

# Spotfire X

Working with F1 race data

## Table of Contents

<b><i>Introduction .....</i></b>	<b><i>3</i></b>
<b><i>Load data SBDF .....</i></b>	<b><i>4</i></b>
<b><i>Validate data .....</i></b>	<b><i>4</i></b>
<b><i>Start visualizing the data .....</i></b>	<b><i>5</i></b>
<b><i>K-Means Clustering.....</i></b>	<b><i>8</i></b>
<b><i>Streaming Data .....</i></b>	<b><i>12</i></b>
<b><i>F1 Circuits.....</i></b>	<b><i>17</i></b>

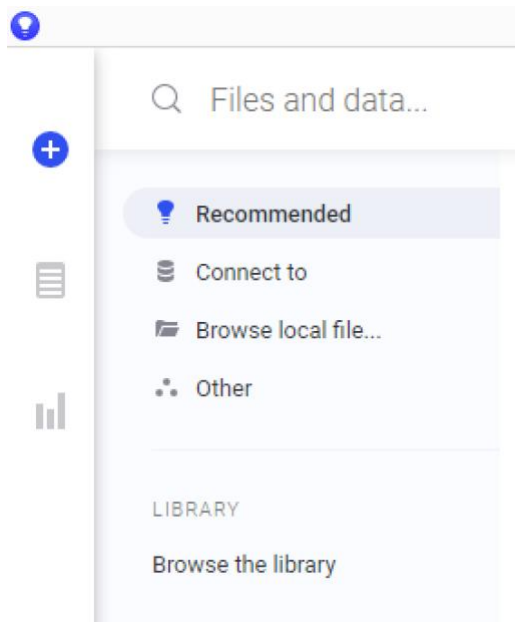
## Introduction

During this workshop we will work with some historical data from the F1. The data contains information about times, positions, distances as well as a lot of information about the car, such as current gear, pedal positions, temperatures, speed, etc.


During the workshop we'll try to figure out what circuit the data is from. We'll visualize the track and will look at clustering the data to create different clusters of data.


## Load data SBDF

Drag and drop the file “Historic Performance Data.sbdf” onto your Spotfire Analyst, or click on the + sign on the left top of the analyst client and choose Browse local file to locate you file.



## Validate data

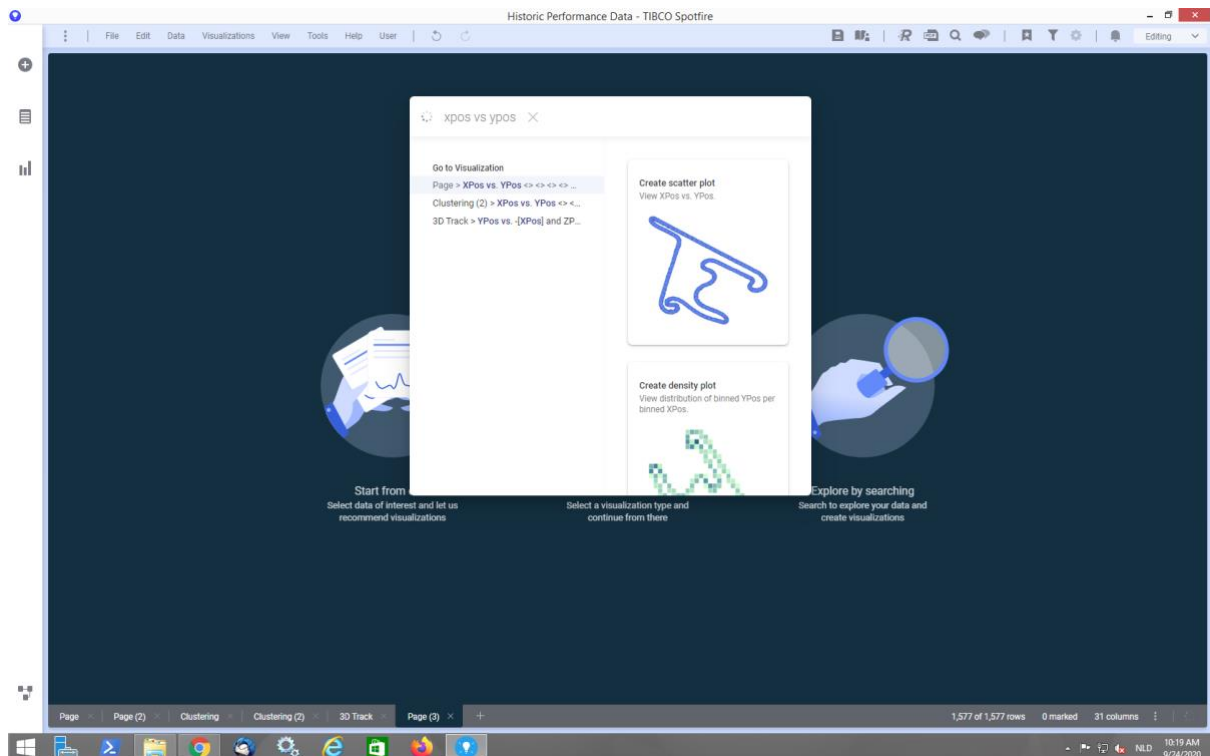
Let start looking at thee data by opening the Data Canvas, do this by clicking the  button on the bottom of the sidebar. This opens the data canvas where you can preview the data table, as well as see the number of rows and columns, and any transformation applied to the data.

It is good practice to look at the data in the data panel as well. Click on the second item from the top of the sidebar  in order to open the data panel. If you select the gear symbol it will provide all kinds of information about the data in order to perform univariate analysis. Is the number of values as expected (in case of categorical data) or is the minimum, maximum and distribution as expected (for continuous variables). The data panel also helps to clean and transform the data in case this is required. What to do with empty values? What is there are trailing spaces? The recommendation engine will assist you in spotting these kind of data issues.

## Start visualizing the data

Let's start by visualizing the data and specifically the circuit. As a picture says more than a thousand words, let's see if we can see from the data what track the data is about.

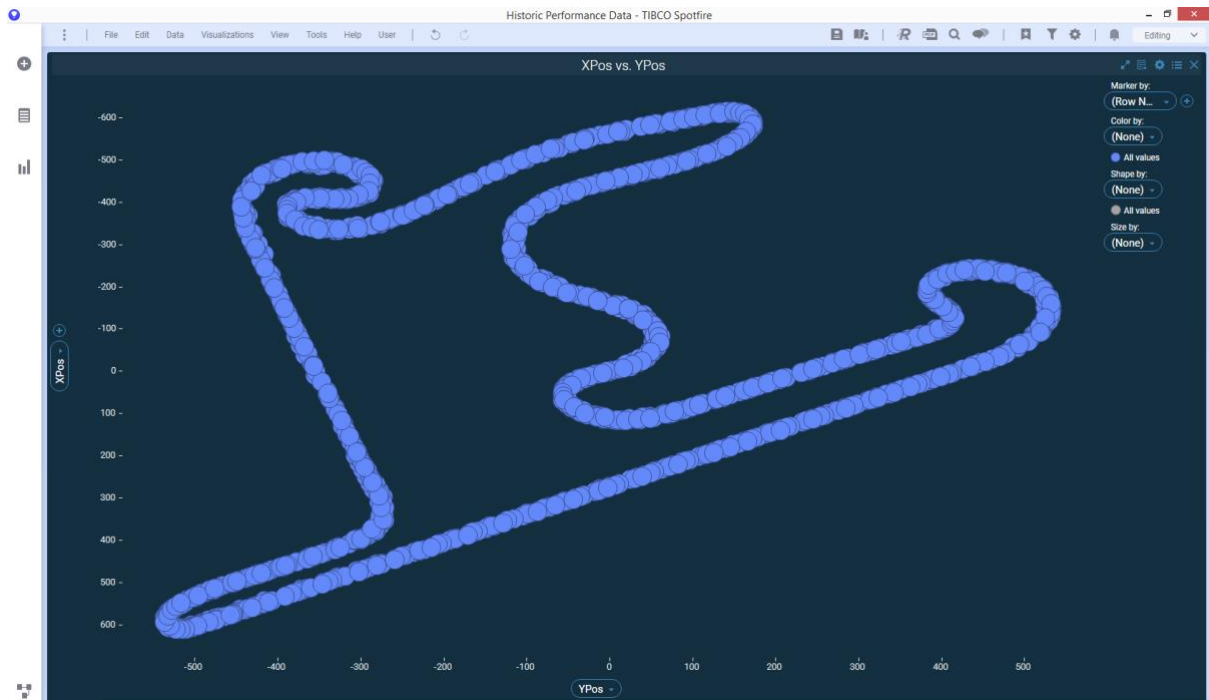
In order to do so, click the right icon on the start screen that says "Explore by searching" and type "xpos vs ypos" in the search bar. This will suggest a number of visualizations.




If you browse across the suggested visualization you get a number of suggestions on how to best visualize this type of data. For our purpose the first suggestion seems to make sense. Let's go ahead and suggest that visualization.

For the F1 fans you should be able to see the racing track that we are analyzing right? If not, take a look at the tracks on the last page of this document.

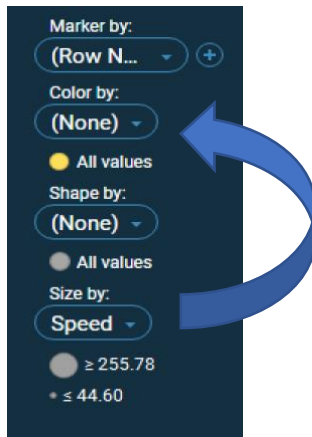
If you compare the track you see that the image is mirrored. In order to fix this, right click on the xpos axis and choose "Reverse Scale" now your visual should look like the one depicted below.



Let's add another visual to this page that shows both the altitude and speed over the distance of the lap. Instead of using the search let's now use the data panel. Click on the second item from the top of the sidebar  in order to open the data panel. Now select the fields we are interested in, i.e. zpos, speed and lapdistance, and wait for the recommendations to show up.

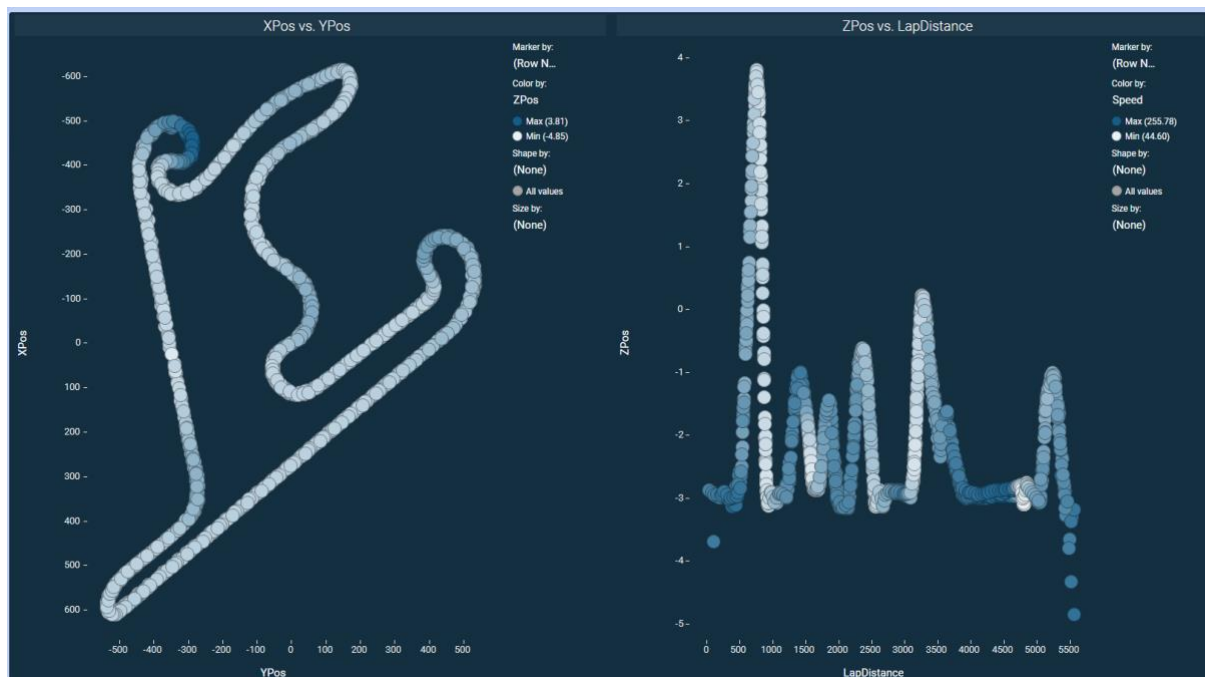
The screenshot shows the Tibco Spotfire interface. On the left is the 'Data in analysis...' sidebar with a search bar. Under 'Historic Performance Data', several fields are listed: CurrentGear, EngineSpeed, FuelLevel, Heading, Lap, LapDistance, LapTime, PitchAngle, RaceTime, RollAngle, Speed, TotalDistance, TyreTemperatureFrontLeft, TyreTemperatureFrontRight, TyreTemperatureRearLeft, TyreTemperatureRearRight, XPos, YPos, and ZPos. 'LapDistance', 'Speed', and 'ZPos' are highlighted with blue bars. Below these are 'CATEGORIES' including BrakePedalPosition, ClutchPedalPosition, GasPedalPosition, K-means Clustering, and Marked 9:58:48 PM. At the bottom of the sidebar is a 'Clear selections' button. On the right, there are four recommendation cards. The first card is titled 'View LapDistance vs. Speed' and shows a red line graph. The second card is titled 'View ZPos vs. LapDistance' and shows a yellow line graph. The third card is titled 'View ZPos vs. Speed' and shows a green scatter plot. The fourth card is titled 'View Speed vs. LapDistance' and shows a purple line graph.

