Capstone Project of Applied Data Science (Week 5):

The Battle of Neighbourhoods

Table of contents

1. Introduction: Business Problem

My friend Jimmy is living in a neighborhood in The Woodland, Texas. He likes the living environment of the neighborhood, mainly because of all the great amenities, such as beautiful parks, theatres, schools and so on.

He recently accepted a job offer from a company in Austin, Texas. Therefore, he will move to Austin soon. He would like to find a neighborhood there similar to his current neighborhood in The Woodland, and asked me for help. After discussing with Jimmy, I understand that he prefers a neighborhood having the same types of amenities as his current neighborhood, and that he hopes the new house is as near to his new company as possible.

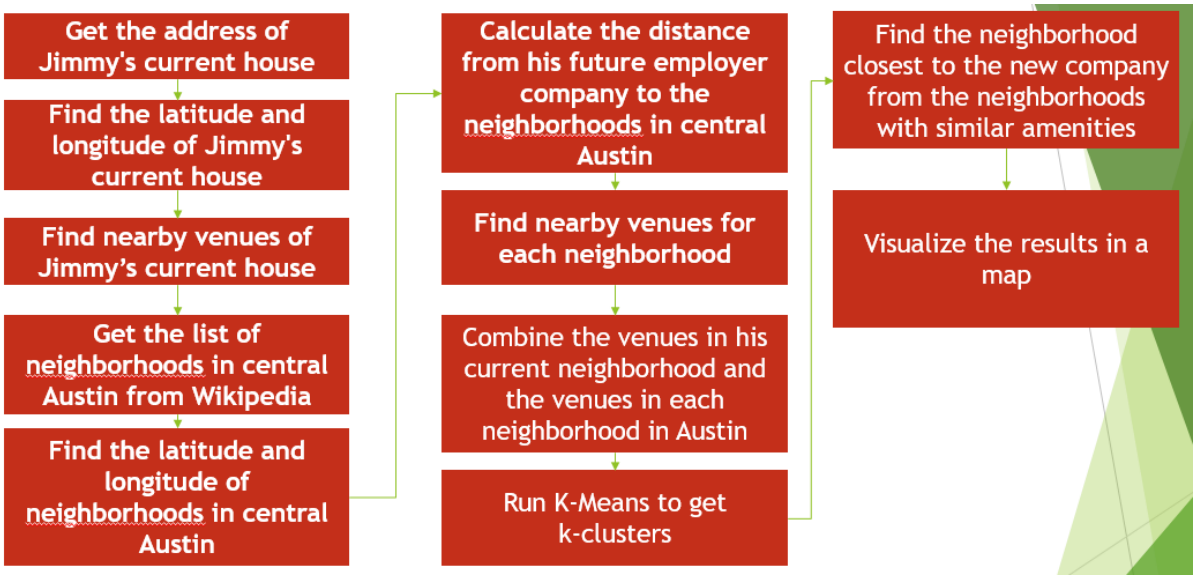
2. Data Sources

To help my friend Jimmy find the neighborhood per his requirement, I will need the following data (data source):

#	Data	Data Source
1	Address of his current house	Jimmy
2	Latitude and longitude of his current house	Foursquare API
3	The borough and neighbourhood of his current house	Foursquare API
4	Amenities in his current neighbourhood, and types of these amenities	Foursquare API
5	Address of his new company	Jimmy
6	Latitude and longitude of his new company	Foursquare API
7	Boroughs near his new company	Foursquare API
8	Neighbourhoods in Austin	Wikipedia
9	Amenities in the neighbourhoods in the nearby boroughs, and types of these amenities	Foursquare API

