

Comparing Neighborhoods in Kansas City, Missouri

1. Introduction

1.1 Background

Kansas City is the largest city in the US state of Missouri. It has many neighborhoods with varying properties, there are residential, industrial and retail areas. In this project we examine where is the best neighborhood to open a restaurant serving contemporary American cuisine.

1.2 Problem

It will be advantageous to analyze data in order to determine what the 'best' neighborhood to open this new restaurant. We will use data to determine where the city is most responsive to service requests in order to gauge which neighborhoods are well-kept by the city. Furthermore, we will check which neighborhoods have the highest crime rate focusing on crimes against persons and property crimes in order to consider the safety of each neighborhood. We will also look at which neighborhoods have similar restaurants to determine which neighborhoods would be most receptive to a new contemporary American restaurant.

2. Data Acquisition and Cleaning

2.1 Data sources

Data will be collected from Kansas City's open data website. This website has a database of 311 call center service requests to the city. We will also use data from Foursquare API to determine which restaurants already exist in each neighborhood. I also used, from Kansas City's open data site, crime data. I then used shapely to determine in which neighborhood each crime was committed in 2018. On KC's open data website they also provide a neighborhood boundary geoson which I relied on heavily.

2.2 Data Cleaning

Now, the data I collected was not formatted correctly for the analysis. So first I grouped the data in the 311 call center data by neighborhood and calculated the share of calls closed within the expected timeframe and included this, and the number of days to close as well as the total number of calls in the final dataframe.

The police crime data provided latitude and longitude information for each crime, but it did not tell us in which neighborhood each crime occurred. So in order to group the crime data by neighborhood we need to determine this. Now, the neighborhood geojson file contains a feature that defines the vertices of a polygon. I input this data as a polygon in shapely, then I used the shapely contain method to determine in which neighborhood each crime was committed.

Since the neighborhood data was given as polygons, I needed to calculate the centroid of each polygon in order to have one point associated with each neighborhood to check for nearby venues using Foursquare API, so I wrote the code to calculate the centroid of each neighborhood, then added this to the data frame.

Finally, I used Foursquare API to find similar venues within one kilometer of the centroid of each neighborhood. While this won't give the total number of each type of restaurant in each neighborhood, it will give us a solid picture of what is close.