

6. What do clusters 7, 8, and 11 have in common?

 $\bigcirc \ \ \, \text{They capture weather patterns associated with warm and dry days}$

TO PASS 80% or higher

Keep Learning

grade 100%

1/1 point

Cluster Analysis in Spark Quiz

157812 / 158726 = 99.4% 157237 / 158726 = 99.1% There is not enough information to determine this ✓ Correct Why is it necessary to scale the data (Step 4)? Since the values of the features are on different scales, all features need to be scaled so that the cluster centers can be displayed on the same plot for easier analysis. Since the values of the features are on different scales, all features need to be scaled so that no one feature dominates the clustering results. Since the values of the features are on different scales, all features need to be scaled so that no one feature dominates the clustering results. Since the values of the features are on different scales, all features need to be scaled so that all values will be positive. ✓ Correct If we wanted to create a data subset by taking every 5th sample instead of every 10th sample, how many samples would be in that subset? 3 17.452 1 1.587.257 158,726 ✓ Correct What is the significance of "seed-17" This line of code creates a k-means model with 12 clusters: 1 kmeans = KMeans (k=12, seed=1)? This specifies that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed. This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 9	ATEST SUBMISSION GRADE	
This line of code creates a k-means model with 12 clusters: 1 ** **Linear** = **Kmans (k-12, seed-1)** **What is the significance of **seed-1?* **This line of code creates a k-means model with 12 clusters: 1 **Linear** = **Kmans (k-12, seed-1)** **What is the significance of **seed-1?* **This specifies that the first cluster centeroid is set to sample #1 **Correct* **Correct** **This specifies that the first cluster centeroid is set to sample #1 **Correct** **Correct** **Linear** = **Kmans (k-12, seed-1)** **This specifies that the first cluster centeroid is set to sample #1 **Correct** **Linear** = **Kmans (k-12, seed-1)** **Linear** = **Line	. What percentage of samples have 0 for rain_accumulation?	1/1 point
There is not enough information to determine this ✓ Correct Vity is it necessary to scale the data (Step 4)?	157812 / 158726 = 99.4%	
Why is it necessary to scale the data (Step 4)? Since the values of the features are on different scales, all features need to be scaled so that the cluster centers can be displayed on the same plot for easier analysis. Since the values of the features are on different scales, all features need to be scaled so that no one feature dominates the clustering results. Since the values of the features are on different scales, all features need to be scaled so that all values will be positive. ✓ correct If we wanted to create a data subset by taking every 5th sample instead of every 10th sample, how many samples would be in that subset? ③ 317.62 ③ 15.87.257 □ 158.726 ✓ correct What is the significance of "seed=17" ⑥ This sets the seed to a specific value, which is necessary to reproduce the k-means results □ this means that this is the first iteration of k-means. The seed value is incremented by 1 every-time k-means is executed □ This specifies that the first cluster centroid is set to sample #1 ✓ correct Just by looking at the values for the cluster centers, which duster contains samples with the lowest relative humidity? ④ Cluster 4 □ Cluster 9 □ Cluster 9 □ Cluster 9	157237 / 158726 = 99.1%	
Why is it necessary to scale the data (Step 4)? Since the values of the features are on different scales, all features need to be scaled so that the cluster centers can be displayed on the same plot for easier analysis. Since the values of the features are on different scales, all features need to be scaled so that no one feature dominates the clustering results. Since the values of the features are on different scales, all features need to be scaled so that all values will be positive. ✓ Correct If we wanted to create a data subset by taking every 5th sample instead of every 10th sample, how many samples would be in that subset? ③ 317/452 ① 1,587.257 ① 158,726 ✓ Correct This line of code creates a k-means model with 12 clusters: 1	There is not enough information to determine this	
Since the values of the features are on different scales, all features need to be scaled so that the cluster centers can be displayed on the same plot for easier analysis. Since the values of the features are on different scales, all features need to be scaled so that no one feature dominates the clustering results. Since the values of the features are on different scales, all features need to be scaled so that all values will be positive. Correct If we wanted to create a data subset by taking every 5th sample instead of every 10th sample, how many samples would be in that subset? 3 317,452 1 1,587,257 158,726 Correct This line of code creates a k-means model with 12 clusters: 1	✓ Correct	
be displayed on the same plot for easier analysis. Since the values of the features are on different scales, all features need to be scaled so that no one feature dominates the clustering results. Since the values of the features are on different scales, all features need to be scaled so that all values will be positive. Correct If we wanted to create a data subset by taking every 5th sample instead of every 10th sample, how many samples would be in that subset? 3 17,452 1,587,257 158,726 Correct This line of code creates a k-means model with 12 clusters: 1 kmeans = KMeans (k-12, seed-1) What is the significance of seed-1?? This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 3 Cluster 9	. Why is it necessary to scale the data (Step 4)?	1/1 point
