The Battle of Neighborhoods

(Yet another capstone project)

Business problem

Background - As a product owner moving to San Francisco on early 2020, I have at heart to find a safe location for myself and my family.

Business Problem - This led to a sort of cheek in tongue idea: how about finding the correlation between criminality and what we could get out of Foursquare, a service absolutely not designed to return geopolitical and societal insights. This would serve multiple purposes:

Primary

- Getting to know the type of neighborhood with a glance at the venues in the streets.

Secondary

- Shouldering the idea that information is not data but what we get out of it.
- The difference between correlation and causation in action.
- Working with nonoptimal/indirectly correlated data for a case, as most real-life projects.
- Showing the importance of ethics and a way fake news could be born.

Baptiste Marien 2019/11/27

Data Selection

To complete this task and validate our assumptions:

- We will have access to the venues database in Foursquare (source - 2019 data)
  Foursquare API Documentation
- We will use the Census Tract boundaries (source - 2010 data)
  GeoJson file for San Francisco in 2010
- We will use the Average Income per Household (source - 2013-2017)
  Average Income per Household
- We will reuse the crime report CSV file (source - 2016)
- And finally, the geolocation tool provided by the Census Bureau will come handy
  Documentation

We will try to correlate crimes with a type a venue and plot it all on the map.
Then we might want to find an inverse correlation with another type of venue.
Finally we will compare it to a first hand correlation with crimes, namely a poverty index (source).

Note: the average income per household comes as a mean value per census block which is a fine grain administrative subdivision for neighborhood, a complex polygon. Making use of the geolocation tool, we will be able to attribute coordinates to a census block, but the reverse process is excessively complex. Here is a map of the census blocks in San Francisco in 2010: