

THE XINOMAVRO PROJECT

WINE BAR ANALYSIS IN GREECE

INTRODUCTION

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](#)”.

This project is inspired from the love for the good Wine and the places you can find it, drink it and have amazing time with friends, the Wine Bars.

The main purpose of this analysis is to help people who are interested on opening a new wine bar in Greece to get an idea which is the current situation in the biggest Greek Cities and to make easier for them to understand where and why it's better to invest their money and open their Wine Bar.

We will involve in our analysis data like i) the population of the cities to have an indication of the size of each one of them, ii) the number of the Wine Bars, iii) the number of the Bistros as they are both wine related places and iv) metrics that characterize each Wine Bar like Foursquare Rating, Price Range and Number of Likes.

DATA

The following data will be used in the analysis:

1. GREEK CITIES POPULATION AND COORDINATES

DATA EXTRACTION

Data Source: [GREECE City & Town Population Geography Population Map cities coordinates location - Tageo.com](http://www.tageo.com/index-e-gr-cities-GR.htm) (<http://www.tageo.com/index-e-gr-cities-GR.htm>)

Extraction Method: Scrap HTML to get the table containing the Greek Cities, population and coordinates from the relevant table of the Tageo.com page

Number of rows: 60

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 SAMPLE

	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

2. GREEK CITIES VENUES (WINE BARS + BISTROT)

DATA EXTRACTION

Data Source: FOURSQUARE API : search Endpoint (GET <https://api.foursquare.com/v2/venues/search>)

Extraction Method: Call the API for each city found in dataset [1. GREEK CITIES POPULATION AND COORDINATES] by passing the latitude and longitude of each city and the specific category id of the "Wine bars" (4bf58dd8d48988d123941735) and "Bistro" (52e81612bcbc57f1066b79f1) as found in Foursquare documentation (<https://developer.foursquare.com/docs/build-with-foursquare/categories/>)

Number of rows for Wine Bars: 462

Number of rows for Bistro: 241

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. In addition by using exactly the same method the venues of the category 'Bistro' were retrieved.

Having the number of Wine Bars per City will be able to create a very important KPI by combining the Number of Winebars and the Population of each city: the Number of Winebars per Person. The same KPI was created for Bistrot showing the Number of Bistro per Person.

DATA QUALITY, CLEANSING AND TRANSFORMATION

- The major problem in this data source is that the same venue can be returned for more than 1 cities. This is possible to metropolitan areas like Athens and Thessaloniki because many cities consist the metropolitan area. The result is that there are venues having distance less than 1500m from multiple city centers. 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, as this is our main interest, and were removed from the Bistro data set.
- It should be noted that not all cities have wine bars.

DATA SAMPLE

	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

3. WINE BAR DETAILS

DATA EXTRACTION

Data Source: FOURSQUARE API : details Endpoint (GET https://api.foursquare.com/v2/venues/VENUE_ID)

Extraction Method: Call the API for each venue found after processing of venues returned from data set [2. GREEK CITIES VENUS (WINE BARS + BISTRO)]

Number of rows: 389

DATA DESCRIPTION

The data source contains details about each Winebar venue. We will use it to get important data for each venue like:

- Rating (score from 1 to 10 with 1 decimal point)
- Price Range (from 1 (least pricey) - 4 (most pricey))
- Number of Likes

We will use the Average Rating, Price Range and Number of Likes for the Winebars of each city, to the analysis and the clustering of the cities.

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:
 - o Missing Price = Average Price Value rounded in the closer integer
 - o Missing Rating = Average Rating Value rounded in 1 decimal

DATA SAMPLE

	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