1. Data Understanding and Preprocessing

- **Loading and Understanding Data**
    - **Importing datasets (e.g., CSV, SQL, JSON) effectively.**
    - **Summarizing datasets: Checking the shape, columns, data types, missing values, and basic statistics.**
    - **Demonstrating domain understanding through exploratory data analysis (EDA).**
- **Handling Missing Data**
    - **Identifying missing or incorrect data and applying appropriate imputation techniques (mean, median, mode, interpolation, etc.).**
- **Data Cleaning**
    - **Removing duplicates, correcting data entry errors, and ensuring uniformity in categorical variables.**
- **Feature Encoding and Transformation**
    - **Encoding categorical variables (e.g., one-hot encoding, label encoding).**
    - **Scaling and normalizing numerical features (e.g., MinMaxScaler, StandardScaler).**
- **Feature Engineering**
    - **Creating new features from existing ones to improve model performance.**

---

2. Exploratory Data Analysis (EDA) and Visualization

- **Descriptive Statistics**
    - **Summarizing data using measures like mean, median, variance, and correlation.**
- **Visualizations**
    - **Creating insightful plots to reveal trends, patterns, and anomalies:**
        - **Histograms, box plots, scatter plots, and pair plots.**
        - **Advanced plots: Heatmaps, time-series plots, or geospatial visualizations.**

        - **Interactive visualizations using Plotly, Tableau, or Dash (optional).**
- **Insights**
    - **Clearly communicating findings from EDA (e.g., trends, correlations, anomalies).**

---

3. Model Development and Implementation

- **Model Selection**
    - **Choosing the appropriate algorithm(s) based on the problem statement (e.g., regression, classification, clustering).**
    - **Justifying the choice of the model with reasoning.**
- **Implementation**
    - **Writing clean, modular, and well-documented code for model training and evaluation.**
    - **Using frameworks such as Scikit-learn, TensorFlow, PyTorch, or XGBoost.**
- **Hyperparameter Tuning**
    - **Demonstrating techniques like GridSearchCV, RandomizedSearchCV, or Bayesian Optimization for improving model performance.**

---

4. Model Evaluation and Validation

- **Performance Metrics**
    - **Properly calculating and interpreting relevant metrics, such as:**
        - **Regression: RMSE, MAE, R^2.**
        - **Classification: Accuracy, Precision, Recall, F1-score, ROC-AUC.**
        - **Clustering: Silhouette score, Davies-Bouldin index.**
- **Cross-Validation**
    - **Implementing K-Fold or other cross-validation techniques to ensure robustness.**

- **Error Analysis**
    - **Analyzing incorrect predictions and understanding areas of model weakness.**

---

5. Explainability and Interpretation

- **Feature Importance**
    - **Using techniques like SHAP, LIME, or built-in feature importance to explain model predictions.**
- **Model Transparency**
    - **Providing insights into how the model arrives at predictions (especially for black-box models).**

---

6. Scalability and Deployment (Optional only for DA Student)

For advanced students, additional marks can be awarded for implementing solutions that are scalable and deployable.

- **Deployment Readiness**
    - **Building a simple web app using Flask, Django, or Streamlit for model predictions.**
    - **Creating APIs for the trained model.**
- **Scalability Features**
    - **Writing efficient code for large datasets or parallelizing computations.**

---

7. Documentation and Presentation

- **Code Documentation**
    - **Writing meaningful comments and ensuring proper function and variable naming.**
- **Report Submission**
    - **Delivering a concise, well-structured report or presentation covering:**
        - **Problem understanding.**

        - **Key steps taken (EDA, modeling, evaluation).**
        - **Final solution and recommendations.**
- **Presentation Skills**
    - **Explaining the results and code logic effectively during a demo or Q&A session.**

---

8. Innovation and Creativity

- **Novel Approaches**
    - **Implementing innovative techniques or proposing unconventional solutions.**
    - **Effective use of advanced tools, libraries, or methods.**
- **Real-World Applicability**
    - **Connecting their solution to real-world use cases or suggesting practical implications.**