

# Moscow Housing Notes

## Data Overview

### Train Data

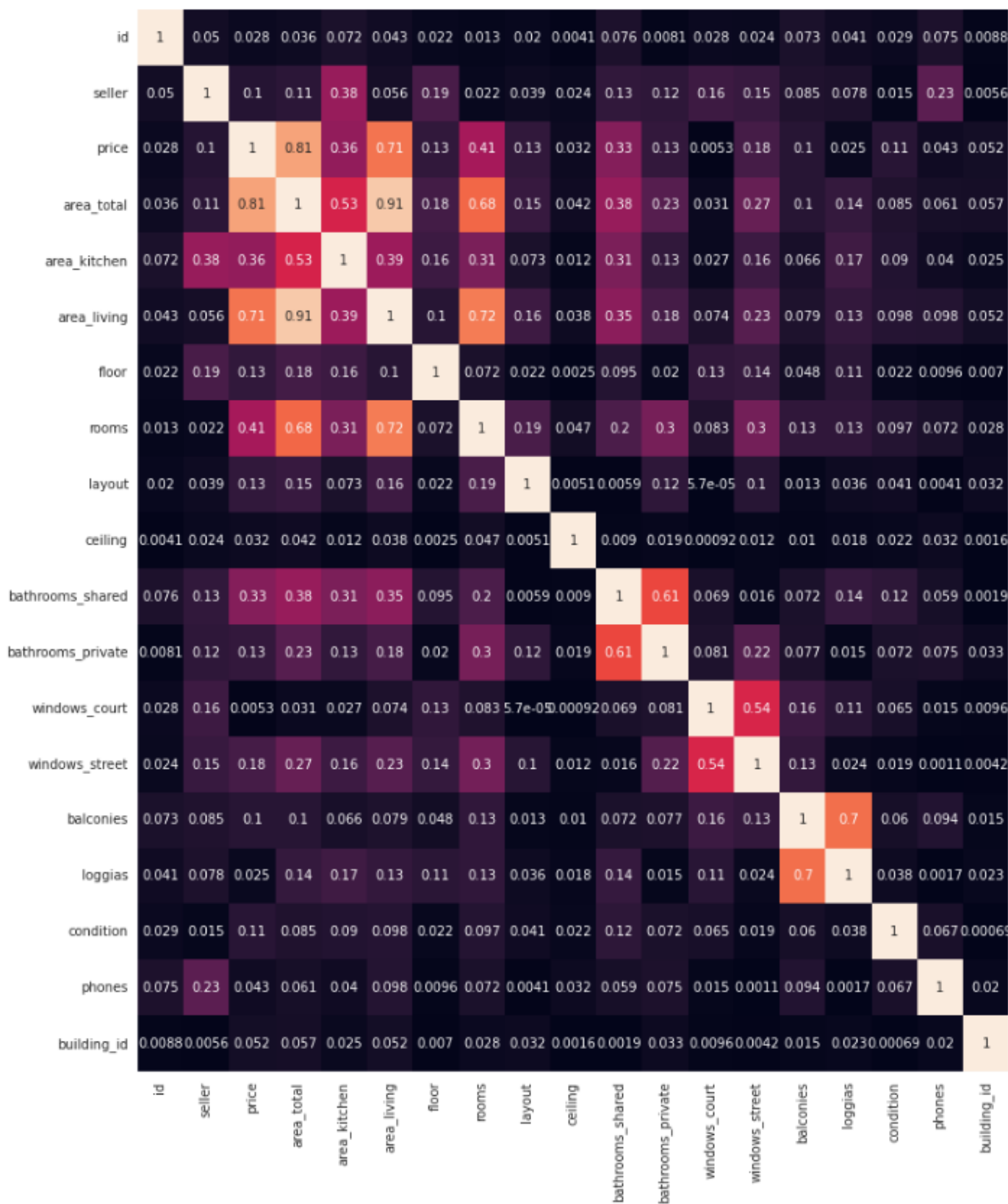
#### Apartments

Basic feature information:

Loaded 23285 apartments

	Type	#NaN	Description	Statistics
id	integer_id	0	Unique ID for apartment	#unique=23285
seller	categorical	8830	The type of apartment seller	Agents(1895), Company(4768), Developer(6185), Owner(1607)
price	real	0	The listed price of the apartment (TARGET)	mean=23556173.5, std=52643927.8, min=900000.0, max=2600000000.0
area_total	real	0	Total area of the apartment	mean=74.5, std=58.7, min=9.3, max=2181.0
area_kitchen	real	4721	Total kitchen area in the apartment	mean=12.6, std=6.6, min=1.0, max=100.0
area_living	real	3882	Total living space area in the apartment	mean=38.7, std=31.4, min=0.0, max=900.0
floor	integer	0	Primary building floor of the apartment	mean=9.0, std=8.3, min=1.0, max=95.0
rooms	integer	0	Number of rooms in the apartment	mean=2.2, std=1.1, min=1.0, max=6.0
layout	categorical	17642	Overall apartment layout	Adjacent(241), Adjacent_isolated(465), Isolated(4937)
ceiling	real	11093	Ceiling height in the apartment	mean=3.3, std=10.8, min=0.0, max=340.0
bathrooms_shared	integer	3872	Number of shared bathrooms	mean=0.7, std=0.8, min=0.0, max=4.0
bathrooms_private	integer	3872	Number of private bathrooms	mean=0.7, std=0.7, min=0.0, max=4.0
windows_court	boolean	8072	Whether the apartment has windows facing a courtyard	False(3086), True(12127)
windows_street	boolean	8072	Whether the apartment has windows facing a street	False(8083), True(7130)
balconies	integer	10520	Number of balconies in the apartment	mean=0.6, std=0.6, min=0.0, max=4.0
loggias	integer	10520	Number of loggias in the apartment (balcony-like, google it)	mean=0.7, std=0.7, min=0.0, max=4.0
condition	categorical	9810	Current condition of the apartment	Decorated(4540), Euro_repair(3296), Special_design(1996), Undecorated(3643)
phones	integer	80	Number of phone numbers associated with the advert	mean=1.2, std=0.5, min=0.0, max=2.0
building_id	integer_id	0	ID used to map apartments to buildings	#unique=6791

Feature correlations:



Buildings

Train

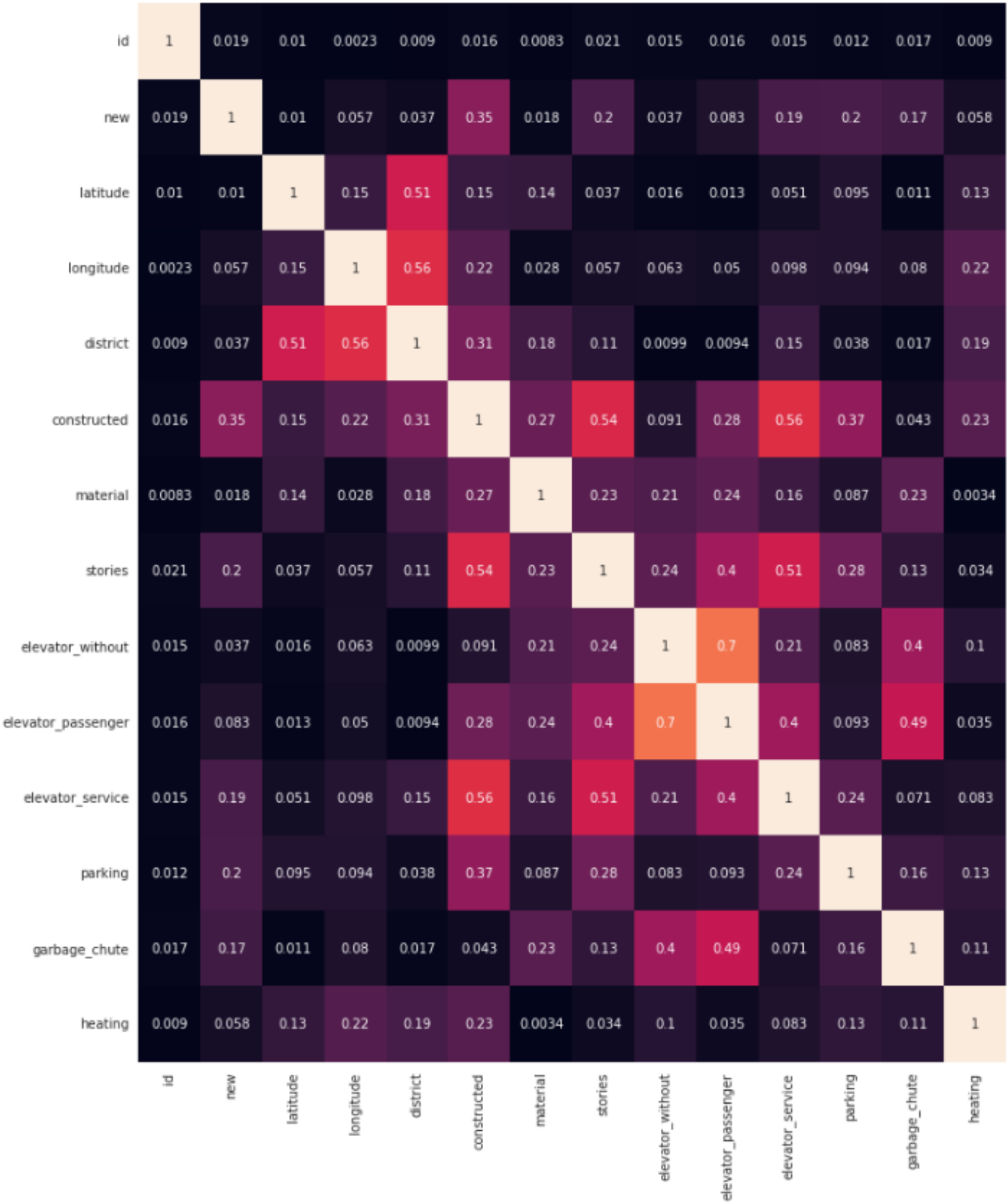
Basic feature information:

Loaded 6791 buildings

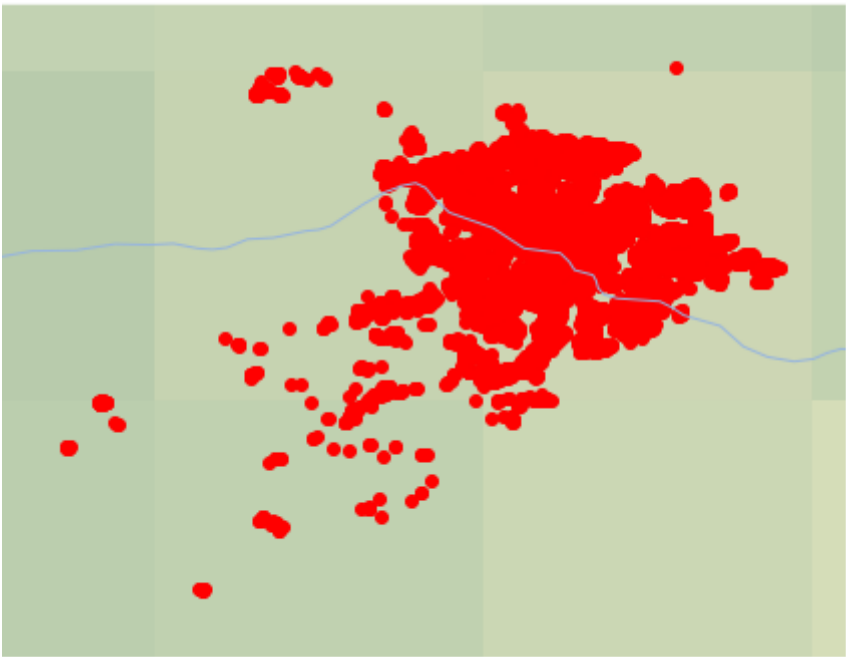
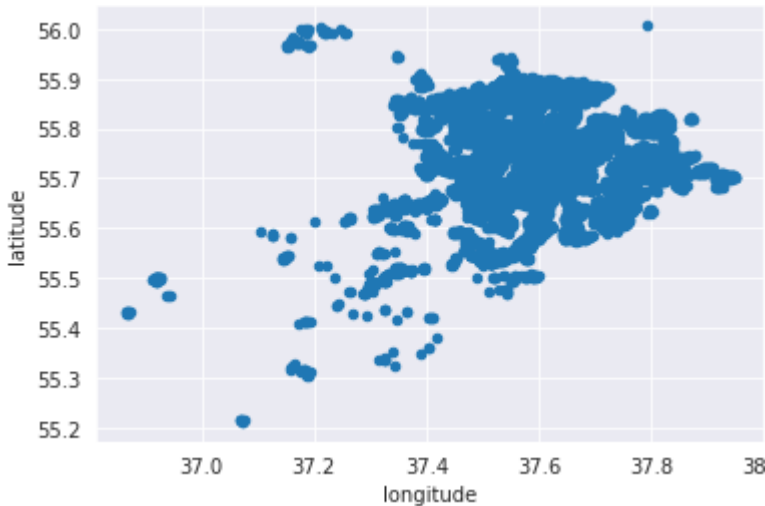
	Type	#NaN	Description	Statistics
id	integer_id	0	Unique ID of building for joining with apartments	#unique=6791
new	boolean	227	Whether it is an old or new building	False(6141), True(423)
latitude	real	0	Latitude coordinate of building	mean=55.7, std=0.1, min=55.2, max=56.0
longitude	real	0	Longitude coordinate of building	mean=37.6, std=0.2, min=36.9, max=38.0
district	categorical	1	Administrative district within Moscow	Central(637), East(956), North(593), North-East(630), North-West(553), Novomoskovsk(344), South(758), South-East(672), South-West(900), Troitsk(146), West(535), Zelenograd(66)
street	string	0	Building street name	#unique=1682
address	string	0	Building address (within street)	#unique=1851
constructed	integer	479	Year when the building was constructed	mean=1985.8, std=24.3, min=1855.0, max=2023.0
material	categorical	1165	Primary building material used in building	Block(679), Bricks(1226), Monolith(1312), Monolithic_brick(19), Panel(2384), Stalin_project(2), Wood(4)
stories	integer	0	Total number of floors in the building	mean=13.0, std=7.0, min=1.0, max=95.0
elevator_without	boolean	305	Whether the building has apartments without elevator access	False(4691), True(1795)
elevator_passenger	boolean	305	Whether the building has apartments with passenger elevator access	False(1099), True(5387)
elevator_service	boolean	305	Whether the building has apartments with service elevator access	False(3452), True(3034)
parking	categorical	3869	Parking options for building	Ground(749), Multilevel(60), Underground(2113)
garbage_chute	boolean	2589	Whether the building has a garbage chute system	False(552), True(3650)
heating	categorical	1437	Primary heating system used in the building	Autonomous_boiler(80), Boiler(35), Central(5101), Individual(138)

