

# Анализ квартир с сайта СІАН в Санкт-Петербурге

Клочков Владислав Сергеевич  
Группа 19.Б08-мм  
st068203@student.spbu.ru

29 апреля 2023

# Сайт

<https://spb.cian.ru/>



Аренда Продажа Новостройки Дома и участки Коммерческая Ипотека Сервисы **ПИК**



+ Разместить объявление

Войти

Санкт-Петербург

## Ваш дом мечты — на Циан

Купить

Снять

Посуточно

Оценить

Ипотека

Подбор риелтора

Разместить объявление

Квартиру в новостройке и вторичке

1, 2 комн.

Цена

Город, адрес, метро, район, ж/д, шоссе или ЖК

Квартиру в новостройке и вторичке

Показать на карте


Найти

# Парсинг сайта cian

Новостройка от застройщика

Василеостровский

- премиум-класс
- первая линия Финского залива
- панорамные окна



Комнат	Потолки	Общая площадь
2	3 м	61.31 м²



Ещё Фото

2-комн. кв., 61,31 м², 6/19 этаж

Секция 10 • Сдача корпуса 1 кв. 2025

Только на Циан

ЖК «Glorax Premium Василеостровский»

сдача ГК: 1 кв. 2025 года

Зенит 5 минут на транспорте

Санкт-Петербург, р-н Василеостровский, Морской, м. Зенит, Глоракс Премиум Василеостровский ЖК

14 629 999 ₽

238 623 ₽/м²

Glorax Premium Василеостровский это роскошный комплекс премиум-класса на первой береговой линии Финского залива. Проект уникален по своему опциональному наполнению: вся необходимая для комфортной городской жизни инфраструктура будет в составе самог...



ЗАСТРОЙЩИК

Glorax

+7 812 213-43-....



три дня назад

# Итоговый датасет

	metro	price	rooms	floor	floors	square	turned
0	0	14629999	2	6	19	61.31	0
1	0	9799999	1	2	19	39.35	0
2	1	15441588	2	15	18	60.60	1
3	1	9350000	1	2	18	37.40	1
4	0	10239999	1	2	18	37.00	0
...	...	...	...	...	...	...	...
7915	0	5859341	1	2	12	35.00	0
7916	1	23324180	1	6	9	54.30	1
7917	0	8779520	1	10	13	37.35	1
7918	0	6449160	1	1	13	35.03	0
7919	1	15156000	1	8	8	42.10	1

7920 rows × 7 columns

**metro** - есть ли пешая доступность к метро

**price** - цена

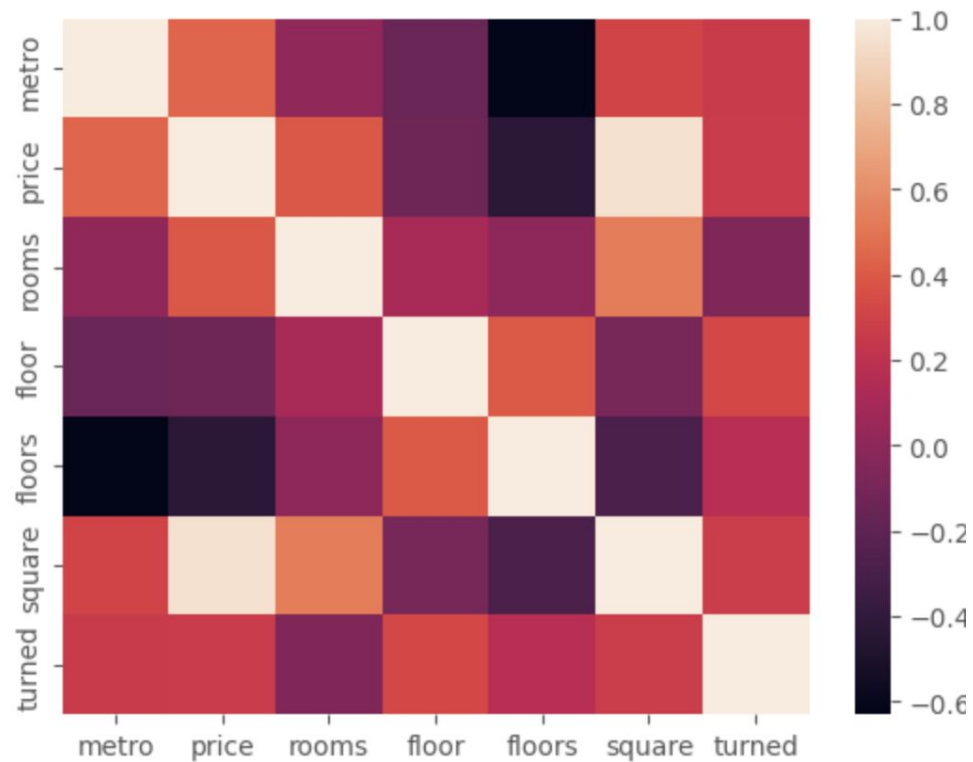
**floor** - этаж

**floors** - этажность дома

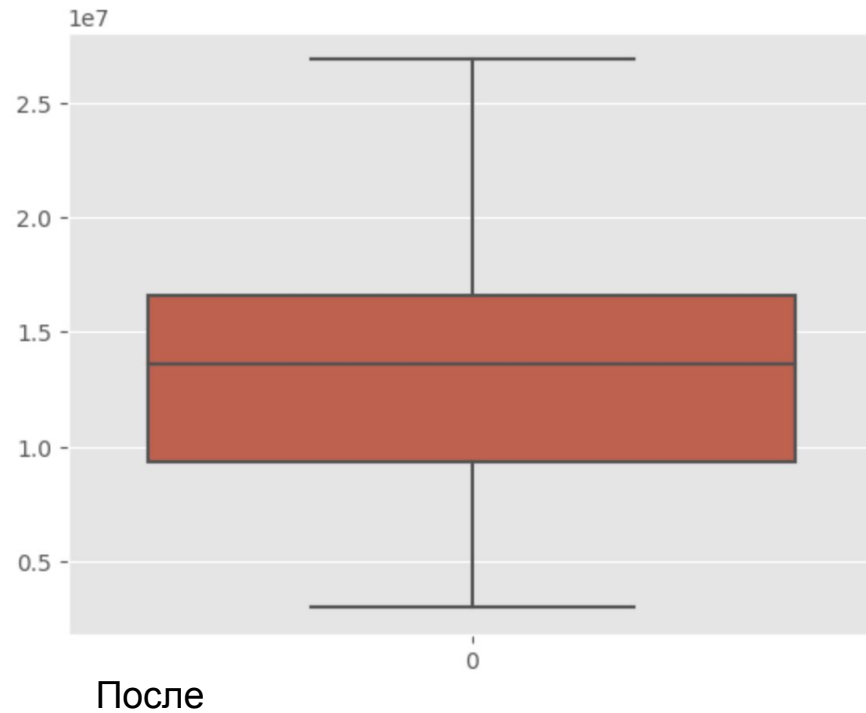
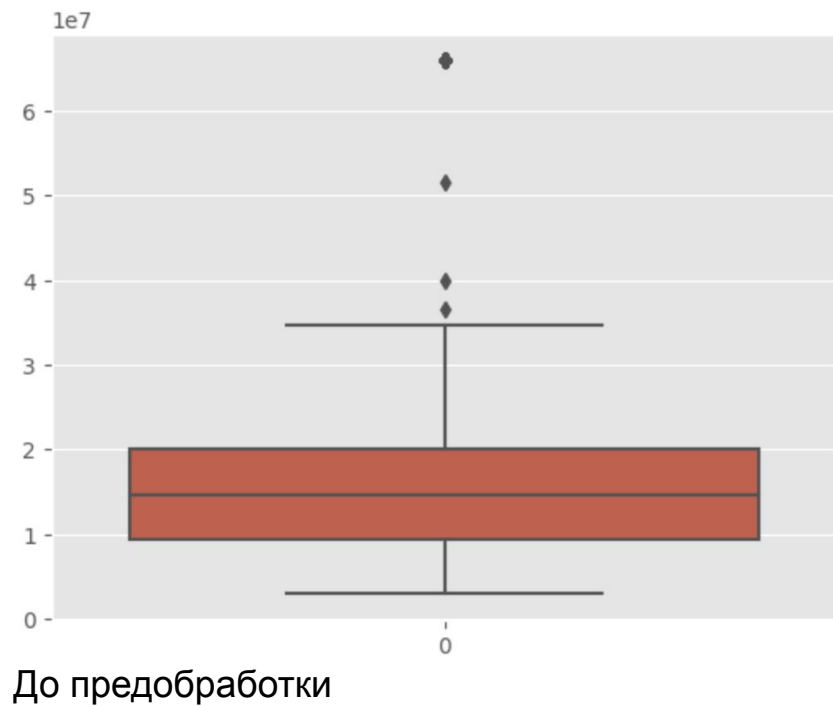
**square** - площадь квартиры

**turned** - сдан ли дом

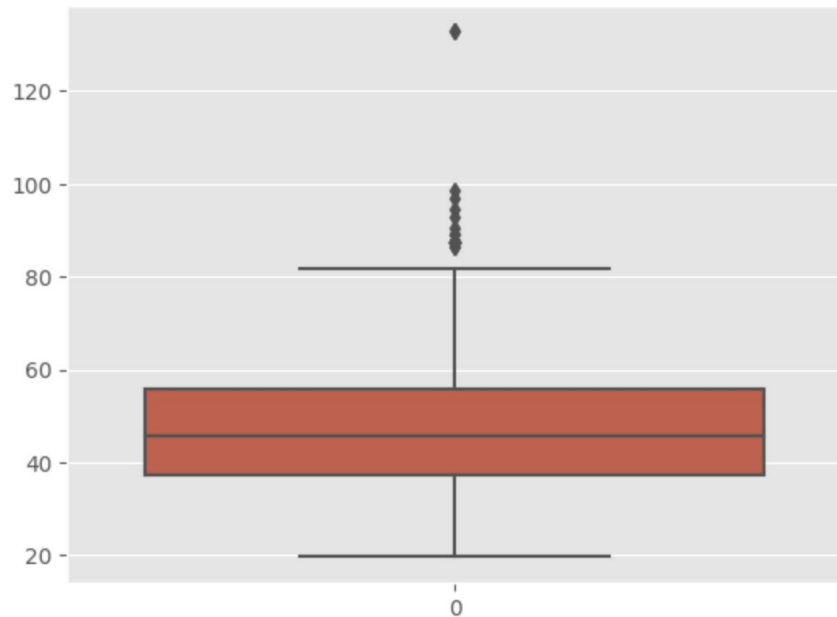
# Корреляция признаков



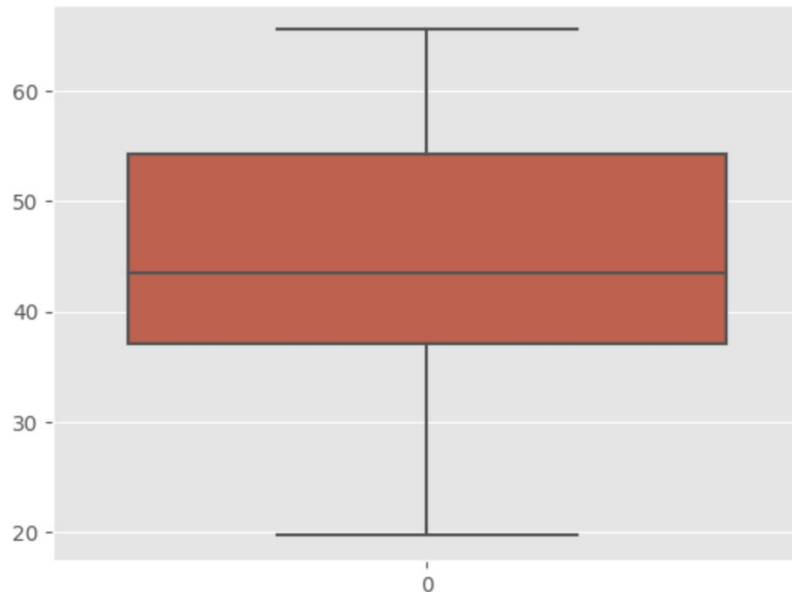
# Распределение цен



# Распределение площади

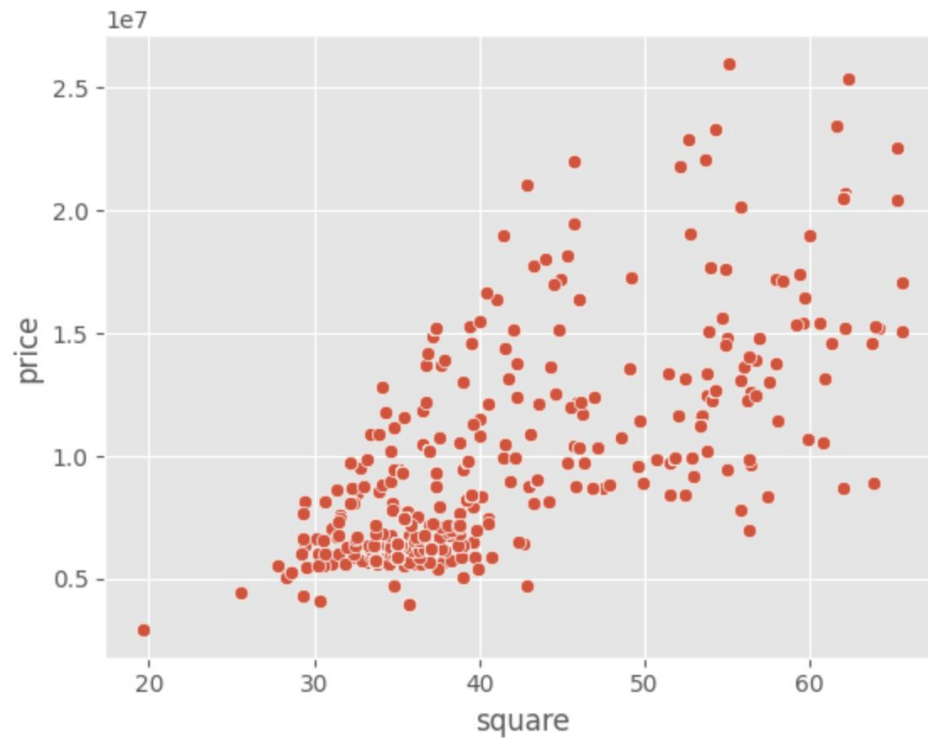


До предобработки



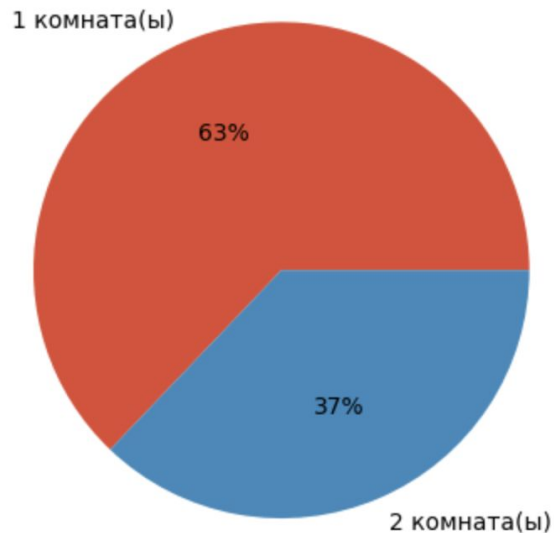
После

# Зависимость цены и площади квартиры





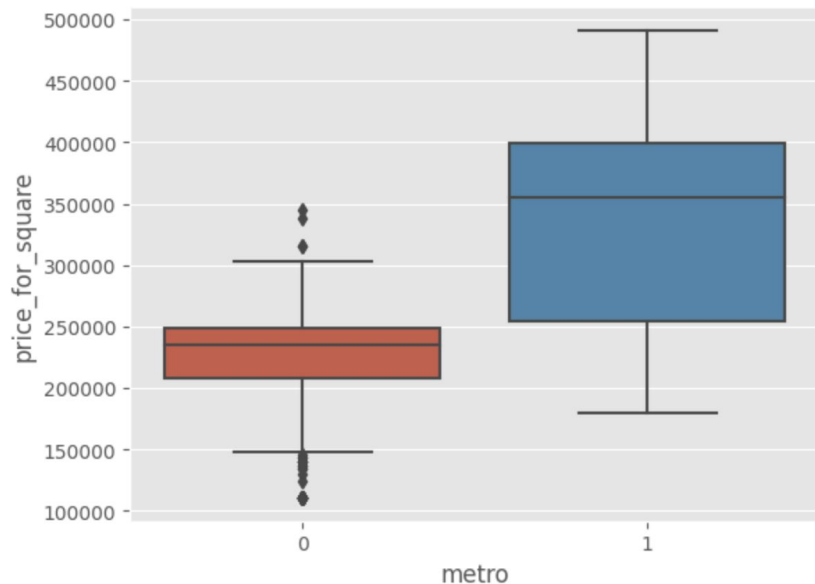
# Зависимость сдачи квартиры от количества комнат



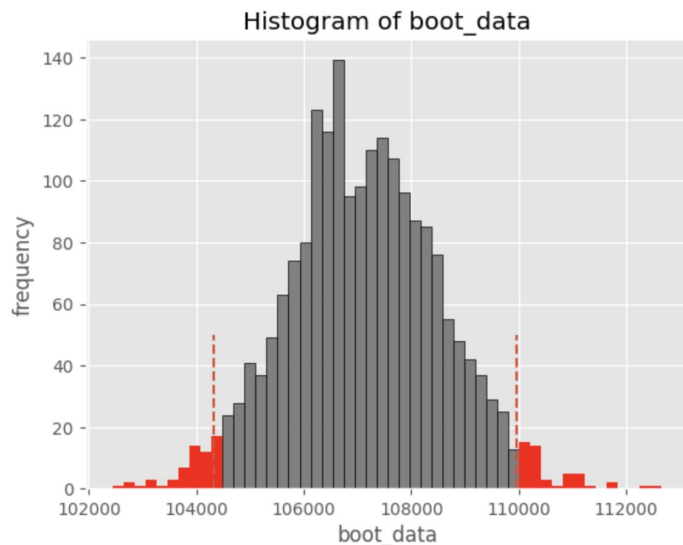
Однокомнатные квартиры сдают чаще

	test	lambda	chi2	dof	pval	cramer	power
0	pearson	1.000000	177.873463	1.0	1.411707e-40	0.162092	1.0
1	cressie-read	0.666667	179.692380	1.0	5.657004e-41	0.162919	1.0
2	log-likelihood	0.000000	184.282811	1.0	5.628184e-42	0.164986	1.0
3	freeman-tukey	-0.500000	188.637208	1.0	6.306736e-43	0.166924	1.0
4	mod-log-likelihood	-1.000000	193.860926	1.0	4.566860e-44	0.169220	1.0
5	neyman	-2.000000	207.310718	1.0	5.304061e-47	0.174991	1.0

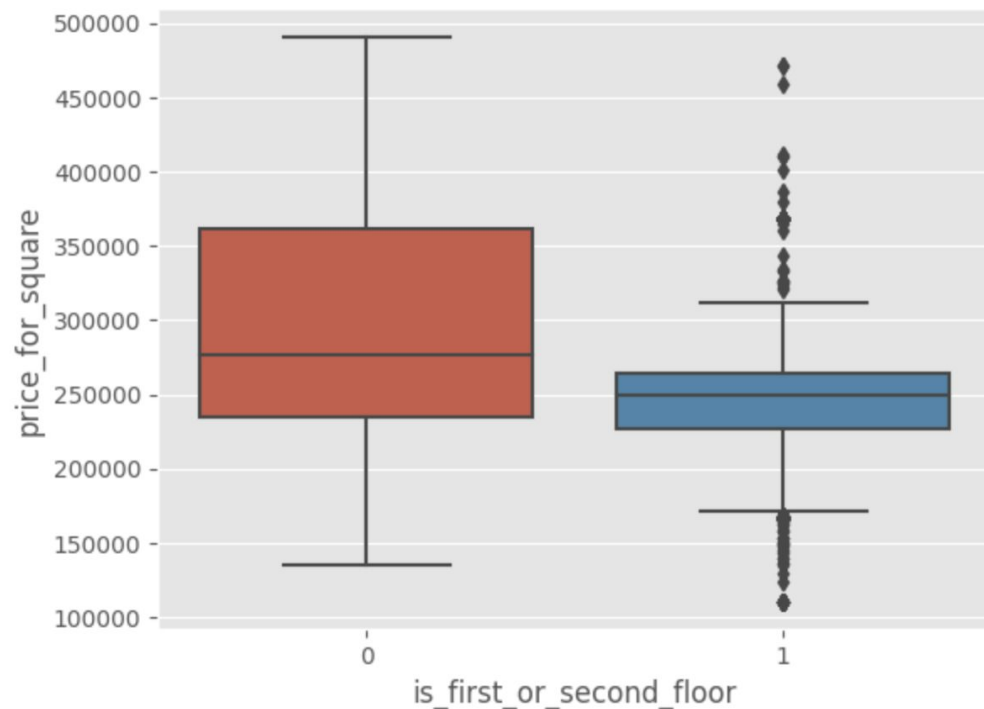
# Влияние наличия метро в шаговой доступности на цену за квадратный метр



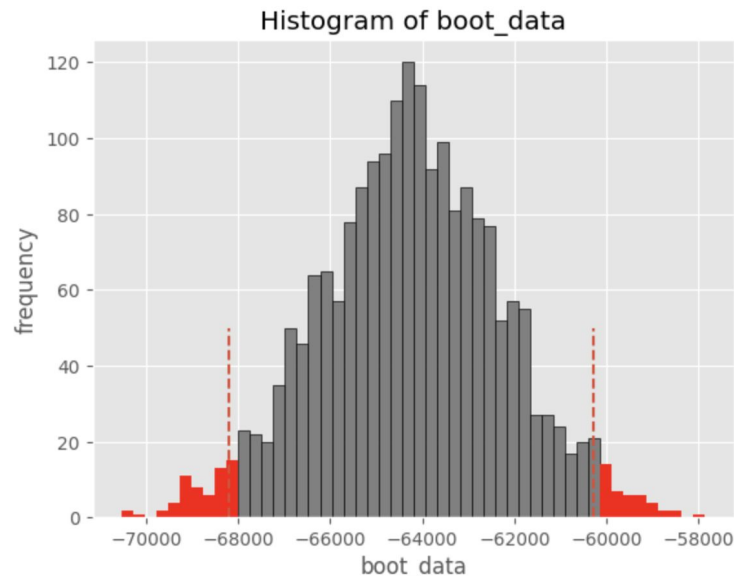
95% доверительный интервал разниц средних не включает 0



# Влияет ли то, что квартира находится на первых двух этажах на цену за квадратный метр?



95% доверительный интервал  
разниц средних не включает 0



# Предсказание цены квартиры

Будем предсказывать по 3 признакам: наличие метро в шаговой доступности, площади квартиры и находится ли квартира на первых двух этажах.

Использовать будем 3 модели:

- 1) Логистическая регрессия
- 2) Решающие деревья
- 3) Random Forest

# Логистическая регрессия

```
logreg.fit(X_train, y_train)
print('r2: ', cross_val_score(logreg, X, y, cv=5, scoring='r2'))
print('mae: ', cross_val_score(logreg, X, y, cv=5, scoring='neg_mean_absolute_error'))
print('mae: ', cross_val_score(logreg, X, y, cv=5, scoring='neg_mean_squared_error'))
```

```
r2: [0.72671016 0.7321032 0.63630391 0.60899283 0.5567388 ]
mae: [-1706793.66322009 -1695953.47932053 -2072445.85450517 -1952611.3042836
      -2212113.51329394]
mae: [-7.15788767e+12 -7.03786976e+12 -9.53146002e+12 -1.03327324e+13
      -1.16940794e+13]
```

```
logreg.score(X_test, y_test)
```

```
0.3599212210733629
```

# Решающие деревья

```
dectrreg.fit(X_train, y_train)
print('r2: ',cross_val_score(dectrreg, X, y, cv=5, scoring='r2'))
print('mae: ',cross_val_score(dectrreg, X, y, cv=5, scoring='neg_mean_absolute_error'))
print('mae: ',cross_val_score(dectrreg, X, y, cv=5, scoring='neg_mean_squared_error'))
```

```
r2: [0.88437716 0.99999942 0.99999967 0.99983613 0.99999959]
mae: [-6.44123903e+05 -5.36357748e+02 -4.10984945e+02 -2.31489311e+03
      -4.56212225e+02]
mae: [-3.02169681e+12 -1.52557588e+07 -8.78104190e+06 -4.28093829e+09
      -1.07218198e+07]
```

```
dectrreg.score(X_test, y_test)
```

```
0.9865428715212297
```

## Подбор гиперпараметров

Лучшая модель:

```
{'min_samples_split': 10,
 'min_samples_leaf': 20,
 'max_leaf_nodes': 100,
 'max_depth': 8}
```

```
dectrreg_2.score(X_test, y_test)
```

```
0.9810854637768409
```

```
print('r2: ',cross_val_score(dectrreg_2, X, y, cv=5, scoring='r2'))
print('mae: ',cross_val_score(dectrreg_2, X, y, cv=5, scoring='neg_mean_absolute_error'))
print('mae: ',cross_val_score(dectrreg_2, X, y, cv=5, scoring='neg_mean_squared_error'))
```

```
r2: [0.88113792 0.99869857 0.99880703 0.99859067 0.99880956]
mae: [-653423.41376314 -84313.33857538 -79231.86431725 -83901.74147835
      -79614.18870237]
mae: [-3.10635126e+12 -3.41899193e+10 -3.13292420e+10 -3.68185149e+10
      -3.11791408e+10]
```

# Random Forest

```
randfreg.fit(X_train, y_train)
print('r2: ', cross_val_score(randfreg, X, y, cv=5, scoring='r2'))
print('mae: ', cross_val_score(randfreg, X, y, cv=5, scoring='neg_mean_absolute_error'))
print('mae: ', cross_val_score(randfreg, X, y, cv=5, scoring='neg_mean_squared_error'))
```

```
r2: [0.885098  0.99999947 0.99999964 0.99992349 0.99999947]
mae: [-6.41398908e+05 -4.99177624e+02 -3.64092955e+02 -2.09517949e+03
      -4.35391664e+02]
mae: [-3.01492932e+12 -1.23973696e+07 -7.68133008e+06 -2.47826178e+09
      -8.35927894e+06]
```

```
randfreg.score(X_test, y_test)
```

0.9895943477701786

## Подбор гиперпараметров

Лучшая модель:

```
{'n_estimators': 1600,
 'min_samples_split': 5,
 'min_samples_leaf': 1,
 'max_depth': 10,
 'bootstrap': True}
```

```
randfreg_2.score(X_test, y_test)
```

0.9896303743981283

```
print('r2: ', cross_val_score(randfreg_2, X, y, cv=5, scoring='r2'))
print('mae: ', cross_val_score(randfreg_2, X, y, cv=5, scoring='neg_mean_absolute_error'))
print('mae: ', cross_val_score(randfreg_2, X, y, cv=5, scoring='neg_mean_squared_error'))
```

```
r2: [0.88502092 0.99999773 0.99999877 0.99998096 0.99999842]
mae: [-643165.6373389 -1383.55528098 -1120.21051319 -2051.52361725
      -1135.46604021]
mae: [-3.00943895e+12 -5.95327681e+07 -3.11557487e+07 -4.41365477e+08
      -3.86741933e+07]
```

# Лучший результат

- Модель: Random Forest Regressor
- Параметры: `{'n_estimators': 1600,  
 'min_samples_split': 5,  
 'min_samples_leaf': 1,  
 'max_depth': 10,  
 'bootstrap': True}`

```
randfreg_2.score(X_test, y_test)
```

0.9896303743981283

- Метрики:

```
print('r2: ',cross_val_score(randfreg_2, X, y, cv=5, scoring='r2'))  
print('mae: ',cross_val_score(randfreg_2, X, y, cv=5, scoring='neg_mean_absolute_error'))  
print('mae: ',cross_val_score(randfreg_2, X, y, cv=5, scoring='neg_mean_squared_error'))
```

```
r2: [0.88502092 0.99999773 0.99999877 0.99998096 0.99999842]  
mae: [-643165.6373389 -1383.55528098 -1120.21051319 -2051.52361725  
      -1135.46604021]  
mae: [-3.00943895e+12 -5.95327681e+07 -3.11557487e+07 -4.41365477e+08  
      -3.86741933e+07]
```



# Анализ квартир с сайта СІАН в Санкт-Петербурге

Клочков Владислав Сергеевич  
Группа 19.Б08-мм  
st068203@student.spbu.ru

29 апреля 2023