Data Analytics (IT – 3006)
Practice Questions (Unit 3)

Q1. Explain the difference between data analysis and data analytics.

Q2. Consider the dataset 1 for a pizza franchise wherein the independent variable depicts the annual franchise fees (X) and the dependent variable (Y) represents start up cost. You have been hired to build a simple linear regression model and to determine the possible relationship among the variables.

Dataset 1

X	Y
1000	1050
1125	1150
1087	1213
1070	1275
1100	1300
1150	1300
1250	1400
1150	1400
1100	1250
1350	1830
1275	1350
1375	1450
1175	1300
1200	1300
1175	1275

Q3. The dataset 2 represents water well from a random sample of wells in Northwest Texas, USA. In the data, X represents pH of well water and Y represents Bicarbonate (parts per million) of well water. With box plot analysis, compare the two variables, determine the correlation coefficient and subsequently, establish the simple linear regression model.

Dataset 2

X	Y
7.6	157
7.1	174
8.2	175
7.5	188
7.4	171
7.8	143
7.3	217
8	190
7.1	142
7.5	190
8.1	215
7	199

7.3	262
7.8	105
7.3	121

Q4. The dataset 3 represents fire and thefts in Chicago, USA. In the data, X represents fires per 1000 housing units and Y represents thefts per 1000 population. Determine outliers for each variable and subsequently, establish the simple linear regression model.

Dataset 3

X	Y
6.2	29
9.5	44
10.5	36
7.7	37
8.6	53
34.1	68
11	75
6.9	18
7.3	31
15.1	25
29.1	34
2.2	14
5.7	11
2	11
2.5	22
4	16
5.4	27
2.2	9
7.2	29
15.1	30

Q5. Explain the importance of data analysis.

Q6. Explain descriptive, predictive, diagnostic, and prescriptive analytics.

Q7. How can the missing value are handled before establishing the simple linear regression?

Q8. How is overfitting different from underfitting?

Q9. The dataset 4 represents cricket chirps vs. temperature, wherein X is chirps/sec for the striped ground cricket and Y is temperature in degrees Fahrenheit. Draw the normal distribution for each variable and subsequently, establish the simple linear regression model.

Dataset 4

X	Y
20	88.59
16	71.59
19.79	93.30
18.39	84.30

17.10	80.59
15.50	75.19
14.69	69.69
17.10	82
15.39	69.40
16.20	83.30
15	79.59
17.20	82.59
16	80.59
17	83.5
14.39	76.30

Q10. How can data quality be ensured for establishing a simple linear regression model?

Q11. Who are the final users of simple linear regression model analysis results?

Q12. How can you create a data-driven class room?

Q13. How do you find the data of independent and dependent variable are meaningful to establish the model?

Q14. How do you keep improving the simple linear regression model?

Q15. What data visualizations should you choose to understand the relationship between independent and dependent variable?

*** The End ***