

Район	Оценка	Стоимость, руб./сут.	Предоплата, руб.	Площадь, м^2	Страховой депозит, руб.	Кол-во постей	Спальные места	Кол-во кроватей	Курение	Вечеринки	Питомцы	балкон / лоджия	Лифт	Парковка	Этаж	Ремонт
Свердловский	6,0	1300	260	29	1000	4	4	2	нет	да	да	нет	нет	нет	Нижний	Косметический
Октябрьский	6,0	2250	450	40	1000	3	2	1	да	да	да	нет	да	да	Нижний	Евроремонт
Свердловский	7,4	3300	660	36	1000	4	4	2	нет	нет	нет	да	да	да	Верхний	Косметический
Свердловский	7,5	2750	690	25	2000	3	2	1	нет	да	да	да	да	нет	Верхний	Евроремонт
Октябрьский	7,8	3200	640	56	2000	4	4	2	нет	да	да	да	да	да	Верхний	Косметический
Свердловский	8,0	2440	450	43	2000	2	2	1	нет	да	да	да	да	да	Верхний	Косметический
Свердловский	8,2	2300	460	28	1000	2	2	1	да	нет	да	да	да	нет	Верхний	Косметический
Правобережный	8,5	2200	440	27	2000	4	4	2	да	да	да	да	да	нет	Верхний	Евроремонт
Правобережный	8,5	2500	625	36	1000	2	1	1	нет	нет	нет	да	нет	нет	Нижний	Косметический
Свердловский	8,6	3300	820	18	1000	2	2	1	да	нет	да	да	да	да	Средний	Евроремонт
Октябрьский	8,6	2700	540	35	1000	3	2	2	нет	нет	да	да	нет	да	Нижний	Дизайнерский
Свердловский	8,7	2600	520	22	1500	2	2	1	нет	да	да	да	да	нет	Верхний	Евроремонт
Свердловский	8,9	3000	600	36	1000	3	3	2	да	да	да	да	да	нет	Верхний	Евроремонт
Свердловский	8,9	2750	550	36	1000	4	4	2	нет	нет	нет	да	да	нет	Верхний	Косметический

Данные на 19 февраля 2024 года

Источник: sutochno.ru



Да/Нет



Да/Нет



Да/Нет



Да/Нет



Да/Нет



Да/Нет

Диаграмма рассеяния

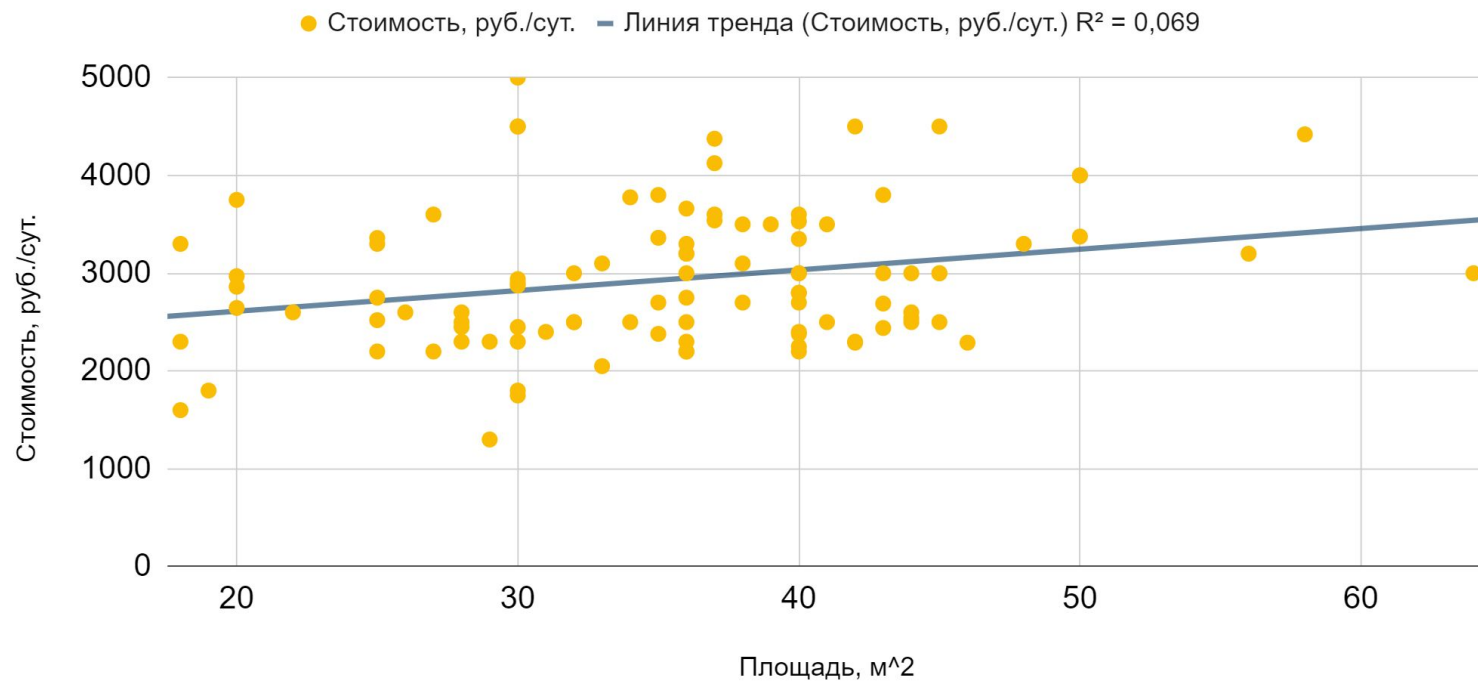


Диаграмма рассеяния

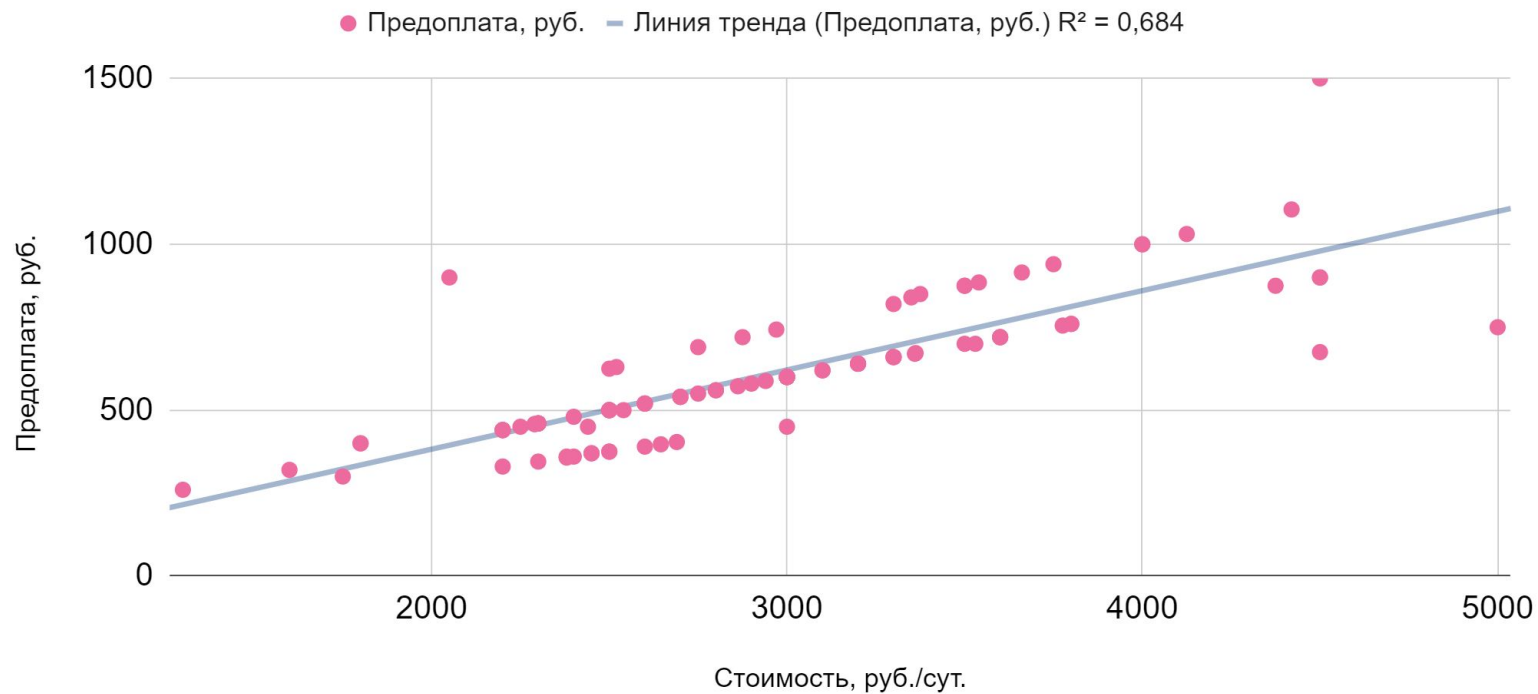
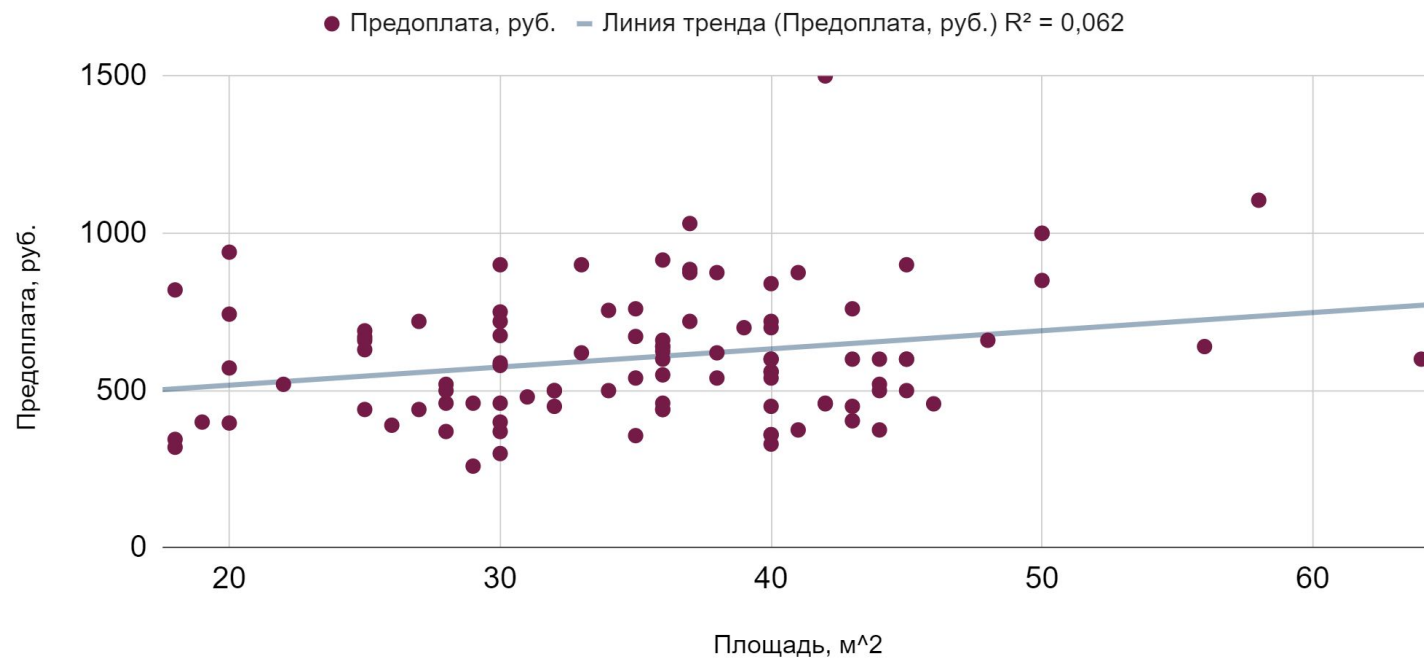
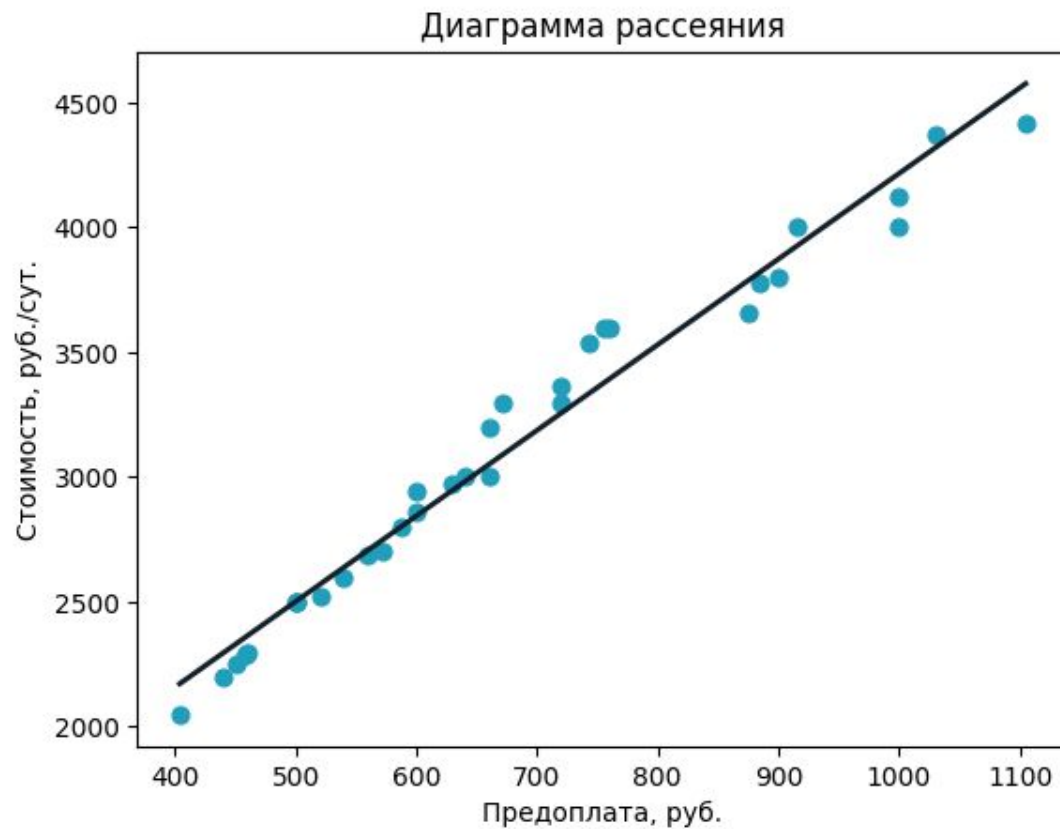


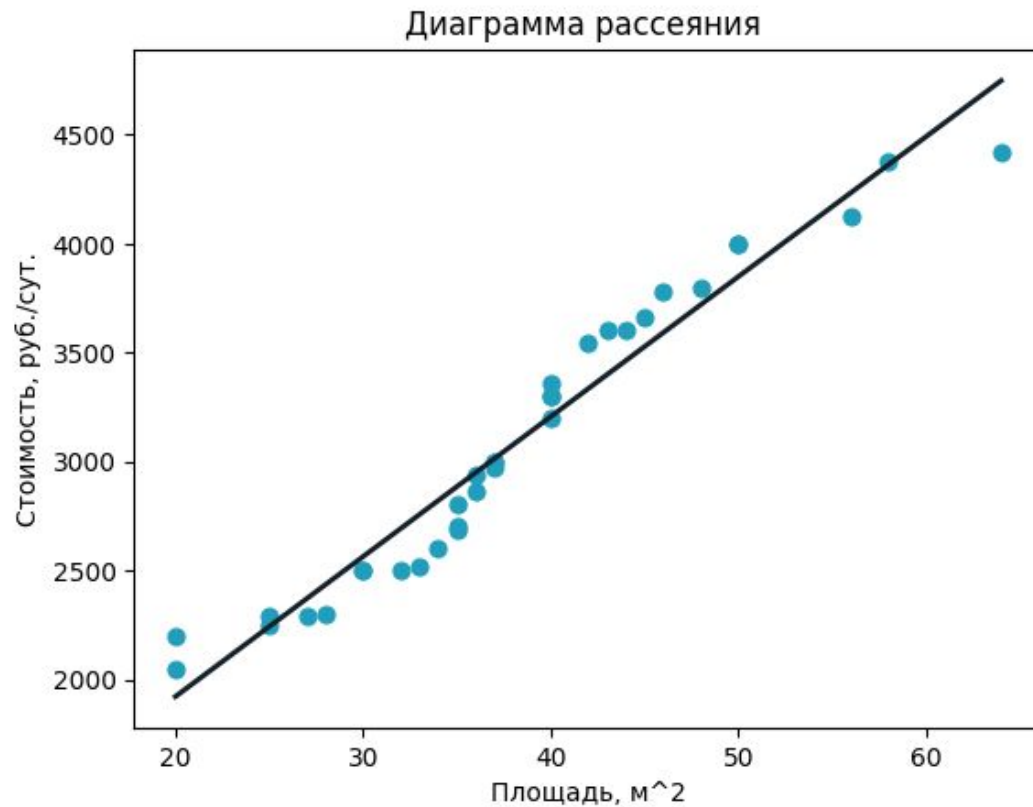
Диаграмма рассеяния



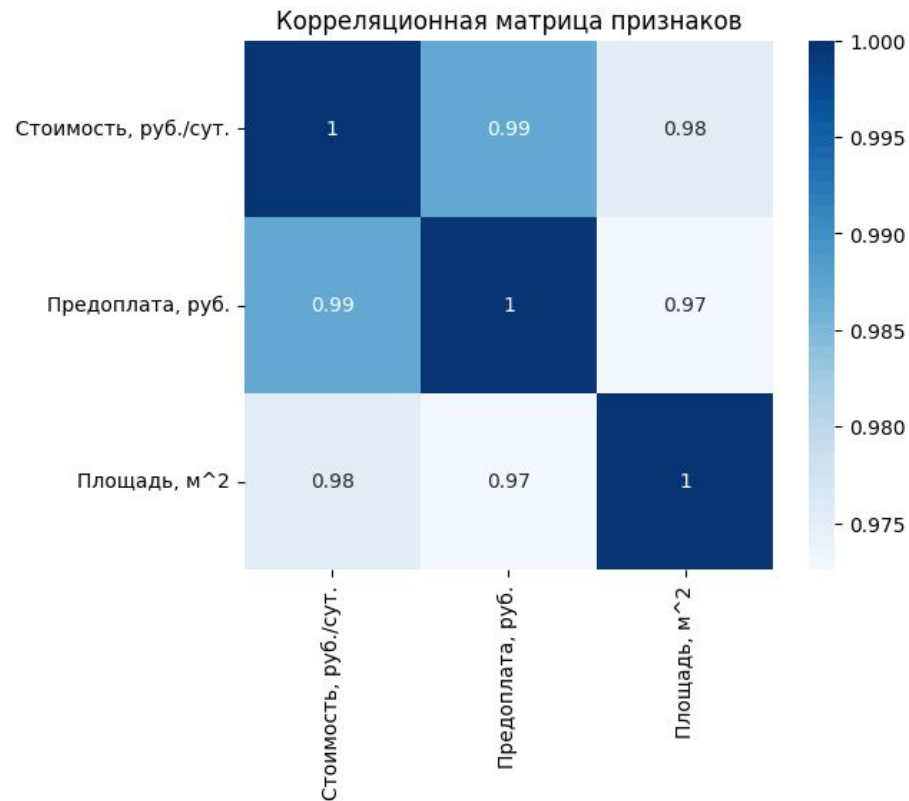
Октябрьский район



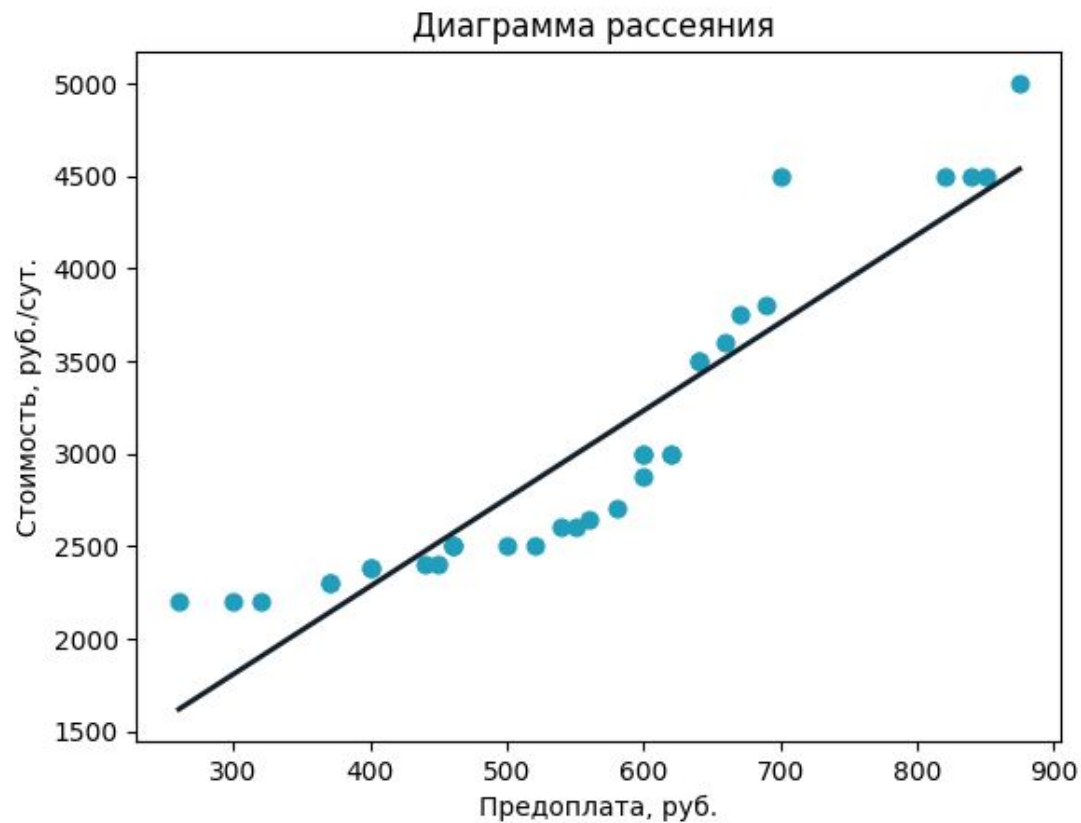
Октябрьский район



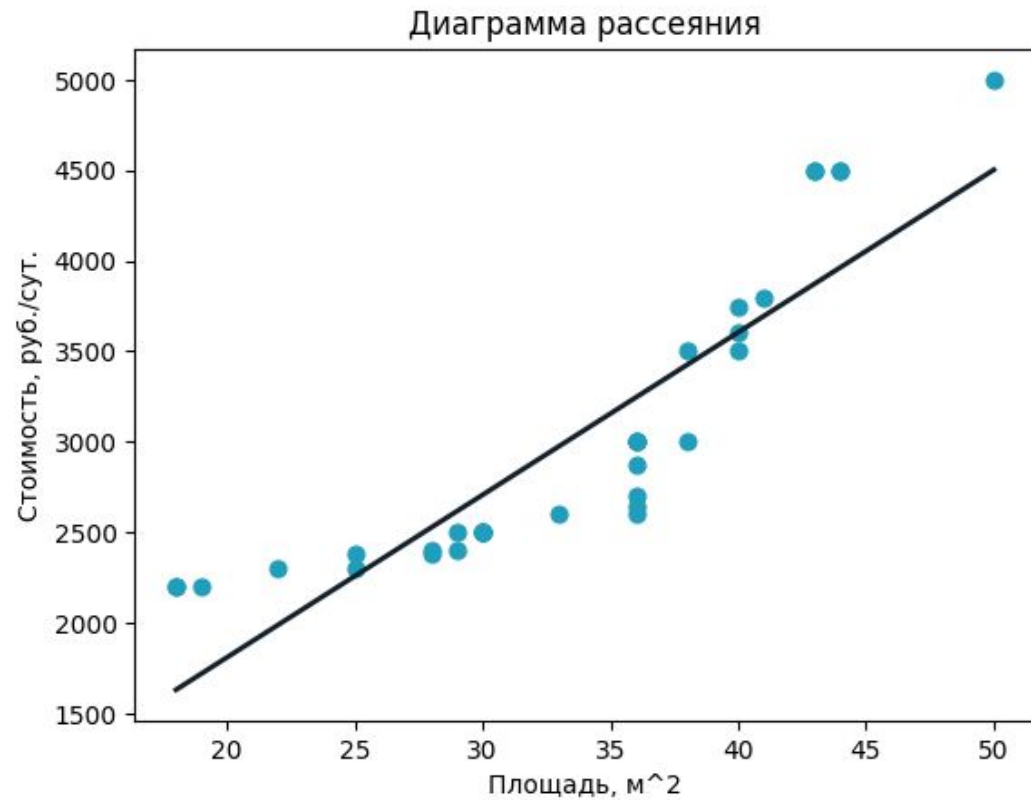
Октябрьский район



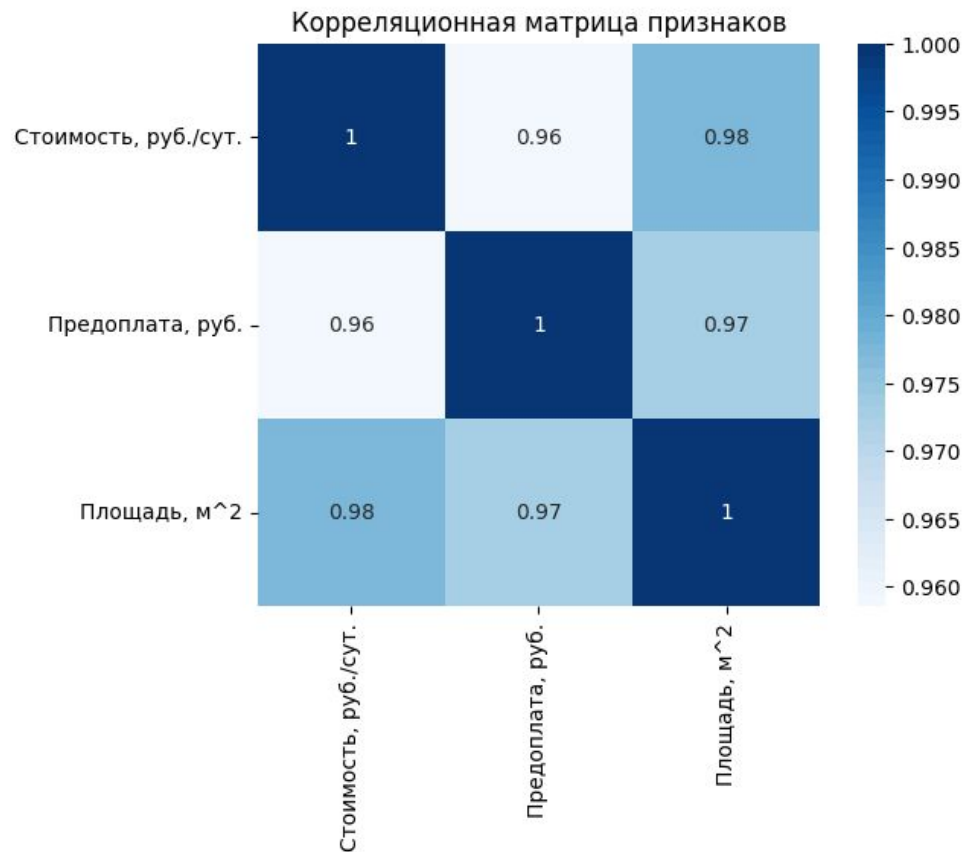
Свердловский район



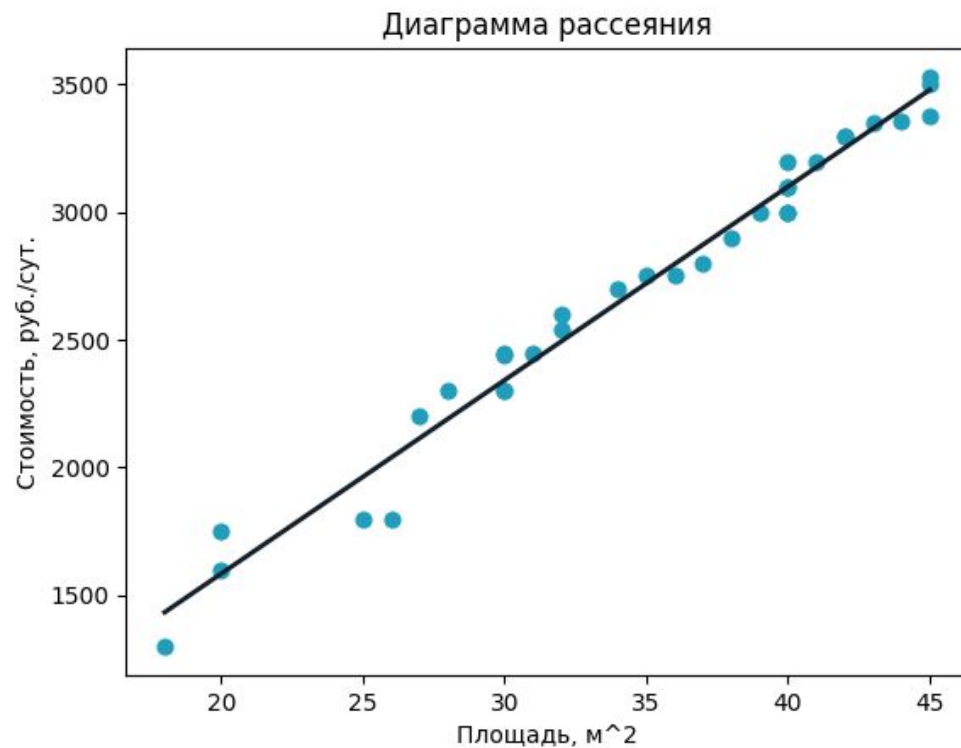
Свердловский район



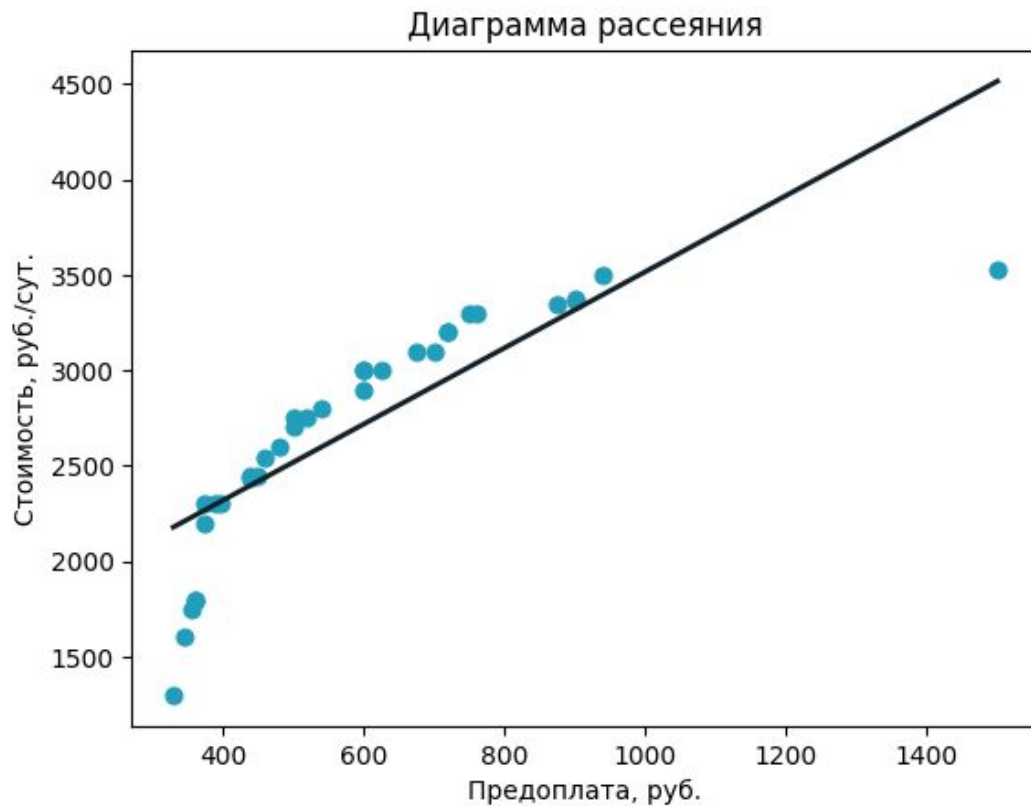
Свердловский район



Правобережный район



Правобережный район



Правобережный район



Разведочный анализ данных

Диаграммы были построены при помощи библиотек Python:

seaborn

numpy

matplotlib

pandas

Описательные статистики получены при помощи надстройки “Анализ данных” ППП Excel

Количественный
показатель

Качественный признак

Стоимость

Район
Этаж

Предоплата

Район
Этаж

Площадь

Район

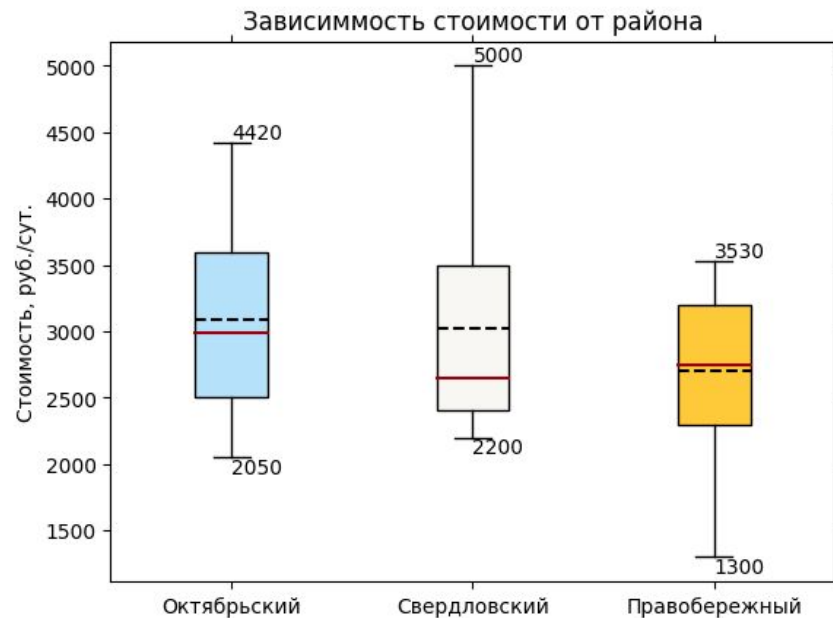
Стоимость - Район

Задача:

Выяснить зависимость стоимости посуточной аренды однокомнатных квартир в г. Иркутске от района, в котором располагается квартира

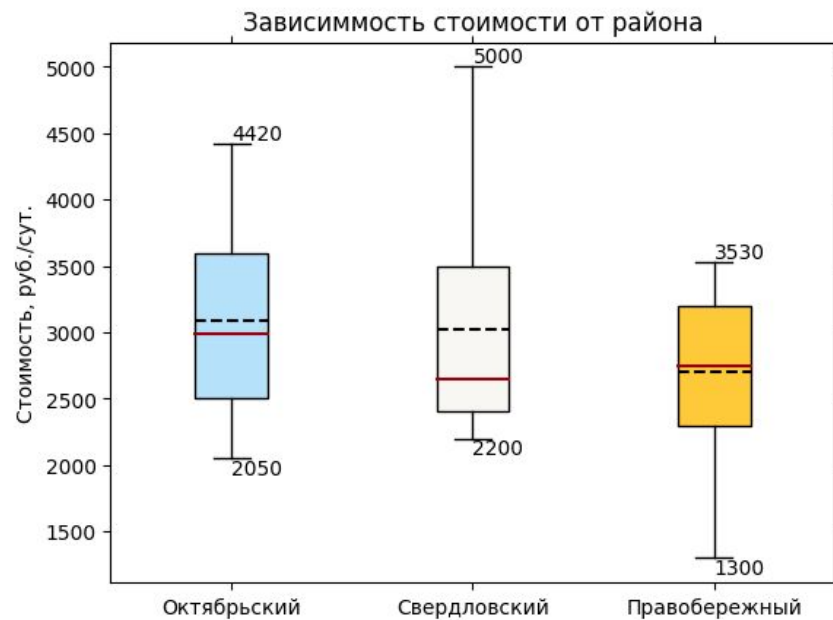
В качестве анализируемых признаков были выбраны стоимость посуточной аренды и район расположения (Свердловский, Октябрьский, Правобережный). Объем выборки составляет 33, 34 и 33 наблюдения соответственно

Стоимость - Район

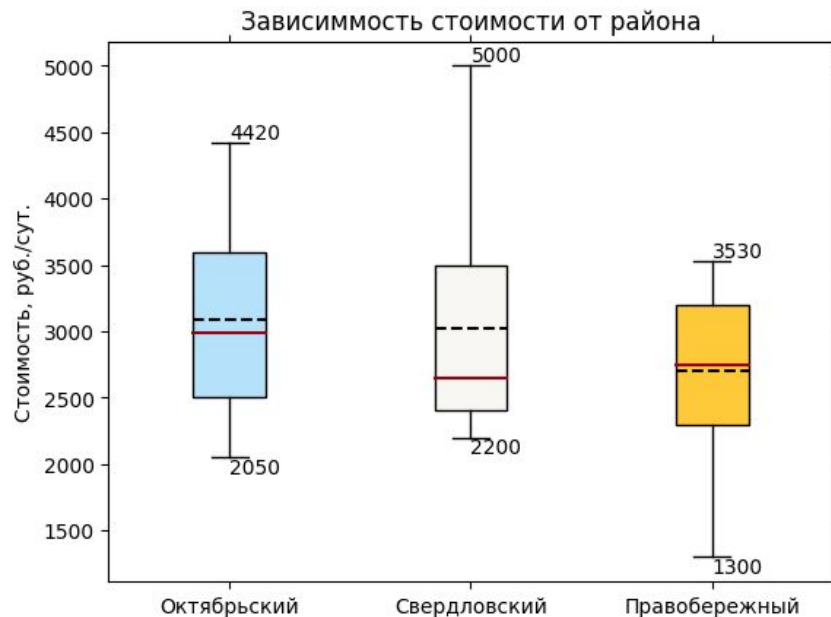


	Октябрьский	Правобережный	Свердловский
Среднее	3088,85	3025,15	2698,33
Медиана	2985,00	2645,00	2750,00
Стандартное отклонение	672,83	814,56	590,50
Асимметричность	0,36	1,07	-0,62
Минимум	2050,00	2200,00	1300,00
Максимум	4420,00	5000,00	3530,00
Коэффициент вариации	0,22	0,27	0,22
Q3	3615,25	3550,00	3200,00
Q1	2500,00	2400,00	2300,00
Вариационный размах	1115,25	1150,00	900,00

Стоимость - Район



Стоимость - Район



Наиболее близким к нормальному является распределение стоимости по Октябрьскому району

В распределении стоимости по Правобережному району наблюдается асимметрия

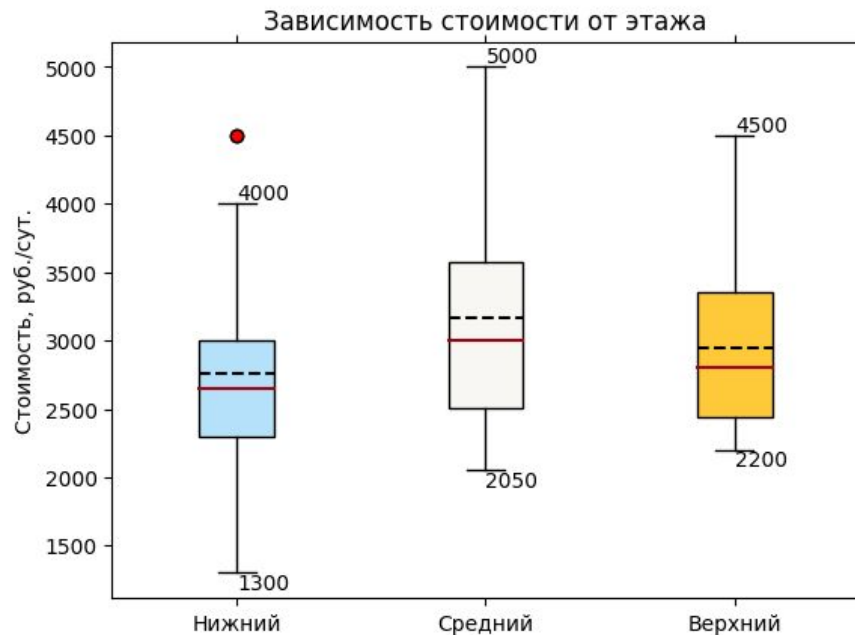
Стоимость - Этаж

Задача:

Выяснить зависимость стоимости посуточной аренды однокомнатных квартир в г. Иркутске от уровня этажа, на котором располагается квартира

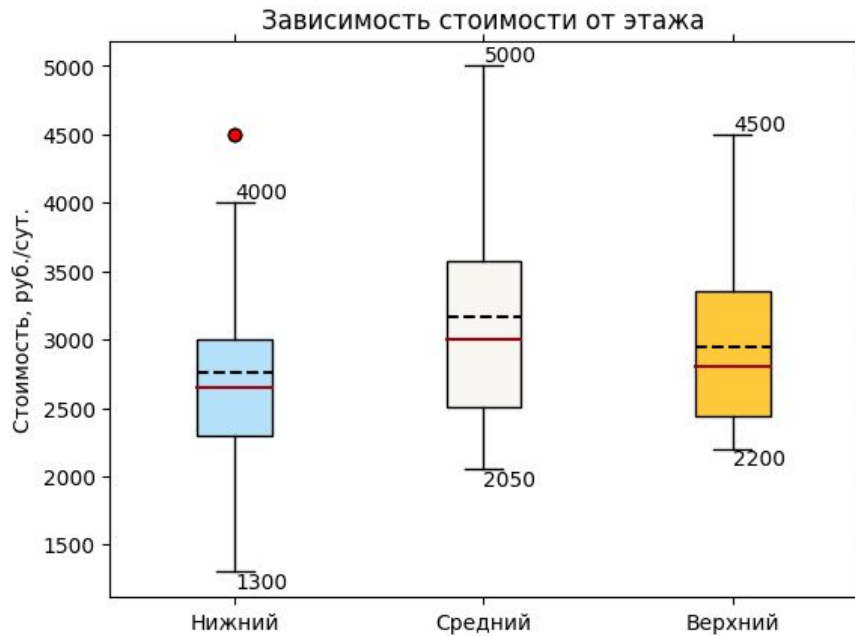
В качестве анализируемых признаков были выбраны стоимость посуточной аренды и этаж (Нижний, Средний, Верхний). Объем выборки составляет 32, 26 и 41 наблюдение соответственно

Стоимость - Этаж



	Нижний	Средний	Верхний
Среднее	2757,88	3162,54	2942,93
Медиана	2645,00	3000,00	2800,00
Стандартное отклонение	762,00	744,19	621,31
Асимметричность	0,52	0,78	0,76
Минимум	1300,00	2050,00	2200,00
Максимум	4500,00	5000,00	4500,00
Коэффициент вариации	0,28	0,24	0,21
Q3	3150,00	3615,25	3356,00
Q1	2300,00	2500,00	2420,00
Вариационный размах	850,00	1115,25	936,00

Стоимость - Этаж



Стоимость - Этаж

Среднее	2757,88
Стандартная ошибка	132,65
Медиана	2645,00
Мода	2500,00
Стандартное отклонение	762,00
Дисперсия выборки	580637,55
Эксцесс	0,30
Асимметричность	0,52
Интервал	3200,00
Минимум	1300,00
Максимум	4500,00
Сумма	91010,00
Счет	33,00
Наибольший(1)	4500,00
Наименьший(1)	1300,00
Уровень надежности(95,0%)	270,19

Критерий трёх сигм

$$\mu + 3\sigma = 5043,87$$

$$4500 < 5043,87$$

По критерию трёх сигм проверяемое значение не является выбросом

Стоимость - Этаж

Среднее	2757,88
Стандартная ошибка	132,65
Медиана	2645,00
Мода	2500,00
Стандартное отклонение	762,00
Дисперсия выборки	580637,55
Эксцесс	0,30
Асимметричность	0,52
Интервал	3200,00
Минимум	1300,00
Максимум	4500,00
Сумма	91010,00
Счет	33,00
Наибольший(1)	4500,00
Наименьший(1)	1300,00
Уровень надежности(95,0%)	270,19

Критерий Граббса

$$D_n = \frac{x_{max} - \bar{x}}{S} = \frac{4500 - 2757,88}{762} \approx 2,29$$

$$D_{\text{крит.}} = 2,83$$

$$D_n < D_{\text{крит.}}$$

По критерию Граббса проверяемое значение не является выбросом

Стоимость - Этаж

Среднее	2757,88
Стандартная ошибка	132,65
Медиана	2645,00
Мода	2500,00
Стандартное отклонение	762,00
Дисперсия выборки	580637,55
Эксцесс	0,30
Асимметричность	0,52
Интервал	3200,00
Минимум	1300,00
Максимум	4500,00
Сумма	91010,00
Счет	33,00
Наибольший(1)	4500,00
Наименьший(1)	1300,00
Уровень надежности(95,0%)	270,19

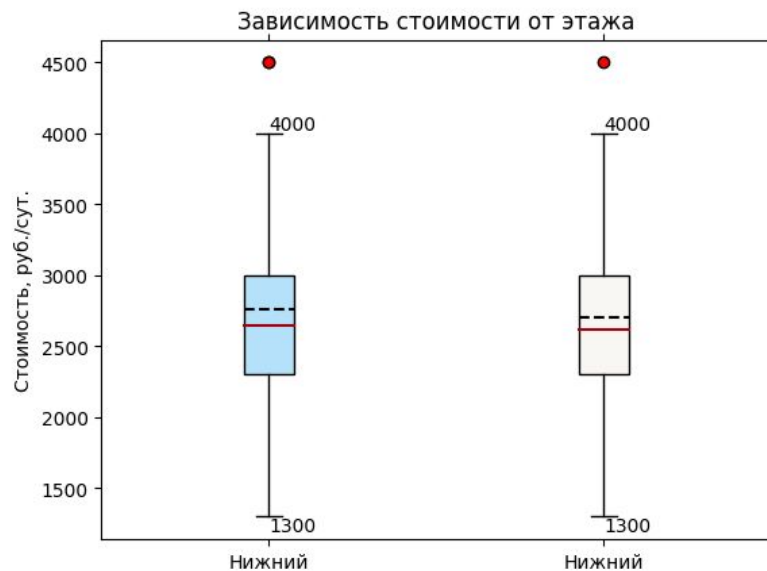
Критерий интерквартильных
расстояний

$$IQR = 3150 - 2300 = 850$$

$$Q3 + 1,5 * IQR = 4425$$

$$4500 > 4425$$

Стоимость - Этаж



	До исключения выбросов	После исключения выбросов
Среднее	2757,88	2703,44
Медиана	2645,00	2622,50
Мода	2500,00	2500,00
Стандартное отклонение	762,00	705,98
Дисперсия выборки	580637,55	498405,54
Эксцесс	0,30	0,41
Асимметричность	0,52	0,40
Минимум	1300,00	1300,00
Максимум	4500,00	4500,00
Счет	33,00	32,00
Наибольший(1)	4500,00	4500,00
Наименьший(1)	1300,00	1300,00
Уровень надежности(95,0%)	270,19	254,53
Коэффициент вариации	0,28	0,26

Стоимость - Этаж

Среднее	2703,44
Стандартная ошибка	124,80
Медиана	2622,50
Мода	2500,00
Стандартное отклонение	705,98
Дисперсия выборки	498405,54
Эксцесс	0,41
Асимметричность	0,40
Интервал	3200,00
Минимум	1300,00
Максимум	4500,00
Сумма	86510,00
Счет	32,00
Наибольший(1)	4500,00
Наименьший(1)	1300,00
Уровень надежности(95,0%)	254,53

Критерий трёх сигм

$$\mu + 3\sigma = 4821,37$$

$$4500 < 4821,37$$

По критерию трёх сигм проверяемое значение не является выбросом

Стоимость - Этаж

Среднее	2703,44
Стандартная ошибка	124,80
Медиана	2622,50
Мода	2500,00
Стандартное отклонение	705,98
Дисперсия выборки	498405,54
Эксцесс	0,41
Асимметричность	0,40
Интервал	3200,00
Минимум	1300,00
Максимум	4500,00
Сумма	86510,00
Счет	32,00
Наибольший(1)	4500,00
Наименьший(1)	1300,00
Уровень надежности(95,0%)	254,53

Критерий Граббса

$$D_n = \frac{x_{max} - \bar{x}}{S} = \frac{4500 - 2703,44}{705,98} \approx 2,54$$

$$D_{\text{крит.}} = 2,818$$

$$D_n < D_{\text{крит.}}$$

По критерию Граббса проверяемое значение не является выбросом

Стоимость - Этаж

Среднее	2703,44
Стандартная ошибка	124,80
Медиана	2622,50
Мода	2500,00
Стандартное отклонение	705,98
Дисперсия выборки	498405,54
Эксцесс	0,41
Асимметричность	0,40
Интервал	3200,00
Минимум	1300,00
Максимум	4500,00
Сумма	86510,00
Счет	32,00
Наибольший(1)	4500,00
Наименьший(1)	1300,00
Уровень надежности(95,0%)	254,53

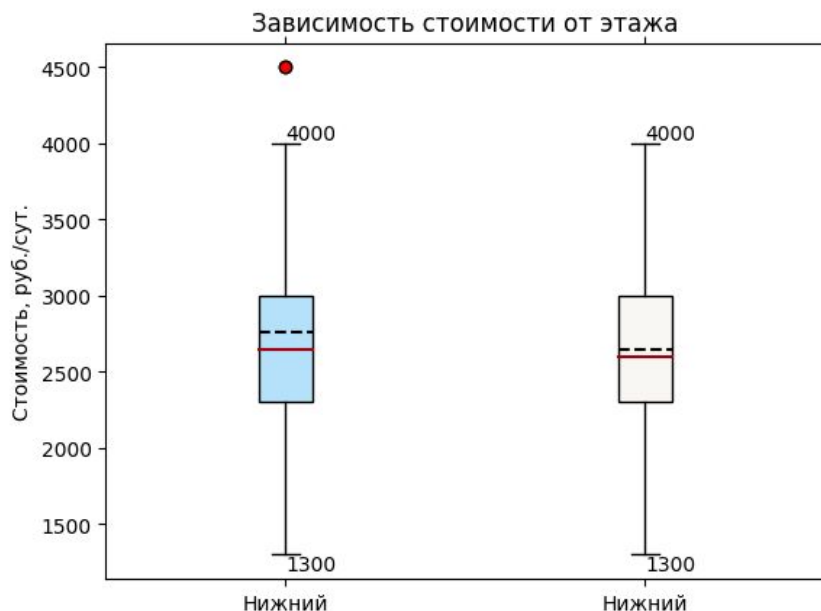
Критерий интерквартильных
расстояний

$$IQR = 3000 - 2300 = 700$$

$$Q3 + 1,5 * IQR = 4050$$

$$4500 > 4050$$

Стоимость - Этаж



	До исключения выбросов	После исключения выбросов
Среднее	2757,88	2645,48
Медиана	2645,00	2600,00
Мода	2500,00	2500,00
Стандартное отклонение	762,00	635,58
Дисперсия выборки	580637,55	403960,59
Эксцесс	0,30	-0,07
Асимметричность	0,52	0,06
Минимум	1300,00	1300,00
Максимум	4500,00	4000,00
Счет	33,00	31,00
Наибольший(1)	4500,00	4000,00
Наименьший(1)	1300,00	1300,00
Уровень надежности(95,0%)	270,19	233,13
Коэффициент вариации	0,28	0,24

