THE BATTLE OF NEIGHBORHOODS

Applied Data Science Capstone

New restaurant location

INTRODUCTION

The problem

- Choosing a location for a new restaurant (or other any other business) is crucial for the success of the business
- To anyone planning to open a new business it is crucial to make a data based, strategic decision on a new business location
- However, many different neighborhoods have a lot of different factors differentiating them from each other, therefore we need a systematic data science approach

The background

- A few factors to consider are:
  - *Foot traffic- what areas are generally heavily frequent area*
  - *Feeder traffic- area there other businesses in the area that attract people that are also in need of your service (food)*
  - *Proximity to areas where people live or work*
  - *Parking or public transportation*
  - *Competition- are there a lot of other businesses offering the same service in a given area*
  - *Many others….*

DATA

Description of data

- I will use data from Foursquare:
  - *Queries of businesses around existing restaurants*
  - *I will use an area in the east of the San Francisco bay area. While less densely populated area than the city of San Francisco, it is still highly populated and therefore represents a good market for a restaurant. However, because of the decreased density car travel is most common and the selection of a location becomes more crucial for an emerging business to gain popularity.*

Usage of data

- To determine which factors are relevant to locate a successful new restaurant I will analyze what businesses are most prevalent in proximity to existing restaurants
- Those data will be used to develop a scoring matrix. Proximity to other restaurant of the same kind will likely be a penalizing factor as that might represent unnecessary competition (e.g.: when opening a Thai restaurant right next to another Thai restaurant). On the other hand other types of restaurants might be a positive factor in moderation. Other businesses that drive entertainment traffic (e.g. movie theater, Mini-golf, etc.) will most likely factor in positively.
- I will feed that data to k-means to identify areas that fit the required profile of the top 5 parameters determined to be relevant