

Perform Data Cleaning

Criteria	Excellent	Good	Needs Improvement
Remove duplicate or irrelevant observations	Thoroughly identifies and removes all duplicates and irrelevant data, ensuring dataset integrity and relevance.	Identifies and removes most duplicates and irrelevant data with minimal errors.	Fails to adequately identify and remove duplicates and irrelevant data, leaving significant issues.
Fix structural errors	Corrects all structural errors e.g., typos, inconsistent formats to ensure uniform data presentation.	Corrects most structural errors, but some minor inconsistencies remain.	Makes minimal or ineffective attempts to correct structural errors, resulting in significant inconsistencies.
Fix unwanted outliers	Accurately identifies and appropriately handles all unwanted outliers, improving data quality.	Identifies and handles most outliers, but some are missed or inappropriately addressed.	Fails to identify or properly handle unwanted outliers, impacting data quality.
Handle missing data	Effectively addresses all missing data using appropriate techniques, ensuring minimal impact on analysis.	Handles most missing data adequately, but some issues remain that may affect analysis.	Inadequately addresses missing data, leading to significant gaps and potential analysis issues.

Exploratory Data Analysis Rubric

Criteria	Excellent	Good	Satisfactory	Needs Improvement	Unsatisfactory
What is the highest JOBID in the web industry?	Accurately identifies and justifies the highest paying job	Correctly identifies with minor justification issues .	Identifies the highest paying job, but with some inaccuracies or lacks justification	Attempts to identify but has significant errors or lacks justification .	No attempt or completely incorrect
Rank the top job roles with the highest salary for all industries.	Correctly ranks top job roles with clear justification and methodology	Correctly ranks most jobs with minor issues .	Correctly ranks some jobs with several issues .	Attempts ranking but has significant errors .7	No attempt or completely incorrect
Which of the industries has the highest average salary?	Accurately identifies and justifies the highest paying industry	Correctly identifies with minor justification issues .	Identifies the industry but with some inaccuracies or lacks justification	Attempts to identify but has significant errors or lacks justification .	No attempt or completely incorrect
Which job role has the lowest pay?	Accurately identifies and justifies the lowest paying job	Correctly identifies with minor justification issues .	Identifies the lowest paying job, but with some inaccuracies or lacks justification	Attempts to identify but has significant errors or lacks justification .	No attempt or completely incorrect
Is there anything interesting in the data for the job role Janitor?	Accurately identifies and justifies it	Correctly identifies with minor justification issues .	Identifies the industries but with some inaccuracies or lacks justification	Attempts to identify but has significant errors or lacks justification .	No attempt or completely incorrect
Which industry has the highest percentage of	Correctly identifies and justifies the	Correctly identifies with minor issues in	Identifies the industry but with several	Attempts to identify but has significant errors	No attempt or completely incorrect

people who are below the median salary?	industry with clear analysis	analysis or justification .	inaccuracies or lacks clear analysis .	or lacks justification	
What are the job roles that are above the median salary?	Correctly lists and justifies the job roles with clear analysis	Correctly lists most job roles with minor issues in analysis or justification .	Lists some job roles but with several inaccuracies or lacks clear analysis	Attempts to list but has significant errors or lacks justification .	No attempt or completely incorrect
Determine if there is a relationship between job role and salary.	Thorough analysis with clear visualizations and justification	Good analysis with minor issues in visualizations or justification .	Basic analysis with several issues in visualizations or justification .	Attempts analysis but with significant errors or lacks justification	No attempt or completely incorrect
Is there a relationship between distance from CBD and job role?	Thorough analysis with clear visualizations and justification	Good analysis with minor issues in visualizations or justification .	Basic analysis with several issues in visualizations or justification .	Attempts analysis but with significant errors or lacks justification	No attempt or completely incorrect
Does the major they studied affect the salary?	Thorough analysis with clear visualizations and justification	Good analysis with minor issues in visualizations or justification .	Basic analysis with several issues in visualizations or justification .	Attempts analysis but with significant errors or lacks justification	No attempt or completely incorrect

