

Reducing Uncertainty in Residential Real Estate

Estimating Russian Home Prices Through Regression and Variable Selection

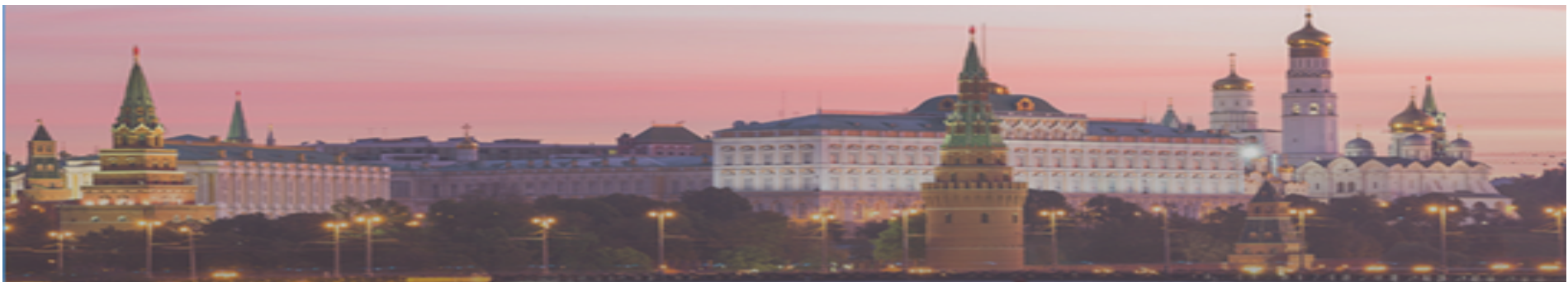
Allen Crane

Brock Friedrich

MSDS 6372 Applied Statistics - Modeling and Inference

Dr. Anthony Tanaydin - Fall 2018

2018-10-07



Introduction

For our project, we have been asked to assist the Russian bank Sberbank to predict real estate prices with more certainty, so they can provide more certainty to their customers and value to their shareholders. The goal of this analysis is to build 2 models for the prediction of Russian real estate property. One model will predict the individual property price given some or all of the variables, and another model to predict the mean price of the properties. Both models will use the TRAIN data set to build the model and the PREDICTION data set for the test model.

Objective #1 : Predicting Home Value with Regression

The TRAIN data set was used for this modeling effort. The original file consists of 30,471 observations with 292 variables each. Appendix A contains the complete data dictionary and explanation of variables.

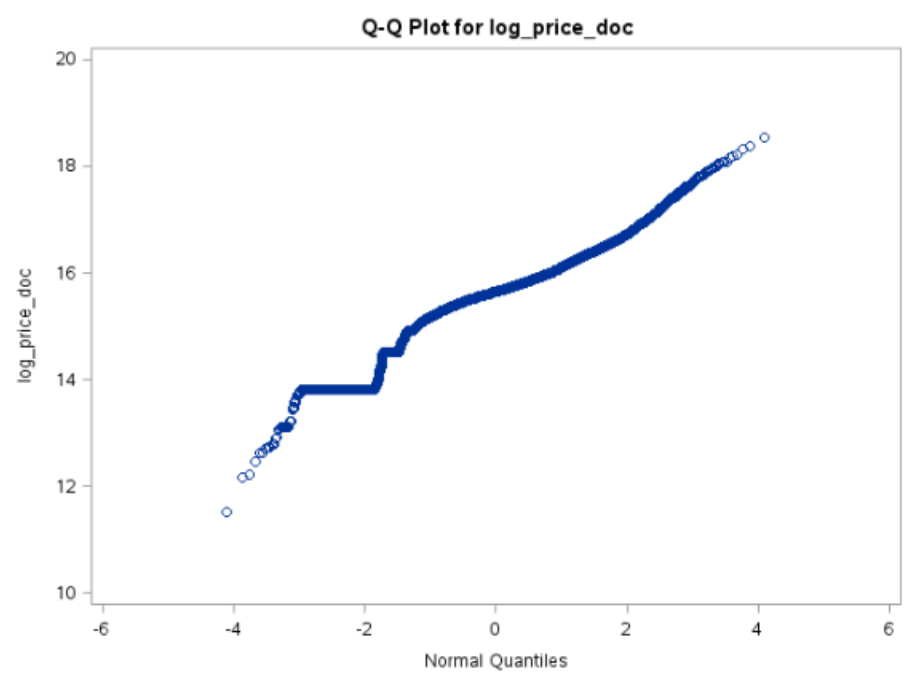
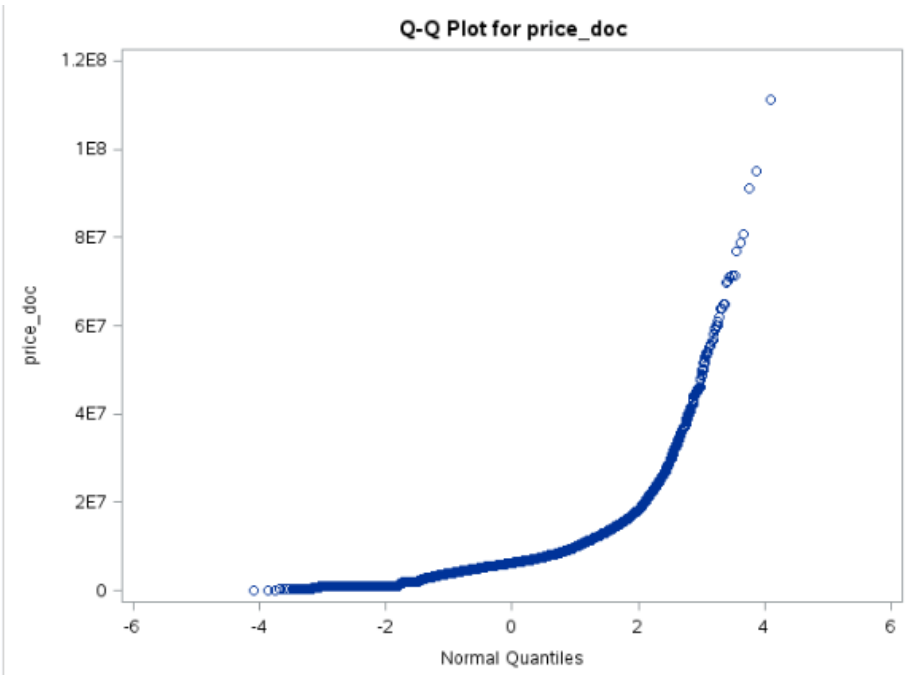
Data Wrangling

The following variable names exceeded 32 characters, and therefore needed to be trimmed to shorter variable names:

Before	After
preschool_education_centers_raion	ps_educ_centers_raion
school_education_centers_top_20_raion	s_educ_centers_top20_raion
raion_build_count_with_material_info	raion_bld_cnt_material_info
public_transport_station_min_walk	pub_trans_stn_min_walk

Some variables exhibited significant curvature in the QQ plot, so the following variables were created using a log transformation on the original data:

Before	After
area_m	log_area_m = log(area_m)
price_doc	log_price_doc=log(price_doc)



In SAS, variables cannot begin with numeric values. As a result, for those of us coding in SAS, we needed rename the variables to begin with non-numeric values. Using “group_” to proceed the variable names, we have the following:

Before	After
0_6_all	group_0_6_all
0_6_male	group_0_6_male
0_6_female	group_0_6_female
7_14_all	group_7_14_all
7_14_male	group_7_14_male
7_14_female	group_7_14_female
0_17_all	group_0_17_all
0_17_male	group_0_17_male
0_17_female	group_0_17_female
16_29_all	group_16_29_all
16_29_male	group_16_29_male
16_29_female	group_16_29_female
0_13_all	group_0_13_all
0_13_male	group_0_13_male
0_13_female	group_0_13_female

In SAS, variables cannot contain dashes/hyphens (“-”). As a result, for those of us coding in SAS, we needed rename the variables not include dashes/hyphens (“-”). Using an underscore (“_”) to replace (“-”), we have the following:

Before	After
build_count_1921-1945	build_count_1921_1945
'build_count_1946-1970	build_count_1946_1970
'build_count_1971-1995	build_count_1971_1995

Several numeric variables--assumed to be candidates for regression analysis--contained “NA” values. This caused them to be considered by SAS as character variables when the data was loaded into a SAS data set. There is probably a way to do the following in one step, but three steps were used to do this data processing step:

Before	After
build_year	n_build_year
kitch_sq	n_kitch_sq
material	n_material
max_floor	n_max_floor
num_room	n_num_room
state	n_state

Then we drop the original character variables:

Before	After
Drop build_year	
Drop kitch_sq	
Drop material	
Drop max_floor	
Drop num_room	
Drop state	

Then we finally rename the “n_*” variables back to the original names:

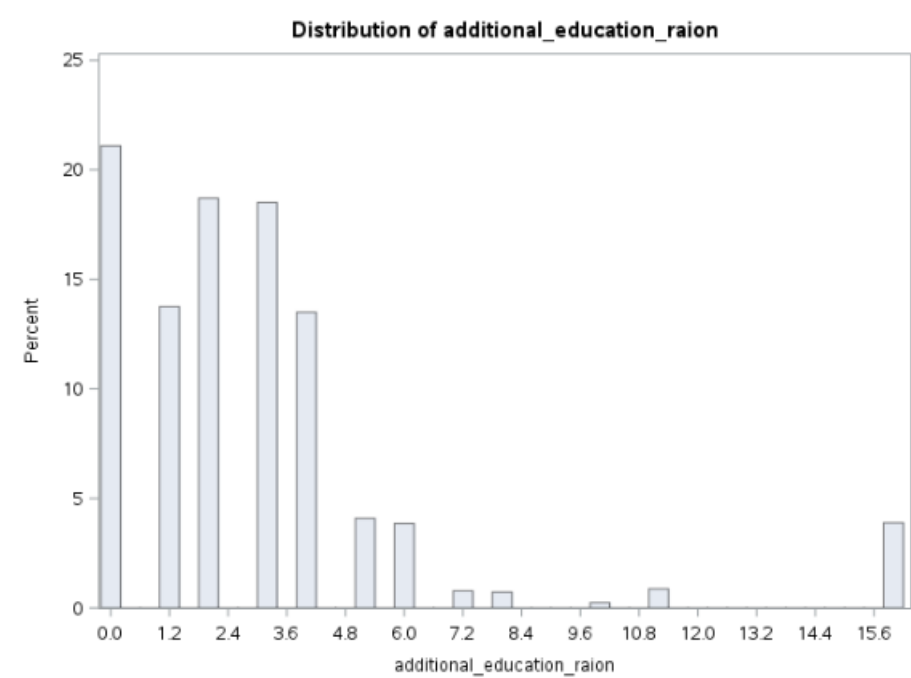
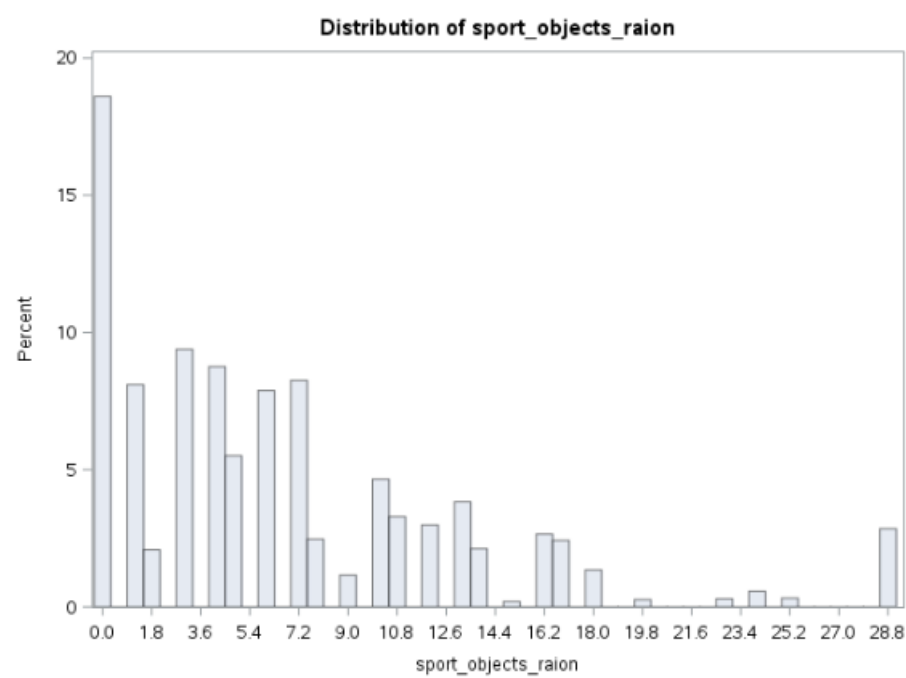
Before	After
Rename	n_build_year build_year
Rename	n_kitch_sq kitch_sq
Rename	n_ material material
Rename	n_max_floor max_floor
Rename	n_num_room num_room
Rename	n_state state

Exploratory Data Analysis

After some initial univariate testing, we began to see some patterns emerge. For example, several variables that end in “raion” (Russian for “neighborhood”) had the same pattern. These are all countable variables. Note how there are many zero values, followed by an approximate normal distribution, with a notable increase at the far right side of these graphs:

Outlier Identification and Handling

The same is true for many of the variables ending in “_raion” (“neighborhood”, in Russian). Further investigation shows that these neighborhoods (a.k.a. “sub_area” in our data) correspond to affluent and government sections of greater Moscow, Presneskoe (Presnensky) and Tversekoe (Tverskoy). These data show that there are some sub_areas that are culturally and administratively dense enough to legitimize these data findings. Therefore, we will not omit these variables due to simply being outliers, else we may risk losing material information about their predictive power for property values in certain areas.



Presnensky District> is an upscale area known for the State Museum of Oriental Art and quiet parks like Patriarch’s Ponds. Child-friendly attractions include the Moscow Zoo, and hands-on exhibits at the nearby Moscow Planetarium. A variety of restaurants serving European and fusion fare line Bolshaya Nikitskaya Street, which is also home to classic repertory at the Mayakovsky Theater.

Tverskoy District> is a district of Central Administrative Okrug of the federal city of Moscow, Russia. Population: 75,378; 75,955. The district extends from Kitai-gorod northwest to Belorussky and Savyolovsky Rail Terminals.



Presnensky District



Tverskoy District

Missing Data and Imputation

For our analysis, several numeric variables contained non-trivial percentages of missing data (percent of the data with “NA” listed as the value). Since the variable with the highest missing value rate (hospital_beds_raion) is 47%, throwing out the observations with missing values would limit us to just over half of the data set. Similarly, removing the entire variable(s) with missing values could cause us to miss significant coefficients for our predictive models. The following is a list of all variables in the TRAIN data set with missing values, and the missing value percentage of all observations for each variable)

Variable	Missing %
hospital_beds_raion	47%
build_year	45%
state	44%
cafe_sum_500_min_price_avg	44%
cafe_sum_500_max_price_avg	44%
cafe_avg_price_500	44%
max_floor	31%
material	31%
num_room	31%
kitch_sq	31%
preschool_quota	22%
school_quota	22%
cafe_sum_1000_min_price_avg	21%

According to Schlomer, Bauman, and Card, “This method is one of several maximum likelihood (ML) approaches. In all ML strategies, the missing values are assumed to be missing at random (MAR).”

observed data are used to estimate parameters, which are then used to estimate the missing scores. These ML strategies have demonstrated superiority to deletion, nonstochastic imputation, and stochastic regression imputation methods (Roth, 1994) for multivariate normal distributions. The EM strategy is based on a recursive process: The missing data have information that is useful in estimating various parameters, and the estimated parameter has information that is useful in finding the most likely value of the missing data (Bennett, 2001). Thus, the EM method is an iterative procedure with two steps in each iteration: In the expectation step, the process is similar to the regression-based imputation. First, starting values for the parameters (e.g., means, covariances) are obtained with available data. Regression methods are used to impute, on the basis of these initial values, the values for the missing data. When this step is completed, in the maximization step new values for the parameters are calculated with the newly imputed data along with the original observed data. Then the process starts over with the expectation step and continues until the estimates change very little from one iteration to the next (i.e., until the estimates converge; Allison, 2001).”

(Source: Gabriel L. Schlomer, Sheri Bauman, and Noel A. Card, “Best Practices for Missing Data Management in Counseling Psychology”, *Journal of Counseling Psychology* 2010, Vol. 57, No. 1, 1-10)

Below is the generic code for the PROC MI command: (see SAS code in Appendix for detail)

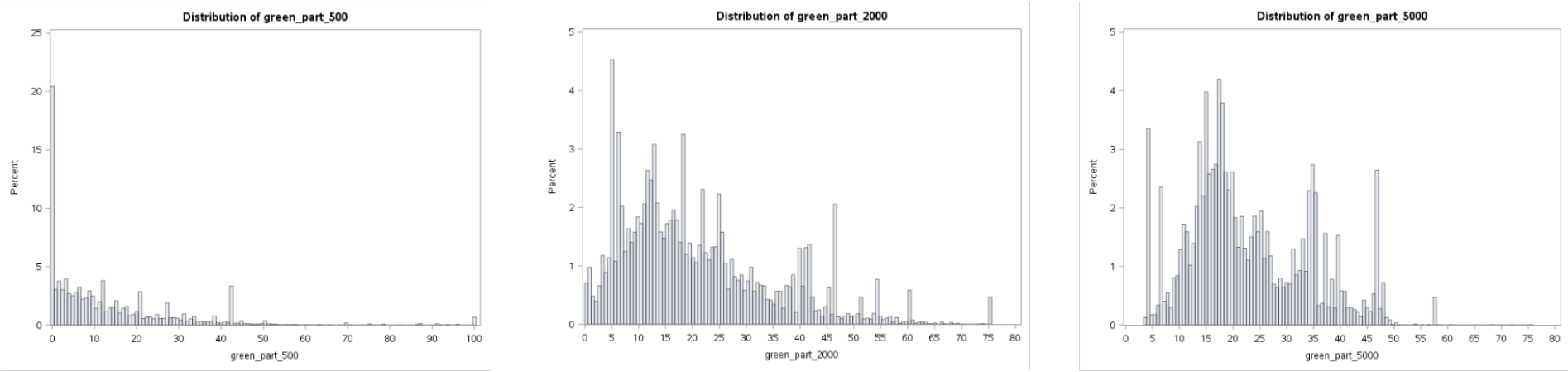
```
proc mi data=train2 seed=501213
mu0=0 out=train4;
    var n_build_year
...
n_prom_part_5000;
run;
```

cafe_sum_1000_max_price_avg	21%
cafe_avg_price_1000	21%
life_sq	21%
raion_build_count_with_material_info	16%
build_count_block	16%
build_count_wood	16%
build_count_frame	16%
build_count_brick	16%
build_count_monolith	16%
build_count_panel	16%
build_count_foam	16%
build_count_slag	16%
build_count_mix	16%
raion_build_count_with_builddate_info	16%
build_count_before_1920	16%
build_count_1921-1945	16%
build_count_1946-1970	16%
build_count_1971-1995	16%
build_count_after_1995	16%
cafe_sum_2000_min_price_avg	6%
cafe_sum_2000_max_price_avg	6%
cafe_avg_price_2000	6%
cafe_sum_3000_min_price_avg	3%
cafe_sum_3000_max_price_avg	3%
cafe_avg_price_3000	3%
cafe_sum_5000_min_price_avg	1%
cafe_sum_5000_max_price_avg	1%
cafe_avg_price_5000	1%
prom_part_5000	1%
floor	1%

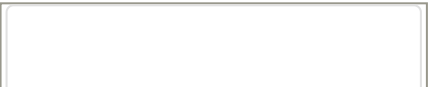
Missing Data Patterns

The PROC MI procedure outputs a list of the missing data patterns for our dataset. We can see in the example below that these data are neither MCAR (Missing Completely At Random) or MAR (Missing At Random). As you can see in the below, our data has “holes” in it that are missing in entire blocks, often with adjacent/similar variables. Since predicted values cannot be computed if a model coefficient (explanatory variable) has missing values, we found out after significant trial and error that it is important to impute as many missing values as possible, so that there will be less of a chance of having a missing predicted value. **Missing value patterns in the pre-imputation data set, showing Not Missing At Random patterns.**

In looking at the data, we can see several groups of variables that appear similar. Many of these are in the form of a distance or countable measure between the property (measured in increments of 500, 1000, 1500, 2000, 3000, and 5000 meters) and some item of significance (distance to a cafe with a certain min/max/average price, the number of large churches within an area, distance to leisure areas and markets, etc). An example of one series (“green_part_X”) is shown below:



To address these, we used the TOLERANCE option in the PROC GLM procedure. Tolerance is a measure of multicollinearity (tolerance = 1/VIF, where VIF is the Variance Inflation Factor). After the



model ran, we eliminated the variables that had tolerance levels less than 0.1 (which corresponds to a VIF of 10 or greater, thus implying multicollinearity). We then removed these variables and ran the reduced model below. Additionally, we used the White test for homoscedasticity, which is represented generically below. A value of $> \alpha = 0.05$ indicates homoscedasticity, and therefore no transformation is needed. Alternatively, a value of $< \alpha = 0.05$ indicates heteroscedasticity, which would indicate that some form of transformation is needed.

```
proc model data=data;
  parms b0 b1;
  y = b0 + b1* x;
  fit y/white;
Run;
```

Modeling

1. A model with OLS parameter estimates. You may choose the variables with or without the use of a variable selection technique (forward, backward, stepwise) (You may have done this already in the EDA.)

- See Appendix 3: Model 1.1 OLS detail

2. A model with LASSO estimation and selection.

- See Appendix 3: Model 1.2 LASSO detail

3. A model of your choice. This may be using another OLS or LASSO model or custom model, etc.

- See Appendix 3: Model 1.2 LARS detail

Model	R2	AIC	SBC
OLS (GLM)	0.4207		
GLMSELECT (AIC)	0.4207	23778403	23019085
LASSO	0.3325	23886070	23124732
LAR	0.3325	23886070	23124732

Prediction

A. Identify which model you feel is the best and discuss why.

Based on the above statistics, we elected to use the reduced GLM SELECT model that optimizes for AIC (/selection=stepwise(select=SL SLE=0.1 SLS=0.08 choose=AIC)). We chose this model because it achieved the highest R2 value, with the fewest number of variables. Hence it was both accurate and more parsimonious than the initial model.

B. Use this model to predict the price of each of the 7662 properties in the prediction data set (predictionData.csv).

See data set named “results_final.csv”. Note: due to PROC MI, we have many thousand more observations than we were asked to produce for the test/prediction data.

C. Submit these predictions in the zip file with your final paper. Please see the sample submission file. SampleSubmission.csv

See data set named “results_final.csv”. Note: due to PROC MI, we have many thousand more observations than we were asked to produce for the test/prediction data.

Source code for Objective #1 can be found in the /src/ folder, titled "Project - 1 *.*.sas"

Objective #2 : Forecasting Mean Home Price using Time Series Analysis

The housing market in Moscow, Russia has seen relative stability in recent years. After crashing in Q4 of 2012, home prices rallied to show consistent improvement, averaging +2.12% (apx. \$19 K USD) month-over-month between Q1 2013 and Q1 2015. Despite the recent historical stability, complex interactions between a large number of housing characteristics add complexity to the task of predicting future prices. As such, the central bank of Russia, Sberbank, has tasked us with shedding some light on the uncertainty that convolutes such broad questions.

Our analysis will examine the forecasting of mean residential home prices from July 2015 to July 2016, emphasising the identification and remediation of serial effects.

Data Exploration

Wrangling the dataset to prepare it for analysis was a straitforward effort. Using Python, the data was read into a Pandas dataframe. Once in the dataframe, it was just a matter of applying the timestamp as the dataframe's index, then resampling the dataframe to the desired frequency. In this case, the data is downsampled from it's original frequency of days ('D') to months ('M'), and calculating the mean for each month.

```
cols = ['timestamp'
        , 'year'
        , 'month'
        , 'day'
        , 'price_doc']
subtrain = train[cols].copy(deep = True)
subtrain = subtrain.set_index('timestamp')
subtrain = subtrain.resample('M').mean()
subtrain = subtrain.sort_index()
subtrain.describe()
```

price_log	year	2011	2015
	count	753	3239
	mean	15.37	15.75
	std	0.75	0.58
	min	12.74	13.12
	25%	15.27	15.49
	50%	15.52	15.78
	75%	15.78	16.07
	max	17.43	18.37

Additional tables in appendix

Now that the data is aggregated into regularized, monthly intervals, the next step is to add an integer sequence to numerically represent the timestamps in a form that will be readily compatable with our regression model.

```
train['yearmonth'] = train.timestamp.apply(lambda x: ('{:04d}{:02d}'.format(x.year, x.month)))
```

Vizualizing the timeseries reveals the short-term relationship between price and number of months. The localized positive trend after month 18 is evident. A variety of factors could contribute to explaining the apparent trend, however, full explanation of the relationships between housing features is outside of the scope of this study. We wish only to investigate and quantify potential serial effects that may be present.



Assumptions

- Normality

The real estate prices do not appear to be normally distributed. A significant right skew is apparent in both the histogram and boxplot of Price. Such a large deviation from normality would likely have a detrimental effect on our regression equation, as well as biasing any inferences, due to the substantial effect that variability has on residuals

Homogeneity of Variance

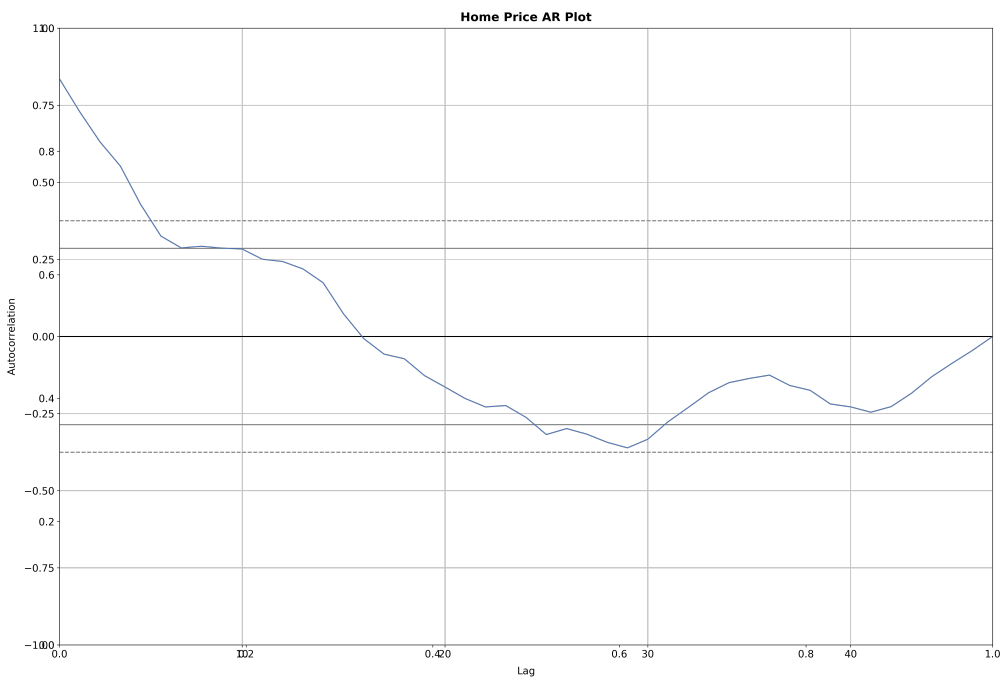
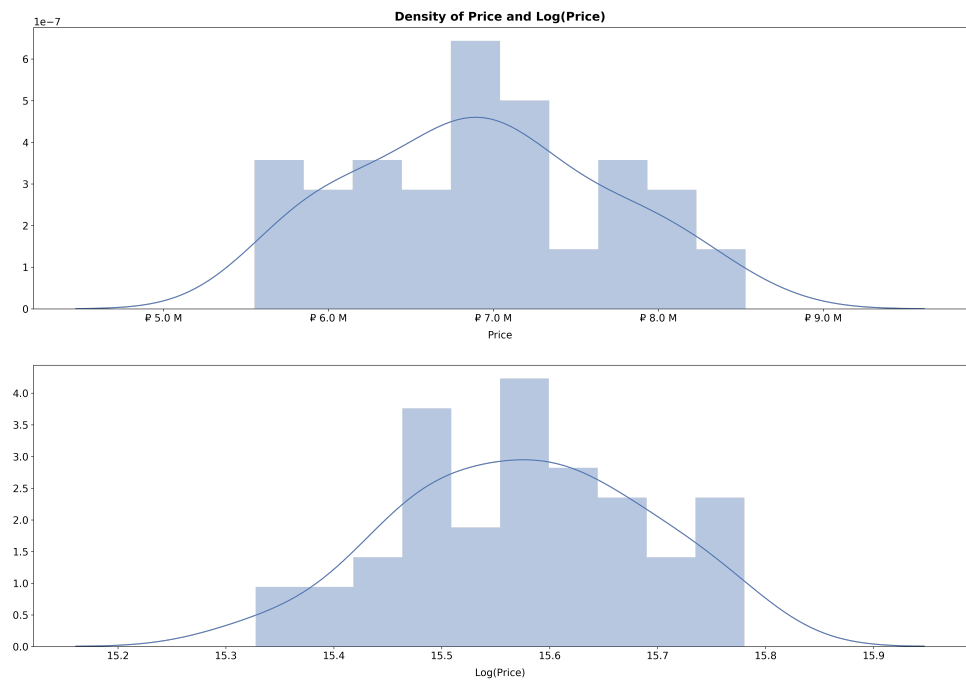
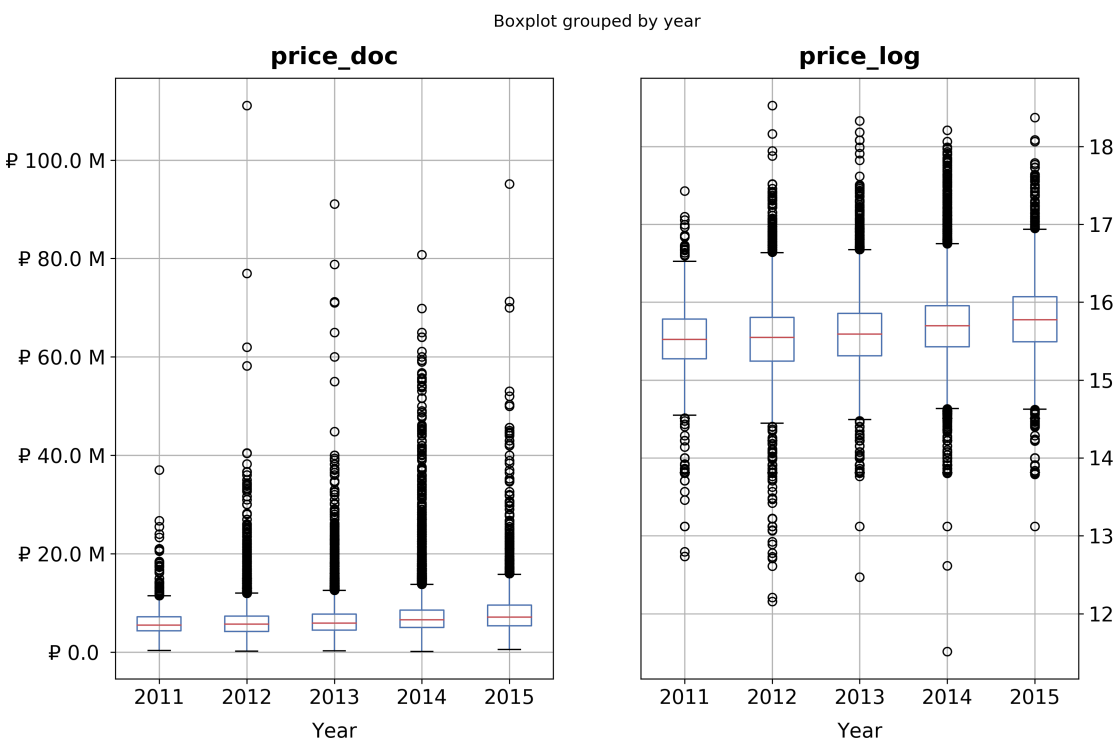
Superficially, the variation across the dataset appears consistent enough to not warrant much concern. There is moderate cause to watchful for a homoscedastic trend as the date increases, but futher insights will have to wait until we analyze the residuals.

Mitigation

We cannot yet make determinations as to the state of the price feature's observational independence or linearity. Mitigating the possibility of bias introduced through unmet distributional assumptions would undoubtedly aid our ability to identify any serial effects, however, we will proceed with caution and examine the price feature on its original scale.

AR Plot

As suspected, an autoregressive plot firmly indicates potential violations of normaliry for lag values < 10, specifically near lag=1. This gives us a good starting point to begin investigating the roots of this potential serial correlation. This cursory AR plot will give us an idea of where to begin our analysis



Initial Regression Model

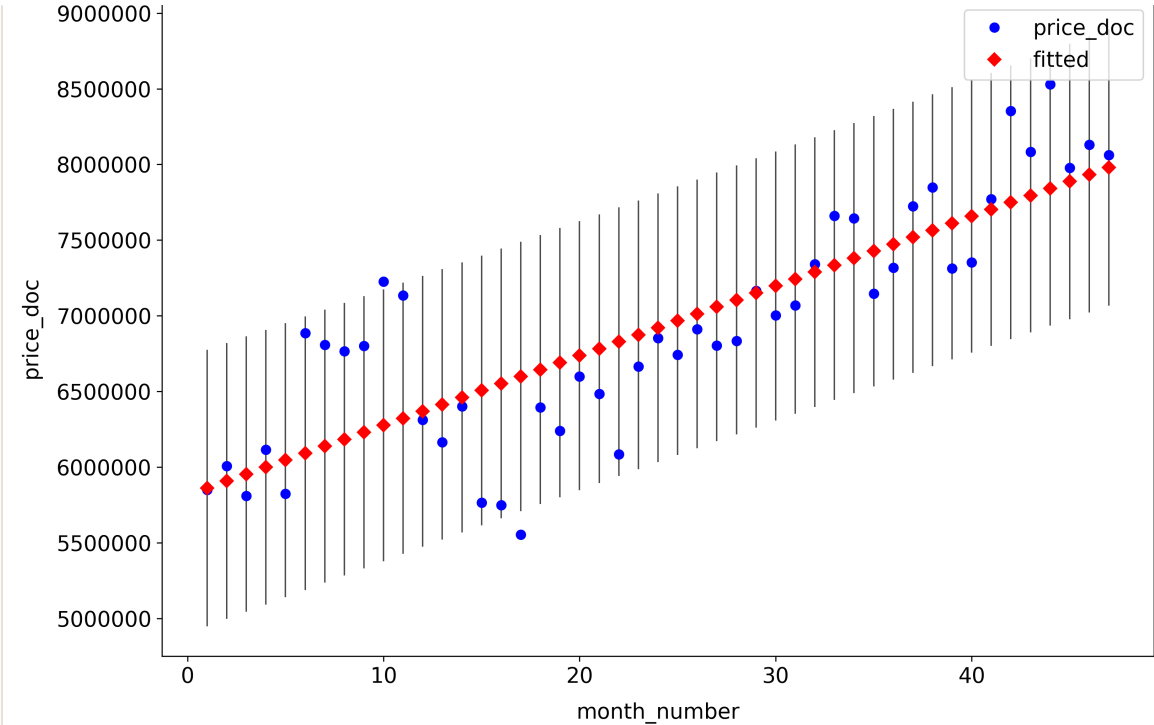
Fitting the data to a simple linear regression using Ordinary Least Squares (OLS) will provide us with an initial set of residuals to examine for temporal bias. We will use the most reduced model available, as it will provide the clearest path to interpretation:

$$Y = \beta_0 + \beta_1 * \log(price) + \mu \text{ where } \mu \sim N(0, \Sigma)$$



OLS Model

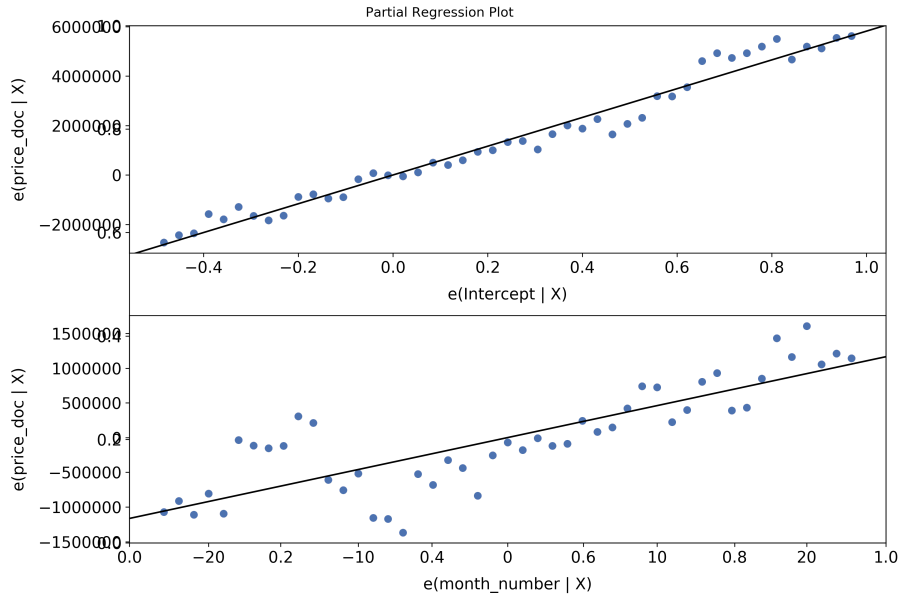
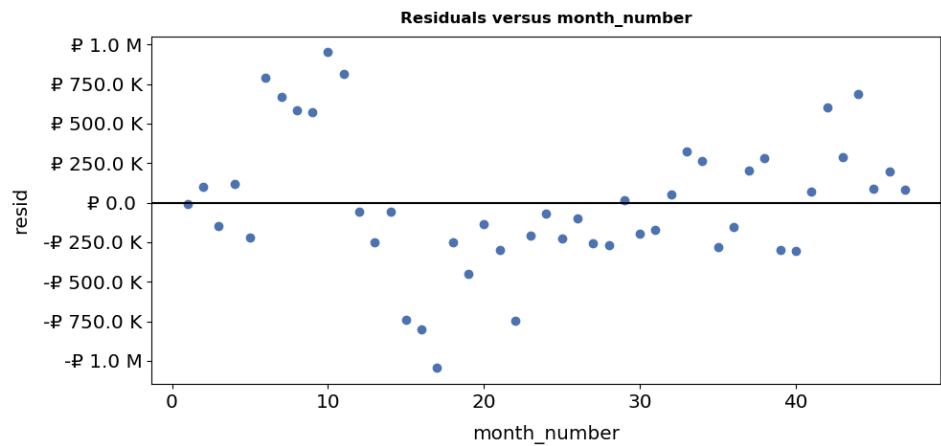
The first order OLS model produced a Durbin-Watson score of 0.736. Durbin-Watson scores < 2 indicate the presence of a positive autocorrelation. The score produced by this model is in alignment with the AR plot produced previously. A residual plot (below) supports this score, showing distinct lack of randomness across the residuals.



Fit statistics				
Durbin-Watson:	0.736	AIC:	1356	
Adj. R-squared:	0.675	BIC:	1359	
No. Observations:	47	Df Model:	1	
F-statistic:	96.35	Df Residuals:	45	
P > F	< 0.001	R-squared:	0.682	
Estimates				
Coef.	Estimate	Std.Err.	t	P> t
Intercept	5815308	129324	44.97	< 0.001
month_number	46046	4691	9.82	< 0.001

Additional diagnostics in appendix.

Residuals and Partial



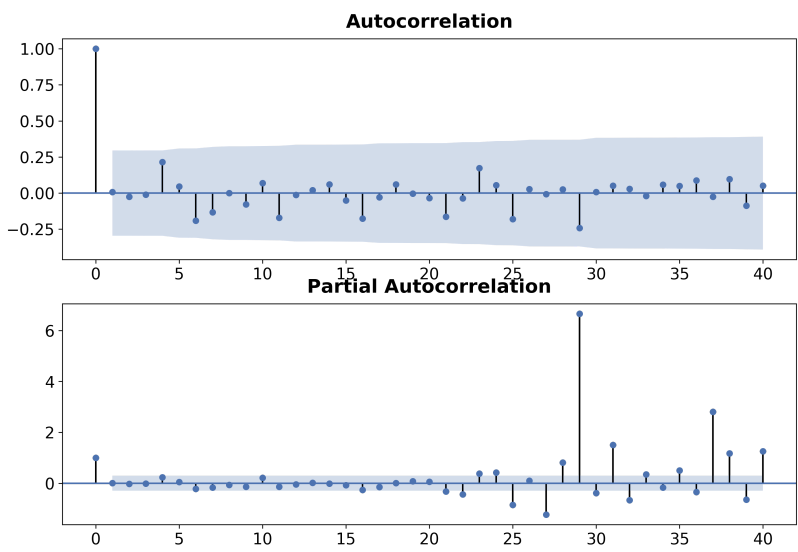
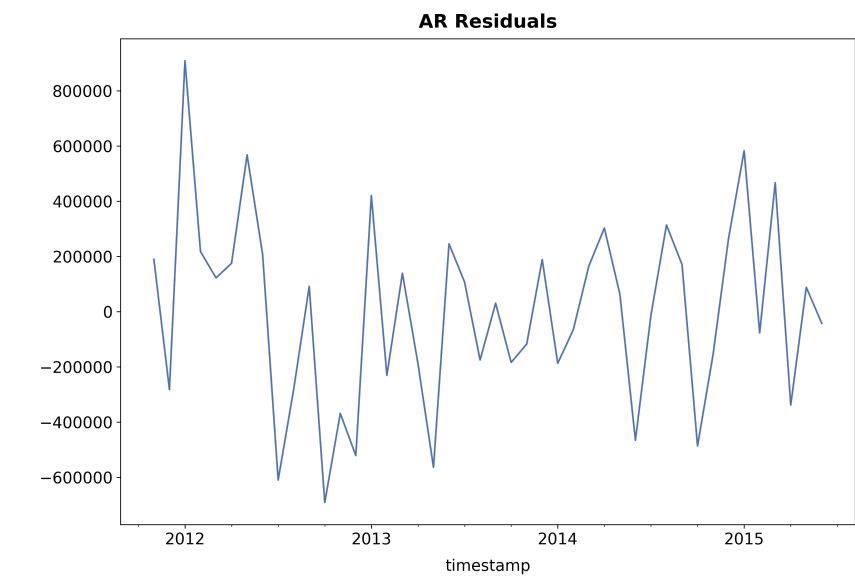
Predicting Mean Home Price

AR Model Summary

Information Criterion		Estimate	t	p > t	Lower CI	Upper CI	SE	
AIC	25.70	const	4014.56	0.07	0.94	-101091.43	109120.55	53626.49

BIC	25.90
Durbin-Watson	1.98

L1.resid	0.59	3.73	< 0.001	0.28	0.90	0.16
L2.resid	0.08	0.43	0.67	-0.28	0.43	0.18
L3.resid	-0.00	-0.02	0.98	-0.31	0.31	0.16



Predictions

A CSV file of the prediction data can be found in the /docs/ folder under the name p2preds.csv.

mean	mean_se	mean_ci_lower	mean_ci_upper	obs_ci_lower	obs_ci_upper
7979477	125261	7727187	8231767	7065329	8893624
8025523	129324	7765050	8285995	7109083	8941962
8071569	133428	7802831	8340307	7152746	8990392
8117615	137569	7840536	8394695	7196318	9038912
8163661	141745	7878171	8449151	7239800	9087522
8209707	145952	7915744	8503670	7283193	9136222
8255753	150188	7953259	8558248	7326497	9185010
8301800	154450	7990721	8612878	7369713	9233886
8347846	158736	8028135	8667557	7412843	9282849
8393892	163044	8065503	8722281	7455886	9331897
8439938	167373	8102830	8777046	7498844	9381032
8485984	171721	8140118	8831850	7541718	9430250

Source code for Obective 2 can be found in the problem2.py.

Appendix: Feature Summary

price_doc	sale price (this is the target variable)
id	transaction id
timestamp	date of transaction
full_sq	total area in square meters, including loggias, balconies and other non-residential areas
life_sq	living area in square meters, excluding loggias, balconies and other non-residential areas
floor	for apartments, floor of the building
max_floor	number of floors in the building
material	wall material
build_year	year built
num_room	number of living rooms
kitch_sq	kitchen area
state	apartment condition
product_type	owner-occupier purchase or investment
sub_area	name of the district

The dataset also includes a collection of features about each property’s surrounding neighbourhood, and some features that are constant across each sub area (known as a Raion). Most of the feature names are self explanatory, with the following notes. See below for a complete list.

full_all	subarea population
male_f, female_f	subarea population by gender
young_*	population younger than working age
work_*	working-age population
ekder_*	retirement-age population
n_m_{all	male
build_count_*	buildings in the subarea by construction type or year
x_count_500	the number of x within 500m of the property
x_part_500	the share of x within 500m of the property
sqm	square meters
cafe_count_d_price_p	number of cafes within d meters of the property that have an average bill under p RUB
trc_	shopping malls
prom_	industrial zones
green_	green zones
metro_	subway
avto	distances by car
mkad_	Moscow Circle Auto Road
ttk_	Third Transport Ring

full_all	subarea population
sadovoe_	Garden Ring
bulvar_ring_	Boulevard Ring
kremlin_	City center
zd_vokzaly_	Train station
oil_chemistry_	Dirty industry
ts_	Power plant

