**Analysis of Railroad Fuel Consumption**

*Carlos Sanchez Valle*

**1. Background and Problem to Explore**

The fuel efficiency in any transportation company is extremely important. Nevertheless, it is critical to rail freight transportation because fuel and power costs follow labor as the second largest contributor of their cost structure and source of comparative advantage. It is contended that “Railways cost structure are at their most highly competitive when railways can operate large trains, well-loaded with traffic, over a heavily used network …”

<https://ppiaf.org/sites/ppiaf.org/files/documents/toolkits/railways_toolkit/ch2_1_2.html>

We seek to corroborate or disprove this statement with regards to fuel consumption. As with any vehicles fuel consumption, the bigger the engine power the higher the fuel consumption. Other reasons can be attributed to higher fuel consumption – Type of Service, Type of Locomotive, Direction, Sector, Track grade or slope, etc. Analysis will show us to what extent these attributes affect fuel consumption. It could also reveal relationships between independent variables and fuel consumption that are not evident from the raw data.

In this project, we will focus only train consumption since shunting locomotive data does not provide with consistent distance, time or weight data. The exploration will include:

1. Summary of consumption by Train Type/SubType, Locomotive Model, Track and Direction.
2. I will show the consumption of locomotives of the same model.
3. What factors affect consumption the most.
4. To compare different types of locomotives I will normalize the consumption by comparing liters per GTK (Weight-Distance)
5. Does time change the consumption due to better equipment (Locomotive Model) or better practices (Engineers, Firemen).
6. Have the factors that affect consumption changed since 2010? Could data help us determine the magnitude of the change.
7. The final product would be a predictive model using different machine learning techniques and compare their results.

.

**2. Potential clients:**

The main client for this model and insights that result would be Operation Planning Managers of a railroad to aid in cost effective locomotive assignment on trains and yards based on demand.

There could also be use for the model in long term planning to determine what models of fuel efficient locomotives to acquire.

**3. Data and Approach:**

The raw datasets are from the Operations System (WebRail) of Ferroviaria Oriental S.A. and are provided as stored procedures in SQL that are accessed via Excel sheets. I will convert these into csv files, that will be imported as pandas dataframes. The data wrangling, clean up, consolidation will be done in python using pandas and numpy.

To create a complete data story, we will need to use visualizations such as scatter plots, violin plots, and histograms using seaborn. For the predictive component, I will be using appropriate machine learning methods from scikit-learn and compare their results.

**4. Deliverables:**

My deliverables will include:

• Model to predict fuel consumption given locomotive, freight weight, track section (gradient/slope), equipment type, crew, distance, time, direction.

• Identify fuel efficient of each type of locomotive.

• Proposal of method to update equipment assignment to include the model result findings as new data is incorporated.

• Flag the least fuel-consuming assignment decisions to alert cost differences.

All of these summarized in: 1) exposition of the project in the form of a paper, 2) code outlining the steps housed in a github repository, and 3) slide deck to present the most noteworthy findings of the project.