Московский авиационный институт

(национальный исследовательский институт)

Факультет «Прикладная математика и физика»

Кафедра вычислительной математики и программирования

**ОТЧЕТ**

**по ЛР №1 по курсу “Искусственный Интеллект”**

VI семестр

*Студент:* Скуридин Алексей

*Группа:*  М8О-304Б-17

*Оценка:*

*Дата:*

**Москва, 2020**

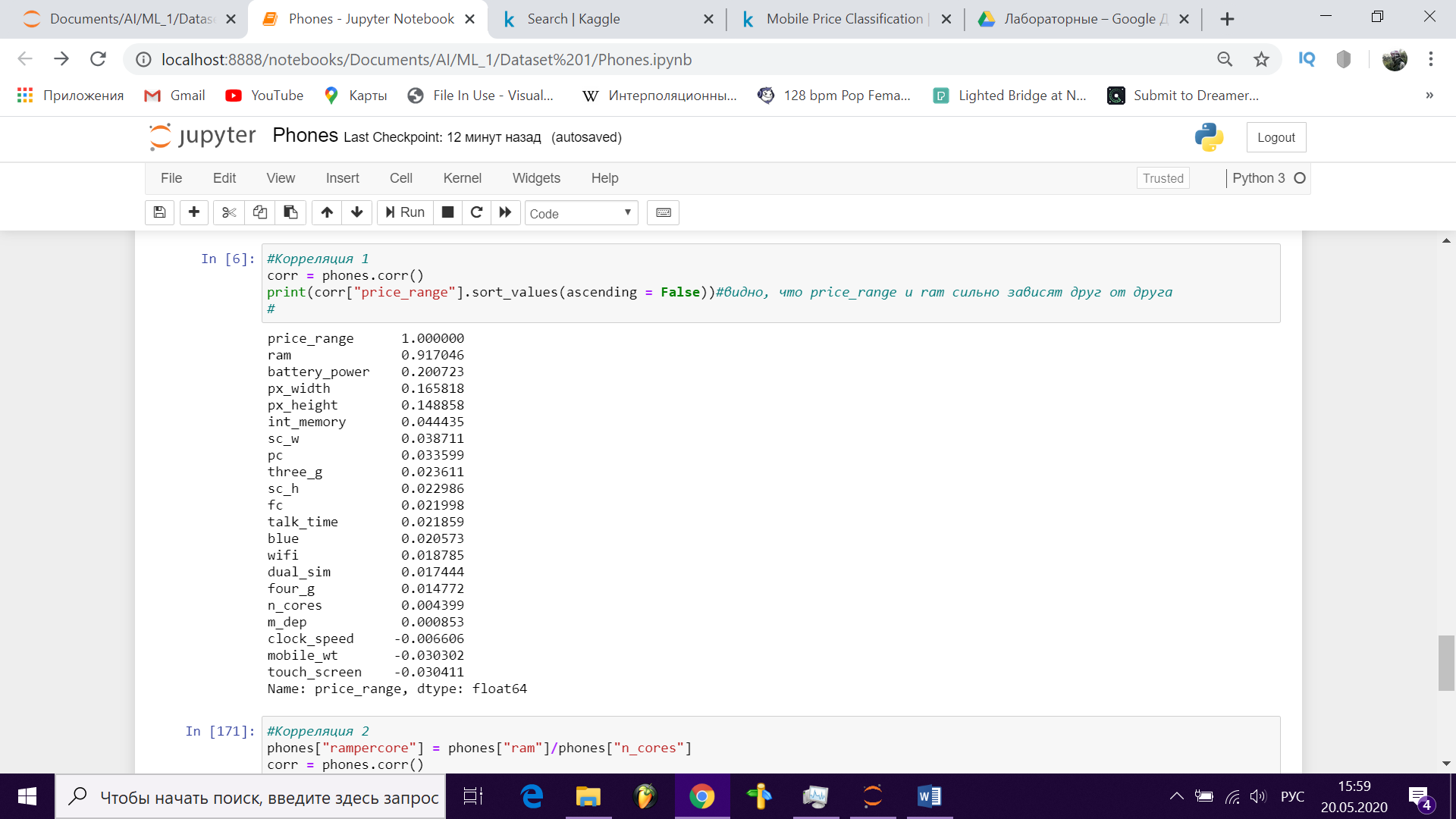
**Задача 1**

Рассматривается датасет мобильных телефонов. Всего в датасете 1000 телефонов. В нем ассматриваются следующие признаки: мощность батареи, наличие Bluetooth, частота ЦП, возможность использования двух сим-карт, разрешение фронтальной камеры, наличие 4G, объем встроенной памяти, толщина телефона, вес телефона, число ядер процессора, разрешение основной камеры, длина экрана в пикселях, ширина экрана в пикселях, количество оперативной памяти (RAM), длина экрана в см, ширина экрана в см, время работы телефона в режиме разговора, наличие 3G, наличие сенсорного экрана, наличие Wi-Fi, ценовая категория.

