Coursera Capstone –

Mask Dispenser

Introduction

This project aims on building a simple recommender for neighbourhoods where to place Mask dispenser. It should take data about how the neighbourhood should look like and return the neighbourhoods that fit the criteria best. The information that is provided to the recommender will contain features like hospitals, schools, train stations ect. The latter will be taken from the foursquare places API.

Problem

With the growth of the number of covid cases in france speacially in Paris, one must think of ways to help the community to fight this virus. The idea consists of placing Mask Dispenser in critical places which we consider such as hospitals, schools, train stations and more.

To do a better work dispersing these dispensers, we will have to creat groups of neighborhoods in such a way that they match the density of the venues.

This will be done by creating a jupyter notebook that shows how to build a recommender based on KNN algorithm.

Data Inputs

The data that we will feed to the algorithm mainly consists of the venues we get from the Foursqare API, a dataset that contains all the Medical analysis laboratory located in paris and a table containing the 20 departement of paris which we scrapp from a website using Beatifulsoup that we add to it the cordinates from Geocoder API.

Dataset :<https://opendata.paris.fr/explore/dataset/laboratoires-danalyses-medicales/download/?format=csv&timezone=Europe/Berlin&lang=fr&use_labels_for_header=true&csv_separator=%3B>

WebPage to be scrapped: <https://www.annuaire-administration.com/code-postal/departement/paris.html>

Methodology

The first step of this project was to get paris "Arrondissment" and add coordinates to it. Once done, we call our laboratories dataset which we preprocess since the latitude and longitutde are in the same column. So we split this column into two 'lat' and 'lng' and we drop the 'coordinates' column afterwards since we wont be needing it.

After that, and as we said in the introduction, we want a certain type of venue such as schools, train stations, universities and these medical labs which are the ones where the virus propagates more. To do so we call the Foursquare API and we precise the category to it in the url link.

We then group all our results in one single dataset, and proceed to a one hot encoding so that we can perform a KNN Clustering so that we can distribut these masks equally between our 20 clusters.

Results

The results of the clustering is seen in the following dataset:



Afterwards, we tried to visualize our venues in the map with 20 different color:



Discussion:

Some of the results obtained were good, however we can notice that some points in some clusters are too far from each other which will cause a huge problem when deciding to put these dispensers. Since some neighbourhoods which has more venues will overuse the dispensers as others will not be used too much.

Conclusion:

This project would be far better if we could get the density of the population in each neighborhood and so we might have gotten better clustering and a more realistic one. However, i wasnt able to find data which provides the population of each of the neighborhoods in our lab dataset, so instead i tried to think of a different way to do so.

All in all, this was a fun project. Manipulating API, Scrapping webpages and combining all this with a dataset has giving me a taste of what a data scientist does.