Cyclades Cluster Clustered

A Data Science project about Greek beaches.

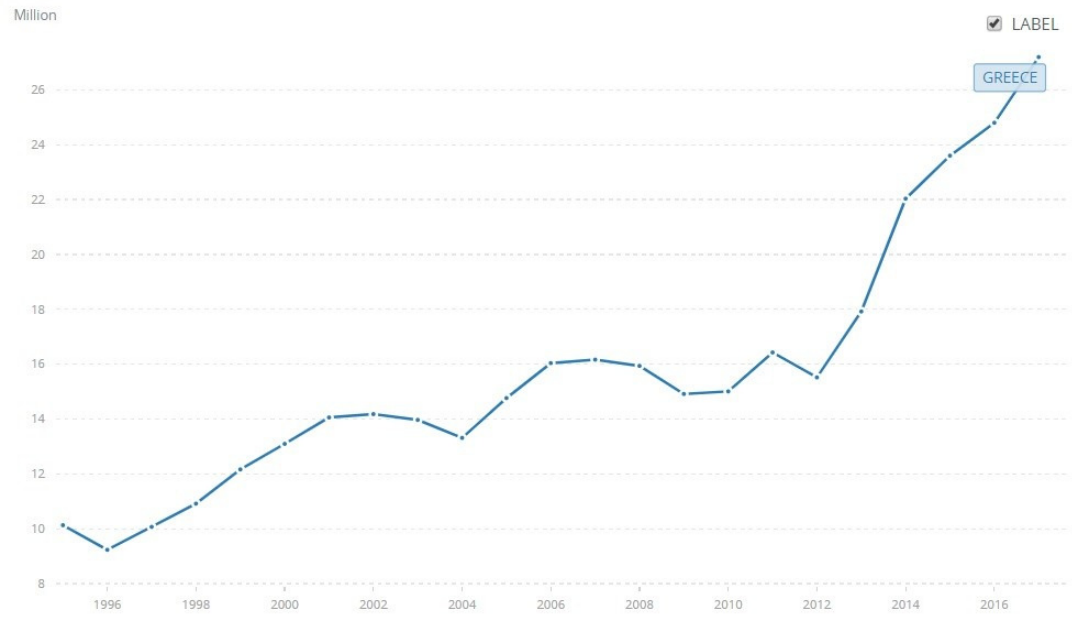
Vakkas Dimitrios

September 18, 2019

1. Introduction

1.1 Background

Greece's tourism industry has boomed over the last 8 years, rising non-stop, breaking many relevant records and becoming very attractive to investors.



Most fruitful period of tourism in Greece is during the summer, mainly because of the warm climate and the picturesque, welcoming beaches with calm waters. Therefore, it comes as no surprise that most visited region of Greece is Southern Aegean.

As we can see in the tables with 2017-2018 data, Southern Aegean received almost 20% of the total tourist visitation while contributing almost 30% of the total incoming tourism revenue for that season.

Region	% Revenue distribution of incoming tourism 2018	Proportion of direct tourism expenditure 2018 - in €m	% Region GDP distribution 2018 - in € m	Direct Contribution of tourism to Region GDP with 2018 data	per capita GDP 2016 - in €
S. Aegean	28,2%	6.080	6.260	97,1%	17.769
Crete	20,0%	4.317	9.139	47,2%	13.811
Attica	14,6%	3.139	87.782	3,6%	22.204
Central Macedonia	14,5%	3.134	25.367	12,4%	12.880
Ionian Islands	10,8%	2.329	3.270	71,2%	15.182
Peloponnese	2,7%	572	8.245	6,9%	13.579
Thessaly	1,7%	373	9.643	3,9%	12.662
Eastern Macedonia & Thrace	2,1%	443	7.223	6,1%	11.432
Epirus	1,4%	306	4.145	7,4%	11.785
Western Greece	1,4%	292	8.406	3,5%	12.058
N. Aegean	1,0%	226	2.570	8,8%	12.266
Central Greece	1,2%	267	8.567	3,1%	14.727
Western Greece	0,4%	83	4.098	2,0%	14.361
Total Country	**100,0%**	**21.562**	**184.714**	**11,7%**	**16.378**

