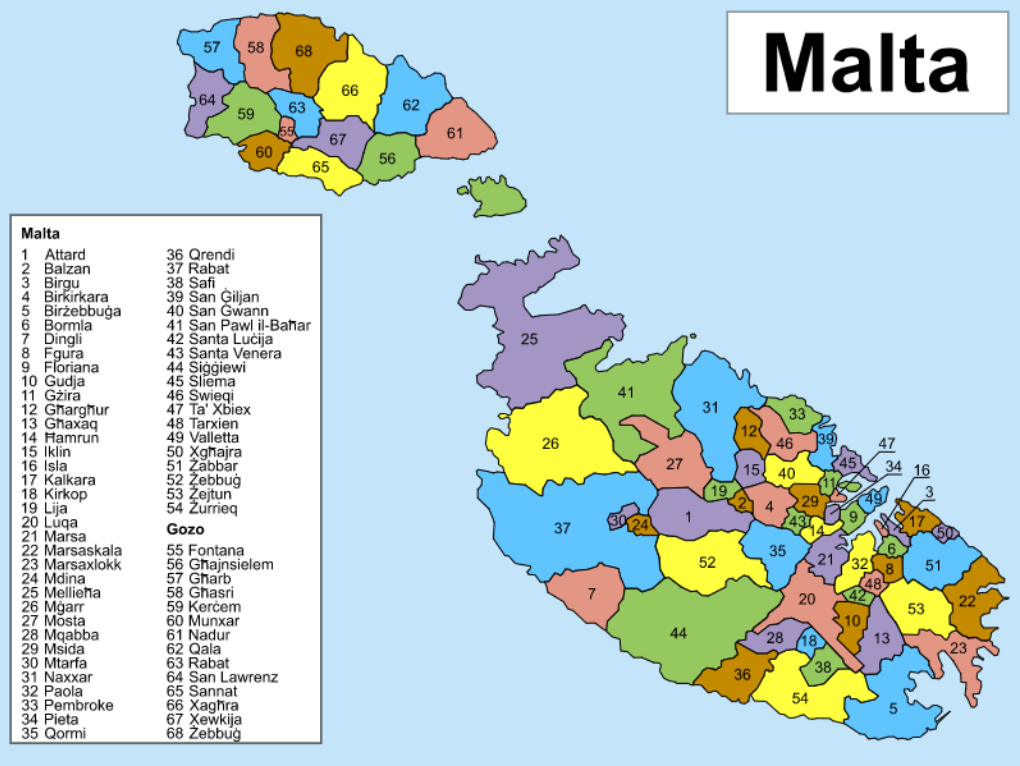
Capstone Project – Restaurants in Malta

Saad Sharif

A. Introduction and Business Problem

Malta is an island nation located in the Mediterranean Sea which lies 80 kilometres south of Sicily. It has an approximate population of **515,000** over an area of **316 square kilometres**. The nation consists of the mainland Malta, Gozo island and a smaller island Comino. Malta has **68 local councils**, across the three islands, all with varying population densities. It is a popular holiday destination with tourists frequently visiting the country every year for its culture, history, leisure activities and weather.



We will be using geographical data on existing venues in Malta, and how close they are to nearby historical sites, so that we can identify the ideal location to open a new restaurant. We will also analyse whether there is a relationship between the proximity to historical sites and the approximate population density of the area. Our analysis will be useful to stakeholders who are interested in getting into the restaurant business in a popular holiday destination and would like to identify optimal places that will positively influence their restaurant's rating.

