The Best City for Tech Sector Professionals
In the West Valley of Phoenix, Arizona

Introduction

The technology sector is wide reaching industry with products and services spanning most aspects of business and personal life. This sector continues to grow with the "Big 5;" Google, Amazon, Apple, Facebook, and Microsoft, leading the sector in growth and innovation. However, there are thousands of other tech sector businesses of all sizes driving additional growth and innovation. As these businesses continue to grow and recruit talent in a marketplace with nearly full employment, they must offer a compelling employee value proposition (EVP) to potential candidates. One aspect of that EVP are the options available for candidates to relocate.

For a tech company in the West Valley of Phoenix, presenting a compelling case for relocating to the region could be challenging. Generally, tech sector companies chose the East Valley of Phoenix, including Scottsdale, Tempe, and Mesa. However, projections show that almost half of the growth in the region will occur in the West Valley (https://www.westmarc.org/about-westmarc/). As such, tech companies in the West Valley must be able to articulate to candidates the benefits of relocating to the West Valley of Phoenix.

Factors that may be considered when determining desirable living locations include cost of living, demographics, crime rates, and activities in the area. Areas that are more likely to attract top talent in the tech sector will likely have a wide range of activities and lower crime rates. The demographics are most likely to be working age adults, both with and without families, and housing costs are likely to be mid- to upper-range. Locating near regions with high quality of life, reasonable living costs, and a lively social environment will help tech sector companies in the West Valley attract top talent to the region. This project will determine which of ten cities in the region offer the greatest benefits to tech sector professionals for the purpose of supporting recruitment marketing.

Data Plan

Defining the West Valley

To begin data compilation, it was important to define the West Valley of Phoenix. Several sources offer lists of the communities considered to be part of the West Valley. The best-defined list is from WESTMARC, the Western Maricopa Coalition. They identify 15 cities and communities in the West Valley. For the purposes of this project, 10 communities and cities located near major highways, including I-10, Loop 101, and Loop 303 were included. The remaining 5 communities and cities were excluded due to either being too great a distance from Central Phoenix or being too small of a community to have separate demographics available through the US Census Bureau. The 10 cities included in this analysis are Avondale, Buckeye, El Mirage, Glendale, Goodyear, Litchfield Park, Peoria, Surprise, Tolleson, and Youngtown.

Sources of Data

Multiple sources of data were utilized to understand each of the 10 communities in the West Valley. Demographic data was downloaded from the US Census Bureau Fast Facts website. This data includes population, median household income, median housing price, population in the

workforce, and a wide range of additional data points. Crime data was gathered from the FBI Crime Report available here. Finally, venue data in the region was gathered from the foursquare API.

Selecting the Data

The data sources noted above provided a great deal of data, which needed to be culled for the purposes of this project. In reviewing the US Census data, several key data points lined up with the purpose of this project and others could be dropped. The data points that were included in this analysis were the estimated 2019 population of each city, the percent of households in each city occupied by the owner, the median housing value, the percent of the population in each city with at least a bachelor's degree, the percent of the community over the age of 16 in the workforce, and median household income. Each of these data points are common predictors of the socioeconomic status of a community. Given that this project is intended to identify the cities the tech professionals would be most likely to find desirable, the factors related to socioeconomic status are the best fit.

Methodology

Exploratory Analysis

The initial step in analyzing the data was an exploratory data analysis. The goal of this analysis was to gain an understanding of the basic features of the data. In particular, I was examining the data for normal distributions and outliers. Understanding these features of the data allowed me to determine if any further wrangling was necessary prior to attempting to cluster the cities.

As a result of this review, one of the original cities included in the definition of the West Valley was dropped from the analysis due to an extreme skew in crime rates. The city of Tolleson's crime rates were more than 600% greater than any other city in the region. In addition, the FBI crime reporting data indicated that these were estimates. As a result of the incredibly large skew and the fact that the data was estimates, I decided to drop Tolleson from the analysis. The resulting data set is normally distributed for most variables, with some skew still remaining due to another city (Litchfield Park) with a comparatively high crime rate.

Machine Learning

For this project, I completed two K-means clustering algorithms. The first K-means was completed using the complete data set, inclusive of both demographic and venue data. This K-means was completed to identify how the cities cluster based on both the socioeconomic features and the social/entertainment features, as both are import to candidates in selecting a city for relocation. In reviewing the clusters, there were some groupings that did not make intuitive sense based on the socioeconomics of the cities, which will be discussed below. As such, I opted to run a second K-means using only the socioeconomic data for the nine cities. This analysis resulted in 4 clusters that were better matched on the socioeconomic data and still have strong similarities in popular venues. The results of the second K-means clustering are discussed in the results section below.

Results

The initial K-means clustering resulted in clusters with wide variability in important socioeconomic information. As a result, the clusters are not helpful to identify which West Valley cities are most likely to appeal to tech professionals. For example, one cluster grouped the two lowest socioeconomic status cities, El Mirage and Youngtown, with one of the middle-income cities, Surprise. Upon review of the cluster, it appears that the popular venues for the three cities outweighed the socioeconomic factors. While the available options for recreation in a region are an important factor for many professionals, the overlap between all the cities and several of the most popular venues was such that it was reasonable to run a second K-means with only the socioeconomic data.

Upon review the results of this second K-means clustering, the city groupings are better aligned on socioeconomic factors and present a clear picture of each cluster. The results of the second K-means cluster are included the table below.

Table 1: Socioeconomic Data for Each Cluster

Cluster	Included Cities	Median Housing Value	Median Household Income	Bachelors %	Workforce %	Crime Rate per 10,000
1	Litchfield Park	$294,900	$80,988	49%	56%	308
2	Avondale Buckeye Goodyear Surprise	$223,125	$67,592	24%	59%	252
3	Glendale Peoria	$223,150	$62,677	27%	64%	299
4	El Mirage Youngtown	$146,650	$48,044	14%	64%	384

From these results, there are 4 primary types of cities in the West Valley region. The first is Cluster 1, which includes only one city, but is the highest socioeconomic status city in the region. This cluster has the highest median house value, highest median household income, and the highest rate of citizens with at least a bachelors degree. Clusters two and three are very similar on socioeconomic factors and represents middle socioeconomic status cities. The key difference between these two clusters are the popular venues, which are distinguished by more family-oriented venues such as American and Mexican restaurants, pizza places, ice cream shops, video stores, and fast food options. The third cluster has many of the same types of venues, but often ranked somewhat lower. In addition, the third cluster also included parks, gyms, bars, and cosmetics stores. The final cluster includes the two cities ranking lowest on socioeconomic factors, including household income, housing value, and education level.

Discussion

There are two key questions to be discussed from the above results. The first question centers on what factors impacted the clustering most and how does this inform future analyzes of this nature. The second question is which cities would be the most appealing for tech companies to use to attract new talent to the region.

In terms of informing future analyzes, it is evident that venues can skew clustering efforts that include limited socioeconomic features. It would be worthwhile, in future analyzes, to either include additional socioeconomic factors or run the K-means clustering on just the socioeconomic factors. Future residents are going to be more likely to select a city based on factors related to housing values, crime rates, and demographics similar to themselves, than on factors such as whether gyms are more popular venues in the area than pizza places. In areas with limited recreational options, available venues may rise higher on the list of important considerations. However, the West Valley has a wide selection of eateries, entertainment, and recreation. In similar regions, the socioeconomic and demographic data will be more salient to the analysis and venue information can add additional context.

Regarding the second question, and purpose of this project, the second cluster of cities offer the best case for attractive cities to relocate to when recruiting for tech professionals. The four cities in the second cluster – Avondale, Buckeye, Goodyear, Surprise – have an affordable median household value and the second highest median household income in the region. In addition, these cities have the lowest crime rates and a high concentration of family-friendly recreational and entertainment venues. They are all located within easy driving distance of Central Phoenix, but also somewhat outside of the heart of the metro region. All four cities are near major highways, making any commute to a workplace easy. In addition, each city has green space and is located near outdoor activities. Based on the results of this analysis, I would recommend to tech companies considering the West Valley for office space to locate near these four cities and use the amenities and demographics of these cities to help recruit talent to the company and the region.

Conclusion

In this analysis, I considered what cities in the West Valley of Phoenix would be appealing to tech professionals and tech companies seeking a new location. I gathered data from the US Census Bureau, the FBI Crime Reporting, and FourSquare. In wrangling the data, I identified a significant outlier, which resulted in dropping one city from the analysis. In addition, I ran an initial K-means clustering, which gave a result that did not make intuitive sense in terms of how the cities grouped on socioeconomic factors. As a result, I ran a second K-means using only the socioeconomic data and considered the venue data as a secondary source to better understand each cluster. This resulted in a significantly better clustering that allowed for a clear distinction between the clusters. After considering the result, while many of the cities of the West Valley have much to offer tech professionals, it is evident that the second cluster offered more than the others. The final outcome is that I would recommend tech companies considering locations in the West Valley look at locations in or near the four cities in this cluster - Avondale, Buckeye, Goodyear, Surprise.