

Предсказание цен на недвижимость в Москве

Цветов Алексей Эдуардович

О себе

- Цветов Алексей
- Высшее (Нижегородский Государственный Технический Университет НГТУ им. Р.Е. Алексеева, направление Информатика и Вычислительная техника, магистр 2019г.)
- Опыт в Сбере (4 года, Центр защиты конфиденциальной информации, старший инженер, сбор и анализ информации по инцидентам с целью выработки и принятия решений и мер по обеспечению защиты информации и эффективному использованию средств автоматического контроля, обнаружения возможных каналов утечки сведений, представляющих персональную, государственную, служебную и коммерческую тайну.
- г. Нижний Новгород
- tsvetov.a.ed@sberbank.ru, 8-910-888-42-58

Описание проекта

Поскольку все большее число граждан и предприятия вовлекается в процессы на рынке недвижимого имущества, появляется необходимость определенно охарактеризовать свойства недвижимого имущества как особого вида товара. Одним из таких свойств является стоимость.

Рынок жилой недвижимости Российской Федерации представляет собой сложную структуру, состоящую из миллионов квартир, характеризующихся множеством признаков. При этом любые изменения на рынке могут стать поводом для спекуляций и преднамеренного увеличения цен на недвижимость. Поэтому так важно понимать, где реальная стоимость квартиры, а где завышенная. В большинстве случаев самостоятельная оценка стоимости невозможна, привлекаются специалисты со стороны, что ведет к затратам денег и времени.

Описание проекта

С этой целью Сбербанк в 2017 г. открывает соревнование на площадке Kaggle для разработки алгоритмов, которые используют широкий спектр функций для прогнозирования цен на недвижимость.

Объектом исследования данной работы выступает рынок жилой недвижимости в г. Москва.

Предметом исследования является использование методов машинного обучения для прогнозирования стоимости объектов жилой недвижимости.

Цель данной работы - создание модели прогнозирования стоимости жилой недвижимости, созданной с применением методов машинного обучения.

Соревнование на площадке Kaggle

Данные:

<https://www.kaggle.com/c/sberbank-russian-housing-market/data>

Объем данных:

Train – 30 471

Test – 7 662

Метрика:

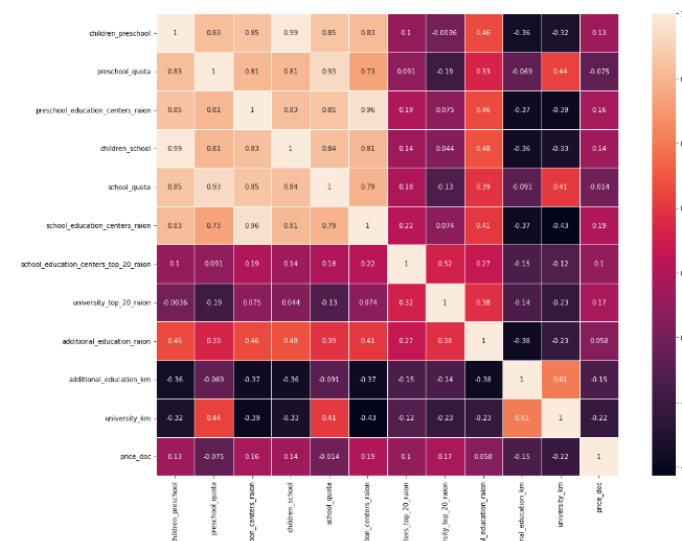
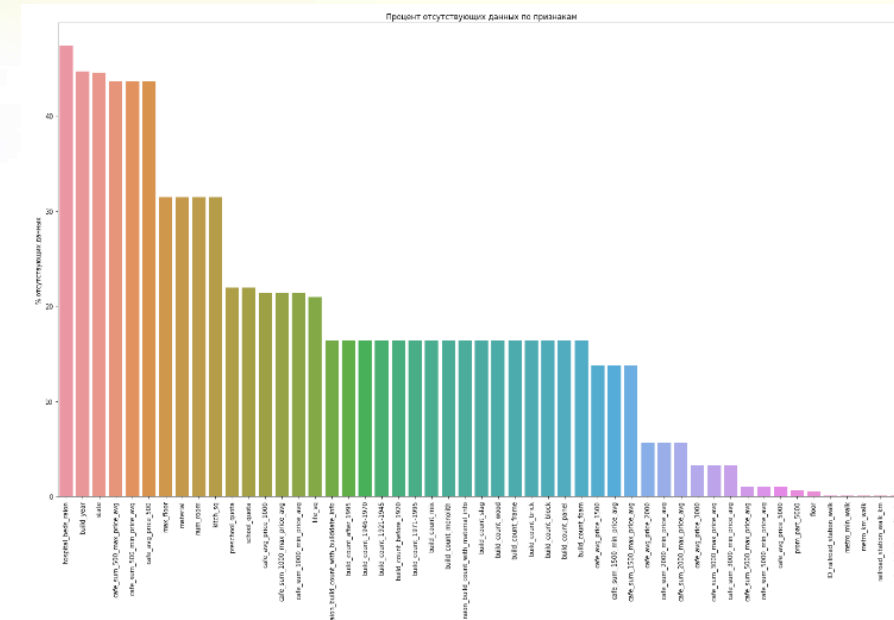
RMSLE(Среднеквадратичная логарифмическая ошибка (Root Mean Squared Logarithmic Error))

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(\hat{y}_i + 1) - \log(y_i + 1))^2}$$

Бизнес-логика

1. Анализ данных:

- просмотр общей информации о выборке
- пропуски в данных
- зависимость целевой переменной от других признаков для квартиры
- корреляция цены от признаков



Ссылка на репозиторий:

https://github.com/AlexTsvetov/School_DS_TsvetovAL/tree/main/Project

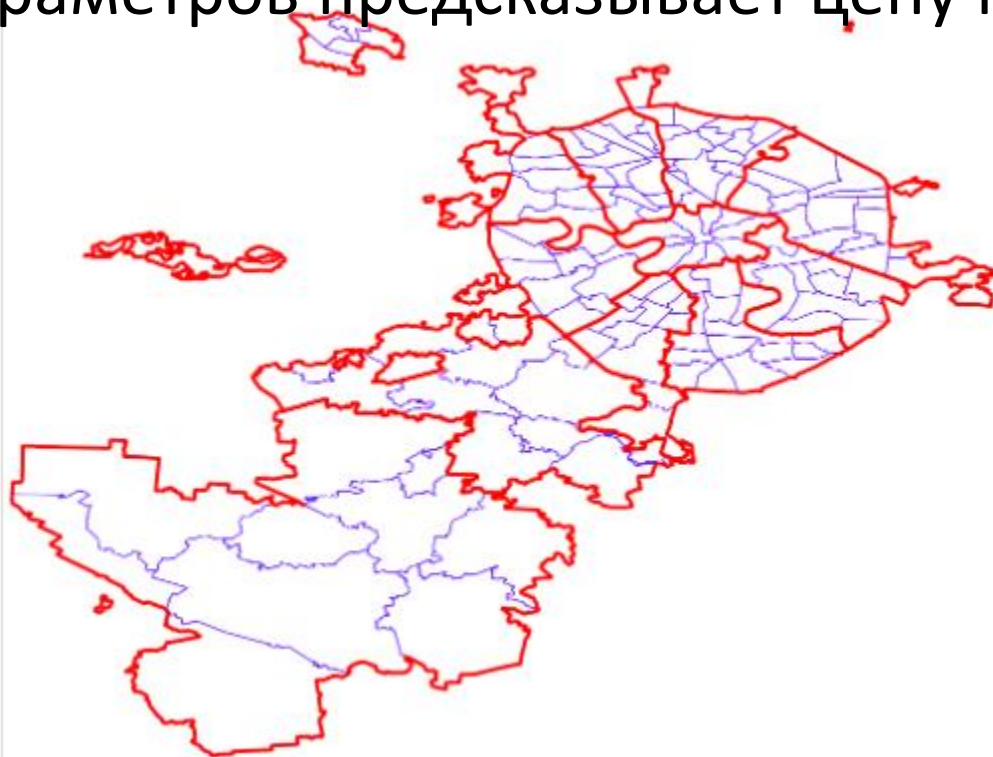
Бизнес-логика



2. Разделение и очистка данных

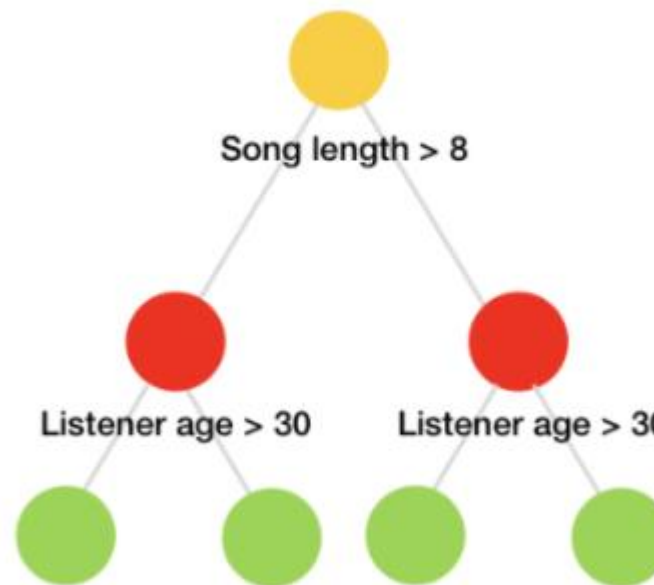
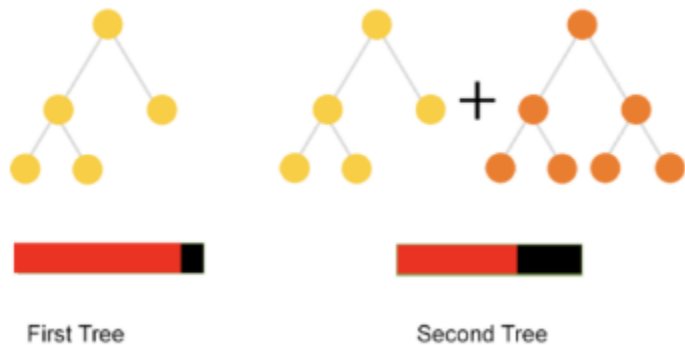
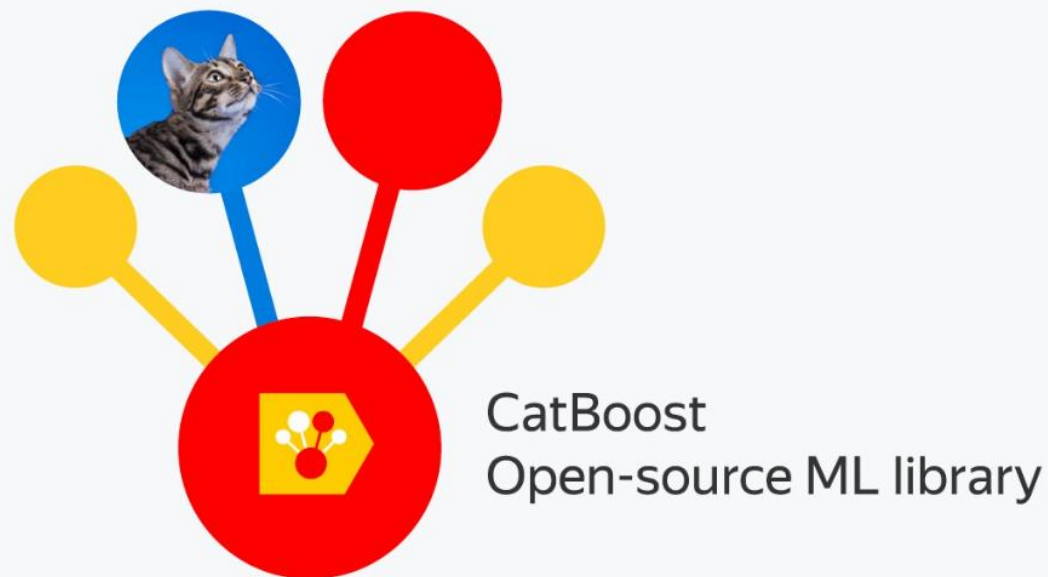
3. Подготовка и обучение модели

4. Создание веб приложения(Dash), которое при указании параметров предсказывает цену квартиры



Модель данных

CatBoost – библиотека, которая была разработана Яндексом в 2017 году, представляет разновидность семейства алгоритмов Boosting и является усовершенствованной реализацией Gradient Boosting Decision Trees (GBDT). CatBoost имеет поддержку категориальных переменных и обеспечивает высокую точность.



Веб приложение по предсказанию на Dash

Расчет стоимости недвижимости в Москве

Укажите параметры квартиры, для расчета ее стоимости:

Общая площадь квартиры:

Жилая площадь:

Этаж в доме:

Количество этажей в доме:

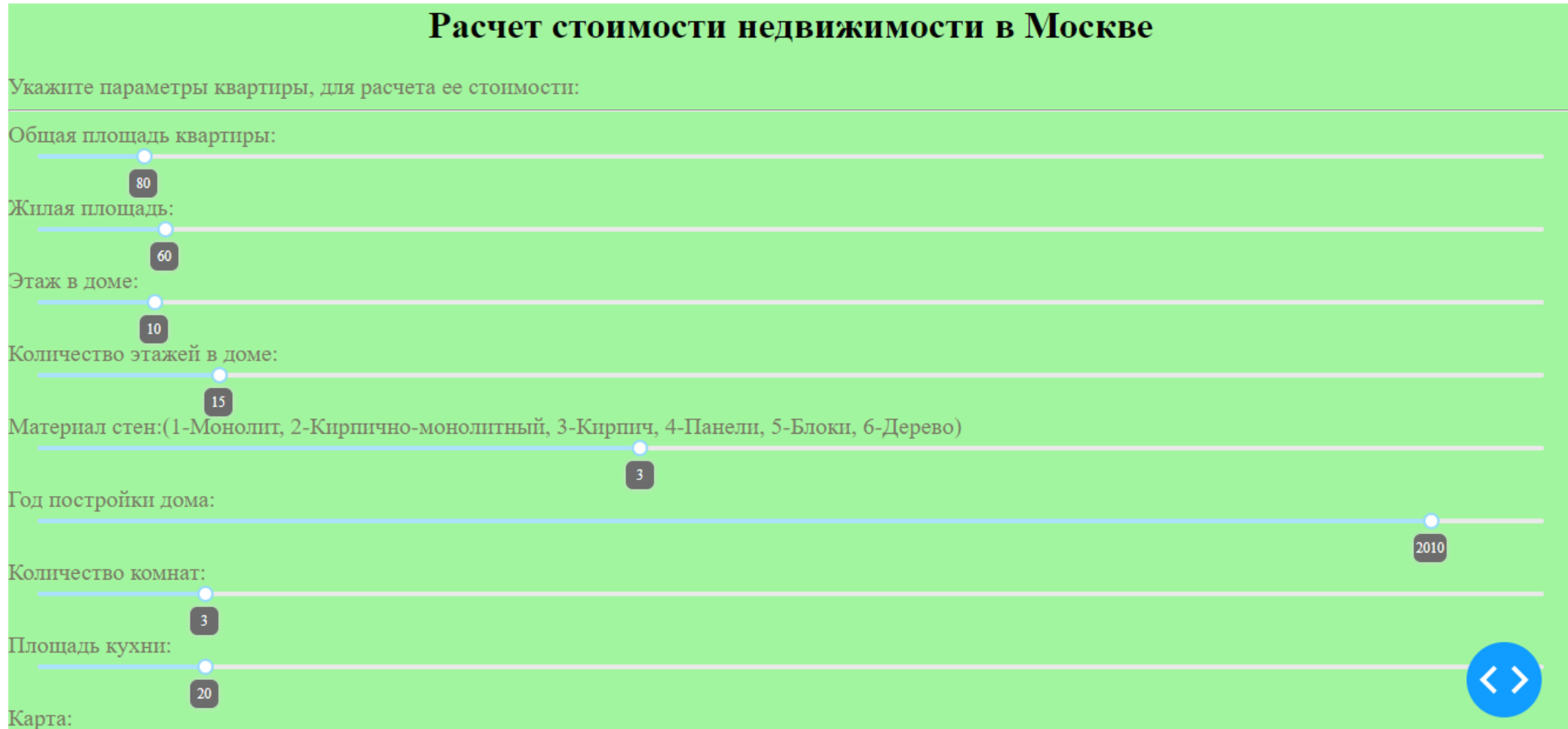
Материал стен:(1-Монолит, 2-Кирпично-монолитный, 3-Кирпич, 4-Панели, 5-Блоки, 6-Дерево)

Год постройки дома:

Количество комнат:

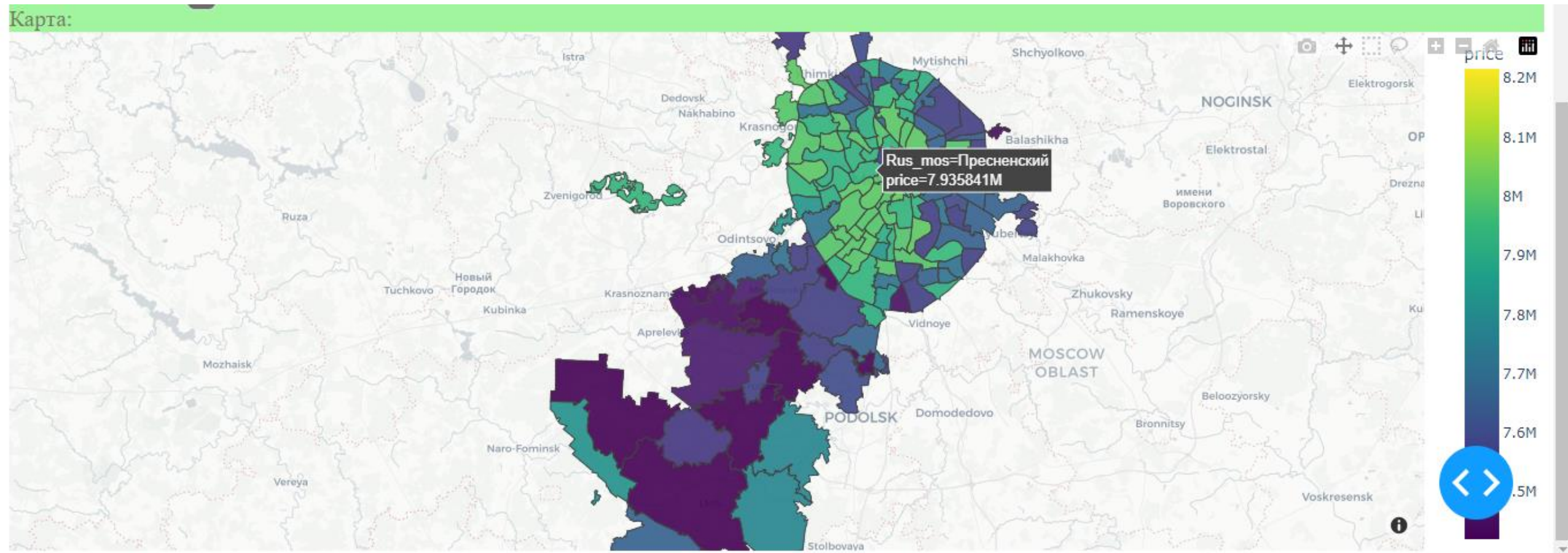
Площадь кухни:

Карта:



Параметр	Значение
Общая площадь квартиры:	80
Жилая площадь:	60
Этаж в доме:	10
Количество этажей в доме:	15
Материал стен:	3 (Кирпич)
Год постройки дома:	2010
Количество комнат:	3
Площадь кухни:	20
Карта:	

Веб приложение по предсказанию на Dash



Оценка модели на Kaggle

Sberbank Russian Housing Market
Can you predict realty price fluctuations in Russia's volatile economy?

 Sberbank · 3,264 teams · 6 years ago

\$25,000
Prize Money

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [Submissions](#) [Late Submission](#) [...](#)

Submissions

You selected 0 of 2 submissions to be evaluated for your final leaderboard score. Since you selected less than 2 submission, Kaggle auto-selected up to 2 submissions from among your public best-scoring unselected submissions for evaluation. The evaluated submission with the best Private Score is used for your final score.

0/2

☒ Submissions evaluated for final score




All

Successful

Selected

Errors

Recent ▾

Submission and Description		Private Score ⓘ	Public Score ⓘ	Selected
 prediction_test.csv Complete (after deadline) · 10d ago		0.32536	0.31665	<input type="checkbox"/>
 prediction_80.csv Complete (after deadline) · 19d ago		0.32788	0.31896	<input type="checkbox"/>
 prediction.csv Complete (after deadline) · 24d ago		0.32014	0.3157	<input type="checkbox"/>

Спасибо за внимание!