Coursera Final Capstone Project

Author: Kyle Stewart

Table of Contents

Introduction

Background

It is estimated that approximately 500,000 people moved from California to Arizona from 2010 to 2018. This may be for a variety of reasons, but arguably one contributing factor is due to California's high cost of living. One of the more popular areas for Californians to settle in Arizona is the greater Phoenix metropolitan area.

Problem

Suppose a resident of California decides to make this move across the state line. In this hypothetical scenario, our California citizen, whom we will name "John Doe", currently lives in California, but he finds a remote data science jobs that would allow him to relocate to another state where the cost of living is more within his budget. He has heard good things about the Phoenix, AZ metropolitan area, and decides to make the move.

However, in our scenario, suppose John Doe really likes "the feel" of his current city, in terms of available venues nearby: parks, stores, the variety of restaurants, etc. The question before John is: "what specific city/community in the greater Pheonix area would be the best match to his current city in California?

Note: in solving this problem, the notebook will be designed to take in **any** input city (from California), to run the comparison.

Interest

Interested parties who might like to know the best solution to this problem obviously includes any Californian resident planning to move to Arizona, in the Phoenix area. However, those cities in and around Phoenix may also be interestd in the results, to know how better to focus any efforts to recruit new residents and grow their commerce.

Data Acquisition and Cleaning

Data Required

1. Location of "target" city in California (where John Doe currently lives)
2. Population of "target" city
3. Venue information for target city
4. List of cities in greater Phoenix, AZ area
5. Venue information for each of these cities
6. Latitude and Longitude of each city

Data Sources

1. Location of "target" city is **user-provided**.
2. Population of "target" city is provided by manual **Google-search**
3. Venue information for "target" city is obtained with **Foursquare API**

4. Publicly available lists of cities in the greater Phoenix, AZ area, along with their estimates populations, will be convered to a more easily machine-readable csv file. Data originates fro**m:** **https://en.wikipedia.org/wiki/Phoenix_metropolitan_area**
5. Venue information for each city in the list is obtained with **Foursquare API**
6. Latitute and longitude of each city (for use in Foursquare API) is obtained by the OpenSteetMap API (via geocoder library for python)

Data Cleaning and Feature Selection

1. City names were collected into a dataframe. The name of each city and state was then entered into the geocoder routine to retrive the longitude and latitude from OpenStreetMap API.
2. Foursquare API was called for each city to compile list of venues, saving the "venue category" for each venue found. Venues were confined to within 2 km (2000 m) of the city center long/lat coordinates, and limited to the 100 top results for each city.
3. One hot encoding was utilized to create a numeric dataframe including all venue category information for all cities.
4. This was finally grouped by city, aggregate into a single pandas dataframe the following: city, state, latitude, longitude, population, and top 15 most popular venues within each city (by venue category count retreived from Foursquare API). This dataframe was used for visual inspection of venue categories.

Methodology:

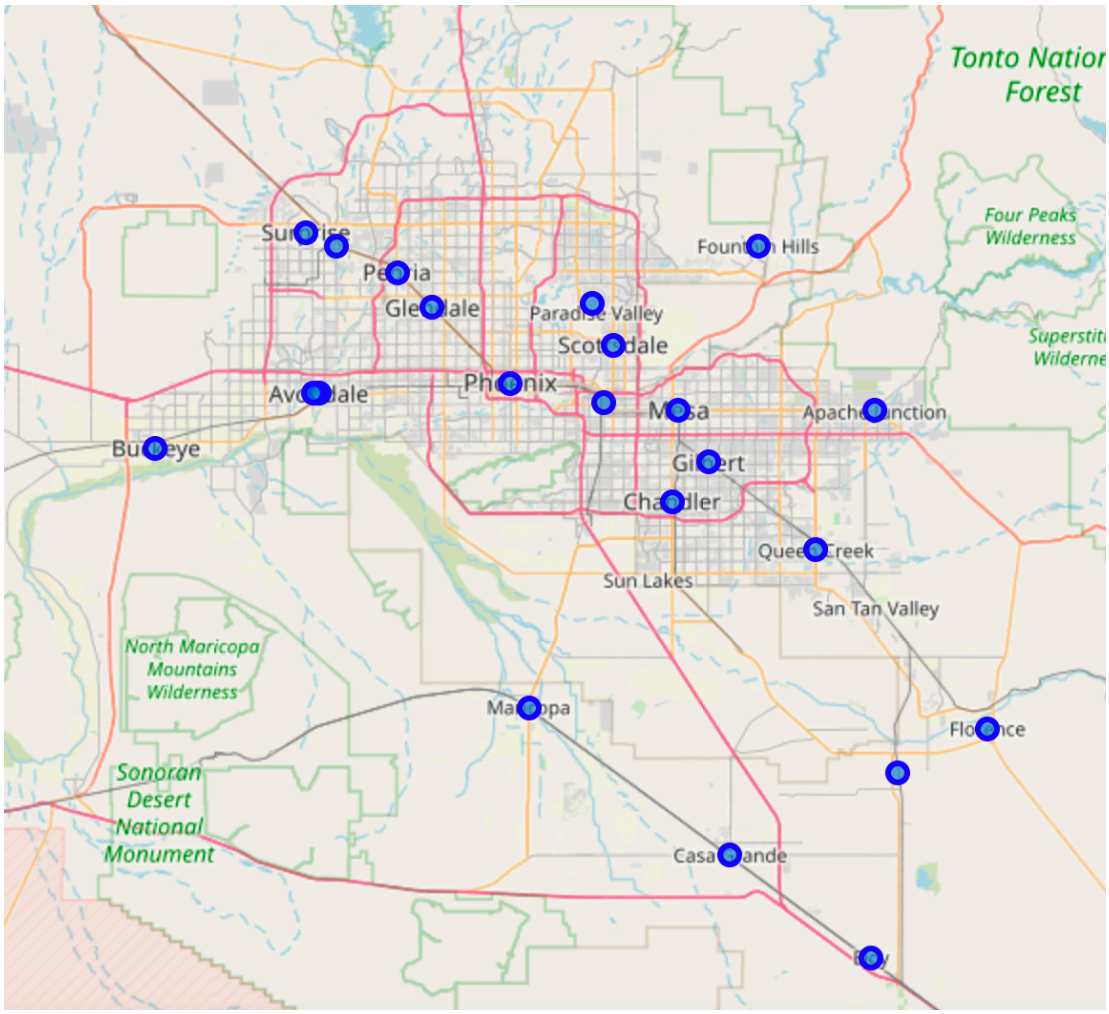
Data Exploration and Visualization

List of # of venues found in Foursquare API from each location (including the CA location of "Rancho Cucamonga, CA"):

City	# of Venues (<2 km)
Apache Junction	38
Avondale	22
Buckeye	17
Casa Grande	13
Chandler	50
Coolidge	12
El Mirage	12
Eloy	9
Florence	16
Fountain Hills	23
Gilbert	46
Glendale	39
Goodyear	7
Maricopa	21
Mesa	55
Paradise Valley	4
Peoria	22
Phoenix	100
Queen Creek	68
Rancho Cucamonga	82
Scottsdale	100

```
Surprise              16
Tempe                100
```

Visualization of target cities in Phoenix, AZ area:



Aggregate Venue Category DataFrame :