The second recommendation seems to meet our needs. So let's drag and drop that onto the canvas in a positions where we would like the visual. Personally I prefer the speed to be presented in color rather than size. So I will drag and drop the speed from the size by to thee color by axis.




Now we have a good feeling of the shape of the track, the difference in height and the speed that the cars are going over the track distance. Obviously we can also choose to color the track based on z-pos or speed to get some more information in our track overview picture too.

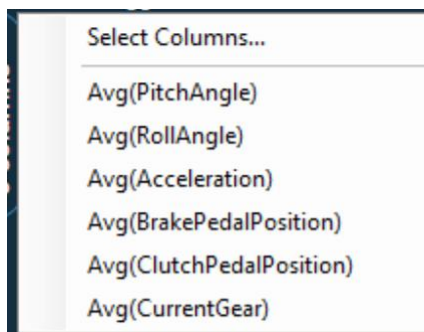
You should now have two visuals that look similar the thee ones below.



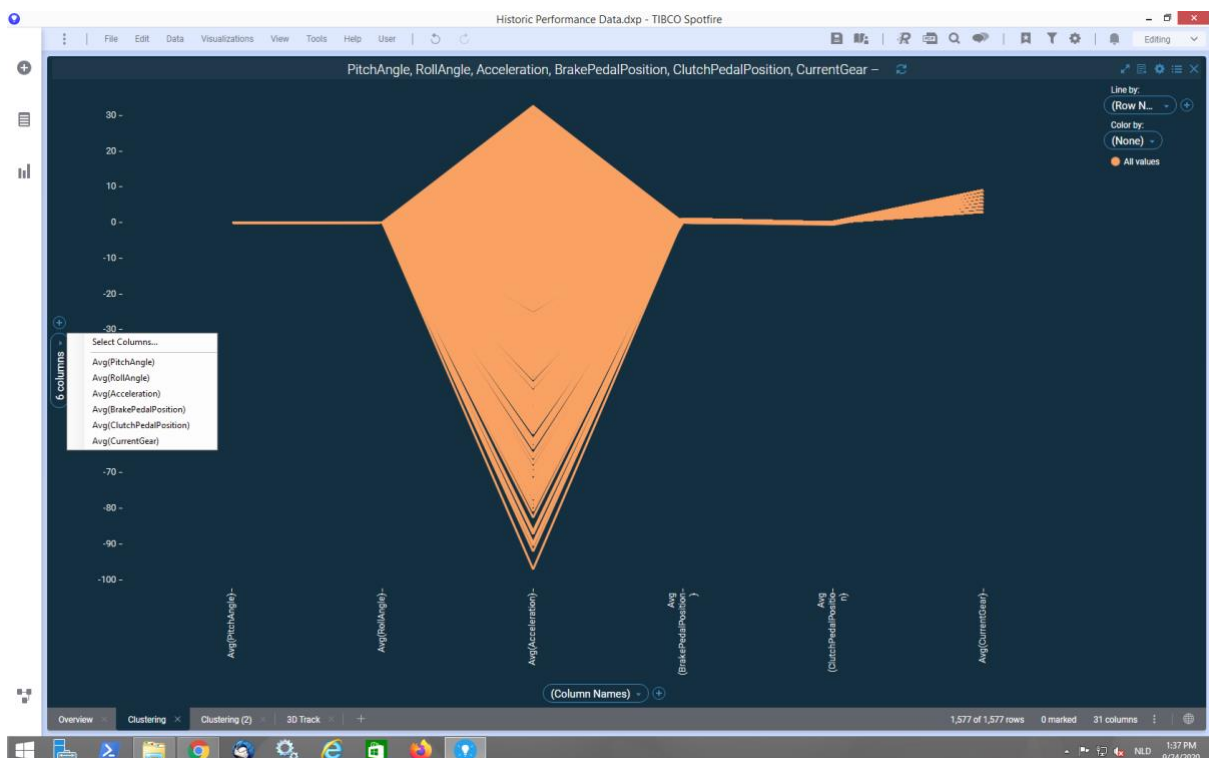
## K-Means Clustering

Let's now look a little closer at the data. Each line represents a moment in time, representing the position and conditions of the car. Let's see if we can find a way of grouping the similar lines together as clusters. Let's add a second page to the analysis file by clicking on the + sign next to the Page tab. Next we will create new line chart on this new page. This can be done by clicking on the bottom icon of the three icons on the top of the sidebar.  Then choose the line chart from the available visualization types.

Then on the line by axis choose row number, so that each row represents 1 line in the chart. On the Y-axis we are going to select the following.



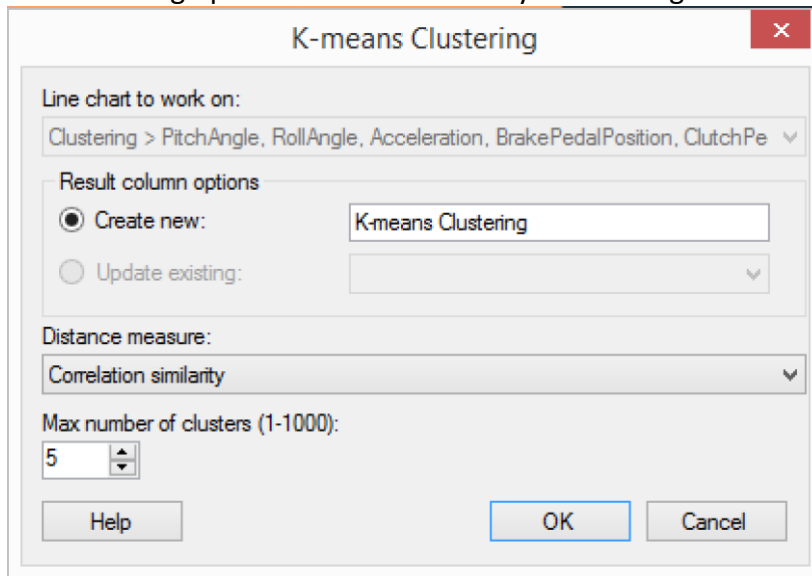
On the X-axis we will now choose "(Column Names)". Your visualization should now look similar to the below:





We will now create cluster of lines that show similar patterns. In order to do so, right-click on the visual and choose k-means clustering.

This will bring up a window that allows you to change some of the settings.

