Assignment No - 1

Q.1 a) What is Descriptive statastics? list out the types of descriptive Statastics.

Ans:- 1) Descriptive statastics describe the characterstics or properties of the data. It helps to summarize the data in a meaningful data in a meaningful way.

2) It allows important patterns to emerge from the data. It also helps to understanding the distribution of data

There are two types of descriptive Statastics :-

1) Measures of Central Tendency :-

It is a single value that attempts to describe a set of data by identifying the central possition within that set of data. Include mean, median and mode.

2) Measures of Spread or dispersion :-

It is way of summarizing a group of data by describing how scores are spread out.

It include range, quantities, variance and standard deviation.

Q.1 b> Wtite in brief
i> Mean ii> Median iii> Mode

Ans :- 1> Mean :-

1> The mean (or average) is the most popullar and well Known measures of central tendency. It can be used with both distance and Contineous data.

2> The mean is equal to the sum of all the values in the data set divided by number of values in the data set.

$$\boxed{\bar{x} = (x_1, x_2 \ldots\ldots x_n/n}$$

2> Median :-

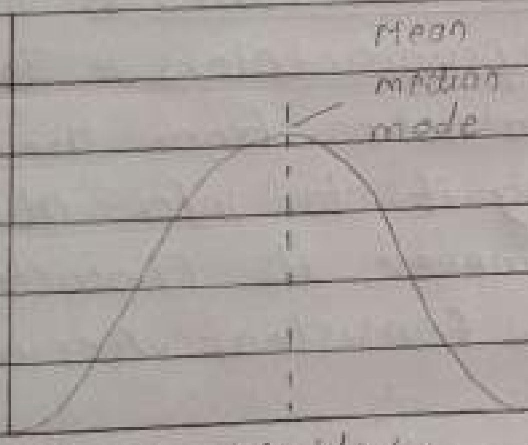
1> The median is the middle Score for a set of data that has been arranged in order of magnitude.

2> The median is less affected by outliers and Skewed data. It is a holistic measure. It is easy method of approximation of median Values of a large data set.

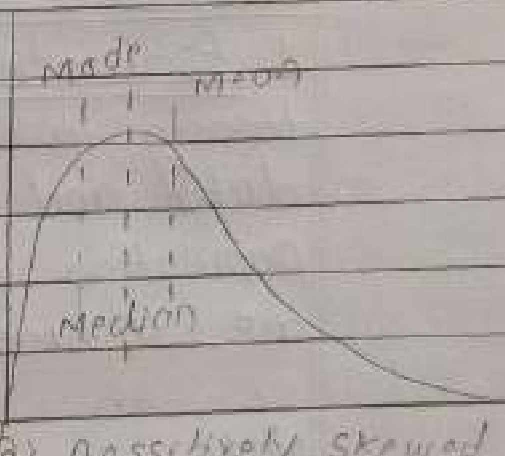
3) Mode :-

1) The mode is the most frequent score in our data set. The mode is used for categorical data where we want to know which is the most common category occuring in the population.

2) The mode is the values that appears most frequently in a data set.



a) symmetric data



b) possitively skewed data

Q.2 a) What is Inferential Statastics ? Explain main two areas of inferential statastics ?

Ans :- 1) Inferential statastics is genaally used when the user needs to make a conclusion about the whole popultation at hand.

2) Inferential statastics use statastical models to compare sample data to other samples or to previous research.

There are main two types :-

1. Estimating parameteters :- This means taking a statastics from the sample data and using it to infer about a population parameter. Its characterstics are unbiased, consistent, Accurecy.

2) Hypothesis tests :-

1) This is where sample data can be used to answer reasearch questions.

2) For example, we might be interested in knowing if a new cancer drug is effective.

Q 2 b) Explain in detail about the Statastical hypothesis.

Ans:-

1) A statastical hypothesis is a formal claim about a State of nature Structured within the framework of a statastical model.

2) A Statastical hypothesis is defined as a statement, which may or may not be true about the population parameter or about the probablity distribution of the parameter that we wish to validate on the basis of Sample information.

3) Most times, experiments are performed with random sample instead of the entire population.

4) In order to have an accurate or more precise interface, the chance factor should be ruled out.

Null hypothesis :-

The probability of chance occurence of the observed is examined by the null hypothesis (HO). Null hypothesis is a statement of no differences.

In ~~Consists Const~~ Contrast to null hypothesis the alternative hypothesis proposes that.

1) The two samples belong to two different populations.

2) Their means are estimates of two different parametric means of the respective population.

3) There is a significant difference between their sample means.

Q.3 Explain in detail Data manipulation.

