**Отчёт по итоговому домашнему заданию  
по предмету “**[**Прикладные методы математической статистики**](https://lms.hse.ru/student.php?lessons_ID=143541)**”**

Выполнен студентами группы БПИ204:  
Пеганов Никита  
Куляхтина Алёна  
Завалинский Даниил

**Описание данных и особенности работы с ними**

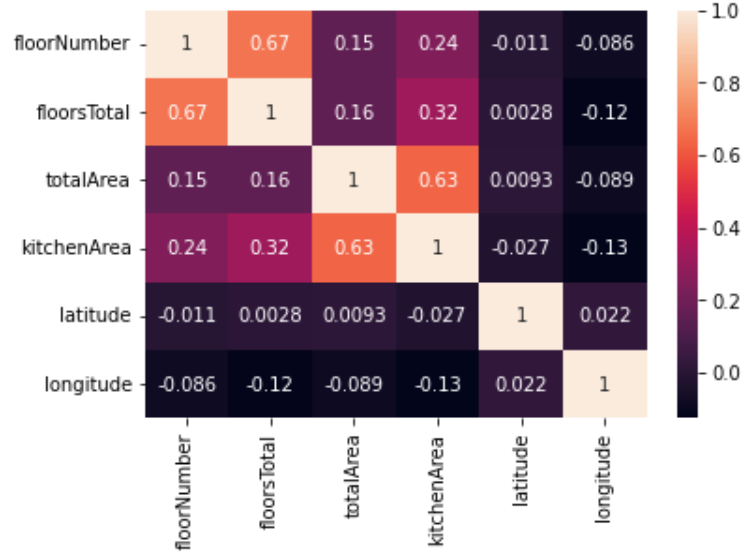
Выбранные нами данные взяты из статьи <https://medium.com/@max.bobkov/machine-learning-moscow-flats-appraising-25a1e9f171db>. По словам автора, они получены в результате парсинга сайта [www.cian.ru](http://www.cian.ru), специализирующегося на продаже квартир. Также данные очищены «от дублей, объявлений с одинаковым текстом в описании, от строк с нулевой площадью кухни, без материала стен и этажей в доме».

Для оценки полученных результатов данные были разделены на две выборки: обучающую и тестовую в пропорции 70 к 30. Обработка данных происходила в среде jupyter notebook на языке python. Код решения можно найти в файле processing.ipynb.

**Описание переменных, используемых в анализе**

Рассмотрим переменные, использованные в датасете:  
wallsMaterial — **материал стен**floorNumber — **на каком этаже расположена квартира**  
floorsTotal — общее число этажей в доме  
totalArea — общая площадь квартиры  
kitchenArea — площадь кухни  
latitude — широта расположения дома на карте  
longitude — долгота расположения дома на карте  
price — цена квартиры

Проверим объясняющие переменные на мультиколлинеарность. Для этого составим матрицу корреляции числовых переменных:



Матрица корреляции показывает, что значительных попарных корреляций между переменными в данных нет. Они не превосходят 0,7. Это означает, что возможных проблем с качеством получаемых оценок не предвидится.

**Описание полученных регрессионных моделей**

**Описание проведенных тестов**