INDEX

CONTAINS

A statistical study on demand of various characteristics Of a laptop for future laptop price prediction

--- INTRODUCTION---

Laptop pc is very useful gadget in era of today. Maximum work in education ,communication ,business ,official etc are done through the personal computer but laptop pc and knowledge of it than we can consider they are ahead a step to over others . It can be used by businessman , scientist , students etc. to complete their urgent computer related work even while travelling .Laptop is necessary gadget for modern era of today.

Laptop computers, also known as notebooks, are portable computers that you can take with you and use in different environments. They include a screen, keyboard, and a trackpad or trackball, which serves as the mouse. Because laptops are meant to be used on the go, they have a battery which allows them to operate without being plugged into a power outlet. Laptops also include a power adapter that allows them to use power from an outlet and recharges the battery.

--- Objectives of study---

The market of laptop is very competitive at present . In order to gain a competitive edge over other manufacturers ,many features are being added day by day by manufactures in laptop ,and the characteristics of laptop are changing respectively by their changes (RAM, processor, hard disk ,graphic-card ,ports,sensors ,webcom, battery power and weight etc).Many features are very useful for the user which is given by the manufactures but particular type of user can use these additional features which is required or given(i.e. 8 GB RAM, 1 TB Hard Disk , i-7 processor and very low weighted etc.) but all features provided may not be useful for all kinds of users. The use of particular features is

useful for the particular purpose of the users "(educational,multipurpose ,official ,commercial etc.)

---purpose of the study---

- To identify if different categories of person would prefer to have different features in a laptop .
- The problem statement is that if any user wants to buy a laptop then our application should be compatible to provide a tentative price of laptop according to the user configurations.
- The problem statement is that if any user wants to buy a laptop then our application should be compatible to provide a tentative price of laptop according to the user configurations.

---LAPTOP—

A Laptop (also known as a Notebook), is an all-in-one computer that uses batteries or AC power that can last for several hours. Unlike the Desktop, the Laptop computer is easily transported and can be used as long as the battery lasts.

However it is becoming very common for public places to have powering units available for phones and Laptops so they can be used almost everywhere! Laptops come in various sizes with various specifications. Like the Desktop, the type of computer depends on the consumer needs.

A Laptop has a built-in monitor, keyboard, and typically a touchpad (or trackball). External peripherals can also be connected using different cable connections, depending on what ports are provided by manufacturer. Essentially, the Laptop can be used as a fully functional unit without any connected devices or power.

As technology advances, the Laptop is becoming more popular with the average consumer and the preferred computer for business users, especially those who travel for work.

Similarities--

- Most Desktops and Laptops are sold with pre-installed software and operating systems like Windows 7, or if it is an Apple computer, then the Mac OS.
- Both computer units come with port connections (varying between each make and model), and a built-in CD/DVD component, however this is being phased out in some of the newer Laptop models.
- Peripheral devices can be connected such as external hard drives, printers, cameras, and phones, etc.
- For both the Desktop and Laptop, key specifications to consider before purchasing is
  - CPU
  - Memory (RAM)
  - Hard drive capacity
  - Graphics card.
  - These specs determine the computer's limits so if video production is the primary task and the computer has a low end graphics card, then the computer (laptop or desktop) will not be well suited for the job.

---characteristics of laptop---

a.How Many Types of Laptops Are?

With advanced information technology, people are going to be more professional. Gone are the days of the monitor, keyboard, and CPU. Now people want everything personal and private. Laptops provide one of the easiest and most convenient ways for all those netizens who often carry their system with their own. Laptops are classified according to their uses and processors, the major types of laptops are given below:

1. Notebook

2. Chromebook
3. Netbook
4. Ultrabook
5. MacBook Air
6. Convertible 2 in 1
7. MacBook

Q1.What is a Notebook?

The notebook laptop computer is simply named a mini-computer that is smaller than other computers. This laptop is designed in such a way that consists of a monitor screen, touchpad, and battery along with enough storage of RAM and ROM. The average weight of a notebook is around 6 pounds which can be carried easily like a briefcase.

Its VGA screen resolution provides the best display comfort to show sharp images. The advanced edition of the notebook computer consists of long battery life so that you can operate it for a long time without plugging in the charger.

Q2.What is a Chromebook?

The Chromebook laptop computer is basically a product of Google which is why it is known as the Google Chromebook. This laptop computer is solely based on the fastest programming that enables users to do work fast and smoothly. If you will compare this laptop to another computer then it is much better than another normal computer.

Moreover, still, a lot of people are confused about the difference between a laptop and Chromebook. While the major difference is price and operating system. Chrome OS is operated by Google itself and its price varies on System Configuration. The first Chromebook was launched in 2011 and after that users liked to operate it for various tasks.

Q3.What is Netbook?

The netbook computer is also known as a portable laptop computer or mini laptop that is similar to a notebook computer. Now the point is, that a netbook just looks like a notebook but it is slightly different from a notebook. You can carry this laptop computer easily just because of its compact size.

In the early phase of this computer, it was designed for students, bloggers, and web professionals. But now as the demand has been hiked, people from various sectors are using a netbook laptop computer. The latest version of the netbook runs on Linux, Windows XP, and Chrome OS. If you are a professional blogger or working in any field operations then netbook is only for you.

Q4.What is an Ultrabook?

An ultrabook is a kind of personal computer that is the same as a laptop. The design of an ultrabook laptop is very thin and light in weight so that anyone can carry it easily. On the other hand, it is much better than tablets or notebooks and is business-friendly for users. Currently, Ultrabook is available in various versions according to its size, battery life, operating system, and price.

Most people are unable to detect the basic difference between an ultrabook and a netbook. For the betterment of a smooth operating system, an ultrabook laptop computer gives you better performance in terms of video performance, RAM, and visual screen size.

Q5.What is the difference between MacBook vs MacBook Air?

Both MacBook and MacBook Air are products of Apple and slightly different from each other. If you are on the fence about what to get and what not to get, then you should be aware of the specifications and price of these laptop computers.

Recently, Apple has added a lot of features to the Mac devices. So it would be better for you to do an inspection first than by anyone of your choice. MacBook and MacBook Air are similar to each other but their specification is different in terms of battery, storage, memory, screen size, retina screen resolution, etc.

Q6.What is a Convertible 2 in 1 Laptop?

Convertible 2 in 1 is the same as the personal portable computer that operates for two computers.

Whether you are a student or a working professional convertible 2 in 1 laptop allows you to remove extra accessories. With the help of this laptop, you can combine two devices in one which is why its name suggests convertible 2 in 1. By means of transport, you can carry this laptop computer very easily after folding it.

The laptop is the renowned version of the computer. Gone are the days when people used to operate desktop systems that carry a lot of additional tools. Since laptops have come forward everything has changed. Now people can choose a laptop according to their work needs or in terms of processor.

B. What is CPU on a laptop?

The processor, also known as the CPU, provides the instructions and processing power the computer needs to do its work. The more powerful and updated your processor, the faster your computer can complete its tasks. By getting a more powerful processor, you can help your computer think and work faster.

All kinds of computing devices such as tablets, PCs, or laptops feature a brain-like unit called the central processing unit or CPU. Your computer's CPU calculates and interprets instructions while you're surfing the web, creating documents, playing games, or running software programs.

CPU speed is measured in gigahertz (GHz), and a CPU speed of 3.5 GHz is more than enough for most users to run your preferred software. For gaming, video editing, and other applications that need several cores, aim for a CPU speed of 3.5 GHz to 4.0 GHz for best results.

C. RAM

RAM is a form of temporary computer storage that allows stored data to be received and read almost instantaneously. When you fire up a program, it

becomes temporarily stored in your computer's memory (or RAM) for easy access, as opposed to being written on the permanent hard drive.

For anyone looking for the bare computing essentials, 4GB of laptop RAM should be sufficient. If you want your PC to be able to flawlessly accomplish more demanding tasks at once, such as gaming, graphic design, and programming, you should have at least 8GB of laptop RAM.

RAM**,** the computer's memory, and the processor both affect how fast your laptop runs. The higher the number for each, the faster the speed. For instance, 4GB of RAM run faster than 2GB. The processor speed is measured in gigahertz, and a 2 GHz machine runs faster than a 1 GHz.

C. PROCESSOR

The processor, also known as the CPU, provides the instructions and processing power the computer needs to do its work. The more powerful and updated your processor, the faster your computer can complete its tasks. By getting a more powerful processor, you can help your computer think and work faster.

The different types of processors are microprocessor**,** microcontroller, embedded processor, digital signal processor and the processors can be varied according to the devices. The important elements of the CPU are labelled as heart elements of the processor and system.

D. Hard disk

An HDD is a data storage device that lives inside the computer. It has spinning disks inside where data is stored magnetically. The HDD has an arm with several "heads" (transducers) that read and write data on the disk. The number of supported hard drives on a PC largely depends on there are how many available SATA slots on the motherboard. The majority of motherboards come with 4 or 6 SATA slots. It means that you can install 4-6 hard drives to your computer.

E. Graphic card

A graphics card (also called a video card, display card, graphics adapter, VGA card/VGA, video adapter, display adapter, or mistakenly GPU) is an expansion card which generates a feed of output images to a display device, such as a computer moniter. Graphics cards are sometimes called discrete or dedicated graphics cards to emphasize their distinction to integrated card. A graphic processing unit that performs the necessary computations is the main component of a graphics card, but the acronym "GPU" is sometimes also used to refer to the graphics card as a whole.
Usually, a graphics card comes in the form of a printed circuit board (expansion board) which are to be inserted into an expansion slot. Others may have dedicated enclosures, and they are connected to the computer via a docking station or a cable. These are known as external GPUs (eGPUs).

F. Battery

A battery is a hardware component that supplies power to a device, enabling that device to work without a power cord. Batteries are often capable of powering a laptop computer for several hours depending on how much power it requires. Today, many high-end devices like computer laptops and Cell phones use rechargeable batteries that allow a user to recharge the battery when depleted of energy. The picture below is an example of a laptop battery with a close-up of the battery rating.

G. Operating system

The operating system (OS) is the system software used to manage the computer's software, hardware, and resources. The OS is needed to coordinate common services and provide a user interface for interacting with the program and hardware. Operating systems are important since we can't use computers without them. Some examples of operating systems include Apple macOS, Microsoft Windows, Google's Android OS, Linux Operating System, and Apple iOS.

H. Ports

A port is an attachment for a notebook computer that allows a number of devices such as a printer ,large monitor ,and keyboard to be simultaneously connected .Each of the devices is attached to the port and when the notebook user wants access to one or more of the devices,the user simply attaches the port rather than having to connect one device at a time .The port duplicates each of the notebook's port ,including parallel and serial ports . In addition ,it may provide some extra ports for devices such as joysticks or musical instrument digital interface devices.

I.GPU

A GPU allows a computer to render images processed by the CPU. You can get a firm grasp on the GPU's power when you run graphics-intensive processes, such as gaming, video editing, and 3D rendering. The short answer here is a definitive yes. No matter what you're going to be doing on your laptop you will need a GPU, as it is solely responsible for creating all the images you see on screen

---Collection of data---

For the study primary and secondary data was collected .The secondary data regarding various characteristics of laptop has been collected from various data sources such as https://docs.aws.amazon.com/sagemaker/latest/dg/data-wrangler-import.html , https://www.amazon.in/laptop/s?k=laptop . Most of the columns in a dataset are noisy and contain lots of information. But with feature engineering ,we will get more good results. The only problem is we are having less data but we will obtain a good accuracy over it.

CHAPTER=2

STATISTICAL TECHNIQUES USED:

While studying the perception of the respondent regarding laptop different statistical tools and techniques was used:

- correlation
- Exploratory data analysis
- Linear regression
- Lasso regression
- Ridge regression
- Random forest

---Correlation ---

Correlation is statistical measure that indicates the extent to which two or more variables fluctuate together .A positive correlation indicates the extent to which those variables increase or decrease in parallel ; a negative correlation indicates the extent to which one variable increases as the other decreases.

A correlation coefficient is a statistical measure of the degree to which changes to the values of one variable predict change to the value of another ,when the fluctuation of one variable reliably predicts a similar fluctuation in another variable ,there's often a tendency to think that means that the change in one causes the change in the other .

Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. For example, an electrical utility may produce less power on a mild day based on the correlation between electricity demand and weather. In this example, there is a casual relationship, because extream whether causes people to use more electricity for heating or cooling. However, in general, the presence of a correlation is not sufficient to infer the presence of a causal relationship (i.e., correlation does not imply caution).

---Exploratory data analysis---

Exploratory analysis is a process to explore and understand the data and data relationship in a complete depth so that it makes feature engineering and machine learning modeling steps smooth and streamlined for prediction. EDA involves Univariate, Bivariate, or Multivariate analysis. EDA helps to prove our assumptions true or false. In other words, it helps to perform hypothesis testing. We will start from the first column and explore each column and understand what impact it creates on the target column. At the required step, we will also perform preprocessing and feature engineering tasks. our aim in performing in-depth EDA is to prepare and clean data for better machine learning modeling to achieve high performance and generalized models. so let's get started with analyzing and preparing the dataset for prediction

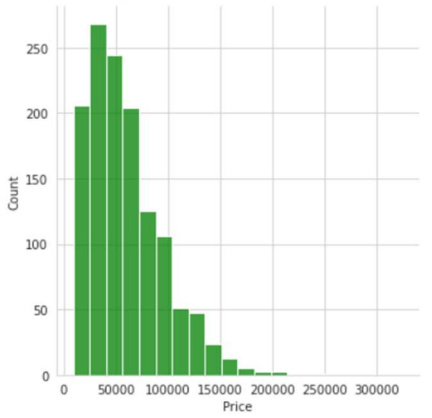
Working with regression problem statement target column distribution is important to understand.

. The distribution of the target variable is skewed and it is obvious that commodities with low prices are sold and purchased more than the branded ones.

1. Distribution of target column (price)

```
# Distribution of target column (price)
sns.set_style('whitegrid')
sns.displot(df['Price'],color='green',bins=20)
```

```
<seaborn.axisgrid.FacetGrid at 0x7f7a0806bd60>
```



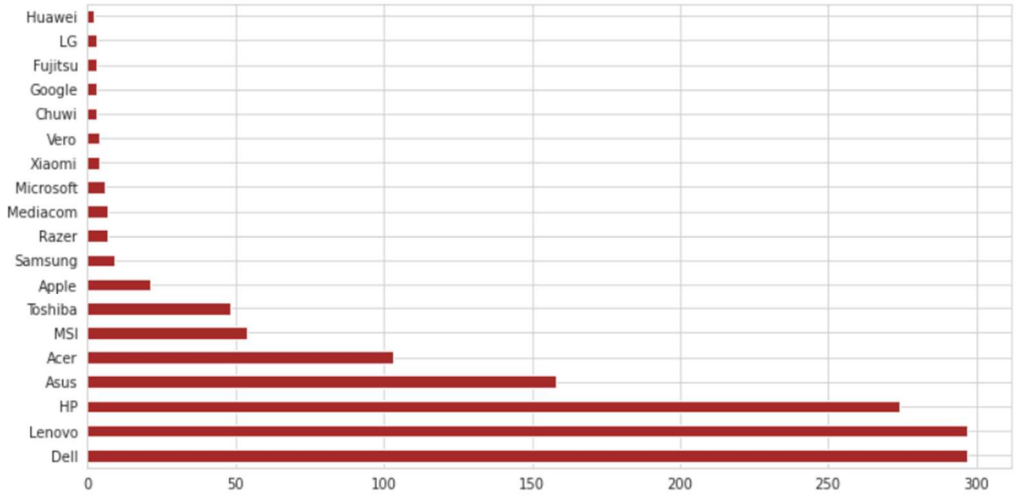

```
[143] #how does brand name impacts the laptop sells
      df['Company'].value_counts().plot(kind='bar')
```

0s completed at 12:43 AM

2. How does brand name impacts the laptop sells?

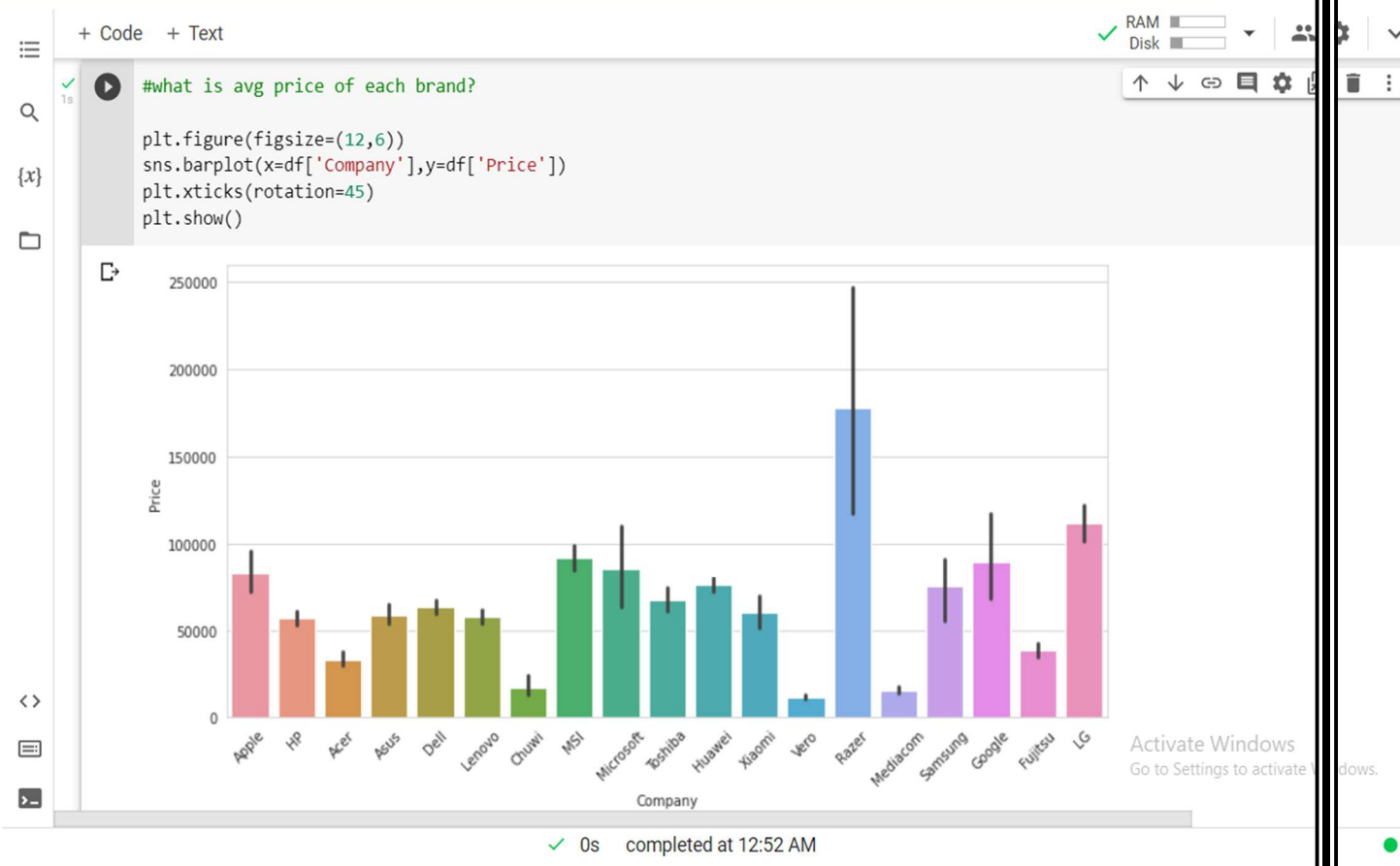
```
#how does brand name impacts the laptop sells
plt.figure(figsize=(12,6))
df['Company'].value_counts().plot(kind='barh',color='brown')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7a0747e9a0>
```

