

# *Анализ цен на квартиры в Волгограде*

*Выполнили:  
студентки группы андэк  
Назаренко Анастасия,  
Ламеева Анна*

# Описание данных

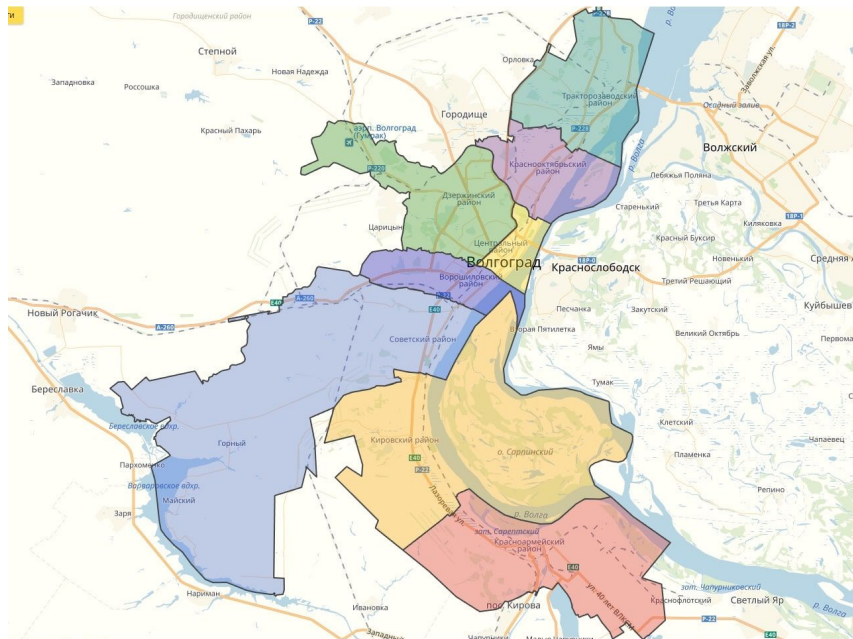
Источник данных: <https://www.cian.ru/>

Парсер: requests + BeautifulSoup

Итого: 3894 наблюдений (без повторов)

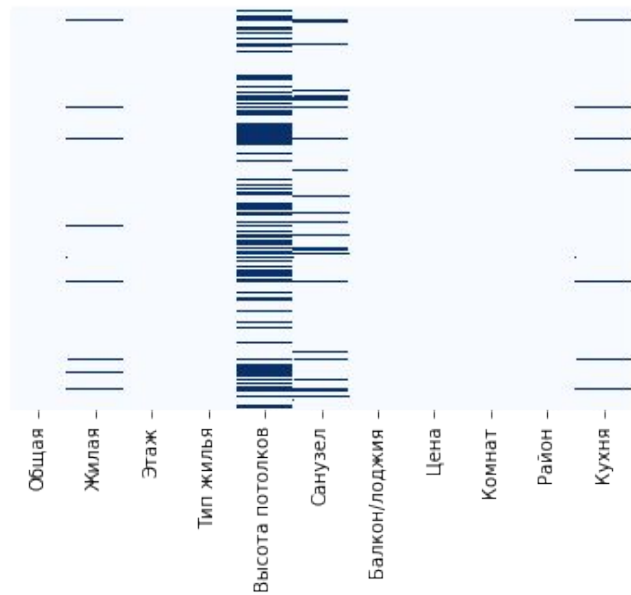
Переменные:

- Цена квартиры,
- Общая площадь,
- Жилая площадь,
- Площадь кухни,
- Кол-во комнат (1, 2, 3),
- Этаж,
- Высота потолков,
- Район,
- Санузел (совмещенный, отдельный),
- Тип жилья (Вторичка/ Новостройка),
- Наличие балкона/ лоджии

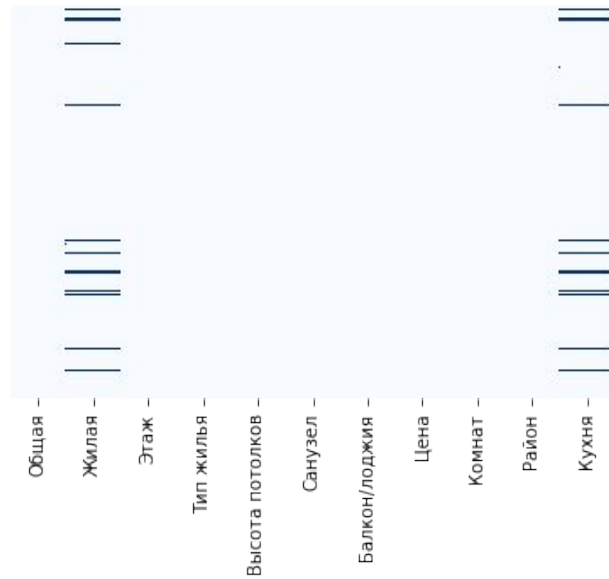


Были исключены из-за обилия пропусков: Срок сдачи, Вид из окон, Тип ремонта, Тип отделки, Планировка

