Q.1: What is the difference between supervised and unsupervised machine learning

Supervised learning When an algorithm is trained on a labelled dataset—that is, when the input data used for training is paired with corresponding output labels—it is referred to as supervised learning. Supervised learning aims to find a mapping or relationship between the input variables and the desired output, which enables the algorithm to produce precise predictions or classifications when faced with fresh, unobserved data.

An input-output pair training set is given to the algorithm during a supervised learning process. For every example in the training set, the algorithm iteratively modifies its parameters to minimize the discrepancy between its predicted output and the actual output (the ground truth). This procedure keeps going until the algorithm performs at an acceptable level.

Supervised learning can be divided into two main types:

1. Regression: In regression problems, the goal is to predict a continuous output or value. For example, predicting the price of a house based on its features, such as the number of bedrooms, square footage, and location.

2. Classification: In classification problems, the goal is to assign input data to one of several predefined categories or classes. Examples include spam email detection, image classification (e.g., identifying whether an image contains a cat or a dog), and sentiment analysis.

Supervised Learning Example: Suppose there is a basket which is filled with some fresh fruits, the task is to arrange the same type of fruits in one place. Also, suppose that the fruits are apple, banana, cherry, and grape. Suppose one already knows from their previous work (or experience) that, the shape of every fruit present in the basket so, it is easy for them to arrange the same type of fruits in one place. Here, the previous work is called training data in Data Mining terminology. So, it learns things from the training data. This is because it has a response variable that says y that if some fruit has so and so features then it is grape, and similarly for every fruit. This type of information is deciphered from the data that is used to train the model. This type of learning is called Supervised Learning. Such problems are listed under classical Classification Tasks.

Unsupervised Learning Unsupervised learning is a type of machine learning where the algorithm is given input data without explicit instructions on what to do with it. In unsupervised learning, the algorithm tries to find patterns, structures, or relationships in the data without the guidance of labelled output.

The main goal of unsupervised learning is often to explore the inherent structure within a set of data points. This can involve identifying clusters of similar data points, detecting outliers, reducing the dimensionality of the data, or discovering patterns and associations.

There are several common types of unsupervised learning techniques:

1. Clustering: Clustering algorithms aim to group similar data points into clusters based on some similarity metric. K-means clustering and hierarchical clustering are examples of unsupervised clustering techniques.

2. Dimensionality Reduction: These techniques aim to reduce the number of features (or dimensions) in the data while preserving its essential information. Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) are examples of dimensionality reduction methods.

3. Association: Association rule learning is used to discover interesting relationships or associations between variables in large datasets. The Apriori algorithm is a well-known example used for association rule learning.

Unsupervised Learning example Again, Suppose there is a basket and it is filled with some fresh fruits. The task is to arrange the same type of fruits in one place. This time there is no information about those fruits beforehand, it's the first time that the fruits are being seen or discovered So how to group similar fruits without any prior knowledge about them? First, any physical characteristic of a particular fruit is selected. Suppose colour. Then the fruits are arranged based on the color. The groups will be something as shown below:

RED COLOR GROUP: apples & cherry fruits. GREEN COLOR GROUP: bananas & grapes. So now, take another physical character say, size, so now the groups will be something like this. RED COLOR AND BIG SIZE: apple. RED COLOR AND SMALL SIZE: cherry fruits. GREEN COLOR AND BIG SIZE: bananas. GREEN COLOR AND SMALL SIZE: grapes.

Difference between Supervised and Unsupervised Learning The distinction between supervised and unsupervised learning depends on whether the learning algorithm uses pattern-class information. Supervised learning assumes the availability of a teacher or supervisor who classifies the training examples, whereas unsupervised learning must identify the pattern-class information as a part of the learning process.

Supervised learning algorithms utilize the information on the class membership of each training instance. This information allows supervised learning algorithms to detect pattern misclassifications as feedback to themselves. In unsupervised learning algorithms, unlabeled instances are used. They blindly or heuristically process them. Unsupervised learning algorithms often have less computational complexity and less accuracy than supervised learning algorithms.

Q.2: Explain the concept of clustering using different methods in unsupervised learning

Clustering: It is basically a type of unsupervised learning method. An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

Clustering is very much important as it determines the intrinsic grouping among the unlabelled data present. There are no criteria for good clustering. It depends on the user, and what criteria they may use which satisfy their need. For instance, we could be interested in finding representatives for homogeneous groups (data reduction), finding “natural clusters” and describing their unknown properties (“natural” data types), in finding useful and suitable groupings (“useful” data classes) or in finding unusual data objects (outlier detection). This algorithm must make some assumptions that constitute the similarity of points and each assumption make different and equally valid clusters.