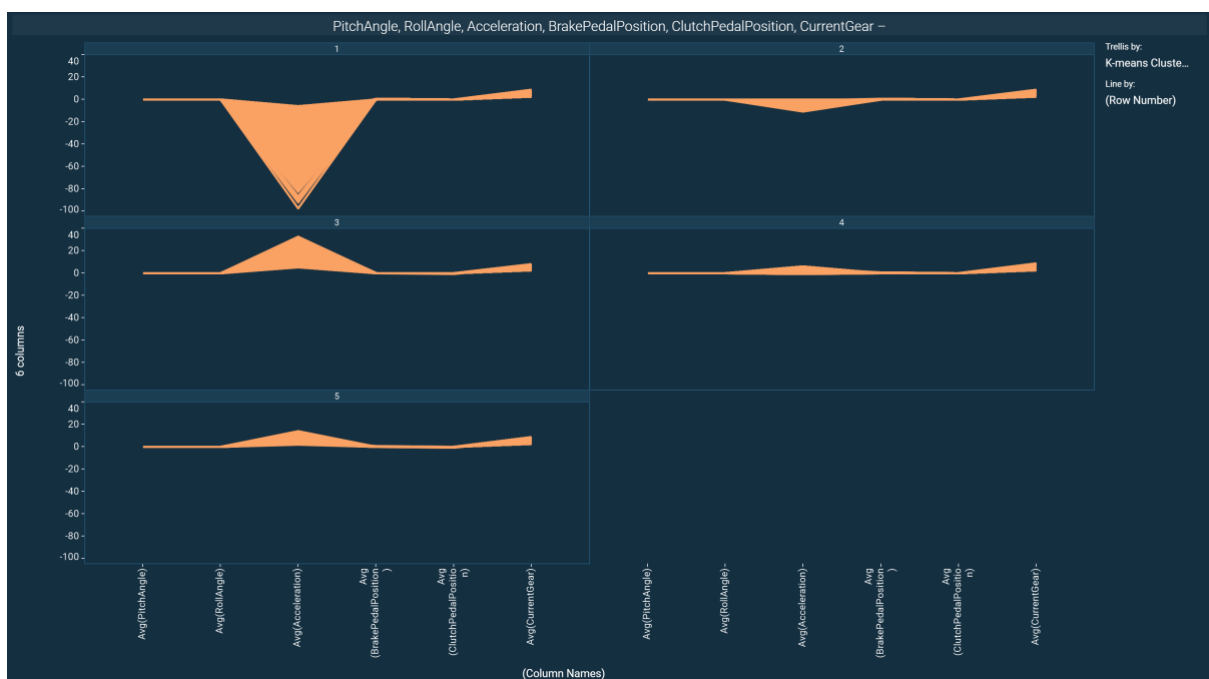


The image shows a 'K-means Clustering' dialog box with the following settings:

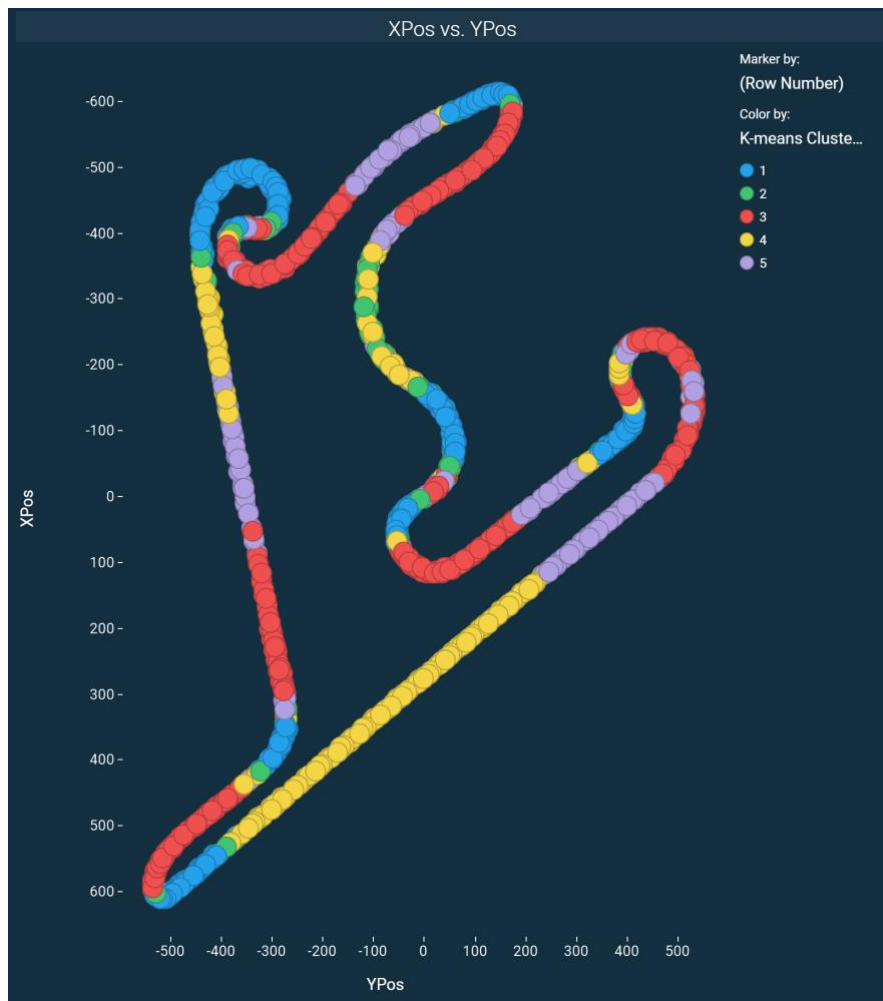
- Line chart to work on:** Clustering > PitchAngle, RollAngle, Acceleration, BrakePedalPosition, ClutchPe
- Result column options:**
  - ☒ Create new: K-means Clustering
  - ☐ Update existing:
- Distance measure:** Correlation similarity
- Max number of clusters (1-1000):** 5
- Buttons: Help, OK, Cancel

For now we will use Correlation similarity as the distance measure and a maximum number of clusters of 5. If you would like more background on k-means clustering, click on the help button for more details. Click ok to execute the clustering.

This will add an additional column to the dataset, indicating the cluster that the line of data falls into. Also, the visualization will update and be trellised by cluster as shown below.



Now let's add another version of the track visual we created on the first page by using copy visualization and then on the page with the line chart paste the visualization (CTRL V or via edit menu choose paste) now change the color by axis to the K-means cluster column.