B. Data and Variables

To tackle the business problem, we have obtained data from the following sources:

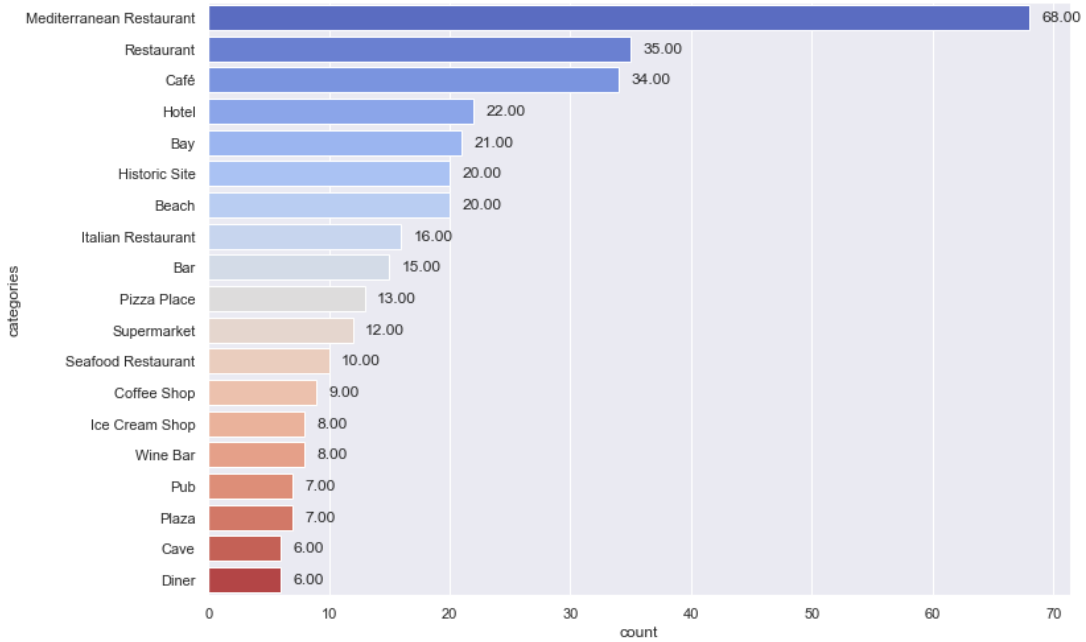
- **Foursquare API** has been used to extract common venues in Malta [1]
- Population Density of each Local Council has been extracted from **Wikipedia** [2]
- List of coordinates from the **European Environment Agency** has been used to test our model against [3]

C. Methodology

The Foursquare API returned **527 distinct venues** in Malta after running it for 100 evenly spread coordinates with a radius of 10 kilometres. The API also pulls the name of each venue and what category it falls in, alongside the latitude and longitude.

	id	name	categories	lat	lng	Label
0	4f3cf1e7e4b02d5ef815d185	The Cliffs	Mediterranean Restaurant	35.853782	14.379652	2
1	4d207fab756e8cfac6d36c54	Bobbyland Bar & Restaurant	Mediterranean Restaurant	35.858796	14.372551	2
2	52ff4d4d498e3e88d44c3eaf	Diar il-Bniet	Mediterranean Restaurant	35.862120	14.384453	2
3	4b6d45f2f964a520b06e2ce3	Ħaġar Qim Temples	Historic Site	35.827639	14.442086	1
4	4dd91ef852b1a5c644753665	Mnajdra Temples	Historic Site	35.826643	14.436346	1
5	4f93ebe4e4b03cc49d3b7963	Hagar Qim Restaurant	Diner	35.826932	14.443263	2
6	502d2f7be4b0eed9c27ec2c1	Talbot & Bons	Wine Bar	35.851223	14.494721	2
7	5974c26cf96b2c032000cdd8	Districtfive	Modern European Restaurant	35.825526	14.530570	2
8	4bca13c70687ef3b7135dbcc	Ferretti Restaurant	Restaurant	35.832365	14.533964	2
9	524d4859498ea4a1ee258b1e	Zen Sushi to go	Sushi Restaurant	35.850970	14.494697	2
10	558eafec498e241c914fe3e7	Vecchia Napoli	Italian Restaurant	35.851031	14.494636	2

The most common occurring category is “Mediterranean Restaurant” with 68 occurrences, followed by just “Restaurant” with 35 occurrences and “Café” with 34 occurrences. For the purpose of our analysis, we will consider categories that match on terms like “Restaurant”, “Café”, “Bar”, “Diner”, “Pizza” and “Coffee” by the definition of **“Restaurant”** as these will most likely fall into the definition of food category.