# Работа с пропущенными значениями

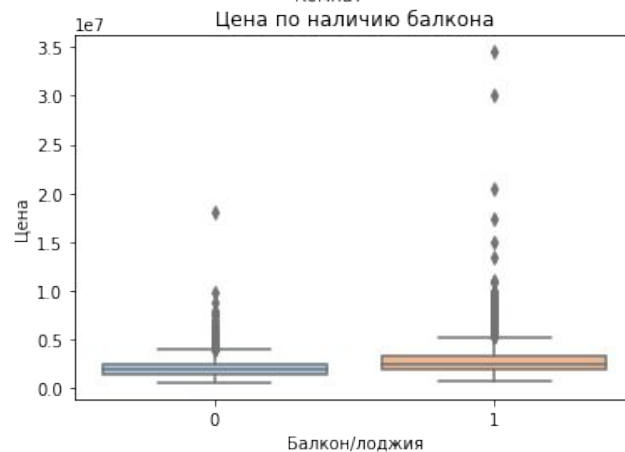
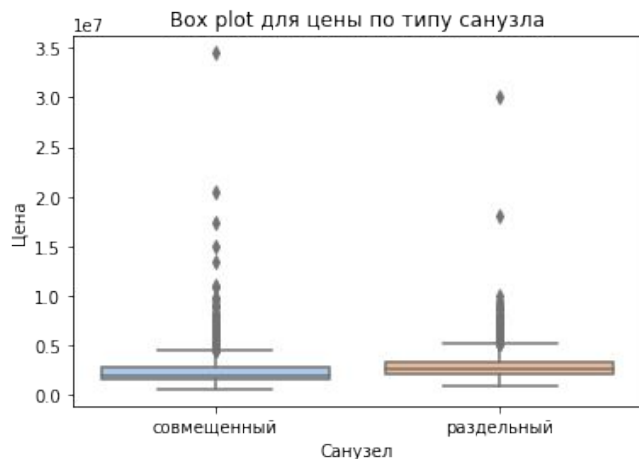
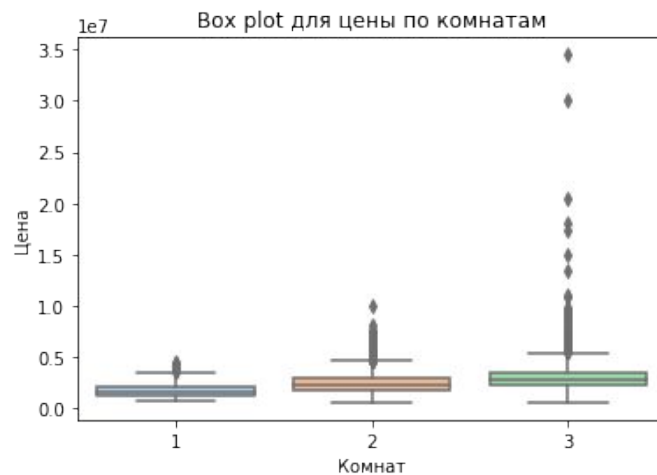
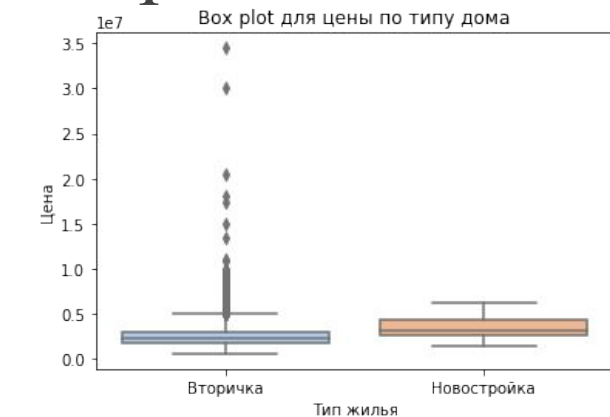


