Task 1:- Data Overview

Objective: Understand the dataset structure.

```
from google.colab import drive
drive.mount('/content/drive')
```

```
     Mounted at /content/drive
```

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
import pandas as pd
import statsmodels.api as sm
from sklearn.model_selection import train_test_split
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import plotly.express as exp
import statsmodels.formula.api as smf
```

```
data=pd.read_excel("/content/COGNIFYZ.xlsx")
data
```

	gender	age	Investment_Avenues	Mutual_Funds	Equity_Market	Debentures	Government_Bonds	Fixed_Deposits	PPF	Gold	...	Duration	Invest_M
0	Female	34	Yes	1	2	5	3	7	6	4	...	1-3 years	
1	Female	23	Yes	4	3	2	1	5	6	7	...	More than 5 years	
2	Male	30	Yes	3	6	4	2	5	1	7	...	3-5 years	
3	Male	22	Yes	2	1	3	7	6	4	5	...	Less than 1 year	
4	Female	24	No	2	1	3	6	4	5	7	...	Less than 1 year	
5	Female	24	No	7	5	4	6	3	1	2	...	1-3 years	
6	Female	27	Yes	3	6	4	2	5	1	7	...	3-5 years	
7	Male	21	Yes	2	3	7	4	6	1	5	...	3-5 years	
8	Male	35	Yes	2	4	7	5	3	1	6	...	1-3 years	
9	Male	31	Yes	1	3	7	4	5	2	6	...	3-5 years	
10	Female	35	Yes	2	4	7	5	3	1	6	...	3-5 years	
11	Male	29	Yes	2	5	7	6	3	1	4	...	1-3 years	
12	Female	21	No	1	2	3	4	5	6	7	...	1-3 years	

13	Female	28	Yes	2	3	7	4	5	1	6	...	1-3 years
14	Female	25	Yes	2	3	7	5	4	1	6	...	1-3 years
15	Male	27	Yes	2	3	7	5	4	1	6	...	1-3 years
16	Female	28	Yes	3	2	7	5	4	1	6	...	1-3 years
17	Male	27	Yes	3	2	7	4	5	1	6	...	1-3 years
18	Male	29	Yes	3	2	7	4	5	1	6	...	1-3 years
19	Male	26	Yes	3	4	6	5	1	2	7	...	3-5 years
20	Male	29	Yes	2	4	7	5	3	1	6	...	3-5 years
21	Female	24	Yes	2	4	5	6	3	1	7	...	3-5 years
22	Male	27	Yes	3	4	6	5	2	1	7	...	3-5 years
23	Male	25	Yes	2	4	6	5	3	1	7	...	3-5 years
24	Female	26	Yes	2	3	7	5	4	1	6	...	3-5 years
25	Female	32	Yes	3	4	7	5	1	2	6	...	3-5 years
26	Male	26	Yes	3	4	6	5	1	2	7	...	3-5 years

```
27    Male  31   Yes   2   3   7   6   4   1   5   ...   1-3 years
28    Male  29   Yes   2   3   6   5   1   4   7   ...   1-3 years
29  Female  34   Yes   5   4   3   2   7   1   6   ...   3-5 years
30    Male  27   Yes   4   5   1   2   7   3   6   ...   1-3 years
31  Female  31   Yes   2   4   7   6   3   1   5   ...   3-5 years
32    Male  27   Yes   2   4   7   5   1   3   6   ...   3-5 years
33    Male  26   Yes   2   3   6   4   1   5   7   ...   1-3 years
34    Male  27   Yes   2   3   6   5   4   1   7   ...   1-3 years
35    Male  30   Yes   1   4   6   5   3   2   7   ...   3-5 years
36    Male  30   Yes   2   4   7   5   1   3   6   ...   1-3 years
37    Male  25   Yes   5   4   7   6   1   2   3   ...   3-5 years
38    Male  31   Yes   2   4   7   5   3   1   6   ...   1-3 years
39    Male  29   Yes   4   3   5   7   2   1   6   ...   3-5 years
40 rows × 24 columns
```

```
data.head()
```

	gender	age	Investment_Avenues	Mutual_Funds	Equity_Market	Debentures	Government_Bonds	Fixed_Deposits	PPF	Gold	...	Duration	Invest_Mc
0	Female	34	Yes	1	2	5	3	7	6	4	...	1-3 years	M
1	Female	23	Yes	4	3	2	1	5	6	7	...	More than 5 years	\
2	Male	30	Yes	3	6	4	2	5	1	7	...	3-5 years	
3	Male	22	Yes	2	1	3	7	6	4	5	...	Less than 1 year	
4	Female	24	No	2	1	3	6	4	5	7	...	Less than 1 year	

5 rows × 24 columns

```
data.shape
```

```
(40, 24)
```

Interpretation:-

The dataset has 40 rows and 24 columns.

There are 40 entries or observations in the dataset, and each observation has 24 attributes or features.

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40 entries, 0 to 39
Data columns (total 24 columns):
 #   Column                             Non-Null Count  Dtype 
---  ------                             --------------  ----- 
 0   gender                             40 non-null     object
 1   age                                40 non-null     int64 
 2   Investment_Avenues                 40 non-null     object
 3   Mutual_Funds                       40 non-null     int64 
 4   Equity_Market                      40 non-null     int64 
 5   Debentures                         40 non-null     int64 
 6   Government_Bonds                   40 non-null     int64 
 7   Fixed_Deposits                     40 non-null     int64 
 8   PPF                                40 non-null     int64 
 9   Gold                               40 non-null     int64 
 10  Stock_Marktet                      40 non-null     object
 11  Factor                             40 non-null     object
 12  Objective                          40 non-null     object
 13  Purpose                            40 non-null     object
 14  Duration                           40 non-null     object
 15  Invest_Monitor                     40 non-null     object
 16  Expect                             40 non-null     object
 17  Avenue                             40 non-null     object
 18  What are your savings objectives?  40 non-null     object
 19  Reason_Equity                      40 non-null     object
 20  Reason_Mutual                      40 non-null     object
 21  Reason_Bonds                       40 non-null     object
 22  Reason_FD                          40 non-null     object
 23  Source                             40 non-null     object
dtypes: int64(8), object(16)
memory usage: 7.6+ KB
```

Interpretation:-

The data types include integers (int64), objects (object, typically representing strings), and categorical variables such as gender, investment avenues, etc.

```
data.columns
```

```
Index(['gender', 'age', 'Investment_Avenues', 'Mutual_Funds', 'Equity_Market',
       'Debentures', 'Government_Bonds', 'Fixed_Deposits', 'PPF', 'Gold',
       'Stock_Marktet', 'Factor', 'Objective', 'Purpose', 'Duration',
       'Invest_Monitor', 'Expect', 'Avenue',
       'What are your savings objectives?', 'Reason_Equity', 'Reason_Mutual',
       'Reason_Bonds', 'Reason_FD', 'Source'],
      dtype='object')
```

The data.columns gives us the Information about the total number of columns and their Names.

```
data.isnull().sum()
```

```
gender                                 0
age                                    0
Investment_Avenues                     0
Mutual_Funds                           0
Equity_Market                          0
Debentures                             0
Government_Bonds                       0
Fixed_Deposits                         0
PPF                                    0
Gold                                   0
Stock_Marktet                          0
Factor                                 0
Objective                              0
Purpose                                0
Duration                               0
Invest_Monitor                         0
Expect                                 0
Avenue                                 0
What are your savings objectives?      0
Reason_Equity                          0
Reason_Mutual                          0
Reason_Bonds                           0
Reason_FD                              0
Source                                 0
dtype: int64
```

From the above result , we get to know that there are no null values included in the datset. Hence can proceed with Further Statistical Analysis.

```
print(data.describe())
```

```
             age  Mutual_Funds  Equity_Market  Debentures  Government_Bonds  \
count  40.000000     40.000000      40.000000   40.000000         40.000000   
mean   27.800000      2.550000       3.475000    5.750000          4.650000   
std     3.560467      1.197219       1.131994    1.675617          1.369072   
min    21.000000      1.000000       1.000000    1.000000          1.000000   
25%    25.750000      2.000000       3.000000    5.000000          4.000000   
```

```
50%    27.000000      2.000000       4.000000    6.500000          5.000000
75%    30.000000      3.000000       4.000000    7.000000          5.000000
max    35.000000      7.000000       6.000000    7.000000          7.000000

       Fixed_Deposits        PPF       Gold
count       40.000000  40.000000  40.000000
mean         3.575000   2.025000   5.975000
std          1.795828   1.609069   1.143263
min          1.000000   1.000000   2.000000
25%          2.750000   1.000000   6.000000
50%          3.500000   1.000000   6.000000
75%          5.000000   2.250000   7.000000
max          7.000000   6.000000   7.000000
```

Interpretations:- The Descriptive Statistics is given above.The descriptive statistics include Total Count , Mean, Standard Deviation , Minimum, Maximum and the Quantiles (25%,50%,75%).