Battle of the Neighbourhood – Oxford City Food Choices

Final Report

A L McKie

1. Introduction

The purpose for this project is to explore the various food or restaurant venues in Oxford, UK. It aims to help identify various venues based on their price and ratings and allow people living in or visiting Oxford to make an informed decision of where to go eat.

There are many people, especially students, living in Oxford or in the surrounding areas of Oxfordshire. Oxford is a world famous university city which it draws hundreds of visitors each year. It would be beneficial to have a selection of venues whereby people can explore the various food or restaurant choices and determine if they would like to try these places out by primarily looking at the average prices that will fit their budget and on the venue ratings. As many students live in the area and are mostly on a tight budget, it is would be good to know where to go for a good inexpensive meal.

Problem to be solved:

The purpose of this project is to help visitors to Oxford to identify and explore food venues close to the Oxford city centre based on their location, price, rating and likes so they can make an informed decision on where to visit.

2. Data Selection

2.1 Location

Oxford city is in the county of Oxfordshire, UK, the postcode areas pertaining to Oxford is scraped from https://en.wikipedia.org/wiki/OX_postcode_area using requests and BeautifulSoup libraries and the information converted to a pandas dataframe. Four postcode areas are a part of Oxford. The postcode areas and Oxford city centre is used to get the location coordinates (latitude and longitude) of Oxford and postcode areas using the Geopy Geocoder library. Folium library is be used to bind the location data of Oxford and the postcode areas onto an interactive map.

	Postcode district	Post town	Coverage	Local authority area(s)	Latitude	Longitude
0	OX1	OXFORD	Central and South Oxford, Kennington, Boars Hi...	Oxford, Vale of White Horse	51.739440	-1.256203
1	OX2	OXFORD	North and West Oxford, Botley, North Hinksey, ...	Oxford, Vale of White Horse, Cherwell	51.760397	-1.273345
2	OX3	OXFORD	North East Oxford, Beckley, Headington, Marsto...	Oxford, South Oxfordshire, Cherwell	51.764686	-1.217932
3	OX4	OXFORD	East Oxford, Cowley, Blackbird Leys, Littlemor...	Oxford, South Oxfordshire	51.730411	-1.210317

Figure 1: Postcode areas in Oxford and location coordinates.

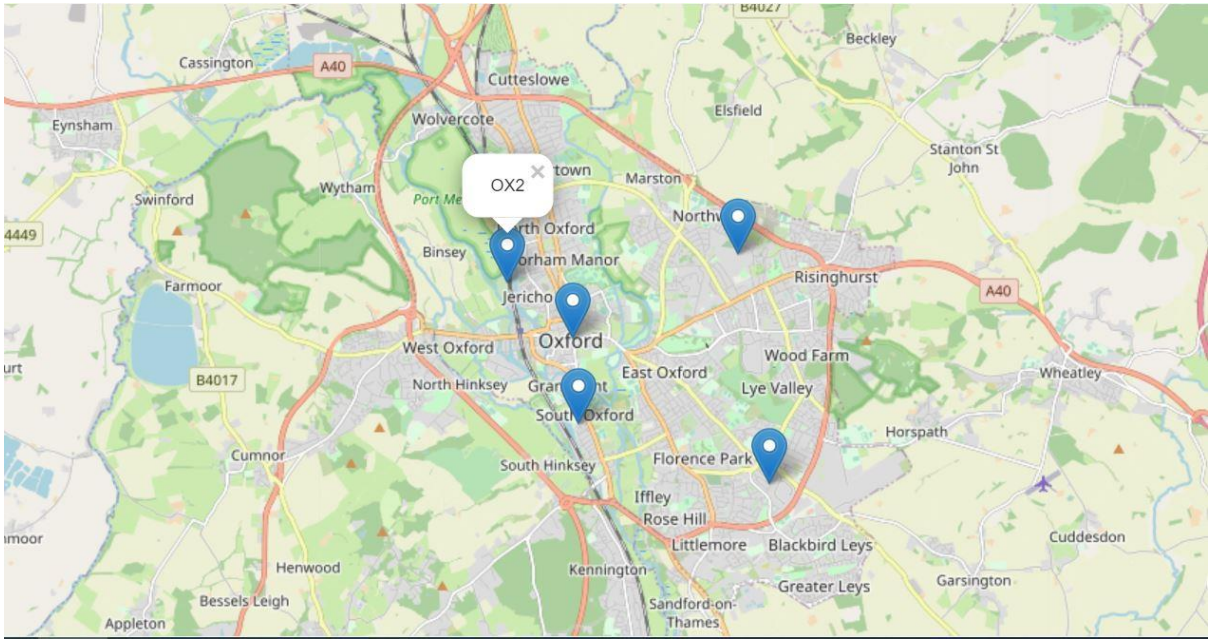

Figure 2: Map of Oxford City and Postcode districts areas.

2.2 Foursquare API

This project uses the Foursquare API as its prime data gathering source as it has a database of millions of places, especially their Places API that provides the ability to perform location search, location sharing and details about various venues. Requests are made to the API to search and explore venue recommendation data nearby the targeted location coordinates (Oxford centre). The information retrieved would be the venue name, venue ID, venue latitude and longitude and venue categories. The venue ID is used to explore more in depth into the venue details and retrieve the venue ratings, price tier and the number of likes. These details will be used to explore which places are good to visit based on their price tier and ratings and the number of likes previous customers have given to the venue. The dataframes for the venues (name, categories, location (lat & long) and venue id) and venue details (price, rating and likes) are merged into one dataframe. Venues were represented on the Oxford map using Folium.

To use the Foursquare API to request data, client credentials (client id and client secret) are required; this is acquired from the [Foursquare developers website] once you sign up. Using the credentials and the foursquare requests URL, requests for nearby venue information can be mined. Request limitations for the number of places per neighbourhood parameter would be set to 100, the radius parameter would be set to 2200m and the request version date will be the current date.

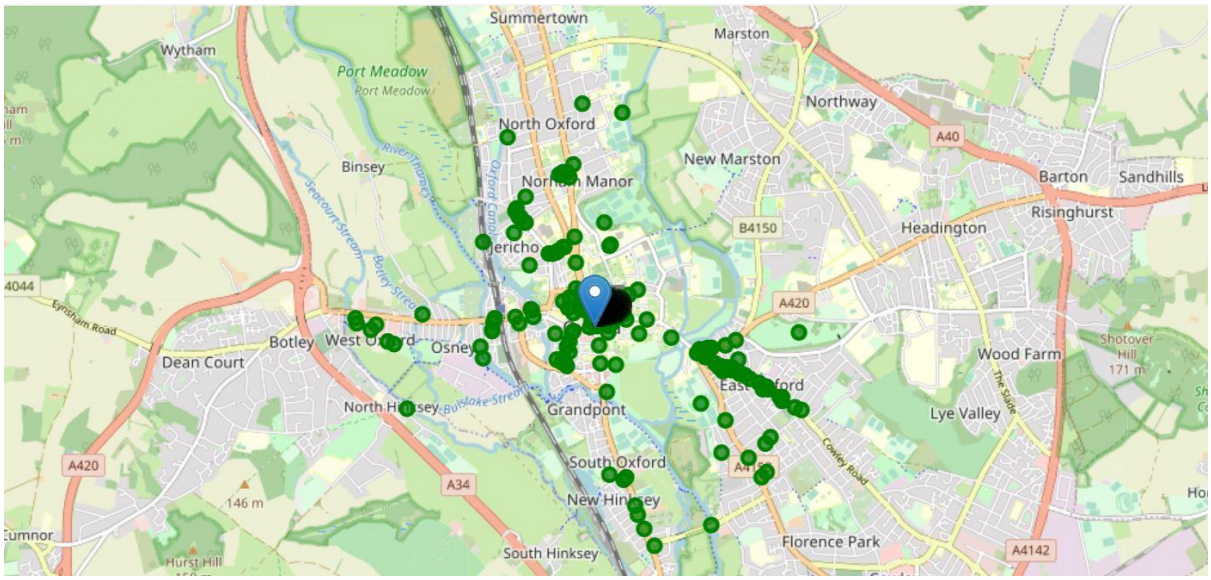

Figure 3: Map of Oxford and Venues retrieved from the Foursquare API.

The number of venues retrieved was 162, but after data cleaning by removing any venues that does not contain any ratings, there was 104 venues.

	name	categories	lat	lng	ID	Likes	Rating	Tips	Price
0	Ben's Cookies	Dessert	51.752419	-1.256478	4b7da82ef964a520b2cc2fe3	78	9.0	23	1
2	The Handle Bar	NaN	51.753216	-1.260391	4df0e3cd7d8ba370a0113561	127	8.7	60	2
3	Society Cafe	Coffee	51.753504	-1.259350	56eaad58498ed08ece59c010	101	8.6	32	1
4	George & Danver	Ice Cream	51.750517	-1.256960	4b1f7212f964a520522624e3	102	8.6	35	2
9	Paul	Bakery	51.752091	-1.257252	54957435498e2fe76c0d33ee	39	8.5	3	1
11	The Bear Inn	Pub	51.751578	-1.255764	4b8948f0f964a520782832e3	123	8.5	26	1
12	Mowgli Street Food	Indian	51.750871	-1.260887	5aaaa05360255e5d97b0b8c2	13	8.5	4	2
15	The Vaults & Garden Café	Café	51.753019	-1.253589	4bb610002ea195213911ab2f	168	8.4	63	2
17	The White Horse	Pub	51.754501	-1.255583	4b672755f964a520203e2be3	50	8.3	10	1
18	Fernando's Cafe	Brazilian	51.752022	-1.258028	4f3a4428e4b0b6d3bb919636	32	8.2	17	2
19	The White Rabbit	Pizza Place	51.754432	-1.260884	4b647488f964a52087b42ae3	88	8.4	21	1
22	The Alternative Tuck Shop	Sandwich Place	51.755106	-1.251797	4c7511c56f789c748b97484c	36	8.7	11	1
23	Moo-Moo's	Juice Bar	51.752415	-1.256812	4b6ad828f964a52025e32be3	32	8.1	12	1
24	George Street Social	Coffee	51.753596	-1.261043	56bdadd7498e51112716eb51	49	8.2	14	1
25	Colombia Coffee Roasters	Café	51.752469	-1.256510	58496119d6fe9020cb1da0f2	12	8.0	4	1
27	The Turf Tavern	Pub	51.754657	-1.253032	4af6e596f964a520f60322e3	422	8.3	104	2
29	Leon	NaN	51.753249	-1.258501	57ff5bd7d67cd3f0ca48353e	20	8.0	3	2
30	Jericho Coffee Traders	Coffee	51.752301	-1.255252	558bf729498e1eeee39b6683	34	8.0	6	1

Figure 4: Cleaned Oxford venues data.

3. Data Analysis and Exploration

Analysis and exploration of the data was undertaken using various visualisation techniques to better understand the venue categories, ratings, price tiers and likes.

3.1 Venue Categories

The venues were grouped by categories and the number of venues per category determined. The most number of venues is Pubs (24) followed by Coffee shops (7) and Cafés (6).

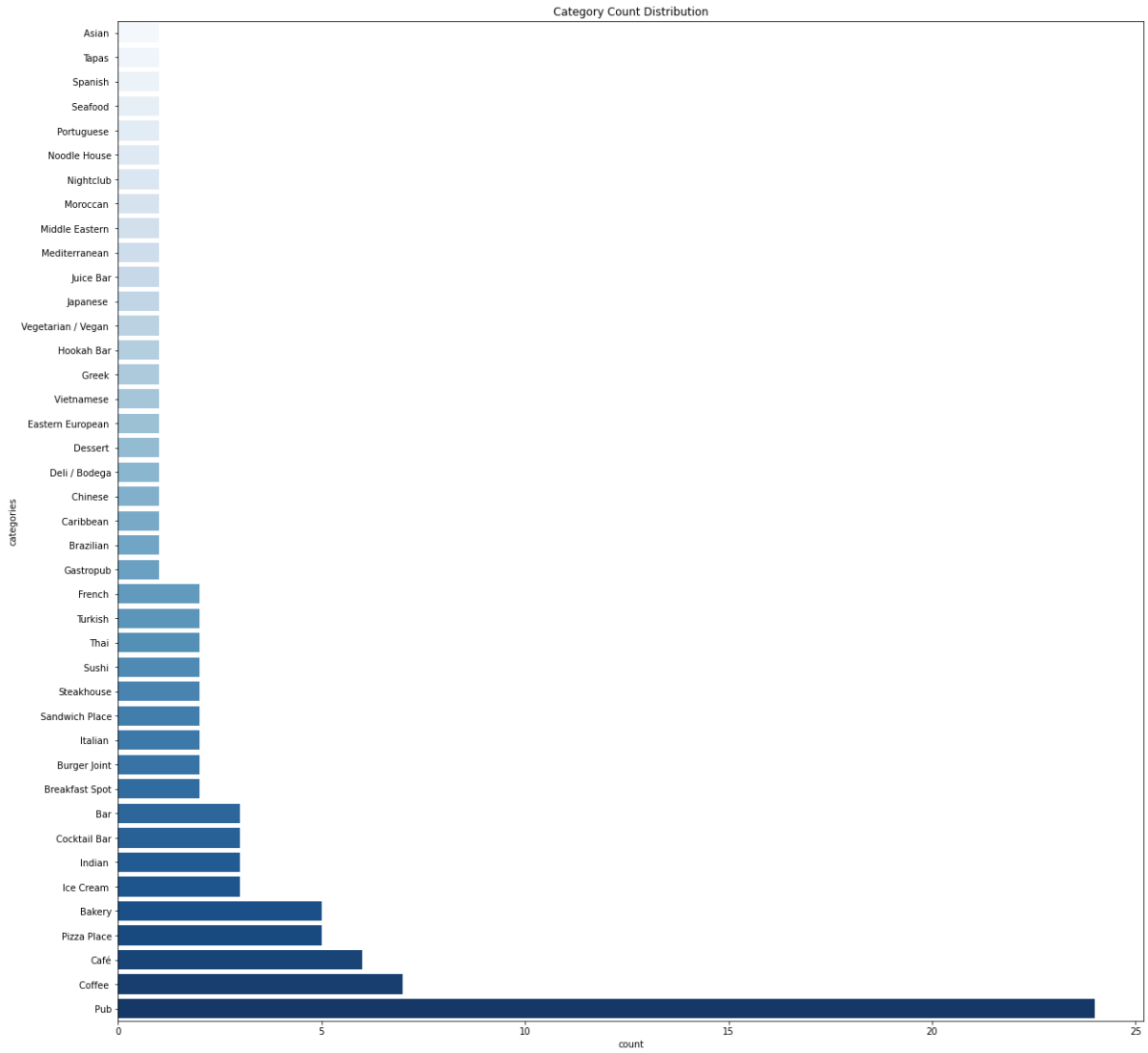

Figure 5: Number of Venue by Category.

A WordCloud was used to visually representation of text data in the categories column. It displays a list of words, the importance of each shown with font size or colour. This format is useful for quickly perceiving the most prominent terms. This also shows that Pubs, Coffee, Bars and Cafés are prominent venue categories.

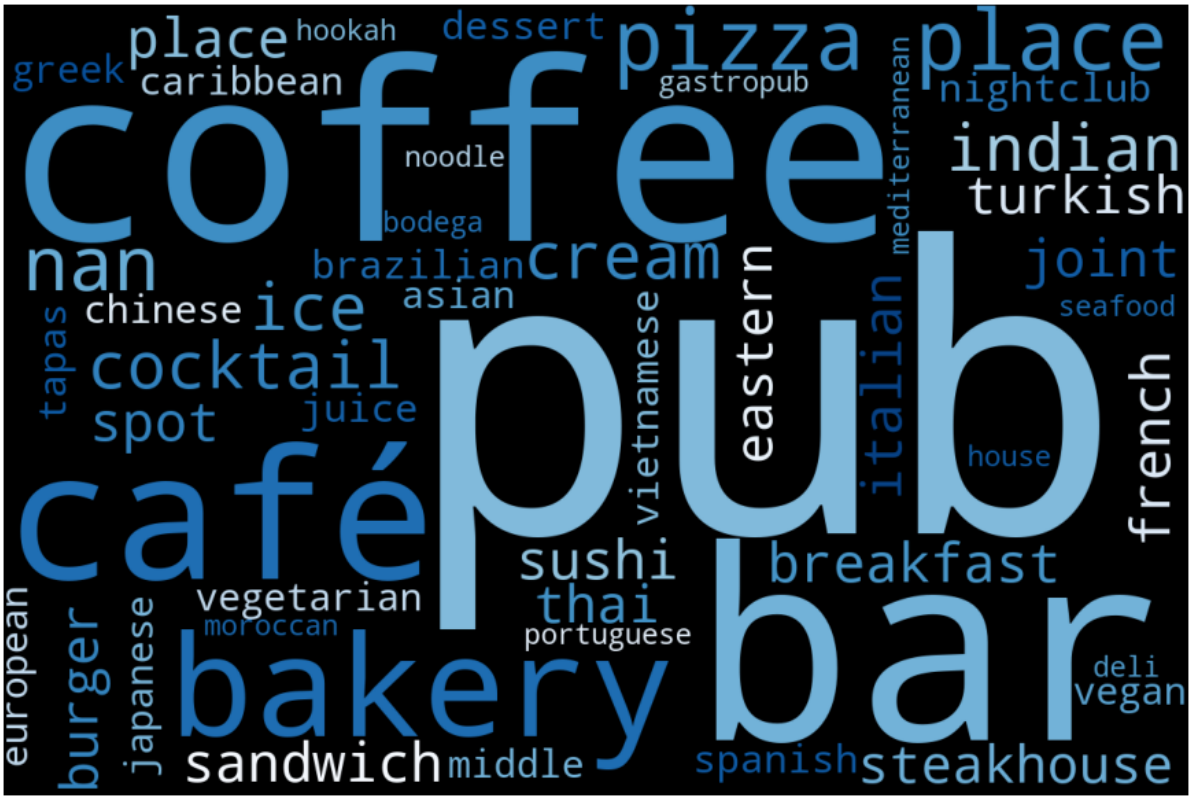

Figure 6: WordCloud showing most prominent words in the Categories type.

3.2 Venue Rating

The venue ratings was analysed, the average ratings is 7.84, with a Min of 6.1 and Max of 9.1. This number of venues with a given rating is represented below.

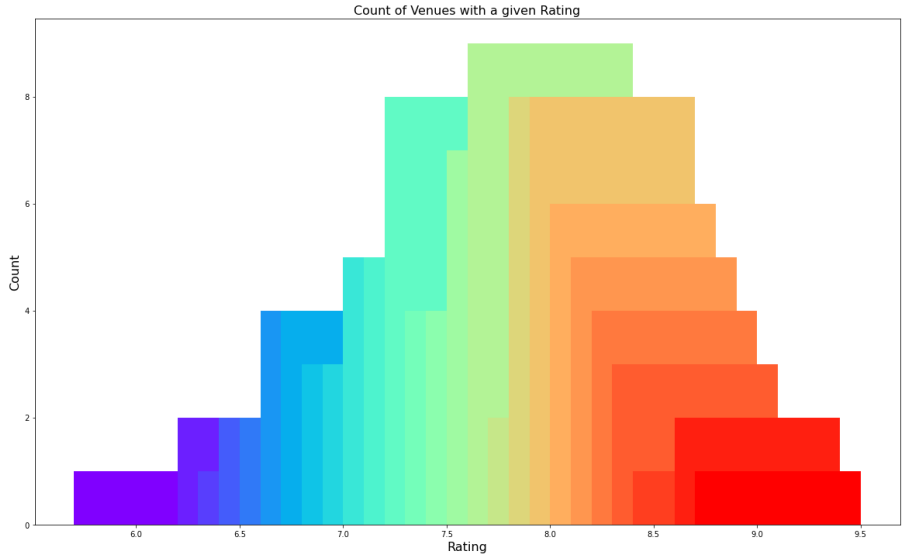

Figure 7: Number of Venues by Rating.

The venue ratings were binned into 5 bin sizes representing, ‘Low’, ’Okay’, ‘Good’, ‘Very good’ and ‘Excellent’ ratings. The number of venues in rating range is shown below, most of the venues fall into the Good and Very Good rating bin. The venues can be represented on the Oxford map by colour coding the venues by ratings, a popup marker with the name, category and rating of the venue.

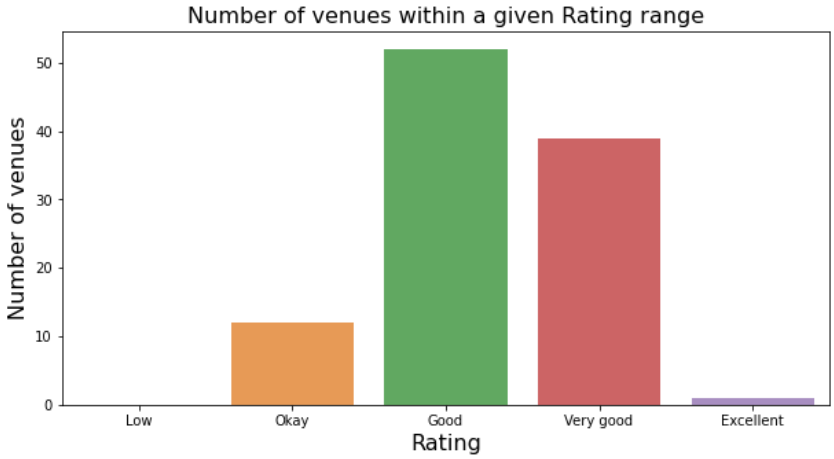

Figure 8: Number of Venues within a given Rating bin.

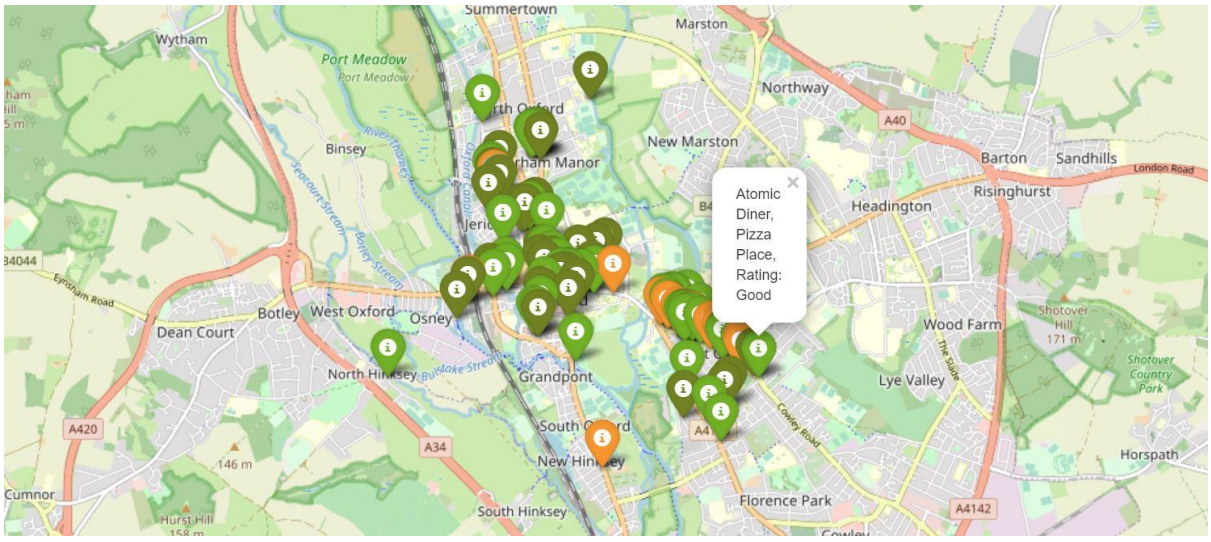

Figure 9: Map of Oxford showing Venues based on their Ratings; Low – Red, Okay – Orange, Good – Green, Very Good – Dark green and Excellent – Light Blue.

The Top 10 venues and lowest 10 venues is represented below, with Branca (Italian) being the top rated venue and Kite Inn being the lowest rated venue.

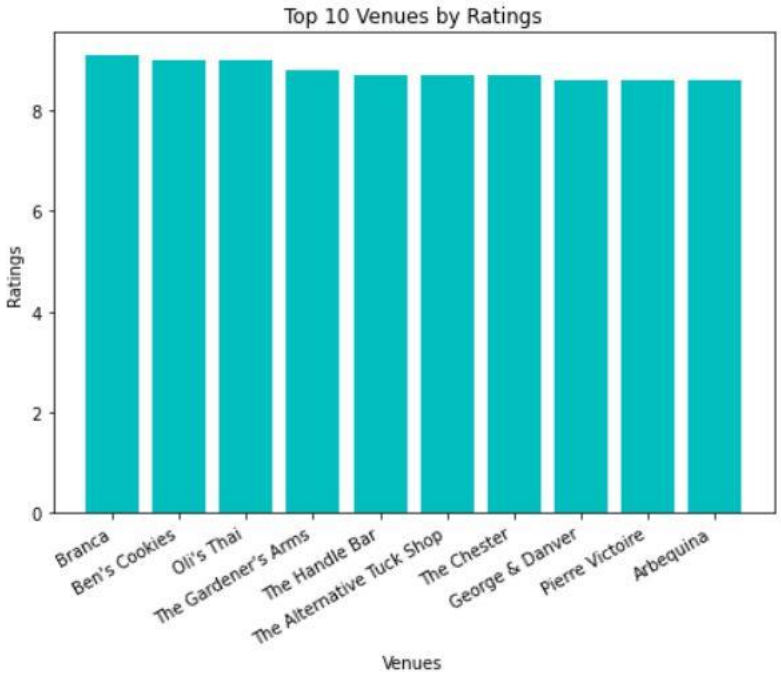

Figure 10: Top 10 Venues by Rating.

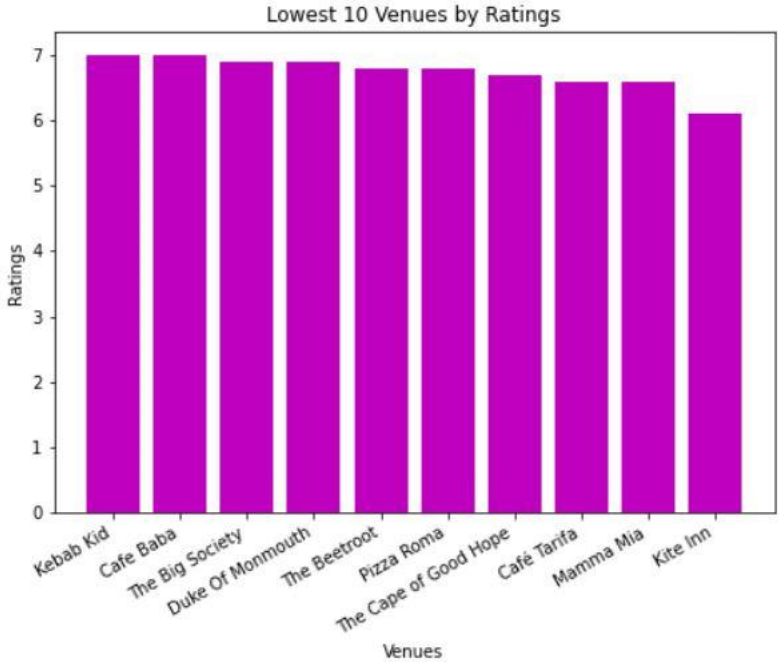

Figure 11: Lowest 10 Venues by Rating.

3.3 Venue Price

The venue price tier data was analysed, the price tier is represented by four tiers (1 cheapest – 4 most expensive). The venue Ratings per Price tier and the number of venues is represented below, both showing that largest amount of venues are in Price tier 1 and has the largest rating variability.

Compared to Price tier 4, which has the lowest number of venues (2) and lowest rating variability. The price tier can be represented on the Oxford map by colour coding, where the Fishes Gastropub and the Porterhouse & Grill Steakhouse are the most expensive restaurants.

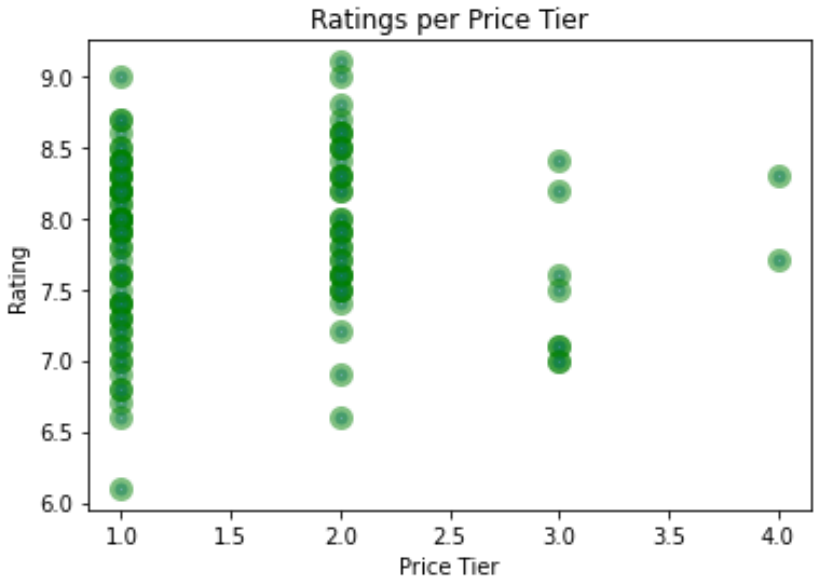

Figure 12: Comparing the Ratings and Price Tier.

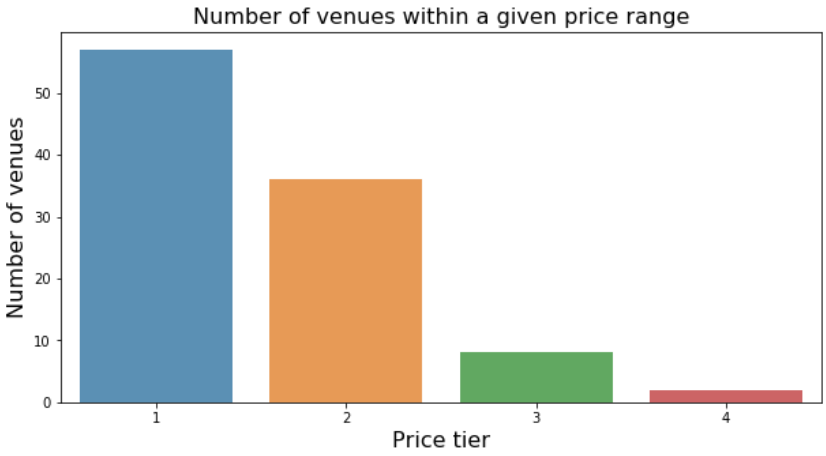

Figure 13: Number of Venues in a given Price Tier; 1 – cheapest and 4 – most expensive.

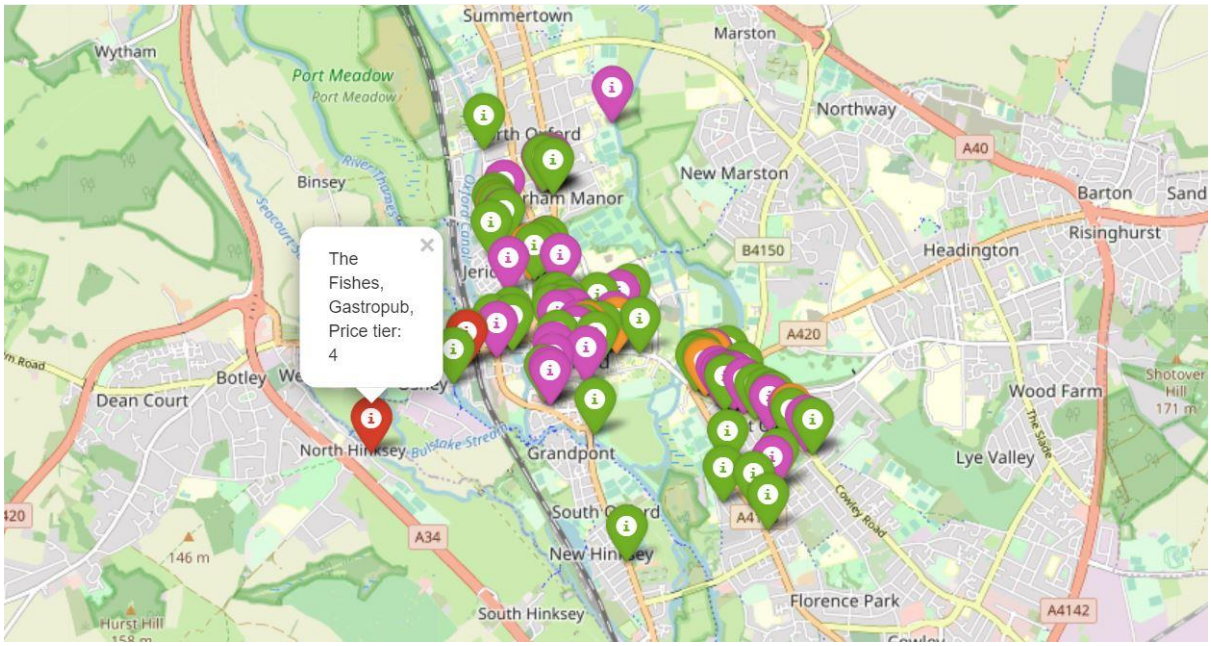

Figure 14: Map of Oxford showing Venue by Price tier; 1 – Green, 2 – Purple, 3 – Orange and 4 – Red.

3.4 Venue Likes

The venues can be analysed by the number of likes, the top and lowest 10 venues is shown in the figures below, showing that the Turf Tavern Pub is the most popular, followed by the Eagle & Child Pub. The least liked venue is the Roma Pizza followed by Kite Inn.

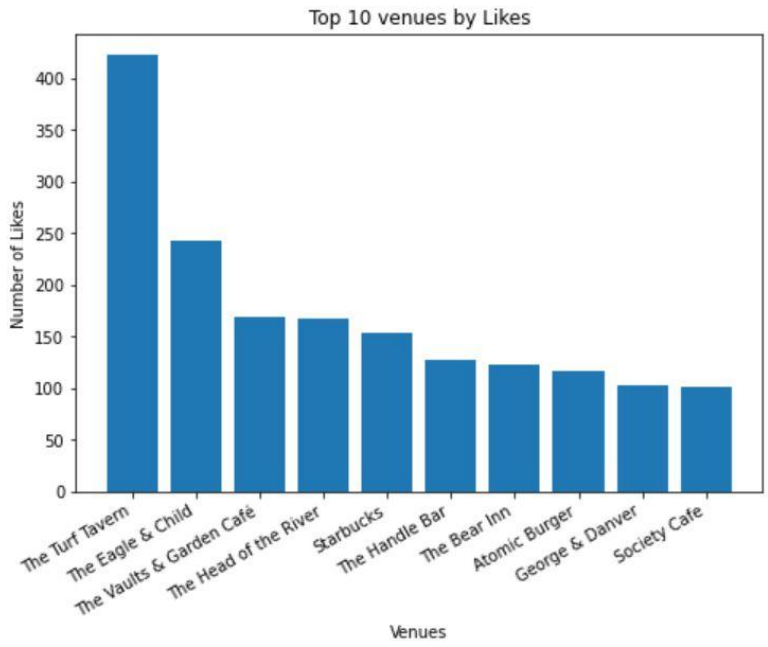

Figure 15: Top 10 Venues by Likes.

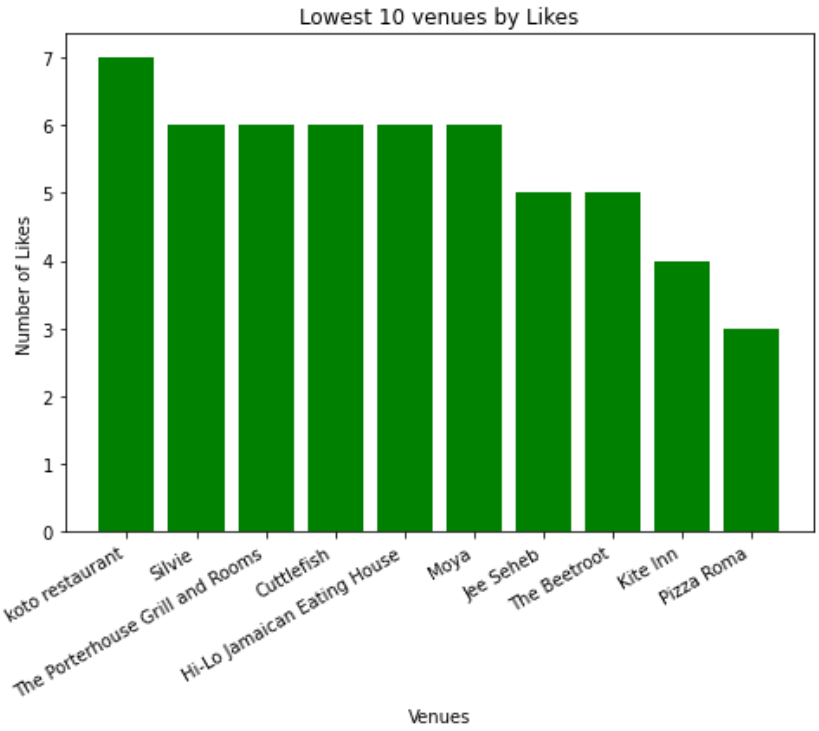

Figure 16: Lowest 10 Venues by Likes.

3.5 Clustering

The venue data can be clustered into various groups based on similarities in the data. Using a form of unsupervised machine learning, the K-Means clustering algorithm method imported from the Scikit library can be used to cluster the venue information. Venues are clustered by the location, ratings, price and likes into three clusters, Cluster 0, Cluster 1 and Cluster 2. The clustered venues is represented on the Oxford map by colour, Cluster 0 - Green, Cluster 1 - Red and Cluster 2 – Purple, cluster centres are larger circles.

Cluster 0 represents venues with low number of likes, relatively cheap (average = 1.53) and has an average rating of 7.75.

Cluster 1 represents venues with a high number of likes, medium priced (average = 2.0) and has an average rating of 7.95.

Cluster 2 represents venues with an medium number of likes, medium priced (average = 1.62) and has the highest average rating of 8.32.

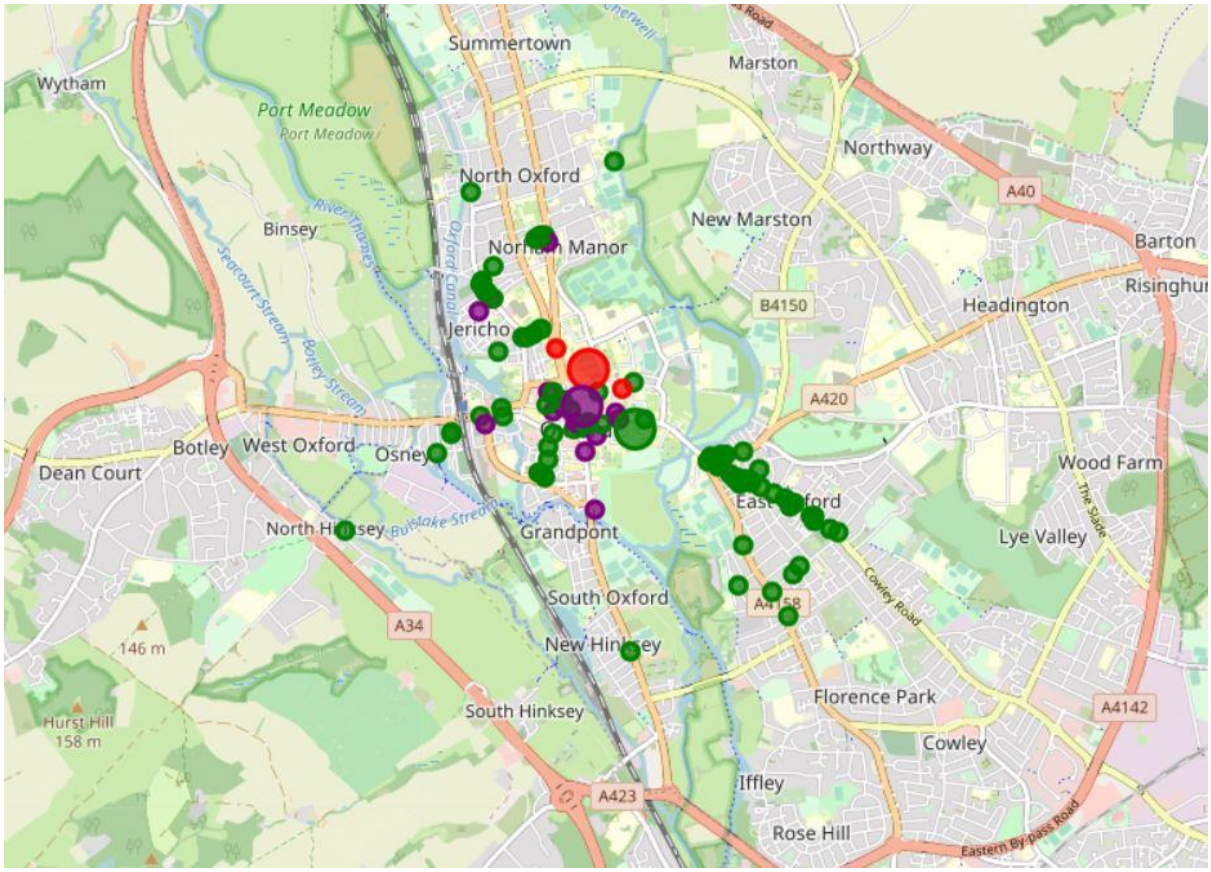

Figure 17: Map of Oxford showing Venues clustered (Cluster 0 - Green, Cluster 1 – Red and Cluster 2 - Purple).

```
result = final_venues2[final_venues2['cluster_labels'] == 0]
print("Cluster 0 : ")
result.head(10).reset_index(drop = True)
```

```
Cluster 0 :
```

	cluster_labels	name	categories	lat	lng	ID	Likes	Rating	Tips	Price	Rating_bin
0	0	Paul	Bakery	51.752091	-1.257252	54957435498e2fe76c0d33ee	39	8.5	3	1	Very good
1	0	Mowgli Street Food	Indian	51.750871	-1.260887	5aaaa05360255e5d97b0b8c2	13	8.5	4	2	Very good
2	0	The White Horse	Pub	51.754501	-1.255583	4b672755f964a520203e2be3	50	8.3	10	1	Very good
3	0	Fernando's Cafe	Brazilian	51.752022	-1.258028	4f3a4428e4b0b6d3bb919636	32	8.2	17	2	Very good
4	0	The Alternative Tuck Shop	Sandwich Place	51.755106	-1.251797	4c7511c56f789c748b97484c	36	8.7	11	1	Very good
5	0	Moo-Moo's	Juice Bar	51.752415	-1.256812	4b6ad828f964a52025e32be3	32	8.1	12	1	Very good
6	0	George Street Social	Coffee	51.753596	-1.261043	56bdadd7498e51112716eb51	49	8.2	14	1	Very good
7	0	Colombia Coffee Roasters	Café	51.752469	-1.256510	58496119d6fe9020cb1da0f2	12	8.0	4	1	Good
8	0	Leon	NaN	51.753249	-1.258501	57ff5bd7d67cd3f0ca48353e	20	8.0	3	2	Good
9	0	Jericho Coffee Traders	Coffee	51.752301	-1.255252	558bf729498e1eeee39b6683	34	8.0	6	1	Good

```
print("These venues for cluster 0 have mean price tier of {:.02f} and rating spread around {:.02f}".
      format(result['Price'].mean(), result['Rating'].astype(float).mean()))
```

```
These venues for cluster 0 have mean price tier of 1.53 and rating spread around 7.75
```

```
result = final_venues2[final_venues2['cluster_labels'] == 1]
print("Cluster 1 : ")
result.head(10).reset_index(drop = True)
```

```
Cluster 1 :
```

	cluster_labels	name	categories	lat	lng	ID	Likes	Rating	Tips	Price	Rating_bin
0	1	The Turf Tavern	Pub	51.754657	-1.253032	4af6e596f964a520f60322e3	422	8.3	104	2	Very good
1	1	The Eagle & Child	Pub	51.757285	-1.260039	4b4ddde0f964a520c6d926e3	243	7.6	58	2	Good

```
print("These venues for cluster 1 have mean price range of {:.02f} and rating spread around {:.02f}".
      format(result['Price'].mean(), result['Rating'].astype(float).mean()))
```

```
These venues for cluster 1 have mean price range of 2.00 and rating spread around 7.95
```

```
result = final_venues2[final_venues2['cluster_labels'] == 2]
print("Cluster 2 : ")
result.head(10).reset_index(drop = True)
```

```
Cluster 2 :
```

	cluster_labels	name	categories	lat	lng	ID	Likes	Rating	Tips	Price	Rating_bin
0	2	Ben's Cookies	Dessert	51.752419	-1.256478	4b7da82ef964a520b2cc2fe3	78	9.0	23	1	Very good
1	2	The Handle Bar	NaN	51.753216	-1.260391	4df0e3cd7d8ba370a0113561	127	8.7	60	2	Very good
2	2	Society Cafe	Coffee	51.753504	-1.259350	56eaad58498ed08ece59c010	101	8.6	32	1	Very good
3	2	George & Danver	Ice Cream	51.750517	-1.256960	4b1f7212f964a520522624e3	102	8.6	35	2	Very good
4	2	The Bear Inn	Pub	51.751578	-1.255764	4b8948f0f964a520782832e3	123	8.5	26	1	Very good
5	2	The Vaults & Garden Café	Café	51.753019	-1.253589	4bb610002ea195213911ab2f	168	8.4	63	2	Very good
6	2	The White Rabbit	Pizza Place	51.754432	-1.260884	4b647488f964a52087b42ae3	88	8.4	21	1	Very good
7	2	Starbucks	Coffee	51.752461	-1.257859	4afc10daf964a520af1f22e3	154	7.8	21	2	Good
8	2	Branca	Italian	51.760642	-1.266777	4bad08e9f964a520d3263be3	84	9.1	25	2	Excellent
9	2	The Head of the River	Pub	51.746693	-1.255910	4b86b3cef964a520479731e3	167	7.9	41	1	Good

```
print("These venues for cluster 2 have mean price range of {:.02f} and rating spread around {:.02f}".
      format(result['Price'].mean(), result['Rating'].astype(float).mean()))
```

```
These venues for cluster 2 have mean price range of 1.62 and rating spread around 8.32
```

4. Discussion

From the above analysis, there are several observations for someone visiting the City of Oxford in England, UK. From 162 venues retrieved from the Foursquare database in a 2.2km radius from the centre of Oxford only 104 venues had any ratings. The most number of venues are Pubs, this is not surprising, as the British love their pubs, and Coffee shops and Cafes follow this. This result is confirmed in the WordCloud generated from the category type, which shows the most prominent words is Pub, Coffee, Bar and Café.

Venues were analysed based on their ratings, price tier and likes. Venue ratings are typically in the range of 6.1 to 9.1 out of 10, with an average of 7.85, using the ratings, venues were binned into 5 groups; Low, Okay, Good, Very Good and Excellent. Most venues fell in the Good or Very Good rating category, while only 1 venue (Branca, Italian) had the highest rating of 9.1. The lowest rated venue was the Kite Inn with 6.1.

On price level, there were 4 price tiers with 1 being the lowest (cheapest) and 4 highest (expensive) price. A significant portion of the venues fell into tier 1 with only 2 venues in tier 4, the Fishes

(Gastropub) and the Porterhouse Grill & Rooms (Steakhouse). There was a wide range of ratings for the different tier levels.

The venue with the most likes is the Turf Tavern (Pub), followed by the Eagle & Child (Pub) and the Vaults & Garden Café. Venues with the least amount of likes is Roma Pizza and Kite Inn.

Finally, venues were clustered based on their location, ratings, price tier and likes into three clusters using the K-Means clustering algorithm.

5. Conclusion

The purpose of this project was to explore the venues food options anyone visiting Oxford city, England in the United Kingdom can enjoy. Several venues have been identified using the Foursquare API and have been plotted on an interactive map. Using different visualisation and analysis methods, a range of options was presented for people to choose from while exploring Oxford; these are based on ratings, price and likes.

6. References

Oxfordshire postcode areas - https://en.wikipedia.org/wiki/OX_postcode_area

Foursquare developer's website