|  |  |
| --- | --- |
| Proposal clearly states whether the problem is a classification or a regression problem | 2.5% |
| Input variables are correctly identified/described | 2.5% |
| Response variable(s) are correctly identified/described | 2.5% |
| Proposal discusses which variables SHOULD NOT be included in the model and explains WHY | 2.5% |
| Exploratory analysis shows the following:   * the number of rows and columns * columns names and their associated data types * number of missing values for each column * number of unique values for each column | 5% |
| Marginal distributions for ALL variables in the data are appropriately visualized | 10% |
| Variable relationships are appropriately visualized:   * Categorical-to-categorical relationships (COMBINATIONS) * Categorical-to-continuous relationships * Continuous-to-continuous relationships | 15% |
| Relationship changes across groups are appropriately visualized | 10% |
| Cluster analysis performed correctly using k-means clustering | 20% |
|  |  |
| Visualizations are used appropriately to interpret clustering results | 15% |
| Discussion includes detailed explanations/interpretations of clustering results | 15% |