dominates the clustering results. Since the values of the features are on different scales, all features need to be scaled so that all values will be positive. Correct If we wanted to create a data subset by taking every 5th sample instead of every 10th sample, how many samples would be in that subset? 317,452 1,587,257 158,726 Correct This line of code creates a k-means model with 12 clusters: 1 kmeans = KMeans (k-12, seed-1) What is the significance of "seed-1?? This sets the seed to a specific value, which is necessary to reproduce the k-means results This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 9		n
y Correct If we wanted to create a data subset by taking every 5th sample instead of every 10th sample, how many samples would be in that subset? 3 117,452 1,587,257 158,726 Correct This line of code creates a k-means model with 12 clusters: 1 k-means = K-Means (k-12, seed=1) What is the significance of "seed=1"? This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 9 Cluster 9		
If we wanted to create a data subset by taking every 5th sample instead of every 10th sample, how many samples would be in that subset? ③ 317,452 ① 1,587,257 ① 158,726 ✔ Correct 1 kmeans = KMeans (k-12, seed-1) What is the significance of "seed=1"? ④ This sets the seed to a specific value, which is necessary to reproduce the k-means results □ This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed □ This specifies that the first cluster centroid is set to sample #1 ✔ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? ⑥ Cluster 4 ○ Cluster 9 Cluster 9		
be in that subset? ③ 317,452 ① 1,587,257 ① 158,726 ✓ Correct 1 kmeans = KMeans (k=12, seed=1) What is the significance of "seed=1?? ④ This sets the seed to a specific value, which is necessary to reproduce the k-means results ① This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed ① This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? ⑥ Cluster 4 ○ Cluster 3 ○ Cluster 9	✓ Correct	
1,587,257 158,726 ✓ Correct This line of code creates a k-means model with 12 clusters: 1 kmeans = KMeans (k=12, seed=1) What is the significance of "seed=1"? This sets the seed to a specific value, which is necessary to reproduce the k-means results This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 9 Cluster 9		ld 1/1 point
 ✓ Correct This line of code creates a k-means model with 12 clusters: 1 kmeans = KMeans (k=12, seed=1) What is the significance of "seed=1"? This sets the seed to a specific value, which is necessary to reproduce the k-means results This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 9 	317,452	
✓ Correct This line of code creates a k-means model with 12 clusters: 1 kmeans = KMeans (k=12, seed=1) What is the significance of "seed=1"? This sets the seed to a specific value, which is necessary to reproduce the k-means results This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 9 Cluster 9	O 1,587,257	
This line of code creates a k-means model with 12 clusters: 1 kmeans = KMeans (k=12, seed=1) What is the significance of "seed=1"? This sets the seed to a specific value, which is necessary to reproduce the k-means results This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 9 Cluster 9	158,726	
1 kmeans = KMeans (k=12, seed=1) What is the significance of "seed=1"? This sets the seed to a specific value, which is necessary to reproduce the k-means results This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 3 Cluster 9	✓ Correct	
1 kmeans = KMeans (k=12, seed=1) What is the significance of "seed=1"? This sets the seed to a specific value, which is necessary to reproduce the k-means results This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 3 Cluster 9	This line of code creates a k-means model with 12 clusters:	1 / 1 noint
What is the significance of "seed=1"?		17 1 point
 This sets the seed to a specific value, which is necessary to reproduce the k-means results This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? ● Cluster 4 Cluster 3 Cluster 9 	1 kmeans = KMeans (k=12, seed=1)	
This means that this is the first iteration of k-means. The seed value is incremented by 1 every time k-means is executed This specifies that the first cluster centroid is set to sample #1 ✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 3 Cluster 9	What is the significance of "seed=1"?	
executed This specifies that the first cluster centroid is set to sample #1 Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 3 Cluster 9	This sets the seed to a specific value, which is necessary to reproduce the k-means results	
✓ Correct Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 3 Cluster 9		
 Just by looking at the values for the cluster centers, which cluster contains samples with the lowest relative humidity? Cluster 4 Cluster 3 Cluster 9 	This specifies that the first cluster centroid is set to sample #1	
Cluster 4Cluster 3Cluster 9	✓ Correct	
Cluster 4Cluster 3Cluster 9		
Cluster 3 Cluster 9		1/1 point
Cluster 9		
✓ Correct	Cluster 9	
	✓ Correct	

	They capture weather patterns associated with high air pressure	
	They capture weather patterns associated with very strong winds	
	✓ Correct	
7.	If we perform clustering with 20 clusters (and seed = 1), which cluster appears to identify Santa Ana conditions (lowest humidity and highest wind speeds)?	1 / 1 point
	Cluster 12	
	Cluster 1	
	Cluster 16	
	✓ Correct	
8.	We did not include the minimum wind measurements in the analysis since they are highly correlated with the average wind measurements. What is the correlation between min_wind_speed and avg_wind_speed (to two decimals)? (Compute this using one-tenth of the original dataset, and dropping all rows with missing values.)	1/1 point
	● 0.97	
	O -0.12	
	O 0.62	
	✓ Correct	