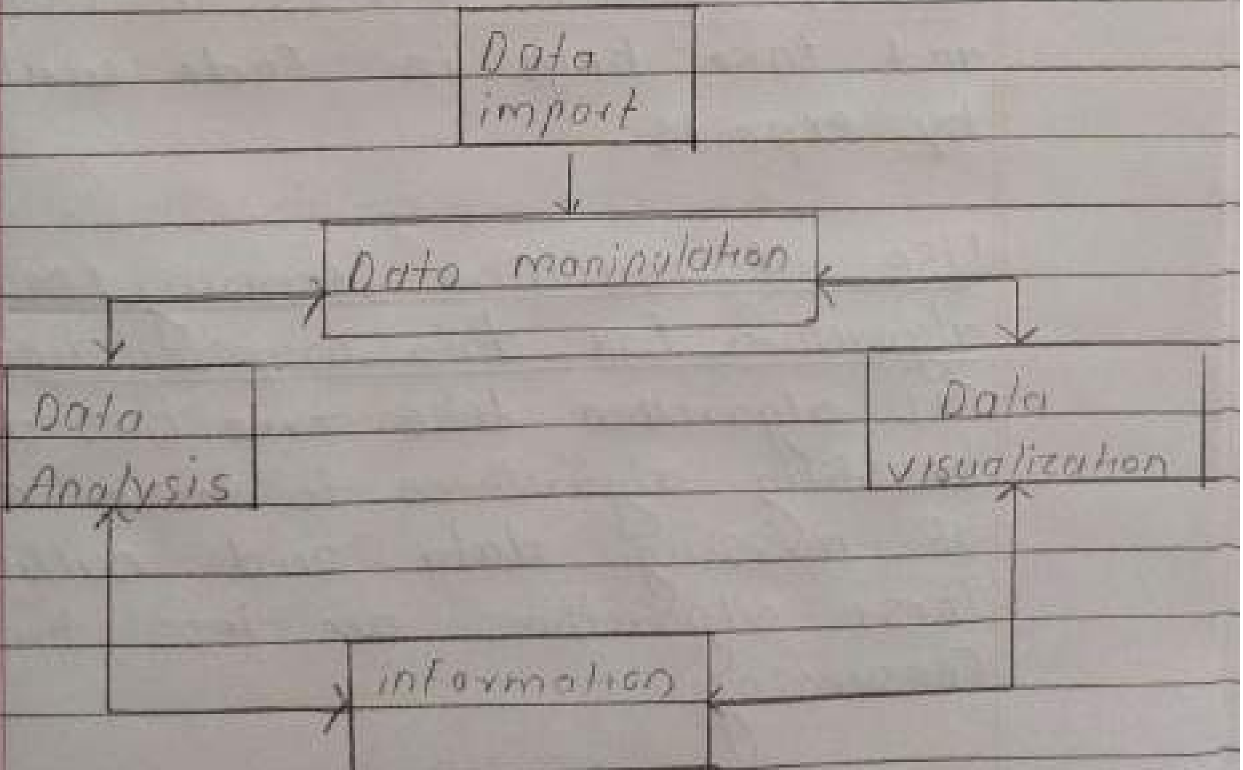
Ans :- 1> Data manipulation is an important phase of predictive modeling.

2> It involves manipulating data using available set of variables. This is done to enhance accuracy and precision associated with data.

3> The data collection proces can have many loopholes.

4> There are various uncontrollable factors which lead to inaccuracy in data such as mental situation of respondents, personal biases, difference / error in readings of machines etc.

Useful data manipulation :-



Q.3 b> Explain different ways to manipulate data.

Ans:- 1.> Manipulating data using inbuilt base R functions :-

This is the first step, but is often repetitive and time consuming Hence it is a less efficient way to solve the problem.

2> Use of packages for data manipulation :-

CRAN has more than 8000 packages available today. These packages are a collection of pre-written commonly used piece of codes. They helps to perform the repetitive tasks fasts, reduce errors in coding and take help in code written by experts.

Use of machine learning (ML) algorithm for data manipulation :-

ML algorithm like tree based boosting algorithms to take care of missing data and outliers. These algorithms are less time consuming.

Q.4 a> Explain the following packages
i> dplyr. Package ii> data.table package.

Ans :- i> dplyr package :-

1> This package is created and maintained by Hadley wickham. This package has everything (almost) to accelerate data manipulation efforts. It known best for data exploration and transformation.

2> It chaining syntax makes it highly adaptive to use. It includes 5 major data manipulation commands.

1. filter
2. select
3. arrange
4. mutate
5. summarise

Use of dplyr package

```
> library (dplyr)
> data ("mtcars")
> data ('iris')
> my data <- mtcars
# read data
> head (mydata).
```

2) data.table package :-

1) This package allows to perform faster manipulation in a data set. A datatable has 3 parts mainly DT[i,j,by].

2) We can tell R to subset the rows using 'i' to calculate 'j' which is grouped by 'by'. Most of the times 'by' relates to categorical variable.

Use of data.table package :-

```
> library(datatable)
```

Q.4. Explain the following with limitation and advantages
i> Scatter plot ii> Histogram.

Ans:- 1> Scatter plot :-

1> A Scatter plot is a graph in which the values of two variables are plotted along two axes, the pattern of the resulting points revaluating any Correlation present.

- Limitations of scatter diagram :-

1> With Scatter diagrams we cannot get exact extent of Correlation.

2> Quantitative measures of the relationship between the variable cannot be viewed. only Show quantitat-~~ative~~ -tive expression.

- Advantages of a Scatter diagram :-

1> Relationship between two variable can be viewed.

2) For a non linear pattan, this is the best method

3> Maximum and minimum value, can be easily determined.

4) plotting diagram very Simple

Date
Page

11> **Histogram :-**

Histogram represents the Frequency distribution of Contineous variables. while, a bar graph is diagramatic Comparison of discrete variables.

Limitations of Histogram :-

1> A Histogram can present data that is misleading as it has many bars.

2> Only two sets of data are used, but to analyze certain types of statastical data, more than two sets of data are necessary.

Advantages of Histogram :-

1> Histogram helps to identify different data. the Frequency of the data occuring in the dataset and categories which are difficult to interpret in a tabular form. It helps visualize the distribution of the data.

Q.5 a) What is data type ? list out the types of data types with example

Ans :- Data type is a collection or grouping of data values, usually specified by a set of possible values, a set of allowed operations on these values.

List of data types :-

1) Numeric
2) Sequence type
3) String
4) Dictionary
4) set.

1) Numeric :- Numeric data type is used to hold numeric values. It include int, Float , Complex

ex
```
num = 5
    print(num1, 'is of type', type(num1))
    num = 2.0
    print(num2, 'is of type', type(num2))
```

2) List :- List is an ordered collection of similar or different types of seperated by commaan.

```
list = ['Name', 'Roll no' ]
```

3) String :- Sequence of of characters represented by either single or double quotes.

ex

```
name = 'python'
print (name)
```

output

```
python
```

4) Dictonary :- python dictonary is an ordered Collection of items.

```
dict = {'Name':'Mayur','Roll no':
        '131'}
print(dict)
```

Output :- {'Name;'Mayur','Roll No':'131'}

5) Set :- set is an unordered Collection of unique items.

ex

```
Student-id = {12, 14, 16}
print (Student-id)
print(type(Student-id)
```

output :-

```
{12, 14, 16}
<class 'set'>
```

Q.5 b> Enumerate the list and its methods with example.

Ans:- The list method is used to define mutiple data in python. The value of any list item can be changed any time.

There are some methods of lists :-

1> python append () :- Used to adding element in list

Syntax : list.append (element)

```
ex  List = ['Math', 'Biology', 1997]
    list.append (20455)
    print (list)
```

2> Python insert () :- Used to inserts element at the specfied position.

Syntax :- list.item (<position, element)

```
ex  list ['Chemistry, 'Math', 2000]
    List.insert ( 2, 10875 )
    print(list)
```

3> Python extend () :- Add constant to List2 to the end of list1

Syntax :- list1 .extend(list2)

```
ex    list1 = [ 1, 2 ]
      List2 = [ 2  ]
    List1.extend (list2)
    print (list1)
```

Q. 6 a) ~~Eluciating~~ Elucidate the String and its methods with example

Ans :- A String is a data Structure in python that represents a Sequence of characters.

The methods of String :-

1) lower() :- Converts all upercase characters in a String into lowercase

2) upper() :- Converts all lowercase characters in a String into uppercase

3) title() :- Convert String to title case

4) Capitalize() : Convert the first Character of a String to uppercase

ex :-

```
text = " geeks For geEks"
print("\n Converted String:")
print( text. ~~upp~~ lower())
print(" \n Converted String:")
print( text. upper())
print("\n Converted String:")
print( text. title())
print(" \n Converted String:")
print( text. ~~title ()~~ Capitalize())
```

output :-

```
geeks for geeks
GEEKS FOR GEEKS
Geeks For Geeks
GEEKS FOR GEEKS.
```

Q.6 b> What is dictionary ? Explain methods available in dictionary ?

Ans :- Dictionary ~~are~~ is mutable data structures that allow you to store key - value pairs. The dictionary can be created using the dict().

The methods available in dictionary are:-

1> Key() :- It use return list of all the available keys in the dictionary

ex :- dict = {'Name': Mayur, 'Rollno': 131}
print (dict . keys())
output: dict-keys ('Name', 'Rollno')

2> values() :- It use returns list of dictionary value from the key value pairs

ex: dict = {'Name': Mayur, 'Rollno': 131}
print (dict . values())
output :- dict-values ('Mayur', '131')

3> Copy() :- This method returns a shallow copy of dictionary.

~~de~~ ex
dict = {'Name': 'Mayur', 'Rollno': '131'}
dict-new = dict . copy()
print (dict-new)
Output :- {'Name': 'Mayur', 'Rollno': '131'}

4) __Update()__ :- The update() inserts new item to the dichonary

__example__:-

```
dict={'Name':'Mayur', 'Rollno':
            '131'}
dict.update({'age': 22})
 print (dict)
```

__Output__:-

```
{'Name':'Mayur', 'Roll No':'131'
  'Age': 22}
```