

# TABLE OF CONTENTS

<b>Introduction :</b>	<b>3</b>
Problem statement	3
Target audience and Utility	3
<b>Data section :</b>	<b>3</b>
Dataset fabrication:	3
Area of Interest and Description:	4
Relevant Location types:	5
Example on next page --->	5
<b>Area of Interest:</b>	<b>7</b>
<b>Dividing the Circular region into a grid:</b>	<b>8</b>
Cartesian Implementation:	8
Unrefined Grid:	8
Refined Grid:	9
Map Implementation:	9
<b>Finding the critical size:</b>	<b>11</b>
<b>Making the grid on AOI:</b>	<b>12</b>
<b>Converting json file to clean dataframe:</b>	<b>13</b>
<b>Calling the API on all cells:</b>	<b>13</b>
<b>Creating informative labels:</b>	<b>14</b>
Event time:	14
Event Scale:	14
Event Type:	14
<b>After Grouping:</b>	<b>16</b>
<b>Conclusion:</b>	<b>16</b>

# **1. Introduction :**

## **1.1. Problem statement**

You are a performing arts organizing agency whose mainstream source of revenue is organizing sensational events/gigs for public entertainment from conception to execution. You aim to organize hundreds of events both highly frequent small scale ones and occasional large scale ones. You work with all kinds of performance artists like actors, comedians, dancers, magicians, circus artists, musicians, and singers, songwriters, choreographers, and so on.

One of the major decisions that have to be made is the selection of an appropriate venue for a performance. For this, the agency needs not only a holistic list of venues in an area but also to group it with tags like nightlife, small scale, relevant performance art that can be performed here, etc. So that an optimal decision can be arrived at.

Once the database for the area of interest is built, a simple K-means clustering would reveal the best zones for further classification and analysis.

## **1.2. Target audience and Utility**

The database once built and analyzed would be of tremendous utility to all companies/agencies involved in any sub-domain of the performing arts ecosystem. A few scenarios would help in grasping of the underlying value of this project:

- 1) An individual comic/ group of comics scouting for popular locations to perform and showcase their gigs throughout a particular city/area.
- 2) Musical Bands/ individual performers looking for crowded venues and popular locations to perform frequently to increase ticket sales.
- 3) Play actors/ dance performers and a host of other performing artists aiming to maximize the reach, popularity, and ticket sales for their art.
- 4) Supporting staff like choreographers, songwriters, scriptwriters, cameramen, etc looking for employment opportunities.
- 5) Not only the supply side but the demand side too, for instance, art enthusiasts and staple audiences of each of the above-mentioned arts can benefit too.

## **2. Data section :**

### **2.1. Dataset fabrication:**

The aim is to construct a comprehensive, all-encompassing database of venues, complete with useful labels like enables night shows, the scale of the event that can be hosted, etc.

To construct such a dataset, one would have to partition the geographical area of interest into non-overlapping subsets of the critical size (see next para). The query will be run on the center of the geographical confine and relevant properties of the results will be used to populate a conspicuous, intuitive dataset for further use.

The foursquare API for venue search has a cap of 30 results for a given query, implying that all the popular performing arts venues might not be revealed if the area size exceeds a critical size (at which any further decrement in size would decrease the query results to below 30).

### **2.2. Area of Interest and Description:**

For this project, we would be looking at the very heart of the culture-rich city of Milan.

Milan is the capital of Lombardia, along with being the richest and most populous region of Italy, it is the second-largest in the country, and is best known as an economic and financial center but it also has its fair share of cultural and architectural attractions.

Zone 1 of Milan city comprises within it, districts of Brera and Porta Sempione, which bubble with its richness of Art and Nightlife.

Brera houses the Brera Academy of Fine Arts and the Brera Art Gallery, which prominently contributed to the development of Brera as an artists' neighborhood and a place of bohemian atmosphere, sometimes referred to as "the Milanese Montmartre".

Other features that contribute to the character of Brera include restaurants, bars, night clubs, antique and art shops, colorful street markets, as well as fortune tellers' booths.

The whole area is one of the centers of the so-called Milanese Moviada (nightlife), with a number of bars, pubs, restaurants, and discos; since the RAI public television company, as well as some major radio stations, have their headquarters in the area, frequent appearances of celebrities contribute to the popularity of the Sempione's nightlife venues.

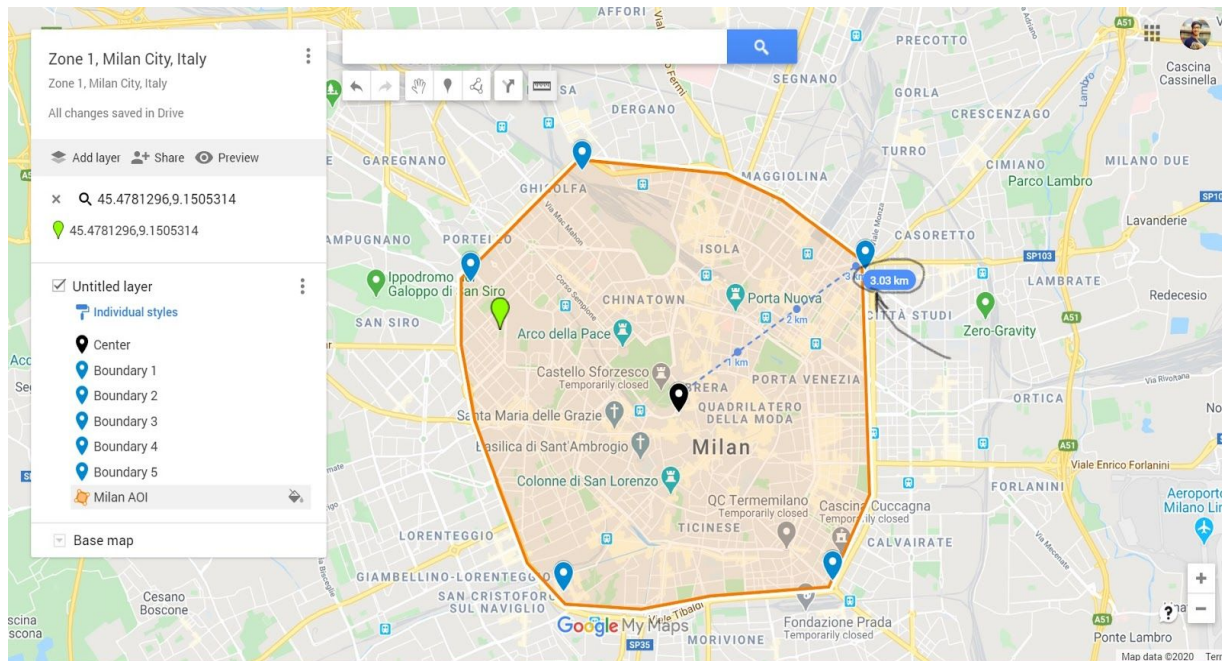
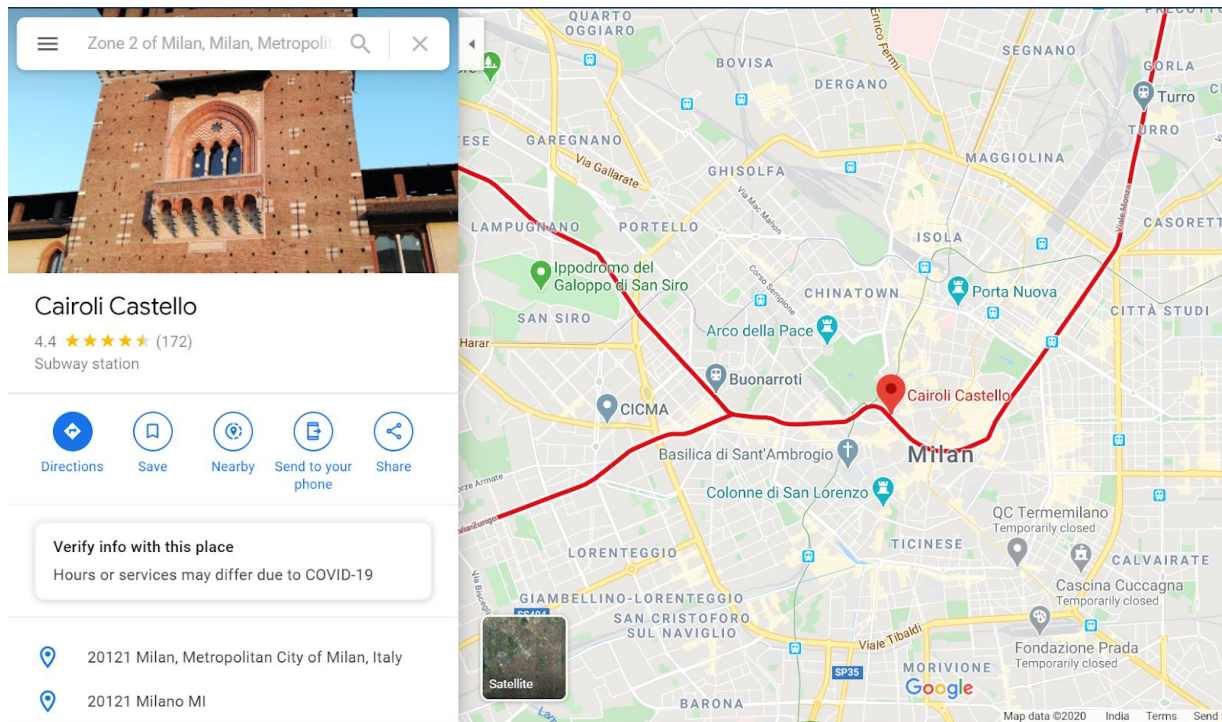
## 2.3. Relevant Location types:

1. ***Amphitheater***  
56aa371be4b08b9a8d5734db
2. ***Comedy Club***  
4bf58dd8d48988d18e941735
3. ***Concert Hall***  
5032792091d4c4b30a586d5c
4. ***General Entertainment***  
4bf58dd8d48988d1f1931735
5. ***Opera House***  
4bf58dd8d48988d136941735
6. ***Theater***  
4bf58dd8d48988d137941735
7. ***Lounge***  
4bf58dd8d48988d121941735
8. ***Nightclub***  
4bf58dd8d48988d11f941735
9. ***Music Venue***  
4bf58dd8d48988d1e5931735

	name	categories	address	lat	lng	labeledLatLngs	postalCode	cc	neighborhood	city
0	The Club	Nightclub	Corso Garibaldi, 97	45.477683	9.184380	[{"label": 'display', 'lat': 45.47768290977855...	20121	IT	Moscova, Milano, Lombardia	Milano
1	Messapika	Lounge	Via Alserio 3	45.491462	9.184598	[{"label": 'display', 'lat': 45.491462, 'lng':...	20159	IT	NaN	Milano
2	Barbarella Club	Nightclub	Bastioni di porta nuova 12	45.480217	9.187761	[{"label": 'display', 'lat': 45.480217, 'lng':...	NaN	IT	NaN	Milano
3	Trottoir (Le)	Nightclub	2 Via Tivoli	45.471117	9.184165	[{"label": 'display', 'lat': 45.471117, 'lng':...	20121	IT	NaN	Milano
4	whatsapp in a bak	Comedy Club	NaN	45.470039	9.179305	[{"label": 'display', 'lat': 45.47003936767578...	NaN	IT	NaN	NaN
5	Heineken Social Networking	General Entertainment	NaN	45.482116	9.187207	[{"label": 'display', 'lat': 45.482116, 'lng':...	NaN	IT	NaN	NaN
6	Boucheron Paris-Press Cocktail	General Entertainment	NaN	45.468656	9.193360	[{"label": 'display', 'lat': 45.468656, 'lng':...	NaN	IT	NaN	NaN
7	Pastrengo	Comedy Club	via pastrengo 12	45.486816	9.186635	[{"label": 'display', 'lat': 45.48681638000001...	20159	IT	NaN	Milano
8	Babbo Running	General Entertainment	Piazza Castello	45.469361	9.181685	[{"label": 'display', 'lat': 45.46936058971595...	NaN	IT	NaN	Milano
9	Milano Film Festival	General Entertainment	NaN	45.471724	9.178391	[{"label": 'display', 'lat': 45.47172396229453...	NaN	IT	NaN	NaN
10	Viberoom	Nightclub	Viale Monte Grappa 14	45.480963	9.189688	[{"label": 'display', 'lat': 45.48096250431438...	NaN	IT	NaN	Milano

### 3. Area of Interest:

*Cairoli Castello* is a subway station right at the **absolute center**(see Map) of zone 1 of Milan City. Let's call a **3.0 Km**(See Map) region from this center as our **AOI (Area of Interest)**.



