BATTLE OF NEIGHBORHOOD REPORT

(Version 1.0)

By,

Janaranjani Kannan

1 Table of Contents

1 Introduction

STEM is an acronym for the fields of Science, Technology, Engineering, and Mathematics, commonly used among educators, policymakers, and government officials. The demand for STEM students increases day by day. To fill academic gaps, parent seek tutoring centers to improve their student's grades, raise college exam test scores and gain academic confidence. In this report, we are going to see how an entrepreneur who is in STEM tutoring industry uses Data Science to find the perfect neighborhood to establish his premises.

1.1 Business Problem

A successful entrepreneur who owns multiple tutoring centers in North East US, is looking to expand in south. To start with, he wants to open a tutoring center in Dallas, Texas hence approached the Data Science team to help him locate the best neighborhood where he can establish a tutoring center and run it successfully.

1.2 Introduction about the city

With an estimated population of 1,345,076 and still growing, Dallas is the ninth most-populous city in the U.S. Dominant sectors of its diverse economy include defense, financial services, information technology, telecommunications, and transportation. Dallas is home to 9 Fortune 500 companies within the city limits. Over 41 colleges and universities are located within its metropolitan area, which is the most of any metropolitan area in Texas. The city has a population from a myriad of ethnic and religious backgrounds.

1.3 Target Audience

The learning center franchises such as Kumon, Mathnasium, Huntington etc or any tutoring center owner can use the data analysis discussed in this report to find the best neighborhood to establish their premises.

With the information on hand, lets move on to Data collection.

2 Data

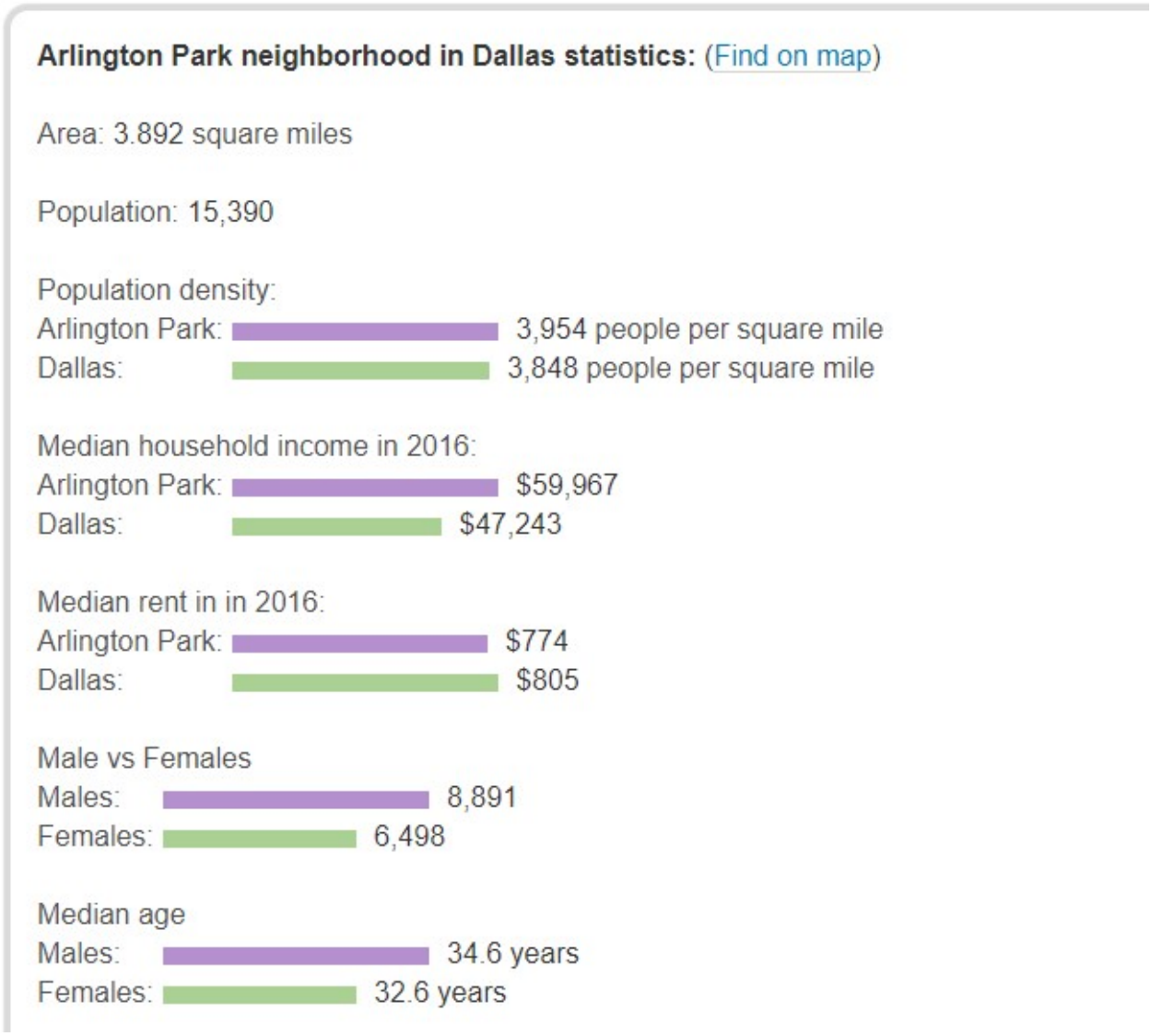
The entrepreneur is looking for middle aged, medium to high income families and populous neighborhood who will be interested in using tutoring centers to help their children academically. We are tasked with finding a neighborhood that satisfies entrepreneur's criteria.

2.1 Data Source

i) The portal **'www.city-data.com'** has detailed, informative profiles for every city in the United States. We are looking for Dallas's neighborhoods lists, population in each neighborhood, Male/Female age and Household income which can be collected from this portal.

ii) **Four-Square API** will allow us to collect details about venues around each neighborhood so that we can single out one neighborhood which doesn't have any tutoring center near it so that the success rate of the entrepreneur's tutoring center will be high.

Raw data: Sample data from **'www.city-data.com'.**



Arlington Park neighborhood in Dallas statistics: (Find on map)

Area: 3.892 square miles

Population: 15,390

Population density:
Arlington Park: 3,954 people per square mile
Dallas: 3,848 people per square mile

Median household income in 2016:
Arlington Park: $59,967
Dallas: $47,243

Median rent in in 2016:
Arlington Park: $774
Dallas: $805

Male vs Females
Males: 8,891
Females: 6,498

Median age
Males: 34.6 years
Females: 32.6 years

2.2 Data Collection Method

1. Data related to Dallas can be scraped from the portal: 'www.city-data.com'.
2. Use 'geopy' module to extract latitude and longitude of each neighborhood.
3. FourSquare API provides venue details for any mentioned radius based on the extracted latitude and longitude of each neighborhood.

2.3 Example Dataset

	City	Neighborhood Name	Population	Male Avg Age	Female Avg Age	Median Household Income	Latitude	Longitude
0	Dallas	Arlington Park	15390.0	34.6	32.6	59967.0	32.817605	-96.857609
1	Dallas	Belmont	5085.0	35.5	32.6	86718.0	32.813733	-96.782253
2	Dallas	Bent Tree	31951.0	42.0	42.3	100680.0	32.973411	-96.826306