Шаг 1

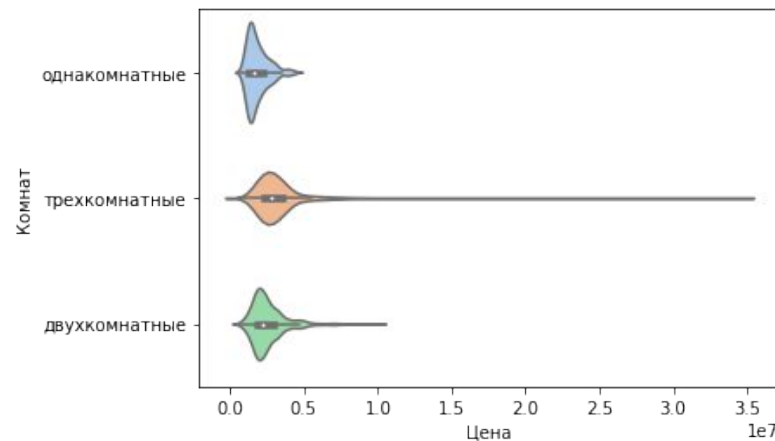
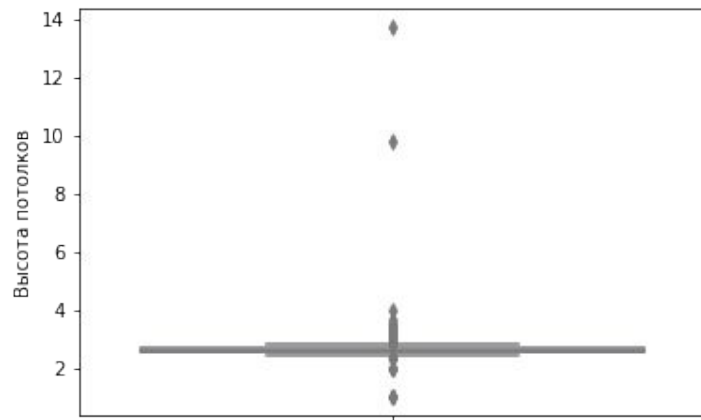
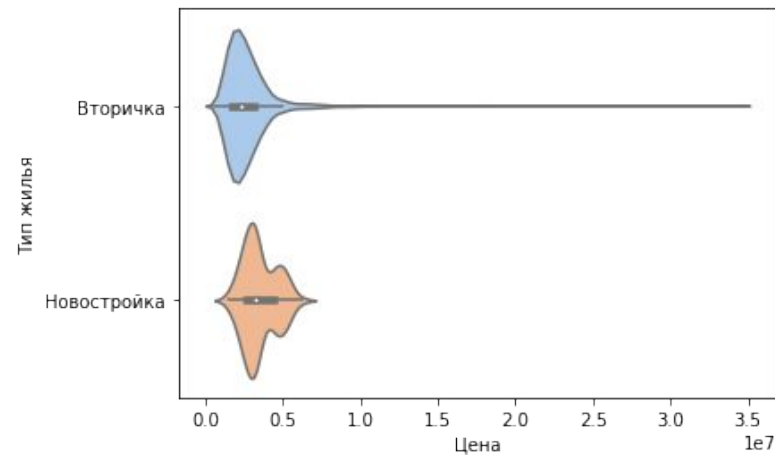
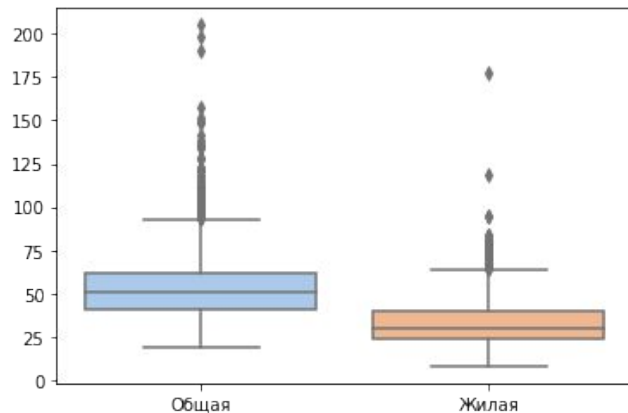


Шаг 2

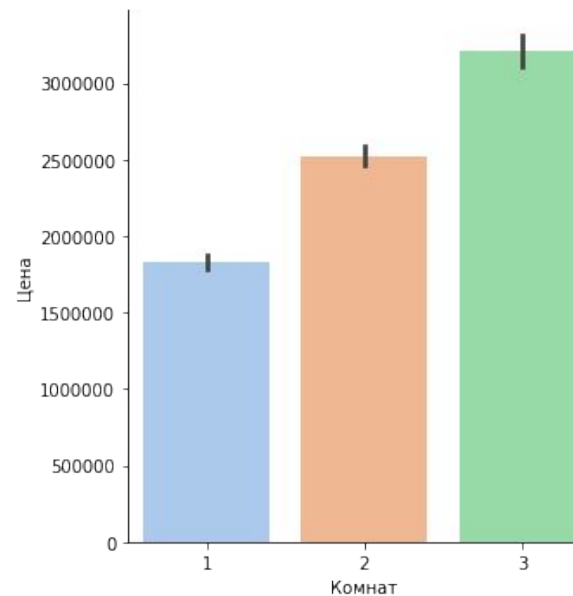
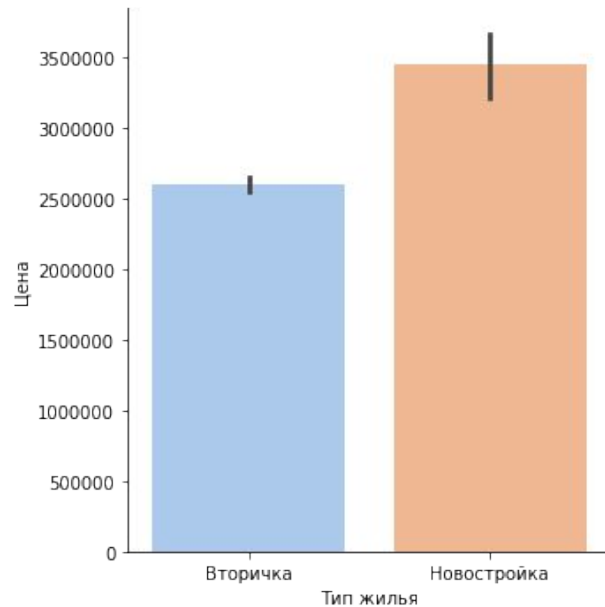
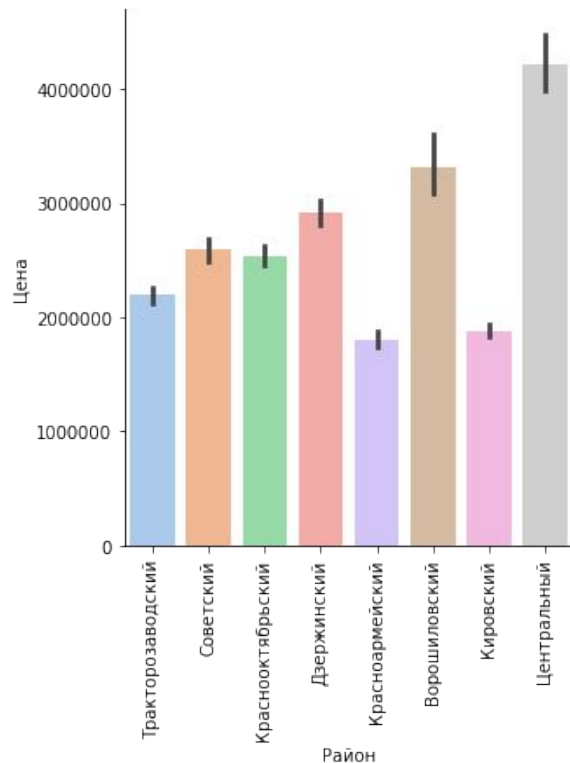
# Предварительный анализ данных



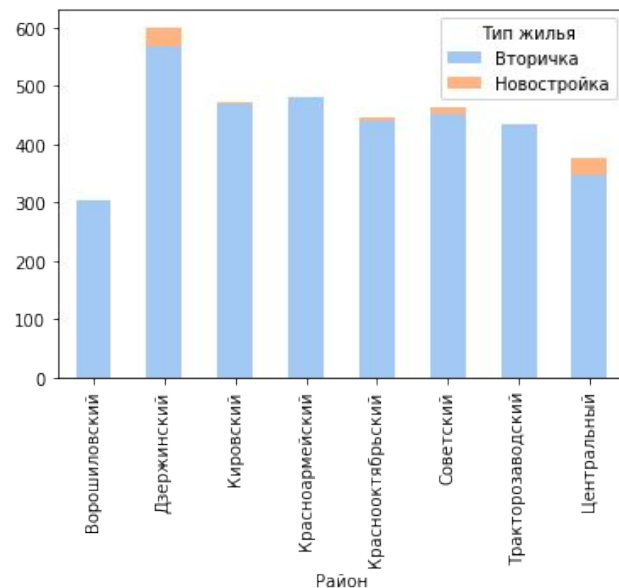
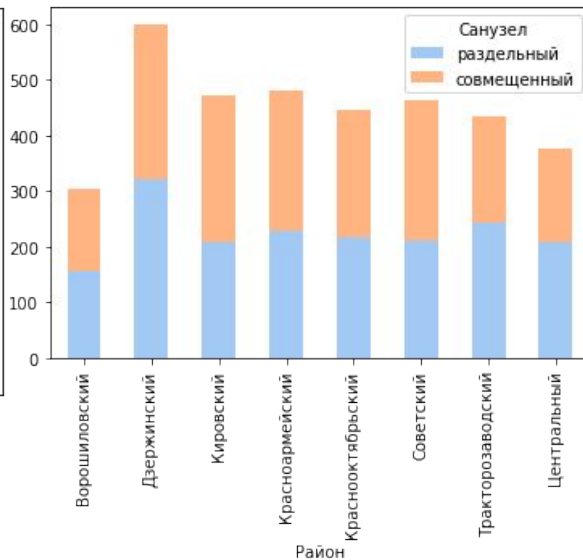
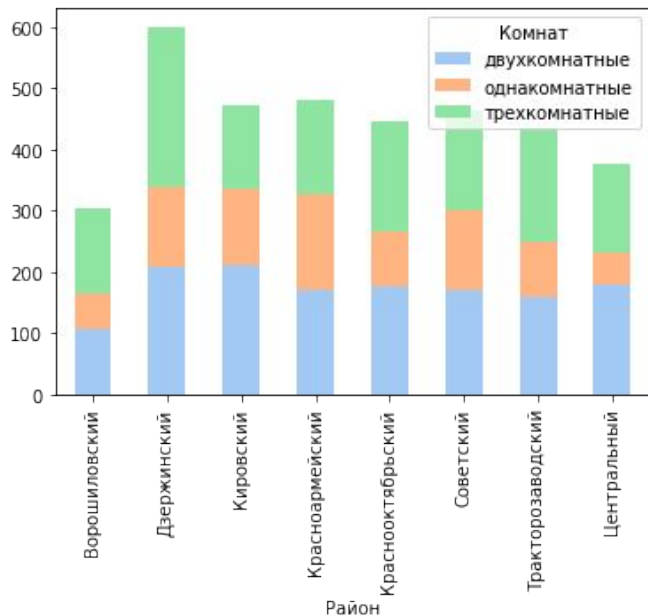
# Предварительный анализ данных



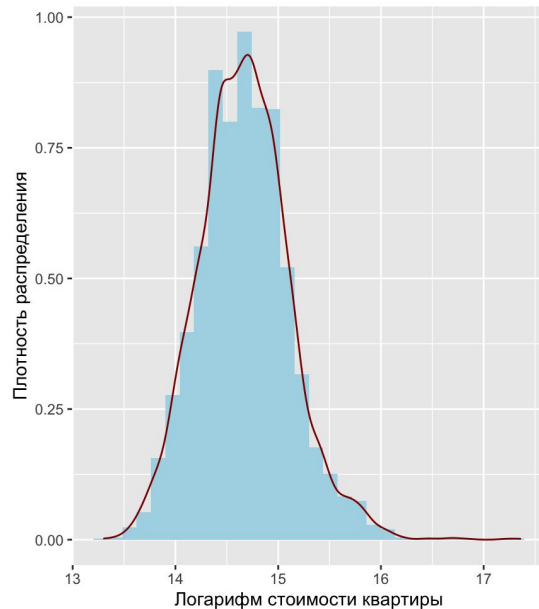
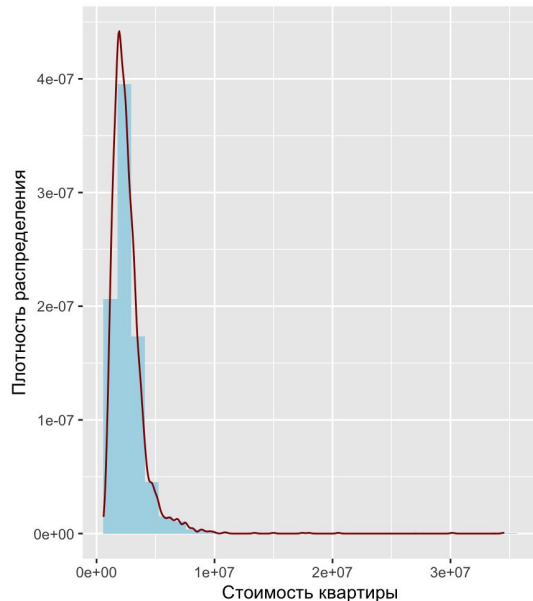
# Предварительный анализ данных



