Section	Ask	Points	What good looks like	What average looks like	What poor looks like	What nothing looks like	Weightage
		60	80-100%	60-80%	<60%	0	100.00%
Exploratory Data Analysis	- Problem definition - Univariate analysis - Bivariate analysis - Use appropriate visualizations to identify the patterns and insights - Key meaningful observations on individual variables and the relationship between variables	**8**	1) Defined problem statement, Observations on shape of data, data types of various attributes, statistical summary. (2) 2) Univariate and Bivariate analysis (variable distributions, interactions between variables) to understand the relationships in data beyond the set of questions already provided (2) 3) The analysis includes answers for Insight-based questions asked (2) **Illustrate the insights based on EDA: (2)** 1) Commented on range of attributes, outliers of various attributes. 2) Commented on the distribution of the variables and potential fixes for better modelling. 3) Key insights shared from univariate and bivariate analysis.	1) Definition of problem, Observations on shape of data, data types of various attributes, statistical summary. 2) The analysis includes answers for Insight-based questions. **Illustrate the insights based on EDA:** 1) Commented on range of attributes, outliers of various attributes 2) Insights provided are very brief or missing for plots in univariate and bivariate analysis	1) Definition of problem (as per given problem statements) and Observations on data types of various attributes, statistical summary. **Illustrate the insights based on EDA:** 1 - Univariates done with no comments 2 - A few random bivariate done with little commentary	1) No EDA is done 2) Did not comment on range of attributes, outliers of various attributes 3) Did not provide insights for any plot	**13.33%**
Data Preprocessing	- Prepare the data for analysis - Missing Value Detection and Treatment (if needed with rationale) - Outlier Detection and Treatment (if needed with rationale) - Feature Engineering (if needed with rationale) - Prepare data for modeling	**5**	1) Dropped case_id column with a reason for the same. (1) 2) Identified and treated the negative values in number of employees column, (2) 3) Identified outliers and provided comments on it - Treatment is not necessary but a valid reasoning should be provided for treatment/non-treatment. (1) 4) Prepared the data by dividing it into train, test, and validation sets (1)	1) Dropped case_id column with a reason for the same. 2) Identified and treated the negative values in number of employees column. 3) Identified outliers but no comments are provided	1) Dropped case_id column with a reason for the same. 2) Identified and treated the negative values in number of employees column.	No data pre-processing done	**8.33%**
Model Building - Original Data	- Build atleast 5 classification models (Using decision trees, random forest, bagging classifier and boosting methods) * You can choose not to build XGBoost if you are facing issues with installation	**6**	1) Choose the metric of choice with proper rationale [1] 2) Build 5 classification models and train them using the original train data [1x5]	1) Build 3-5 classification models and trained them using the original train data	1) Build < 3 classification models and trained them using the original train data	1) Did not train models with default hyperparameters or 2) Trained regressors instead of classifiers or 3) Did not run the cell/ result is an error	**10.00%**
Model Building - Oversampled Data	- Build atleast 5 classification models using oversampled train data (Using decision trees, random forest, bagging classifier and boosting methods) * You can choose not to build XGBoost if you are facing issues with the installation	**6**	1) Oversample the train data [1] 2) Build 5 classification models and train them using oversampled data [1x5]	1) Oversampled the train data 2) Build 3-5 classification models and trained them using oversampled data	1) Oversampled the train data 2) Build less than 3 classification models and trained them using oversampled data	1) Did not train models with default hyperparameters or 2) Trained regressors instead of classifiers or 3) Did not run the cell/ result is an error 4) Did not choose any scorer (or chose accuracy) to tune the models	**10.00%**

Model Building - Undersampled Data	- Build atleast 5 classification models using undersampled train data (Using decision trees, random forest, bagging classifier and boosting methods) * You can choose not to build XGBoost if you are facing issues with the installation	**6**	1) Undersample the train data [1] 2) Build 5 classification models and train them using undersampled data [1x5]	1) Undersampled the train data 2) Build 3-5 classification models and trained them using undersampled data	1) Undersampled the train data 2) Build less than 3 classification models and trained them using undersampled data	1) Did not train models with default hyperparameters or 2) Trained regressors instead of classifiers or 3) Did not run the cell/ result is an error	**10.00%**
Hyperparameter Tuning	- Choose atleast 3 best performing models among all the models built previously (Mention the reason for the choices made) - Tune the chosen models - Check the performance of the tuned models	**10**	1) Choose atleast 3 best performing models among all the models built previously and mention proper reasoning for choosing those models [1] 3) Tune the chosen models with metric of interest [2.5x3] 4) Check the performance of the tuned models [0.5x3]	1) Choose atleast 3 models 2) Tune the chosen models	1) Choose and tune less than 3 models obtained above with metric of interest	1) Did not train models with default hyperparameters or 2) Trained regressors instead of classifiers or 3) Did not run the cell/ result is an error 4) Did not choose any scorer (or chose accuracy) to tune the models	**16.67%**
Model Performances	- Compare performances of the tuned models and choose a final model. - Check the performance of final model on test data.	**5**	1) Compare model performances of tuned models [2] 2) Choose best model [1] 3) Find test performance [1] 4) Metric of choice > 0.8 on the test set [1]	1) Compare model performances of tuned models 2) Choose best model 3) Find test performance 4) 0.75 < Metric of choice < 0.8 on test data	1) Choose best model 2) Metric of choice < 0.75 on test data		**8.33%**
Actionable Insights & Recommendations	- Compare model performance on various metrics. - Conclude with the key takeaways for the business	**6**	1) 3-4 Conclusions provided on the model performance comparison and EDA, key takeaways in form of important features identified. (4) 2) 2-3 Recommendations mentioned (2) [Recommendations can also include points on additional data sources for further analysis, model implementation in real world, potential business benefits from improving the model, etc.]	1) 1-2 Conclusions provided on the model performance comparison and EDA, key takeaways in form of important features identified. 2) 1 Recommendation mentioned	1) Any Conclusions provided on the model performance comparison and EDA 2) No recommendations	None of the steps are performed.	**10.00%**

Presentation / Notebook - Overall quality	- Structure and flow - Crispness - Visual appeal - Conclusion and Business Recommendations OR - Structure and flow - Well commented code - Conclusion and Business Recommendations	**8**	1) Clear structure and flow - everything sits well in a story 2) Crispness (Not too many words, just enough to keep the focus on key things/points) 3) Visual appeal (Use of charts, colors, diagrams, format, symmetry, informative visualizations that are easy to interpret) OR 1) Well structured notebook with a logical flow 2) Clean and well commented code	1) There is structure and flow but some bits are jumbled 2) Points are made but in too many words 3) Lesser charts, format is not the cleanest OR 1) There is structure and flow but some bits are missing 2) Some of the code is commented * If any section is missing and points have been deducted for the same in a previous section, then no points should be deducted in this section	1) No structure or flow 2) Only a few points are covered, story is not complete 3) Not many visuals used OR 1) no structure or flow 2) no comments in the code	No report OR No code	**13.33%**