

THE XINOMAVRO* PROJECT

WINE BAR ANALYSIS IN GREECE

ILIAS KAPETANAKIS 2021

*Xinomavro:" (Greek: Ξινόμαυρο [ksi'nomavro], lit. 'sour black') is the principal red wine grape of the uplands of Naousa in the regional unit of Imathia, and around Amyntaio, in Macedonia, Greece.".

Introduction

- ▶ Wine Bars current picture in the 60 biggest Greek Cities
- ▶ Where and why it's better to invest money and open a Wine Bar.
- ▶ Data involved:
 - ▶ the population of the cities to have an indication of the size of each one of them,
 - ▶ the number of the Wine Bars,
 - ▶ the number of the Bistros as they are both wine related places
 - ▶ metrics that characterize each Wine Bar like Foursquare Rating, Price Range and Number of Likes.



Data: Greek Cities Data

- ▶ **Data Source:** [GREECE City & Town Population Geography Population Map cities coordinates location – Tageo.com](#)
- ▶ **Extraction Method**
 - ▶ Scrap HTML to get the table containing the Greek Cities, population and coordinates
- ▶ **Data Description**
 - ▶ The data Source contains the 60 biggest Greek cities in terms of population, the population of each one of them and the Latitude/Longitude.
- ▶ **Data Quality, Cleansing and Transformation**
 - ▶ Remove unnecessary columns like Rank
 - ▶ Fix Header of the Dataset (Header is appearing as 1st row)
 - ▶ Convert numeric fields – Population
 - ▶ Handle Duplicate City Names: Iraklion Crete duplicate with Iraklion Attica



Data: Greek Wine Bars & Bistro

- ▶ **Data Source:** [FOURSQUARE API](#) : search Endpoint (GET
<https://api.foursquare.com/v2/venues/search>)
- ▶ **Extraction Method**
 - ▶ Call the API for each Greek city and the specific category id of the “Wine bars” and “Bistro”
- ▶ **Data Description**
 - ▶ The data set contains all the venues belonging to the WineBar category for each city in a radius of 1500m from the City Coordinates. The information retrieved includes the Venue ID, Venue Name, Venue Latitude, Venue Longitude, Venue Distance from the City Coordinates and Venue Category. The same was used for the category ‘Bistro’ to be retrieved
- ▶ **Data Quality, Cleansing and Transformation**
 - ▶ The same venue can be returned for more than 1 cities. In cases like this the minimum distance will be used to keep each venue only one time and assign it to the city that is closer to the venue.
 - ▶ Venues belonging both to Wine Bars and Bistro categories were kept only as Wine Bars

Data: Greek Wine Bars Details

- ▶ **Data Source:** [FOURSQUARE](#) API : details Endpoint (GET
https://api.foursquare.com/v2/venues/VENUE_ID)
- ▶ **Extraction Method:**
 - ▶ Call the API for each winbar venue found from the search API call
- ▶ **Data Description**
 - ▶ The data source contains details about each Wine bar venue:
 - ▶ Rating (score from 1 to 10 with 1 decimal point)
 - ▶ Price Range (from 1 (least pricey) – 4 (most pricey))
 - ▶ Number of Likes
- ▶ **Data Quality, Cleansing and Transformation**
 - ▶ In many venues price and rating info were missing from the data
 - ▶ It was decided to replace the missing values as follows:
 - ▶ Missing Price = Average City Price Value rounded in the closer integer
 - ▶ Missing Rating = Average City Rating Value rounded in 1 decimal
 - ▶ Missing Likes = The Median of total Likes

Data Samples

▶ Greek Cities Data

	City	Population	Latitude	Longitude
1	Athinai	762100	37.980	23.730
2	Thessaloniki	372100	40.640	22.940
3	Piraeus	179600	37.960	23.640
4	Patrai	164000	38.240	21.730
5	Peristerion	141000	38.020	23.700

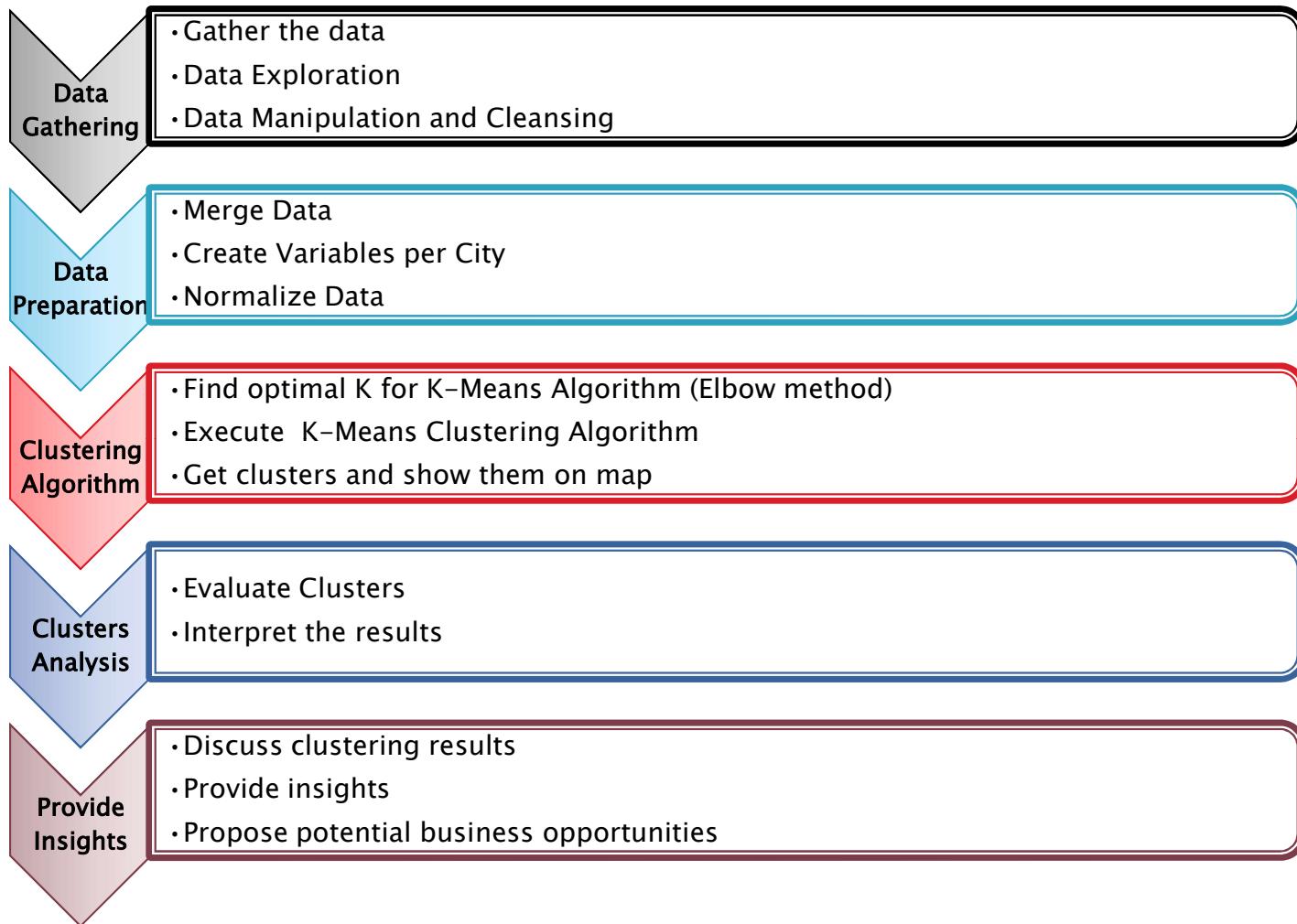
▶ Greek Wine Bars & Bistro

	City	Population	City_Latitude	City_Longitude	Venue_Id	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Distance	Venue_Category
0	Athinai	762100	37.980	23.730	50c37cd7e4b04a2d9cd2a324	Harvest	37.979581	23.728421	146	Café
1	Athinai	762100	37.980	23.730	53880882498e450b0a1fbd04	Delight	37.979810	23.732515	221	Café
2	Athinai	762100	37.980	23.730	5b0af271f193c0002c3c53d2	Kalimeres	37.978510	23.723880	562	Wine Bar
3	Athinai	762100	37.980	23.730	56a2a79b498ef61d4b2b036a	Wine O'Clock	37.967770	23.729855	1361	Wine Bar
4	Athinai	762100	37.980	23.730	56733fe7498e0010a67a750b	Acropaul's	37.966752	23.728304	1482	Wine Bar

▶ Greek Wine Bars Details

	Venue_ID	Venue_Name	Price	Likes	Rating
0	4adcdadef964a520f15721e3	Peacock Roof Garden Restaurant	2	4	6.7
1	4adcdadff964a520215821e3	CAFÉ & BISTROT VIENNA	1	15	7.5
2	4b5b2f7ff964a52087e928e3	Gala 1985	2	59	7.7
3	4b684c8bf964a52057702be3	Chocolat Royal	2	355	6.9
4	4b8589b7f964a5201f6431e3	Scala Vinoteca	3	229	9.2
5	4ba296a8f964a520f00638e3	Franco's	1	30	6.7
6	4bb25ff3eb3e9521a8f0c90a	Thema Coffees and Drinks	2	281	7.9