Furthermore, “Historic Site” will be used as the venues representing historical points of interests for tourists. This category appears 20 times in the dataset. We can calculate how close a restaurant is to historical sites using the following formula:

Historic Proximity: $\frac{n^2}{\sum_{i=1}^{n} |x-h_i|}$

Where **x** represents a restaurant, **h** represents a historical site, **n** is the number of historical sites nearby, and **|x-h|** is the distance between the restaurant and each historical site.

In layman's terms, the formula calculates the average distance of nearby historic sites to a given restaurant, within a given radius, divided again by the number of historic sites. It then calculates the inverse of this value, so that a high number means that a venue has many historic sites nearby.

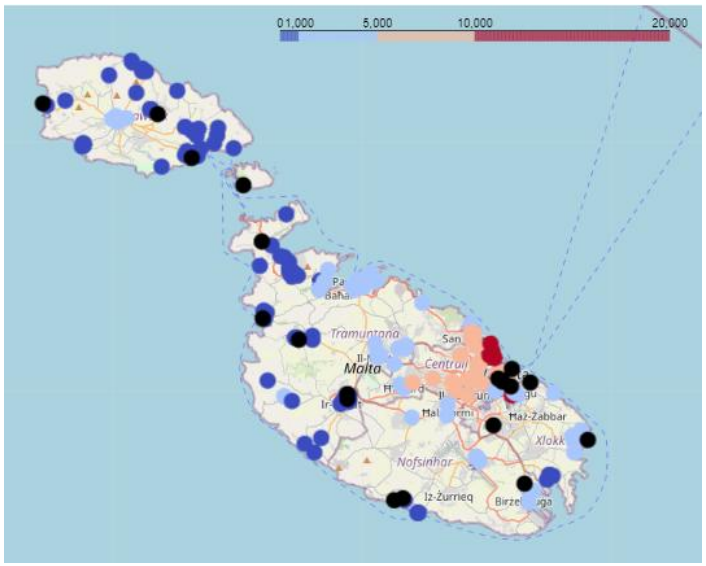
Additionally, the population density of each local council has been used to get an approximate value for the area each venue is in. Longitude and latitude of each Council has been obtained from Wikipedia and the population density to the closest council coordinates has been used for each restaurant. As a result, this is an “approximation” and can be further fine-tuned with more available and accurate population data.

After reducing the data to “food categories” and historical sites, we get **263 venues**. Once we get the local council, population density and historic proximity, the table now looks like the below:

id	name	categories	lat	lng	Label	Historic_Proximity	Council	Population_Density
4f3cf1e7e4b02d5ef815d185	The Cliffs	Mediterranean Restaurant	35.853782	14.379652	2	0.971451	Dingli	651.052632
4d207fab756e8cfac8d36c54	Bobbyland Bar & Restaurant	Mediterranean Restaurant	35.858796	14.372551	2	0.966896	Dingli	651.052632
52ff4d4d498e3e88d44c3eaf	Diar il-Bniet	Mediterranean Restaurant	35.862120	14.384453	2	1.064673	Dingli	651.052632
4f93ebe4e4b03cc49d3b7963	Hagar Qim Restaurant	Diner	35.826932	14.443263	2	1.426688	Qrendi	571.428571
502d2f7be4b0eed9c27ec2c1	Talbot & Bons	Wine Bar	35.851223	14.494721	2	2.262681	Gudja	1384.347826
5974c26cf96b2c032000cdd8	Districtfive	Modern European Restaurant	35.825526	14.530570	2	1.590556	Birzebbuga	1403.804348
4bca13c70687ef3b7135dbcc	Ferretti Restaurant	Restaurant	35.832365	14.533964	2	1.710038	Birzebbuga	1403.804348
524d4859498ea4a1ee258b1e	Zen Sushi to go	Sushi Restaurant	35.850970	14.494697	2	2.257499	Gudja	1384.347826
558eafec498e241c914fe3e7	Vecchia Napoli	Italian Restaurant	35.851031	14.494636	2	2.258572	Gudja	1384.347826
4e3991d8e4cd799aaef2817a	Al Fresco	Restaurant	35.832379	14.530586	2	1.736452	Birzebbuga	1403.804348

