Venues and living in San jose

Capstone Project - The Battle of Neighborhoods

Samaa Gazzaz

Applied Data Science Capstone

IBM Data Science Professional Certificate

# Introduction/Business Problem

Every year, more that 7 million Americans move to a different state within the US [[1](https://www.governing.com/gov-data/residents-moving-to-new-state-demographics-population-statistics.html)]. Along with it, comes the responsibility of finding a good location to lay roots and settle. One of the major indicators that people relay on when deciding where to move is the availability of venues that fit their needs and lifestyle. Having moved to San Jose lately myself, I discuss in this report the different venues in San Jose neighborhoods, their types, distribution and clusters that they fall into. By the end of the report, the knowledge produced shall be sufficient in guiding the moving decision based on venues nearby, and choosing the best fit neighborhood depending on one’s needs.

# Data

In this project, we focus on the city of San Jose and its neighborhood. Wikipedia provides a comprehensive list of San Jose’s neighborhoods that we will be utilizing for this project [2]. As you can notice, the list is organized alphabetically, but other than that not much information is provided in this page other than the neighborhood names. Luckily, each neighborhood name links to a page dedicated to that neighborhoods information from where we shall acquire the latitude and longitude information of the center of that neighborhood.

There are multiple challenges with these data: First, some of the neighborhoods don’t provide latitude and longitude information; thus, we eliminate such neighborhoods from our experiment. Next, the distances between the neighborhoods’ centers are not normalized which causes duplicate venue entries when retrieving nearby venues. We discuss the solution to this problem in the methodology.

Other data used in this project, is the data retrieved using the Foursquare API. Specifically, nearby venues to each neighborhood are recorded and utilized in understanding and qualifying the neighborhoods.

# Methodology

section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, and what machine learnings were used and why.

# Results

section where you discuss the results.

# Discussion

section where you discuss any observations you noted and any recommendations you can make based on the results.

# Conclusion

section where you conclude the report.

# References

[1] New State Residents Statistics, Demographic Data. Retrieved April 4, 2019, from https://www.governing.com/gov-data/residents-moving-to-new-state-demographics-population-statistics.html

[2] Category:Neighborhoods in San Jose, California. (2017, November 15). Retrieved April 4, 2019, from https://en.wikipedia.org/wiki/Category:Neighborhoods\_in\_San\_Jose,\_California