Elevator NaNs are all the same entries. They usually also miss much more data.

Feature correlations:



Locations from longitude and latitude:



Train and Test

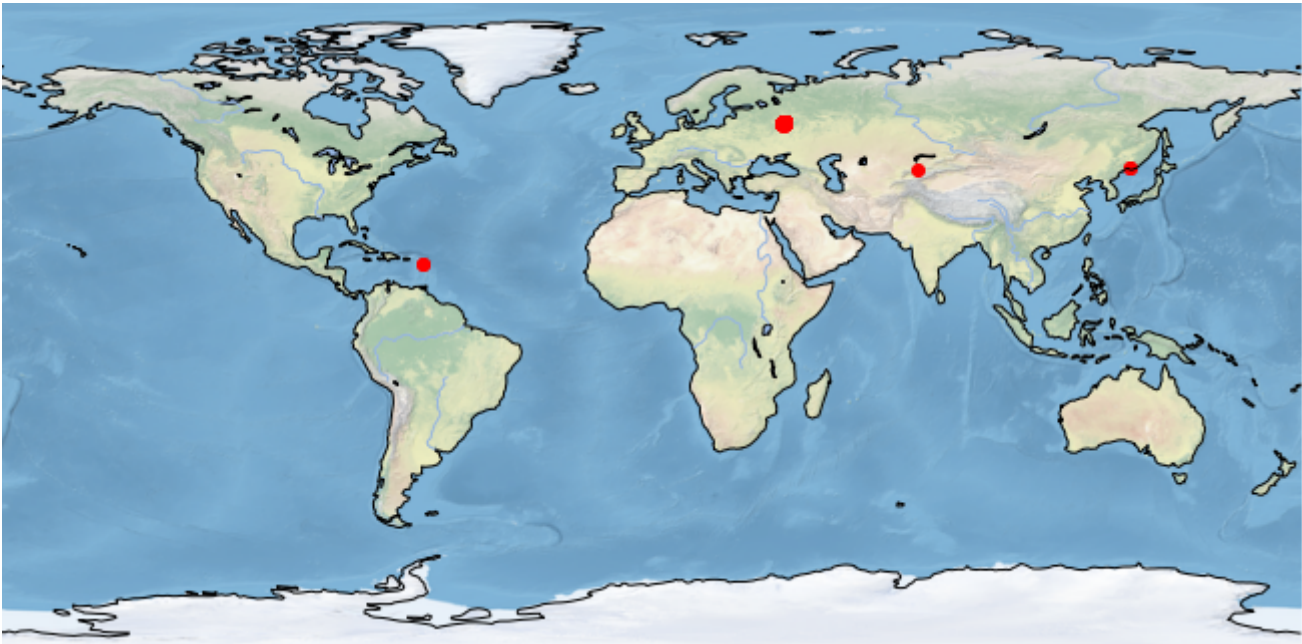
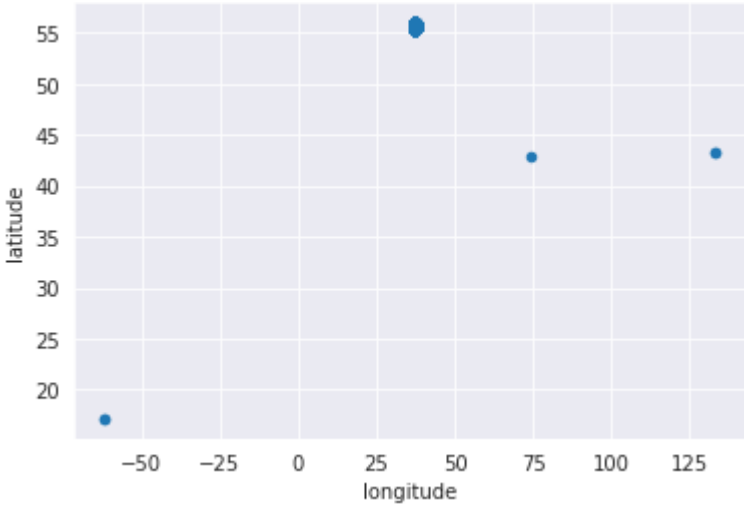
Total NaNs:

id	0
new	325
latitude	1
longitude	1
district	14
street	0
address	0
constructed	708
material	1710
stories	0
elevator_without	447
elevator_passenger	447
elevator_service	447
parking	5507
garbage_chute	3691
heating	2081

Correlations

id	1	0.0066	0.00012	0.0094	0.01	0.012	0.0035	0.025	0.0077	0.0068	0.01	0.016	0.009	0.015
new	0.0066	1	0.057	0.053	0.054	0.35	0.014	0.21	0.038	0.085	0.19	0.2	0.18	0.05
latitude	0.00012	0.057	1	0.41	0.51	0.051	0.039	0.014	0.0077	0.013	0.013	0.05	0.026	0.14
longitude	0.0094	0.053	0.41	1	0.55	0.028	0.0032	0.02	0.014	0.017	0.0033	0.018	0.034	0.22
district	0.01	0.054	0.51	0.55	1	0.32	0.17	0.11	0.016	0.014	0.16	0.032	0.013	0.19
constructed	0.012	0.35	0.051	0.028	0.32	1	0.28	0.55	0.093	0.28	0.56	0.36	0.039	0.23
material	0.0035	0.014	0.039	0.0032	0.17	0.28	1	0.24	0.21	0.25	0.17	0.087	0.24	0.0016
stories	0.025	0.21	0.014	0.02	0.11	0.55	0.24	1	0.24	0.41	0.51	0.28	0.11	0.039
elevator_without	0.0077	0.038	0.0077	0.014	0.016	0.093	0.21	0.24	1	0.71	0.22	0.067	0.41	0.098
elevator_passenger	0.0068	0.085	0.013	0.017	0.014	0.28	0.25	0.41	0.71	1	0.4	0.095	0.49	0.04
elevator_service	0.01	0.19	0.013	0.0033	0.16	0.56	0.17	0.51	0.22	0.4	1	0.24	0.073	0.076
parking	0.016	0.2	0.05	0.018	0.032	0.36	0.087	0.28	0.067	0.095	0.24	1	0.15	0.14
garbage_chute	0.009	0.18	0.026	0.034	0.013	0.039	0.24	0.11	0.41	0.49	0.073	0.15	1	0.1
heating	0.015	0.05	0.14	0.22	0.19	0.23	0.0016	0.039	0.098	0.04	0.076	0.14	0.1	1
id	new	latitude	longitude	district	constructed	material	stories	elevator_without	elevator_passenger	elevator_service	parking	garbage_chute	heating	

Locations:



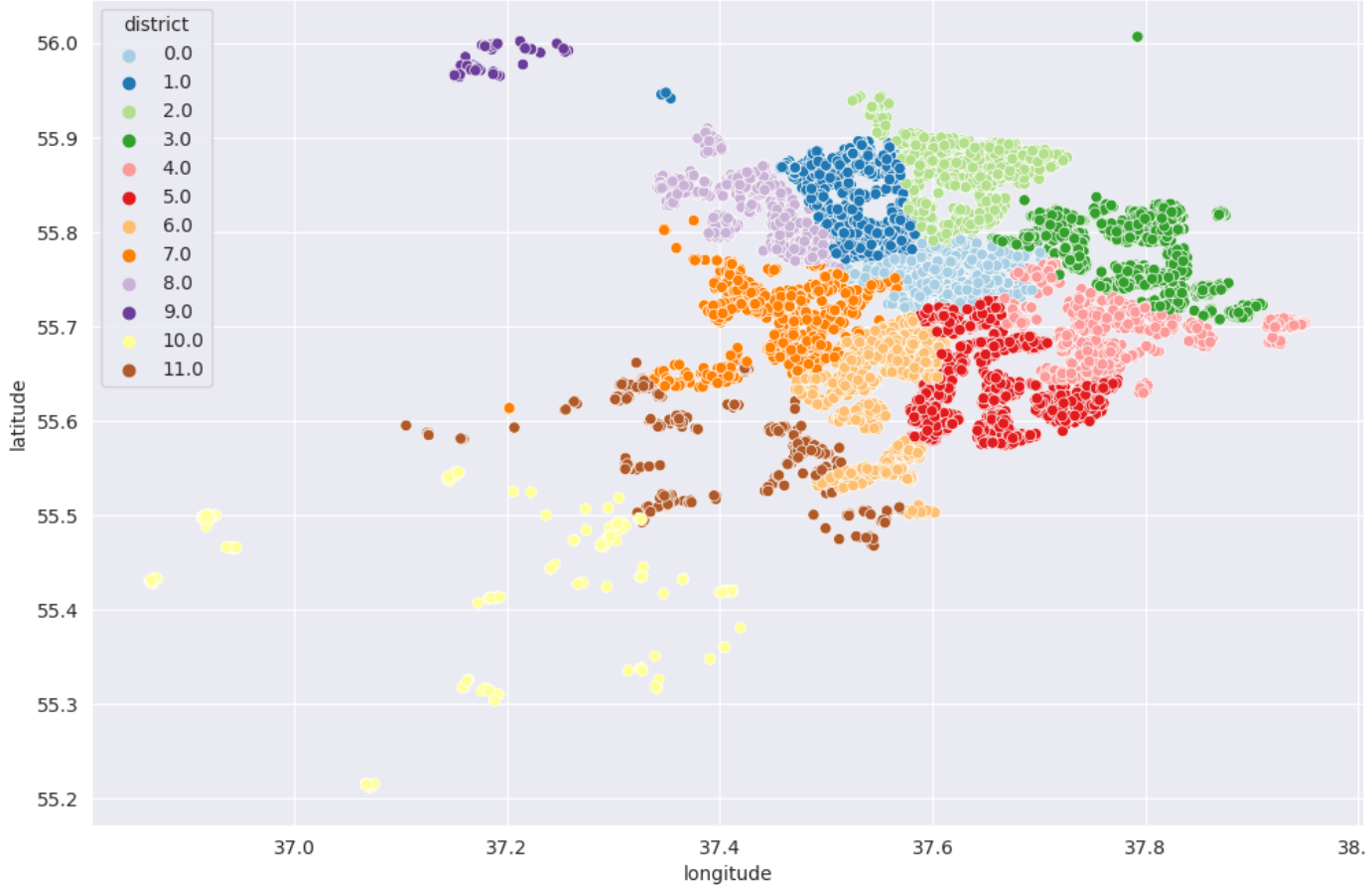
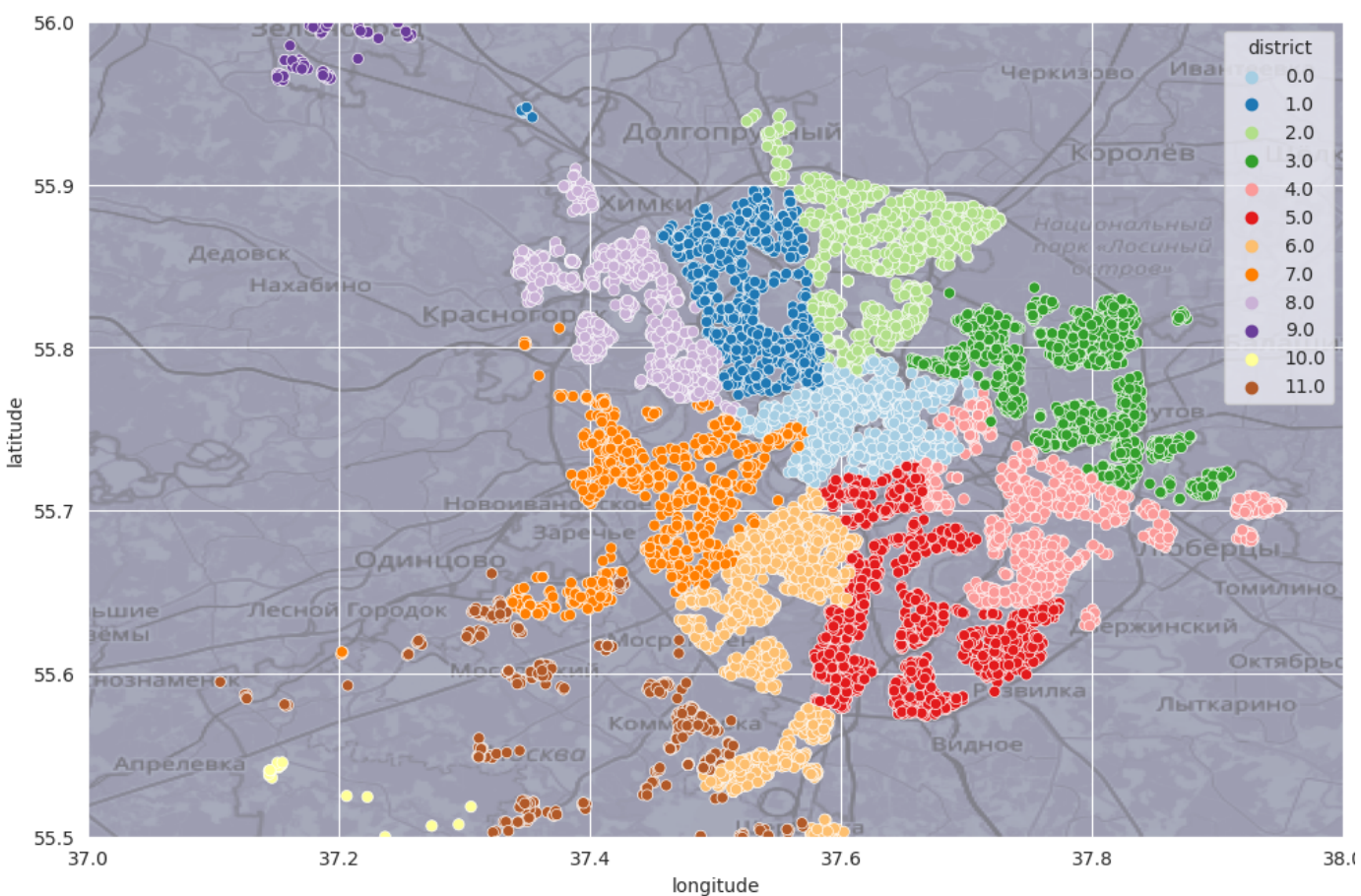
There are some corrupted data-points.

Data-points with coordinates outside of Moscow:

	id	new	latitude	longitude	district	street	address	constructed	material	stories	elevator_without	elevator_passenger	elevator_service	parking	garbage_chute	heating
7	3803	1.000000	nan	nan	nan	пос. Коммунарка	Москва А101 ЖК	2019.000000	2.000000	12.000000	1.000000	1.000000	1.000000	nan	nan	nan
182	5667	nan	43.352968	132.759204	nan	улица Центральная	48	nan	nan	2.000000	nan	nan	nan	nan	nan	nan
1108	4412	1.000000	17.141734	-61.790500	nan	Бунинские Луга ЖК	к2/2/2	2021.000000	nan	9.000000	0.000000	1.000000	1.000000	nan	nan	nan
2270	4202	nan	42.914700	74.517184	nan	улица 1-я Линия	57	2016.000000	nan	3.000000	nan	nan	nan	nan	nan	nan
2721	8811	0.000000	43.353545	132.768097	nan	улица Центральная	75	2007.000000	nan	5.000000	0.000000	1.000000	1.000000	1.000000	1.000000	nan
2838	4636	1.000000	17.141734	-61.790500	nan	Бунинские Луга ЖК	к2/2/1	2021.000000	3.000000	17.000000	1.000000	0.000000	0.000000	nan	0.000000	nan



District clustered data-points:



Unclear definition of feature "new":

	id	new	latitude	longitude	district	street	address	constructed	material	stories	elevator_without	elevator_passenger	elevator_service	parking	garbage_chute	heating
5369	5448	0.000000	55.704697	37.945466	4.000000	13-й кв-л	к16	2021.000000	nan	17.000000	0.000000	1.000000	1.000000	nan	nan	nan
13	7066	0.000000	55.553422	37.464284	11.000000	улица Василия Ощепкова	2	2021.000000	3.000000	15.000000	1.000000	1.000000	1.000000	0.000000	0.000000	nan
1347	973	1.000000	55.840907	37.515227	1.000000	Михалковская улица	46к3	1917.000000	3.000000	5.000000	1.000000	0.000000	0.000000	nan	nan	0.000000



the oldest building with 'new'=True is from 1917 and the newest with 'new'=False is from 2021  
so ether we define the feature our self or we just drop it

## Joined sets

Correlations between apartment and building features:

building_id	0.0088	0.0056	0.052	0.057	0.025	0.052	0.007	0.028	0.032	0.0016	0.0019	0.033	0.0096	0.0042	0.015	0.023	0.00069	0.02
new	0.048	0.79	0.089	0.077	0.26	0.16	0.11	0.11	0.01	0.025	0.033	0.084	0.15	0.082	0.11	0.092	0.11	0.22
latitude	0.033	0.033	0.054	0.062	0.029	0.068	0.058	0.073	0.0059	0.013	0.034	0.02	0.036	0.07	0.087	0.03	0.024	0.019
longitude	0.058	0.041	0.034	0.079	0.1	0.08	0.055	0.055	0.01	0.013	0.068	0.0038	0.045	0.059	0.0062	0.024	0.018	0.0073
district	0.013	0.011	0.18	0.12	0.0094	0.097	0.027	0.1	0.05	0.029	0.048	0.043	0.012	0.065	0.027	0.018	0.091	0.022
constructed	0.028	0.58	0.048	0.097	0.38	0.0021	0.26	0.024	0.0031	0.015	0.16	0.06	0.14	0.12	0.18	0.24	0.077	0.12
material	0.023	0.1	0.1	0.14	0.1	0.15	0.0016	0.14	0.067	0.025	0.11	0.0091	0.019	0.071	0.1	0.06	0.066	0.011
stories	0.0023	0.31	0.086	0.17	0.25	0.12	0.67	0.079	0.036	0.0084	0.15	0.014	0.14	0.17	0.1	0.18	0.0014	0.0022
elevator_without	0.012	0.13	0.16	0.16	0.068	0.16	0.072	0.096	0.01	0.02	0.14	0.035	0.00096	0.044	0.054	0.049	0.039	0.049
elevator_passenger	0.0084	0.21	0.022	0.05	0.16	0.021	0.18	0.021	0.09	0.019	0.043	0.05	0.065	0.086	0.14	0.19	0.02	0.038
elevator_service	0.0025	0.33	0.025	0.093	0.3	0.044	0.24	0.026	0.04	0.0013	0.13	0.037	0.13	0.14	0.11	0.18	0.00081	0.045
parking	0.034	0.19	0.2	0.26	0.22	0.24	0.23	0.18	0.056	0.04	0.27	0.035	0.0061	0.072	0.097	0.11	0.0097	0.0026
garbage_chute	0.023	0.23	0.029	0.017	0.1	0.028	0.018	0.012	0.015	0.0079	0.098	0.048	0.032	0.03	0.0071	0.041	0.065	0.018
heating	0.0068	0.14	0.087	0.13	0.17	0.14	0.012	0.074	0.021	0.011	0.1	0.02	0.029	0.0096	0.057	0.08	0.049	0.0093
id	seller	price	area_total	area_kitchen	area_living	floor	rooms	layout	ceiling	hrooms_shared	hrooms_private	windows_court	windows_street	balconies	loggias	condition	phones	