



Coursera IBM Data Science Capstone

Battle of the Neighborhoods –
Determining Ideal Location for New
Gym in Austin, TX

May 5, 2020

Prepared By: A.J. Murphy



Introduction

- Obesity is a pandemic in the United States
 - US Adult Obesity Rates at 30.9% in 2018
 - US Adolescent Obesity Rate at 14.8% in 2017
- Obesity has major social and economic costs
 - Annual medical cost of ~\$147 billion in 2008
 - Average additional annual medical cost of \$1,429 for obese person



Business Problem

- The CDC recommends prevention as a primary method to combat obesity
 - 150 minutes of physical activity per week for adults
 - Equates to 30 minutes per day, 5 days a week
- Gym memberships in US have risen by 86% from 2000 to 2017



Business Problem (cont'd)

- The obesity rate in Texas was 34.8% in 2018
- The population in Austin, TX is constantly growing, with a 22% increase from 2010 to 2018
- Given the obesity pandemic and the rise in popularity of gym memberships, the goal is to identify an ideal location for a gym in Austin, TX



The Data

- Multiple datasets were utilized from the internet as the basis for this study
 - CDC Dataset regarding Adult and Adolescent obesity rates in the US
 - Census data regarding population data
 - US Zip Code dataset containing zip codes and associated geographic details
- Foursquare API used to find venue data centralized around area of interest using these data sets



Methodology

- Jupyter Notebook used to harness Python 3 code
 - Multiple Python libraries such as Pandas, Matplotlib, Scikit Learn & GeoPy used to collect, manipulate, and visualize the data
 - Foursquare API used to find location specific venues and help determine current gym locations and potential new gym locations



Analysis

- After installing and importing all dependencies, pandas was used to create curated dataframes that were used throughout the study

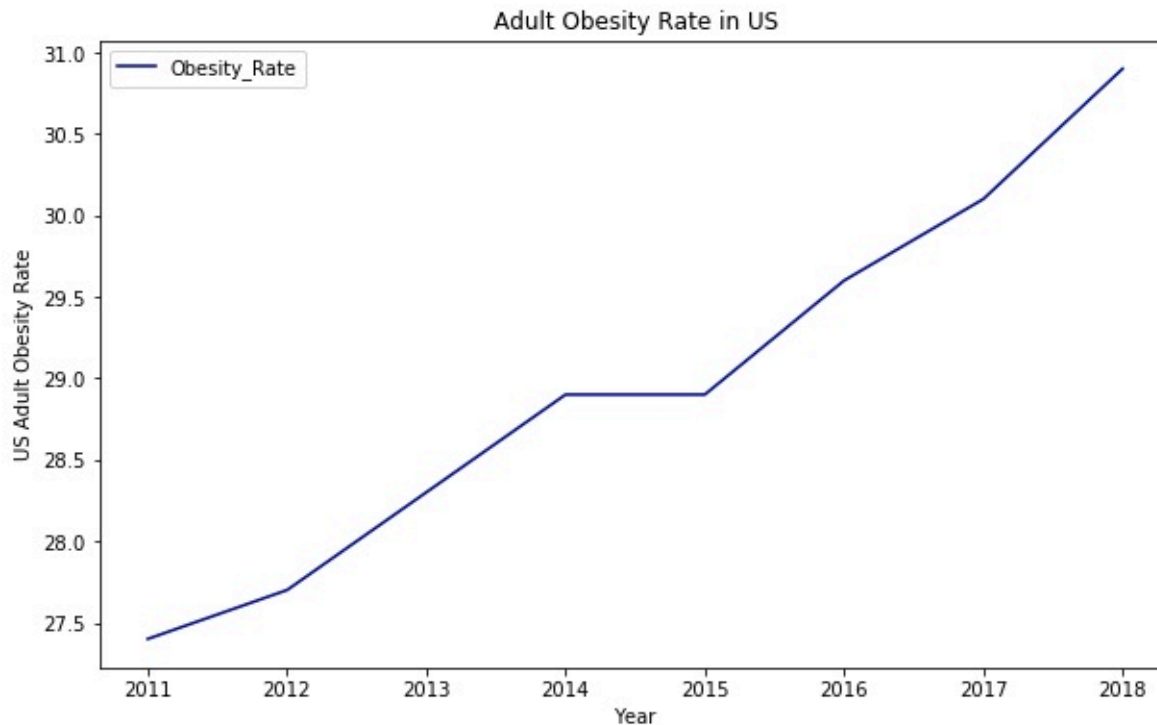
Analysis (cont'd)

- Pandas library used to create and clean dataframe containing zip codes and pertinent geographic information for Austin, TX:

	Zip_Code	City	State	Latitude	Longitude	Population
0	78701	Austin	Texas	30.27049	-97.74235	9427
1	78702	Austin	Texas	30.26327	-97.71432	23389
2	78703	Austin	Texas	30.29409	-97.76571	20890
3	78704	Austin	Texas	30.24315	-97.76537	48486
4	78705	Austin	Texas	30.29437	-97.73855	33948
5	78712	Austin	Texas	30.28502	-97.73477	860
6	78717	Austin	Texas	30.48988	-97.75371	30218
7	78719	Austin	Texas	30.14483	-97.67083	1815
8	78721	Austin	Texas	30.27005	-97.68365	12492
9	78722	Austin	Texas	30.28997	-97.71465	7110
10	78723	Austin	Texas	30.30427	-97.68570	34569
11	78724	Austin	Texas	30.29440	-97.61415	24779
12	78725	Austin	Texas	30.23581	-97.60837	7886
13	78726	Austin	Texas	30.42949	-97.84207	13867
14	78727	Austin	Texas	30.42950	-97.71741	29509
15	78728	Austin	Texas	30.45655	-97.68986	21480
16	78729	Austin	Texas	30.45842	-97.75595	29315
17	78730	Austin	Texas	30.36489	-97.83731	9186
18	78731	Austin	Texas	30.34736	-97.76847	27175
19	78732	Austin	Texas	30.37912	-97.89310	17849
20	78733	Austin	Texas	30.32323	-97.87609	8611
21	78734	Austin	Texas	30.37853	-97.94961	18745
22	78735	Austin	Texas	30.26590	-97.86658	17923
23	78736	Austin	Texas	30.26110	-97.95944	9047
24	78737	Austin	Texas	30.18779	-97.95966	16160
25	78738	Austin	Texas	30.31942	-97.95838	15589
26	78739	Austin	Texas	30.17845	-97.88869	20312
27	78741	Austin	Texas	30.23049	-97.71401	52716
28	78742	Austin	Texas	30.24413	-97.65830	828
29	78744	Austin	Texas	30.18277	-97.72920	48969
30	78745	Austin	Texas	30.20685	-97.79738	62771
31	78746	Austin	Texas	30.29729	-97.81054	28495
32	78747	Austin	Texas	30.12653	-97.74017	20166
33	78748	Austin	Texas	30.16538	-97.82343	50997
34	78749	Austin	Texas	30.21376	-97.85821	37774
35	78750	Austin	Texas	30.41828	-97.80246	30847
36	78751	Austin	Texas	30.31082	-97.72274	15805
37	78752	Austin	Texas	30.33180	-97.70426	21324
38	78753	Austin	Texas	30.38204	-97.67361	59085
39	78754	Austin	Texas	30.35575	-97.64482	24408
40	78756	Austin	Texas	30.32227	-97.74017	8323
41	78757	Austin	Texas	30.35158	-97.73252	24823
42	78758	Austin	Texas	30.38799	-97.70684	47470
43	78759	Austin	Texas	30.40268	-97.76105	42524

Analysis (cont'd)

- Matplotlib used to create plot of historical US Adult Obesity Rate:



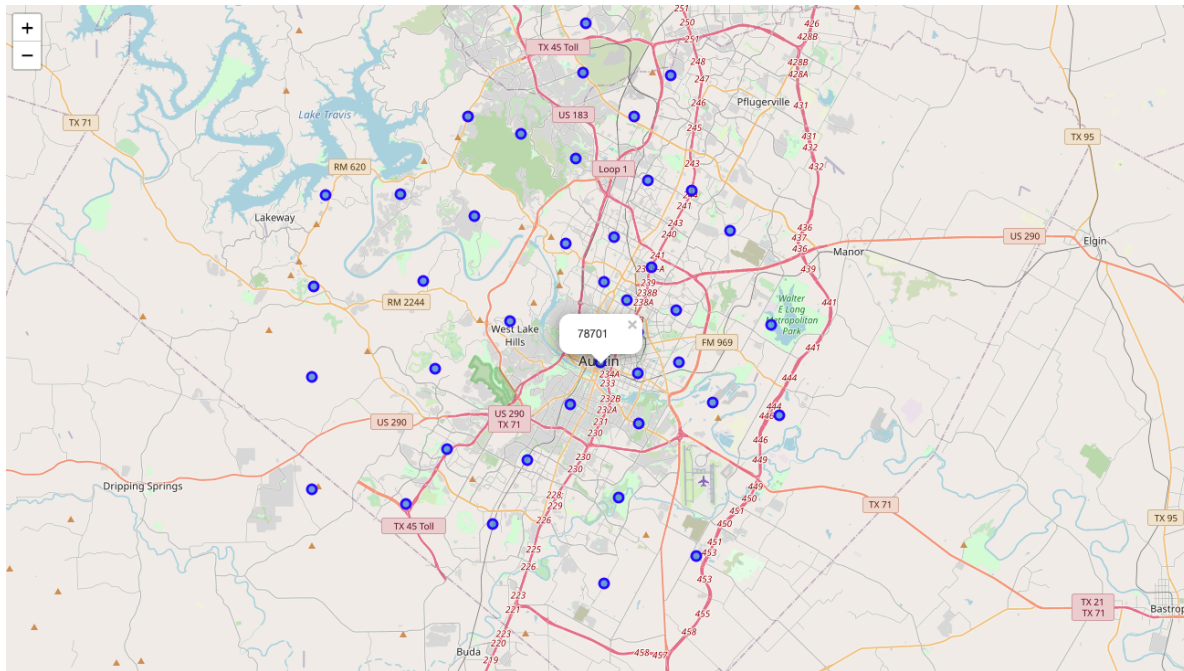
Analysis (cont'd)

- Matplotlib used to create graph of historical US Gym Memberships:



Analysis (cont'd)

- GeoPy and Folium used to create map of Austin, TX with Zip Code markers:



Analysis (cont'd)

- Foursquare API used to find top 100 venues per zip code. Head function used to display first 5 venues for zip code 78701:

	Zip_Code	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	78701	30.27049	-97.74235	Perry's Steakhouse	30.269374	-97.743676	Steakhouse
1	78701	30.27049	-97.74235	Paramount Theatre	30.269457	-97.742077	Movie Theater
2	78701	30.27049	-97.74235	Chi'lantro BBQ	30.270600	-97.741928	Food Truck
3	78701	30.27049	-97.74235	Caffé Medici	30.270119	-97.742154	Coffee Shop
4	78701	30.27049	-97.74235	The Townsend	30.269611	-97.742448	Lounge

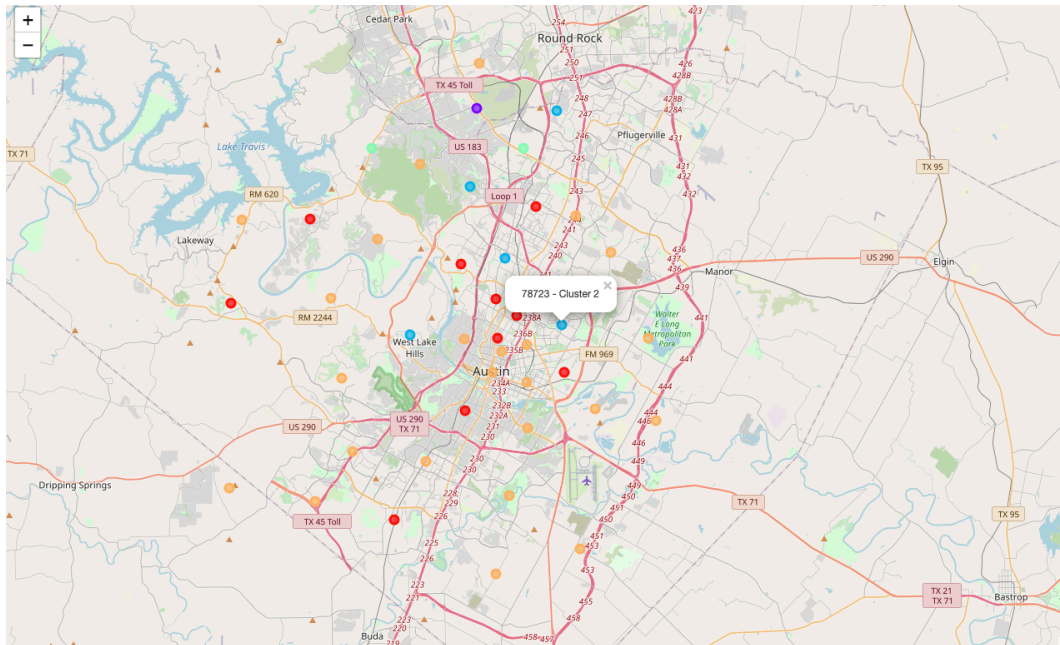
Analysis (cont'd)

- Scikit learn library used to run a k-means clustering of Austin Zip Code, using 5 clusters. Pandas used to create new dataframe with cluster labels included:

	Zip_Code	Gyms	Cluster Labels	City	State	Latitude	Longitude	Population
39	78756	0.030000	0	Austin	Texas	30.32227	-97.74017	8323
3	78704	0.030000	0	Austin	Texas	30.24315	-97.76537	48486
4	78705	0.020000	0	Austin	Texas	30.29437	-97.73855	33948
24	78738	0.033333	0	Austin	Texas	30.31942	-97.95838	15589
41	78758	0.015385	0	Austin	Texas	30.38799	-97.70684	47470

Analysis (cont'd)

- Folium used to create updated map of Austin, with the zip code markers color-coded by their cluster:





Results and Discussion

- Cluster 4 contained 25 of 43 zip codes in Austin, and Cluster 0 had 10 of 43
- Clusters 1, 2, 3 had 1, 5, and 2 zip codes assigned to them, respectively
- Cluster 4 would be ideal given the lower frequency of gyms

Results and Discussion

- Given the large number of zip codes in Cluster 4, the population data was evaluated to suggest the three highest populated zip codes within the cluster as the best potential candidates for a location for a new gym
 - 78745, 78753, OR 78741
- Additional data such as demographics, median income, etc could be used in future studies to refine these results



Conclusion

- Obesity in the US is a Pandemic
- Gyms are a great way to promote obesity prevention
- In this study, Jupyter Notebook, Python and Python libraries such as Pandas, Matplotlib, GeoPy, and Scikit Learn were used in conjunction with Foursquare API to identify and propose the best locations for a new gym in Austin as 78745, 78753 or 78741

References

- 1: Centers for Disease Control and Prevention. (2020, February 27). *Adult Obesity Facts*. <https://www.cdc.gov/obesity/data/adult.html>
- 2: Harvard T.H.Chan School of Public Health. (2020). *Obesity Prevention Source: Obesity Consequences – Economic Costs*. <https://www.hsph.harvard.edu/obesity-prevention-source/obesity-consequences/economic/>
- 3: Centers for Disease Control and Prevention. (2020, April 10). *Nutrition, Physical Activity, and Obesity: Data, Trends, and Maps*. https://nccd.cdc.gov/dnpao_dtm/rdPage.aspx?rdReport=DNPAO_DTM.ExploreByLocation&rdRequestForwarding=Form
- 4: Centers for Disease Control and Prevention. (2019, October 23). *Overweight & Obesity Strategies to Prevent Obesity*. <https://www.cdc.gov/obesity/strategies/index.html>
- 5: Centers for Disease Control and Prevention. (2020, April 10). *Physical Activity Basics - Adults*. <https://www.cdc.gov/physicalactivity/basics/adults/index.htm>
- 6: Statista. (2020). *Sports & Recreation – Sports & Fitness*. <https://www.statista.com/statistics/236123/us-fitness-center--health-club-memberships/>
- 7: SimpleMaps Interactive Maps & Data. (2020, February 27). *US Zip Codes Database*. <https://simplemaps.com/data/us-zips>
- 8: Foursquare Developers. (2020). *Foursquare Developers – My Apps*. <https://foursquare.com/developers/apps>