1. Introduction

A local business owner is looking at opening a coffee shop in LA. The stakeholder is looking for an area with a diverse population that shows promise to hold multicultural events. Finding similarities in venues around the LA area to identify the appropriate place and culture for the coffee shop.

Since LA have a number of different areas with a variety of venues, the stake hold wants to understand what area has a diversity of shops with no coffee shop around. As well as the area the stake holder is interested in knowing the already existing shops to understand the culture of the area.

The Data will be used to find the most frequent location in each neighborhood. A K-nearest neighbor algorithm will be used to determined similarities in c and determine the best place for a local coffee shop.

2. Data

The data use to run the K nearest neighbor algorithm will be from the four-square API. The data product will be transformed to show the frequency of each shop type for 110 LA neighborhoods. Along with the neighborhood data the city demographics, average income, schools, average age and percent house ownership will helps identify more details about the areas.

3. Methodology

The first steps was to collect the data form the LA wiki page. Location, demographics, average income, schools, average age and percent house ownership was collected. Identify the types of venues in each to the neighborhood. Second, identify all the area on a map to show where the center of the neighborhoods taken form the wiki data.

Once the map data has been created, the longitude and latitude was used to pull the venue data from the four square API. Once the venue data is pulled and made into a data frame the data is counted by venue Category and normalized. Then K-nearest neighbor algorithm was set to k=5.

4. Results

The venues were categized by likeness using the K nearest neighbor algorithm and a map was generated with the grouping. Figure 1 should the resulting groupings. The map show that venues were mostly clustered in to similar areas. This could be because the are a few areas in LA that are large outliers.

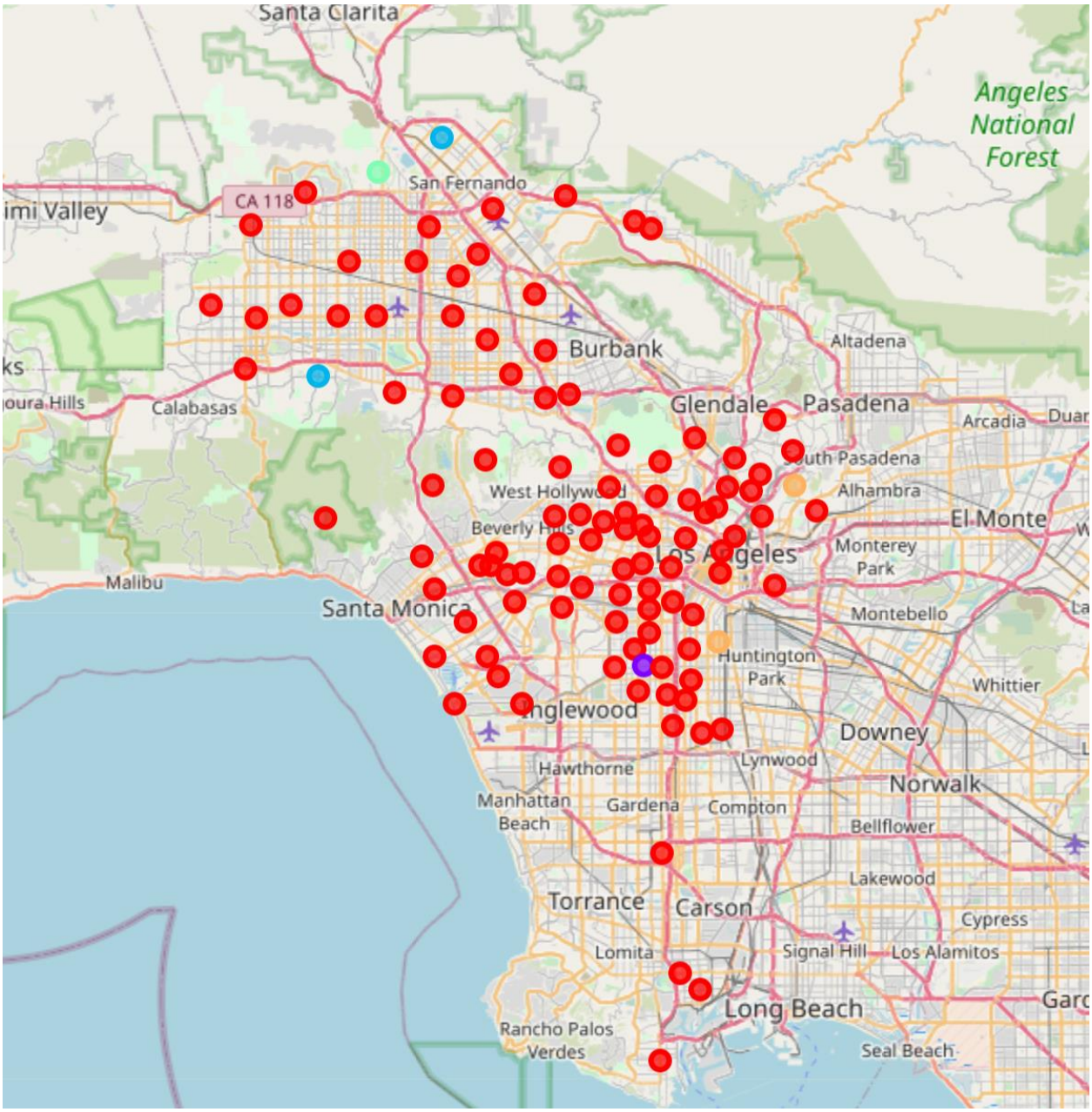

Figure 1

Bal-Air and Granada_Hills were then removed to see if the clustering improved. A second data product was created and then K nearest neighbor algorithm was ran and populated on the map (figure 2)

