User Guide

This dashboard provides comprehensive data on passenger train track sections in the Netherlands, allowing users to analyze and visualize various characteristics. This step-by-step guide will guide the user through the sections of the dashboard.

The Left Menu

Include/Exclude Columns

Select or deselect characteristics from the dataset to customize your analysis. Use the checkboxes to include or exclude specific data. To include a characteristic, ensure the checkbox is selected in the left menu (see Figure 1). To exclude a characteristic, uncheck the checkbox (see Figure 2).

Figure 1: Include Characteristic

Figure 2: Exclude Characteristic

Filtering Section

In the left menu, you can filter the dataset based on specific characteristics using multiple options:

1. Select from list: Filter features based on their presence in the train track section (see Figure 3).
2. Sliders: Set a range for numerical features to filter the dataset (see Figure 4).

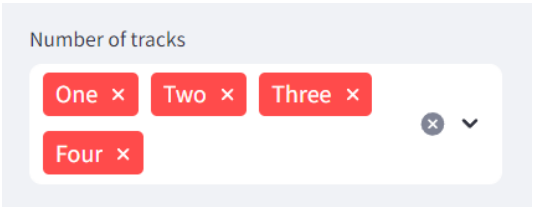

Figure 3: Select from List

Figure 4: Slider Function

Train Track Section Analysis

This section of the dashboard explains its functionalities and displays an image of the defined track sections when the expander is clicked (see Figure 5).

Figure 5: Expander to see Map

Filtering Section

This section shows results based on the filters applied in the left menu. By clicking on the expander in Figure 6, it shows the tracks that match the selected criteria in a tabular format.

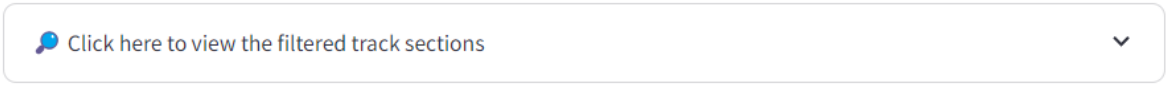

Figure 6: Expander Showing Table of Matching Tracks

Additional visualization options can be found under the 'Visualization Options' as shown in Figure 7.

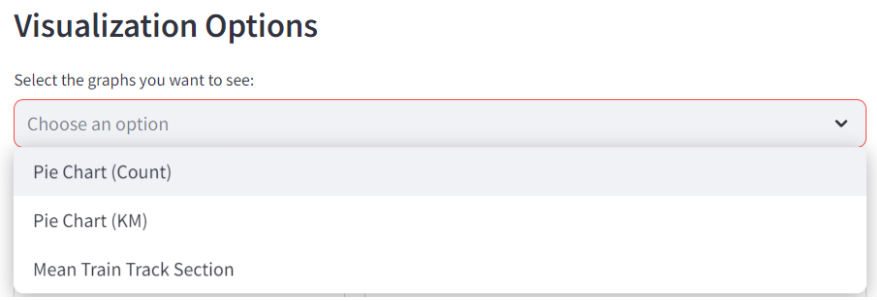

Figure 7: Visualization Options

1. **Pie Chart (Count):**
   Displays the percentage of train track sections that match the selected filters (see Figure 9).

2. **Pie Chart (KM):**
   Displays the kilometers of train track sections that match the selected filters (see Figure 10).

3. **Pie Chart (Track Length):**
   Displays the length of track that match the selected filters (see Figure 11).

   Based on the selection you make in Figure 8, the second pie chart is displayed either in kilometers of track, or in track length.

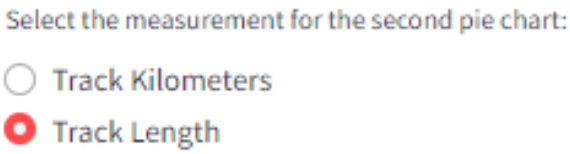

Figure 8: Select the measurement for the second pie chart

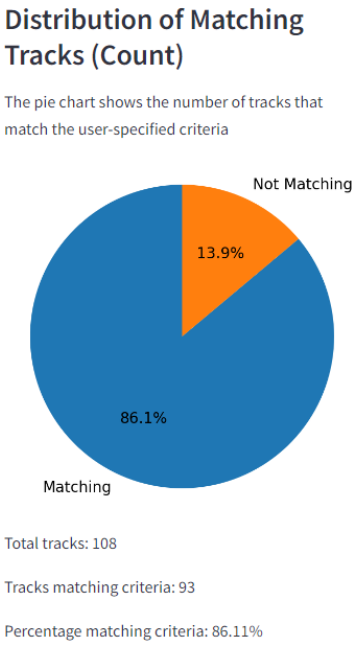

Figure 9: Pie Chart Showing Matching Tracks (Count)

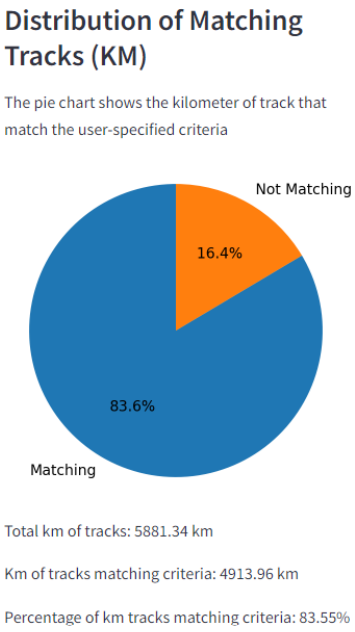

Figure 10: Pie Chart Showing Matching Tracks (KM)

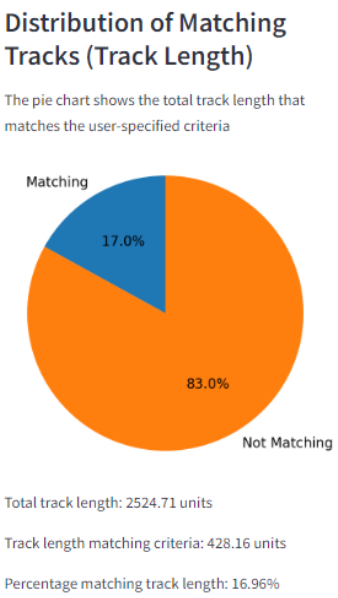

Figure 11: Pie Chart Showing Matching Tracks (Track Length)

4. **Mean Track Section:**
   Shows mean values for numerical characteristics and mode values for categorical characteristics based on user selections (see Figure 12 and Figure 13).

Numerical Columns

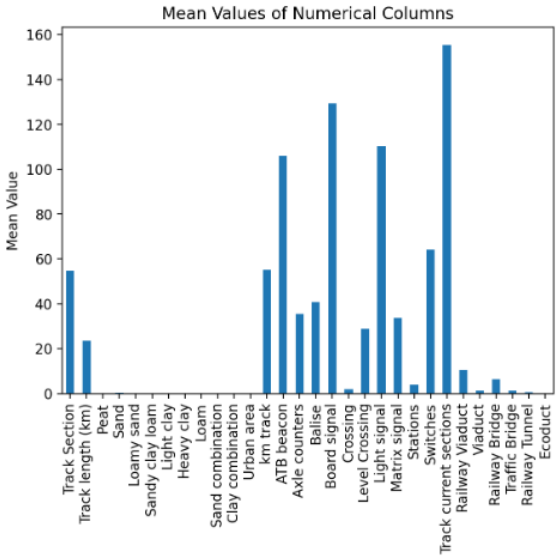

Figure 12: Numerical Mean Values

Categorical Columns

	0
Detection system	Circuit
ERTMS in 2031	No
Emplacement	No
Number of tracks	Two
Safety System	ATB-EG
Tranche 1 ERTMS	No
Travelers per day	10000-25000
Type of track	primary
Urban/Regional/Suburban	Regional

Figure 13: Mode Non-Numerical Features

A detailed explanation is available by clicking the expander (see Figure 14). For statistical insights, another expander (see Figure 15) shows the distribution of numerical values in Figure 16.

