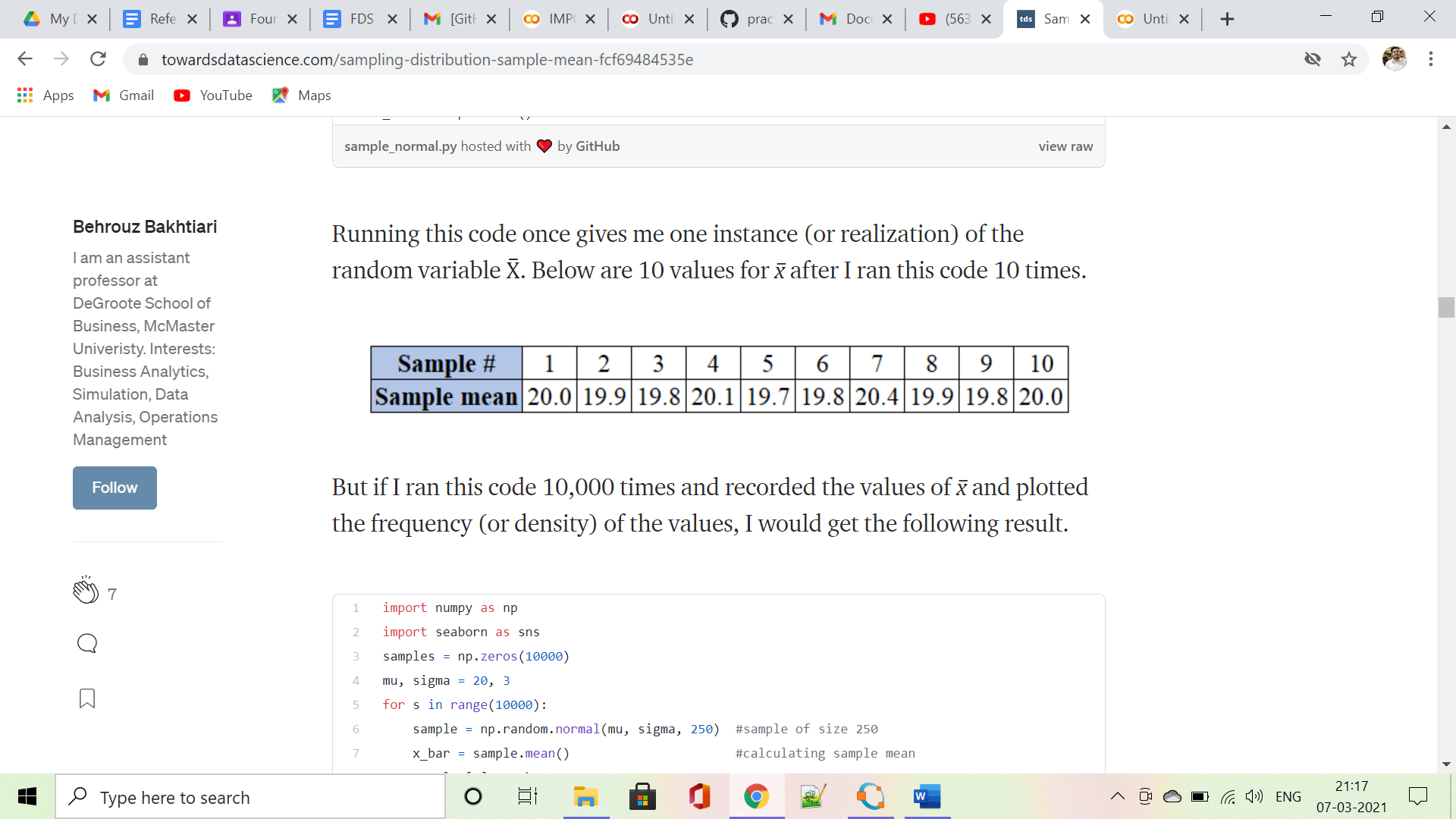
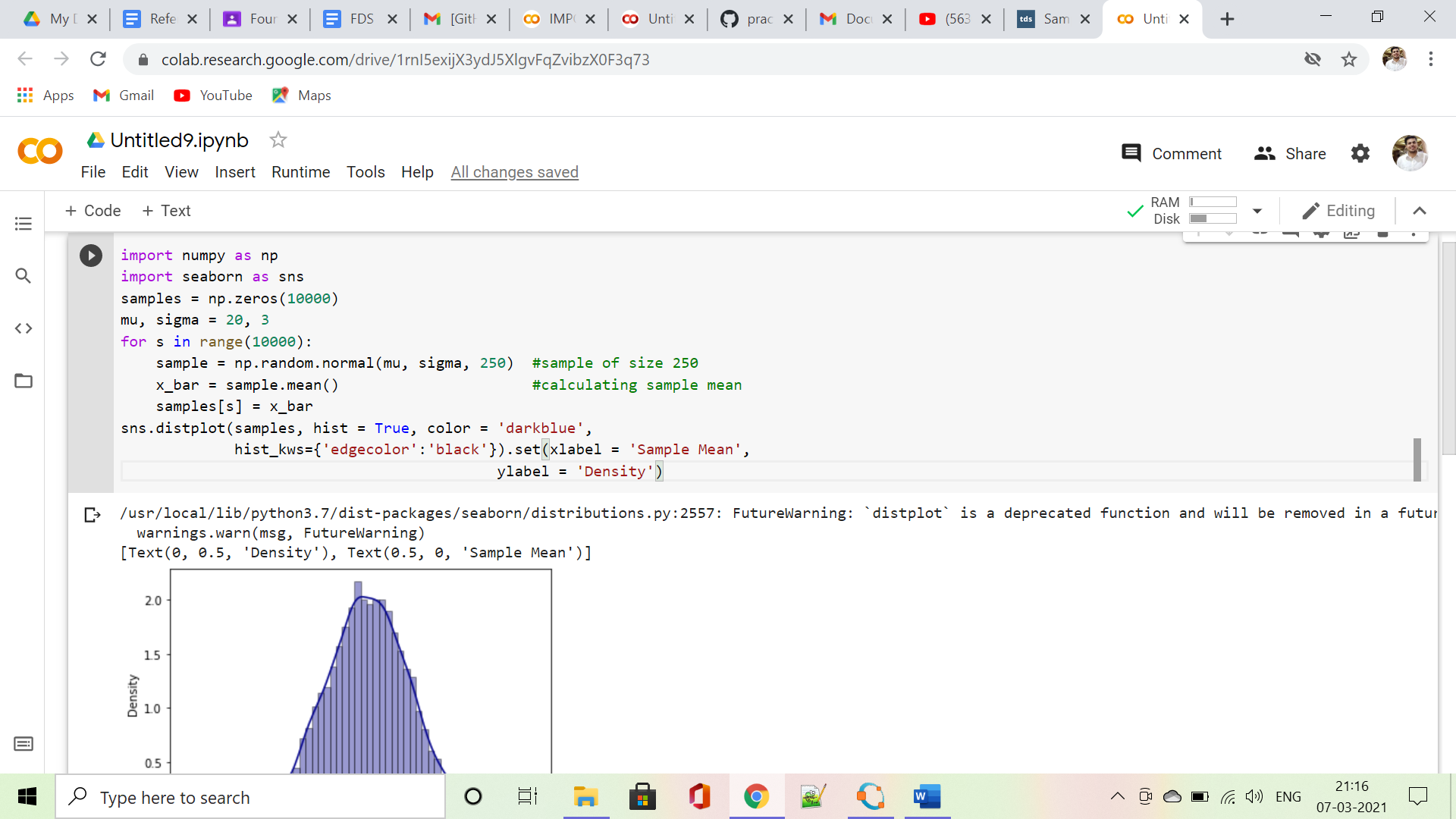
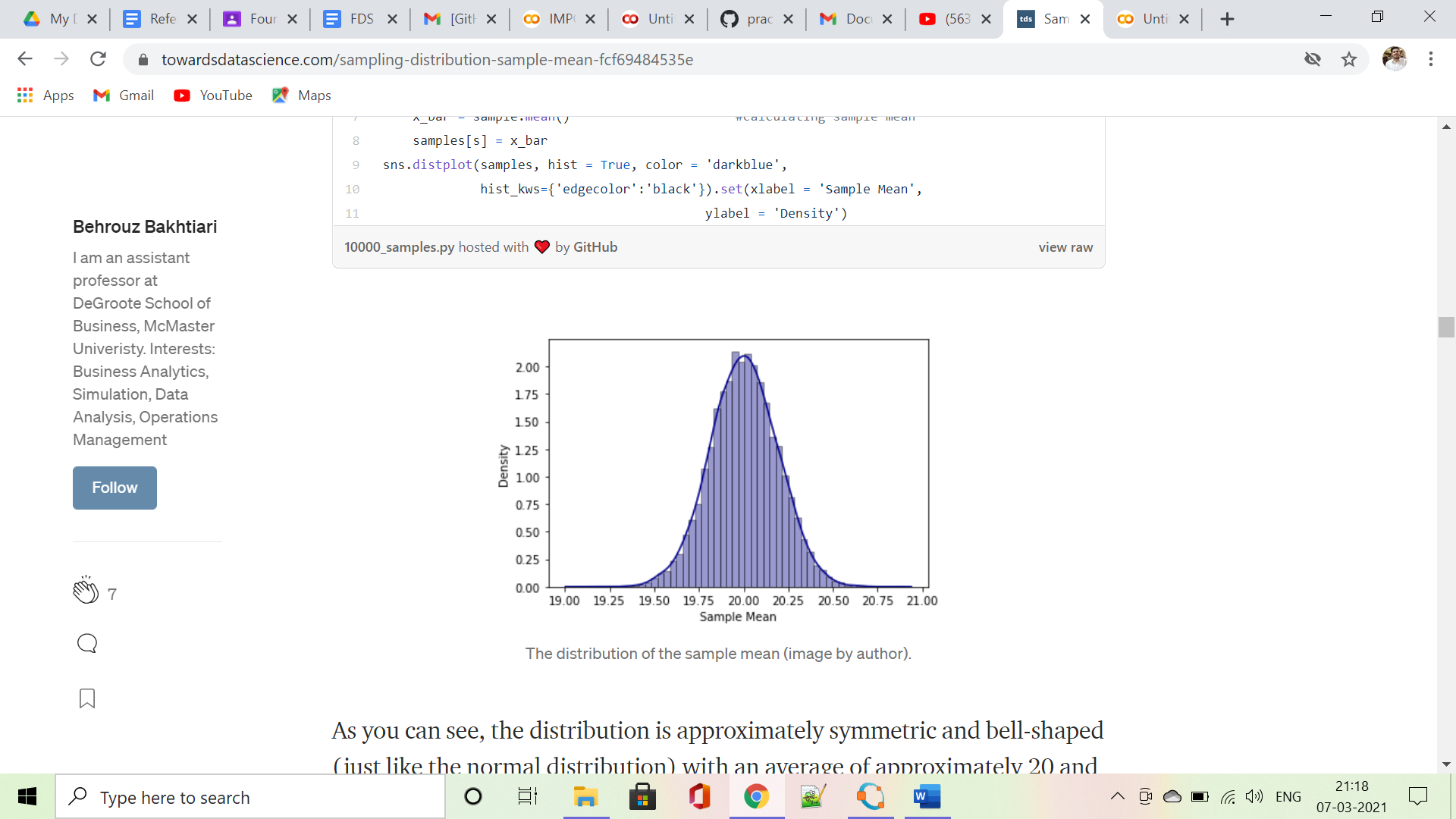
*Sampling Distributions:*

* One of the most important concepts discussed in the context of inferential data analysis is the idea of sampling distributions. Understanding sampling distributions helps us better comprehend and interpret results from our descriptive as well as predictive data analysis investigations. Sampling distributions are also frequently used in decision making under uncertainty and hypothesis testing.
* Sampling distributions describe the distribution for a specific statistic. That is, sampling distributions are a subset (sample) of the full data set, with which you can play, explore and simulate *statistics*like averages, variance and skew.





**OUTPUT:**



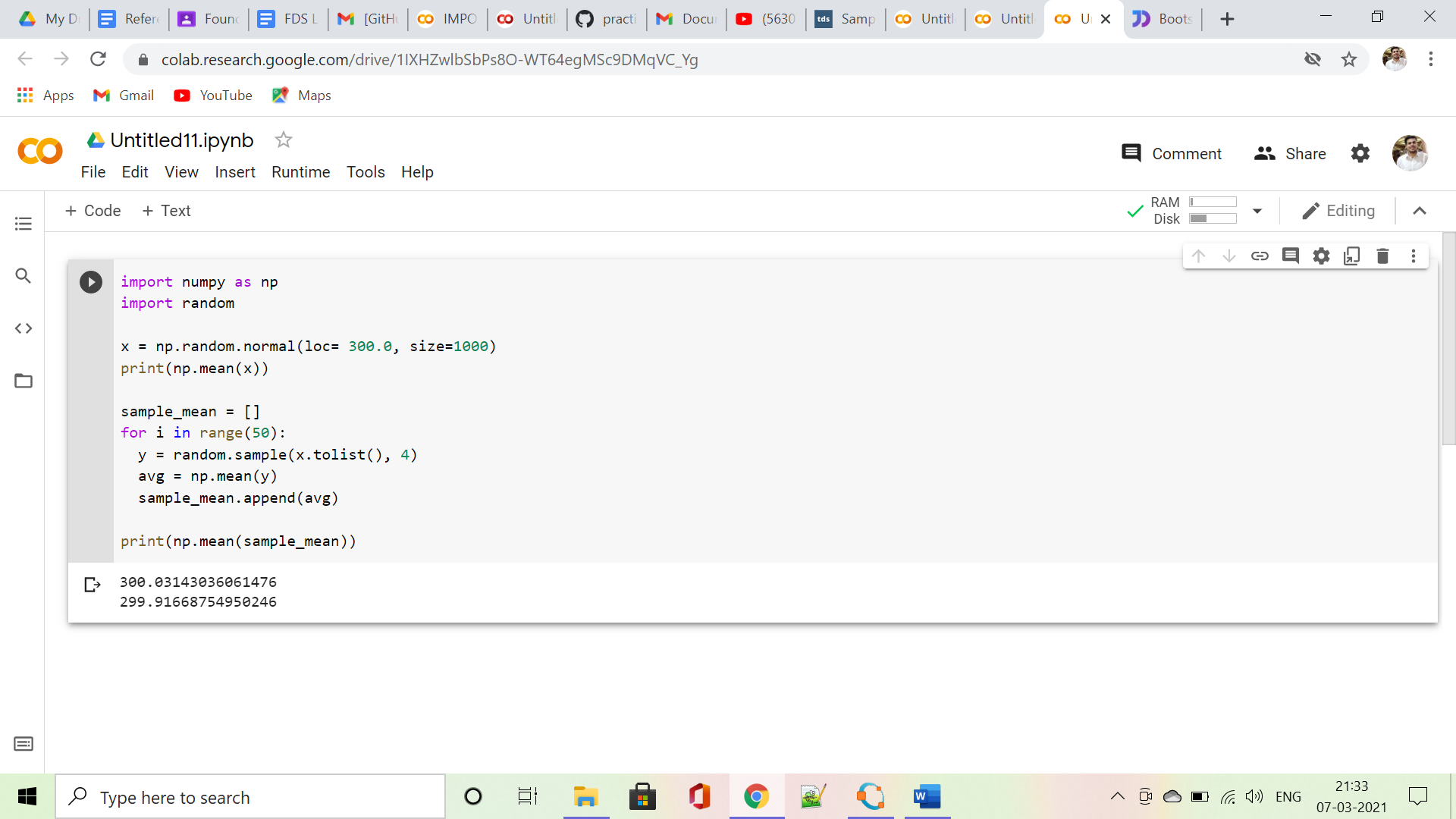
As you can see, the distribution is approximately symmetric and bell-shaped (just like the normal distribution) with an average of approximately 20 and a standard error that is approximately equal to 3/sqrt(250) = 0.19.