ML Model Development Rubric

Criteria	Excellent	Good	Needs Improvement
Model Selection and Justification	Thorough explanation of model choice, considering data characteristics and ML theory. Welljustified and detailed reasoning.	Good explanation with minor gaps in reasoning or missing considerations.	Inadequate or incorrect explanation with significant gaps in reasoning.
Data Preparation and Feature Engineering	Comprehensive and appropriate data preprocessing and feature engineering.	Good preprocessing and feature engineering with minor issues or omissions.	Incomplete or incorrect preprocessing and feature engineering.
Model Training and Tuning	Correctly trains model, applies hyperparameter tuning, and uses crossvalidation.	Good model training and tuning with minor issues or limited crossvalidation.	Inadequate model training and tuning with significant errors.
Model Evaluation and Interpretation	Thorough evaluation with appropriate metrics, clear interpretation of results.	Good evaluation with minor issues in metrics or interpretation.	Incomplete or incorrect evaluation and interpretation of results.

Comparison between PySpark and NonPySpark Workflow

Criteria	Excellent	Good	Needs Improvement
Comparison of Speed	Thorough comparison of execution speed with clear metrics and examples.	Good comparison with minor issues or less detailed metrics.	Inadequate or superficial comparison with significant gaps.
Comparison of Ease of Use	Comprehensive evaluation of ease of use, considering user experience, learning curve, and coding complexity.	Good evaluation with minor issues or less detailed analysis.	Inadequate evaluation with significant gaps or superficial analysis.
Comparison of Efficiency	Detailed comparison of efficiency, including resource utilization and scalability.	Good comparison with minor issues or less detailed analysis.	Inadequate comparison with significant gaps or superficial analysis.
Recommendation and Justification	Clear, welljustified recommendation based on the comparison, tailored to the specific use case.	Good recommendation with minor issues or less detailed justification.	Inadequate or incorrect recommendation with insufficient justification.

Recommend Strategies for Career Development

Criteria	Excellent	Good	Needs Improvement
Clarity and Practicality of Recommendations	Provides clear, practical, and actionable strategies. Recommendations are easy to understand and implement.	Provides practical recommendations but may lack some clarity or detail, making them slightly less effective.	Provides vague or unclear recommendations, making them difficult to understand and implement.
Justification and Relevance of Strategies	Thoroughly justifies each strategy, demonstrating clear relevance to career development and personal growth.	Justifies most strategies, but some may lack detail or clear relevance to career development.	Provides inadequate justification with significant gaps or unclear relevance to career development.
Originality and Insightfulness	Recommendations demonstrate originality and provide deep insight into career development, offering unique perspectives.	Provides recommendations that are good but lack some originality or deep insight, making them less distinctive.	Provides recommendations that lack originality or insight, offering common or superficial advice.

Presentation

Criteria	Excellent	Good	Needs Improvement
Content and Relevance	Presentation effectively communicates findings relevant to the government organization's needs. All key points are covered comprehensively.	Presentation covers most key points relevant to the government organization's needs, but may lack depth in some areas.	Presentation lacks coherence and relevance to the government organization's needs, with significant gaps in content.
Clarity and Structure	Presentation is wellstructured, clear, and easy to follow. Each section flows logically and transitions smoothly.	Presentation is generally clear and structured, but may have minor issues with organization or flow.	Presentation lacks clarity and structure, making it difficult to follow. Transitions between sections are abrupt or unclear.
Engagement and Delivery	Presenter engages the panel effectively throughout the presentation. Delivery is confident, dynamic, and holds the panel's attention.	Presenter engages the panel adequately, but may lack dynamism or confidence in delivery. Delivery is somewhat engaging but could be improved.	Presenter struggles to engage the panel, with a monotone delivery or lack of confidence. Presentation lacks energy and fails to hold the panel's attention.
Use of Visual Aids	Visual aids slides, charts, graphs, etc. are used effectively to enhance understanding and support key points.	Visual aids are used adequately but may be overly cluttered or lack visual appeal. They support key points but could be more impactful.	Visual aids are poorly designed or used ineffectively, detracting from the presentation's effectiveness. They fail to support key points or are confusing.
Question and Answer	Presenter effectively interacts with panel members, addressing questions and concerns thoughtfully and thoroughly.	Presenter interacts with panel members but may struggle to address all questions or concerns adequately. Responses are generally thoughtful but could be more thorough.	Presenter struggles to interact with panel members, providing incomplete or unclear responses to questions and concerns.