**Module 4 – Performance based clustering for alliance selection**

Welcome to Module 4!  Here’s what’s included:

* **Video Tutorial** – Step-by-step walkthrough of the analysis process
* **Dataset** – Raw data for hands-on practice
* **RapidMiner Process File** – Pre-built workflow to replicate the analysis

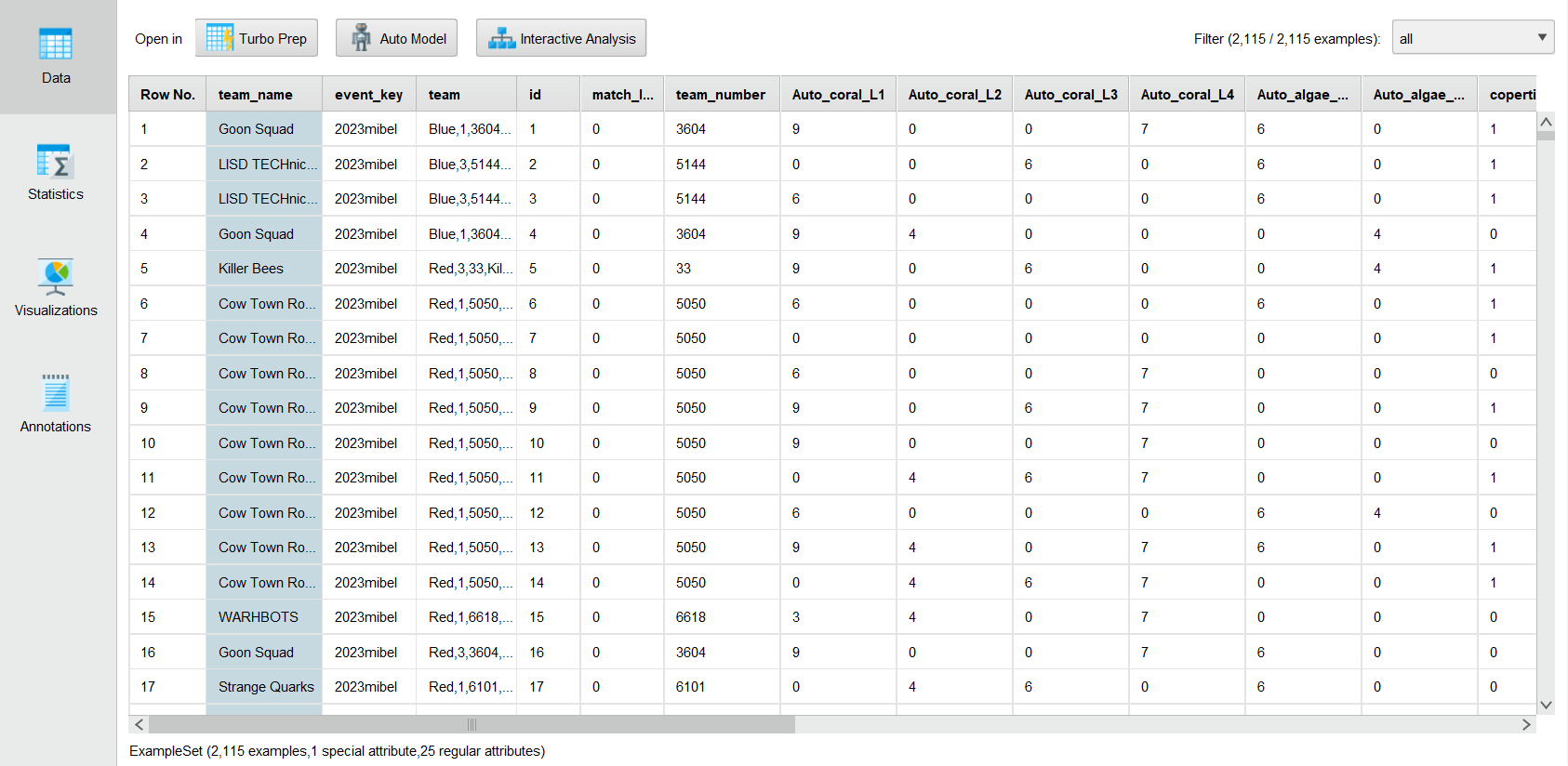
Get started by watching the video and exploring the dataset.