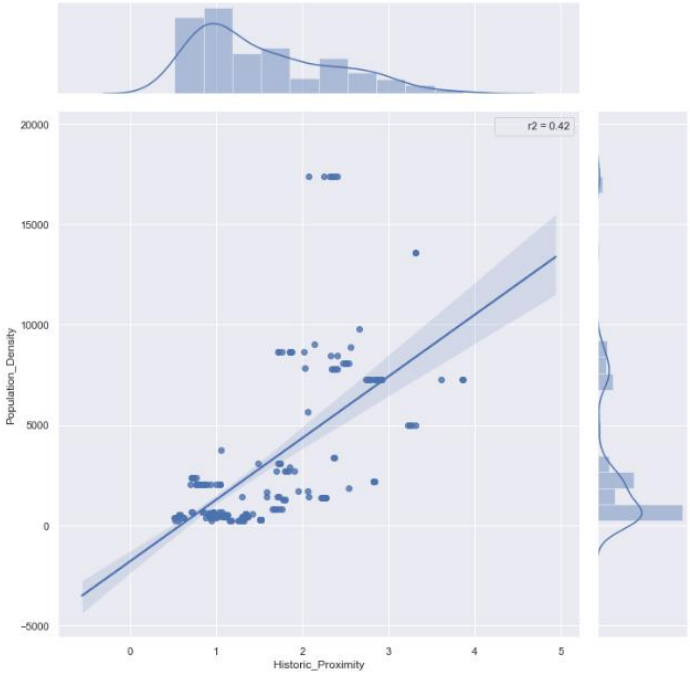
Initial Plot

Here we have plotted a geographical map which illustrates where each restaurant is located in Malta with colours to indicate the approximate population density. Additionally, the historic sites have also been plotted as black circles. It is interesting to note that the area with the most historic sites, near the capital **Valletta**, is also associated to the highest population density in Malta.