Clustering Methods:

Density-Based Methods: These methods consider the clusters as the dense region having some similarities and differences from the lower dense region of the space. These methods have good accuracy and the ability to merge two clusters. Example DBSCAN (Density-Based Spatial Clustering of Applications with Noise), OPTICS (Ordering Points to Identify Clustering Structure), etc. Hierarchical Based Methods: The clusters formed in this method form a tree-type structure based on the hierarchy. New clusters are formed using the previously formed one. It is divided into two category Agglomerative (bottom-up approach) Divisive (top-down approach) Examples CURE (Clustering Using Representatives), BIRCH (Balanced Iterative Reducing Clustering and using Hierarchies), etc.

Partitioning Methods: These methods partition the objects into k clusters and each partition forms one cluster. This method is used to optimize an objective criterion similarity function such as when the distance is a major parameter example K-means, CLARANS (Clustering Large Applications based upon Randomized Search), etc. Grid-based Methods: In this method, the data space is formulated into a finite number of cells that form a grid-like structure. All the clustering operations done on these grids are fast and independent of the number of data objects example STING (Statistical Information Grid), wave cluster, CLIQUE (CLustering In Quest), etc.

Clustering Algorithms: K-means clustering algorithm – It is the simplest unsupervised learning algorithm that solves clustering problem.K-means algorithm partitions n observations into k clusters where each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster.

Applications of Clustering in different fields:

1. Marketing: It can be used to characterize & discover customer segments for marketing purposes.
2. Biology: It can be used for classification among different species of plants and animals.
3. Libraries: It is used in clustering different books on the basis of topics and information.
4. Insurance: It is used to acknowledge the customers, their policies and identifying the frauds.
5. City Planning: It is used to make groups of houses and to study their values based on their geographical locations and other factors present.

6. Earthquake studies: By learning the earthquake-affected areas we can determine the dangerous zones.
7. Image Processing: Clustering can be used to group similar images together, classify images based on content, and identify patterns in image data.
8. Genetics: Clustering is used to group genes that have similar expression patterns and identify gene networks that work together in biological processes.
9. Finance: Clustering is used to identify market segments based on customer behavior, identify patterns in stock market data, and analyze risk in investment portfolios.
10. Customer Service: Clustering is used to group customer inquiries and complaints into categories, identify common issues, and develop targeted solutions.
11. Manufacturing: Clustering is used to group similar products together, optimize production processes, and identify defects in manufacturing processes.
12. Medical diagnosis: Clustering is used to group patients with similar symptoms or diseases, which helps in making accurate diagnoses and identifying effective treatments.
13. Fraud detection: Clustering is used to identify suspicious patterns or anomalies in financial transactions, which can help in detecting fraud or other financial crimes.
14. Traffic analysis: Clustering is used to group similar patterns of traffic data, such as peak hours, routes, and speeds, which can help in improving transportation planning and infrastructure.
15. Social network analysis: Clustering is used to identify communities or groups within social networks, which can help in understanding social behavior, influence, and trends.
16. Cybersecurity: Clustering is used to group similar patterns of network traffic or system behavior, which can help in detecting and preventing cyberattacks.
17. Climate analysis: Clustering is used to group similar patterns of climate data, such as temperature, precipitation, and wind, which can help in understanding climate change and its impact on the environment.
18. Sports analysis: Clustering is used to group similar patterns of player or team performance data, which can help in analyzing player or team strengths and weaknesses and making strategic decisions.
19. Crime analysis: Clustering is used to group similar patterns of crime data, such as location, time, and type, which can help in identifying crime hotspots, predicting future crime trends, and improving crime prevention strategies.

```
```

