Statistical Modelling with Python

Zuwa Ojefua
August, 2023

The problem

Statement	Context
Estimating the available number of bikes at a given bike station in a city, Dublin.	Based on characteristics of Places of Interest (POIs) such restaurants and bars within 1000m of the bike station

Process Outline

Step 1

CityBikes API

To obtain Bike station information including:

- Latitude
- Longitude
- Number of bikes.

Step 2

Foursquare and Yelp APIs

Using coordinates obtained from step 1, extract features of POIs within 1000m radius of each bike station.

Step 3

Joining Data and EDA

- Creation of new DataFrame
- Use of visualization to explore the data
- Storage of data in SQLite database.

Process Outline

Step 4

Model Building

To demonstrate relationship between number of bikes at a station and characteristics of POIs around the station.

Step 5

Interpretation of results

Derivation of insights from the model.

CityBikes results

	Station Name	Latitude	Longitude	Number of Bikes
0	GEORGES LANE	53.350230	-6.279696	17
1	NORTH CIRCULAR ROAD (O'CONNELL'S)	53.357841	-6.251557	28
2	PHIBSBOROUGH ROAD	53.356307	-6.273717	2
3	SMITHFIELD NORTH	53.349562	-6.278198	23
4	CLONMEL STREET	53.336021	-6.262980	3
...	...	...	...	...
109	JAMES STREET EAST	53.336597	-6.248109	0
110	PARNELL SQUARE NORTH	53.353742	-6.265301	1

Yelp results

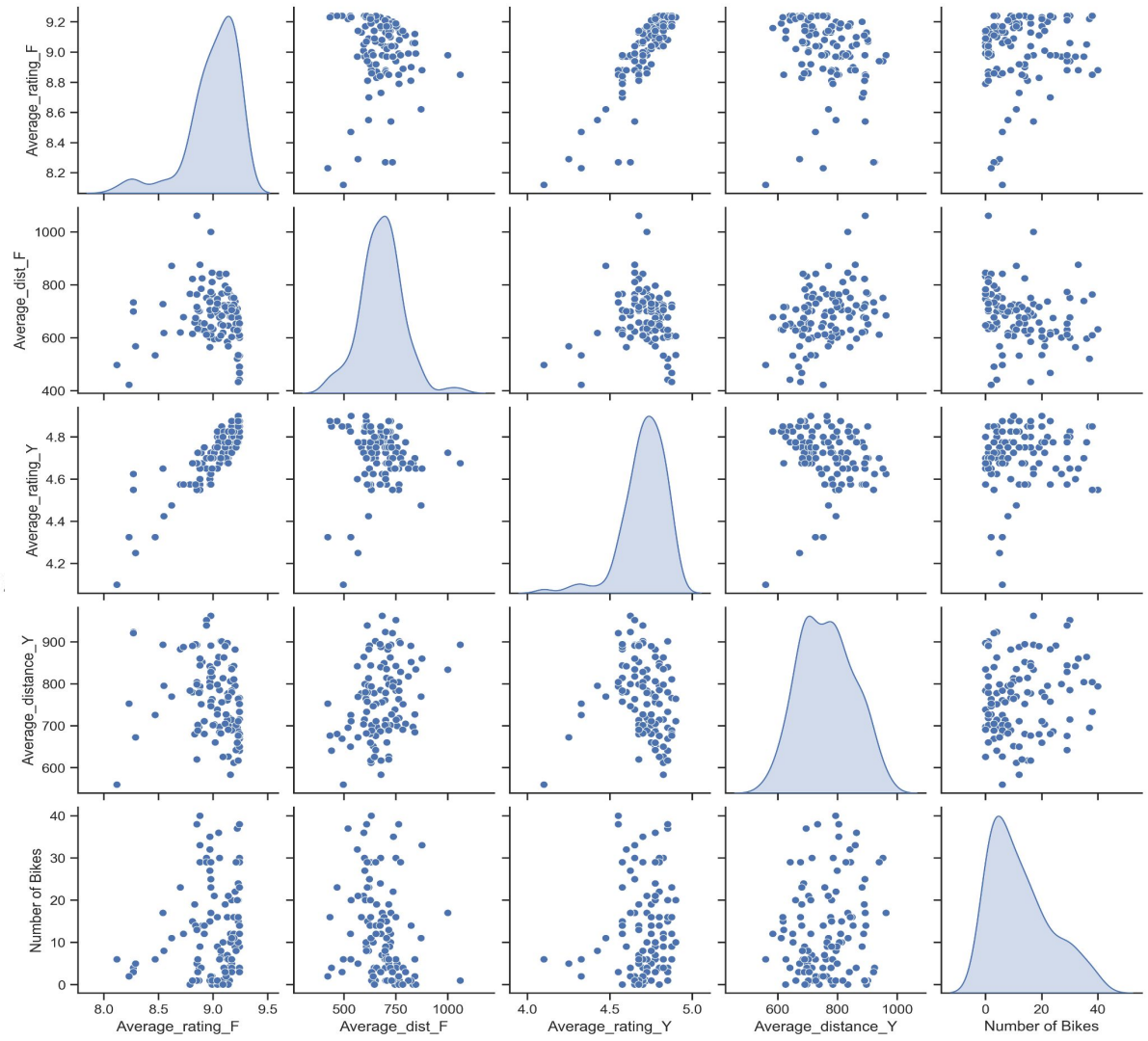
	Latitude	Longitude	Yelp Name	Distance_Y	Review count	Rating_Y	Price
0	53.350230	-6.279696	0 Bar 1661 1 La Pausa Caf...	0 625.352533 1 1144.067227 2 991...	0 7 1 5 2 6 3 17 4 33 5...	0 5.0 1 5.0 2 5.0 3 5.0 4 ...	0 NaN 1 € 2 NaN 3 €€ 4 ...
1	53.357841	-6.251557	0 La Pausa Cafe 1 T...	0 952.731369 1 1060.959965 2 626...	0 5 1 7 2 6 3 15 4 ...	0 5.0 1 5.0 2 5.0 3 5.0 4 ...	0 € 1 NaN 2 NaN 3 €€ 4 ...
2	53.356307	-6.273717	0 Thai Jasmine Grab & Go 1 ...	0 1200.229622 1 808.969104 2 535...	0 11 1 7 2 5 3 15 4 4 5...	0 5.0 1 5.0 2 5.0 3 5.0 4 ...	0 € 1 NaN 2 € 3 €€ 4 ...
3	53.349562	-6.278198	0 Thai Jasmine Grab & Go 1 ...	0 1013.261503 1 518.209459 2 1112...	0 11 1 7 2 5 3 6 4 17 5...	0 5.0 1 5.0 2 5.0 3 5.0 4 ...	0 € 1 NaN 2 € 3 NaN 4 ...

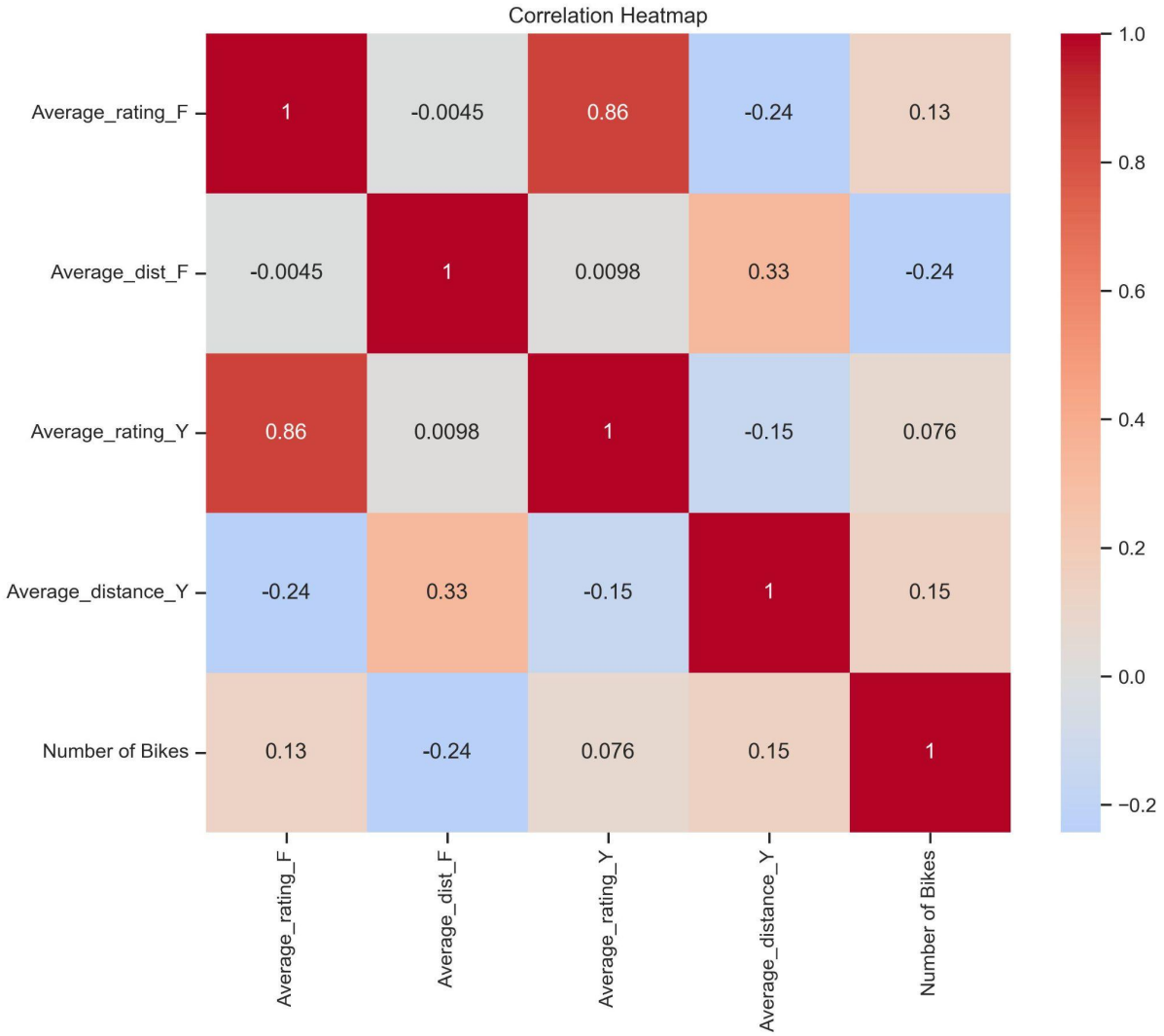
Foursquare results

	Latitude	Longitude	Foursquare Name	Distance	Popularity	Rating	Price
0	53.350230	-6.279696	0 Urbanity Coffee 1 The Porterhous...	0 380 1 985 2 286 3 825 4 658 5...	0 0.992885 1 0.995971 2 0.958540 3 ...	0 9.3 1 9.2 2 9.1 3 9.1 4 9.1 5...	0 1.0 1 2.0 2 NaN 3 2.0 4 1.0 5...
1	53.357841	-6.251557	0 147 Deli 1 Gate The...	0 736 1 869 2 305 3 942 4 1...	0 0.977284 1 0.929310 2 0.997200 3 ...	0 9.4 1 9.0 2 9.0 3 9.0 4 8.9 5...	0 1.0 1 NaN 2 NaN 3 1.0 4 1.0 5...
2	53.356307	-6.273717	0 147 Deli 1 Lill...	0 977 1 983 2 929 3 915 4 604 5...	0 0.977284 1 0.958540 2 0.920853 3 ...	0 9.4 1 9.1 2 9.1 3 9.1 4 9.0 5...	0 1.0 1 NaN 2 1.0 3 NaN 4 1.0 5...
3	53.349562	-6.278198	0 St Patrick's Park 1 Urbanity Coffe...	0 1130 1 308 2 861 3 386 4 ...	0 0.999257 1 0.992885 2 0.995971 3 ...	0 9.3 1 9.3 2 9.2 3 9.1 4 9.1 5...	0 NaN 1 1.0 2 2.0 3 NaN 4 2.0 5...
4	53.336021	-6.262980	0 St. Stephen's Green 1 ...	0 346 1 764 2 745 3 948 4 846 5	0 0.999829 1 0.997400 2 0.999257 3	0 9.5 1 9.4 2 9.3 3 9.2 4 9.2 5	0 NaN 1 NaN 2 NaN 3 NaN 4 NaN 5

Merged DataFrame

	Latitude	Longitude	Average_dist_F	Average_popularity_F	Average_rating_F	Average_distance_Y	Average_review_count_Y	Average_rating_Y	Station Name	Number of Bikes
0	53.350230	-6.279696	606.1	0.967403	9.09	813.634974	21.05	4.725	GEORGES LANE	17
1	53.357841	-6.251557	824.0	0.963300	8.91	852.302705	28.95	4.675	NORTH CIRCULAR ROAD (O'CONNELL'S)	28
2	53.356307	-6.273717	999.3	0.969428	9.01	839.916357	19.25	4.725	PHIBSBOROUGH ROAD	2
3	53.349562	-6.278198	584.4	0.972062	9.13	777.232101	16.45	4.750	SMITHFIELD NORTH	23
4	53.336021	-6.262980	709.7	0.990946	9.20	689.425498	29.90	4.775	CLONMEL STREET	3
...	...	...	...	...	...	...	...	...	...	...
109	53.336597	-6.248109	782.5	0.976231	9.13	690.606529	18.15	4.675	JAMES STREET EAST	0
110	53.353742	-6.265301	646.9	0.967628	9.10	625.914671	17.00	4.800	PARNELL SQUARE NORTH	1


Average_rating_F
Average_dist_F
Average_rating_Y
Average_distance_Y
Number of Bikes
9.2
9.0
8.8
8.6
8.4
8.2
1000
800
600
400
4.8
4.6
4.4
4.2
900
800
700
600
40
30
20
10
0
8.0
8.5
9.0
9.5
500
750
1000
4.0
4.5
5.0
600
800
1000
0
20
40


Correlation Heatmap
Average_rating_F
Average_dist_F
Average_rating_Y
Average_distance_Y
Number of Bikes
1
-0.0045
0.86
-0.24
0.13
-0.0045
1
0.0098
0.33
-0.24
0.86
0.0098
1
-0.15
0.076
-0.24
0.33
-0.15
1
0.15
0.13
-0.24
0.076
0.15
1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Model results

OLS Regression Results

Dep. Variable:	Number of Bikes	R-squared:	0.154
Model:	OLS	Adj. R-squared:	0.131
Method:	Least Squares	F-statistic:	6.656
Date:	Mon, 28 Aug 2023	Prob (F-statistic):	0.000358
Time:	12:49:01	Log-Likelihood:	-422.25
No. Observations:	114	AIC:	852.5
Df Residuals:	110	BIC:	863.4
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-74.2776	40.032	-1.855	0.066	-153.612	5.057
Average_distance_Y	0.0371	0.012	3.208	0.002	0.014	0.060
Average_rating_F	9.2016	4.143	2.221	0.028	0.991	17.412
Average_dist_F	-0.0358	0.010	-3.671	0.000	-0.055	-0.016

Omnibus:	11.459	Durbin-Watson:	1.902
Prob(Omnibus):	0.003	Jarque-Bera (JB):	12.309
Skew:	0.801	Prob(JB):	0.00212
Kurtosis:	3.158	Cond. No.	4.42e+04

Interpretation

Overall, the model suggests that the characteristics of the POIs have a statistically significant impact on the estimated Number of bikes for a given station.

However the Adj. R-squared indicating the goodness of fit of the model shows that around 13% of the variability in Number of bikes can be explained by the independent variables.

Challenges

Project timeframe was shortened due to API request rate limits.

Request failed: Status code: 429.

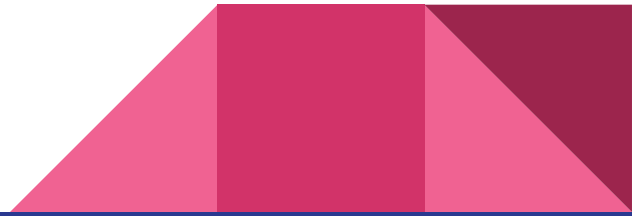
Due to request restrictions on the Yelp API

Future Goals

Extract different characteristics for modelling.

Different type of model.

Include more categories in API request for POIs to enrich dataset.



Thank you.