Выводы

1. Распределение стало более близким к нормальному
2. Выборка стала более однородной
3. Уменьшилась дисперсия

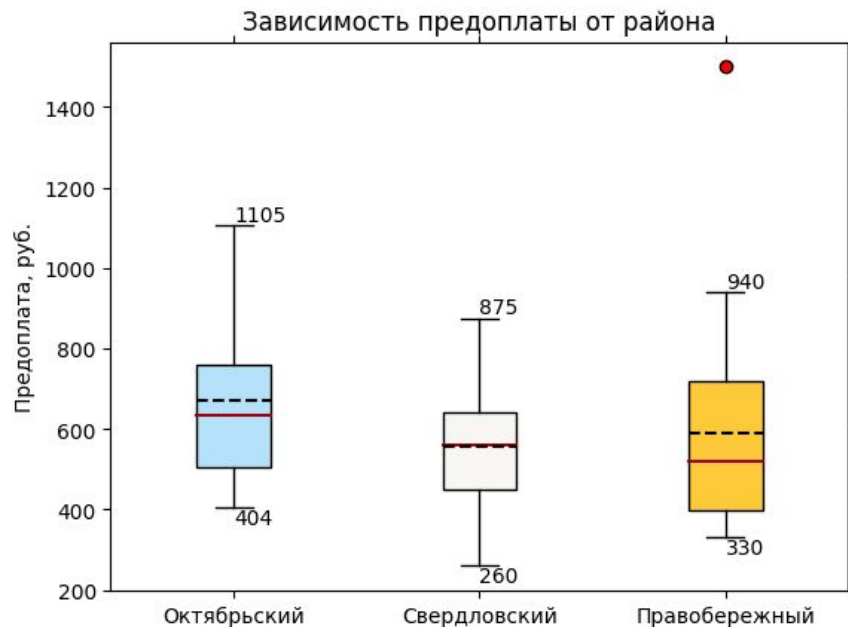
Предоплата - Район

Задача:

Выяснить зависимость размера предоплаты за посуточную аренду однокомнатных квартир в г. Иркутске от района, в котором располагается квартира

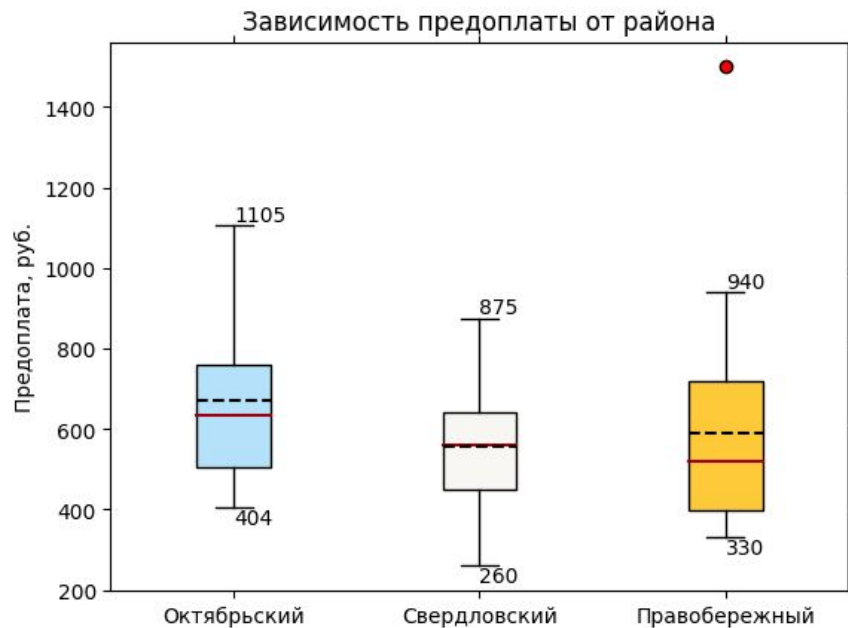
В качестве анализируемых признаков были выбраны размер предоплаты посуточной аренды и район (Октябрьский, Свердловский, Правобережный). Объем выборки составляет 34, 33 и 33 наблюдений соответственно

Предоплата - Район



	Октябрьский	Свердловский	Правобережный
Среднее	671,21	556,52	590,42
Медиана	635,00	560,00	520,00
Мода	500,00	460,00	600,00
Стандартное отклонение	193,63	159,44	244,17
Дисперсия выборки	37490,96	25422,63	59618,19
Эксцесс	-0,55	-0,41	4,68
Асимметричность	0,67	0,21	1,77
Минимум	404,00	260,00	330,00
Максимум	1105,00	875,00	1500,00
Q3	788,75	650,00	720,00
Q1	500,00	445,00	393,50
Вариационный размах	288,75	205,00	326,50
Коэффициент вариации	0,29	0,29	0,41

Предоплата - Район



Предоплата - Район

Среднее	590,42
Стандартная ошибка	42,50
Медиана	520,00
Мода	600,00
Стандартное отклонение	244,17
Дисперсия выборки	59618,19
Эксцесс	4,68
Асимметричность	1,77
Интервал	1170,00
Минимум	330,00
Максимум	1500,00
Сумма	19484,00
Счет	33,00
Наибольший(1)	1500,00
Наименьший(1)	330,00
Уровень надежности(95,0%)	86,58

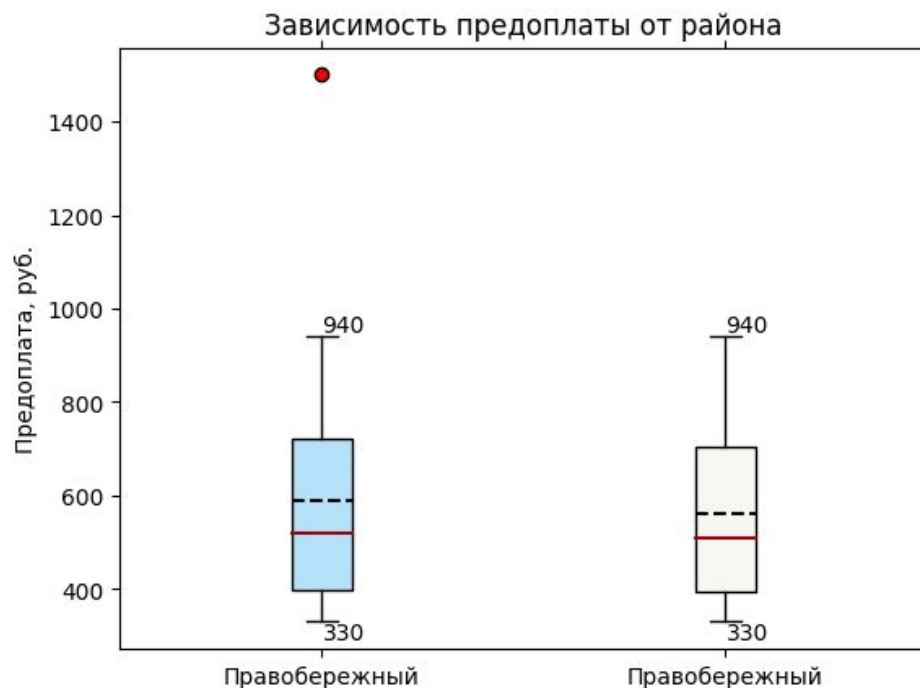
Критерий трёх сигм

$$\mu + 3\sigma = 1322,93$$

$$1500 > 1322,93$$

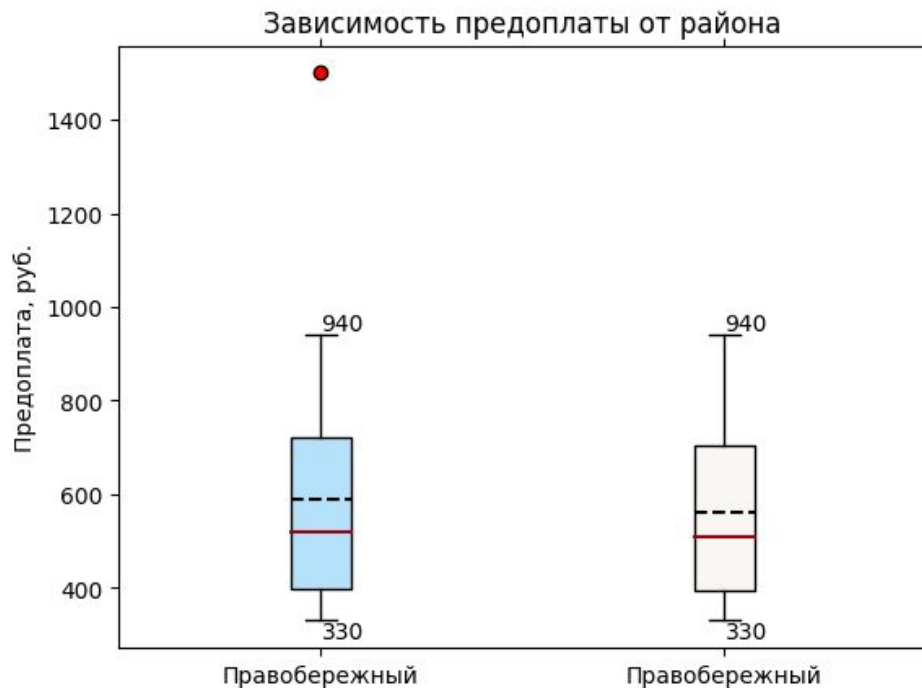
По критерию трёх сигм проверяемое значение является выбросом

Предоплата - Район



	До исключения выбросов	После исключения выбросов
Среднее	590,42	562,00
Медиана	520,00	510,00
Мода	600,00	600,00
Стандартное отклонение	244,17	184,44
Дисперсия выборки	59618,19	34019,35
Экссесс	4,68	-0,75
Асимметричность	1,77	0,61
Минимум	330,00	330,00
Максимум	1500,00	940,00
Сумма	19484,00	17984,00
Счет	33,00	32,00
Наибольший(1)	1500,00	940,00
Наименьший(1)	330,00	330,00
Уровень надежности(95,0%)	86,58	66,50
Коэффициент вариации	0,41	0,33