	City	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	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue
0	Apache Junction	Fast Food Restaurant	Pharmacy	Chinese Restaurant	Breakfast Spot	American Restaurant	Convenience Store	Pizza Place	Gas Station	Liquor Store	Salon / Barbershop	Sandwich Place	Sports Bar	Park	Discount Store	Bar
1	Avondale	Mexican Restaurant	Bakery	Health & Beauty Service	Diner	Social Club	Locksmith	Fast Food Restaurant	Burger Joint	Seafood Restaurant	Auto Garage	Discount Store	Automotive Shop	Fish & Chips Shop	American Restaurant	Gun Shop
2	Buckeye	Mexican Restaurant	Discount Store	Park	Pizza Place	Business Service	Bank	Café	BBQ Joint	Sandwich Place	Pool	Gas Station	Grocery Store	Convenience Store	Coffee Shop	Dog Run
3	Casa Grande	Construction & Landscaping	Pizza Place	City Hall	Football Stadium	Burger Joint	Burrito Place	New American Restaurant	Café	Chinese Restaurant	Mexican Restaurant	American Restaurant	Ice Cream Shop	Dessert Shop	Eastern European Restaurant	Financial or Legal Service
4	Chandler	Mexican Restaurant	Taco Place	Italian Restaurant	Beer Bar	Bar	Brewery	Café	Lounge	Intersection	Juice Bar	Korean Restaurant	Donut Shop	Sandwich Place	Library	Liquor Store
5	Coolidge	Pizza Place	Mexican Restaurant	Convenience Store	Greek Restaurant	Post Office	Discount Store	Café	Business Service	Cupcake Shop	Cosmetics Shop	Comic Shop	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant
6	El Mirage	Mexican Restaurant	Convenience Store	Grocery Store	Locksmith	Baseball Field	Bakery	Food & Drink Shop	Market	Pizza Place	Cupcake Shop	Fabric Shop	Flower Shop	Fish & Chips Shop	Concert Hall	Financial or Legal Service
7	Eloy	Convenience Store	Post Office	Discount Store	Mexican Restaurant	Park	Burger Joint	Grocery Store	Real Estate Office	Falafel Restaurant	Farmers Market	Fabric Shop	English Restaurant	Fast Food Restaurant	Eastern European Restaurant	Donut Shop
8	Florence	Mexican Restaurant	Sandwich Place	Fast Food Restaurant	Greek Restaurant	Bar	Gas Station	General Travel	Arts & Crafts Store	Chinese Restaurant	Convenience Store	Food & Drink Shop	Hardware Store	Taco Place	Falafel Restaurant	Dive Bar
9	Fountain Hills	Bank	Pharmacy	Pizza Place	Grocery Store	Gym / Fitness Center	Plaza	Hardware Store	Salon / Barbershop	Sandwich Place	Shipping Store	Donut Shop	Mexican Restaurant	Spanish Restaurant	Sports Bar	Japanese Restaurant
10	Gilbert	Mexican Restaurant	Plaza	Ice Cream Shop	Sandwich Place	Italian Restaurant	Coffee Shop	New American Restaurant	Bridal Shop	Café	Restaurant	Candy Store	Pool	Gym	Pizza Place	Convenience Store
11	Glendale	Mexican Restaurant	Coffee Shop	Bar	American Restaurant	Pizza Place	Bank	Steakhouse	Sports Bar	Italian Restaurant	Bed & Breakfast	Comic Shop	Tea Room	Restaurant	Boutique	Fast Food Restaurant
12	Goodyear	Intersection	Home Service	Fish & Chips Shop	Airport	Automotive Shop	Burger Joint	Sandwich Place	Yoga Studio	Eastern European Restaurant	English Restaurant	Falafel Restaurant	Fabric Shop	Dog Run	Farmers Market	Fast Food Restaurant
13	Maricopa	American Restaurant	Fast Food Restaurant	Discount Store	Coffee Shop	BBQ Joint	Clothing Store	Convenience Store	Donut Shop	Flower Shop	Lawyer	Gas Station	Pizza Place	Train Station	Wings Joint	Gym / Fitness Center
14	Mesa	Mexican Restaurant	Sandwich Place	Bank	Brewery	Hotel	Bookstore	Fast Food Restaurant	Bakery	Arts & Crafts Store	Pizza Place	Performing Arts Venue	Paper / Office Supplies Store	Music Store	Museum	Antique Shop
15	Paradise Valley	Football Stadium	Pawn Shop	Trail	Tour Provider	Yoga Studio	Dive Bar	Financial or Legal Service	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Fabric Shop	English Restaurant	Eastern European Restaurant	Donut Shop	Dog Run
16	Peoria	Fast Food Restaurant	Theater	Coffee Shop	Park	Clothing Store	Chinese Restaurant	Donut Shop	Mexican Restaurant	Breakfast Spot	Big Box Store	Latin American Restaurant	Bar	Automotive Shop	Pharmacy	Home Service
17	Phoenix	Coffee Shop	Hotel	American Restaurant	Lounge	Pizza Place	Music Venue	Salon / Barbershop	Basketball Stadium	Bar	Cocktail Bar	Shopping Mall	Rock Club	Pub	Park	New American Restaurant
18	Queen Creek	Mexican Restaurant	Pizza Place	Sandwich Place	American Restaurant	Fast Food Restaurant	Bakery	Park	Coffee Shop	Pharmacy	Chinese Restaurant	Ice Cream Shop	Fried Chicken Joint	Burger Joint	Big Box Store	Convenience Store
19	Rancho Cucamonga	Mexican Restaurant	Clothing Store	Japanese Restaurant	Burger Joint	Coffee Shop	Bakery	Pizza Place	Seafood Restaurant	Fast Food Restaurant	Financial or Legal Service	Bank	Pharmacy	Discount Store	Mobile Phone Shop	Asian Restaurant
20	Scottsdale	Coffee Shop	Bar	American Restaurant	Pizza Place	New American Restaurant	Hotel	Taco Place	Mexican Restaurant	Italian Restaurant	Café	Burger Joint	Sporting Goods Shop	Sushi Restaurant	Plaza	Ice Cream Shop
21	Surprise	Baseball Field	Baseball Stadium	Dog Run	Fast Food Restaurant	Sushi Restaurant	Lake	Hotel	Park	Mexican Restaurant	Coffee Shop	Locksmith	Pool	Accessories Store	Tennis Court	Eastern European Restaurant
22	Tempe	Coffee Shop	Pizza Place	Sandwich Place	Breakfast Spot	American Restaurant	Mexican Restaurant	Bar	Yoga Studio	Dessert Shop	Seafood Restaurant	Brewery	Chinese Restaurant	Beer Garden	Burger Joint	Mediterranean Restaurant