Q.3: Load the Customer_Churn dataset. a. Build the kmeans algorithm on top of ' customer features '. For the model, the number of clusters should be 3 b. Calculate the clustering vector values for the monthly charges column form the customer_features c. Bind the monthly charges column to the clustering vector and store that data in month_group d. Separate all the 3 clusters with their values e. Write interference how k mean is different from KNN from above result

```
import pandas as pd
import numpy as np
```

```
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')

from sklearn.cluster import KMeans
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier

df = pd.read_csv("customer_churn.csv")
df.head()
```

```
   customerID  gender  SeniorCitizen Partner Dependents  tenure
PhoneService  \
0  7590-VHVEG  Female              0     Yes         No       1
No
1  5575-GNVDE    Male              0      No         No      34
Yes
2  3668-QPYBK    Male              0      No         No       2
Yes
3  7795-CFOCW    Male              0      No         No      45
No
4  9237-HQITU  Female              0      No         No       2
Yes

      MultipleLines InternetService OnlineSecurity  ...
DeviceProtection  \
0  No phone service             DSL             No  ...
No
1                No             DSL            Yes  ...
Yes
2                No             DSL            Yes  ...
No
3  No phone service             DSL            Yes  ...
Yes
4                No     Fiber optic             No  ...
No

  TechSupport StreamingTV StreamingMovies        Contract
PaperlessBilling  \
0          No          No              No  Month-to-month
Yes
1          No          No              No        One year
No
2          No          No              No  Month-to-month
Yes
3         Yes          No              No        One year
```

```
No
4          No          No               No  Month-to-month
Yes

               PaymentMethod MonthlyCharges  TotalCharges Churn
0           Electronic check          29.85         29.85    No
1               Mailed check          56.95        1889.5    No
2               Mailed check          53.85        108.15   Yes
3  Bank transfer (automatic)          42.30       1840.75    No
4           Electronic check          70.70        151.65   Yes

[5 rows x 21 columns]
```

```
df.info
```

```
<bound method DataFrame.info of       customerID  gender
SeniorCitizen Partner Dependents  tenure  \
0     7590-VHVEG  Female              0     Yes         No       1
1     5575-GNVDE    Male              0      No         No      34
2     3668-QPYBK    Male              0      No         No       2
3     7795-CFOCW    Male              0      No         No      45
4     9237-HQITU  Female              0      No         No       2
...          ...     ...            ...     ...        ...     ...
7038  6840-RESVB    Male              0     Yes        Yes      24
7039  2234-XADUH  Female              0     Yes        Yes      72
7040  4801-JZAZL  Female              0     Yes        Yes      11
7041  8361-LTMKD    Male              1     Yes         No       4
7042  3186-AJIEK    Male              0      No         No      66

     PhoneService     MultipleLines InternetService
OnlineSecurity  ...  \
0              No  No phone service             DSL
No  ...
1             Yes                No             DSL
Yes  ...
2             Yes                No             DSL
Yes  ...
3              No  No phone service             DSL
Yes  ...
4             Yes                No     Fiber optic
No  ...
...           ...               ...             ...            ...  ..
.
7038          Yes               Yes             DSL
Yes  ...
7039          Yes               Yes     Fiber optic
No  ...
7040           No  No phone service             DSL
Yes  ...
7041          Yes               Yes     Fiber optic
```

```
No  ...
7042          Yes                 No     Fiber optic
Yes  ...

     DeviceProtection TechSupport StreamingTV StreamingMovies
Contract  \
0                  No          No          No              No  Month-
to-month
1                 Yes          No          No              No
One year
2                  No          No          No              No  Month-
to-month
3                 Yes         Yes          No              No
One year
4                  No          No          No              No  Month-
to-month
...               ...         ...         ...             ...
...
7038              Yes         Yes         Yes             Yes
One year
7039              Yes          No         Yes             Yes
One year
7040               No          No          No              No  Month-
to-month
7041               No          No          No              No  Month-
to-month
7042              Yes         Yes         Yes             Yes
Two year

     PaperlessBilling              PaymentMethod MonthlyCharges
TotalCharges  \
0                 Yes           Electronic check          29.85
29.85
1                  No               Mailed check          56.95
1889.5
2                 Yes               Mailed check          53.85
108.15
3                  No  Bank transfer (automatic)          42.30
1840.75
4                 Yes           Electronic check          70.70
151.65
...               ...                        ...            ...
...
7038              Yes               Mailed check          84.80
1990.5
7039              Yes    Credit card (automatic)         103.20
7362.9
7040              Yes           Electronic check          29.60
346.45
```

```
7041               Yes                Mailed check           74.40
306.6
7042               Yes  Bank transfer (automatic)         105.65
6844.5

     Churn
0       No
1       No
2      Yes
3       No
4      Yes
...    ...
7038    No
7039    No
7040    No
7041   Yes
7042    No

[7043 rows x 21 columns]>
```

```
df.shape
```

```
(7043, 21)
```

```
df['gender'].value_counts()
```

```
gender
Male      3555
Female    3488
Name: count, dtype: int64
```

```
df['SeniorCitizen'].value_counts()
```

```
SeniorCitizen
0    5901
1    1142
Name: count, dtype: int64
```

```
df['Dependents'].value_counts()
```

```
Dependents
No     4933
Yes    2110
Name: count, dtype: int64
```

```
df['PhoneService'].value_counts()
```

```
PhoneService
Yes    6361
No      682
Name: count, dtype: int64
```

```
df['MultipleLines'].value_counts()
```

```
MultipleLines
No                  3390
Yes                 2971
No phone service     682
Name: count, dtype: int64
```

```
df['InternetService'].value_counts()
```

```
InternetService
Fiber optic    3096
DSL            2421
No             1526
Name: count, dtype: int64
```

```
df['Contract'].value_counts()
```

```
Contract
Month-to-month    3875
Two year          1695
One year          1473
Name: count, dtype: int64
```

```
df['PaymentMethod'].value_counts()
```

```
PaymentMethod
Electronic check             2365
Mailed check                 1612
Bank transfer (automatic)    1544
Credit card (automatic)      1522
Name: count, dtype: int64
```

```
df['Churn'].value_counts()
```

```
Churn
No     5174
Yes    1869
Name: count, dtype: int64
```

```
df.isnull().sum()
```

```
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
```

```
OnlineBackup         0
DeviceProtection     0
TechSupport          0
StreamingTV          0
StreamingMovies      0
Contract             0
PaperlessBilling     0
PaymentMethod        0
MonthlyCharges       0
TotalCharges         0
Churn                0
dtype: int64
```

```
df = df.drop(["customerID"], axis = 1)
df.head()
```

```
   gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0  Female              0     Yes         No       1           No
1    Male              0      No         No      34          Yes
2    Male              0      No         No       2          Yes
3    Male              0      No         No      45           No
4  Female              0      No         No       2          Yes

      MultipleLines InternetService OnlineSecurity OnlineBackup  \
0  No phone service             DSL             No          Yes
1                No             DSL            Yes           No
2                No             DSL            Yes          Yes
3  No phone service             DSL            Yes           No
4                No     Fiber optic             No           No

  DeviceProtection TechSupport StreamingTV StreamingMovies
Contract  \
0               No          No          No              No  Month-to-
month
1              Yes          No          No              No        One
year
2               No          No          No              No  Month-to-
month
3              Yes         Yes          No              No        One
year
4               No          No          No              No  Month-to-
month

  PaperlessBilling              PaymentMethod  MonthlyCharges
TotalCharges  \
0              Yes           Electronic check           29.85
29.85
1               No               Mailed check           56.95
1889.5
2              Yes               Mailed check           53.85
```