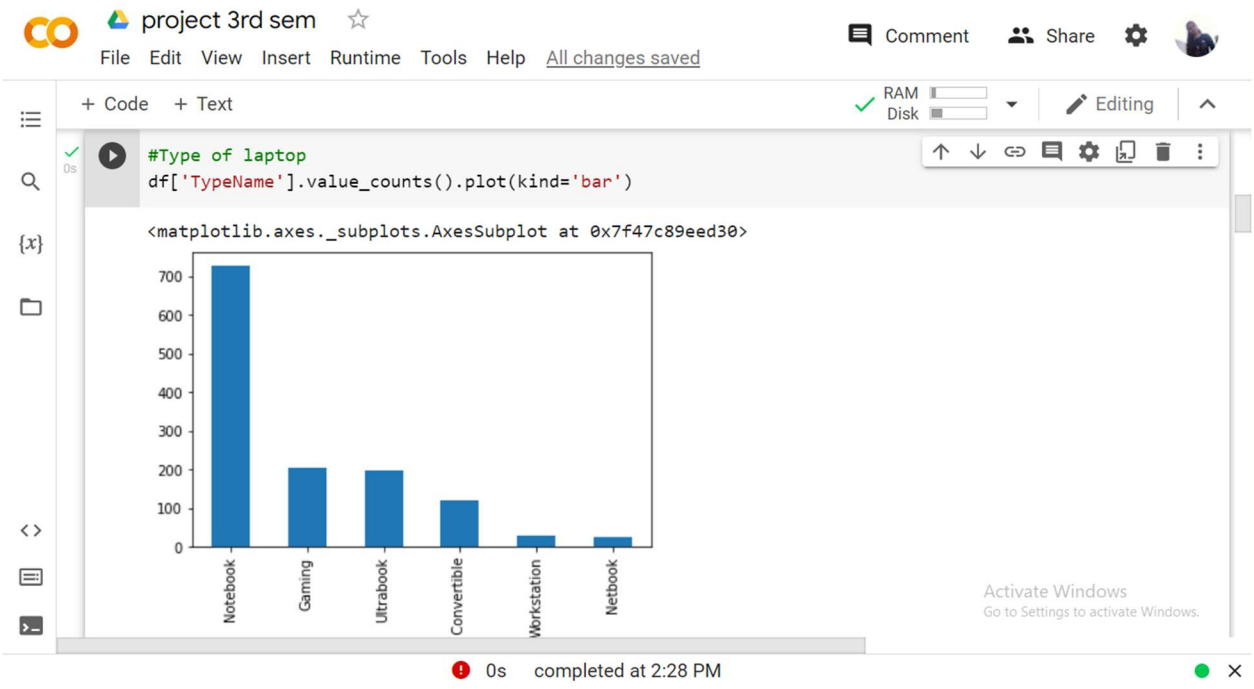


0s completed at 12:50 AM

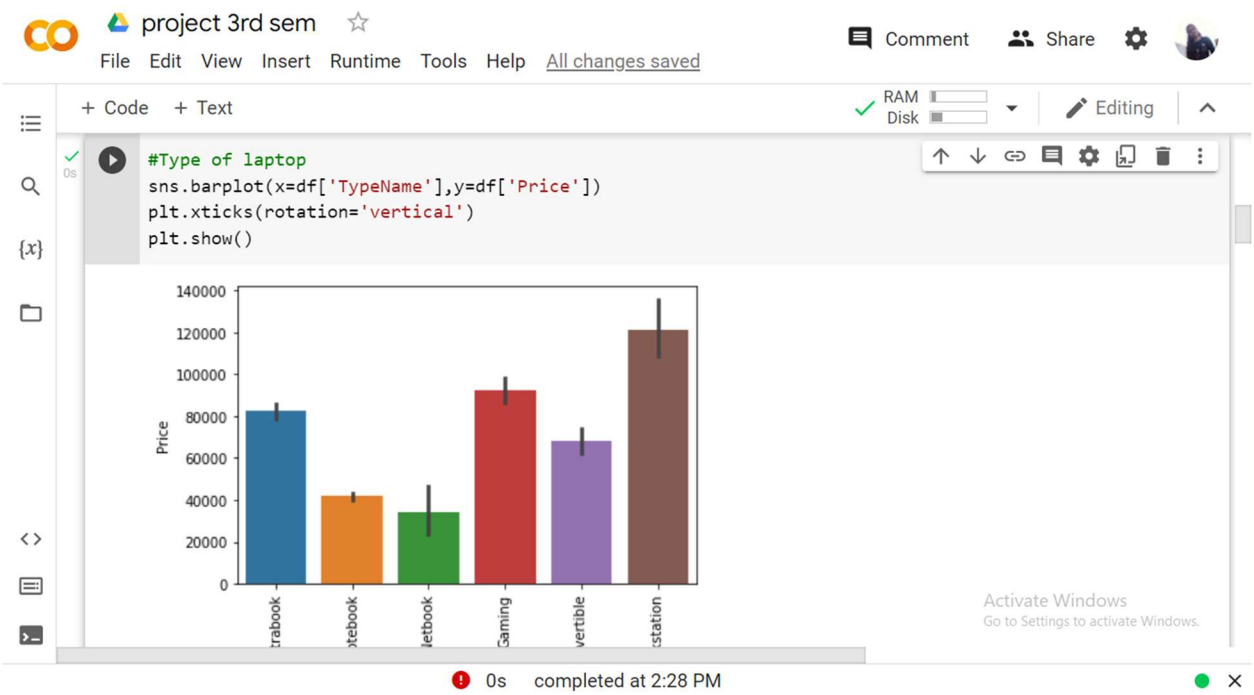
3. what is avg price of each brand?



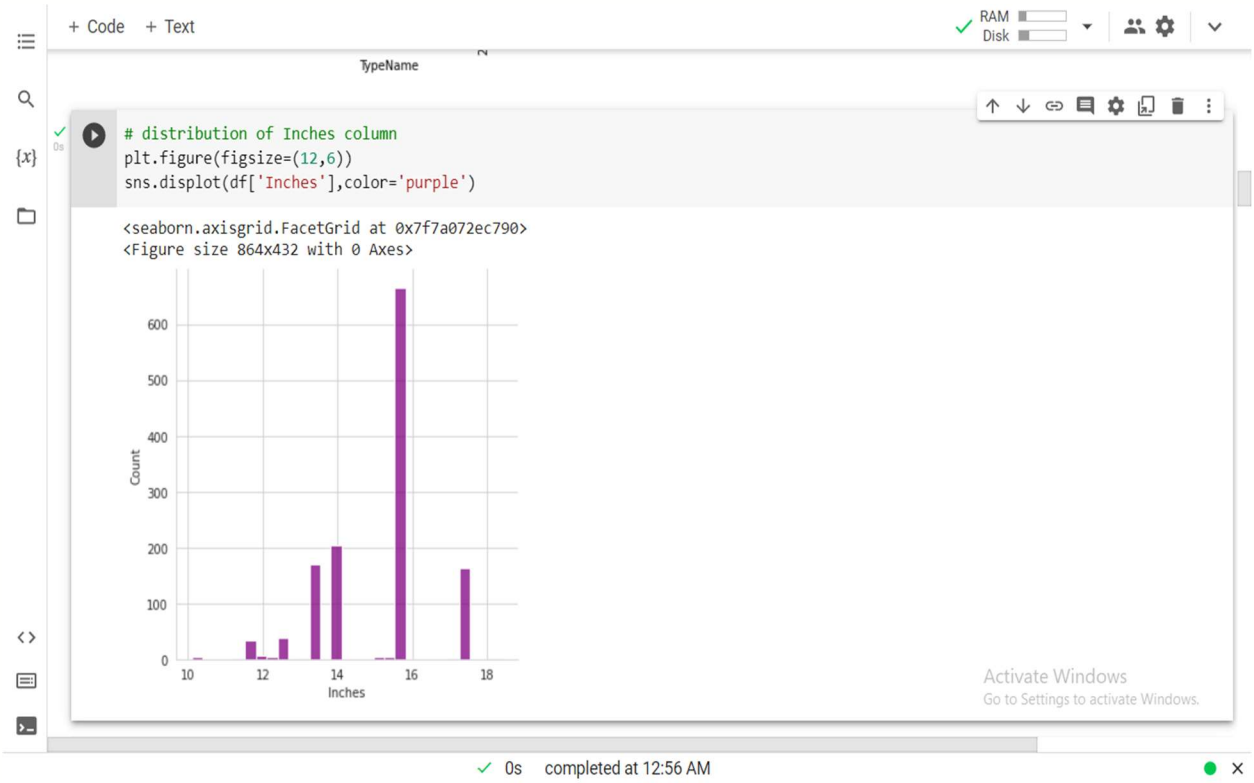
4. Type of laptop



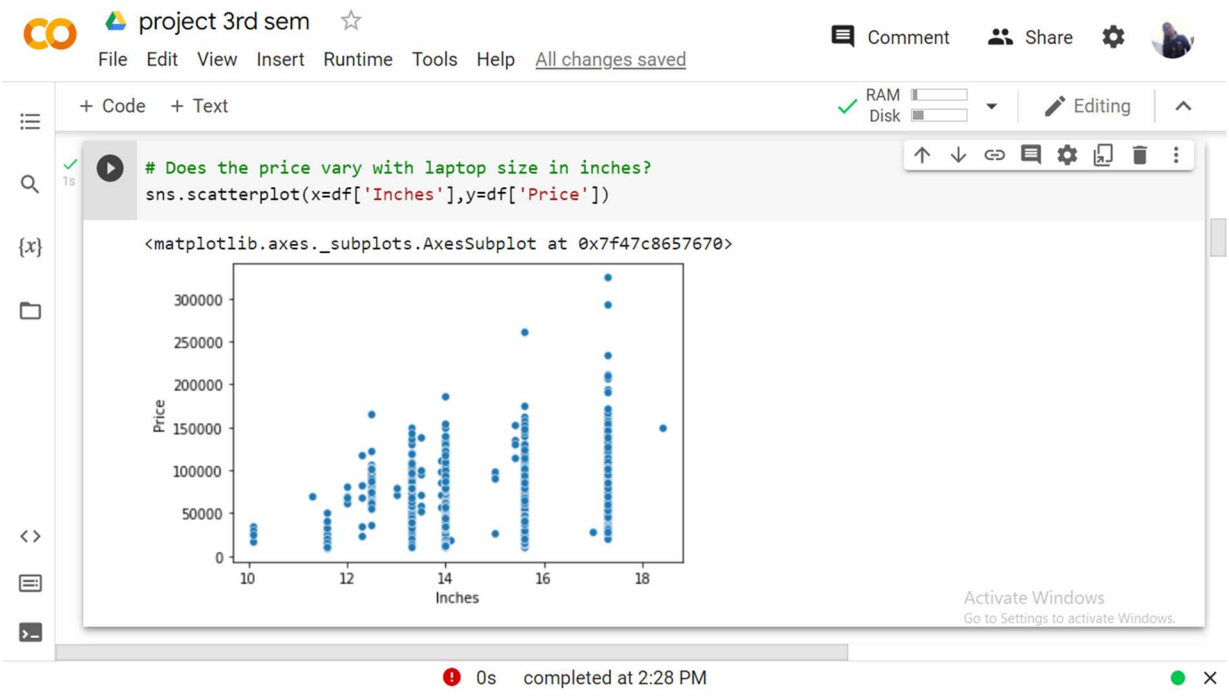
5. Type of laptop and its name



6. Distribution of Inches column



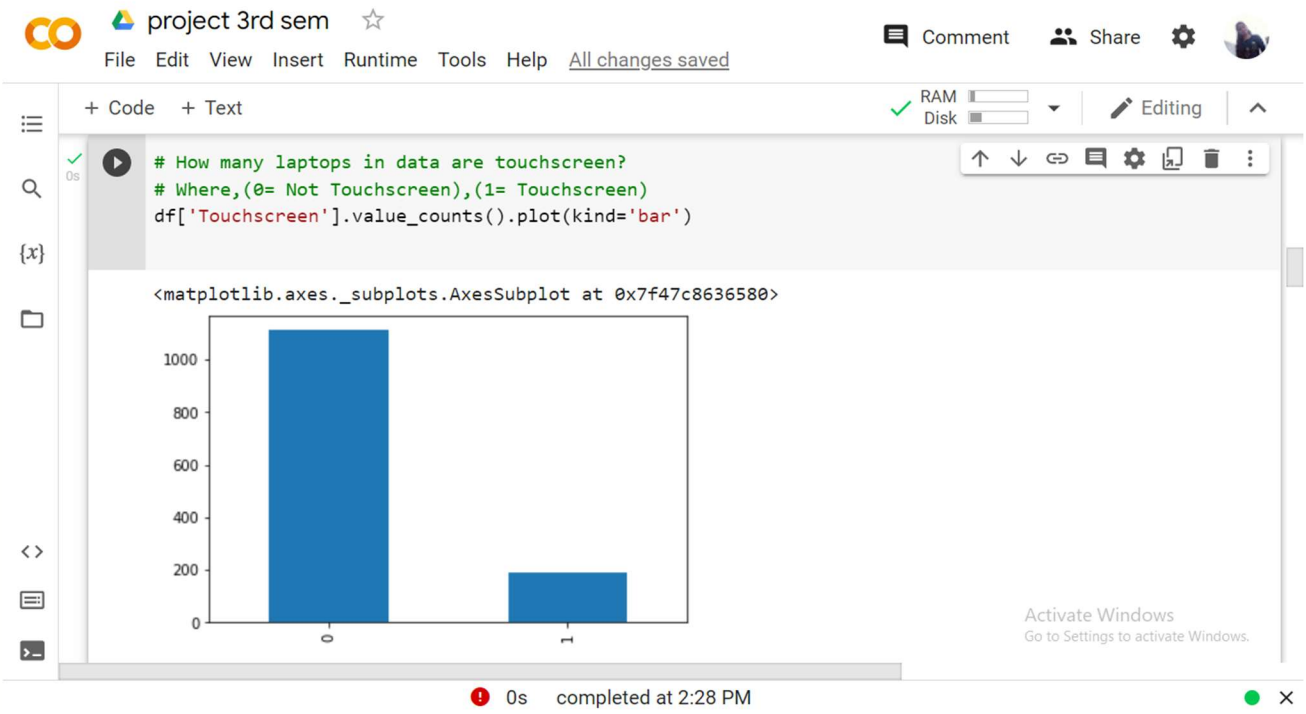
8. Does the price vary with laptop size in inches?



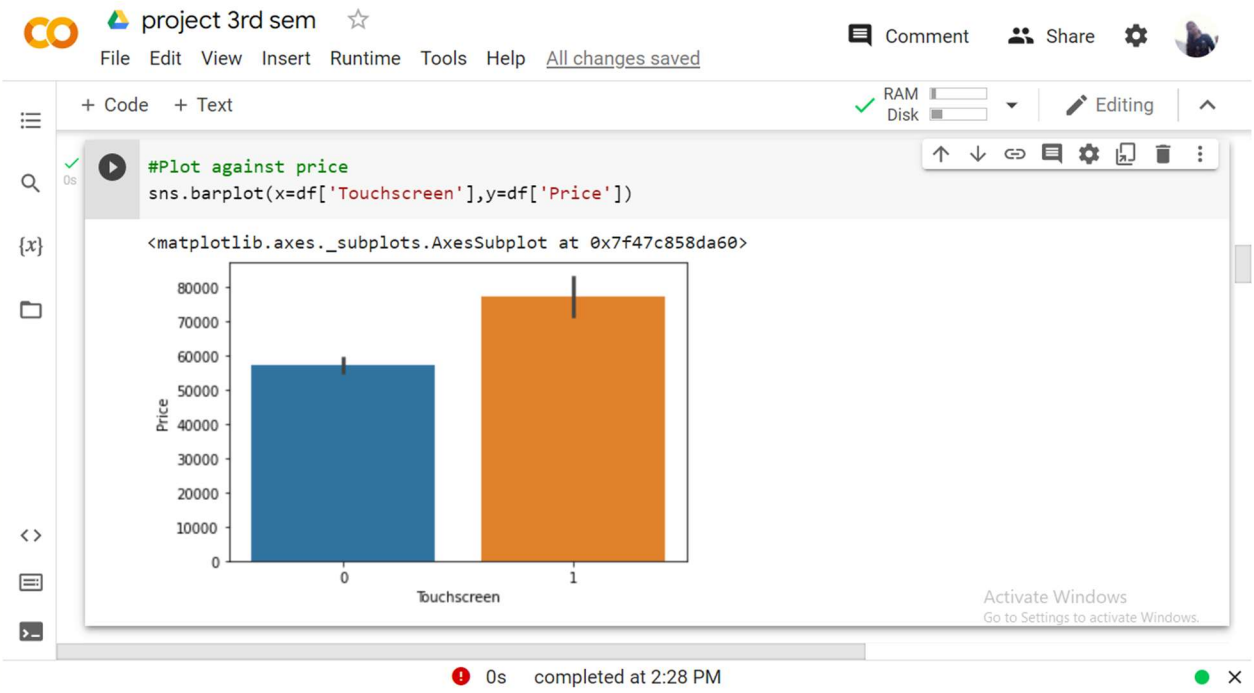
There is a relationship ,but not a strong relationship between the price and Inches

9. How many laptops in data are touchscreen?

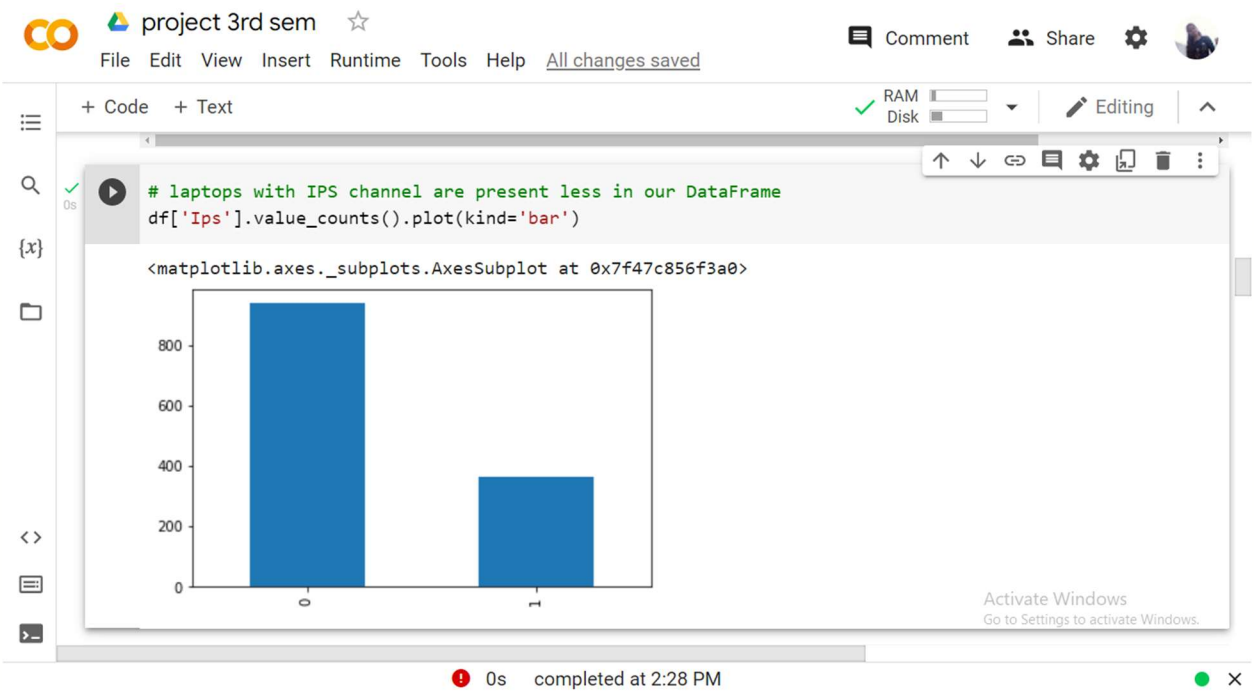
Where,(0= Not Touchscreen),(1= Touchscreen)



9. Touchscreen against price

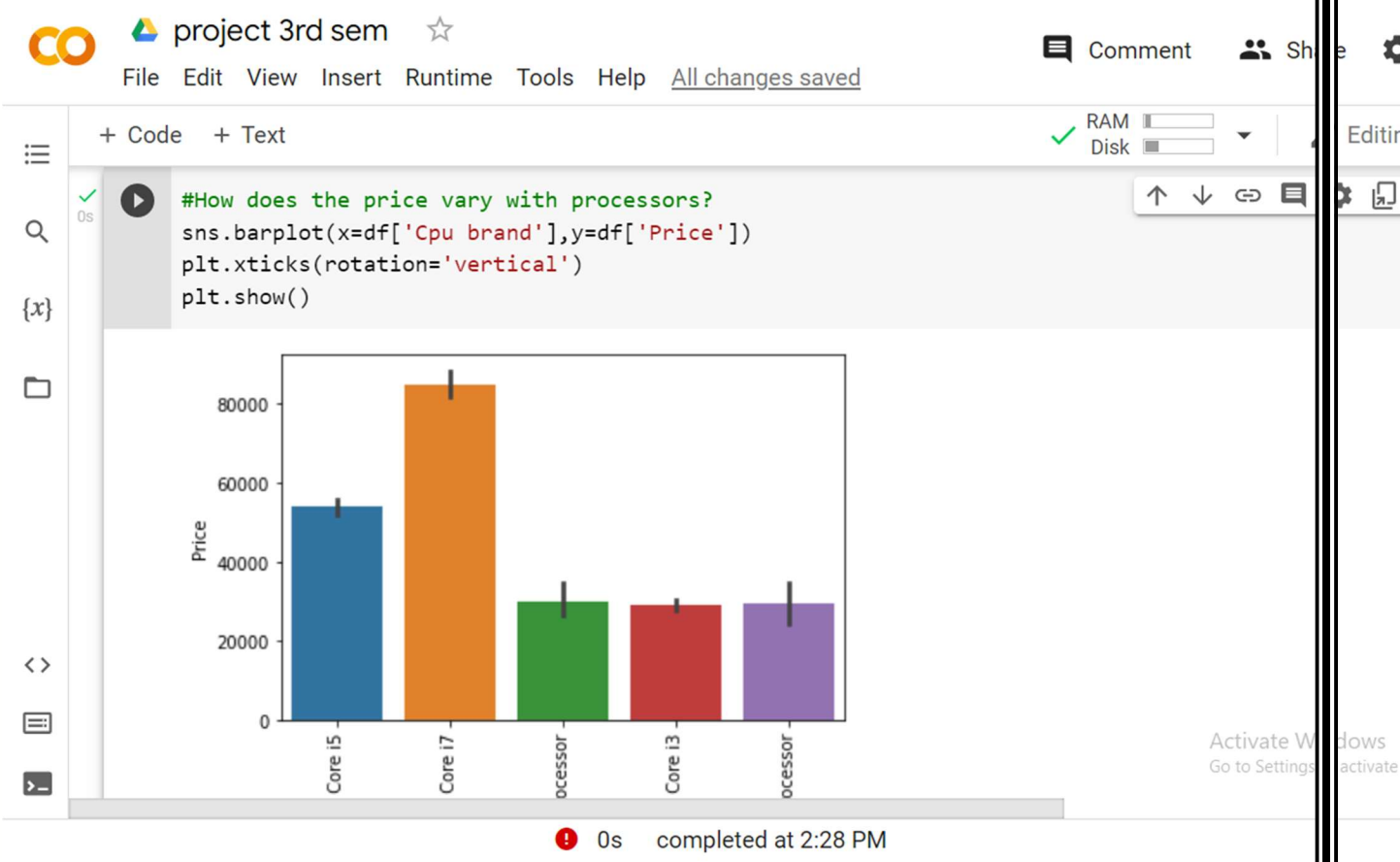


10. laptops with IPS channel are present less in our DataFrame

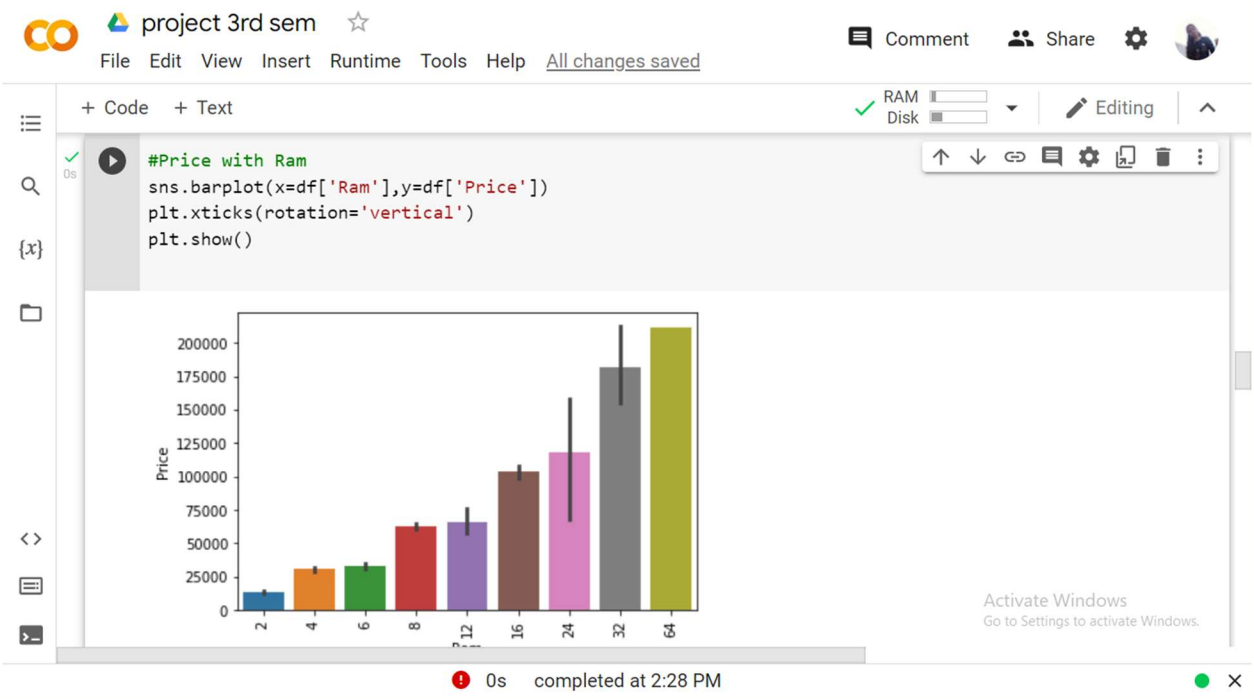


11. Plot IPS against price

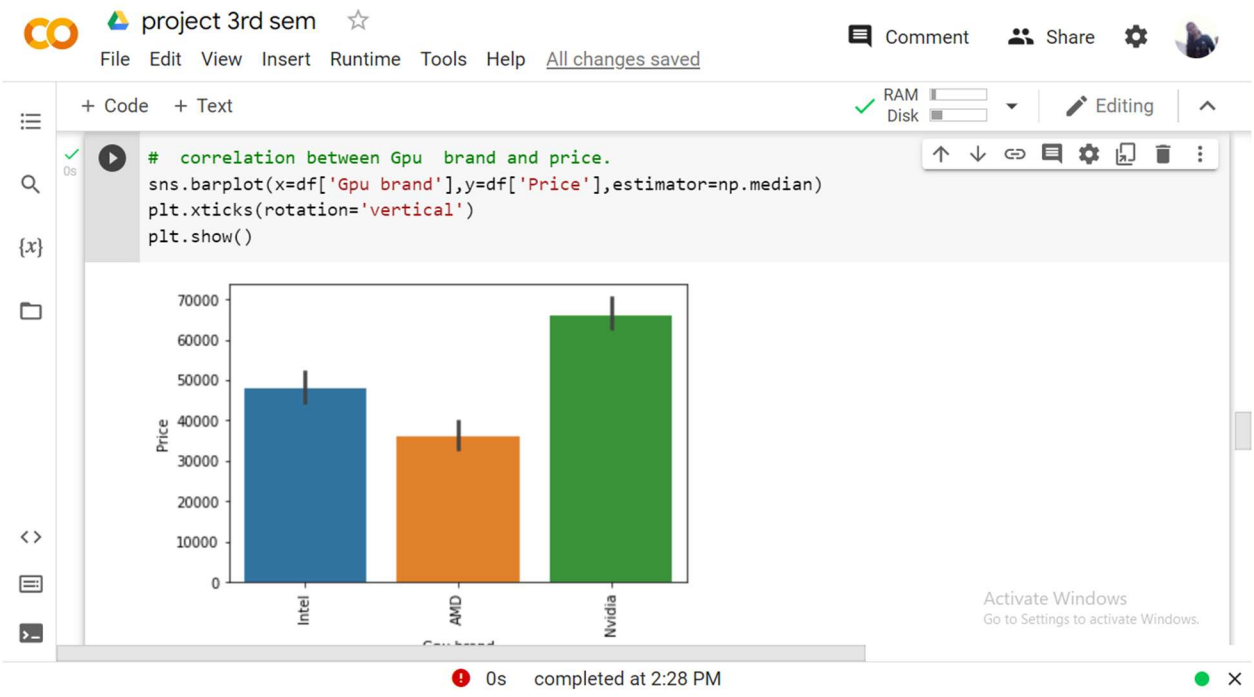
12. How does the price vary with cpu/processor?



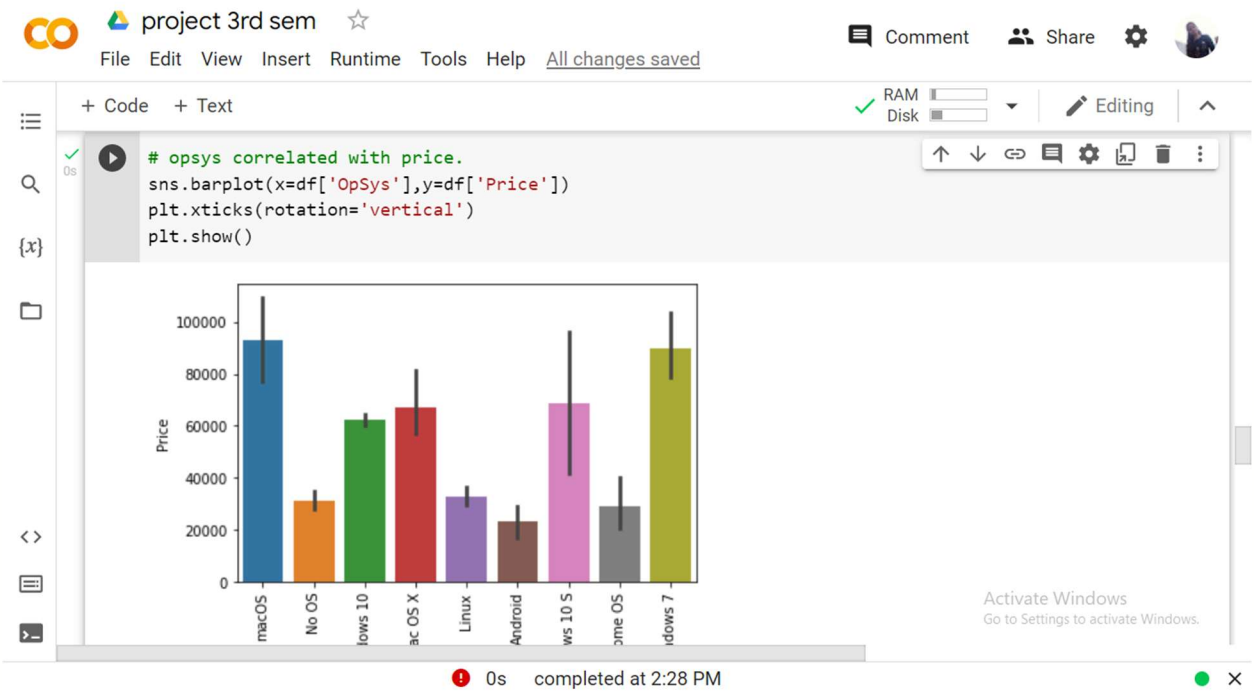
13. Plot Price against Ram



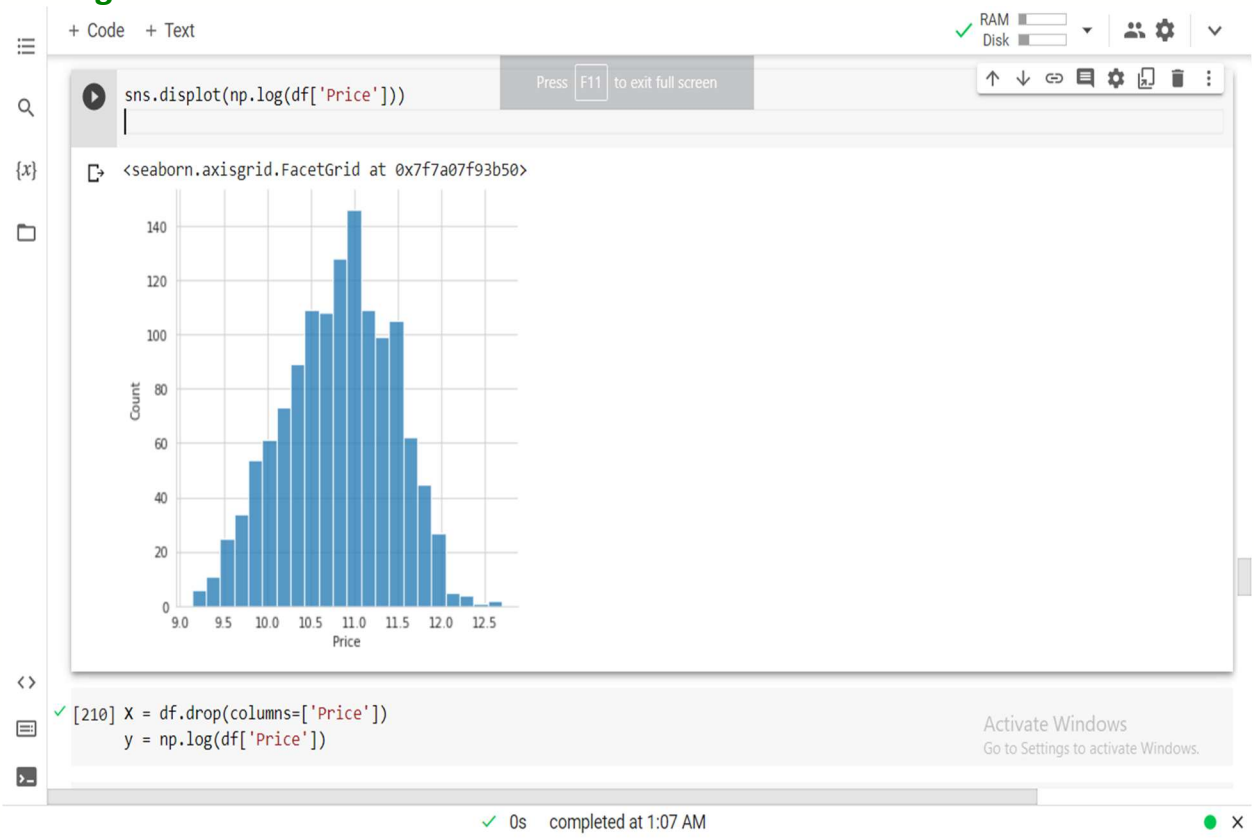
14. correlation between Gpu brand and price.



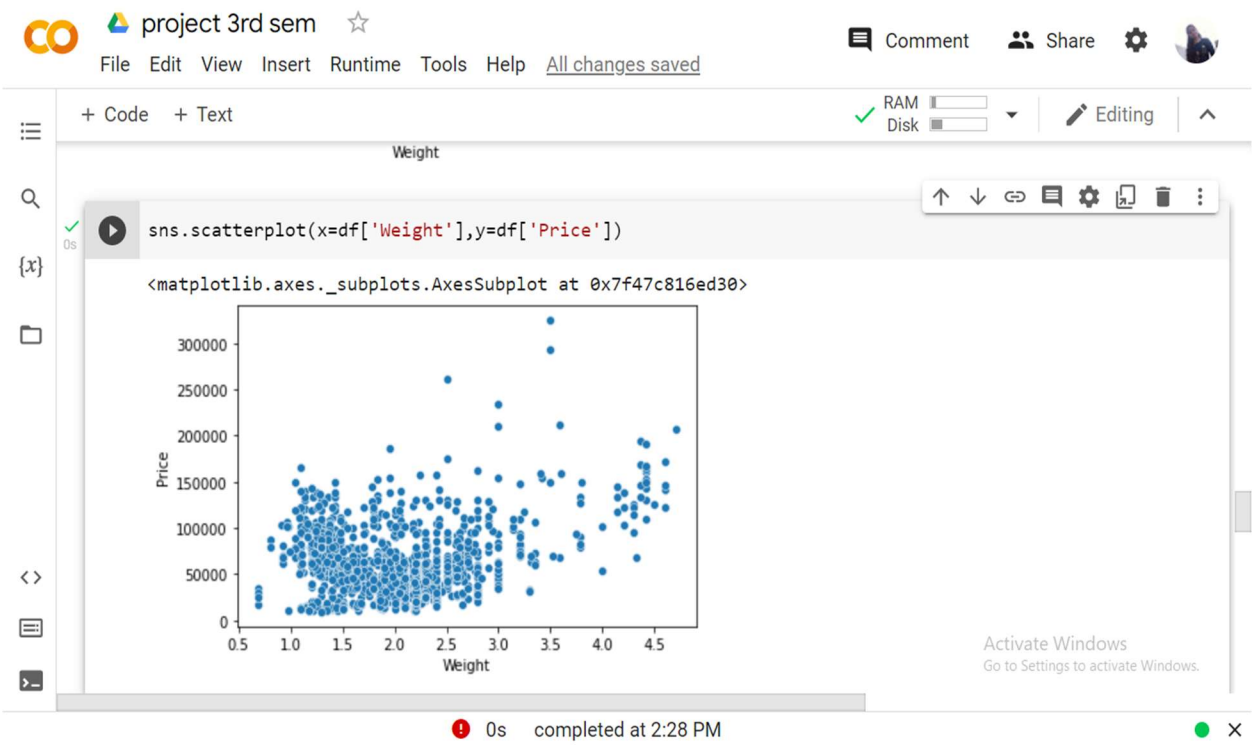
15. operating system correlated with price .



16. Log-Normal Transformation

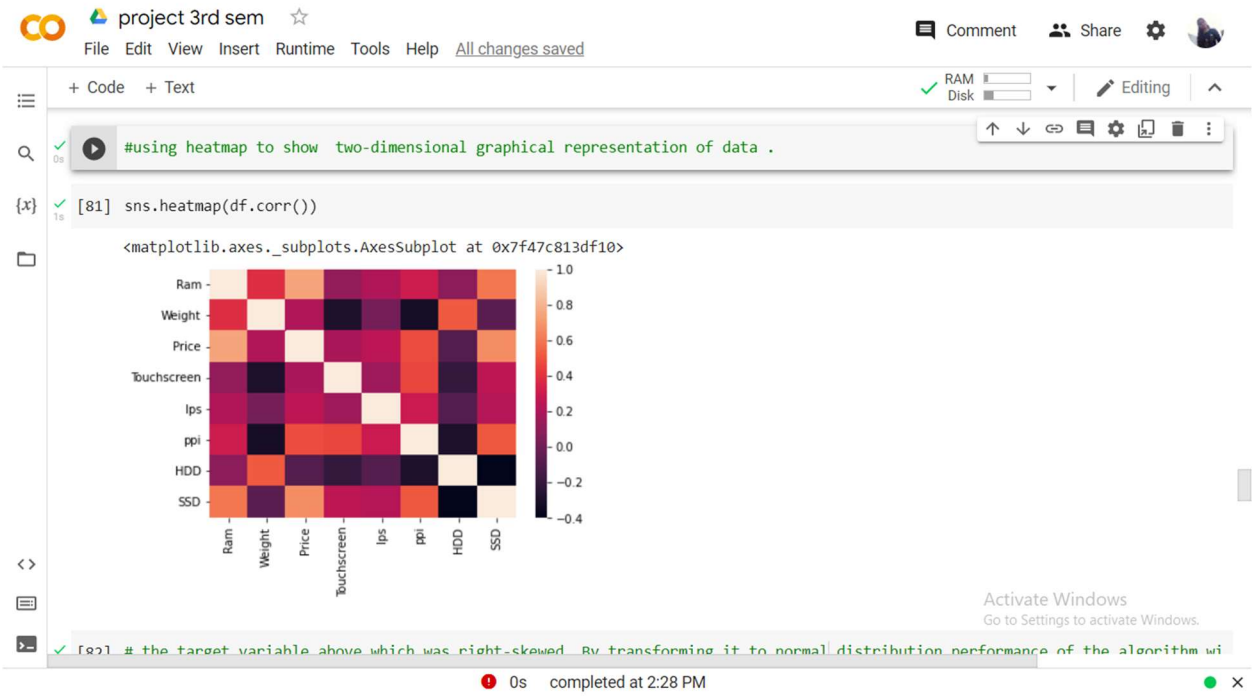


17. correlation between weight and price

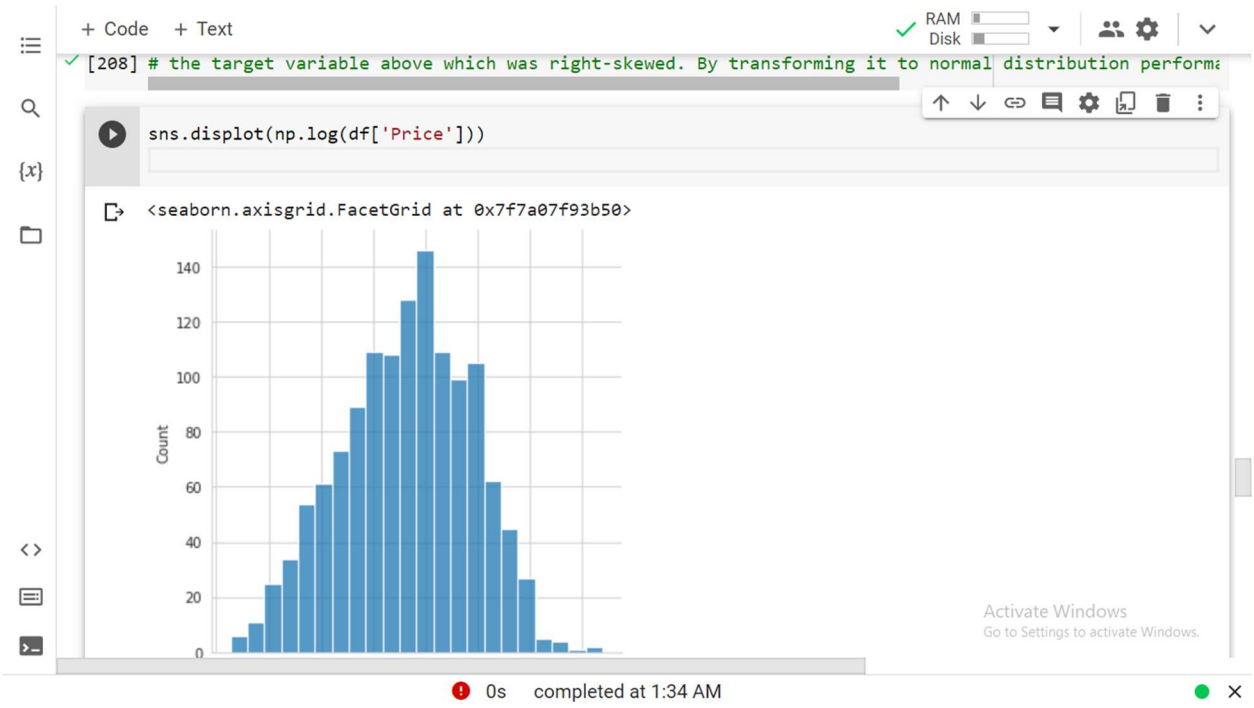


There is a weak correlation between weight and price.

18. using heatmap to show two-dimensional graphical representation of data .



19. **The target variable above which was right-skewed. By transforming it to normal distribution performance of the algorithm will increase.**



----REGRESSION----

Regression is defined as a statistical method that helps us to analyze and understand the relationship between two or more variables of interest. The process that is adapted to perform regression analysis helps to understand which factors are important, which factors can be ignored, and how they are influencing each other.

In regression, we normally have one dependent variable and one or more independent variables. Here we try to "regress" the value of the dependent variable "Y" with the help of the independent variables.

In other words, we are trying to understand, how the value of 'Y' changes w.r.t change in 'X'.

- For the regression analysis is be a successful method, we understand the following terms: **Dependent Variable:** This is the variable that we are trying to understand or forecast.
- **Independent Variable:** These are factors that influence the analysis or target variable and provide us with information regarding the relationship of the variables with the target variable.

1.What is Regression Analysis?

Regression analysis is used for prediction and forecasting. This has substantial overlap with the field of machine learning. This statistical method is used across different industries such as,

- Financial Industry- Understand the trend in the stock prices, forecast the prices, and evaluate risks in the insurance domain
- Marketing- Understand the effectiveness of market campaigns, and forecast pricing and sales of the product.
- Manufacturing- Evaluate the relationship of variables that determine to define a better engine to provide better performance
- Medicine- Forecast the different combinations of medicines to prepare generic medicines for diseases.

----LINEAR REGRESSION----

The simplest of all regression types is Linear Regression which tries to establish relationships between Independent and Dependent variables. The Dependent variable considered here is always a continuous variable.

a predictive model used for finding the linear relationship between a dependent variable and one or more independent variables.Here, 'Y' is our dependent variable, which is a continuous numerical and we are trying to understand how 'Y' changes with 'X'.

So, if we are supposed to answer, the above question of "What will be the GRE score of the student, if his CCGPA is 8.32?" our go-to option should be linear regression.

Examples of Independent & Dependent Variables*:*

• Here x is Rainfall and y is Crop Yield

• Secondly, x is Advertising Expense and y is Sales

• At last, x is sales of goods and y is GDP

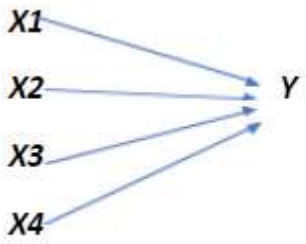
If the relationship with the dependent variable is in the form of single variables, then it is known as Simple Linear Regression

Simple Linear Regression

X —–> Y

If the relationship between Independent and dependent variables is multiple in number, then it is called Multiple Linear Regression

Multiple Linear Regression



Simple Linear Regression Model

As the model is used to predict the dependent variable, the relationship between the variables can be written in the below format.

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Where,

Yi – Dependent variable

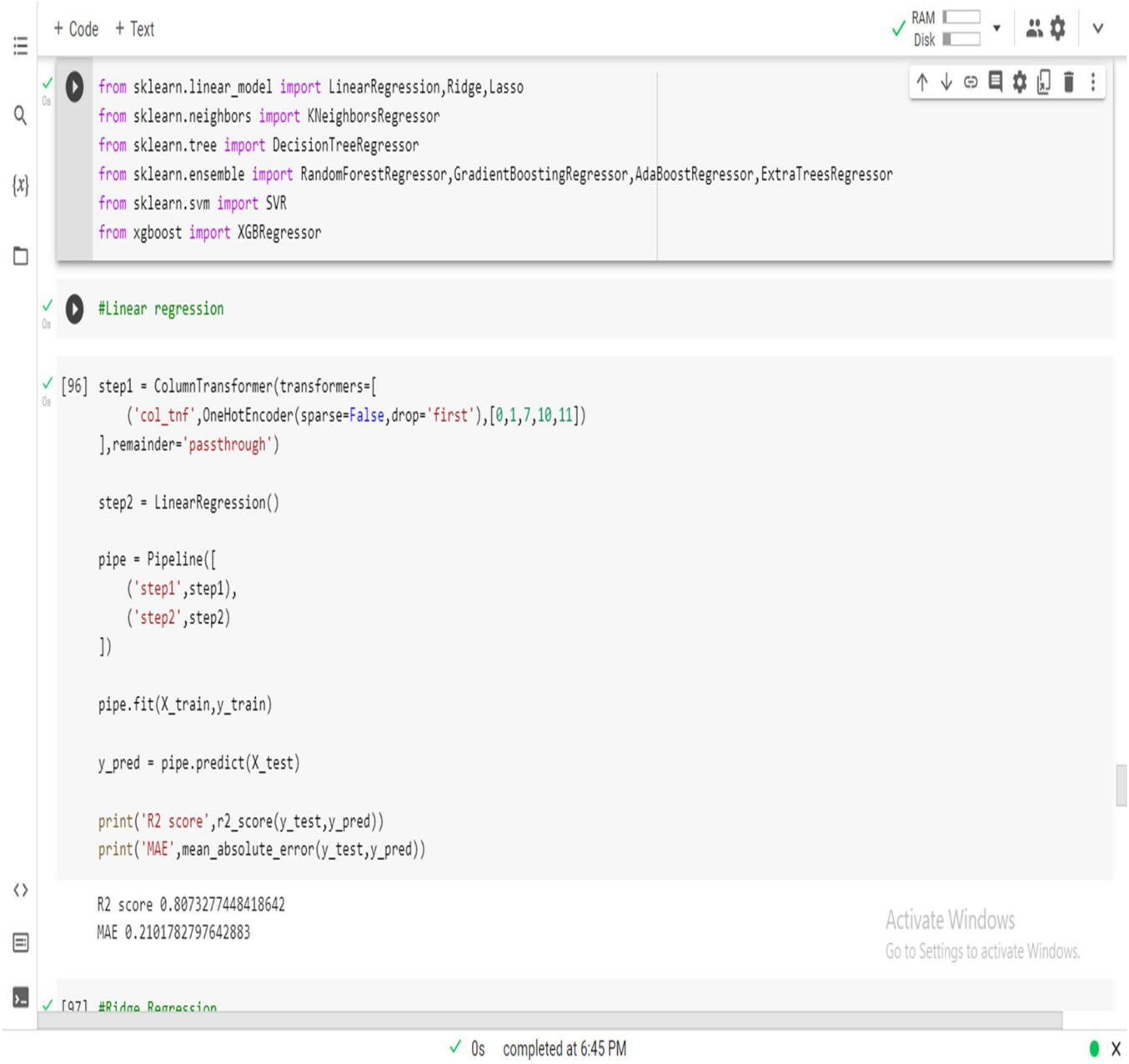
β0 -- Intercept

β1 – Slope Coefficient

Xi – Independent Variable

εi – Random Error Term

Implement Pipeline for training and testing



```
+ Code  + Text                                                    RAM  Disk

from sklearn.linear_model import LinearRegression,Ridge,Lasso
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor,GradientBoostingRegressor,AdaBoostRegressor,ExtraTreesRegressor
from sklearn.svm import SVR
from xgboost import XGBRegressor

#Linear regression

[96] step1 = ColumnTransformer(transformers=[
        ('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
    ],remainder='passthrough')

    step2 = LinearRegression()

    pipe = Pipeline([
        ('step1',step1),
        ('step2',step2)
    ])

    pipe.fit(X_train,y_train)

    y_pred = pipe.predict(X_test)

    print('R2 score',r2_score(y_test,y_pred))
    print('MAE',mean_absolute_error(y_test,y_pred))

    R2 score 0.8073277448418642
    MAE 0.2101782797642883

[97] #Ridge Regression

✓ 0s  completed at 6:45 PM
```

----Regularization----

Regularization is an important concept that is used to avoid overfitting of the data, especially when the trained and test data are much varying.
Regularization is implemented by adding a “penalty” term to the best fit derived from the trained data, to achieve a *lesser variance* with the tested data and also restricts the influence of predictor variables over the output variable by compressing their coefficients.

In regularization, what we do is normally we keep the same number of features but reduce the magnitude of the coefficients. We can reduce the magnitude of the coefficients by using different types of regression techniques which uses regularization to overcome this problem.

There are two main regularization techniques, namely Ridge Regression and Lasso Regression. They both differ in the way they assign a penalty to the coefficients.
In this blog, we will try to understand more about Lasso Regularization technique.

----Lasso regression----

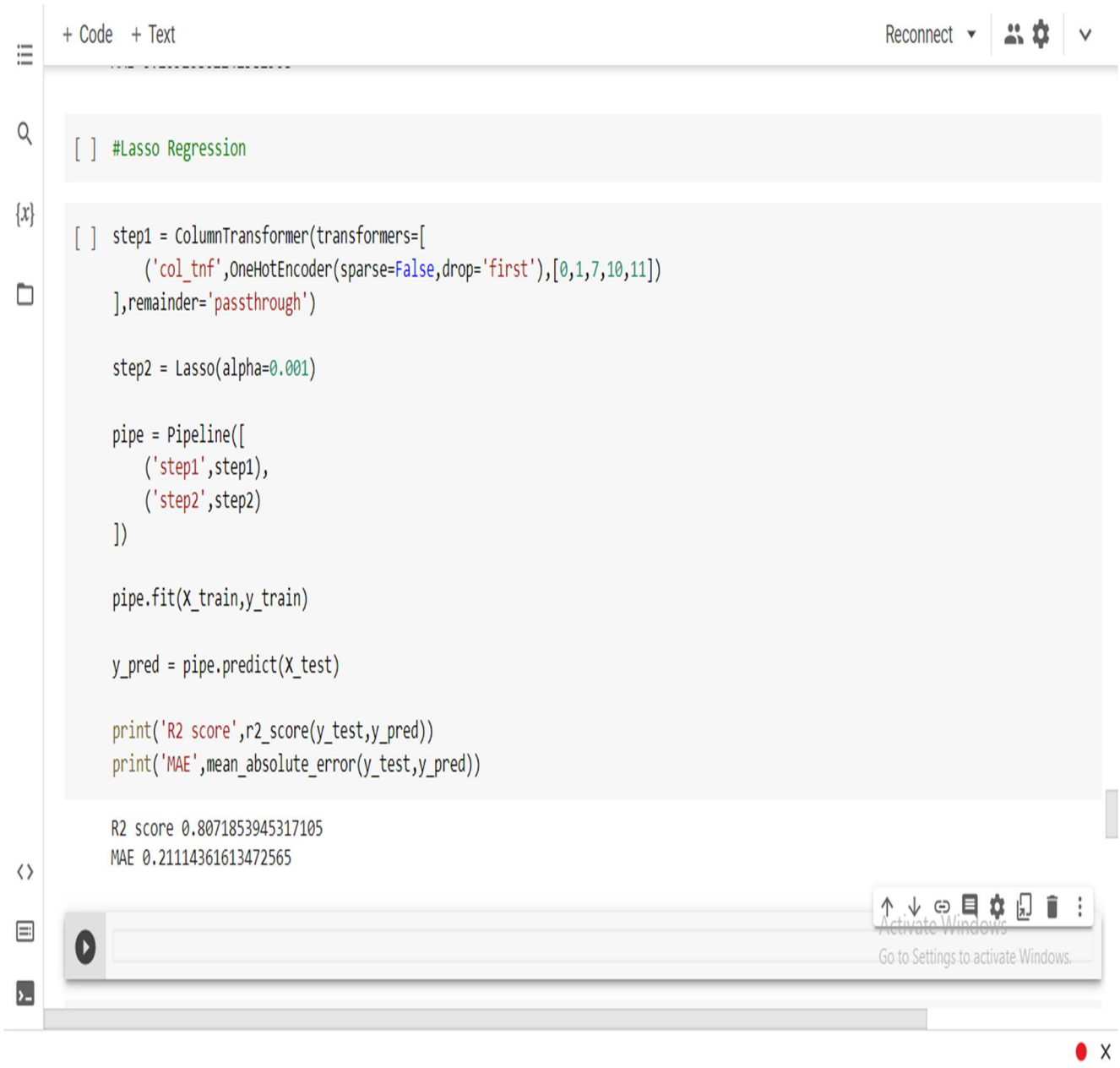
Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

Lasso Regression uses L1 regularization technique . It is used when we have more features because it automatically performs feature selection. The word “LASSO” stands for Least Absolute Shrinkage and Selection Operator. It is a statistical formula for the regularisation of data models and feature selection.
The cost function for lasso regression

Min(||Y - X(theta)||^2 + λ||theta||)

λ is the hypermeter, whose value is equal to the alpha in the Lasso function

It is generally used when we have more number of features because it automatically does feature selection.


+ Code
+ Text
Reconnect
[ ] #Lasso Regression
[ ] step1 = ColumnTransformer(transformers=[
('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')
step2 = Lasso(alpha=0.001)
pipe = Pipeline([
('step1',step1),
('step2',step2)
])
pipe.fit(X_train,y_train)
y_pred = pipe.predict(X_test)
print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
R2 score 0.8071853945317105
MAE 0.21114361613472565
Activate Windows
Go to Settings to activate Windows.

----Ridge regression----

Ridge regression is a model tuning method that is used to analyse any data that suffers from multicollinearity. This method performs L2 regularization. When the issue of multicollinearity occurs, least-squares are unbiased, and variances are large, this results in predicted values being far away from the actual values.

The cost function for ridge regression:

Min(||Y – X(theta)||^2 + λ||theta||^2)

Lambda is the penalty term. λ given here is denoted by an alpha parameter in the ridge function. So, by changing the values of alpha, we are controlling the penalty term. The higher the values of alpha, the bigger is the penalty and therefore the magnitude of coefficients is reduced.

It shrinks the parameters. Therefore, it is used to prevent multicollinearity

It reduces the model complexity by coefficient shrinkage

Check out the free course on regression analysis.

For any type of regression machine learning model, the usual regression equation forms the base which is written as:

Y = XB + e

Where Y is the dependent variable, X represents the independent variables, B is the regression coefficients to be estimated, and e represents the errors are residuals.

Once we add the lambda function to this equation, the variance that is not evaluated by the general model is considered

Assumptions of Ridge Regressions

The assumptions of ridge regression are the same as that of linear regression: linearity, constant variance, and independence. However, as ridge regression does not provide confidence limits, the distribution of errors to be normal need not be assumed.

----Random forest----

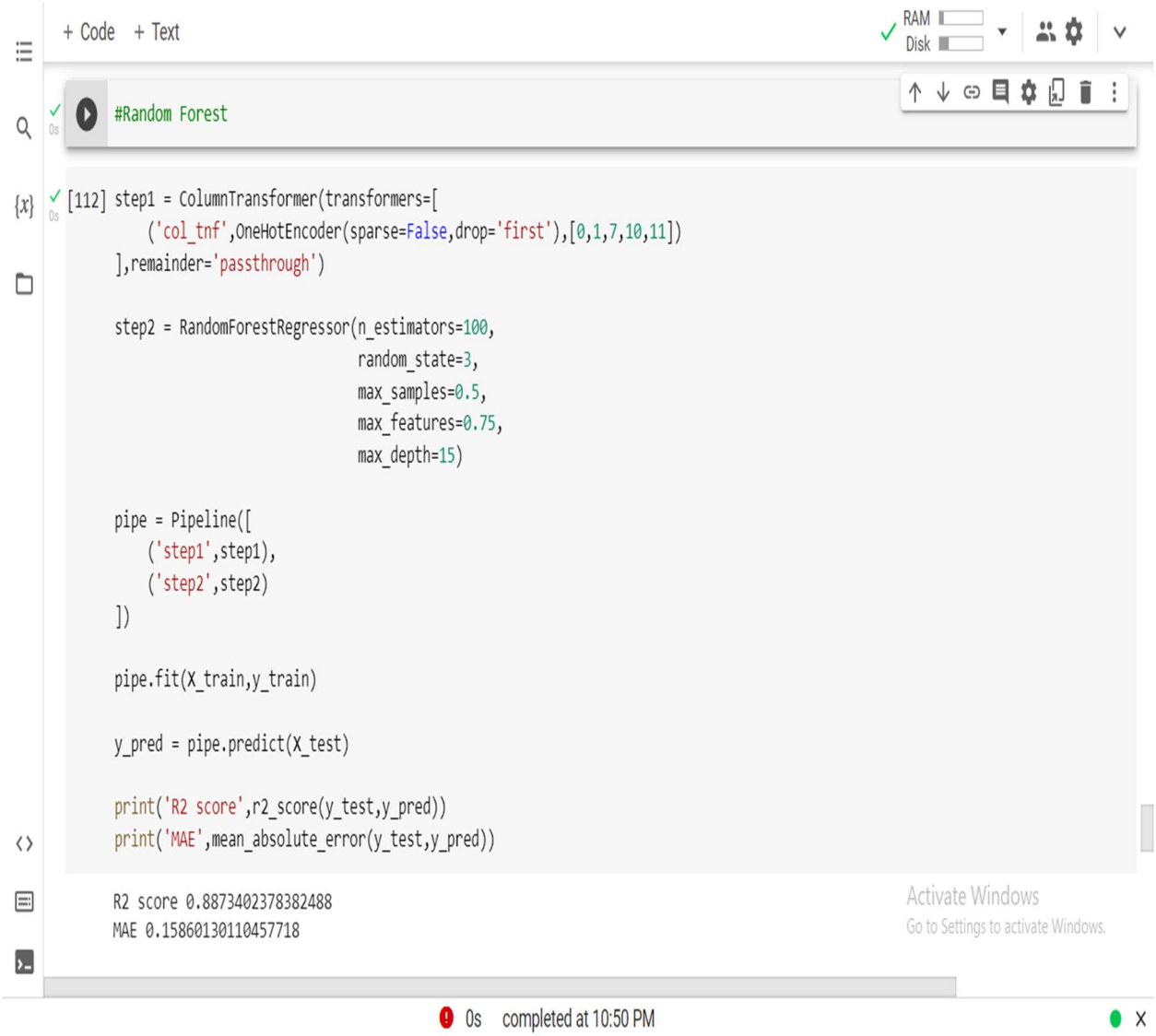
Random forests (Breiman, 2001, Machine Learning 45: 5–32) is a statistical- or machine-learning algorithm for prediction.
The random forest algorithm is made up of a collection of decision trees, and each tree in the ensemble is comprised of a data sample drawn from a training set with replacement, called the bootstrap sample.

Random Forest is a supervised machine learning algorithm made up of decision trees. Random Forest is used for both classification and regression—for example, classifying whether an email is “spam” or “not spam”.

It can perform both regression and classification tasks. A random forest produces good predictions that can be understood easily. It can handle large datasets efficiently. The random forest algorithm provides a higher level of accuracy in predicting outcomes over the decision tree algorithm.

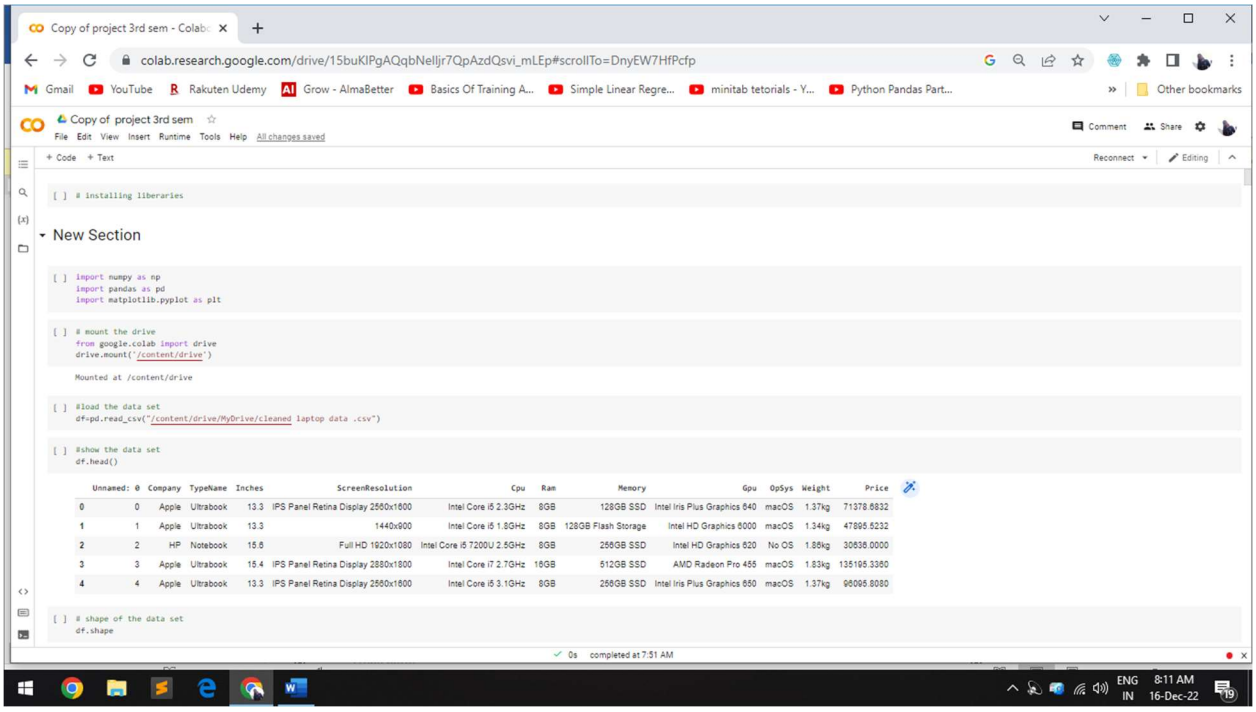
Random forests consist of multiple single trees each based on a random sample of the training data. They are typically more accurate than single decision trees. The following figure shows the decision boundary becomes more accurate and stable

as more trees are added.

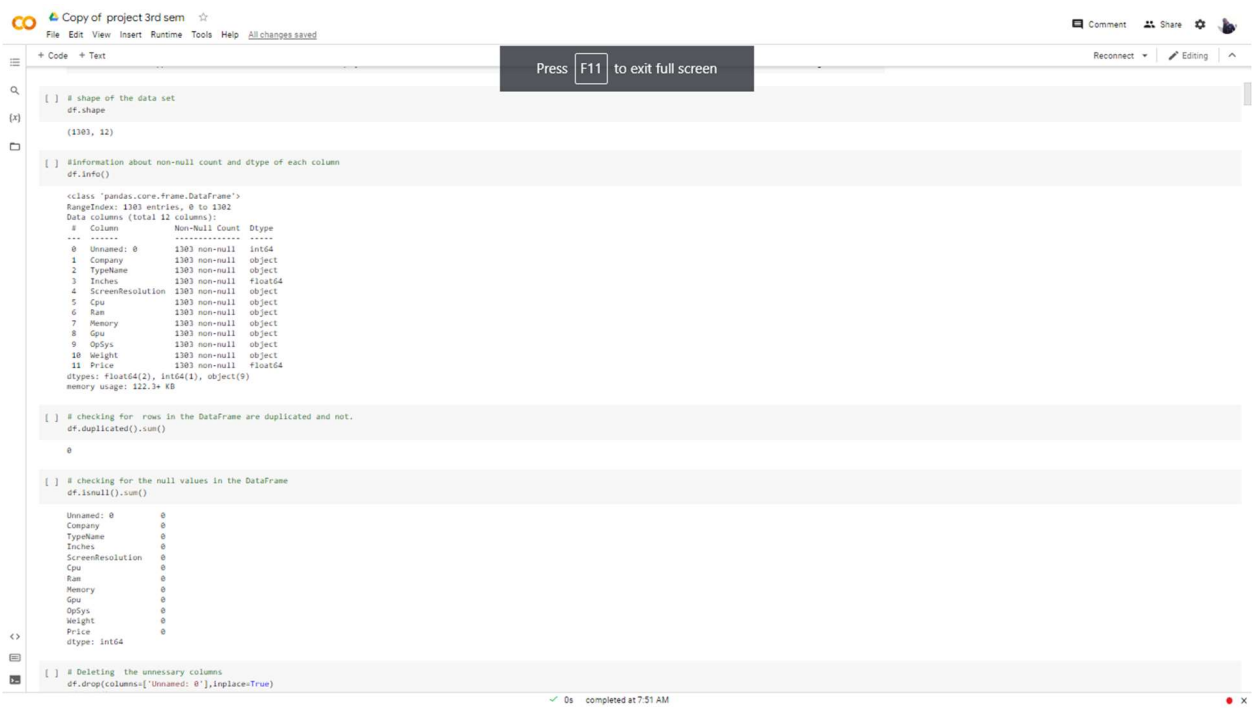


---CHAPTER-4---

--- NOTEBOOK---



	Unnamed: 0	Company	TypeName	Inches	ScreenResolution	Cpu	Ram	Memory	Gpu	OpSys	Weight	Price
0	0	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 2.3GHz	8GB	128GB SSD	Intel Iris Plus Graphics 640	macOS	1.37kg	71378.6832
1	1	Apple	Ultrabook	13.3	1440x900	Intel Core i5 1.8GHz	8GB	128GB Flash Storage	Intel HD Graphics 6000	macOS	1.34kg	47895.5232
2	2	HP	Notebook	15.6	Full HD 1920x1080	Intel Core i5 7200U 2.5GHz	8GB	256GB SSD	Intel HD Graphics 620	No OS	1.86kg	30636.0000
3	3	Apple	Ultrabook	15.4	IPS Panel Retina Display 2880x1800	Intel Core i7 2.7GHz	16GB	512GB SSD	AMD Radeon Pro 455	macOS	1.83kg	135195.3360
4	4	Apple	Ultrabook	13.3	IPS Panel Retina Display 2560x1600	Intel Core i5 3.1GHz	8GB	256GB SSD	Intel Iris Plus Graphics 650	macOS	1.37kg	96095.8080



Copy of project 3rd sem

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

Press F11 to exit full screen

Reconnect Editing

```
[ ] # shape of the data set
    df.shape
```

```
(1303, 12)
```

```
[ ] #information about non-null count and dtype of each column
    df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1303 entries, 0 to 1302
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   Unnamed: 0        1303 non-null   int64  
 1   Company           1303 non-null   object 
 2   TypeName          1303 non-null   object 
 3   Inches            1303 non-null   float64
 4   ScreenResolution  1303 non-null   object 
 5   Cpu               1303 non-null   object 
 6   Ram               1303 non-null   object 
 7   Memory            1303 non-null   object 
 8   Gpu               1303 non-null   object 
 9   OpSys             1303 non-null   object 
 10  Weight            1303 non-null   object 
 11  Price             1303 non-null   float64
dtypes: float64(2), int64(1), object(9)
memory usage: 122.3+ KB
```

```
[ ] # checking for  rows in the DataFrame are duplicated and not.
    df.duplicated().sum()
```

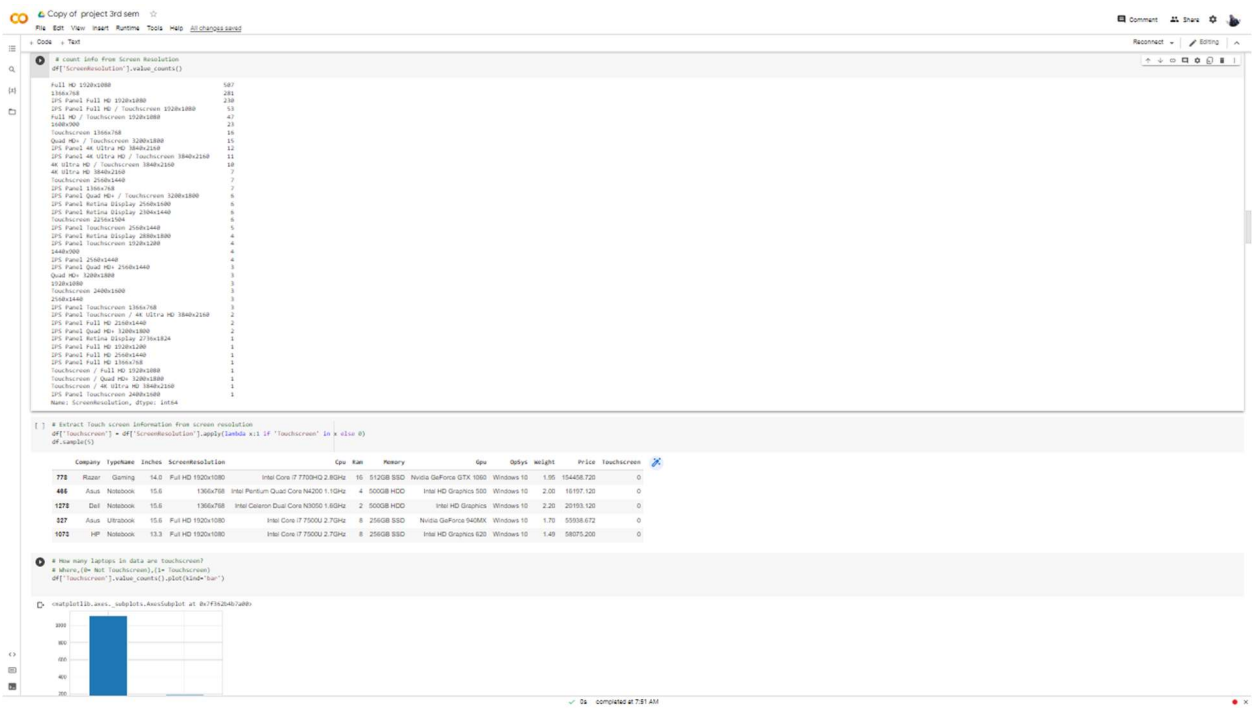
```
0
```

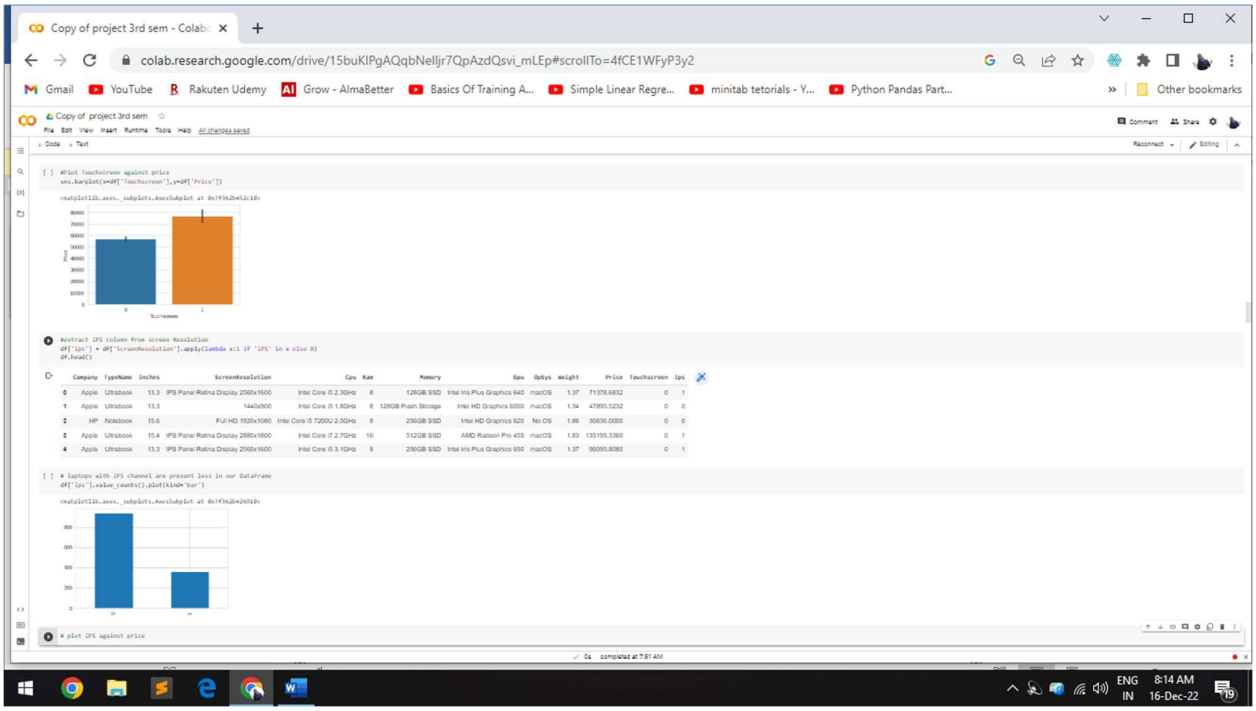
```
[ ] # checking for the null values in the DataFrame
    df.isnull().sum()
```

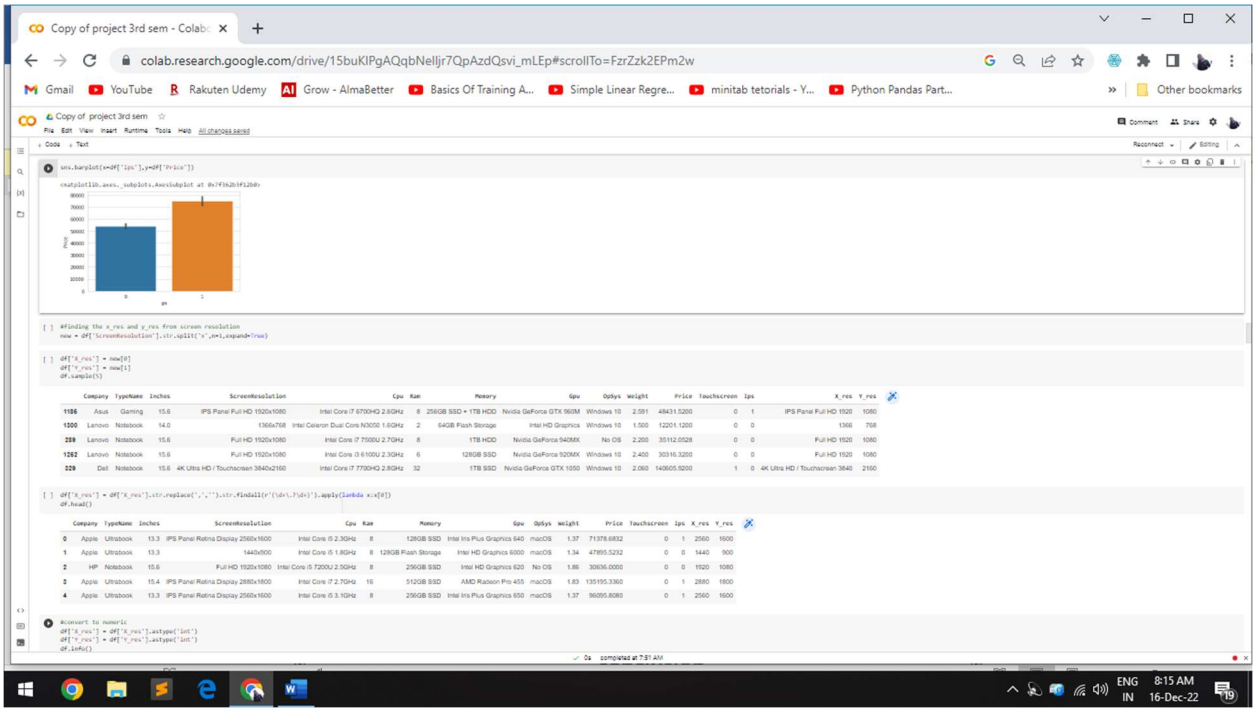
```
Unnamed: 0          0
Company             0
TypeName            0
Inches              0
ScreenResolution    0
Cpu                 0
Ram                 0
Memory              0
Gpu                 0
OpSys               0
Weight              0
Price               0
dtype: int64
```

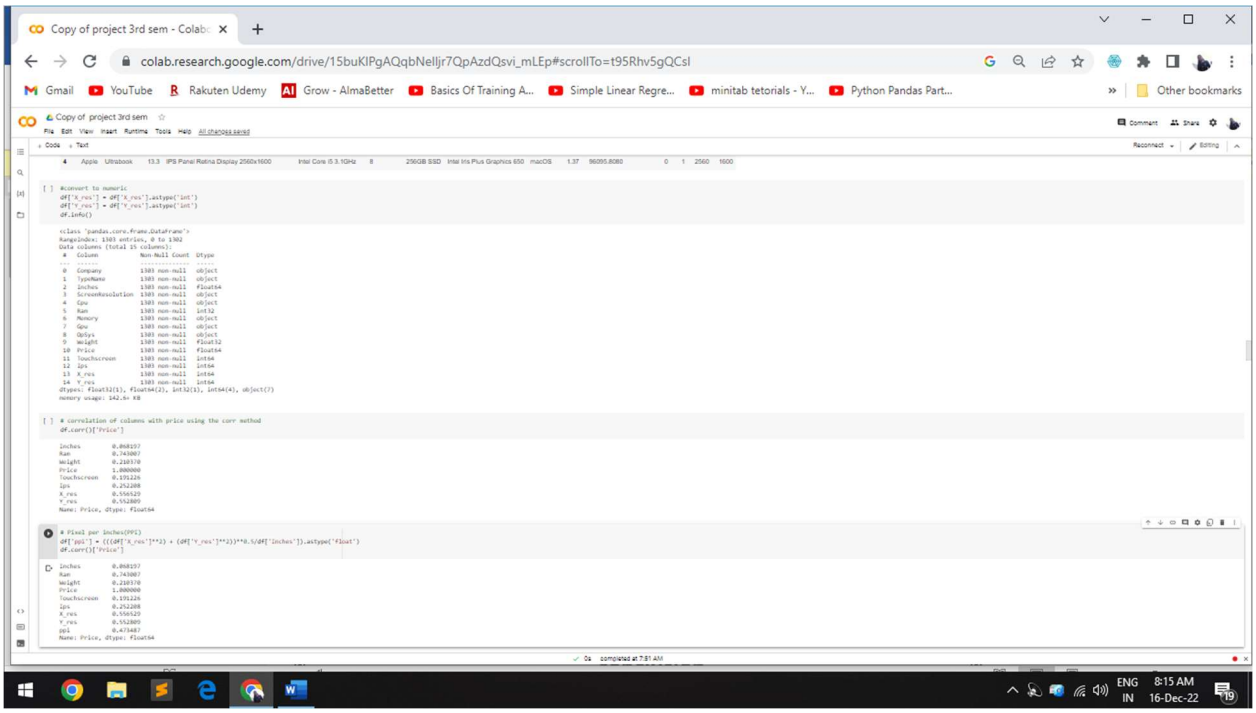
```
[ ] # Deleting  the unnessary columns
    df.drop(columns=['Unnamed: 0'],inplace=True)
```

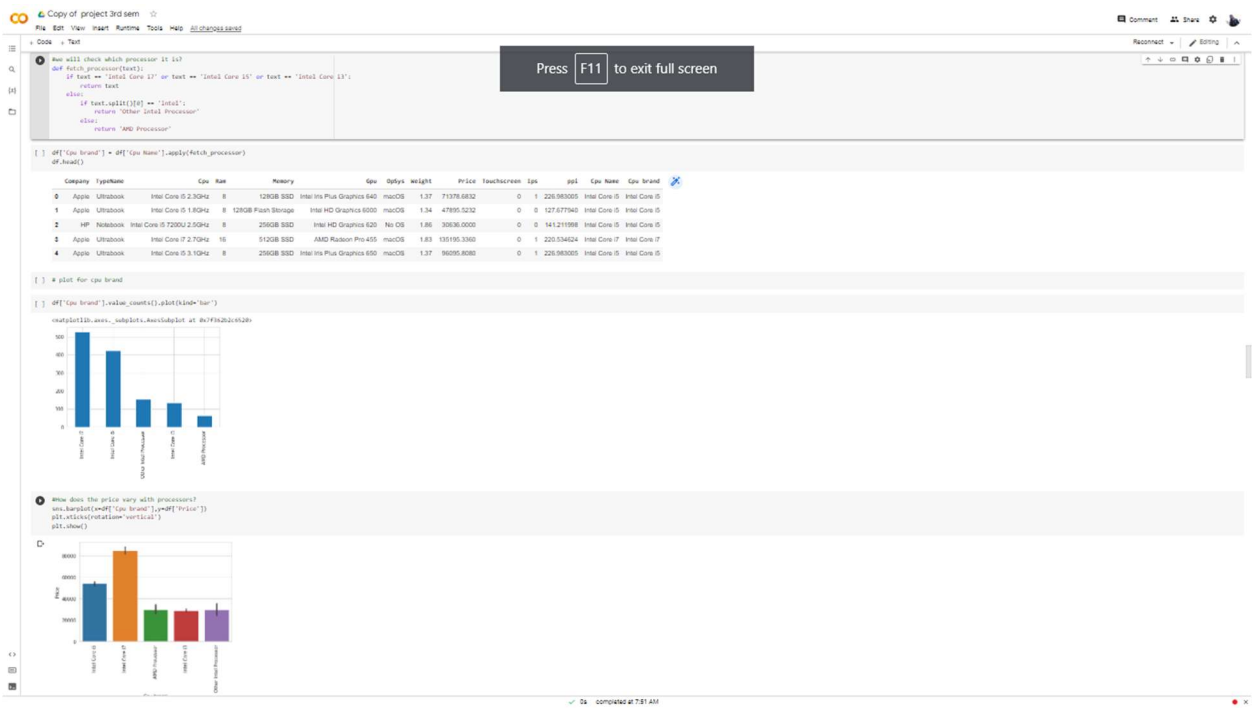
0s completed at 7:51 AM

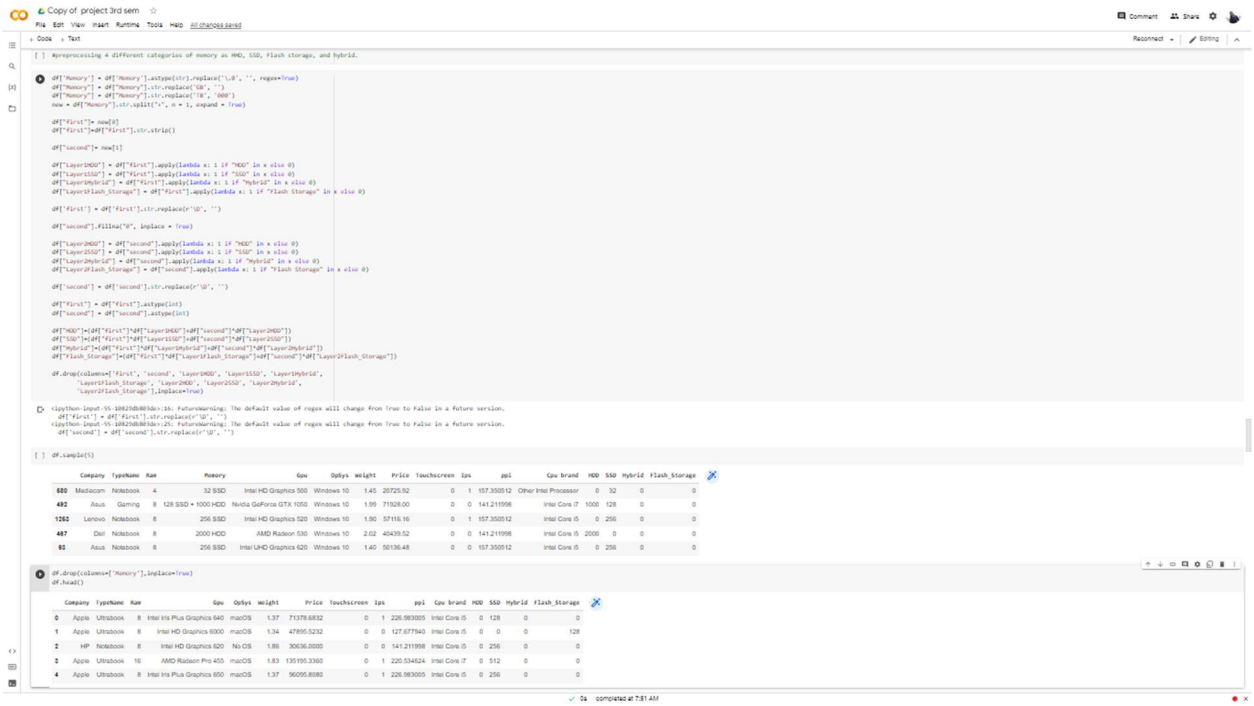


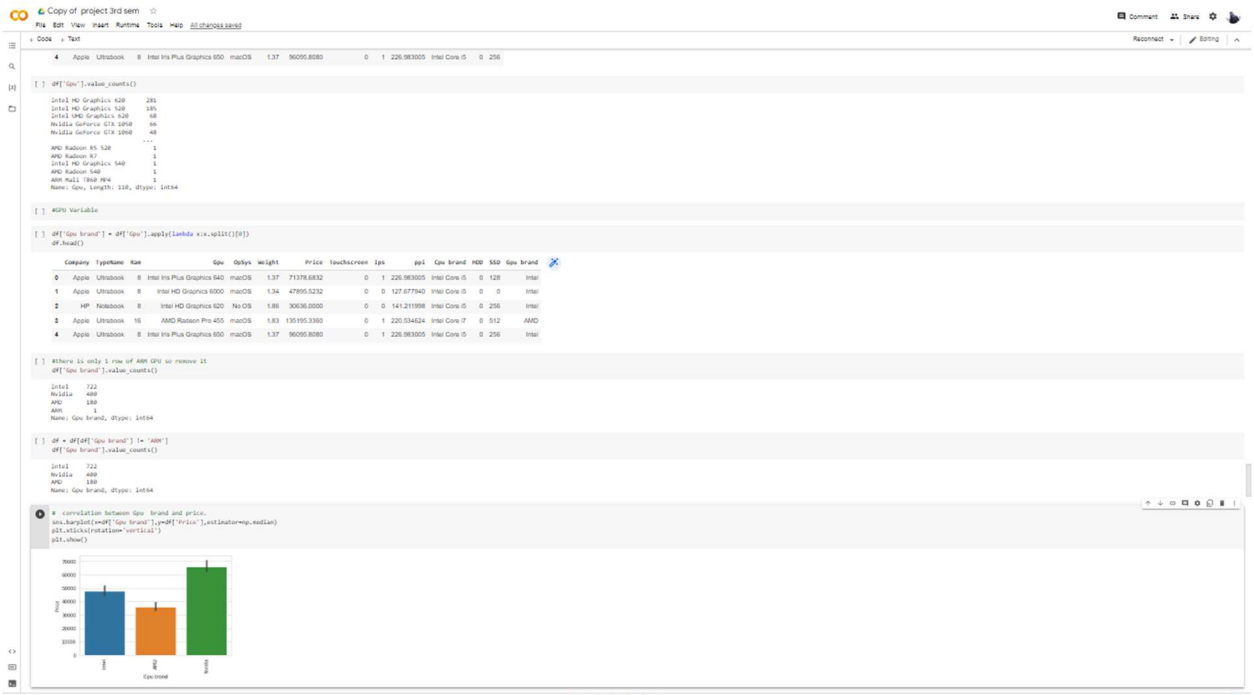


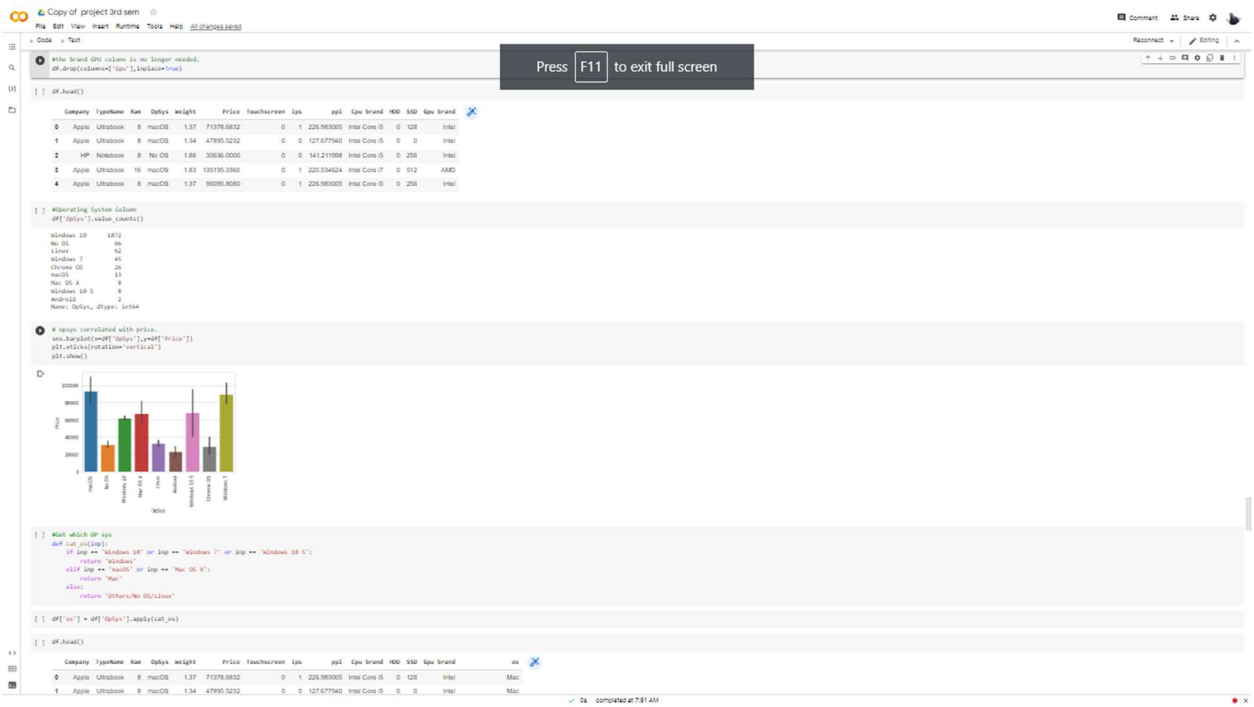

Copy of project 3rd sem - Colab
colab.research.google.com/drive/15buKlPgAQqbNelljr7QpAzdQsvi_mLEp#scrollTo=FzrZzk2EPm2w
Gmail
YouTube
Rakuten Udemy
Grow - AlmaBetter
Basics Of Training A...
Simple Linear Regre...
minitab tetorials - Y...
Python Pandas Part...
Other bookmarks
Copy of project 3rd sem
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Reconnect
Editing
sns.barplot(x=df['ips'],y=df['Price'])
matplotlib.axes._subplots.AxesSubplot at 0x7f3b2b1f12b0
#finding the x_res and y_res from screen resolution
new = df['ScreenResolution'].str.split('x',n=1,expand=True)
df['X_res'] = new[0]
df['Y_res'] = new[1]
df.sample(5)
Company TypeName Inches ScreenResolution Cpu Ram Memory Gpu OpSys Weight Price Touchscreen Ips X_res Y_res
1106 Asus Gaming 15.6 IPS Panel Full HD 1920x1080 Intel Core i7 8700HQ 2.6GHz 8 256GB SSD + 1TB HDD Nvidia GeForce GTX 960M Windows 10 2.591 48431.5200 0 1 IPS Panel Full HD 1920 1080
1300 Lenovo Notebook 14.0 1366x768 Intel Celeron Dual Core N3050 1.6GHz 2 64GB Flash Storage Intel HD Graphics Windows 10 1.500 12201.1200 0 0 1366 768
289 Lenovo Notebook 15.6 Full HD 1920x1080 Intel Core i7 7500U 2.7GHz 8 1TB HDD Nvidia GeForce 940MX No OS 2.200 35112.0528 0 0 Full HD 1920 1080
1262 Lenovo Notebook 15.6 Full HD 1920x1080 Intel Core i3 6100U 2.3GHz 6 128GB SSD Nvidia GeForce 920MX Windows 10 2.400 30316.3200 0 0 Full HD 1920 1080
329 Dell Notebook 15.6 4K Ultra HD / Touchscreen 3840x2160 Intel Core i7 7700HQ 2.8GHz 32 1TB SSD Nvidia GeForce GTX 1050 Windows 10 2.060 140605.9200 1 0 4K Ultra HD / Touchscreen 3840 2160
df['X_res'] = df['X_res'].str.replace(',','').str.findall(r'(\d+\.?\d+)').apply(lambda x:x[0])
df.head()
Company TypeName Inches ScreenResolution Cpu Ram Memory Gpu OpSys Weight Price Touchscreen Ips X_res Y_res
0 Apple Ultrabook 13.3 IPS Panel Retina Display 2560x1600 Intel Core i5 2.3GHz 8 128GB SSD Intel Iris Plus Graphics 640 macOS 1.37 71378.6832 0 1 2560 1600
1 Apple Ultrabook 13.3 1440x900 Intel Core i5 1.8GHz 8 128GB Flash Storage Intel HD Graphics 6000 macOS 1.34 47895.5232 0 0 1440 900
2 HP Notebook 15.6 Full HD 1920x1080 Intel Core i5 7200U 2.5GHz 8 256GB SSD Intel HD Graphics 620 No OS 1.86 30636.0000 0 0 1920 1080
3 Apple Ultrabook 15.4 IPS Panel Retina Display 2880x1800 Intel Core i7 2.7GHz 16 512GB SSD AMD Radeon Pro 455 macOS 1.83 135195.3360 0 1 2880 1800
4 Apple Ultrabook 13.3 IPS Panel Retina Display 2560x1600 Intel Core i5 3.1GHz 8 256GB SSD Intel Iris Plus Graphics 650 macOS 1.37 96095.8080 0 1 2560 1600
#convert to numeric
df['X_res'] = df['X_res'].astype('int')
df['Y_res'] = df['Y_res'].astype('int')
df.info()
completed at 7:51 AM
ENG IN
8:15 AM 16-Dec-22

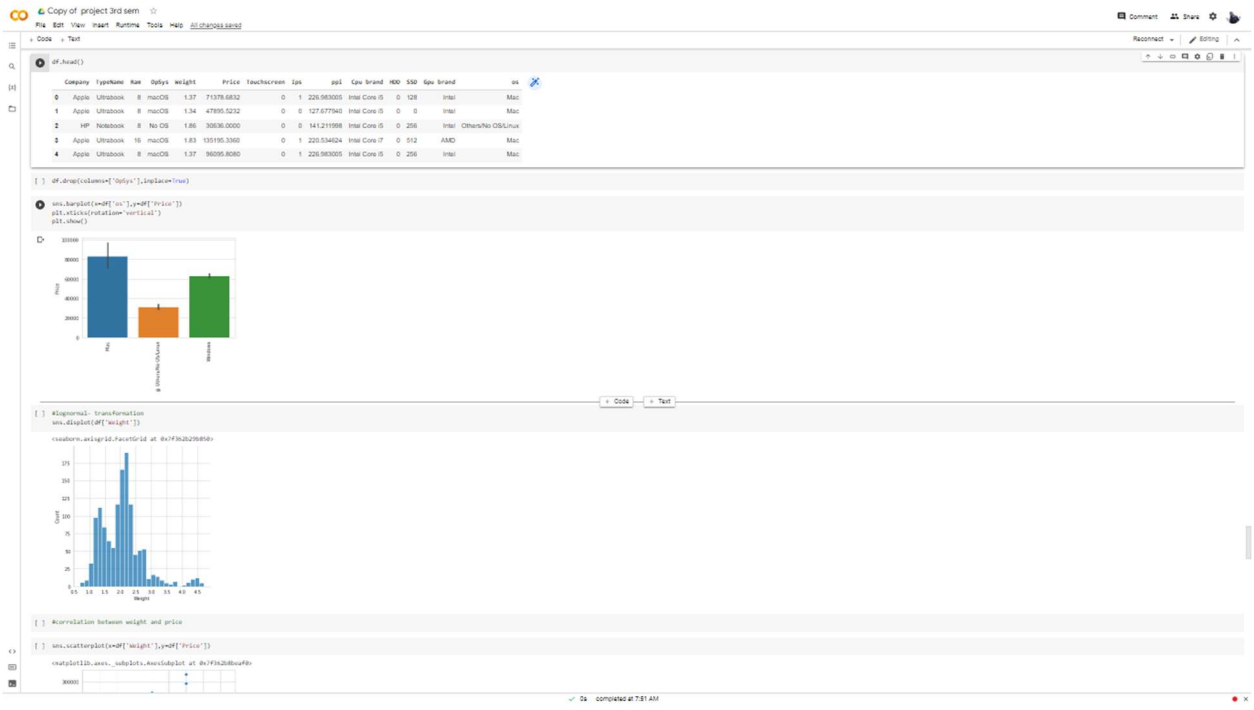

Copy of project 3rd sem - Colab
colab.research.google.com/drive/15buKlPgAQqbNelljr7QpAzdQsvi_mLEp#scrollTo=t95Rhv5gQCsl
Gmail
YouTube
Rakuten Udemy
Grow - AlmaBetter
Basics Of Training A...
Simple Linear Regre...
minitab tetorials - Y...
Python Pandas Part...
Other bookmarks
Copy of project 3rd sem
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Reconnect
Editing
4 Apple Ultrabook 13.3 IPS Panel Retina Display 2560x1600 Intel Core i5 3.1GHz 8 256GB SSD Intel Iris Plus Graphics 650 macOS 1.37 96095.8080 0 1 2560 1600
#convert to numeric
df['X_res'] = df['X_res'].astype('int')
df['Y_res'] = df['Y_res'].astype('int')
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1303 entries, 0 to 1302
Data columns (total 15 columns):
Column Non-Null Count Dtype
0 Company 1303 non-null object
1 TypeName 1303 non-null object
2 Inches 1303 non-null float64
3 ScreenResolution 1303 non-null object
4 Cpu 1303 non-null object
5 Ram 1303 non-null int32
6 Memory 1303 non-null object
7 Gpu 1303 non-null object
8 OpSys 1303 non-null object
9 Weight 1303 non-null float32
10 Price 1303 non-null float64
11 Touchscreen 1303 non-null int64
12 Ips 1303 non-null int64
13 X_res 1303 non-null int64
14 Y_res 1303 non-null int64
dtypes: float32(1), float64(2), int32(1), int64(4), object(7)
memory usage: 142.6+ KB
correlation of columns with price using the corr method
df.corr()['Price']
Inches 0.068197
Ram 0.743007
Weight 0.210370
Price 1.000000
Touchscreen 0.191226
Ips 0.252208
X_res 0.556529
Y_res 0.552809
Name: Price, dtype: float64
Pixel per Inches(PPI)
df['ppi'] = (((df['X_res']**2) + (df['Y_res']**2))**0.5/df['Inches']).astype('float')
df.corr()['Price']
Inches 0.068197
Ram 0.743007
Weight 0.210370
Price 1.000000
Touchscreen 0.191226
Ips 0.252208
X_res 0.556529
Y_res 0.552809
ppi 0.473487
Name: Price, dtype: float64
completed at 7:51 AM
ENG IN
8:15 AM 16-Dec-22

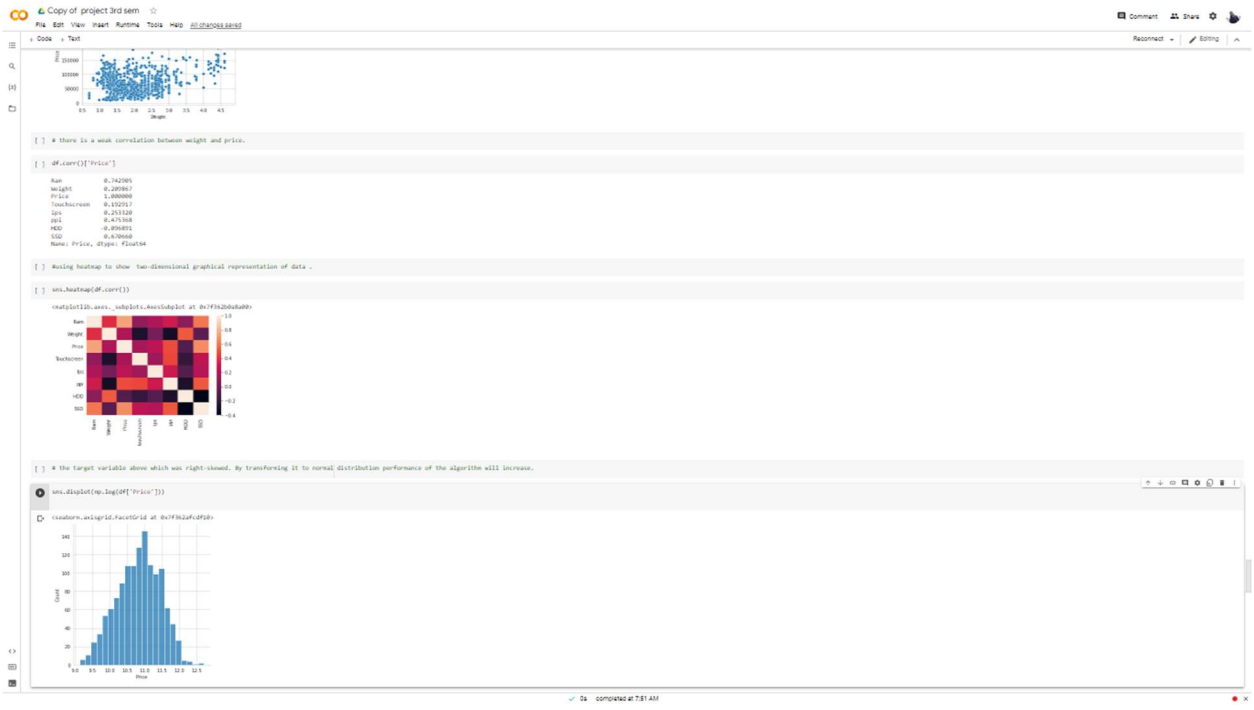

Copy of project 3rd sem
Press F11 to exit full screen

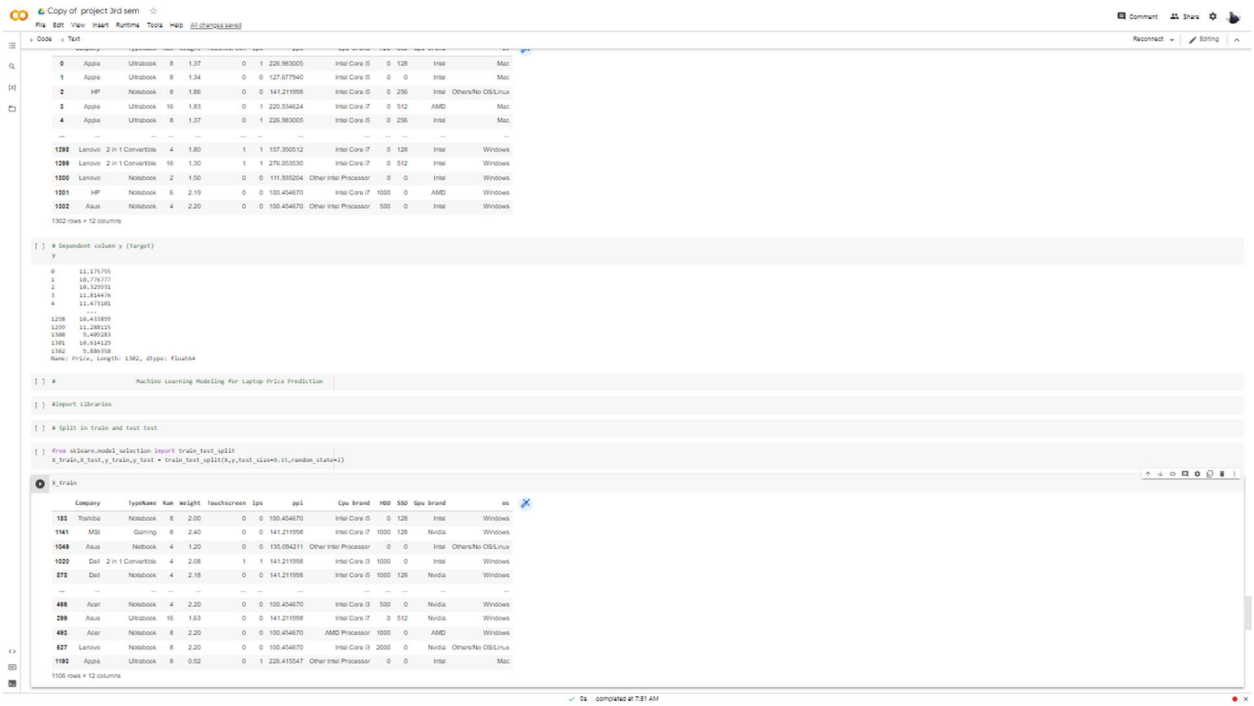

Copy of project 3rd sem




Press F11 to exit full screen


Copy of project 3rd sem
df.head()
df.drop(columns=['OpSys'],inplace=True)
sns.barplot(x=df['os'],y=df['Price'])
plt.xticks(rotation='vertical')
plt.show()
#lognormal transformation
sns.displot(df['Weight'])
#correlation between weight and price
sns.scatterplot(x=df['Weight'],y=df['Price'])


Copy of project 3rd sem
there is a weak correlation between weight and price.
df.corr()['Price']
Ram 0.742905
Weight 0.209867
Price 1.000000
Touchscreen 0.192917
Ips 0.253320
ppi 0.475368
HDD -0.096891
SSD 0.670660
Name: Price, dtype: float64
#using heatmap to show two-dimensional graphical representation of data .
sns.heatmap(df.corr())
# the target variable above which was right-skewed. By transforming it to normal distribution performance of the algorithm will increase.
sns.displot(np.log(df['Price']))



Copy of project 3rd sem

	Company	TypeName	Ram	Weight	Touchscreen	Ips	ppi	Cpu brand	HDD	SSD	Gpu brand	os
0	Apple	Ultrabook	8	1.37	0	1	226.983005	Intel Core i5	0	128	Intel	Mac
1	Apple	Ultrabook	8	1.34	0	0	127.677940	Intel Core i5	0	0	Intel	Mac
2	HP	Notebook	8	1.86	0	0	141.211998	Intel Core i5	0	256	Intel	Others/No OS/Linux
3	Apple	Ultrabook	16	1.83	0	1	220.534624	Intel Core i7	0	512	AMD	Mac
4	Apple	Ultrabook	8	1.37	0	1	226.983005	Intel Core i5	0	256	Intel	Mac
...	...	...	...	...	...	...	...	...	...	...	...	...
1298	Lenovo	2 in 1 Convertible	4	1.80	1	1	157.350512	Intel Core i7	0	128	Intel	Windows
1299	Lenovo	2 in 1 Convertible	16	1.30	1	1	276.053530	Intel Core i7	0	512	Intel	Windows
1300	Lenovo	Notebook	2	1.50	0	0	111.935204	Other Intel Processor	0	0	Intel	Windows
1301	HP	Notebook	6	2.19	0	0	100.454670	Intel Core i7	1000	0	AMD	Windows
1302	Asus	Notebook	4	2.20	0	0	100.454670	Other Intel Processor	500	0	Intel	Windows

1302 rows × 12 columns

```
# Dependent column y (target)
y
```

```
0       11.175755
1       10.776777
2       10.329931
3       11.814476
4       11.473101
          ...
1298    10.433899
1299    11.288115
1300     9.409283
1301    10.614129
1302     9.886358
Name: Price, Length: 1302, dtype: float64
```

```
#                    Machine Learning Modelling for Laptop Price Prediction
```

```
#import Libraries
```

```
# Split in train and test test
```

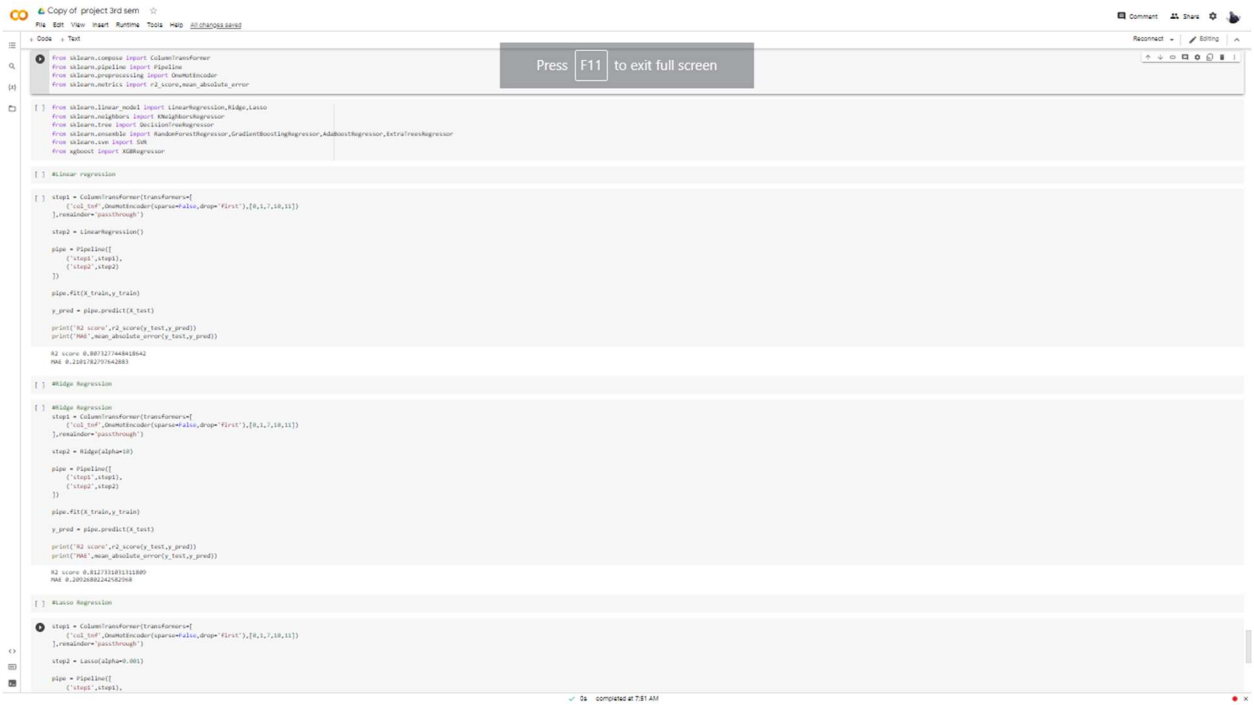
```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.15,random_state=2)
```

```
X_train
```

	Company	TypeName	Ram	Weight	Touchscreen	Ips	ppi	Cpu brand	HDD	SSD	Gpu brand	os
183	Toshiba	Notebook	8	2.00	0	0	100.454670	Intel Core i5	0	128	Intel	Windows
1141	MSI	Gaming	8	2.40	0	0	141.211998	Intel Core i7	1000	128	Nvidia	Windows
1049	Asus	Netbook	4	1.20	0	0	135.094211	Other Intel Processor	0	0	Intel	Others/No OS/Linux
1020	Dell	2 in 1 Convertible	4	2.08	1	1	141.211998	Intel Core i3	1000	0	Intel	Windows
878	Dell	Notebook	4	2.18	0	0	141.211998	Intel Core i5	1000	128	Nvidia	Windows
...	...	...	...	...	...	...	...	...	...	...	...	...
466	Acer	Notebook	4	2.20	0	0	100.454670	Intel Core i3	500	0	Nvidia	Windows
299	Asus	Ultrabook	16	1.63	0	0	141.211998	Intel Core i7	0	512	Nvidia	Windows
493	Acer	Notebook	8	2.20	0	0	100.454670	AMD Processor	1000	0	AMD	Windows
527	Lenovo	Notebook	8	2.20	0	0	100.454670	Intel Core i3	2000	0	Nvidia	Others/No OS/Linux
1193	Apple	Ultrabook	8	0.92	0	1	226.415547	Other Intel Processor	0	0	Intel	Mac

1106 rows × 12 columns

0s completed at 7:51 AM



Copy of project 3rd sem

Press F11 to exit full screen

```
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import OneHotEncoder
from sklearn.metrics import r2_score,mean_absolute_error
```

```
from sklearn.linear_model import LinearRegression,Ridge,Lasso
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor,GradientBoostingRegressor,AdaBoostRegressor,ExtraTreesRegressor
from sklearn.svm import SVR
from xgboost import XGBRegressor
```

```
#Linear regression
```

```
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')

step2 = LinearRegression()

pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])

pipe.fit(X_train,y_train)

y_pred = pipe.predict(X_test)

print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
```

```
R2 score 0.8073277448418642
MAE 0.2101782795642883
```

```
#Ridge Regression
```

```
#Ridge Regression
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')

step2 = Ridge(alpha=10)

pipe = Pipeline([
    ('step1',step1),
    ('step2',step2)
])

pipe.fit(X_train,y_train)

y_pred = pipe.predict(X_test)

print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
```

```
R2 score 0.8127331013111889
MAE 0.2092688224262968
```

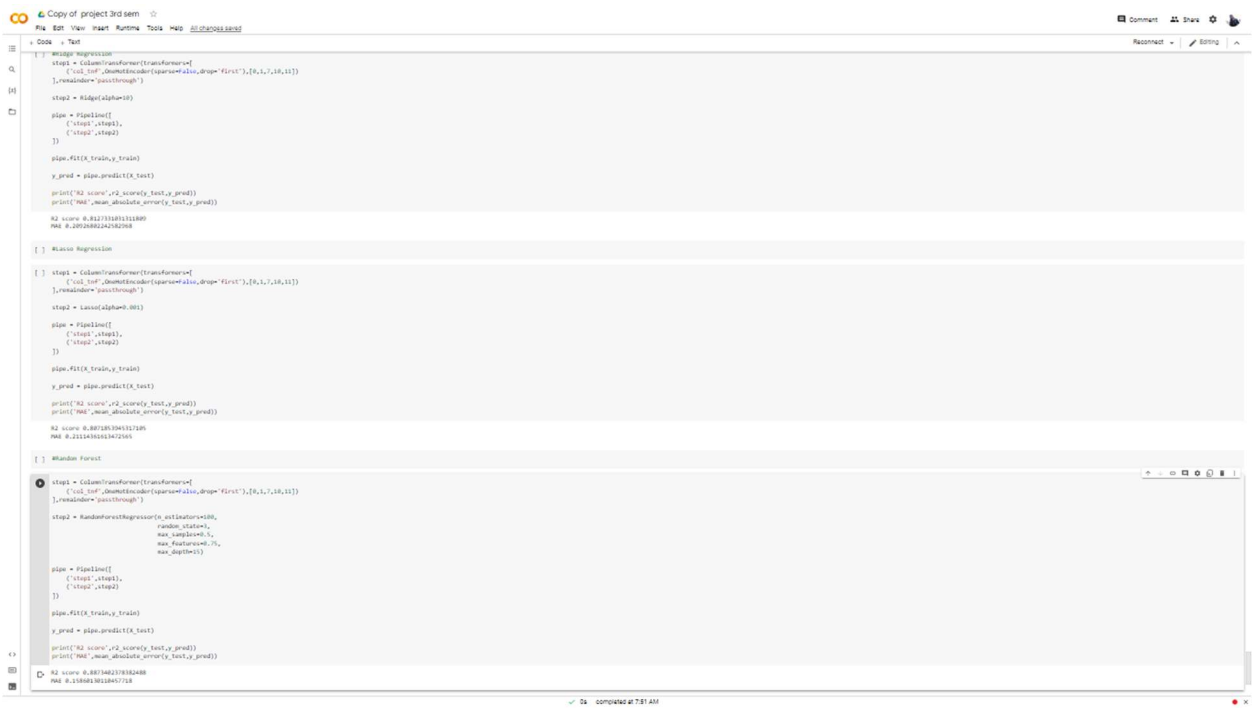
```
#Lasso Regression
```

```
step1 = ColumnTransformer(transformers=[
    ('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')

step2 = Lasso(alpha=0.001)

pipe = Pipeline([
    ('step1',step1),
```

0s completed at 7:51 AM


Copy of project 3rd sem
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
Reconnect Editing
+ Code + Text
#ridge regression
step1 = ColumnTransformer(transformers=[
('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')
step2 = Ridge(alpha=10)
pipe = Pipeline([
('step1',step1),
('step2',step2)
])
pipe.fit(X_train,y_train)
y_pred = pipe.predict(X_test)
print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
R2 score 0.8127331013118895
MAE 0.20926802242582968
#Lasso Regression
step1 = ColumnTransformer(transformers=[
('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')
step2 = Lasso(alpha=0.001)
pipe = Pipeline([
('step1',step1),
('step2',step2)
])
pipe.fit(X_train,y_train)
y_pred = pipe.predict(X_test)
print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
R2 score 0.8071853945317105
MAE 0.21114361613472565
#Random Forest
step1 = ColumnTransformer(transformers=[
('col_tnf',OneHotEncoder(sparse=False,drop='first'),[0,1,7,10,11])
],remainder='passthrough')
step2 = RandomForestRegressor(n_estimators=100,
random_state=3,
max_samples=0.5,
max_features=0.75,
max_depth=15)
pipe = Pipeline([
('step1',step1),
('step2',step2)
])
pipe.fit(X_train,y_train)
y_pred = pipe.predict(X_test)
print('R2 score',r2_score(y_test,y_pred))
print('MAE',mean_absolute_error(y_test,y_pred))
R2 score 0.8873402378382488
MAE 0.1586019010457718
0s completed at 7:51 AM

----conclusion---

- According to the model applied Linear Regression gives 80% (R=0.8073277448418642) accuracy ,and mean absolute error(MAE 0.2101782797642883).

- The Ridge Regression model gives 81% (R2 score 0.8127331031311809) accuracy ,and mean absolute error (MAE 0.20926802242582968) .

- The lesso Regression model gives 80% (R2 score 0.8071853945317105) accuracy ,and mean absolute error (MAE 0.21114361613472565) .

- The Random forest Regression model gives 88% (R2 score 0.8873402378382488) accuracy ,and mean absolute error (MAE 0.15860130110457718).

- Hence Random Forest will give best accuracy among all. That means the prediction based on these model will be accurate and near to the actual value.