# Предварительный анализ данных



# Предварительный анализ данных

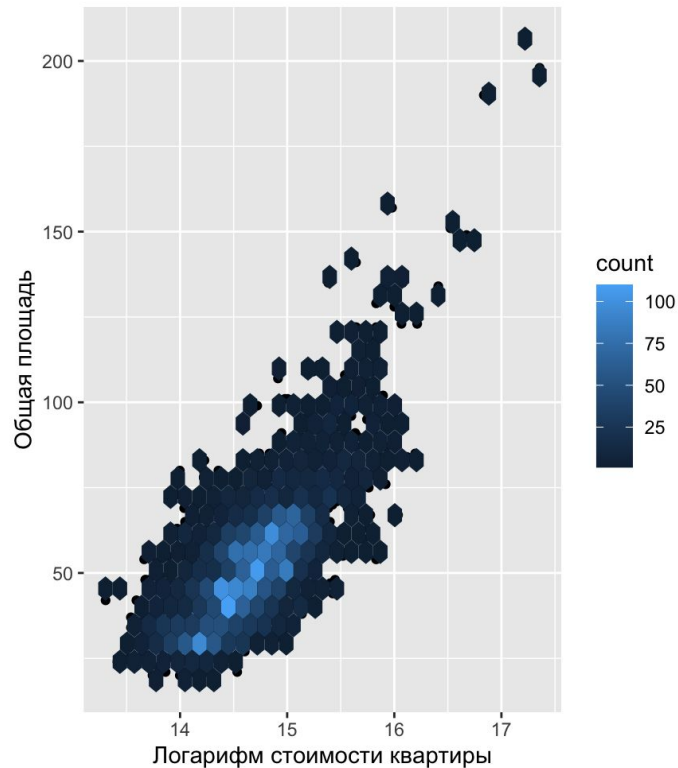
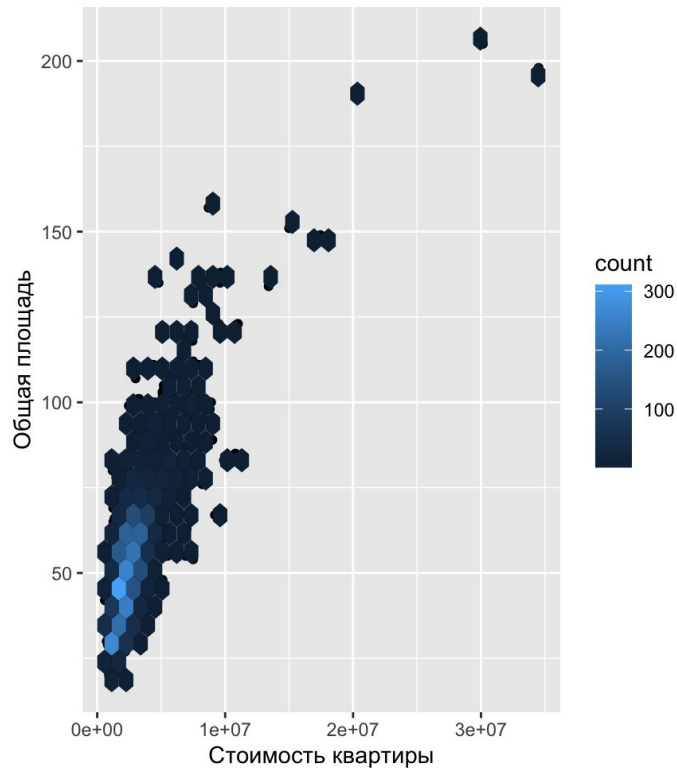


Корреляционная матрица

	Общая	Жилая	Этаж	Высота потолков	Цена	Кухня
Общая	1	0.9	0.21	0.11	0.75	0.51
Жилая	0.9	1	0.093	0.079	0.6	0.22
Этаж	0.21	0.093	1	-0.017	0.33	0.27
Высота потолков	0.11	0.079	-0.017	1	0.053	0.08
Цена	0.75	0.6	0.33	0.053	1	0.53
Кухня	0.51	0.22	0.27	0.08	0.53	1



# Регрессионный анализ



*Зависимость становится более наглядная при взятии логарифма от цены*

# Регрессионный анализ

Изначально были построены 3 модели:

Классическая линейная; Полулогарифмическая; Логарифмическая

Зависимая переменная:

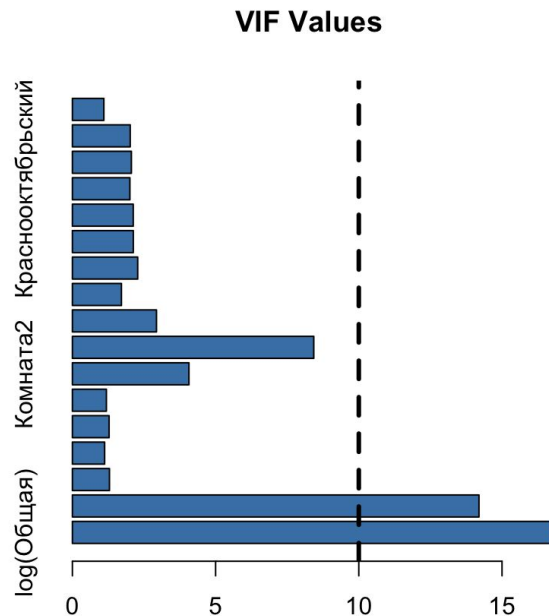
	Цена	log(Цена)	log(Цена)
Общая	67,240.810 <sup>***</sup>	0.015 <sup>***</sup>	
	(2,755.864)	(0.001)	
Жилая	17,593.440 <sup>***</sup>	-0.002	
	(3,847.148)	(0.001)	
log(Общая)			0.931 <sup>***</sup>
			(0.047)
log(Жилая)			-0.003
			(0.036)
Этаж	42,726.960 <sup>***</sup>	0.016 <sup>***</sup>	0.014 <sup>***</sup>
	(4,630.051)	(0.001)	(0.001)
Высота.потолков	-455,218.500 <sup>***</sup>	-0.097 <sup>***</sup>	
	(96,234.820)	(0.026)	
log(Высота.потолков)			-0.368 <sup>***</sup>
			(0.066)
Санузел	225,249.000 <sup>***</sup>	0.004	0.031 <sup>***</sup>
	(30,515.280)	(0.008)	(0.008)
Балкон.лоджия	179,580.500 <sup>***</sup>	0.109 <sup>***</sup>	0.098 <sup>***</sup>
	(32,196.570)	(0.009)	(0.009)
Кухня	34,892.590 <sup>***</sup>	0.015 <sup>***</sup>	
	(6,122.978)	(0.002)	
Ворошиловский	-396,997.500 <sup>***</sup>	-0.132 <sup>***</sup>	-0.142 <sup>***</sup>
	(63,650.960)	(0.017)	(0.017)
Дзержинский	-789,268.000 <sup>***</sup>	-0.249 <sup>***</sup>	-0.252 <sup>***</sup>
	(54,513.170)	(0.015)	(0.015)

Кировский	-1,279,096.000 <sup>***</sup>	-0.498 <sup>***</sup>	-0.500 <sup>***</sup>
	(58,094.650)	(0.016)	(0.016)
Красноармейский	-1,385,110.000 <sup>***</sup>	-0.558 <sup>***</sup>	-0.558 <sup>***</sup>
	(57,806.390)	(0.016)	(0.015)
Краснооктябрьский	-936,567.100 <sup>***</sup>	-0.273 <sup>***</sup>	-0.282 <sup>***</sup>
	(58,057.000)	(0.016)	(0.016)
Советский	-922,886.600 <sup>***</sup>	-0.294 <sup>***</sup>	-0.303 <sup>***</sup>
	(57,561.810)	(0.016)	(0.015)
Тракторозаводский	-1,189,173.000 <sup>***</sup>	-0.409 <sup>***</sup>	-0.414 <sup>***</sup>
	(58,728.570)	(0.016)	(0.016)
Вторичка	607,485.200 <sup>***</sup>	0.077 <sup>***</sup>	0.084 <sup>***</sup>
	(93,346.210)	(0.026)	(0.025)
Комната2	-543,413.600 <sup>***</sup>	0.063 <sup>***</sup>	-0.052 <sup>***</sup>
	(45,160.140)	(0.012)	(0.015)
Комната3	-1,260,020.000 <sup>***</sup>	0.029	-0.125 <sup>***</sup>
	(65,230.660)	(0.018)	(0.022)
log(Кухня)			0.184 <sup>***</sup>
			(0.019)
Constant	36,321.070	14.134 <sup>***</sup>	11.192 <sup>***</sup>
	(272,908.400)	(0.075)	(0.103)
AIC	107277	-617.339	-774.01
BIC	107394	-499.9	-656.5
Observations	3,57	3,57	3,57
R <sup>2</sup>	0.726	0.753	0.764

## Тестирование логарифмической модели (Модель 2)

\_\_\_\_\_

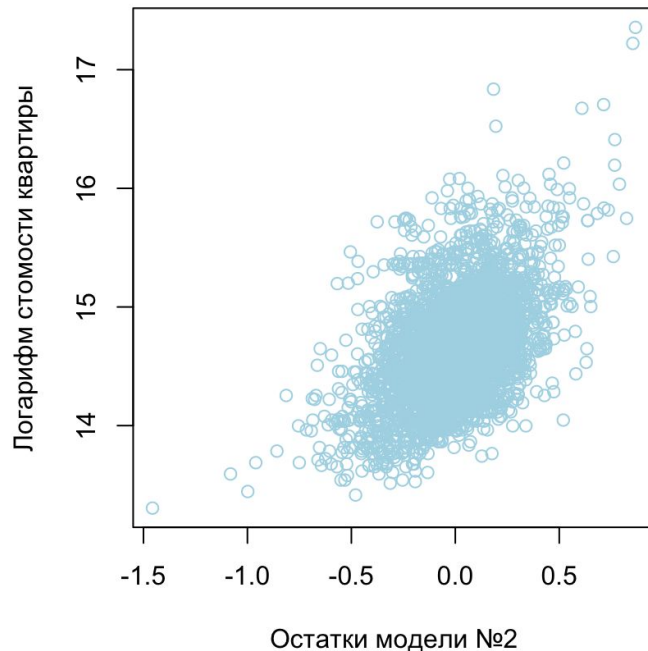
- 1) Тест на мультиколлинеарность с помощью VIFs. Большие значения наблюдаются для  $\log(\text{Жилая площадь})$  и  $\log(\text{Общая площадь})$ ;
- 2) Тест Жарка-Берра на нормальное распределение остатков. P-value = 0,000. Остатки не подчиняются нормальному распределению
- 3) Результаты теста Рамсея для логарифмической модели показывают, что RESET = 3.7451, p-value = 0.02373, значит, гипотеза о правильной спецификации не отвергается на 1% уровне значимости.



# Тестирование на гетероскедастичность

— — —

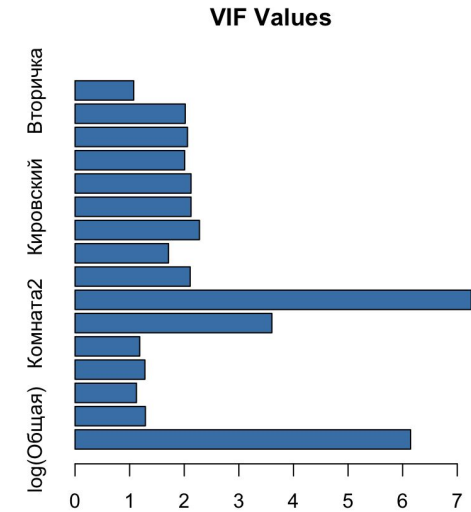
Гетероскедастичность проверялась с помощью теста Бреуша-Пагана. P-value в данной модели 0,000, что меньше уровня значимости 0,05, а значит, у нас присутствует гетероскедастичность в модели



Удаляем жилую площадь

# Wald-test p-value = 0.9243 - регрессор  
действительно можно выбросить

$\overline{\text{AIC}} = -776$   
 $\text{BIC} = -664.8$

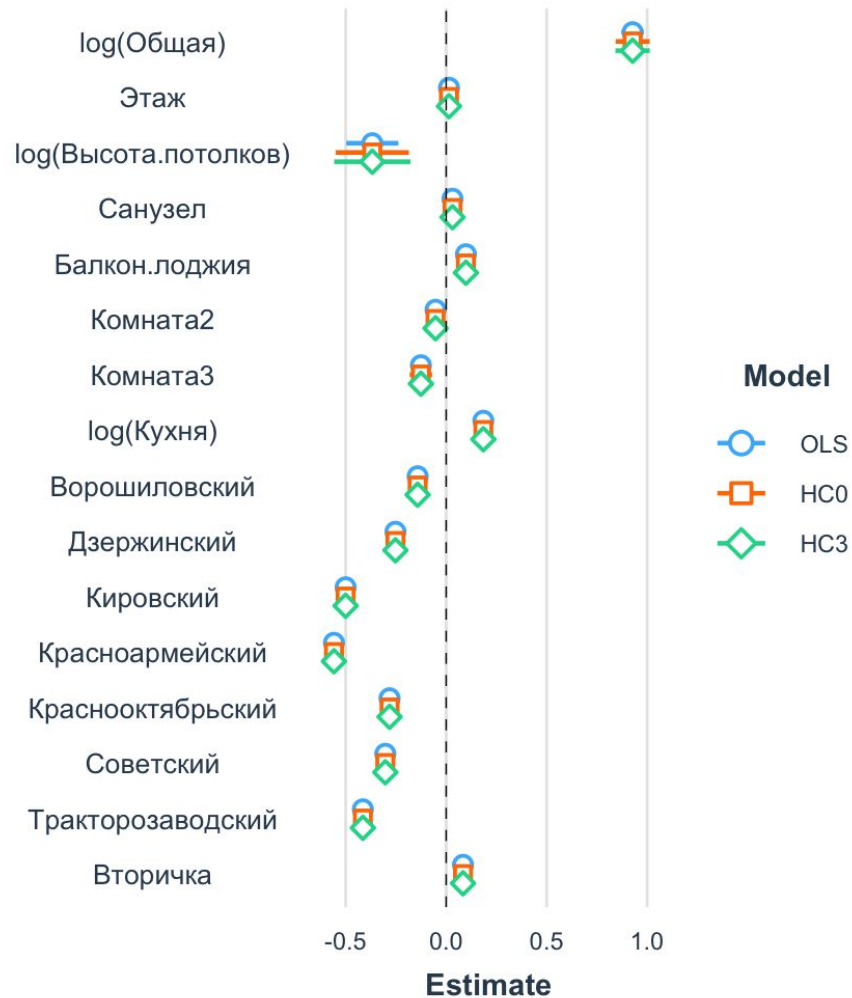


MODEL INFO:  
*Observations:* 3570  
*Dependent Variable:* log(Цена)  
*Type:* OLS linear regression

