

✔ Congratulations! You passed!

Grade
received 80%

Latest Submission
Grade 80%

To pass 70% or
higher

Go to next item

1. What is the main benefit of visualizing the course titles in a word cloud?

1 / 1 point

- ☒ The word cloud provides a quick visualization of the popular learning topics across all the courses.
- ☐ The word cloud displays tallies for the most frequently used words in all of the course titles.
- ☐ The word cloud provides deeper insight into similarity among course titles than does a histogram or bar chart.
- ☐ The word cloud can be displayed to the user to allow them to select from the most popular courses.

✔ **Correct**
Correct. A word cloud allows us to visually inspect popular learning topics among all courses.

2. In the Exploratory Data Analysis lab, how can we find the course enrollment counts for each user using Pandas dataframe?

1 / 1 point

- ☒ Use the `groupby()` on the user column and use the `size()` method to count the courses for each user.
- ☐ Use the `groupby()` method on the rating column then use the `slice()` method on the user column.
- ☐ Use the `groupby()` method on the user column then use the `slice()` method on the rating column.

✔ **Correct**
Correct. Apply Pandas' `groupby()` and `size()` method on the user column to aggregate the rating count for each user, then report the total number of users after aggregation.

3. In the Exploratory Data Analysis lab, why do we need to plot a histogram that shows the number of how many courses users are enrolled in (i.e user enrollment)?

1 / 1 point

- ☐ To identify when students enrolled in a particular course
- ☒ To illustrate the distribution of course enrollment
- ☐ To identify the number of students enrolled in each course
- ☐ To illustrate the top-20 most popular courses

✔ **Correct**
Correct. A histogram allows us to see the distribution of course enrollment across all users.

4. In the Exploratory Data Analysis lab, which percentage range do the 20 highest rated courses fall into when compared to the total number of ratings?

0 / 1 point

- ☐ 0%-24%
- ☐ 50%-74%
- ☒ 75%-100%
- ☐ 25%-49%

✘ **Incorrect**
Incorrect. Review the Exploratory Data Analysis lab.

5. Which of the following best describes a “Bag of Words” (BoW) feature?

1 / 1 point

- ☒ An array containing the frequency that words appear in a course's title and description
- ☐ A random sample of words from the course titles and descriptions
- ☐ An indexed token dictionary
- ☐ The distribution of the most frequently used words from all course titles and descriptions

✔ **Correct**
Correct. In general, BoW features are essentially the counts or frequencies of each word that appears in a list of words.

6. In the Extract BoW Features lab, what does the `stopwords.words()` method do?

1 / 1 point

- ☐ Filters out a list of commonly used but unimportant words from a Bag of Words
- ☐ Keeps a list of important words from a Bag of Words
- ☒ Retrieves a list of commonly used but unimportant words
- ☐ Allows you to enter a list of commonly used but unimportant words into a dictionary

✔ **Correct**
Correct. The `stopwords.words()` method returns a list of commonly used but unimportant words.

7. In the Extract BoW Features lab, what does the method `tokens_dict.doc2bow()` do?

1 / 1 point

- ☐ Creates a token dictionary
- ☐ Generates a Bag of Words feature given a course's title and description
- ☒ Generates a Bag of Words feature from a tokenized list
- ☐ Counts the number of nouns in a Bag of Words

✔ **Correct**
Correct. The method `tokens_dict.doc2bow()` generates a Bag of Words feature vector from a tokenized course.

8. Which of the following could NOT be a cosine similarity measurement?

1 / 1 point

- ☐ 1
- ☐ 0.25
- ☐ 0
- ☒ -0.25

✔ **Correct**
Correct. Cosine similarity measurements cannot be negative.

9. Which format of the Bag of Words feature can be used directly to compute the cosine similarity?

0 / 1 point

- ☐ Horizontal/sparse
- ☐ Set
- ☒ A Dense/vertical
- ☐ Dictionary

✘ **Incorrect**
Incorrect. Review the Calculate Course Similarity lab.

10. When comparing two course's Bag of Words features you find the cosine similarity to be 0.72. Which of the following is a true statement about this measurement?

1 / 1 point

- ☒ The two courses can be considered relatively similar to each other.
- ☐ The two courses have less than 28% of their Bag of Words features in common.
- ☐ The two courses have exactly 72% of their words Bag of Words features in common.
- ☐ The two courses are not very similar to each other.

✔ **Correct**
Correct. The closer the cosine similarity is to 1, the more similar the courses are to each other.