**Rent apartment price predict**

Turmanova Saule

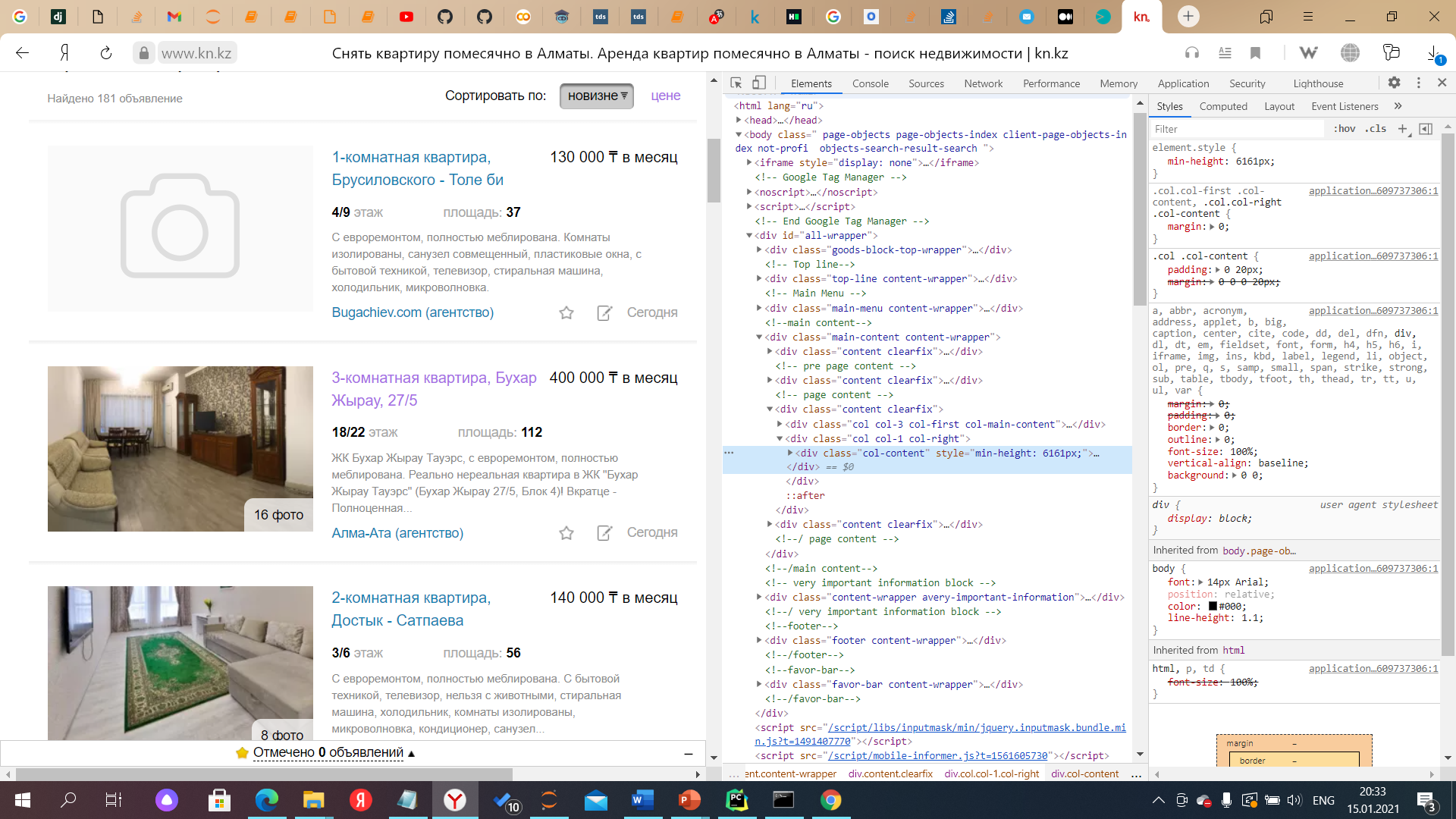
**Introduction**

Renting or buying a new home, whether you are a university student or a middle-class family, is always a complex process that often seems impulsive. As a student, I faced the problem of finding a decent rental housing. Thanks to internet sites nowadays we can search for an apartment on our own, rental services provide data through their websites.

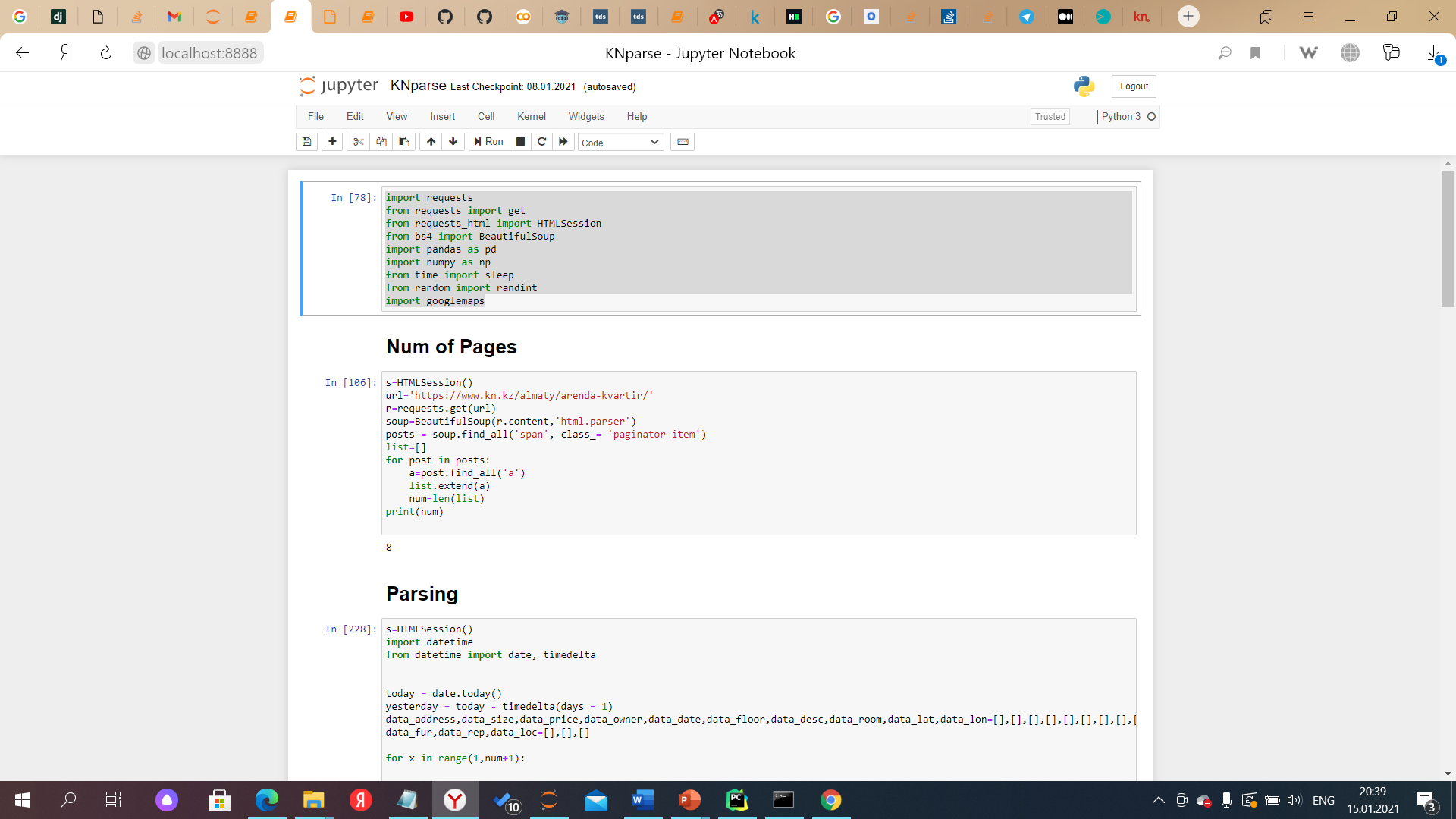
In this project I’m going make algorithm that will predict price of rental places in Almaty based on other information like the number of rooms, distance from center of Almaty, area, etc.

