Business Problem

I live in Paris, FR. I will be moving to Sydney, AU in the coming months.
I love the neighborhood I'm living in Paris. It's named Montmartre. It's a really convenient place with a lot of cafés and restaurants. I really like to find a similar surrounding Sydney.

Due to the COVID lockdown, we cannot wonder too much around, but I still need to find a place.
Also, I just Brock my leg and I will not be able to walk too much for at least one year.

So I need to use data science to narrow my apartment research. I'd like to know the areas in Sydney similars to the Paris I live in now.

I'd like to target a few areas to visits and see the venues around to be able to have a good idea where to get my apartment.

Data and data sources

For the data source, I will only use data from Foursquare. It does give enough information to pick up the right area. Also, the data is widely used, reliable, and updated frequently. It's the perfect data source to do so.

Methodology

I will get the data for Montmartre, the Paris neighbourhood.
Then I will get the dataset for Sydney. In order to overcome the Foursquare 100 venues limit, I preselect 4 areas because it's near by the water (bay or ocean). I'd like to move near the water.
I will look for venues in those areas with a wide radius.

Get the data

In [33]:
```
#Get the data for Montmatre, a parisian neighborhood
MontmartreLat=48.8857636427167
MontmartreLng=2.3349730916262725
DataFromParis = GetVenuesNearByInDataFrame(MontmartreLat,MontmartreLng,111,300)
```


```
k_', 'suffix': '.png'}, 'primary': True}]
[{'id': '4bf58dd8d48988d1e5931735', 'name': 'Music Venue', 'pluralName': 'Music Venues', 'shortName': 'Mus
v2/arts_entertainment/musicvenue_', 'suffix': '.png'}, 'primary': True}]
[{'id': '4bf58dd8d48988d16a941735', 'name': 'Bakery', 'pluralName': 'Bakeries', 'shortName': 'Bakery', 'ic
_', 'suffix': '.png'}, 'primary': True}]
[{'id': '4bf58dd8d48988d110941735', 'name': 'Italian Restaurant', 'pluralName': 'Italian Restaurants', 'sh
ategories_v2/food/italian_', 'suffix': '.png'}, 'primary': True}]
[{'id': '4bf58dd8d48988d10c941735', 'name': 'French Restaurant', 'pluralName': 'French Restaurants', 'short
gories_v2/food/french_', 'suffix': '.png'}, 'primary': True}]
[{'id': '4bf58dd8d48988d164941735', 'name': 'Plaza', 'pluralName': 'Plazas', 'shortName': 'Plaza', 'icon':
plaza_', 'suffix': '.png'}, 'primary': True}]
[{'id': '4bf58dd8d48988d110941735', 'name': 'Italian Restaurant', 'pluralName': 'Italian Restaurants', 'sh
ategories_v2/food/italian_', 'suffix': '.png'}, 'primary': True}]
[{'id': '4bf58dd8d48988d10c941735', 'name': 'French Restaurant', 'pluralName': 'French Restaurants', 'short
gories_v2/food/french_', 'suffix': '.png'}, 'primary': True}]
[{'id': '52e81612bcbc57f1066b79f1', 'name': 'Bistro', 'pluralName': 'Bistros', 'shortName': 'Bistro', 'ico
```


I will prepare the dataset to fit a K-means model. It's the most appropriate model for Clustering.
In order to do so, I'm cleaning up the data set, transform the venue categories to numerical data, normalise it.

0	Boulangerie Alexine	48.886141	2.334477	Paris	75018	0	0	0	0	0	...	0	0
1	Guilo Guilo	48.885942	2.337048	Paris	75018	0	0	0	0	0	...	0	0
2	Terrass'' Hotel	48.886532	2.333159	Paris	75018	0	0	0	0	0	...	0	0
3	Al Caratello	48.885248	2.336002	Paris	75018	0	0	0	0	0	...	0	0
4	Le Grenier à Pain	48.885283	2.336718	Paris	75018	0	0	0	0	0	...	0	0

5 rows × 116 columns

[164]:
```
#Group the data by PostalCode
city_grouped = Data_hot.groupby(['postalCode']).mean().reset_index()
city_grouped.head()
```

Out[164]:

	postalCode	lat	lng	Antique Shop	Argentinian Restaurant	Art Gallery	Australian Restaurant	BBQ Joint	Bagel Shop	Bakery	...	Tennis Court	Thai Restaurant	Tl
0	2008	-33.886771	151.200181	0.0	0.00	0.062500	0.062500	0.0625	0.0	0.062500	...	0.0	0.00	
1	2010	-33.888539	151.209600	0.0	0.04	0.000000	0.000000	0.0400	0.0	0.000000	...	0.0	0.04	
2	2014	-33.891864	151.200566	0.0	0.00	0.000000	0.000000	0.0000	0.0	0.000000	...	0.0	0.00	
3	2015	-33.897807	151.198286	0.0	0.00	0.000000	0.000000	0.0000	0.0	0.000000	...	0.0	0.25	
4	2016	-33.892259	151.203755	0.0	0.00	0.026316	0.026316	0.0000	0.0	0.052632	...	0.0	0.00	

5 rows × 114 columns

Then I will fit the data into a K-means model. I took me couple of K setting to find an appropriate cluster. By appropriate, I mean not too many areas (remember I need to walk around with cruches) , not too few, I still want to have couple of choices.

```
k=6
k_means = KMeans(init="k-means++", n_clusters=k, n_init=21)
k_means.fit(X)
labels = k_means.labels_
k_means
```

]: KMeans(n_clusters=6, n_init=21)

```
labels
city_grouped["Label"] = labels
city_grouped.head()
```

]:

Then I select the Sydney neighbourhood in the same cluster than Montmartre.

```
#Get the Sydney Neighborhood similar to Montmatre  (Sydney postal codes are 4 digits. Paris = 5 digits)
SydneyAreasLikeParis = city_grouped.loc[(city_grouped['postalCode'].astype('int64') <=10000) & (city_grouped['Label'] ==MontmatreCluster[0])  ]
AllSydney = city_grouped.loc[(city_grouped['postalCode'].astype('int64') <=10000) ]

SydneyAreasLikeParis.head()
```

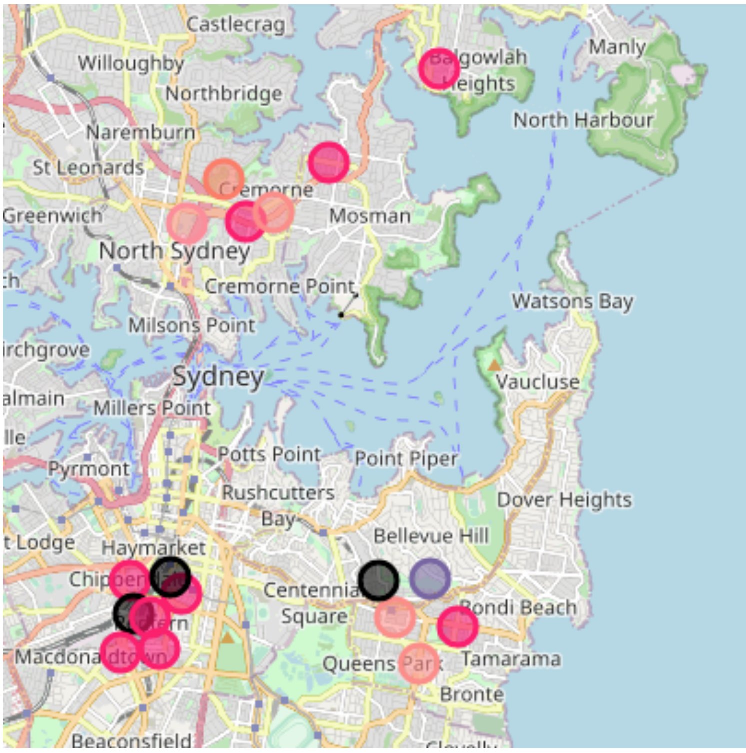
0]:

	postalCode	lat	lng	Antique Shop	Argentinian Restaurant	Art Gallery	Australian Restaurant	BBQ Joint	Bagel Shop	Bakery	...	Theater	Turkish Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Whisky Bar
6	2022	-33.892258	151.248545	0.0	0.0	0.0	0.0	0.0	0.0	0.020833	...	0.0	0.0	0.020833	0.0	0.0
8	2024	-33.899519	151.253908	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.000000	0.0	0.0
16	2090	-33.829548	151.228729	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.000000	0.0	0.0

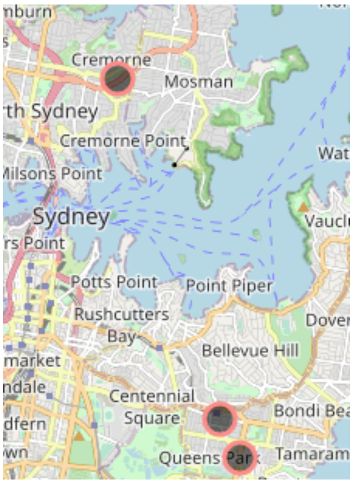
3 rows × 116 columns

I will show 3 maps:

- 1 map with the result of all the Sydney areas. We can see all the areas considered.



- 1 map showing the areas “like Montmartre”.We can see all the areas selected to be “like Montmartre”



- 1 map with all the venues in the areas “like Montmartre”. I can already have a feeling of what is near by: interest points, bakeries, restaurants…

Results

3 areas are selected with a wide variety of venues and interests. It does make my life a way easier to move.

Discussion

Even if with a small amount of data (500 points), the analysis is relevant.
I probably need to refine the analysis and pick up the venues categories I like the most in order to weight it according my tastes.

Conclusion

I'm pretty impress by the exercice. It's actually a true story and I end up to move in the Cremone area. This area was pick up by the system. Spotted on!

References

https://developer.foursquare.com/