# Recommendation system for visit Cali, Colombia

Miguel Lemos

July 20, 2021

## Introduction

In this project we will try to find an optimal location for activities in Washington.

Since there are lots of activities as restaurant, theaters and bars near from his hotel in Washington we will try to detect locations where they can make more activities as close to city center as possible, assuming that first two conditions are met.

A group of scientists will come to Cali - Colombia, to a biologic workshop. They will arrive to Four Sheraton hotel, but they don't know where they can go to eat, and they would like go to see some theater or live music nearby.

The hotel it's allocate in Street 18 North # 4N - 08 in Cali, and them free hours is between 17 and 23.

1.2. Problem

Scientist never have come to the city, and don’t know somebody that recommend some sites that will help to visit the city.

1.3. Interest

Cali, it’s a big city, and they will have trouble with those activities when have arrive, and like know the city near from they will be.

## Data Acquisition and cleaning

Based on definition of our problem, factors that will influence our decission are:

number of existing restaurants in the neighborhood (any type of activities)

number of and distance to Italian restaurants in the neighborhood, if any

distance of neighborhood from city center

We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding

number of activities and their type and location in every neighborhood will be obtained using Foursquare API

coordinate of Washington center will be obtained using Google Maps API geocoding of well-known Washington location

2.1. Data sources

All data will be scrapped from Foursquare, locating the most important places in the city for the eating, culture and hobbies.

2.2. Data cleaning

Before the extraction from Foursquare, this data will be sorted by rank of quality, nearby, and cluster group of point that information will provide to the scientist which site of the city it’s more important and visit more sites in the city

2.3. Feature selection

The principal features for the model, will be the calcifications of the places and his nearby one by other.

## Exploratory Data Analysis

In this project we will direct our efforts on detecting areas of W. DC, where the scientist can visit fast and secure.

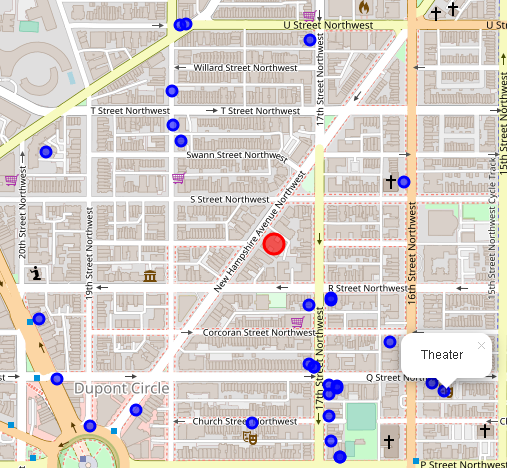
In first step we have collected the required data: location and type (category) of every restaurant within 6km Washinton. We have also identified Italian restaurants (according to Foursquare categorization).

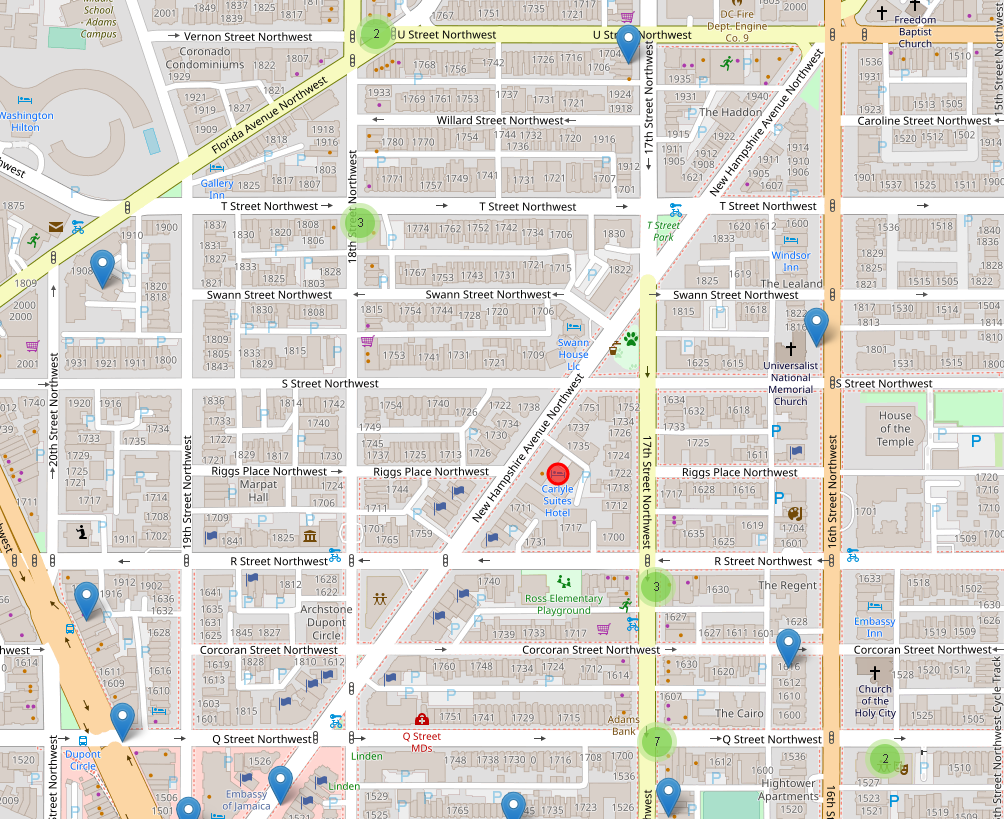
Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Berlin - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general (and no Italian restaurants in vicinity) and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two restaurants in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

## Predictive Modeling

Data obtained show to us around of 30 venues to have interest in the city. Of them We make a cluster by nearby using k-means and folium plugin, making fast, and best decision to visit venues in the city





## Conclusions

One venue of interest is around of 250mts and near of this, around of 100mts, would visit another 6 sites, distributed, this is the best options in the map.

## Future directions

Always than they visit a new place can use the model to visit the nearest places in USA hotels and optimize their time