Course: **Data Mining** 08-Jan-2023 **(Fall 2022)**

Resource Person: **Dr. Muhammad Faheem**

ASSIGNMENT–1 (Introduction and Getting to know Data)

Total Points: 30
Submission Due: Saturday Jan 14, 2023
(Google Classroom Course Page)

Instructions: Please Read Carefully!

- **This is an individual assignment. Everyone is expected to complete the given assignment on their own, without seeking any help from any website or any other individual.** There will be strict penalties for any work found copied from any source and the university policy on plagiarism will be strictly enforced.
- Assignment is to be submitted via Google Classroom.
    - You should already have created your account on Google Classroom as per my earlier email. If not, then follow the link in that email to create your account.
    - Submit your assignment on or before due date. **No late submissions will be possible**.
    - The viva of this assignment will be conducted.

Question 1: **[5]**

How is a data warehouse different from a relational database? Explain your justification with a set of examples.

Question 2: **[5]**

Compare and discuss the role of different statistical description techniques for data preprocessing step. Discuss each measure of central tendency and dispersion of data and elaborate which one should be preferred in a specific situation.

Question 3: **[10]**

Suppose that the values for a given set of data are grouped into intervals. The intervals and corresponding frequencies are as follows:

Work Experience	**Frequency**
1-2	100
3-4	350
5-6	600
7-8	400
9-10	300
11-12	400
13-14	15

Compute an approximate median value for the data.

Question 4: **[10]**

Suppose that a hospital tested the age and body fat data for 18 randomly selected adults with the following results:

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2

age	52	54	27	27	39	41	47	49	50
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

a) Draw the boxplots for age and %fat.
b) Draw a scatter plot and q-q plot based on these two variables.