**Capstone Project  
Where could I open my next business place in London?**

**Andi Breshanaj**

**Introduction to the problem**

In this project I want to analyze a dataset of London City to find where a person could open a place for his/her business.

In particular with the Foursquare location data I will try to figure out the main problem: which is the best place to open a business?

**Dataset**

The information that I'm going to use for the Capstone project are located in this [site](https://www.doogal.co.uk/london_postcodes.php).

Basically the dataset is quite huge (about 135 MB and 320426 rows with 44 attribute), so for this reason I will use basic info such as:

1. District
2. District Code
3. Ward (Neighborhood)

Before to start the analyze I decided to filter some of the rows with the "In Use?" attribute that defines if a postcode is it used or not.  
After this filtering I'm going to use an Unsupervised Algorithm: K-Means.Unfortunately the latitude and the longitude of the dataframe are for every postcode so I'm going to use some external services to find the lat and lon of each district code. At the end for each district I will try to find with the Foursquare data the best place to open a business for each district.

In the below lines there are some info about the dataset: number of rows, the first lines of the df etc.

* The people that could be interested are everybody: someone that have enough money to open a business or maybe a graduated student from university that want to start a new challenge

A close up of a map

Description automatically generatedIn the following picture there is a list of London postcodes districts

Some info of the dataset:

* 44 attributes
* 320426 rows (177967 after the filtering)
* 33 District used
* 691 Neibourhoods

**Methodology**

* For this project I used an unsupervised algorithm: Kmeans with k = 4
* In the next slide there is a picture there is a map of district of the city of London
* There are 207 uniques categories.
* ISSUE: I didn’t find any resource in the Internet to find latitude and longitude of the District. For this reason I used a service at this site <https://www.latlong.net/>

A close up of a map

Description automatically generated

The above picture represent the map of the 33 districts (using Folium Library). In the next page there is the map with the clustered districts (based on the venues).

A picture containing text, map

Description automatically generated

**Results**

This are the main results:

* Coffee shop is the best place to open a business in London
* The best places as the above slide will explain better are situated in the city centre

**Discussions and Conclusions**

The study of the places and the most venues in London was an exciting project. I found some issues especially:

* Where to find a good daset with consistent info about the city of London
* Find the Latitudes and longitudes of the districts

The analysis suggest that the best business to open in London are:

* 'Coffee Shop’: 22%,
* 'Café’: 16%,
* 'Pub’: 16%

The best place to open one of the 3 business above is in the city centre.