MMAI 5400 Assignment 3 -- Aspect-Based Sentiment Analysis

In this assignment you will work with the same dataset as in assignment 2, that is, reviews from Trustpilot.com. However, instead of simple Sentiment Analysis (SA) you will do Aspect-Based Sentiment Analysis (ABSA). You will use PyABSA to run ABSA on the `reviews.csv` data.

Submission

The assignment should be submitted as Python 3 code and uploaded to Canvas as a single `PY` file (**not** a Jupyter Notebook) and the trained model. The due date is on August 1 at 8:30 am.

The code will be tested and should produce the output specified below.

Data

Use the file `reviews.csv`, containing 1920 reviews. To read `reviews.csv` with `pandas` it might be good to specify tab `\t` as the delimiter, as follows:

```
data = pd.read_csv('reviews.csv', delimiter='\t')
```

or

```
data = pd.read_csv('reviews.csv', sep='\t')
```

The two are equivalent.

Task

Your task is to run ABSA of the reviews. You will use a pre-trained BERT-based model (use `checkpoint='multilingual'`). For each **sentence** you need to return the aspect tokens and their corresponding sentiment.

The resulting table should look as follows:

Table 1

	Aspect	Sentiment	review_ID	Name
1	`staff`	`Neutral`	`0`	`La Marquesita`
2	`food`	`Positive`	`0`	`La Marquesita`
3	`atmosphere`	`Positive`	`1`	`La Marquesita`
...	...	...	...	...

Your code will be tested on a set of unseen reviews and should produce a table that is consistent with that produced by my reference code.

Deliverable

You need to submit a single Python file (`PY` **NOT** `IPYNB`) with a method called `absa()`. The method takes as argument a list of sentences (strings), performs ABSA and returns a dataframe with the format specified in **Table 1**.

Your code should follow the PEP 8 style guide. See the original PEP 8 style guide, an easier to read version, or a short PEP 8 YouTube intro. Practically, adding a PEP 8 plugin to your text editor (e.g. Falke8) will make it easier to follow to style guide.

Grading

For grading the model will be evaluated on unseen data. For full marks, the code has to be bug-free, PEP8 formatted and the returned dataframe has to be at least 80% consistent with the reference code's output.

Good luck!