**Ex No: 2**

**Exploratory Data Analysis** **(EDA)**

**AIM**:

To perform exploratory data analysis for the dataset and obtain Measures of Central Tendency, Measure of Dispersion, Descriptive Statistics, Skewness and Kurtosis, and correlation using python.

**Dataset Description:**

The dataset contains information on over 2,000 mobile phones from different brands. It includes details such as the storage capacity, RAM, screen size, camera specifications, battery capacity, and price of each device.

The dataset is structured as a CSV file with 7 columns:

* Brand: The brand name of the mobile phone.
* Model: The model name of the mobile phone.
* Storage: The amount of storage space available on the mobile phone in GB.
* RAM: The amount of random access memory available on the mobile phone in GB.
* Screen Size: The size of the mobile phone's screen in inches.
* Camera: The quality of the mobile phone's cameras, measured in megapixels.
* Battery Capacity: The amount of battery life the mobile phone has in mAh.
* Price: The price of the mobile phone in USD.

**Problem Statement**

The mobile phone price prediction problem is to develop a model that can predict the price of a mobile phone given a set of features. The target variable is the price of the mobile phone in USD. The goal of the problem is to develop a model that can accurately predict the price of a mobile phone given its features. This model can be used by a variety of stakeholders, including:

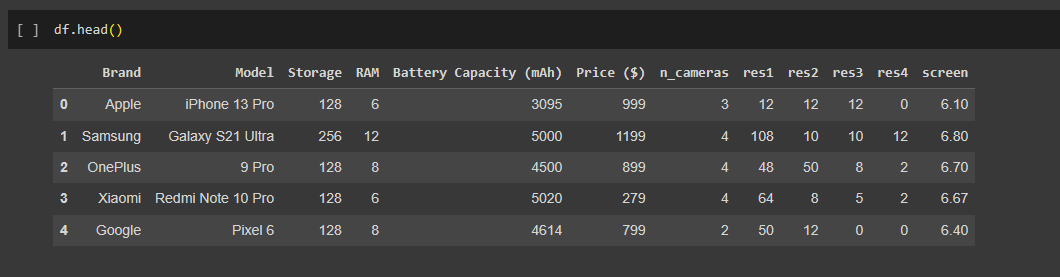
Mobile phone manufacturers: Manufacturers can use the model to develop a pricing strategy for their products. They can also use the model to identify the features that are most important to consumers and to determine how much they should charge for their phones based on those features.

Retailers: Retailers can use the model to set prices for mobile phones in their store. They can also use the model to compare the prices of different phones from different manufacturers and to ensure that they are charging a competitive price.

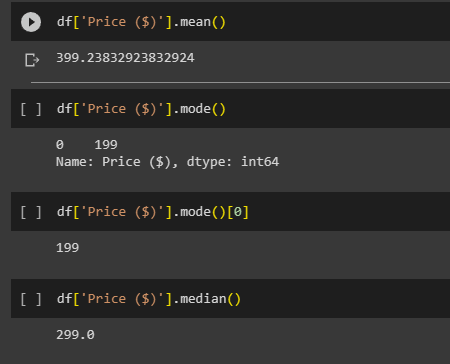
Consumers: Consumers can use the model to make informed decisions about which mobile phone to buy. They can use the model to compare the prices of different phones with different features and to find the best value for their money.

**PROGRAMS WITH OUTPUT:**

df.head()