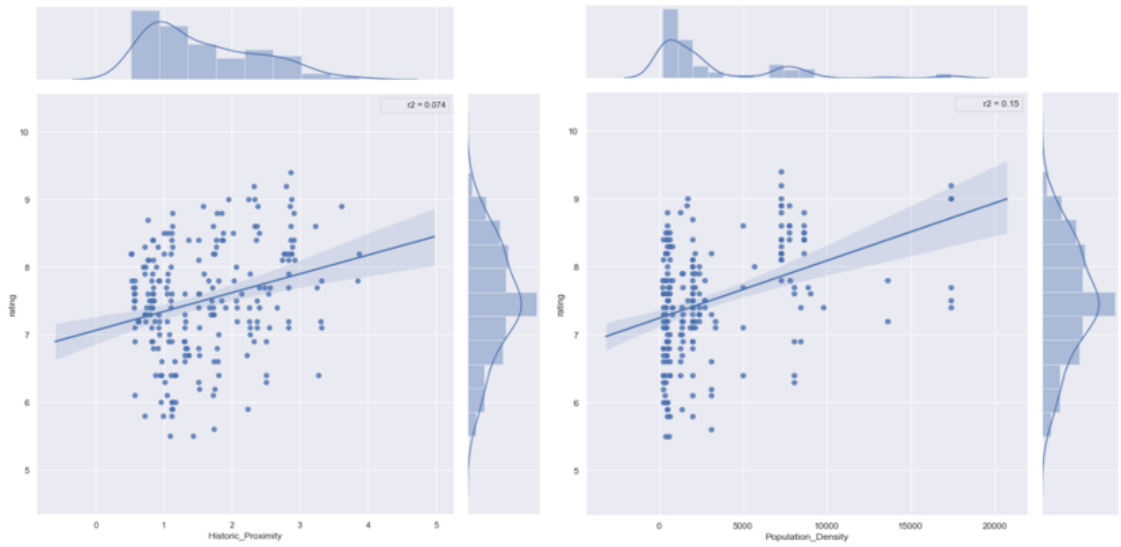
Regression Analysis

Let's observe the relationship between the historic proximity and the population density. Below we have created a scatter plot with a regression line so that the relationship can be easily evaluated. Clearly there is, although not strong, a positive correlation between the two variables as the **R-squared is 0.42**. This makes sense as these points of interest would attract tourists to the area and therefore cause the population density to rise. Outliers to this pattern could be tourist attractions that may not necessarily represent historical sites, such as beaches and water parks.



The following two scatter plots analyse the correlation between the historic proximity and population density against the dependent variable restaurant ratings.

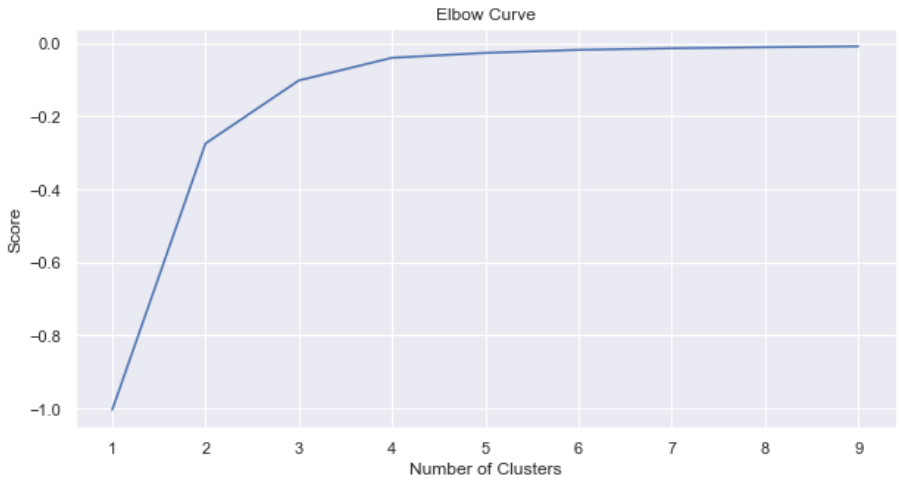
We can see that there is a very slight positive correlation in both graphs but the R-squared values are very small. The population density has a slightly steeper gradient which indicates its influence on the restaurant ratings is marginally greater. Furthermore, the R-squared value between the population density and the rating is **0.15** which is also slightly greater than the R-squared between the historic proximity and ratings which is only **0.074**.