**1)Scrapping website and DataCleaning**

As a data source for the analysis, I used the Kazakh real estate website kn.kz. For scrapping I used BeautifulSoup library to pull data from the html code of the page

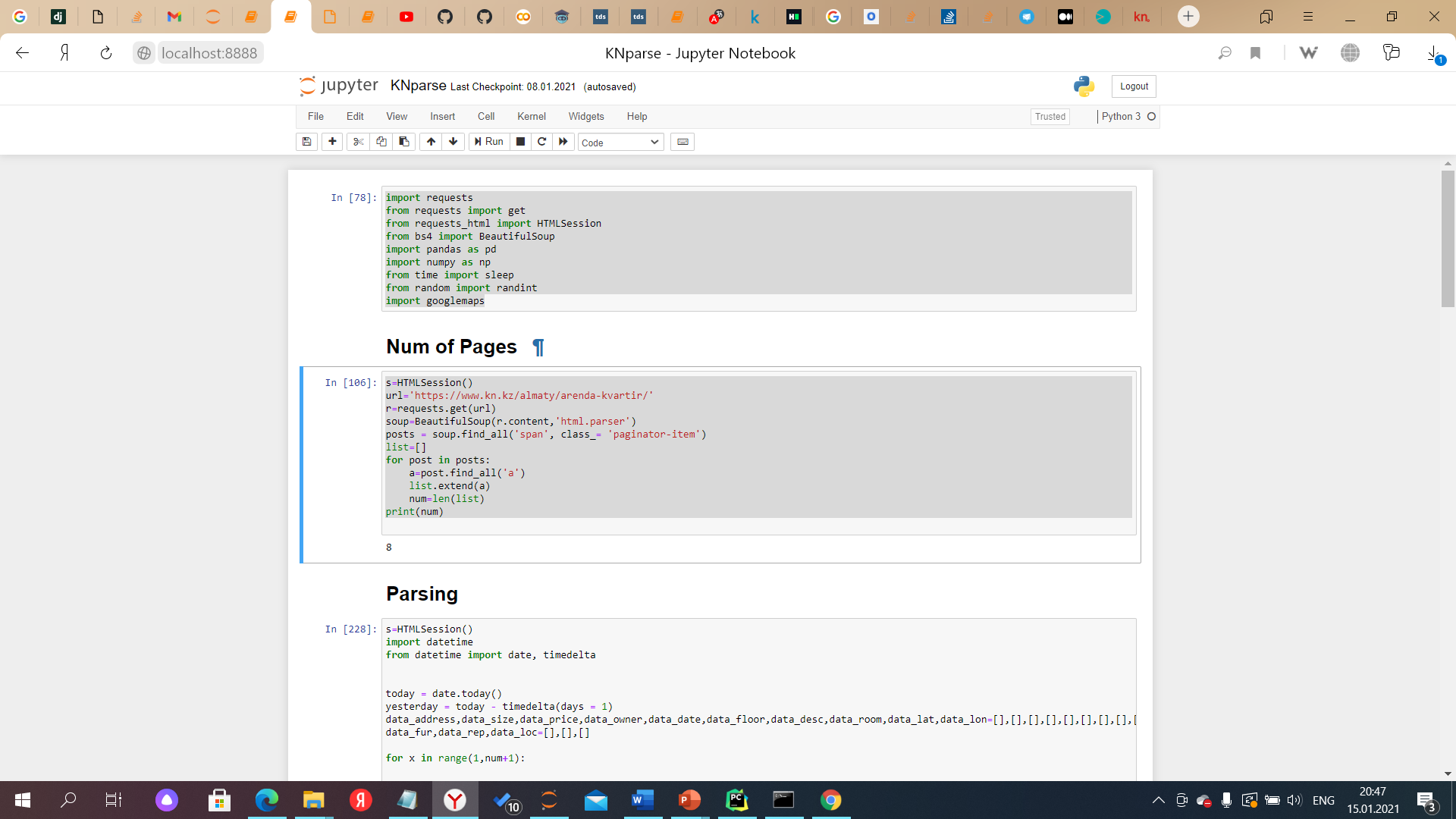


**Site from inside**

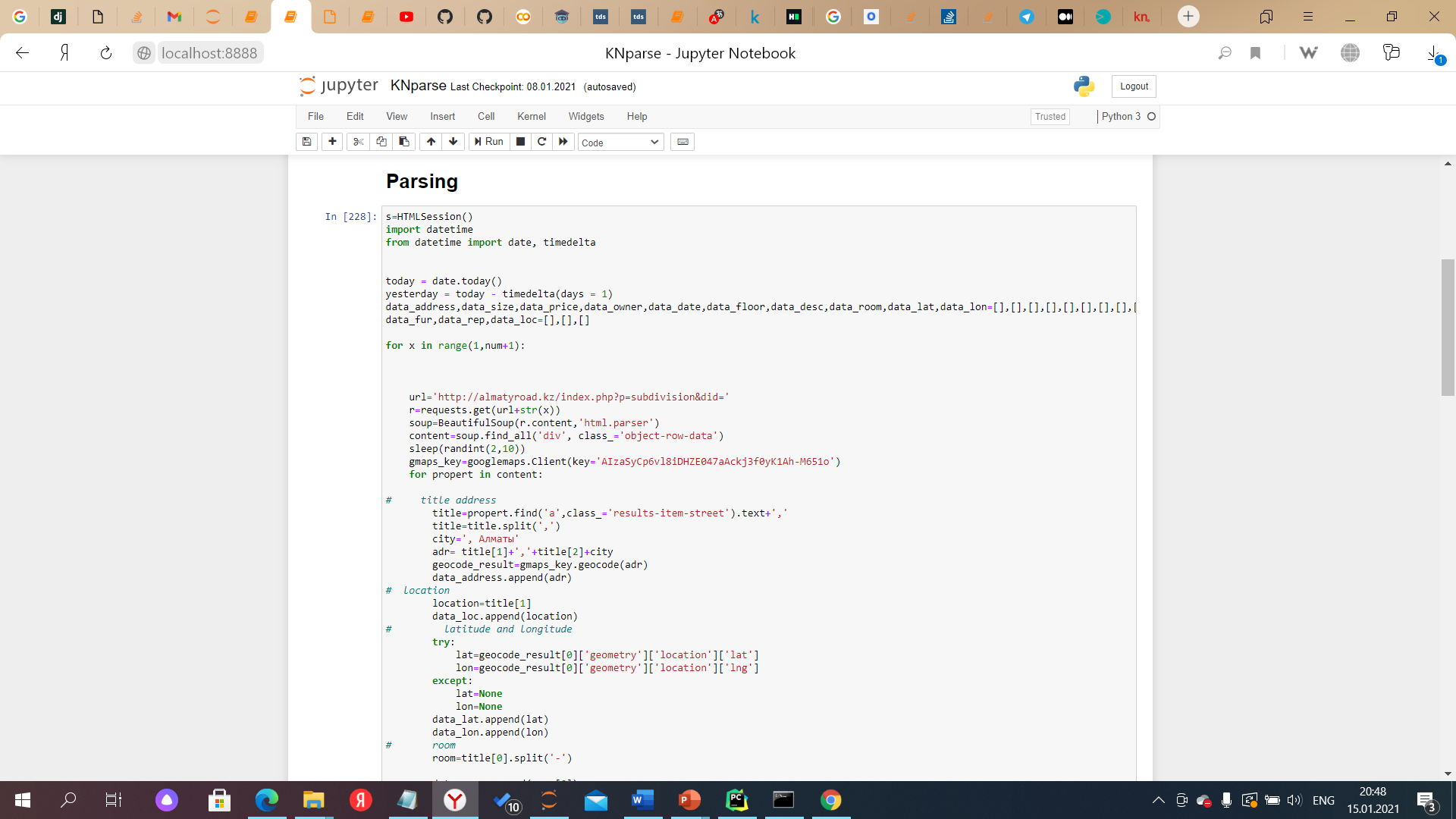


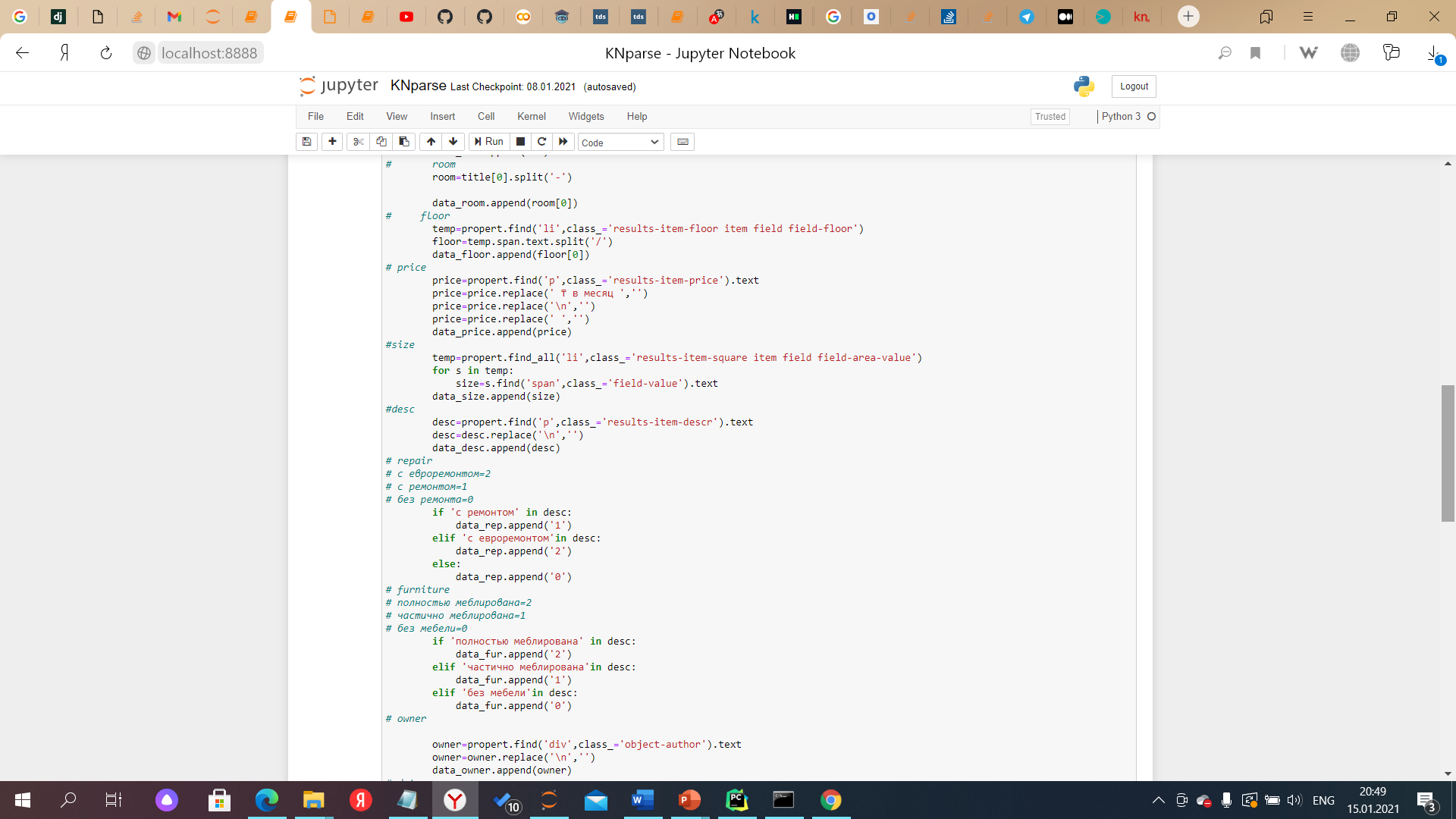
**Importing future necessary libraries**

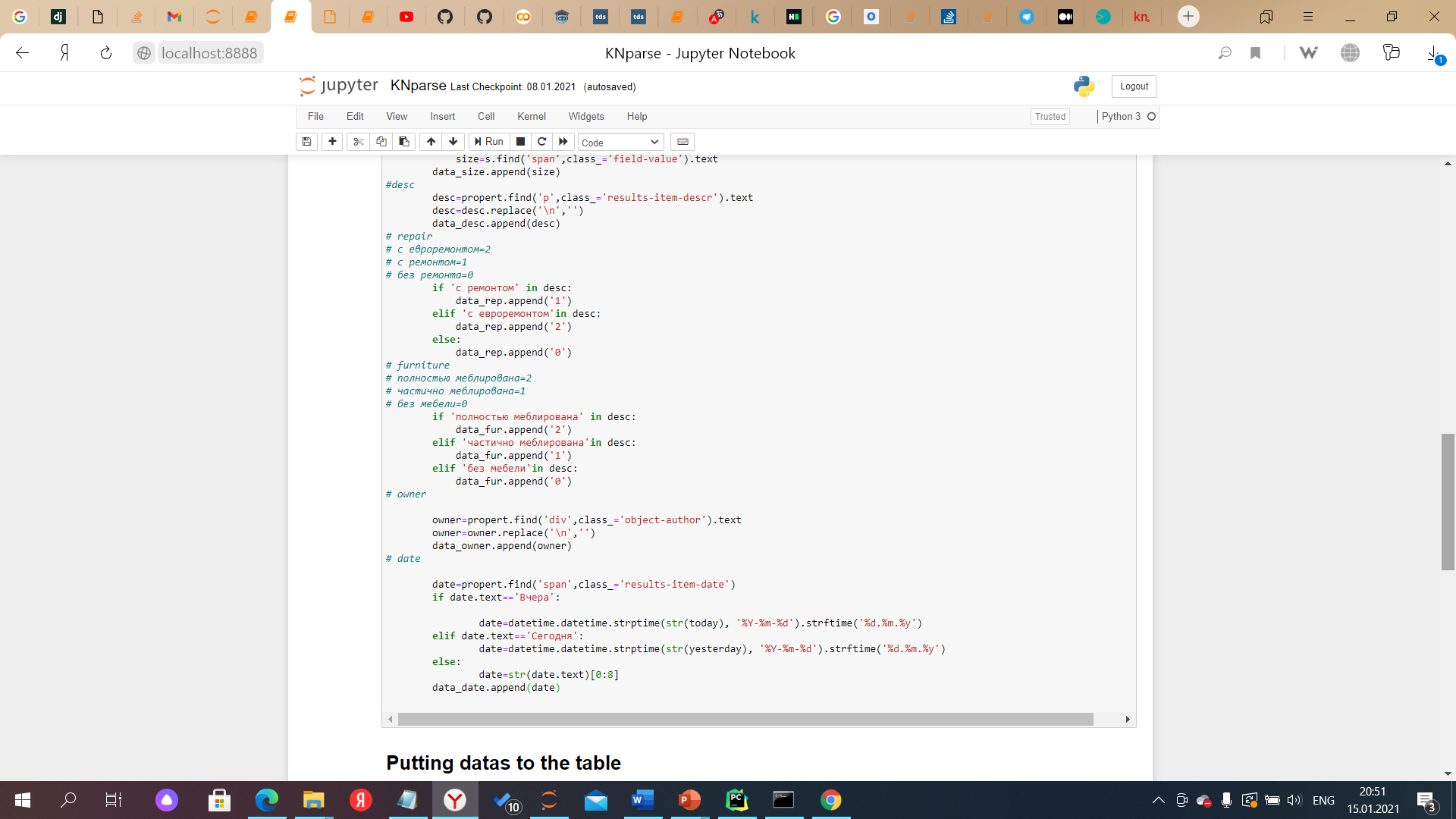
Then, real coding is starting.



**Here I count number of pages**







**full parsing of data from the site kn.kz .**

Data on the date, agency, description of the apartment, floor , and area were directly taken from the site as descriptive information. In the very process of parsing I cleaned up redistributed the data to the desired columns, saved the text from unnecessary garbage such as ("[],(),$,/,\").The titles were divided into the address and the number of rooms. Then I found the latitude and longitude of the addresses with which I will search for the distance from the selected center(Abay-Auezov intersection) . From the description line, I subtracted data on the availability and quality of repairs, as well as on the availability of furniture. In these lines I replaced words with numbers .

**Furniture:**

“полностью меблированa”=2;

“ частично меблирована”=1;

“без мебели”=0;