Complete description of neighbourhood features

Feature	Description
area_m	Area mun. area, sq.m.
raion_popul	Number of municipality population. district
green_zone_part	Proportion of area of ♂♂greenery in the total area
indust_part	Share of industrial zones in area of ♂♂the total area
children_preschool	Number of pre-school age population
preschool_quota	Number of seats in pre-school organizations
preschool_education_centers_raion	Number of pre-school institutions
children_school	Population of school-age children
school_quota	Number of high school seats in area
school_education_centers_raion	Number of high school institutions
school_education_centers_top_20_raion	Number of high schools of the top 20 best schools in Moscow
hospital_beds_raion	Number of hospital beds for the district
healthcare_centers_raion	Number of healthcare centers in district
university_top_20_raion	Number of higher education institutions in the top ten ranking of the Federal rank
sport_objects_raion	Number of higher education institutions
additional_education_raion	Number of additional education organizations
culture_objects_top_25	Presence of the key objects of cultural heritage (significant objects for the level of the RF constituent entities, city)
culture_objects_top_25_raion	Number of objects of cultural heritage
shopping_centers_raion	Number of malls and shopping centres in district
office_raion	Number of malls and shopping centres in district
thermal_power_plant_raion	Presence of thermal power station in district
incineration_raion	Presence of incinerators
oil_chemistry_raion	Presence of dirty industries
radiation_raion	Presence of radioactive waste disposal
railroad_terminal_raion	Presence of the railroad terminal in district
big_market_raion	Presence of large grocery / wholesale markets
nuclear_reactor_raion	Presence of existing nuclear reactors
detention_facility_raion	Presence of detention centers, prisons
full_all	Total number of population in the municipality
male_f	Male population
female_f	Female population
young_all	Population younger than working age
young_male	Male population younger than working age
young_female	Feale population younger than working age
work_all	Working-age population
work_male	Male working-age population
work_female	Female working-age population

Feature	Description
ekder_all	Population older than working age
ekder_male	Male population older than working age
ekder_female	Female population older than working age
0_6_all	Population aged 0-6
0_6_male	Male population aged 0-7
0_6_female	Female population aged 0-8
7_14_all	Population aged 7-14
7_14_male	Male population aged 7-14
7_14_female	Female population aged 7-14
0_17_all	Population aged 0-17
0_17_male	Male population aged 0-17
0_17_female	Female population aged 0-17
16_29_all	Population aged 16-19
16_29_male	Male population aged 16-19
16_29_female	Female population aged 16-19
0_13_all	Population aged 0-13
0_13_male	Male population aged 0-13
0_13_female	Female population aged 0-13
raion_build_count_with_material_info	Number of building with material info in district
build_count_block	Share of block buildings
build_count_wood	Share of wood buildings
build_count_frame	Share of frame buildings
build_count_brick	Share of brick buildings
build_count_monolith	Share of monolith buildings
build_count_panel	Share of panel buildings
build_count_foam	Share of foam buildings
build_count_slag	Share of slag buildings
build_count_mix	Share of mixed buildings
raion_build_count_with_builddate_info	Number of building with build year info in district
build_count_before_1920	Share of before_1920 buildings
build_count_1921-1945	Share of 1921-1945 buildings
build_count_1946-1970	Share of 1946-1970 buildings
build_count_1971-1995	Share of 1971-1995 buildings
build_count_after_1995	Share of after_1995 buildings
7_14_male	Male population aged 7-14
7_14_female	Female population aged 7-14
0_17_all	Population aged 0-17
0_17_male	Male population aged 0-17
0_17_female	Female population aged 0-17
16_29_all	Population aged 16-19
16_29_male	Male population aged 16-19
16_29_female	Female population aged 16-19
0_13_all	Population aged 0-13
0_13_male	Male population aged 0-13
0_13_female	Female population aged 0-13
metro_min_avto	Time to subway by car, min.
metro_km_avto	Distance to subway by car, km
metro_min_walk	Time to metro by foot
metro_km_walk	Distance to the metro, km
kindergarten_km	Distance to kindergarten
school_km	Distance to high school

Feature	Description
park_km	Distance to park
green_zone_km	Distance to green zone
industrial_zone_km	Distance to industrial zone
water_treatment_km	Distance to water treatment
cemetery_km	Distance to the cemetery
incineration_km	Distance to the incineration
railroad_station_walk_km	Distance to the railroad station (walk)
railroad_station_walk_min	Time to the railroad station (walk)
ID_railroad_station_walk	Nearest railroad station id (walk)
railroad_station_avto_km	Distance to the railroad station (avto)
railroad_station_avto_min	Time to the railroad station (avto)
ID_railroad_station_avto	Nearest railroad station id (avto)
public_transport_station_km	Distance to the public transport station (walk)
public_transport_station_min_walk	Time to the public transport station (walk)
water_km	Distance to the water reservoir / river
water_1line	First line to the river (150 m)
mkad_km	Distance to MKAD (Moscow Circle Auto Road)
ttk_km	Distance to the TTC (Third Transport Ring)
sadovoe_km	Distance to the Garden Ring
bulvar_ring_km	The distance to the Boulevard Ring
kremlin_km	Distance to the city center (Kremlin)
big_road1_km	Distance to Nearest major road
ID_big_road1	Nearest big road id
big_road1_1line	First line to the road (100 m for highwys, 250 m to MKAD)
big_road2_km	The distance to next distant major road
ID_big_road2	2nd nearest big road id
railroad_km	Distance to the railway / Moscow Central Ring / open areas Underground
railroad_1line	First line to the railway (100 m)
zd_vokzaly_avto_km	Distance to train station
ID_railroad_terminal	Nearest railroad terminal id
bus_terminal_avto_km	Distance to bus terminal (avto)
ID_bus_terminal	Nearest bus terminal id
oil_chemistry_km	Distance to dirty industries
nuclear_reactor_km	Distance to nuclear reactor
radiation_km	Distance to burial of radioactive waste
power_transmission_line_km	Distance to power transmission line
thermal_power_plant_km	Distance to thermal power plant
ts_km	Distance to power station
big_market_km	Distance to grocery / wholesale markets
market_shop_km	Distance to markets and department stores
fitness_km	Distance to fitness
swim_pool_km	Distance to swimming pool
ice_rink_km	Distance to ice palace
stadium_km	Distance to stadium
basketball_km	Distance to the basketball courts
hospice_morgue_km	Distance to hospice/morgue
detention_facility_km	Distance to detention facility
public_healthcare_km	Distance to public healthcare
university_km	Distance to universities
workplaces_km	Distance to workplaces
shopping_centers_km	Distance to shopping centers

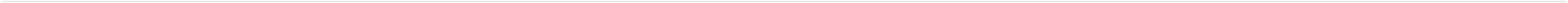
Feature	Description
office_km	Distance to business centers/ offices
additional_education_km	Distance to additional education
preschool_km	Distance to preschool education organizations
big_church_km	Distance to large church
church_synagogue_km	Distance to Christian chirches and Synagogues
mosque_km	Distance to mosques
theater_km	Distance to theater
museum_km	Distance to museums
exhibition_km	Distance to exhibition
catering_km	Distance to catering
ecology	Ecological zone where the house is located
green_part_500	The share of green zones in 500 meters zone
prom_part_500	The share of industrial zones in 500 meters zone
office_count_500	The number of office space in 500 meters zone
office_sqm_500	The square of office space in 500 meters zone
trc_count_500	The number of shopping malls in 500 meters zone
trc_sqm_500	The square of shopping malls in 500 meters zone
cafe_count_500	The number of cafes or restaurants in 500 meters zone
cafe_sum_500_min_price_avg	Cafes and restaurant min average bill in 500 meters zone
cafe_sum_500_max_price_avg	Cafes and restaurant max average bill in 500 meters zone
cafe_avg_price_500	Cafes and restaurant average bill in 500 meters zone
cafe_count_500_na_price	Cafes and restaurant bill N/A in 500 meters zone
cafe_count_500_price_500	Cafes and restaurant bill, average under 500 in 500 meters zone
cafe_count_500_price_1000	Cafes and restaurant bill, average 500-1000 in 500 meters zone
cafe_count_500_price_1500	Cafes and restaurant bill, average 1000-1500 in 500 meters zone
cafe_count_500_price_2500	Cafes and restaurant bill, average 1500-2500 in 500 meters zone
cafe_count_500_price_4000	Cafes and restaurant bill, average 2500-4000 in 500 meters zone
cafe_count_500_price_high	Cafes and restaurant bill, average over 4000 in 500 meters zone
big_church_count_500	The number of big churches in 500 meters zone
church_count_500	The number of churches in 500 meters zone
mosque_count_500	The number of mosques in 500 meters zone
leisure_count_500	The number of leisure facilities in 500 meters zone
sport_count_500	The number of sport facilities in 500 meters zone
market_count_500	The number of markets in 500 meters zone
green_part_1000	The share of green zones in 1000 meters zone
prom_part_1000	The share of industrial zones in 1000 meters zone
office_count_1000	The number of office space in 1000 meters zone
office_sqm_1000	The square of office space in 1000 meters zone
trc_count_1000	The number of shopping malls in 1000 meters zone
trc_sqm_1000	The square of shopping malls in 1000 meters zone
cafe_count_1000	The number of cafes or restaurants in 1000 meters zone
cafe_sum_1000_min_price_avg	Cafes and restaurant min average bill in 1000 meters zone
cafe_sum_1000_max_price_avg	Cafes and restaurant max average bill in 1000 meters zone
cafe_avg_price_1000	Cafes and restaurant average bill in 1000 meters zone
cafe_count_1000_na_price	Cafes and restaurant bill N/A in 1000 meters zone
cafe_count_1000_price_500	Cafes and restaurant bill, average under 500 in 1000 meters zone
cafe_count_1000_price_1000	Cafes and restaurant bill, average 500-1000 in 1000 meters zone
cafe_count_1000_price_1500	Cafes and restaurant bill, average 1000-1500 in 1000 meters zone
cafe_count_1000_price_2500	Cafes and restaurant bill, average 1500-2500 in 1000 meters zone
cafe_count_1000_price_4000	Cafes and restaurant bill, average 2500-4000 in 1000 meters zone
cafe_count_1000_price_high	Cafes and restaurant bill, average over 4000 in 1000 meters zone

Feature	Description
big_church_count_1000	The number of big churches in 1000 meters zone
church_count_1000	The number of churches in 1000 meters zone
mosque_count_1000	The number of mosques in 1000 meters zone
leisure_count_1000	The number of leisure facilities in 1000 meters zone
sport_count_1000	The number of sport facilities in 1000 meters zone
market_count_1000	The number of markets in 1000 meters zone
green_part_1500	The share of green zones in 1500 meters zone
prom_part_1500	The share of industrial zones in 1500 meters zone
office_count_1500	The number of office space in 1500 meters zone
office_sqm_1500	The square of office space in 1500 meters zone
trc_count_1500	The number of shopping malls in 1500 meters zone
trc_sqm_1500	The square of shopping malls in 1500 meters zone
cafe_count_1500	The number of cafes or restaurants in 1500 meters zone
cafe_sum_1500_min_price_avg	Cafes and restaurant min average bill in 1500 meters zone
cafe_sum_1500_max_price_avg	Cafes and restaurant max average bill in 1500 meters zone
cafe_avg_price_1500	Cafes and restaurant average bill in 1500 meters zone
cafe_count_1500_na_price	Cafes and restaurant bill N/A in 1500 meters zone
cafe_count_1500_price_500	Cafes and restaurant bill, average under 500 in 1500 meters zone
cafe_count_1500_price_1000	Cafes and restaurant bill, average 500-1000 in 1500 meters zone
cafe_count_1500_price_1500	Cafes and restaurant bill, average 1000-1500 in 1500 meters zone
cafe_count_1500_price_2500	Cafes and restaurant bill, average 1500-2500 in 1500 meters zone
cafe_count_1500_price_4000	Cafes and restaurant bill, average 2500-4000 in 1500 meters zone
cafe_count_1500_price_high	Cafes and restaurant bill, average over 4000 in 1500 meters zone
big_church_count_1500	The number of big churches in 1500 meters zone
church_count_1500	The number of churches in 1500 meters zone
mosque_count_1500	The number of mosques in 1500 meters zone
leisure_count_1500	The number of leisure facilities in 1500 meters zone
sport_count_1500	The number of sport facilities in 1500 meters zone
market_count_1500	The number of markets in 1500 meters zone
green_part_2000	The share of green zones in 2000 meters zone
prom_part_2000	The share of industrial zones in 2000 meters zone
office_count_2000	The number of office space in 2000 meters zone
office_sqm_2000	The square of office space in 2000 meters zone
trc_count_2000	The number of shopping malls in 2000 meters zone
trc_sqm_2000	The square of shopping malls in 2000 meters zone
cafe_count_2000	The number of cafes or restaurants in 1500 meters zone
cafe_sum_2000_min_price_avg	Cafes and restaurant min average bill in 2000 meters zone
cafe_sum_2000_max_price_avg	Cafes and restaurant max average bill in 2000 meters zone
cafe_avg_price_2000	Cafes and restaurant average bill in 2000 meters zone
cafe_count_2000_na_price	Cafes and restaurant bill N/A in 2000 meters zone
cafe_count_2000_price_500	Cafes and restaurant bill, average under 500 in 2000 meters zone
cafe_count_2000_price_1000	Cafes and restaurant bill, average 500-1000 in 2000 meters zone
cafe_count_2000_price_1500	Cafes and restaurant bill, average 1000-1500 in 2000 meters zone
cafe_count_2000_price_2500	Cafes and restaurant bill, average 1500-2500 in 2000 meters zone
cafe_count_2000_price_4000	Cafes and restaurant bill, average 2500-4000 in 2000 meters zone
cafe_count_2000_price_high	Cafes and restaurant bill, average over 4000 in 2000 meters zone
big_church_count_2000	The number of big churches in 2000 meters zone
church_count_2000	The number of churches in 2000 meters zone
mosque_count_2000	The number of mosques in 2000 meters zone
leisure_count_2000	The number of leisure facilities in 2000 meters zone
sport_count_2000	The number of sport facilities in 2000 meters zone

Feature	Description
market_count_2000	The number of markets in 2000 meters zone
green_part_3000	The share of green zones in 3000 meters zone
prom_part_3000	The share of industrial zones in 3000 meters zone
office_count_3000	The number of office space in 3000 meters zone
office_sqm_3000	The square of office space in 3000 meters zone
trc_count_3000	The number of shopping malls in 3000 meters zone
trc_sqm_3000	The square of shopping malls in 3000 meters zone
cafe_count_3000	The number of cafes or restaurants in 1500 meters zone
cafe_sum_3000_min_price_avg	Cafes and restaurant min average bill in 3000 meters zone
cafe_sum_3000_max_price_avg	Cafes and restaurant max average bill in 3000 meters zone
cafe_avg_price_3000	Cafes and restaurant average bill in 3000 meters zone
cafe_count_3000_na_price	Cafes and restaurant bill N/A in 3000 meters zone
cafe_count_3000_price_500	Cafes and restaurant bill, average under 500 in 3000 meters zone
cafe_count_3000_price_1000	Cafes and restaurant bill, average 500-1000 in 3000 meters zone
cafe_count_3000_price_1500	Cafes and restaurant bill, average 1000-1500 in 3000 meters zone
cafe_count_3000_price_2500	Cafes and restaurant bill, average 1500-2500 in 3000 meters zone
cafe_count_3000_price_4000	Cafes and restaurant bill, average 2500-4000 in 3000 meters zone
cafe_count_3000_price_high	Cafes and restaurant bill, average over 4000 in 3000 meters zone
big_church_count_3000	The number of big churches in 3000 meters zone
church_count_3000	The number of churches in 3000 meters zone
mosque_count_3000	The number of mosques in 3000 meters zone
leisure_count_3000	The number of leisure facilities in 3000 meters zone
sport_count_3000	The number of sport facilities in 3000 meters zone
market_count_3000	The number of markets in 3000 meters zone
green_part_5000	The share of green zones in 5000 meters zone
prom_part_5000	The share of industrial zones in 5000 meters zone
office_count_5000	The number of office space in 5000 meters zone
office_sqm_5000	The square of office space in 5000 meters zone
trc_count_5000	The number of shopping malls in 5000 meters zone
trc_sqm_5000	The square of shopping malls in 5000 meters zone
cafe_count_5000	The number of cafes or restaurants in 1500 meters zone
cafe_sum_5000_min_price_avg	Cafes and restaurant min average bill in 5000 meters zone
cafe_sum_5000_max_price_avg	Cafes and restaurant max average bill in 5000 meters zone
cafe_avg_price_5000	Cafes and restaurant average bill in 5000 meters zone
cafe_count_5000_na_price	Cafes and restaurant bill N/A in 5000 meters zone
cafe_count_5000_price_500	Cafes and restaurant bill, average under 500 in 5000 meters zone
cafe_count_5000_price_1000	Cafes and restaurant bill, average 500-1000 in 5000 meters zone
cafe_count_5000_price_1500	Cafes and restaurant bill, average 1000-1500 in 5000 meters zone
cafe_count_5000_price_2500	Cafes and restaurant bill, average 1500-2500 in 5000 meters zone
cafe_count_5000_price_4000	Cafes and restaurant bill, average 2500-4000 in 5000 meters zone
cafe_count_5000_price_high	Cafes and restaurant bill, average over 4000 in 5000 meters zone
big_church_count_5000	The number of big churches in 5000 meters zone
church_count_5000	The number of churches in 5000 meters zone
mosque_count_5000	The number of mosques in 5000 meters zone
leisure_count_5000	The number of leisure facilities in 5000 meters zone
sport_count_5000	The number of sport facilities in 5000 meters zone
market_count_5000	The number of markets in 5000 meters zone

- [train.csv and test.csv](#)

- [Complete description of neighbourhood features](#)



Appendix: Feature Summary

Feature Summary

	count	mean	std	min	25%	50%	75%	max
id	30471.00	15237.92	8796.50	1.00	7620.50	15238.00	22855.50	30473.00
full_sq	30471.00	54.21	38.03	0.00	38.00	49.00	63.00	5326.00
life_sq	24088.00	34.40	52.29	0.00	20.00	30.00	43.00	7478.00
floor	30304.00	7.67	5.32	0.00	3.00	6.50	11.00	77.00
max_floor	20899.00	12.56	6.76	0.00	9.00	12.00	17.00	117.00
material	20899.00	1.83	1.48	1.00	1.00	1.00	2.00	6.00
build_year	16866.00	3068.06	154387.80	0.00	1967.00	1979.00	2005.00	20052009.00
num_room	20899.00	1.91	0.85	0.00	1.00	2.00	2.00	19.00
kitch_sq	20899.00	6.40	28.27	0.00	1.00	6.00	9.00	2014.00
state	16912.00	2.11	0.88	1.00	1.00	2.00	3.00	33.00
area_m	30471.00	17657051.42	20649607.99	2081627.76	7307410.57	10508030.10	18036436.51	206071809.20
raion_popul	30471.00	84056.43	57871.29	2546.00	21819.00	83502.00	122862.00	247469.00
green_zone_part	30471.00	0.22	0.18	0.00	0.06	0.17	0.34	0.85
indust_part	30471.00	0.12	0.12	0.00	0.02	0.07	0.20	0.52
children_preschool	30471.00	5140.03	3816.63	175.00	1706.00	4857.00	7103.00	19223.00
preschool_quota	23783.00	3271.27	2169.76	0.00	1874.00	2854.00	4050.00	11926.00
preschool_education_centers_raion	30471.00	4.07	2.99	0.00	2.00	4.00	6.00	13.00
children_school	30471.00	5354.27	3989.64	168.00	1564.00	5261.00	7227.00	19083.00
school_quota	23786.00	8324.97	4289.73	1012.00	5782.00	7377.00	9891.00	24750.00
school_education_centers_raion	30471.00	4.71	3.45	0.00	2.00	5.00	7.00	14.00
school_education_centers_top_20_raion	30471.00	0.11	0.33	0.00	0.00	0.00	0.00	2.00
hospital_beds_raion	16030.00	1190.74	1057.02	0.00	520.00	990.00	1786.00	4849.00
healthcare_centers_raion	30471.00	1.32	1.49	0.00	0.00	1.00	2.00	6.00
university_top_20_raion	30471.00	0.14	0.44	0.00	0.00	0.00	0.00	3.00
sport_objects_raion	30471.00	6.64	6.57	0.00	1.00	5.00	10.00	29.00
additional_education_raion	30471.00	2.90	3.29	0.00	1.00	2.00	4.00	16.00
culture_objects_top_25_raion	30471.00	0.29	1.51	0.00	0.00	0.00	0.00	10.00
shopping_centers_raion	30471.00	4.20	4.74	0.00	1.00	3.00	6.00	23.00
office_raion	30471.00	8.25	23.54	0.00	0.00	2.00	5.00	141.00
full_all	30471.00	146306.15	283025.12	2546.00	28179.00	85219.00	125111.00	1716730.00
male_f	30471.00	67207.79	129444.56	1208.00	13522.00	39261.00	58226.00	774585.00
female_f	30471.00	79098.66	153630.89	1341.00	15031.00	45729.00	67872.00	942145.00
young_all	30471.00	11178.81	8287.96	365.00	3459.00	10988.00	14906.00	40692.00
young_male	30471.00	5723.85	4275.44	189.00	1782.00	5470.00	7597.00	20977.00
young_female	30471.00	5455.01	4020.55	177.00	1677.00	5333.00	7617.00	19715.00
work_all	30471.00	53667.91	37483.56	1633.00	13996.00	52030.00	77612.00	161290.00
work_male	30471.00	27253.59	18939.15	863.00	7394.00	26382.00	38841.00	79622.00
work_female	30471.00	26414.41	18643.13	771.00	6661.00	26092.00	37942.00	81668.00
ekder_all	30471.00	19209.71	13174.72	548.00	4695.00	20036.00	29172.00	57086.00
ekder_male	30471.00	5811.62	4073.83	156.00	1331.00	6180.00	8563.00	19275.00
ekder_female	30471.00	13398.17	9144.33	393.00	3365.00	13540.00	20165.00	37811.00
0_6_all	30471.00	5140.03	3816.63	175.00	1706.00	4857.00	7103.00	19223.00
0_6_male	30471.00	2631.20	1973.22	91.00	862.00	2435.00	3523.00	9987.00
0_6_female	30471.00	2508.96	1846.60	85.00	844.00	2390.00	3455.00	9236.00

	count	mean	std	min	25%	50%	75%	max
7_14_all	30471.00	5354.27	3989.64	168.00	1564.00	5261.00	7227.00	19083.00
7_14_male	30471.00	2743.47	2054.52	87.00	821.00	2693.00	3585.00	9761.00
7_14_female	30471.00	2610.95	1941.38	82.00	743.00	2535.00	3534.00	9322.00
0_17_all	30471.00	12540.61	9253.05	411.00	3831.00	12508.00	16727.00	45170.00
0_17_male	30471.00	6423.20	4769.59	214.00	1973.00	6085.00	8599.00	23233.00
0_17_female	30471.00	6117.57	4491.73	198.00	1858.00	6185.00	8549.00	21937.00
16_29_all	30471.00	31316.29	60381.52	575.00	5829.00	17864.00	27194.00	367659.00
16_29_male	30471.00	15369.43	29298.65	308.00	2955.00	8896.00	13683.00	172958.00
16_29_female	30471.00	15946.98	31108.98	253.00	2874.00	9353.00	14184.00	194701.00
0_13_all	30471.00	9841.10	7290.01	322.00	3112.00	9633.00	13121.00	36035.00
0_13_male	30471.00	5037.27	3760.75	166.00	1600.00	4835.00	6684.00	18574.00
0_13_female	30471.00	4803.83	3536.91	156.00	1512.00	4667.00	6699.00	17461.00
raion_build_count_with_material_info	25480.00	328.66	277.90	1.00	180.00	273.00	400.00	1681.00
build_count_block	25480.00	50.30	46.94	0.00	13.00	42.00	72.00	223.00
build_count_wood	25480.00	40.85	126.74	0.00	0.00	0.00	7.00	793.00
build_count_frame	25480.00	4.95	14.93	0.00	0.00	0.00	1.00	97.00
build_count_brick	25480.00	108.05	129.12	0.00	10.00	67.00	156.00	664.00
build_count_monolith	25480.00	12.04	19.28	0.00	2.00	6.00	13.00	127.00
build_count_panel	25480.00	107.24	88.26	0.00	35.00	92.00	157.00	431.00
build_count_foam	25480.00	0.17	1.13	0.00	0.00	0.00	0.00	11.00
build_count_slag	25480.00	4.49	13.23	0.00	0.00	0.00	2.00	84.00
build_count_mix	25480.00	0.57	1.53	0.00	0.00	0.00	0.00	9.00
raion_build_count_with_builddate_info	25480.00	328.33	277.92	1.00	178.00	271.00	400.00	1680.00
build_count_before_1920	25480.00	18.90	60.98	0.00	0.00	0.00	3.00	371.00
build_count_1921-1945	25480.00	26.72	62.51	0.00	0.00	2.00	20.00	382.00
build_count_1946-1970	25480.00	141.43	125.43	0.00	14.00	135.00	216.00	845.00
build_count_1971-1995	25480.00	80.13	57.87	0.00	37.00	71.00	111.00	246.00
build_count_after_1995	25480.00	61.15	114.22	0.00	14.00	24.00	57.00	799.00
ID_metro	30471.00	72.48	54.90	1.00	27.00	53.00	108.00	223.00
metro_min_avto	30471.00	4.96	6.55	0.00	1.72	2.80	4.83	61.44
metro_km_avto	30471.00	3.70	5.81	0.00	1.04	1.78	3.78	74.91
metro_min_walk	30446.00	42.74	69.30	0.00	11.48	20.45	45.32	711.22
metro_km_walk	30446.00	3.56	5.78	0.00	0.96	1.70	3.78	59.27
kindergarten_km	30471.00	0.98	2.12	0.00	0.20	0.35	0.97	29.09
school_km	30471.00	1.32	3.06	0.00	0.27	0.47	0.89	47.39
park_km	30471.00	3.10	3.97	0.00	0.97	1.80	3.40	47.35
green_zone_km	30471.00	0.30	0.30	0.00	0.10	0.21	0.42	1.98
industrial_km	30471.00	0.77	0.72	0.00	0.29	0.58	1.04	14.05
water_treatment_km	30471.00	11.17	7.04	0.27	5.30	10.38	16.79	47.59
cemetery_km	30471.00	2.31	1.45	0.00	1.33	1.97	3.09	15.78
incineration_km	30471.00	10.88	6.79	0.20	6.22	10.32	13.39	58.63
railroad_station_walk_km	30446.00	4.39	3.84	0.03	1.93	3.24	5.15	24.65
railroad_station_walk_min	30446.00	52.64	46.04	0.34	23.18	38.83	61.77	295.84
ID_railroad_station_walk	30446.00	38.86	27.77	1.00	18.00	33.00	53.00	133.00
railroad_station_avto_km	30471.00	4.59	3.91	0.03	2.12	3.43	5.39	24.65
railroad_station_avto_min	30471.00	6.09	4.79	0.04	3.24	4.94	7.30	38.69
ID_railroad_station_avto	30471.00	45.71	34.39	1.00	19.00	34.00	73.00	138.00
public_transport_station_km	30471.00	0.41	1.27	0.00	0.10	0.16	0.28	17.41
public_transport_station_min_walk	30471.00	4.97	15.27	0.03	1.22	1.92	3.34	208.96
water_km	30471.00	0.69	0.43	0.01	0.34	0.62	0.96	2.83
mkad_km	30471.00	6.27	5.14	0.01	2.63	5.47	8.18	53.28

	count	mean	std	min	25%	50%	75%	max
ttk_km	30471.00	11.32	8.07	0.00	5.34	9.84	15.68	66.03
sadovoe_km	30471.00	14.06	8.34	0.00	8.35	12.75	18.72	68.85
bulvar_ring_km	30471.00	15.02	8.48	0.00	9.26	13.61	19.95	69.98
kremlin_km	30471.00	16.04	8.44	0.07	10.46	14.88	20.67	70.74
big_road1_km	30471.00	1.88	1.30	0.00	0.78	1.72	2.81	7.00
ID_big_road1	30471.00	11.52	11.35	1.00	2.00	10.00	14.00	48.00
big_road2_km	30471.00	3.40	1.87	0.00	2.10	3.21	4.32	13.80
ID_big_road2	30471.00	22.41	17.39	1.00	4.00	21.00	38.00	58.00
railroad_km	30471.00	1.89	1.97	0.00	0.66	1.24	2.52	17.39
zd_vokzaly_avto_km	30471.00	17.21	10.30	0.14	9.99	14.76	24.06	91.22
ID_railroad_terminal	30471.00	51.67	35.34	5.00	32.00	50.00	83.00	121.00
bus_terminal_avto_km	30471.00	9.99	7.33	0.06	5.21	7.45	13.28	74.80
ID_bus_terminal	30471.00	6.70	3.79	1.00	3.00	8.00	9.00	14.00
oil_chemistry_km	30471.00	17.40	10.40	0.51	8.72	16.70	23.42	70.41
nuclear_reactor_km	30471.00	10.95	7.25	0.31	5.24	8.97	16.37	64.26
radiation_km	30471.00	4.41	5.60	0.00	1.23	2.44	4.69	53.89
power_transmission_line_km	30471.00	3.49	4.55	0.03	0.98	1.90	4.93	43.32
thermal_power_plant_km	30471.00	7.34	5.62	0.40	3.77	5.89	9.82	56.86
ts_km	30471.00	4.93	4.80	0.00	2.06	3.97	5.55	54.08
big_market_km	30471.00	13.28	8.50	0.66	7.53	11.91	16.56	59.50
market_shop_km	30471.00	3.96	3.45	0.00	1.54	2.93	5.49	41.10
fitness_km	30471.00	1.15	1.54	0.00	0.36	0.66	1.33	26.65
swim_pool_km	30471.00	4.23	4.63	0.00	1.71	2.88	5.37	53.36
ice_rink_km	30471.00	6.12	4.67	0.00	3.04	5.55	7.96	46.04
stadium_km	30471.00	9.44	7.64	0.11	4.02	6.97	13.59	83.40
basketball_km	30471.00	4.79	5.40	0.01	1.31	2.88	6.36	56.70
hospice_morgue_km	30471.00	2.65	2.95	0.00	1.12	1.90	3.30	43.69
detention_facility_km	30471.00	14.55	11.08	0.04	5.67	11.31	24.88	89.37
public_healthcare_km	30471.00	3.36	4.34	0.00	1.28	2.34	3.98	76.06
university_km	30471.00	6.86	6.73	0.00	2.20	4.34	9.38	84.86
workplaces_km	30471.00	3.93	4.91	0.00	1.02	2.03	5.42	55.28
shopping_centers_km	30471.00	1.51	2.25	0.00	0.48	0.84	1.55	26.26
office_km	30471.00	2.01	2.31	0.00	0.56	1.05	3.05	18.96
additional_education_km	30471.00	1.33	1.53	0.00	0.47	0.90	1.57	24.27
preschool_km	30471.00	1.35	3.05	0.00	0.29	0.49	0.94	47.39
big_church_km	30471.00	2.33	2.99	0.00	0.86	1.49	2.92	45.67
church_synagogue_km	30471.00	0.97	0.75	0.00	0.53	0.86	1.25	15.62
mosque_km	30471.00	7.74	5.78	0.01	3.77	6.54	10.05	44.85
theater_km	30471.00	9.64	7.11	0.03	4.23	8.61	13.46	87.60
museum_km	30471.00	7.06	5.70	0.01	2.88	5.64	10.33	59.20
exhibition_km	30471.00	5.55	5.30	0.01	2.24	4.11	6.97	54.43
catering_km	30471.00	0.69	0.83	0.00	0.21	0.41	0.84	12.16
green_part_500	30471.00	13.38	16.05	0.00	1.48	8.38	19.92	100.00
prom_part_500	30471.00	5.72	11.54	0.00	0.00	0.00	5.76	98.77
office_count_500	30471.00	0.74	2.28	0.00	0.00	0.00	0.00	34.00
office_sqm_500	30471.00	13983.38	42610.02	0.00	0.00	0.00	0.00	611015.00
trc_count_500	30471.00	0.56	1.25	0.00	0.00	0.00	1.00	8.00
trc_sqm_500	30471.00	21797.26	81580.16	0.00	0.00	0.00	120.00	1500000.00
cafe_count_500	30471.00	3.87	11.28	0.00	0.00	1.00	3.00	120.00
cafe_sum_500_min_price_avg	17190.00	741.32	341.10	300.00	500.00	666.67	954.84	4000.00
cafe_sum_500_max_price_avg	17190.00	1247.02	526.54	500.00	1000.00	1166.67	1500.00	6000.00

	count	mean	std	min	25%	50%	75%	max
cafe_avg_price_500	17190.00	994.17	432.52	400.00	750.00	916.67	1250.00	5000.00
cafe_count_500_na_price	30471.00	0.34	1.36	0.00	0.00	0.00	0.00	13.00
cafe_count_500_price_500	30471.00	0.99	3.06	0.00	0.00	0.00	1.00	33.00
cafe_count_500_price_1000	30471.00	0.98	2.27	0.00	0.00	0.00	1.00	39.00
cafe_count_500_price_1500	30471.00	0.84	2.32	0.00	0.00	0.00	1.00	29.00
cafe_count_500_price_2500	30471.00	0.54	2.14	0.00	0.00	0.00	0.00	22.00
cafe_count_500_price_4000	30471.00	0.14	0.69	0.00	0.00	0.00	0.00	14.00
cafe_count_500_price_high	30471.00	0.03	0.18	0.00	0.00	0.00	0.00	3.00
big_church_count_500	30471.00	0.29	1.19	0.00	0.00	0.00	0.00	11.00
church_count_500	30471.00	0.58	2.16	0.00	0.00	0.00	0.00	17.00
mosque_count_500	30471.00	0.00	0.07	0.00	0.00	0.00	0.00	1.00
leisure_count_500	30471.00	0.07	0.39	0.00	0.00	0.00	0.00	9.00
sport_count_500	30471.00	0.91	1.64	0.00	0.00	0.00	1.00	11.00
market_count_500	30471.00	0.12	0.40	0.00	0.00	0.00	0.00	4.00
green_part_1000	30471.00	16.96	14.51	0.00	6.31	13.04	24.18	100.00
prom_part_1000	30471.00	8.78	11.52	0.00	0.00	4.02	12.62	72.20
office_count_1000	30471.00	3.09	8.88	0.00	0.00	0.00	2.00	91.00
office_sqm_1000	30471.00	62267.22	143852.96	0.00	0.00	0.00	54500.00	2244723.00
trc_count_1000	30471.00	1.98	3.25	0.00	0.00	1.00	3.00	20.00
trc_sqm_1000	30471.00	65880.79	150279.65	0.00	0.00	7800.00	67183.00	1500000.00
cafe_count_1000	30471.00	15.41	47.63	0.00	1.00	4.00	11.00	449.00
cafe_sum_1000_min_price_avg	23947.00	710.92	224.43	300.00	543.17	669.23	839.29	2500.00
cafe_sum_1000_max_price_avg	23947.00	1206.62	343.13	500.00	1000.00	1142.86	1400.00	4000.00
cafe_avg_price_1000	23947.00	958.77	282.63	400.00	750.00	912.50	1120.00	3250.00
cafe_count_1000_na_price	30471.00	1.02	3.17	0.00	0.00	0.00	1.00	28.00
cafe_count_1000_price_500	30471.00	4.15	13.86	0.00	0.00	1.00	3.00	112.00
cafe_count_1000_price_1000	30471.00	3.94	9.20	0.00	0.00	1.00	4.00	107.00
cafe_count_1000_price_1500	30471.00	3.52	10.50	0.00	0.00	1.00	3.00	104.00
cafe_count_1000_price_2500	30471.00	1.95	7.51	0.00	0.00	0.00	1.00	79.00
cafe_count_1000_price_4000	30471.00	0.77	3.96	0.00	0.00	0.00	0.00	40.00
cafe_count_1000_price_high	30471.00	0.06	0.33	0.00	0.00	0.00	0.00	7.00
big_church_count_1000	30471.00	0.81	2.49	0.00	0.00	0.00	1.00	27.00
church_count_1000	30471.00	1.82	5.11	0.00	0.00	1.00	1.00	38.00
mosque_count_1000	30471.00	0.02	0.14	0.00	0.00	0.00	0.00	1.00
leisure_count_1000	30471.00	0.47	1.82	0.00	0.00	0.00	0.00	30.00
sport_count_1000	30471.00	2.91	3.33	0.00	0.00	2.00	4.00	25.00
market_count_1000	30471.00	0.38	0.73	0.00	0.00	0.00	1.00	6.00
green_part_1500	30471.00	19.20	14.20	0.00	8.47	14.95	26.69	90.41
prom_part_1500	30471.00	10.60	10.98	0.00	1.52	7.82	15.34	63.00
office_count_1500	30471.00	7.31	20.97	0.00	0.00	1.00	4.00	173.00
office_sqm_1500	30471.00	140371.19	301684.43	0.00	0.00	16650.00	117300.00	2908344.00
trc_count_1500	30471.00	3.73	4.48	0.00	0.00	3.00	5.00	27.00
trc_sqm_1500	30471.00	127714.86	212631.16	0.00	0.00	49410.00	154590.00	1533000.00
cafe_count_1500	30471.00	32.46	97.45	0.00	2.00	10.00	23.00	784.00
cafe_sum_1500_min_price_avg	26272.00	714.06	192.26	300.00	585.71	692.31	821.43	2500.00
cafe_sum_1500_max_price_avg	26272.00	1206.01	291.30	500.00	1000.00	1166.67	1366.67	4000.00
cafe_avg_price_1500	26272.00	960.04	240.67	400.00	795.00	926.32	1093.75	3250.00
cafe_count_1500_na_price	30471.00	2.11	5.98	0.00	0.00	1.00	2.00	54.00
cafe_count_1500_price_500	30471.00	8.23	24.89	0.00	0.00	2.00	6.00	195.00
cafe_count_1500_price_1000	30471.00	8.76	22.13	0.00	1.00	3.00	8.00	177.00
cafe_count_1500_price_1500	30471.00	7.88	24.13	0.00	0.00	2.00	6.00	183.00

	count	mean	std	min	25%	50%	75%	max
cafe_count_1500_price_2500	30471.00	3.83	13.51	0.00	0.00	1.00	2.00	127.00
cafe_count_1500_price_4000	30471.00	1.46	6.74	0.00	0.00	0.00	0.00	58.00
cafe_count_1500_price_high	30471.00	0.19	0.91	0.00	0.00	0.00	0.00	12.00
big_church_count_1500	30471.00	1.97	6.48	0.00	0.00	1.00	1.00	44.00
church_count_1500	30471.00	3.66	9.74	0.00	1.00	1.00	3.00	75.00
mosque_count_1500	30471.00	0.04	0.19	0.00	0.00	0.00	0.00	1.00
leisure_count_1500	30471.00	0.95	3.35	0.00	0.00	0.00	1.00	44.00
sport_count_1500	30471.00	5.87	5.76	0.00	1.00	5.00	9.00	37.00
market_count_1500	30471.00	0.77	1.12	0.00	0.00	0.00	1.00	7.00
green_part_2000	30471.00	20.84	14.19	0.01	10.16	17.63	28.33	75.30
prom_part_2000	30471.00	11.22	9.64	0.00	3.12	8.80	16.21	56.10
office_count_2000	30471.00	13.32	38.72	0.00	0.00	2.00	7.00	250.00
office_sqm_2000	30471.00	246497.20	503636.39	0.00	0.00	58411.00	207193.00	3602982.00
trc_count_2000	30471.00	5.97	6.34	0.00	1.00	5.00	9.00	37.00
trc_sqm_2000	30471.00	212758.76	290831.52	0.00	12065.00	117300.00	286681.00	2448300.00
cafe_count_2000	30471.00	55.03	161.63	0.00	3.00	18.00	37.00	1115.00
cafe_sum_2000_min_price_avg	28746.00	719.98	199.37	300.00	607.69	683.33	791.67	2166.67
cafe_sum_2000_max_price_avg	28746.00	1210.91	302.83	500.00	1000.00	1156.25	1321.62	3500.00
cafe_avg_price_2000	28746.00	965.44	250.01	400.00	823.47	919.23	1057.24	2833.33
cafe_count_2000_na_price	30471.00	3.59	10.22	0.00	0.00	1.00	3.00	70.00
cafe_count_2000_price_500	30471.00	13.60	39.16	0.00	1.00	4.00	10.00	278.00
cafe_count_2000_price_1000	30471.00	15.23	39.26	0.00	1.00	6.00	13.00	262.00
cafe_count_2000_price_1500	30471.00	13.27	39.21	0.00	1.00	4.00	9.00	261.00
cafe_count_2000_price_2500	30471.00	6.65	23.23	0.00	0.00	1.00	3.00	170.00
cafe_count_2000_price_4000	30471.00	2.32	9.72	0.00	0.00	0.00	1.00	81.00
cafe_count_2000_price_high	30471.00	0.38	1.69	0.00	0.00	0.00	0.00	16.00
big_church_count_2000	30471.00	3.25	10.00	0.00	0.00	1.00	2.00	70.00
church_count_2000	30471.00	6.20	15.49	0.00	2.00	3.00	5.00	108.00
mosque_count_2000	30471.00	0.09	0.28	0.00	0.00	0.00	0.00	1.00
leisure_count_2000	30471.00	1.92	7.19	0.00	0.00	0.00	1.00	55.00
sport_count_2000	30471.00	9.85	9.22	0.00	2.00	9.00	14.00	54.00
market_count_2000	30471.00	1.17	1.43	0.00	0.00	1.00	2.00	8.00
green_part_3000	30471.00	22.73	14.02	0.31	12.15	20.26	30.36	74.02
prom_part_3000	30471.00	10.98	7.94	0.00	4.24	9.66	15.73	45.10
office_count_3000	30471.00	29.37	81.35	0.00	0.00	5.00	17.00	493.00
office_sqm_3000	30471.00	543261.56	1056127.70	0.00	0.00	130303.00	494706.50	6106112.00
trc_count_3000	30471.00	11.83	10.91	0.00	2.00	11.00	17.00	66.00
trc_sqm_3000	30471.00	438131.57	469731.55	0.00	41100.00	294350.00	659453.00	2654102.00
cafe_count_3000	30471.00	110.87	290.13	0.00	6.00	41.00	78.00	1815.00
cafe_sum_3000_min_price_avg	29480.00	765.90	224.01	300.00	650.00	711.11	815.63	1833.33
cafe_sum_3000_max_price_avg	29480.00	1283.33	353.30	500.00	1102.27	1211.54	1333.33	3000.00
cafe_avg_price_3000	29480.00	1024.62	288.16	400.00	875.80	961.11	1083.33	2416.67
cafe_count_3000_na_price	30471.00	7.27	18.24	0.00	0.00	3.00	6.00	119.00
cafe_count_3000_price_500	30471.00	27.78	71.83	0.00	1.00	9.00	22.00	449.00
cafe_count_3000_price_1000	30471.00	30.45	69.53	0.00	2.00	14.00	26.00	441.00
cafe_count_3000_price_1500	30471.00	26.69	70.86	0.00	2.00	10.00	17.00	446.00
cafe_count_3000_price_2500	30471.00	13.32	41.28	0.00	1.00	3.00	6.00	266.00
cafe_count_3000_price_4000	30471.00	4.65	16.88	0.00	0.00	1.00	2.00	113.00
cafe_count_3000_price_high	30471.00	0.71	2.78	0.00	0.00	0.00	0.00	23.00
big_church_count_3000	30471.00	6.12	15.41	0.00	1.00	2.00	5.00	102.00
church_count_3000	30471.00	12.29	25.79	0.00	3.00	6.00	10.00	164.00

	count	mean	std	min	25%	50%	75%	max
mosque_count_3000	30471.00	0.20	0.45	0.00	0.00	0.00	0.00	2.00
leisure_count_3000	30471.00	3.88	13.19	0.00	0.00	0.00	2.00	85.00
sport_count_3000	30471.00	20.24	18.84	0.00	5.00	18.00	29.00	100.00
market_count_3000	30471.00	2.32	2.27	0.00	0.00	2.00	4.00	10.00
green_part_5000	30471.00	22.77	11.21	3.52	14.78	19.76	31.41	75.46
prom_part_5000	30293.00	10.35	5.67	0.21	6.05	8.98	14.00	28.56
office_count_5000	30471.00	71.36	155.34	0.00	2.00	15.00	53.00	789.00
office_sqm_5000	30471.00	1401057.39	2303052.49	0.00	85159.00	432438.00	1433847.00	12702114.00
trc_count_5000	30471.00	30.13	23.92	0.00	6.00	31.00	43.00	120.00
trc_sqm_5000	30471.00	1173871.15	1004809.88	0.00	262000.00	1075495.00	1683836.00	4585477.00
cafe_count_5000	30471.00	265.53	514.54	0.00	20.00	108.00	222.00	2645.00
cafe_sum_5000_min_price_avg	30174.00	765.10	152.41	300.00	670.87	721.74	816.66	1875.00
cafe_sum_5000_max_price_avg	30174.00	1278.28	236.21	500.00	1144.23	1211.95	1346.09	3000.00
cafe_avg_price_5000	30174.00	1021.69	194.12	400.00	909.38	966.67	1091.67	2437.50
cafe_count_5000_na_price	30471.00	17.81	33.27	0.00	1.00	8.00	15.00	174.00
cafe_count_5000_price_500	30471.00	66.20	125.93	0.00	4.00	28.00	59.00	650.00
cafe_count_5000_price_1000	30471.00	73.44	126.17	0.00	8.00	36.00	69.00	648.00
cafe_count_5000_price_1500	30471.00	63.47	124.08	0.00	6.00	24.00	51.00	641.00
cafe_count_5000_price_2500	30471.00	32.06	73.47	0.00	2.00	8.00	21.00	377.00
cafe_count_5000_price_4000	30471.00	10.78	28.39	0.00	1.00	2.00	5.00	147.00
cafe_count_5000_price_high	30471.00	1.77	5.42	0.00	0.00	0.00	1.00	30.00
big_church_count_5000	30471.00	15.05	29.12	0.00	2.00	7.00	12.00	151.00
church_count_5000	30471.00	30.25	47.35	0.00	9.00	16.00	28.00	250.00
mosque_count_5000	30471.00	0.44	0.61	0.00	0.00	0.00	1.00	2.00
leisure_count_5000	30471.00	8.65	20.58	0.00	0.00	2.00	7.00	106.00
sport_count_5000	30471.00	52.80	46.29	0.00	11.00	48.00	76.00	218.00
market_count_5000	30471.00	5.99	4.89	0.00	1.00	5.00	10.00	21.00
price_doc	30471.00	7123035.28	4780111.33	100000.00	4740002.00	6274411.00	8300000.00	11111112.00
year	30471.00	2013.45	0.96	2011.00	2013.00	2014.00	2014.00	2015.00
month	30471.00	6.75	3.52	1.00	4.00	6.00	10.00	12.00
day	30471.00	16.44	8.69	1.00	9.00	17.00	24.00	31.00

:)