MODEL FIT:  
 $F(16,3553) = 718.68, p = 0.00$   
 $R^2 = 0.76$   
 $Adj. R^2 = 0.76$

*Standard errors: OLS*

	Est.	2.5%	97.5%	t val.	p	VIF
(Intercept)	11.19	10.99	11.39	108.90	0.00	
log(Общая)	0.93	0.87	0.98	32.50	0.00	6.14
Этаж	0.01	0.01	0.02	11.20	0.00	1.29
log(Высота.потолков)	-0.37	-0.50	-0.24	-5.55	0.00	1.12
Санузел	0.03	0.02	0.05	3.83	0.00	1.28
Балкон.лоджия	0.10	0.08	0.12	11.38	0.00	1.18
Комната2	-0.05	-0.08	-0.03	-3.74	0.00	3.61
Комната3	-0.13	-0.17	-0.09	-6.26	0.00	7.26
log(Кухня)	0.18	0.15	0.22	11.73	0.00	2.11
Ворошиловский	-0.14	-0.18	-0.11	-8.32	0.00	1.71
Дзержинский	-0.25	-0.28	-0.22	-17.24	0.00	2.28
Кировский	-0.50	-0.53	-0.47	-32.07	0.00	2.12
Красноармейский	-0.56	-0.59	-0.53	-36.06	0.00	2.12
Краснооктябрьский	-0.28	-0.31	-0.25	-18.12	0.00	2.01
Советский	-0.30	-0.33	-0.27	-19.57	0.00	2.06
Тракторозаводский	-0.41	-0.45	-0.38	-26.34	0.00	2.02
Вторичка	0.08	0.03	0.13	3.36	0.00	1.07



## Коэффициенты регрессии

### MODEL INFO:

Observations: 3570

Dependent Variable: log(Цена)

Type: OLS linear regression

### MODEL FIT:

$F(16,3553) = 718.68$ ,  $p = 0.00$

$R^2 = 0.76$

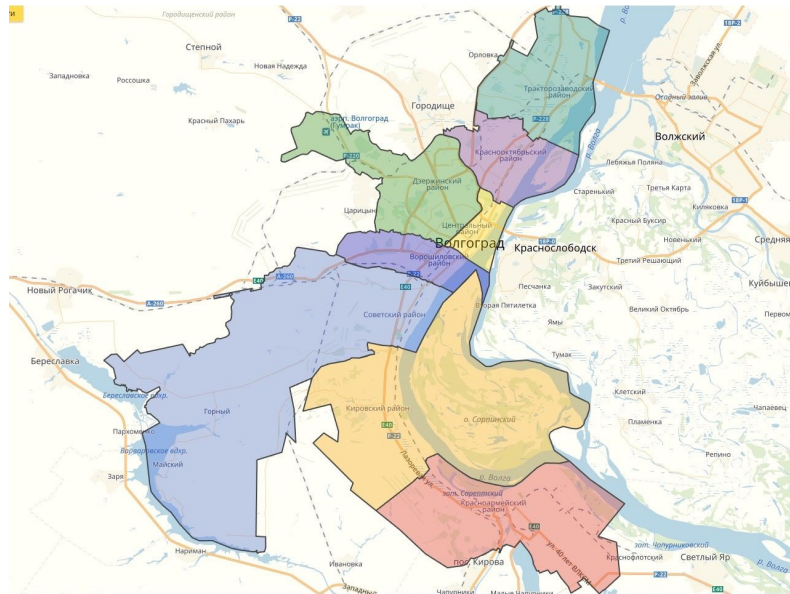
Adj.  $R^2 = 0.76$

Standard errors: Robust, type = HC3

	Est.	2.5%	97.5%	t val.	p	VIF
(Intercept)	11.19	10.91	11.47	78.00	0.00	
log(Общая)	0.93	0.84	1.01	21.25	0.00	6.14
Этаж	0.01	0.01	0.02	11.83	0.00	1.29
log(Высота.потолков)	-0.37	-0.56	-0.18	-3.80	0.00	1.12
Санузел	0.03	0.01	0.05	3.65	0.00	1.28
Балкон. лоджия	0.10	0.08	0.12	10.83	0.00	1.18
Комната2	-0.05	-0.09	-0.02	-2.81	0.01	3.61
Комната3	-0.13	-0.18	-0.07	-4.38	0.00	7.26
log(Кухня)	0.18	0.14	0.23	8.14	0.00	2.11
Ворошиловский	-0.14	-0.18	-0.11	-8.25	0.00	1.71
Дзержинский	-0.25	-0.28	-0.22	-17.91	0.00	2.28
Кировский	-0.50	-0.53	-0.47	-29.43	0.00	2.12
Красноармейский	-0.56	-0.59	-0.53	-36.13	0.00	2.12
Краснооктябрьский	-0.28	-0.31	-0.25	-19.41	0.00	2.01
Советский	-0.30	-0.33	-0.27	-19.78	0.00	2.06
Тракторозаводский	-0.41	-0.44	-0.38	-27.10	0.00	2.02
Вторичка	0.08	0.04	0.12	3.99	0.00	1.07

# Выводы

- 1) Увеличение общей площади на 1% увеличивает цену квартиры на 0,93%;
- 2) Увеличение площади кухни на 1% увеличивает цену квартиры на 0,18%;
- 3) Квартиры в Центральном районе самые дорогие. Расположение в каком - либо другом районе уменьшает цену квартиры



*Спасибо за внимание!*