

Data Analysis and Visualization (DS3001)

Date: September 21st 2024

Course Instructor(s)

Ms. Umme Ammarah

Ms. Esha Tur Razia Babar

Final Exam

Total Time (Hrs): 1

Total Marks: 40

Total Questions: 3

Roll No

Section

Student Signature

Do not write below this line

- Attempt all parts of a question together and in order.
- Solve question # 2 on Question paper.
- Question#1 and 3, should be solved on answer sheet, show all the working to get full marks.

CLO #1:

Q1: The following data set is used to understand the relationship of different variables with a person's per hour salary in USD.

$$y = mx + c$$

ID	Year of Birth	Gender	Weight (kg)	Working hours per day	Work Experience (months)	Salary (USD) (per hour)
1	1998	M	94 <u>1</u>	10	<u>(-29)</u> <u>Outliers</u>	12
2	1999	F	54 <u>0.46</u>	8	12	18
3	2000	F	63 <u>0.58</u>	16	18	15
4	2002	M	55 <u>0.48</u>	24	0	NA <u>52.20</u>
5	1994	M	93 <u>0.98</u>	13	46	25
6	1998	NA	45 <u>0.34</u>	12	5	17
7	1998	M	63 <u>0.58</u>	16	18	25
8	1998	F	41 <u>0.29</u>	13	<u>21</u>	NA <u>34.31</u>
9	1999	M	19 <u>0</u>	14	25	45
10	2000	F	27 <u>0.10</u>	15	6	85
11	2001	M	42 <u>0.30</u>	10	<u>(129)</u>	62

$$m = 1.626 \quad C = 13.178$$

- 1) Suppose there is a linear relationship between working hours per day and Salary (USD) while all other variables are redundant. Use linear regression imputation to impute missing values in the salary (USD) column. Show all working to obtain full marks. (10 marks)
- 2) Apply IQR methods on Work Experience column to find outliers. (5 marks) IQR = 20 LB = -25 UB = 55
- 3) Apply scaling on weight column to map all weight values to the range of [0,1]. (5 marks)

$$\left(\frac{V - \min}{\max - \min} \right) (\text{new max} - \text{new min}) + \text{new min}$$

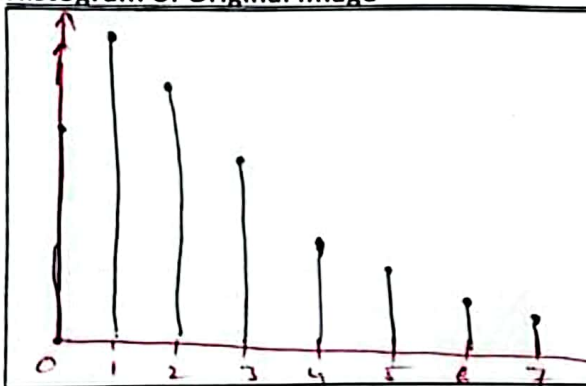
$$\frac{V - 19}{94 - 19} = \frac{V - 19}{75}$$

CLO # 1

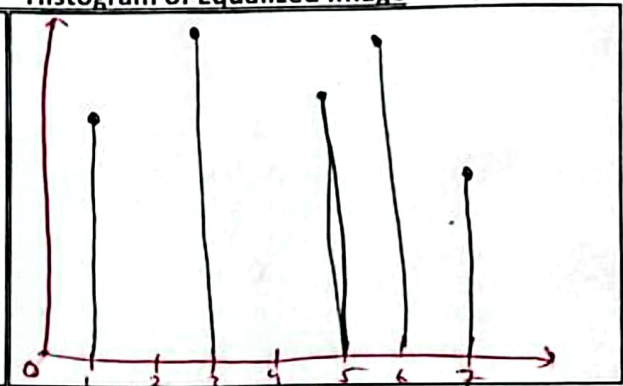
Q2: Consider a digital 3-bit image of size 64 x 64. Frequency distribution of grey-levels are given below. You are required to perform histogram equalization of this image by filling the table given. Draw histograms of the image before and after the equalization process. [10 marks]

r_k	n_k	$P_r(r_k)$	$S_k(CDF)$	Equalized Levels	Final Values
0	790	0.19	0.19	1.33	1
1	1023	0.25	0.44	3.08	3
2	850	0.21	0.65	4.55	5
3	656	0.16	0.81	5.67	6
4	329	0.08	0.89	6.23	6
5	245	0.06	0.95	6.65	7
6	122	0.03	0.98	6.86	7
7	81	0.02	1	7	7

Histogram of Original Image



Histogram of Equalized Image



CLO #: 1

Q3: You are given an image of size 4x4 and a weighted median filter of size 3x3. Apply filter on the image, output and input image should be of same size. Use mirror/reflection padding. [10marks]

3	7	6	2
2	5	6	2
4	7	2	0
3	5	6	3

4	3	1
5	6	2
1	3	2

5	5	6	6
5	5	6	2
5	5	5	2
5	5	5	2