WEST VALLEY CITIES FOR TECH PROFESSIONALS

Coursera IBM Data Science Capstone Project

INTRODUCTION TO ANALYSIS

- The West Valley of Phoenix is a rapidly growing region
- Tech companies are locating in the Phoenix region at a high rate
- The West Valley should be consider by tech companies, but understanding what the region has to offer for professionals is important to recruiting efforts
- Identifying the cities that would be most appealing to professionals will help tech companies identify locations for the company.

WHAT FACTORS TO CONSIDER

- Socioeconomic Data
  - Median Household Income
  - Median Housing Value
  - Education Levels
- Social Factors
  - Crime Rates
  - Available Entertainment & Recreation

FINDING THE DATA

- Three primary sources were utilized for this analysis.
  - The US Census Bureau Fast Facts provided socioeconomic and demographic data for each city in the region
  - The FBI Crime Reporting system provided data on crime rates in each city from 2019
  - FourSquare offered data regarding popular venues in the area.

WRANGLING THE DATA

- US Census Data provided 60+ data points
  - Culled that data set to the most salient
    - Median Housing Value
    - Median Household Income
    - Bachelor's degree or higher %
    - % in the Workforce

- This data set required cleaning to make data types appropriate, syntax accurate, eliminate unnecessary data, etc.

WRANGLING THE DATA

- The crime data from the FBI also required some cleaning.
- One city in the original definition of the West Valley only had estimates of crime rates and these were 300% higher than every other city in the region.
  - As a result of the skew in this data set and the fact that they were not actual data, but estimates, this city was dropped from the analysis

- The remaining data points across both the Census and FBI data were within normal distributions.

VENUE DATA

- FourSquare provided data on the top 100 venues in each city, which was then reduced to the 10 most common venues in that top 100.

- While there was variability in the venues, there were more similarities across the 9 cities.

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
Avondale	Mexican Restaurant	Pizza Place	Fast Food Restaurant	American Restaurant	Convenience Store	Japanese Restaurant	Grocery Store	Coffee Shop	Pharmacy
Buckeye	Mexican Restaurant	Pizza Place	Gas Station	Fast Food Restaurant	Sandwich Place	Discount Store	Video Store	Convenience Store	Café
El Mirage	Mexican Restaurant	Fast Food Restaurant	Grocery Store	Convenience Store	Coffee Shop	Pharmacy	Sandwich Place	Ice Cream Shop	American Restaurant
Glendale	Mexican Restaurant	Sandwich Place	Park	Convenience Store	Grocery Store	Fast Food Restaurant	Bar	Candy Store	Pizza Place
Goodyear	Mexican Restaurant	Pizza Place	American Restaurant	Japanese Restaurant	Convenience Store	Seafood Restaurant	Bar	Fast Food Restaurant	Pharmacy
Litchfield Park	Coffee Shop	Mexican Restaurant	Pizza Place	Grocery Store	Gym	Pharmacy	American Restaurant	Japanese Restaurant	Golf Course
Peoria	Pizza Place	Chinese Restaurant	Mexican Restaurant	Grocery Store	Convenience Store	Fast Food Restaurant	Gym / Fitness Center	Cosmetics Shop	Coffee Shop
Surprise	Pharmacy	Coffee Shop	Pizza Place	Grocery Store	Fast Food Restaurant	Mexican Restaurant	Sandwich Place	American Restaurant	Italian Restaurant
Youngtown	Pizza Place	Mexican Restaurant	Pharmacy	Fast Food Restaurant	Convenience Store	Golf Course	Discount Store	Video Store	Intersection

K-MEANS CLUSTERING

City	Owner Housing Rate	Median Housing Value	% Bachelors	% in Workforce	Median Household Income	Total Crime per 10,000	Violent Crime per 10,000	Property Crime per 10,000
Avondale	0.54	196000	0.18	0.68	58938	355	34	321
Goodyear	0.75	274500	0.30	0.57	80336	278	25	253
Litchfield Park	0.83	294900	0.49	0.56	80988	308	73	235
Peoria	0.73	248700	0.32	0.64	73039	213	23	190

City	Owner Housing Rate	Median Housing Value	% Bachelors	% in Workforce	Median Household Income	Total Crime per 10,000	Violent Crime per 10,000	Property Crime per 10,000
El Mirage	0.62	151500	0.15	0.69	54646	283	22	261
Surprise	0.75	225200	0.27	0.56	65160	166	12	154
Youngtown	0.52	141800	0.13	0.58	41441	485	134	350

- First K-means included both socioeconomic & venue data, which resulted in groupings that showed too much within-group variability.
- As the charts show, there is wide variability within the groups on multiple domains.
- As a result, a second K-means was run without the venue data.

K-MEANS CLUSTERING

City	Owner Housing Rate	Median Housing Value	% Bachelors	% in Workforce	Median Household Income	Total Crime per 10,000	Violent Crime per 10,000	Property Crime per 10,000
Avondale	0.54	196000	0.18	0.68	58938	355	34	321
Buckeye	0.71	196800	0.19	0.55	65932	207	18	189
Goodyear	0.75	274500	0.30	0.57	80336	278	25	253
Surprise	0.75	225200	0.27	0.56	65160	166	12	154

City	Owner Housing Rate	Median Housing Value	% Bachelors	% in Workforce	Median Household Income	Total Crime per 10,000	Violent Crime per 10,000	Property Crime per 10,000
Glendale	0.55	197600	0.21	0.62	52314	454	47	408
Peoria	0.73	248700	0.32	0.64	73039	213	23	190

City	Owner Housing Rate	Median Housing Value	% Bachelors	% in Workforce	Median Household Income	Total Crime per 10,000	Violent Crime per 10,000	Property Crime per 10,000
El Mirage	0.62	151500	0.15	0.69	54646	283	22	261
Youngtown	0.52	141800	0.13	0.58	41441	485	134	350

- While there is still some areas of variability within the groups, there is far greater consistency than in the first K-means clustering.
- Each cluster can be meaningfully described:
  - Avondale Cluster – middle socioeconomic status, slightly more family-friendly venues
  - Glendale Cluster - middle socioeconomic status, slightly greater young, urban venues
  - El Mirage Cluster – Lowest socioeconomic status of all the cities in the region

CONCLUSIONS

- The data selected for clustering can skew the groupings and must be examined to coherence
- Cities that are more likely to attract tech talent can be identified and this information should be part of how tech companies select areas for their offices.
- Other data that could be useful in future analyzes would include school system quality, tax rates, and cost of living