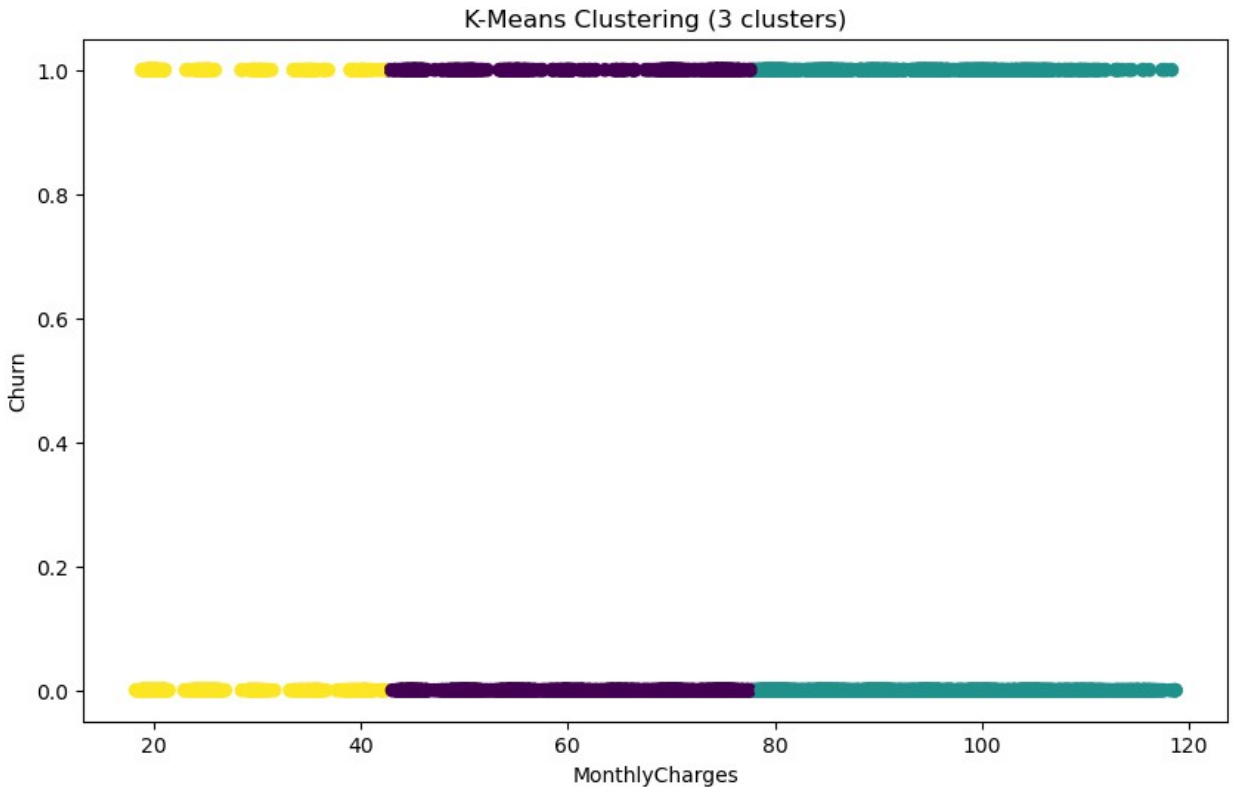
```
108.15
3               No  Bank transfer (automatic)           42.30
1840.75
4              Yes           Electronic check           70.70
151.65

  Churn
0    No
1    No
2   Yes
3    No
4   Yes
```

