

DATA

In this project, I will use the dataset provided by American Community Survey and hosted by University of Southern California. The data set includes median rent price from 2010 to 2016. Median rent price measures gross rent, which includes rent price plus the cost of utilities like electricity, water, gas, and sewage. Many factors can affect the rent price of a housing unit, including its age, condition, size, and design. Additionally, neighborhood characteristics like proximity to amenities, school district performance, and crime rates and safety have a sizable impact on rent prices. The general strength of the economy and housing markets can also cause rent prices to rise or fall. Because the rate of homeownership has declined since the Great Recession, an increased demand for rental units has caused rent prices to rise. Median Gross Rent, MGR, is a measure of the average level of housing affordability in an area. It is a useful tool for comparing affordability on both a neighborhood and national scale.

To this project, I have used only the MGR in 2016. Since there are multiple MGRs for each neighborhood in 2016, the maximum rate has been used for clustering. Each neighborhood has its location in a tuple which includes latitude and longitude. Therefore, we had to separate these values and put them in their own columns.

The workplace neighborhood is selected randomly from all the neighborhoods in the dataset. Next, is to calculate the miles distance of each neighborhood from workplace. We do this using the Haversine formula and then normalizing and K-Means clustering the data with scikit-learn. The cluster numbers is set to 5 and only MGR and Distance are used to do the clustering.

After clustering all the neighborhoods in 5 distinct group, we select the top two clusters with lowest distance to workplace and then the one that has a lower MGR as the cluster of choice for relocation. Then, using the foursquare API we will get all the venues in these neighborhoods and define top 10 most common venues in each neighborhood. In next step, user enters three venues of their choice that are most important to them.

Using a loop, all the neighborhoods are examined and assigned with a score based on the three venues that the user has selected. Finally, the neighborhood with the highest score is selected as a best neighborhood to move in and will be shown on the map. If more two or more neighborhoods gain the highest score, a list of those will be recommended to move in.