

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**School of Computing Science and Engineering****Master of Computer Applications  
Mid Term Examination - Mar 2024****Duration : 90 Minutes  
Max Marks : 50****Sem IV - MCAN2412 - Data Analytics with Python**General Instructions*Answer to the specific question asked**Draw neat, labelled diagrams wherever necessary**Approved data hand books are allowed subject to verification by the Invigilator*

- 1) The probability that at least one of the events Q and R occur is 0.6. If Q and R have probability of occurring together as 0.2, then what will be  $P(Q) + P(R)$ ? K2 (2)
- 2) Describe the sources of data commonly used in data analytics. Provide examples of each source and explain their significance in data analysis K1 (3)
- 3) Discuss the characteristics of data that make it suitable for analysis. How do these characteristics differ across structured, semi-structured, and unstructured data? K2 (4)
- 4) If two dice are thrown , what is the probability that the sum is a) Greater than 8 b) less than 6 c) neither 7 nor 11 K2 (6)
- 5) Provide examples of real-world applications of data analytics across different industries (e.g., retail, healthcare, finance). How do these applications demonstrate the value of data analytics? K3 (6)
- 6) A Survey conducted by a bank revealed that 40% of the accounts are savings accounts and 35% of the accounts are current accounts and the balance are loan accounts. K3 (9)

What is the probability that an account taken at random is a loan account ?

What is the probability that an account taken at random is NOT savings account ?

What is the probability that an account taken at random is NOT a current account

What is the probability that an account taken at random is a current account or a loan account?

- 7) Outline the stages involved in the analytic process, from data collection to insights generation. Discuss the role of various tools and technologies at each stage. K4 (8)
- 8) A clothing company produces men's jeans. The jeans are made and sold with either a regular cut or a boot cut. In an effort to estimate the proportion of their men's jeans market in Oklahoma City that prefers boot-cut jeans, the analyst takes a random sample of 212 jeans sales from the company's two Oklahoma City retail outlets. Only 34 of the sales were for boot-cut jeans. Construct a 90% confidence interval to estimate the proportion of the population in Oklahoma City who prefer boot-cut jeans. K4 (12)

**OR**

An RTO department has observed that on an highway more number of accidents have taken place in the early hours of the day than other timings and wish to associate the outcome of the accidents with timings. A survey findings shows that, of the 400 accident cases studied, 280 had met with accident in the early hours and 99 of them were fatal. Further, those who met with accident in the early hours and died was 80. Does this data indicate any association between the time of accident and fatality of the accident. Use  $\alpha = 0.05$ . K4 (12)