

**BUSINESS ANALYTICS  
(DSC3132)**

**Time Allotted : 2½ hrs**

**Full Marks : 60**

*Figures out of the right margin indicate full marks.*

*Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.*

*Candidates are required to give answer in their own words as far as practicable.*

**Group – A**

1. Answer any twelve:

**12 × 1 = 12**

*Choose the correct alternative for the following*

- (i) What challenge is associated with moving beyond descriptive analytics?
  - (a) The inability to summarise trends effectively
  - (b) The absence of user engagement metrics
  - (c) The lack of historical data analysis
  - (d) The need for robust data collection
- (ii) What is the ideal outcome for a theory regarding Type I and Type II errors?
  - (a) Balanced Type I and Type II errors
  - (b) Minimal Type I errors and frequent Type II errors
  - (c) No Type I or Type II errors
  - (d) Frequent Type I errors and minimal Type II errors
- (iii) Which example best illustrates implicit data collection?
  - (a) Asking users to fill out demographic forms
  - (b) Tracking user navigation paths and page visits
  - (c) Collecting product review submissions
  - (d) Conducting customer feedback surveys
- (iv) What is the primary purpose of combining explicit and implicit data techniques?
  - (a) To focus solely on engagement levels of users
  - (b) To simplify the process of data collection
  - (c) To provide comprehensive insights into user behaviour
  - (d) To identify only the reasons behind feature adoption drops
- (v) Which of the following is not a key concept essential to every web product?
  - (a) Acquisition
  - (b) Attrition
  - (c) Engagement
  - (d) Revenue
- (vi) What type of test is used to evaluate click-through rates (CTR)?
  - (a) Regression analysis for click rates
  - (b) T-test for independent samples
  - (c) ANOVA test for categorical variables
  - (d) Pearson chi-squared test

- (vii) Why is randomization important in research?
  - (a) It removes the need for treatment and control groups
  - (b) It eliminates the need for statistical significance testing
  - (c) It helps create comparable populations without knowing all confounding factors
  - (d) It ensures equal sample sizes in all research studies
- (viii) How does the Average Treatment Effect on the Treated (ATT) mitigate selection bias?
  - (a) By comparing treated users with similar control users
  - (b) By focusing only on untreated individuals in the study
  - (c) By comparing treated users with entirely randomised groups
  - (d) By eliminating the need for control groups altogether
- (ix) What problem can arise when multiple treated observations match to a single control?
  - (a) It can lead to imperfect matches, limiting further statistical analysis
  - (b) It eliminates the need for verifying the matching process in the analysis
  - (c) It ensures perfect matches but reduces the sample size significantly
  - (d) It balances all variables but complicates causality verification
- (x) What is the most practical approach currently for estimating causal effects in uplift analysis?
  - (a) Predicting subgroup causal effects
  - (b) Estimating average treatment effects
  - (c) Estimating subpopulation treatment effects
  - (d) Predicting individual treatment effects

*Fill in the blanks with the correct word*

- (xi) Churn rate measures \_\_\_\_\_.
- (xii) \_\_\_\_\_ data collection captures user preferences indirectly.
- (xiii) Continuous variables converted into categories during data preparation through \_\_\_\_\_.
- (xiv) For comparison, the regression discontinuity (RD) design depends on \_\_\_\_\_.
- (xv) \_\_\_\_\_ regression is commonly used in uplift modelling for straightforward interpretation.

### **Group - B**

- 2. (a) Differentiate between quantitative and qualitative analytics. [[CO1](Apply/10CQ)]
- (b) Mention two use cases each for quantitative and qualitative analytics. [[CO1](Apply/10CQ)]
- (c) How can quantitative and qualitative analytics be combined? [[CO2](Apply/10CQ)]
- 6 + 4 + 2 = 12**
- 3. (a) Using an example scenario, differentiate between Type I and Type II errors. [[CO5](Apply/10CQ)]
- (b) “Type I error is associated with false positives, while Type II error is associated with false negatives.” Justify or refute this statement with reasoning. [[CO5](Apply/10CQ)]

- (c) What is confirmation bias and which desirable quality of a theory does it negatively impact?

