

Purple State Plunge: A Deep Dive into Pennsylvania's Cities

Joseph J. Kamedze III
Coursera Capstone Project

Data

The data used for this project consists of census data from a website hosted by the state of Pennsylvania, a database of location coordinates for each city in Pennsylvania, and Foursquare location data. The census data was pulled from the website, "data.pa.gov", and consists of population data for all municipalities in Pennsylvania from the 2010 census. This database also contains population estimates of each municipality from 2010 to 2018. The data from this source, which contains population data for 2201 municipalities, will be filtered to focus only on the state's 57 cities. This data will be used to calculate each city's population growth over 2 and 10 years, the overall growth rate, and the growth rate for each year. From this data, a histogram was created and shows the population growth rate from 2010-2018 as shown in Figure 1.

The data for the city coordinates exists in an Excel spreadsheet and was not readily available in database format online. The website, "data.pa.gov", provides a .json file that contains shape files for each of the state's municipalities, but that file only contains coordinate data for the municipality boundaries, rather than coordinates for the city itself. As a result, the database was required to be built manually by inputting the coordinate values onto an Excel spreadsheet. That spreadsheet was then uploaded to GitHub. A Folium map of the cities using this location database is shown in Figure 2.

Finally, location data from Foursquare was used to pull the most common venues for each city. A 5-mile radius was around each city's location coordinates was used to search for venues, and the top 10 most common venues were returned for each city.

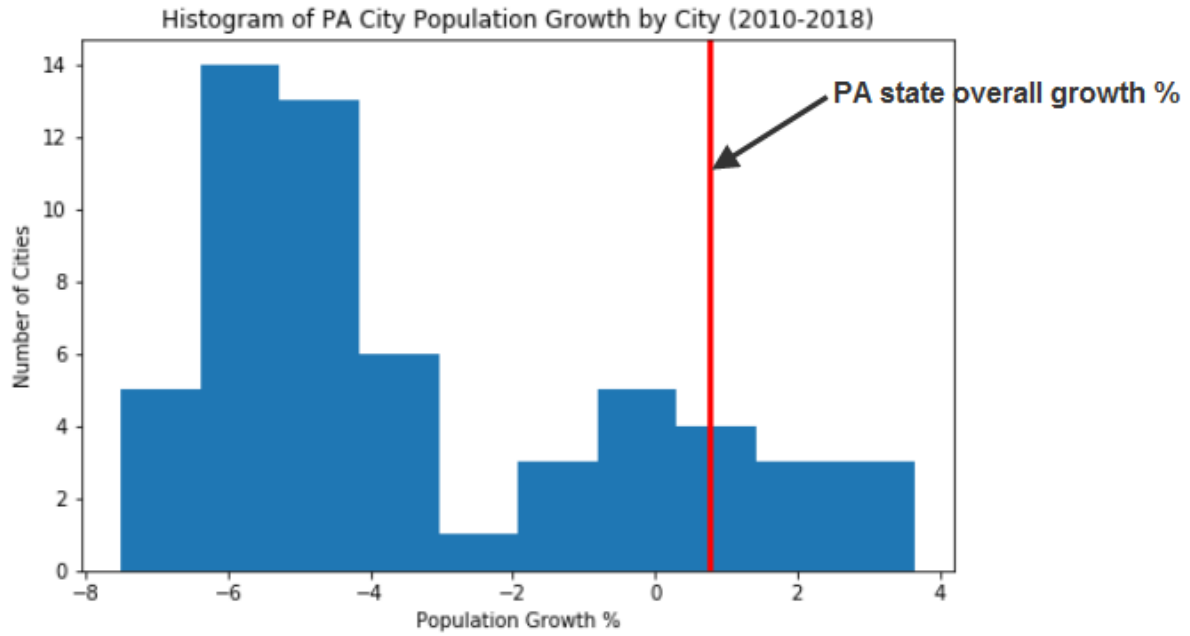


Figure 1: Histogram

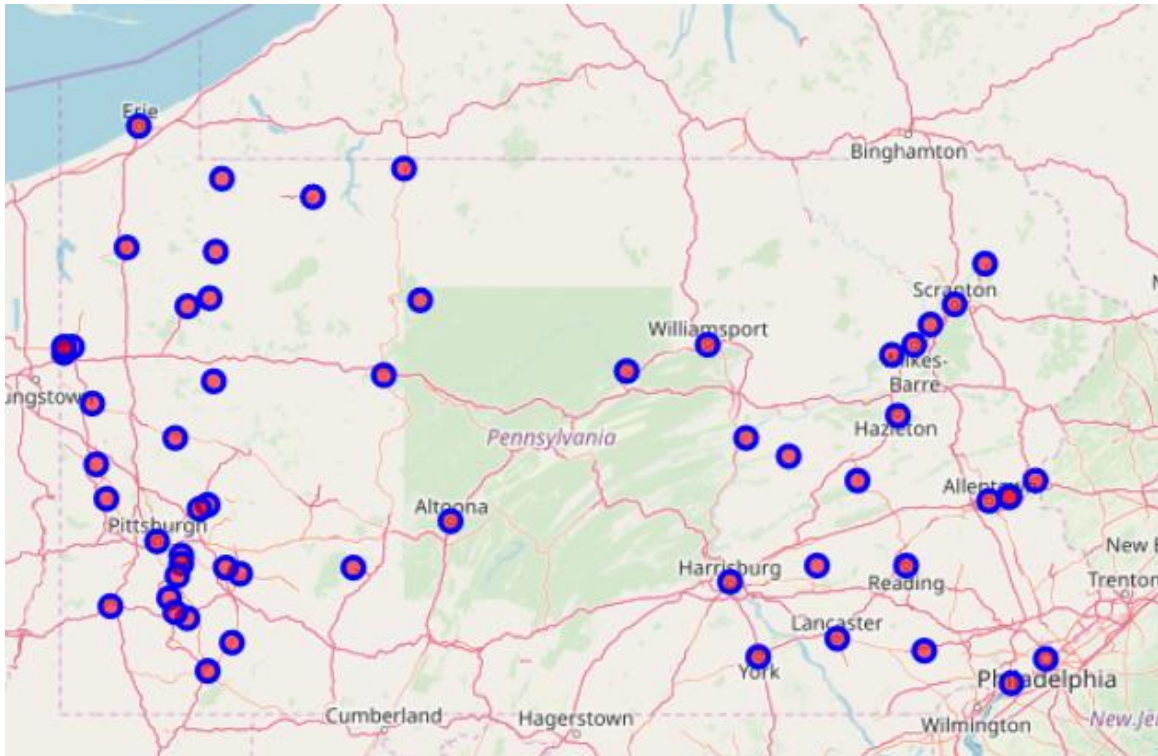


Figure 2: Folium map of Pennsylvania cities