**Applied Data Science Capstone**

**Assignment: Capstone Project - The Battle of Neighborhoods**

**Final Report**

Stefano Paoli, November 2019

# Introduction/Business Problem

A (hypothetical) client is a dentist who would like to open a new dental office in the New York area.

He has all the expertise and capital to open the clinic and is asking us to recommend one or more places where to establish the clinic.

The client has two requirements:

1. The business must be located in an area that presents good business opportunities and allow him to get enough patients
2. Ideally, the location would be (or not far from) a nice place to live for a family, as he would like to live near his work.

To address the first requirement, we look for locations that are relatively underserved, i.e. have a lower number of existing dental offices per capita. We also factor in economic data, like the median household income levels, as we can expect that a more affluent population creates better business conditions for a dental office.

To address the second requirement, we look at various aspects like: the presence of schools, parks, shopping and restaurants, and we look for lower levels of criminality.

Besides the borough of NYC we will also compare with a small town in the NY area, to see if it presents a better option.

# Data

The primary source for this analysis is the Foursquare Places API, from which we get data about existing dental offices and neighborhood data like the presence of schools, shopping, parks etc.

For other demographic and social data we use government sources.

We will use the data to identify places that have the lowest level of dental office per capita, but the highest level of income. As a second step we will look at which one also offer the best family living environment by offering higher number of schools, shopping and restaurants, but lowest level of criminality.

The following section describes sources and data in more detail.

## Data sources

Foursquare:

* We use the Venues endpoint group to get number of venues for the categories of interest: Dentist's Office, Arts & Entertainment, Parks, etc.

NYC Open Data:

* We use this source to get population data
* Source: <https://data.cityofnewyork.us/City-Government/New-York-City-Population-By-Neighborhood-Tabulatio/swpk-hqdp/data>
* File: New\_York\_City\_Population\_By\_Neighborhood\_Tabulation\_Areas.csv

NYU Furman Center:

* We use this source to get social and economic data
* Source: [http://app.coredata.nyc](http://app.coredata.nyc/?mlb=false&ntii=hh_inc_med_adj&ntr=Borough&mz=14&vtl=https%3A%2F%2Fthefurmancenter.carto.com%2Fu%2Fnyufc%2Fapi%2Fv2%2Fviz%2F98d1f16e-95fd-4e52-a2b1-b7abaf634828%2Fviz.json&mln=true&mlp=true&mlat=40.718&ptsb=&nty=2018&mb=roadmap&pf=%7B%22subsidies%22%3Atrue%7D&md=table&mlv=false&mlng=-73.996&btl=Borough&atp=neighborhoods)
* Income: borough-medianhouseholdincome2018.csv
* Poverty rate: borough-povertyrate.csv
* Unemployment: borough-unemploymentrate.csv
* Crime rate: borough-seriouscrimerateper1000residents.csv
* Population density: borough-populationdensity1000personspersquaremile.csv

We may use additional sources for the small town in the NY area we want to use as comparative option. In general we can use the official USA census data source <https://data.census.gov/cedsci/>.

## Data cleaning

This section will be completed as part of week 2 assignment.

# Methodology

To be completed.

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, if any, and what machine learnings were used and why.

# Results

To be completed.

# Discussion

To be completed.

Discuss any observations you noted and any recommendations you can make based on the results.

# Conclusion

To be completed.