# Exercise 7 - Data Manipulation with Pandas

Load in the dataset renfe\_trains\_cleaned.csv

### Pivot Tables

Use a pivot table to explore how price differs with respect to the type of fare for each destination

### Working with Time Series

Convert departure & arrival to a more appropriate datatype

Calculate the duration of each train journey and add it as a column called duration

Make departure the index of the DataFrame

Sort the index low to high (earlier to later). This will make slicing possible later.

Select all journeys which departedon 07/05/19

Select all journeys which departed on 07/05/19 to 11/05/19

Add one year to each date in the index of the DataFrame (but do not save it!)

Create a subset of the DataFrame called madrid\_to\_barca which contains only journeys with origin as MADRID and destination as BARCELONA

Select only those tickets in madrid\_to\_barca which are in the Promo category for fare and Turista for vehicle\_class and update madrid\_to\_barca to only contain those

Compute a 7 day rolling average for price for the madrid\_to\_barca DataFrame. Add it as a column called rolling.

Plot the rolling average vs. the actual values of price

### Combining Tables

Read in the fare\_conditions.csv file. It contains the conditions for the type of ticket that has been purchased.

Add the fare conditions to the original df DataFrame