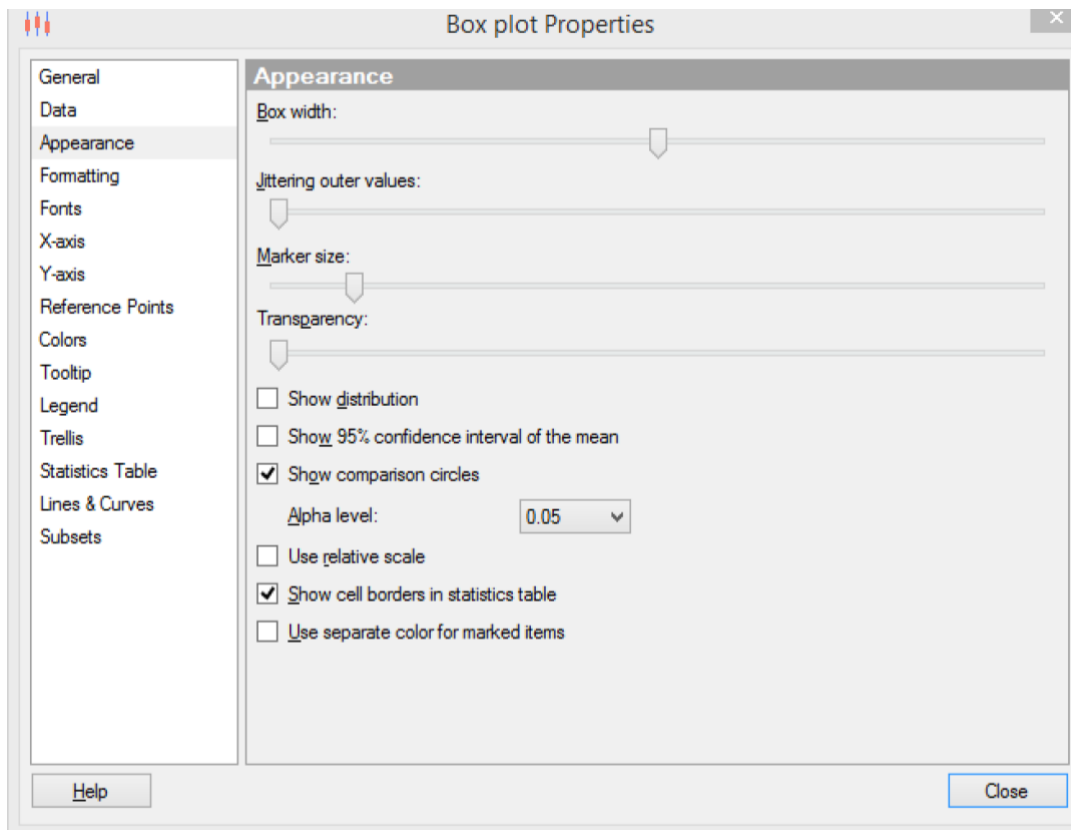
You will see that the data is now clustered into different categories, i.e.

- 1) Going into a bend
- 2) Intermediate areas
- 3) Coming out of a bend
- 4) High speed (straight(er) stretches
- 5) Accelerating after coming out of a bend

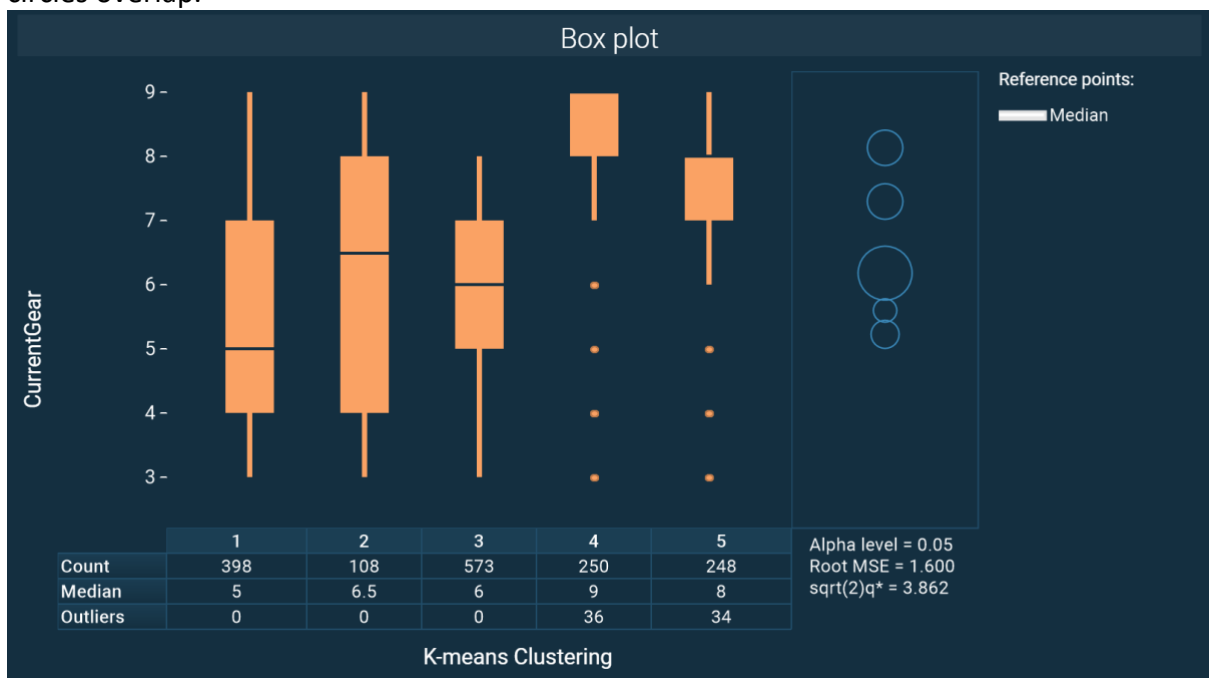
As the cluster is added as an additional column, we can now use this information in any part of the analysis. A nice feature to analyze the cluster is to use box plots.

Let's add a boxplot to the page by clicking on the add visualization icon in the sidebar and choose the boxplot.

Configure the x-axis to show the clusters and then set the y-axis to represent current gear. Finally go into the properties of the boxplot on the Appearance tab and select the checkbox in front of Show comparison circles.



