K. K. Wagh Institute of Engineering Education & Research, Nashik
(An Autonomous Institute From A.Y. 2022-23)

SUMMER-2024	
Exam Seat No.:	
Academic Year:2023-2024	Semester:III
Class:PG-II	Program:MCA
Branch Code:M.C.A.	Pattern:2022
Name of Course:Elective II:C: Industry Elective	Course Code:MCA223004C
Max. Marks:60	Duration:2.30 Hrs.

Instructions: Candidates should read carefully the instructions printed on the Question Paper and on the cover page of the Answer Book, which is provided for their use.

1. This question paper contains __page(s).
2. Answer to each new question is to be started on a new page.
3. Assume suitable data wherever required, but justify it.
4. Draw the neat labelled diagrams, wherever necessary.
5. The last columns indicates the Course Outcome and level of Blooms Taxonomy of the Question/sub-question.

Question No. 1 Attempt following Question

1a)   Explain the Sources of Data and Challenges in Data Processing   (6)   CO1

Question No. 2 Attempt following Question

2a)   Explain the data manipulation terms with suitable examples   (6)   CO2

    i)   Labelling   ii) Ordering   iii) Slicing   iv) Filtering

Question No. 3 Attempt following Question

3a)   Illustrate the methods using python libraries with sample codes   (8)   CO3

    i) Import Data Set and   ii) Explore the Data Set

OR

3b)   Illustrate the methods for   (8)   CO3

    i) Finding dirty data   ii) Finding Duplicate data   iii) Rebuilding missing data.

3c)   Predict the importance of Data cleaning and data processing.   (8)   CO3

OR

3d)   Demonstrate a simple shell script to build a data cleaning pipeline using python libraries.   (8)   CO3

Question No. 4 Attempt following Question

4a)	Predict the challenges in data visualization.	(8)	CO4

OR

4b)	Examine the conventional tools used for data visualization.	(8)	CO4

4c)	Illustrate different techniques used for visual data representation.	(8)	CO4

OR

4d)	Examine the Case Study of Analysis of a business problem of Zomato using visualization.	(8)	CO4

Question No. 5 Attempt following Question

5a)	Illustrate Apache Spark Architecture with labelled diagram.	(8)	CO5

OR

5b)	Identify and classify the computational trade-offs in a Spark application.	(8)	CO5

5c)	Illustrate the statistical and machine learning methods for Modelling data.	(8)	CO5

OR

5d)	Analyse the features of Microsoft Power BI.	(8)	CO5

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX