End Term (Odd) Semester Examination November 2025

Roll no...............................

Name of the Course and semester: MBA III Semester
Name of the Paper: Data Science with Python
Paper Code: MB 303 (BA1)

Time: 3 hour   Maximum Marks: 100

Note:
  (i)   This question paper contains two Sections-Section A and B
  (ii)   Both Sections are compulsory
  (iii) Answer any two sub questions from a, b & c in each main question of Section A.
         Each sub question carries 10 marks.
  (iv)   Section B, consisting of a case study, is compulsory. It is of 20 Marks.

Section A

Q1.   (2X10=20 Marks)

a. "Analytics has transformed managerial decision-making across industries." Analyze the importance and scope of Data Analytics in modern business environments. Illustrate your answer using at least two industry-specific examples. (CO1)

b. You have been hired as a Data Scientist in an FMCG company. Explain the key roles and responsibilities you would perform across the data lifecycle, and analyze how each responsibility helps improve business outcomes. (CO1)

c. A banking firm wants to use predictive models to identify potential loan defaulters. Analyze how different types of analytics would be applied in this context and justify why Data Science is crucial for building such models. (CO1)

Q2.   (2X10=20 Marks)

a. "Successful Data Analytics requires a systematic and disciplined approach." Analyze the Data Analytics process and explain how each step contributes to solving a real business problem such as customer churn or declining sales. (CO2)

b. Python relies heavily on operators and basic functions for analytical computations. Explain how arithmetic, comparison, logical operators, and built-in functions (like len(), type(), range()) support data manipulation, and apply them to a real analytics scenario (e.g., sales trend calculation). (CO2)

c. A retail company wants to categorize, sort, and store customer information efficiently. Apply your understanding of Python data structures to recommend the most suitable structure(s) for different components of customer data, justifying your choices. (CO2)

Graphic Era
HILL UNIVERSITY
Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)
University under section 2(f) of UGC Act, 1956

End Term (Odd) Semester Examination November 2025

Q3. (2X10=20 Marks)

a. Analyze the properties of the four measurement scales (nominal, ordinal, interval, ratio). Explain how wrong identification of the scale can lead to incorrect statistical analysis and misleading conclusions. (CO3)

b. Compare probability and non-probability sampling techniques. Analyze how the choice of sampling method influences the accuracy and bias of research findings. Provide suitable applications for each. (CO3)

c. Analyze how Python packages such as NumPy, Pandas, SciPy, Matplotlib, and Seaborn support the statistical analysis process. Explain how using these tools improves accuracy, speed, and interpretation of results. (CO3)

Q4. (2X10=20 Marks)

a. Analyze the role of Exploratory Data Analysis (EDA) in identifying patterns, anomalies, and assumptions before modeling. Explain how poor EDA can lead to incorrect model selection or inaccurate insights. (CO4)

b. Analyze the assumptions of linear regression (linearity, homoscedasticity, independence, normality). Explain how violating any two assumptions affects model accuracy with examples. (CO4)

c. Compare graphical EDA techniques (boxplot, histogram, heatmap) with numerical EDA techniques (summary statistics, skewness, kurtosis). Analyze how combining both improves the quality of insights. (CO5)

Section B

Q5. Case Study (20 Marks)

A healthcare analytics company, MediInsight, is analyzing patient recovery data from multiple hospitals to identify treatment effectiveness. Analysts must use tabulation, graphical analysis, and visualization tools in Python to present insights clearly. They plan to use bar plots, pie charts, scatter plots, box-whisker plots, histograms, and line charts to compare treatment durations, patient demographics, and medicine effectiveness.

However, the analysts face challenges: incomplete data, varying formats, and limited expertise in advanced tools. They must apply skills in data cleaning, statistical reasoning, and interpretation. To ensure reliability, the team uses measures of central tendency, measures of position, and normality checks.

For treatment comparison, they apply hypothesis testing (t-test, chi-square). MediInsight also needs to interact with other platforms (Excel, Python and SQL) for seamless data flow.

Answer the following questions:

1. Analyze how MediInsight can use different visualization tools (bar plot, scatter plot, box-whisker plot, histogram, line chart, pie chart) to identify treatment effectiveness and patient trends. (CO5)

2. Discuss the skills and tools required for MediInsight's analysts to manage data cleaning, tabulation, graphical modeling, and interaction with external platforms like Excel, Python and SQL. (CO4)

3. Explain how measures of central tendency, measures of position, and normal distribution checks can help MediInsight interpret patient recovery data more accurately. (CO4)

4. Evaluate how hypothesis testing (such as t-tests and chi-square tests) can assist MediInsight in determining whether differences in treatment outcomes across hospitals are statistically significant. (CO5)