European Union (EU): Diversity and Business Opportunities

Table of Contents

Introduction

The European Union (EU) is a unique economic and political entity consisted of 27 EU countries that cover much of the European continent (**Source:** EU's official web page, https://europa.eu/european-union/about-eu/eu-in-brief_en, Accessed 20th November, 2020). The predecessor, European Economic Community (EEC), was created in 1958 after the Second World War, and initially increasing economic cooperation between six countries: Belgium, Germany, France, Italy, Luxembourg and the Netherlands. Since then 22 more countries joined the EU. Initially, the EEC was begun as a purely economic union and it has been evolved into an organization expanding policy areas, from climate, environment and health to external relations and security, justice and migration. In 1993, the name EEC was changed to EU. The EU consists 27 countries now after the exiting of Great Britain in January, 2020.The Schengen Area, an area without internal borders, an area within which citizens, many non-EU nationals, business people and tourists can freely travel without being subjected to border checks.

Aim

Because EU countries, under the Schengen Area, allows free circulation of EU citizens, and non-citizens within the EU, there is a diverse population that belongs to different ethnic groups, religions, and cultures. About 400 million of the population within EU belong to 12 major nationalities, and about 60 million belong to minor ethnic groups (based on nationality). Current trends in migration and working opportunities indicate that the European populations will be becoming more and more ethnically and culturally diverse on the basis of current trends. Therefore

The aim of this project is twofold. **Aim 1:** To analyze some available data in these countries to evaluate similarities and/or dissimilarities between the cities within EU countries. A collection of 93 cities are used in the analysis. **Aim 2:** To evaluate venue data between select cities (based on higher population) to determine a suitable city to open a restaurant and which type of cuisine based on city's cultural diversity depending on current cuisine trends in these countries.

Aim 1 will help understand the diversity within European countries while the **Aim2** will help making informed decisions in one of the business opportunities: Cuisine. This project can be useful for business owners and entrepreneurs who are looking to invest in the restaurant industry within the European Union. The main objective of this project is to analyze appropriate data and to provide recommendations for the stakeholders.

Data

Data that is required for this project have been obtained from different sources (mainly through web scraping), analyzed as outlined under the **methodology** below.

A1. European demographic data was scraped from Wikipedia to obtain some population data to investigate about population in European countries. Analysis of demographic data will help understand the diversity within EU countries.

A2. The population within city limits has been obtained from Wikipedia.

B. Forbes Best Countries for Business in 2018 indicates certain European countries are among the best countries for business: Sweden (#2), Netharlands (#4), Denmark (#7), and Switzerland (#10). This will also be helpful in determining a city to open a restaurant.

Sources:

https://en.wikipedia.org/wiki/Demographics_of_the_European_Union#Ethnic_composition, Accessed 20th November, 2020)(https://en.wikipedia.org/wiki/List_of_cities_in_the_European_Union_by_population_within_city_limits, Accessed 20th November, 2020)https://www.forbes.com/best-countries-for-business/list/#tab:overall, , Accessed 20th November, 2020)

Initial observation of EU population data is that Germany (18.5%), France (15%), Italy (13.6%, Spain (10.4%), and Poland (8.5%) (**Figure 1**) are among the highest populated countries; however, their population density is not the highest, meaning that there are some countries with less area having a higher population. For example, Malta has the highest population density among EU countries, and Malta has the highest % net migration of 35% (**Figure 2, Figure 3**). EU immigration data also shows that there are some countries with higher net migration whereas some countries has a negative trend. Further investigation of these data may help understanding the uniqueness or diversity in some cities and/ or countries.

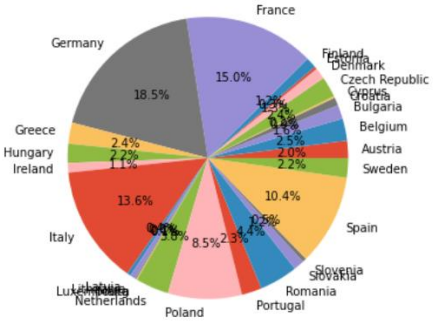

Figure 1. Population amongst EU countries

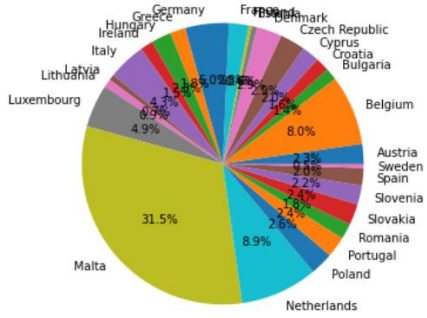

Figure 2. Population density amongst EU countries

Country	Pop_Growth_Pct	Nat_Growth_Pct	Net_Migr_Pct
Malta	36.8	1.6	35.3
Luxembourg	19.6	3.2	16.3
Cyprus	13.4	4.1	9.3
Ireland	15.2	6.2	9.0
Sweden	10.8	2.3	8.5

Figure 3. Net migration amongst top 5 EU countries

Eventually, these analyses will help better inform investors and stakeholders for decision making with respect to business opportunities as well as the trends in these countries.

Methods

To achieve these goals above, the following strategy was used:

A. Data Collection and Wrangling

1. EU city data has been collected from Wikipedia
    - Link: https://en.wikipedia.org/wiki/Demographics_of_the_European_Union#Population_by_country
    - Web scrape tables (country, city data, which include additional demographic data.)
2. Data wrangling to create a clean Pandas data frames for European countries and cities data.
3. BeautifulSoup package was used for extracting data tables.

B. Geographical Coordinates and Venue Data

1. Use Geo Location data
    - The geographical coordinates for European cities have been obtained from the GeoPy library in python. This data is used for plotting the map of European cities using the Folium library in python. The geocoder library in python has been used to obtain latitude and longitude data for various neighborhoods in European cities.
2. Use Foursquare API to extract Venue data
    - Venue data will be used for comparing the 93 cities.
    - Additional efforts will be made to evaluate other information available from Foursquare API such as residence etc.

C. Cluster Analysis

1. Compare cities through K-means clustering.
    - Venue data will be used to compare the 93 cities
    - Compare select 2-3 cities (dissimilar and/ or similar) that have been clustered together (for similarity) or differently (dissimilarity).
    - Efforts will be made to compare cities within the same country if they will be found to be in different clusters
2. Identify a city and which type of restaurant that can be opened in the city based on venue data clustered in #5.

D. Other Analysis

Statistical analysis methods such as multivariate data analysis will be explored by using additional EU data.

Results and Discussion

EU city data have been obtained different web pages, initially inspected, and analyzed through Foursquare API, followed by K-means cluster analysis.

Initial Observations

The following has been observed:

- EU cities have varying population as well as population density.
- Germany has the highest population whereas Malta has the highest population density.
- Net population growth was impacted by net migration rate in countries including **Malta (35.3%)**, **Luxumberg (16.3%)**, **Ireland (9.0%)**, and **Sweden (8.5%) (Figure 2, Figure 3)**.

EU cities selected for these analyses are shown in the map in Figure 4.

Figure 4. Selected EU cities in the analysis

Analysis results

- We have observed that there are a) similarities, b) dissimilarities between cities in different EU countries, c) similarities and dissimilarities between cities within the same country.

- K-means cluster analysis of venue data have shown 5 clusters of cities within the EU (Figure 5). There are remarkable findings in our analysis. There are some countries that have been clustered separately: **Spain, Italy, a city in Poland.** Italian cities have been clustered together with one exception: Milan.

- In addition, some cities of western European countries (France, Germany) are clustered with those of Nordic (Sweden, Denmark, and Finland) and Baltic Countries (Latvia, Estonia). Dublin, Ireland was clustered together with some western and eastern EU countries.

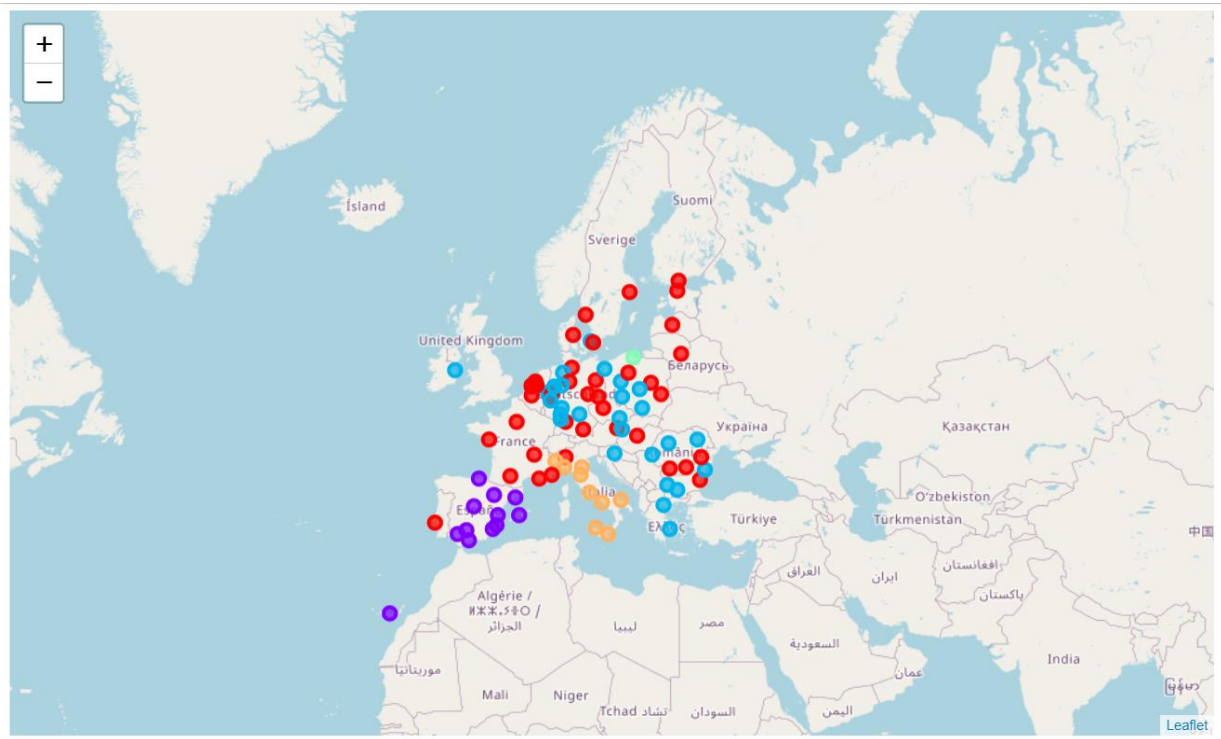

Figure 5. K-Means clustering of venue data obtained from the Foursquare API for the EU cities show that there are 5 clusters of cities based on venue data

- Write something about different venues

- Milan belongs to a different cluster compared to other Italian cities that were studied. **Milan can be a candidate city to open a restaurant business compared to other Italian cities because of the following reasons**: It is the second-most populous city in Italy after Rome. It has the second-largest economy among EU cities after Paris and is the wealthiest among EU non-capital cities. Therefore, Milan is **our candidate city** to open a restaurant. We know that Italian cuisine has attracted food enthusiasts all around the world; however, the exploration of a more unique cuisine, French Cuisine for example, can offer a nuanced discussion about the impact of location on the success of a food business.

- Based on available data, 15 central and shopping districts (Figure 6) were selected and 200 venues within 1000 meters (1 km) were analyzed.

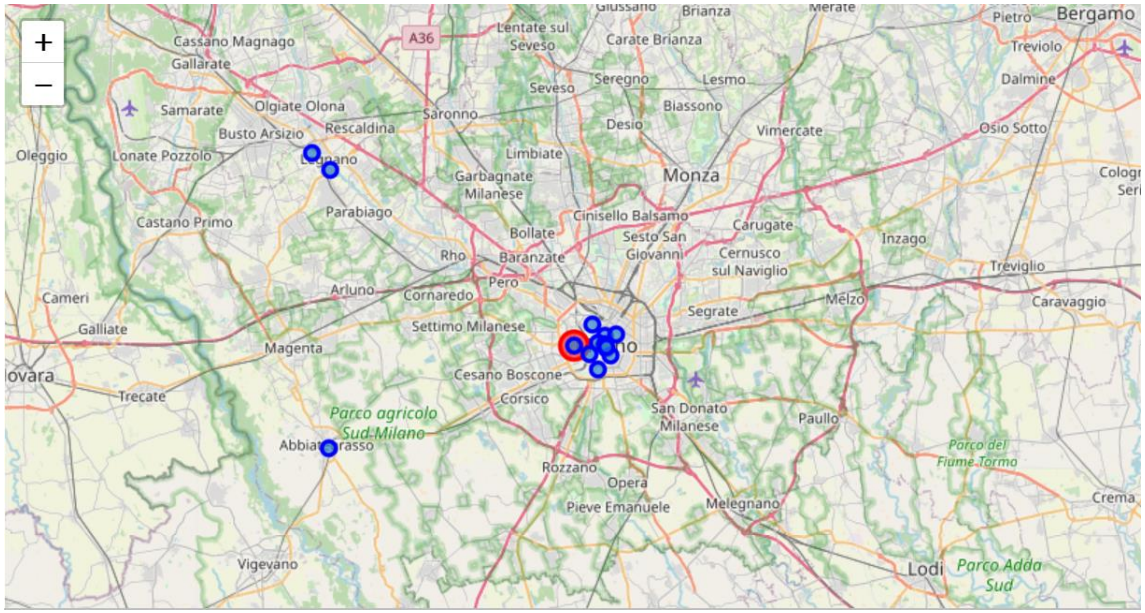

Figure 6. Central and shopping districts in Milan selected for venue analysis

- A set of 5 district clusters (Figure 7, Table 1-5) were observed. Clustering pattern observed was as follows: seven districts, 2x three districts, 2x individual districts. The 10 most common venues in the 5 clustered districts are shown in Table 1-5.

