Practice Quiz

1. For the Exploratory Data Analysis lab, did you find out which courses may be of interest to you?
2. For the Exploratory Data Analysis lab, did you use the Pandas dataframe sum() and sort\_values() methods to generate a sorted course count per genre?
3. For the Exploratory Data Analysis lab, did you use the Seaborn barplot() or other plot methods to plot course genre counts using a bar chart?
4. For the Exploratory Data Analysis lab, did you apply the Pandas groupby() and size() methods on the user column to aggregate the rating count for each user, then report the total number of users after aggregation?
5. For the Exploratory Data Analysis lab, did you use describe() to report the statistics summary of the user enrollments?
6. For the Exploratory Data Analysis lab, did you plot the histogram of user interactions?
7. For the Exploratory Data Analysis lab, did you find the top 20 rated courses?
8. For the Exploratory Data Analysis lab, did you use Pandas merge() method to join the course\_df (contains the course title column)?
9. For the Exploratory Data Analysis lab, did you get the rating sum of the top 20 courses?
10. For the Extract BoW Features lab did you use provided tokenize\_course() method to tokenize all courses in courses\_df['course\_texts']?
11. For the Extract BoW Features lab did you use gensim.corpora.Dictionary(tokenized\_courses) to create a token dictionary?
12. For the Extract BoW Features lab did you use tokens\_dict.doc2bow(course) to generate BoW features for each tokenized course?
13. For the Extract BoW Features lab did you create a new course\_bow dataframe based on the extracted BoW feature?
14. For the Calculate Course Similarity Using BoW lab did you find courses that are similar to course Machine Learning with Python (ML0101ENv3)?

Answer to all above questions is ‘yes’

Quiz

1. What is the main benefit of visualizing the course titles in a word cloud?
   1. The word cloud provides a quick visualization of the popular learning topics across all the courses.
2. In the Exploratory Data Analysis lab, how can we find the course enrollment counts for each user using Pandas dataframe?
   1. ?
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. To illustrate the distribution of course enrollment
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?
   1. 50%-74%
5. Which of the following best describes a “Bag of Words” (BoW) feature?
   1. An array containing the frequency that words appear in a course’s title and description
6. In the Extract BoW Features lab, what does the stopwords.words() method do?
   1. Retrieves a list of commonly used but unimportant words
7. In the Extract BoW Features lab, what does the method tokens\_dict.doc2bow() do?
   1. Generates a Bag of Words feature from a tokenized list
8. Which of the following could NOT be a cosine similarity measurement?
   1. -0.25
9. Which format of the Bag of Words feature can be used directly to compute the cosine similarity?
   1. Horizontal/sparse
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. The two courses can be considered relatively similar to each other.