Section	Ask	Points	What good looks like	What average looks like	What poor looks like	What nothing looks like	
		60	80-100%	60-80%	<60%	0%	% Weightage
Define the problem and perform an Exploratory Data Analysis	- Problem definition, questions to be answered - Data background and contents - Univariate analysis - Bivariate analysis	5	1) Definition of problem (as per given problem statement with additional views) 2) Observations on shape of data, data types of all the attributes, conversion of categorical attributes to 'category', missing value detection, statistical summary. 3) Univariate Analysis (distribution plots of all the continuous variable(s), barplots/countplots of all the categorical variables) 4) Bivariate Analysis (Relationships between important variables such as conversion status and time spent on the page, preferred language and time spent on the page, landing page type and time spent on the page)	1) Definition of problem (as per given problem statements) 2) Observations on data types of various attributes, missing value detection, statistical summary. 3) Univariate Analysis (distribution plots of few continuous variable(s), barplots/countplots of few categorical variables) 4) Bivariate Analysis (Relationships between few unimportant variables)	1) Definition of problem (as per given problem statements) 2) Observations on data types of various attributes, statistical summary. 3) Univariate and Bivariate Analysis (any plot)		8.33
Illustrate the insights based on EDA	Key meaningful observations on individual variables and the relationship between variables	5	1) Comments on range of attributes, outliers of various attributes 2) Comments on the distribution of the variables and relationship between them 3) Comments for each univariate and bivariate plots	1) Comments on range of attributes, outliers of various attributes 2) Comments for some univariate and bivariate plots	1) A few random univariate done with little commentary 2) A few random bivariates done with little commentary		8.33
Do the users spend more time on the new landing page than the old landing page?	Perform the hypothesis test (weightage) - Visual analysis (1) - Hypothesis formulation (2) - Select the appropriate test (1) - Data collection and preparation(2) - Find the p-value(2) - Conclusion based on the p-value(2)	10	1) Visual analysis 2) Correct null and alternative hypothesis 3) Selection of appropriate test 4) Appropriate data collection 5) Correct p-value 6) Appropriate conclusion and inference based on the p-value	1) Correct null and alternative hypothesis 2) Selection of appropriate test 3) Partial data collection 4) Correct p-value 5) Appropriate conclusion and inference based on the p-value	1) Partially correct null and alternative hypothesis 2) Selection of appropriate test 3) Partial data collection 4) Incorrect p-value 5) Appropriate conclusion and inference based on the p-value		16.67
Is the conversion rate (the proportion of users who visit the landing page and get converted) for the new page greater than the conversion rate for the old page?	Perform the hypothesis test (weightage) - Visual analysis (1) - Hypothesis formulation (2) - Select the appropriate test (1) - Data collection and preparation(2) - Find the p-value(2) - Conclusion based on the p-value(2)	10	1) Visual analysis 2) Correct null and alternative hypothesis 3) Selection of appropriate test 4) Appropriate data collection 5) Correct p-value 6) Appropriate conclusion and inference based on the p-value	1) Correct null and alternative hypothesis 2) Selection of appropriate test 3) Partial data collection 4) Correct p-value 5) Appropriate conclusion and inference based on the p-value	1) Partially correct null and alternative hypothesis 2) Selection of appropriate test 3) Partial data collection 4) Incorrect p-value 5) Appropriate conclusion and inference based on the p-value		16.67

Criteria	Weightage	Points	Full	Partial	Lower	Lowest	Score
Does the converted status depend on the preferred language? [Hint: Create a contingency table using the pandas.crosstab() function]	Perform the hypothesis test (weightage) - Visual analysis (1) - Hypothesis formulation (2) - Select the appropriate test (1) - Data collection and preparation(2) - Find the p-value(2) - Conclusion based on the p-value(2)	10	1) Visual analysis 2) Correct null and alternative hypothesis 3) Selection of appropriate test 4) Appropriate data collection 5) Correct p-value 6) Appropriate conclusion and inference based on the p-value	1) Correct null and alternative hypothesis 2) Selection of appropriate test 3) Partial data collection 4) Correct p-value 5) Appropriate conclusion and inference based on the p-value	1) Partially correct null and alternative hypothesis 2) Selection of appropriate test 3) Partial data collection 4) Incorrect p-value 5) Appropriate conclusion and inference based on the p-value		16.67
Is the mean time spent on the new page same for the different language users?	Perform the hypothesis test (weightage) - Visual analysis (1) - Hypothesis formulation (2) - Select the appropriate test (1) - Data collection and preparation(2) - Find the p-value(2) - Conclusion based on the p-value(2)	10	1) Visual analysis 2) Correct null and alternative hypothesis 3) Selection of appropriate test 4) Appropriate data collection 5) Correct p-value 6) Appropriate conclusion and inference based on the p-value	1) Correct null and alternative hypothesis 2) Selection of appropriate test 3) Partial data collection 4) Correct p-value 5) Appropriate conclusion and inference based on the p-value	1) Partially correct null and alternative hypothesis 2) Selection of appropriate test 3) Partial data collection 4) Incorrect p-value 5) Appropriate conclusion and inference based on the p-value		16.67
Presentation - Overall quality	- Structure and flow(0.5) - Crispness (0.5) - Visual appeal (0.5) - Key insight and recommendation based on EDA (1) - Formulated hypothesis and inference based on the p-value (2.5)	5	- Clear structure and flow - everything sits well in a story - (Problem - Data overview - Solution overview - Findings - Recommendations) - Crispness - not too many words - just enough to keep the focus on key things/points - Visual appeal - use of charts, colors, diagrams, format, symmetry - informative visualizations that are easy to interpret - All key insights and recommendations covered - all key ones from EDA are stressed upon & important insights are not left just in the notebook - The null and alternative hypotheses are covered - Inference based on the p-value is presented	- There is structure and flow but some bits are missing / jumbled - Points are made but in too many words - lesser charts, format is not the cleanest - key insights are covered but some points are missed - The null and alternative hypotheses are covered - Inference based on the p-value is presented	- no structure or flow - only a few points are covered - story is not complete - not many visuals used - final insights and recommendations are missing - Conclusion of hypothesis test is missing	- no presentation	8.33
Notebook - Overall	- Structure and flow - Well commented code	5	-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	8.33

Section(s)	Important Notes for the Evaluators						
Is the conversion rate (the proportion of users who visit the landing page and get converted) for the new page greater than the conversion rate for the old page?	- For this problem, Two proportions z-test/Chi-square test of independence/Fishers' exact test can be used to draw the inference. - We have included the two proportions z-test in our sample solution. But, the learners may use either of the three tests to draw the conclusion. The p-value might be different depending on the test. Marks should not be deducted in that case.						
All the sections where Hypothesis Testing has been used.	- Please follow the weightage given in this rubric to deduct marks for the hypothesis testing. - Some learners may use non-parametric tests (advanced tests not covered in the course) to answer the questions. In that case, marks should not be deducted if the approach is correct. - Marks should not be deducted if the learner uses the Rejection Region approach to perform hypothesis testing instead of the p-value approach.						