Good morning, everyone!

I will present to you the use case that Eduardo and I have built for the Data Science Engine.

So, in order to sustain the fierce com**pe**titiveness of the market, companies are motivated by the search for more optimized and economical processes. In this context, the maintenance sector becomes a key component, as all maintenance processes tend to be costly and time-consuming. Therefore, efficient management of all resources involved in maintenance ensures time and financial savings, as well as better utilization of machines.

The problem that I bring to this case is that not all data obtained during these maintenance processes are used in planning calculations, either because these data are still collected on paper, or because extracting useful information from these data is not obvious. More specifically, in our case, we will focus on maintenance of wagon equipment.

Railway transportation is a fundamental component of the transportation infrastructure in many countries, bringing efficiency, sustainability, and connectivity to the economy and society. Its highlight is the transport of heavy loads over long distances, providing a gainful alternative compared to other means of transportation, while also having a low environmental impact.

The wagons, also known as railroad cars, are vehicles specially designed for **trans**porting loads on railways. They are essential for the efficient railway operation and have a variety of designs to meet different cargo requirements. Wagons are usually composed of a metal box supported by wheelsets., which are these structures. The wheelsets allow the wagons to roll smoothly along the tracks and play a crucial role in the stability of the wagons.

Thus, proper maintenance of wagons and wheelsets is essential to ensure the safety and efficiency of railway operations. There are three **ca**tegories of maintenance: corrective, which consists of performing repairs as soon as the equipment fails, being less efficient and resulting in an undesirable production stoppage period; preventive, which is performed at fixed periods, but may be performed unnecessarily or not prevent the failure; and predictive, based on (AI) and/or Machine Learning, is capable of predicting the occurrence of a failure based on the machine history. This method will be the topic addressed in this case.

Thus, our system will be structured as follows:

Maintenance tests will be conducted in laboratories. The data generated from the results of these tests will build the maintenance history. Based on this historical data, the algorithm is capable of predicting the remaining useful life of our wheelsets. This will enable more efficient and accurate decision-making regarding the necessary actions to avoid failures and maximize production.

The choice of our prediction algorithm is based on the work of Mathew et al. (2017). In their work, the remaining useful life of an engine is estimated. The authors compared the effectiveness of different Machine Learning methods. From the results obtained, it is possible to see that Random Forest presented the smallest error among the other methods.

For our implementation, we used a data repository provided by (NASA), including sensor measurements from a Turbofan engine in a degraded state. These data were divided into sets for training, testing, and verifying the accuracy of the prediction system.

Each row of the database represents a different cycle over time, which we can int**e**rpret as different maintenance interventions. And each wheelset has a unique ID. Therefore, the data is presented for a complete life cycle for each of the bogies. For example, for wheelset with ID 1, we have data for 192 cycles from the beginning of its life until its failure. So for each line we can calculate the remaining useful life of that wheelset.

Each of the analyzed engines has 21 sensors collecting different types of data related to the engine. But for implementation purposes, not all data were used. To determine which sensor information would be included in the project and which would be excluded, it was necessary to analyze the relationship between the data and the target variable, as well as the relevance of each feature in constructing the prediction model.

In "Register Intervention," a project is created for a combination of wagon and maintenance type. All configurations have been carried out in order to avoid duplicate data

In the "Wagon Info" tab, you define the wagon number that will undergo maintenance, as well as the type of intervention. After saving the project, the information about the wheelset serials for the selected wagon is displayed in the "Wheel Data" tab.

Additionally, during the saving process, four samples are logged in the project, one for each wheelset position. The PRODUCT field is defined similarly for all samples, using the value WAGONWHEELS, which contains all the maintenance tests to be performed. Furthermore, the SERIAL and CYCLE fields are defined for each sample based on the wheelset position and the maintenance history of each serial, respectively.

### VW Wheels Laboratory Home

#### View Samples to Enter Results

Folder with Incomplete and In Progress samples for entering results. In the VW section of the Folder, we can view some information about the current prediction model and update it if needed.

#### View Completed Samples

Samples with completed maintenance and all results inputted.

### VW RUL Model

Visual Workflow with metrics information about the prediction models.

#### Bottom Refresh the Model

Runs the script for generating the prediction models.

#### Bottom Visualize Decision Tree

Opens file with visualization of the decision tree generated by the random forest algorithm.

#### Confusion Matrix

A confusion matrix is used to evaluate the performance of a classification model. It allows the understanding of the performance of a classification model by measuring the number of correct and incorrect predictions for each class.

#### Metrics

Accuracy: The overall accuracy of the model, calculated as (TP + TN) / (TP + TN + FP + FN).

Precision: The proportion of correctly predicted positive instances out of all instances predicted as positive, calculated as TP / (TP + FP).

#### Feature Importances

Helps in understanding which features have the most significant impact on the model's performance.

#### Remaining Useful Life Actual vs. Predicted

#### Evolution of the most important features along with the evolution with RUL