Time: 3hours   **Marks: 80**

NB.
1) Question number 1 is compulsory
2) Attempt any three out of the remaining five questions.
3) Assume suitable data if necessary and justify the assumptions.
4) Figures to the right indicate full marks

Q1   Answer the following   **20**
a)   Explain the significance of data science considering Volume and Dimensions of Data.
b)   Explain performance evaluation with respect to Time series forecasting
c)   Write a note on measure of spread
d)   Write a note on Applications of Data Science

Q2   a)   Explain Cross Validation, K-fold cross validation in detail   **10**
b)   Explain the Data Science Process in detail   **10**

Q3   a)   What is Data Visualization? What is Multivariate data? Explain with an example   **10**
Scatter plot.
b)   What are Outliers and their Causes? Can Statistics be used to detect Outliers if   **10**
yes, Explain.

Q4   a)   Explain SMOTE in detail   **10**
b)   Calculate the Bowley's coefficient of skewness from the following distribution.   **10**

Income	NO. of persons
30 - 40	8
40 - 50	24
50 - 60	48
60 - 70	68
70 - 80	30
80 -90	13
90 - 100	9

Q5   a)   Explain Smoothing methods in Time Series Forecasting   **10**
b)   A stenographer claims that she can type at the rate of 120 words per minute. Can   **10**
we reject her claim on the basis of 100 trails in which she demonstrates a mean of
116 words with a standard deviation of 15 words? Use 5 % level of significance,
$Z_\alpha = 1.96$.

Q6   a)   What are Recommendation engines? Explain.   **10**
b)   Time series analysis using linear regression   **10**

****