**Fastest Growing Cities in the U.S.**

Tom Allen

02/07/2020

IBM Data Science Certification Capstone

1. **Introduction**
   1. **Background**

There has never been an easier time to move and experience a different area, whether it’s relocating for a different job, starting a new business or moving for fun. Technology has allowed users to explore areas much easier and quicker all while sitting in the comfort of their own home.

**1.2 Problem**

As cities become denser, problems start to occur including lack of infrastructure to support the masses, increase in housing costs and outdated public transportation. I wanted to see what cities are experiencing the most growth and what characteristics these areas have. For entrepreneurs, re-locating to a fast-growing city could be the kickstart they need to get their business going in the right direction.

1. **Data**
   1. **Data Collection**

I was able to find a public [dataset](https://public.opendatasoft.com/explore/dataset/1000-largest-us-cities-by-population-with-geographic-coordinates/export/?sort=-rank) that comprised of the 1,000 largest cities in the U.S. This dataset included City, Rank, State, Growth from 2003 to 2013, Population and each city’s geographical coordinates. The first 5 rows of this dataset are displayed in Figure 1.

***Figure 1***

First 5 rows of the dataset


I was also able to download a .csv file that had contained a state’s region that I could eventually merge with the first dataset. This would be helpful to find what regions of the U.S. were growing the fastest. The first 5 rows of this dataset are displayed in Figure 2.

***Figure 2***

A screenshot of a cell phone

Description automatically generated

* 1. **Data Cleaning**

The datasets were relatively clean to begin with and had no missing values. The only cleaning that was done was for formatting purposes to get the data in a format that I could easily analyze.

First, I had to separate the Coordinates column into Latitude and Longitude because it would be much easier to call a city’s geographical location if they were separated columns.

Second, I merged the dataset containing the city’s growth and coordinates with a dataset that contained each states region. I joined them on the key: “State”.

Third, I created two different subsets of the merged data frame to get 1) data frame containing the top 10 most populous cities and 2) data frame containing the top 10 fastest growing cities. This would allow me to compare these city’s locations and see if there were any patterns between them. See the top 10 fastest growing cities data frame in Figure 3.

**Figure 3**

A screenshot of a cell phone

Description automatically generated

1. **Exploratory Data Analysis**
   1. **Plotting the Fastest Growing States**

I first made a quick bar chart, see Figure 4, that displayed the fastest growing states, so the viewer would not have to manually check each state’s growth rate. This plot grouped by each state and aggregated the mean of each city’s growth rate. The viewer could then clearly see that Arizona, Idaho, Nevada and Texas were the fastest growing states.

***Figure 4***

A picture containing object, antenna

Description automatically generated

**3.2 Plotting the Fastest Growing Regions**

I made another bar chart, see Figure 5, that displayed what regions were growing the fastest. This time, I grouped by region and aggregated the mean of each city’s growth rate. This chart displayed the Mountain region had the fastest average growth rate, with West South Central in second.

***Figure 5***

A screenshot of a social media post

Description automatically generated

* 1. **Plotting Using a Folium Map**

Using a Folium Map, I could plot each city on a map to clearly see the locations or areas the fastest growing or most populous cities were located in, see Figure 6. I plotted the most populous cities with blue dots and the fastest growing cities with red dots. The map could easily display the locations of each of these cities on a readable U.S. map. It clearly showed, most of the fastest growing cities were located in the Mountain region and California. One interesting inference from this chart, was that most of the fastest growing cities were located just outside or near major cities.

***Figure 6***

A close up of a map

Description automatically generated

1. **Getting City Data**
   1. **Foursquare**

Using the Foursquare API, I could extract data from nearby venues around these city’s coordinates. I could return venue names, tips, frequency of visits, etc. I first had to set up an account in order to get client identification and client secret numbers in order for the API call to work.

* 1. **Getting Nearby Venues**

Using a function that used the Foursquare API, I was able to pass in all the fastest growing city’s coordinates to get venues nearby. This function returned a data frame, see Figure 7, consisting of the city, location and venues that were in a 500-meter radius. For each venue, it had its official name and a category it belonged to (i.e. Domino’s Pizza and Pizza Shop).

***Figure 7***

A screenshot of a cell phone

Description automatically generated

* 1. **Venue Category Frequency**

Instead of looking at the specific venue names, grouping by the category they belonged to could tell a much better story about each city. After encoding each venue by the frequency of it being nearby to that city, I was able to get tables that displayed the most frequent categories of venues for each city. This gave me a quick idea of which types of venues were most popular in that city. Figure 8, shows the first 2 cities and the most frequent venues.

***Figure 8***

A screenshot of a cell phone

Description automatically generated

1. **Conclusion**

Through this analysis, I was able to gather valuable information regarding city growth and what types of venues surround these areas. This could be valuable to anyone looking to move or an entrepreneur looking to start a new business.

One observation was that the Mountain region was the fastest growing area and Arizona was a big part in that growth. Are people wanting to move to a warmer climate or is businesses starting to re-locate to this area?

Another observation was that the fastest growing cities were located around the most populous cities. Personally, I expected this because after living in major city, traffic is a major problem, housing prices are increasing to low supply and more demand and the aging infrastructure is not able to keep up with the amount of people living there. It seems people are starting to move just outside of the city, to have a better cost of living or to get away from the clustering of people.

1. **Recommendations**
   1. **Opening a Mexican Restaurant**

It is clear that in the Mountain region, Mexican Restaurants are extremely popular and there are many cities that are growing fast giving the business an opportunity to succeed.

* 1. **Moving Outside of Major Cities**

As noted in conclusion section, some of the fastest growing cities tend to be near major cities for reasons that may include cost of living, congestion, city taxes, etc. By getting a head start and moving to a faster growing city, a homeowner can by for cheap, and experience growth in the property’s value in order to make a solid investment.