

# **DESCRIPTIVE STATISTICS**

**A Project Report**

**Submitted for the partial fulfilment of the**

**requirements for Semester-III**

**of BSc (Information Technology / ~~Computer Science~~)**

**By**

**Prajapati Gaurav Pravinbhai**

**(Roll No: 347)**

**Under the Guidance of**

**Dr. Sindhu P. M.**



**Department of Computer and Information Science**

**Nagindas Khandwala College (Autonomous)**

**Mumbai**

**2022 – 2023**

<CERTIFICATE>

This is to certify that the project titled, "<sup>Types of</sup> Mobile users", is bonafied work of Gaurav Prajapati  
(Name of the student) bearing Seat No.: 347 submitted in partial fulfilment of the  
requirements for the Semester-III of BSc (Information Technology / Computer Science) from  
Nagindas Khandwala College (Autonomous), Mumbai.

  
<Signature of the Guide>



19 SEP 2022

## AKNOWLEDGEMENT

<This is a sample. Make changes as required>

I am very thankful to <sup>NKC</sup> college, the Principal Prof. Dr. Moushum Datta for allowing me to take up this project and to provide the resources for this work. This project would not be possible without the patient support and guidance from my supervisor Dr. Sindhu P. M., who extended her valuable time in explaining the new concepts, providing the reference materials and also by motivating and encouraging me to go ahead with this work. This helped me to learn the research skills and presentation skills too. I am indebted to my supervisor, which is beyond the words.

I am grateful to all the teachers in the Department of Computer and Information Science for their support and cooperation. I take this opportunity to thank my classmates, friends and parents for the support, cooperation and encouragement that I receive always. Last but not the least, I am thankful to the almighty for timely completion of this project.

Ganraj Prajapati. Ganraj.

(Name and Signature of the student)

12-9-22  
Date

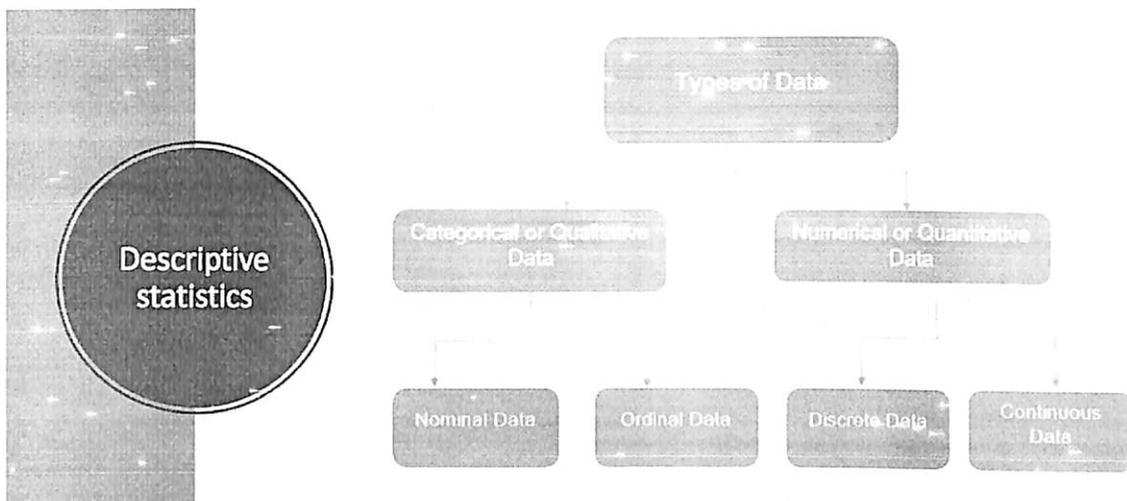
Contents

- I. Introduction..... 1
  - Ordinal Data ..... 1
  - Quantitative Data ..... 2
  - Motivation..... 2
  - Terminology..... 3
  - Objectives ..... 3
  - Problem Statement..... 4
- II. Data Collection and Software Used ..... 5
- III. Data Visualization ..... 6
- IV. Data Analysis and Results..... 8
- V. Conclusion..... 11
  - Summary ..... 11
  - Future Enhancement ..... 11
- References ..... 12

DESCRIPTIVE STATISTICS

**I. Introduction**

Data is the new oil.” Today data is everywhere in every field. Whether you are a data scientist, marketer, businessman, data analyst, researcher, or you are in any other profession, you need to play or experiment with raw or structured data. This data is so important for us that it becomes important to handle and store it properly, without any error. While working on these data, it is important to know the types of data to process them and get the right results. There are two types of data: Qualitative and Quantitative data, which are further classified into four types of data: nominal, ordinal, discrete, and Continuous.



