Exeter Eco-Cafe Project

Introduction

- I want to open an eco-friendly cafe in my home city of Exeter in the UK.
- The city has a wide variety of cafes and eateries.
- None of them climate change thematic.
- Very large student population.
- An eco-friendly cafe could be popular.

- Consider the cafe's carbon footprint.
- Demonstrate comparatively low emissions to its peers.
- Find a commercial with low building emissions or potential for infrastructure upgrades.

- Use machine learning to cluster the districts of Exeter into three zones.

Data

- Non-domestic Energy Performance Certificates (EPCs) from the Energy Performance of Buildings Register via Ministry of Housing, Communities & Local Government's Non-Domestic Energy Performance Certificates API.
- Business venue data by category from Foursquare Developer's Places API.
- Lower Layer Super Output Area (LSOA) boundaries from a GitHub repository maintained by martinjc.
- Postcode to Output Area to Lower Layer Super Output Area to Middle Layer Super Output Area to Local Authority District (February 2020) Lookup in the UK published by the Office for National Statistics.

exeter_pcd_to_oa_lookup DataFrame

- Loaded the lookup table into a DataFrame called exeter_pcd_to_oa_lookup.

	postcode	oa_code	lsoa_code	msoa_code	la_code
0	EX1 1AA	E00171931	E01020016	E02004156	E07000041
1	EX1 1AB	E00171932	E01020016	E02004156	E07000041
2	EX1 1AD	E00171931	E01020016	E02004156	E07000041
3	EX1 1AE	E00171931	E01020016	E02004156	E07000041
4	EX1 1AF	E00171932	E01020016	E02004156	E07000041

venues DataFrame

- Used custom function to get data from Foursquare Places API.
- Categories:

'bakery':'4bf58dd8d48988d16a941735'
'cafe':'4bf58dd8d48988d16d941735'
'coffee_shop':'4bf58dd8d48988d1e0931735'

- Built a DataFrame called venues using selected features from the JSON.

	id	name	postcode	categories	coordinate	lsoa_code
0	52b71919498e7539a262bf4b	Artigiano Espresso Bar	EX4 3PZ	[Café]	(50.72521043944179, -3.5281112803418457)	E01020016
1	4b533692f964a520889227e3	Starbucks	EX4 3LF	[Coffee Shop]	(50.725003, -3.528024)	E01020016
2	4bb5ee5f2ea19521d8e5aa2f	Tea On The Green	EX1 1EZ	[Café]	(50.723252852007754, -3.5301092031773256)	E01020016
3	4c066b785753c9283ad83af1	Costa Coffee	EX1 1LL	[Coffee Shop]	(50.72373054265367, -3.528447185596214)	E01020016
4	601ea846a1c89144b3bd34e9	crankhouse coffee	EX4 3JQ	[Coffee Shop]	(50.721998, -3.533819)	E01020016
5	4f16b1d7e4b0ac7374106bef	Coffee #1	EX1 1GN	[Coffee Shop]	(50.723934788136646, -3.527016812594657)	E01020016
6	4bdade93c4e5952150c26a08	Caffè Nero	EX4 3RS	[Coffee Shop]	(50.72435338401233, -3.531534249563054 4)	NaN
7	5738db5ecd10aa5138b269a1	Starbucks	EX2 8JD	[Coffee Shop]	(50.707949, -3.5371)	E01019972
8	58ab428d5cab2f6649f6c7c8	Puerto Lounge	EX2 4AE	[Café]	(50.718438, -3.532146)	E01020016
9	52f273c7498eb3bd7af0fbc5	Patisserie Valerie	EX4 3DU	[Café]	(50.72301845183361, -3.5321274284828124)	E01020018
10	5790af08cd10f3265885202e	Chococo	EX4 3LS	[Café]	(50.724914, -3.531498)	E01020018
11	4e57a0b01fc7d639d0d2e34e	Cafe Espresso	EX4 3PT	[Café]	(50.725485116386736, -3.5293834472469388)	E01020016
12	57d06fab498e5f51b7443237	boatyard café & Bakery	EX2 8DP	[Bakery]	(50.7165, -3.5298545)	E01019971
13	610e8c270f8ad442ed338131	Cake Or Death	EX4 3AJ	[Bakery]	(50.720348, -3.536105)	E01020018
14	5dd00b53827925008fafabe	Zukis Caffetteria	EX4 3SB	[Coffee Shop]	(50.725765, -3.5331385)	E01020017
15	5d25c2a9b05fbc00239d0d47	Grow Coffee House	EX1 1EG	[Coffee Shop]	(50.720812, -3.530923)	E01020016
16	6120bfce3ce4590c2cd906ae	Lilac Bakery	EX4 1HR	[Bakery]	(50.71514, -3.543866)	E01020030
17	4f64ca2fe4b09fcf62a6de62	Morrisons Exeter Cafe	EX4 7DY	[Café]	(50.733702329928576, -3.5044559275217675)	E01020000
18	5bb23f17f1936e002cc72db5	Sacred Grounds	EX4 3AN	[Café]	(50.72057, -3.535264)	E01020018
19	5c9769feb6eedb002ccce6df	The Common Beaver	EX2 4TA	[Café]	(50.721066, -3.520341)	E01019996
20	5863ebc513af1c32d0c6cbfa	Board	EX1 1EQ	[Café]	(50.721373, -3.531493)	E01020016
21	611a6d65d1ed4f5a46a898d1	Pura Vida	EX4 6QR	[Coffee Shop]	(50.728538, -3.524216)	E01020020
22	58e763873f5a5c09059f4bc1	March Coffee	EX1 1EQ	[Coffee Shop]	(50.72192874191416, -3.5319864749908447)	E01020016
23	6117de5289aea3606cfd102f	twisters cafe	EX4 6NS	[Café]	(50.727098, -3.523739)	E01020022
24	4b55f191f964a520ebf727e3	Boston Tea Party	EX4 3RP	[Café]	(50.72463506829964, -3.5317258717099715)	E01020018
25	4b893ed2f964a520ab2532e3	Caffè Nero	EX4 3DT	[Coffee Shop]	(50.722733856053054, -3.532398901833983)	E01020016
26	4db6bdeb1e7248d135ce75f7	Trailways Cafe	EX4 8JF	[Café]	(50.736699581976644, -3.477773666381836)	E01020005
27	58f0b862149946327a1b7381	No 1 Polsloe	EX1 2HL	[Café]	(50.730053, -3.514011)	E01020008
28	4cbeacea73ef199c7b3ad8a2	Binelli's Café-Bar-Restaurant, Exeter	EX1 1EU	[Coffee Shop]	(50.72367092459626, -3.5297776758670807)	E01020016

epc_data_by_pcd DataFrame

	postcode	lsoa_code	energy_ratings	building_levels	main_heating_fuels	floor_areas	building_emissions
0	EX1 1AF	E01020016	[C, C]	[3, 3]	[Natural Gas, Natural Gas]	[261, 264]	[62.78, 61.7]
1	EX1 1AL	E01020016	[B]	[5]	[Grid Supplied Electricity]	[711]	[26.27]
2	EX1 1AP	E01020016	[C, C]	[3, 3]	[Natural Gas, Natural Gas]	[978, 338]	[39.25, 58.34]
3	EX1 1AR	E01020016	[C, F]	[3, 3]	[Grid Supplied Electricity, Grid Supplied Elec...	[198, 195]	[95.08, 228.59]
4	EX1 1BA	E01020016	[F, E]	[3, 3]	[Natural Gas, Grid Supplied Electricity]	[60, 50]	[187.64]

- Used custom function to get data from Ministry of Housing, Communities & Local Government's Non-Domestic Energy Performance Certificates API.
- Data retrieved by postcode if available.
- Built a DataFrame called epc_data_by_pcd using selected features from the JSON.

venues_by_lsoa DataFrame

- Built a DataFrame called venues_by_lsoa with all distinct LSOA codes and # venues recorded for each.

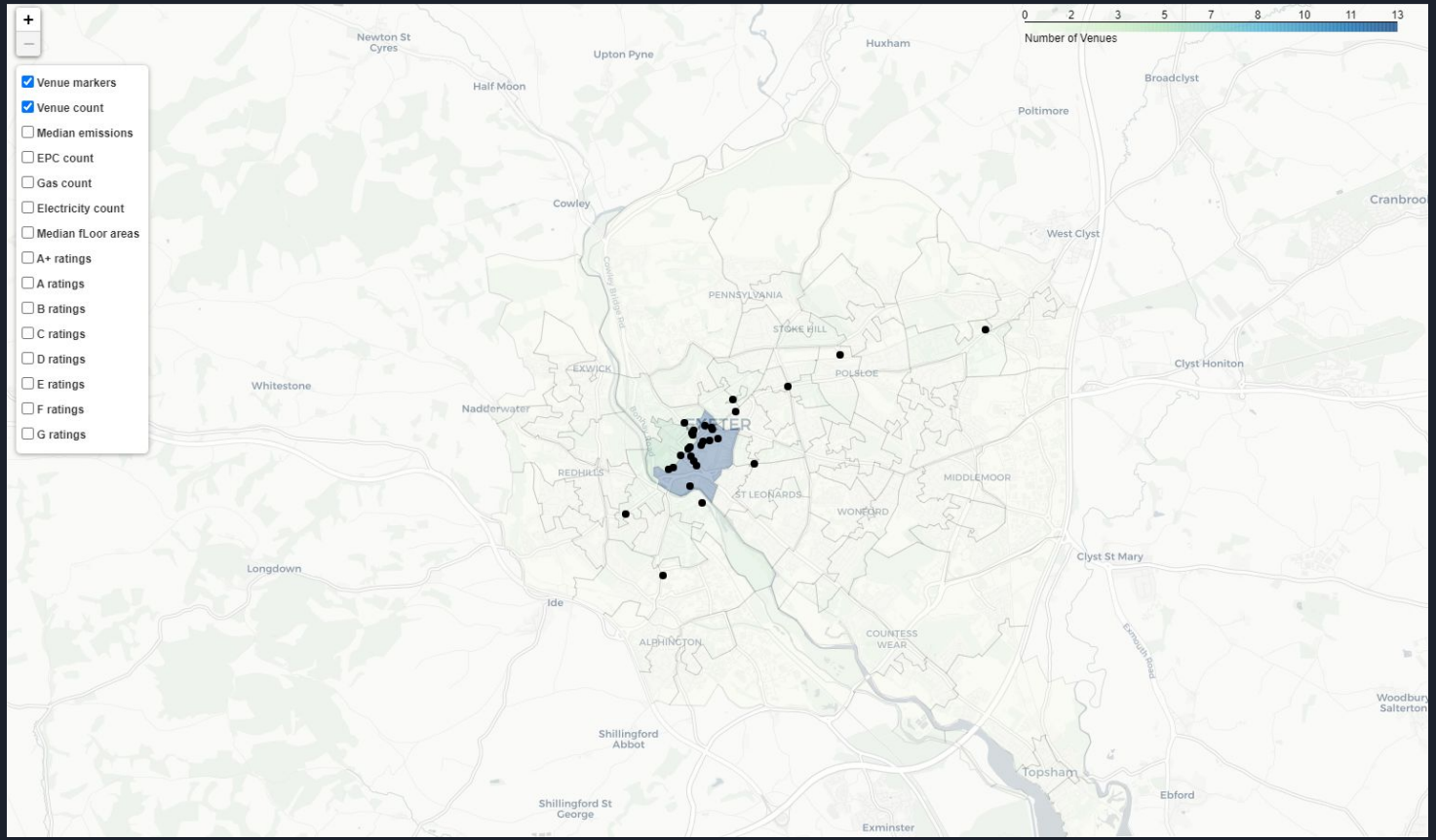
lsoa_code	venue_count
E01020016	13
E01019995	0
E01019996	1
E01019987	0
E01019994	0

Methodology

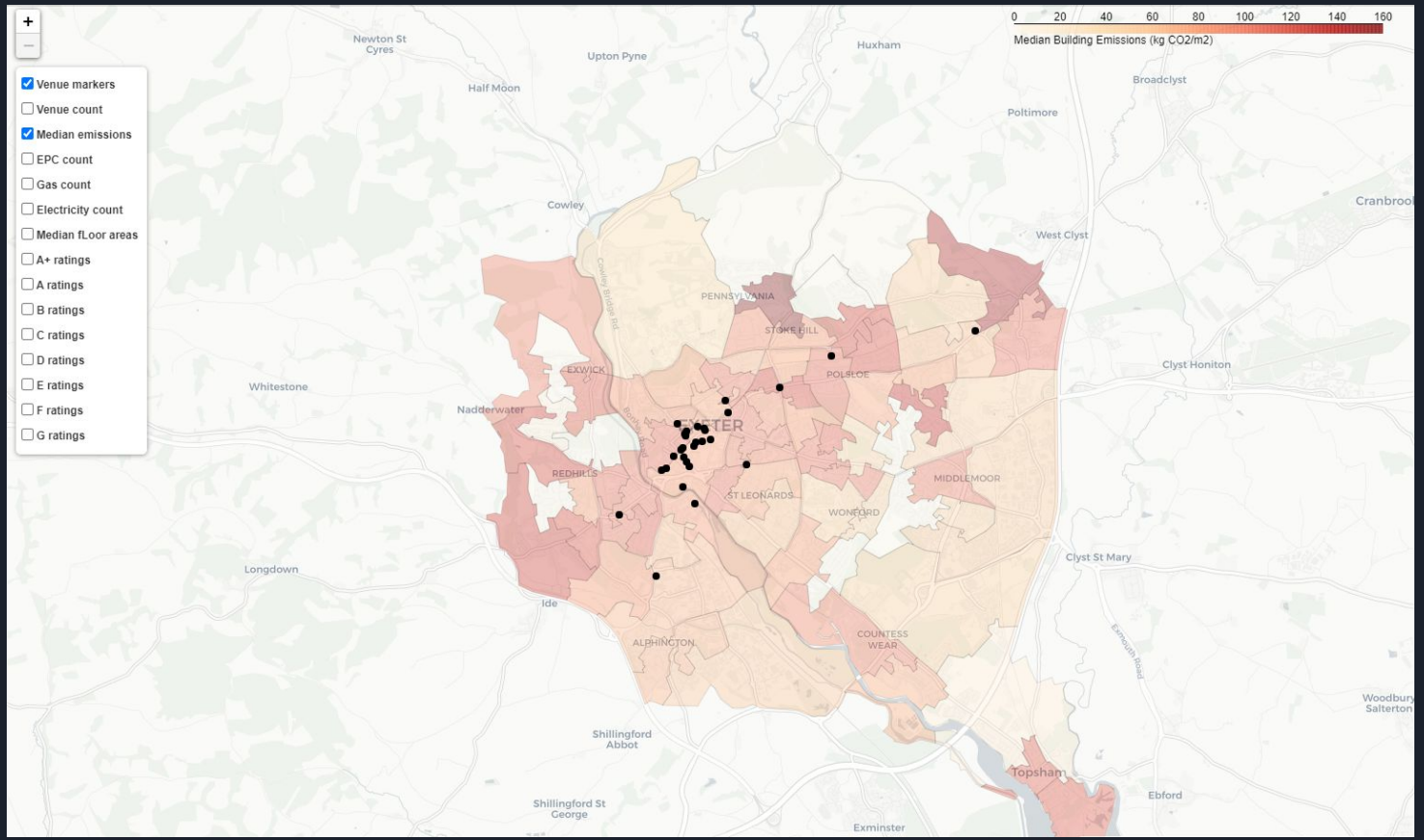
lsoa_code	epc_count	epc_mode	rating_A+	rating_A	rating_B	rating_C	rating_D	rating_E	rating_F	rating_G	fuels_mode	gas_count	electricity_count	floor_areas_median	emissions_median
E01020016	547	D	0	4	25	135	168	95	38	82	Grid Supplied Electricity	200	340	217.0	81.960
E01019995	130	D	0	1	12	28	40	22	11	16	Grid Supplied Electricity	50	78	139.0	99.650
E01019996	43	C	0	0	3	14	13	7	4	2	Natural Gas	24	17	173.0	94.865
E01019987	53	C	0	0	7	17	13	7	4	5	Grid Supplied Electricity	24	29	111.0	102.140
E01019994	20	D	0	1	2	6	6	2	0	3	Natural Gas	12	8	103.0	87.965

- Need to create choropleth maps to analyse the data.
- Built a DataFrame called epc_data_by_lsoa from epc_data_by_pcd by grouping by LSOA.
- Engineered 15 features to map via choropleth:

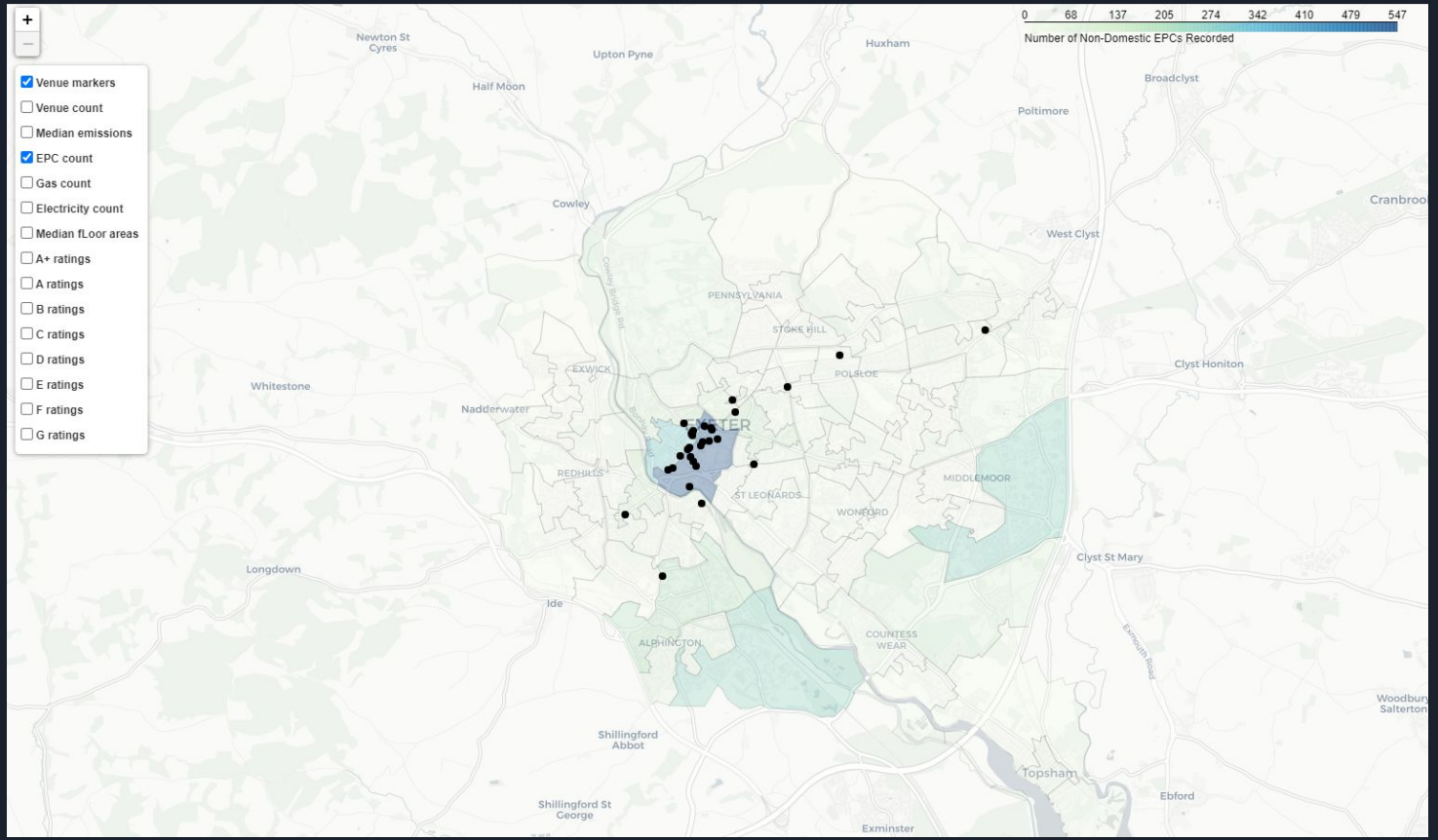
Epc_count ● epc_mode ● rating_A ● rating_A ● rating_B ● rating_C ● rating_D ● rating_E ● rating_F ● rating_G ● fuels_mode ● gas_count ● electricity_count ● floor_areas_median ● emissions_median



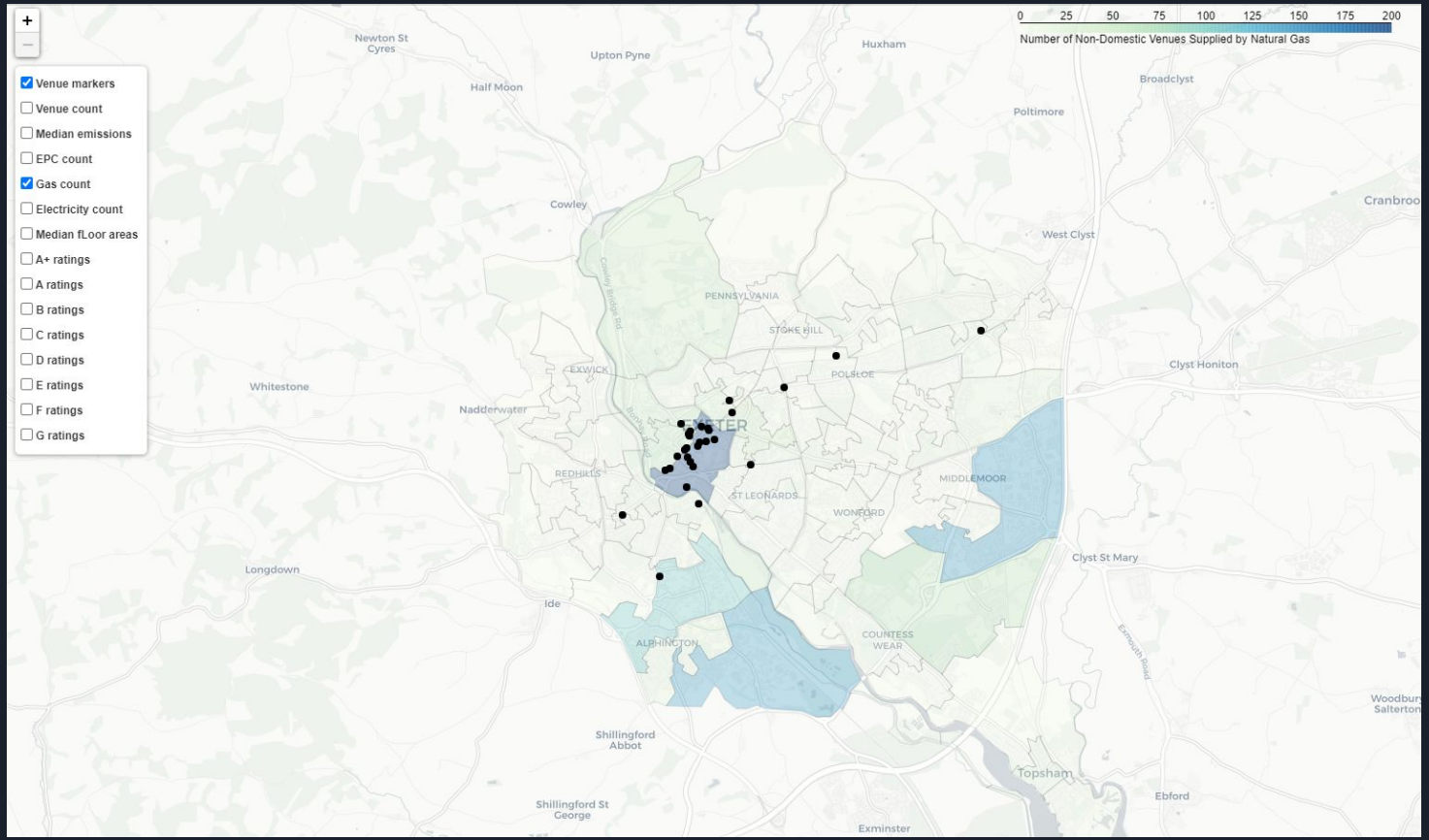
Venue count choropleth with competition venue markers.



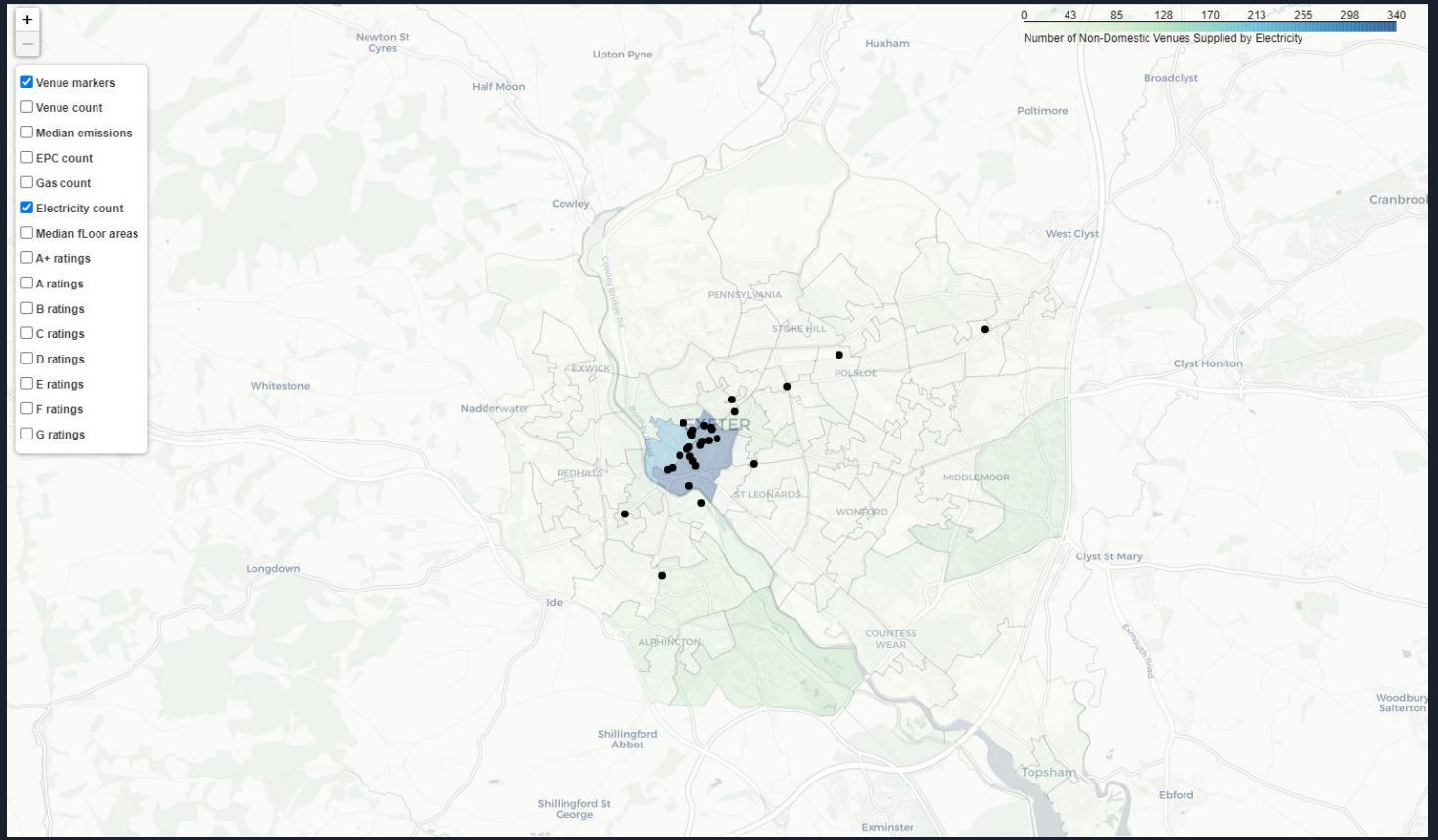
Median emissions (kg co^2/m^2) choropleth with competition venue markers.



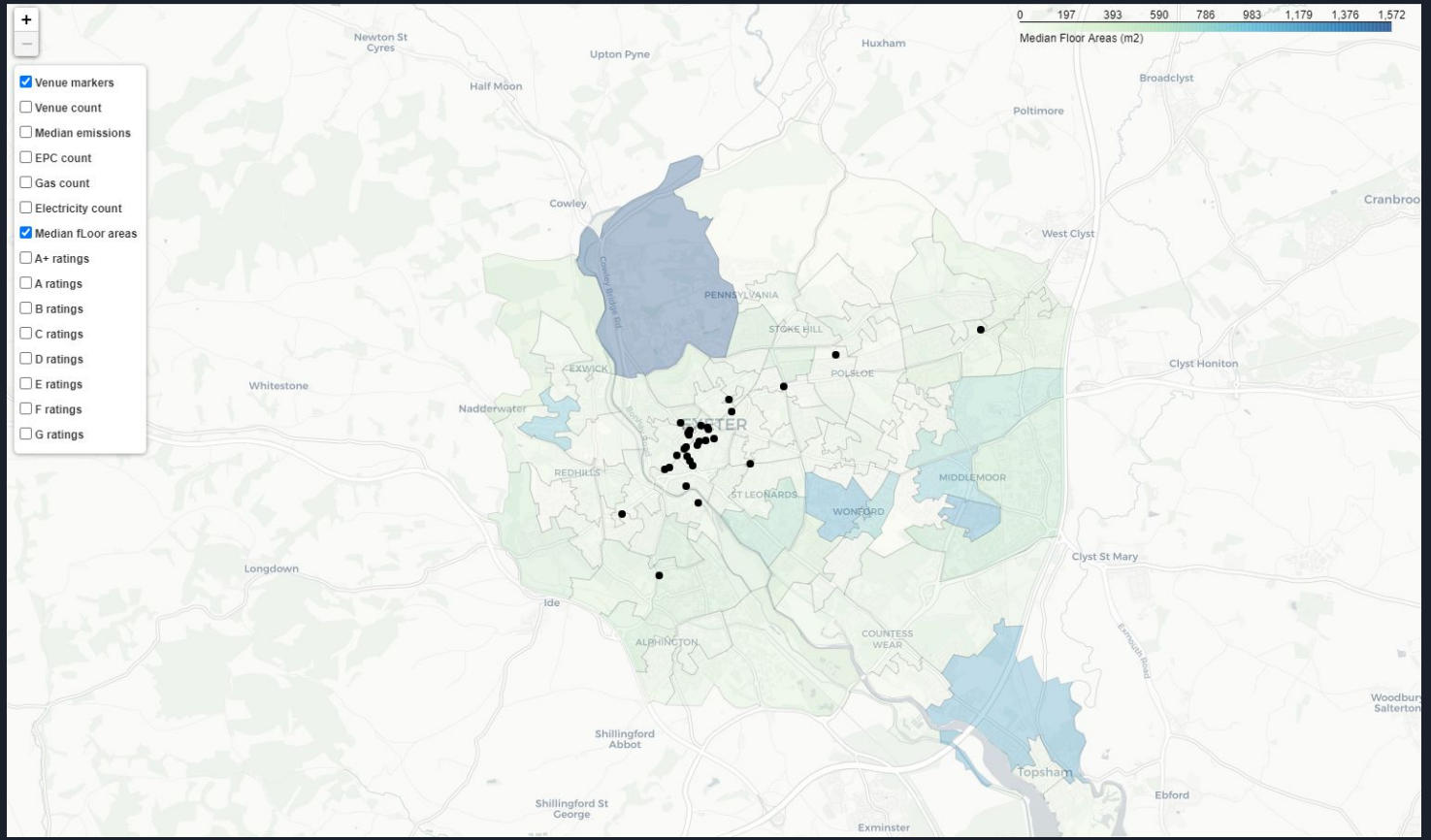
EPC count choropleth with competition venue markers.



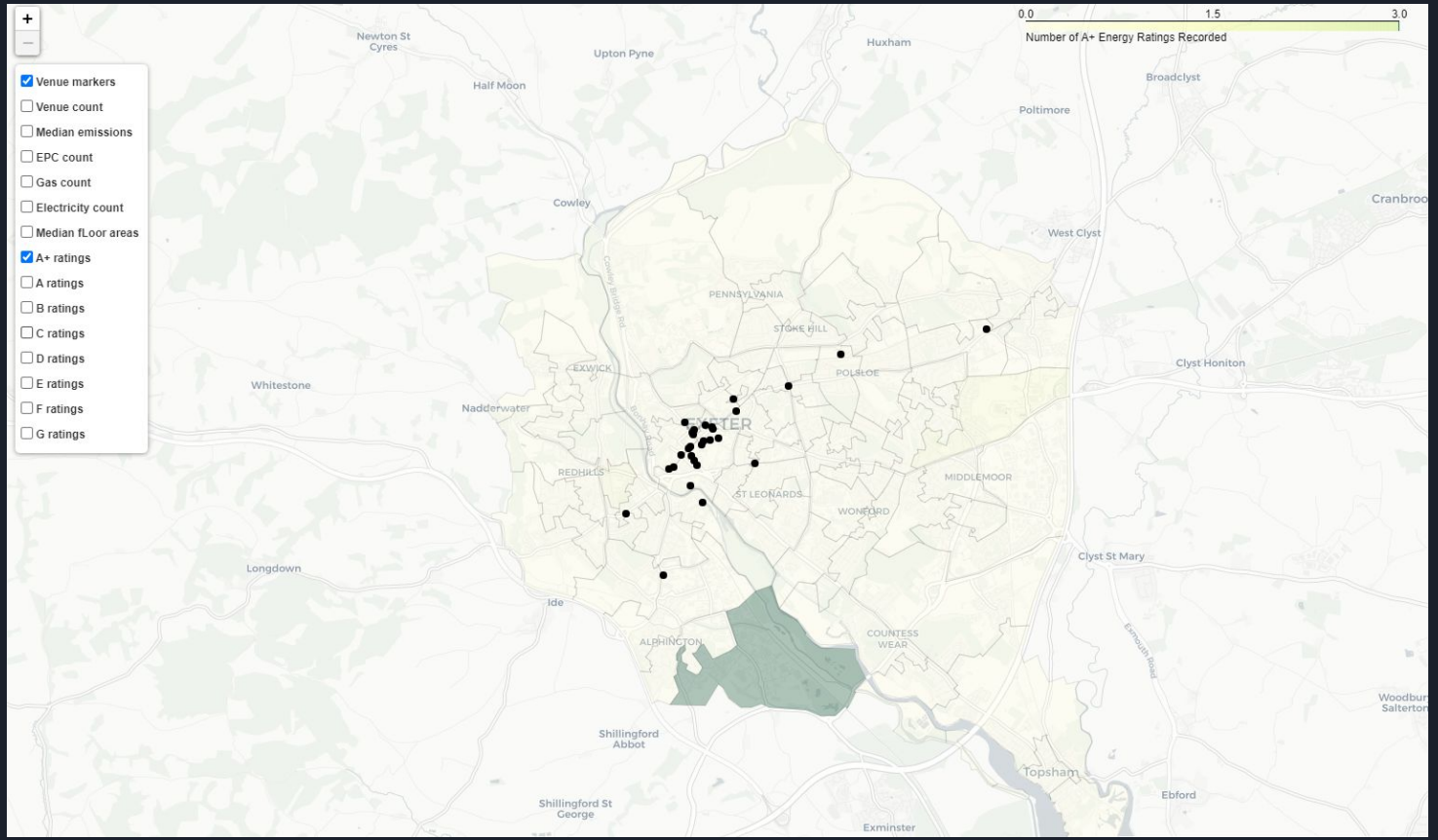
Gas supply count choropleth with competition venue markers.



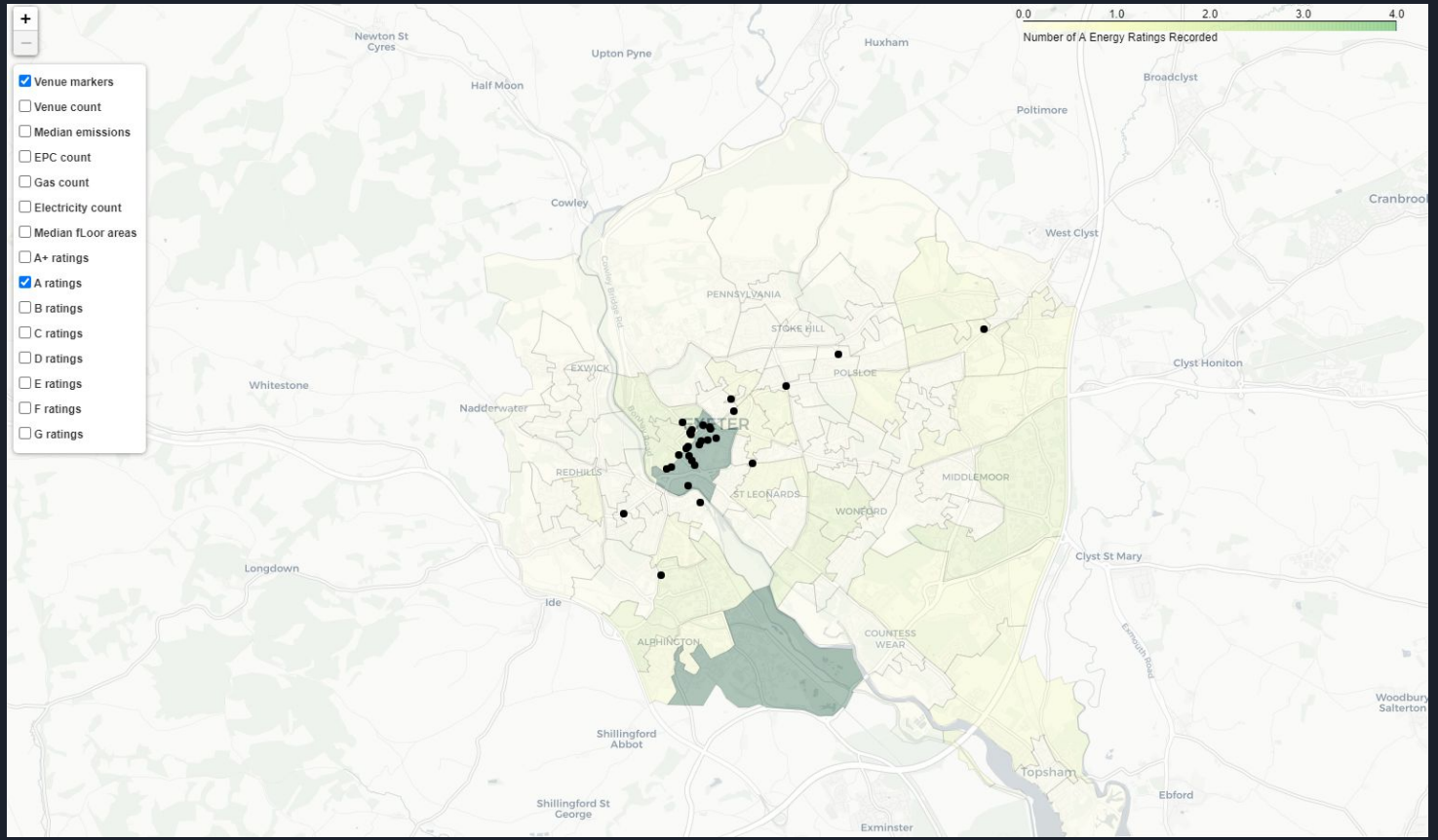
Electricity supply count choropleth with competition venue markers.



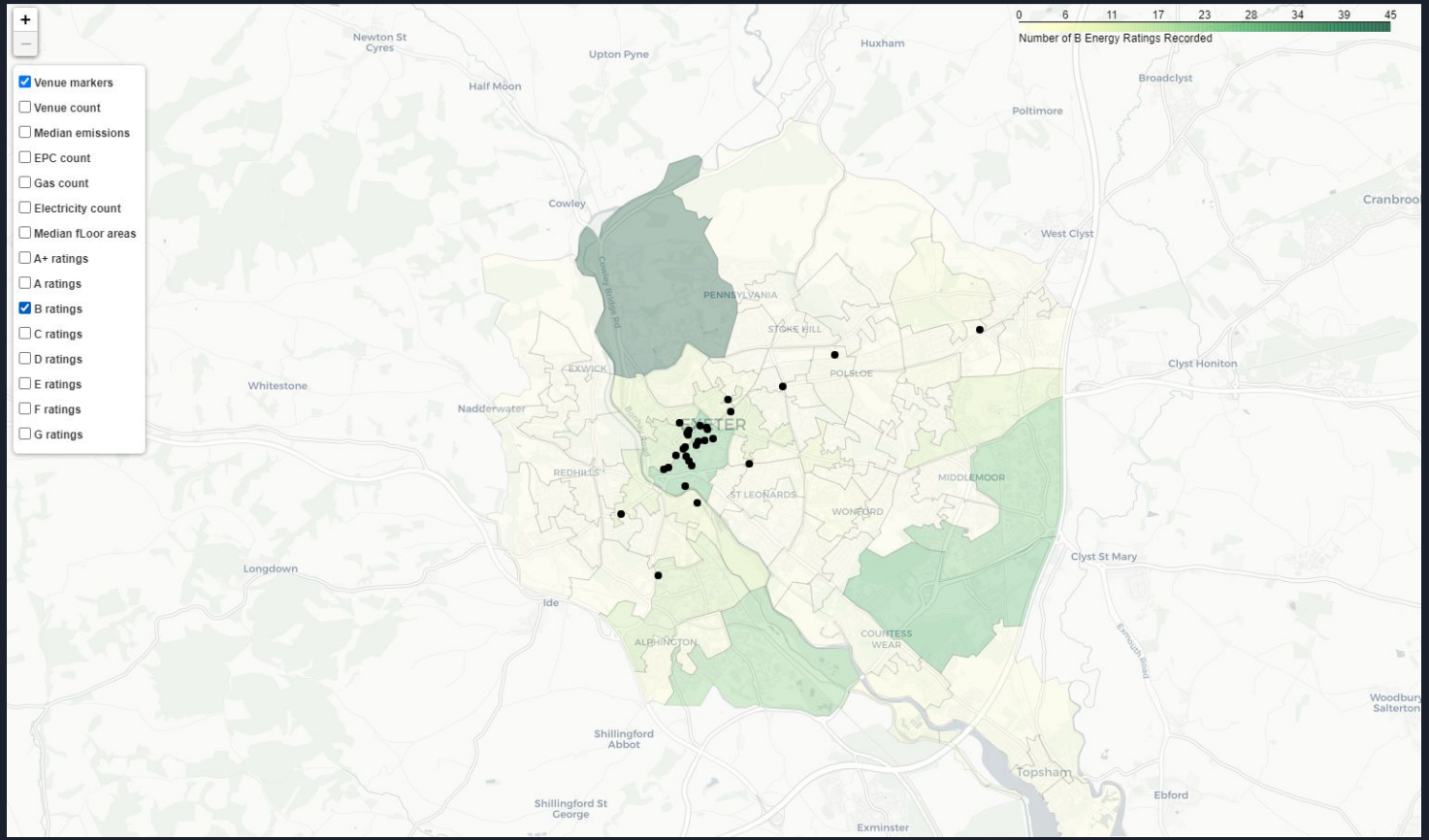
Median floor areas (m^2) choropleth with competition venue markers.



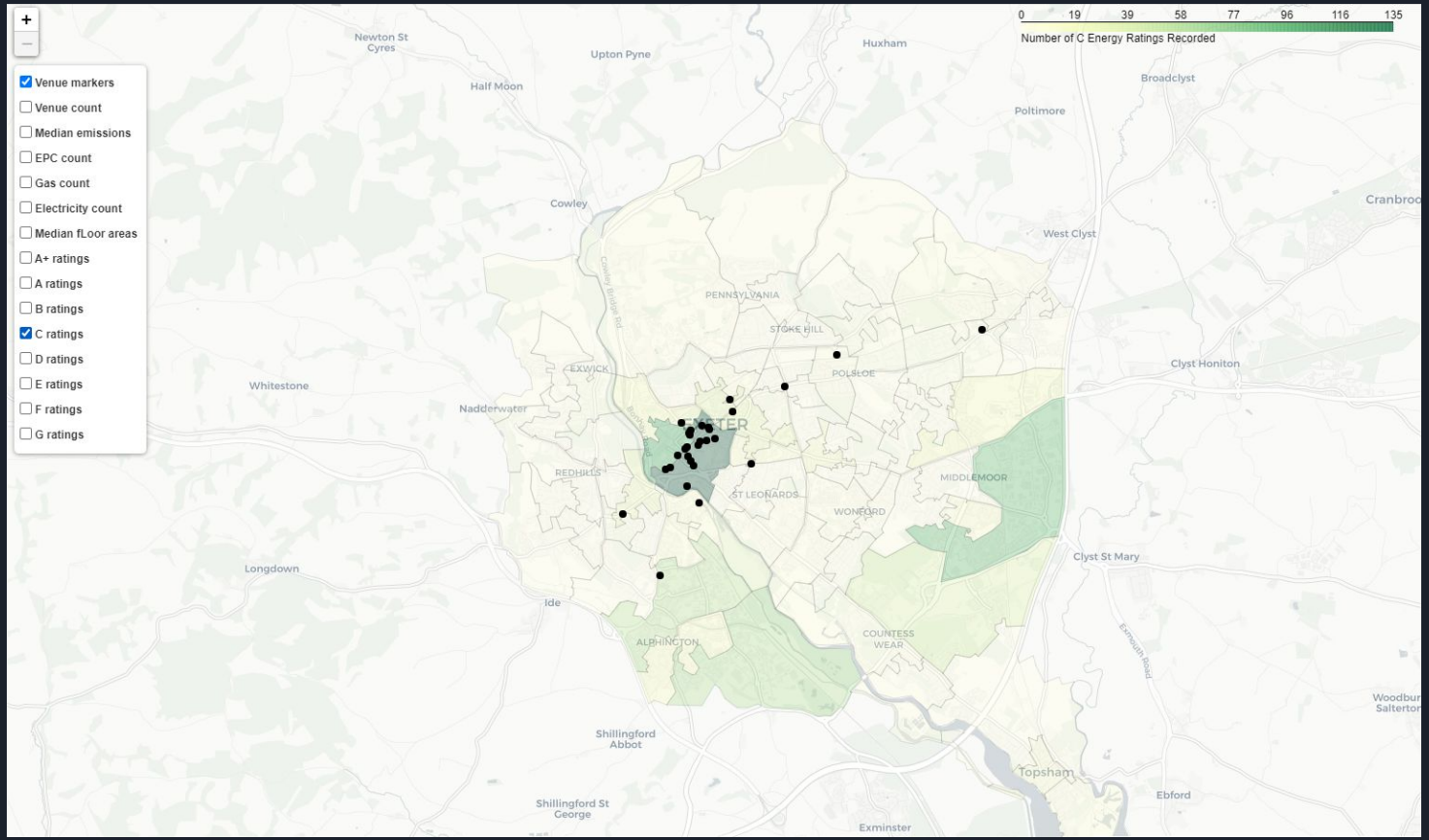
A+ ratings choropleth with competition venue markers.



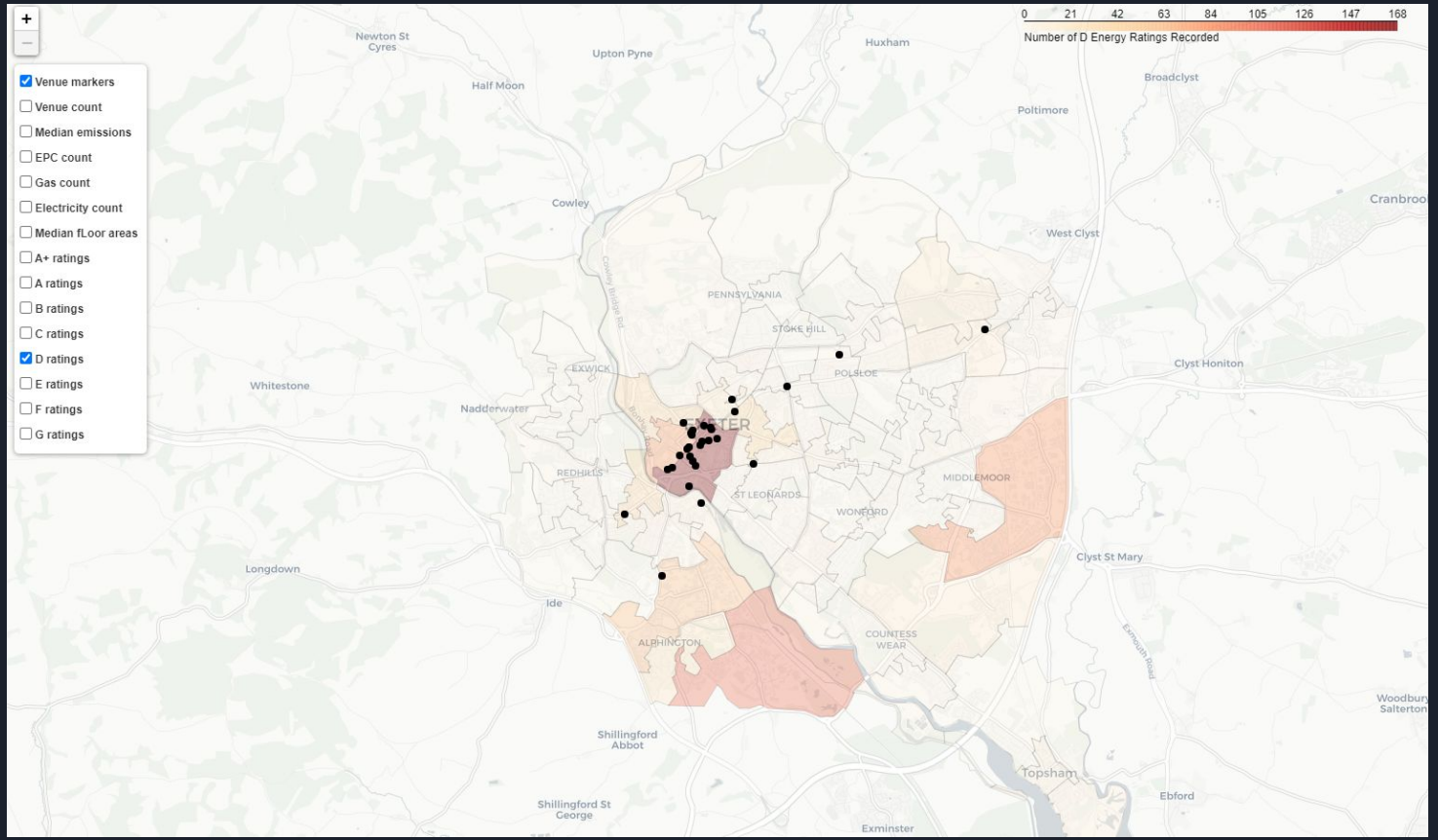
A ratings choropleth with competition venue markers.



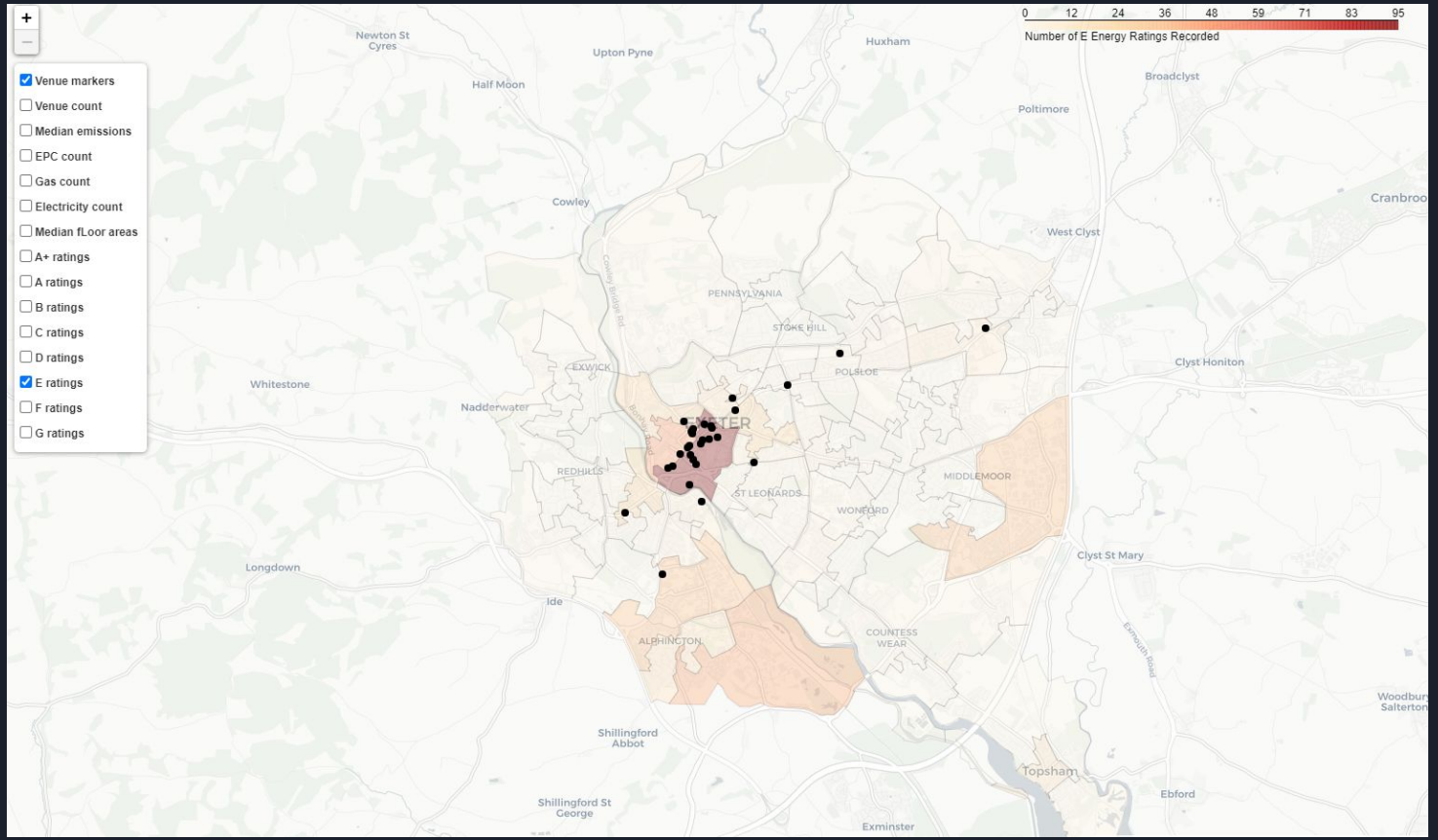
B ratings choropleth with competition venue markers.



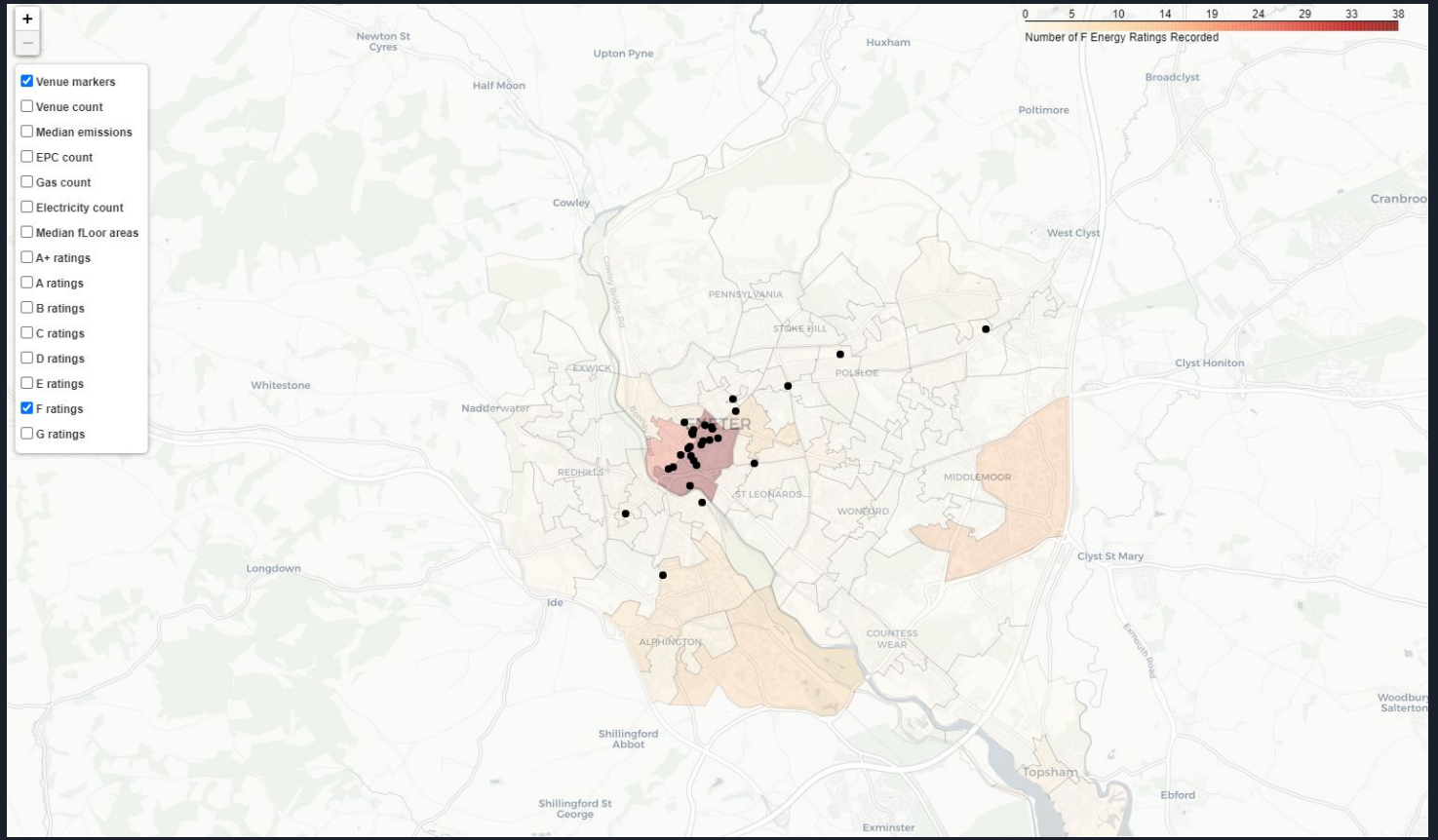
C ratings choropleth with competition venue markers.



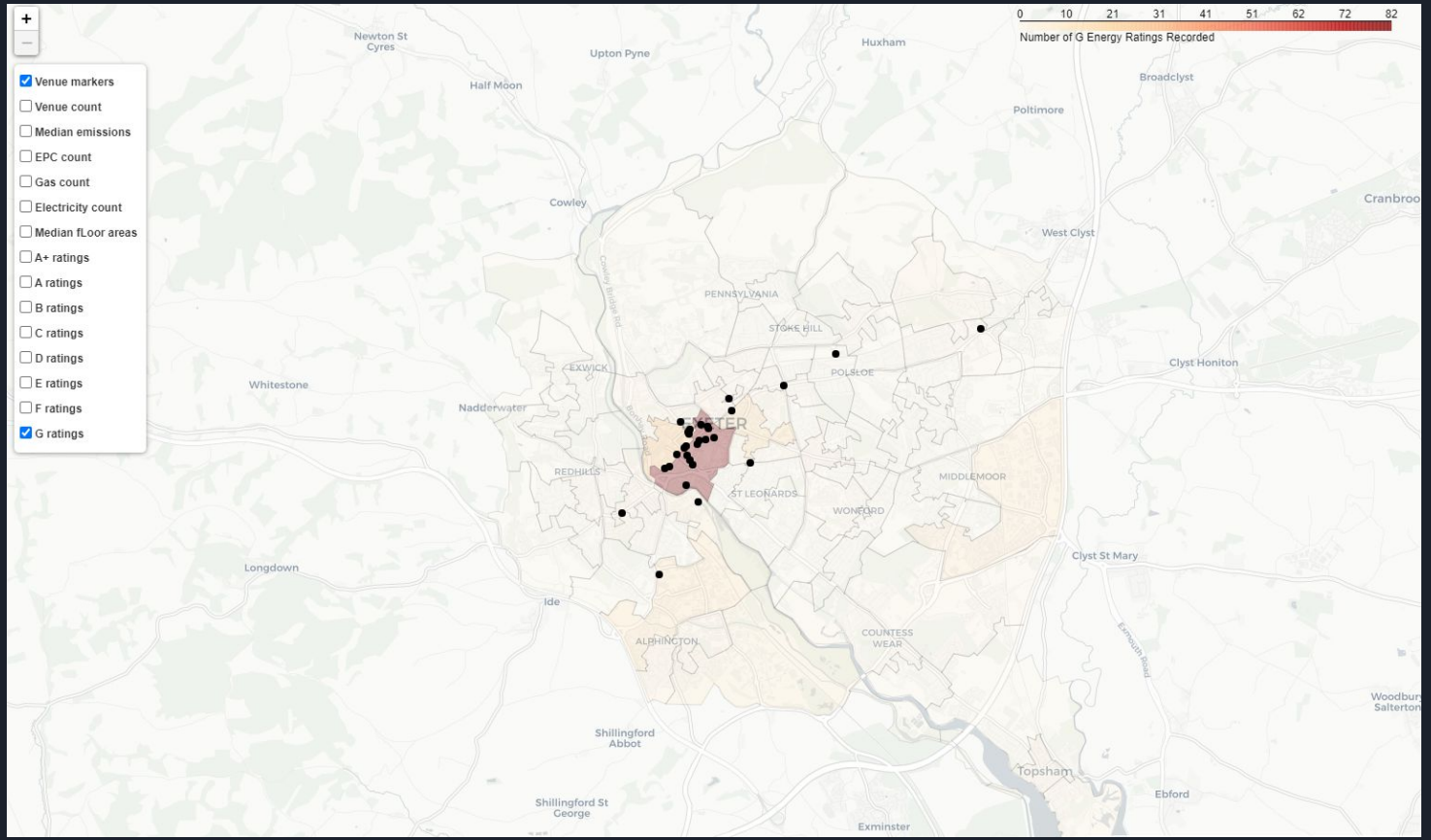
D ratings choropleth with competition venue markers.



E ratings choropleth with competition venue markers.



F ratings choropleth with competition venue markers.



G ratings choropleth with competition venue markers.

exeter_by_lsoa DataFrame

lsoa_code	venue_count	floor_areas_median	emissions_median	mode_B	mode_C	mode_D	mode_E	mode_Grid Supplied Electricity	mode_Natural Gas
E01020016	13	217.0	81.960	0	0	1	0	1	0
E01019995	0	139.0	99.650	0	0	1	0	1	0
E01019996	1	173.0	94.865	0	1	0	0	0	1
E01019987	0	111.0	102.140	0	1	0	0	1	0
E01019994	0	103.0	87.965	0	0	1	0	0	1

- Need a DataFrame to perform k-means clustering.
- Selective features only:

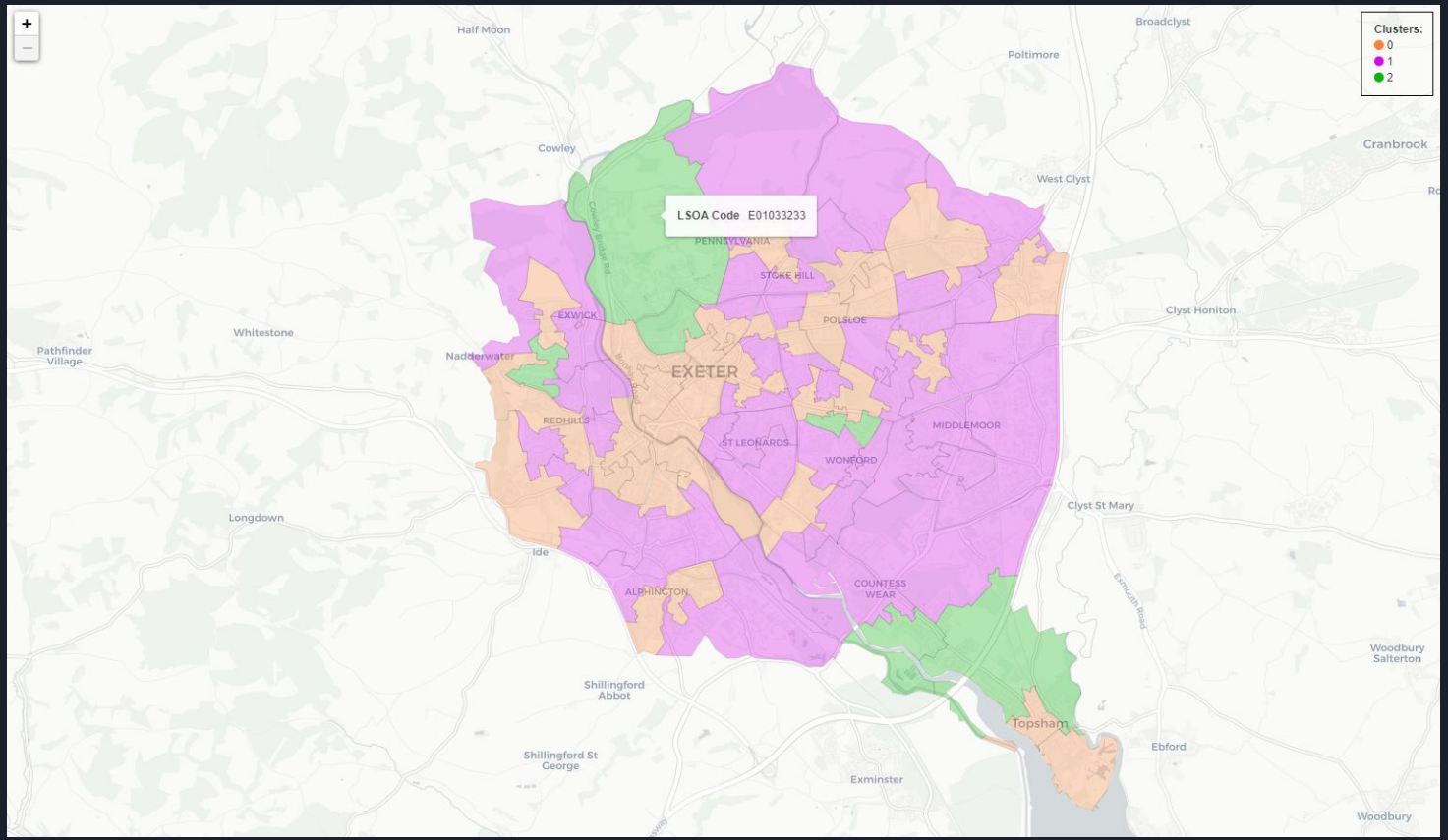
competition (venue_count) ● building footprint (floor_areas_median) ● building emissions (emissions_median) ● most common EPC rating (epc_mode) ● most common main fuel supply (fuels_mode)

- One-hot encoding required.
- Built a DataFrame called exeter_by_lsoa from venues_by_lsoa and epc_data_by_lsoa.

K-Means Clustering

lsoa_code	k_means_label	venue_count	floor_areas_median	emissions_median	mode_B	mode_C	mode_D	mode_E	mode_Grid Supplied Electricity	mode_Natural Gas
E01020016	0	13	217.0	81.960	0	0	1	0	1	0
E01019995	0	0	139.0	99.650	0	0	1	0	1	0
E01019996	1	1	173.0	94.865	0	1	0	0	0	1
E01019987	0	0	111.0	102.140	0	1	0	0	1	0
E01019994	1	0	103.0	87.965	0	0	1	0	0	1

- K = 3
- Clustered LSOAs into three zones.
- Assigned labels to k_means_label in exeter_by_lsoa



K-means clustering results choropleth map.

clusters_average_statistics DataFrame.

- Built a DataFrame called clusters_average_statistics using aggregated features from exeter_by_lsoa.
- Transposed for readability.
- Characteristics of each cluster/zone.

'Venue_count':'sum'
'Floor_areas_median':'mean'
'Emissions_median':'mean'
'mode_B':'sum'
'mode_C':'sum'
'mode_D':'sum'
'mode_E':'sum'
'mode_Grid Supplied Electricity':'sum'
'mode_Natural Gas':'sum'

k_means_label	0	1	2
venue_count	25.000000	3.000000	0.000000
floor_areas_median	200.533333	320.026316	681.916667
emissions_median	97.038333	64.421184	52.980833
mode_B	0.000000	0.000000	6.000000
mode_C	11.000000	17.000000	0.000000
mode_D	16.000000	10.000000	0.000000
mode_E	3.000000	4.000000	0.000000
mode_Grid Supplied Electricity	30.000000	0.000000	0.000000
mode_Natural Gas	0.000000	31.000000	6.000000
lsoa_count	30.000000	38.000000	6.000000

Discussion & Conclusion

Cluster 0	Cluster 1	Cluster 2
✅ High competition ✅ Small footprint ✅ High emissions ✅ Low ratings ✅ Electricity supplied	✅ Low competition ✅ Medium footprint ✅ Medium emissions ✅ Medium ratings ✅ Gas supplied	✅ Low competition ✅ Large footprint ✅ Low emissions ✅ High ratings ✅ Gas supplied

- Different challenges depending on which of these zones the venue might be located.
- High competition is okay—venue concept is niche.
- Small m^2 footprints = low upgrade investment costs.
- Gas supply is preferable for a cafe venue.
- Low emissions & high ratings usually mean buildings are specialised.