

You've been asked to clean the attached dataset (billings_europe.csv), which contains billings data in

Euros grouped by various categories:

- Row 5: Segment and Period (separated by a " – ")
- Row 6: Type
- Row 7: Subtype

For the purposes of this exercise, assume the data in the spreadsheet are correct (e.g., that missing columns or rows are supposed to be missing, and can be excluded).

Exercise

Using Python only (do not do any cleaning steps in Excel), produce the following:

- The sum of billings by segment
- The sum of billings by "Period" (see row 5, above) for which the type is "Market" for 2016 to present (including January 1, 2016)
- Summary statistics by Segment
- Find the unique and there count for subtype
- Find out there are outliers in values columns
- Count nan values in values columns and fill it with appropriate values
- Apply descriptive stats on values column and describe the output

Suggestions:

- Clean the dataset so that each of the headers in rows 5-7 is listed in a separate column next to date.

The information in row 5 should be split into two columns to create Segment and Period.

o If the Type is missing, take the last non-missing value to the left.

o See the example output below:

Output

Please return the following materials:

- The attached output spreadsheet ("Python Exercise Output") with:
- The code you used to clean the dataset.
- The code for exercise

segment	period	subtype	values	date
Austria	On Trailing EY	AT	10148.6	31-Dec-74
Austria	On Forward EY	AT	NaN	31-Dec-74
Austria	On Trailing DY	AT	9333.4	31-Dec-74
Austria	On Forward DY	AT	NaN	31-Dec-74
Austria	On Trailing BVY	AT	7724.2	31-Dec-74
...
Austria	On Trailing EY	AT	5880.4	1-Sep-17
Austria	On Forward EY	AT	2128.6	1-Sep-17
Austria	On Trailing DY	AT	7084.0	1-Sep-17
Austria	On Forward DY	AT	6332.8	1-Sep-17
Austria	On Trailing BVY	AT	10598.4	1-Sep-17