**The Qualitative data are further classified into two parts:**

**Nominal Data**

Nominal Data is used to label variables without any order or quantitative value. The colour of hair can be considered nominal data, as one colour can't be compared with another colour.

**Ordinal Data**

Ordinal data have natural ordering where a number is present in some kind of order by their position on the scale. These data are used for observation like customer satisfaction, happiness, etc., but we can't do any arithmetical tasks on them.

## DESCRIPTIVE STATISTICS

**Quantitative Data**

Quantitative data can be expressed in numerical values, which makes it countable and includes statistical data analysis. These kinds of data are also known as Numerical data. It answers the questions like, “how much,” “how many,” and “how often.” For example, the price of a phone, the computer’s ram, the height or weight of a person, etc., falls under the quantitative data.

**What is Data Visualization?**

Data visualization is the graphical representation of information and data. By using visual elements, graphs, pie charts, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.

In the world of Big Data, data visualization tools and technologies are essential to analyse massive amounts of information and make data-driven decisions.

**Univariate Data Analysis**

**Univariate analysis** is the simplest form of analysing data. “Uni” means “one”, so in other words your data has only one variable. It doesn’t deal with causes or relationships and its major purpose is to describe; It takes data, summarizes that data and finds patterns in the data.

**Bivariate Data Analysis**

Bivariate analysis lets you study the relationship that exists between two variables. This has a lot of use in real life. It helps to find out if there is an association between the variables and if yes then what is the strength of association.

**Motivation**

It is interesting to know more about the field of Mobile phones and the specifications we have to see while buying a phone for example brand, memory, storage, colour and price.

This work will be beneficial to the society to get the acknowledgement about the products and variants in products. When planning to purchase a cell phone, there are a number of factors that one must take into consideration. One may easily end up with a product that does not meet expectations because of the failure to plan. In this paper, the researcher will look at concepts that should be considered when purchasing a cell phone.

This application area motivated me to take up this works ahead.

DESCRIPTIVE STATISTICS

**Terminology**

- i. To get an Overview about Mobile Marketing their growth analysis, inflation & deflation in the market. Covering the major difference between Original Price and Selling Price.
- ii. Formula we are going apply to perform this,
  - A) Standard deviation:  $s = \sqrt{\sum (x_i - \bar{x})^2 / n - 1}$
  - B) Mean, Median, Mode:  $(\text{sum of total numbers}) / N$ ,  $((n/2) + (n/2 + 1)) / 2$ ,  $\text{mode} = \sum x / n$ .
  - C) Variance, Covariance:  $s^2 = \sum (x_i - \bar{x})^2 / n - 1$
  - D) Kurtosis:  $= \text{KURT}(\text{ARRAY}, \text{ARRAY})$

**Objectives**

- 1. To find the major difference between Original Price and Selling Price & specification interests difference between male and female buyers.
- 2. To find the interest of people towards Mobile Phone Specifications.



DESCRIPTIVE STATISTICS

## II. Data Collection and Software Used

Collection method: Secondary data

We have used Secondary data to access the information and status of Specifications in India.

Secondary data are the primary data which are collected, and readily available for the data analysis.

Explanation about the data: Size of data is 10 \* 200 matrix.

Data types in dataset: Quantitative and Qualitative.

Type of dataset we are using:

name	gender	brand	model	color	memory	storage	rating	selling_pr	original_price
Yasin	Male	OPPO	A53	Moonligh	4 GB	64 GB	4.5	11990	15990
Prit	Male	OPPO	A53	Mint Crea	4 GB	64 GB	4.5	11990	15990
Meet	Male	OPPO	A53	Moonligh	6 GB	128 GB	4.3	13990	17990
Drashti	Female	OPPO	A53	Mint Crea	6 GB	128 GB	4.3	13990	17990
Saloni	Female	OPPO	A53	Electric Bl	4 GB	64 GB	4.5	11990	15990
Hinal	Female	OPPO	A53	Electric Bl	6 GB	128 GB	4.3	13990	17990
Jay	Male	OPPO	A12	Deep Blue	4 GB	64 GB	4.4	10490	11990
Darshana	Female	OPPO	A12	Black	3 GB	256 GB	4.4	9490	10990
Hardik	Male	OPPO	A12	Blue	3 GB	256 GB	4.4	9490	10990

Used coding platform: Colab.

Colab:

Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

Excel:

It is widely used for using Excel for statistical data analysis. Regression is a process of establishing a relationship among many variables; to establish a relationship between dependant variables and independent variables. Regression is great for use for using Excel for statistical data analysis.

R programming:

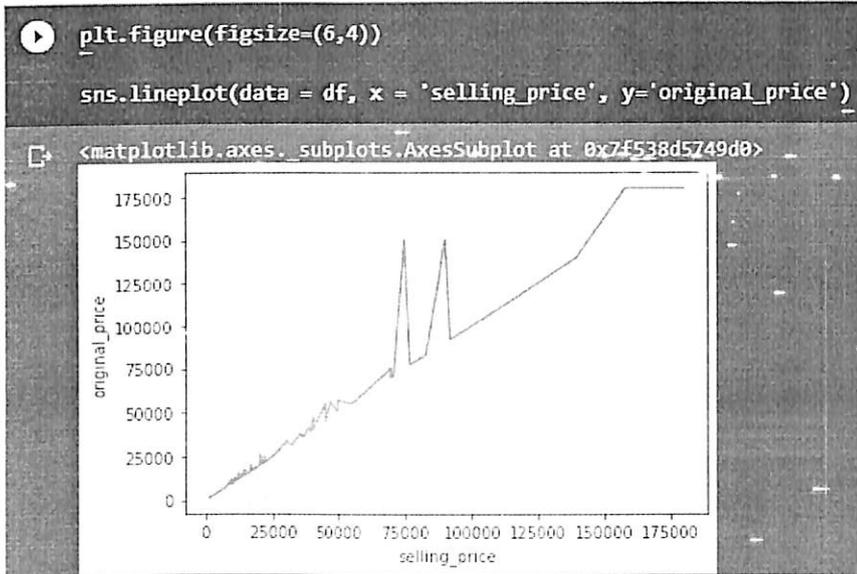
R is programming language for statistical computing and graphics that you can use to clean, analyze, and graph your data. It is widely used by researchers from diverse disciplines to estimate and display results and by teachers of statistics and research methods.

DESCRIPTIVE STATISTICS

### III. Data Visualization

Line Graph: Bivariate analysis

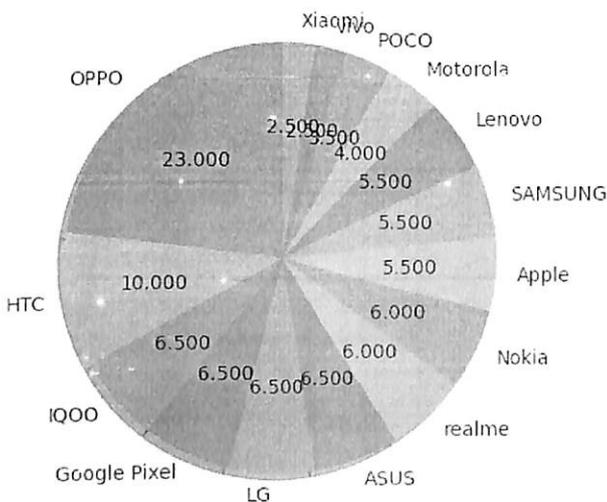
Here the line graph is used to check the relation between our data. The relation in our data is positive.



Pie chart: Univariate analysis

Here this pie chart shows us the optimization of the brands, which has higher demand than each other

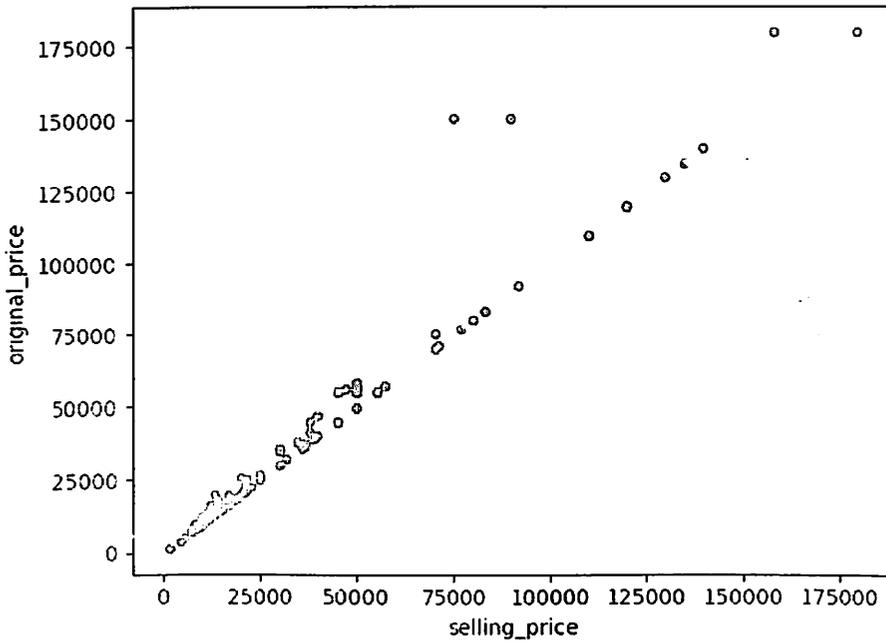
```
plt.figure(figsize=(5,5))
plt.pie(df['brand'].value_counts(),startangle=90,autopct='%.3f',labels=['OPPO','HTC','IQOO','Google Pixel','LG','ASUS','realme','Nokia','Apple','SAMSUNG','Lenovo','Motorola','POCO','vivo','Xiaomi'],shadow=True)
```



DESCRIPTIVE STATISTICS

Scatter plot: Bivariate analysis

Scatter plots are used to plot data points on a horizontal and a vertical axis in the attempt to show how much one variable is affected by another. Each row in the data table is represented by a marker whose position depends on its values in the columns set on the X and Y axes.



DESCRIPTIVE STATISTICS

**IV. Data Analysis and Results**

**Quantitative Data Analysis:**

Skewness is used to obtain approximate probabilities and quantiles of distributions.

	selling_price	original_price
Skewness	4.68911828	3.67028857

Kurtosis is used as a measure to define, how greater will be the deviation.

	selling_price	original_price
Krutosis	25.95138142	15.53747152

**1) Univariate analysis**

Uni means one and variate means variable, so in univariate analysis, there is only one dependable variable. The objective of univariate analysis is to derive the data, define and summarize it, and analyze the pattern present in it. In a dataset, it explores each variable separately. It is possible for two kinds of variables- Categorical and Numerical.

**A) Central tendency:** A measure of central tendency is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution.

- i) **Mean:** Mean is the average of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers.  
Formula:  $\text{mean} = (\text{sum of total numbers})/n$
- ii) **Mode:** A mode is defined as the value that has a higher frequency is a given set of values.  
Formula:  $\text{mode} = \{x/n\}$
- iii) **Median:** A median is the value separating the higher half from the lower half of a data sample.  
Formula:  $\text{median} = ((n/2) + (n/2+1))/2$

**Results for Central Tendency for our Dataset**

	selling_price	original_price
Mean	7231.08475	11974.65345
Median	12	128
Mode	128	128

**B) Dispersion:** Dispersion is the state of getting dispersed or spread. Statistical dispersion means the extent to which numerical data is likely to vary about an average value.

- i) **Range:** A series of things in a line a range of mountains.

DESCRIPTIVE STATISTICS

Ex: 10 to 20

```
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
```

- ii) Variance: A measures variability from the average or mean. It is the expected value of the squared variation of a random variable from its mean value, in probability and statistics.  
Formula:  $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
- iii) Standard Deviation: The square root of variance. A Standard deviation on either side of the mean accounts for 68% of the data.  
Formula:  $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Results for Dispersion for our Dataset

	selling_price	original_price
Range	4.01631E+11	
Variance	419002144.2	672043360.5
Standard Deviation	20469.54187	25923.79911

2) Bivariate analysis

Bi means two and variate means variable, so here there are two variables. The analysis is related to cause and the relationship between the two variables. There are three types of bivariate analysis.

A) Multiple Regression: Multiple regression is a statistical technique that can used to analyze the relationship between a single dependent variable and several independent variables.

Formula:  $y = b_1x_1 + b_2x_2 + \dots + b_nx_n + c$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.979992686							
R Square	0.960305664							
Adjusted R Square	0.960184576							
Standard Error	6500.122577							
Observations	199							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	2.01791E+11	2.01791E+11	4775.947181	4.4803E-140			
Residual	197	8123563922	42251593.51					
Total	198	2.10111E+11						
Coefficients		Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	833.4640067	614.8013845	1.3556385	0.176758197	-378.9728533	2045.901027	-378.9728533	2045.901027
X Variable 1	0.906063434	0.0311079	89.10811908	4.4803E-140	0.88020792	0.931913947	0.88020792	0.931913947

B) Correlation coefficients: Correlation coefficients are used to measure how strong ( robust relation) a relationship is between two variables.

Positive correlation, negative correlation, zero correlation.

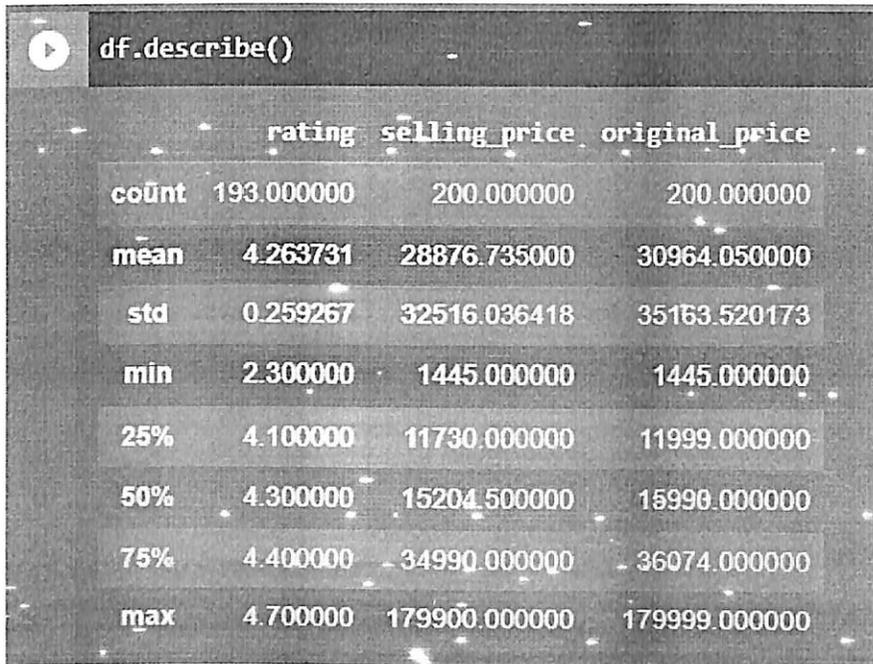
Correlation Coefficient                      0.979992686

3) Categorical Data Analysis

## DESCRIPTIVE STATISTICS

Categorical data is a collection of information that is divided into groups. I.e, if an organisation or agency is trying to get a biodata of its employees, the resulting data is referred to as categorical. This data is called categorical because it may be grouped according to the variables present in the biodata such as sex, state of residence, etc.

Results In python:



```
df.describe()
```

	rating	selling_price	original_price
count	193.000000	200.000000	200.000000
mean	4.263731	28876.735000	30964.050000
std	0.259267	32516.036418	35163.520173
min	2.300000	1445.000000	1445.000000
25%	4.100000	11730.000000	11999.000000
50%	4.300000	15204.500000	15990.000000
75%	4.400000	34990.000000	36074.000000
max	4.700000	179900.000000	179999.000000

DESCRIPTIVE STATISTICS

## **V. Conclusion**

### **Summary**

Above Data Analysis show how much we need to know about the specifications of a mobile phone so that we can buy a perfect product while buying a phone. Technologies are growing rapidly and we should update ourselves to buy a better product for long term use. They will also save money from this knowledge.

### **Future Enhancement**

Write a few points on how to take this work ahead: mention about how to increase/widen the scope of this work.