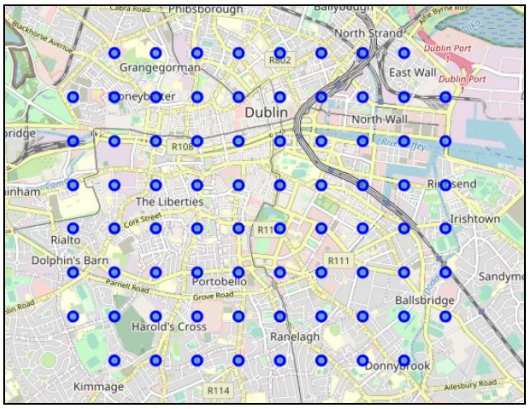
DATA

Firstly we will create a mesh of points on top of Dublin city due to asymmetrical boroughs shape, so we can use each point in order to explore the area leveraging the Foursquare API. Although we will get extra duplicates, these are easy to handle using pandas on python. The advantage is the lack of missing venues if we were using the API in the conventional way.



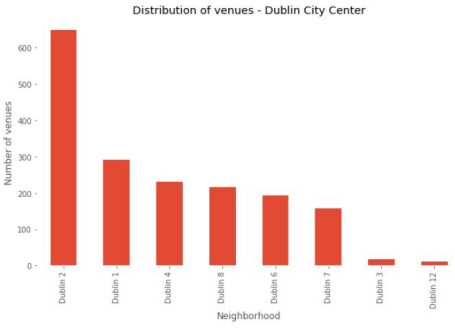
Mesh with 76 points and coordinates, district and postcode information for each single one.

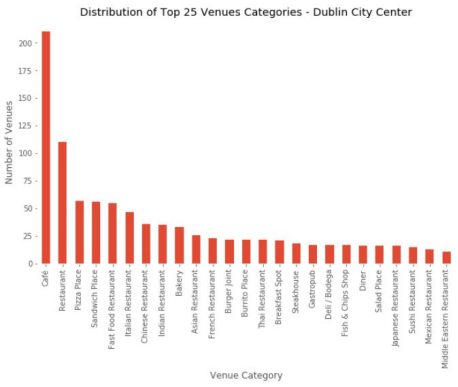
	Latitude	Longitude	Coordinates	District	PostCode
0	53.322556	-6.287982	53.322556, -6.287982	Kimmage C ED	Dublin 6
1	53.322556	-6.280417	53.322556, -6.280417	Kimmage C ED	Dublin 6
2	53.322556	-6.272853	53.322556, -6.272853	Rathmines West F ED	Dublin 6
3	53.322556	-6.265288	53.322556, -6.265288	Rathmines and Rathgar West ED	Dublin 6
4	53.322556	-6.257723	53.322556, -6.257723	Rathmines and Rathgar West ED	Dublin 6

- ***Research Dublin City using Foursquare API***
  First I used the explore API endpoint in order to get venues around each point of the mesh. After that, I repeated it with a specific category parameter, like food, hotel, etc.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Dublin 12	53.322556	-6.291964	Pickles Deli	53.321021	-6.293015	Deli / Bodega
1	Dublin 12	53.322556	-6.291964	Monto Cafe	53.320752	-6.292583	Diner
2	Dublin 12	53.322556	-6.291964	Apache Pizza	53.322971	-6.295267	Pizza Place
3	Dublin 12	53.322556	-6.291964	Matt The Rashers	53.320459	-6.291814	Café
4	Dublin 6	53.322556	-6.281214	Craft Restaurant	53.322698	-6.279316	Restaurant

Neighborhood	Venue
Dublin 2	649
Dublin 1	291
Dublin 4	230
Dublin 8	216
Dublin 6	194
Dublin 7	158
Dublin 3	18
Dublin 12	12





- ***Research Dublin City using Google Places API***
  Although Foursquare gave us quite a bit of information, due to Foursquare limitations and also because in a Data Science project like this the data to work with is crucial in order to get satisfactory results, I used Google API extending the datasets we already had obtained.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue id	Venue price level	Venue Rating	Total user ratings	Venue Category
Dublin 3	53.356207	-6.242594	Dublin	53.349805	-6.260310	71ccb3b82cc1f534c1da9998f8610d85a6d7c598	0	0.0	0	[locality, political]
Dublin 3	53.356207	-6.242594	Jurys Inn Christchurch Dublin	53.342832	-6.270520	2ccbdecd28694cc8b74bfe56279db7fb52187101	0	4.2	1320	[lodging, restaurant, food, point_of_interest,...
Dublin 3	53.356207	-6.242594	Kinlay House Dublin	53.343726	-6.269898	7511e298b05c19078d13d98bb64909a1e62549f7	0	3.9	1110	[lodging, point_of_interest, establishment]
Dublin 3	53.356207	-6.242594	Radisson Blu Royal Hotel, Dublin	53.340853	-6.268325	5245b21949a3fad5785f03ff1427446a7339d412	0	4.3	1363	[lodging, point_of_interest, establishment]

- ***Web scraping Restaurant ranking list and details***

First we get the list of the ranking of all restaurants in Dublin City from TA website.

Name	Ranking	Link
Tang Cafe	1	/Restaurant_Review-g186605-d10387074-Reviews-T...
Taza	2	/Restaurant_Review-g186605-d17181958-Reviews-T...
Glovers Alley	3	/Restaurant_Review-g186605-d13477650-Reviews-G...

Once the list is retrieved I wrote a different script for web scraping the details of each single restaurant creating a dataframe with all the information. We also use Geopy libraries in order to get its coordinates. For the restaurant that we could not get this information we will use Google API.

	Name	Ranking	Reviews	Rating	Price	Cuisines	Address	Phone	Link	Geopy Latitude	Geopy Longitude
0	Tang Cafe	1	502	5.0	Cheap	['Cafe', 'European', 'Healthy']	23c Dawson Street, Dublin D02 PW18 Ireland	+353 86 391 5401	/Restaurant_Review-g186605-d10387074-Reviews-T...	53.348720	-6.258399
1	Taza	2	178	5.0	Average	['Indian', 'Asian', 'Pakistani']	2 Ardcollum Avenue Artane, Dublin D05 XW88 Ire...	+353 1 558 2866	/Restaurant_Review-g186605-d17181958-Reviews-T...	NaN	NaN
2	Glovers Alley	3	193	5.0	Expensive	['Irish', 'European', 'Vegetarian Friendly']	128 Stephen's Green Fitzwilliam Hotel, Dublin ...	+353 1 244 0733	/Restaurant_Review-g186605-d13477650-Reviews-G...	53.339644	-6.263466

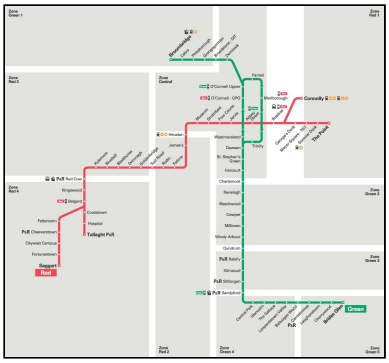
After that, we just remove restaurants outside the project scope boundaries based on latitude and longitude coordinates.

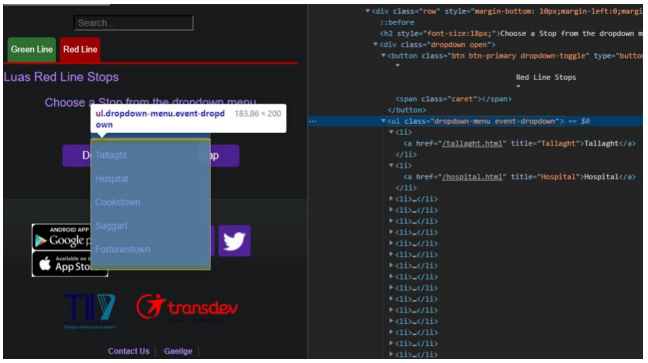
We'll compare the correlation between Rating and Ranking in order to filter some samples that have outliers, like high Rating but low Ranking due to inconsistent data from the website from where we had collected the data . After that, we'll define Rating as the KPI to measure the performance of a business, therefore, as the target with the predictive models we'll use in the project.

After cleaning the restaurant's dataset, we obtained 1188 samples.

- **Feature set:**
  - **Luas stops (city tram)**

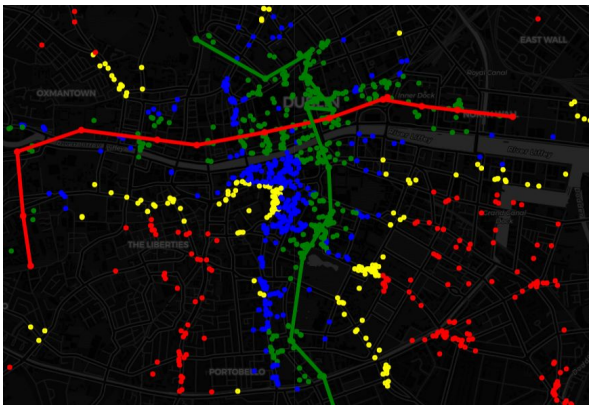
Data has been collected by web scraping the official website.





	Name	Address	Latitude	Longitude
0	Tallaght Luas stop	Oldbawn, Dublin, Ireland	53.287367	-6.374592
1	Saggart Luas stop	Saggart Luas Stop, Fortunestown, Saggart, Co. ...	53.284641	-6.437762
2	Fortunestown Luas stop	Fortunestown, Tallaght, Dublin, Ireland	53.284210	-6.424610
3	Citywest Campus Luas stop	Citywest Campus Luas Stop, Cooldown Commons, D...	53.287800	-6.418820
4	Cheeverstown Luas stop	Cheeverstown Luas Stop, Tallaght, Dublin, Ireland	53.291036	-6.406877

Once data had been collected, we calculated the distance between each restaurant to the closest Luas stop, fixing the minimum distance as 100 meter. This means, a restaurant will get the same weight value once it has a Luas stop in a range of 100 meters.



	Latitude	Longitude	Distances min
count	1188.000000	1188.000000	1188.000000
mean	53.341764	-6.260744	396.907407
std	0.007607	0.012098	347.031231
min	53.322580	-6.295349	100.000000
25%	53.337302	-6.265706	155.000000
50%	53.343006	-6.262388	292.500000
75%	53.347322	-6.255040	476.500000
max	53.356205	-6.227568	1834.000000

Statistics of all our samples versus Luas and closest restaurants in green shown in the map.

  - **Liffey (river)**

After choosing 5 points which describe the Liffey trajectory, we divided each segment using python in order to get extra points to calculate distances between each restaurant and the closest part of the river.

Once we calculated this data we normalize it and save it for later use. Similar to the previous feature, we fixed a minimum distance of 200 meters.

	Name	Ranking	Reviews	Rating	Price	Cuisines	Address	Latitude	Longitude	Phone	Link	Distance	Distance Proximity
59	Umi Falafel	73	1008	4.5	Cheap	['Lebanese', 'Fast food', 'Mediterranean']	Callaghan House 13 Dame Street, Dublin D02HX67...	53.324603	-6.265240	+353 1 670 6866	/Restaurant_Review-g186605-d5003310-Reviews-Um...	2344.58	Farest
229	The Brewer's Dining Hall - Guinness Storehouse	267	494	4.5	Average	['Irish', 'European', 'Vegetarian Friendly']	St. James Gate Saint James's Gate, Dublin Dubl...	53.344112	-6.285172	+353 1 408 4800	/Restaurant_Review-g186605-d1554254-Reviews-Th...	284.904	Farest

Statistics of all our samples versus Liffey and closest restaurants in green shown in the map.

	Ranking	Reviews	Rating	Latitude	Longitude	Distance
count	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000
mean	867.664141	277.626263	4.141835	53.341764	-6.260744	762.503896
std	583.001783	555.012944	0.568664	0.007607	0.012098	648.587962
min	1.000000	1.000000	1.000000	53.322580	-6.295349	200.000000
25%	353.750000	12.000000	4.000000	53.337302	-6.265706	252.670755
50%	782.500000	71.000000	4.000000	53.343006	-6.262388	514.833512
75%	1353.750000	299.250000	4.500000	53.347322	-6.255040	1057.797971
max	2045.000000	6164.000000	5.000000	53.356205	-6.227568	2741.908314

- **Companies**

A csv with a list of companies was obtained from IDA Ireland.

	Company	Sector(s)	Product	Nationality	Address
0	Symantec Limited	ICT- ICT - Software	Package Software for PC and Mac and technical ...	United States of America	Ballycoolin Industrial Park Blanchardstown Du...
1	Ig International Management Limited	Financial Services	Asset Management - owned by Power Financial Gr...	Canada	Brooklawn House Shelbourne Road Ballsbridge Du...

After filtering the data for businesses in Dublin, we used Geopy to get the coordinates, and alternatively Google API for those company names or addresses that Geopy could not resolve.

	Company	Address	Latitude	Longitude
0	Symantec Limited	Ballycoolen, Dublin 15, Ireland	53.412959	-6.373212
1	Ig International Management Limited	Brooklawn House, Shelbourne Rd, Ballsbridge, D...	53.332099	-6.230495

- **Several features in range:**

We are taking into account 4 different ranges for features such as **Pubs**, **Tourism**, **Cafes**, **Universities**, **Accommodations**, **Museums** and **Restaurants**. We calculated the amount of each category within that range versus every single restaurant in our sample data.

- < 250 meters
- 250 to 500 meters
- 500 to 1000 meters
- 1000 to 2000 meters

Once calculated the number of venues per range we multiply each one for a fixed number, which is proportional to the proximity to the restaurant. For instance, within the range 250 meters, it'd be 2000/250, and so on.

Link	Hotel weight	Tourism weight	University weight	Cafe weight	Pub weight	Museum weight	Company weight	Restaurant weight
/Restaurant_Review-g186605-d953880-Reviews-The...	381.0	120.0	35.0	394.0	1028.0	93.0	353.0	2711.0

This is the describe method giving us detailed information about the dataframe:

	Ranking	Reviews	Rating	Latitude	Longitude	Hotel weight	Tourism weight	University weight	Cafe weight	Pub weight	Museum weight
count	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000
mean	867.664141	277.626263	4.141835	53.341764	-6.260744	258.089226	89.207071	27.739899	277.725589	652.114478	67.496633
std	583.001783	555.012944	0.568664	0.007607	0.012098	101.944971	39.773485	13.983280	103.940762	302.811160	27.388912
min	1.000000	1.000000	1.000000	53.322580	-6.295349	11.000000	0.000000	0.000000	20.000000	21.000000	1.000000
25%	353.750000	12.000000	4.000000	53.337302	-6.265706	181.000000	60.000000	18.000000	197.000000	405.000000	47.000000
50%	782.500000	71.000000	4.000000	53.343006	-6.262388	274.500000	93.000000	27.000000	292.500000	646.500000	72.000000
75%	1353.750000	299.250000	4.500000	53.347322	-6.255040	345.250000	125.000000	39.000000	383.000000	904.000000	93.000000
max	2045.000000	6164.000000	5.000000	53.356205	-6.227568	410.000000	148.000000	66.000000	424.000000	1204.000000	111.000000

And we normalize the dataframe using scikit learn library *preprocessing.StandardScaler* in order to use them later.

	Ranking	Reviews	Rating	Latitude	Longitude	Hotel weight NOR	Tourism weight NOR	University weight NOR	Cafe weight NOR	Pub weight NOR	Museum weight NOR
count	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000	1188.000000
mean	867.664141	277.626263	4.141835	53.341764	-6.260744	2.532719	2.243822	1.984626	2.673086	2.154442	2.465416
std	583.001783	555.012944	0.568664	0.007607	0.012098	1.000421	1.000421	1.000421	1.000421	1.000421	1.000421
min	1.000000	1.000000	1.000000	53.322580	-6.295349	0.107947	0.000000	0.000000	0.192498	0.069379	0.036527
25%	353.750000	12.000000	4.000000	53.337302	-6.265706	1.776215	1.509178	1.287794	1.896109	1.338030	1.716746
50%	782.500000	71.000000	4.000000	53.343006	-6.262388	2.693763	2.339226	1.931691	2.815288	2.135893	2.629908
75%	1353.750000	299.250000	4.500000	53.347322	-6.255040	3.388057	3.144121	2.790220	3.686343	2.986616	3.396965
max	2045.000000	6164.000000	5.000000	53.356205	-6.227568	4.023471	3.722639	4.721910	4.080965	3.977750	4.054442

- **Price Area**

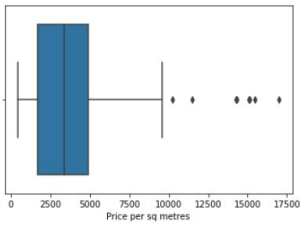
By web scraping the leader website Daft.ie where Estate Agents publish their ads we collected all the ads for commercials in Dublin.

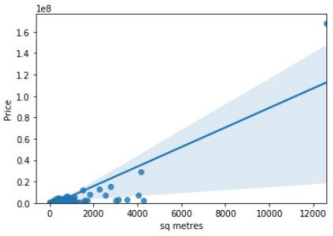
	Name	Commercial Type	Price	sq metres	Neighbourhood
0	11/11A Ormond Quay Lower, Dublin 1	Office To Let or For Sale	1300000	465	Dublin 1
1	21 Ormond Quay Upper, Dublin 1	Restaurant / Bar / Hotel For Sale	2250000	570	Dublin 1
2	32 Lower Ormond Quay, Dublin 1, Dublin 1	Investment Property For Sale	700000	314	Dublin 1
3	308 The Capel Building, Mary Street, Dublin 1	Office For Sale	180000	NaN	Dublin 1
4	Independent House, Talbot Street, Dublin 1	Investment Property For Sale	29000000	NaN	Dublin 1

Data had to be cleaned and corrected due to some advertisements had no price, or some had wrong information published on the website.

	Price	sq metres	Price per sq metres
count	1.730000e+02	173.000000	173.000000
mean	1.373641e+06	585.901734	5624.466243
std	2.076866e+06	1206.434887	26487.273172
min	1.000000e+00	4.000000	0.010000
25%	3.950000e+05	154.000000	1325.300000
50%	6.500000e+05	254.000000	3232.760000
75%	1.400000e+06	546.000000	4828.800000
max	1.500000e+07	13152.000000	350000.000000

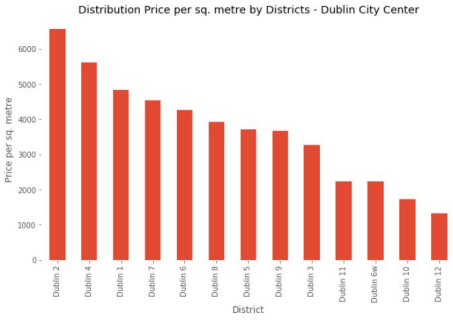
After visualizing the data we could identify some outliers and remove them from the dataframe.





	Name	Commercial Type	Price	sq metres	Neighbourhood	Price per sq metres
0	Cleary's Bar, Lounge & Adjacent Barbers Shop, ...	Restaurant / Bar / Hotel For Sale	1500000.0	301.0	Dublin 1	4983.39
1	21 Ormond Quay Upper, Dublin 1	Restaurant / Bar / Hotel For Sale	2250000.0	570.0	Dublin 1	3947.37

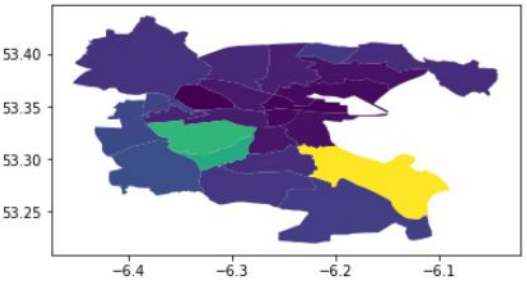
We calculated the means for each district and visualized on a heatmap.



- **Districts Geoshapes**

Dublin Postcodes from [GeoCluster]repository. Converted to shapefile in order to import it with geopandas

```
   id                                           geometry
0   0  POLYGON ((-6.33272 53.37071, -6.33274 53.37072...
1   1  POLYGON ((-6.25925 53.36170, -6.24931 53.35908...
2   2  POLYGON ((-6.26780 53.34588, -6.25454 53.34802...
3   3  POLYGON ((-6.21911 53.35739, -6.22102 53.35491...
4   4  POLYGON ((-6.23737 53.34359, -6.23052 53.34304...
```



Getting the list of restaurants and transforming the dataframe in a geodataframe

	Link	Latitude	Longitude	geometry
0	/Restaurant_Review-g186605-d10387074-Reviews-T...	53.34872	-6.258399	POINT (-6.25840 53.34872)

	Link	District
0	/Restaurant_Review-g186605-d10387074-Reviews-T...	1.0
1	/Restaurant_Review-g186605-d13477650-Reviews-G...	2.0
2	/Restaurant_Review-g186605-d6403998-Reviews-Da...	8.0
3	/Restaurant_Review-g186605-d2239110-Reviews-Mu...	4.0
4	/Restaurant_Review-g186605-d15590976-Reviews-T...	8.0

After we just merge the geodataframes to get district data into our main dataframe. Due to many samples had missing district information, or were wrong, we used the coordinates to tackle this issue.

	Name	Ranking	Reviews	Rating	Price	Cuisines	Address	Latitude	Longitude	Phone	Link	District
0	Tang Cafe	1	502	5.0	Cheap	['Cafe', 'European', 'Healthy']	23c Dawson Street, Dublin D02 PW18 Ireland	53.348720	-6.258399	+353 86 391 5401	/Restaurant_Review-g186605-d10387074-Reviews-T...	1.0
1	Glovers Alley	3	193	5.0	Expensive	['Irish', 'European', 'Vegetarian Friendly']	128 Stephen's Green Fitzwilliam Hotel, Dublin ...	53.339644	-6.263466	+353 1 244 0733	/Restaurant_Review-g186605-d13477650-Reviews-G...	2.0
2	Darkey Kelly's Bar & Restaurant	4	1645	4.5	Average	['Irish', 'Bar', 'European']	Fishamble Street ChristChruch, Dublin Ireland	53.343513	-6.271060	+353 83 346 4682	/Restaurant_Review-g186605-d6403998-Reviews-Da...	8.0
3	Mulberry Garden	5	909	4.5	Expensive	['Irish', 'European', 'Contemporary']	Mulberry Lane Donnybrook, Dublin 04 Ireland	53.322659	-6.236801	+353 1 269 3300	/Restaurant_Review-g186605-d2239110-Reviews-Mu...	4.0
4	The Landmark	6	468	4.5	Average	['Irish', 'Bar', 'European']	The Landmark 40 Wexford Street, Dublin D02 CH6...	53.337441	-6.265903	+353 1 537 9951	/Restaurant_Review-g186605-d15590976-Reviews-T...	8.0

This map with interactive layers can be found within the repository.