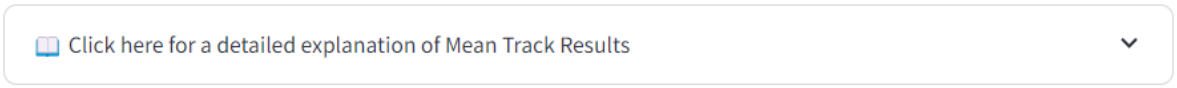

Figure 14: Expander for Detailed Explanation

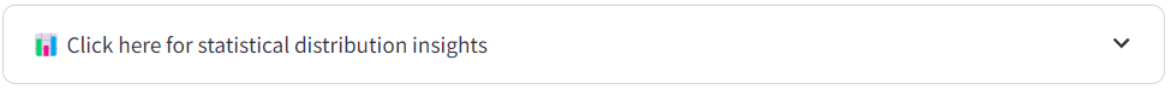

Figure 15: Expander for Distribution of Values

Histograms for Distribution of Numerical Features

Histograms provide a visual representation of the distribution of numerical features in the dataset. They show how the data points are spread across different values, which helps in understanding the underlying patterns and distributions of the data.

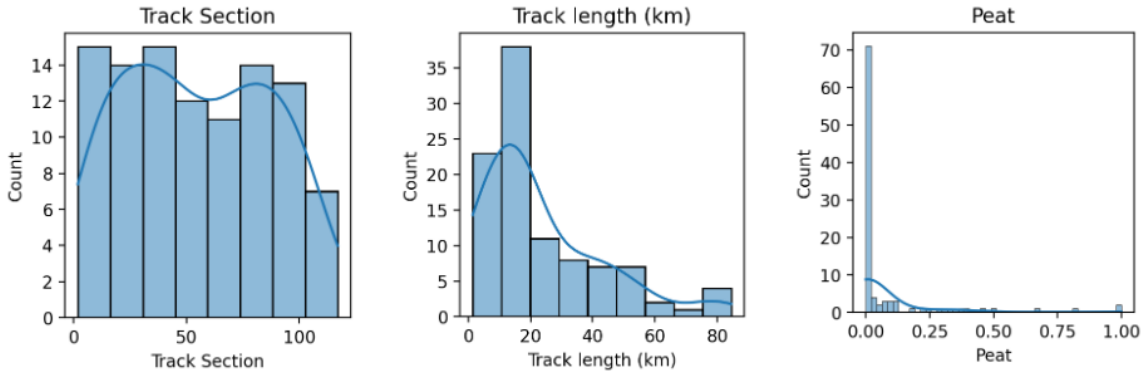

Figure 16: Distributions of Numerical Features

Additionally, you can download the data on the mean tracks to an Excel file using the button shown in Figure 17. Another feature in this part of the dashboard allows you to find the real tracks that match the average track section the most. A top 10 list is displayed when the button in Figure 18 is clicked.

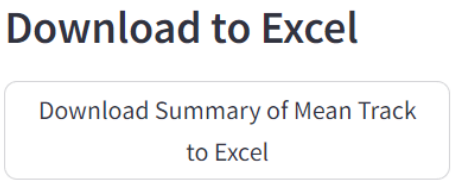

Figure 17: Button to Download Mean Track to Excel

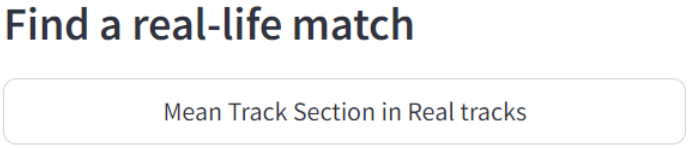

Figure 18: Button to Find a Track Match

Urban/Suburban/Regional Train Tracks

This section of the dashboard displays data results based on the categories of urban, suburban, and regional parts of the Netherlands. Users can choose from the following visualization options as shown in Figure 19:

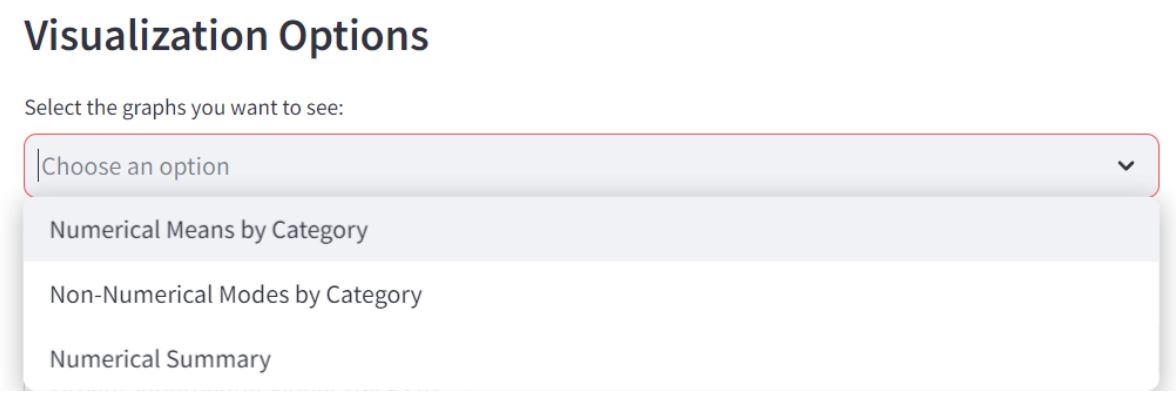

Figure 19: Visualization Options

1. **Numerical Means by Category**
   (see Figure 20): This shows the mean values with standard deviations for each characteristic by category. For more details, users can click the expander in Figure 21 to see the statistical distributions per category as shown in Figure 22.

