

IBM Applied Data Science Capstone Project

Starting a Grocery Store in Dallas, Texas

By Jianing He

May 2020

Introduction

Dallas is the largest city in North Texas. Its estimated population was 1,323,573 in 2019 and is the ninth most-populous city in the U.S. and third in Texas. Dallas is also the main core of the largest metropolitan area in the Southern United States with diverse economy including defense, financial services, information technology, telecommunications and transportation. With this large amount of population and economy, the requirement of grocery stores or supermarkets is huge. In addition, the Dallas City Council encourages grocery-anchored developments with given million-dollar plus subsidies. According to a March 2017 City Council briefing from the Economic Development Department, the Dallas City has given more than \$8.4 million incentive to four grocery-anchored developments and will fund more. Thus, it is a good investment to start a grocery store in Dallas with a development grant offered by the city. There is a lot more to be concerned to start a grocery store for a certainty. Especially, the location of the grocery store is one of the most essential decisions that will determine the target market and the profit.

Business Problem

The objective of this project is to analyze the neighborhoods of Dallas, Texas and to select the best locations to open a grocery store. With data science methodology and K-means clustering algorithm, we will give grocery-anchored developers the recommended locations where to open their new grocery stores.

Data we need to solve this problem

1. List of neighborhoods in Dallas, Texas.

https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Dallas

2. Latitude and longitude coordinates of the neighborhoods in Dallas for map visualization and for Foursquare locations.
3. Venue data from Foursquare API to show how many grocery stores in each neighborhood and then segment and cluster these neighborhoods in Dallas.