Comparing Kansas City and St. Louis Metros

Matthew Perkins

Contents

1. Introduction

1.1 Background

The Kansas City metropolitan area and St. Louis metropolitan area are two of the largest metro areas in the Midwest. Both are home to many well-established businesses and some businesses seeking to expand. With the cost of real estate being cheaper compared to the east and west coast of the United States, businesses are looking to open locations in one of the metro areas or the other or perhaps both. It is important that businesses know the similarities and differences in the cities within each metro area and between each metro area. This will give businesses an idea of which cities to build their locations in based on a city's characteristics. A company's locations are often a consideration for many employees wanting to work there as well. Additionally, for example, if a company had locations in Kansas City and wanted to open another location in St. Louis, the findings could help find a city of similar or better quality compared to their current location in Kansas City. The findings can benefit startups. Startups can decide where to open their business based on the other types of businesses in each group of cities. Finally, this project can serve the interest of individuals or families looking to move from one metro area to another to find a city like their current city.

1.2 Objective

The problem of this project focuses on finding groups of cities in each metro area that share similar characteristics. These characteristics are the cost of the venues in the city, quality of venues in the city, and types of venues in the city. Overall, those three factors will help assess the quality of life in each city. Considering the constant comparisons made between these two metro areas, the goal is to use data from various sources to determine how similar the cities in each metro area are to one another.

2. Data

2.1 Data sources

The data for this project comes from a multitude of sources. The Cities in each metro area will come from their respective Wikipedia pages (Kansas City's page, St. Louis' page). Coordinates for each city were obtained using the MapQuest Geocoding API. Using those coordinates, the top venues will around each city will be obtained through the Foursquare Places API. Finally, the Yelp API will be used to gather information about price, ratings, and number of people who rated each venue.

2.2 Description of data

The data used for analysis will include venue category, price of the venue, rating of the venue, and number of people who rated the venue. Venue categories are categorical variables and would need to be given dummy values in order to perform a clustering algorithm on the data. Price and number of people who rate the venue will be integers, and rating is a float data type since it can have a decimal place. Prices will range from 1 (least pricey) to 4 (most pricey). Ratings will range from 1 to 5 (with 5 being excellent).

2.3 Purpose of the data

This data will be used for clustering (or grouping) the cities in each metro area. Additionally, it will help assess the essential characteristics of a city's venues. Price is used to measure cost of venues, ratings help asses the quality of the venues, and venue category gives the type of venues available in a city. These venue properties will reflect the expense of living in a city in terms of daily activities in the city, the overall quality of the city, and the diversity of choice for activities in the city. The lower the cost, the higher the quality, and the more unique venue options available are, generally, desirable qualities of a city people want to live in.