Capstone Project

The battle of neighborhoods

1. Project description:

As a tourist visiting London the capital of the United Kingdom and the largest city of the Europe Union, it would be helpful having an idea of best places to visit before landing.

Therefore, this project aims to implement a recommender system to suggest the best places to visit giving certain people preferences. This can be done by finding and filtering different venues. Locating and suggesting the most interesting neighborhoods based on the preferences can be performed based on the number of visitors, likes, prices, rating …etc.

2. Data description:

FOURSQUARE PLACES API

In this project we use Foursquare Places API to obtain details about venues in London including location, tips, categories etc… This is possible by the use of a explore call that returns a list of recommended locations in a specified area. After getting a list of all venues, we use then venue calls to get statistics about each one.

OTHER DATA SOURCES

A List of London district names with Postal codes to match with API Foursquare data and are retrieved from: https://en.wikipedia.org/wiki/List_of_areas_of_London

3. Methodology:

There are many existing venues in neighborhoods London

- **Which city part has the most number of venues by category ?**
- **Let suppose someone is interested in food and restaurants. So which is the best part of London to visit and which are the best restaurants to recommend ? The following criteria can then be used:**

- **rating**
- **number of users checked-in**
- **Price category**

- We then segment and cluster neighborhoods based on these criterion using K-means method.

Data collection methodology:

First step: use explore calls from Foursquare Places API to get available venues of each city part based postal codes and GPS Coordinates. Second step: collect each venue statistics based on venue calls. Next we present the results.

4. Results

Top 20 Neighborhoods based on number of venues

city	postalCode	Number of restaurants
London	EC1R 4QE	14
	EC2N 4AY	13
Finchley	N12 0GL	12
London	EC1R 0HA	10
	EC1R 3EA	10
	N3 1RT	10
	N1 8ED	9
Croydon	CR0 1LD	9
Crouch End	N8 8TE	9
London	EC1V 9LA	9
Croydon	CR0 1DP	9
London	EC4M 7DZ	9
	EC1V 9LT	9
	WC1N 3NB	8
	E2 7JE	8
Spitalfields	E1 7LJ	8
London	ec4a 3by	7
	N8 9TE	6
	E2 7DP	6
Ilford	IG2 7RH	6

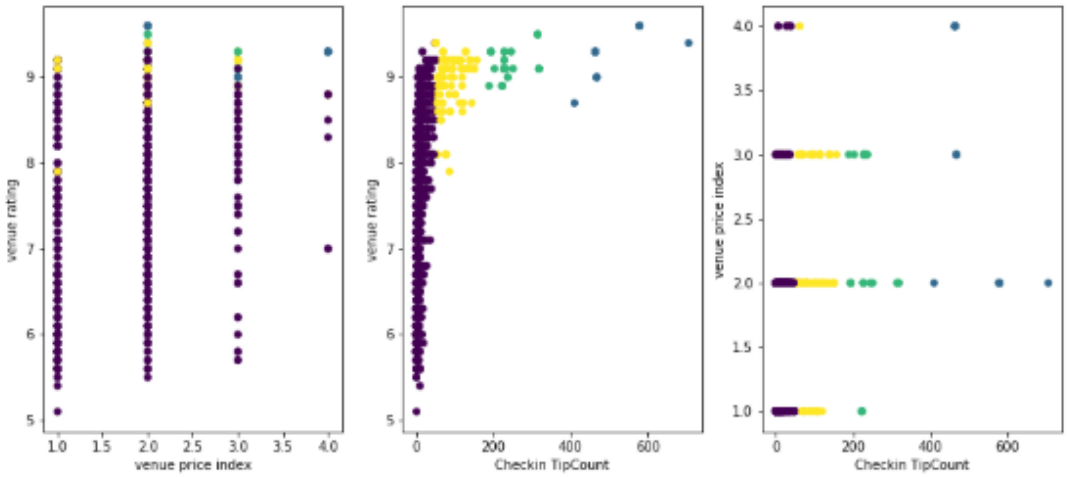
Top 20 Neighborhoods based on number of restaurants

city	postalCode	Number of venues
Brixton	SW9 8PS	47
London	WC1V 7EN	32
	W1T 3NP	30
	EC2Y 8DS	30
Hayes	UB3 3EX	24
London	EC1R 4QE	20
	EC4M 9AF	20
	WC1B 3DG	20
Crystal Palace	SE19 3RY	20
London	W11 2ED	19
Croydon	CR0 1DP	18
London	SE19 1RX	18
	EC1V 9LT	18
Croydon	CR0 1LD	18
London	E1 6BJ	18
	EC1R 3EA	17
	WC1A 1LY	17
	WC1N 3NB	17
	WC2B 5JF	16
	E2 7DJ	16

Top 20 Neighborhoods based on number of restaurants with respectively mean of rating and mean price index

		Number of restaurants	mean rating	mean price index
city	postalCode			
London	EC1R 4QE	14	9.142857	2.000000
	EC2N 4AY	13	9.207692	3.692308
Finchley	N12 0GL	12	7.150000	1.500000
London	EC1R 0HA	10	9.200000	3.000000
	EC1R 3EA	10	9.200000	2.000000
	N3 1RT	10	7.350000	2.000000
	N1 8ED	9	9.100000	2.000000
Croydon	CR0 1LD	9	7.533333	2.000000
Crouch End	N8 8TE	9	8.422222	1.000000
London	EC1V 9LA	9	9.100000	2.000000
Croydon	CR0 1DP	9	6.533333	1.666667
London	EC4M 7DZ	9	9.200000	2.000000
	EC1V 9LT	9	9.200000	3.000000
	WC1N 3NB	8	9.100000	2.000000
	E2 7JE	8	9.600000	2.000000
Spitalfields	E1 7LJ	8	9.100000	2.000000
London	ec4a 3by	7	9.100000	1.000000
	N8 9TE	6	9.000000	2.000000
	E2 7DP	6	9.400000	2.000000
Ilford	IG2 7RH	6	6.500000	1.000000

Clustering of restaurants based price index, checkin tipcount and rating



5. Conclusion:

In this project we started by collecting data from Foursquare API with a limiting regular account. Then we performed city exploration and research of restaurants in london by gathering data, analyzing data, statistic and clustering, in order to make the best possible recommendation based on some given references such as price, rating.

This research can be extended to other venues and by using other types of statistics.