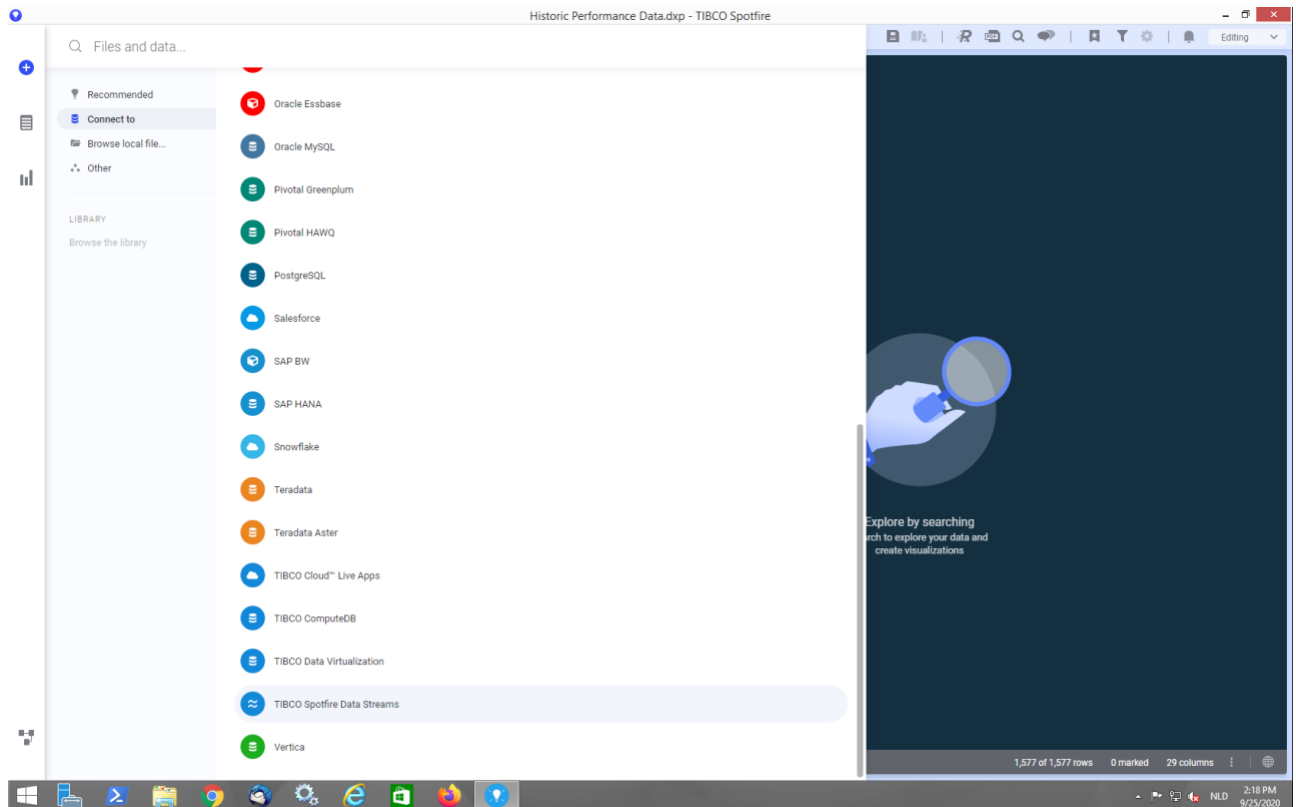
Your visualization should now look similar to the visualization in the image below. The comparison circles are useful to identify where the distribution between the different clusters is significantly different. When marking a circle, the visualization will indicate if the distribution is not significantly different by showing a line below the boxplot when the circles overlap.



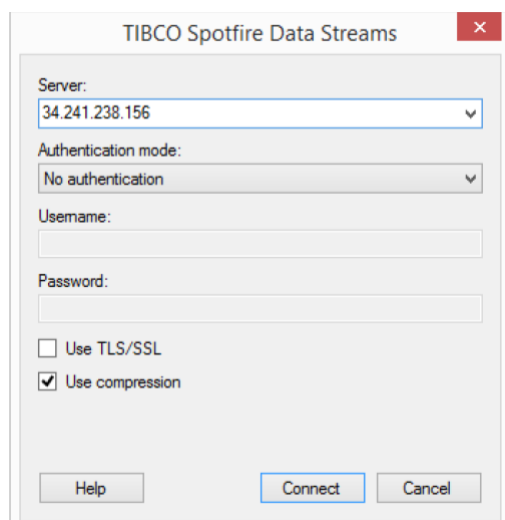
Investigate in the for different measures if this is the case to get a feeling of how the data is clustered.

## Streaming Data

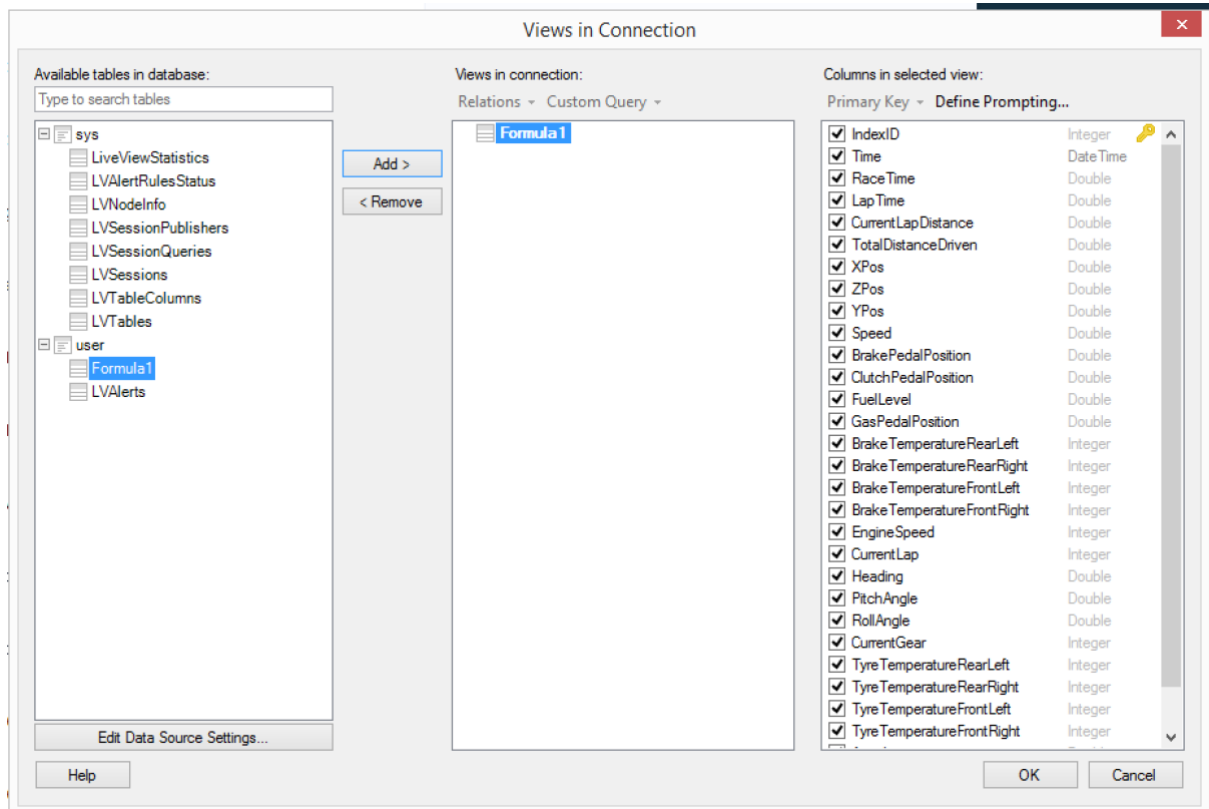
Now let's have a look at adding some steaming data to the dashboard. In order to do so we will add a connection to TIBCO Data Streams by clicking on the + sign on the right top of the sidebar and choosing "Add connection to" and then browsing down to TIBCO Spotfire Data Streams as shown on the image below



This will bring up a window where we need to add the server name or ip address. For today the server will be running at 34.241.238.156



Authentication mode can be left at none and then click on connect. You will see that connecting to a streaming data source is pretty much the same as connecting to any database. Add the table under user with the name Formula1 to the view as shown below.