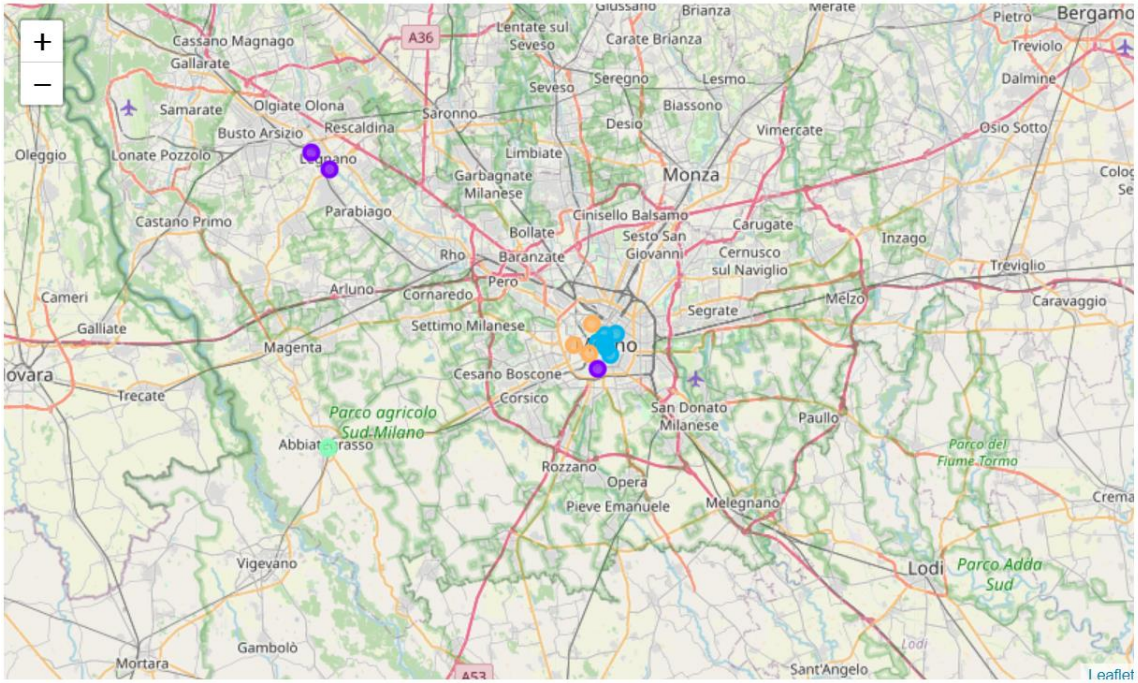

Figure 7. K-Means clustering of central and shopping districts in Milan using venue data obtained from Foursquares API

Table 1. Ten most common venues in the Cluster #1

	District	City	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
12	Buenos Aires	Milan	Pizza Place	Bus Stop	Ice Cream Shop	Gym	Pharmacy	Sandwich Place	Restaurant	Café	BBQ Joint	Italian Restaurant

Table 2. Ten most common venues in the Cluster #2

	District	City	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
5	Corso Magenta	Milan	Café	Bar	Italian Restaurant	Cocktail Bar	Playground	Lounge	Auto Garage	Indie Movie Theater	Ice Cream Shop	Market
7	Ticinese	Milan	Cocktail Bar	Ice Cream Shop	Italian Restaurant	Café	Seafood Restaurant	Japanese Restaurant	Bar	Restaurant	Boutique	Lounge
11	Torino	Milan	Cocktail Bar	Italian Restaurant	Bar	Electronics Store	Restaurant	Basketball Stadium	Pizza Place	Supermarket	Café	Motorcycle Shop

Table 3. Ten most common venues in the Cluster# 3

	District	City	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	Brera	Milan	Hotel	Italian Restaurant	Boutique	Japanese Restaurant	Ice Cream Shop	Dessert Shop	Cocktail Bar	Plaza	Restaurant	Park
1	Castello	Milan	Italian Restaurant	Ice Cream Shop	Plaza	Hotel	Café	Monument / Landmark	Wine Bar	Theater	Gift Shop	Dessert Shop
2	Duomo	Milan	Hotel	Italian Restaurant	Plaza	Boutique	Ice Cream Shop	Bakery	Monument / Landmark	Cocktail Bar	Gift Shop	Dessert Shop
3	Giardini	Milan	Boutique	Hotel	Italian Restaurant	Cocktail Bar	Women's Store	Park	Art Gallery	Dessert Shop	Ice Cream Shop	Men's Store
4	La Scala	Milan	Hotel	Boutique	Plaza	Italian Restaurant	Ice Cream Shop	Cocktail Bar	Monument / Landmark	Bakery	Art Museum	Men's Store
8	Università	Milan	Plaza	Boutique	Hotel	Italian Restaurant	Ice Cream Shop	Monument / Landmark	Furniture / Home Store	Bakery	Coffee Shop	Cocktail Bar

10	Vittorio Emanuele	Milan	Hotel	Boutique	Plaza	Italian Restaurant	Ice Cream Shop	Cocktail Bar	Bakery	Monument / Landmark	Lounge	Men's Store

Table 4. Ten most common venues in the Cluster #4

District	City	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
Monte Napoleone	Milan	Pizza Place	Supermarket	Cocktail Bar	Historic Site	Bar	Tea Room	Restaurant	Dessert Shop	Café	Italian Restaurant

Table 5. Ten most common venues in the Cluster #5

	District	City	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
6	Sant'Ambrogio	Milan	Italian Restaurant	Ice Cream Shop	Café	Pizza Place	Historic Site	Seafood Restaurant	Park	Science Museum	Dessert Shop	Hotel
13	Paolo Sarpi	Milan	Italian Restaurant	Chinese Restaurant	Pizza Place	Seafood Restaurant	Japanese Restaurant	Sushi Restaurant	Café	Bakery	Ice Cream Shop	Hotel
14	Corso Vercelli	Milan	Italian Restaurant	Ice Cream Shop	Plaza	Sandwich Place	Café	Supermarket	Pizza Place	Japanese Restaurant	Clothing Store	Asian Restaurant

Summary: In the clustered districts within a radius of 1 km, the number of restaurants and/ or hotels were found among the 10 most common venues as follows.

Cluster 1: Buenos Aires has 2 restaurants amongst the ten most common venues (4% frequency for restaurants amongst top 5 venues).

Cluster 2: Corso Magenta has 2 restaurants amongst the ten most common venues (7% frequency for restaurants amongst top 5 venues).

Ticinese has 4 restaurants amongst the ten most common venues (9% frequency for restaurants amongst top 5 venues).

Torino has 2 restaurants in the ten most common venues (15% frequency for restaurants amongst top 5 venues).

Cluster 3: Brera has 3 restaurants and a hotel amongst the ten most common venues (13% frequency for restaurants amongst top 5 venues).

Castello has 1 restaurant and a hotel amongst the ten most common venues (10% frequency for restaurants amongst top 5 venues).

Duomo has 1 restaurant and a hotel amongst the ten most common venues (8% frequency for restaurants amongst top 5 venues).

Giardini has 1 restaurant and a hotel amongst the ten most common venues (6% frequency for restaurants amongst top 5 venues).

La Scala has 1 restaurant and a hotel amongst the ten most common venues (7% frequency for restaurants amongst top 5 venues).

Università has 1 restaurant and a hotel amongst the ten most common venues (6% frequency for restaurants amongst top 5 venues).

Vittorio Emanuele has 1 restaurant and a hotel amongst the ten most common venues (8% frequency for restaurants amongst top 5 venues).

Cluster 4: Monte Napoleone has 2 restaurants amongst the ten most common venues (0% frequency for restaurants amongst top 5 venues).

Cluster 5: Sant'Ambrogio has 2 restaurants and a hotel amongst the ten most common venues (19% frequency for restaurants amongst top 5 venues).

Paolo Sarpi has 5 restaurants and a hotel amongst the ten most common venues (27% frequency for restaurants amongst top 5 venues).

Corso Vercelli has 3 restaurants amongst the ten most common venues (7% frequency for restaurants amongst top 5 venues).

Conclusion

- Population in EU is diverse
- Certain countries are higher in population density
- Migration has a significant effect on countries' population
- Immigration will also benefit diversified cuisine types
- Milan is different than other cities studied based on venues
- **Milan found 5 different clusters of districts based on the venues compared to other districts**
- **Milan's Monte Napoleone, Buenos Aires, Giardini districts** are better locations to open a restaurant. It is recommended to open **a restaurant with non-Italian cuisine such as French or Spanish cuisine.**
- Future studies will be needed to perform a grid search to find restaurant density in districts of Milan. Restaurant density will be used - to determine the best place to establish a restaurant.