

Task 1: Data Overview Internship at Cognifyz Technologies

AJINKYA KUTARMARE (BUSINESS ANALYTICS INTERN)





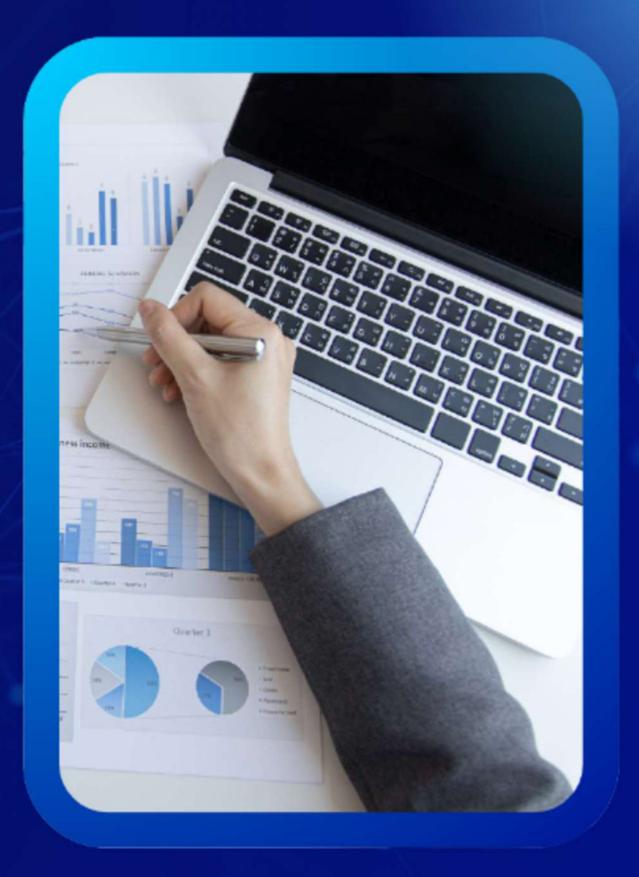
Objective: Understand the dataset structure

The objective of this task was to load a given dataset into a data analysis tool (such as Python with pandas) and perform an initial analysis to understand its structure. This involved importing the dataset, checking for basic information like the number of entries, columns, and data types, and using descriptive statistics to ensure the dataset is well-organized. By performing this analysis, I gained insights into the dataset's composition, which is crucial for further data processing and analysis.



Load the Dataset

- Imported the pandas library to handle and analyze the dataset.
- Used pd.read_csv() to load the dataset.
- Stored the loaded dataset in a pandas DataFrame named df.
- Previewed the first 5 rows of the dataset using df.head() to check its structure and ensure the data loaded correctly.





```
#pandas library
   import pandas as pd
   #dataset
   df = pd.read_csv(r'C:/Users/bhard/Desktop/DATA ANALYST OFFLINE/Python/internship/Dataset.csv')
   #first 5 rows of the dataset
   print(df.head())
         age Investment_Avenues Mutual_Funds Equity_Market Debentures \
  gender
0 Female
                            Yes
  Female
           23
                            Yes
    Male
           30
                            Yes
    Male
           22
                            Yes
4 Female
           24
                            No
  Government_Bonds Fixed_Deposits PPF Gold ...
                                                    Duration \
                                                        1-3 years
                                          7 ... More than 5 years
                                                         3-5 years
                                          5 ... Less than 1 year
                                                  Less than 1 year
                                          7 ...
 Invest Monitor
                              Avenue What are your savings objectives? \
                 Expect
        Monthly 20%-30% Mutual Fund
                                                     Retirement Plan
         Weekly 20%-30% Mutual Fund
                                                         Health Care
          Daily 20%-30%
                              Equity
                                                     Retirement Plan
          Daily 10%-20%
                              Equity
                                                      Retirement Plan
```



Descriptive Statistics

- Imported the pandas library to analyze the dataset's structure and statistics.
- Used df.info() to gather details about the dataset, such as the number of entries, columns, and data types.
- Analyzed the dataset's structure, ensuring all 40 entries had valid data across 24 columns with no missing values.
- Used df.describe() to calculate basic statistical measures (e.g., mean, standard deviation, min, max) for the numeric columns.
- The descriptive statistics helped in understanding the distribution and range of values for numerical attributes in the dataset.





```
df_info = df.info()
   # basic statistical details of the dataset (e.g., mean, std, min, max, etc.)
   df_description = df.describe()
   # Show the output
   print("Dataset Information:")
   print(df_info)
   print("\nDataset Statistical Summary:")
   print(df_description)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40 entries, 0 to 39
Data columns (total 24 columns):
    Column
                                        Non-Null Count Dtype
                                                        object
    gender
                                        40 non-null
                                        40 non-null
                                                        int64
    age
                                        40 non-null
                                                       object
    Investment_Avenues
    Mutual_Funds
                                                        int64
                                        40 non-null
    Equity_Market
                                        40 non-null
                                                        int64
                                        40 non-null
                                                        int64
    Debentures
    Government Bonds
                                                        int64
                                        40 non-null
    Fixed_Deposits
                                                        int64
                                       40 non-null
                                        40 non-null
                                                        int64
    PPF
                                                        int64
    Gold
                                        40 non-null
    Stock Marktet
                                                        object
                                        40 non-null
                                                        object
 11 Factor
                                        40 non-null
 12 Objective
                                        40 non-null
                                                        object
```



Conclusion & Next Steps

Conclusion:

- Task 1: Data Overview completed successfully (Dataset loaded and analyzed using descriptive statistics).
- Task 2: Gender distribution analysis will be the next step, which will include visualization of the dataset's gender distribution.

Next Steps:

 Task 2: Create and visualize gender distribution (bar chart/pie chart).