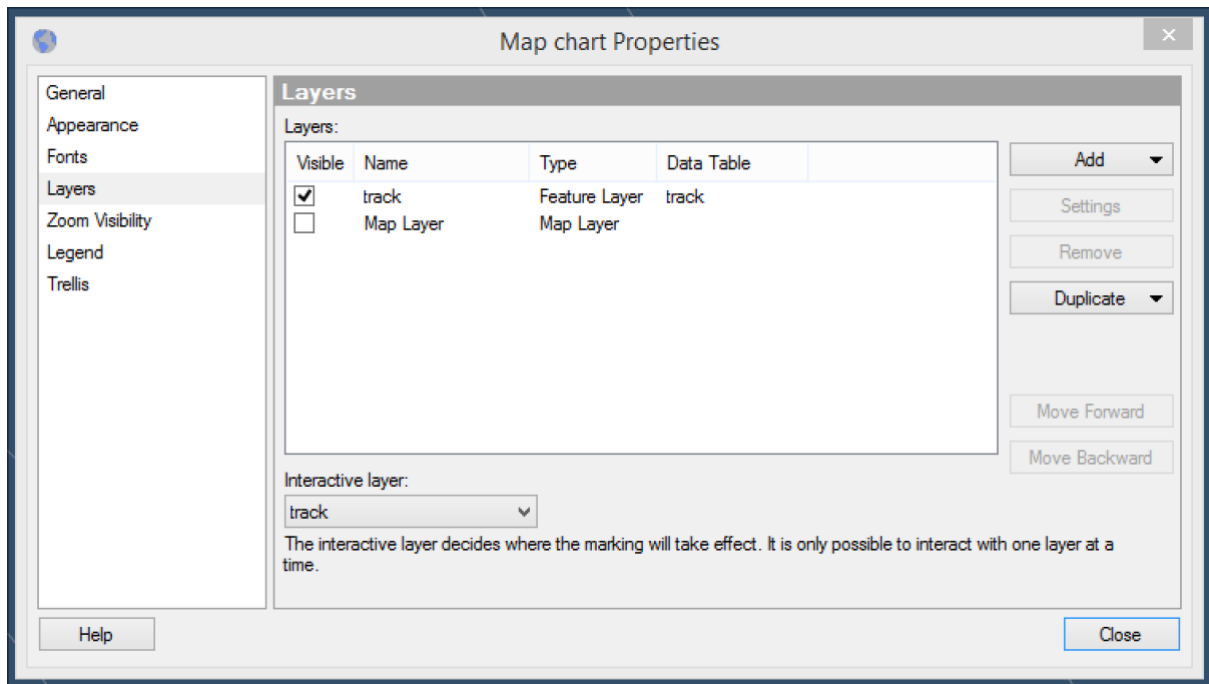


Let's also load some data about the track by clicking on the + sign and choosing browse local file. Let's browse to the file named " track.sbdf" and add it to the analysis.

Let's add another page to the analysis by clicking on the + sign on the tab next to the page(2) tab. On this new page add new mapchart

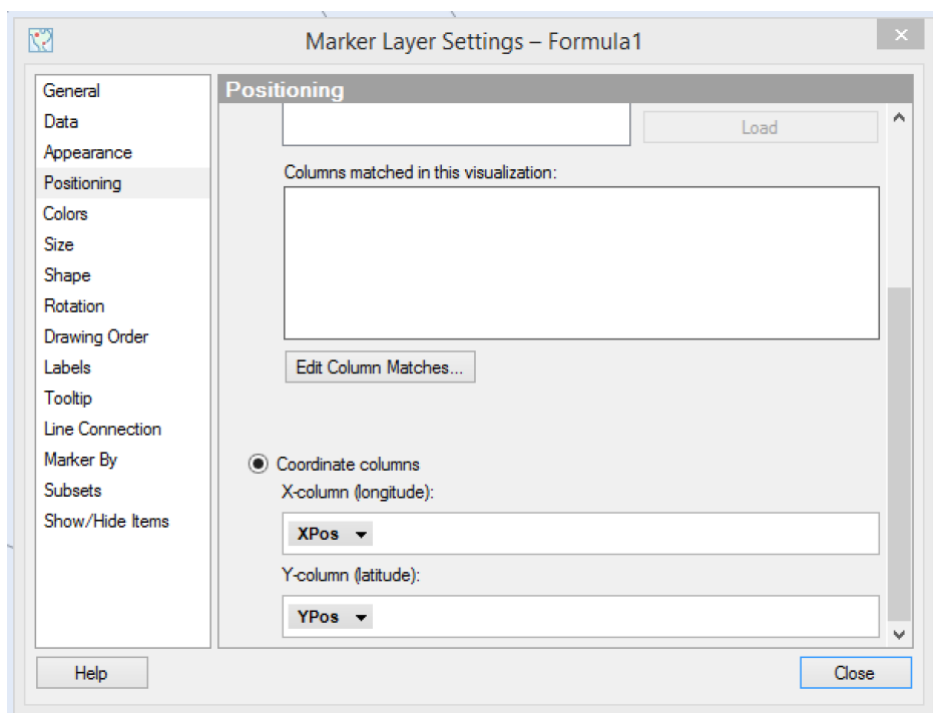
The map chart should show up with the track already visualized, because the track table is the only georeferenced data in the analysis.

Now let's go into the properties of the map chart. Under the Layers section, we can see that there is a map layer by default. In this case we will not require a map, so let's remove that layer by selecting it and clicking on the remove button.



Now we Add a layer for the streaming data to show up. This will be a marker layer.

