**Analyzing neighborhoods in Houston to start a**

**private school**

Applied Data Science Capstone Project

By: Saranya Kamalasekaran

Date: February 6, 2021

Abstract

The purpose of this project is to analyze neighborhoods in Houston to find a best location to start a private school. Houston is a large metropolis in Texas with 2.3 million residents, is the fourth most populous city in the United States, trailing after New York, Los Angeles, and Chicago. The city is the largest in the South and the Southwest.

In this Machine Learning (ML) project I was able to present a clear picture by analyzing the number of private schools in the densely populated suburbs, and by visualizing the data points on Houston map to pin point a specific area to start a new private school with valuable reasoning.

This project will be highly valuable for new business owners and entrepreneurs. In this workflow, automated process for identifying top venues across the neighborhoods are used to find better recommendations for the stakeholders.

Data overview

The following data are used to analyze the problem and quantify the results.

**Private school dataset**: The datasets used in this project is extracted from Kaggle which is originally obtained from the US Department of Homeland Security. It contains information about all private schools with attributes regarding their geographical distribution <https://www.kaggle.com/andrewmvd/us-schools-dataset>.

* **The Houston county data:** There are 9 counties in Houston and this data is used to filter private school datasets pertaining to Houston.

https://en.m.wikipedia.org/wiki/Greater\_Houston#Counties.

**Houston neighborhoods data**: List of Houston Neighborhoods is obtained from <https://en.wikipedia.org/wiki/List_of_Houston_neighborhoods>.

Latitude and longitude of each neighborhood is retrieved using Geocoder from geopy library in python.

**Foursquare API:** Data obtained from this API is used to explore types of venues and their frequencies in each neighborhood. This is then used to cluster neighborhoods based on their developments.