Coffee Shop	Mexican Restaurant	Pizza Place	Grocery Store	Sandwich Place	Fast Food Restaurant	Convenience Store	Bar	Sushi Restaurant	Brewery
9	Maricopa	Glendale	33.538686	-112.185994	3	Mexican Restaurant	Fast Food Restaurant	American Restaurant	Convenience Store	Park	Grocery Store	Pizza Place	Bar	Furniture / Home Store	Sandwich Place
11	Maricopa	Guadalupe	33.363125	-111.962533	3	Grocery Store	Pizza Place	Italian Restaurant	Breakfast Spot	Mexican Restaurant	Ice Cream Shop	Steakhouse	Convenience Store	Coffee Shop	Furniture / Home Store
14	Maricopa	Mesa	33.415112	-111.831477	3	Mexican Restaurant	Brewery	Fast Food Restaurant	Convenience Store	Sandwich Place	Steakhouse	Korean Restaurant	Pizza Place	Grocery Store	Coffee Shop
16	Maricopa	Peoria	33.580612	-112.237294	3	Mexican Restaurant	Pizza Place	Fast Food Restaurant	Grocery Store	American Restaurant	Coffee Shop	Sandwich Place	Convenience Store	Park	BBQ Joint
17	Maricopa	Phoenix	33.448437	-112.074142	3	Mexican Restaurant	Hotel	Coffee Shop	Café	Pizza Place	Taco Place	Park	Grocery Store	Italian Restaurant	Southern / Soul Food Restaurant
18	Maricopa	Queen Creek	33.248386	-111.634158	3	Mexican Restaurant	Pizza Place	Pharmacy	Coffee Shop	Sandwich Place	Salon / Barbershop	Italian Restaurant	Golf Course	Furniture / Home Store	Grocery Store
20	Maricopa	Surprise	33.629227	-112.368019	3	Coffee Shop	Mexican Restaurant	Pizza Place	Grocery Store	Pharmacy	Fast Food Restaurant	American Restaurant	Ice Cream Shop	Golf Course	Italian Restaurant
22	Maricopa	Tolleson	33.450050	-112.259309	3	Coffee Shop	Fast Food Restaurant	Convenience Store	Mexican Restaurant	Grocery Store	Pizza Place	Hardware Store	Burger Joint	Furniture / Home Store	Seafood Restaurant
23	Maricopa	Youngtown	33.593730	-112.303326	3	Mexican Restaurant	American Restaurant	Pizza Place	Grocery Store	Coffee Shop	Sandwich Place	Convenience Store	Pharmacy	Fast Food Restaurant	Lingerie Store

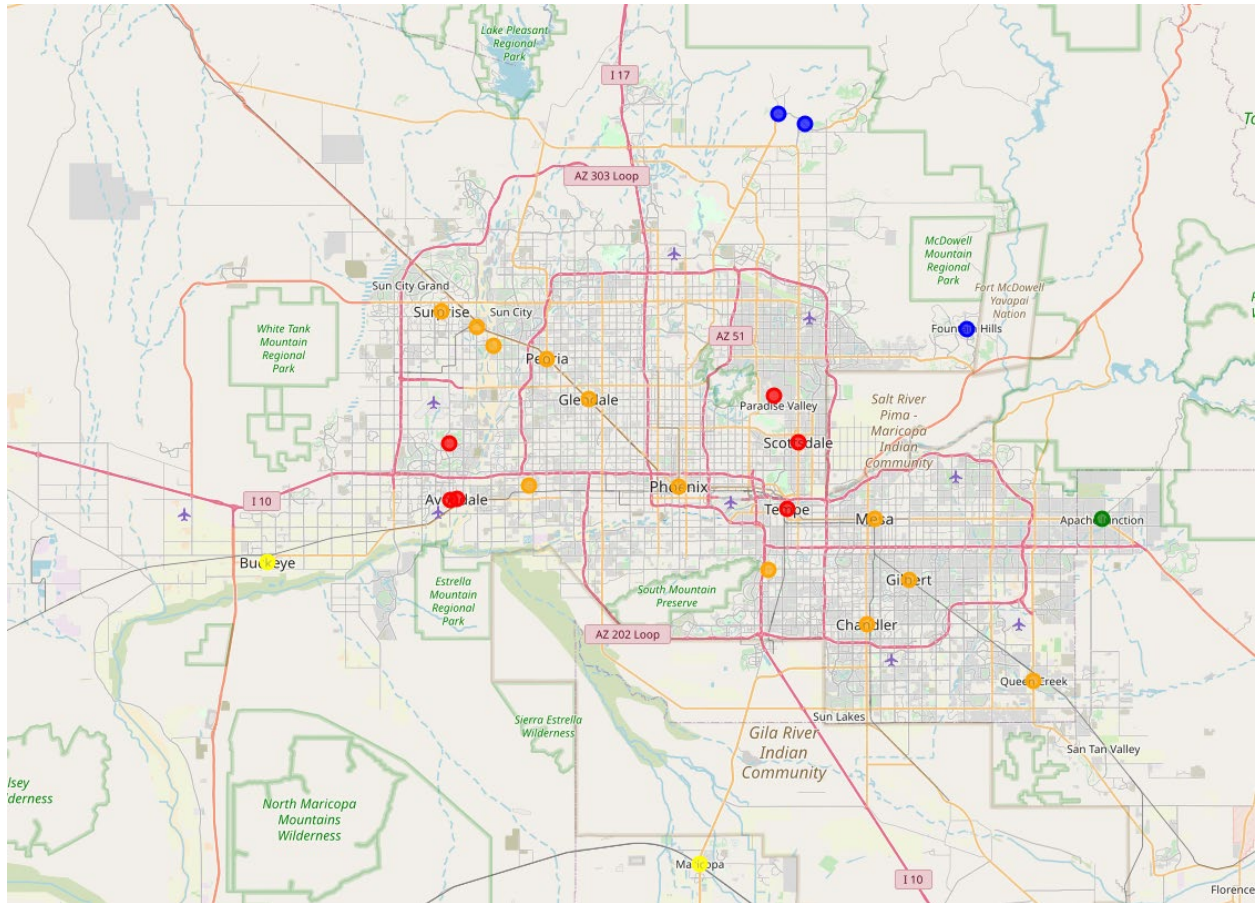
Cluster 4: Neighborhoods with little Coffee Shops