**Repair**

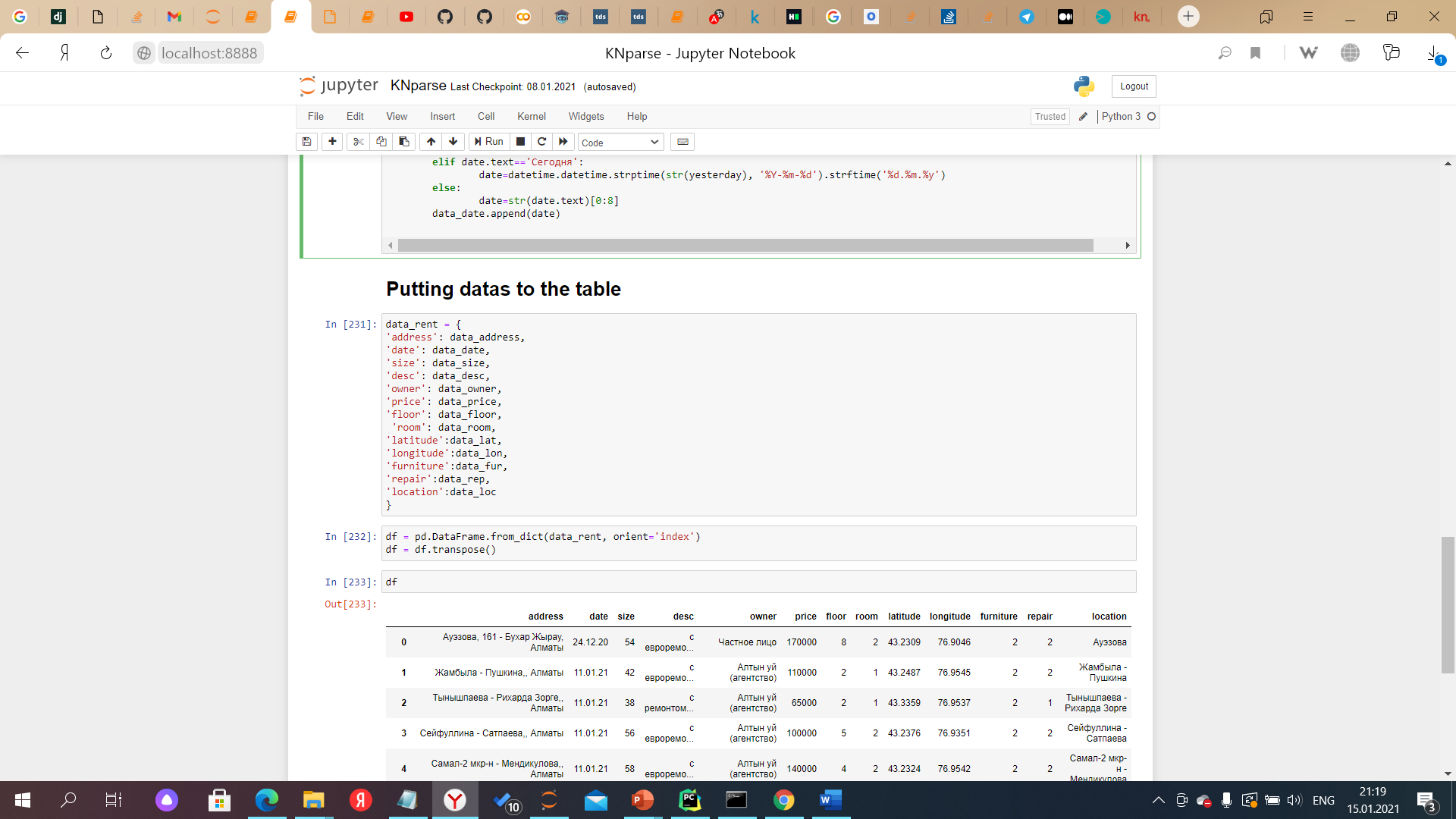
“с евроремонтом”=2;

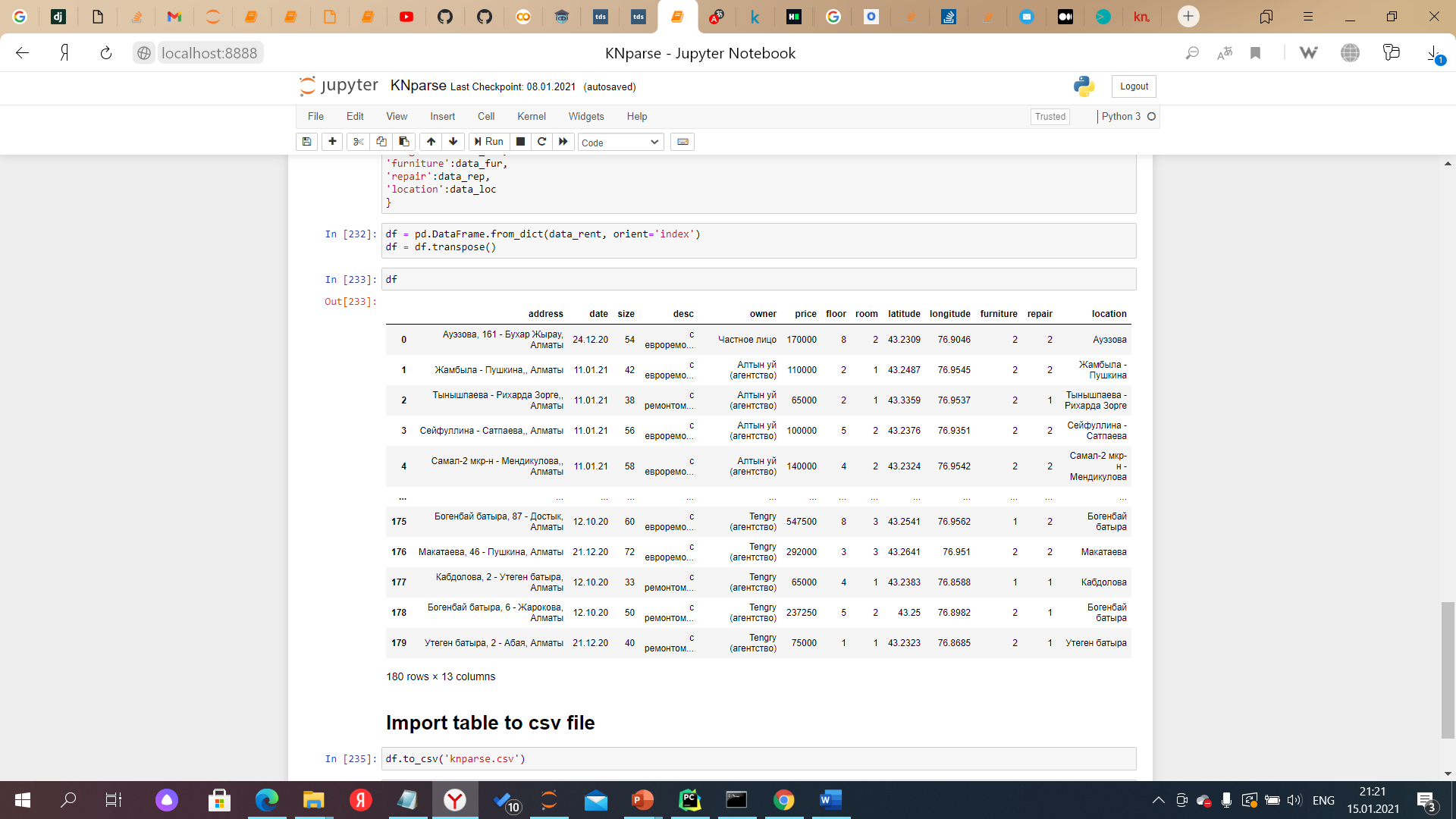
“ с ремонтом”=1;

“без ремонта”=0;

The date column has also been changed. The words "" and" today " were replaced with the corresponding dates.

Then all the data taken was entered in the dictionary data\_rent.

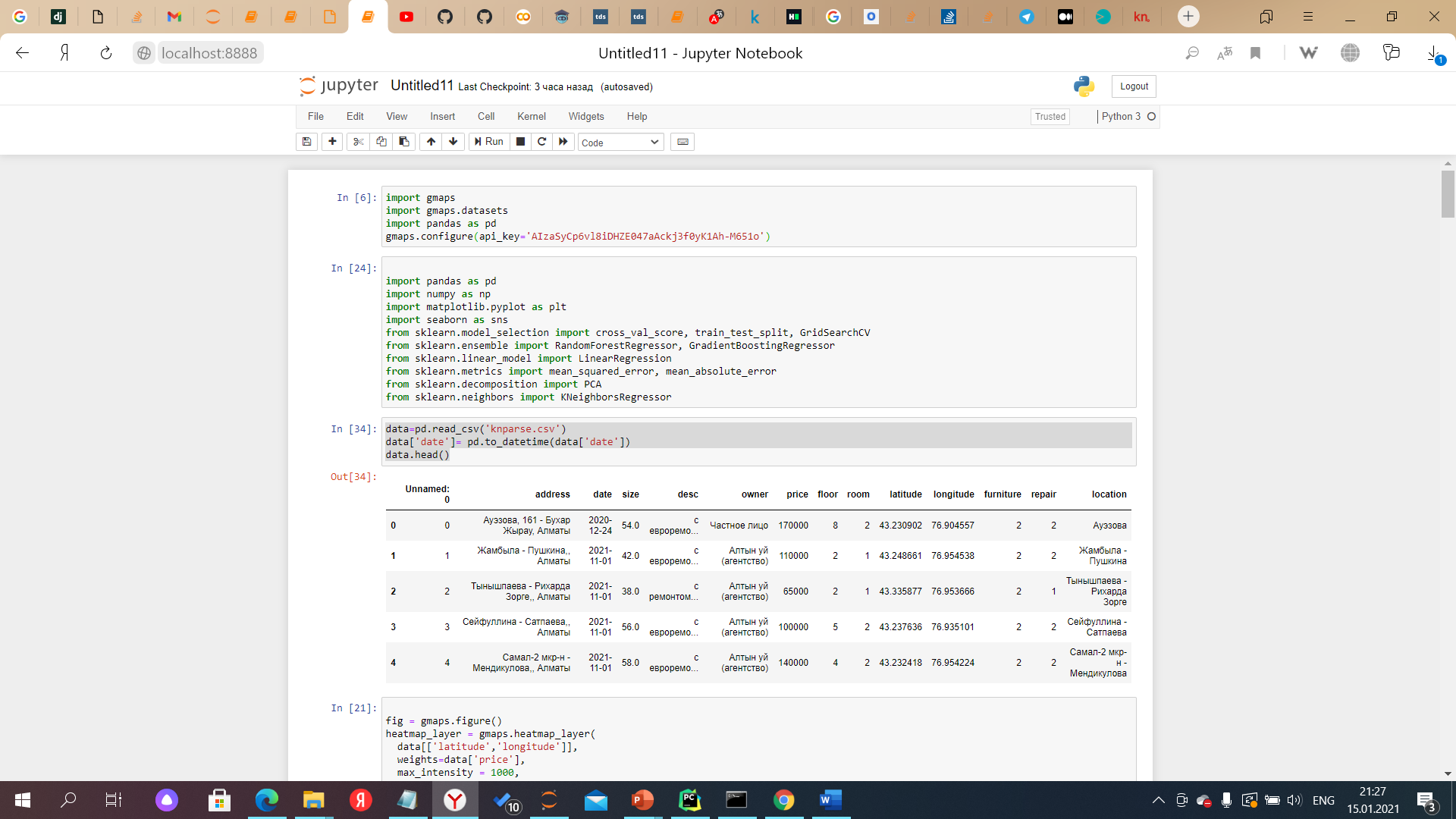


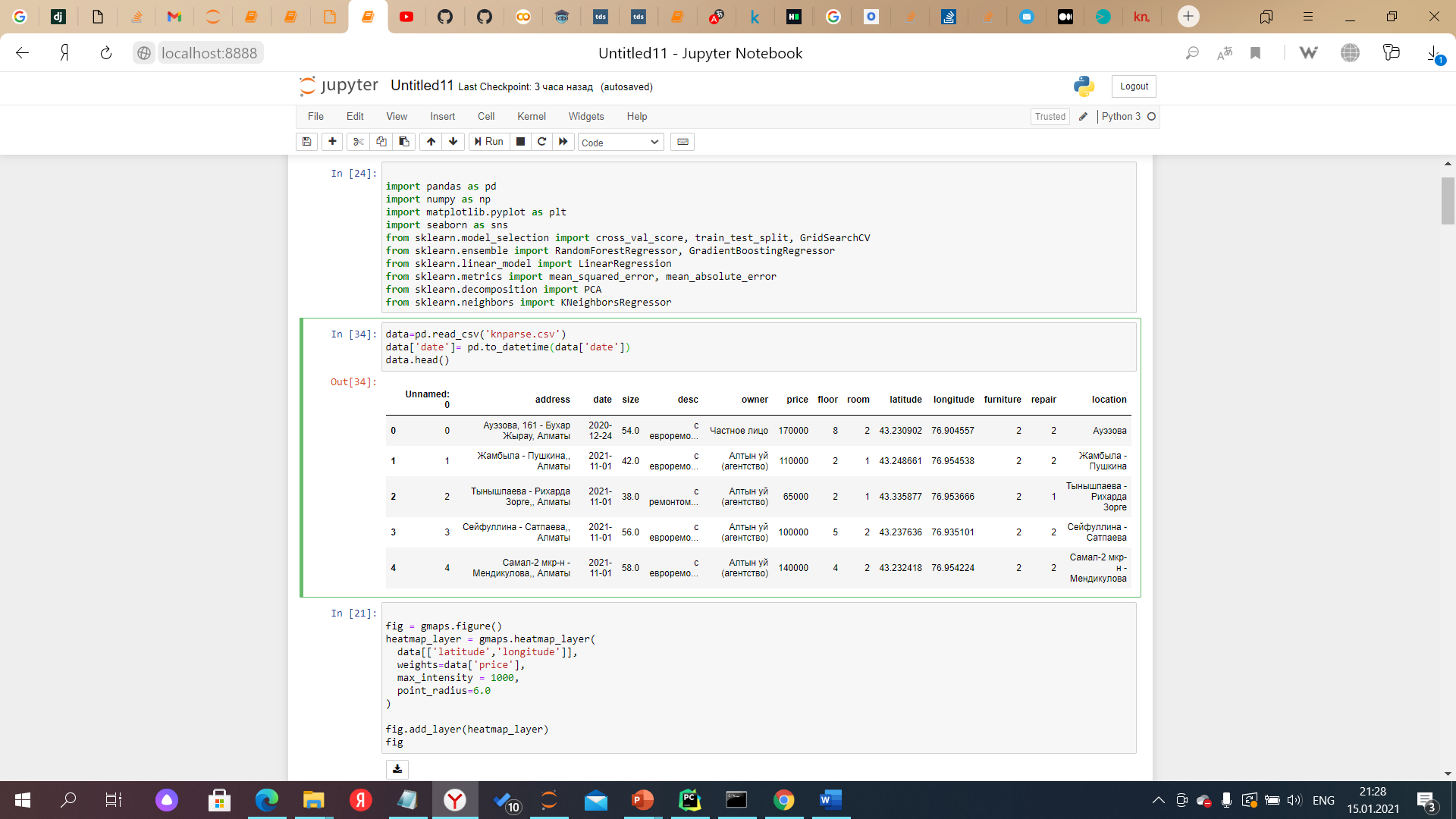


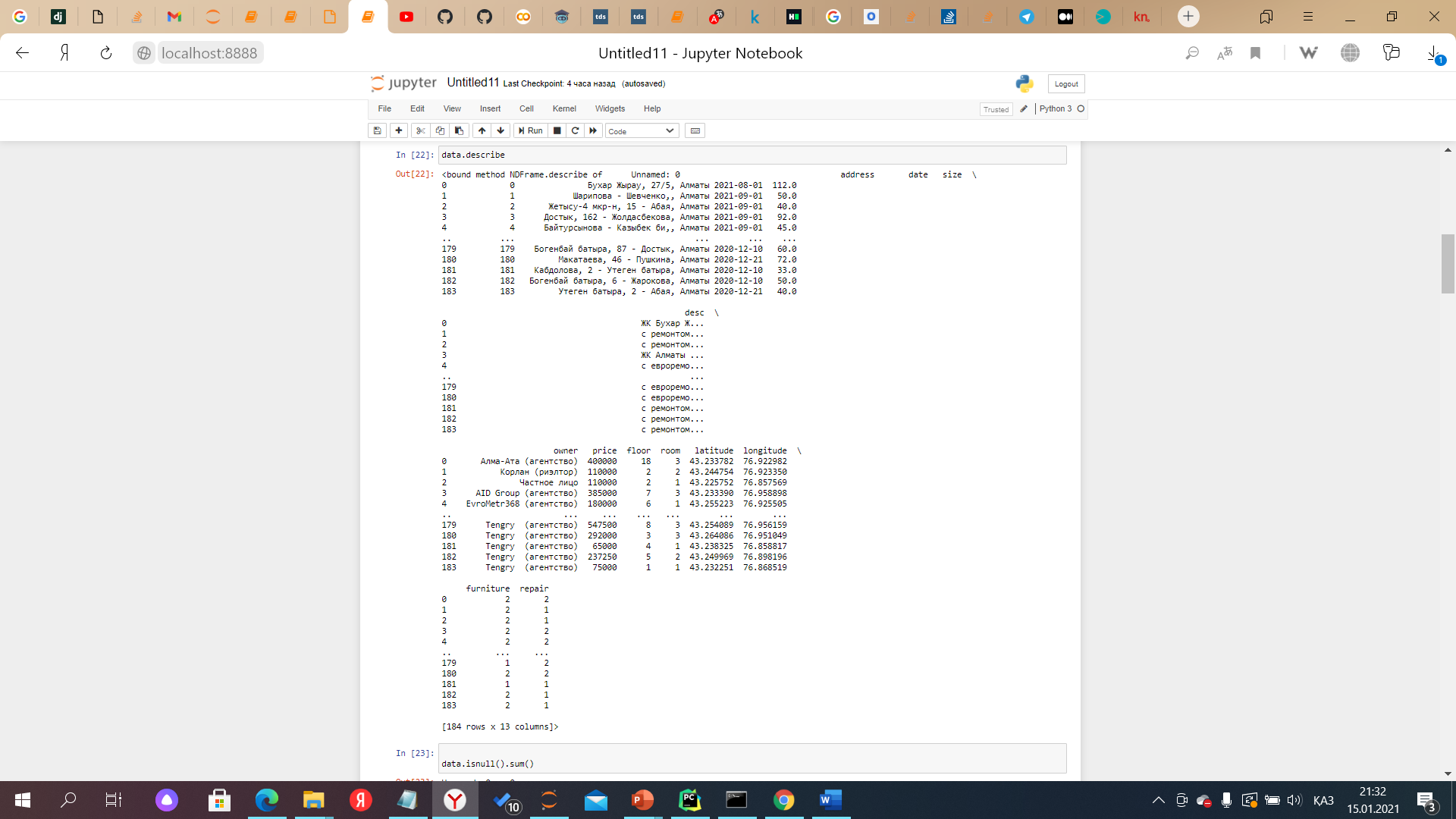
**Table of taken data**

Then a csv was compiled from that dictionary.

