Battle of the Neighbourhoods

IBM Data Specialization Capstone

Agenda Layout

01 **Introduction**
Identifying the purpose

02 **Problem**
Explaining the problem

03 **Data**
Describing which types of data will be used and from where

04 **Methodology**
How the data has been cleaned and analyzed

05 **Results & Conclusion**
Which locations are best

Introduction

Within this presentation, we will look to answer the question; ‘If I were to look to establish a new restaurant or café, where should I place it and what type of venue should it be?

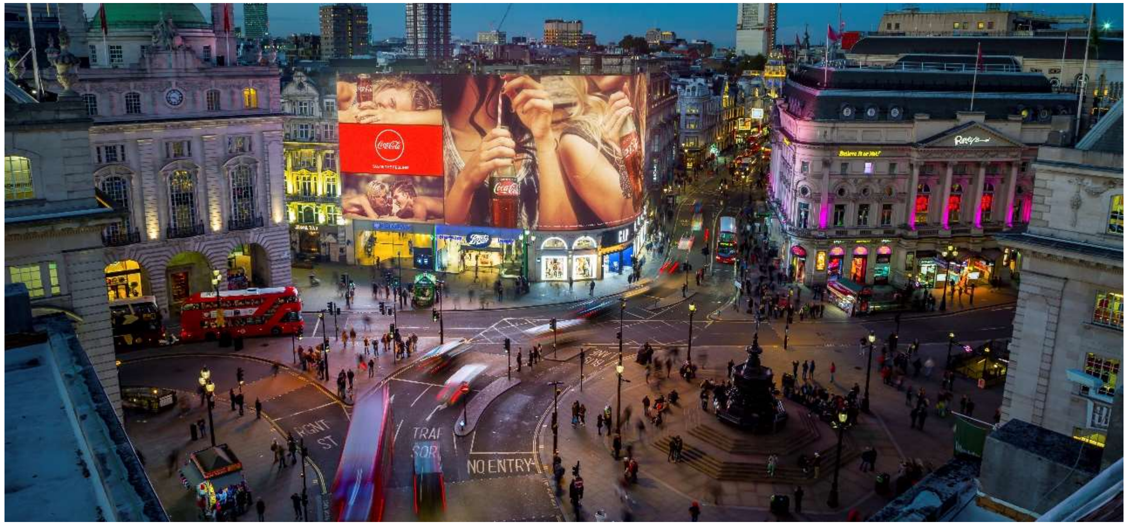
We will be looking to answer this question for the city of **London**

<u>London:</u>

Population: 9.3 million (2020)
Demographics:
White 59.79%, Asian 18.49%, Black 13.32%, Mixed 4.96%, Other 3.44% (Wiki – 2011 cenus)
Economy: £500 billion – 1/3 of UK’s GDP

Problem

London is well known as one of the world leaders in banking, retail, fashion, tourism, transport and within media is one of the most used movie/TV settings. However, it is a dog eat dog world within the streets underneath the sky scrapers, one where business are constantly rising and falling and to that end, any new business venture needs to plan carefully to avoid failure.



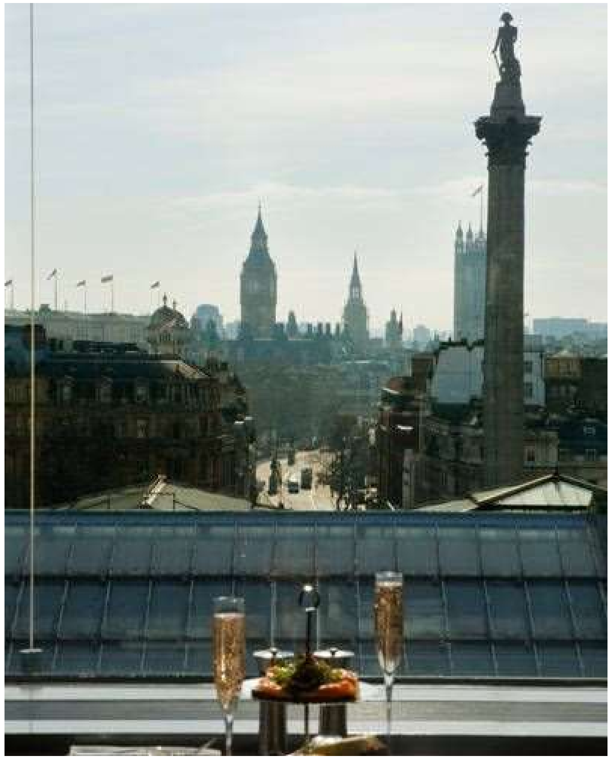
Problem

The restaurant business model is, in theory easy;

- customer orders food
- You provide said food
- customer pays
- repeat

But how can we know what type of restaurant is best to open up and where? London is one of the most diverse places in the world with more than 270 different nationalities and around 300 languag es being spoken within it. With this comes some of the most diverse types of cuisine all within a few mile radius - head to Brick Lane for some of the best curry, head to Piccadilly where China town can serve you up some fresh duck or head over to Knightsbridge/Mayfair for some upscale fine dining.

So what if we wanted to open a Pizza restaurant or a noodle bar? Where would be best to go? This project will tackle this issue head on using data sets on cuisine and demographics throughout London.



Problem

The Plan

To create a successful business model, we are going to look at various factors to gain the best insight into where to select a location for our restaurant. We will look at;

- London Population
- London's Demographics
- Accessibility to nearby markets to ensure costs are low but quality high
- Reviews of similar local eateries
- Who are the competitors in that location?
- Cuisine served / Menu of the competitors
- Segmentation of the Borough
- Untapped markets
- Saturated markets etc

Using the above, we will be able to select the best location for this company to begin their restaurant business in.

Data

For this ‘London’ is referred to as the 32 Boroughs.

The first set of data we will look at is population by Borough.
For this, we will be using a table from Wikipedia ‘London’s 32 Boroughs’ – https://en.wikipedia.org/wiki/List_of_areas_of_London

This data has been cleaned so that any non-London townships have been removed, also any post-code groupings between boroughs have been given their own line.

	Location	Borough	Postcode
0	Abbey Wood	Bexley, Greenwich	SE2
1	Acton	Ealing, Hammersmith and Fulham	W3
2	Acton	Ealing, Hammersmith and Fulham	W4
3	Angel	Islington	EC1
4	Angel	Islington	N1
5	Church End	Brent	NW10
6	Church End	Barnet	N3
7	Clapham	Lambeth, Wandsworth	SW4
8	Clerkenwell	Islington	EC1
9	Colindale	Barnet	NW9
10	Colliers Wood	Merton	SW19
11	Colney Hatch	Barnet	N11
12	Colney Hatch	Barnet	N10
13	Covent Garden	Westminster	WC2
14	Archway	Islington	N19
15	Cricklewood	Barnet, Brent, Camden	NW2
16	Crofton Park	Lewisham	SE4
17	Crossness	Bexley	SE2
18	Crouch End	Haringey	N8
19	Crystal Palace	Bromley	SE19

Data

The second set of data we will look at is demographics.

This will give us a good breakdown of the cultures of each borough and where the type of venues will differ. London demographics; https://en.wikipedia.org/wiki/Demography_of_London

Following this, we decided to look at the South East of London due to the varied ethnicities present.

Local authority	White	Mixed	Asian	Black	Other
Barnet	64.1	4.8	18.5	7.7	4.8
Barking and Dagenham	58.3	4.2	15.9	20	1.6
Bexley	81.9	2.3	6.6	8.5	0.8
Brent	36.3	5.1	34.1	18.8	5.8
Bromley	84.3	3.5	5.2	6	0.9
Camden	66.3	5.6	16.1	8.2	3.8
City of London	78.6	3.9	12.7	2.6	2.1
Croydon	55.1	6.6	16.4	20.2	1.8
Ealing	49	4.5	29.7	10.9	6
Enfield	61	5.5	11.2	17.2	5.1
Greenwich	62.5	4.8	11.7	19.1	1.9
Hackney	54.7	6.4	10.5	23.1	5.3

	Location	Borough	Postcode	Latitude	Longitude
0	Abbey Wood	Bexley, Greenwich	SE2	51.49245	0.12127
1	Crofton Park	Lewisham	SE4	51.46268	-0.03558
2	Crossness	Bexley	SE2	51.49245	0.12127
3	Crystal Palace	Bromley	SE19	51.41990	-0.08808
4	Crystal Palace	Bromley	SE20	51.41009	-0.05683

The third set of data we will use is geo data so we can analyse the geographical data of each location in finer detail.

For this we will be using FourSquare and a Geocoder.

Table left: Head of data after adding co-ordinates and merging tables.

Methodology

We began by cleaning the data and removing any postcode not in SE. This left us with the ‘se_df’ table which displays only SE post code locations and their latitude and longitudes.

	Location	Borough	Postcode	Latitude	Longitude
0	Abbey Wood	Bexley, Greenwich	SE2	51.49245	0.12127
1	Crofton Park	Lewisham	SE4	51.46268	-0.03558
2	Crossness	Bexley	SE2	51.49245	0.12127
3	Crystal Palace	Bromley	SE19	51.41990	-0.08808

We then focused on one single borough to gain an insight into the make-up of venues within that particular area. For this task we selected **Lambeth.**

	name	categories	lat	lng
0	Mercato Metropolitano	Street Food Gathering	51.498318	-0.098162
1	Terry's Cafe	Café	51.500715	-0.098314
2	Southwark Playhouse	Theater	51.497779	-0.098603
3	German Kraft	Brewery	51.498662	-0.098935
4	The Gladstone Arms	Pub	51.500961	-0.095031

	Count
Pub	8
Coffee Shop	8
Hotel	5
Theater	4
Street Food Gathering	4

This data was then counted so we could see the most popular type of venue. As you can see both pub and coffee shop are top of the list. This appeared to be ‘in fit’ with the area as it is a heavy commuter area with Waterloo train station being within the vicinity.

Methodology

This same methodology was applied to multiple locations so we could gain an idea of the most popular venue types.
Again, Pub and coffee shop were in the top 2, but Pub now has almost double the amount of venues as coffee shop.

	Count
Pub	720
Coffee Shop	429
Café	366
Park	349
Grocery Store	302

We then began a clustering process after creating a visual map of the boroughs within the SE post codes.

One hot encoding was also used in order to show where locations were the most popular in which area.

‘Italian Restaurants’ was looked at within hot encoding and from the data it shows it is popular in a number of locations.

Methodology

Using clustering we were able to obtain information on the top 10 most common venues within our locations. A sample is shown below. We can see pub and coffee shop appearing in some form on each location.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Mo Commo Venu
0	Abbey Wood	Supermarket	Grocery Store	Fast Food Restaurant	Pub	Clothing Store	Furniture / Home Store	Pharma
1	Anerley	Pub	Café	Supermarket	Coffee Shop	Pizza Place	Gym / Fitness Center	Groce Sto
2	Bankside	Pub	Coffee Shop	Hotel	Street Food Gathering	Theater	Italian Restaurant	Restaura
3	Beckenham	Pub	Café	Supermarket	Coffee Shop	Pizza Place	Gym / Fitness Center	Groce Sto
4	Bellingham	Grocery Store	Supermarket	Park	Café	Coffee Shop	Pub	Itali Restaura

Methodology

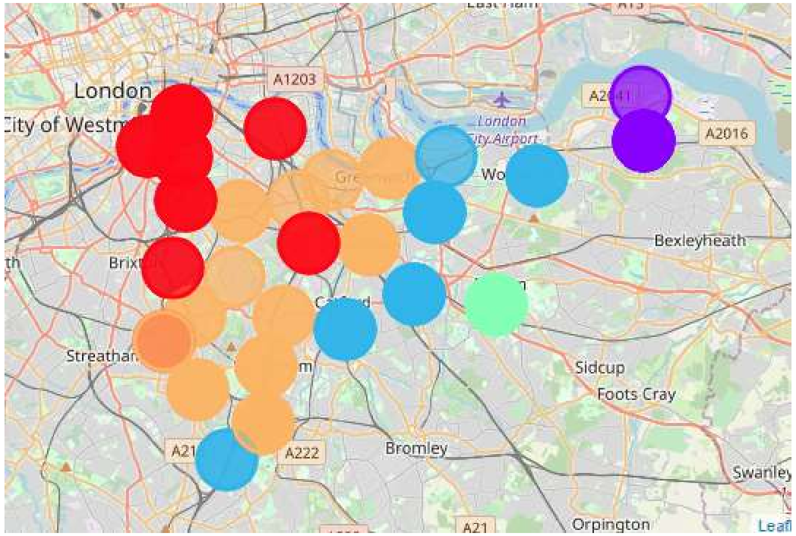
Finally from there we used varying analytic methods in order to discover the most popular locations by type. This was then plotted onto a map.
To the right, a silhouette coefficient has been used.

In [127]:
```
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans

for n_cluster2 in range(2, 10):
    kmeans2 = KMeans(n_clusters = n_cluster2, random_state = 0).fit(se_grouped_c
    label2 = kmeans2.labels_
    sil_coeff = silhouette_score(se_grouped_clustering, label2, metric = 'euclid
    print("Where n_clusters = {}, the Silhouette Coefficient is {}".format(n_clu
```

```
Where n_clusters = 2, the Silhouette Coefficient is 0.6902239335122877
Where n_clusters = 3, the Silhouette Coefficient is 0.6731520102865589
Where n_clusters = 4, the Silhouette Coefficient is 0.6663111887414048
Where n_clusters = 5, the Silhouette Coefficient is 0.695503672171252
Where n_clusters = 6, the Silhouette Coefficient is 0.7798187771292893
Where n_clusters = 7, the Silhouette Coefficient is 0.8037924304533735
Where n_clusters = 8, the Silhouette Coefficient is 0.870253134292505
Where n_clusters = 9, the Silhouette Coefficient is 0.9223740149354457
```

The higher the number, the better the coefficient. Remember we used a cluster value of 5.



Results

After reviewing the above cluster data, we can now make certain observations;

1. Pubs & Coffee shops are the most popular venue type in SE London
2. Pizza places/Italian Restaurants are very popular.
3. Other cultural food places (Indian, Korean etc) are quite far down the list in terms of popularity. But they are still on the list meaning there could be a market for them there?

Conclusion

Based on the data and the visualisation of our dataframes, we can make an assumption that a pub or co
ffee shop/cafe is the best type of venue to open in SE London. London is a commuter heavy location an
d as such, they want quick easy accessible food.
Tourism is lower in SE London than in other parts and as such, less likely to have visitors wanting a sit-d
own meal in it.

That being said, there are some venues of a restaurant experience who feature in the top 10 venues acr
oss most clusters. If you're in Greenwich, then Mediterranean food is a good choice and across the boar
d Italian seems to be a good type to go for.

It would be helpful to know if these particular venues offer takeaway or dine-in only experiences to asse
ss fully if they fit in the assumption of the 'commuter dash'.