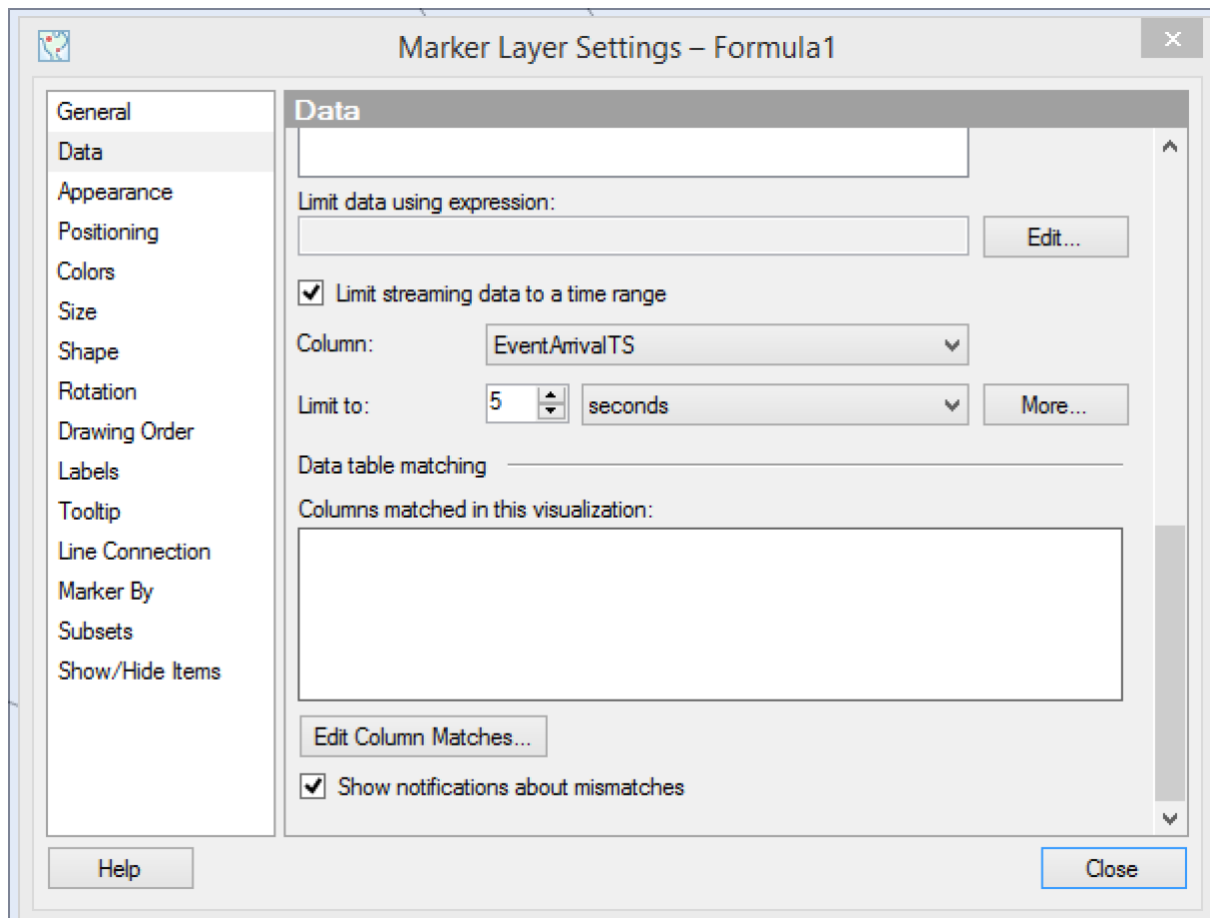
For the position we go to the Positioning area and scroll down to Coordinates columns. Let's add xpos as the X-Column (with aggregation set to none) and ypos with aggregation set to none for the Y-Column



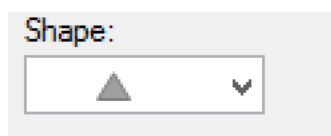
On the marker by tab choose IndexID.

As we don't want to flood the visuals with data we are going to limit the data that is displayed. On the Data Tab Scroll down a little to the area where it says limit streaming data

to a time range and choose the column EventArrivalTS as the column to limit the data on and type 5 in the limit to box and finally choose seconds as the measure.

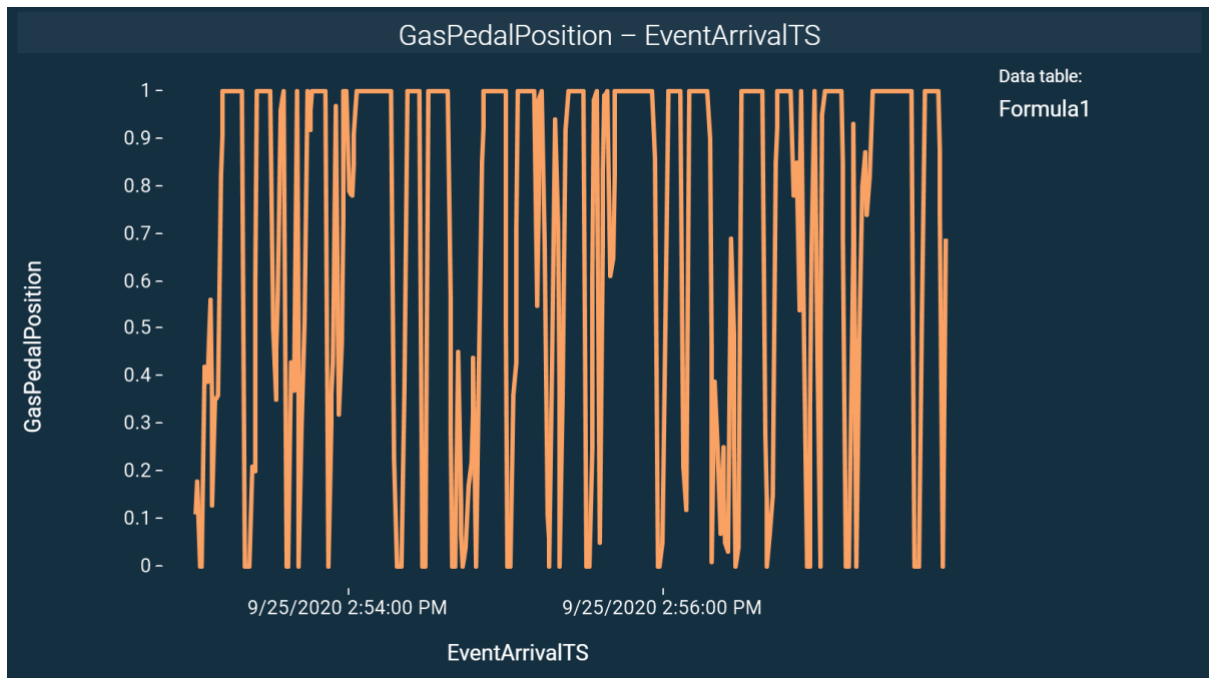


Now as a last step we would like to see the direction the car is moving. In order to do so, choose the triangle as the shape.



And go to the Rotation tab and choose the average as "Heading". Leave the direction as clockwise.

If you finished the streaming map chart, let's try and add one or two more visuals to the page. In this case let's add a line chart that used the Formula1 streaming data table and put EventArrivalTS on the X-axis and GspedalPosition on the Yaxis. Bar in mind that you mind want to limit the data that is shown in the visual.



A last visualization to add is another line chart, that shows the EventArrival on the X-Axis and the Avg Speed on the Y-Axis.



There you go. You have finished your first streaming dashboard in Spotfire!



## F1 Circuits

