**Evaluation Rubric**

**Rubrics**

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Meets expectations** | **Does not meet expectations** |
| **Data understanding, preparation and EDA (25%)** | Data quality checks, if any, are performed, and all data quality issues are addressed in the right way. Explanations for data quality issues are clearly mentioned in comments.    Categorical variables are handled appropriately.        Dummy variables are created properly wherever applicable.        New metrics are derived if applicable and are used for analysis and modelling.        The data is converted to a clean format suitable for analysis. | Data quality checks, if any, are not done, data quality issues are not addressed correctly to an appropriate level.    Categorical variables are not handled appropriately where required.    Dummy variables are not created properly.      New metrics are not derived or are not used for analysis.    The data is not converted to a clean format which is suitable for analysis or is not cleaned using commands. |
| **Model building and evaluation (50%)** | Model parameters are tuned using correct principles and the approach is explained clearly. Both the technical and business aspects are considered while building the model.      Correct variable selection techniques are used. A reasonable number of different models are attempted and the best one is chosen based on key performance metrics.    Residual analysis is performed after model building and the assumptions are validated.    Model evaluation is done using the correct principles and appropriate evaluation metrics are chosen.      The results are on par with the best possible model on the dataset.    The model is interpreted and explained correctly. The commented code includes a brief explanation of the important variables and the model in simple terms. | Parameters are not tuned enough or tuned incorrectly. Relevant business aspects are not considered while model building.    Variable selection techniques are used incorrectly / not conducted. A variety of models are not considered or a sub-optimal one is finalised.    Residual analysis is not performed after model building and the assumptions are not validated.    The evaluation process deviates from correct model selection principles, inappropriate metrics are evaluated or are incorrectly evaluated.      The results are not on par with the best possible model on the dataset.    The model is not interpreted and explained correctly. |
| **Coding Guidelines (5%)** | Appropriate comments are written wherever applicable.        If new variables are created, the names are descriptive and unambiguous.      The code is written concisely wherever possible.        Overall, code readability is good with appropriate indentations. | Comments are not written rendering the code difficult to understand.      Variables are poorly or ambiguously named.      The code is more complex than what is required by the problem.    Code readability is poor because of poor indentation / other reasons. |
| **Subjective Questions (20%)** | The answers are correct, concise, and to the point.    Examples are provided wherever necessary. | The answers are incorrect or unnecessarily long.    Appropriate examples are not provided in places where required. |