Q6

Due Nov 7 at 11:59pm **Points** 1 **Questions** 10
Available after Nov 4 at 11:59pm **Time Limit** 20 Minutes

Instructions

This quiz covers material from: **week 6**

Reminders:

- There are 10 questions.
- You have only one attempt.
- You have 20 minutes.

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	14 minutes	1 out of 1

ⓘ Correct answers are hidden.

Score for this quiz: **1** out of 1
Submitted Nov 7 at 11:09am
This attempt took 14 minutes.

Question 1 — 0.1 / 0.1 pts

Tokenization is often a first step in text analysis. If using `word_tokenize` (as done in the class example), how would you expect the sentence "I ate lunch today" to be tokenized (where each string in a list is a different token)?

- ['ate', 'lunch', 'today']
- ['I', 'ate', 'lunch', 'today']

- ['I ate', 'lunch today']
- ['I ate lunch today']

Question 2

0.1 / 0.1 pts

Why do we remove stop words in text analysis?

- (selected) As they don't carry emotional information, remove them to lower the computational burden.
- As they carry the vast majority of emotional information, remove them to allow for detection of more common words.
- As they don't carry emotional information, remove them to increase the computational burden.
- As they carry the vast majority of emotional information, remove them to allow for detection of less common words.

Question 3

0.1 / 0.1 pts

Sentiment Analysis aims to programmatically infer the _____.

- meaning behind verbose text.

- [x] emotional content of text.
- [ ] most important words in a text.
- [ ] age of a text.

Question 4

0.1 / 0.1 pts

To carry out natural language processing analyses in Python, one popular package specifically designed to carry out these analyses is...

- [ ] tf-idf
- [ ] pandas
- [x] nltk (natural language toolkit)
- [ ] check_word

Question 5

0.1 / 0.1 pts

Which of the following would you expect would be removed from analysis during stop word removal?

- [ ] joyous
- [ ] disgusting
- [ ] heart
- [x] and

Question 6

0.1 / 0.1 pts

In lecture, we tried to find a multilinear relationship between states' teen birth rate and two predictors, namely poverty percentage and violent crime rate. We removed one data point because there was a high correlation between its predictors. Why was this necessary?

- [ ] Statsmodels couldn't handle that much data.
- [x] We can't have multicollinearity when doing linear regression.
- [ ] The data point skewed the distribution of poverty percentage.
- [ ] The data point didn't have tidy data.

Question 7

0.1 / 0.1 pts

Which of the following seaborn functions is useful for quickly visualizing the maximum likelihood line against your data (though isn't the best for statistical interpretation)?

- [ ] sns.lineplot()
- [ ] sns.scatterplot()
- [x] sns.lmplot()
- [ ] sns.histplot()

Question 8

0.1 / 0.1 pts

What is the process during sentiment analysis where we identify the root of each token (ex: jumping vs jumped)?

- [ ] frequency analysis
- [x] stemming
- [ ] tokenization
- [ ] tf-idf

Question 9

0.1 / 0.1 pts

Why do we sometimes want to do a log transformation before linear regression?

- [ ] Because logarithms are really fun to work with.
- [ ] To scale the difference between the outcome and predictors.
- [x] To make the data more approximately Normal.
- [ ] To prove the predictors are independent.

Question 10

0.1 / 0.1 pts

What is a token?

- [ ] A sentence with lots of information.
- [x] A meaningful unit of text.
- [ ] Another way of saying corpus.
- [ ] Any word that doesn't seem useful for sentiment analysis.

Quiz Score: **1** out of 1