Bootstrapping:

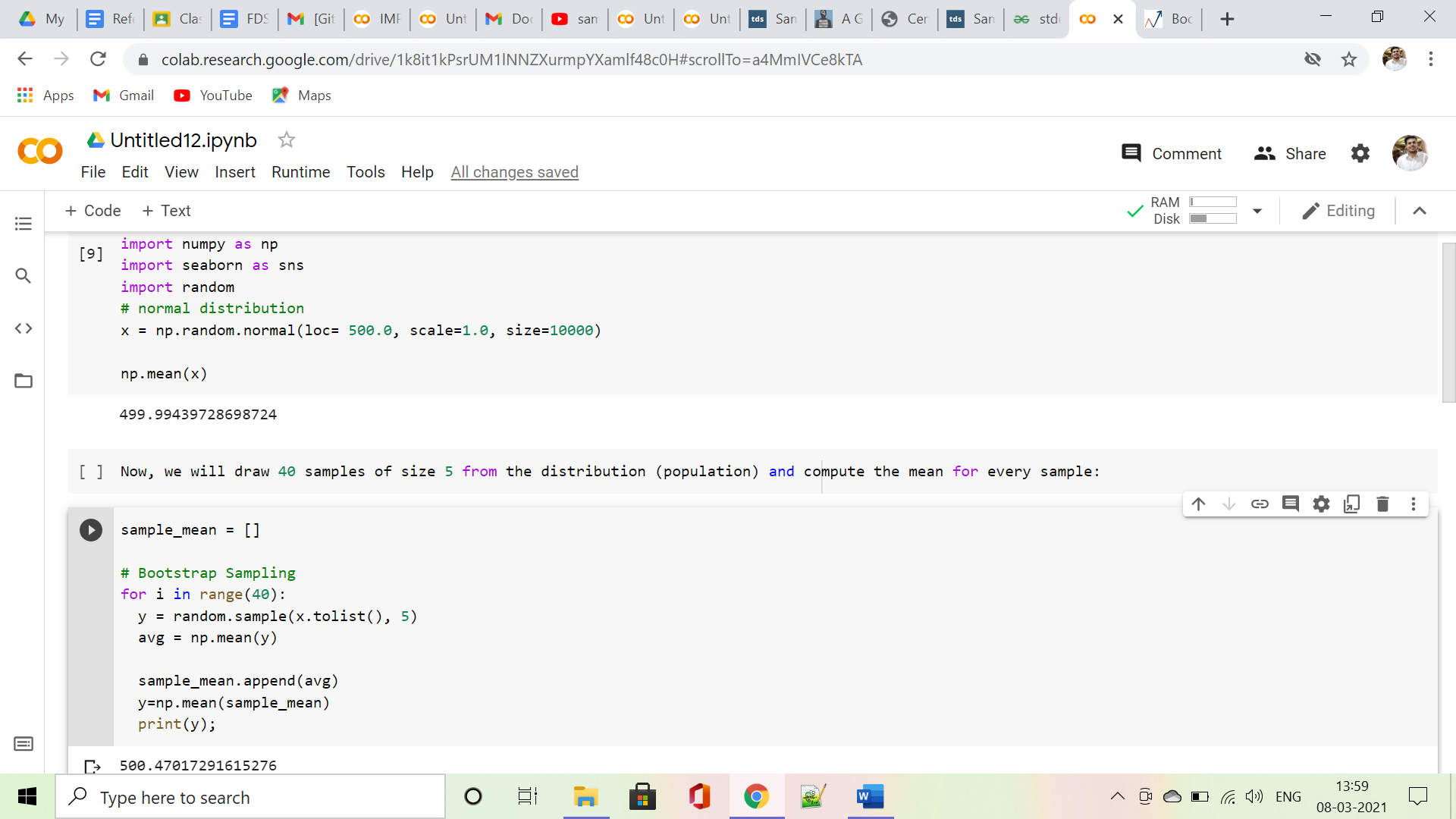
* *In statistics, Bootstrap Sampling is a method that involves drawing of sample data repeatedly with replacement from a data source to estimate a population parameter.*

Let’s break it down and understand the key terms:

* **Sampling:** With respect to statistics, sampling is the process of selecting a subset of items from a vast collection of items (population) to estimate a certain characteristic of the entire population.
* **Sampling with replacement:** It means a data point in a drawn sample can reappear in future drawn samples as well
* **Parameter estimation:** It is a method of estimating parameters for the population using samples. A parameter is a measurable characteristic associated with a population. For example, the average height of residents in a city, the count of red blood cells, etc



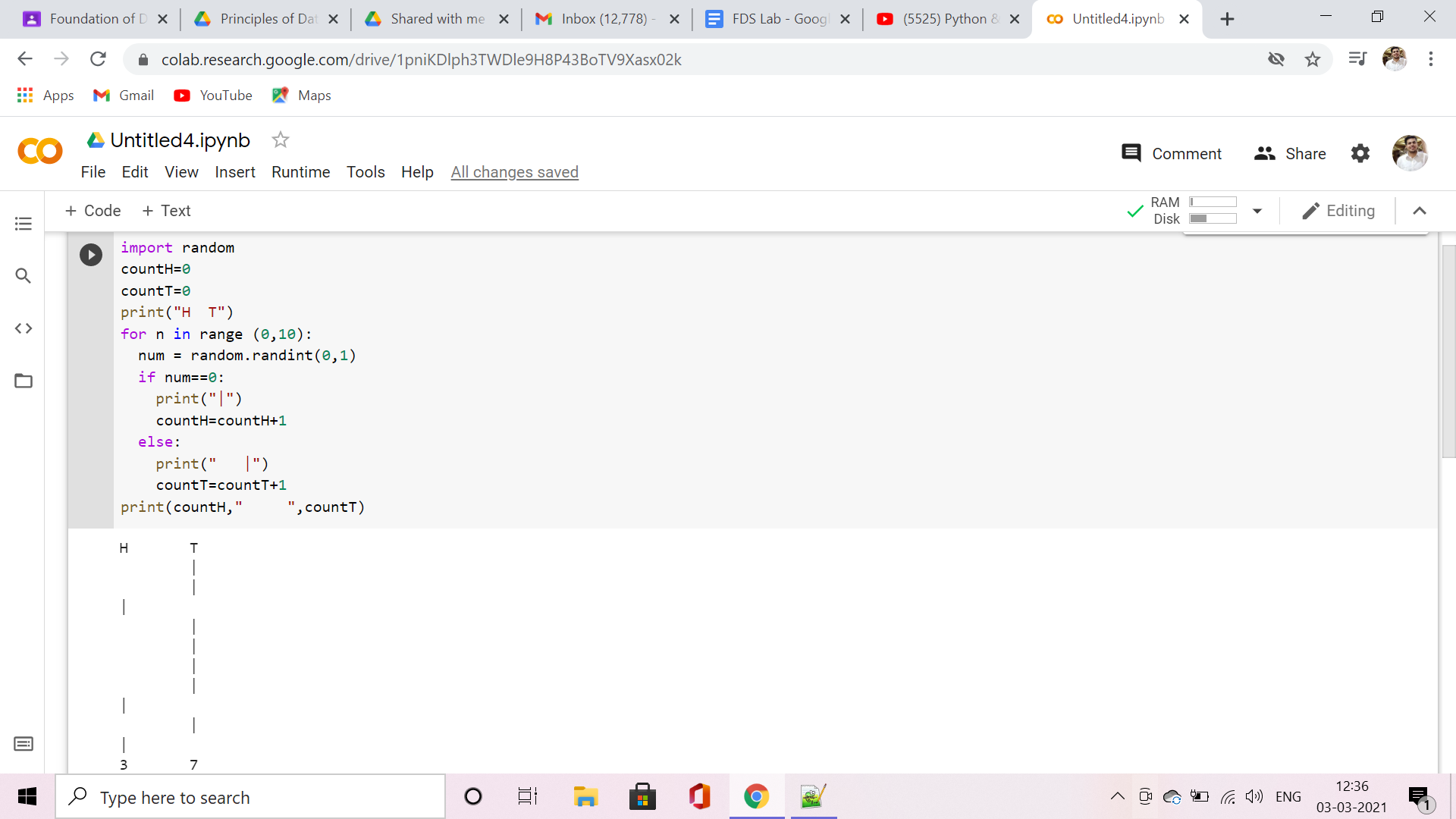
Example 2:



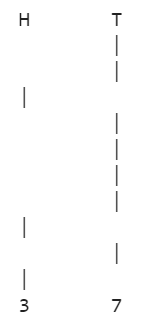
*Law of Large Numbers:*

* *The*[*law of large numbers*](https://en.wikipedia.org/wiki/Law_of_large_numbers)*is a theorem from probability and statistics that suggests that the average result from repeating an experiment multiple times will better approximate the true or expected underlying result.*
* *The law of large numbers explains why casinos always make money in the long run.*

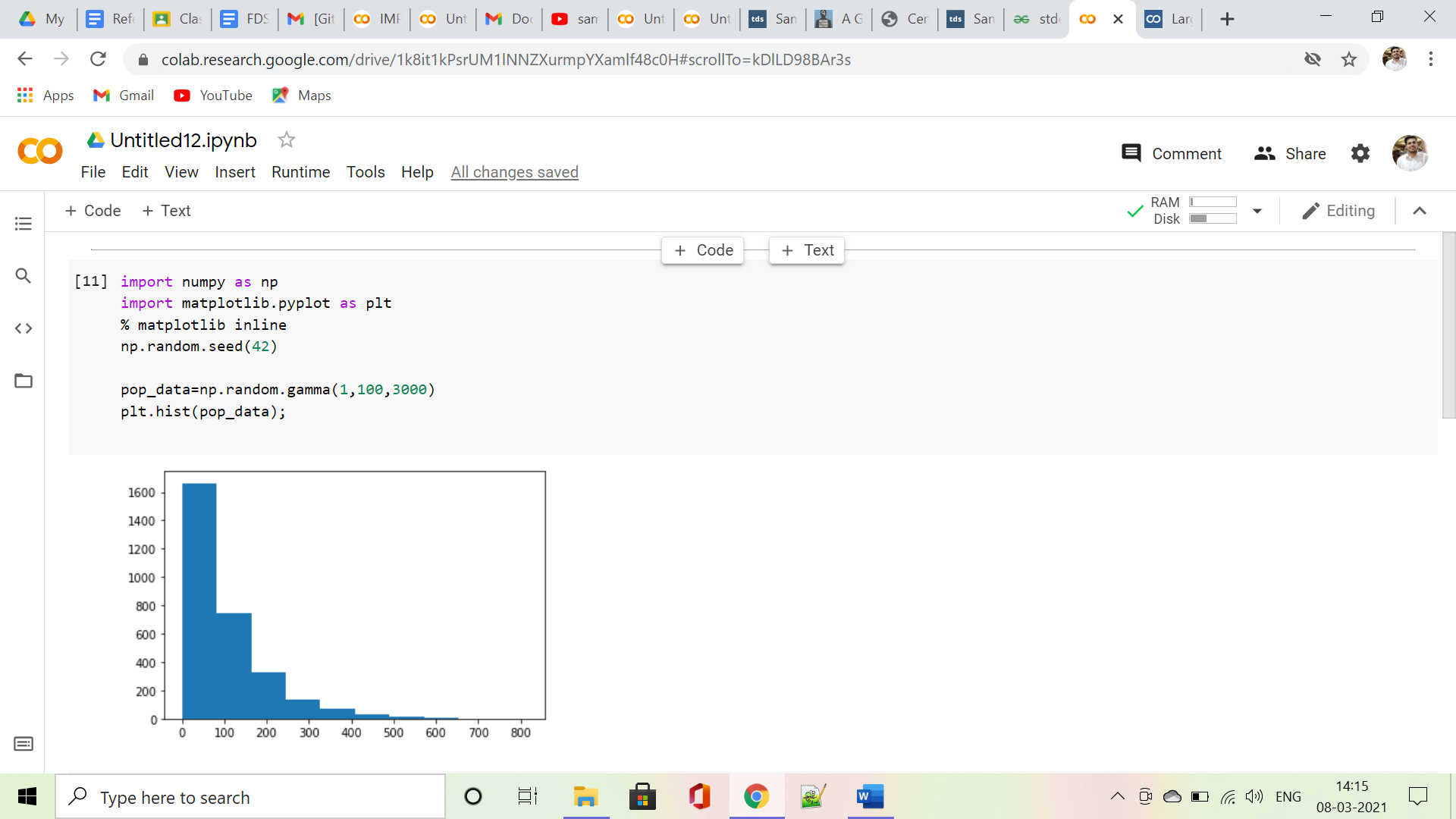
*Example 1:*

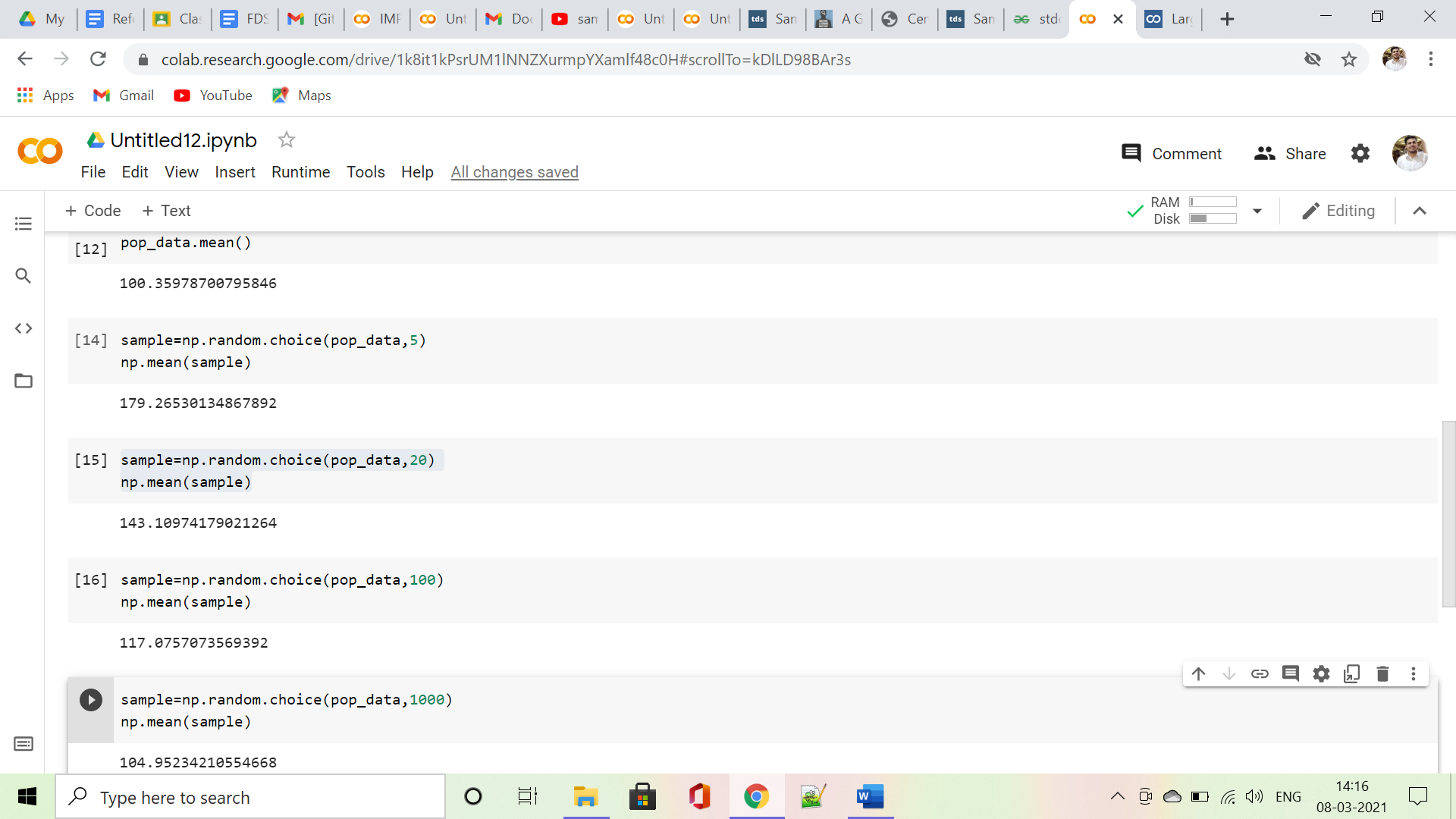


**OUTPUT:**



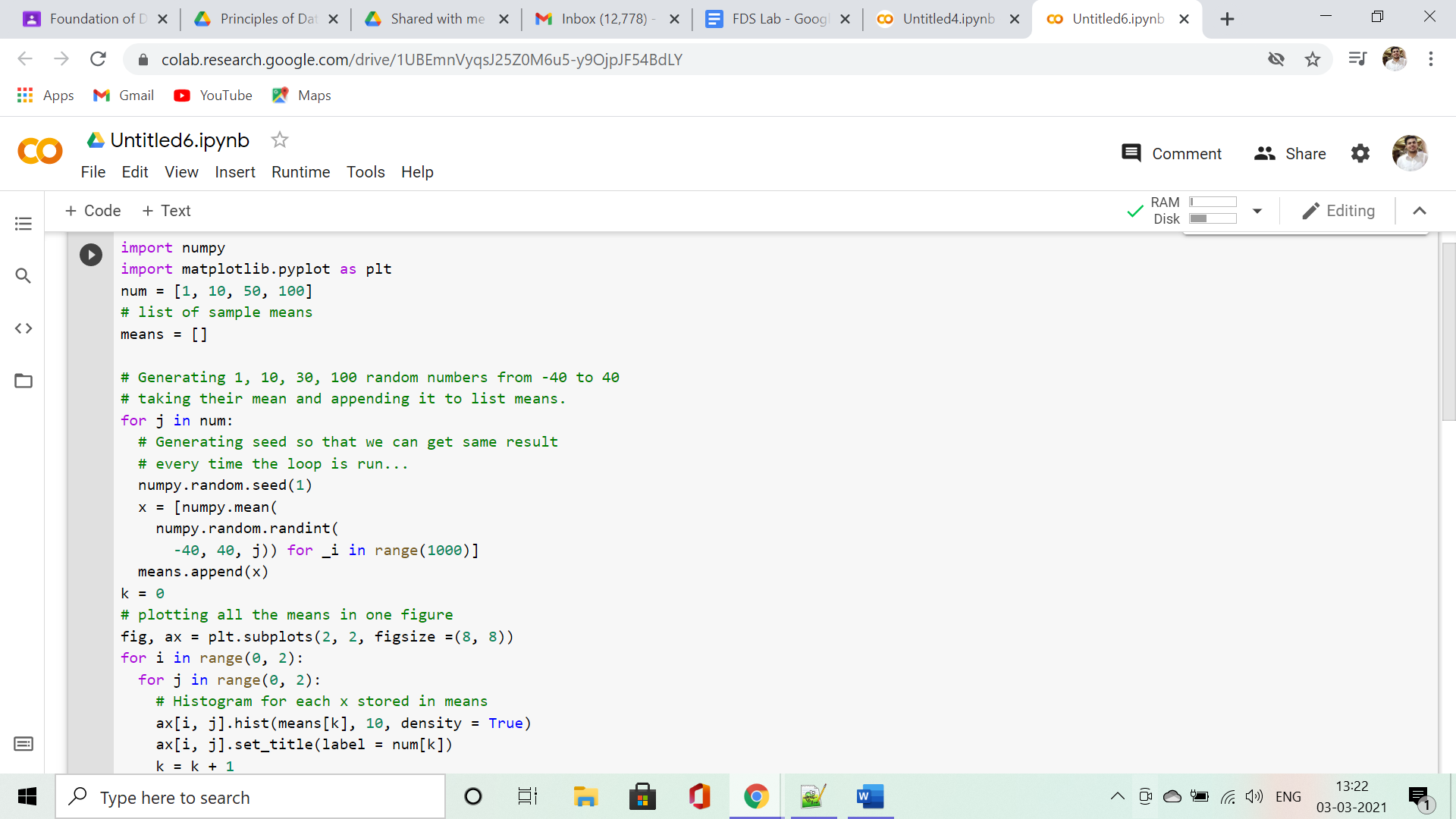
*Example 2: Law of large No.*



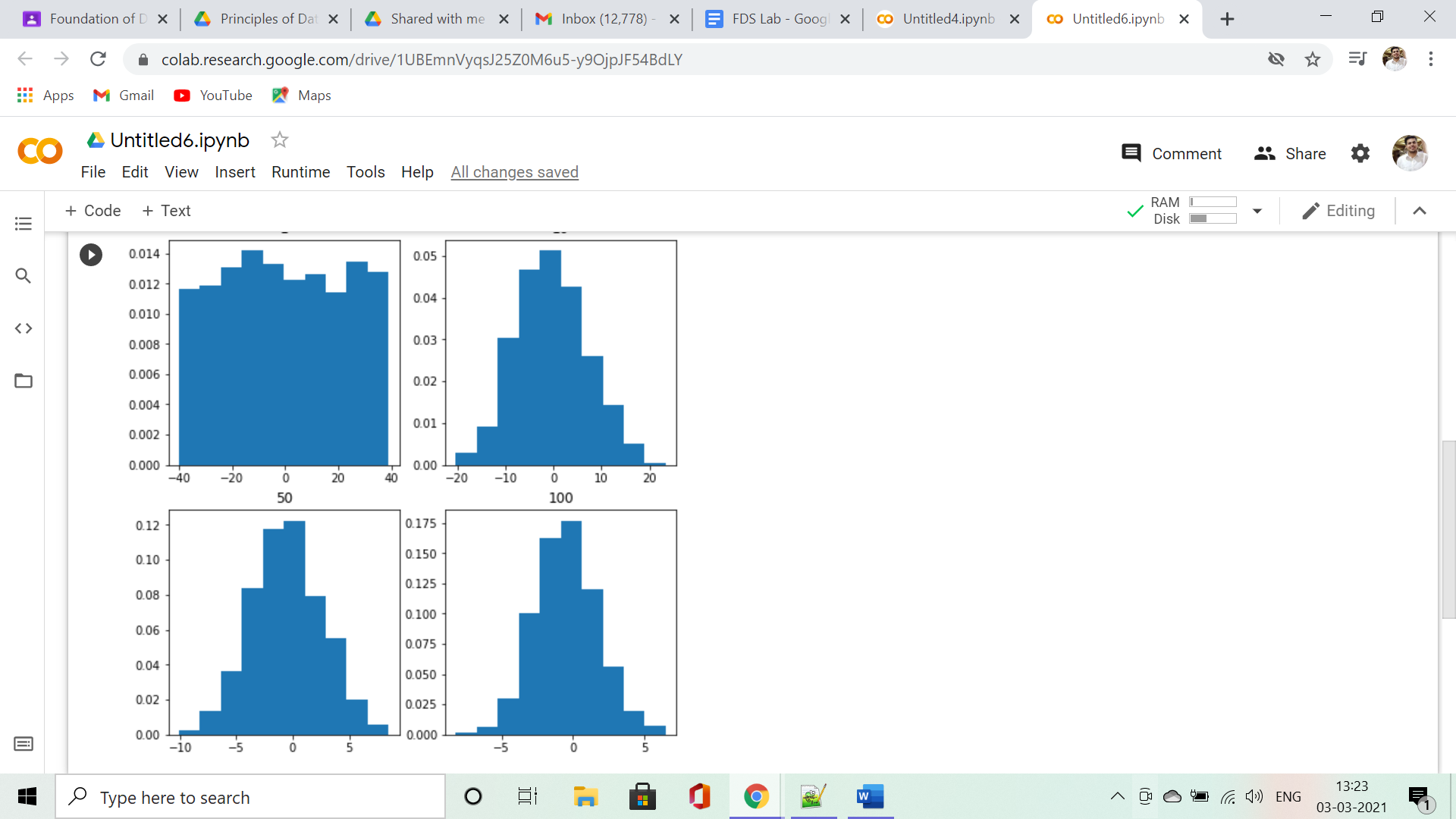


*Central Limit Theorem:*

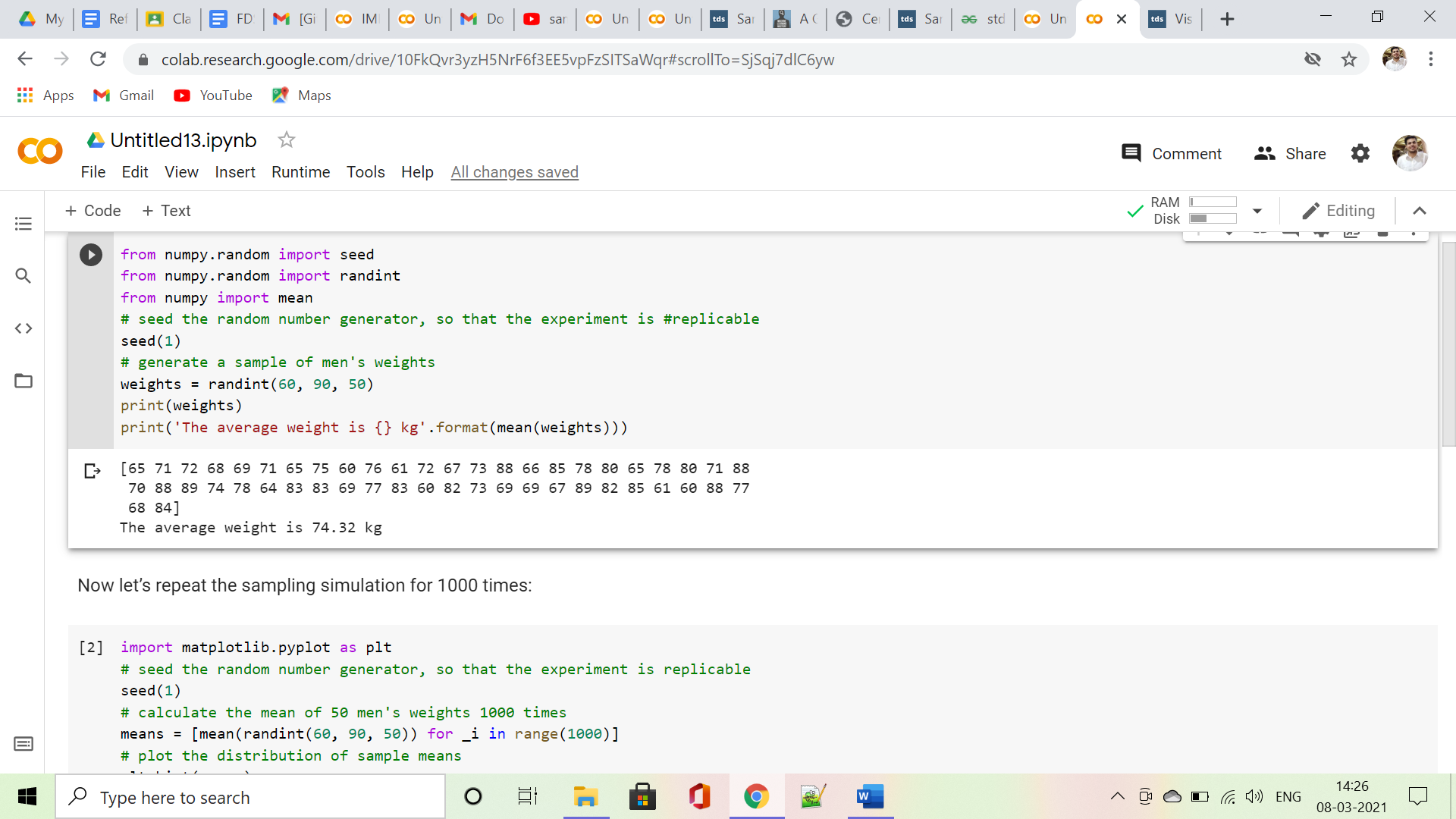
* The central limit theorem states that if you have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population **with replacement**, then the distribution of the sample means will be approximately normally distributed.

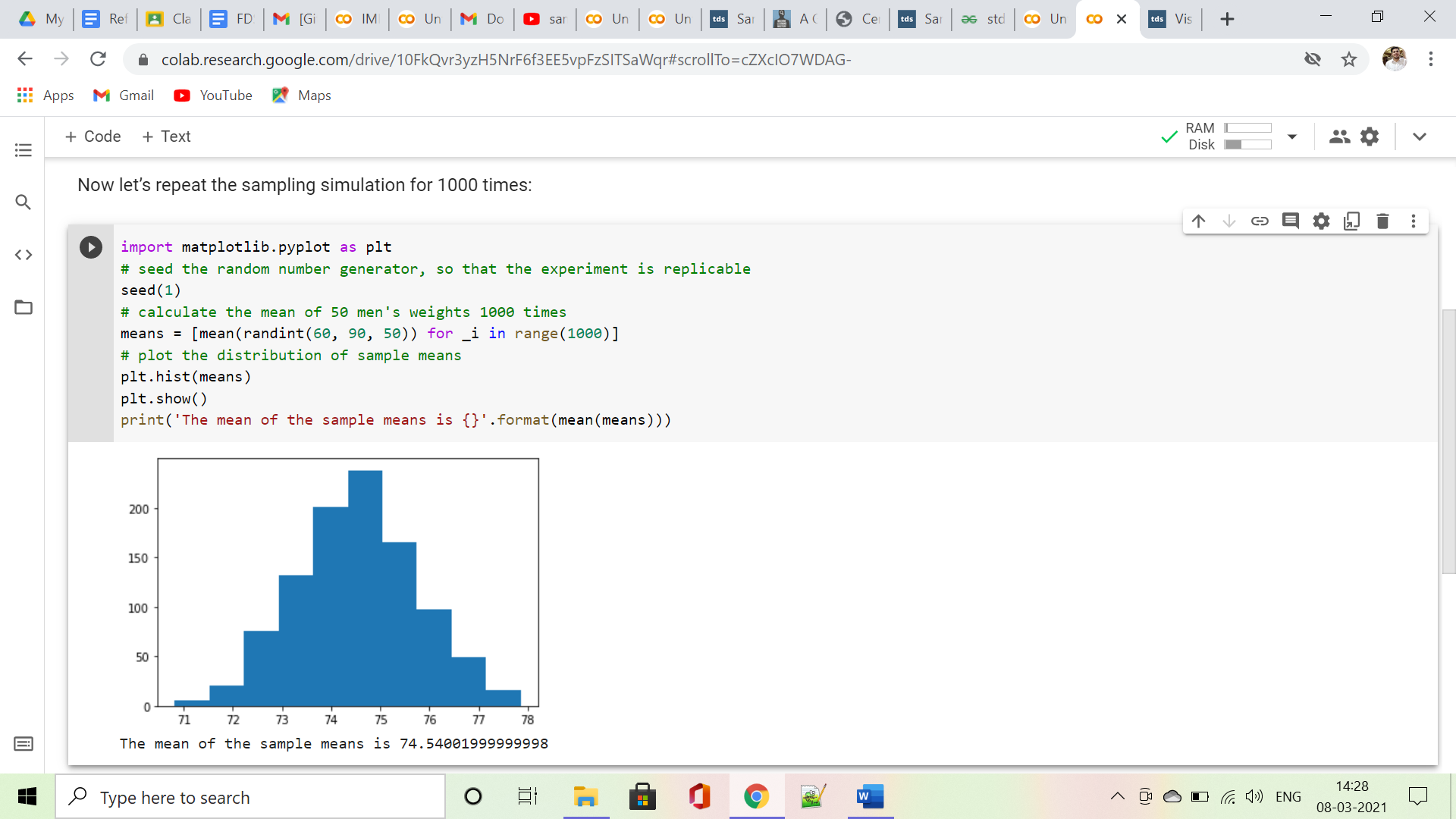


*OUTPUT:*



*Example 2:*

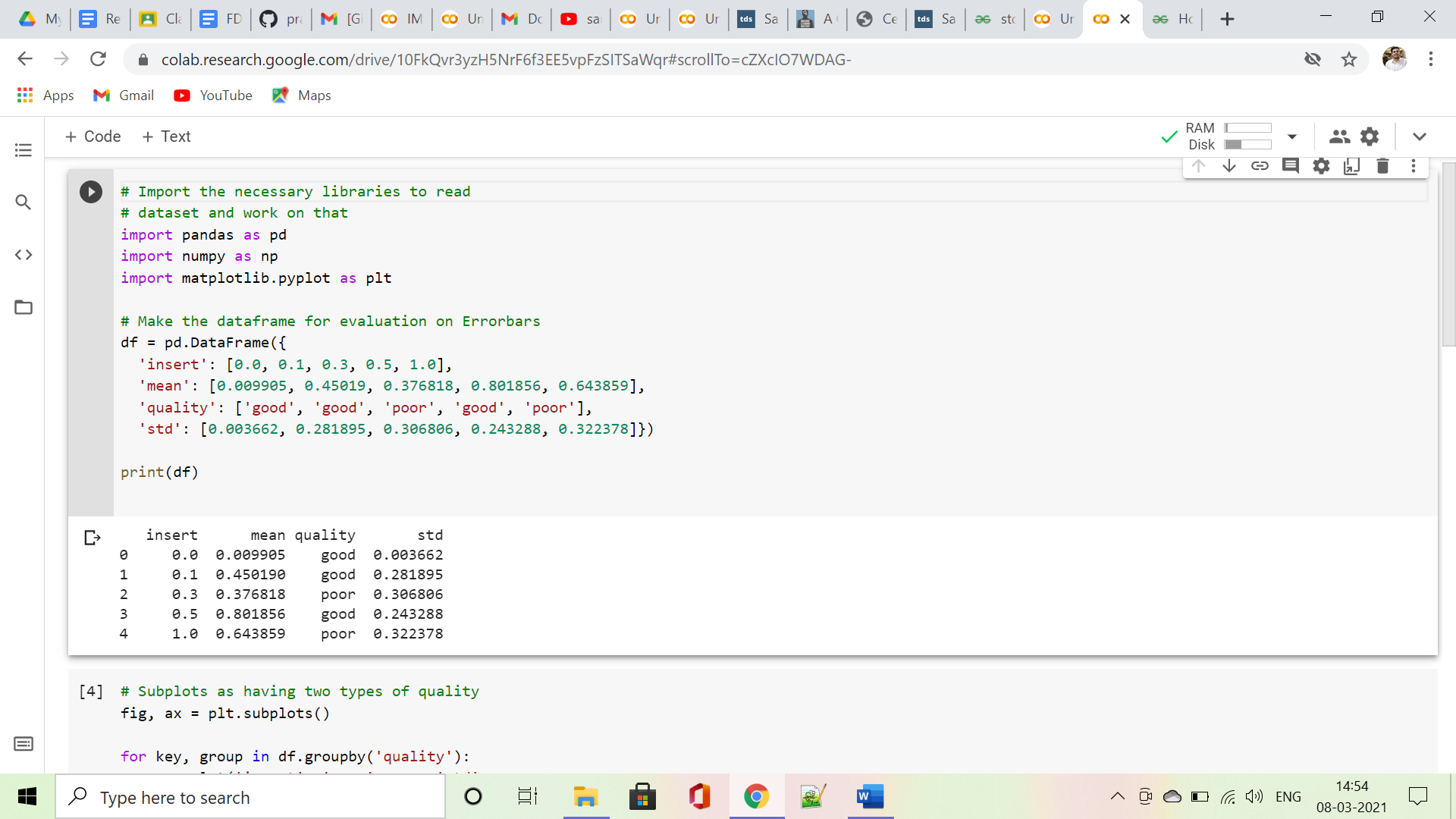




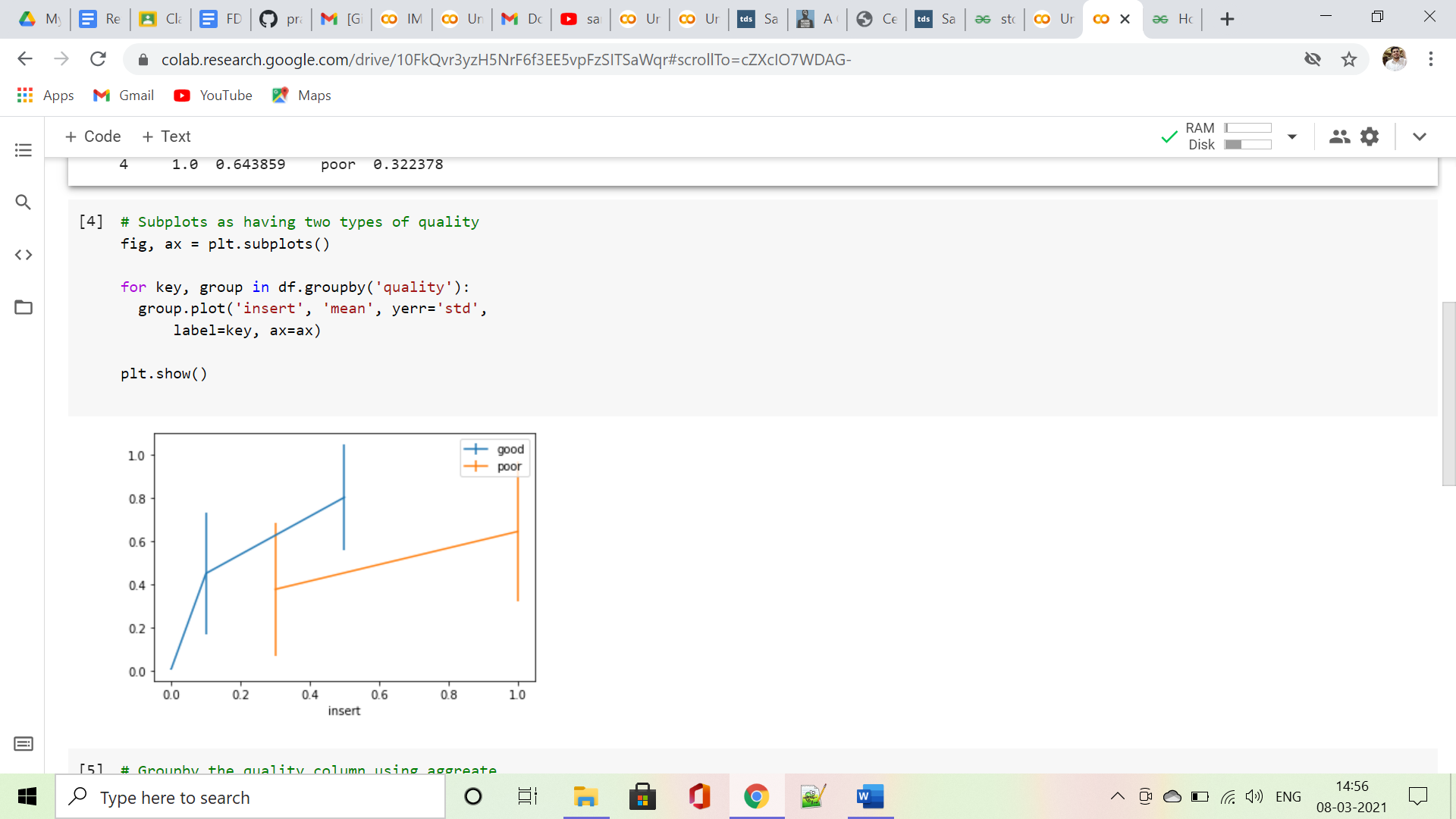
*Standard Deviation:*

* **Standard Deviation** is a measure of spread in Statistics. It is used to quantify the measure of spread, variation of a set of data values. It is very much similar to variance, gives the measure of deviation whereas variance provides the squared value.
* Standard deviation is a number that describes how spread out the values are. A low standard deviation means that most of the numbers are close to the mean (average) value.
* A high standard deviation means that the values are spread out over a wider range.

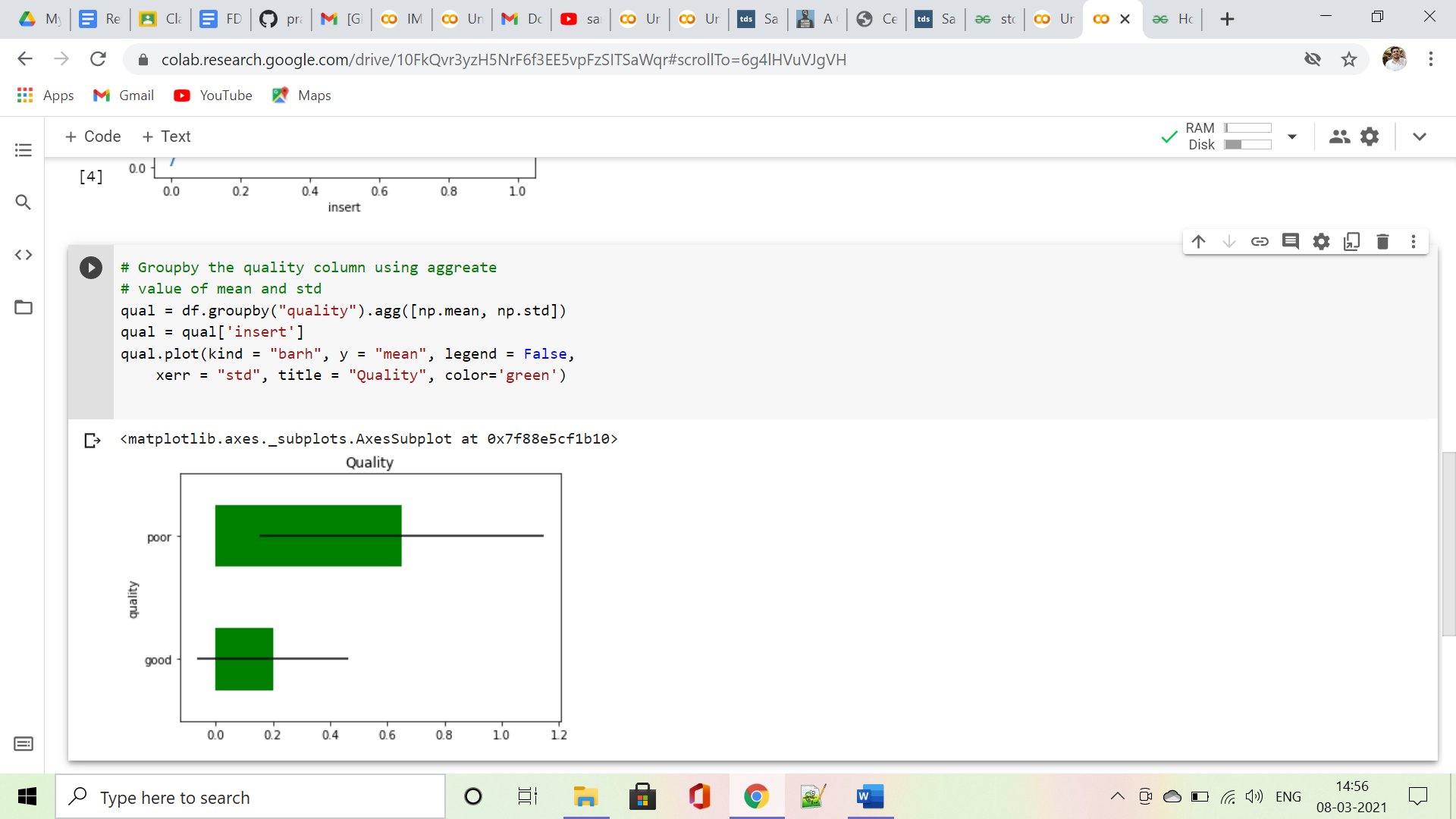
*Example:*



groupby the subplots with mean and std to get error bars:



we see error bars using NumPy keywords of mean and std:



***Confidence Intervals:***

* Confidence intervals are ‘intervals’ that we can create to guess with a certain degree of accuracy where a parameter of interest lies. In the real world, working with an entire populations data can be slow and heavy, but we can use sampling distributions to estimate what a population parameter most probably is.