

Course Name:- Data Cleaning Project Walkthrough

Mission Name:- Data cleaning Walkthrough

* Two ways to go about finding a good topic for project:-

i) Think about what sectors or angles you are really interested in, then find datasets relating to those sectors.

ii) Review several datasets and find one that seems interesting enough to explore.

* When we handle multiple datasets:-

i) Read each of them in a pandas dataframe.

ii) Store all the DataFrames in a dictionary.

* When any dataset is ^{encountered} ~~encountered~~, first we should check for null values and duplicate values.

* $z = \text{pd.concat}([x, y], \text{axis}=0)$

The code above will combine dataframes x and y by especially appending y to the end of x . The combined dataframe z will have the number of rows in x plus the no. of rows in y .

* It's important to pass the keyword argument `errors="coerce"` when we call `pandas.to_numeric()`, so that pandas treats any invalid strings it can't convert to numbers as missing values instead.