**DATA SCIENCE FUNDAMENTALS**

**Assignment-01**

**Date- 07/09/23**

Problem Statement – Analyze any dataset

Dataset of housing from kaggle and performed exploratory data analysis using pandas-profiling and using other python functions.

Dataset link - https://www.kaggle.com/datasets/ashydv/housing-dataset

ydata-profiling delivers an extended analysis of a DataFrame while allowing the data analysis to be exported in different formats such as **html** and **json**.

Here final report of analysis is saved in html format

Repository – [click\_here](https://github.com/jainulhasan/DSF_assignment/tree/main/dsf) (download housing.html file and then open to see report)

Step by Step exploratory data analysis using python functions.

1. The [df.info](http://df.info/)() function will give us the basic information about the dataset

#Basic information

df.info()

#Describe the data- descriptive statistics

df.describe()

1. df.duplicate.sum() function to the sum of duplicate value present if any. It will show the number of duplicate values if they are present in the data.

#Find the duplicates

df.duplicated().sum()

1. find the number of unique values in the particular column using unique() function in python.

df[‘area’].unique()

1. Know the datatypes –

#Datatypes

df.dtypes