Greek Travel

A REPORT ON THE GREEK ISLANDS' POPULARITY

INTRODUCTION

How can we find groups of islands that are similar in terms of their popularity with the tourists.

The island groups are either:

- **very touristic** and therefore appeal to a younger demographic looking for fun and partying or

- **less touristic** and appeal to an older demographic looking for relaxation.

DATA

A list of the islands that we can use can be found online on the Wikipedia page:

https://simple.wikipedia.org/wiki/List_of_Greek_islands

We use the Foursquare database to identify the venues on these islands. Venues are identified based on the name of the island because:

- The latitude and longitude are not readily available. This obstacle could be overridden as we found those data points on a separate dataset that we could merge

- The islands are of different sizes and in close proximity to one another therefore defining a radius around a central location would yield incorrect results. Often times the coordinates of an island are those of its main town ('Chora') and therefore are located near the port and not in the centre of the islands

METHODOLOGY

Step 1: Get the data from the Wikipedia page and identify a normalization factor

Step 2: Request Foursquare data

Step 3: Overview of the data

Step 4: Cluster the data

Step 1: Get the data from the Wikipedia page and identify a normalization factor

Rank by size	Island name	Area (km²)	NormFactor
1	Crete	8.336	1.000000
2	Euboea	3.655	0.438460
3	Lesbos	1.630	0.195537
4	Rhodes	1.398	0.167706
...	...	...	...
39	Kasos	66	0.007917
40	Alonnisos	64	0.007678

Step 2: Request Foursquare data

| Feature	Description	Aggregation Method
isAccomodation	This feature is a measure of how many hotels, bed&breakfast etc are available for accommodation on the island	sum
isFoodEntertainment	This feature is a measure of how many restaurants, coffee shops, bars, etc. are available on the island	sum
Latitude	The mean of the coordinates of the venues on the island. This was included in order to group together islands that are similar but also in close proximity so that one can easily 'hop' from one to the other as part of the same vacation package	mean
Longitude	Same as above	mean

```
In [68]: sum_df_normalised.head()
         sum_df.head()

Out[68]:
```
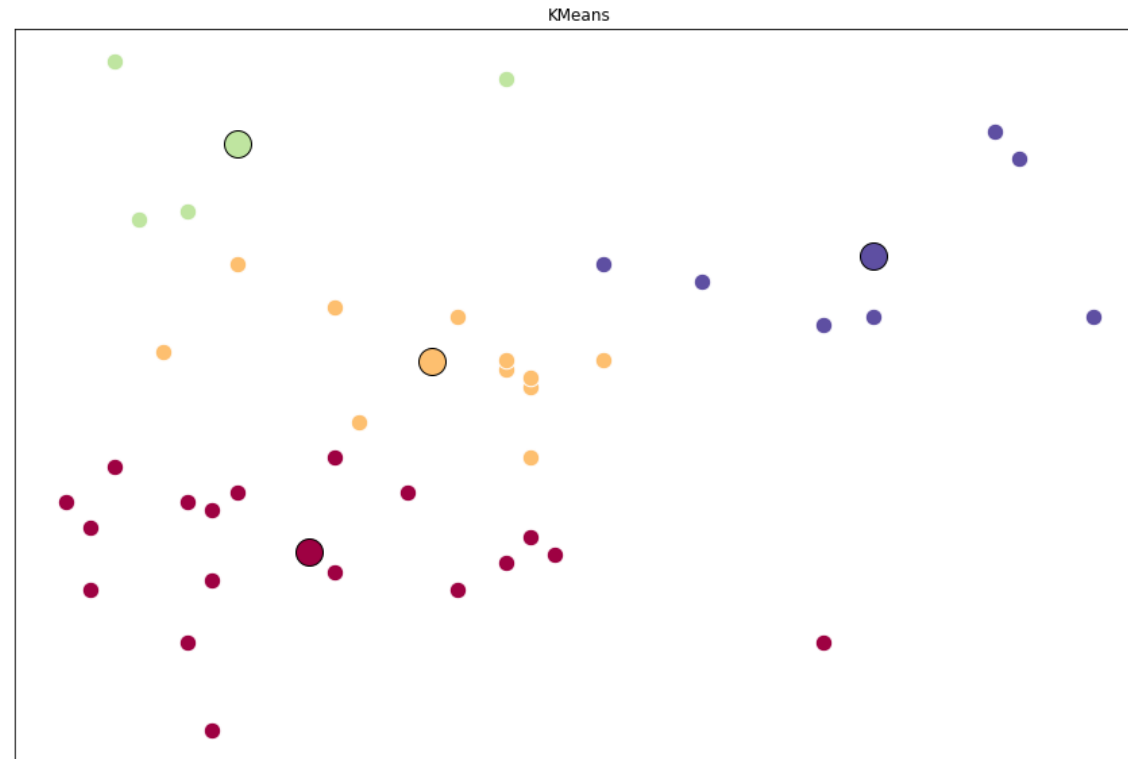
	IslandName	isAccomodation	isFoodEntertainment	Latitude	Longitude
0	Aegina	7	73	37.746538	23.428162
1	Alonnisos	39	85	39.146594	23.859353
2	Amorgos	19	42	36.830276	25.885195
3	Andros	16	67	37.836922	24.933639
4	Astypalaia	33	67	36.547200	26.352486

Step 3: Overview of the data
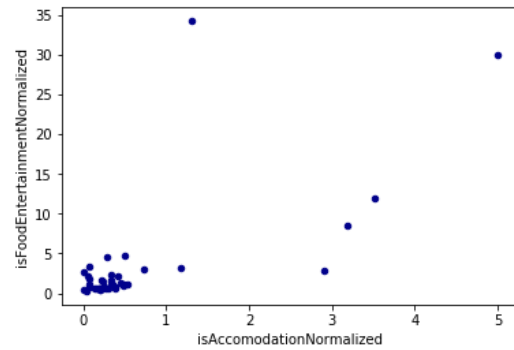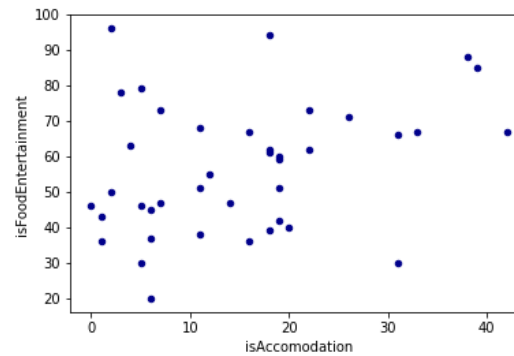
Step 4: Cluster the data

Conclusion

The analysis conducted as part of this exercise has revealed groups of clusters that share common characteristics in term of how popular they are with the tourists and how the infrastructure exists to support those tourists.

Based on the results provided by an unsupervised clustering algorithm we've managed to identify islands that can be grouped together to create holiday packages for customers that come to the travel agency.

Further analysis should be conducted to take into account flights and boat trips between these islands in order to better understand how to structure the packages and provide our customers with the best possible service.