Выводы



1. Распределение стало более близким к нормальному
2. Выборка стала более однородной
3. Дисперсия значительно уменьшилась

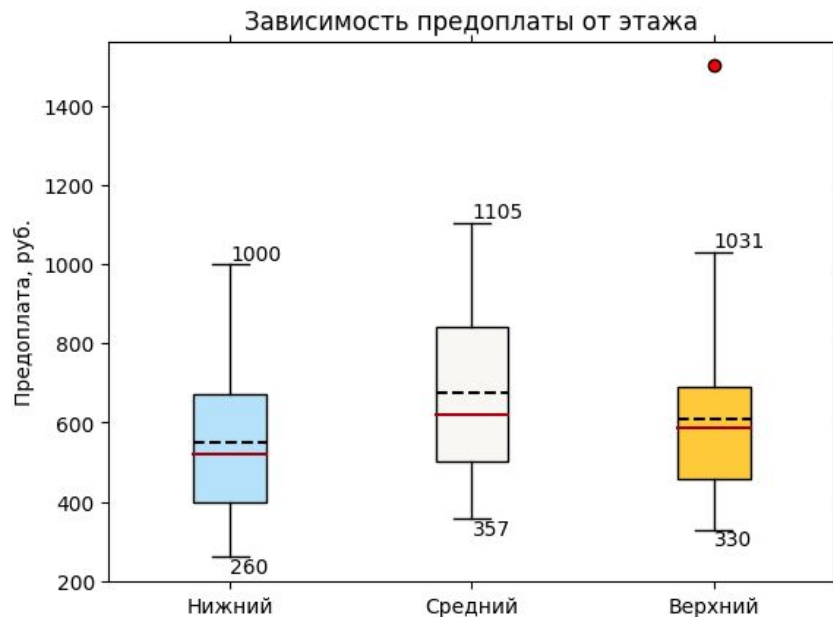
Предоплата - Этаж

Задача:

Выяснить зависимость размера предоплаты посуточной аренды однокомнатных квартир в г. Иркутске от уровня этажа, на котором располагается квартира

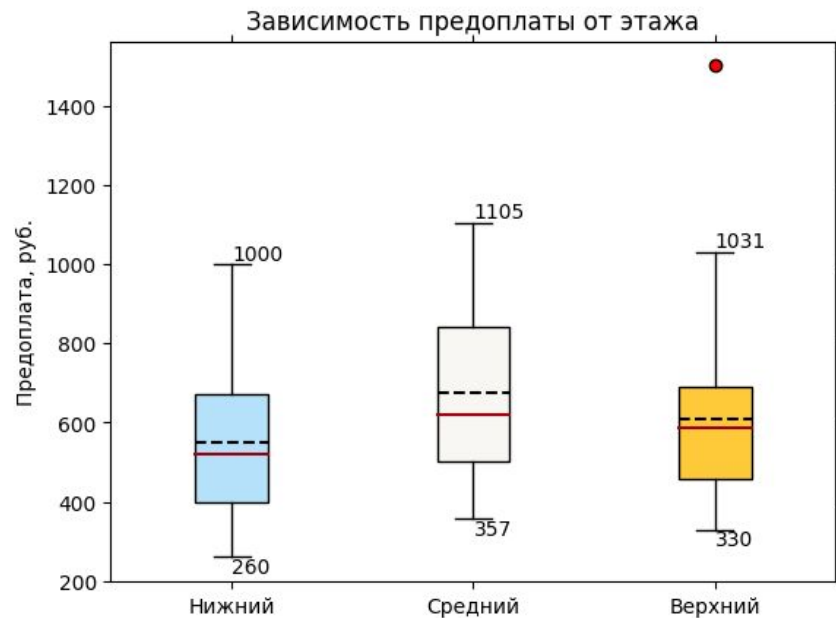
В качестве анализируемых признаков были выбраны размер предоплаты и этаж (Нижний, Средний, Верхний). Объем выборки составляет 32, 26 и 41 наблюдение соответственно

Предоплата - Этаж



	Нижний	Средний	Верхний
Среднее	550,61	674,69	608,73
Медиана	520,00	620,00	588,00
Мода	540,00	500,00	440,00
Стандартное отклонение	182,97	197,89	218,94
Дисперсия выборки	33479,56	39158,86	47932,90
Экссесс	0,15	-0,70	5,73
Асимметричность	0,70	0,40	1,86
Минимум	260,00	357,00	330,00
Максимум	1000,00	1105,00	1500,00
Q3	672,50	856,25	695,00
Q1	400,00	500,00	454,00
Вариационный размах	272,50	356,25	241,00
Коэффициент вариации	0,33	0,29	0,36

Предоплата - Этаж



Предоплата - Этаж

Среднее	608,73
Стандартная ошибка	34,19
Медиана	588,00
Мода	440,00
Стандартное отклонение	218,94
Дисперсия выборки	47932,90
Эксцесс	5,73
Асимметричность	1,86
Интервал	1170,00
Минимум	330,00
Максимум	1500,00
Сумма	24958,00
Счет	41,00
Наибольший(1)	1500,00
Наименьший(1)	330,00
Уровень надежности(95,0%)	69,10

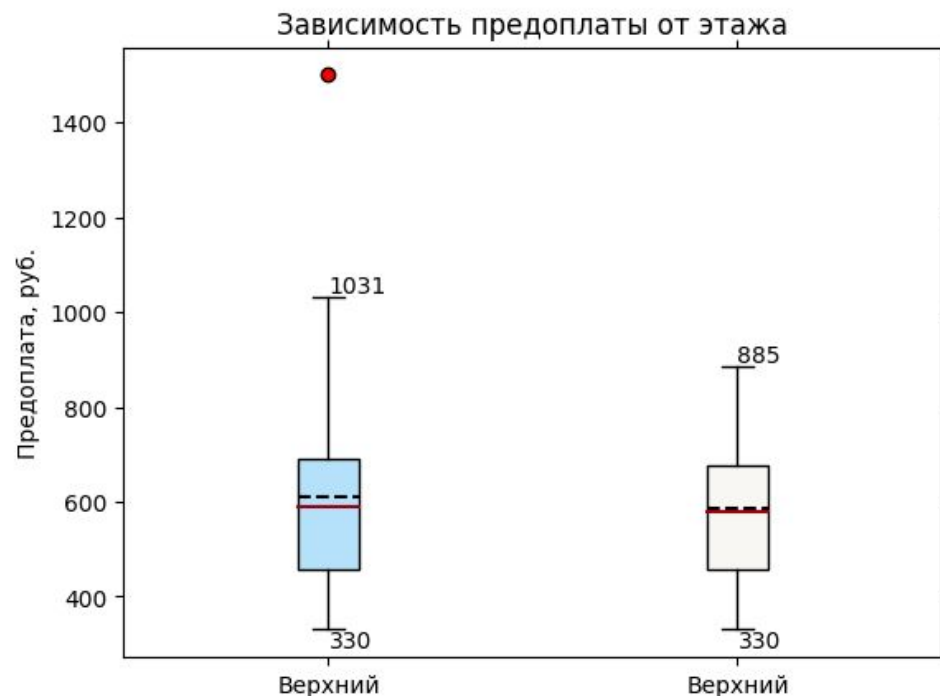
Критерий трёх сигм

$$\mu + 3\sigma = 1265,54$$

$$1500 > 1265,54$$

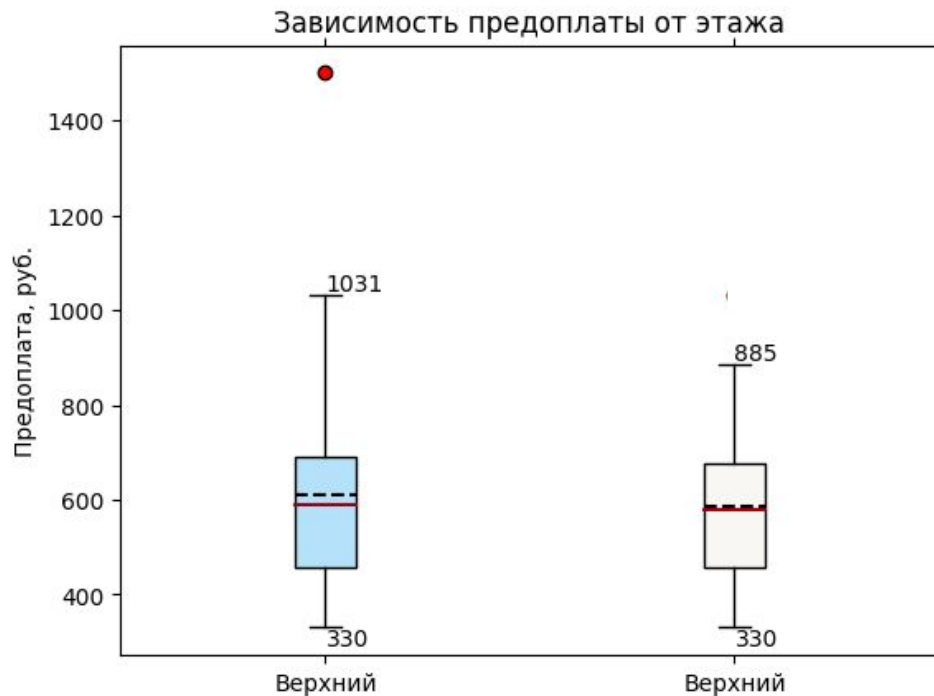
По критерию трёх сигм проверяемое значение является выбросом

Предоплата - Этаж



	До исключения выбросов	После исключения выбросов
Среднее	608,73	586,45
Стандартная ошибка	34,19	26,59
Медиана	588,00	580,00
Мода	440,00	440,00
Стандартное отклонение	218,94	168,18
Дисперсия выборки	47932,90	28284,56
Экссесс	5,73	-0,07
Асимметричность	1,86	0,62
Интервал	1170,00	701,00
Минимум	330,00	330,00
Максимум	1500,00	1031,00
Сумма	24958,00	23458,00
Счет	41,00	40,00
Наибольший(1)	1500,00	1031,00
Наименьший(1)	330,00	330,00
Уровень надежности(95,0%)	69,10	53,79
Коэффициент вариации	0,36	0,29

Выводы



1. Распределение стало более близким к нормальному
2. Выборка стала более однородной
3. Дисперсия значительно уменьшилась

Выводы

Выбросам на двух последних графиках соответствует одна запись из матрицы “Объект-Признак”:

42	Правобережный	4500	1500	42	2	нет	нет	да	1	Верхний	1	да	да	нет	1000	евроремонт
43	Правобережный	3500	700	39	2	нет	нет	нет	2	Верхний	1	да	нет	нет	1000	косметический
44	Правобережный	2500	375	41	3	нет	нет	да	6	Верхний	3	да	да	нет	1000	евроремонт
45	Правобережный	2400	360	40	3	нет	нет	да	4	Верхний	2	да	да	да	1000	евроремонт

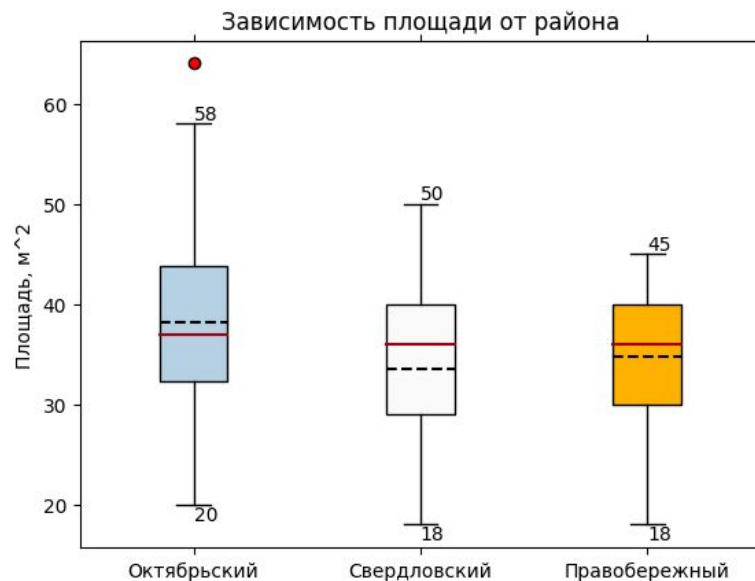
Площадь - Район

Задача:

Выяснить зависимость размера площади однокомнатных квартир, предоставляемых для посуточной аренды в г. Иркутске от района, в котором располагается квартира

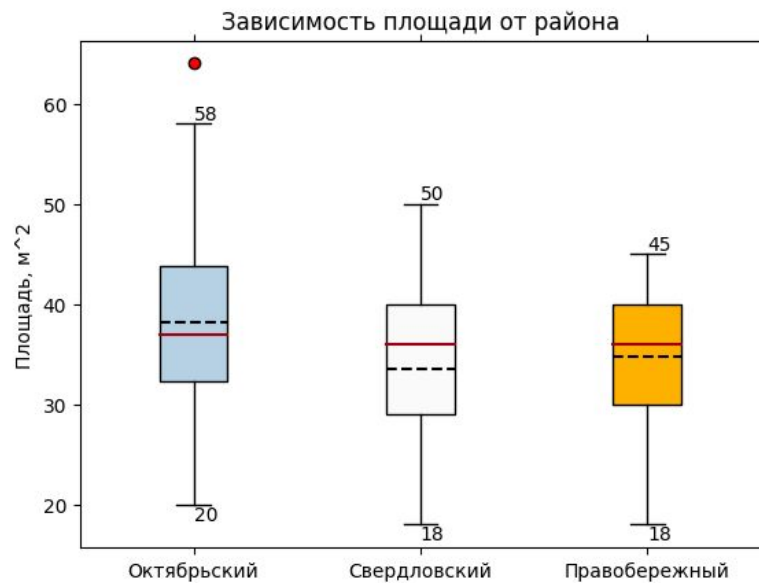
В качестве анализируемых признаков были выбраны площадь квартиры и район расположения (Свердловский, Октябрьский, Правобережный). Объем выборок составляет 33, 34 и 33 наблюдения соответственно

Площадь - Район



	Октябрьский	Свердловский	Правобережный
Среднее	38,18	33,55	34,70
Медиана	37,00	36,00	36,00
Мода	40,00	36,00	40,00
Стандартное отклонение	10,21	8,02	7,70
Дисперсия выборки	104,33	64,32	59,34
Эксцесс	0,29	-0,42	-0,61
Асимметричность	0,47	-0,27	-0,53
Минимум	20,00	18,00	18,00
Максимум	64,00	50,00	45,00
Q3	44,00	41,00	40,00
Q1	32,00	30,00	29,00
Вариационный размах	12,00	11,00	11,00
Коэффициент вариации	0,27	0,24	0,22

Площадь - Район



Площадь - Район

Среднее	38,18
Стандартная ошибка	1,75
Медиана	37,00
Мода	40,00
Стандартное отклонение	10,21
Дисперсия выборки	104,33
Эксцесс	0,29
Асимметричность	0,47
Интервал	44,00
Минимум	20,00
Максимум	64,00
Сумма	1298,00
Счет	34,00
Наибольший(1)	64,00
Наименьший(1)	20,00
Уровень надежности(95,0%)	3,56

Критерий трёх сигм

$$\mu + 3\sigma = 68,82$$

$$64 < 68,82$$

По критерию трёх сигм проверяемое значение не является выбросом

$$D_n \approx 2,53$$

$$D_{\text{крит.}} \approx 2,842$$

$$D_n < D_{\text{крит.}}$$

По критерию Граббса проверяемое значение не является выбросом

Площадь - Район

Среднее	2703,44
Стандартная ошибка	124,80
Медиана	2622,50
Мода	2500,00
Стандартное отклонение	705,98
Дисперсия выборки	498405,54
Эксцесс	0,41
Асимметричность	0,40
Интервал	3200,00
Минимум	1300,00
Максимум	4500,00
Сумма	86510,00
Счет	32,00
Наибольший(1)	4500,00
Наименьший(1)	1300,00
Уровень надежности(95,0%)	254,53

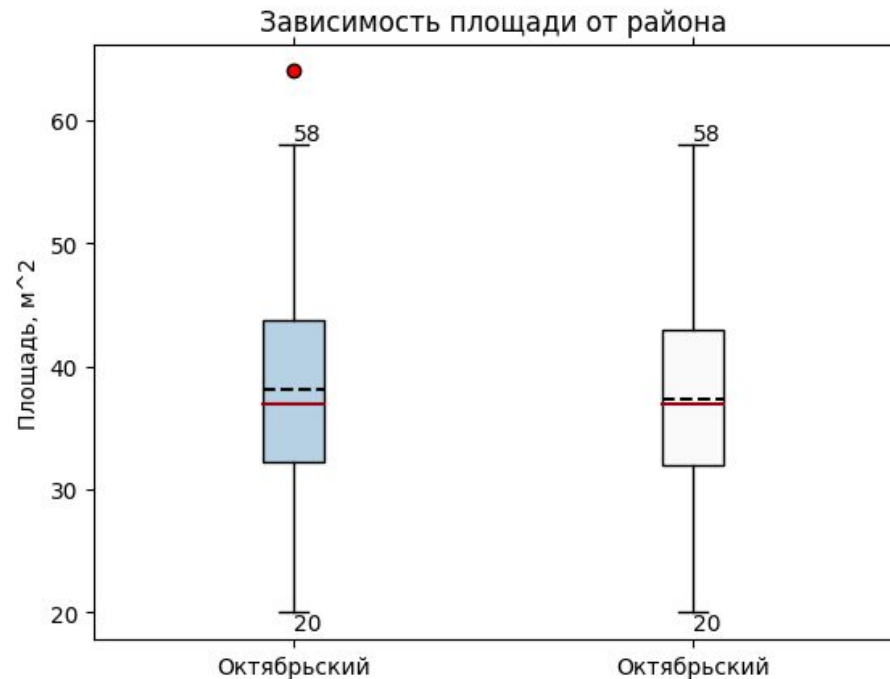
Метод интерквартильных расстояний

$$IQR = 44 - 32 = 12$$

$$Q3 + 1,5 * IQR = 62$$

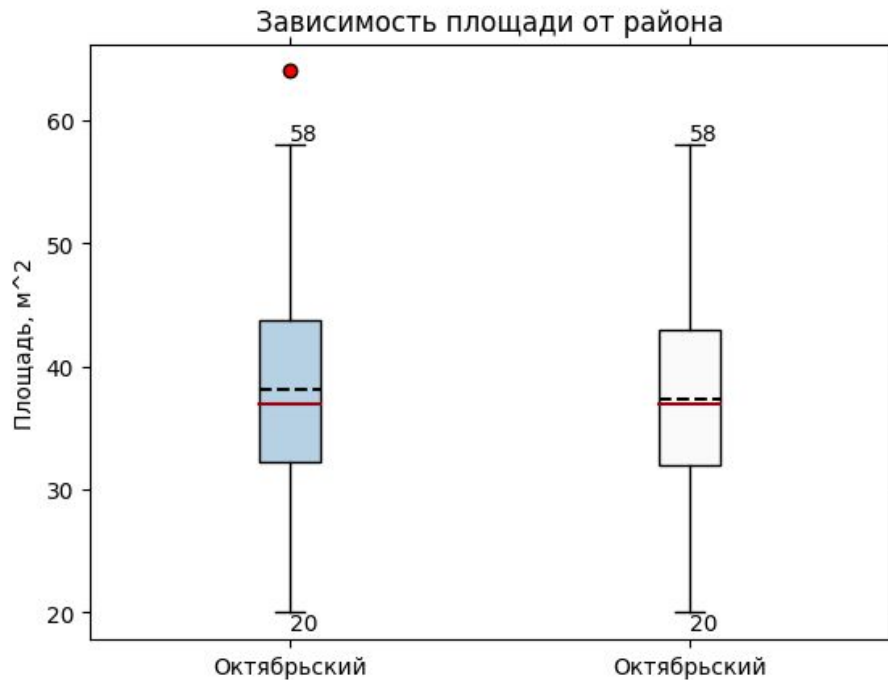
$$64 > 62$$

Площадь - Район



	До исключения выбросов	После исключения выбросов
Среднее	38,18	37,39
Стандартная ошибка	1,75	1,62
Медиана	37,00	37,00
Мода	40,00	40,00
Стандартное отклонение	10,21	9,28
Дисперсия выборки	104,33	86,12
Эксцесс	0,29	-0,09
Асимметричность	0,47	0,19
Интервал	44,00	38,00
Минимум	20,00	20,00
Максимум	64,00	58,00
Сумма	1298,00	1234,00
Счет	34,00	33,00
Наибольший(1)	64,00	58,00
Наименьший(1)	20,00	20,00
Уровень надежности(95,0%)	3,56	3,29
Коэффициент вариации	0,27	0,25

Выводы



1. Распределение стало более близким к нормальному
2. Выборка стала более однородной
3. Дисперсия уменьшилась