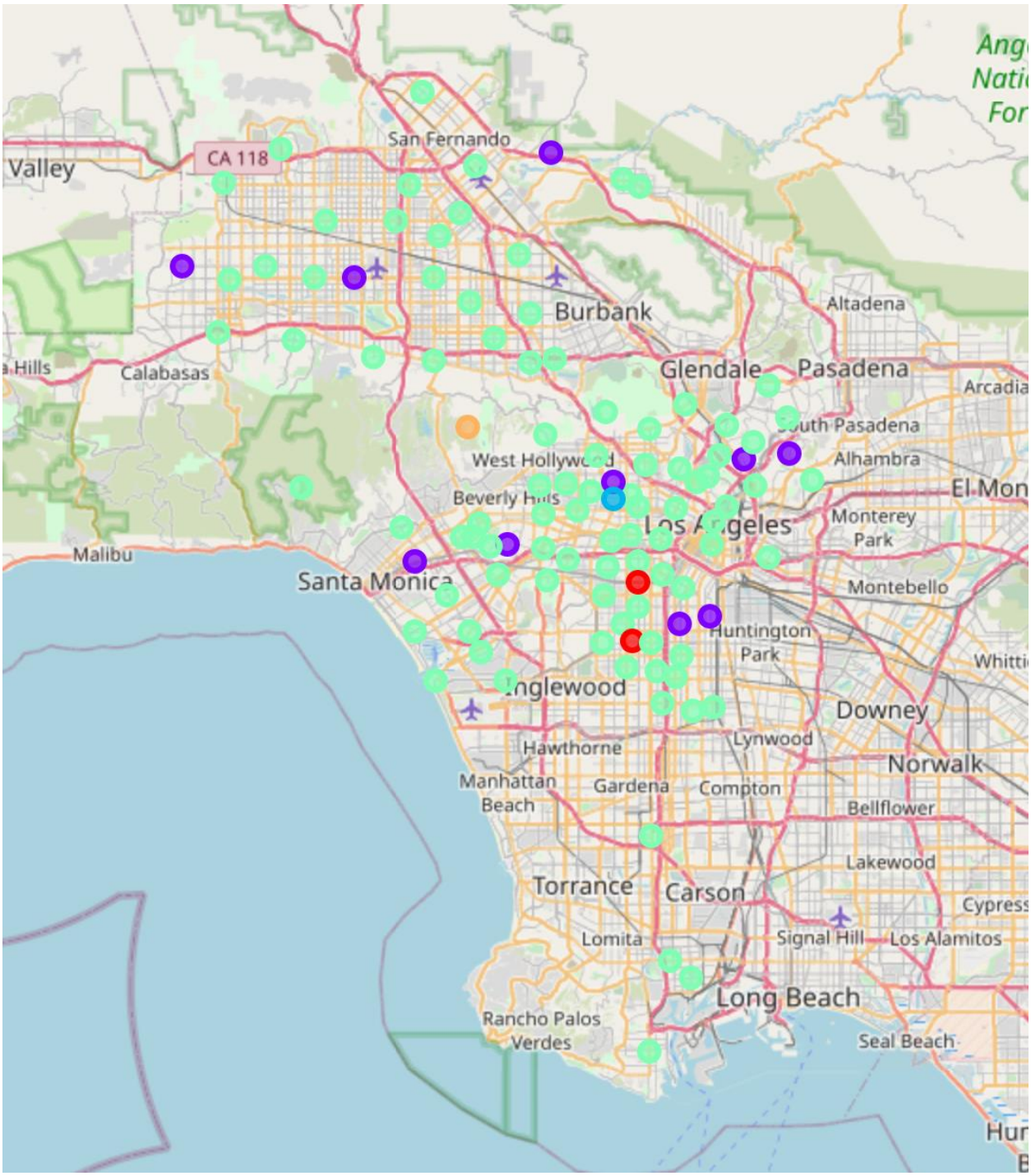

Figure 2

Here there are more of a separating between the neighborhoods showing more equal separation.

5. Discussion

The first clustering shows a that there were a few unique neighborhoods Bal-Air and Granada_Hills. After removing Bal-Air and Granada_Hills. the new clustering shows that even though the cluster reviled more unique areas with one large cluster.

6. Conclusion

The purpose of this data project was to identify possible location for a culturally relevant coffee shop in LA. The K nearest neighbor algorithm relived two unique neighborhoods, Bal-Air and Granada_Hills. After removing the two unique neighbor relived a new clustering.

The unique clusters showed the areas that would be worth exploring. It would be help full to look at other factor such as income, housing ownership as well as ethnicity.