



## K-means Assignment

[support@intellipaat.com](mailto:support@intellipaat.com)

+91-7022374614

US: 1-800-216-8930 (Toll-Free)

## Data Set Information:

This dataset has been obtained from UCI ML Repository.

<https://archive.ics.uci.edu/ml/datasets/Facebook+Live+Sellers+in+Thailand>

The 'Facebook Live Sellers in Thailand' is a dataset present in UCI Machine Learning Datasets. The data includes 7050 observations and twelve attributes. The data is about the live selling attribute on the Facebook platform. Each record consists of details about the time live details of the sale are published to Facebook and engagements in the data. The engagements are regular Facebook dealings such as sharing and emoji reaction.

Column Name	Description
Status ID	Id of the status
Status Type	Media type of the post(video, photo, link, status)
Status Published	Date and time of the status being posted
Num_reactions	Number of reactions given to that post
Num_comments	Number of comments on the post
num_shares	Number of shares made for that post
num_likes	Number of Thumbs up emojis gave to the post
num_loves	Number of heart emojis given to the post
num_wows	Number of wow emojis given to the post
num_hahas	Number of haha emojis given to the post
num_sads	Number of sad emojis given to the post
num_angrys	Number of angry emojis given to the post

## Business Problem:

Live Shopping on Facebook is an interactive way to sell items, connect straight with viewers, and gain likely customers, all in real-time. When you sell products through Live Shopping on Facebook, you are live streaming as you feature and demonstrate your products. Small vendors can now reach a more expansive audience and connect with many clients. You need to implement K-Means clustering to find intrinsic batches within the dataset that depict the same status\_type behavior. The status\_type behavior variable consists of posts of a distinct nature (video, photos, statuses, and links).

## Questions:

1. Which columns contain null data?
  - a. 4
  - b. 5
  - c. 3
  - d. 0
2. How many status ids are unique?
  - a. 4509
  - b. 3984
  - c. 6503
  - d. 6997
3. How many video post has been made for live selling:
  - a. 2334
  - b. 2345
  - c. 1827
  - d. 2871
4. How many likes have been given to the photo posted on 4/19/2018 at 22:26?
  - a. 359
  - b. 379
  - c. 293
  - d. 493
5. How many total shares were made for the video post that was posted on 7/15/2017 at 10:52?
  - a. 153
  - b. 167
  - c. 139
  - d. 173
6. What does k means do?
  - a. Group similar elements
  - b. Finds probability of elements
  - c. Finds relationships between dependent and independent data

- d. None of the above
7. What output will the below code give?

```
colour=['red','blue','green','yellow','cyan']
```

```
n=0
```

```
for i in range(0,len(colour)):
```

```
    plt.scatter(X[df['Clusters(5)']==i],Y[df['Clusters(5)']==i],s=10,c=colour[i])
```

```
plt.scatter(k.cluster_centers_[0],k.cluster_centers_[1], color='black')
```

- a. Gives an error
- b. Gives 2 clusters
- c. Gives 5 clusters
- d. Gives 5 clusters with midpoints
8. Which method is used to decide the optimal number of clusters in k-means clustering?
- a. Elbow method
- b. Append method
- c. Evaluation Method
- d. None
9. In what cases is K Means used?
- a. Regression
- b. Classification
- c. Clustering
- d. None of the above
10. For exploration, do we use Hierarchical Clustering?
- a. True
- b. False
11. Which is the function used for the K Means?
- a. KMeans
- b. Kmeans
- c. Clusters
- d. None of the above
12. How to train the K-means model?
- a. predict()
- b. fit()
- c. train()
- d. train\_test\_split()
13. What does random\_state do?
- a. Shows states
- b. Checks and validate data when code is run multiple times.
- c. Used to split data
- d. All of the above.

14. How to check which data is placed under which cluster?

- a. Labels\_
- b. Intertia\_
- c. Cluster\_
- d. None of the above

15. Why do we perform standard scalar?

- a. Normalizing data
- b. Removing standard values
- c. Adding standard values
- d. None of the above

Perform the following operations on the dataset:

- Drop Status id and status published column.
- Use a label encoder to encode the status type column.
- Standardize the data using min-max scalar
- Create a K means model for 2 clusters.

16. From the above model, what is the inter-cluster variance of the model?

- a. 237.75
- b. 342.43
- c. 210.93
- d. 284.29