# WHY DATA EXPLORATION FOR MACHINE LEARNING:

* Goal: Understanf the content and relationships in the data set
* Make best use of information in dataset
* Avoid problems when training and deploying machine learning models

**MACHINE LEARNING WITHOUT UNDERSTANDING DATA IS A WASTE OF TIME!**

# WHAT ARE YOU LOOKING FOR?

* Characteristics of the data / size, data types, etc.
* Summary statistics of the dataset
* Errors and outliers (valores atipicos)
* Distribution of features and label (Si las etiquetas son numéricas, categóricas, etc, si las etiquetas son numéricas nos fijamos en la distribución, si son categóricas nos fijamos en la frecuencia)
* Relationships between features and label

# DATA EXPLORATION IS AN ITERATIVE PROCESS

* Expect to repeat steps
  + Example, once errors are correct, need to visualize again
  + Before and after data preparation
  + Before and after constructing machine learning model.
  + Need to repeat often!
* Try lots of ideas – expect some to work

**THIS IS AN EXPLORATION PROCESS!**