National Institute of Technology, Warangal
Department of Computer Science and Engineering
Subject - Data Science Fundamentals

Date: 13/09/23
Class Assignment
Mtech I year I Semester (CSE)

Name: Jainul Hasan
Id: 23CSM1R09

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 (download housing.html file and then open to see report)

Step by Step exploratory data analysis using python functions.

1) The df.info() function will give us the basic information about the dataset

```
#Basic information

df.info()

#Describe the data- descriptive statistics

df.describe()
```

2) 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()
```

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

df[‘area’].unique()

4) Know the datatypes –

#Datatypes

df.dtypes

Other than this ydata_profiling gives a detailed descriptions of each attribute, and relation of attributes with other attributes, screenshots of some analysis done by ydata_profiing are given below.