**LAB EXERCISE – 1**

**Descriptive Statistics**

**Aim of the Experiment**

The main aim of this experiment is to explore the given dataset. A sample database is created and is available in the file sample.csv.

Sample Dataset

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | first | last | gender | Marks | selected |
| 1 | Leone | Debrick | Female | 50 | TRUE |
| 2 | Romola | Phinnessy | Female | 60 | FALSE |
| 3 | Geri | Prium | Male | 65 | FALSE |
| 4 | Sandy | Doveston | Female | 95 | FALSE |
| 5 | Jacenta | Jansik | Female | 31 | TRUE |
| 6 | Diane-marie | Medhurst | Female | 45 | TRUE |
| 7 | Austen | Pool | Male | 45 | TRUE |
| 8 | Vanya | Teffrey | Male | 70 | FALSE |
| 9 | Giordano | Elloy | Male | 36 | FALSE |
| 10 | Rozele | Fawcett | Female | 50 | FALSE |

The objectives of this experiment are:

1. Explore all the statistical operations of Pandas and given in Listing 1
2. Use Describe command and explore the dataset as given in Listing 2
3. Use Descriptive Statistics for univariate and bivariate data as given in Listing 3

**Reference to the Textbook**

All the fundamentals are given in Chapter 2 and Appendix 2.

**Listing 1**

import pandas as pd

col\_list=["id","first","last","gender","Marks","selected"]

df = pd.read\_csv("sample.csv",usecols=col\_list)

print(df)

mean1 = df['Marks'].mean()

sum1 = df['Marks'].sum()

max1 = df['Marks'].max()

min1 = df['Marks'].min()

count1 = df['Marks'].count()

median1 = df['Marks'].median()

sd1 = df['Marks'].std()

var1 = df['Marks'].var()

print('Mean Marks\n' + str(mean1))

print('Sum of the Marks\n' + str(sum1))

print('Maximum of the marks\n' + str(max1))

print('Minimum of the marks\n' + str(min1))

print('Count of the marks\n' + str(count1))

print('Standard deviation of the marks\n' + str(sd1))

print('Variance of the marks\n' + str(var1))

print('End of Summary \n\n\n')

**Output**

**A screenshot of a computer

Description automatically generated**

**Listing 2**

import pandas as pd

col\_list=["id","first","last","gender","Marks","selected"]

df = pd.read\_csv("sample.csv",usecols=col\_list)

print(df)

print(df.shape)

print(df.head(5))

print(df.describe())

**Output**

**A screenshot of a computer

Description automatically generated**

Listing 3

# Listing 3

# Illustrates the use of univariate, bivariate Statistics

from scipy import stats

data = [1,2,3,4,5,6,8,8,8,8,9,10,10,10,12,14,18,18,18,22,39,44,55,55,55,55,66,78,79,88]

print("\n Details of the data \n", stats.describe(data))

print("\n The cumulative frequency of the data-n",stats.cumfreq(data))

print("\n The Geometric mean of the data-n",stats.gmean(data))

print("\n The Harmonic mean of the data-n",stats.hmean(data))

print("\n The IQR of the data-n",stats.iqr(data))

print("\n The Zscore of the data-n",stats.zscore(data))

print("\n The skewness of the data-n",stats.skew(data))

print("\n The Kurtosos of the data-n",stats.kurtosis(data))

# Check correlation of the data

data = [1,2,3,6,8,10]

data1 = [2,3,4,5,9,12]

print("\n The Spearman correlation of the data-n",stats.spearmanr(data,data1))

**Output**

**Text

Description automatically generated**