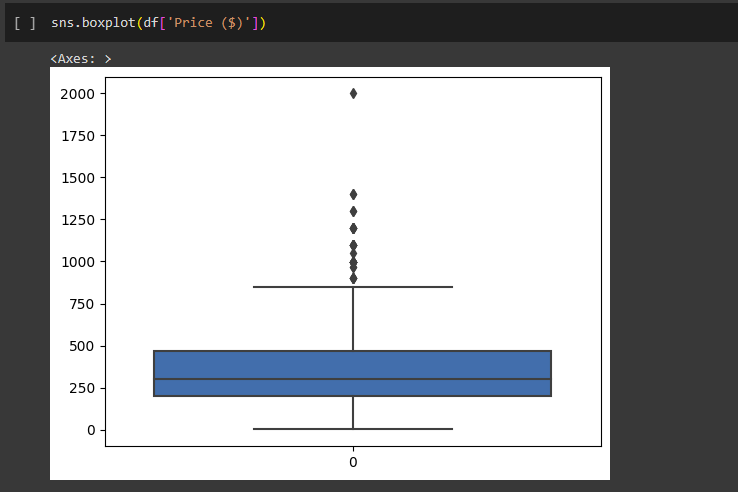
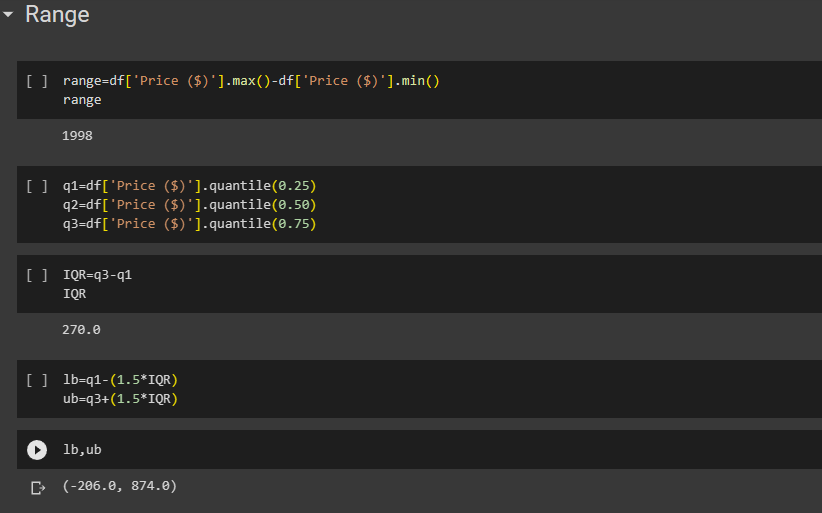
****

