Experiment 1

Student Name:Upwan **UID: 21BCS8105**
Branch:CSE **Section/Group: 20BCS_WM-702A**
Semester: 5th **Date of Performance: 24/08/2022**
Subject Name: Machine Learning Lab **Subject Code: 20CSP-317**

Aim/Overview of the practical:

Implement Exploratory Data Analysis on any data set.

Task to be done:

To implement exploratory data analysis on any data set.

Apparatus/Simulator Used:

- **Google Collab**
- **Python**
- **.csv file**

Code and Output:

```
import pandas as pd
import numpy as np
import seaborn as sns
data=pd.read_csv("/content/employees.csv")
data
```

	First Name	Gender	Start Date	Last Login Time	Salary	Bonus %	Senior Management	Team
0	Douglas	Male	08/06/93	12:42 PM	97308	6.945	True	Marketing
1	Thomas	Male	3/31/1996	6:53 AM	61933	4.170	True	NaN
2	Maria	Female	4/23/1993	11:17 AM	130590	11.858	False	Finance
3	Jerry	Male	03/04/05	1:00 PM	138705	9.340	True	Finance
4	Larry	Male	1/24/1998	4:47 PM	101004	1.389	True	Client Services
...	...	...	...	...	...	...	...	...
995	Henry	NaN	11/23/2014	6:09 AM	132483	16.655	False	Distribution
996	Phillip	Male	1/31/1984	6:30 AM	42392	19.675	False	Finance
997	Russell	Male	5/20/2013	12:39 PM	96914	1.421	False	Product
998	Larry	Male	4/20/2013	4:45 PM	60500	11.985	False	Business Development
999	Albert	Male	5/15/2012	6:24 PM	129949	10.169	True	Sales

1000 rows × 8 columns

```
import pandas as pd
import numpy as np
import seaborn as sns
data=pd.read_csv("/content/employees.csv")
data.describe()
```

	Salary	Bonus %
count	1000.000000	1000.000000
mean	90662.181000	10.207555
std	32923.693342	5.528481
min	35013.000000	1.015000
25%	62613.000000	5.401750
50%	90428.000000	9.838500
75%	118740.250000	14.838000
max	149908.000000	19.944000

```
import pandas as pd
import numpy as np
import seaborn as sns
data=pd.read_csv("/content/employees.csv")
data.shape
```

```
(1000, 8)
```

+ Code + Text

```
import pandas as pd
import numpy as np
import seaborn as sns
data=pd.read_csv("/content/employees.csv")
data.columns
```

```
Index(['First Name', 'Gender', 'Start Date', 'Last Login Time', 'Salary',
       'Bonus %', 'Senior Management', 'Team'],
      dtype='object')
```

```
import pandas as pd
import numpy as np
import seaborn as sns
data=pd.read_csv("/content/employees.csv")
data.nunique()
```

```
First Name           200
Gender                 2
Start Date           972
Last Login Time      720
Salary               995
Bonus %              971
Senior Management      2
Team                  10
dtype: int64
```

Learning outcomes (What I have learnt):

1. **Learnt how to implement exploratory data for analysis**
2. **Learnt about numpy, seaborn, pandas libraries.**
3. **Learnt about different method to implement exploratory data for analysis**

Evaluation Grid :

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.	Student Performance (Conduct of experiment) objectives/Outcomes.		12
2.	Viva Voce		10
3.	Submission of Work Sheet (Record)		8
	Total		30