Methodology

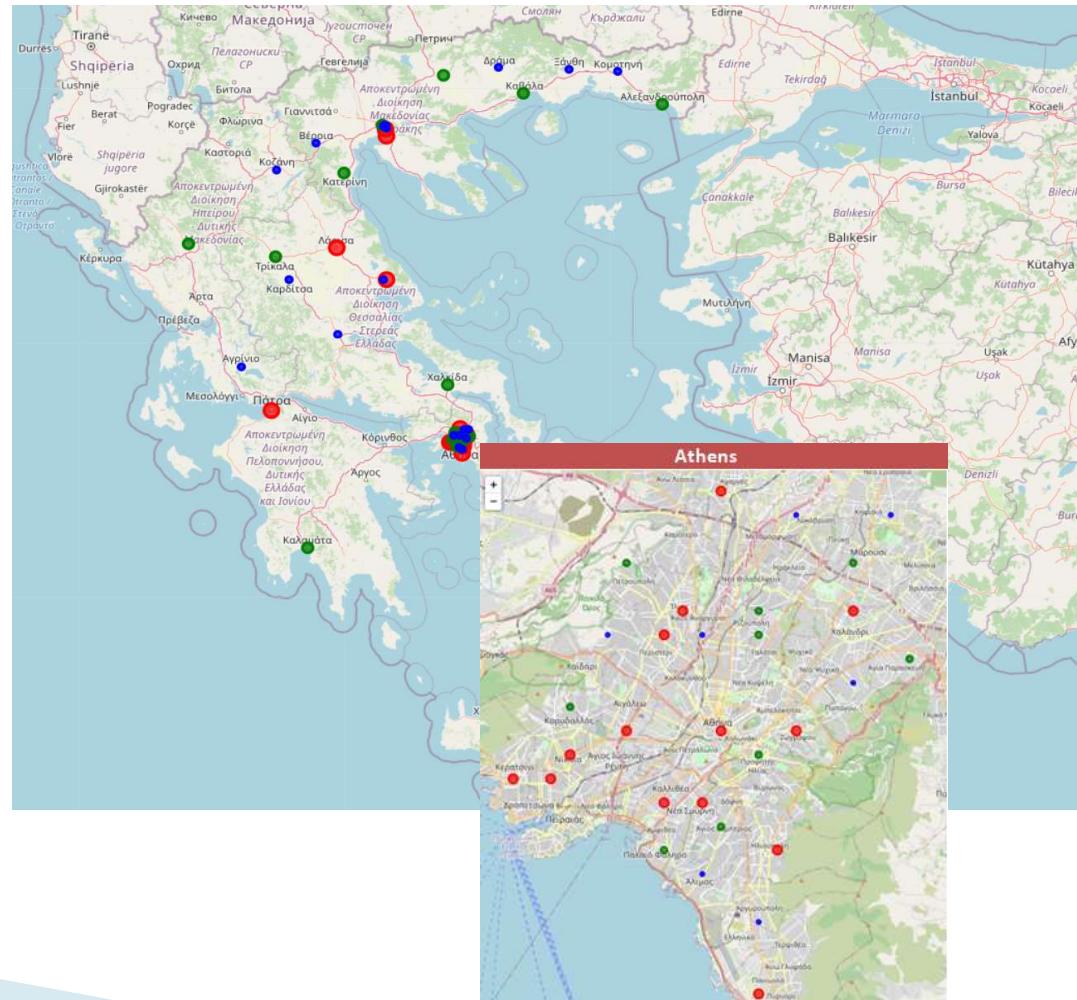


Data Gathering and Exploration: Cities Population and Location

Color	Population Range
Blue	Greater than 73300
Green	Between 49400 and 71000
Red	Lower than 47400

Special care is needed
for Metropolitan areas:

- ▶ Cities are one next to another
- ▶ Attention to cases where same venue is close to more than 1 city.

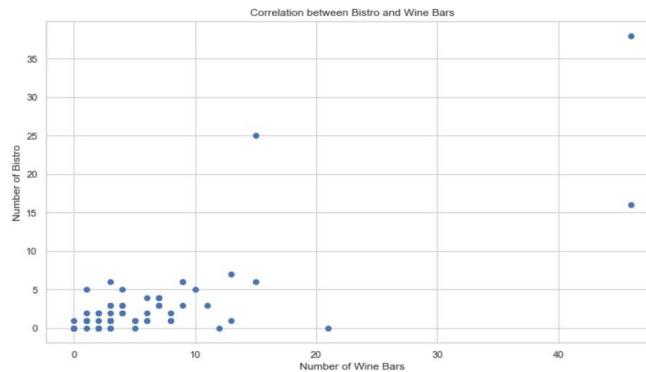


Data Gathering and Exploration: Wine Bars and Bistro Venues

- ▶ Clean duplicate venues by assign it to the minimum distance city i.e

City	Population	City_Latitude	City_Longitude	Venue_Id	Venue_Name	Venue_Latitude	Venue_Longitude	Venue_Distance	Venue_Category
Athinai	762100	37.980	23.730	4d41d00789c3a143abd2f183	Oinoscent	37.974037	23.732125	689	Wine Bar
Viron	62500	37.970	23.750	4d41d00789c3a143abd2f183	Oinoscent	37.974037	23.732125	1631	Wine Bar

- ▶ Correlation between # of Wine Bars and Bistro is strong.



- ▶ Check the outliers

City	Population	Latitude	Longitude	wine	bistro
Athinai	762100	37.980	23.730	46	38
Thessaloniki	372100	40.640	22.940	46	16
Viron	62500	37.970	23.750	15	25

Data Gathering and Exploration: Wine Bars Details–Price, Likes, Rating I

▶ Missing Data

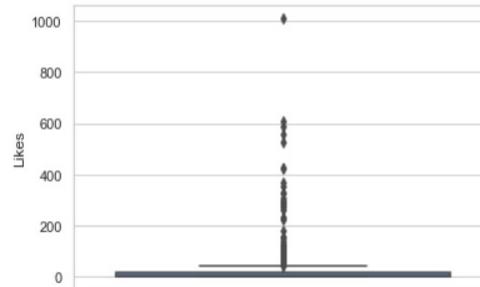
missing_price	missing_rating	missing_likes
14	231	109

▶ Handle Missing Data

Find the Average price and rating per city and then replace the missing values with the average of each city

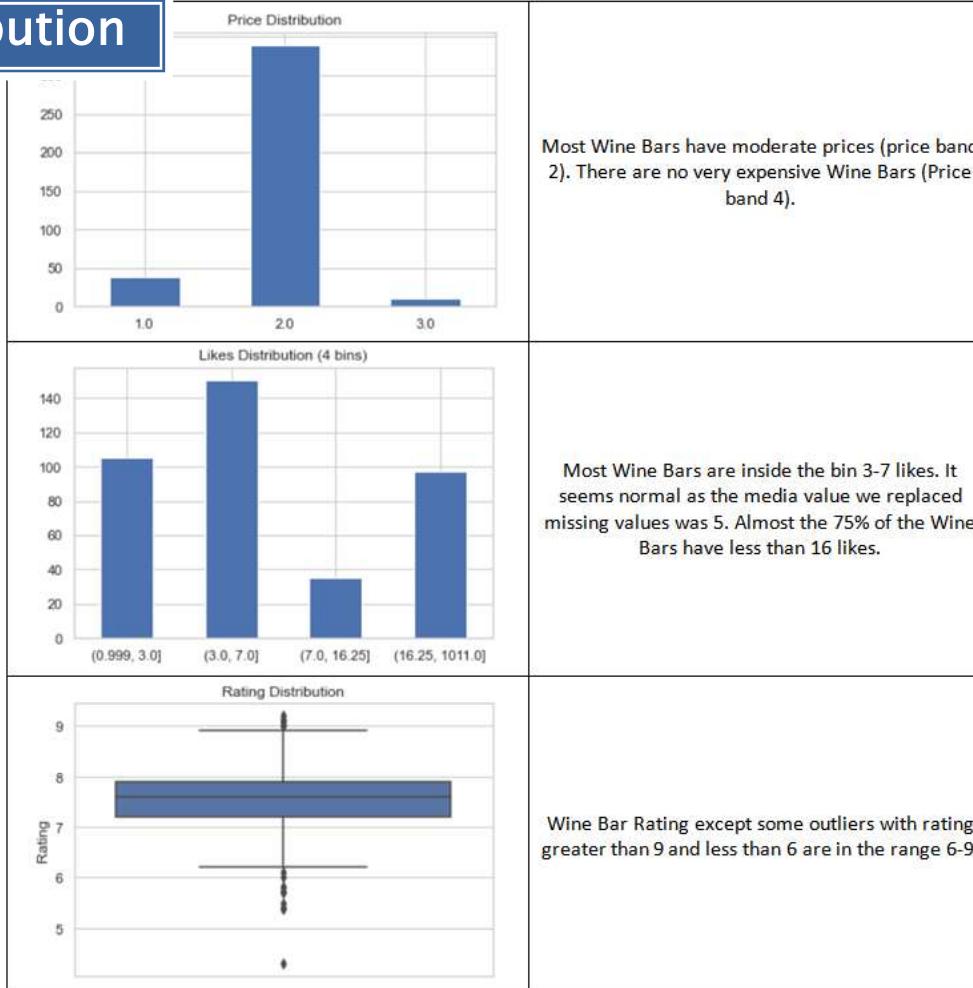
If following step 1 there are still missing values, because there are no data to calculate average for a city, then we replace the missing value with the total average price or rating.

For Likes we will replace the 0 values with the median of all likes. As shown in the boxplot below there are some extreme values and all the others are in the same small range it is much preferable to use the median instead of average.



Data Gathering and Exploration: Wine Bars Details–Price, Likes, Rating II

Variables Distribution

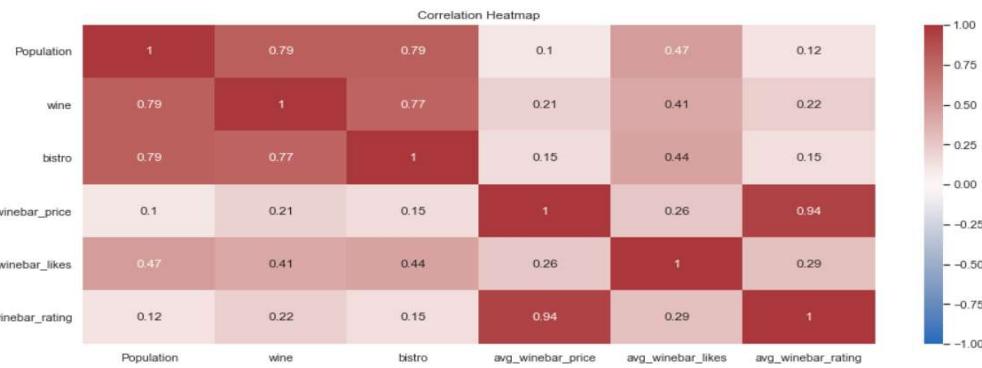


Data Preparation: Merge Data Sets

► Merged Data Set on City Level

	City	Population	Latitude	Longitude	wine	bistro	avg_winebar_price	avg_winebar_likes	avg_winebar_rating
0	Athinai	762100	37.980	23.730	46	38	2.000000	132.413043	7.695652
1	Thessaloniki	372100	40.640	22.940	46	16	1.869565	41.500000	7.602174
2	Piraeus	179600	37.960	23.640	4	2	2.000000	26.500000	9.100000
3	Patrai	164000	38.240	21.730	13	7	2.076923	26.384615	7.500000
4	Peristerion	141000	38.020	23.700	2	2	2.000000	5.500000	7.800000

► Variables Correlation Matrix



- Correlation between the Number of Wine bars and the Number of Bistro
- Correlation between Wine Bar Average Price and Average Rating.
 - In reality the rating of a place isn't straight forward connected with the prices of the place. Explained because most Wine bars (almost 80%) has the same value (2) in Price.

Data Preparation: Final Variables

- ▶ Transform number of wine bars and bistro to a new KPI called Number of Wine Bars per person.

Variable Name	Definition	Business Meaning
population	Population of each city.	It is an indication of the size of each city.
wine_var	Number of Wine Bars per Person.	How a city competes in terms of the number of wine bars compared to other cities.
bistro_var	Number of Bistro per Person.	How a similar business to the wine bars, is doing in the same city.
avg_winebar_price	Average Wine Bar Price Band.	How expensive the Wine Bars of a city are.
avg_winbar_likes	Average Wine Bar Likes.	How much the customers like the Wine Bars of a city on average.
avg_winbar_rating	Average Wine Bar Rating.	The quality of the Wine Bars of a city on average.

- ▶ Data sample

City	population	wine_var	bistro_var	avg_winebar_price	avg_winebar_likes	avg_winebar_rating
Athinai	762100	0.000060	0.000050	2.000000	132.413043	7.695652
Thessaloniki	372100	0.000124	0.000043	1.869565	41.500000	7.602174
Piraeus	179600	0.000022	0.000011	2.000000	26.500000	9.100000
Patrai	164000	0.000079	0.000043	2.076923	26.384615	7.500000
Peristerion	141000	0.000014	0.000014	2.000000	5.500000	7.800000



Data Preparation: Data Normalization

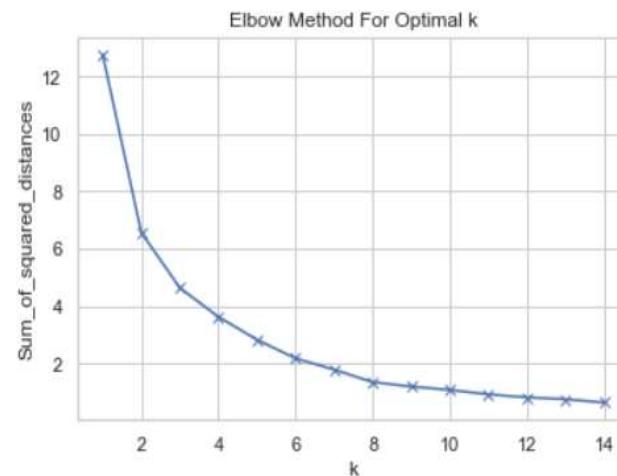
- ▶ Bring the data in a form where a certain variable not to influence the result more than the others
- ▶ By making the ranges consistent between variables, normalization enables a fair comparison between the different features, making sure they have the same impact.
- ▶ Data after applying a MinMax Scaler to normalize the values:

	0	1	2	3	4	5
0	1.000000	0.103473	0.124656	0.800000	0.903843	0.845676
1	0.466192	0.211925	0.107498	0.747826	0.283276	0.835404
2	0.202710	0.038180	0.027840	0.800000	0.180887	1.000000
3	0.181358	0.135889	0.106707	0.830769	0.180100	0.824176
4	0.149877	0.024316	0.035461	0.800000	0.037543	0.857143



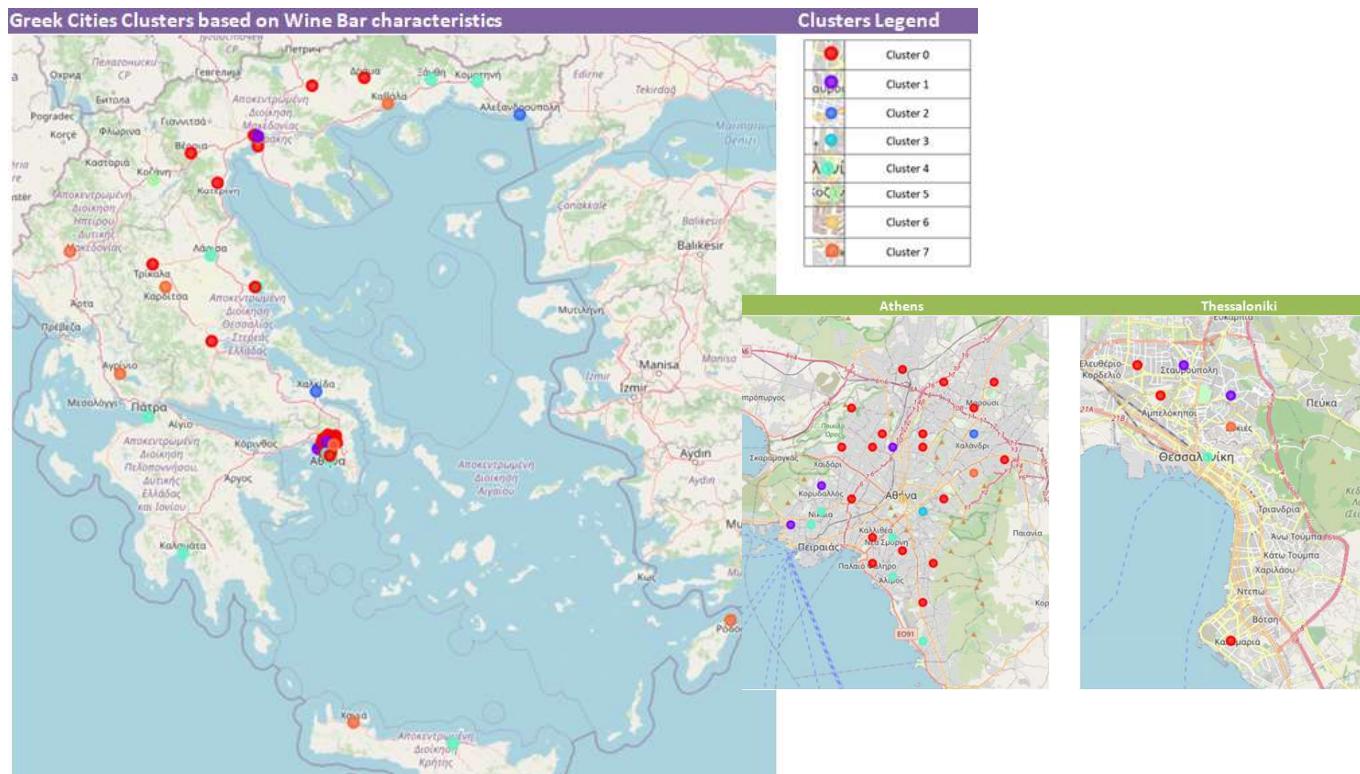
Clustering Algorithm: Find Optimal K for K-Means

- ▶ How many clusters to create or what is the optimal K parameter?
- ▶ The elbow method for our case showing that the optimal K is 8.



Clustering Algorithm: Execute K-Means Clustering

- ▶ K-Means using parameter k=8
- ▶ Each city is assigned to one of the clusters created based on the characteristics of the wine bars of the city.
- ▶ Helpful to identify common patterns and extract useful insights regarding the wine bar business in different cities.



Clusters Evaluation: Clusters Raw Data

- ▶ Aggregate each cluster's cities metrics to cluster level

Cluster	cities	avg_population	avg_winebar	avg_bistro	avg(avg_winebar_price)	avg(avg_winebar_likes)	avg(avg_winebar_rating)
0	28	62921.428571	5.446856	2.548391	1.939838	8.307653	7.453784
1	5	52020.000000	0.000000	0.289855	0.000000	0.000000	0.000000
2	3	59366.666667	8.399810	5.093948	2.023810	112.103175	8.127778
3	1	62500.000000	24.000000	40.000000	1.866667	25.200000	7.213333
4	13	114892.307692	8.951285	2.855173	1.941875	40.294551	7.795584
5	1	36000.000000	58.333333	0.000000	1.904762	5.523810	6.804762
6	1	762100.000000	6.035953	4.986222	2.000000	132.413043	7.695652
7	8	47837.500000	17.596208	8.980689	1.927778	11.463095	7.222153

- ▶ Compare the metrics of each cluster with the total averages

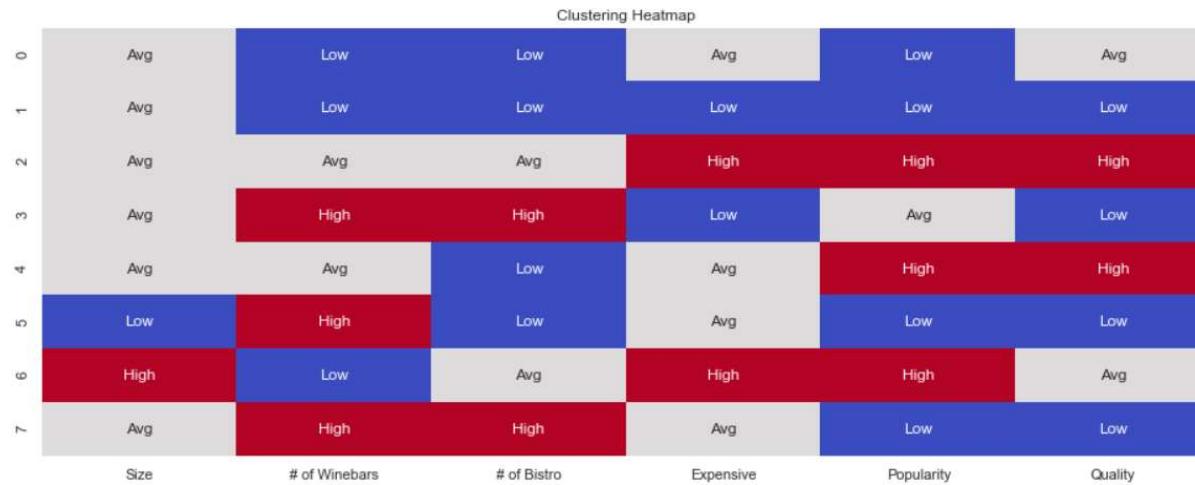
total_avg_population	total_avg_winebar	total_avg_bistro	total_avg_price	total_avg_likes	total_avg_rating
85032.727273	9.513037	4.374287	1.942271	24.50172	7.52587

- ▶ Number of venues per person converted to Number of venues per 100k persons in order to help us with the interpretation and comparison of the results.



Clusters Evaluation: Advanced Clusters Evaluation

- ▶ Difficult to interpret raw results
- ▶ Visualize the evaluation of the clusters as a heatmap.
- ▶ Three main values to use on Heatmap:
 - Low: blue color, the average variable value of the cluster is below a certain threshold of the total average
 - Avg: grey color, the average variable value of the cluster is around the total average
 - High: red color, the average variable value of the cluster is above a certain threshold of the total average



Variables Business Meaning

Population = Size
Price = Expensive
Likes = Popularity
Rating = Quality

Clustering Results Analysis: Overview

- ▶ Based on the data, metrics and the Heatmap, we can summarize the 8 clusters:

Cluster	Cities	Name	Description
0	28	The Not so Popular Average	Small number of wine bars, average quality and price, low popularity
1	5	Dry Cities	No wine bars
2	3	High Performance	Consists of cities with an average number of wine bars and exceptional quality
3	1	Champion with low quality	Many wine bars and bistro with below average quality
4	13	The Popular Average	Small number of wine bars, average price, high in quality and popularity
5	1	Quantity not Quality	Many wine bars but low quality and likes
6	1	The Capital	The biggest in size, not so many wine bars with average quality, popular places and high prices
7	8	Trying to get Improved	Many wine bars with below the average popularity and quality



Clustering Results Analysis: Clusters to Avoid

Cluster 0	
Name	The Not so Popular Average
Summary	High: - Low: Winebars, Bistro, Popularity Average: Size, Expensive, Quality
Comment	<p>The biggest cluster consists of cities with no strong points. We can assume that this is the average having at the same time less number of wine bars per person than the average, that at the same time are not so popular. To be more precise the average Likes of cluster's wine bars are the second worst from the seven clusters that have wine bars. The characteristics of this cluster don't show an opportunity or something special. It's an average after all.</p>
Cluster 1	
Name	Dry Cities
Summary	No Wine Bars
Comment	<p>This is an expected cluster with cities that don't have any wine bars. This fact can be probably explained form the fact that all 5 cities are part of metropolitan areas, so there maybe another center next to them where restaurant and bars exists and social life is more developed. There is no reason to invest in such an area.</p>
Cluster 6	
Name	The Capital
Summary	High: Size, Expensive, Popularity Low: Winebars Average: Bistro, Quality
Comment	<p>The Greek capital is the only city of cluster 6. This is normal as its characteristics are unique. The major difference here is the size (762000 population) that is double than the second biggest city in Greece (Thessaloniki - 372100). At the same time it is an expensive city with wine bars of average quality. With the results we have in our hands now, it is not so attractive to invest there. In order to check the possibility of opening a new wine bar in Athens a different analysis should take place on each neighborhoods or hot areas of the city.</p>



Clustering Results Analysis: Risky Clusters

Cluster 2	
Name	High Performance
Summary	High: Popularity, Expensive, Quality Low: - Average: Size, Winebars, Bistro
Cluster's Cities	
Comment	These 3 cities are top in terms of Wine bars quality and Popularity. It is a bit risky to invest in such an area where people are happy with the existing wine bars.