	Borough	Neighborhood	Latitude	Longitude	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
0	Pinal	Apache Junction	33.415048	-111.549578	4	Fast Food Restaurant	Sandwich Place	Trail	Pharmacy	Grocery Store	Pizza Place	Convenience Store	Mexican Restaurant	Breakfast Spot	Spa

The cluster three is the biggest cluster with 12 of the 24 neighborhoods in the borough Maricopa. Upon closely examining these neighborhoods, we can see that the most common venues in this cluster are Mexican Restaurants, Coffee Shops, Pizza places followed by Fast food and American Restaurant.

The clusters one, two and three are highly populated with coffee shops and the clusters zero and four are sparsely populated with coffee shops. There are good opportunities for us to open up coffee shops in the later mentioned cluster considering the other factors such as population, salary and types of jobs in those neighborhoods.

The results are visualized in the above map with Cluster 0 in yellow color, Cluster 1 in red color, Cluster 2 in blue color, Cluster 3 in orange color and Cluster 4 in green color.



According to this analysis, Cluster zero and four will provide the least competition for an upcoming coffee shop as a pizza place and American restaurants are the most common venues in this area and, the frequency of coffee shops as common venues is very low compared to the remaining neighborhood.

We identified that from the total set of venues, the majority of them were Coffee Shops and Mexican Restaurants. A visitor who loves Coffee/Mexican Restaurants would surely benefit from coming to Phoenix.

LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

In this capstone project, I only take into consideration of one factor - the occurrence/existence of Coffee Shops in each neighborhood. Many factors can be taken into consideration such as population density, the income of residents, rent that could influence the decision to open a coffee shop. However, to put all these data into this project is not possible to do within a short time frame for this capstone project. Future research can take into consideration these factors.

CONCLUSION

To conclude this project, we have got a small glimpse of how real-life data-science projects look like. I've made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore the major neighborhoods of Phoenix and saw the results of segmentation of neighborhoods using Folium map. We further shortlist the neighborhoods based on the common venues and perform machine learning by utilizing k-means clustering to choose a neighborhood that best suits the business problem. Potential for this kind of analysis in a real-life business problem is discussed in great detail.

Finally, since my analysis was mostly concentrated on the possibilities of opening a coffee shop targeting the huge pool of office workers and students around Downtown Phoenix, some of the results obtained are surprisingly exactly what I have expected after staying 2 years in Phoenix. Hopefully, this kind of analysis will provide you initial guidance to take more real-life challenges using data science.

The final decision on optimal coffee shop location will be made by my client based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like the attractiveness of each location (proximity to park or water), levels of noise/proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood.