Applied Data Science Capstone
By Christopher James

Ideal Tech Investment in American Cities

Introduction

It is common knowledge that Silicon Valley is the premier location for all things tech. Thanks to economic agglomeration effects, this is not going to change anytime soon. Despite this, there are a rising number of tech companies, from startups to established titans, along with investors, who are looking for alternatives to Silicon Valley. An example of this is AOL founder Steve Case's Rise of the Rest Fund. Some of these companies are looking to move, while others may just want to open regional locations.

Problem

In this analysis, I will be looking at the number of tech startups in the top 100 biggest American cities, along with the Zillow home value indexes as a way to gauge cost of living. I will also list the percentage of people with bachelor's degrees or higher, because this

industry requires a skilled, educated workforce. I will use this data to try to discover the ideal place to focus investment. Due to the aforementioned agglomeration effects, the ideal place has a high number of tech startups with a low home value index, along with a high percentage of college grads.

Data acquisition and cleaning

Data Sources

The data sources I will be using are as follows: For the tech startup numbers, I will be using the foursquare api, as linked here. For the housing data, I will be using the zillow home value index, as linked here. For the college degree numbers, I will be using this site.

Data Cleaning

I combined data from a few sources into one dataframe. First, I scraped the website List of the 100 largest cities ranked by Educational attainment - persons 25 years and over - percent bachelor's degree or higher (indexmundi.com) for the names of the cities and the percentage of adults over the age of 25 that have bachelor's degrees. I would need the state names for the foursquare api and for narrowing down cities with the same name when finding metro areas. However, since some cities featured in multiple states, I had to manually enter the information after looking each one up.

Next, I read two different csv files into dataframes. One contained the zillow home value index for all American cities. The other contained the same, except for all American metropolitan areas. They both came from Housing Data - Zillow Research.

After that, I matched the cities to the right metro area by comparing the city names and states and added that to the dataframe.

Then I added housing costs by matching the names of the metropolitan areas.

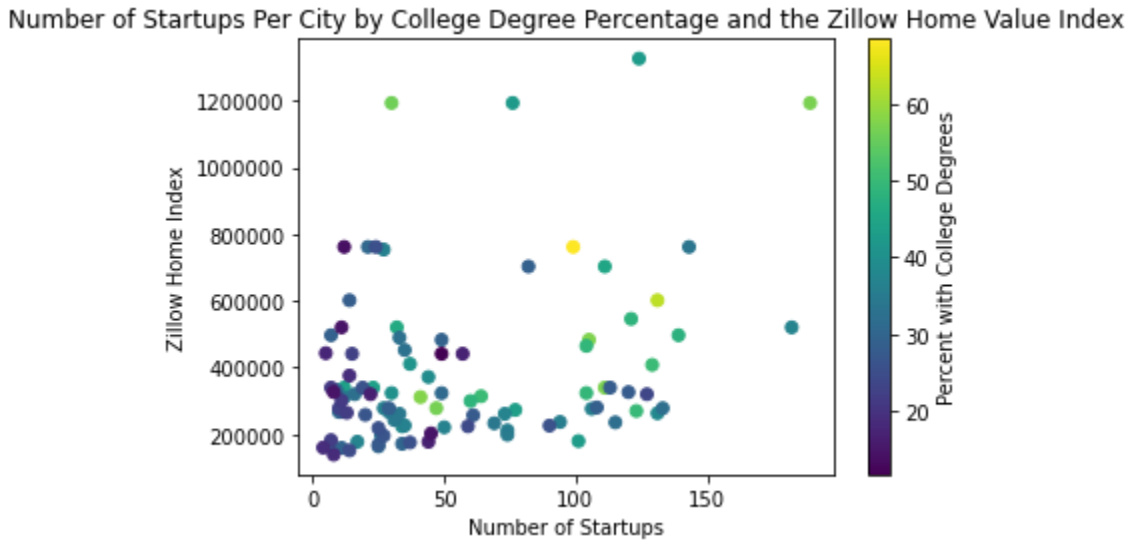
Finally, I used the Foursquare api to download the number of tech startups in each city and added that to the dataframe.

This is a snapshot of what the final dataframe looked like:

	City	State	HomePrice	StartupNumber	DegreePercent	Metro
0	Irvine	CA	761635	99	68.5	Los Angeles-Long Beach-Anaheim
1	Seattle	WA	601735	131	62.8	Seattle-Tacoma-Bellevue
2	Madison	WI	311633	41	57.9	Madison
3	Washington	DC	483254	105	57.6	Washington-Arlington-Alexandria
4	Scottsdale	AZ	339878	111	57.2	Phoenix-Mesa-Scottsdale

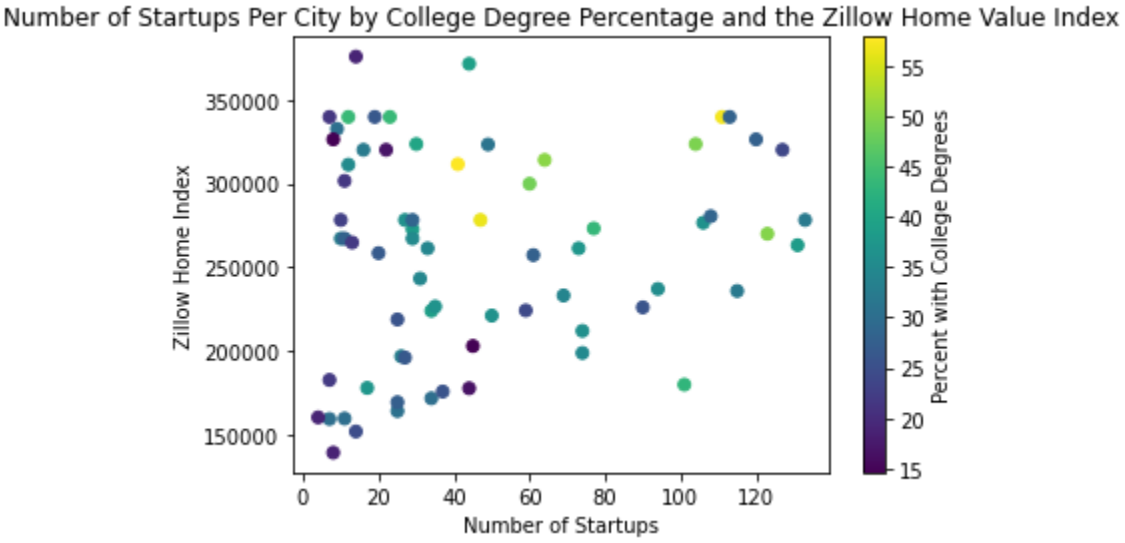
Data Analysis

To get a feel of the overall look of the data, I first chose to plot all of it in a scatterplot.



As expected, most cities are not tech hubs. Most have somewhat lower levels of college graduates, along with lower housing costs and lower startup levels

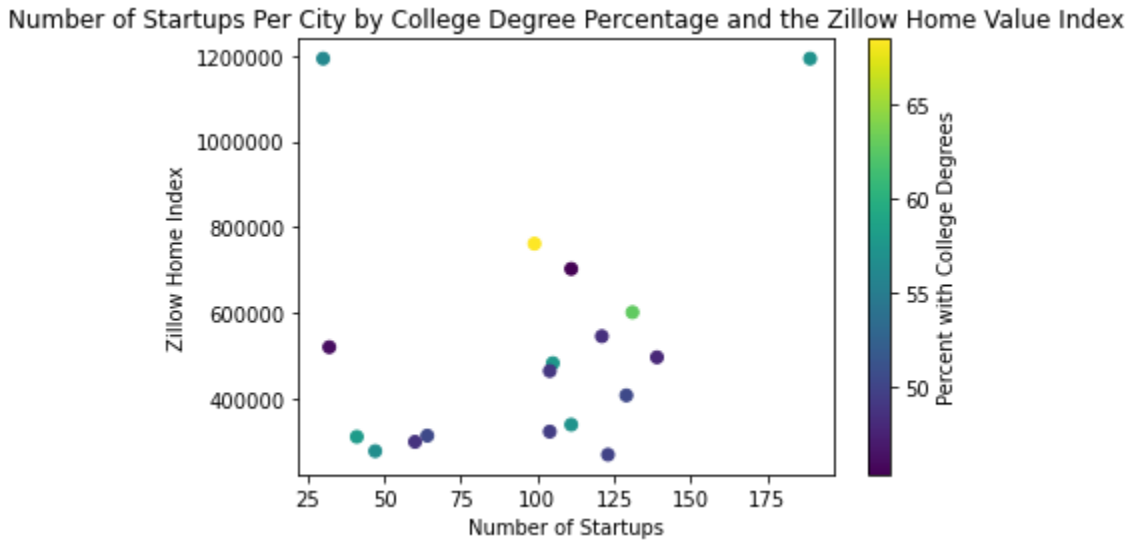
After that, I looked at cities where the Zillow home value index was less than 400000 dollars:



Above: Cities with home values less than 400000 dollars

Cities with homes that have a median value of less than 400k dollars typically have lower numbers of tech startups. So, while there are many cities in this cluster that have somewhat higher levels of college graduates, the low tech startup levels indicate that they may focus on other industries than tech.

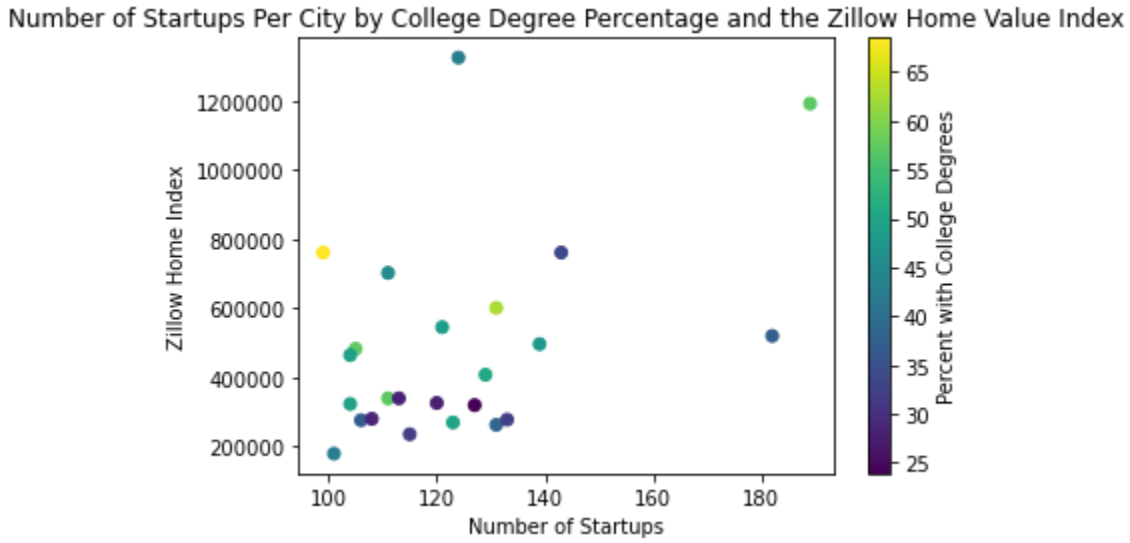
I then created a scatter plot only of cities where at least 45% of adults over the age of 25 had at least bachelor's degrees:



Above: Cities where at least 45% of adults over the age of 25 have at least a bachelor's degree

There are fewer cities in the cluster made up of cities that have at least 45% college graduates than the one with lower housing costs. While there are a few here with higher levels of tech startups, they also have higher housing costs. This may indicate that these markets are already hot, and might not unlock the best value for employers or employees.

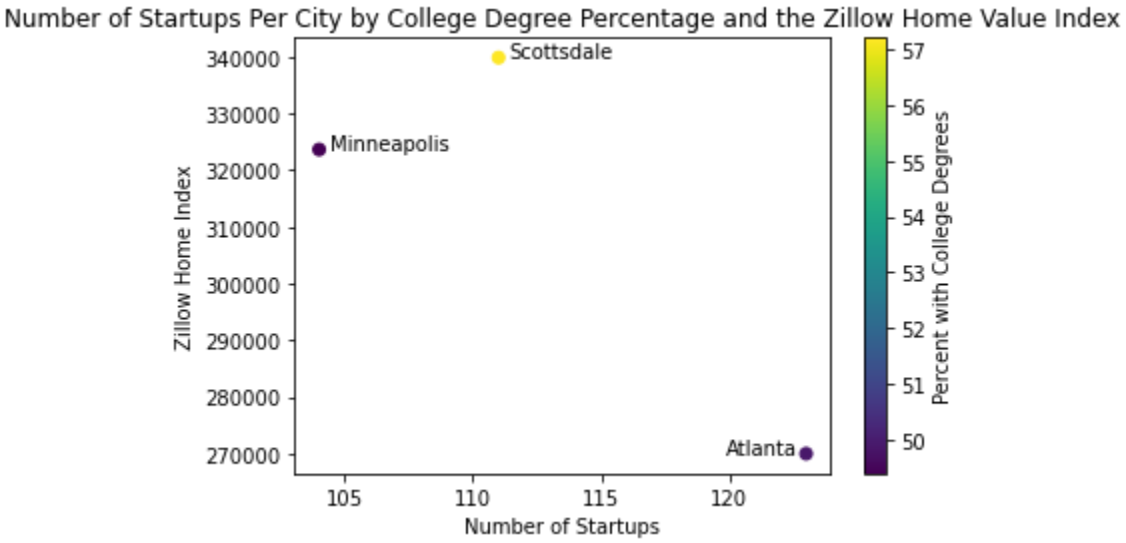
Then, I created a scatter plot containing only cities that had at least 99 tech startups:



Above: Cities that had at least 99 tech startups.

Some cities that have higher numbers of tech startups tend to have lower housing costs, but they also tend to have a very low percentage of people with college degrees. This might indicate problems recruiting top employees to the city not only because of lack of cultural fit, but because of a possible lack of alternative employment opportunities.

Finally, I took a look at only cities that had all 3 of these requirements:



Above: Cities that met all 3 of the above requirements.

Conclusion

The purpose of this project was to identify alternative American cities that would be a good fit for companies to move or venture funds to invest in. The features that were used to perform this analysis were the number of tech startups in the city, the percentage of adults over 25 with at least a bachelor's degree, and the zillow home value index for the metropolitan area that contains the city. Interested stakeholders may also want to take other things into consideration, such as climate, transportation and cultural amenities. They may also assign different weights to

these values depending on their own needs. That being said, My personal recommendation would be Atlanta. While it may have a relatively lower level of college graduates than the other 2, the increase in tech startups and the significantly lower cost of housing make it worth it.