



Data Analysis Competition Track

Test your skills in Data Analysis, Solutioning, and Modelling!

You MUST complete the stage you are on before unlocking the next one!

Stage	Name	Points
1	Exploratory Data Analysis (EDA)	15
2	Solution Proposal	13
3	Model Creation	25
4	Model Serving	12

Stage 1: Exploratory Data Analysis (EDA) (15 points)

Rules

Task	Deliverable	Grading
Explore and analyze the provided UNWTO dataset(s) and provide insights to at least 5 data sets.	<p>Create 5 insightful visualizations.</p> <p>At least 2 of the insights must use a minimum of 2 dataset unions.</p> <p>Explain each insight with a short paragraph (a Google docs page)</p>	<p>For each of the 5 Visualizations:</p> <ul style="list-style-type: none">• Relevance (1 points): Does the visualization directly address a question or trend?• Visualization type (1 points): Is the chosen visualization the most appropriate for the data and the question being addressed?• Clarity (1 points): Is the insight explained clearly in the Google doc? <p><i>TOTAL=3 points each Viz (total of 15)</i></p>

-----CALL A JUDGE -----

Complete Stage 1 to unlock Stage 2-A!



Data Analysis Competition Track

Stage 2-A: Solution Proposal Stage (3 points)

Rules

Task	Deliverable	Grading
Define the problem and propose a compelling solution to address the problem or question, categorizing the solution into the following: <ol style="list-style-type: none">1. Regression2. Classification3. Clustering4. Other	Deliver 3 brief written proposals for solutions on google docs . Each proposal should not exceed 200 words. The Judge will provide you feedback on each solution.	<ul style="list-style-type: none">• 1 point can be awarded per solution provided.• Feedback is provided to the team on each proposal. <i>TOTAL= 3 points for the evaluation</i>

-----CALL A JUDGE -----

Complete Stage 2-A to unlock Stage 2-B!



Data Analysis Competition Track

Stage 2-B: Solution Proposal Stage (25 points)

Rules

Deliverable: Deliver a complete documentation of the selected proposal: 2-3 pages on google docs.

Task	Grading
<p>Select 1 proposal and detail it further addressing the following criteria:</p> <ol style="list-style-type: none">Problem Statement:<ul style="list-style-type: none">Clearly define the problem and need addressed by the proposal.Assess the relevance and importance of the problem.Tools and Models:<ul style="list-style-type: none">Identify the specific models proposed to tackle the problem.Briefly explain the reasoning behind the chosen model types.Data Pipeline and Model Creation:<ul style="list-style-type: none">Outline the proposed steps for data acquisition, processing, and preparation.Describe the planned model creation process, including training and validation.Model Evaluation:<ul style="list-style-type: none">Specify the metrics that will be used to evaluate the model's performance.Outline the planned evaluation strategy.Model Deployment:<ul style="list-style-type: none">Briefly describe how the model will be served to users in a user-friendly manner (e.g., API, web application).	<ul style="list-style-type: none">Problem Statement (1 point): clearly defined problem and need, is it relevant & important?Tools & Models (3 point): What models do they think they will be able to use to tackle this problem? Is the reasoning valid? Have they explained it well?Data Pipeline & Model Creation (3 point): have they outlines the steps for the data pipeline and described the model creation process, including training and validation?Model Evaluation (2 point): Have they defined the evaluation metrics and explained the evaluation strategy?Model Deployment (1 point): Have they explained how the model will be served to the users in a user-friendly manner (API, web app, etc.) <p>TOTAL= 10 points</p>



Data Analysis Competition Track

-----CALL A JUDGE -----

Complete Stage 2 to unlock Stage 3!

Stage 3: Model Creation Stage (25 points)

Rules

Task	Grading
A -Feature Engineering	5 Points
B -Training & Testing Split	5 Points
C -Model Training	5 Points
D -Model Evaluation	10 Points
TOTAL – 25 Points	

A. Feature Engineering

Task	Deliverable	Grading
<ul style="list-style-type: none">Analyze feature importance based on chosen model type (e.g., feature importance in Random Forest).Refine or remove features based on their contribution to the model's performance.Document any feature engineering techniques employed (e.g., scaling, normalization).	<ul style="list-style-type: none">Summarize the analysis of feature importance and the rationale behind any feature selection or modification.Clearly show/ document any feature engineering techniques applied.	<p>Feature Usage (Up to 5 points based on the scale below):</p> <ul style="list-style-type: none">5 points: Feature importance analysis is performed, and relevant features are used in the model.3 points: Feature importance analysis is attempted, but feature selection is not well-justified.



Data Analysis Competition Track

		<ul style="list-style-type: none">• 0 point: No analysis of feature importance is conducted.
--	--	---

B. Training & Testing Split

Task	Deliverable	Grading
<ul style="list-style-type: none">• Implement a method for splitting the data into training and testing sets.• Common approaches include random split, stratified split, or time-based split.• Clearly define the chosen splitting ratio (e.g., 80%/20% train/test).	<ul style="list-style-type: none">• Justify the selected splitting ratio.• Ensure code demonstrates the data split implementation.	<p>Data Splitting (Up to 5 points based on the scale below):</p> <ul style="list-style-type: none">• 5 points: A proper data splitting method (e.g., stratified split) is used with a justified splitting ratio.• 3 points: A basic data splitting method (e.g., random split) is used with a reasonable splitting ratio.• 1 point: Data splitting method is unclear or poorly explained.• 0 points: Data is not split for training and testing.

C. Model Training

Task	Deliverable	Grading
<ul style="list-style-type: none">• Train the model on the prepared training data set.	<ul style="list-style-type: none">• Outline the training process, including hyperparameter tuning if applicable.	<p>Training (Up to 5 points based on the scale below):</p>

Data Analysis Competition Track

<ul style="list-style-type: none"> • Document hyperparameter tuning strategies if applicable to your chosen model. • Track and record training metrics (e.g., loss function, accuracy) during training. 	<ul style="list-style-type: none"> • Present training metrics achieved (e.g., loss curves, accuracy plots). • Ensure code demonstrates model training. 	<ul style="list-style-type: none"> • 5 points: Model training is conducted, and basic training metrics are tracked. • 3 points: Model training is attempted, but tracking of training metrics is missing. • 0 points: Model training is not demonstrated.
---	--	---

D. Model Evaluation

Task	Deliverable	Grading
<ul style="list-style-type: none"> • Evaluate the model's performance on the unseen testing data set. • Calculate relevant evaluation metrics based on the problem type (e.g., classification - accuracy, precision, recall; regression - R-squared, Mean Squared Error). • Visualize the model's performance using techniques like confusion matrices or ROC curves. 	<ul style="list-style-type: none"> • Describe the evaluation metrics used and their relevance to the problem. • Present the calculated evaluation scores for the model on the testing set. • Visualize model performance using appropriate methods (e.g., confusion matrices, ROC curves). • Ensure code for evaluation metric calculation and visualization. 	<p>Evaluation (Up to 10 points based on the scale below):</p> <ul style="list-style-type: none"> • 10 points: Relevant evaluation metrics are chosen and calculated for the model's performance. • 7 points: Some evaluation metrics are used, but their choice might not be optimal for the problem. • 3 points: Evaluation is attempted, but metrics are not clearly explained or calculated incorrectly. • 1 point: Minimal or no evaluation of the model is performed.



Data Analysis Competition Track

-----CALL A JUDGE -----

Complete Stage 2 to unlock Stage 3!

Stage 3: Model Serving Stage (12 points)

Rules

Task	Deliverable	Grading
<p>1. Choose Deployment Method: Select a method for making the model accessible to users. Common options include web applications, APIs, or standalone applications.</p> <p>2. Develop Deployment Platform:</p> <ul style="list-style-type: none">Depending on your chosen method, develop a platform to serve your model.This might involve building a web application using Python frameworks like Flask or Django, or creating an API using libraries like FastAPI. <p>3. Integrate Model:</p> <ul style="list-style-type: none">Integrate your trained model into the chosen deployment platform.	<p>Live Deployment</p> <p>Provide a live link or access to the deployed model for demonstration to the Judges.</p>	<p>Web-based (2 points):</p> <ul style="list-style-type: none">2 points: The model is deployed as a web application or API, allowing easy access for users.1 point: The model is deployed in a less user-friendly format (e.g., standalone script) requiring additional effort for user interaction.0 points: No deployment method is presented. <p>User Friendliness (3 points)</p>



Data Analysis Competition Track

<ul style="list-style-type: none">• This may involve writing code to handle user input, make predictions using the model, and return results.		Results (7 points)
4. User Interface Design: <ul style="list-style-type: none">• Design a user-friendly interface for interacting with the model.• Consider factors like clarity, ease of use, and visual appeal.		<i>TOTAL= 12 points</i>

-----CALL A JUDGE -----