## 4. Dividing the Circular region into a grid:

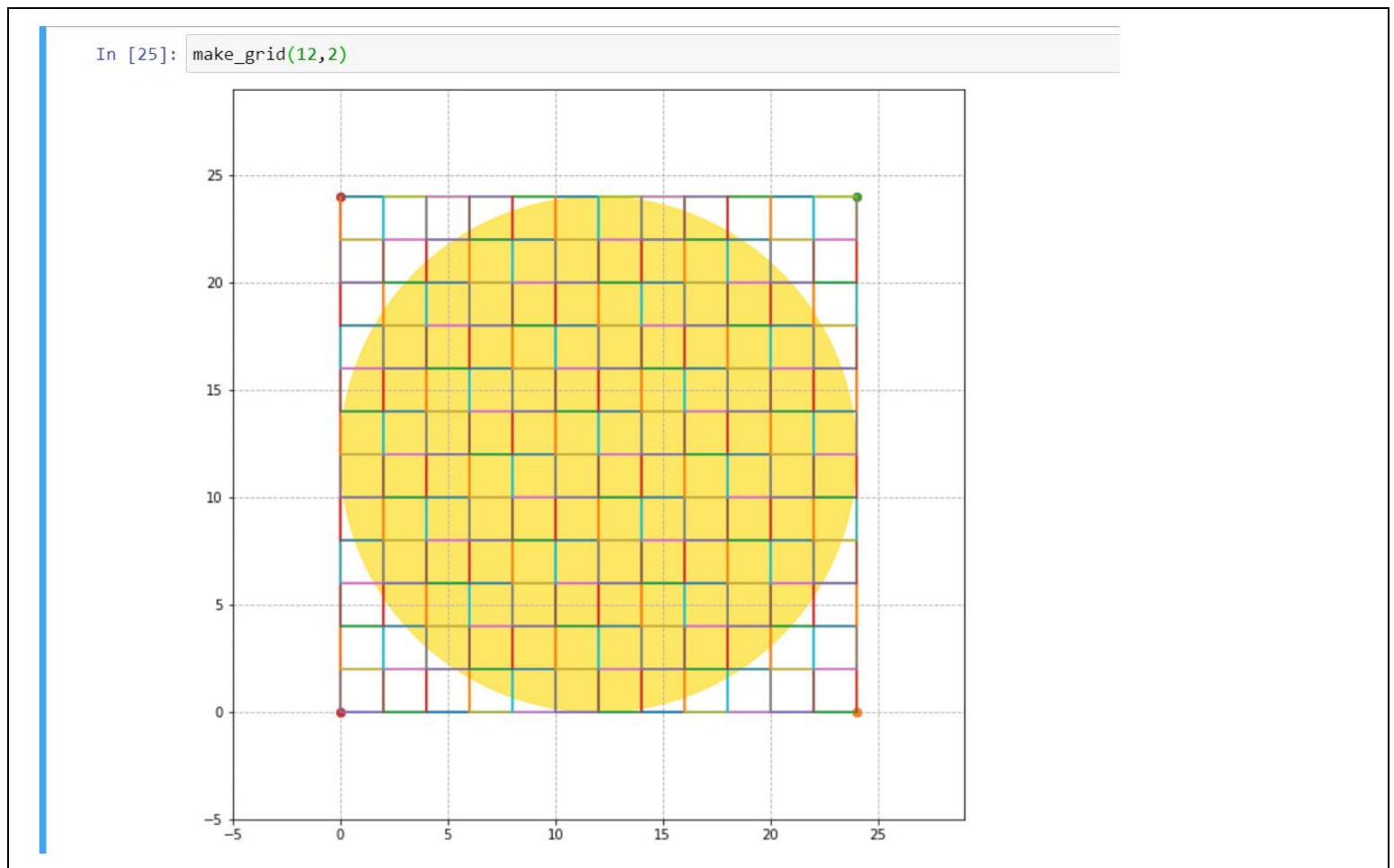
The venues API must be called at **every rectangular cell** of the grid by specifying its South-West and North-East corner. This 3.0Km radius region around the Center must be sliced into grids of cell size of the '**critical size**'.

### 4.1. Cartesian Implementation:

As a first step, let's try to solve this problem on the *Cartesian Plane* and *Visualize* it using Matplotlib.

#### 4.1.1. Unrefined Grid:

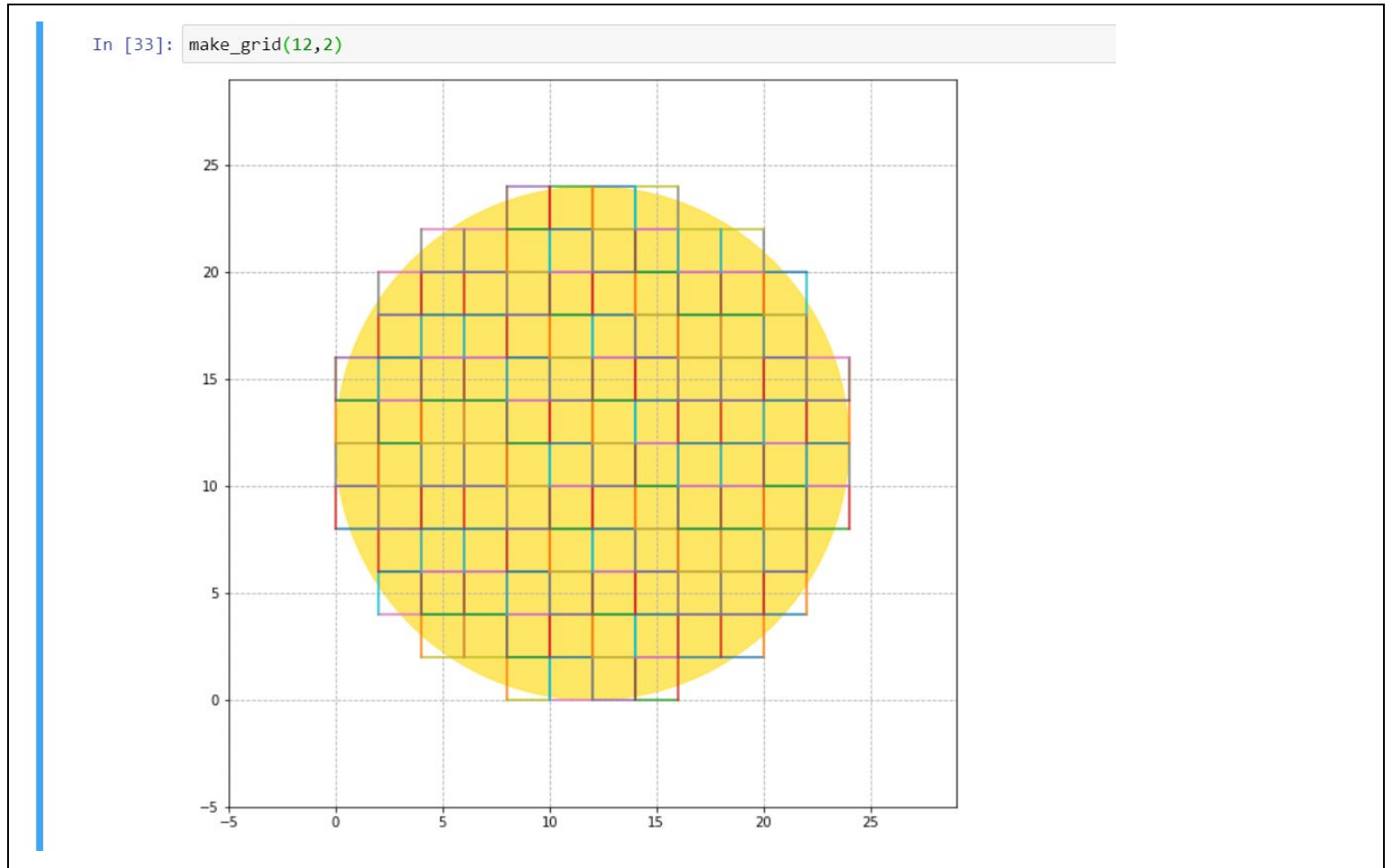
The '**make\_grid(R,S)**' function takes in a radius(R) as an argument and divides the circle and generates a grid of square cells each of side S.



**Short Coming:** Clearly, the API calls made at cells *lying outside the circle* will be *wasted*. Let's add a condition that the *center of the square must lie inside the circle*.

### 4.1.2. Refined Grid:

With the condition described above incorporated, Let's see the output plot of our 'make\_grid' function.



**Result:** Clearly, this is a much better outcome. Without *any significant losses*, we are able to slice *any circle* with a *grid of squares of any side*.

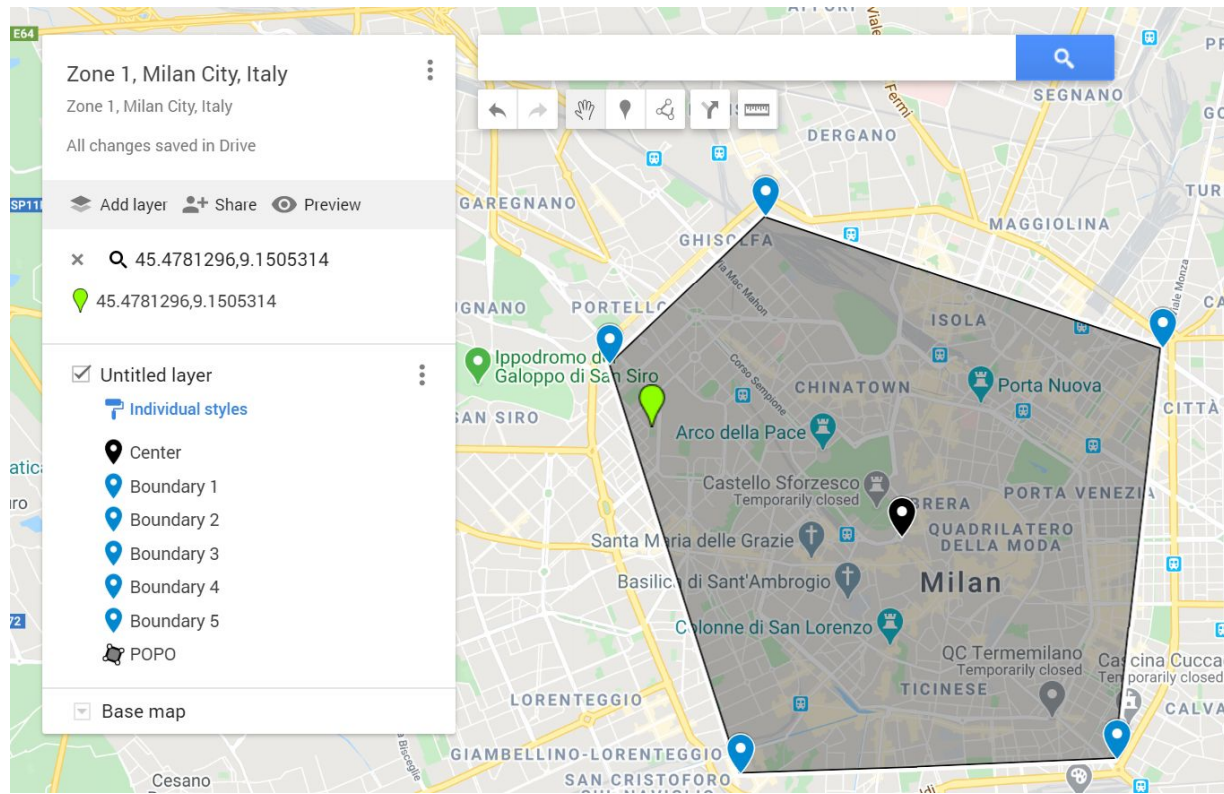
## 4.2. Map Implementation:

Using Cartesian coordinates was straightforward. However, the challenge lies in implementing the same procedure on the *surface of a sphere (Earth)* in a *different coordinate system ([lat,lng])*.

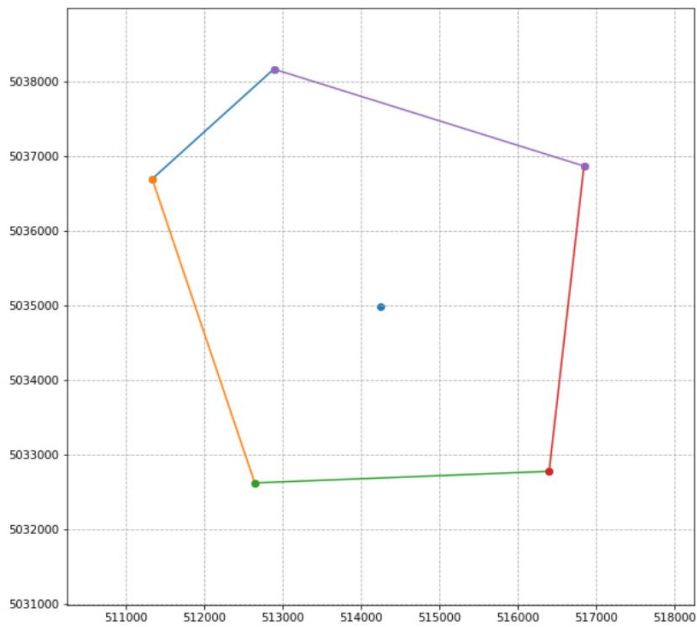
Let's use the *PyProj* package in Python with the *appropriate transformation parameters* to transform latitude and longitudes to x and y coordinates and *vice versa*.

To test, let's visually see both the map and the cartesian transformation of points.





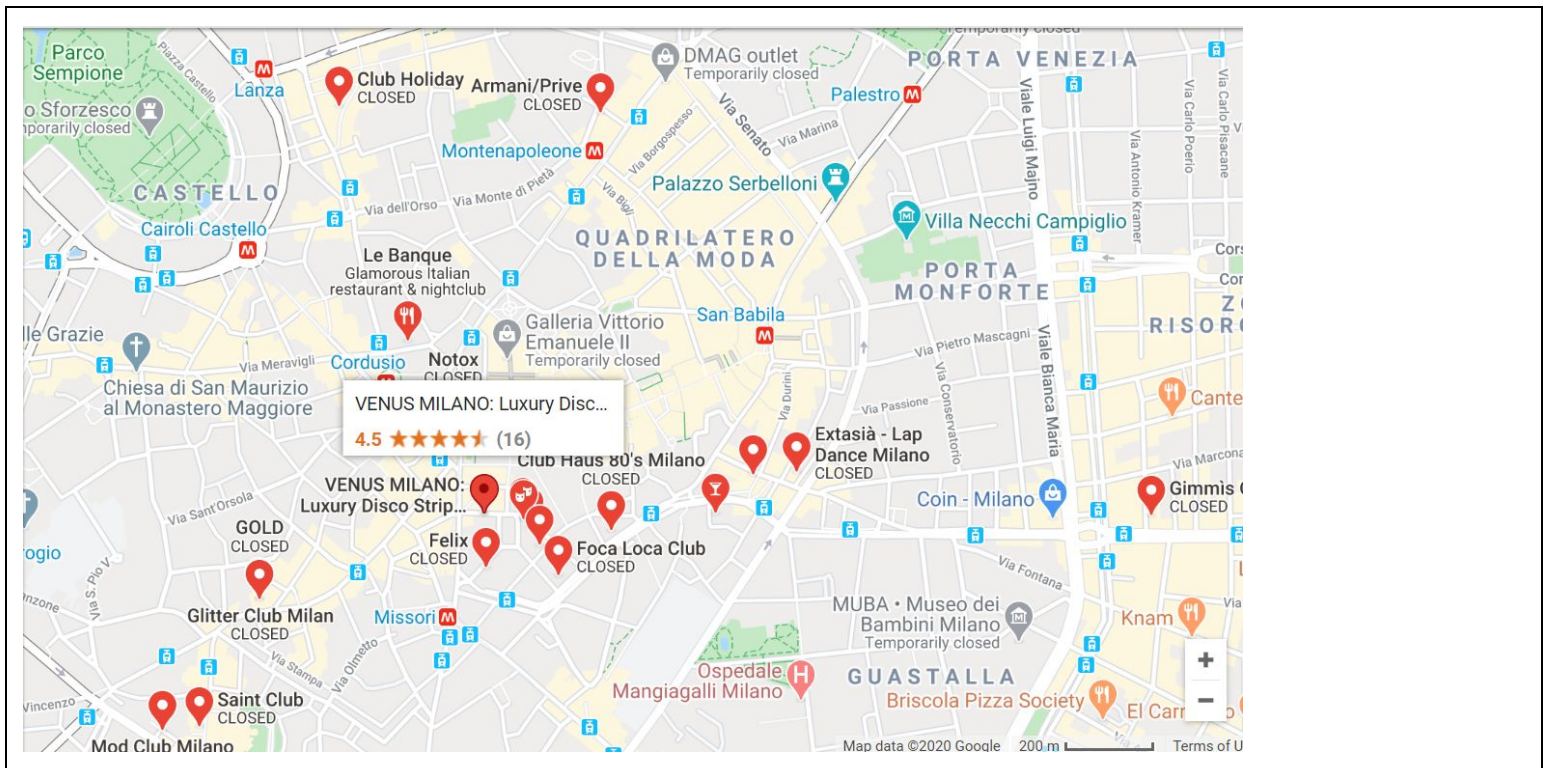
Out[76]: <matplotlib.collections.PathCollection at 0x202bc482688>



**Conclusion:** The transformations were verified and they work flawlessly by producing the intended outcome.

## 5. Finding the critical size:

- The next step in our problem is figuring out the *side length of the square cell* to be used for grid generation.
- The side should be such that when the API is called, it returns *just short of 30 entries*.
- Any decrease in side length would cause lesser than 30 results, and the *result bandwidth* of the API query will be wasted.
- Any more than the critical size, and we *would miss out* on some venues. (eg: 65 venues exist in reality but the API returns only 30 as that is the *cap limit*)
- The *distribution density* of each kind of location mentioned in the 'Relevant Location types' section would *vary* among different regions of our AOI.
- As a solution, we appeal to common sense and select a point where there is a *high density of such venues known beforehand*.
- We iteratively find out the critical size and use that as the length of our square as cell length for the grid.



Clearly, the area near 'VENUS MILANO' has a high density. Let's select this region to find the critical size, SW corner = Missori subway stn and NE corner = San Babila subway stn.

NAME	LOCATION
------	----------

Missori Subway Stn	[45.46058, 9.18819]
San Babila Subway Stn	[45.46642, 9.19767]

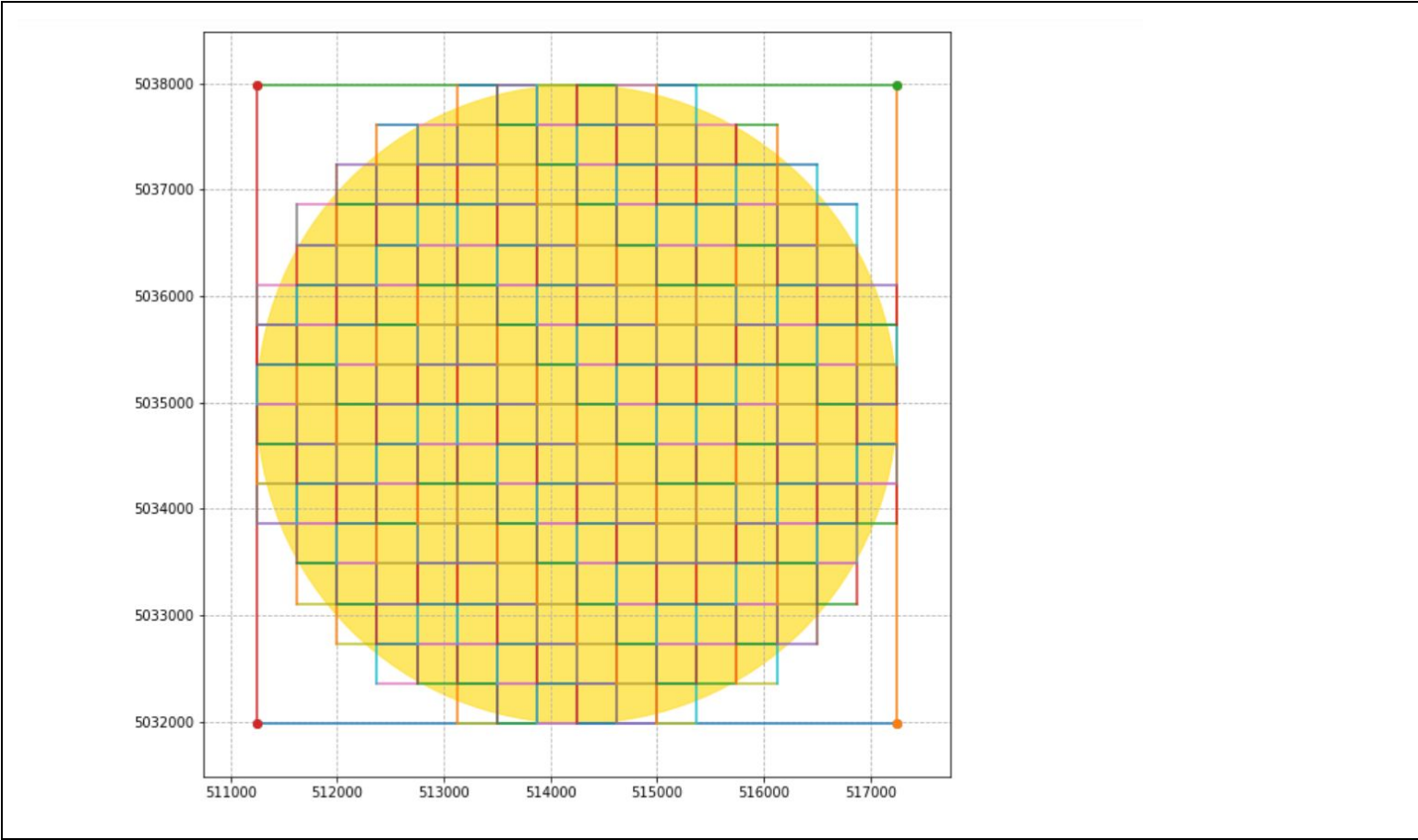
The loop of cell side length was calibrated to run from 700 meters to 100 meters in decrements of 25 meters. The loop was designed to break whenever the length of the results was less than 30 (cap) and the critical side length was found.

**Critical\_side\_length = 375 meters**

## 6. Making the grid on AOI:

Now that the AOI and the critical size has been defined, we must fabricate *a new version of our 'make\_grid(R,S)'* to return the list of *SW corners* and the corresponding *NE corners of each cell* in the grid.

This is one of the most difficult tasks encountered so far.



One can clearly see the 6 Km diameter of the circle. The grid size to AOI scale can be visually seen quite conspicuously.

***The number of cells in the grid = 208 = Number of API calls that will be made***

The list of SouthWest coordinates and NorthEast coordinates of each cell is returned by the function called `'map_grid(epicenter,R,s)'`.

## 7. Converting JSON file to clean data frame:

Once the API has been called and the response result obtained, it has to be converted to a clean looking data frame for further concatenation and classificational analysis.

This job is performed perfectly by the `'venues_to_df()'` function.

This function is to be acted upon every time an API call is made, to get a clean data frame.

## 8. Calling the API on all cells:

The main loop may run into errors, so to be safe, I concatenate the resulting data frames of 20 APIs each and then collate them into a list called `'list_of_df'`.

Let's concatenate all these data frames into one ultimate data frame and write it into a CSV file.

	NAME	CATEGORY	ADDRESS	LAT	LNG	ID
0	Teatro Orione ora Ecoteatro	Theater	Via Fezzan 11	45.459073	9.145781	4d41cf1fbd53f04d480a5115
1	Red's cess	Concert Hall	NaN	45.464245	9.147996	4e3dde9745dd68e327403145
2	Sotterranei Oscuri di Casa Gozzo	Lounge	Via Giuseppe Frua 21/6	45.464475	9.147840	4df3070218a88611c6bb2255
3	Accademia DI Teatroterapia Non Chiamateci Attori	Theater	Via Giovanni Battista Moroni 8	45.465579	9.146489	4dcd458b52b1f8915b78cc68
4	la mia ora di yoga	General Entertainment	NaN	45.468349	9.148692	4f15acfae4b0960d02fe2b1d
5	The Tune	Music Venue	Via Frascati	45.468979	9.145949	4c92423d238c6dcb07e3c655
6	Teatro del buratto	Theater	Via Mosè Bianchi 84	45.475672	9.145051	4ebfddfc0e019a162ac92d67
7	Bea Expo Festival	General Entertainment	Spazio Eventiquattro	45.477550	9.145829	50a63edde4b0451f155359d5
8	The Mule #2	Lounge	Piazza Napoli 30/7	45.452847	9.151729	5831ee2688cfcc313b9871d8
9	Freesound Scuola di Musica	Music Venue	Via Washington 40	45.454724	9.152981	4f5a893fe4b0695cb957ca73

## 9. Creating informative labels:

Categories = [Opera House, General Entertainment, Concert Hall, Comedy Club, Amphitheater, Music Venue, Nightclub, Lounge, Theater] are the unique values in the categories column.

Each of these has to be grouped on the basis of :-

### 9.1. Event time:

For each category, decide a plausible set of times at which an event/gig/performance can take place.

### 9.2. Event Scale:

For each category, decide a plausible set of scales at which an event/gig/performance can take place.

### 9.3. Event Type:

For each category, decide a plausible set of event types that can take place.

#### **Event time:**

{'Theater': 'Day, Evening', 'Concert Hall': 'Day, Evening', 'Lounge': 'Evening, Night', 'General Entertainment': 'Day, Evening, Night, LateNight',  
'Music Venue': 'Day, Evening, Night, LateNight', 'Comedy Club': 'Evening, Night', 'Nightclub': 'Evening, Night',  
'Dance Studio': 'Day, Evening, Night, LateNight',  
'Event Space': 'Day, Evening, Night, LateNight', 'Residential Building (Apartment / Condo)': 'Day, Evening', 'Cocktail Bar': 'Evening, Night', 'Gastropub': 'Evening, Night',  
'Amphitheater': 'Day, Evening', 'Bar': 'Evening, Night', 'Italian Restaurant': 'Day, Evening, Night, LateNight', 'Art Gallery': 'Day, Evening',  
'Strip Club': 'Evening, Night', 'Other Nightlife': 'Evening, Night', 'Road': 'Day, Evening, Night, LateNight', 'Jazz Club': 'Day, Evening, Night, LateNight', 'Office': 'Evening, Night',  
'Miscellaneous Shop': None, 'Radio Station': 'Day, Evening', 'Movie Theater': 'Evening, Night',  
'Electronics Store': None, 'Performing Arts Venue': 'Day, Evening, Night, LateNight', 'Social Club': 'Evening, Night',  
'Art Museum': 'Day, Evening', 'Café': 'Day, Evening, Night, LateNight', 'Building': 'Day, Evening, Night, LateNight',  
'Cultural Center': 'Day, Evening', 'Opera House': 'Day, Evening',  
'Coffee Shop': 'Day, Evening, Night, LateNight', 'Wine Bar': 'Evening, Night', 'Shoe Store': None,  
'Library': None, 'Rock Club': 'Day, Evening, Night, LateNight',  
'College Auditorium': 'Day, Evening, Night, LateNight', 'Modern European Restaurant': 'Day, Evening, Night, LateNight', 'Pub': 'Evening, Night',  
'Coworking Space': 'Evening, Night', 'College Library': None, 'Gay Bar': 'Evening, Night', 'Hotel Bar': 'Evening, Night',  
'Photography Lab': None, 'Hotel': 'Day, Evening, Night, LateNight', 'Piano Bar': 'Evening, Night', 'Indie Theater': 'Evening, Night',  
'General College & University': 'Day, Evening, Night, LateNight', 'Music School': 'Day, Evening, Night, LateNight',  
'Restaurant': 'Day, Evening, Night, LateNight',

'Mediterranean Restaurant': 'Day, Evening, Night, LateNight', 'Diner': 'Day, Evening, Night, LateNight', 'Hookah Bar': 'Evening, Night'}

### **Event Scale:**

{ 'Theater': 'large', 'Concert Hall': 'large', 'Lounge': 'small', 'General Entertainment': 'small',  
'Music Venue': 'small', 'Comedy Club': 'small', 'Nightclub': 'small', 'Dance Studio': 'small',  
'Event Space': 'large', 'Residential Building (Apartment / Condo)': 'small',  
'Piano Bar': 'small', 'Rock Club': 'small', 'Cocktail Bar': 'small', 'Gastropub': 'small',  
'Amphitheater': 'large', 'Bar': 'small', 'Italian Restaurant': 'small', 'Art Gallery': 'small',  
'Strip Club': 'small', 'Other Nightlife': 'small', 'Road': 'None', 'Jazz Club': 'small', 'Office': 'small',  
'Miscellaneous Shop': 'None', 'Radio Station': 'small', 'Movie Theater': 'large',  
'Electronics Store': 'None', 'Performing Arts Venue': 'large', 'Social Club': 'small',  
'Art Museum': 'large', 'Café': 'small', 'Building': 'small', 'Cultural Center': 'large', 'Opera House': 'large',  
'Coffee Shop': 'small', 'Wine Bar': 'small', 'Shoe Store': 'None', 'Library': 'None',  
'College Auditorium': 'large', 'Modern European Restaurant': 'small', 'Pub': 'small',  
'Coworking Space': 'small', 'College Library': 'None', 'Gay Bar': 'small', 'Hotel Bar': 'small',  
'Photography Lab': 'None', 'Hotel': 'small', 'Indie Theater': 'small',  
'General College & University': 'large', 'Music School': 'small', 'Restaurant': 'small',  
'Mediterranean Restaurant': 'small', 'Diner': 'small', 'Hookah Bar': 'small' }

### **Event Type:**

{ 'Theater': 'All', 'Concert Hall': 'All', 'Lounge': 'dance, music, acting, magic,comedy', 'General Entertainment': 'All',  
'Music Venue': 'Musical Theatre, Music', 'Comedy Club': 'Comedy', 'Nightclub': 'dance, music, acting, magic,comedy', 'Dance Studio': 'Musical Theater, Dance, Music',  
'Event Space': 'All', 'Residential Building (Apartment / Condo)': 'All',  
'Piano Bar': 'Musical Theater, Dance, Music, Magic', 'Rock Club': 'Musical Theatre, Music', 'Cocktail Bar': 'dance, music, acting, magic,comedy', 'Gastropub': 'dance, music, acting, magic,comedy',  
'Amphitheater': 'All', 'Bar': 'dance, music, acting, magic,comedy', 'Italian Restaurant': 'dance, music, acting, magic,comedy', 'Art Gallery': 'dance, music, acting, magic,comedy',  
'Strip Club': 'dance, music, acting, magic,comedy', 'Other Nightlife': 'dance, music, acting, magic,comedy',  
'Road': 'All', 'Jazz Club': 'Musical Theatre, Music', 'Office': 'dance, music, acting, magic,comedy',  
'Miscellaneous Shop': 'None', 'Radio Station': 'dance, music, acting, magic,comedy', 'Movie Theater': 'dance, music, acting, magic,comedy',  
'Electronics Store': 'None', 'Performing Arts Venue': 'All', 'Social Club': 'All',  
'Art Museum': 'dance, music, acting, magic,comedy', 'Café': 'dance, music, acting, magic,comedy',  
'Building': 'dance, music, acting, magic,comedy', 'Cultural Center': 'All', 'Opera House': 'All',  
'Coffee Shop': 'dance, music, acting, magic,comedy', 'Wine Bar': 'dance, music, acting, magic,comedy', 'Shoe Store': 'None', 'Library': 'None',  
'College Auditorium': 'All', 'Modern European Restaurant': 'dance, music, acting, magic,comedy', 'Pub': 'dance, music, acting, magic,comedy',  
'Coworking Space': 'dance, music, acting, magic,comedy', 'College Library': 'None', 'Gay Bar': 'dance, music, acting, magic,comedy', 'Hotel Bar': 'dance, music, acting, magic,comedy',  
'Photography Lab': 'None', 'Hotel': 'dance, music, acting, magic,comedy', 'Indie Theater': 'dance, music, acting, magic,comedy',  
'General College & University': 'All', 'Music School': 'Musical Theatre, Music', 'Restaurant': 'dance, music, acting, magic,comedy',  
'Mediterranean Restaurant': 'dance, music, acting, magic,comedy', 'Diner': 'dance, music, acting, magic,comedy',



```
'Hookah Bar': 'dance, music, acting, magic, comedy'}
```

## 10. After Grouping:

After the event type, scale, and time have been decided, the final dataset looks like:

Out[142]:

	NAME	CATEGORY	ADDRESS	LAT	LNG	ID	SCALE	EVENT	TIME_OF_DAY
0	Teatro Orione ora Ecoteatro	Theater	Via Fezzan 11	45.459073	9.145781	4d41cf1fd53f04d480a5115	large	All	Day, Evening
1	Red's ccess	Concert Hall	NaN	45.464245	9.147996	4e3dde9745dd68e327403145	large	All	Day, Evening
2	Sotterranei Oscuri di Casa Gozzo	Lounge	Via Giuseppe Frua 21/6	45.464475	9.147840	4df3070218a88611c6bb2255	small	dance, music, acting, magic, comedy	Evening, Night
3	Accademia Di Teatrotterapia Non Chiamateci Attori	Theater	Via Giovanni Battista Moroni 8	45.465579	9.146489	4dcd458b52b1f8915b78cc68	large	All	Day, Evening
4	la mia ora di yoga	General Entertainment	NaN	45.468349	9.148692	4f15acfae4b0960d02fe2b1d	small	All	Day, Evening, Night, LateNight
5	The Tune	Music Venue	Via Frascati	45.468979	9.145949	4c92423d238c6dcb07e3c655	small	Musical Theatre, Music	Day, Evening, Night, LateNight
6	Teatro del buratto	Theater	Via Mosè Bianchi 84	45.475672	9.145051	4ebfdffc0e019a162ac92d67	large	All	Day, Evening
7	Bea Expo Festival	General Entertainment	Spazio Eventiquattro	45.477550	9.145829	50a63edde4b0451f155359d5	small	All	Day, Evening, Night, LateNight
8	The Mule #2	Lounge	Piazza Napoli 30/7	45.452847	9.151729	5831ee2688cfcc313b9871d8	small	dance, music, acting, magic, comedy	Evening, Night
9	Freesound Scuola di Musica	Music Venue	Via Washington 40	45.454724	9.152981	4f5a893fe4b0695cb957ca73	small	Musical Theatre, Music	Day, Evening, Night, LateNight

## 11. Conclusion:

Thus, the required comprehensive dataset has been built to assist the agency and all of our target audiences in making choices every day about the venue on the basis of performance type, time, and scale in order to increase their reach, spread their art and at the same time increasing ticket sales.