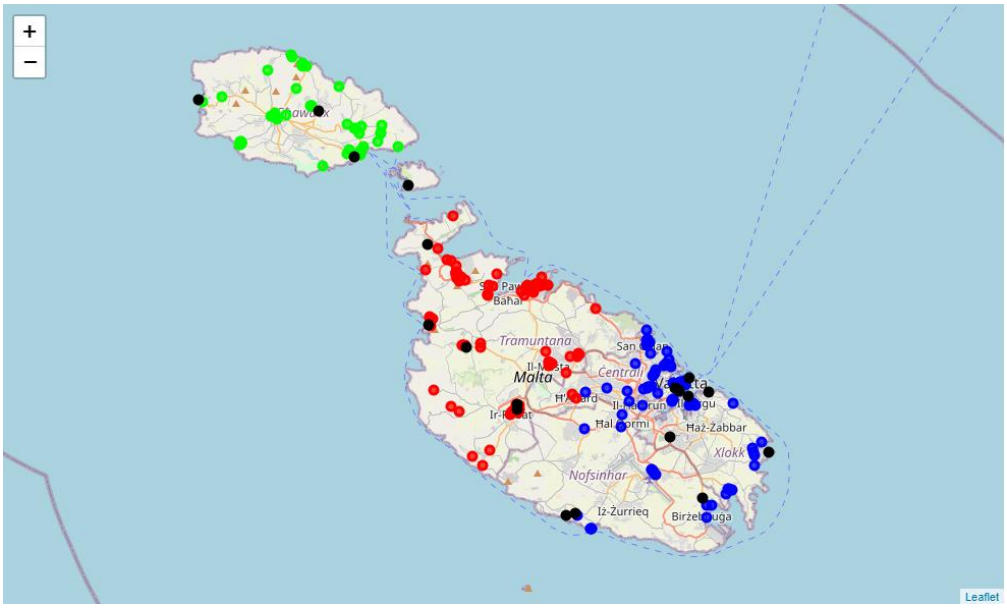
The low R-Squared values mean that the distance from historic sites and the nearby population density are not good variables to be used alone in order to predict the restaurants rating. Perhaps there are a lot more other variables at play here that are not being included in our analysis. This could include factors like the restaurants having a good view, being near beaches, price and services within the restaurant too.

K-Means Clustering

We can apply the K-Means algorithm to see how the restaurants cluster together in relation to historic sites. First, we must plot the **elbow curve** to identify the best number of clusters to use in the algorithm. As we can see, the **inflection point** is at 3. This is where diminishing returns mean any additional clusters are no longer worth their benefit to the model.

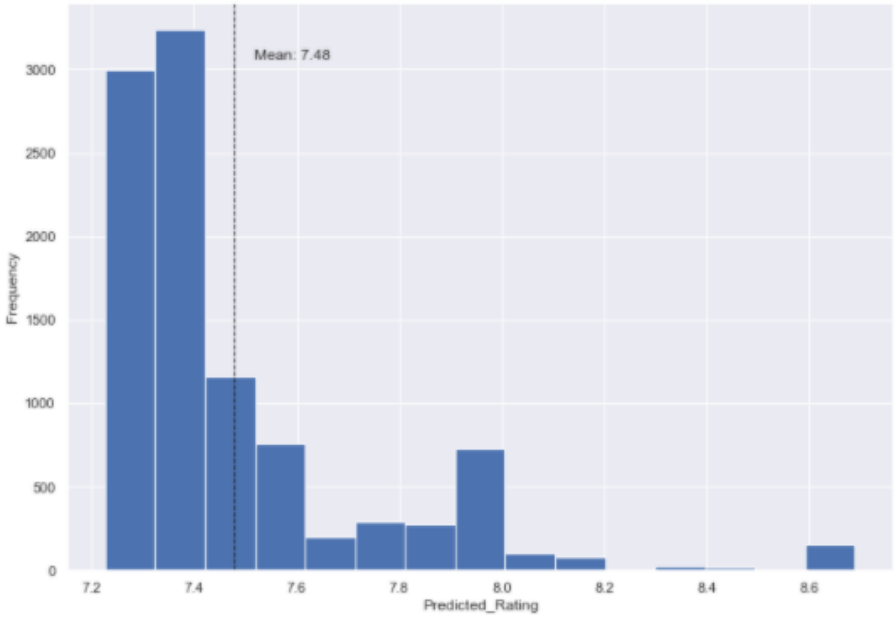


The output of the K-means algorithm is the below geographical plot. It is not surprising that one entire cluster (green) has been found in the island of Gozo. The remaining two clusters are found in the mainland Malta. The second cluster (red) represents restaurants located in the north-western region of the mainland, whilst the third cluster (blue) represents restaurants located in the south-eastern region.



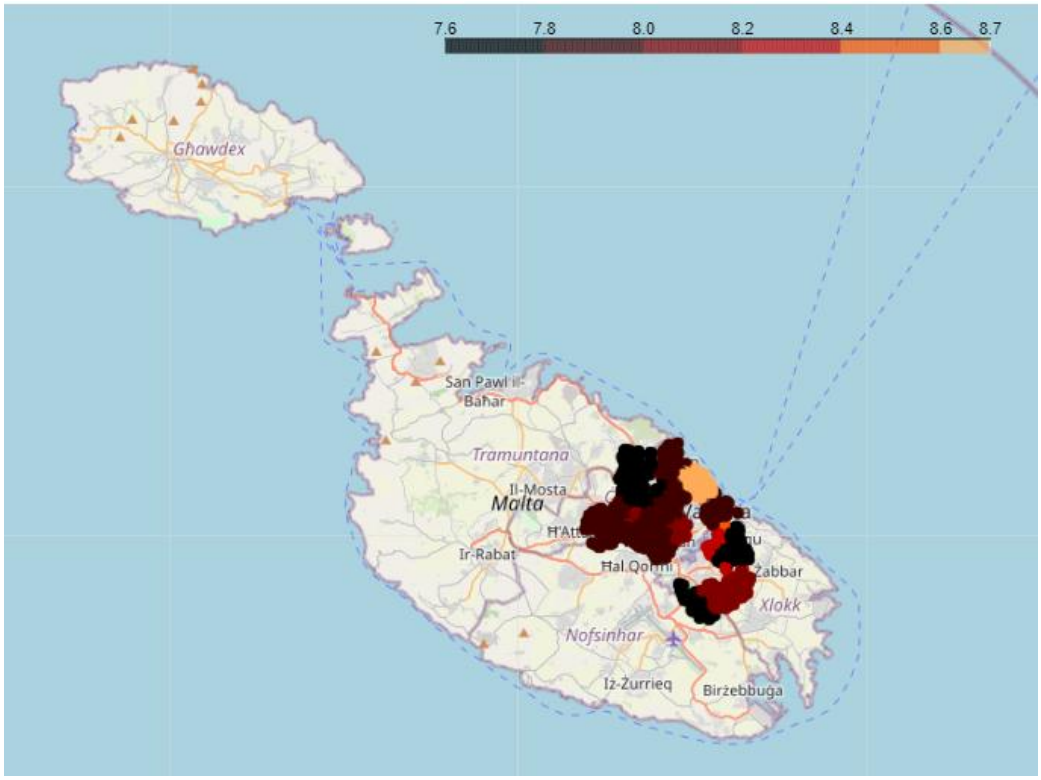
D. Results

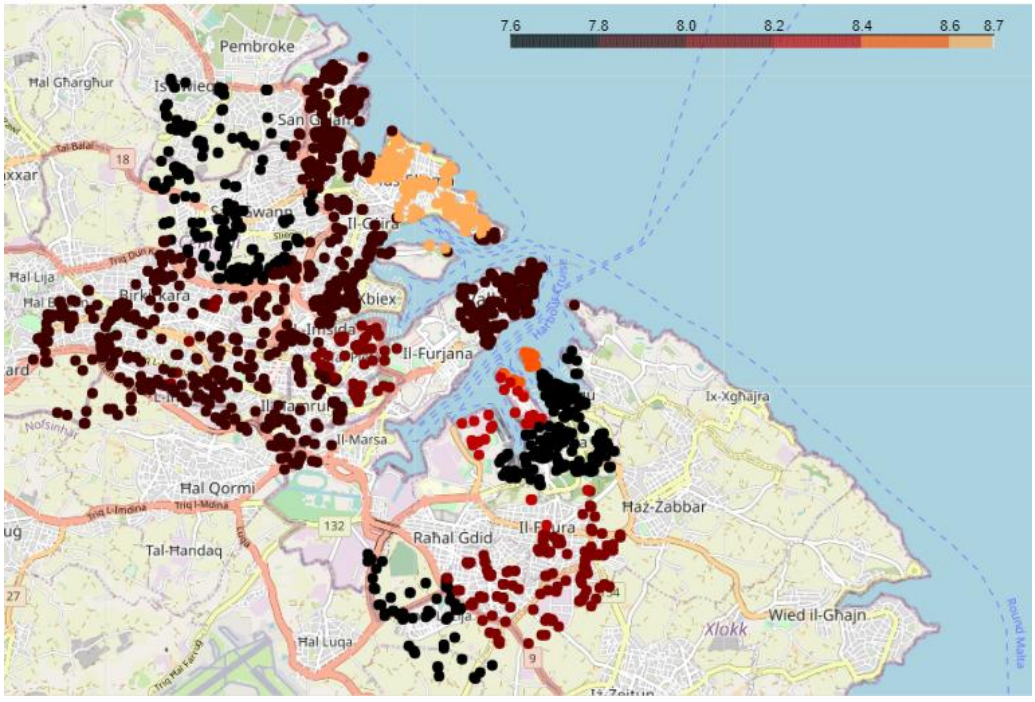
In order to find the optimal locations to open our new restaurant, we can create a **regression model** that takes the historical proximity and population densities and calculated a predicted rating. We can pass **10,000 random coordinates** into the model and select the longitude and latitudes of the ones that return the highest predicted rating.



After creating the model, we can see that the predicted rating values are **positively skewed**, which means that the results are primarily concentrated on the left side, which results in the right tail of the distribution being far longer.

After selecting the results that have a predicted value greater than 7.6, we can now plot the coordinates on the map of Malta to see how they are distributed.





Finally, we have our results! The locations with the highest predicted ratings are concentrated in the council Sliema. The next best places to consider would be Birgu and Isla, which are both just south of the Grand Harbour cruise that runs between them and Valletta. Other possible locations that are also worth looking into are Paola, Tarxien, Fgura, Pieta, Valletta and all the other neighbouring councils that surround Sliema.

It is interesting to note that all of our results have come from the third (blue) cluster in the K-Means analysis. None of the highest-ranking coordinates have come from the northern region of the Mainland Malta or from the island of Gozo.

E. Discussion

The aim of this analysis was to find the optimal places that stakeholders can consider when deciding where to open a restaurant in Malta. Our goal was to select the locations that optimize the predicted rating we expect to receive.

We have analysed the frequency of popular venues that are scattered on the island nation and have shortlisted food categories, such as restaurants, cafes, pizza places and bars, as well as identifying historical sites.

We also explored the relationship between the distance from nearby historical sites and the approximate population density, as well as how these two variables influence the actual restaurant ratings by applying regression analysis.

Furthermore, we used K-Means algorithm to understand how the restaurants cluster together and whether there are any patterns between these clusters and the historical sites.

Finally, we created a regression model where the historical proximity and population density were the independent variables and the restaurant ratings was the dependent variable. We then used 10,000 random coordinates to calculate their associated predicted ratings and plotted the highest-ranking results against the map of Malta.

F. Conclusion

Overall, we were able to gain some understanding on the distribution of food venues on this popular holiday destination island. We shortlisted some local councils that, based on our analysis, are the most promising locations to open a new restaurant.

For a better and accurate results going forward, other factors that have not been discussed, such as the cuisine, prices and restaurant service are all worth taking into consideration when understanding how to receive a good rating.

Additionally, there are other variables, such as rent per square meter, staff wages and advertisements that stakeholders would also want to understand as these would all impact the overall profit margin of this business venture!

Many thanks!

Saad Sharif

G. References

[1] Foursquare API: **https://foursquare.com/**

[2] Malta Councils – Wikipedia: **https://en.wikipedia.org/wiki/Local_councils_of_Malta**

[3] European Environment Agency: **https://www.eea.europa.eu/data-and-maps/data/eea-reference-grids-2/gis-files/malta-shapefile/view**