a. Build the kmeans algorithm on top of ' customer features '. For the model, the number of clusters should be 3 b. Calculate the clustering vector values for the monthly charges column form the customer_features

```
label_encoder = LabelEncoder()
df['Churn'] = label_encoder.fit_transform(df['Churn'])

X = df[['MonthlyCharges', 'Churn']]
kmeans = KMeans(n_clusters=3, random_state=42)
df['Cluster'] = kmeans.fit_predict(X)
monthly_charges_cluster = df.groupby('Cluster')
['MonthlyCharges'].mean().reset_index()
plt.figure(figsize=(10, 6))
plt.scatter(X['MonthlyCharges'], X['Churn'], c=df['Cluster'],
cmap='viridis')
plt.title('K-Means Clustering (3 clusters)')
plt.xlabel('MonthlyCharges')
plt.ylabel('Churn')
plt.show()
print('Clustering Vector Values for Monthly Charges:')
print(monthly_charges_cluster)
```




```
Clustering Vector Values for Monthly Charges:
   Cluster  MonthlyCharges
0        0       61.628808
1        1       94.054258
2        2       23.384619
```

c. Bind the monthly charges column to the clustering vector and store that data in month_group

```
df['month_group'] = df['MonthlyCharges'].astype(str) + '_Cluster_' +
df['Cluster'].astype(str)
print(df[['MonthlyCharges', 'Cluster', 'month_group']].head())
```

```
   MonthlyCharges  Cluster      month_group
0           29.85        2  29.85_Cluster_2
1           56.95        0  56.95_Cluster_0
2           53.85        0  53.85_Cluster_0
3           42.30        2   42.3_Cluster_2
4           70.70        0   70.7_Cluster_0
```

