Data Science Term Question Paper

Course: Data Science (Master's Level)

Total Marks: 100

Section A: Descriptive Statistics (20 marks)

1. Define descriptive statistics and explain their significance in data analysis. (5 marks)
2. Given the following dataset, calculate the mean, median, mode, variance, and standard deviation:
3. (5 marks)
4. Create a histogram and a box plot for the dataset. Interpret the results. (10 marks)

Section B: Inferential Statistics (25 marks)

1. Explain the concept of inferential statistics. (5 marks)
2. Hypothesis Testing:
    - State the null and alternative hypotheses for a scenario of your choice.
    - Perform a t-test or ANOVA (choose one) on the dataset.
    - Calculate p-values and interpret the results. (15 marks)
3. Confidence Intervals:
    - Calculate a 95% confidence interval for the meaning of a specific feature in the dataset. (5 marks)

Section C: Regression Analysis (25 marks)

1. Define linear regression and its assumptions. (5 marks)
2. Regression Problem:
    - Consider predicting house prices based on features like square footage, number of bedrooms, etc.
    - Build a linear regression model using the dataset.
    - Evaluate the model (R-squared, residuals) and interpret the coefficients. (20 marks)

Section D: Graphs and Visualization (15 marks)

1. Discuss the importance of data visualization.
2. Using Matplotlib or Seaborn, create:
    - A scatter plot showing the relationship between two relevant features.
    - A bar chart displaying a summary statistic (e.g., mean) for a categorical variable. (10 marks)
3. Explain the insights gained from the visualizations. (5 marks)

Section E: Error Calculation and Prediction Model (15 marks)

1. Error Metrics:

    - Define Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
    - Calculate these errors for your regression model. (10 marks)
2. Prediction Model:
    - Choose an appropriate prediction model (e.g., ARIMA, ARMA, SARIMA).
    - Evaluate its performance (box plots) using cross-validation. (5 marks)

Section F: Application-Level Questions (10 marks)

1. Relate statistical concepts to real-world scenarios:
2. Discuss the challenges and ethical considerations in applying statistical models to practical problems. (5 marks)

Use any suitable time series in indian context from the following

1. Rainfall data of India
2. Temperature data of india
3. Sentiment analysis on imdb review
4. Wind speed analysis of India
5. Stock market analysis
6. Sales analysis of a company
7. Production of a product in a company
8. Growth of bank loans
9. Agriculture product procurement cost and selling cost

Or any other relevant dataset in Indian context and submit report by 19th august in as a pdf file. No plagiarism is entertained put your efforts and continue the good efforts.