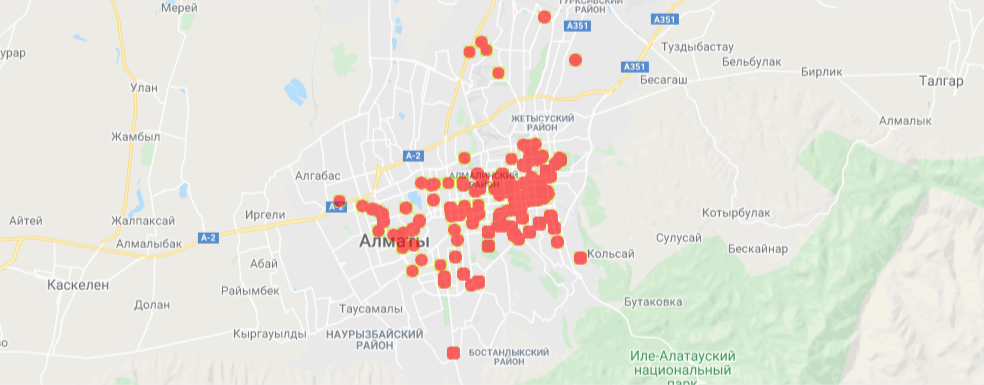
**Another python page**





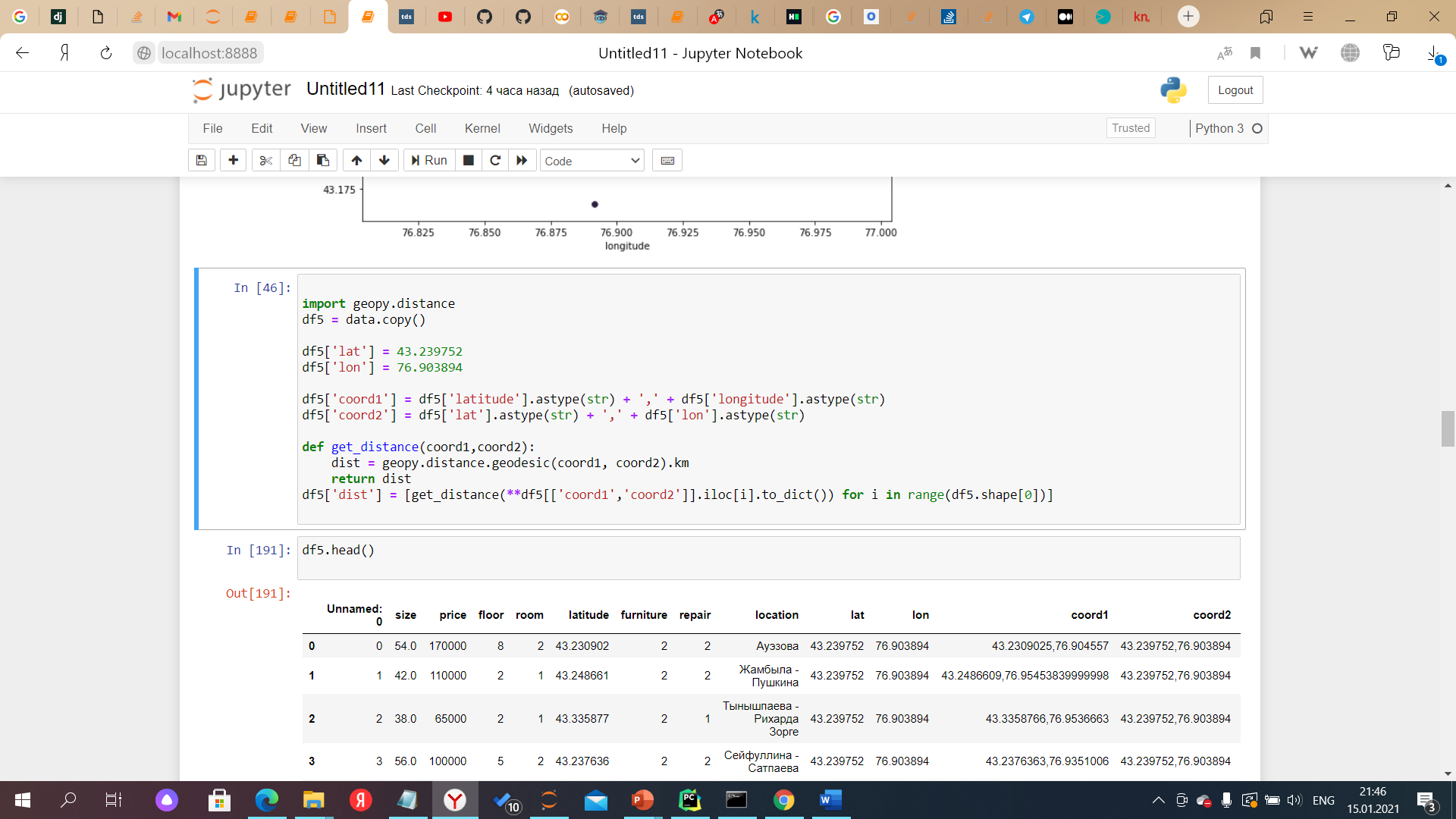




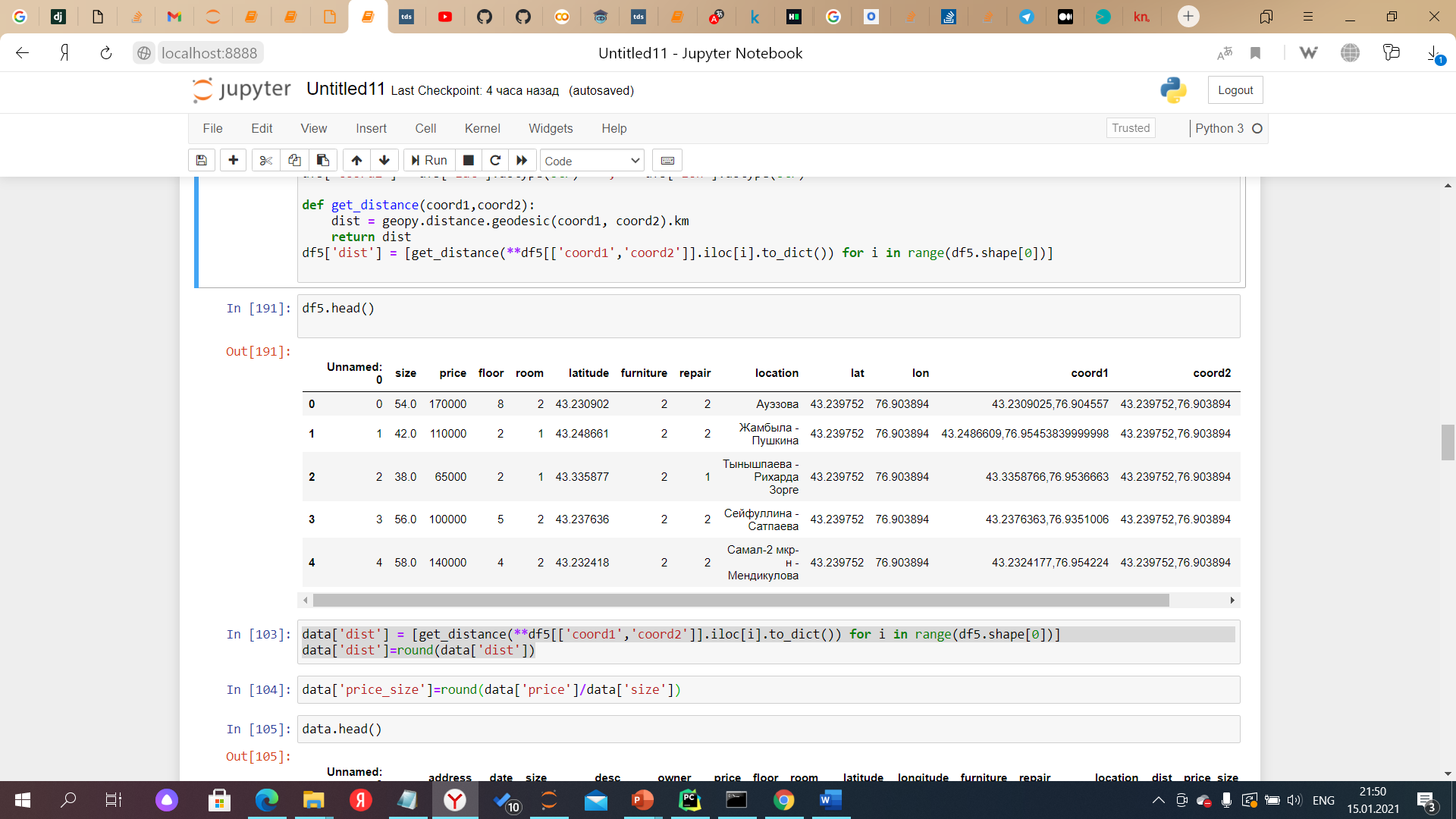
**Location of apartments on the map of Almaty**

**Distance**

The Abay-Auezov intersection was chosen as the center(latitude: 43.239752,longitude: 76.903894)

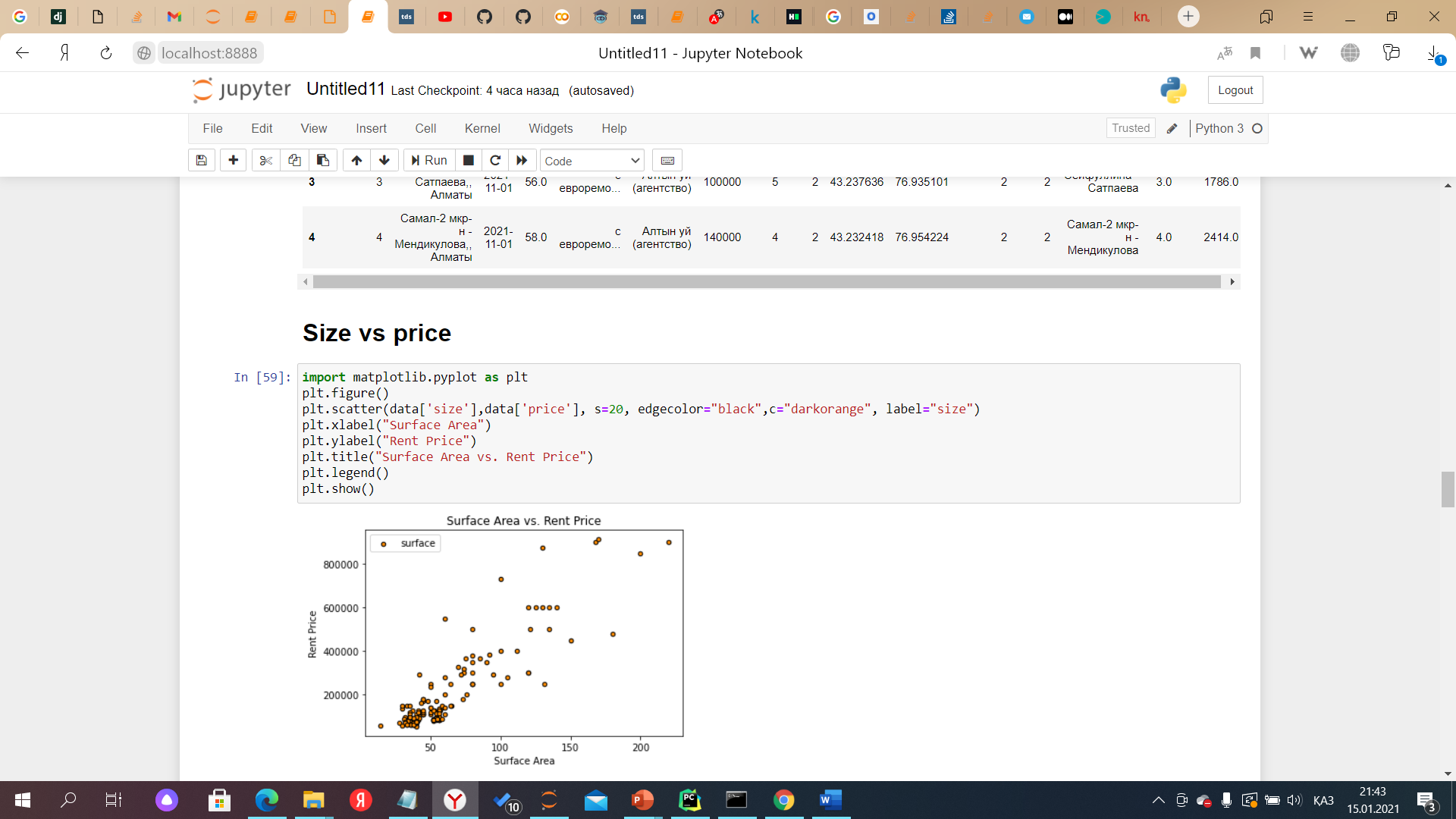


I create data copy for not loosing or messing the original one.



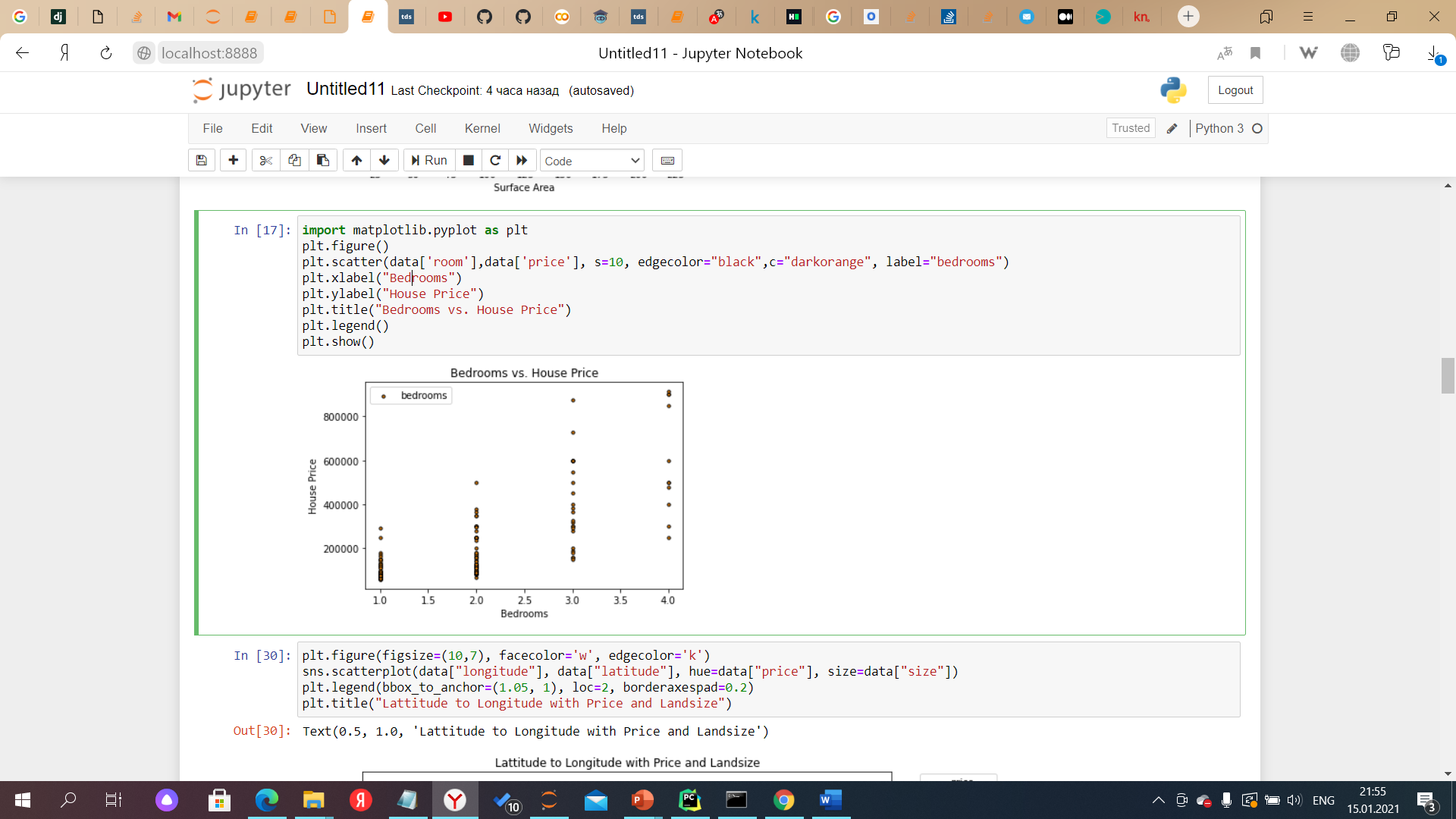
**Visualization**

We should also take a look on how the variables in our data frame relate to house prices. This is the code to check the relationship between house prices and surface area:



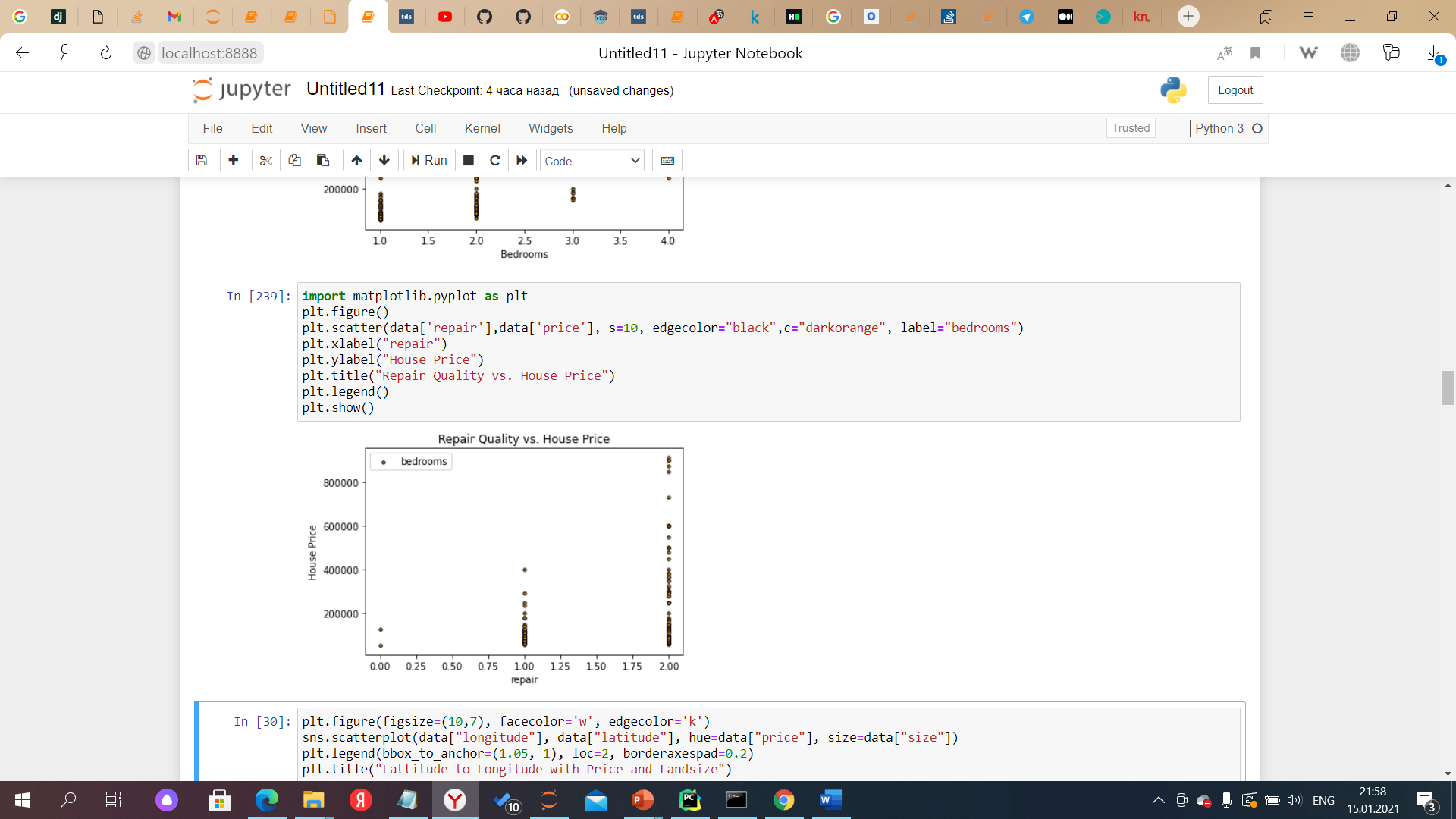
Here we see that among the apartments most common are apartments with an average size of 50 sq m and a price not exceeding 200 thousand.

**Room vs Price**

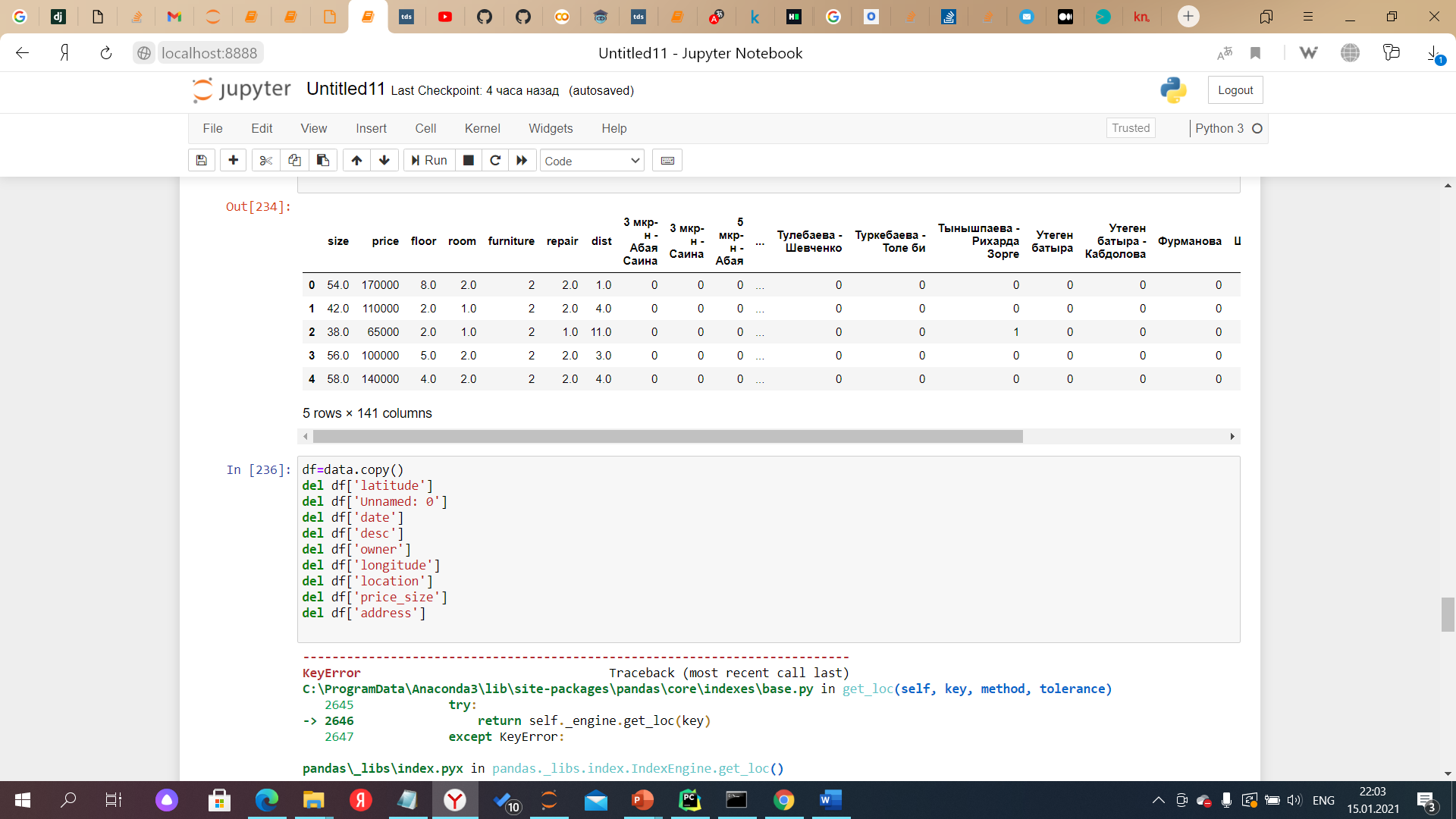


Most of 1-2 rooms apartments cost around 200 000 tg.

**Quality repair vs Price**



It was not surprising that with higher repait there higher prices.



Here we get rif of unnecessari data . And starting implementation of algorithm. For predictinf price we will use linear regression.

**Linear Regression :-**

In easy words a model in statistics which helps us predicts the future based upon past relationship of variables.

**Splitting data into train and test**

