



✓ Congratulations! You passed!

Next Item



1 / 1  
point

1. Make sure you have the latest versions of the notebook and the file `kmeans-arrays.npz` [Read this post](#) if

- ... you downloaded the files before September 10
- ... you created an Amazon EC2 instance before October 1



1 / 1  
point

2. (True/False) The clustering objective (heterogeneity) is non-increasing for this example.



1 / 1  
point

3. Let's step back from this particular example. If the clustering objective (heterogeneity) would ever increase when running K-means, that would indicate: (choose one)



1 / 1  
point

4. Refer to the output of K-means for  $K=3$  and  $\text{seed}=0$ . Which of the three clusters contains the greatest number of data points in the end?



1 / 1  
point

5. Another way to capture the effect of changing initialization is to look at the distribution of cluster assignments. Compute the size (# of member data points) of clusters for each of the multiple runs of K-means.

Look at the size of the largest cluster (most # of member data points) across multiple runs, with seeds 0, 20000, ..., 120000. What is the **minimum** value this quantity takes?



1 / 1  
point

6. Refer to the section "Visualize clusters of documents". Which of the 10 clusters above contains the **greatest** number of articles?



1 / 1  
point

7. Refer to the section "Visualize clusters of documents". Which of the 10 clusters above contains the **least** number of articles?



1 / 1  
point

8. Another sign of too large  $K$  is having lots of small clusters. Look at the distribution of cluster sizes (by number of member data points). How many of the 100 clusters have fewer than 236 articles, i.e. 0.4% of the dataset?