Cluster 4	
Name	The Popular Average
Summary	High: Popularity, Quality Low: Bistro Average: Winebars, Expensive, Size
Comment	The second biggest cluster with 13 cities consists of cities with nothing special in terms of number of wine bars but it performs high in terms of quality and popularity. We could say that this is the alter ego of cluster 0 having similar characteristics but the exact opposite behavior regarding popularity and quality. In addition it is similar to "High Performance" cluster 2 both having an average number of wine bars and high popularity and quality. Their difference is that cluster 2 is much better in popularity with an average of 112 likes vs 40 likes and an average rating of 8.13 vs 7.79. The same comment as in cluster 2 is valid here. Too risky to invest to a high performance city with already popular places with good quality.



Clustering Results Analysis: Proposed Clusters

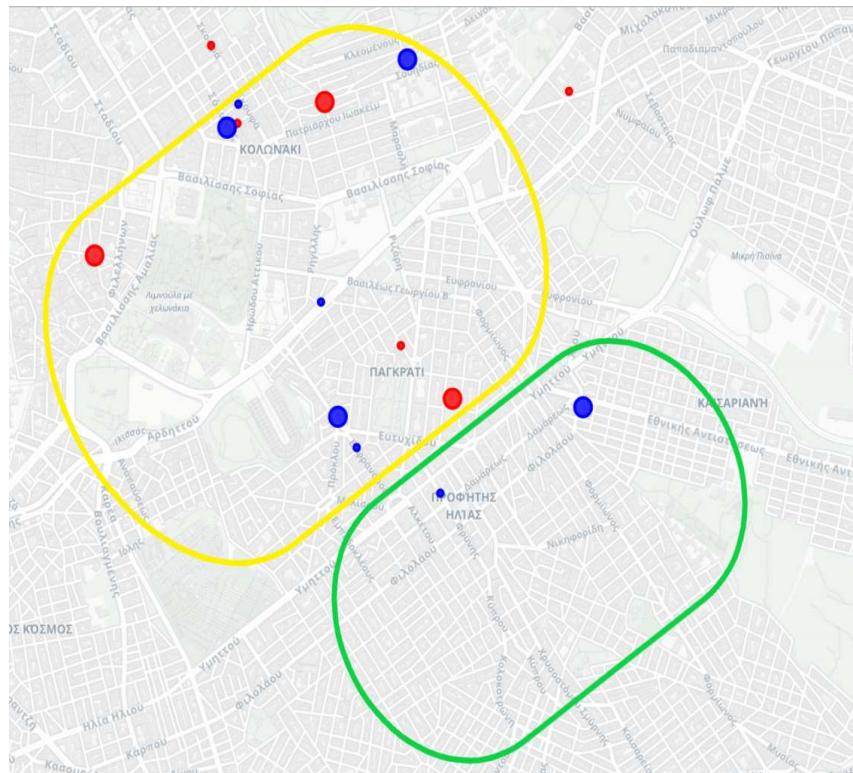
Cluster 3	
Name	Champion with low quality
Summary	High: Winebars, Bistro Low: Expensive, Quality Average: Size, Popularity
Comment	This one city cluster is the only one that has at the same time a high number of wine bars and bistro but the number of bistro (40 per 100k persons) is greater (almost double) than the number of wine bars (24 per 100k persons). It is worth mention that the Quality of the existing wine bars are below average. This city needs more analysis and it is a good candidate to host one more wine bar.
Cluster 5	
Name	Quantity not Quality
Summary	High: Winebars Low: Size, Bistro, Quality Average: Expensive
Comment	One of a kind city and the only one that compose this cluster. Kozani is a small city having an impressive number of wine bars compared to the population with 58 wine bars per 100k persons. On the other hand is very poor when it comes to popularity and quality. This fact makes Kozani a city worth thinking it as the city of your investment.
Cluster 7	
Name	Trying to get Improved
Summary	High: Winebars, Bistro Low: Popularity, Quality Average: Size, Expensive
Comment	The third cluster in size (8 cities) that compares to cluster 0 "The Not so Popular Average" and cluster 4 "The Popular Average", is a cluster with high number of wine bars and bistros, almost the double of the average. This is exactly the difference with cluster 0 and the low popularity and quality is its difference with cluster 4. The fact that the winebars of the cities of this cluster are not so popular and of low quality makes this cluster interesting to examine one by one the 8 cities in order to identify potential business opportunity.



Deeper on Proposed Clusters: Cluster 3 – “Champion with low quality

- ▶ Only one city → Viron
- ▶ Very high number of wine bars and bistro but low quality because of the below the average ratings.
- ▶ Bistros are almost double than the wine bars.
- ▶ A wine friendly city and more wine bars are welcome.
- ▶ Viron borders city of Athens, especially Pagrati area where many bars, café and restaurants exist
- ▶ Part of Athens metropolitan area. The campus of University of Athens is near and many students are living in Pagrati and Viron area.
- ▶ The big red wine circled (High rating with high popularity) are the top performance wine bars of the area while
- ▶ Small blue circles are places that are low performing in terms of popularity and rating.
- ▶ Yellow shape: Wine bars that belongs to Athens city but are closer to Viron.
- ▶ Green area is mainly the Viron area with a part of Pagrati.
- ▶ A wine friendly city. More wine bars are welcome.