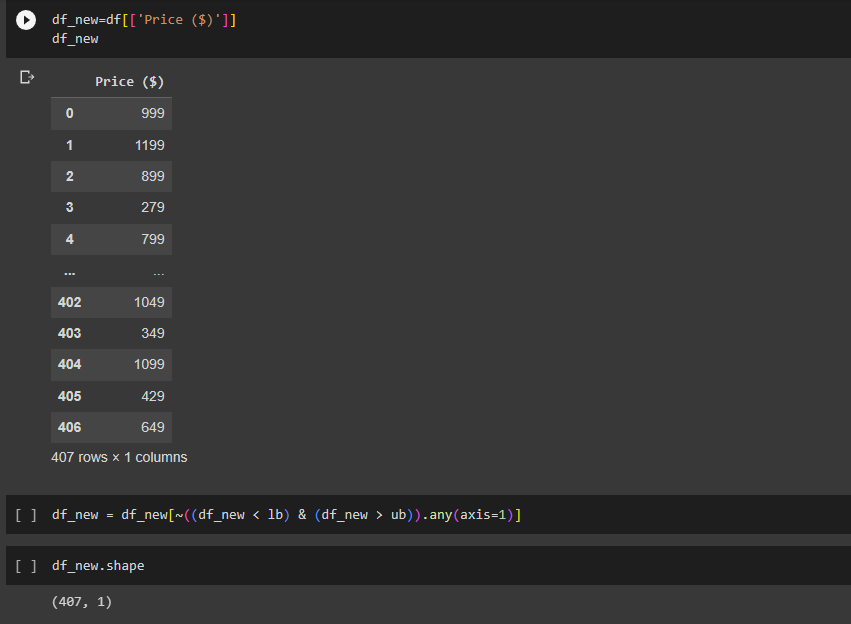
**Measures of Central Tendency**

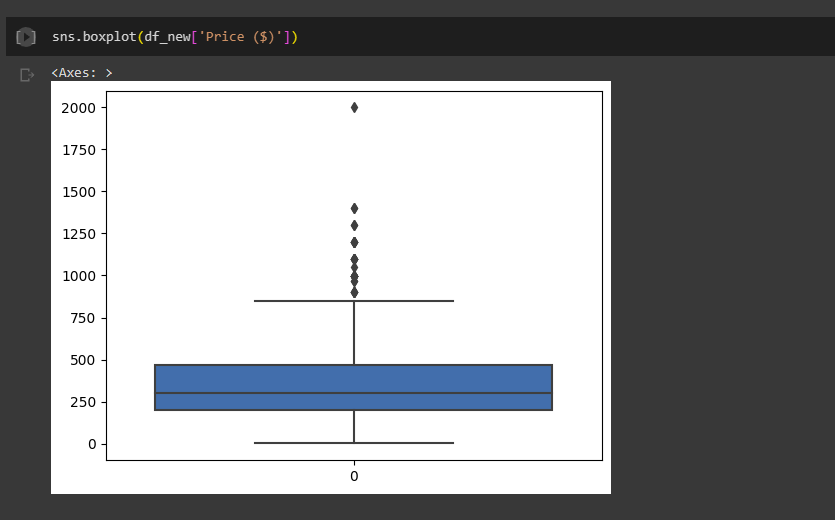


**Measure of Dispersion**

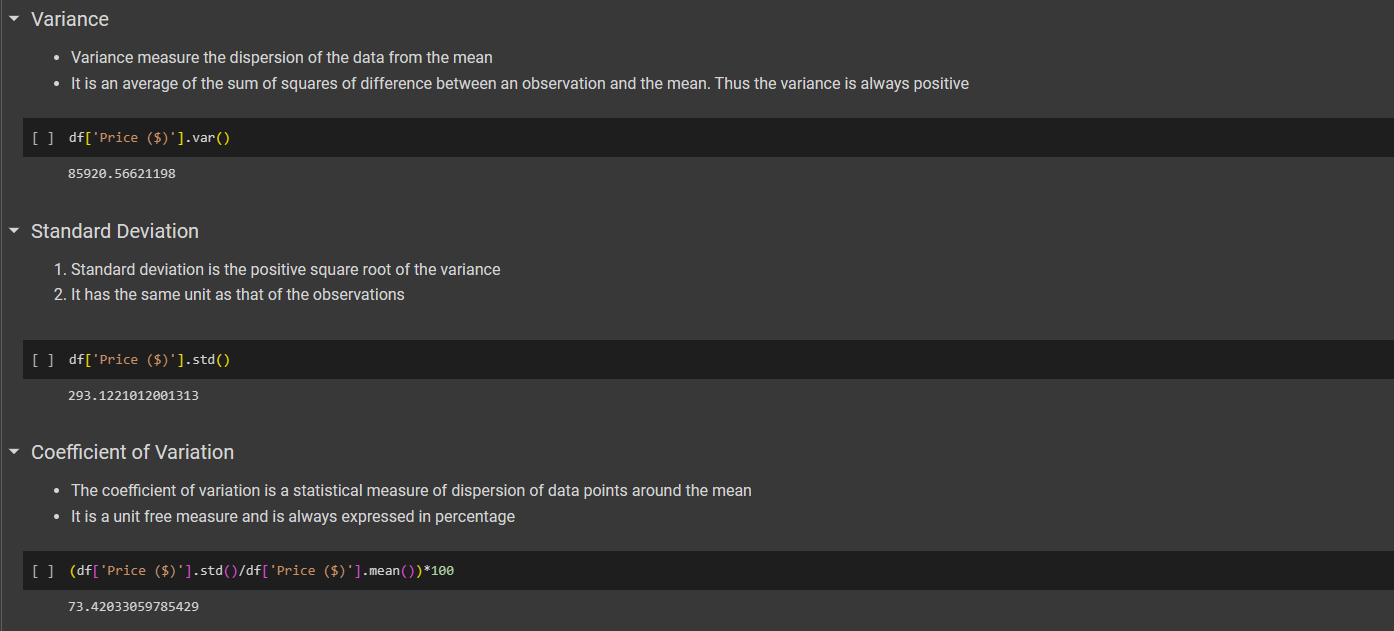
1. **Range**

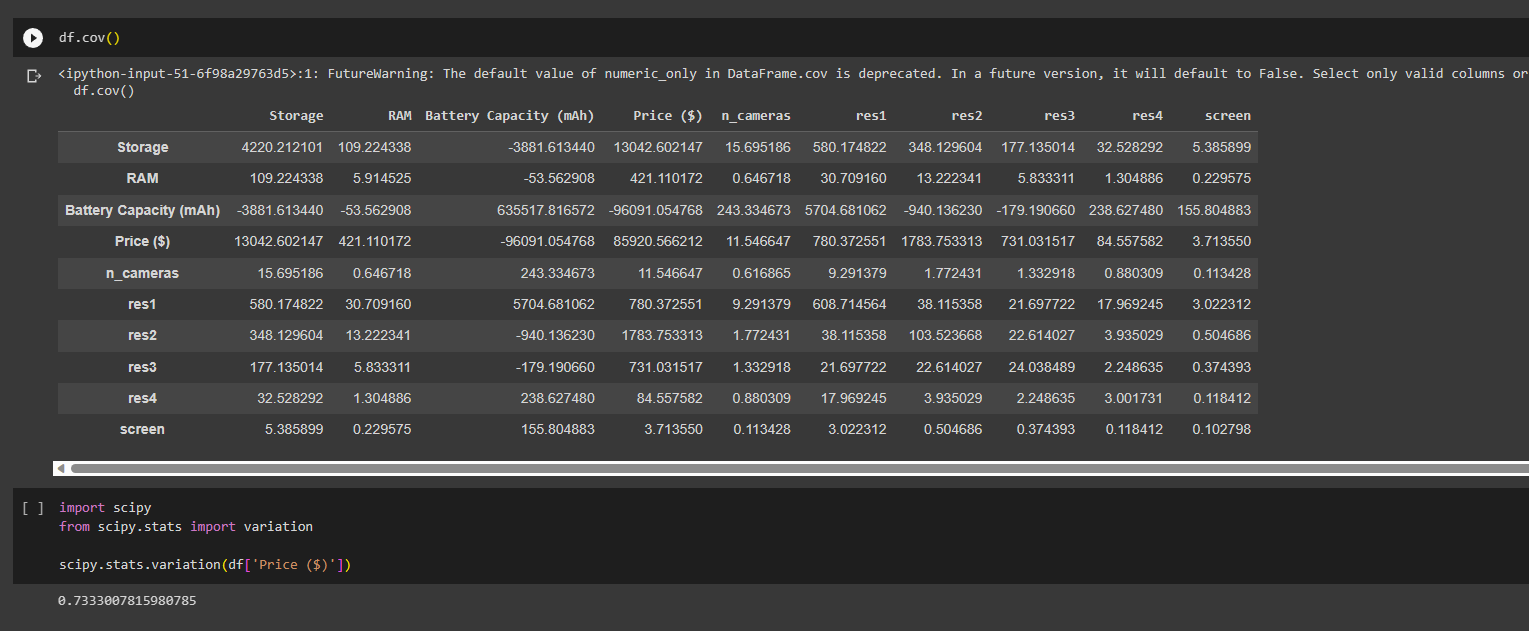




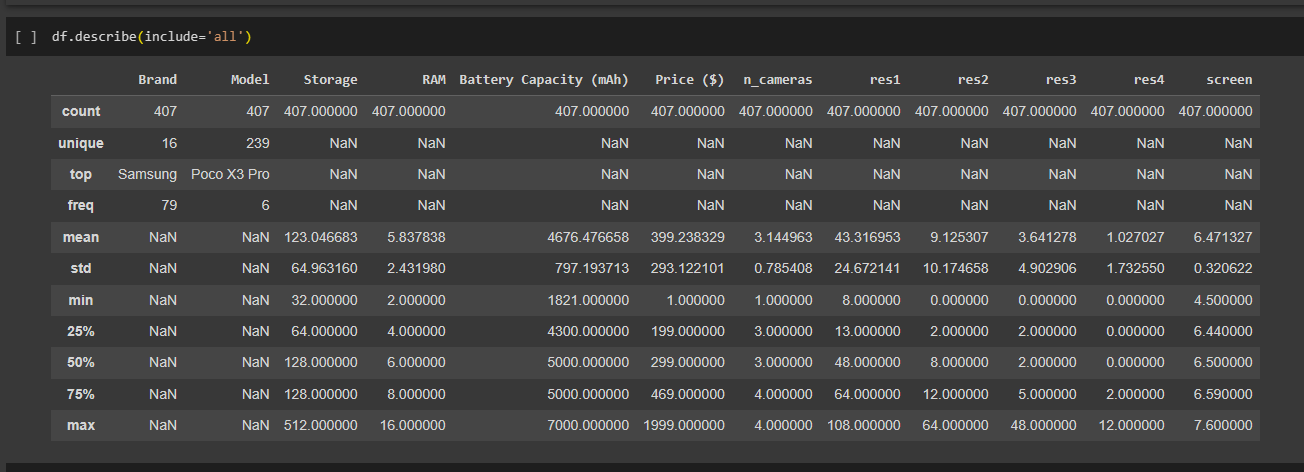


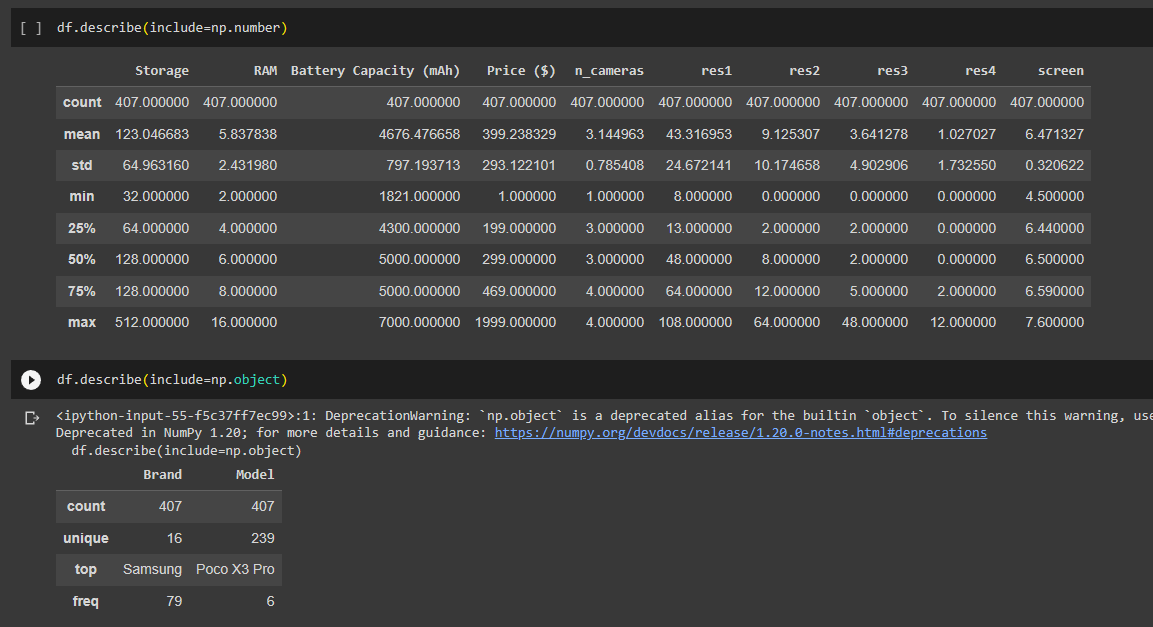
1. **Variance And Standard Deviation**



****

**Descriptive Statistics**

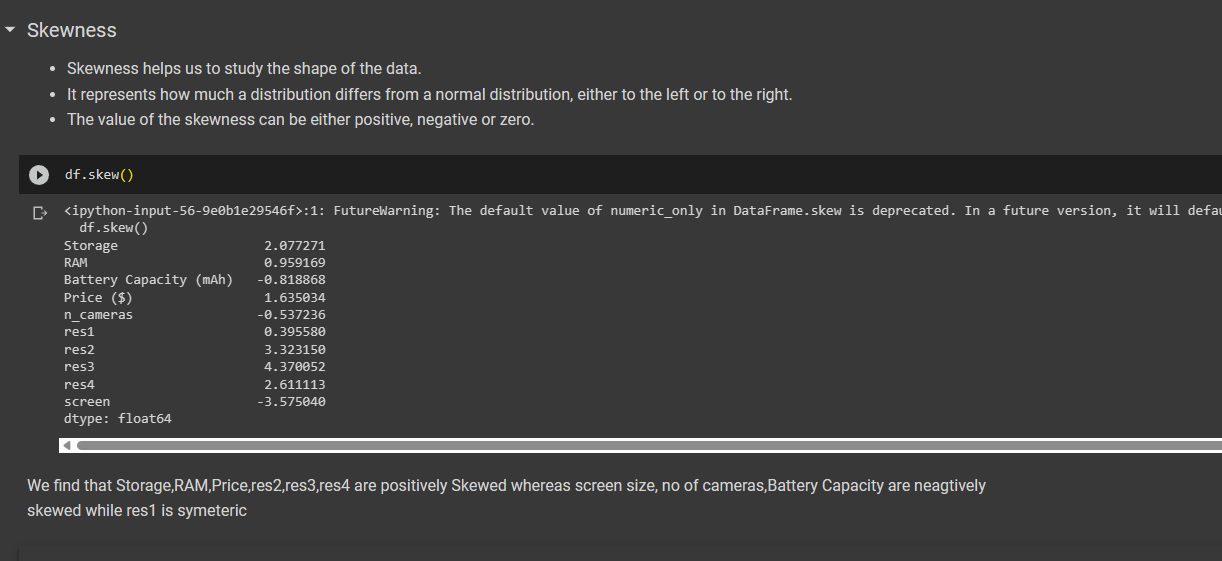
****

****

**Shape of the data**

**1.Skewness**

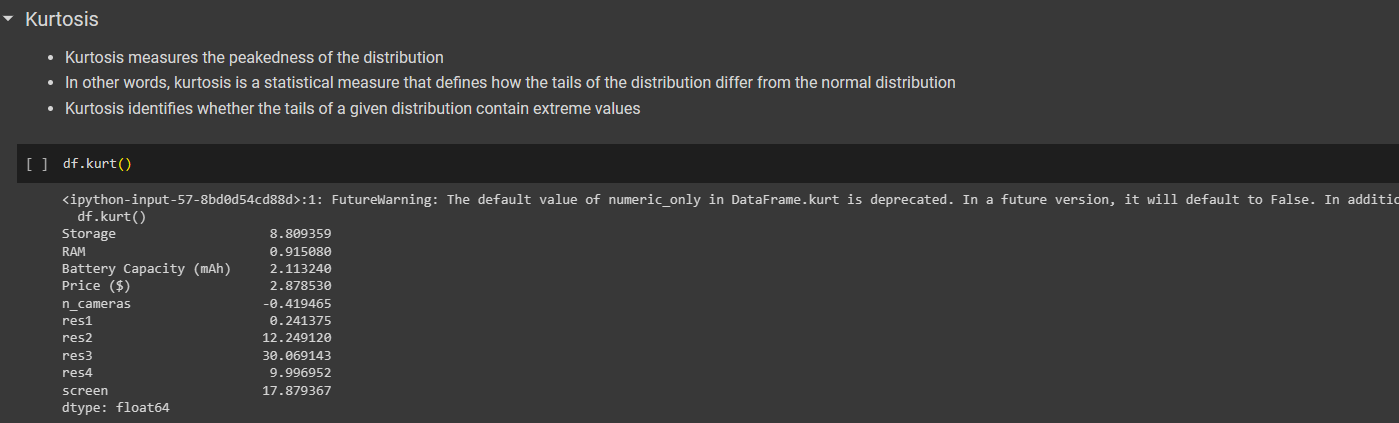
* Skewness helps us to study the shape of the data.
* It represents how much a distribution differs from a normal distribution, either to the left or to the right.
* The value of the skewness can be either positive, negative or zero.

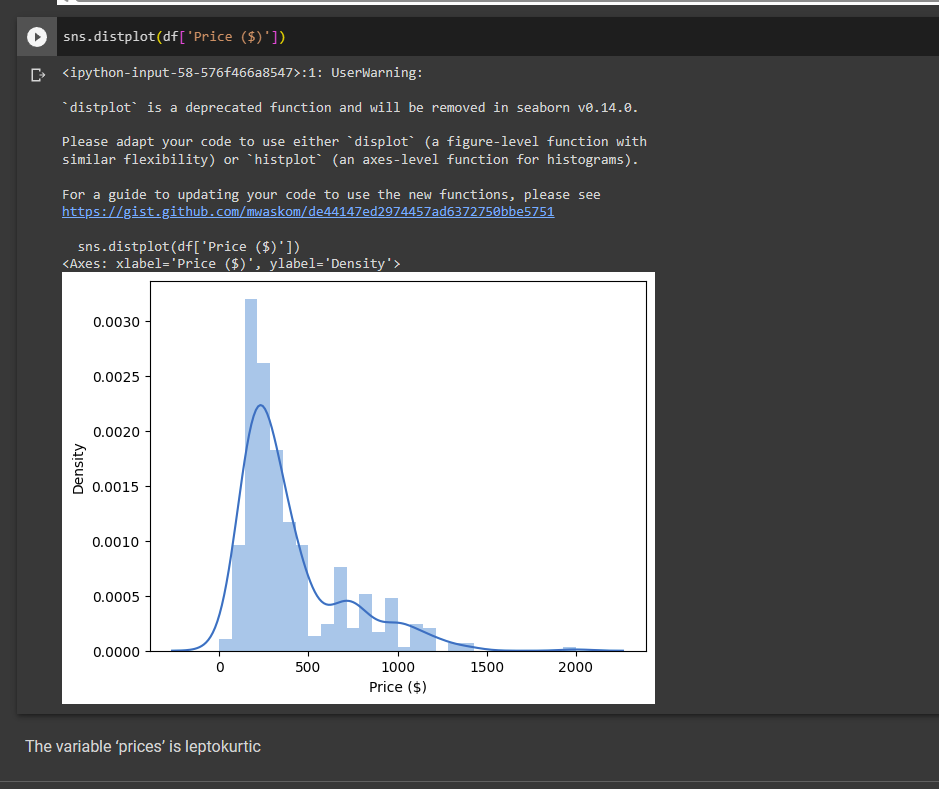
****

We find that Storage,RAM,Price,res2,res3,res4 are positively Skewed whereas screen size, no of cameras,Battery Capacity are neagtively skewed while res1 is symeteric

**2.Kurtosis**

* Kurtosis measures the peakedness of the distribution
* In other words, kurtosis is a statistical measure that defines how the tails of the distribution differ from the normal distribution
* Kurtosis identifies whether the tails of a given distribution contain extreme values

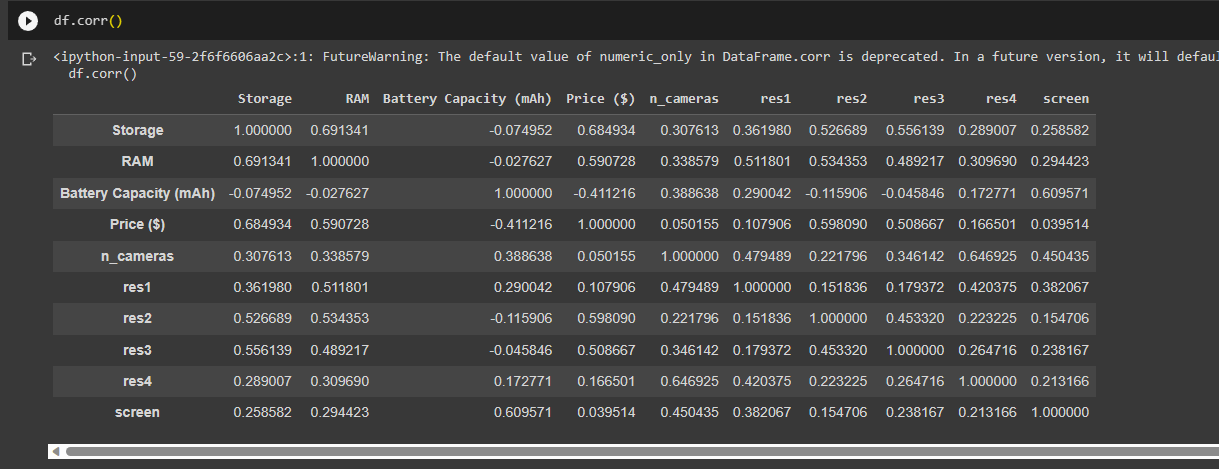


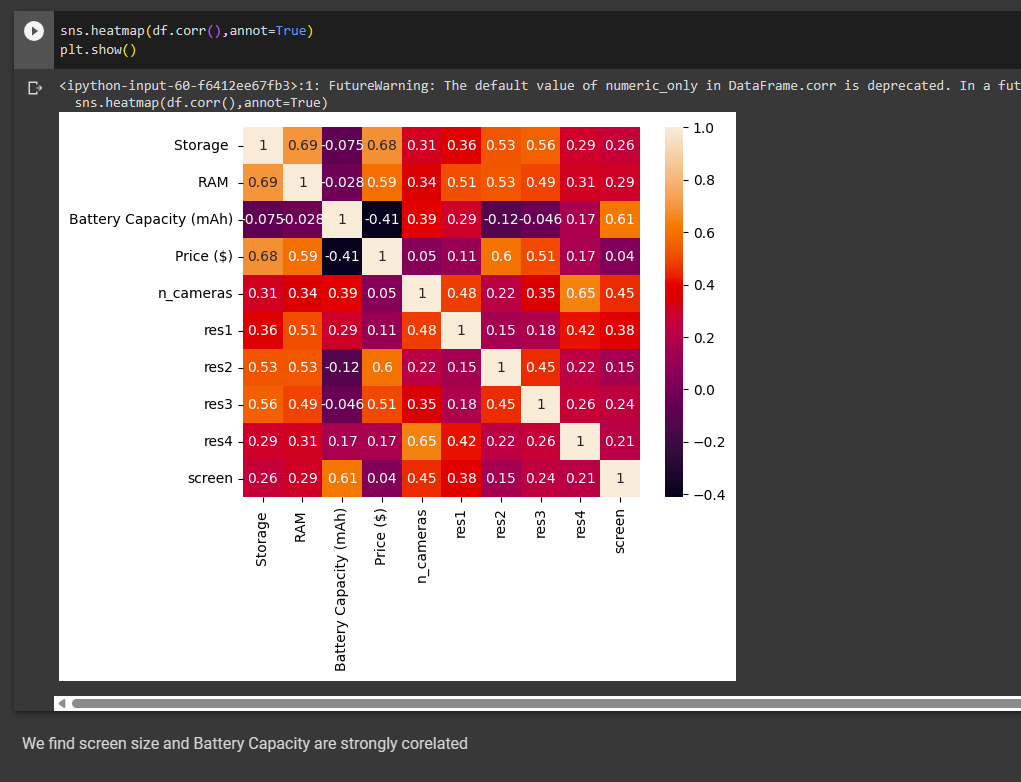


The variable ‘prices’ is leptokurtic

**Correlation**

* It shows whether pairs of variables are related to each other
* If there is correlation, it shows how strong the correlation is
* Correlation takes values between -1 to +1, where values close to +1 represents strong positive correlation while values close to -1 represents strong negative correlation

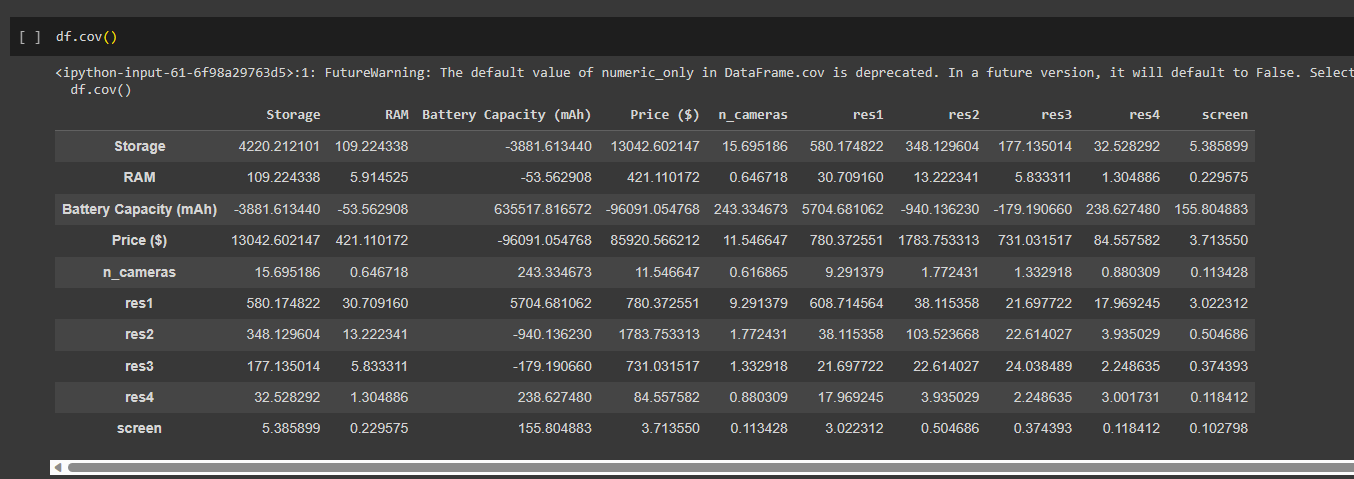


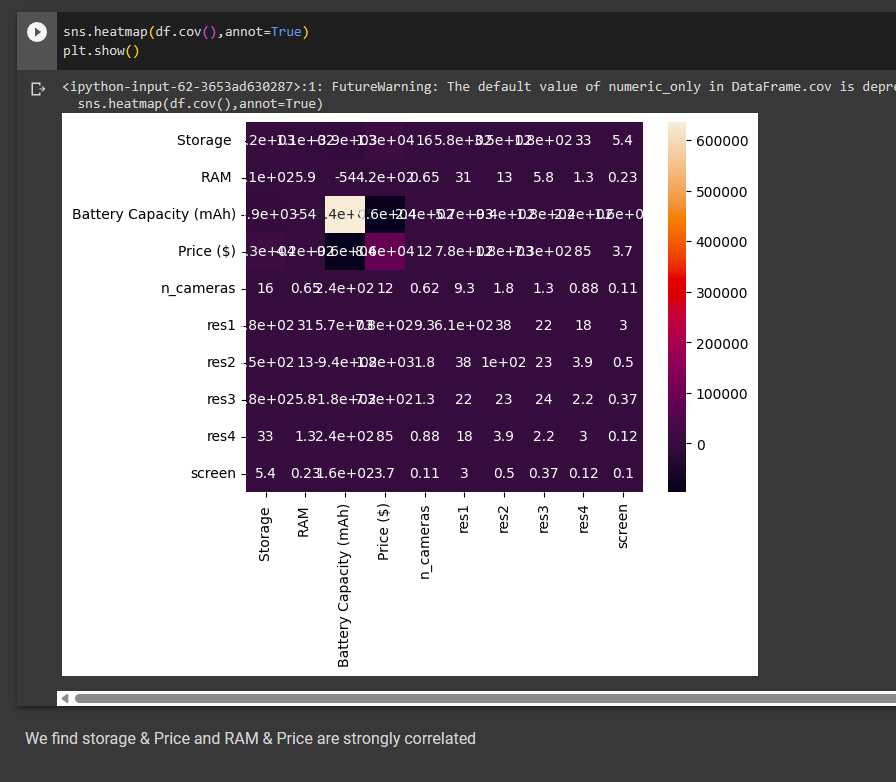


We find screen size and Battery Capacity are strongly corelated

**Covariance**

* It is the relationship between a pair of random variables where change in one variable causes change in another variable
* It can take any value between -infinity to +infinity, where the negative value represents the negative relationship whereas a positive value represents the positive relationship





**CONCLUSION:**

The exploratory data analysis has been done using the given dataset and the results have been analysed using the above visualizations.