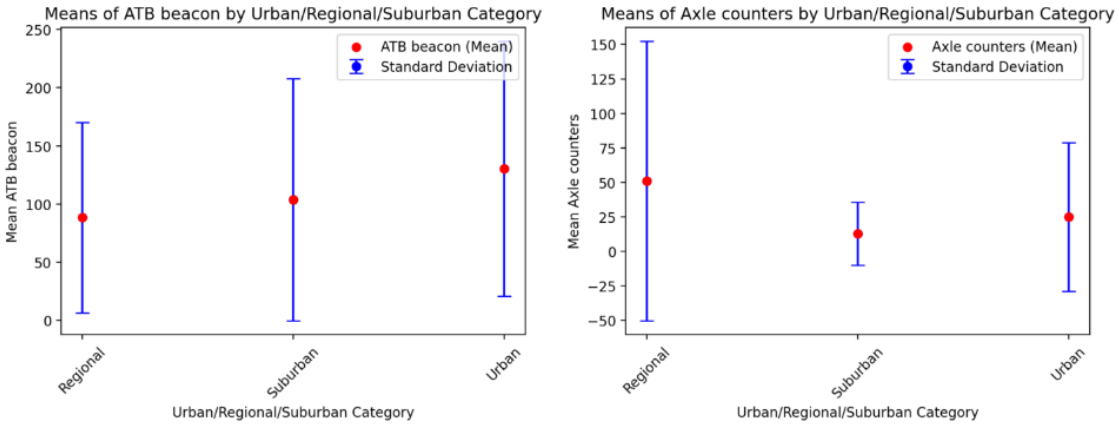

Figure 20: Examples of Means per Category

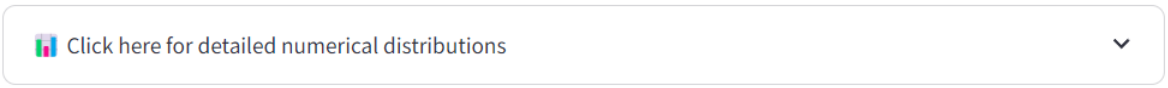

Figure 21: Expander for Details on Distributions

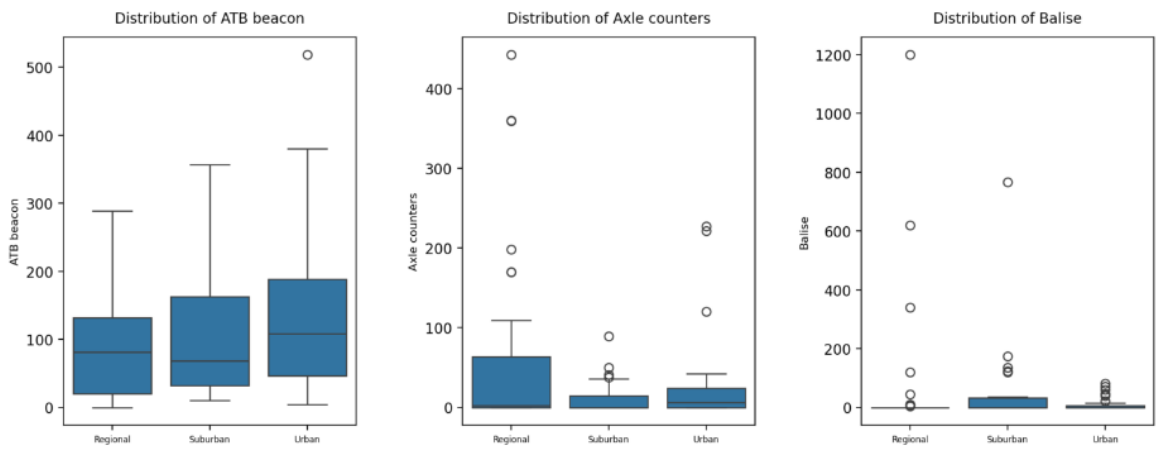

Figure 22: Distributions of Characteristics per Category

2. **Non-Numerical Feature Modes**
   (see Figure 23): This displays the mode of non-numerical features by category. For additional statistical details behind these numbers, users can click on the expander (see Figure 24) to view the distributions shown in Figure 25.

Non-Numerical Feature Modes

Below are the most common values (mode) for the non-numerical features across different track categories.

	Detection system	ERTMS in 2031	Emplacement	Number of tracks	Safety System	Tranche 1 ERTMS	Travelers per day	Type of track	Urban
Regional	Circuit	No	No	Two	ATB-EG	No	10000-25000	primary	Regior
Suburban	Circuit	No	No	Two	ATB-EG	No	10000-25000	primary	Suburl
Urban	Circuit	No	Yes	Two	ATB-EG	No	10000-25000	primary	Urban

Figure 23: Non-Numerical Feature Modes

Click here for detailed non-numerical feature distributions

Figure 24: Expander for Statistical Distributions

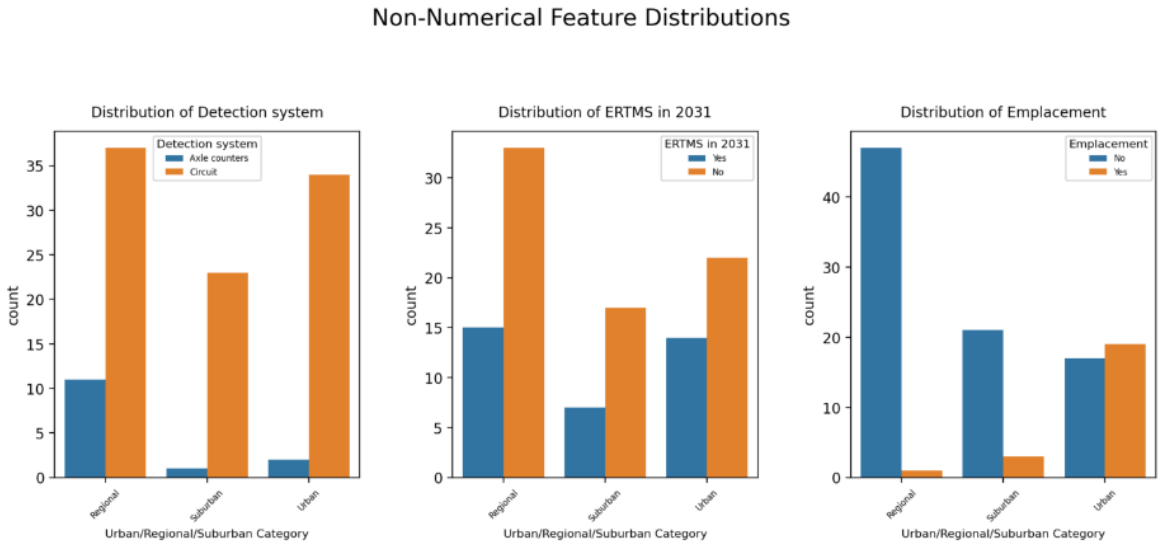

Figure 25: Statistical Distributions per Category for Non-Numerical Features

3. **Summary of Numerical Features by Category** (see Figure 26): This visualization option provides an overall summary of the numerical features of the track per category.

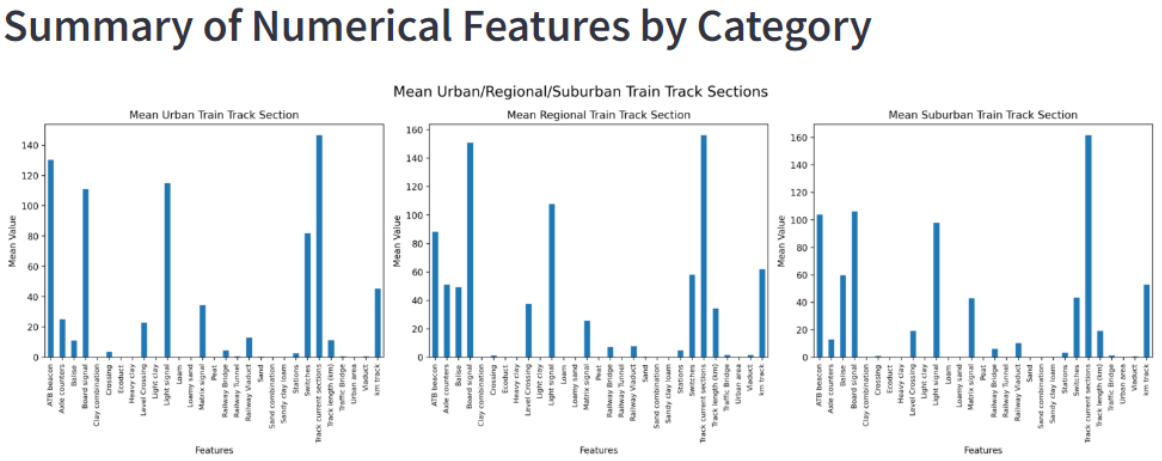

Figure 26: Summary of Numerical Features by Category

Similar to the Filtering Section, you can download the data on the mean tracks to an Excel file using the button shown in Figure 27. Besides this, this part of the dashboard also allows you to find the real tracks that match the average track section per category the most. A top 10 list per category is displayed when the button in Figure 28 is clicked.

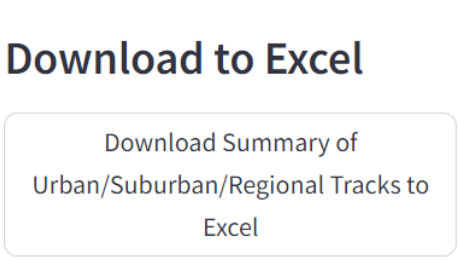

Figure 27: Button to Download Mean Track per Category to Excel

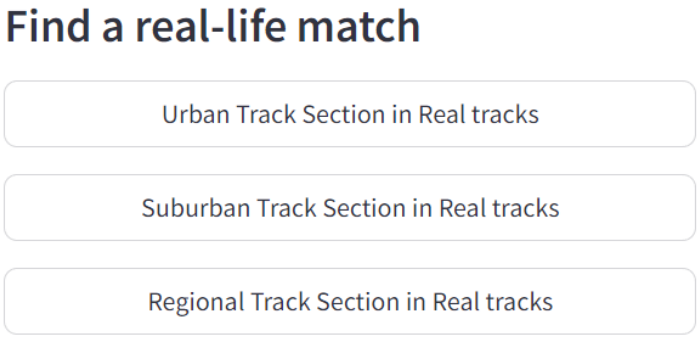

Figure 28: Button to Find a Track Match per Category

K-Means Clustering

This section presents the results of the k-means clustering algorithm. The clustering is based on the user's selections in the left menu. Before exploring the visualization options, let us provide a brief explanation of k-means clustering.

The k-means clustering algorithm groups similar data points into clusters. Each data point is assigned to the nearest cluster center, called a centroid. The algorithm minimizes the distance between the data points and their respective centroids, making the data points in the same cluster similar to each other. In this analysis, the section of the train tracks is grouped into five groups according to the elbow method. You can access more information by clicking the expander in Figure 29.

Figure 29: Expander for more information on k-means clustering

Simple Example:

Imagine that you have a bunch of different colored marbles and you want to group them into clusters based on their colors. Here's how k-means clustering works in this scenario:

Suppose that you have 5 empty bowls (representing 5 clusters). You randomly place one marble in each bowl to start with. For each marble, you look at its color and place it in the bowl that has the most similar colored marble. In this way, marbles of similar colors start to group together. Once all the marbles are placed in bowls, you look at the colors in each bowl and find the average color (this represents the centroid). You move the centroid (average color) to the center of its group. You go through the marbles again and reassign them to the nearest bowl based on the new centroids. Repeat this process until the marbles stay in the same bowls. By the end, each bowl will contain marbles of similar colors, representing a cluster of similar data points.

Users can choose from the following visualization options, as shown in Figure 30.

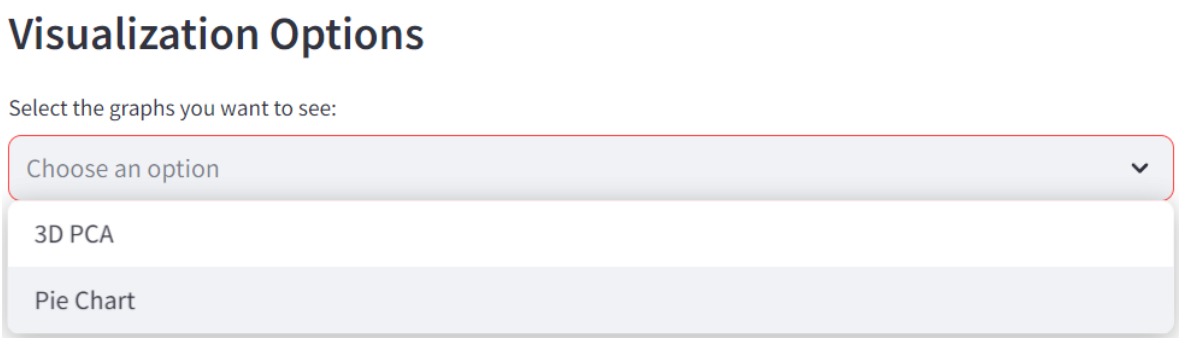

Figure 30: Visualization Options

1. **3D PCA Plot**
   (see Figure 31) This plot displays the clusters found in the data, where each point represents a section of the train track. For more details, users can click on the expander in Figure 33

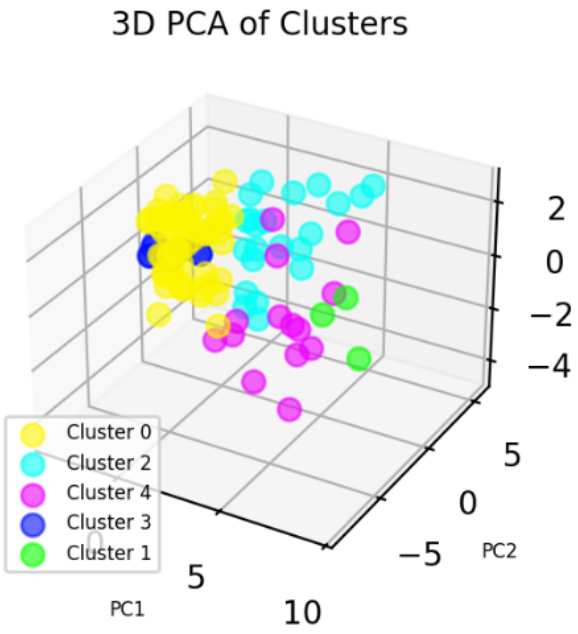

Figure 31: 3D PCA Plot

2. **Pie Chart**
   (see Figure 32). This chart shows the distribution of clusters, each slice of the pie represents a cluster. For more details, users can click on the expander in Figure 33

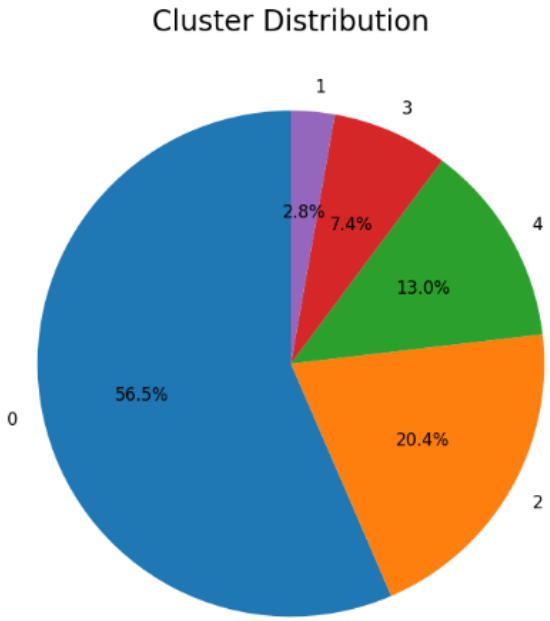

Figure 32: Pie Chart of Clusters

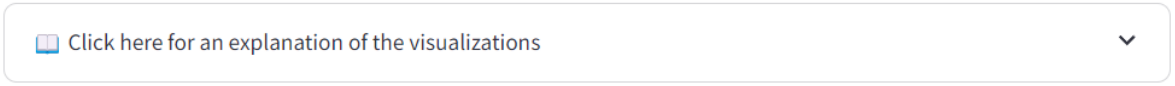

Figure 33: Expander for details on visualization options

Similar to the previous sections, you can download the data on the clustered tracks to an Excel file using the button in Figure 34. Another feature of this part of the dashboard allows you to find the real tracks that match the clusters found the most. A list of the top 10 per cluster is displayed when the button in Figure 35 is clicked.

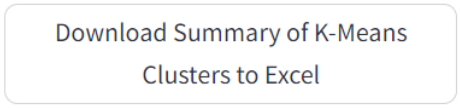

Figure 34: Button to Download Clustered Tracks per to Excel

Figure 35: Button to Find a Track Match per Cluster