Source: BoG – Processing INSETE Intelligence
** Regions GDP is estimated taking into account the percentage distribution of 2016*

Percentage Distribution of Visits, Overnights and Receipts of Inbound Tourits per Region, 2017

	Visits	Ovenight Stays	Receipts
Southern Aegean	19%	22%	26%
Central Macedonia	23%	19%	13%
Crete	15%	19%	23%
Attica	17%	14%	15%
Ionian Islands	10%	12%	12%
Eastern Macedonia & Thrace	4%	3%	2%
Peloponnese	2%	2%	2%
Thessaly	2%	2%	2%
Epirus	2%	2%	2%
Northern Aegean	1%	2%	1%
Western Greece	2%	1%	1%
Central Greece	1%	1%	1%
Western Macedonia	1%	0%	0%

Source: BoG, processing: SETE Intelligence

Southern Aegean consists of two islands clusters, Cyclades and Dodecanese. In this project I analyzed Cyclades as part of the following hypothesis:

1.2 Problem

A major tourist agency, that aims to fulfill its customer's needs and deliver them the best possible services, so they keep coming back, needs to match customer's preferences with appropriate tourist destinations and attractions. The goal of this project is to achieve that for Cyclades.

As mentioned above, Cyclades' main attractions are the beaches. Whether a tourist wants to socialize, have a drink at a loud beach bar, relax by the calm waves, come closer with nature or seeks adrenaline of water sports, there is the right beach for her. The task is to put beaches with similar characteristics in groups, in order to help tourist agency, direct each customer, near the most fitting one. Using powerful, machine learning, clustering algorithm and Foursquare API, I was able to analyze more than 40 beaches and derive useful insights.

1.3 Interest

Apart from tourist agencies, anyone who wishes to invest in a beach related businesses like beach bars, water sports and restaurants will find this project useful.

2. Data acquisition and cleaning

2.1 Data sources

Based on definition of our problem, factors that will influence our decision are:

- number of top nearby venues for each beach, to determine the most popular and thus meaningful for our analysis.
- popular venue categories near each beach, to help us make groups of similar beaches
- venue categories that beach visitors go right after the beach, to help us better identify each group.

Cyclades consists of 33 islands varying in size and character. We decided to include only the top 12 according to the number of incoming ferry passengers during 2018.

Following data sources will be needed to extract/generate the required information:

- top 18 islands, according to the number of incoming ferry passengers during 2018 using online available data from **ELSTAT** (Greek statistical authority)
- coordinates of each island using **Geopy** Python geocoding client
- coordinates of the top beaches of each island, in terms of numbers of nearby venues using **Foursquare API**
- most popular venue categories for each beach using **Foursquare API**
- venue categories that users go after every particular beach using **Foursquare API**

2.2 Data cleaning

As mentioned earlier, Cyclades consists of 33 Islands. In consultation with our stakeholders, we decided to take in consideration only the top 12 most visited, as they have a meaningful market size, to analyze. In order to determine that, we draw online data from ELSTAT. Specifically, we looked at incoming ferry passengers during 2018 for every island and then chose the top 12.

	Port	Incoming_Passengers
0	Thira	1027773.0
1	Paros	924105.0
2	Mykonos	790150.0
3	Naxos	518884.0
4	Tinos	500300.0
5	Syros	350416.0
6	Antiparos	326429.0
7	Gavrio Andros	275900.0
8	Adamas Milos	187132.0
9	Ios	155365.0
10	Kea	144254.0
11	Sifnos	133064.0

We used Geopy geocoder client to determine the islands coordinates. Then, we utilized Foursquare API to find beaches for each island and their coordinates.

We found 119 beaches. After visualizing them to better understand our d ata, we noticed that, in some cases, Foursquare returned many duplicate results for the same beach. Also many beaches are too close together an d can be considered as one for our analysis. We calculated the distances between every result and dropped the unnecessary.
After removing duplicates, we had 86 beaches.

Since we had selected the beaches, we queried Foursquare API and got the top nearby venues for every beach. We got back 1934 results but beaches themselves were also included in the list, so we removed them.

We then checked how many results we got for each beach and saw that there were many beaches that returned only a few venues. These places, were either too isolate or with little to none tourist infrastructures and thus did not help our stakeholders include them in their tourist packages. Consequently, we removed those too and left with 1487 venues. Also, our final number of beaches is 47.

The same procedure was repeated nearby venues that users go after visiting each beach.

3. Exploratory Data Analysis

By using onehot encoding on venue categories and grouping our data by every beach while calculating the mean, we managed to get the frequency of each separate venue category, occurring on each beach.

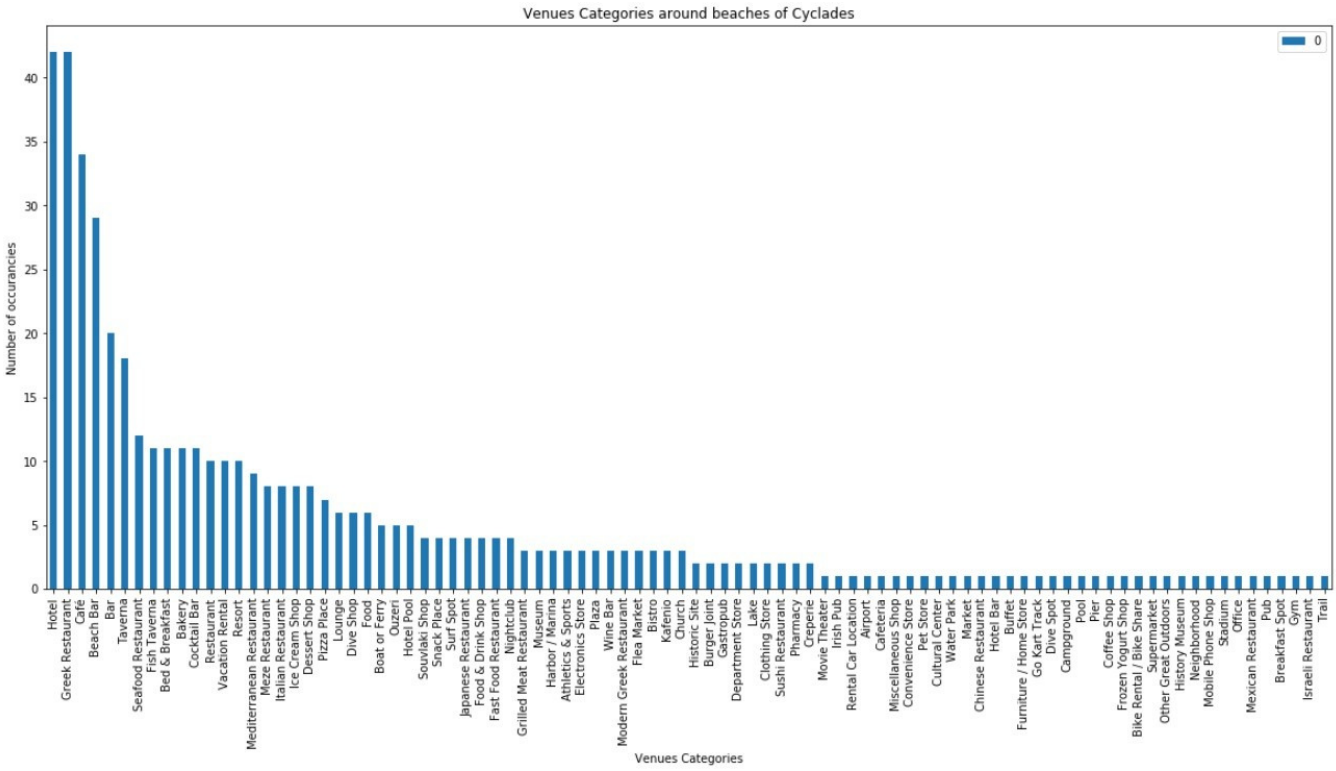
```
----Agios Fokas Beach (Άγιος Φωκάς)----
              venue  freq
0  Greek Restaurant  0.16
1             Café  0.12
2              Bar  0.08
3         Beach Bar  0.08
4            Hotel  0.08
```

```
----Agios Petros Beach (Παραλία Αγίου Πέτρου)----
              venue  freq
0             Hotel  0.13
1  Greek Restaurant  0.11
2         Beach Bar  0.09
3             Café  0.07
4  Seafood Restaurant  0.04
```

```
----Agkathopes Beach (Αγκαθωπές)----
              venue  freq
0  Greek Restaurant  0.20
1             Café  0.15
2          Taverna  0.10
3         Beach Bar  0.05
4   Meze Restaurant  0.05
```

```
----Aneroussa Beach (Παραλία Ανερούσας)----
              venue  freq
0             Hotel  0.14
1             Café  0.12
2  Greek Restaurant  0.12
3           Bakery  0.07
4              Bar  0.07
```

In addition, it was useful to have an overall look of the most frequent venue categories across all beaches with the help of the following bar chart.



We can easily notice that Hotels, Greek restaurants, Cafés and Beach Bars are very popular around these beaches.

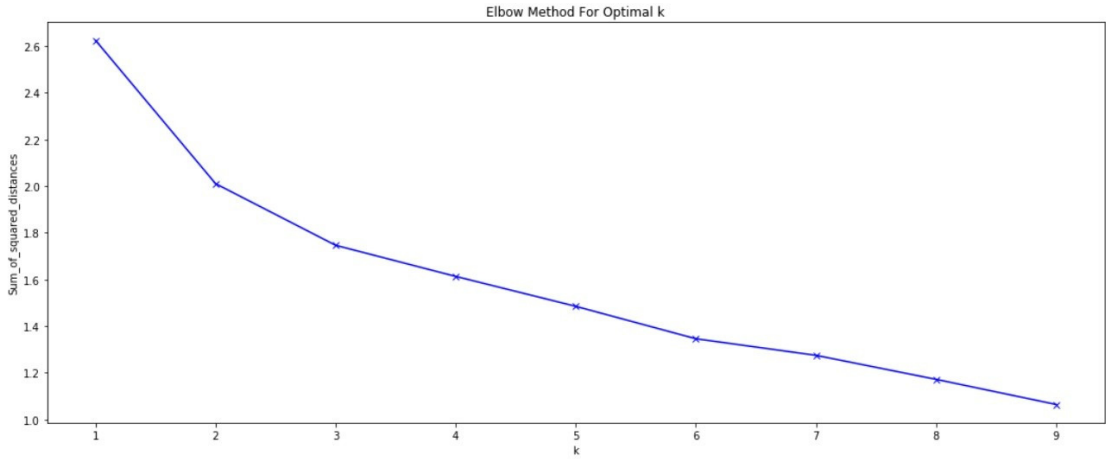
With this information we were able to create a table with top 10 most common venue categories for each beach.

	Beach	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agia Anna Beach (Παραλία Αγίας Άννας)	Greek Restaurant	Hotel	Beach Bar	Café	Taverna	Snack Place	Souvlaki Shop	Ouzeri	Convenience Store	Fish Taverna
1	Agios Fokas Beach (Άγιος Φωκάς)	Greek Restaurant	Café	Bar	Hotel	Beach Bar	Dessert Shop	Coffee Shop	Souvlaki Shop	Cocktail Bar	Pizza Place
2	Agios Petros Beach (Παραλία Αγίου Πέτρου)	Hotel	Greek Restaurant	Beach Bar	Café	Boat or Ferry	Dessert Shop	Souvlaki Shop	Bakery	Seafood Restaurant	Pizza Place
3	Agkathopes Beach (Αγκαθωπές)	Greek Restaurant	Café	Taverna	Kafenio	Restaurant	Snack Place	Meze Restaurant	Bistro	Beach Bar	Sushi Restaurant
4	Aneroussa Beach (Παραλία Ανερούσας)	Hotel	Greek Restaurant	Café	Bakery	Ice Cream Shop	Bar	Bed & Breakfast	Pizza Place	Scenic Lookout	Breakfast Spot

4. K-means Clustering

In order to give our stakeholders insights about Cyclades' beaches, we focused in categorizing those using **k-means** clustering algorithm. Venue categories were used as input to the algorithm. In particular, most popular venue categories for each beach were used to group beaches to clusters.

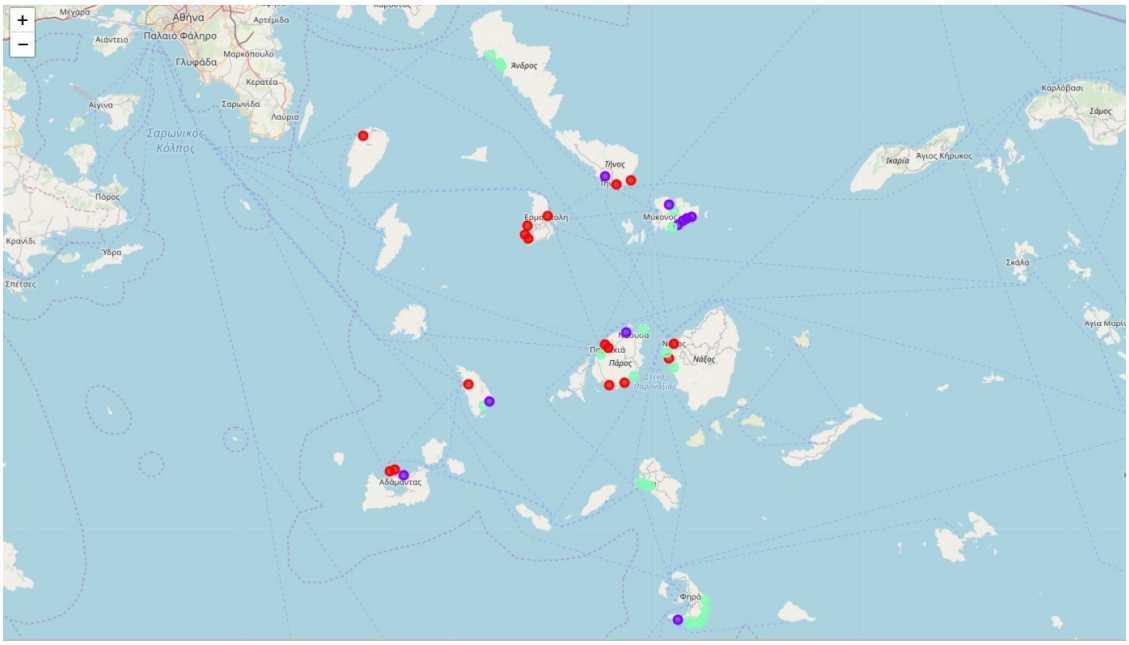
K-means is very efficient with this kind of data but it requires knowing the number of clusters upfront. To find the appropriate number of clusters, I used the **elbow method**. As we can see from the graph, the elbow can be found at k=3 clusters.



After running k-means with 3 clusters the following merged table, containing cluster labels, was derived:

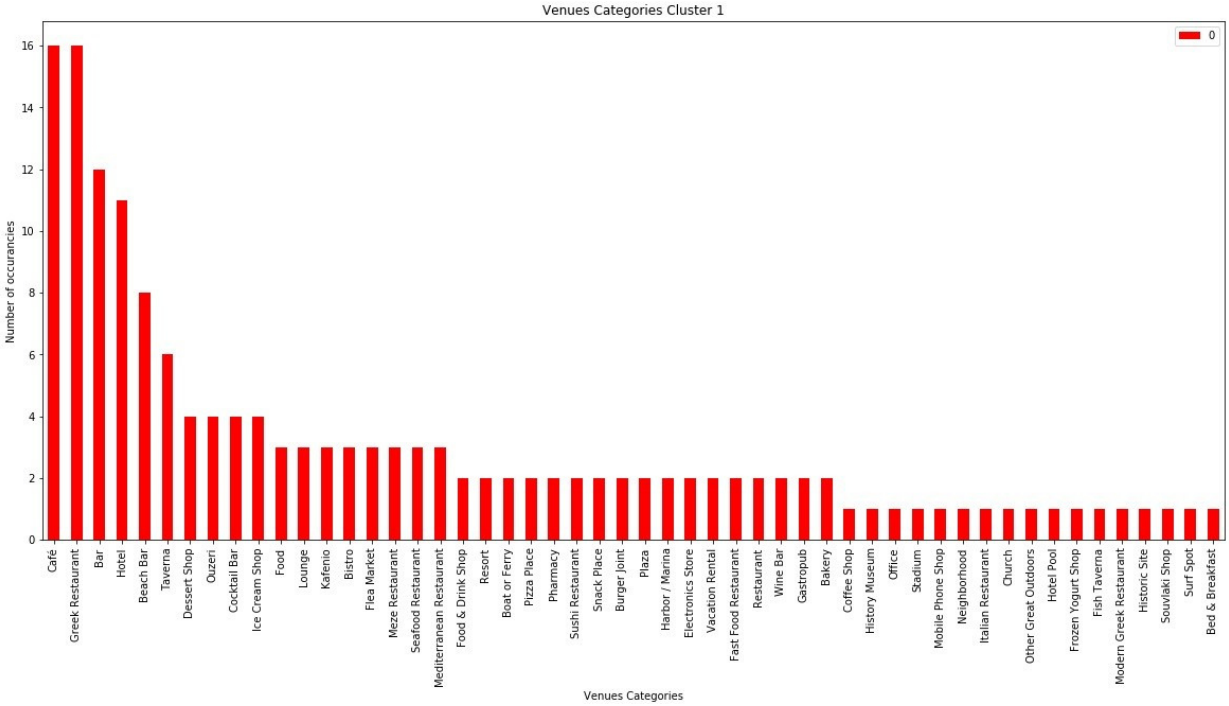
	Island	Island Latitude	Island Longitude	Beach	Beach Latitude	Beach Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Thira	36.407111	25.456664	Kamari Beach (Παραλία Καμαρίου)	36.376584	25.485964	2	Hotel	Greek Restaurant	Café	Taverna	Beach Bar	Mediterranean Restaurant	Grilled Meat Restaurant	Restaurant	Movie Theater	Irish Pub
1	Thira	36.407111	25.456664	Monolithos Beach (Παραλία Μονόλιθου)	36.404254	25.482393	2	Hotel	Rental Car Location	Greek Restaurant	Fish Taverna	Café	Airport	Cafeteria	Miscellaneous Shop	Convenience Store	Pet Store
2	Thira	36.407111	25.456664	Perissa Beach (Παραλία Περίσσας)	36.355556	25.475620	2	Beach Bar	Greek Restaurant	Hotel	Café	Meze Restaurant	Taverna	Bed & Breakfast	Souvlaki Shop	Grilled Meat Restaurant	Italian Restaurant
3	Thira	36.407111	25.456664	Perivolos Beach (Περίβολος)	36.342231	25.459853	2	Beach Bar	Hotel	Greek Restaurant	Bed & Breakfast	Restaurant	Meze Restaurant	Fish Taverna	Café	Taverna	Italian Restaurant
4	Thira	36.407111	25.456664	Vlichada Beach (Παραλία	36.338505	25.432236	2	Hotel	Fish Taverna	Beach Bar	Bed & Breakfast	Boat or Ferry	Greek Restaurant	Taverna	Lounge	Museum	Café

The clusters can be visualized on the map, to give us an idea of how they are distributed.

