Practical No. 1

Aim:- Descriptive Statistics

a. Write an program to find basic descriptive statistics using summary, str, quartile function on mtcars & cars datasets.

b. Write an program to find subset of dataset by using subset (), aggregate () functions on iris dataset.

Theory :-

Descriptive statistics is about describing and summarizing data. It uses two main approaches:

1. **The quantitative approach** describes and summarizes data numerically.
2. **The visual approach** illustrates data with charts, plots, histograms, and other graphs.

You can apply descriptive statistics to one or many datasets or variables. When you describe and summarize a single variable, you're performing **univariate analysis**. When you search for statistical relationships among a pair of variables, you're doing a **bivariate analysis**. Similarly, a **multivariate analysis** is concerned with multiple variables at once.

1. describe() :- The describe() method returns description of the data in the DataFrame.
   If the DataFrame contains numerical data, the description contains these information for each column:
   count - The number of not-empty values.
   mean - The average (mean) value.
   std - The standard deviation.
   min - the minimum value.
   25% - The 25% percentile*.
   50% - The 50% percentile*.
   75% - The 75% percentile*.
   max - the maximum value.

2. info() :-The info() method prints information about the DataFrame. The information contains the number of columns, column labels, column data types, memory usage, range index, and the number of cells in each column (non-null values).

3. quantile() :- The quantile() method calculates the quantile of the values in a given axis.
   Default axis is row.
   By specifying the column axis (axis='columns'), the quantile() method calculates the quantile column-wise and returns the mean value for each row.

4. subset() :- The subset function is used to extract a specific subset of the dataset based on certain conditions.

5. groupby() :- The groupby() function in pandas is a powerful tool for grouping data based on one or more criteria and then applying a function to each group independently.

6. agg() :- The agg() function is used to aggregate data in a DataFrame or a Series. This function is often used in conjunction with groupby() for more complex data aggregation tasks.

Input :-

Program a:-

Usage :- mtcars.csv
The mtcars dataset is a well-known dataset in the field of statistics and data analysis. The dataset provides information about various aspects of different car models, particularly focusing on fuel efficiency and performance.

Format :- A data frame with 32 observations (rows) on 11 numeric variables (columns).

	Variables	Description
1.	mpg	Miles/(US) gallon
2.	cyl	Number of cylinders
3.	disp	Displacement (cu.in.)
4.	hp	Gross horsepower
5.	drat	Rear axle ratio
6.	wt	Weight (1000 lbs)
7.	qsec	1/4 mile time
8.	vs	Engine (0 = V-shaped, 1 = straight)
9.	am	Transmission (0 = automatic, 1 = manual)
10.	gear	Number of forward gears
11.	carb	Number of carburetors

Program b:-

Usage :- iris.csv

The Iris dataset is a classic dataset in the field of machine learning and statistics. The dataset consists of measurements of various features of three different species of iris flowers.

	Variables	Description
1.	sepal_length	Represents the length of the iris flower's sepal (the outermost whorl of a flower) in cm.
2.	sepal_width	Represents the width of the sepal of the iris flower in cm.
3.	petal_length	Represents the length of the petal (inner whorl) of the iris flower in cm.
4.	petal_width	Represents the width of the petal of the iris flower in cm.
5.	species	Categorical column represents the species of the iris flower among Setosa, Versicolor, and Virginica.

Program a :-

```
import pandas as pd

mtcars = pd.read_csv('mtcars.csv')
```

```
print("Summary Statistics for mtcars:")
print(mtcars.describe())

print("\nGeneral Information for mtcars:")
print(mtcars.info())

print("\nQuartile Information for mtcars:")
numeric_columns = mtcars.select_dtypes(include=['number']).columns
print(mtcars[numeric_columns].quantile([0.25, 0.5, 0.75]))
```

Program b :-

```
import pandas as pd

iris = pd.read_csv('iris.csv')

print("Original Iris dataset:")
print(iris.head())

subset_condition = (iris['sepal_length'] > 5.0) & (iris['sepal_width'] > 3.0)
print("\nSubset of Iris Dataset:")
print(subset_condition)

aggregate_result = iris.groupby('species').agg({'petal_length': 'mean'})
print("\nAggregate result - Mean petal length for each species:")
print(aggregate_result)
```

Conclusion :- The program to implement descriptive statistics to preform relevant given operations on iris and cars dataset is executed successfully.