Capstone Project The Battle of Neighborhoods

Prospects of a Coffee Shop close to business areas in Milan, Italy.

Background of the Business Problem

- Milan is a metropolis in Italy's northern Lombardy region. It is a global capital of fashion and design. Home to the national stock exchange, it’s a financial hub also known for its high-end restaurants and shops. Milan is definitely one of the best places to start up a new business.
- During the daytime, especially in the morning and lunch hours, office areas provide huge opportunities for coffee shops. Reasonably priced (one lunch meal 5€) shops are usually always full during the lunch hours (11 am — 2 pm) and, given this scenario, we will go through the benefits and pitfalls of opening a breakfast cum lunch coffee shop in highly dense office places.
- The core of Milano is made of 9 municipalities but, we will later concentrate on 4 most busiest business boroughs of Milan: Centro Storico, Stazione Centrale, Città Studi e Porta Garibaldi to target daily office workers.
- We will go through each step of this project and address them separately. I first outline the initial data preparation and describe future steps to start the battle of neighborhoods in Milan.

Target audience

This project is mainly aimed at group of people:

- Entrepreneurs who wants to invest or open a coffee shop. This analysis will be a comprehensive guide to start or expand coffee shops targeting the large pool of office workers in Milan during lunch hours.
- People with passion for data that want see an example of applied data science methodology

Data collection

1. Milan Municipalities Table from Wikipedia
2. Dataset Coordinates of Milan municipalities: Geopy
3. Average Property Price in Major Municipalities of Milan
4. Foursquare data

Milan Municipalities Table from Wikipedia

Source: https://en.wikipedia.org/wiki/Zones_of_Milan

Web page scrapping:

```
response = requests.get('https://en.wikipedia.org/wiki/Zones_of_Milan').text
soup = BeautifulSoup(response,'lxml')
table = soup.find('table', {'class':'wikitable sortable'})
```

```
table_rows = table.find_all('tr')

res = []
for tr in table_rows:
    td = tr.find_all('td')
    row = [tr.text.strip() for tr in td if tr.text.strip()]
    if row:
        res.append(row)


df = pd.DataFrame(res, columns=["Municipio", "Description", "Area", "Population","Densit
#remove column Districts
df_municip = df.drop(columns=['Districts'])
df_municip
```

	Num	LongName	Area	Population	Density
0	1	Centro storico	9.67	96,315	11,074
1	2	Stazione Centrale, Gorla, Turro, Greco, Cresce...	12.58	153.109	13,031
2	3	Città Studi, Lambrate, Porta Venezia	14.23	141,229	10,785
3	4	Porta Vittoria, Forlanini	20.95	156.369	8,069
4	5	Vigentino, Chiaravalle, Gratosoglio	29.87	123,779	4,487
5	6	Barona, Lorenteggio	18.28	149,000	8,998
6	7	Baggio, De Angeli, San Siro	31.34	170,814	6,093
7	8	Fiera, Gallaratese, Quarto Oggiaro	23.72	181,669	8,326
8	9	Porta Garibaldi, Niguarda	21.12	181,598	9,204

Dataset Coordinates of Milan municipalities

Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.

Source: https://github.com/geopy/geopy

```
from geopy.geocoders import Nominatim
geolocator = Nominatim()
location = geolocator.geocode("Milano, MI, Lom, Italia")
address = []
coord = []
address = "Municipio "+ df['Municipio']+", Milano, MI, Lom, Italy"
coord = address.apply(geolocator.geocode).apply(lambda x: (x.latitude, x.longitude))
df_municip['Coordinates'] = coord
df_municip
```

]:

	Municipio	Description	Area	Population	Density	Name	Coordinates
0	1	Centro storico	9.67	96,315	11,074	Centro storico	(45.4672806, 9.18596201012092)
1	2	Stazione Centrale, Gorla, Turro, Greco, Cresce...	12.58	153.109	13,031	Stazione centrale	(45.4999899, 9.21899862605208)
2	3	Città Studi, Lambrate, Porta Venezia	14.23	141,229	10,785	Città studi	(45.4897995, 9.24103088613553)
3	4	Porta Vittoria, Forlanini	20.95	156.369	8,069	Porta vittoria	(45.446205, 9.23934326456628)
4	5	Vigentino, Chiaravalle, Gratosoglio	29.87	123,779	4,487	Vigentino	(45.4198458, 9.19815254935277)
5	6	Barona, Lorenteggio	18.28	149,000	8,998	Barona	(45.43792895, 9.1456481512419)
6	7	Baggio, De Angeli, San Siro	31.34	170,814	6,093	Baggio	(45.469598, 9.11475732032285)
7	8	Fiera, Gallaratese, Quarto Oggiaro	23.72	181,669	8,326	Fiera	(45.50001535, 9.12277422514706)
8	9	Porta Garibaldi, Niguarda	21.12	181,598	9,204	Porta garibaldi	(45.5077041, 9.17941004494827)

Average Property Price in Municipalities of Milan

- Source: Mercato Immobiliare a Milano
- https://www.mercato-immobiliare.info/lombardia/milano/milano.html
- Web page scrapping

```
response2 = requests.get('https://www.mercato-immobiliare.info/lombardia/milano/milano.html').text
#soup2 = BeautifulSoup(response2,'lxml')
soup2 = BeautifulSoup(response2, 'html.parser')
table2 = soup2.find('table', {'id':'childrentable'})
```

```
table_rows2 = table2.find_all('tr')

res = []
for tr in table_rows2:
    td = tr.find_all('td')
    row = [tr.text.strip() for tr in td if tr.text.strip()]
    if row:
        res.append(row)


df_price = pd.DataFrame(res, columns=["Name", "Price","Link"])
df_price = df_price.drop(columns=['Link'])
```

	Shortname	Price
0	Centro storico	€ 6.500 /m²
1	Città studi	€ 3.950 /m²
2	Porta garibaldi	€ 5.500 /m²
3	Stazione centrale	€ 4.550 /m²

Foursquare data

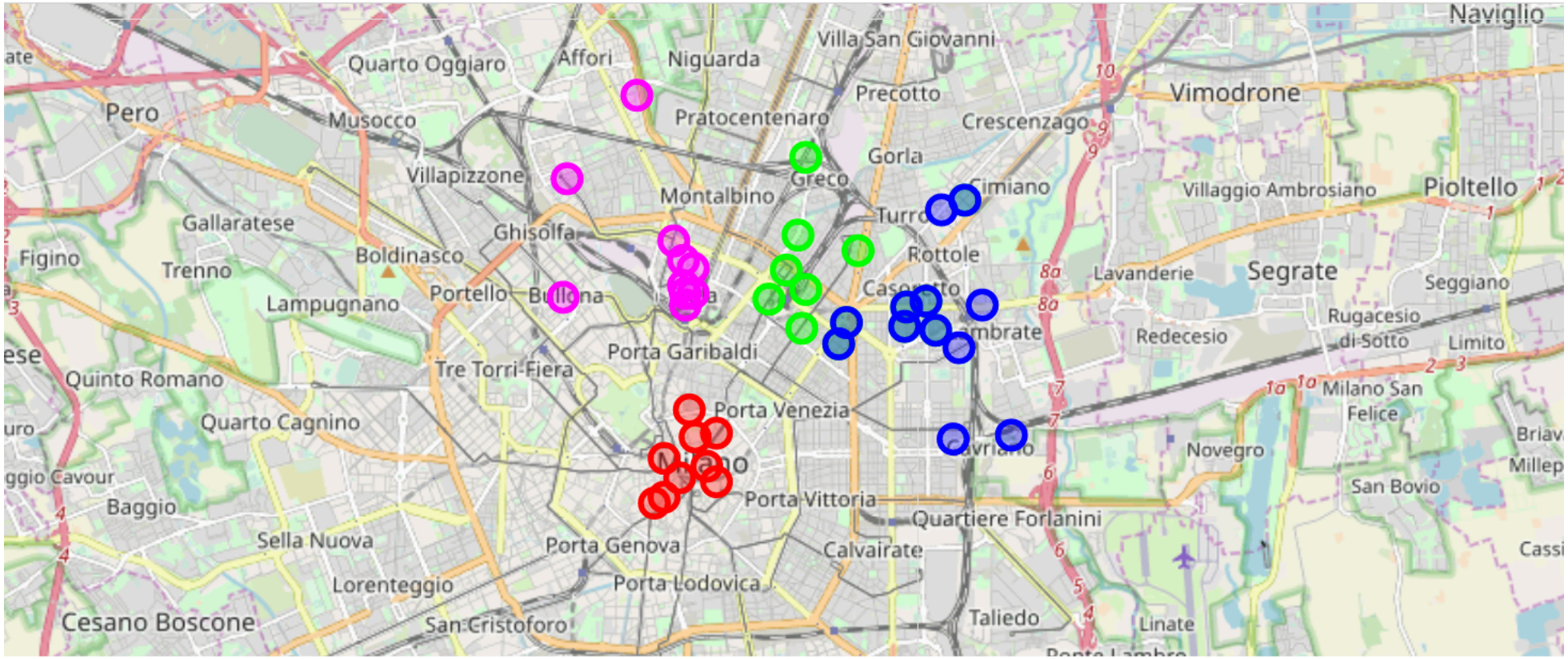
Empower users to explore millions of points-of-interests via your app or website by powering rich location data features like geo-tagging and venue search.

https://developer.foursquare.com

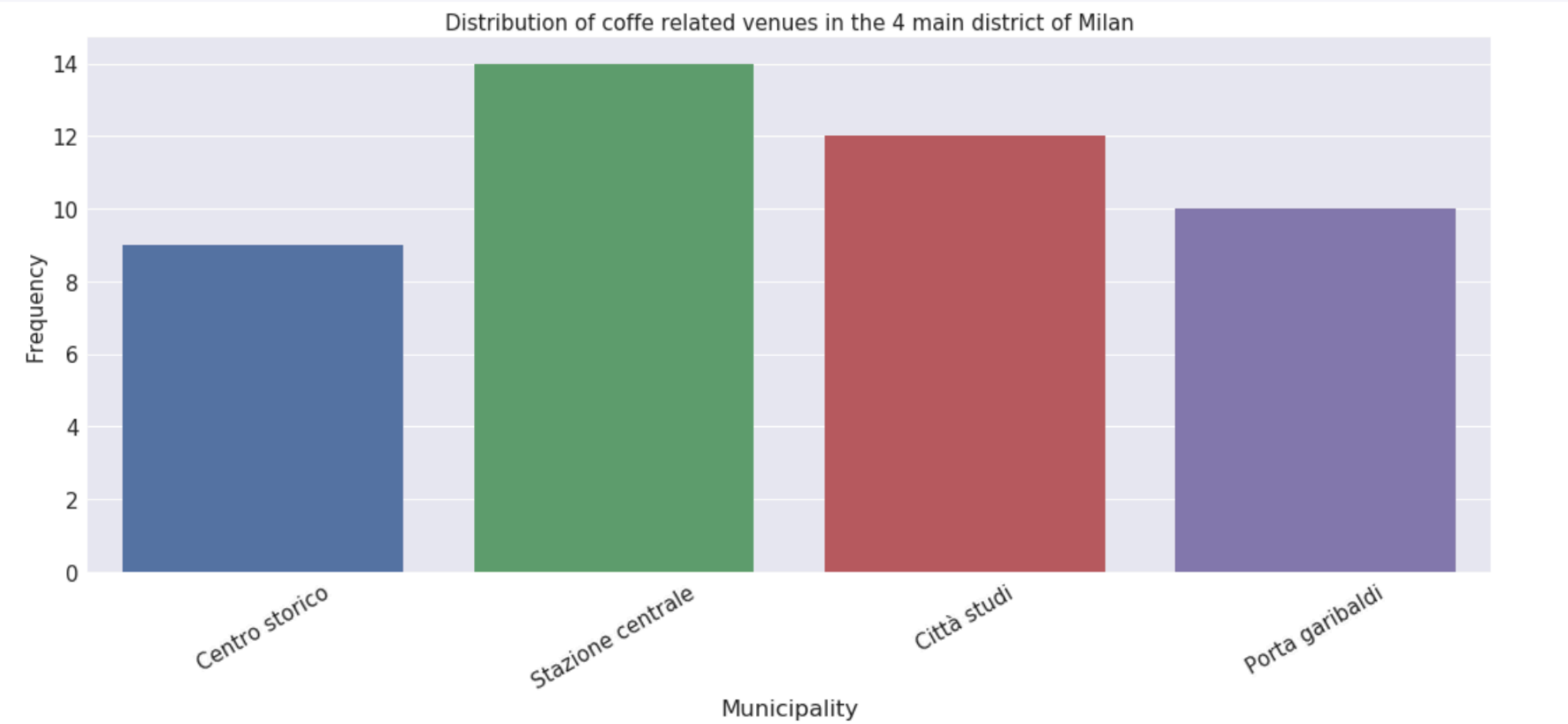
	District	Dist_Latitude	Dist_Longitude	Venue	Venue_Lat	Venue_Long	Venue_Category
300	Porta garibaldi	45.507704	9.17941	Virgin Active	45.502018	9.182590	Gym / Fitness Center
301	Porta garibaldi	45.507704	9.17941	Al Paradiso Della Pizza	45.511351	9.175416	Pizza Place
302	Porta garibaldi	45.507704	9.17941	Birrificio La Ribalta	45.507038	9.173219	Brewery
303	Porta garibaldi	45.507704	9.17941	Istanbul Kebab	45.510421	9.176128	Kebab Restaurant
304	Porta garibaldi	45.507704	9.17941	Esselunga	45.512380	9.173461	Supermarket
305	Porta garibaldi	45.507704	9.17941	Total Natural Training	45.509142	9.194221	Gym
306	Porta garibaldi	45.507704	9.17941	Spirit de Milan	45.506678	9.159744	Ballroom
307	Porta garibaldi	45.507704	9.17941	Il Bucatino con Giardino	45.502088	9.165959	Italian Restaurant
308	Porta garibaldi	45.507704	9.17941	Sushi Ran	45.500348	9.170942	Japanese Restaurant
309	Porta garibaldi	45.507704	9.17941	Hotel La Residenza	45.512413	9.178770	Hotel
310	Porta garibaldi	45.507704	9.17941	Piazza Dergano	45.504034	9.175839	Plaza
311	Porta garibaldi	45.507704	9.17941	Palestra McFIT	45.504708	9.198828	Gym
312	Porta garibaldi	45.507704	9.17941	Biologic Bar & Restaurant	45.512302	9.178709	Hotel Bar
313	Porta garibaldi	45.507704	9.17941	MiSciolgo	45.499214	9.165308	Ice Cream Shop
314	Porta garibaldi	45.507704	9.17941	Teatro della Cooperativa	45.515042	9.190504	Theater

Data Visualization – map leaflet

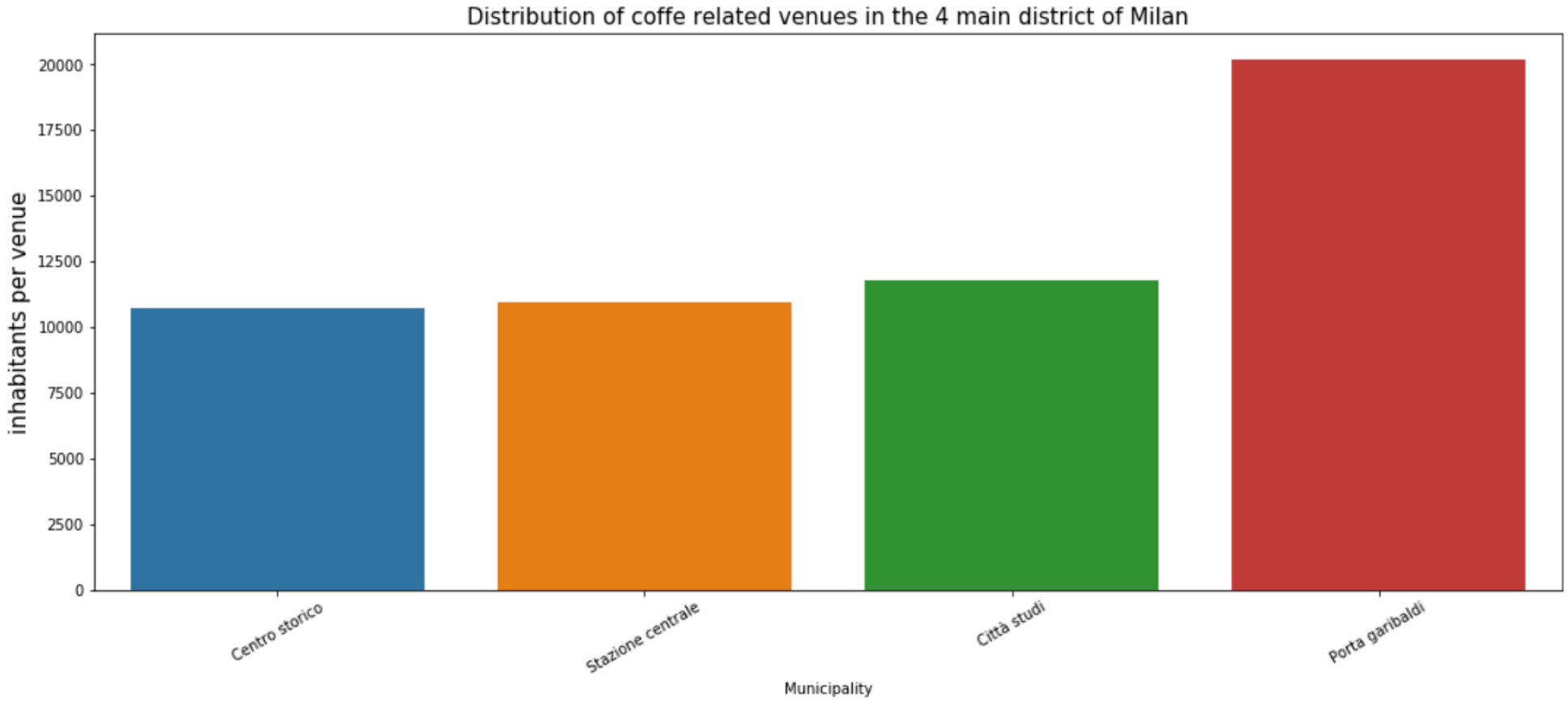
Foilum map



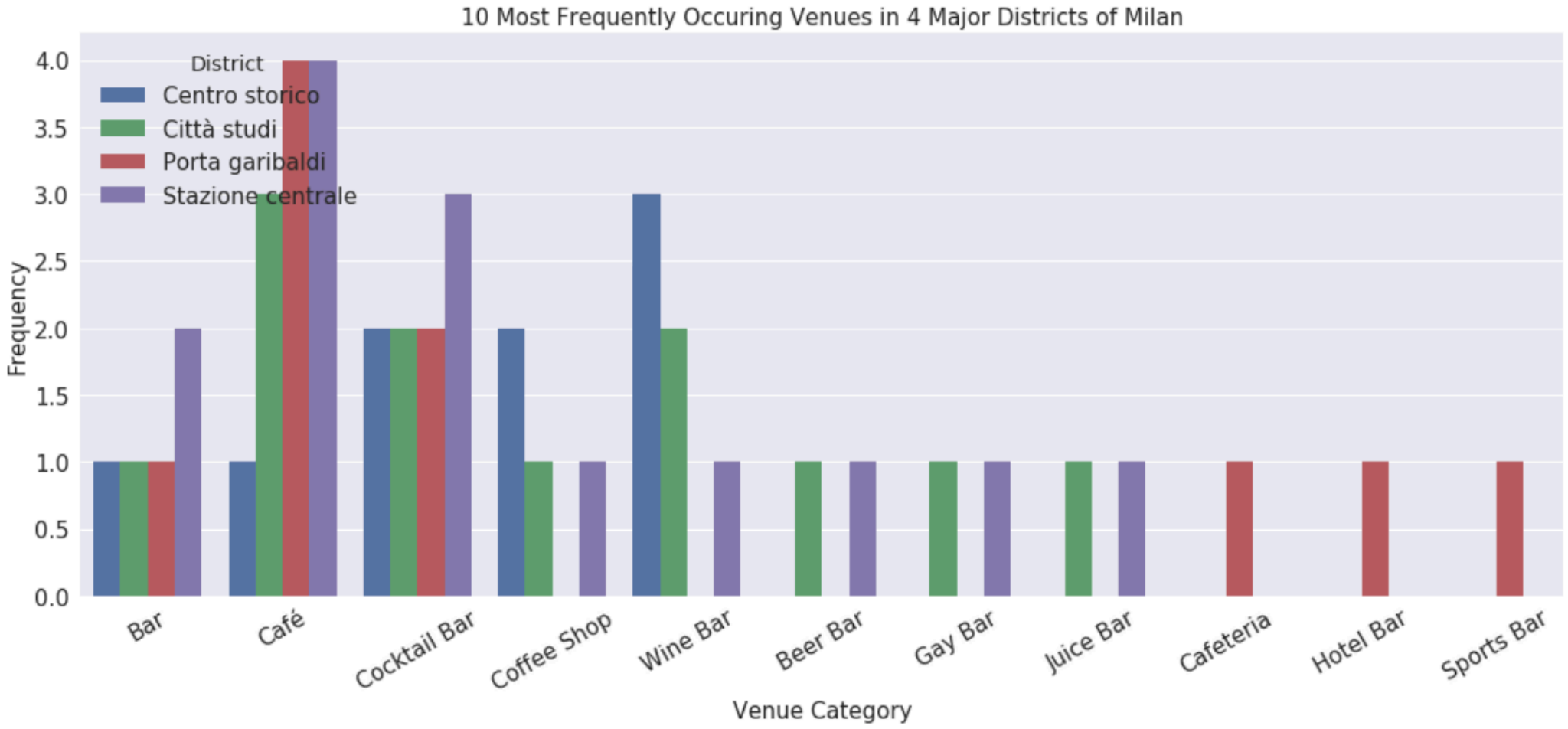
Data Visualization – Barplot 1



Data Visualization – Barplot 2



Data Visualization – Barplot 3



Result and discussion

- We limited our analysis to the 4 main district (municipalities) of Milan: Centro storico, Stazione centrale, Porta Garibaldi, Città Studi (highest average of property price).
- Foursquare data resulting from our analysis shows that the number of Cafe, Bar and Coffee shop is quite low considering the population and the size of the area.
- The most common category is Cafè and they are concentrated in Stazione centrale and Porta Garibaldi.
- Porta Garibaldi is the district with the highest ratio inhabitants per venue (over than 20 thousands). As this district also is one of the most expanding business areas in Milan it looks as best location for opening a new Café or Coffee shop.
- Anyway the accuracy of data purely depends on the data provided by FourSquare, as there could be venues that are not registered there.
- Also recommended district should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.