**PROJECT-1**

**EXPLORATORY ANALYSIS**

#### DATASET: Supermarket store branches sales analysis

DATASET- In the dataset, You'll get data of different stores of a supermarket company as per their store IDs which for ease has been converted to positive integers.

1)Store ID: (Index) ID of the particular store.

2)Store Area: Physical Area of the store in yard square.

3)Items Available: Number of different items available in the corresponding store.

4)Daily Customer Count: Number of customers who visited to stores on an average over month.

5)Store Sales: Sales in (US $) that stores made.

PROBLEM:-

Analysing the performances of stores in the past on basis of which will try to rectify defects as well as to leverage the positives.

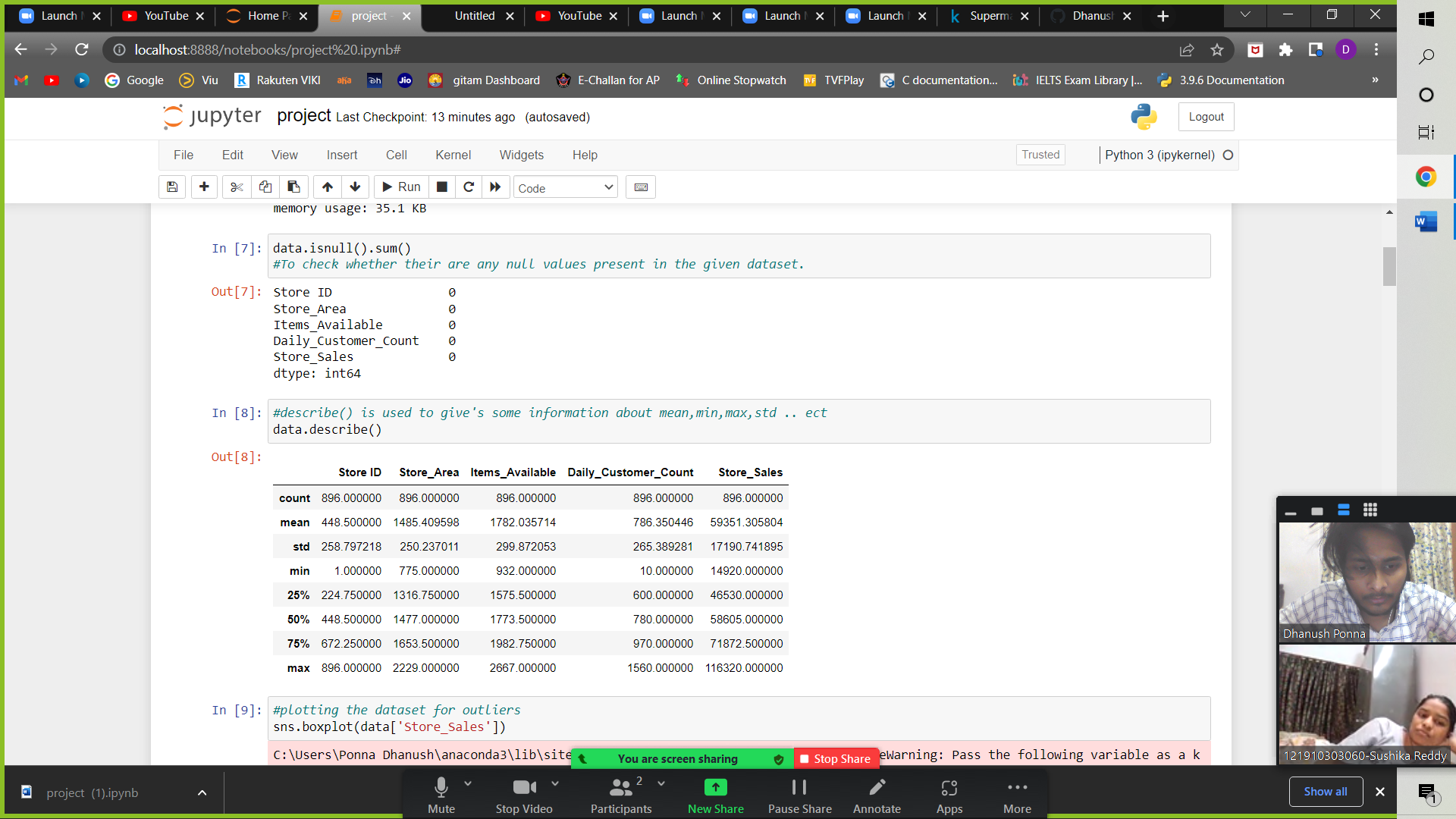
WHY I CHOOSE?

A supermarket is a self-service shop offering a wide variety of food, beverages and household products, organized into sections. This kind of store is larger and has a wider selection than earlier grocery stores, but is smaller and more limited in the range of merchandise than a hypermarket or big-box market. All things considered super market is easy and well categorised shop than some other road-side markets.

Data preparation :

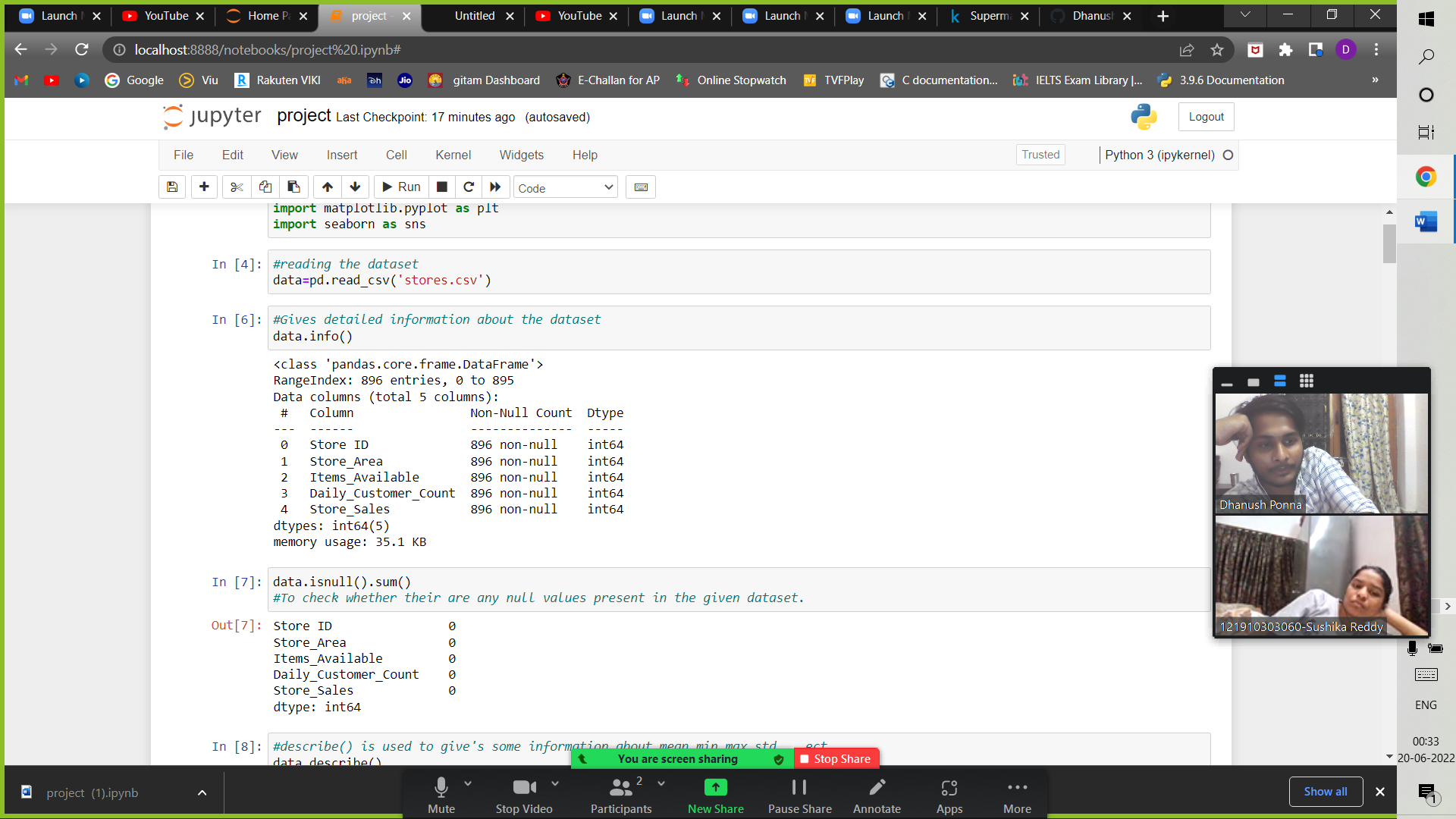
As there are no null values present in my dataset.

There is no need to replace or re-organize any of my attribute.



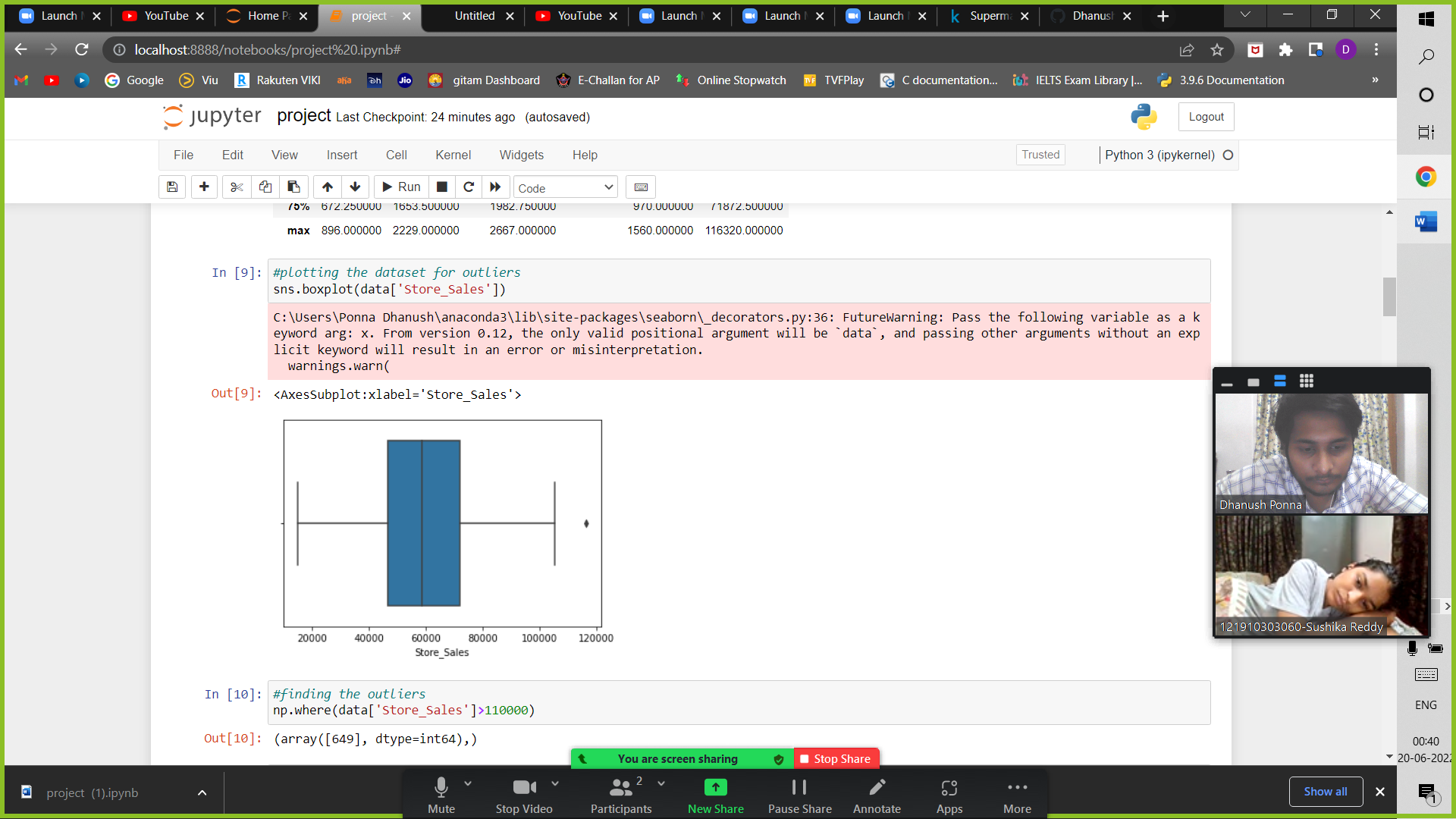
Gathering the information of the attributes based on the data

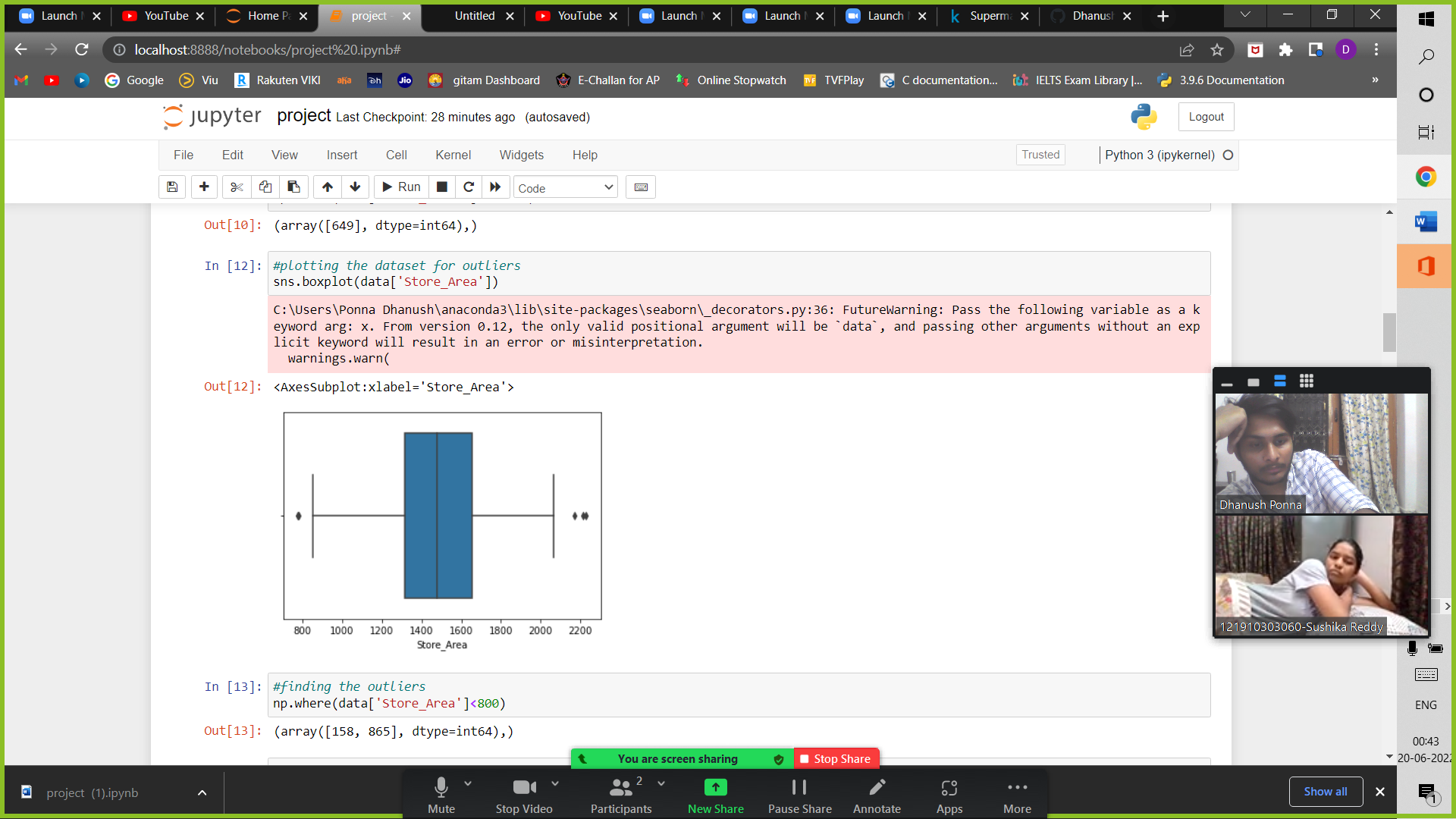
respectively

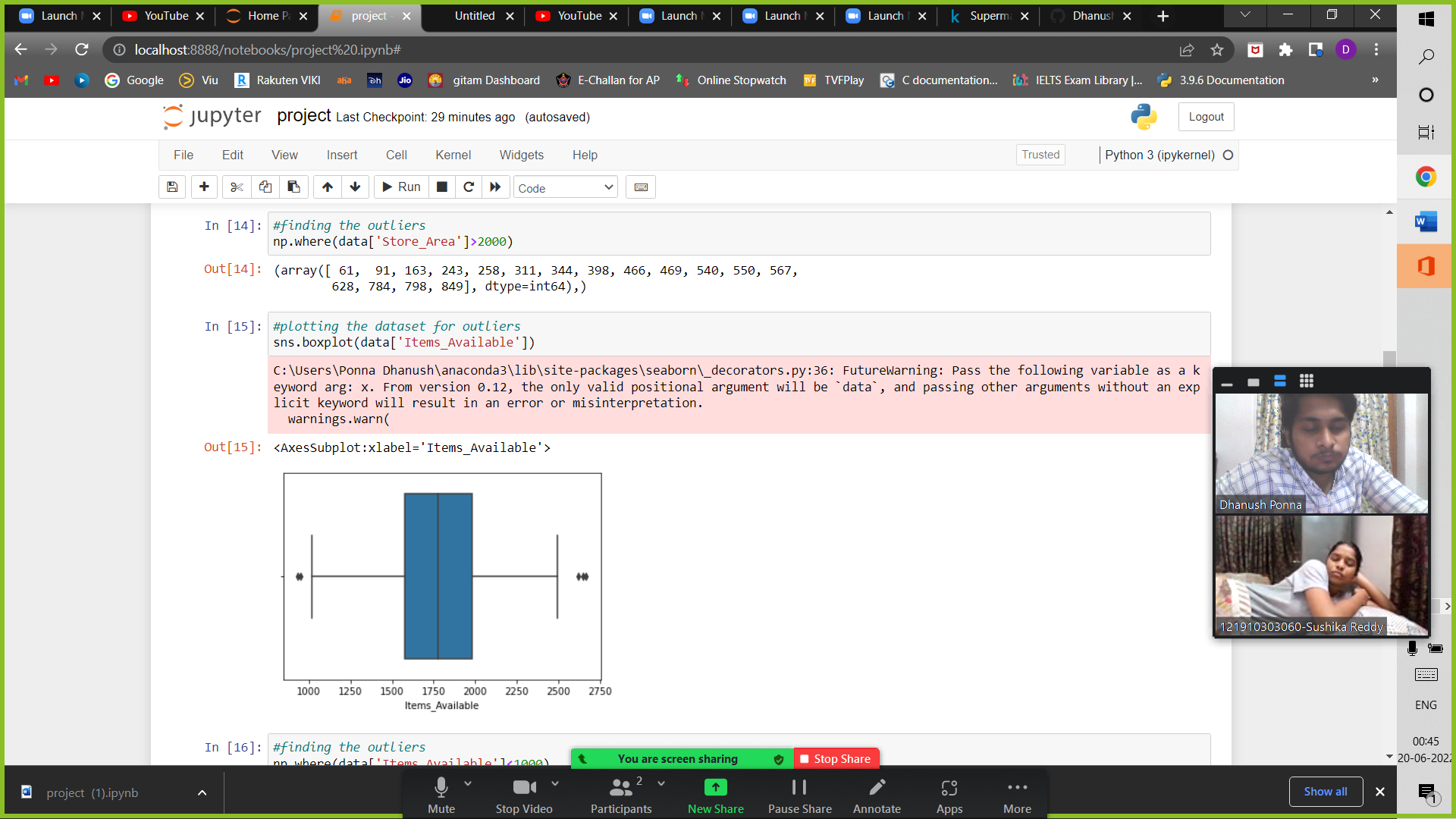


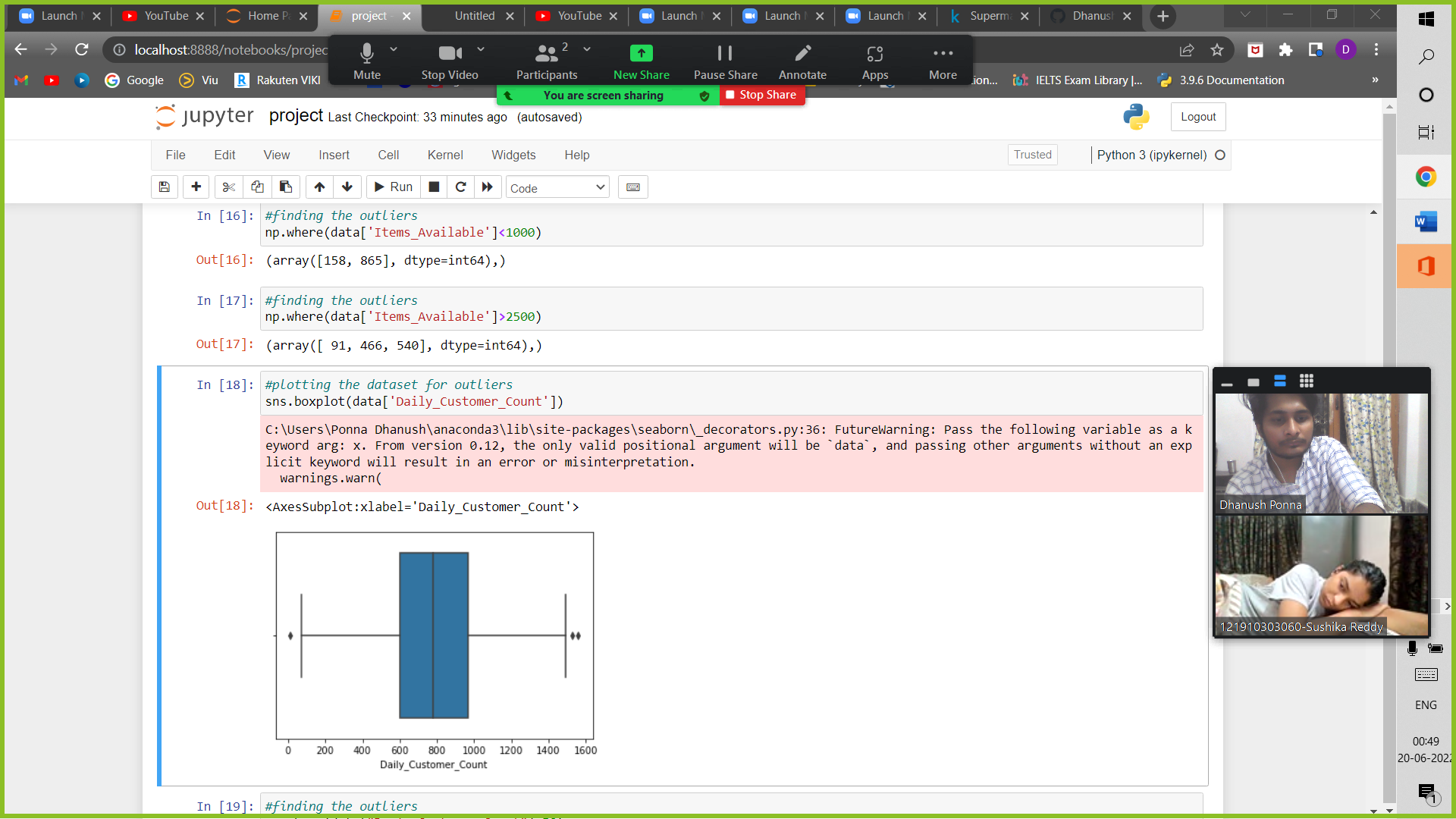
Doing the outlier analysis

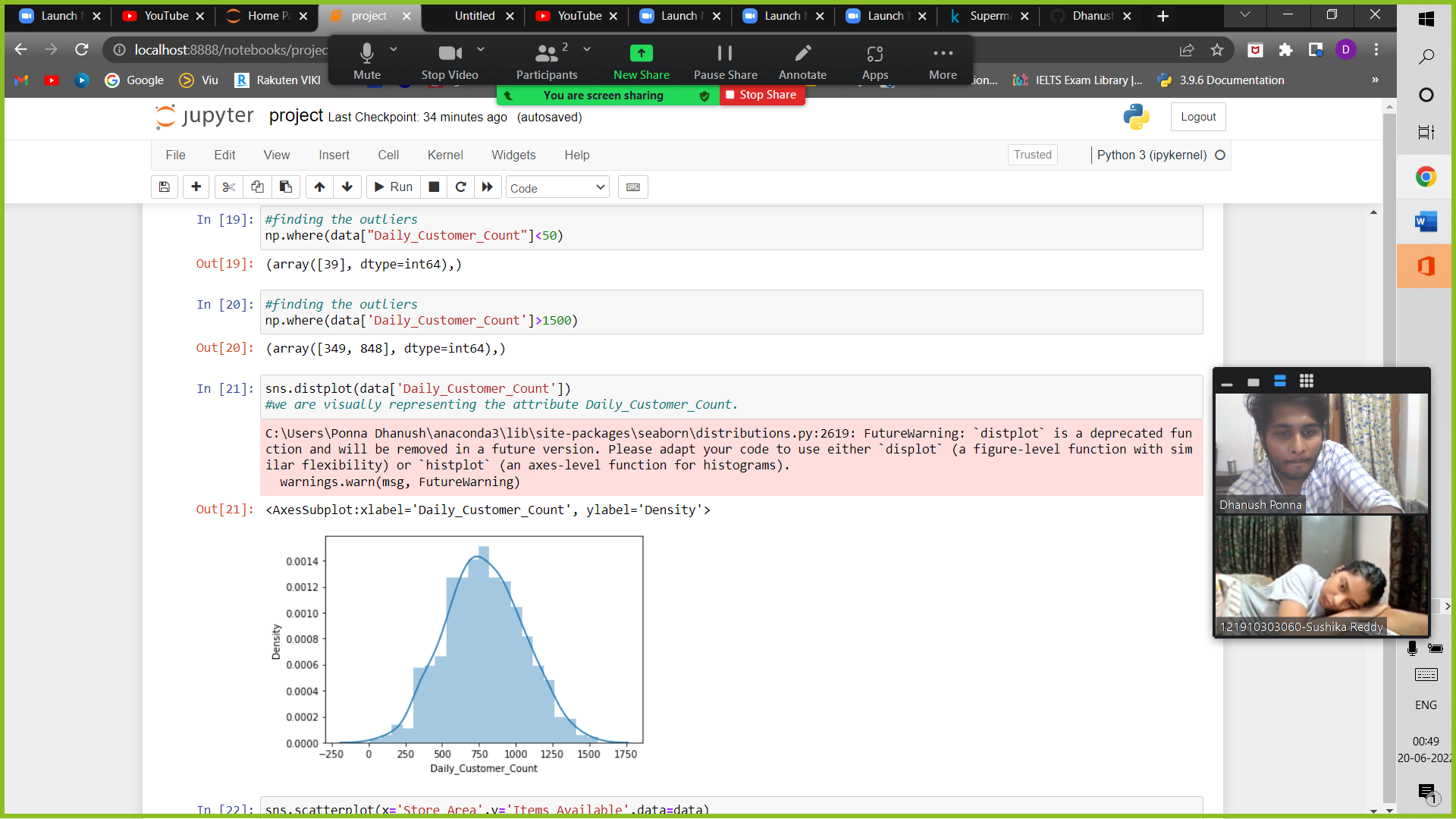
Outlier analysis is carried forward by plotting a boxplot. Inspite of having the outliers it does not make any noticeable changes to the approach. So, we can even ignore the of the attributes as it does not possess any significant change to the analysis.





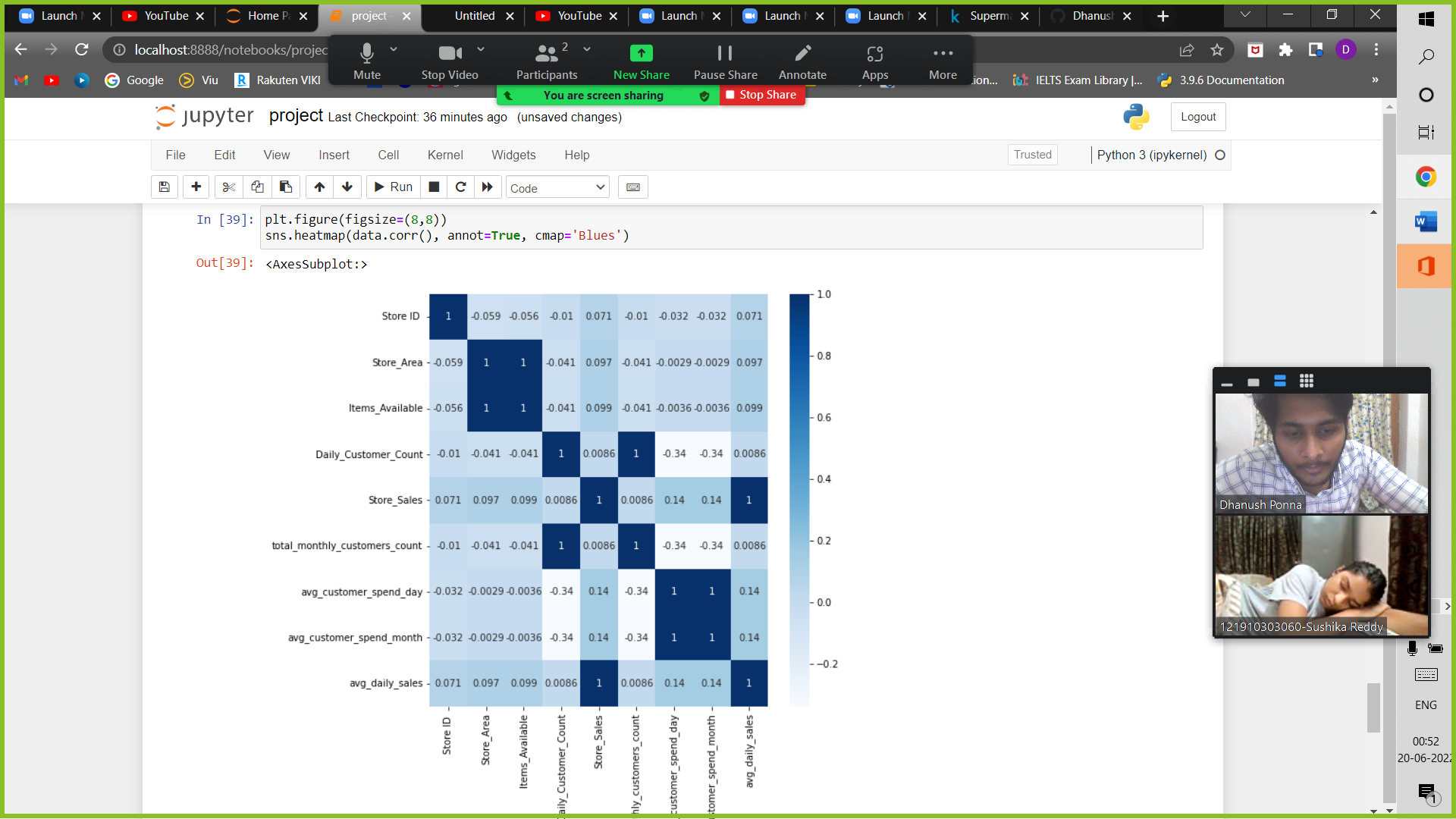




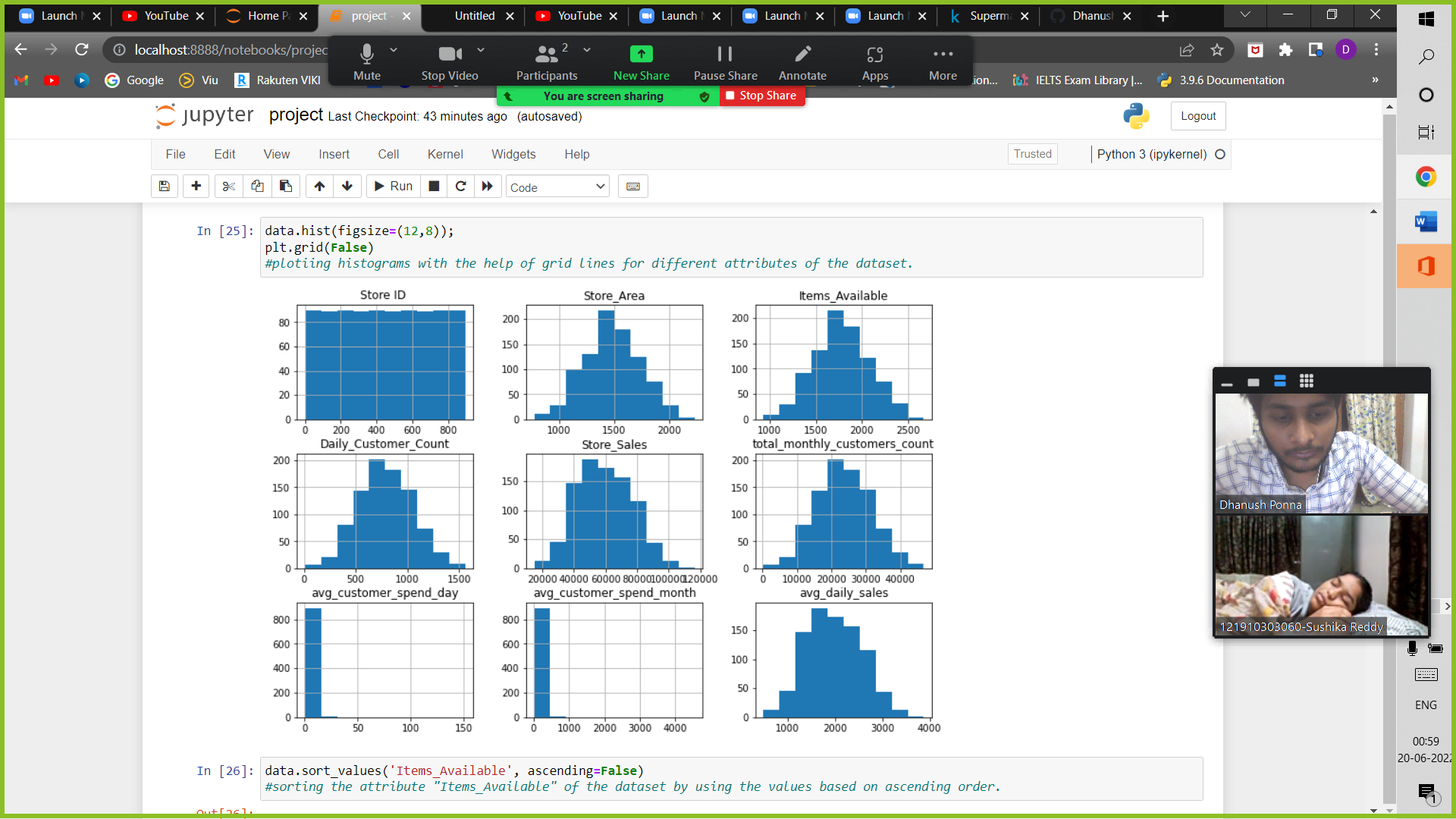


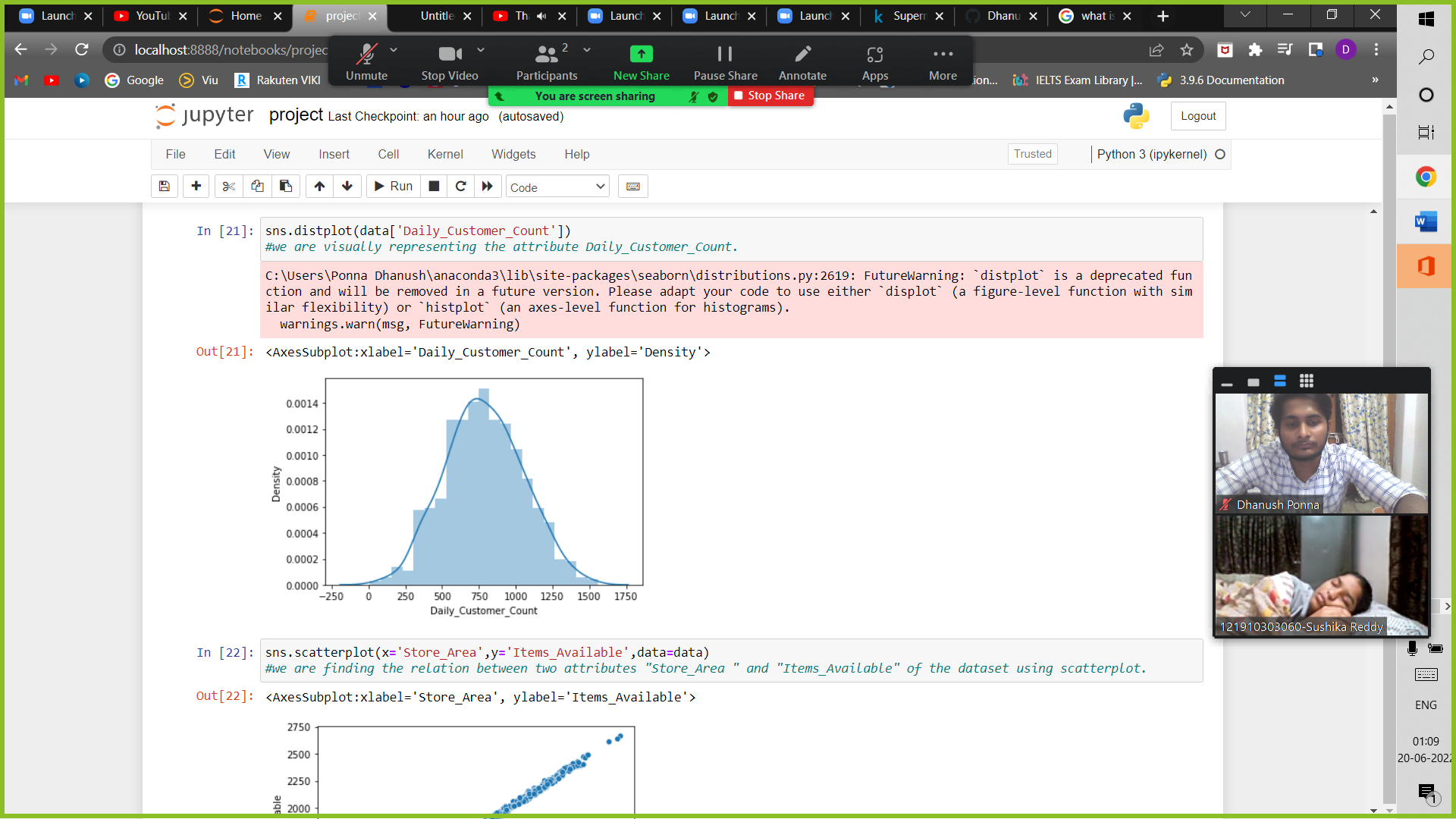
Exploratory analysis:

Plotting a heatmap showing the correlation between the attributes.

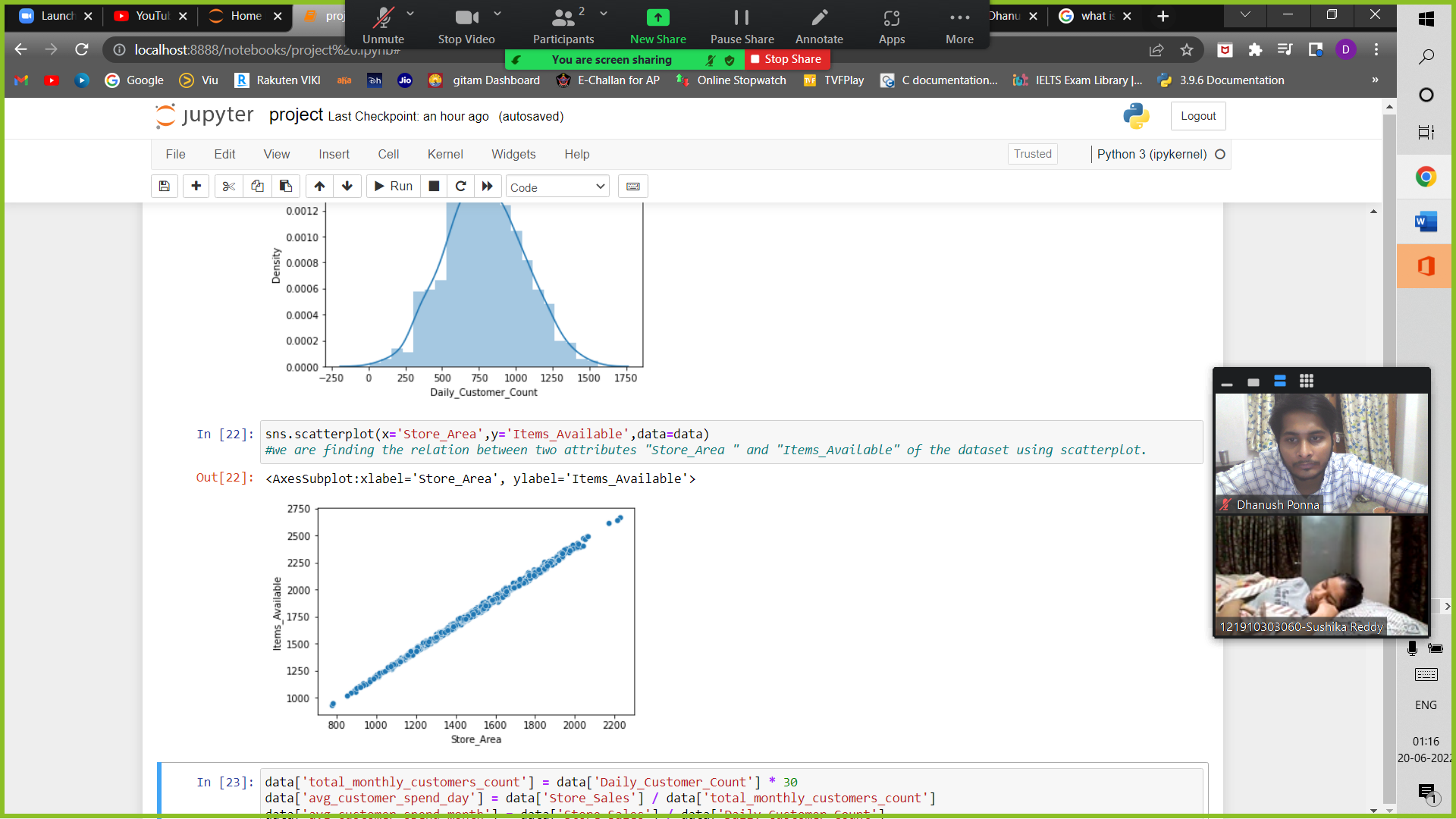


We see high correlation between Items Available and Store Sales, and Items Available with Avg Daily Sales (i.e. it’s almost about 0.099)

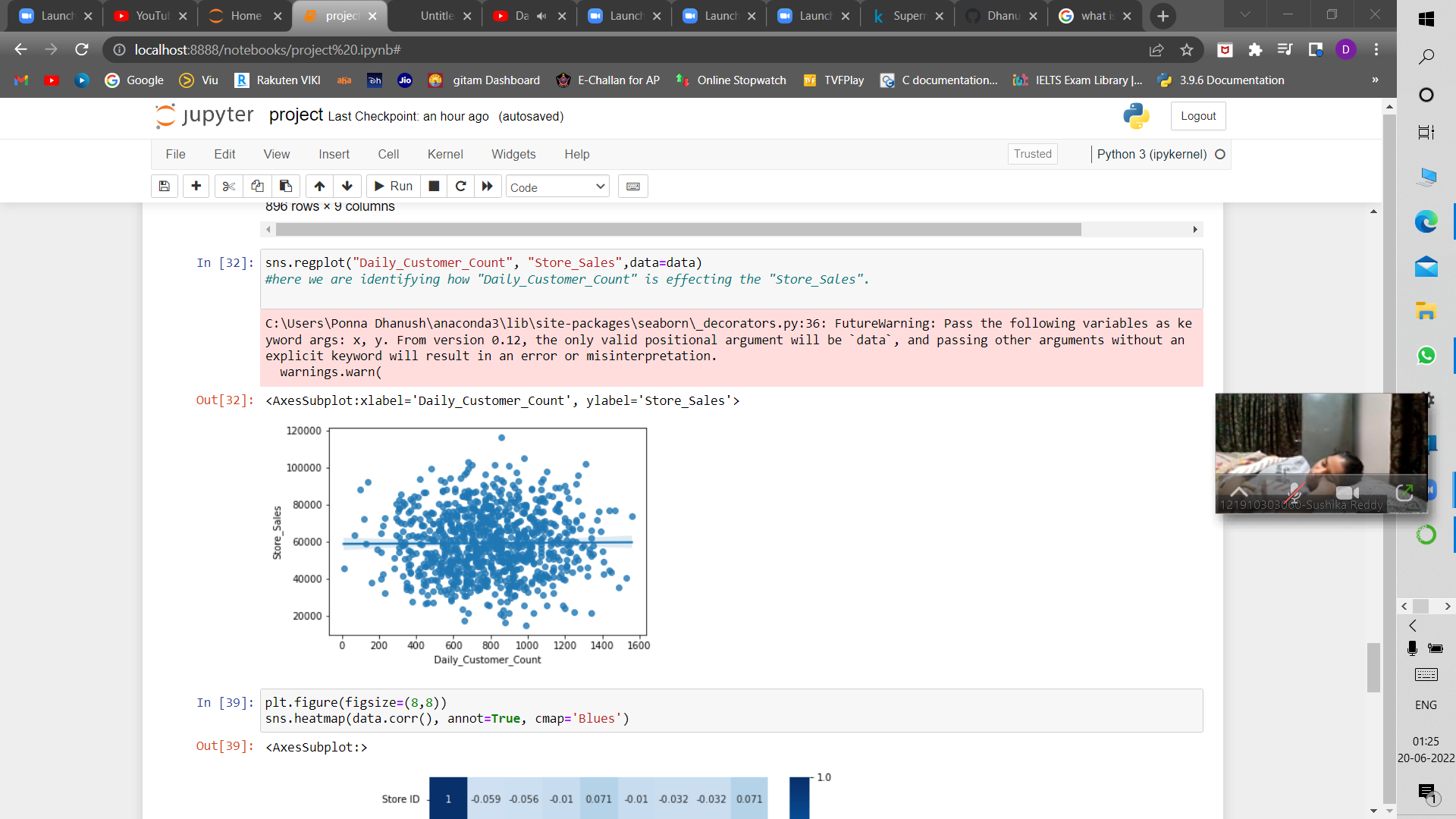
* Plotting and visualizing the data in the form of a histogram.
* We get to observe that daily customer count is mostly lies between (600-800) heads.



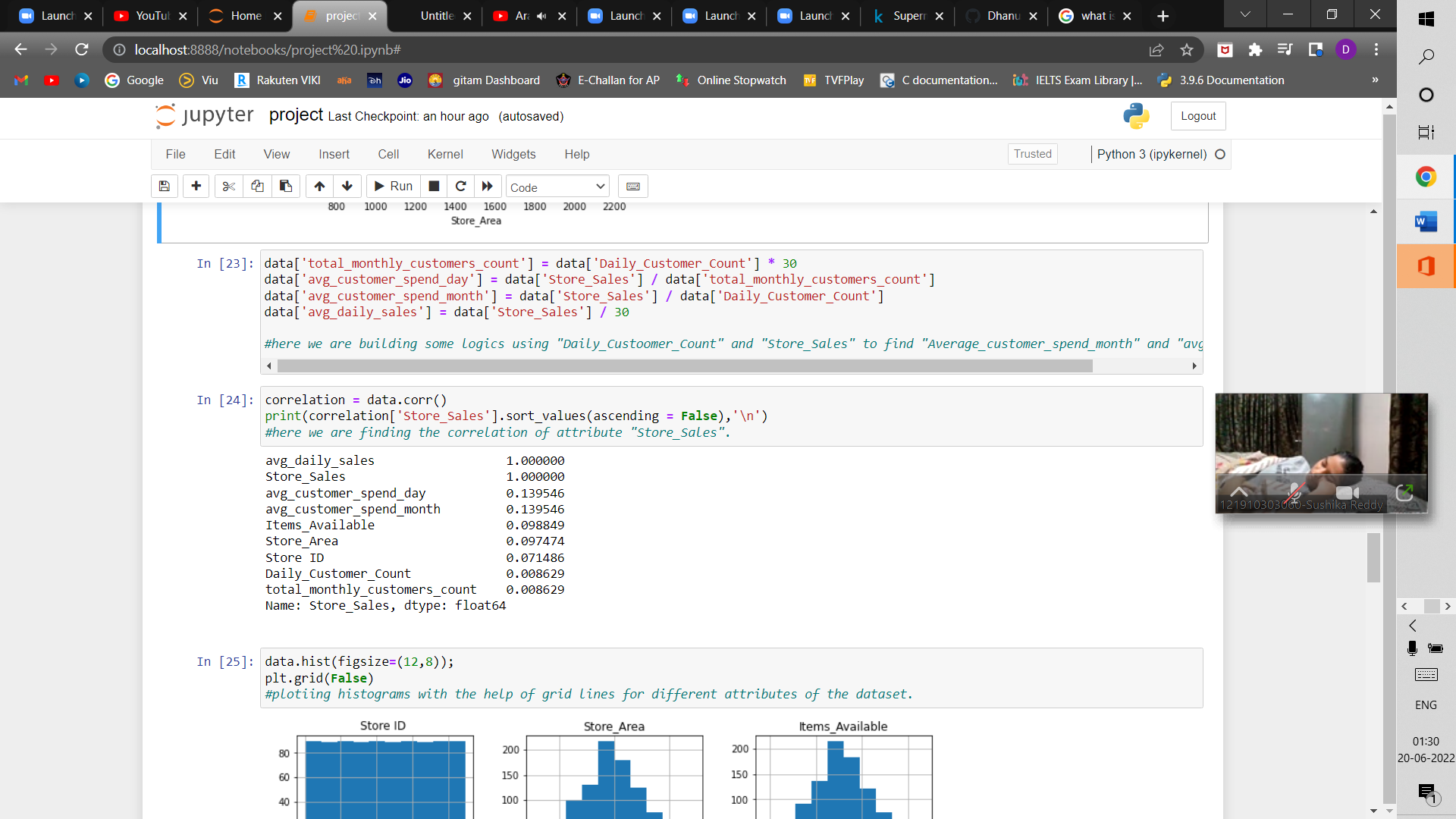
* As we can see that the graph between Store Area and Items Available is steadily increasing, which is clearly evident that store area plays a major role in items available for that particular branch.



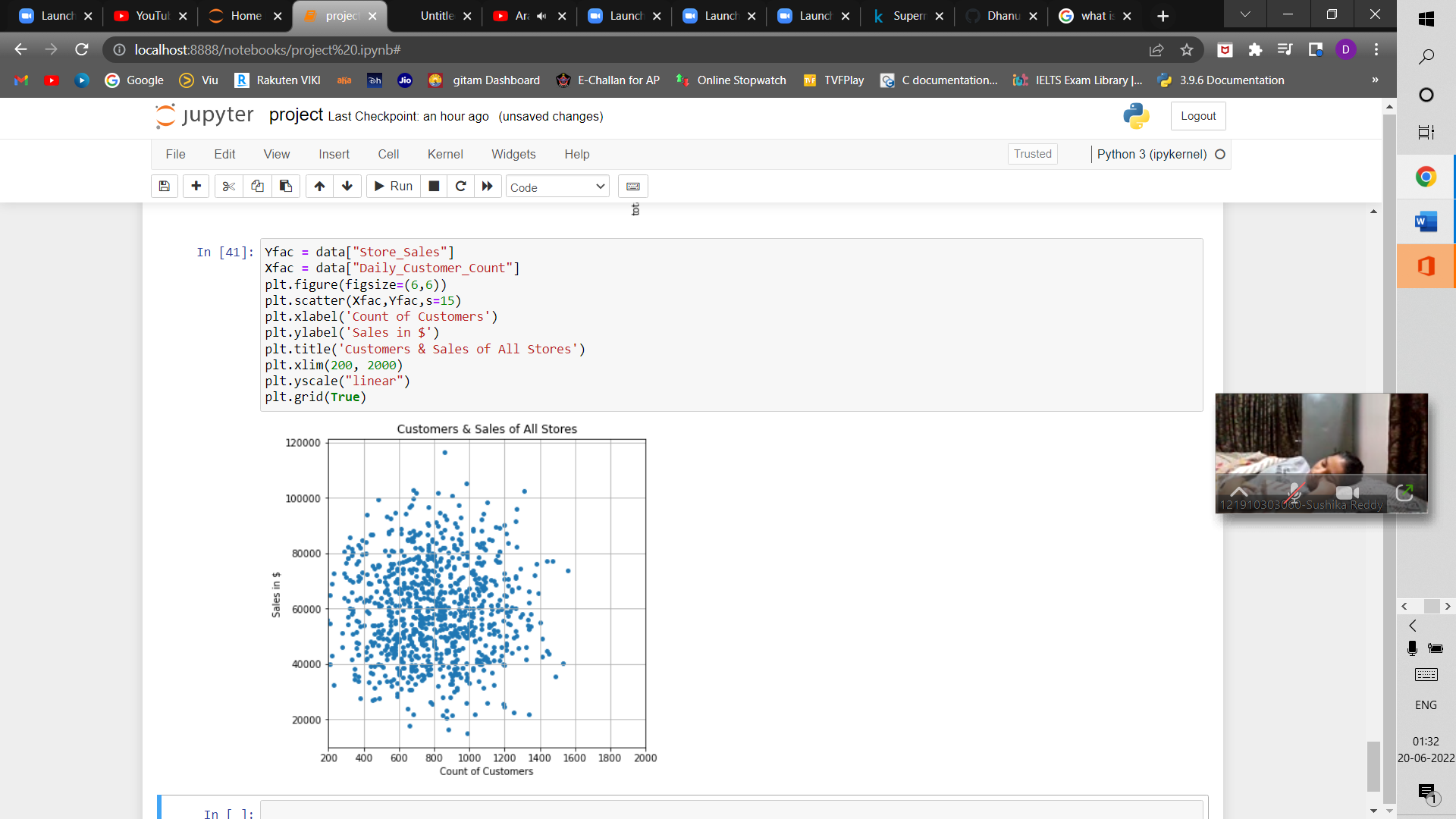
* It is clear that even the daily customer count increases ,store sales remain consistent (i.e. it’s about 60,000$)



* Here we are building some statistical logics using "Daily Customer Count" and "Store Sales" to find "Average customer spend month" and "avg daily sales".



* Finally , I want to conclude that even there is raise in daily customer count the total store sales are consistent .



At last, we can draw this analysis by saying that most of the customers are just going to the store and returning with empty hands which leads to increase in daily customer count and consistent store sales which are bought by regular customers.

Advice:-

We should introduce a new innovative idea which leads to high store sales like “it is mandatory for every customer to buy at least one product from store.