

Capstone Project Final Report

Introduction

In recent years, the city of Tulsa in the mid-east of Oklahoma has been rapidly changing. In the past, it has been famous for its abundance of wealthy oil companies which made the city prosper. However, throughout the 90's the city's oil industry began to decline, and more and more companies started to move south to Texas. This left the city of Tulsa in a very low state. The population was aging, the economy was declining and as it had very little to offer, it felt like a dying city. To combat this trend, local wealthy philanthropists started investing in the city, and with an excellent multi-year plan of the city's leadership it began reviving.

As a result, Tulsa is now a great place to live. Housing prices are relatively very low (almost half of the national average), and new restaurants, retailers and other venues for past time activities are constantly popping up. The city is full of green lawns, beautiful parks and lush trees throughout. Tulsa is now experiencing a growth period, and construction is evident in many places. As such, opportunities to capitalize on this momentum are plentiful, and the real estate market is booming.

This project will focus on assisting a local construction and development company in Tulsa, Oklahoma. The data will help the company determine recent places of growth, despite the ongoing Covid-19 pandemic, where new homeowners would best utilize mixed use real estate, urban density, walkability to retail and food/beverage concepts, along with other amenities. Their aim would ultimately be double: To identify prime locations for investment in housing, and to find opportunities for developing the surrounding up-and-coming areas.

Data

To complete the project, a few sources of data were needed. However, Tulsa doesn't have readily available data like New York or Toronto. While there are sources to find Tulsa's list of neighborhoods, none (that I could find) come in table form or include their coordinates. As such, I had to resort to using zip code data, instead. This meant that the area represented by each row in the dataframe will be bigger, but at least it could be shown on a map. It also allowed me to find much more information on it.

As for the sources I used:

1. Realtor.com – This website provided most of the usable columns for my main dataset. This includes such columns as "median listing price", "median listing price per square feet", "new listing count" and "total listing count". Additionally, Realtor.com had historic data for each month in the last 4 years. This historic information was used to show trends in each zip code.

2. Zipatlas.com – This website provided pinpoint coordinates of each zip code in Tulsa. It was also where I found the population and population density for each zip code area.
3. OpenDataDE on Github – had a list of json files listing the borders of all the zip codes in each US state. Obviously, only the file for Oklahoma was relevant for this project.
4. FourSquare API – By using FS I obtained the list of most popular venues in each zip code area, in a 1,000 meter radius (approximately 0.62 US miles).