Задача классификации заключается в следующем. Существует четыре ценовые категории. 0 – самые дешевые телефоны, 3 – самые дорогие телефоны. Необходимо по заданным параметрам телефона отнести его к одному из классов.

**Анализ данных и отбор нужных признаков**

Сначала необходимо определить ряд признаков, которые я буду рассматривать. Для этого я вычислил корреляцию для price\_range и всех других признаков.



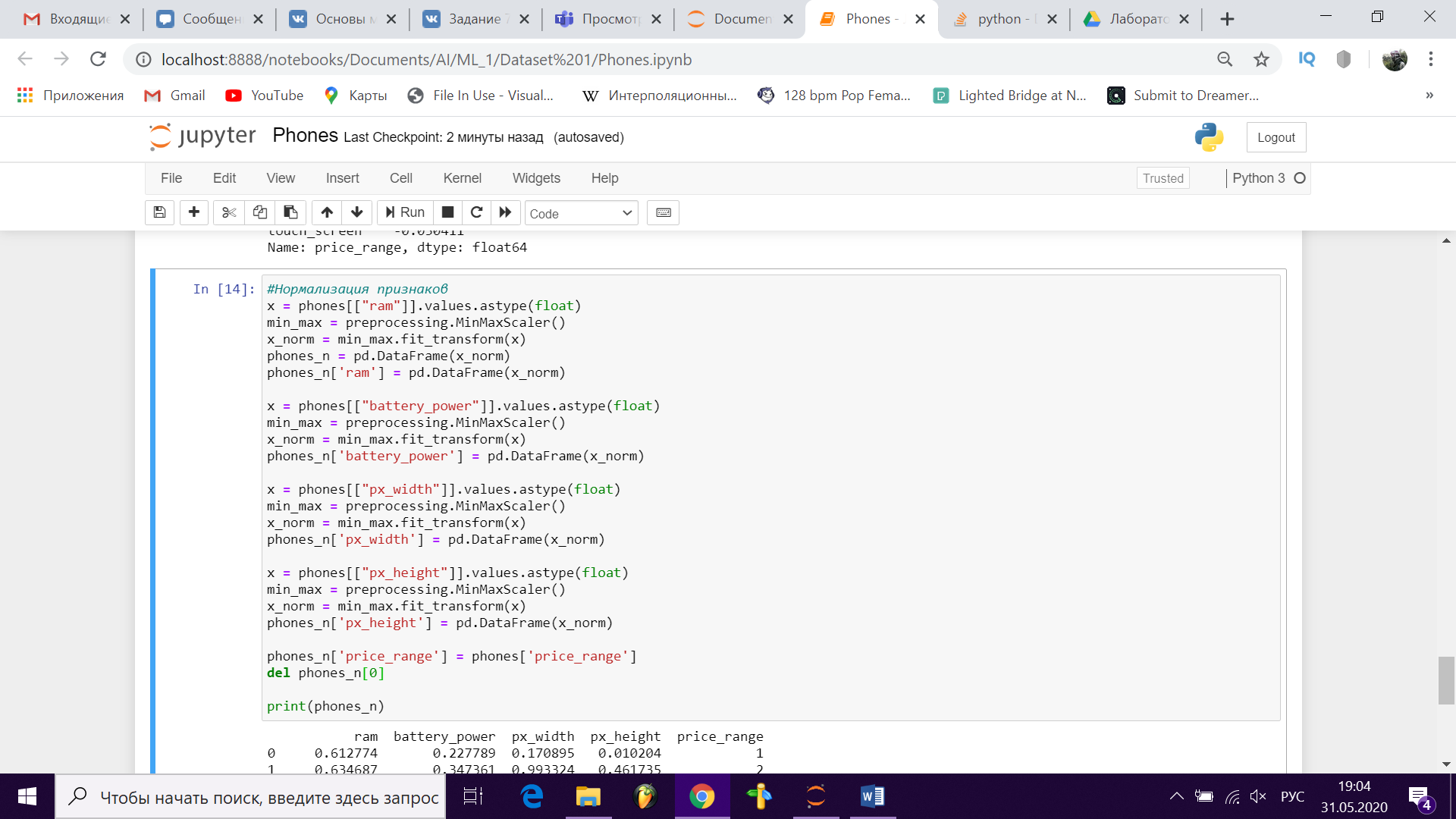
Видно, что корреляция price\_range и ram = 0.917046. Поэтому этот признак очень хорошо подходит для решаемой задачи. Также неплохо показывает себя признак battery\_power, корреляция равна 0.2. К этим признакам можно добавить px\_width и px\_height, с корреляцией 0.165818 и 0.148858 соответственно. Остальные признаки не подходят для решения поставленной задачи.

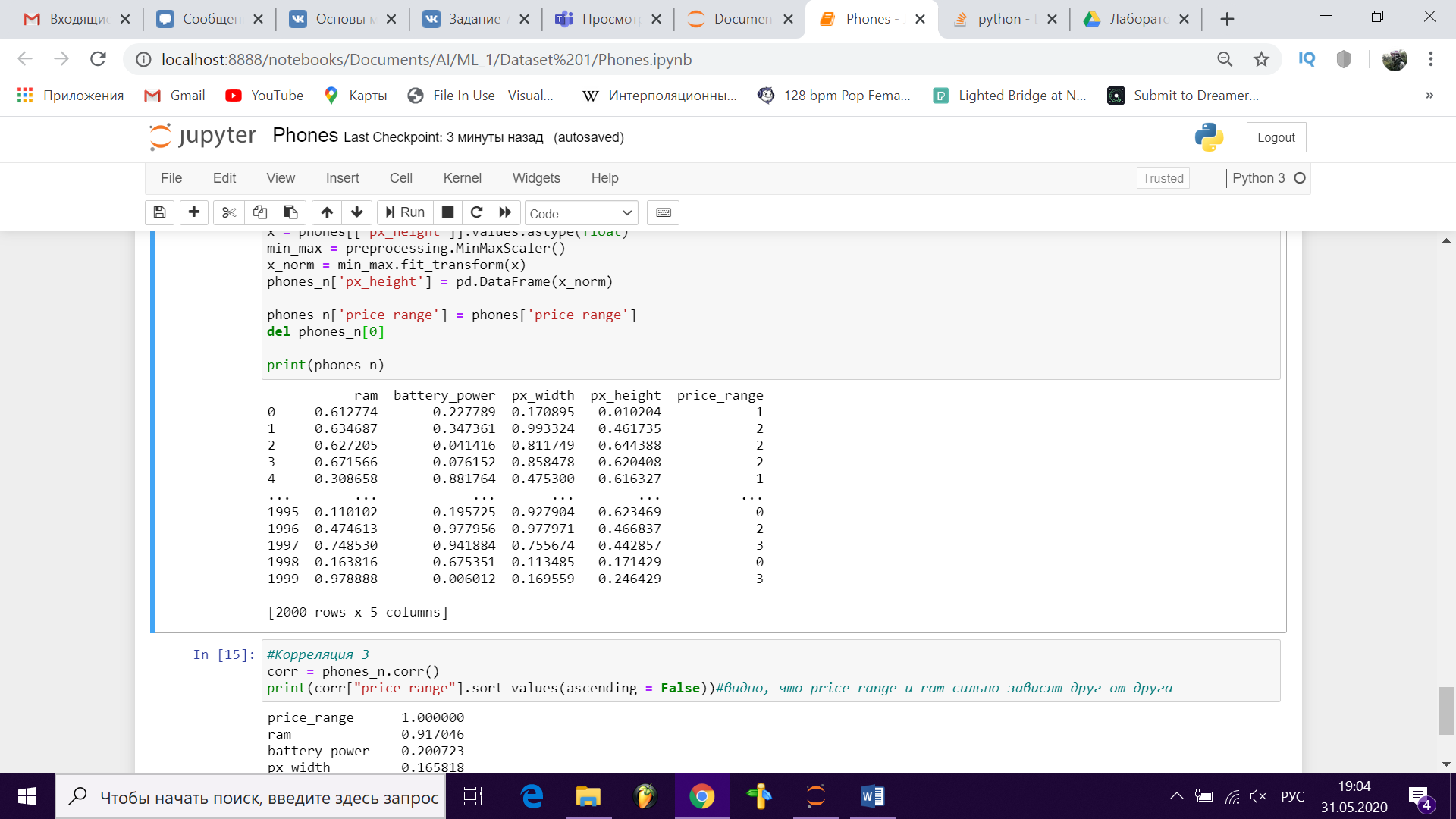
Интересно, что такие важные признаки, как разрешение камеры и частота ЦП не сильно отражаются на цене устройства.

Далее я попробовал разделить признак n\_cores (число ядер ЦП) на отдельные бинарные признаки (наличие 1 ядра, наличие 2 ядер и т.д.), но корреляция для полученных признаков была все равно маленькой.

Поскольку корреляция отображает зависимость нормально распределенных признаков, я решил рассматривать ещё несколько признаков с маленькой корреляцией, но которые обычно влияют на цену смартфона – объем встроенной памяти (int\_mem) и разрешение основной камеры (pc).

На всякий случай, я решил нормировать выбранные мной признаки.



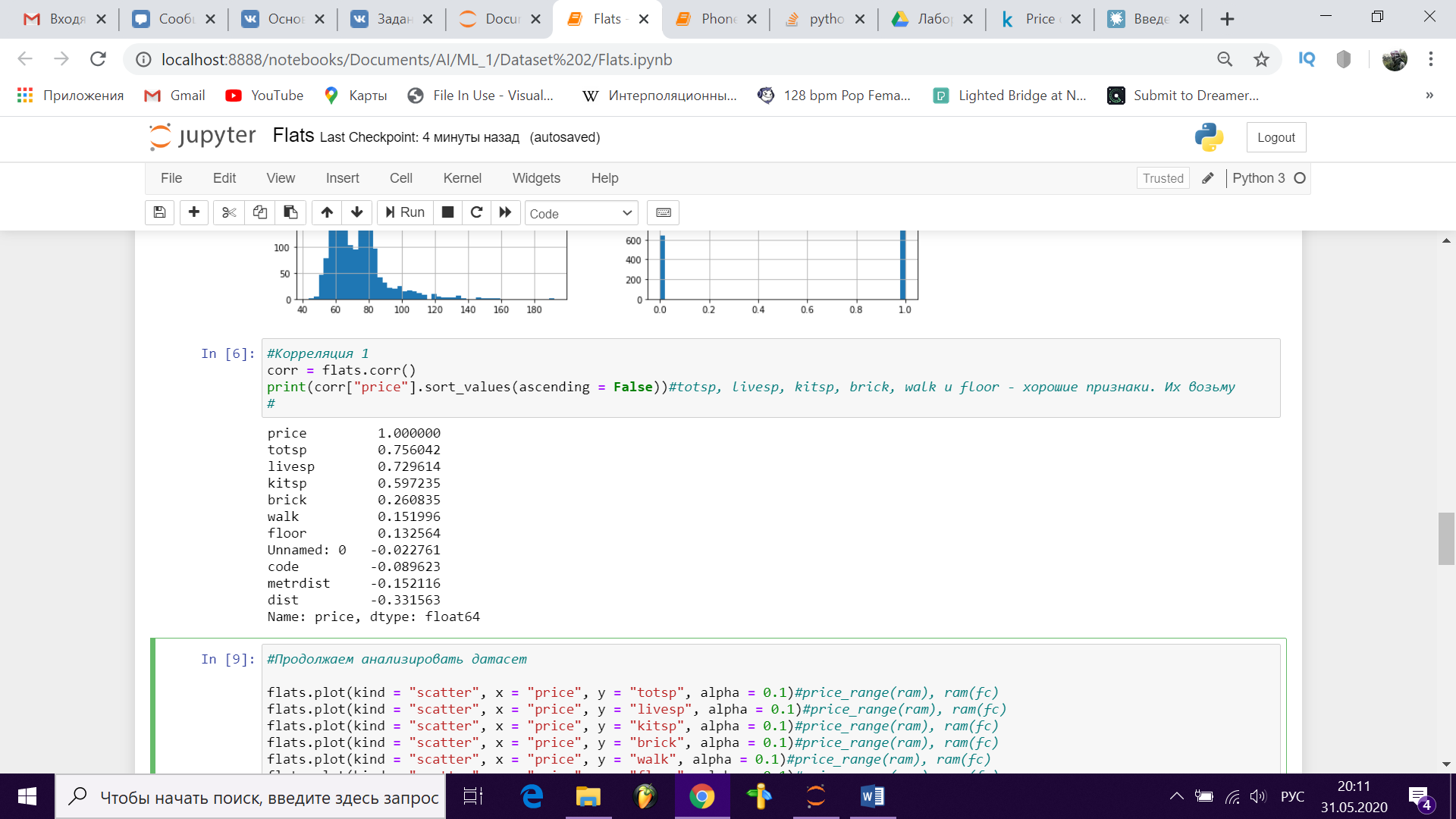


**Задача 2**

Датасет – квартиры в Москве. В нем рассматриваются следующие признаки: price (цена квартиры в тысячах $), totsp (общая площадь квартиры), livesp (жилая площадь квартиры), kitsp (площадь кухни), dist (расстояние от центра квартиры), metrdist (расстояние до ближайшей станции метро в минутах), walk – (бинарный признак, показывает в шаговой ли доступности метро), brick (кирпичный монолит или другой), floor (1 – этаж кроме первого и последнего, 0 – иначе).

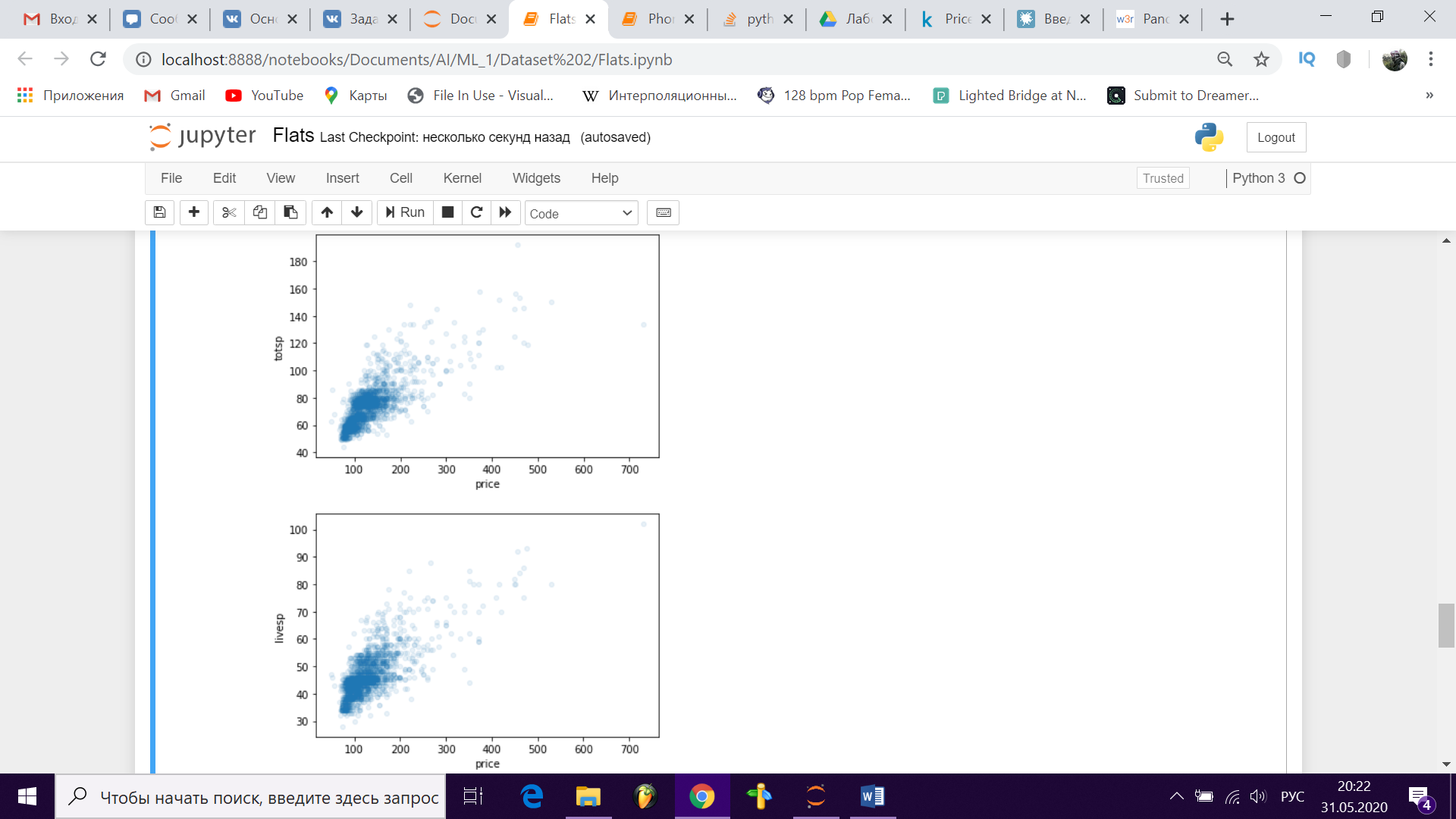
Мною рассматривается задача регрессии. Конкретно я хочу, чтобы программа предсказывала цену квартиры по заданным параметрам.

**Анализ данных и отбор нужных признаков**

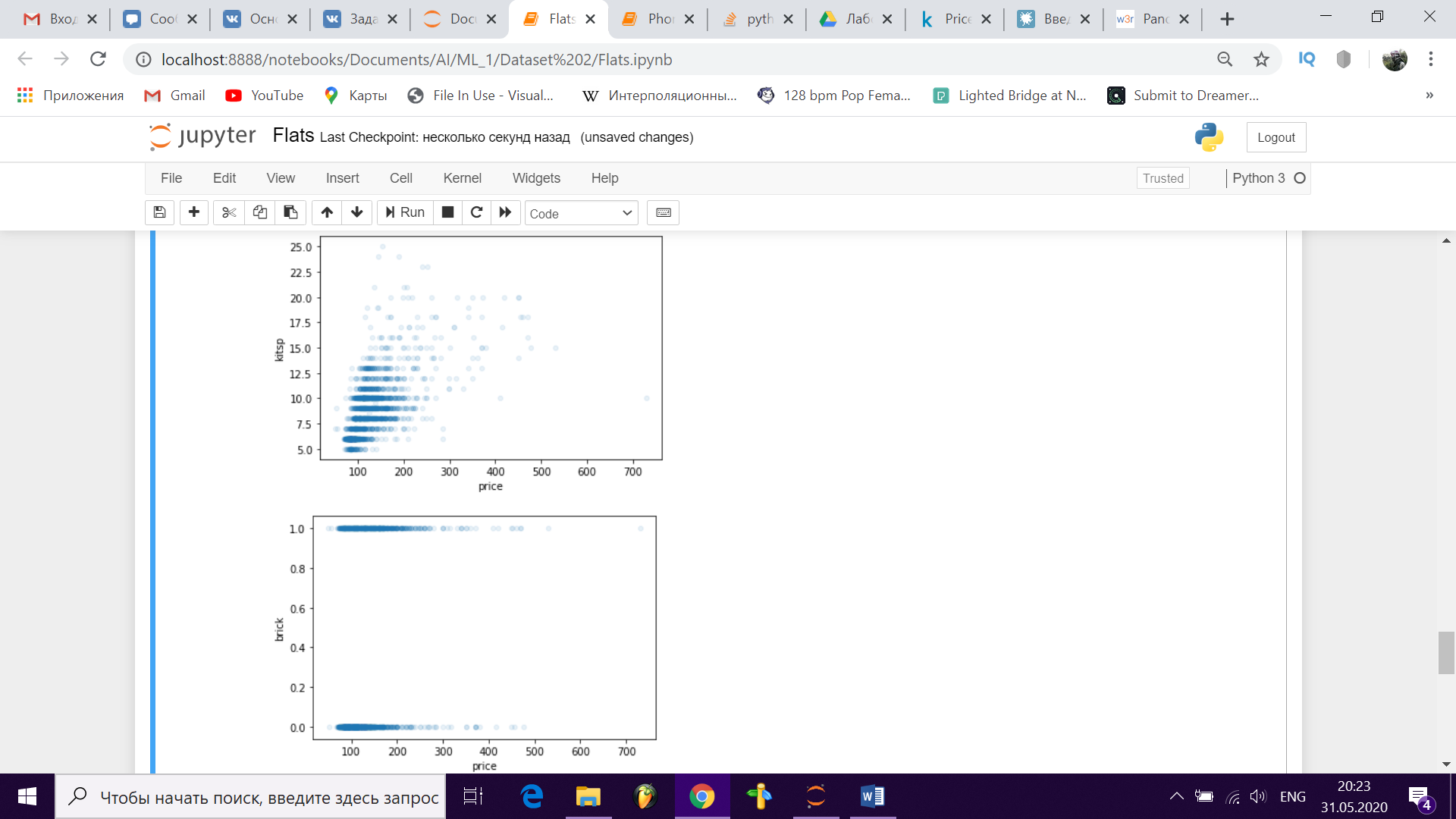


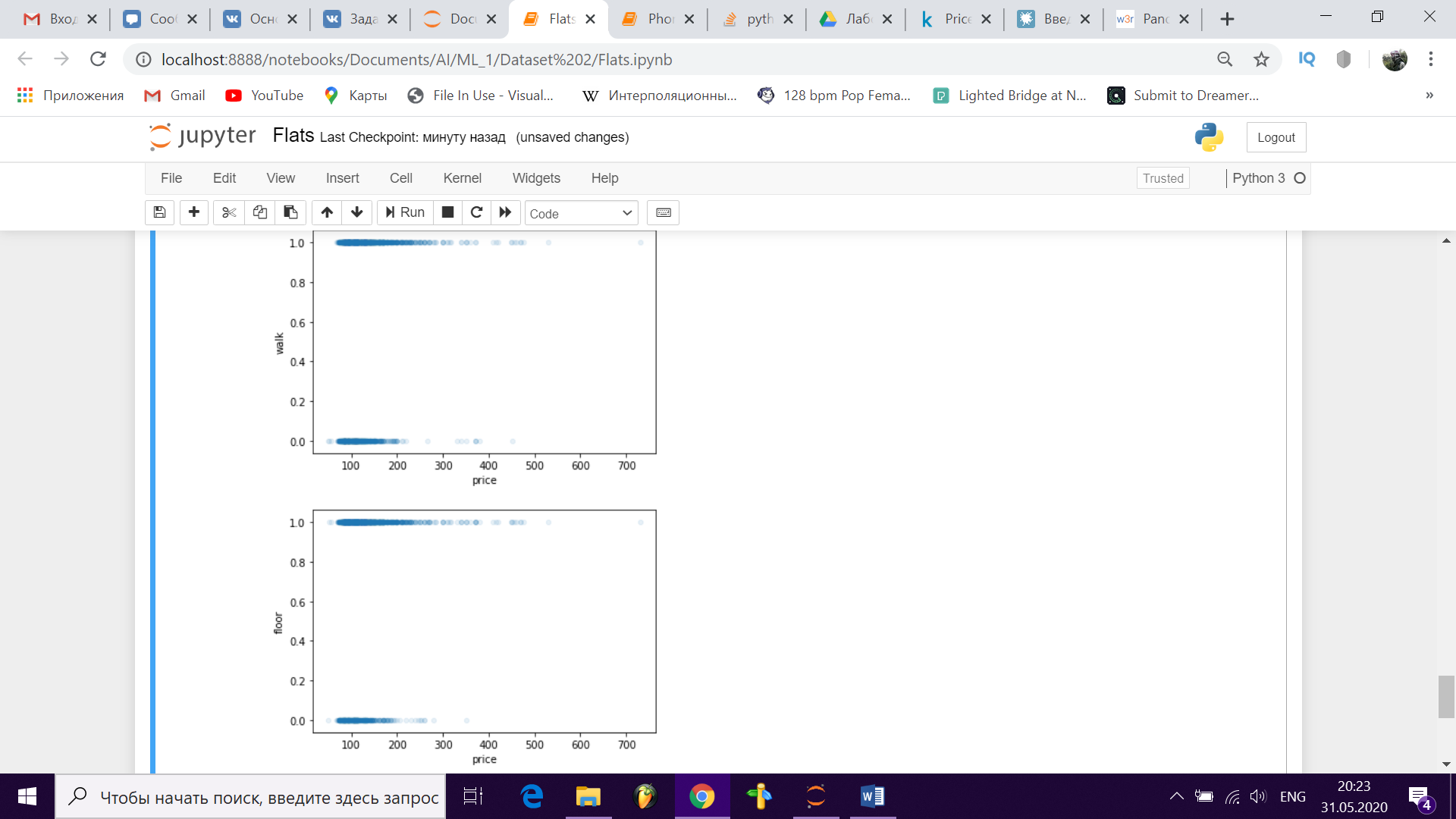
Значения корреляции оказались неплохими для признаков totsp, livesp, kitsp, brick, walk. Отсюда можно сделать вывод, что цена на квартиры в первую очередь зависит от таких факторов, как площадь квартиры. Dist – тоже неплохой признак, но корреляция у него отрицательная. Поэтому я возьму модуль этого признака. Все признаки я нормализовал.

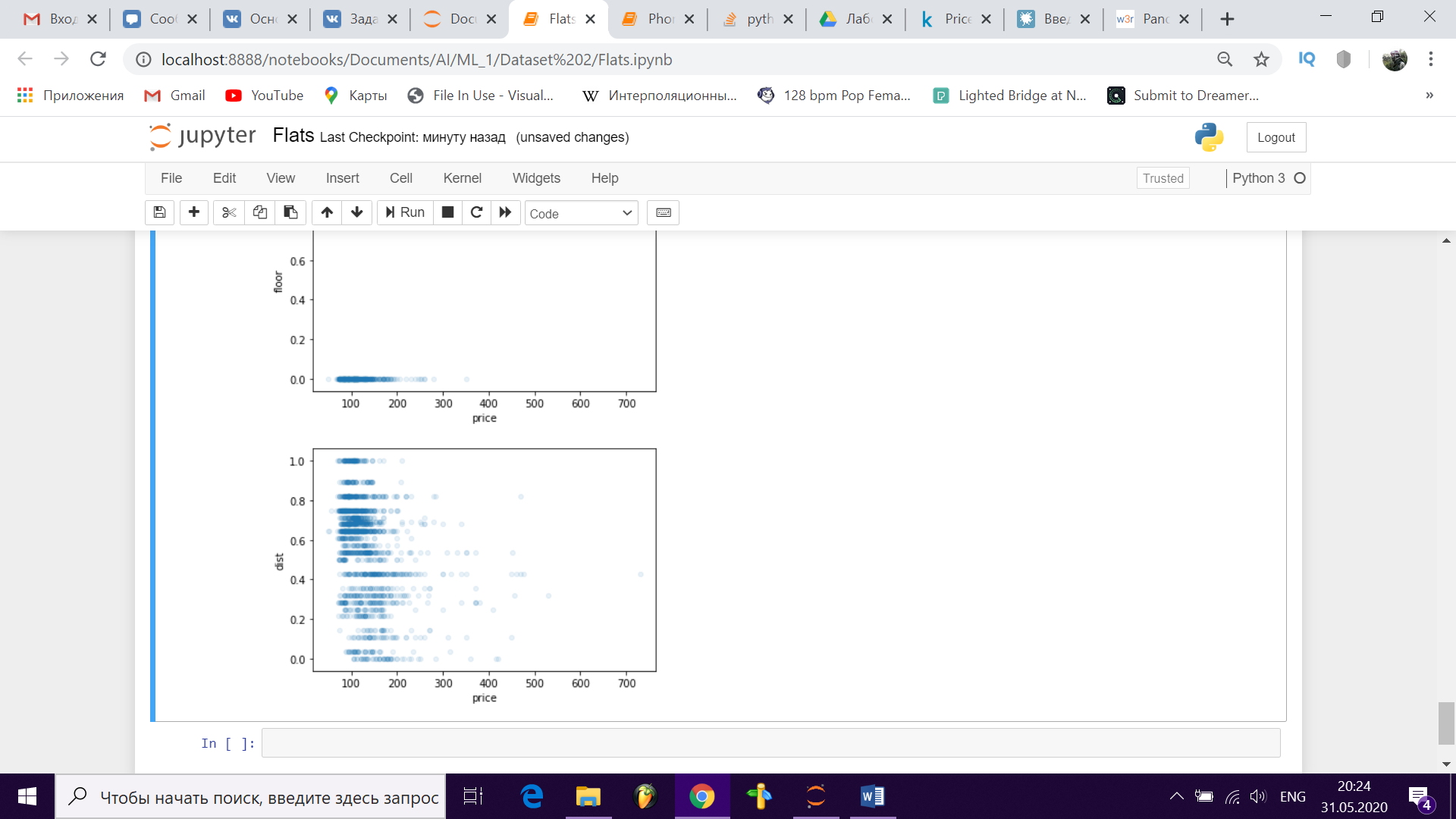
Далее я решил визуализировать зависимости



По большей части, все выглядит неплохо







Однако я не уверен насколько хорошо будут работать бинарные признаки, график получился не сильно понятный. Но я попробую поработать с этими признаками, посмотрю, что получится.