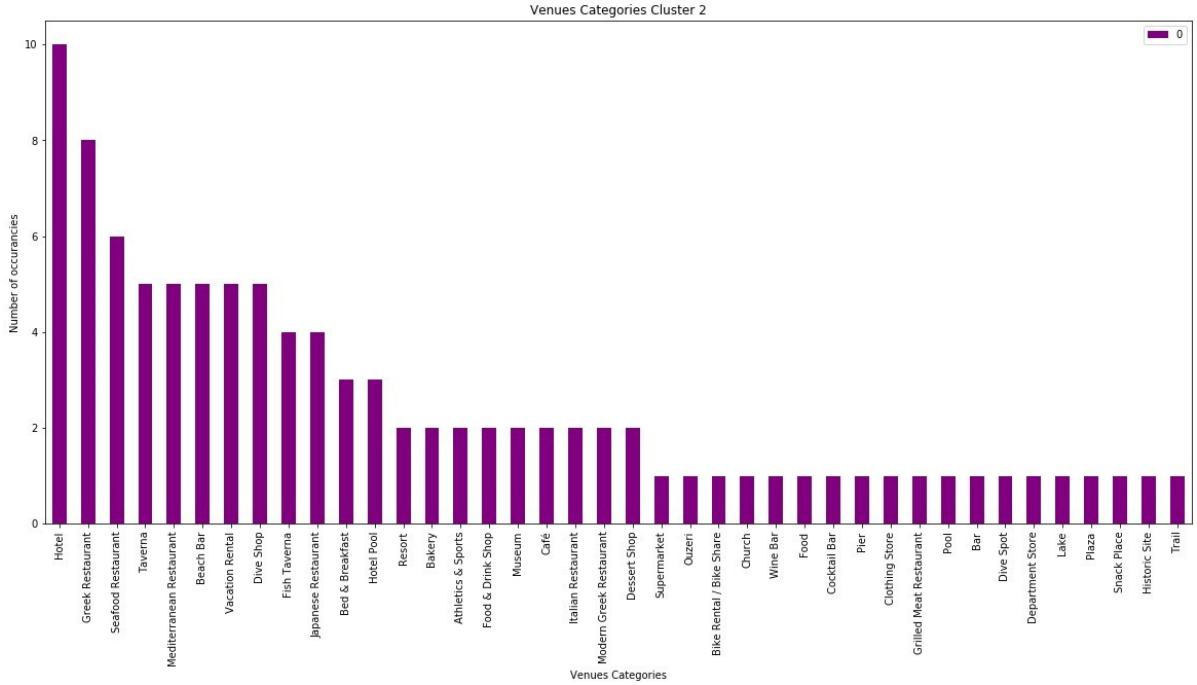


We can also see the most common venue categories for each cluster.

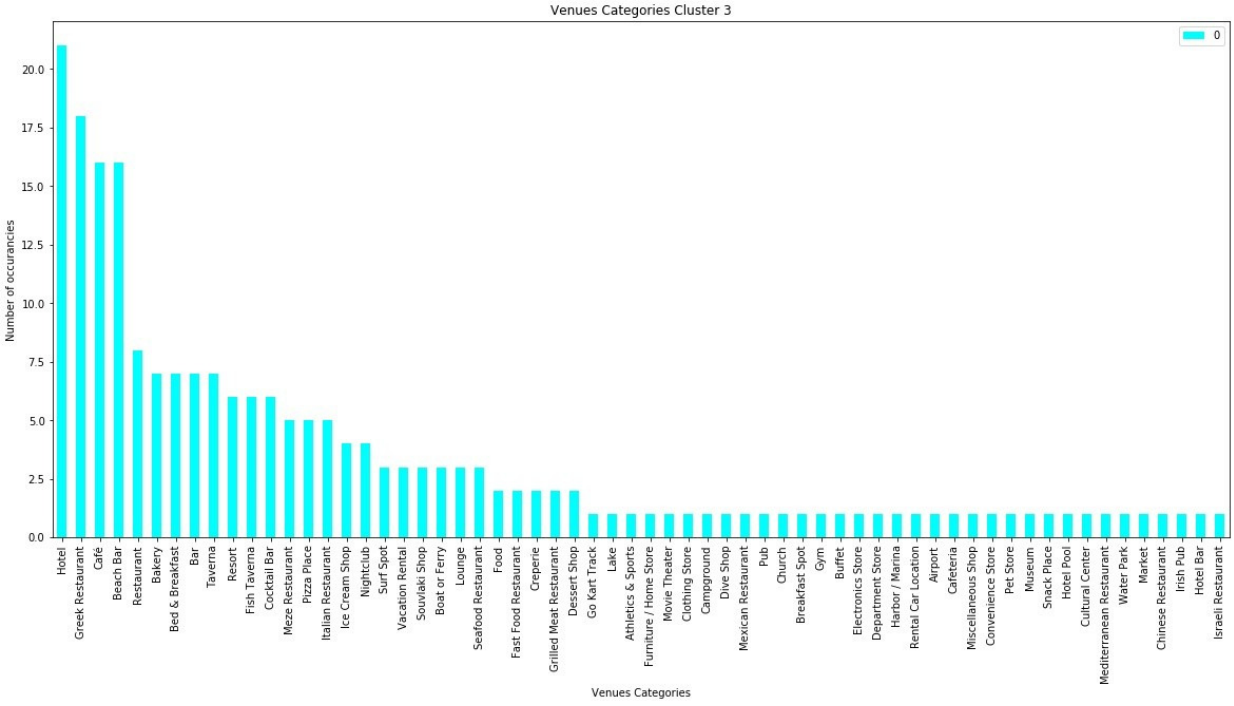
Cluster 1 - red



Cluster 2 – purple



Cluster 3 – Cyan



5. Conclusions

By studying all the available information on each cluster, common patterns and characteristics can be derived:

Cluster 1:

- In the first cluster, amongst the most popular venues, we notice a lot of Greek related food places, like Greek restaurants, Ouzeri, Kafenio, Meze restaurants and tavernas, as well as Cafés. Greeks are known avid coffee drinkers. A research about this cluster's beaches confirms that are preferred by Greeks. If we try to build the profile of this cluster, we could state: **Traditional - Localish - Family - Populous**
- The customer of this cluster is **visiting with his family, relates to Modern Greek way of life and wants to meet the most popular places of a destination**.

Cluster 2:

- In the second cluster we can point out hotels as the most common venue in every single beach, plus there are a lot of hotel pools. Visitors of this cluster seem to prefer to spend time in their hotels and most of them, go for food to a nearby restaurant, after enjoying the beach. Moreover, there are a lot of Dive shops and spots as well as fish tavernas and/or seafood restaurants. Also we notice that, many of the beaches are close to scenic venues like lake, river, trail and historic site. We could characterize this cluster as: **Peaceful - Relaxing - Scenic**
- All the previous observations, lead us to the conclusion that this cluster would appeal mostly to **middle-aged+ customer, with good income, that appreciates nature of the calm waters of Cyclades and seek comfort.**

Cluster 3:

- In the final cluster, bars, nightclubs, beach bars, pubs and cocktail bars stand out amongst the most popular venues as well as venues that are visited after the beach. The food related venues seem inclined to more fast and cheap choices like bakeries, Breakfast spot, pizza place, creperie, sandwich place, snack place and souvlaki spots. In addition, accommodation venues seem inclined to cheaper choices too, as, except hotels, we notice Bed & Breakfast, Vacation Rentals and Campground. Lastly, a few surf spots, Athletics & Sports and a go kart track have made it to the top 10 most common venues. This cluster can be tagged as: **Romance - Active - Party**

- The visitor of this cluster is **young, desires to have fun and socialize, loves nightlife and parties by the sea.**

Clusters	islands #	Tags	Customer Profile
1 - red	16	Traditional - Localish - Family - Populous	visiting with his family, relates to modern Greek way of life and wants to meet the most popular places of a destination
2 - purple	10	Peaceful - Relaxing - Scenic	middleaged+ customers, with good income, that appreciate nature of the calm waters of Cyclades and seek comfort
3 - cyan	21	Romance - Active - Party	young, desires to have fun and socialize, loves nightlife and parties by the sea

Finally, very interesting observations can be made about each island and geographic distribution of clusters.

- **Thira** (or Santorini) and **Mykonos** appear to have both chill and wild aspects.
- **Andros**, **Naxos** and **Ios** have a Greek traditional taste but know how to party.
- **Paros** and **Sifnos** are versatile and can support all kind of tastes.
- **Kea**, **Milos**, **Syros** and **Tinos** tend to have a more quiet and family friendly character.

** Please note that all **Antiparos**' beaches were ruled out during our analysis process.*

Islands	Clusters	Comments
Thira (or Santorini), Mykonos	2, 3	both chill and wild aspects
Andros, Naxos, Ios	1, 3	Greek traditional taste, know how to party
Paros, Sifnos	1, 2, 3	versatile
Kea, Milos, Syros, Tinos	1, 2	quiet, family friendly character

To summarize, this project covered a wide portion of Southern Aegean beaches and managed to extract useful insights that can be utilized by stakeholders, which in this hypothesis was a travel agency company. we grouped beaches locations to clusters, based on nearby venues frequency. With this information, stakeholders can target customers more efficiently by matching their expectations with appropriate destinations. Moreover, project's results can be used as groundwork for anyone who wishes to invest in a beach related businesses like beach bars, water sports and restaurants.