[[CO5](Evaluate/HOCQ)]

**4 + 4 + 4 = 12**

### Group - C

4. (a) What are implicit data collection techniques? Give at least three examples.  
 (b) What are the strengths of implicit data collection techniques?  
 (c) What are the limitations of implicit data collection techniques?
5. (a) Why is data preparation critical in analytics?  
 (b) Discuss the steps in the data preparation process.  
 (c) What are the common issues in raw datasets?

[[CO5](Remember/LOCQ)]

[[CO5](Remember/LOCQ)]

[[CO5](Remember/LOCQ)]

**4 + 4 + 4 = 12**

[[CO2](Understand/LOCQ)]

[[CO2](Remember/LOCQ)]

[[CO2](Understand/LOCQ)]

**4 + 4 + 4 = 12**

### Group - D

6. (a) What is an A/B test and how is it different from randomized controlled trial (RCT)?  
 (b) In the context of an A/B test, define the following: treatment, treated group, control group.  
 (c) Why is A/B testing important?
7. (a) Using difference-in-difference (DID) modelling, suppose you want to explore the effects of a promotional TV campaign in one designated market area (DMA) versus another, but you are unable to randomize the viewers. From the table below, find out the average treatment effect on the treated (ATT); show the steps in your calculation, and mention what you conclude from the answer.

[[CO5](Understand/LOCQ)]

[[CO5](Remember/LOCQ)]

[[CO5](Understand/LOCQ)]

**4 + 4 + 4 = 12**

Treatment	Before Commercial	After Commercial
DMA 1 (treated)	2,100 downloads	4,230 downloads
DMA 2 (control)	2,300 downloads	3,548 downloads

[[CO6](Evaluate/HOCQ)]

- (b) In the context of part (a) of this question, suppose you are using an ordinary least squares (OLS) regression based approach to find out the causal effect, using the following equation:

$$y = \text{inter} + b1 \cdot \text{time} + b2 \cdot \text{treatment} + b3 \cdot (\text{time} \cdot \text{treatment}) + \text{error}$$

Interpret the following results in terms of the quanta of causal effect and its statistical significance. Explain your answer.

[[CO5](Evaluate/HOCQ)]

Coefficients	Estimate	Std. Error	t Value	Pr(>  t )
(Intercept)	2,298.95	2.42	949.9	<2e-16 ***
treated	-199.88	3.423	-58.4	<2e-16 ***
time	1,248.413	3.423	364.7	<2e-16 ***
treated:time	882.568	4.841	182.3	<2e-16 ***

- (c) What is a placebo test? [[C05](Remember/LOCQ)]  
**6 + 4 + 2 = 12**

### Group - E

8. (a) In the context of uplift modelling, what is a lift? Explain with an example. [[C06](Apply/IOCQ)]
- (b) What is uplift modelling and how is it different from other types of causal inference from observational data? [[C06](Understand/IOCQ)]
- (c) “The goal of uplift modelling is to place users within four groups.” Identify and explain the four groups. [[C06](Understand/LOCQ)]  
**4 + 4 + 4 = 12**
9. (a) Give two real-world scenarios where the individual treatment effect can vary significantly from the individual treatment effect. From the causal inference viewpoint, is there any difficulty in such situations? If yes, why? If no, why not? [[C06](Analyse/IOCQ)]
- (b) Discuss what prevents subpopulation treatment effect (STE) from enabling true causal inference. [[C06](Understand/LOCQ)]
- (c) With an example, distinguish between main effect and heterogenous treatment effect. [[C06](Understand/LOCQ)]  
**4 + 4 + 4 = 12**

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	52.1	33.3	14.6