

Section	Ask	Points	What good looks like	What average looks like	What poor looks like	What nothing looks like
		40	80-100%	60-80%	<60%	0
Define the problem and perform an Exploratory Data Analysis	<ul style="list-style-type: none"> <li>- Univariate analysis</li> <li>- Bivariate analysis</li> <li>- Use appropriate visualizations to identify the patterns and insights</li> <li>- Come up with a customer profile (characteristics of a customer) of the different packages</li> <li>- Any other exploratory deep dive</li> </ul>	6	1) Definition of problem (as per given problem statement with additional views) and Observations on shape of data, data types of various attributes, missing values, statistical summary. (1) 2) Univariate Analysis (boxplots, histograms, distribution plots for continuous variables like Age, Duration of Pitch, Monthly Income and countplots for categorical variables) (1) 3) Bivariate Analysis (correlation matrix heatmap, pairplot, plots between independent variables and the target variable like Designation VS ProfTaken, Age VS ProdTaken, etc.) (2) 4) Customer Profiling (Profile on the customers who purchased a package for each of the 5 packages - Basic, Standard, Deluxe, Super Deluxe, and King. Provide 3-4 points on distinguish features for each package) (2)	1) Definition of problem (as per given problem statement with additional views) and Observations on shape of data, data types of various attributes, missing values, statistical summary. 2) Univariate Analysis (boxplots, histograms, distribution plots for continuous variables like Age, Duration of Pitch, Monthly Income and countplots for categorical variables) 3) Bivariate Analysis (correlation matrix heatmap, pairplot, plots between independent variables and the target variable like Designation VS ProfTaken, Age VS ProdTaken, etc.)	1) Definition of problem (as per given problem statements) and Observations on data types of various attributes, statistical summary.  2) Univariate and Bivariate Analysis (any plot)	No EDA is done in the notebook
Illustrate the insights based on EDA	Key meaningful observations on individual variables and the relationship between variables	3	1) Comments on range of attributes, outliers of various attributes 2) Comments on the distribution of the variables and meaningful insights extracted from univariate and bivariate plots	1) Comments on range of attributes, outliers of various attributes 2) Insights are very brief or missing for plots in univariate and bivariate analysis	1 - Univariates done with no comments 2 - A few random bivariate done with little commentary	1) No comments on range of attributes, outliers of various attributes 2) No Insights provided for any plot
Data pre-processing	<ul style="list-style-type: none"> <li>- Prepare the data for analysis</li> <li>- Missing value Treatment,</li> <li>- Outlier Detection(treat, if needed- why or why not ),</li> <li>- Feature Engineering,</li> <li>- Prepare data for modeling</li> </ul>	6	1) Dropped ID column with a reason for the same. (0.5) 2) Replaced data error i.e. "Fe male" with "Female" in gender column (0.5) 2) Identify and impute missing values in every column (median for continuous variables and mode for categorical variables) (2) 3) Identify and treat extreme values (drop or cap these values) in variables - DurationofPitch and Monthly Income, and NumberofTrips with reasoning. (1) 4) Drop columns 'DurationOfPitch', 'NumberOfFollowups', 'ProductPitched', 'PitchSatisfactionScore' with elaborated reason for the same (2)	1) Dropped ID column with a reason for the same and replaced data error i.e. "Fe male" with "Female" in gender column 2) Identify and impute missing values in every column (median for continuous variables and mode for categorical variables) 3) Identify and treat extreme values (drop or cap these values) in variables - DurationofPitch and Monthly Income, and NumberofTrips	1) Dropped ID column with a reason for the same and replaced data error i.e. "Fe male" with "Female" in gender column 2) Identify and impute missing values in every column (median for continuous variables and mode for categorical variables)	No data pre-processing done in the notebook
Model building - Bagging	<ul style="list-style-type: none"> <li>- Build Bagging classifier, Random Forest, and Decision Tree.</li> <li>- Comment on model performance</li> </ul>	3	1) Train the decision tree classifier with default parameters and comment on model performance(1) 2) Train the bagging classifier with default parameters and comment on model performance (1) 3) Train the random forest classifier with default parameters and comment on model performance(1)	1) Train the decision tree classifier with default parameters 2) Train the bagging classifier with default parameters 3) Train the random forest classifier with default parameters	Trained only one or two out of the three models - Decision Tree, Bagging, Random Forest	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
Model Improvement - Bagging	<ul style="list-style-type: none"> <li>- Comment on which metric is right for model performance evaluation and why?</li> <li>- Comment on the model performance after tuning the Decision Tree, Bagging, and Random Forest classifier to improve the model performance.</li> </ul>	7	1) Choose and provide reasoning for the best metric to evaluate the performance of the model and for scoring in GridSearchCV. <b>Recall should be selected.</b> (1) 2) Try to improve the performance of the decision tree classifier using hyperparameter tuning with GridSearchCV and comment on model performance (2) 3) Try to improve the performance of the bagging classifier using hyperparameter tuning with GridSearchCV and comment on model performance(2) 4) Try to improve the performance of the random forest classifier using hyperparameter tuning with GridSearchCV and comment on model performance (2)	1) Choose the best metric to evaluate the performance of the model and for scoring in GridSearchCV. <b>Recall or F1-Score</b> should be selected. 2) Try to improve the performance of the decision tree classifier using hyperparameter tuning with GridSearchCV 3) Try to improve the performance of the bagging classifier using hyperparameter tuning with GridSearchCV 4) Try to improve the performance of the random forest classifier using hyperparameter tuning with GridSearchCV	1) Choose the best metric to evaluate the performance of the model and for scoring in GridSearchCV. <b>Precision is selected.</b> 2) Tuning used for only or two models out of the three models - Decision Tree, Bagging, Random Forest	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
Model building - Boosting	<ul style="list-style-type: none"> <li>- Build Adaboost, GradientBoost, XGBoost, and Stacking classifiers</li> <li>- Comment on model performance</li> </ul>	4	1) Train the adaboost classifier with default parameters and comment on model performance(1) 2) Train the gradient boosting classifier with default parameters and comment on model performance (1) 3) Train the xgboost classifier with default parameters and comment on model performance (1) 4) Train the stacking classifier and comment on model performance (1)	1) Train the adaboost classifier with default parameters 2) Train the gradient boosting classifier with default parameters 3) Train the xgboost classifier with default parameters 4) Train the stacking classifier	Trained only two or three out of the four models - Adaboost, Gradient Boosting, Xgboost, Stacking	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
Model Improvement - Boosting	<ul style="list-style-type: none"> <li>- Comment on which metric is right for model performance evaluation and why?</li> <li>- Comment on the model performance after tuning the AdaBoost, and Gradient Boosting classifier on the appropriate metric to improve the model performance.</li> </ul> <p>* Please note XGBoost can take a significantly longer time to run, so if you have time complexity issues then you can avoid tuning XGBoost.</p>	4	1) Try to improve the performance of the adaboost classifier using hyperparameter tuning with GridSearchCV and comment on model performance(2) 2) Try to improve the performance of the gradient boosting classifier using hyperparameter tuning with GridSearchCV and comment on model performance (2)	1) Try to improve the performance of the adaboost classifier using hyperparameter tuning with GridSearchCV 2) Try to improve the performance of the gradient boosting classifier using hyperparameter tuning with GridSearchCV	1) Tuning used for only or one of models out of the two models - Adaboost, Gradient Boosting	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
Actionable Insights & Recommendations	<ul style="list-style-type: none"> <li>- Compare model performance on various metrics.</li> <li>- Conclude with the key takeaways</li> <li>- What would your advice be to grow the business?</li> </ul>	4	1) Compare model performances and provide conclusion on the model performance comparison, key takeaways in form of important features identified. (2)  3) Recommendations - such that the business can take action upon them. (2)	1) Compare model performances and comment  2) Recommendations - such that the business can take action upon them.	1) Compare model performances and comment.	None of the steps are performed.
Notebook - Overall	<ul style="list-style-type: none"> <li>- Structure and flow</li> <li>- Well commented code</li> </ul>	3	-Well structured notebook with a logical flow -Clean and well commented code	- There is structure and flow but some bits are missing - Some of the code is commented	- no structure or flow - no comments in the code	- no code