Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

VII SEMESTER B. TECH (ELECTRICAL & ELECTRONICS ENGINEERING)

END SEMESTER EXAMINATIONS, NOVEMBER 2022

DATA ANALYTICS [ELE 4077]

REVISED CREDIT SYSTEM

Time: 3 Hours **Date: 21 NOVEMBER 2022** **Max. Marks: 50**

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data may be suitably assumed.

1A. In an excel sheet, column A contains the names of 25 restaurants in and around Manipal, column B has their rating out of 5 and Column C contains their contact number. Write separate excel formulas to answer the following questions.

a) Some of the contact numbers are missing in column C. How many contact numbers are missing?

b) How many eateries have a rating higher than the average score?

c) How many eateries are above 50th percentile of their rating? ***(4)***

1B. A dataset has mean of 35 and standard deviation of 5. Assume that all the data points in the dataset are normally distributed. Find the probabilities of the following random variable X

i) $25 < X < 45$

ii) $25 < X < 50$

iii) $X < 40$ ***(3)***

1C. The amount of paracetamol specified by a drug regulatory authority is 500 mg with an allowed error of 10%. From one of the manufacturing units of a paracetamol manufacturing company, a drug inspector collects 100 samples having average amount of paracetamol as 530 mg with standard deviation of 100 mg. For which confidence interval is the paracetamol tablets from this manufacturer consumable?

Confidence Interval	Z^*
90%	± 1.65
95%	± 1.96
99%	± 2.58

(3)

2A. A manufacturer claims that the average life of its product is 36 months. An auditor selects a sample of 49 units which has average life of 34.5 months and standard deviation of 4 months. Develop a hypothesis and perform z-test to test the hypothesis. ***(4)***

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07
+0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790
+0.1	.53983	.54380	.54776	.55172	.55567	.55966	.56360	.56749
+0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642
+0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431
+0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082
+0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566
+0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857
+0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935
+0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785
+0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398
+1	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769
+1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900
+1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796
+1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466
+1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922
+1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179
+1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254
+1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164
+1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926
+1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558
+2	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077
+2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500
+2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840

2B. Briefly discuss the two types of errors in hypothesis testing with an example. ***(3)***

2C. Discuss bias-variance tradeoff? ***(3)***

3A. The table below contains the categorization of cricket players and non-cricket players based on gender and age. Using Gini Index find the most homogenous split. Verify your result by using entropy.

	Age < 50	**Age > 50**
Gender: Female	Playing: 10 Non-playing: 390	Playing: 0 Non-playing: 100
Gender: Male	Playing: 250 Non-playing: 50	Playing: 50 Non-playing: 150

(4)

3B. What is truncation of decision trees? Also briefly discuss its various methods. ***(3)***

3C. Discuss the steps involved in bagging of random forests? ***(3)***

4A. Discuss the steps involved in creating *K* clusters. ***(3)***

4B. Discuss briefly one-step and multi-step validations using suitable example. ***(4)***

4C. Explain briefly the various methods employed in handling missing values in a time series dataset. ***(3)***

5A. Briefly discuss simple exponential smoothing technique. ***(3)***

5B. How are autoregressive models different from moving average models? ***(4)***

5C. Briefly explain autocorrelation and partial-autocorrelation functions. ***(3)***