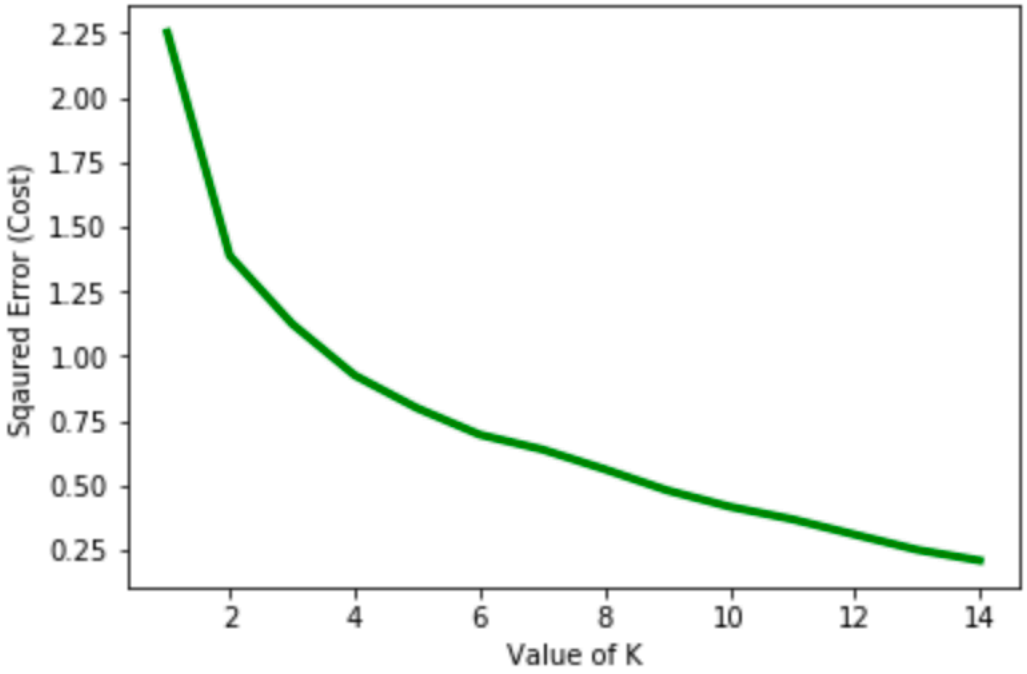
Results

Cluster Modeling

Deciding on k_clusters (cluster number)

The cluster modeling implemented was k-means clustering on the final list of cities, one-hot-encoded venues, and scaled population size of each city (simple scaling). This resulted in 16 features for the model.

