Coursera Capstone Project

Francesco M.

May 2020

A Guide to Milan for Tourists and Investors

Introduction

The city of Milan is constantly evolving, from high end fashion neighbourhoods to restaurants, cocktail bars and nightclubs. The city of Milan is world renowned for its high end fashion, millions of tourist trail the streets of Milan each year in search of its boutiques and fashion shows. Another key sector of the Milan environment and economy are the restaurants, pubs, cocktail bars and nightclubs.

The aim of this project is to localize these distinct areas in Milan's city center. This is essential for tourists, who through this insight can easily reach their favourite restaurants, shops or clubs in such a big city, but also very important to investors or whoever wants to open a business in Milan. Through the categorization of the various areas in Milan's city center it will become evident that there is a clear distinction between the restaurants, boutiques and nightlife, this is a vital information when choosing where to open a new business.

Data

To provide such an analysis, the project shall use the geospatial data provided by the official website of the municipality of Milan:

- https://dati.comune.milano.it/dataset/ds634-numeri-civici-coordinate

The following website presents a dataset with all the addresses in the city of Milan, more than 63,000 entries. The website supports two different types of formats for download, JSON and CSV, both were originally zipped, so had to be unzipped before importing. Once the data is correctly acquired and cleaned, it can be transformed in a dataset.

As by Figure 1, the dataframe presents the district (Milan has a total of 9 ditricts) the postal code, the road name and the longitude and latitude. This dataframe shall be reduced to only the city center (district 1), and further reduced by means of *KMeans* clustering to generate a series of 100 evenly spread out point's of interest that shall characterize the city center, as shown in Figure 3.

	District	Postal Code	Road Name	Longitude	Latitude
0	1	20121	Bastioni DI PORTA NUOVA	9.189394	45.480053
1	1	20121	Bastioni DI PORTA VENEZIA	9.202396	45.475062
2	1	20121	Bastioni DI PORTA VOLTA	9.182029	45.479434
3	1	20121	Corso DI PORTA NUOVA	9.191710	45.475896
4	1	20121	Corso GIACOMO MATTEOTTI	9.195200	45.466907

Figure 1: Data imported in Pandas DataFrame

A clustering technique has been used in this early stage of data formatting to ensure that the distribution of data was evenly spread out along Milan's city center. In fact, randomly selecting points from the dataset would have resulted in a patchy distribution of datapoints, making the analysis less effective. This methodology shall be further explored in the following section.

	Cluster Labels	Postal Code	Road Name	Longitude	Latitude
0	0	20121	Via LUIGI ALBERTINI	9.182543	45.474717
1	1	20145	Via GIOVANNI RANDACCIO	9.167483	45.476815
2	2	20154	Piazza ERCOLE LUIGI MORSELLI	9.174736	45.478654
3	3	20129	Corso VENEZIA	9.204959	45.474350
4	4	20123	Via PIETRO AZARIO	9.168378	45.460085

Figure 2: District 1 DataFrame containing 100 evenly spread points

Once the 100 datapoints are located, a call shall be made to the Foursquare API, one for each datapoint, to explore all the venues around each datapoint. The endpoint shall be:

- https : //api.foursquare.com/v2/venues/explore

This call shall return venues of all categories around each datapoint. These can be added to a dataframe, then grouped for each point of interest, as shown in Figure 3

	Road Name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bastioni DI PORTA VENEZIA	Italian Restaurant	Hotel	Pizza Place	Art Gallery	African Restaurant
1	Corso DI PORTA VIGENTINA	Restaurant	Wine Bar	Pizza Place	Italian Restaurant	Bistro
2	Corso VENEZIA	Italian Restaurant	Pizza Place	Café	African Restaurant	Art Gallery
3	Corso VITTORIO EMANUELE II	Boutique	Plaza	Italian Restaurant	Sporting Goods Shop	Monument / Landmark
4	Foro BUONAPARTE	Italian Restaurant	Café	Plaza	Ice Cream Shop	Platform

Figure 3: Top categories per point of interest

Moreover, these can be clustered to observe the distribution of different categories in Milan's city center. A final analysis can be done observing the two main categories illustrated in the introduction, the nightlife (bars, pubs and clubs) and fashion (boutiques and clothing stores). This will give the desired insight in the distinction between these two different areas of Milan. These clusters shall be amply explained in the following sections.

Code and Methodology

Results

Observations

Conclusion