CAPSTONE PROJECT THE BATTLE OF THE NEIGHBORHOODS

Report by Student: Cynthia

INTRODUCTION

Many companies relocate jobs to improve their profitability position on a regular basis. Job relocation for employees can be an upsetting scenario for them as well as their family. It would be helpful for affected employees who have decided to relocate with job to be able to evaluate the surrounding cities to determine which is most like the current city they live in to determine best city that would cause the least disruption to family and what they are accustom to.

The solution to this problem would be beneficial to any employee whose job is being relocated to a new state. This information can minimize the effects of this level of change to a family's life.

The company's relocation plans will go smoother when employees can make a quick and comfortable adjustment to a new location as well as the employee's family, therefore the employer would benefit as well to this being an offering to their employees.

DATA

The following data will be considered in this project: We will look at the top five cities located a minimum of 10 miles from current job location as well as the relocation destination.

The top five cities will be determined by reviewing this site: https://www.distantias.com.

The location details for all cities used in the project will be found on site:

http://www.lat-long.com

	Town	mi.	Area
0	Newcastle	10.47	California
1	Elverta	11.49	California
2	Rio Linda	11.55	California
3	El Dorado Hills	11.71	California
4	Pilot Hill	12.13	California

	Town	mi.	Area
0	Scottsdale	12.95	Arizona
1	Queen Creek	15.38	Arizona
2	Sacaton	15.56	Arizona
3	Paradise Valley	17.25	Arizona
4	Fountain Hills	19.95	Arizona

This project will look at the Four Square site data to acquire venue details for the top five surrounding cities. Comparisons will be made between the venues found in both cities to determine similarities with the original and job relocation job city.

Sample Venue Data


```
{'meta': {'code': 200, 'requestId': '5e0a3928f7706a001b68ffb4'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'},
    {'name': '$-$$$$', 'key': 'price'}]},
  'headerLocation': 'Downtown Chandler',
  'headerFullLocation': 'Downtown Chandler, Chandler',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 38,
  'suggestedBounds': {'ne': {'lat': 33.310722304500004,
    'lng': -111.83595734802682},
   'sw': {'lat': 33.3017222955, 'lng': -111.84670605197317}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4b8d77aaf964a5206cfd32e3',
       'name': 'Chandler Center for the Arts',
       'location': {'address': '250 N Arizona Ave',
        'crossStreet': 'at E. Chandler Blvd.',
        'lat': 33.30745402671831,
        'lng': -111.84215246969697,
        'labeledLatLngs': [{'label': 'display',
          'lat': 33.30745402671831,
          'lng': -111.84215246969697}],
        'distance': 156,
        'postalCode': '85225',
        'cc': 'US',
        'city': 'Chandler',
        'state': 'AZ',
```

METHODOLOGY

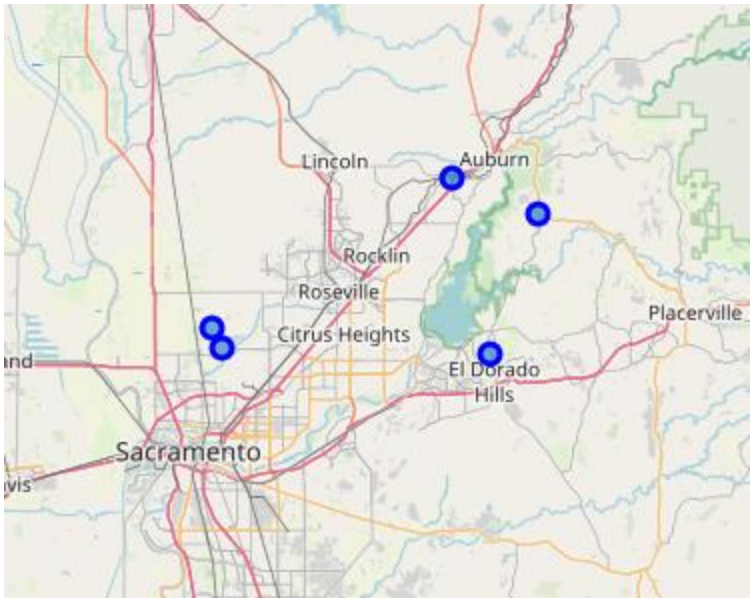
This problem looks at a scenario of an insurance company with jobs relocating from Roseville, CA to Chandler, AZ.

The first step is to map the Roseville location and determine the top five surrounding cites. This is done by scraping information from the site distantias.com. We only look at cities that are a minimum 10 miles from the original job location city.

A data frame is manually created with their location detail for the five cities. Once the location details for each city is determined, this data frame is joined with the top five cities data frame.

	Town	mi.	Area	Latitude	Longitude
0	Newcastle	10.47	California	38.874066	-121.133281
1	Elverta	11.49	California	38.713791	-121.462733
2	Rio Linda	11.55	California	38.691013	-121.448563
3	El Dorado Hills	11.71	California	38.686295	-121.081268
4	Pilot Hill	12.13	California	38.834900	-121.014388

These cities are displayed via a folium map to provide a visual of their vicinity.



These previous steps are repeated for the destination location, Chandler AZ.

Venue data for the surrounding cities is used to find similarities in the surrounding city for the job to and from city. This data is acquired via a search of the Four Square location data. It will provide a listing for the most popular venues in the top five cities.

In the foursquare search screen, we ask for all venues In a 10 mile radius and allow up to 200 results returned. In our example, no city return results above 100.

```
# create the API request URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.

    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    lat,
    lng,
    radius,
    200)
```

The venue data was converted to a onehot coded dataframe where categorical values are mapped into integer values.

	Neighborhood	ATM	American Restaurant	Art Gallery	Asian Restaurant	Auto Dealership	Automotive Shop	BBQ Joint	Bakery	Bank	...	Trail	Train Station	Video Game Store	Vietnamese Restaurant
0	Newcastle	0	0	0	0	0	0	0	0	0	...	0	0	0	0
1	Newcastle	0	0	0	0	0	0	0	0	0	...	0	0	0	0
2	Newcastle	0	0	0	0	0	0	0	0	0	...	0	0	0	0
3	Newcastle	0	0	0	0	0	0	0	0	0	...	0	0	0	0
4	Newcastle	0	0	0	0	0	0	0	0	0	...	0	0	0	0

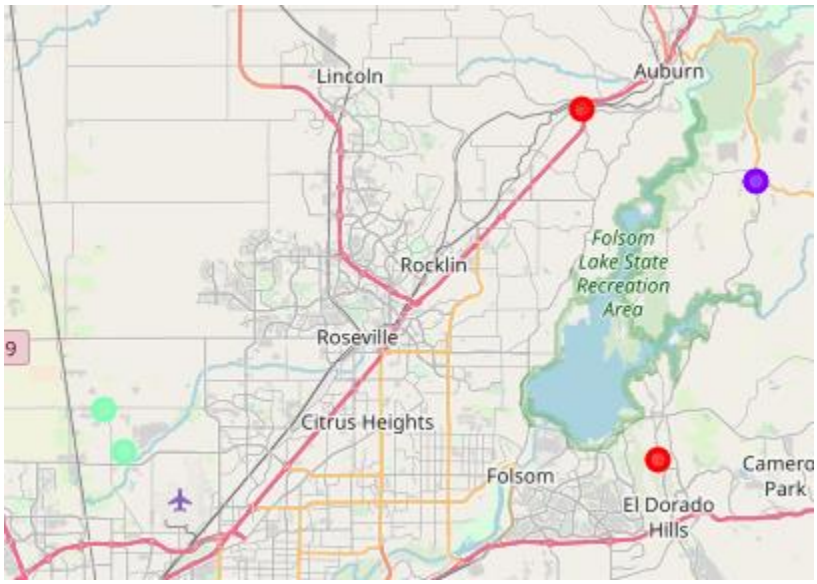
First we look at the top five venues and their frequency in the top five cities for both the to and from job locations.

```
----El Dorado Hills----
                      venue  freq
0               Coffee Shop  0.08
1             Grocery Store  0.05
2           Sushi Restaurant  0.05
3  New American Restaurant  0.04
4                       Gym  0.04
5       American Restaurant  0.04
6               Burger Joint  0.04
7                       Park  0.03
8        Mexican Restaurant  0.03
9                      Café  0.03
```