The value k (number of clusters) was varied from 1-15 to find a best-value to use, following the standard guidelines to look for the “elbow” of the accuracy curve.



While no **clear** “elbow” was present (it seemed more like a smooth curve on visual inspection of the above graphe), for the sake of increased accuracy, **I chose a value of k=10**

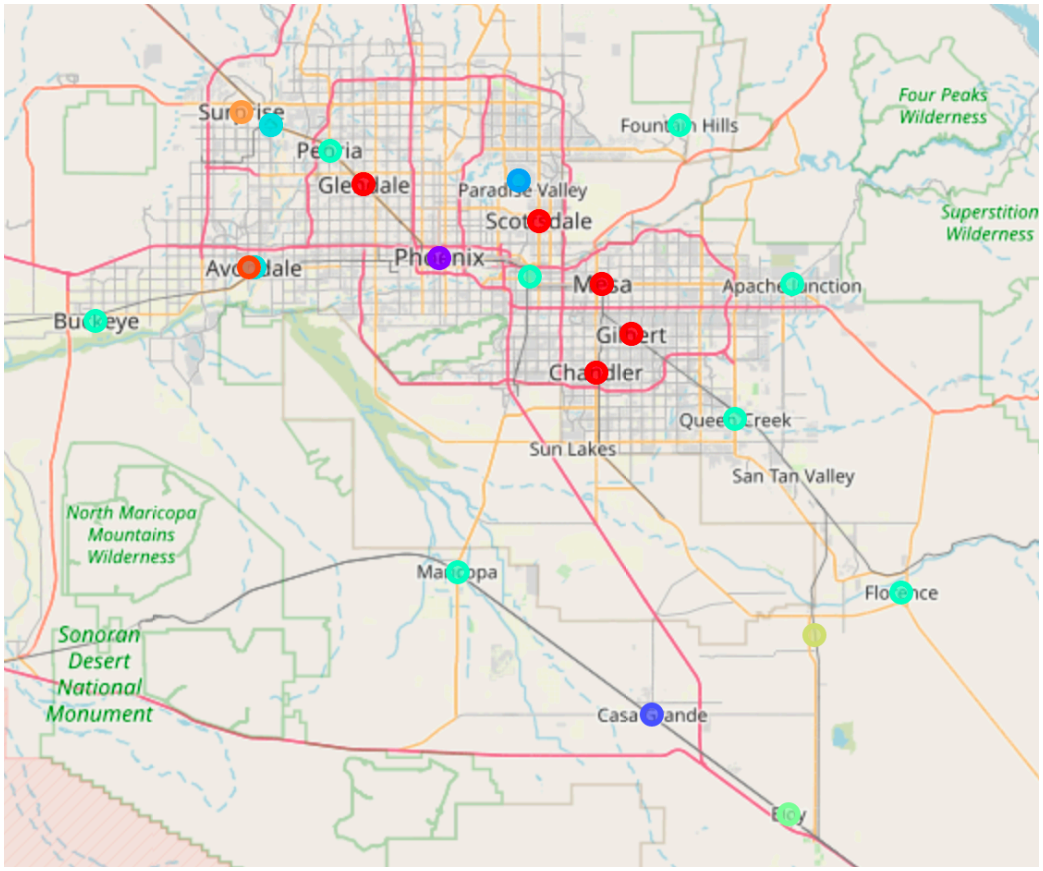
Clustering Model Results

As a quick manual verification of the cluster modelling, the top 2 most common venues were printed out for all members of each cluster, with the **target city being grouped into Cluster 5**:

```
for Cluster  0
1st:  ['Mexican Restaurant' 'Mexican Restaurant' 'Coffee Shop'
 'Mexican Restaurant' 'Mexican Restaurant']
2nd:  ['Sandwich Place' 'Taco Place' 'Bar' 'Bar' 'Plaza']

for Cluster  1
1st:  ['Coffee Shop']
2nd:  ['Hotel']

for Cluster  2
1st:  ['Construction & Landscaping']
2nd:  ['Pizza Place']

for Cluster  3
1st:  ['Football Stadium']
2nd:  ['Pawn Shop']

for Cluster  4
1st:  ['Mexican Restaurant' 'Mexican Restaurant']
2nd:  ['Bakery' 'Convenience Store']

for Cluster  5
1st:  ['Mexican Restaurant' 'Coffee Shop' 'Fast Food Restaurant'
 'Mexican Restaurant' 'American Restaurant' 'Fast Food Restaurant'
 'Mexican Restaurant' 'Mexican Restaurant' 'Grocery Store']
2nd:  ['Clothing Store' 'Pizza Place' 'Theater' 'Discount Store'
 'Fast Food Restaurant' 'Pharmacy' 'Pizza Place' 'Sandwich Place'
 'Pharmacy']

for Cluster  6
1st:  ['Convenience Store']
2nd:  ['Grocery Store']

for Cluster  7
1st:  ['Pizza Place']
2nd:  ['Mexican Restaurant']

for Cluster  8
1st:  ['Baseball Field']
2nd:  ['Baseball Stadium']

for Cluster  9
1st:  ['Burger Joint']
2nd:  ['Fish & Chips Shop']
```

Cluster Map

The clusters were also mapped out, with our target cluster (Cluster 5) shown in red on the map below.



Discussion

Further Communication with client

Based on the custering results, there are still 7 cities in the Phoenix area that were clustered as similar to our target city of Rancho Cucamonga, CA.

At this point in our business model, we would want to discuss with our client and see if there were any other considerations he wanted to consider. We have clustered on Foursquare venues, but are there any other attributes of the Phoenix, AZ, city that "John Doe" would like to keep in common with his current town?

Frequent communication and updating of goals is a necesarry part of producing the desired deliverable. In our hypothetical scenario, **suppose John Doe, our client, responds back to our query with the following additional follow-up criteria**:

> ***"Well, if the models show multiple cities near Phoenix with similar types of venues, can you also find a town that has a similar population and total number of venues? I'm used to having lots of options when I go into town, so I don't want to move somewhere with fewer venues."***

Follow-up additional analysis

Returning to the data, here is the final list of Cluster 5 cities, sorted by population size.

	City	Population	State
19	Fountain Hills	24583	AZ
18	Florence	26074	AZ
16	Queen Creek	39184	AZ
15	Apache Junction	40538	AZ
14	Maricopa	48007	AZ
12	Buckeye	68453	AZ
8	Peoria	168181	AZ
0	Rancho Cucamonga	177751	CA
7	Tempe	185038	AZ

Reviewing the list, the most similar in population size to Rancho Cucamonga, CA are **Peoria, AZ** (slightly smaller population: 168,181) and **Tempe, AZ** (slightly larger population: 185,038).

Returning to our previous results of Number of venues (<2 km from city center), we see that the number of venues are:

City	# of Venues (<2 km)
Peoria	22
Rancho Cucamonga	82
Tempe	100

We can see that even though it is similar in population, Peoria has far fewer total number of venues, contrary to the desires of the client.

As a result, our modeling, combined with further communication of goals with the client, results in a “best-match” result of **Tempe, AZ**.

Conclusion

Our goal in this project was to give a best-match recommendation for a client looking to move from a target town in California to a city in the greater Phoenix, AZ metropolitan area. Using Foursquare venue data in combination with public data about cities near Phoenix, we were succesfully able to provide a recommendation based on venue variety, total number of venues, and population size similarity, in accordance with the client’s wishes.

While I would not recommend that any client uses this modelling as 100% of their decision-making in choosing a new city to move to, sight-unseen, hopefully this would provide our client with a good “starting place” to begin a search. Venue categories, venue number, and city population size all recommend that Tempe, AZ would be a good city to start looking for someone interested in a “similar” feel to that of Rancho Cucamonga, CA.