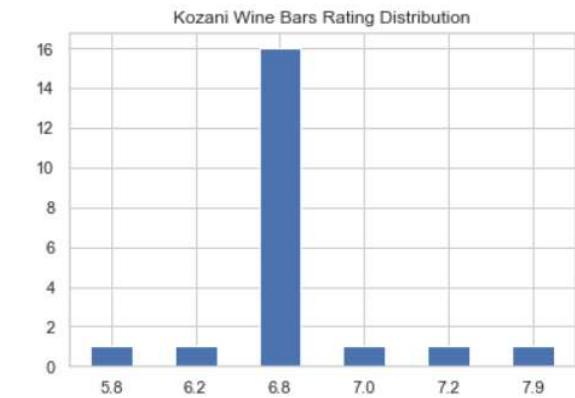
Cluster	City	Population	Latitude	Longitude	wine	bistro	avg_winebar_price	avg_winebar_likes	avg_winebar_rating
3	Viron	62500	37.970	23.750	15	25	1.866667	25.2	7.213333



Deeper on Proposed Clusters:

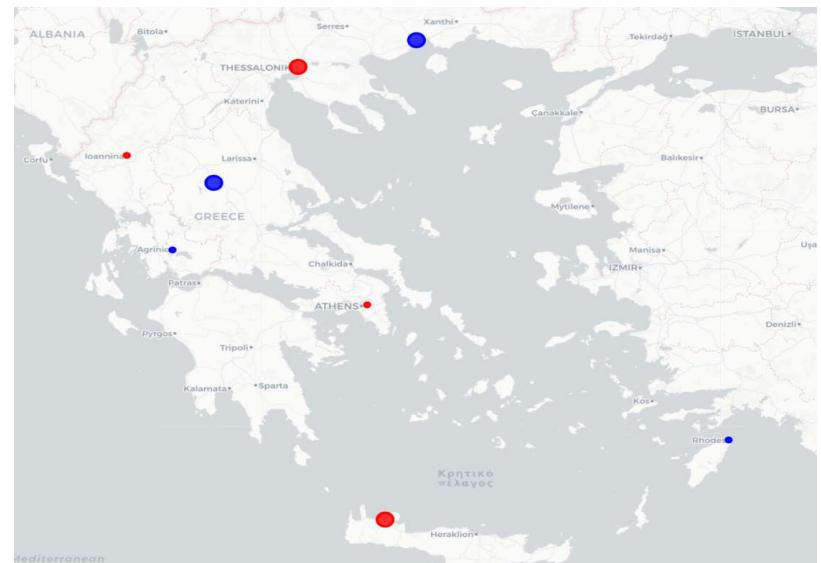
Cluster 5 – “Quantity not Quality”

- ▶ Only one city → Kozani
- ▶ A small Greek city located in northern Greece.
- ▶ Impressive number of wine bars with an average of 58 wine bars per 100k persons and
- ▶ At the same time it is on the bottom 5 cities in the average rating of the area's winebars with 6.8/10.
- ▶ There are no bistros in the area.
- ▶ Only 3 out of 21 wine bars have rating more than 7 and only 1 is above the Greek total average of 7.53.
- ▶ Only 3 out of 21 wine bars have rating more than 7 and only 1 is above the Greek total average of 7.53.
- ▶ Economy:
 - ▶ A large part of the population works in the the largest power plant in Greece. The [Ptolemaida Basin](#) hosts the [Western Macedonia Lignite Center](#), which is accountable for the production of 40% of the electric energy of the country.
 - ▶ Other famous products are [marble](#), [saffron](#) ([Krokos](#), Kozanis), [fruits](#), local [wines](#) and specialized arts and crafts industry.
- ▶ A city that people like hanging out to wine bars but they are not very happy with them.
- ▶ A good quality wine bar serving local wines and products with below the average prices has good possibilities to win the locals and become the new hot spot of the area.



Deeper on Proposed Clusters: Cluster 7 – “Trying to get Improved”

- ▶ 8 cities placed all over Greece
- ▶ High number of wine bars
- ▶ Poor popularity and quality.
- ▶ The big red circles are the top performance cities of the cluster while the small blue circles are low performing in terms of popularity and rating.
- ▶ The two small blue circles are cities Agrinio and Rhodes
- ▶ Agrinio is an agriculture area mostly known for the tobacco crops.
- ▶ Rhodes is a famous very touristic island of the Aegean sea.
- ▶ The city of Rhodes has 9 wine bars with an average rating below 7.
- ▶ The economy is tourist-oriented, and the most developed sector is service.
- ▶ With almost 2 millions of tourists every year (in 2015 the arrival number of tourists was about 1,901,000) a new wine bar with good quality (missing from the island) around or inside the historic center serving mainly greek wines from all over Greece it is the right thing to do.
- ▶ With the power of social networks and travel guides/place recommendation sites like Instagram, Foursquare, Trip Advisor etc the new wine bar will quickly travel all over the world.



Conclusion

- ▶ Analysis of the wine bar status of the 60 biggest cities in Greece via clustering
- ▶ Where and why someone could open a new wine bar .
- ▶ Greek Population and location info combined with Wine bars data taken from Foursquare API.
- ▶ 6 main variables used: Population, Number of Winebars / Person, Number of Bistro / Person, Wine Bar average Price Range, average Number of Likes and Rating average.
- ▶ Explore, Transform and prepare the Data for K-Means Clustering.
- ▶ Clustering algorithm produced 8 clusters
- ▶ Analyze clustering results and propose 3 clusters on which business opportunity was identified.



"Numeric data and analysis are a very strong tool in our hands, but understanding the culture of the people of a city, talk to them, walk to the streets and feel the vibes and the atmosphere is equally important before making your next business move"

Ilias Kapetanakis 2021