CMR INSTITUTE OF TECHNOLOGY

USN | | | | | | | | | | |

Internal Assessment Test II – Dec 2025

Sub:	Business Data Analytics						Sub Code:		MMCA311C
Date:	17.12.25	Duration:	90 min's	Max Marks:	50	Sem:	III	Branch:	MCA

Note : Answer FIVE FULL Questions, choosing ONE full question from each Module

	PART I	MARKS	OBE CO	OBE RBT
1	Provide real-world examples for descriptive, predictive, and prescriptive analytics, explaining their purposes. OR	[10]	CO1	L2
2	Explain how organizations use BI and BA differently for reporting, trend identification, and forecasting.	[10]	CO2	L3
3	PART II Explain four methods for handling missing values and three techniques for outliers with suitable examples. OR	[10]	CO2	L2
4	Discuss how histograms, boxplots, and heatmaps help uncover business insights in exploratory analysis.	[10]	CO2	L3
5	PART III Explain Logistic Regression and describe two business applications involving classification. OR	[10]	CO3	L2
6	Discuss model evaluation metrics such as RMSE, MAE, Accuracy, Precision, and Recall.	[10]	CO3	L2
7	PART IV Explain Sensitivity Analysis and Scenario Analysis, highlighting differences, purpose, and business relevance. OR	[10]	CO3	L3
8	Explain Decision Trees and Business Rules as decision-making tools with use cases like customer targeting and loan approval	[10]	CO3	L3
9	PART V Explain dashboards, their characteristics, and differences between operational, analytical, and strategic dashboards. OR	[10]	CO3	L2
10	Explain Data-Driven Storytelling and describe the structure of a compelling analytical story with an example.	[10]	CO3	L3

1. Provide real-world examples for descriptive, predictive, and prescriptive analytics, explaining their purposes.

1. Descriptive Analytics
Purpose:
Descriptive analytics focuses on summarizing historical data to understand *what has happened* in the past. It helps organizations monitor performance and identify trends and patterns.
Real-World Example:
A retail company uses sales reports and dashboards to analyze:

- Daily and monthly sales
- Best-selling products
- Revenue by region or store

Explanation:
By analyzing past sales data using charts, totals, and averages, managers can understand customer buying behavior and seasonal trends. This information helps in evaluating business performance but does not predict future outcomes.

2. Predictive Analytics
Purpose:
Predictive analytics uses historical data, statistical models, and machine learning to forecast *what is likely to happen* in the future.
Real-World Example:
An e-commerce platform predicts customer demand and product sales during festive seasons.

Explanation:
Using past sales data, customer behavior, and seasonal trends, predictive models forecast future demand. This helps the company plan inventory, avoid stockouts, and reduce excess inventory.

3. Prescriptive Analytics
Purpose:
Prescriptive analytics recommends optimal actions by analyzing data and considering constraints to answer *what should be done*.
Real-World Example:
A ride-sharing company uses prescriptive analytics to decide dynamic pricing and driver allocation.
Explanation:
Based on demand forecasts, traffic conditions, and driver availability, the system suggests optimal ride prices and driver locations. This maximizes revenue while ensuring customer satisfaction and efficient resource utilization.

2. Explain how organizations use BI and BA differently for reporting, trend identification, and forecasting.

1. Use of BI and BA for Reporting
Business Intelligence (BI):
- BI is primarily used for standardized reporting and performance monitoring.
- It answers the question: "What has happened?"
- Reports are generated from historical and current data stored in data warehouses.

Example:
A retail organization uses BI dashboards to generate:
- Daily sales reports
- Inventory status reports
- Profit and loss statements

Purpose:
To provide managers with a clear, accurate view of current operations and past performance for routine decision-making.

Business Analytics (BA):
- BA goes beyond reporting by analyzing data relationships.
- It answers: "Why did it happen?"
- Uses statistical analysis and data mining techniques.

Example:
Analyzing why sales dropped in a particular region by examining customer behavior, pricing, and promotions.
Purpose:
To support deeper insights rather than routine status reporting.

2. Use of BI and BA for Trend Identification
Business Intelligence (BI):
- BI identifies basic trends and patterns through historical data visualization.
- Trends are usually descriptive in nature.

Example:
A telecom company uses BI charts to observe:
- Monthly increase in customer churn
- Seasonal usage patterns

Purpose:
To recognize patterns and highlight areas that require attention.

Business Analytics (BA):
- BA identifies hidden and complex trends using advanced techniques.
- It explains *relationships and root causes* behind trends.

Example:
Using clustering analysis to identify customer segments most likely to churn.
Purpose:
To understand *why trends occur* and how different factors influence them.

3. Use of BI and BA for Forecasting
Business Intelligence (BI):
- BI supports basic forecasting using historical comparisons and summaries.
- Forecasts are often simple and rule-based.

Example:
Forecasting next quarter sales based on the same quarter's performance in previous years.
Purpose:
To support short-term planning with limited analytical depth.

Business Analytics (BA):
- BA is extensively used for advanced forecasting.

- Uses predictive models, machine learning, and statistical techniques.

Example:
Predicting future demand using regression models and time-series analysis.
Purpose:
To enable accurate forecasting and proactive decision-making.

3. Explain four methods for handling missing values and three techniques for outliers with suitable examples.

Methods for Handling Missing Values (Any Four)
1. Deletion Method
Explanation:
Missing records are removed from the dataset. This can be done using:

- Listwise deletion (remove entire row)
- Pairwise deletion (remove only for specific analysis)

Example:
In a customer dataset, if only 2 out of 10,000 records have missing age values, those records can be deleted without affecting results.
Use Case:
Suitable when missing values are very few and occur randomly.

2. Mean / Median / Mode Imputation
Explanation:
Missing values are replaced using:

- Mean for normally distributed numerical data
- Median for skewed data
- Mode for categorical data

Example:
If the average salary in a dataset is ₹40,000, missing salary values are replaced with ₹40,000.
Use Case:
Simple and fast method, useful for large datasets.

3. Forward Fill / Backward Fill
Explanation:
Missing values are filled using the previous or next available value.
Example:
In a daily temperature dataset, if Tuesday's value is missing, it can be filled using Monday's temperature (forward fill).
Use Case:
Commonly used in time-series data.

4. Predictive Imputation
Explanation:
Missing values are predicted using statistical or machine learning models like regression or k-NN.
Example:
If a student's marks in one subject are missing, a regression model using marks from other subjects can estimate the missing value.
Use Case:
Provides more accurate results when relationships exist among variables.

Techniques for Handling Outliers (Any Three)
1. Z-Score Method
Explanation:
Outliers are detected using standard deviation. Values with Z-score greater than ±3 are considered outliers.
Example:
In a salary dataset, most salaries are between ₹20,000–₹80,000, but one salary is ₹5,00,000, which becomes an outlier.
Use Case:
Effective when data is normally distributed.

2. Interquartile Range (IQR) Method
Explanation:
Outliers are values below
Q1−1.5×IQRQ1 - 1.5 \times IQRQ1−1.5×IQR or above
Q3+1.5×IQRQ3 + 1.5 \times IQRQ3+1.5×IQR
Example:
In house price data, extremely high-priced luxury villas are identified as outliers.
Use Case:
Works well for skewed distributions.

3. Capping (Winsorization)

Explanation:
Extreme values are capped at a fixed percentile (e.g., 5th and 95th percentiles).
Example:
Customer spending above the 95th percentile is replaced with the 95th percentile value.
Use Case:
Prevents loss of data while reducing the effect of extreme values.

4. Discuss how histograms, boxplots, and heatmaps help uncover business insights in exploratory analysis.

Exploratory Data Analysis (EDA) uses visual tools such as histograms, boxplots, and heatmaps to understand data characteristics, identify patterns, and uncover meaningful business insights before applying advanced analytics. Each visualization serves a distinct purpose in analyzing business data.

1. Histograms
Role in Exploratory Analysis:
Histograms display the frequency distribution of numerical data by grouping values into bins.
Business Insights Uncovered:
- Identify data distribution (normal, skewed, or bimodal)
- Detect concentration of customers or transactions
- Reveal seasonal or demand patterns

Business Example:
A bank uses histograms to analyze loan amounts issued to customers. The histogram may show that most loans fall in the ₹1–5 lakh range, indicating the primary customer segment and helping design targeted loan products.
Decision Impact:
Supports pricing strategies, segmentation, and risk assessment.

2. Boxplots
Role in Exploratory Analysis:
Boxplots summarize data using median, quartiles, range, and outliers, offering a compact view of data variability.
Business Insights Uncovered:
- Identify outliers that may indicate anomalies or special cases
- Compare performance across groups
- Understand data spread and consistency

Business Example:
An HR department uses boxplots to compare employee salaries across departments. Outliers may indicate unusually high executive salaries or data errors.
Decision Impact:
Helps in pay equity analysis, anomaly detection, and policy review.

3. Heatmaps
Role in Exploratory Analysis:
Heatmaps use color intensity to represent values in a matrix, making complex relationships easy to interpret.
Business Insights Uncovered:
- Identify correlations between variables
- Spot high and low activity zones
- Reveal patterns and relationships at a glance

Business Example:
An e-commerce company uses a heatmap to analyze product sales across regions. Darker colors in certain regions indicate high demand, guiding inventory allocation.
Decision Impact:
Supports regional strategy, cross-selling, and operational optimization.

5. Explain Logistic Regression and describe two business applications involving classification.

Logistic Regression is a widely used classification technique in statistics and machine learning, especially for problems where the outcome is categorical, most commonly binary (yes/no, 0/1).

Logistic Regression
Explanation
- Logistic Regression models the probability that a given input belongs to a particular class.
- Unlike linear regression, it uses the logistic (sigmoid) function to ensure predicted values lie between 0 and 1.

Sigmoid Function:

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}}$$

- The output probability is converted into a class label using a threshold (commonly 0.5).
- If probability $\geq 0.5 \rightarrow$ Class 1
- If probability $< 0.5 \rightarrow$ Class 0

Key Characteristics

- Suitable for binary classification
- Interpretable coefficients (odds ratios)
- Efficient and easy to implement
- Can be extended to multinomial logistic regression

Business Applications of Logistic Regression
1. Credit Risk Assessment (Loan Approval)
Application:
Banks use logistic regression to classify whether a loan applicant is likely to default or not.
Inputs (Features):

- Income level
- Credit score
- Employment status
- Existing liabilities

Output (Class):

- Default (1) / Non-default (0)

Business Value:
Helps financial institutions reduce credit risk, improve loan approval decisions, and comply with regulatory requirements.

2. Customer Churn Prediction
Application:
Telecom and subscription-based companies use logistic regression to classify customers as likely to churn or stay.
Inputs (Features):

- Usage patterns
- Customer complaints
- Subscription duration
- Payment history

Output (Class):

- Churn (1) / Retain (0)

Business Value:
Enables companies to take proactive retention actions, such as personalized offers or discounts, improving customer lifetime value.

6. Discuss model evaluation metrics such as RMSE, MAE, Accuracy, Precision, and Recall.

Model evaluation metrics are used to assess the performance of predictive and classification models. Different metrics are used depending on whether the problem is regression or classification.

1. RMSE (Root Mean Square Error)
Type: Regression metric
Definition:
RMSE measures the square root of the average of squared differences between actual and predicted values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Interpretation:

- Lower RMSE indicates better model performance
- Penalizes large errors more heavily

Example:
Used in sales or demand forecasting to evaluate how close predictions are to actual sales.
Use Case:
Preferred when large prediction errors are particularly costly.

2. MAE (Mean Absolute Error)
Type: Regression metric
Definition:
MAE calculates the average of absolute differences between actual and predicted values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Interpretation:

- Easy to understand
- Treats all errors equally

Example:
Used in price prediction models to measure average prediction error.
Use Case:
Useful when robustness to outliers is required.

3. Accuracy
Type: Classification metric
Definition:
Accuracy is the proportion of correctly classified instances out of total instances.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Interpretation:
- Simple and intuitive
- Can be misleading with imbalanced datasets

Example:
Used in spam email classification.
Use Case:
Effective when class distribution is balanced.

4. Precision
Type: Classification metric
Definition:
Precision measures the proportion of true positive predictions among all positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

Interpretation:
- High precision means fewer false positives

Example:
Used in fraud detection, where falsely flagging a transaction is costly.
Use Case:
Important when false positives must be minimized.

5. Recall (Sensitivity / True Positive Rate)
Type: Classification metric
Definition:
Recall measures the proportion of actual positives that are correctly identified.

$$Recall = \frac{TP}{TP + FN}$$

Interpretation:
- High recall means fewer false negatives

Example:
Used in medical diagnosis, where missing a disease is dangerous.
Use Case:
Important when false negatives are critical.

7. Explain Sensitivity Analysis and Scenario Analysis, highlighting differences, purpose, and business relevance.

   Sensitivity Analysis and Scenario Analysis are important analytical techniques used in decision-making and optimization models to evaluate uncertainty and risk. They help managers understand how changes in inputs affect outcomes and support better business planning.

   1. Sensitivity Analysis
   Definition
   Sensitivity Analysis examines how changes in one input variable at a time affect the output of a model, while keeping all other variables constant.
   Purpose

- To identify critical variables that have the greatest impact on results
- To test the robustness of a model
- To understand risk and uncertainty

   Business Example

A manufacturing company analyzes how changes in raw material cost affect total profit while keeping sales volume and labor costs constant.

Business Relevance

- Helps in pricing decisions
- Supports cost control
- Identifies variables that require close monitoring

Key Characteristics

- One-variable-at-a-time analysis
- Focuses on impact magnitude
- Often used in optimization and financial models

2. Scenario Analysis

Definition

Scenario Analysis evaluates the impact of simultaneous changes in multiple variables by creating different future scenarios such as best-case, worst-case, and most-likely-case.

Purpose

- To evaluate possible future outcomes
- To support strategic planning
- To assess overall business risk

Business Example

A retail company evaluates profitability under:

- Best-case: High demand, low costs
- Worst-case: Low demand, high costs
- Most-likely: Average demand and costs

Business Relevance

- Supports long-term strategic decisions
- Helps in risk management
- Useful for investment and expansion planning

Key Characteristics

- Multiple variables change together
- Focus on possible futures
- Strategic in nature

8. Explain Decision Trees and Business Rules as decision-making tools with use cases like customer targeting and loan approval

Decision Trees and Business Rules are widely used decision-making tools in analytics and information systems. They help organizations make consistent, transparent, and data-driven decisions, especially in classification and rule-based scenarios such as customer targeting and loan approval.

1. Decision Trees

Explanation

A Decision Tree is a graphical and analytical model that represents decisions in a tree-like structure consisting of:

- Root node – starting decision point
- Decision nodes – conditions or tests on attributes
- Branches – outcomes of tests
- Leaf nodes – final decisions or class labels

Decision trees work by recursively splitting data based on attributes that best separate the target classes (e.g., using information gain or Gini index).

Use Case 1: Customer Targeting

Scenario:

A retail company wants to identify customers likely to respond to a marketing campaign.

Decision Tree Example Rules:

- If *Age < 35* and *Online Purchase Frequency = High* → Target customer
- If *Age ≥ 35* and *Income > ₹8 lakh* → Target customer
- Else → Do not target

Business Value:

- Improves campaign response rate

- Reduces marketing costs
- Enables personalized marketing

Use Case 2: Loan Approval
Scenario:
A bank uses a decision tree to classify loan applications.
Decision Criteria:
- Credit score
- Annual income
- Employment stability

Decision Tree Outcome:
- If *Credit Score > 750* and *Income > ₹5 lakh* → Approve loan
- If *Credit Score < 600* → Reject loan

Business Value:
- Faster and consistent decisions
- Reduced human bias
- Regulatory transparency

2. Business Rules
Explanation
Business Rules are explicit IF–THEN statements that define organizational policies, constraints, and operational decisions. They are usually derived from domain expertise, regulations, or company strategy.
Example Format:
IF condition THEN action

Use Case 1: Customer Targeting
Example Rules:
- IF customer made a purchase in the last 30 days AND cart value > ₹3,000
  THEN offer a 10% discount
- IF customer is inactive for 6 months
  THEN send re-engagement email

Business Value:
- Simple and transparent logic
- Easy to modify without retraining models
- Ensures policy consistency

Use Case 2: Loan Approval
Example Rules:
- IF applicant age < 21 OR income < ₹2 lakh
  THEN reject loan
- IF applicant has default history
  THEN reject loan

Business Value:
- Ensures regulatory compliance
- Reduces risk
- Provides explainable decisions

9. Explain dashboards, their characteristics, and differences between operational, analytical, and strategic dashboards

Dashboards are essential tools in Business Intelligence (BI) that visually present key data to support monitoring, analysis, and decision-making at different levels of an organization. They provide quick insights by integrating metrics, charts, and indicators in a single view.

1. Dashboards – Meaning
A dashboard is a visual interface that displays key performance indicators (KPIs), metrics, and trends using charts, graphs, and tables. Dashboards help decision-makers quickly understand performance and take timely actions.

2. Characteristics of Dashboards
1. Visual and Interactive
   Use charts, gauges, heat maps, and filters for easy interpretation.
2. Real-time or Near Real-time Data
   Especially important for operational dashboards.
3. Role-based Design
   Tailored for executives, managers, or operational staff.
4. Single-Screen View
   Presents critical information without clutter.

5. Drill-down Capability
   Allows users to explore detailed data.
6. Performance-focused
   Displays KPIs aligned with business objectives.

3. Types of Dashboards
A. Operational Dashboards
Purpose:
Monitor day-to-day operations and immediate performance.
Users:
Front-line staff, supervisors, operations managers
Key Features:
- Real-time or near real-time data
- Alerts and thresholds
- Highly detailed

Example:
A call center dashboard showing:
- Number of calls waiting
- Average call handling time
- Agent availability

Decision Support:
Helps in quick corrective actions.

B. Analytical Dashboards
Purpose:
Support data analysis and trend identification.
Users:
Business analysts, middle-level managers
Key Features:
- Historical data
- Advanced visualizations
- Drill-down and slice-and-dice analysis

Example:
A sales analytics dashboard showing:
- Monthly sales trends
- Customer segmentation
- Product-wise performance

Decision Support:
Helps in tactical decisions and performance improvement.

C. Strategic Dashboards
Purpose:
Support long-term strategic planning.
Users:
Top executives, senior management
Key Features:
- High-level KPIs
- Aggregated data
- Periodic updates (weekly/monthly)

Example:
A CEO dashboard showing:
- Revenue growth
- Market share
- Customer satisfaction index

Decision Support:
Guides strategic decisions and goal alignment.

4. Differences Between Dashboard Types

Aspect	Operational	Analytical	Strategic
Focus	Daily operations	Analysis & trends	Long-term strategy
Users	Staff, supervisors	Analysts, managers	Executives
Data Type	Real-time	Historical	Aggregated
Detail Level	Very detailed	Moderate to detailed	Highly summarized
Update Frequency	Real-time	Daily/weekly	Monthly/quarterly

10. Explain Data-Driven Storytelling and describe the structure of a compelling analytical story with an example.

Data-Driven Storytelling is the practice of communicating insights derived from data in a clear, engaging, and persuasive narrative so that decision-makers can understand the insights and act on them. It combines data, analytics, visuals, and narrative to turn numbers into meaningful business stories.

1. Data-Driven Storytelling – Explanation
Data-driven storytelling goes beyond presenting charts or dashboards. It explains:

- What the data shows
- Why it matters
- What action should be taken

It bridges the gap between analytical results and business decisions, ensuring insights are understandable to both technical and non-technical stakeholders.

2. Structure of a Compelling Analytical Story
A strong analytical story typically follows a clear structure, similar to a narrative.

1. Context / Business Problem
Purpose:
Sets the background and defines the problem to be solved.
Example:
"A retail company observed a decline in quarterly sales despite increased marketing spend."

2. Data and Assumptions
Purpose:
Explains data sources, scope, and assumptions to build credibility.
Example:
"Sales data from 150 stores, customer demographics, and promotion records over the last 2 years were analyzed."

3. Analysis and Key Insights
Purpose:
Presents insights supported by data and visualizations.
Example:
"Analysis showed that sales dropped mainly in urban stores, and discounts had minimal impact on repeat customers."

4. Visualization and Narrative
Purpose:
Uses charts and storytelling to make insights easy to understand.
Example:
"A line chart highlighted declining footfall, while a heatmap showed poor promotion performance in metro regions."

5. Recommendations / Actions
Purpose:
Converts insights into actionable decisions.
Example:
"Shift marketing focus to loyalty programs and improve in-store experience in urban locations."

6. Impact and Outcome (Optional)
Purpose:
Shows expected or achieved benefits.
Example:
"Implementing these changes increased customer retention by 12% within six months."

3. Example of Data-Driven Storytelling (End-to-End)
Scenario: Customer Churn in a Telecom Company

- Context: Rising customer churn rate
- Data: Call records, billing data, complaint history
- Insights: Customers with frequent service issues churned 3× more
- Visualization: Bar charts and churn heatmaps
- Action: Improve network quality and proactive customer support
- Outcome: 15% reduction in churn