

QPCode

Reg.No :

Name :

MAHATMA GANDHI UNIVERSITY, KOTTAYAM
MGU-UGP (HONOURS) DEGREE EXAMINATION October 2026

Semester V

DSE - MG5DSEMATA01 - ADVANCED DATA ANALYSIS USING PYTHON
(2024 ADMISSION ONWARDS)

Duration: 2 Hours

Maximum Marks: 70

Remember(K), Understand(U), Apply(A), Analyse(An), Evaluate(E), Create(C), Skill(S), Interest(I) and Appreciation(Ap)

Students should attempt at least one question from each course outcome to enhance their overall outcome attainability.

Part A

MCQ

Answer any **10** questions
Each question carries **3** marks

1. The non-parametric alternative for one-sample test is
 - a. Kolmogorov-Smirnov
 - b. Mann-Whitney
 - c. Shapiro-Wilk
 - d. Kruskal-Wallis[K] / [CO1]
2. A grayscale image contains noise, and you want to smooth the image while reducing high-frequency noise using SciPy. Which ndimage function should be applied?
 - a. ndimage.gaussian_filter()
 - b. ndimage.label()
 - c. ndimage.find_objects()
 - d. ndimage.center_of_mass()[A] / [CO1]
3. Which scipy.special function calculates the number of combinations of N objects taken k at a time?
 - a. perm()
 - b. comb()
 - c. gamma()
 - d. cbart()[K] / [CO1]
4. Which Python syntax is used to perform a paired t-test?
 - a. stats.ttest_ind(before, after)
 - b. stats.f_oneway(before, after)
 - c. stats.ttest_rel(before, after)
 - d. stats.anova(before, after)[K] / [CO1]
5. Which SQL statement retrieves all columns from a table named cars?
 - a. SELECT cars FROM *
 - b. SELECT ALL FROM cars
 - c. SELECT * FROM cars
 - d. GET * FROM cars[K] / [CO2]
6. Which method returns one row from the query result?
 - a. fetchall()
 - b. fetchone()
 - c. fetchrow()
 - d. getone()[K] / [CO2]
7. Which SQL statement retrieves only the name and price columns from the cars table?
 - a. SELECT * FROM cars
 - b. SELECT name, price FROM cars
 - c. SELECT cars.name, price *
 - d. GET name, price FROM cars[A] / [CO2]
8. A monthly sales time series shows a clear upward trend, meaning its mean changes over time. What should be investigated before fitting an ARIMA model?[A] / [CO3]

- | | | |
|--|---|--|
| a. Whether the series is stationary | b. Whether the data is categorical | |
| c. Whether the column names are alphabetical | d. Whether all observations are identical | |
9. In Python, the ARIMA model is available in which package?
- | | | |
|----------------|---------------|-------------|
| a. statsmodels | b. pandas | |
| c. numpy | d. matplotlib | [K] / [CO3] |
10. Which method can be used to check whether a time series is stationary?
- | | | |
|---------------------------------------|--------------------|-------------|
| a. K-Means clustering | b. Chi-square test | |
| c. Augmented Dickey-Fuller (ADF) test | d. Rolling() | [K] / [CO3] |
11. Exponential smoothing is computed using which method?
- | | | |
|--------------------|-----------------|-------------|
| a. Rolling method | b. EWMA method | |
| c. adfuller method | d. ARIMA method | [K] / [CO3] |
12. What is the primary purpose of dimensionality reduction in machine learning?
- | | | |
|------------------------------------|--|-------------|
| a. Increase the number of features | b. Reduce the number of features while retaining important information | |
| c. Remove all observations | d. Convert numerical data into categorical data | [U] / [CO4] |
13. Which algorithm divides observations into a predefined number of clusters based on similarity?
- | | | |
|-----------------------|----------------------|-------------|
| a. PCA | b. Factor Analysis | |
| c. K-Means Clustering | d. Linear Regression | [K] / [CO4] |
14. Which clustering technique creates a hierarchical tree-like structure (dendrogram) of clusters?
- | | | |
|--|--------------------|-------------|
| a. K-Means | b. PCA | |
| c. Agglomerative Hierarchical Clustering | d. Factor Analysis | [K] / [CO4] |
15. Which technique is primarily used to identify underlying latent factors that explain relationships among observed variables?
- | | | |
|-----------------------|--------------------|-------------|
| a. K-Means Clustering | b. Factor Analysis | |
| c. Linear Regression | d. Decision Tree | [K] / [CO4] |
- [3x10 = 30]**

Part B

One or Two Sentences

Answer any **4** questions

Each question carries **5** marks

- | | |
|--|--------------|
| 16. Discuss the different ways to determine the normality of a sample. | [U] / [CO1] |
| 17. Explain the utility of GROUP BY and ORDER BY clause with an example. | [U] / [CO2] |
| 18. A time series has the values 20, 24, 22, 26, and 30. Apply a 3-period moving average and calculate the available smoothed values. | [A] / [CO3] |
| 19. Explain the purpose of using seasonal decomposition on time series data. | [U] / [CO3] |
| 20. A dendrogram is obtained from hierarchical clustering. Explain how you would analyze the dendrogram to decide the number of clusters | [An] / [CO4] |
| 21. Write a Python program to load an image and perform image rotation: | [A] / [CO1] |
| a. Rotate the image by 45°. | |

- b. Rotate the image by 90°.
- c. Display the original and rotated images.
- d. Explain the effect of the reshape parameter.
- e. Read a number and find cube root and find gamma value of the cube root

[5x4 = 20]

Part C

Short Essay Questions

Answer any **2** questions

Each question carries **10** marks

- 22. Differentiate between parametric and non-parametric statistical tests with respect to their assumptions. For each type of test, select one appropriate statistical test, provide a suitable example, write the corresponding Python code to perform the test. [A] / [CO1]
- 23. Consider the mtcars dataset stored in a database table named cars1. Using SQLAlchemy, write Python code to connect to the database and perform the following SQL operations: [A] / [CO2]
 - a. Retrieve all columns from the cars1 table and display the first two records.
 - b. Retrieve the mpg, cyl, and hp columns for cars having mpg > 20.
 - c. Retrieve cars having either cyl = 4 or cyl = 6 using the IN operator.
 - d. Retrieve cars in descending order of mpg and display only the first five records.
 - e. Explain the purpose of engine.connect(), text(), execute(), and fetchall() in the given SQLAlchemy code.
- 24. Explain the complete process of time series forecasting, from checking stationarity to applying ARIMA. [U] / [CO3]
- 25. Explain Factor Analysis (FA) as an unsupervised machine learning technique. Discuss the types of FA, factorability tests (Bartlett's test and KMO test), methods for determining the number of factors, factor loadings [U] / [CO4]

[10x2 = 20]