Quiz 1 **Unlimited Attempts Allowed**

Available: Oct 14, 2024 5:40pm until Oct 14, 2024 11:30pmAvailable: Oct 14, 2024 5:40pm until Oct 14, 2024 11:30pm

**Details**

1). 5 pt

https://github.com/jackiekazil/data-wrangling/blob/master/data/chp3/data-text.csv

[Links to an external site.](https://snap.stanford.edu/data/web-FineFoods.html)Read this data as CSV and print the first 1000 samples.

* Use csv reader and pandas read\_csv - both.
* Prepare another file where you save the middle 500 samples (of those 1000 samples) as a tab separated value.
* However, in the newly created file, ignore the following columns: Low, High and Comments. For the remaining columns, replace each column title with their lower-case value.
* Submit the newly created file as part of the zip submission

2) 5 pt

Import Twitter Corpus from NLTK.

For each of the tweets in the tweets.20150430-223406.json corpus, apply tokenization and POS tagging.

* Try using both NLTK and Spacy

Find the total number of adjectives and nouns (can be with or without duplicate), and output them.

Find the top-10 (in terms of frequency) nouns and output them.

3) 5 pt

Build a binary classifier using Naive Bayes using this dataset. Use Gaussian Naive Bayes. <https://archive.ics.uci.edu/dataset/53/iris>

[Links to an external site.](https://archive.ics.uci.edu/dataset/53/iris)Test the accuracy on 80-20 split. Print macro-F1 accuracy.

Submit your python file and output file (in txt).

[5% penalty if you submit after 6:30 pm, but by 6:45 pm]

[10% penalty if you submit after 6:45, but before 7:00 pm]

15% penalty after 7:00 pm

20% penalty after 7:30 pm

50% penalty after 8:00 pm]