

ROLL NO. _____

ENROLLMENT NO. _____

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, SCHOOL OF ENGINEERING
SCIENCES & TECHNOLOGY, JAMIA HAMDARD**

SUPPLEMENTARY EXAMINATION 2025

MCA PE221

Data Science and Analytics

TIME: 2.5 HRS**MAXIMUM MARKS: 60**

Note: Do not write anything on paper except Roll No. & Enrollment No. Use of mobile phone is not allowed. Please follow proper numbering while answering the questions Proper steps are required to be specified for all questions

SECTION A

*All questions of Section A are compulsory (4*5)*

S.NO	QUESTIONS	MARKS	CO	BL															
Q1.	At an e-commerce platform, how would you classify fruits and vegetables from the image data, ask questions, make assumptions and preprocessing, specify end to end data science workflow and perform conclusion.	5	CO2	L3															
Q2.	2. The following results are from a prospective study that considered predictors of mammography use in women. The investigators used logistic regression to analyze their data. Table 3 Results of a logistic regression predicting annual mammography use <table><tr><th>Variable name</th><th>Parameter estimate (std error)</th><th>Significance value</th></tr><tr><td>Family history-associated risk group</td><td>.14 (.09)</td><td>not sig.</td></tr><tr><td>Age</td><td>-.04 (.02)</td><td><0.05</td></tr><tr><td>Worry</td><td>-.04 (.01)</td><td><0.05</td></tr><tr><td>Worryx family history-associated risk group</td><td>-.05 (.02)</td><td><0.05</td></tr></table> a. What is the odds ratio for getting a mammogram for every 10-year increase in age?	Variable name	Parameter estimate (std error)	Significance value	Family history-associated risk group	.14 (.09)	not sig.	Age	-.04 (.02)	<0.05	Worry	-.04 (.01)	<0.05	Worryx family history-associated risk group	-.05 (.02)	<0.05	5	CO2	L1
Variable name	Parameter estimate (std error)	Significance value																	
Family history-associated risk group	.14 (.09)	not sig.																	
Age	-.04 (.02)	<0.05																	
Worry	-.04 (.01)	<0.05																	
Worryx family history-associated risk group	-.05 (.02)	<0.05																	
Q3.	You receive a dataset containing customer transaction data over the past year. The dataset is incomplete with numerous missing values and some duplicate entries. How would you go about cleaning this data to prepare it for analysis?	5	CO1	L4															
Q4.	A box contains 10 blue balls, 15 green balls and x yellow balls. If the probability of selecting a yellow ball is $\frac{2}{3}$, what is the total number of balls in the box? Given below the premises, specify whether the Conclusion "Taking appropriate medicine, ailment D lasts longer than ailment A" is True or False?	5	CO4	L5															

SECTION B

*Answer any 4 questions in Section B (4*10)*

S.NO	QUESTIONS	MARKS	CO	BL
Q1.	There are many types of vehicles and they are classified using different sets of criteria. Determine whether the stated conclusion is valid based on the given information. If not, write invalid. Explain your reasoning using a Venn diagram.	10	CO3	L2

Q2.	Imagine you're tasked with evaluating the effectiveness of two different marketing campaigns. What statistical test would you use to determine which campaign was more successful, and why?	10	CO3	L3
Q3.	Describe how you evaluate the performance of a data science model. Can you give an example of how you've applied these evaluation techniques in a past project?	10	CO1	L1
Q4.	What are some of the techniques used for sampling? What is the main advantage of sampling?	10	CO4	L4
Q5.	Through an example specify in detail conceptual framework of Data Analysis problems.	10	CO5	L5
Q6.	How would you analyze the effectiveness of a new LinkedIn chat feature that shows a "green dot" for active users?	10	CO2	L6

BLOOMS TAXANOMY LEVELS (BL)

1-REMEMBERING, 2-UNDERSTANDING, 3-APPLYING, 4-ANALYSING,
5-EVALUATING, 6-CREATING

8

8