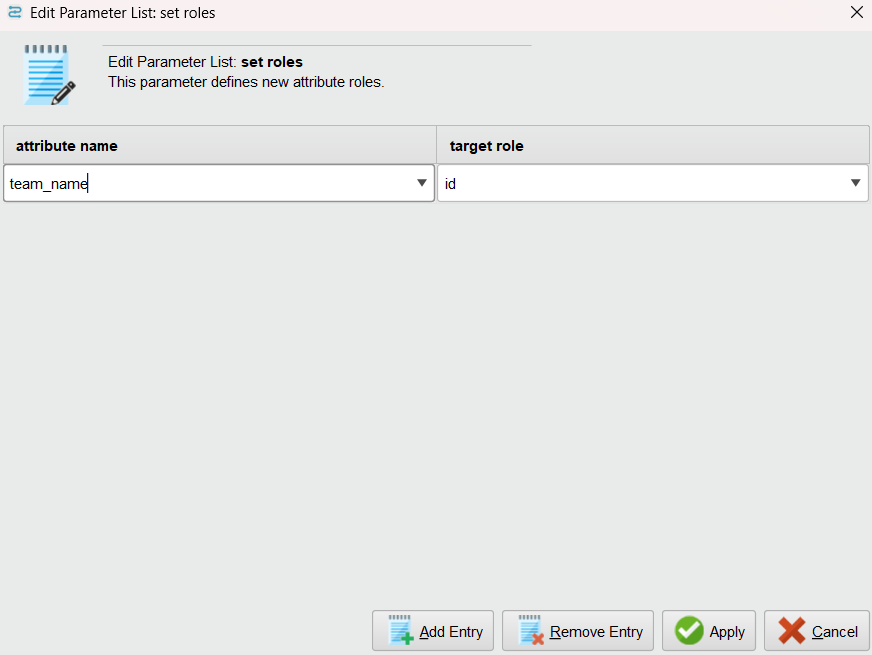
**Objective**

This workflow aims to optimize alliance selection using K-Means Clustering to group teams based on key performance metrics like auto score, teleop score, etc. The goal is to identify the best possible alliances by ensuring that teams with complementary strengths are paired together, improving overall competition performance. This process involves data preprocessing, feature engineering, clustering analysis, and selecting top teams within each cluster.

**Step-by-Step Workflow in AI Studio**

1. **Launch AI Studio**.
2. **Import** the provided dataset into a repository of your choice.
3. Use the **Retrieve** operator to access and load the dataset into your process.



1. Use the **Select Attributes** to include all the necessary attributes.
2. Bring in the **Set Role** operator to set the role of attribute **team\_name** as **id** .  
   
3. Navigate to the **Statistics** tab to check for missing values.
4. Use the **Replace Missing Values** operator to handle missing data in selected attributes. Replace missing values with the string "NA" in selected attributes.

A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

1. Use the **Filter Examples** operator to narrow down the dataset to a single event (if needed).   
     
   A screenshot of a computer

   AI-generated content may be incorrect.

9. Use the **Normalize** operator with the **default Z-transformation** option to scale all performance metrics, centering the data around zero with a standard deviation of one for consistency and balanced analysis (Use other transformation methods as needed which is best suited for your data).

A screenshot of a computer

AI-generated content may be incorrect.

10. Use the **Aggregate** operator to group performance attributes by Team Number and Event Key, and compute the average of each attribute.

A screenshot of a computer

AI-generated content may be incorrect.

11. Use the **Rename** operator to update the attribute names of the aggregated results, making them more easier to interpret.

12. Use the **Generate Attributes** operator to create key metrics such as Total Auto Coral, Total Teleop Coral, Total Auto Algae, and Total Teleop Algae and Team Reliability using the following logic for each.

* Total\_auto\_coral : [average(Auto\_coral\_L1)]+[average(Auto\_coral\_L2)]+[average(Auto\_coral\_L3)]+[average(Auto\_coral\_L4)]
* Total\_teleop\_coral : [average(Teleop\_coral\_L1)]+[average(Teleop\_coral\_L2)]+[average(Teleop\_coral\_L3)]+[average(Teleop\_coral\_L4)]
* Total\_auto\_Algae : [average(Auto\_algae\_net)]+[average(Auto\_algae\_proc)]
* Total\_teleop\_Algae : [average(Teleop\_algae\_net)]+[average(Teleop\_algae\_proc)]
* Team reliability : (0.30\*(Total\_auto\_coral +Total\_auto\_Algae)+0.30\*(Total\_teleop\_coral+Total\_teleop\_Algae ) + (0.25\* total\_barge\_score)+0.15\*(bonus\_points))

13. Use the **Select Attributes** operator to choose key metrics like Team Name, Total Auto Coral, and Total Teleop Coral for forming clusters, or optionally select other generated metrics such as Total Auto Algae based on your analysis focus.

14.Use the **K-Means Clustering** operator to group teams into clusters, setting the number of clusters (K) to 3 while keeping other parameters at their default values.

A screenshot of a computer

AI-generated content may be incorrect.

15. Analyze the **Cluster Model** and **Centroid Table** to identify performance patterns.

16. Use the **Multiply** operator to duplicate the original dataset before attribute selection, allowing you to join it later with the clustering results for a more comprehensive analysis.

17. Use the **Join** operator to perform an **inner join** on the duplicated original dataset and the clustering results using Team Name as the key attribute.

A screenshot of a computer

AI-generated content may be incorrect.

18. Use the **Filter Examples** operator to narrow down the dataset by displaying only the results from Cluster 0 for deeper analysis.

A screenshot of a computer

AI-generated content may be incorrect.

19. Use the built-in **Detect Outliers** operator to identify anomalies in the dataset, configuring the lower and upper bounds to define the neighborhood size for comparison, and applying the default Euclidean distance function for measuring similarity.

A screenshot of a computer

AI-generated content may be incorrect.

20. Use the **Filter Examples** operator again to select only those teams with an outlier score less than 1.

A screenshot of a computer

AI-generated content may be incorrect.

21. Use the **Select Attributes** operator to display only the necessary attributes within the selected teams in the cluster for further analysis and visualizations.

22. Finally go to the **Visualizations tab** and use column bar charts or any other charts as per need to analyze shortlisted teams in the cluster .Continue incorporating more attributes to gain a comprehensive view of team performance.