Morgan Stanley
Python First Round Exercise

Background

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 country
- 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

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.
  - If the Type is missing, take the last non-missing value to the left.
  - See the example output below:

Date	Segment	Period	Type	Subtype	Value
1974-12-31	Austria	On Trailing BVY	Countries	AT	7724.2
1974-12-31	Austria	On Trailing DY	Countries	AT	9333.4
1974-12-31	Austria	On Trailing EY	Countries	AT	10148.6

Output

Please return the following materials:

- The attached output spreadsheet ("Python Exercise Output") with:
  - The sum of billings by country and sum of billings by period in the "Output" tab
  - Sector summary stats in the "Sector Summary Stats" tab
- The code you used to clean the dataset.