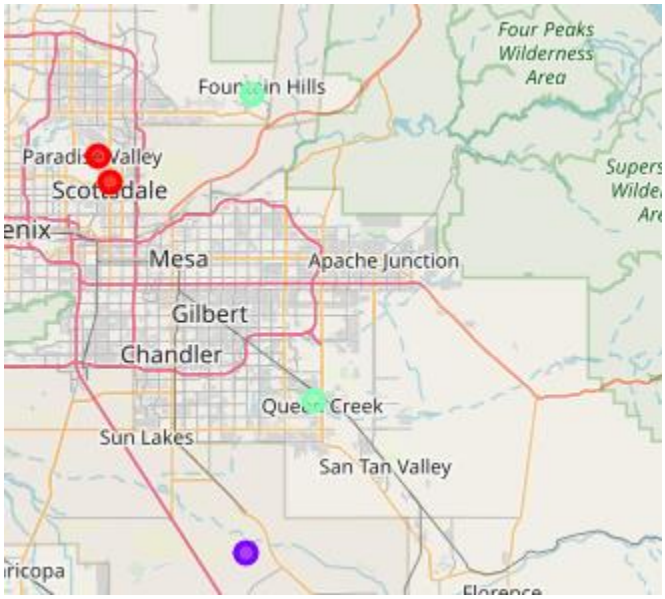
K-means clustering was used since the data is unsupervised. We used k-means to cluster the venue data to determine what types of venues were found and how they are both similar between locations and dissimilar between each other.

The venues results are clustered using k=3. This k-value was randomly chosen. Three clusters where created and displayed on the folium map for both the to and from job location.

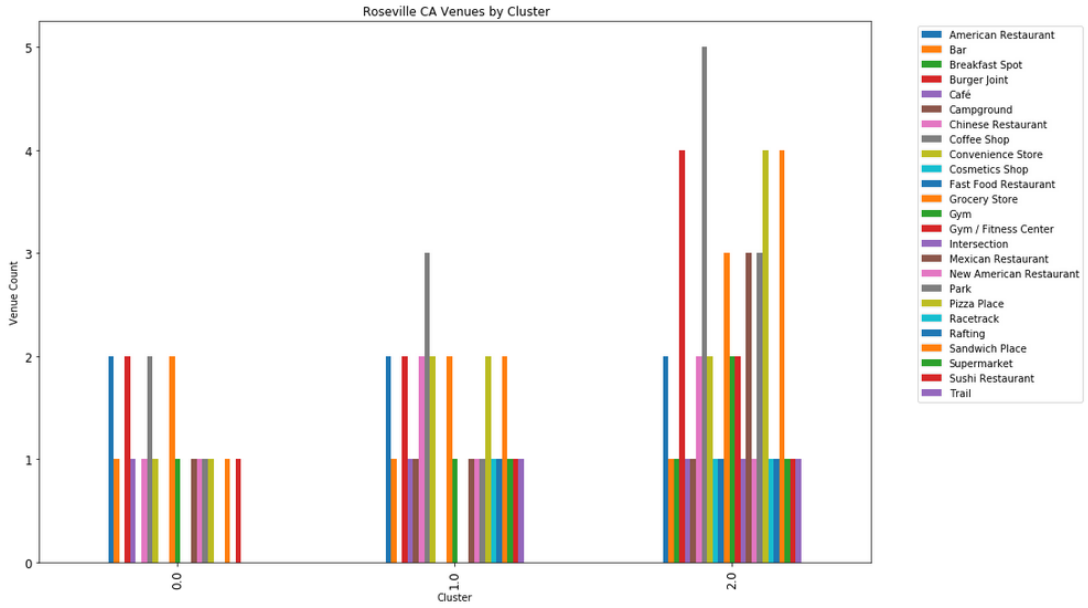
Roseville, CA

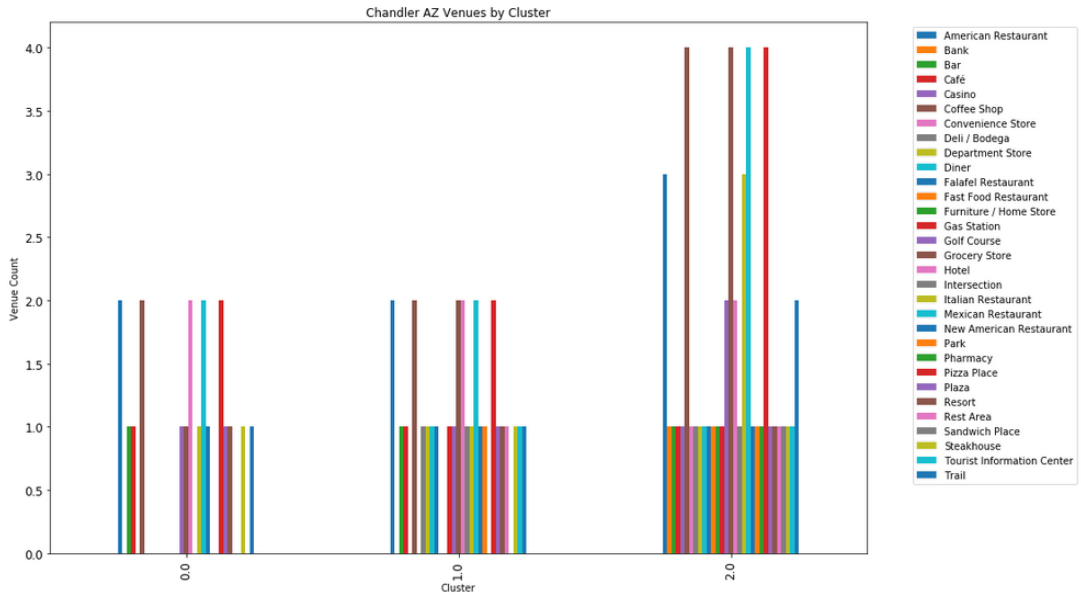


Chandler, AZ



A bar chart was created for each location show the venue categories in each cluster for both the to and from location cities.





RESULTS

The foursquare search returned the following total of results for all surrounding cities of the to and from job location.

Job Location City	Total # of venues	Total # of unique venues
Roseville, CA	404	120
Chandler, AZ	305	115

When we look at venues for each city, they return similar counts of unique categories as well as total count of venues. There are 14 venues in common between the two locations.

When comparing the top ten venues for the individual cities right away we can see that a few venues that don't show up i.e gym, discount store, & deli in the two location so if these are favorite venues for the employee, they may have some challenges.

There will also be some new categories of venues that the employee will have access to i.e. hotels, casinos and steak house that are available in the job location to state.

When reviewing our clusters, the look like the following for our to and from job locations:

<u>Roseville, CA</u>

Cluster 1 Grocery & Gyms & Bars

Cluster 2 Restaurants & Brewery

Cluster 3 Restaurants, Parks, Gyms, Business services

Chandler, AZ

Cluster 1 Resorts/hotels/Tourist venues,

Cluster 2 Pharmacy & wine venues,

Cluster 3 Furniture, Restaurants

DISCUSSION

The analysis show that there are many venues available in the surrounding area of the job relocation city. They also show where there may be challenges for example with looking for a gym location if that is a common venue used by the employee. There are some new venues that the employee may like to take advantage of as well. Form the analysis we can recommend that the employee move in the area between fountain Hills and Queen creek because these would offer the most access to the same venues found in home location, Roseville, CA. There are three clusters identifying possible suitable locations for housings. These areas should be the starting point to look for a new residence. The employee can choose to drill into the venues and exclude venues that are not in their preference list and therefore determine that a cluster from one of the other areas makes the most sense for their family.

CONCLUSION

The goal of this project was to determine best candidate cities to move to in the vicinity of a job relocation in order to minimize the disruption caused to someone as an employee and a financial resource to their family. By looking at foursquare venue information for surrounding cities of the job relocation state, we give employee a visual of possible cities that offer the employee access to the same venues they are accustom to in the area of their home city. The expected outcome is that this will ease the adjustment of a move and quickly acclimate and employee and their family to job relocation state.