

INTRODUCTION – Analyzing neighborhoods in Dallas, TX for starting a restaurant (The Battle of the Neighborhoods)

Table of contents

1. [Introduction: Business Problem](#)
2. [Data Collection](#)
 - 2.1. [Neighborhoods data](#)
 - 2.2 [House property value data](#)
 - 2.3 [Geographical coordinates](#)
 - 2.4 [Shape of Dallas boundary](#)
 - 2.5 [Venue data](#)

1. Introduction: Business Problem

Dallas is the business and financial services center for the state of Texas and has evolved into a major high-tech hub. Dallas has become a popular migrant destination, attracting residents from abroad as well as from other states. Located in North Texas, the city of Dallas is the main core of the largest metropolitan area in the Southern United States and the largest inland metropolitan area in the U.S. that lacks any navigable link to the sea. It is the most populous city in the Dallas–Fort Worth metroplex, the fourth-largest metropolitan area in the country at 7.5 million people. Given the rapid rise of this multicultural and financially booming city, it is home to numerous cuisines and flavors from all over the world. This multicultural hotspot drives people to explore new cuisines with friends from the world over and so it is no surprise that the restaurant industry is flourishing here. Having lived in the Dallas-Fort Worth multiplex (DFW) for two years during Master's, I have grown fond of the city and its people, and through this project I hope to provide potential restaurant business owners an exploratory analysis on which areas of Dallas to target to open a restaurant.

Given my love for Indian and Chinese cuisine, this project will be targeted for stakeholders interested in opening an Indian, Asian, or Chinese restaurant in Dallas.

2. Data Collection

The data required for this project has been collected from multiple sources. A summary of the data required for this project is given below.

2.1 Neighborhoods data

The names of the neighborhoods and areas in Dallas is scraped from https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Dallas. The data is read into a pandas data frame using the `read_html()` method. The neighborhoods will be referred to as a combination of Area and Neighborhood i.e., Area_Neighborhood henceforth, to avoid any chance of neighborhoods containing the same name.

2.2 House property value data

The latest (July 2021) house property prices of mid-tier properties is obtained from: <https://www.zillow.com/research/data/>

The ZHVI is defined below:

Zillow Home Value Index (ZHVI): A smoothed, seasonally adjusted measure of the typical home value and market changes across a given region and housing type. It reflects the typical value for homes in the 35th to 65th percentile range.

2.3 Geographical coordinates

The geographical coordinates for Dallas data has been obtained from the GeoPy library. This data is relevant for plotting the map of Dallas using the Folium library. The geocoder library has been used to obtain latitude, longitude, zipcode data for various neighborhoods in Dallas. The geographical coordinates are then further used for plotting using the Folium library in python.

2.4 Shape of Dallas boundary

The outline of the Dallas boundary on the world map is obtained from a shapely file obtained from: https://www2.census.gov/geo/tiger/TIGER2019/ZCTA5/tl_2019_us_zcta510.zip. The accompanying files need to be in the same folder when reading the .shp file. Beyond that, the .shp file is converted to geojson to ensure that it can be used to plot the polygon, multipolygon and other shapes forming the Dallas map boundary.

This will be useful when augmented with HZVI data in determining which regions (by zipcode) have cheaper property prices to open a restaurant in Dallas.

2.5 Venue data

The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighborhoods in Dallas and is used to study the popular venues of different neighborhoods.