

Реферат 2

Шуман С.А.

1. Источники данных:

1.1 Данные были взяты с сайта Банка России с раздела: "Статистика". Они отражают числовые показатели выдачи ипотечных кредитов за 2024 г. (ссылка: <https://www.cbr.ru/statistics/>). Были взяты:

- Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых);
- Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы квартир), первичный рынок (руб.);
- Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации (млн руб.)

1.2 Рейтинговые данные были взяты с сайта «Риа Рейтинг». Был взят: «Рейтинг регионов по доступности покупки жилья в ипотеку» (ссылка: <https://riarating.ru/infografika/20241007/630270859.html>). Рейтинг состоит из следующих показателей:

- Рейтинговое место субъекта
- Доля семей, которые могут купить квартиру в ипотеку по рыночной процентной ставке (%)
- Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льготных программ (%)

2. Предварительная подготовка данных:

2.1 Регионы были отсортированы по показателю «Рейтинговое место субъекта» по возрастанию, затем были сформированы 5 групп регионов (по 17 субъектов в каждой группе $\Rightarrow 85/5$), чтобы получить категориальное значение. Пропусков в данных нет. Получили такие данные:

Исходные данные:

Группа	Регион	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по рыночной процентной ставке (%)	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льготных программ (%)	Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых)	Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы квартир), первичный рынок (руб.)	Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации (млн руб.)
1	в том числе Ненецкий автономный округ	39,4	66,8	7,79	113 034	4,2
1	в том числе Ханты-Мансийский автономный округ - Югра	37,2	60,4	8,92	126 782	5,1
1	Магаданская область	35,7	65,3	7,87	161 000	5,7
1	Мурманская область	32,4	50,7	9,82	95 537	5,3
1	в том числе Ямало-Ненецкий автономный округ	32	58,1	9,00	182 697	5,5
1	Чукотский автономный округ	31,7	56,3	8,91	246 224	4,5
1	Республика Коми	26,6	44	9,78	120 226	5,4
1	Красноярский край	24,2	43,3	8,94	131 269	5,4
1	Республика Саха (Якутия)	23,7	53,4	5,48	152 534	6,1
1	Сахалинская область	23,5	54,6	6,58	156 142	6,5
1	Кемеровская область - Кузбасс	21,4	39,3	9,82	116 230	4,8
1	Камчатский край	21,4	45,2	8,92	161 266	6,0
1	Иркутская область	20,5	39,3	9,38	138 054	4,7
1	Архангельская область	18,7	47,1	6,00	123 482	4,5
1	Челябинская область	18,5	36	9,78	118 910	4,8
1	Пермский край	18,5	37,9	8,90	132 956	4,9
1	Еврейская автономная область	18,1	40,9	6,76	123 540	5,8
2	Оренбургская область	18	37	9,50	79 520	4,4
2	Вологодская область	17,5	35,5	9,34	89 338	5,0
2	Свердловская область	17,3	37	8,81	146 968	4,7
2	Хабаровский край	17,2	40,3	6,78	153 203	5,9
2	Томская область	16,6	35,2	8,80	130 403	5,7
2	Республика Хакасия	16,5	37	8,56	111 886	4,7
2	Тюменская область	16,5	37,4	8,54	132 048	4,9

2.2 Загружаем 1 файл в Оранж:

File - Orange

Source

File: Реферат 2\Данные.xlsx

Sheet свод

URL:

File Type

Automatically detect type

Info

85 instances
7 features (no missing values)
Data has no target variable.
1 meta attribute

Columns (Double click to edit)

	Name	Type	Role	Values
1	Место	N numeric	skip	
2	Группа	C categorical	target	
3	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по рыночной процентной ставке (%)	N numeric	skip	
4	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льготных программ (%)	N numeric	feature	
5	Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых)	N numeric	feature	
6	Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы квартир), первичный рынок (руб.)	N numeric	feature	
7	Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации (млн руб.)	N numeric	feature	
8	Регион	S text	meta	

Reset

Apply

Browse documentation datasets

85

2.3 Загружаем 2 файл в Оранж:

File (1) - Orange

FileViewWindowHelp

Source

File:

Реферат 2\Данные.xlsx

Sheet

свод

URL:

File Type

Automatically detect type

Info

85 instances
7 features (no missing values)
Data has no target variable.
1 meta attribute

Columns (Double click to edit)

	Name	Type	Role	Values
1	Место	N numeric	skip	
2	Группа	N numeric	skip	
3	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по рыночной процентной ставке (%)	N numeric	target	
4	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льготных программ (%)	N numeric	feature	
5	Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых)	N numeric	feature	
6	Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы квартир), первичный рынок (руб.)	N numeric	feature	
7	Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации (млн руб.)	N numeric	feature	
8	Регион	S text	meta	

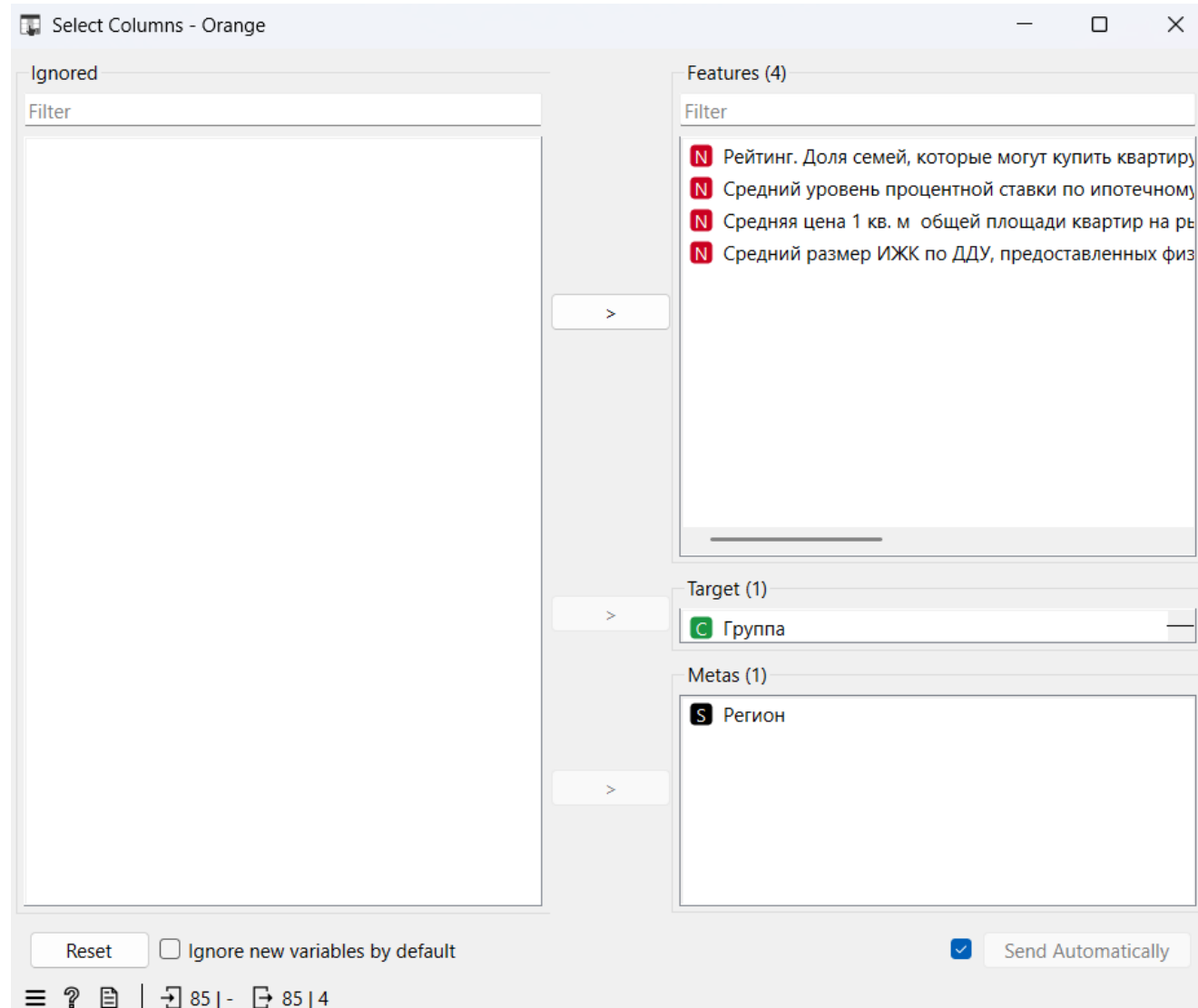
Reset

Apply

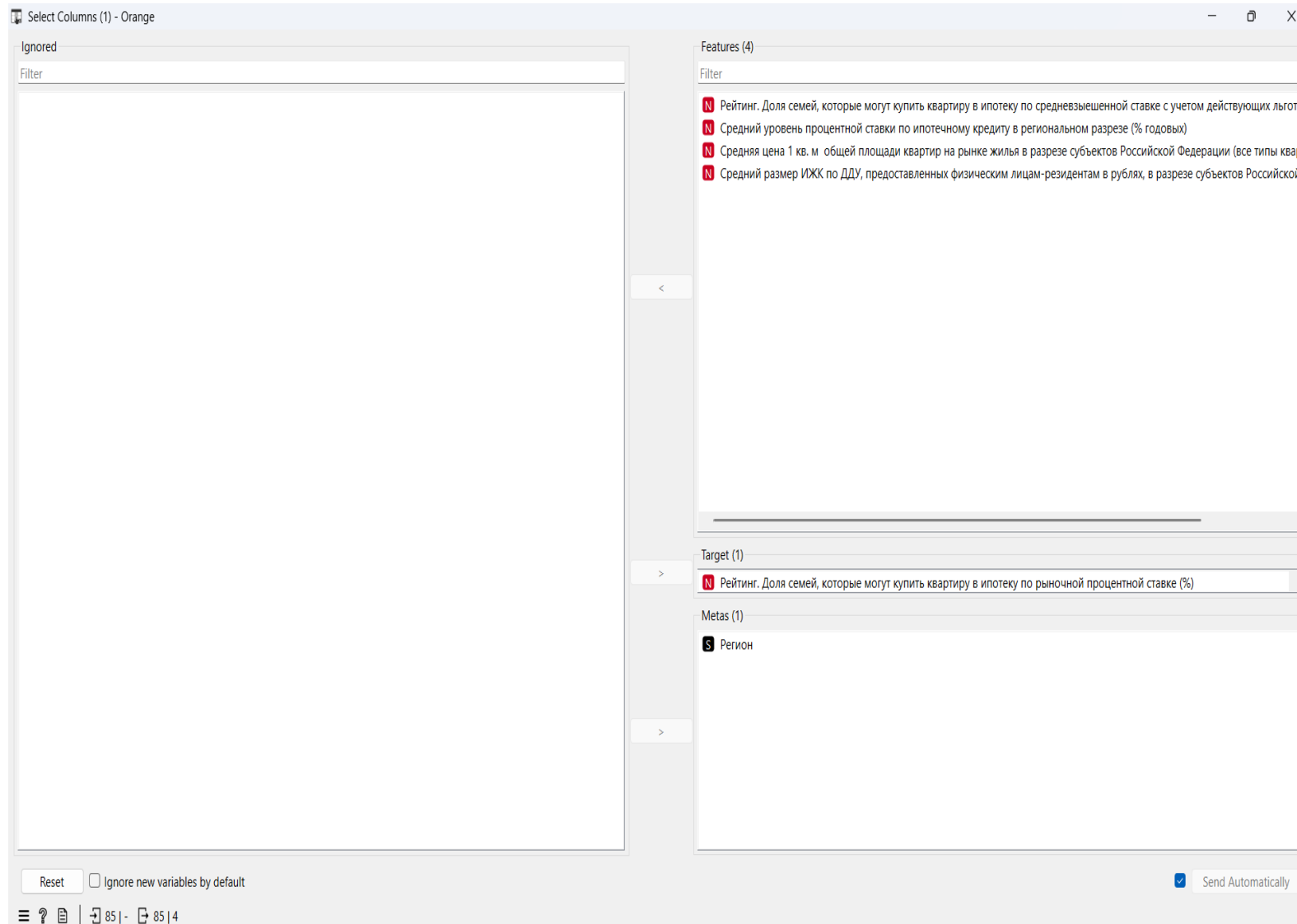
Browse documentation datasets

85

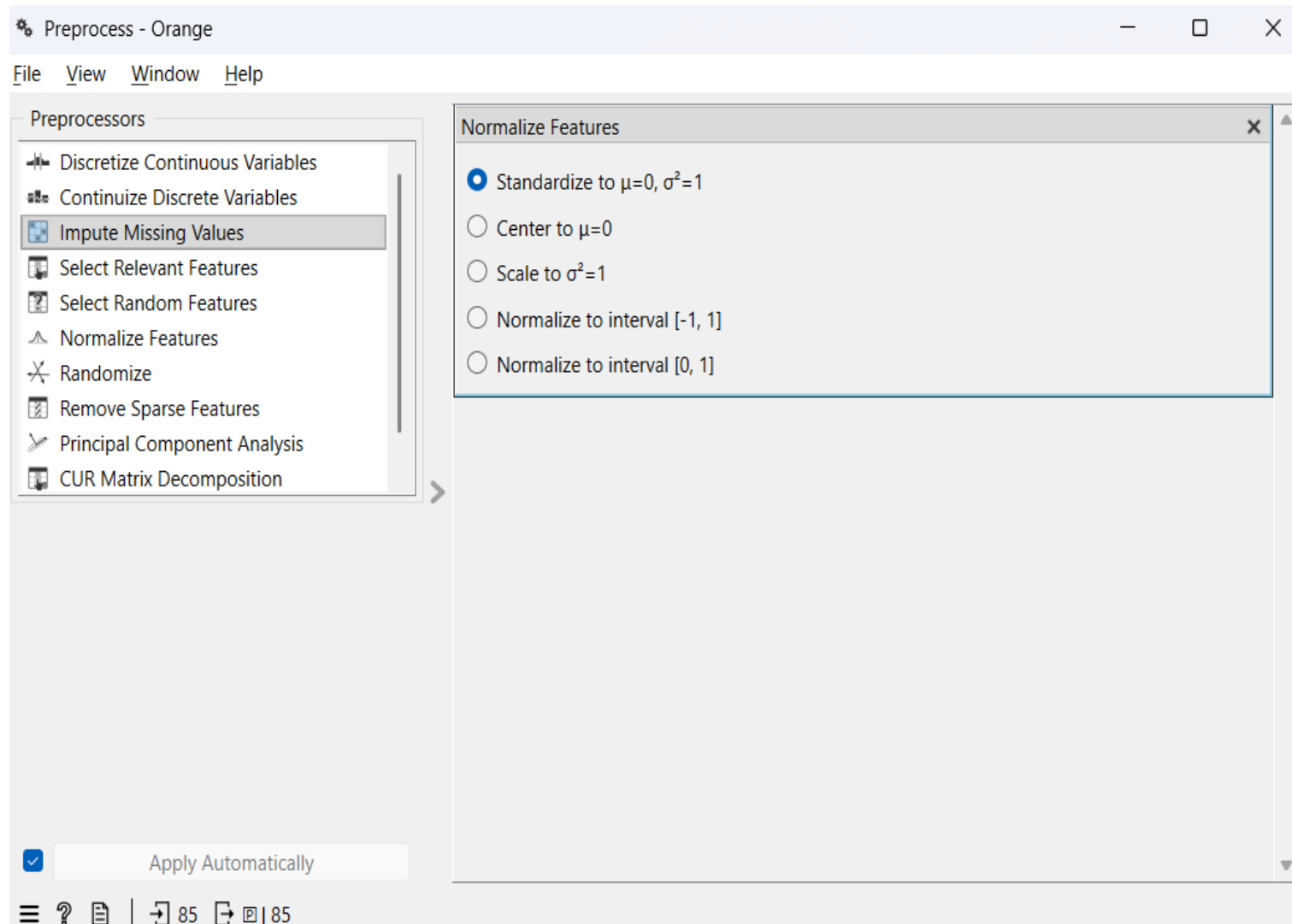
2.4 Добавляем Select Columns для 1 файла:



2.5 Добавляем Select Columns для 2 файла:



2.6 Стандартизируем данные для двух файлов:



3. Корреляционный анализ: Корреляция данных в 1 файле

Correlations - Orange

File View Window Help

Pearson correlation

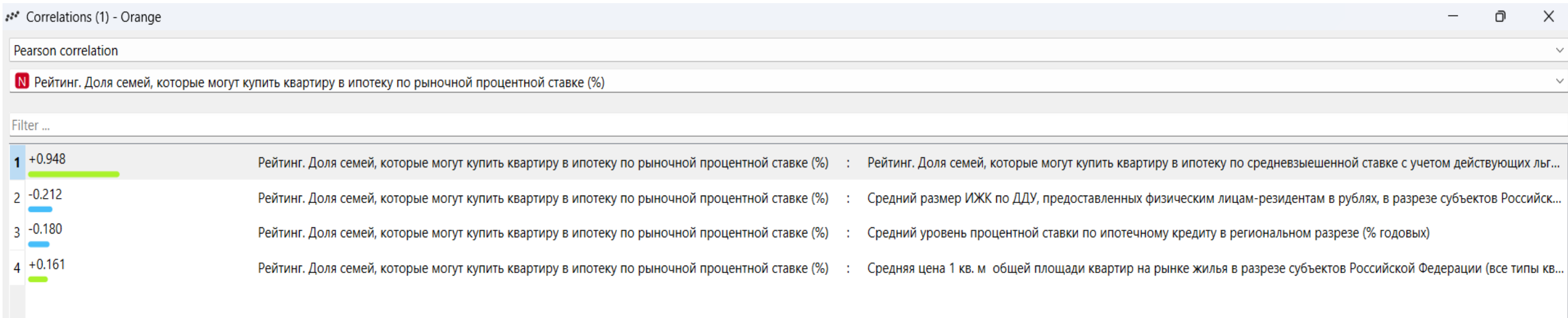
(All combinations)

Filter ...

1	-0.404	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льгот...	: Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых)
2	-0.326	Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых)	: Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы кв...
3	+0.289	Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации...	: Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых)
4	-0.248	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льгот...	: Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации...
5	+0.225	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льгот...	: Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы кв...
6	+0.216	Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации...	: Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы кв...





- Существует отрицательная корреляция (-0,40) между «долей семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке» со «средним уровнем процентной ставки по ипотечному кредиту»;
- Существует отрицательная корреляция (-0,33) между «средним уровнем процентной ставки по ипотечному кредиту» со «средней ценой 1 кв. м.».

3.2 Корреляция данных в 2 файле



- Присутствует сильная положительная корреляция (0,95) между «долей семей, которые могут купить квартиру в ипотеку по рыночной процентной ставке» с «долей семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льгот в регионе».

4. Анализ рангов: Анализ рангов 1 файла

Rank - Orange				
Scoring Methods				
<input type="checkbox"/>	Information Gain			
<input checked="" type="checkbox"/>	Information Gain Ratio			
<input checked="" type="checkbox"/>	Gini Decrease			
<input type="checkbox"/>	ANOVA			
<input type="checkbox"/>	χ^2			
<input type="checkbox"/>	ReliefF			
<input type="checkbox"/>	FCBF			
		#	Gain ratio	Gini
1	 Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льготных программ (%)		0.641	0.360
2	 Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых)		0.119	0.062
3	 Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы квартир), первичный рынок (руб.)		0.089	0.049
4	 Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации (млн руб.)		0.065	0.034

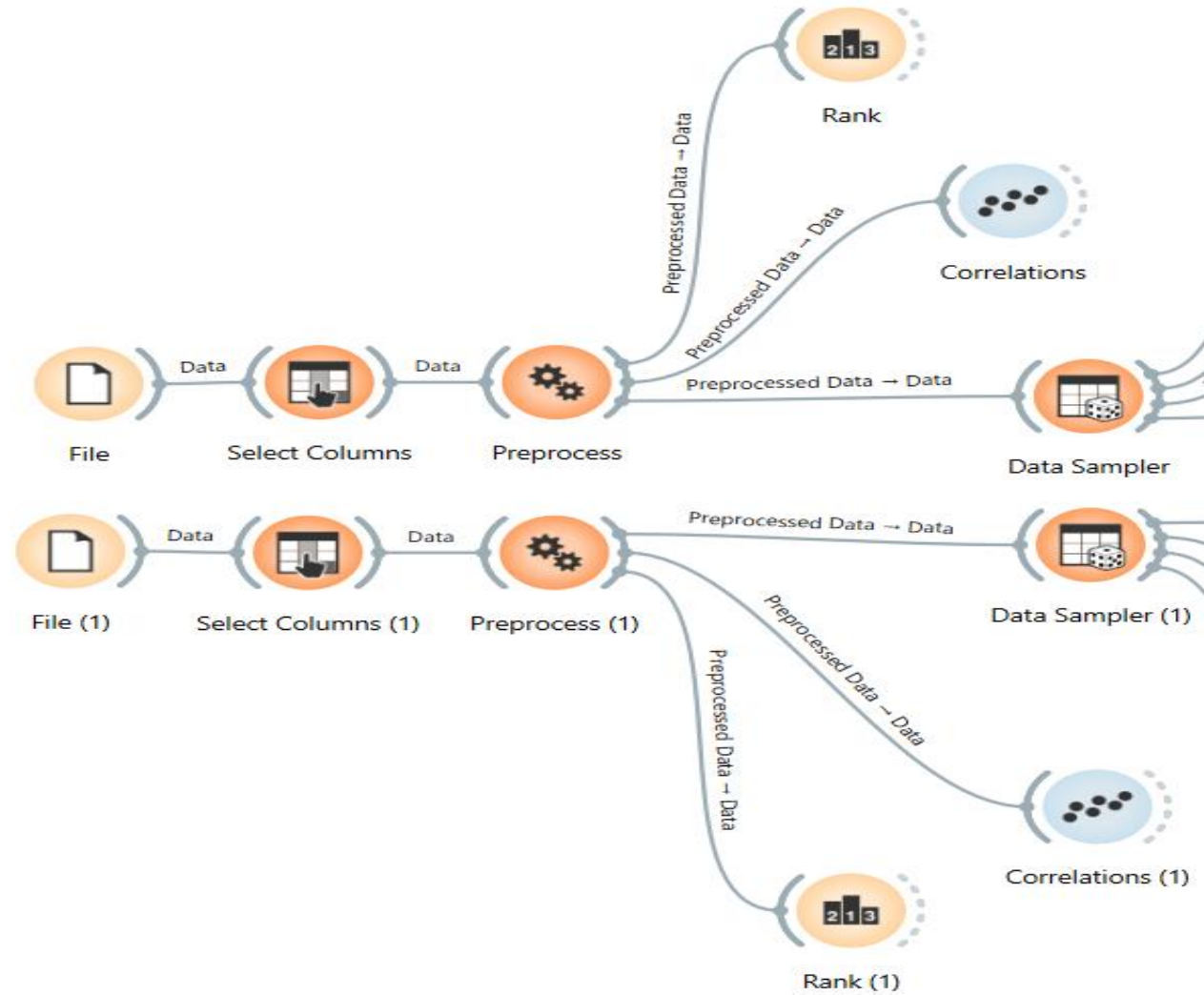
- Наиболее влияющей переменной является «доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льгот в регионе», затем «средний уровень процентной ставки по ипотечному кредиту».

4. Анализ рангов: Анализ рангов 2 файла

Rank (1) - Orange				
Scoring Methods				
<input checked="" type="checkbox"/> Univariate Regression				
<input checked="" type="checkbox"/> RReliefF				
		#	Univar. reg.	RReliefF
1	Рейтинг. Доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льготных программ (%)		729.036	0.248
2	Средний размер ИЖК по ДДУ, предоставленных физическим лицам-резидентам в рублях, в разрезе субъектов Российской Федерации (млн руб.)		3.891	0.126
3	Средний уровень процентной ставки по ипотечному кредиту в региональном разрезе (% годовых)		2.783	0.107
4	Средняя цена 1 кв. м общей площади квартир на рынке жилья в разрезе субъектов Российской Федерации (все типы квартир), первичный рынок (руб.)		2.195	0.089

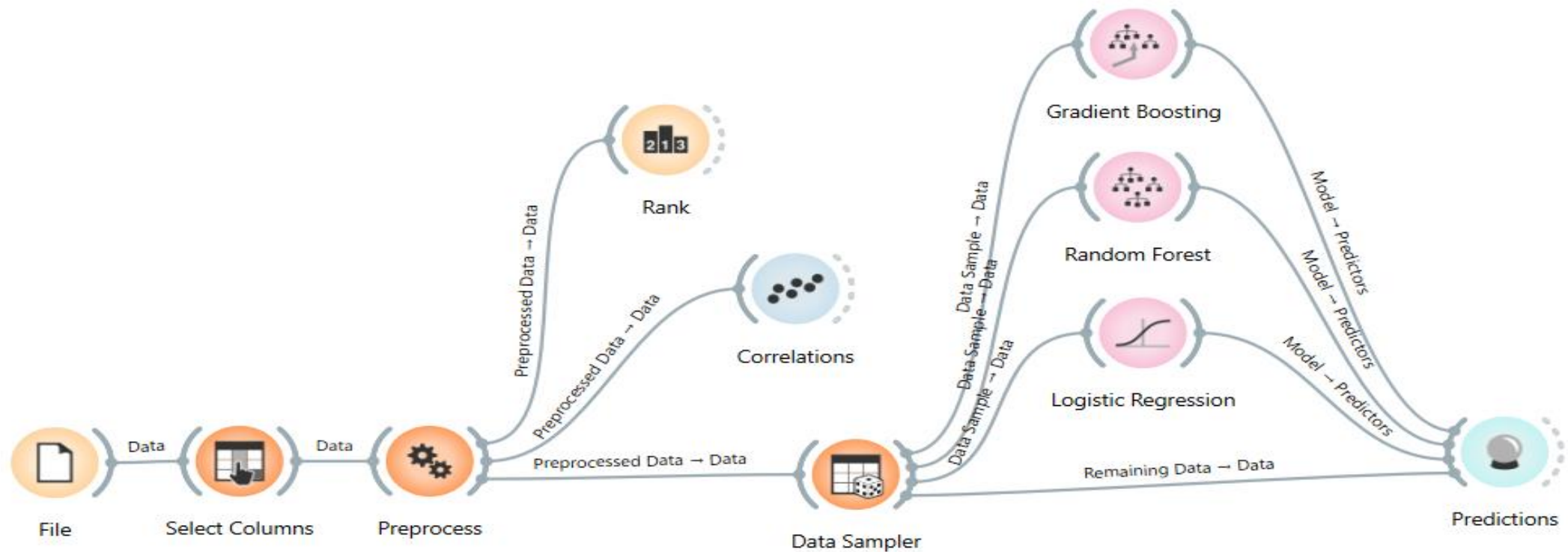
- Наиболее влияющей переменной является «доля семей, которые могут купить квартиру в ипотеку по средневзвешенной ставке с учетом действующих льгот в регионе», затем «средний размер ИЖК по ДДУ».

5. Разделение на тестовую и обучающую выборки:



6. Строим модели, создаем прогноз и подключаем тестовые данные:

6.1 Прогноз для 1 файла:



6. Строим модели, создаем прогноз и подключаем тестовые данные:

6.1 Прогноз для 1 файла:

Predictions - Orange

FileViewWindowHelp

Show probabilities forClasses in dataShow classification errors

	Gradient Boosting	error	Random Forest	error	Logistic Regression	error
1	0.00 : 0.00 : 1.00 : 0.00 : 0.00 → 3	0.000	0.00 : 0.00 : 0.97 : 0.03 : 0.00 → 3	0.033	0.07 : 0.37 : 0.33 : 0.15 : 0.08 → 2	0.674
2	0.00 : 0.00 : 0.00 : 0.00 : 1.00 → 5	0.000	0.17 : 0.03 : 0.00 : 0.10 : 0.70 → 5	0.300	0.01 : 0.11 : 0.23 : 0.27 : 0.37 → 5	0.626
3	1.00 : 0.00 : 0.00 : 0.00 : 0.00 → 1	0.000	0.97 : 0.03 : 0.00 : 0.00 : 0.00 → 1	0.025	0.97 : 0.03 : 0.01 : 0.00 : 0.00 → 1	0.032
4	0.00 : 0.00 : 0.00 : 0.00 : 1.00 → 5	0.000	0.17 : 0.03 : 0.00 : 0.10 : 0.70 → 5	0.300	0.00 : 0.04 : 0.12 : 0.25 : 0.58 → 5	0.418
5	1.00 : 0.00 : 0.00 : 0.00 : 0.00 → 1	0.000	0.80 : 0.06 : 0.10 : 0.05 : 0.00 → 1	0.202	0.99 : 0.00 : 0.00 : 0.00 : 0.00 → 1	0.006
6	0.00 : 0.00 : 0.00 : 0.00 : 1.00 → 5	0.000	0.00 : 0.03 : 0.06 : 0.17 : 0.75 → 5	0.252	0.00 : 0.00 : 0.04 : 0.31 : 0.65 → 5	0.345
7	0.00 : 1.00 : 0.00 : 0.00 : 0.00 → 2	0.000	0.05 : 0.42 : 0.36 : 0.07 : 0.10 → 2	0.579	0.20 : 0.39 : 0.29 : 0.10 : 0.02 → 2	0.613
8	0.00 : 0.00 : 1.00 : 0.00 : 0.00 → 3	0.000	0.00 : 0.03 : 0.97 : 0.00 : 0.00 → 3	0.033	0.07 : 0.20 : 0.38 : 0.25 : 0.10 → 3	0.623
9	0.00 : 0.00 : 0.00 : 1.00 : 0.00 → 4	0.000	0.00 : 0.20 : 0.37 : 0.43 : 0.00 → 4	0.567	0.01 : 0.34 : 0.27 : 0.15 : 0.22 → 2	0.855
10	0.00 : 0.00 : 0.00 : 0.00 : 1.00 → 5	0.001	0.00 : 0.03 : 0.06 : 0.17 : 0.75 → 5	0.252	0.00 : 0.00 : 0.04 : 0.14 : 0.82 → 5	0.185
11	0.00 : 0.00 : 0.00 : 0.00 : 1.00 → 5	0.002	0.00 : 0.03 : 0.06 : 0.17 : 0.75 → 5	0.252	0.00 : 0.00 : 0.03 : 0.55 : 0.42 → 4	0.577
12	0.00 : 0.00 : 0.00 : 0.00 : 1.00 → 5	0.004	0.10 : 0.00 : 0.00 : 0.27 : 0.63 → 5	0.367	0.00 : 0.02 : 0.05 : 0.11 : 0.82 → 5	0.177
13	0.00 : 0.84 : 0.14 : 0.02 : 0.00 → 2	0.161	0.11 : 0.44 : 0.31 : 0.04 : 0.10 → 2	0.562	0.31 : 0.30 : 0.27 : 0.10 : 0.02 → 1	0.703
14	0.00 : 0.00 : 0.17 : 0.83 : 0.00 → 4	0.169	0.00 : 0.00 : 0.40 : 0.60 : 0.00 → 4	0.400	0.02 : 0.14 : 0.33 : 0.27 : 0.24 → 3	0.732
15	0.00 : 0.78 : 0.22 : 0.00 : 0.00 → 2	0.224	0.00 : 0.35 : 0.62 : 0.03 : 0.00 → 3	0.645	0.14 : 0.36 : 0.32 : 0.14 : 0.04 → 2	0.638
16	0.00 : 0.00 : 0.20 : 0.05 : 0.76 → 5	0.244	0.00 : 0.03 : 0.06 : 0.17 : 0.75 → 5	0.252	0.00 : 0.01 : 0.05 : 0.14 : 0.80 → 5	0.199
17	0.00 : 0.04 : 0.73 : 0.24 : 0.00 → 3	0.275	0.17 : 0.25 : 0.20 : 0.38 : 0.00 → 4	0.800	0.06 : 0.23 : 0.35 : 0.27 : 0.09 → 3	0.653
18	0.00 : 0.00 : 0.88 : 0.12 : 0.00 → 3	0.881	0.05 : 0.22 : 0.25 : 0.48 : 0.00 → 4	0.517	0.19 : 0.01 : 0.20 : 0.57 : 0.04 → 4	0.434
19	0.00 : 0.02 : 0.98 : 0.00 : 0.00 → 3	0.985	0.08 : 0.33 : 0.50 : 0.08 : 0.00 → 3	0.667	0.23 : 0.16 : 0.35 : 0.23 : 0.04 → 3	0.844
20	0.00 : 0.00 : 1.00 : 0.00 : 0.00 → 3	0.999	0.05 : 0.12 : 0.38 : 0.45 : 0.00 → 4	0.550	0.10 : 0.07 : 0.35 : 0.39 : 0.10 → 4	0.613
21	0.00 : 0.00 : 1.00 : 0.00 : 0.00 → 3	0.999	0.00 : 0.14 : 0.68 : 0.17 : 0.00 → 3	0.827	0.07 : 0.30 : 0.34 : 0.19 : 0.10 → 3	0.810
22	0.00 : 0.97 : 0.03 : 0.00 : 0.00 → 2	1.000	0.00 : 0.48 : 0.48 : 0.03 : 0.00 → 2	0.967	0.22 : 0.12 : 0.34 : 0.24 : 0.09 → 3	0.758
23	1.00 : 0.00 : 0.00 : 0.00 : 0.00 → 1	1.000	0.61 : 0.26 : 0.10 : 0.03 : 0.00 → 1	0.743	0.29 : 0.43 : 0.22 : 0.05 : 0.01 → 2	0.569

Restore Original Order

Группа	Регион	у по средневзые	ипотечному кред	зрезе субъектов	цам-резидентам
3	Республика Марий Эл	-0.0850	0.0049	-0.57587	-0.793
5	Алтайский край	-0.8908	0.5476	-0.0271211	-0.686
1	в том числе Ханты-Мансийский автономный ок...	2.3000	0.1406	0.0250902	-0.258
5	Республика Алтай	-1.3662	0.5317	0.83739	-0.900
1	Магаданская область	2.6948	-0.6975	0.800386	0.384
5	Республика Ингушетия	-1.9222	3.3971	-1.51962	2.632
2	Тульская область	0.3420	-0.2425	-0.291763	-0.686
3	Тверская область	-0.0770	0.1645	-0.555168	0.063
4	Псковская область	-0.5201	0.2124	-1.16749	-1.114
5	Кабардино-Балкарская Республика	-1.6240	1.0026	-1.41745	0.919
5	Чеченская Республика	-1.9786	4.3788	-1.02618	4.452
5	Карачаево-Черкесская Республика	-1.7368	1.4096	-1.37589	-0.472
2	Тюменская область	0.4468	-0.1627	0.144401	-0.472
4	Орловская область	-0.3670	-0.0350	-0.787921	0.063
2	Ярославская область	0.1567	0.1406	-0.404444	-0.579
5	Республика Северная Осетия - Алания	-1.5676	0.6674	-0.961768	0.384
3	Самарская область	-0.3026	0.9148	0.0422945	-0.472
4	г. Санкт-Петербург	-0.1092	-0.3064	3.41349	1.347
2	Томская область	0.2695	0.0448	0.107128	0.384
4	Ленинградская область	-0.0528	-0.2825	0.534373	0.812
4	Кировская область	-0.1575	-0.0430	-0.268259	-0.686
4	Приморский край	0.2292	-2.1501	1.14086	0.063
2	Удмуртская Республика	0.5999	-0.6815	-0.339894	-0.900

Show performance scoresTarget class: (Average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Gradient Boosting	0.850	0.680	0.671	0.775	0.680	0.622
Random Forest	0.896	0.680	0.679	0.700	0.680	0.602
Logistic Regression	0.889	0.600	0.607	0.660	0.600	0.509

25 | 3x25

6. Строим модели, создаем прогноз и подключаем тестовые данные:

6.1 Прогноз для 1 файла:

- Random Forest показывает наилучшие результаты по AUC (0,896) и F1-score (0,679), а также имеет высокую точность (Precision = 0,700) и полноту (Recall = 0,680). Это говорит о его сбалансированной производительности.
- Gradient Boosting демонстрирует наивысшие Precision (0,775) и MCC (0,622), что указывает на хорошее качество классификации, особенно если важна минимизация ложных срабатываний.
- Logistic Regression уступает по всем метрикам, кроме AUC, где она близка к Random Forest.

Если выбирать одну модель, то **Random Forest** выглядит предпочтительнее благодаря высокому AUC и сбалансированности метрик. Однако если критически важна **точность (Precision)**, то можно рассмотреть **Gradient Boosting**.

6.2 Регионы с наибольшими ошибками по модели «Gradient Boosting»

	Gradient Boosting	error	Random Forest	error	Logistic Regression	error	Группа	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	1.00 : 0.00 : 0.00 : 0.00 : 0.00 → 1	1.000	0.80 : 0.06 : 0.10 : 0.05 : 0.00 → 1	0.943	0.47 : 0.10 : 0.27 : 0.14 : 0.02 → 1	0.900	2	Хабаровский край	0.6804	-1.5675	0.623729	0.598
2	0.00 : 1.00 : 0.00 : 0.00 : 0.00 → 2	1.000	0.10 : 0.82 : 0.05 : 0.03 : 0.00 → 2	0.900	0.25 : 0.36 : 0.27 : 0.11 : 0.01 → 2	0.750	1	Челябинская область	0.3340	0.8270	-0.15328	-0.579

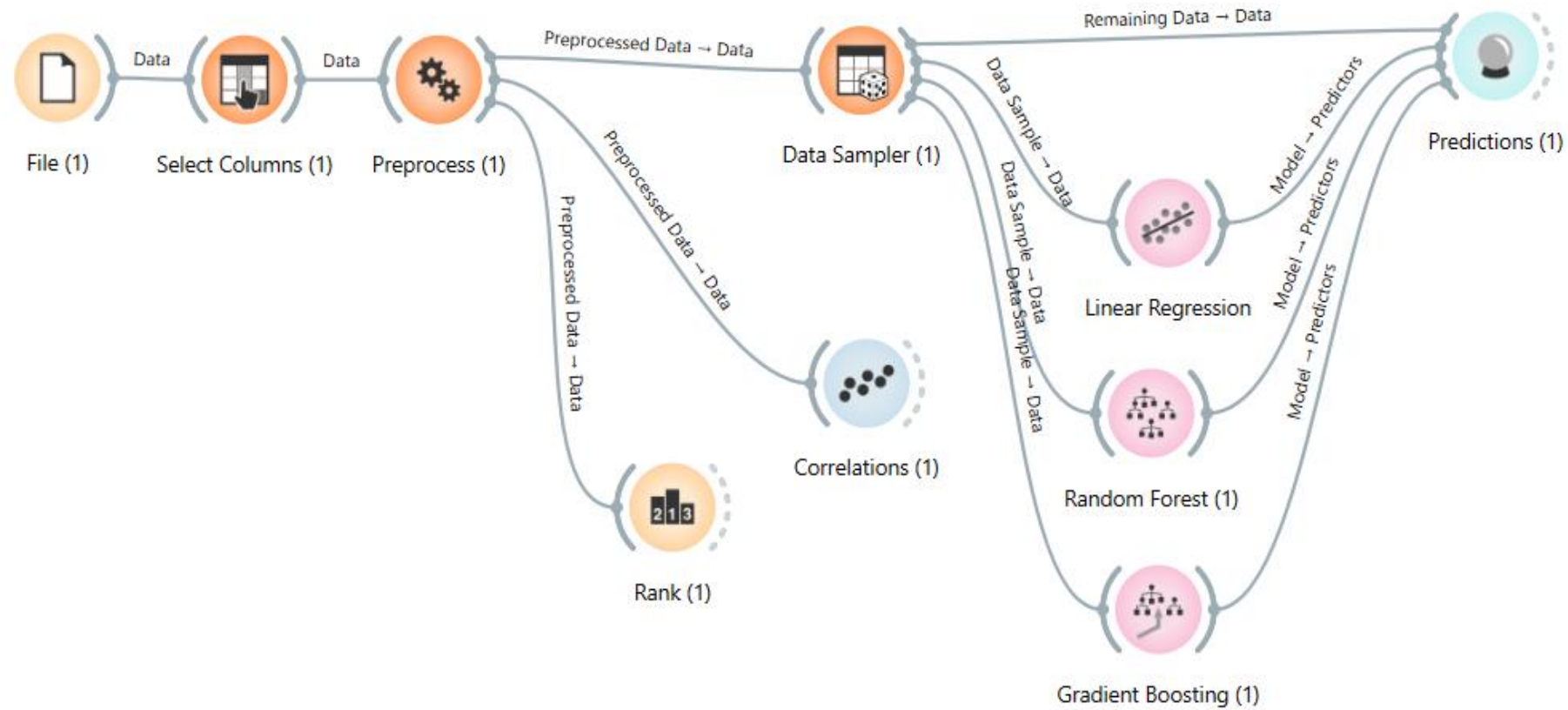
6.3 Регионы с наибольшими ошибками по модели «Random Forest»

	Gradient Boosting	error	Random Forest	error	Logistic Regression	error	Группа	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	0.00 : 0.97 : 0.03 : 0.00 : 0.00 → 2	1.000	0.00 : 0.48 : 0.48 : 0.03 : 0.00 → 2	0.967	0.22 : 0.12 : 0.34 : 0.24 : 0.09 → 3	0.758	4	Приморский край	0.2292	-2.1501	1.14086	0.063
2	1.00 : 0.00 : 0.00 : 0.00 : 0.00 → 1	1.000	0.80 : 0.06 : 0.10 : 0.05 : 0.00 → 1	0.943	0.47 : 0.10 : 0.27 : 0.14 : 0.02 → 1	0.900	2	Хабаровский край	0.6804	-1.5675	0.623729	0.598

6.4 Регионы с наибольшими

	Gradient Boosting	error	Random Forest	error	Logistic Regression	error	Группа	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	1.00 : 0.00 : 0.00 : 0.00 : 0.00 → 1	1.000	0.80 : 0.06 : 0.10 : 0.05 : 0.00 → 1	0.943	0.47 : 0.10 : 0.27 : 0.14 : 0.02 → 1	0.900	2	Хабаровский край	0.6804	-1.5675	0.623729	0.598
2	0.00 : 0.00 : 0.00 : 1.00 : 0.00 → 4	0.000	0.00 : 0.20 : 0.37 : 0.43 : 0.00 → 4	0.567	0.01 : 0.34 : 0.27 : 0.15 : 0.22 → 2	0.855	4	Псковская область	-0.5201	0.2124	-1.16749	-1.114

6.5 Прогноз для 2 файла:



6.5 Прогноз для 2 файла:

Predictions (1) - Orange

Shown regression error: Difference

	Random Forest (1)	error	Gradient Boosting (1)	error	Linear Regression	error
1		12.4 0.5		12.4 0.5		13.7 1.8
2		15.5 -0.7		15.6 -0.6		16.6 0.4
3		11.6 -1.5		12.5 -0.6		13.7 0.6
4		4.3 -0.2		3.6 -0.9		4.8 0.3
5		4.3 -0.8		3.6 -1.5		3.9 -1.2
6		14.3 3.6		11.6 0.9		12.4 1.7
7		13.9 -0.5		14.0 -0.4		15.8 1.4
8		4.3 -0.2		4.0 -0.5		3.0 -1.5
9		10.0 -0.4		11.1 0.7		10.6 0.2
10		12.1 2.3		10.0 0.2		11.0 1.2
11		4.3 -1.0		3.8 -1.5		2.6 -2.7
12		13.0 -0.9		12.6 -1.3		13.7 -0.2
13		22.8 7.6		20.0 4.8		17.9 2.7
14		30.6 -6.6		30.5 -6.7		32.4 -4.8
15		12.3 1.1		12.3 1.1		13.5 2.3
16		14.0 -2.6		15.4 -1.2		16.3 -0.3
17		29.2 -6.5		33.8 -1.9		34.0 -1.7
18		15.7 -0.8		17.3 0.8		17.6 1.1
19		5.4 0.9		4.7 0.2		5.2 0.7
20		5.3 0.0		5.4 0.1		2.8 -2.5
21		20.5 3.3		18.0 0.8		16.6 -0.6
22		11.8 1.7		12.5 2.4		13.2 3.1
23		8.1 0.7		7.9 0.5		8.5 1.1
24		16.1 -2.4		16.7 -1.8		18.5 -0.0

Restore Original Order

№ квартиры в инс	Регион	у по средневзые	ипотечному кред	зрезе субъектов	цам-резидентам
11.9	Республика Марий Эл	-0.0850	0.0049	-0.57587	-0.793
16.2	Тульская область	0.3420	-0.2425	-0.291763	-0.686
13.1	Самарская область	-0.3026	0.9148	0.0422945	-0.472
4.5	Чеченская Республика	-1.9786	4.3788	-1.02618	4.452
5.1	Республика Ингушетия	-1.9222	3.3971	-1.51962	2.632
10.7	Приморский край	0.2292	-2.1501	1.14086	0.063
14.4	Ярославская область	0.1567	0.1406	-0.404444	-0.579
4.5	Карачаево-Черкесская Республика	-1.7368	1.4096	-1.37589	-0.472
10.4	Псковская область	-0.5201	0.2124	-1.16749	-1.114
9.8	Орловская область	-0.3670	-0.0350	-0.787921	0.063
5.3	Кабардино-Балкарская Республика	-1.6240	1.0026	-1.41745	0.919
13.9	Тверская область	-0.0770	0.1645	-0.555168	0.063
15.2	Удмуртская Республика	0.5999	-0.6815	-0.339894	-0.900
37.2	в том числе Ханты-Мансийский автономный округ - Югра	2.3000	0.1406	0.0250902	-0.258
11.2	г. Санкт-Петербург	-0.1092	-0.3064	3.41349	1.347
16.6	Томская область	0.2695	0.0448	0.107128	0.384
35.7	Магаданская область	2.6948	-0.6975	0.800386	0.384
16.5	Тюменская область	0.4468	-0.1627	0.144401	-0.472
4.5	Республика Алтай	-1.3662	0.5317	0.83739	-0.900
5.3	Республика Северная Осетия - Алания	-1.5676	0.6674	-0.961768	0.384
17.2	Хабаровский край	0.6804	-1.5675	0.623729	0.598
10.1	Ленинградская область	-0.0528	-0.2825	0.534373	0.812
7.4	Алтайский край	-0.8908	0.5476	-0.0271211	-0.686
18.5	Челябинская область	0.3340	0.8270	-0.15328	-0.579

Show performance scores

Model	MSE	RMSE	MAE	MAPE	R2
Random Forest (1)	7.863	2.804	1.897	13.472	0.882
Gradient Boosting (1)	3.785	1.946	1.280	10.378	0.943
Linear Regression	3.268	1.808	1.435	14.813	0.951

25 | 25 | 3x25

6.5 Прогноз для 2 файла:

Gradient Boosting демонстрирует наиболее сбалансированные результаты:

- Лучшие показатели по MAE и MAPE;
- Очень высокий R^2 (0.943);
- Хорошие значения MSE и RMSE.

Linear Regression показывает отличный R^2 , но имеет худший MAPE, что может указывать на проблемы с процентной ошибкой прогнозирования.

Random Forest значительно уступает по всем метрикам.

Таким образом, **Gradient Boosting** является оптимальным выбором, так как сочетает высокую объясняющую способность (R^2) с наименьшими ошибками в абсолютном (MAE) и процентном (MAPE) выражении.

6.6 Регионы с наибольшими ошибками по модели «Random Forest»

	Random Forest (1)	error	Gradient Boosting (1)	error	Linear Regression	error	ъ квартиру в ипс	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	22.8	7.6	20.0	4.8	17.9	2.7	15.2	Удмуртская Республика	0.5999	-0.6815	-0.339894	-0.900
2	14.3	3.6	11.6	0.9	12.4	1.7	10.7	Приморский край	0.2292	-2.1501	1.14086	0.063

	Random Forest (1)	error	Gradient Boosting (1)	error	Linear Regression	error	ъ квартиру в ипс	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	30.6	-6.6	30.5	-6.7	32.4	-4.8	37.2	в том числе Ханты-Мансийский автономный округ - Югра	2.3000	0.1406	0.0250902	-0.258
2	29.2	-6.5	33.8	-1.9	34.0	-1.7	35.7	Магаданская область	2.6948	-0.6975	0.800386	0.384

6.7 Регионы с наибольшими ошибками по модели «Gradient Boosting»

	Random Forest (1)	error	Gradient Boosting (1)	error	Linear Regression	error	ъ квартиру в ипс	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	22.8	7.6	20.0	4.8	17.9	2.7	15.2	Удмуртская Республика	0.5999	-0.6815	-0.339894	-0.900
2	11.8	1.7	12.5	2.4	13.2	3.1	10.1	Ленинградская область	-0.0528	-0.2825	0.534373	0.812

	Random Forest (1)	error	Gradient Boosting (1)	error	Linear Regression	error	ъ квартиру в ипс	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	30.6	-6.6	30.5	-6.7	32.4	-4.8	37.2	в том числе Ханты-Мансийский автономный округ - Югра	2.3000	0.1406	0.0250902	-0.258
2	29.2	-6.5	33.8	-1.9	34.0	-1.7	35.7	Магаданская область	2.6948	-0.6975	0.800386	0.384

6.8 Регионы с наибольшими ошибками по модели «Linear Regression»

	Random Forest (1)	error	Gradient Boosting (1)	error	Linear Regression	error	ъ квартиру в ипс	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	11.8	1.7	12.5	2.4	13.2	3.1	10.1	Ленинградская область	-0.0528	-0.2825	0.534373	0.812
2	22.8	7.6	20.0	4.8	17.9	2.7	15.2	Удмуртская Республика	0.5999	-0.6815	-0.339894	-0.900

	Random Forest (1)	error	Gradient Boosting (1)	error	Linear Regression	error	ъ квартиру в ипс	Регион	у по средневые	ипотечному кред	зрезе субъектов	цам-резидентам
1	30.6	-6.6	30.5	-6.7	32.4	-4.8	37.2	в том числе Ханты-Мансийский автономный округ - Югра	2.3000	0.1406	0.0250902	-0.258
2	4.3	-1.0	3.8	-1.5	2.6	-2.7	5.3	Кабардино-Балкарская Республика	-1.6240	1.0026	-1.41745	0.919

7. Финальный проект:

