

SESSION PLAN

Session Name

EDA & Data Preprocessing

Learning Outcomes

- Detect and treat missing values, skewness, and outliers in data
- Normalize and aggregate data
- Encode categorical variables for applying models
- Explore univariate and multivariate data to get insights

Prerequisites for the Student

- Go through the concept and solve the tasks and assessments.

Student Activities

- Discussion with Mentor what they have learned.
- Overview of EDA and Data Preprocessing
 - Data Cleaning - Outlier Detection, Handling of Outliers, Detecting missing data, handling missing data
 - Data Transformation - Transforming skewed data, standardization and normalization, encoding categorical variables.
- Good video on Standard Scaling -
<https://www.youtube.com/watch?v=4zX-iBDRn38>
- Good Blog on Outlier removal -
<https://towardsdatascience.com/ways-to-detect-and-remove-the-outliers-404d16608dba>
- Code Along
- Questions and Discussion on doubts - AMA

Next Session

- Concept - Logistic Regression (30 min)
- Key topics to be highlighted - highlight where they would need to spend more time and importance w.r.t Data Science.
 - Difference between Linear and Logistic Regression
 - Sigmoid function
 - Cost function with gradient descent
 - Evaluation metrics