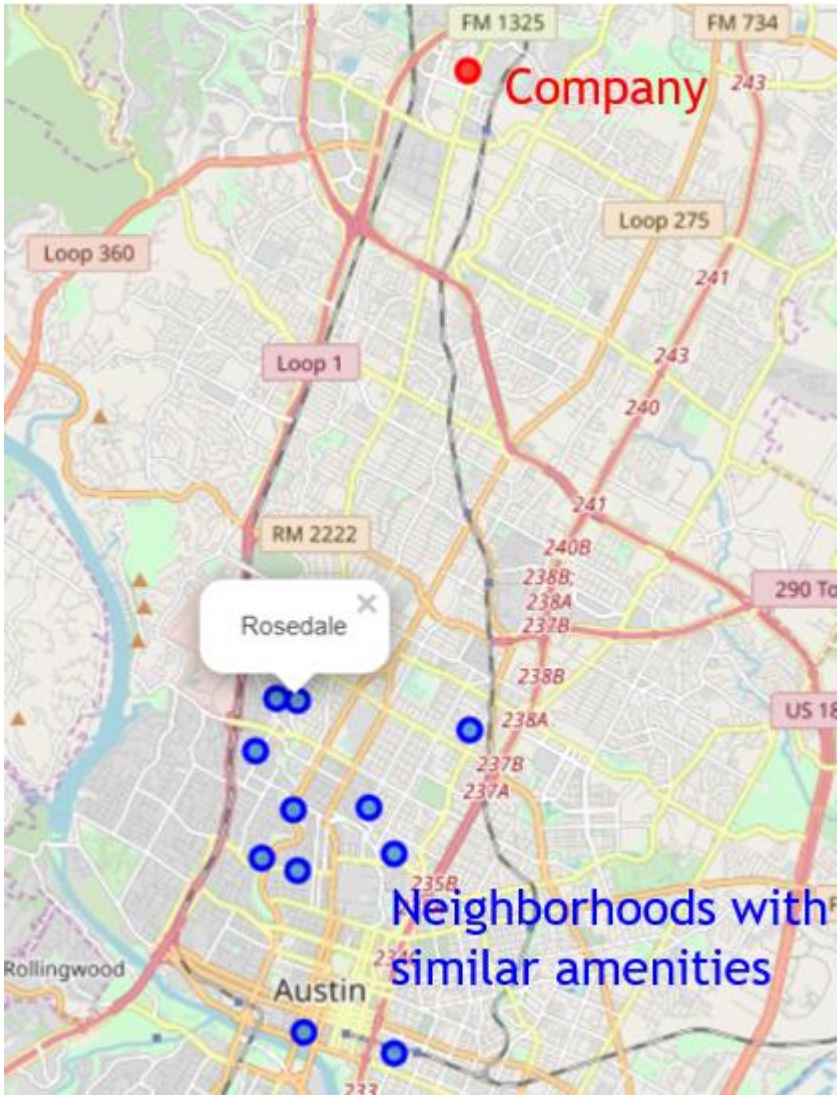
3. Methodology

The methodology is described as the following workflow.



4. Results

According to Jimmy's requirements: (1) similar amenities as his current neighbourhood; (2) as close to his new company as possible, we find that the best neighbourhood is **"Rosedale"**.



To test the result's dependence on the number of clusters and initial choice of centroids, we changed the number of clusters from 4 to 6, and running the code multiple times, **Rosedale** turns out to be the result most of the time.

5. Discussion

In this study, we find the ideal neighbourhood based on two criteria: amenities and distance. To find the optimal neighbourhood, we qualify all the neighbourhoods in Austin in two rounds. In the first round of qualification, we filtered out the neighbourhoods whose amenities are dissimilar to his current neighbourhood; in the second round of qualification, we picked the one closest to his company from all the remaining neighbourhoods.

Our customer (in this case, my friend Jimmy) may have more complicated requirement for the neighbourhood, for example, low crime rate, good kindergarten, major race of residents, and so on. In that case, our workflow (multiple runs of qualification) still applies; however, we need different types of data source than the Foursquare API.

6. Conclusion

I suggest my friend Jimmy to find a house in Rosedale, Austin, where he can find similar amenities as his current neighbourhood, and he will not waste a lot of time in transportation to company. I am happy that what I learnt from the course can be applied in real life.