d. Separate all the 3 clusters with their values

```
cluster_0 = df[df['Cluster'] == 0]
cluster_1 = df[df['Cluster'] == 1]
cluster_2 = df[df['Cluster'] == 2]
```

```
print('Cluster 0:')
print(cluster_0.head())

print('\nCluster 1:')
print(cluster_1.head())

print('\nCluster 2:')
print(cluster_2.head())
```

```
Cluster 0:
    customerID  gender  SeniorCitizen Partner Dependents  tenure
PhoneService  \
1   5575-GNVDE    Male              0      No         No      34
Yes
2   3668-QPYBK    Male              0      No         No       2
Yes
4   9237-HQITU  Female              0      No         No       2
Yes
9   6388-TABGU    Male              0      No        Yes      62
Yes
10  9763-GRSKD    Male              0     Yes        Yes      13
Yes

   MultipleLines InternetService OnlineSecurity  ... StreamingTV  \
1             No             DSL             Yes  ...          No
2             No             DSL             Yes  ...          No
4             No     Fiber optic              No  ...          No
9             No             DSL             Yes  ...          No
10            No             DSL             Yes  ...          No

   StreamingMovies        Contract PaperlessBilling  \
1               No        One year               No
2               No  Month-to-month              Yes
4               No  Month-to-month              Yes
9               No        One year               No
10              No  Month-to-month              Yes

                PaymentMethod MonthlyCharges TotalCharges Churn
Cluster  \
1                Mailed check          56.95       1889.5     0
0
2                Mailed check          53.85       108.15     1
0
4            Electronic check          70.70       151.65     1
0
9   Bank transfer (automatic)          56.15      3487.95     0
0
10               Mailed check          49.95       587.45     0
0
```

```
        month_group
1   56.95_Cluster_0
2   53.85_Cluster_0
4    70.7_Cluster_0
9   56.15_Cluster_0
10  49.95_Cluster_0

[5 rows x 23 columns]

Cluster 1:
    customerID  gender  SeniorCitizen Partner Dependents  tenure
PhoneService  \
5   9305-CDSKC  Female              0      No         No       8
Yes
6   1452-KIOVK    Male              0      No        Yes      22
Yes
8   7892-POOKP  Female              0     Yes         No      28
Yes
12  8091-TTVAX    Male              0     Yes         No      58
Yes
13  0280-XJGEX    Male              0      No         No      49
Yes

   MultipleLines InternetService OnlineSecurity  ... StreamingTV  \
5            Yes     Fiber optic             No  ...         Yes
6            Yes     Fiber optic             No  ...         Yes
8            Yes     Fiber optic             No  ...         Yes
12           Yes     Fiber optic             No  ...         Yes
13           Yes     Fiber optic             No  ...         Yes

   StreamingMovies        Contract PaperlessBilling  \
5              Yes  Month-to-month              Yes
6               No  Month-to-month              Yes
8              Yes  Month-to-month              Yes
12             Yes        One year               No
13             Yes  Month-to-month              Yes

                 PaymentMethod MonthlyCharges TotalCharges Churn
Cluster  \
5            Electronic check          99.65        820.5     1
1
6     Credit card (automatic)          89.10       1949.4     0
1
8            Electronic check         104.80      3046.05     1
1
12    Credit card (automatic)         100.35       5681.1     0
1
13  Bank transfer (automatic)         103.70       5036.3     1
1
```

```
         month_group
5    99.65_Cluster_1
6     89.1_Cluster_1
8    104.8_Cluster_1
12  100.35_Cluster_1
13   103.7_Cluster_1

[5 rows x 23 columns]

Cluster 2:
    customerID  gender  SeniorCitizen Partner Dependents  tenure
PhoneService  \
0   7590-VHVEG  Female              0     Yes         No       1
No
3   7795-CFOCW    Male              0      No         No      45
No
7   6713-OKOMC  Female              0      No         No      10
No
11  7469-LKBCI    Male              0      No         No      16
Yes
16  8191-XWSZG  Female              0      No         No      52
Yes

       MultipleLines InternetService       OnlineSecurity  ...  \
0   No phone service             DSL                   No  ...
3   No phone service             DSL                  Yes  ...
7   No phone service             DSL                  Yes  ...
11                No              No  No internet service  ...
16                No              No  No internet service  ...

            StreamingTV      StreamingMovies        Contract
PaperlessBilling  \
0                   No                   No  Month-to-month
Yes
3                   No                   No        One year
No
7                   No                   No  Month-to-month
No
11  No internet service  No internet service        Two year
No
16  No internet service  No internet service        One year
No

                PaymentMethod MonthlyCharges TotalCharges Churn
Cluster  \
0            Electronic check          29.85        29.85     0
2
3   Bank transfer (automatic)          42.30      1840.75     0
2
```

```
7                Mailed check          29.75         301.9     0
2
11    Credit card (automatic)          18.95         326.8     0
2
16               Mailed check          20.65       1022.95     0
2

        month_group
0   29.85_Cluster_2
3    42.3_Cluster_2
7   29.75_Cluster_2
11  18.95_Cluster_2
16  20.65_Cluster_2

[5 rows x 23 columns]
```

e. Write interference how k mean is different from KNN from above result

K-Means (k-means clustering) and K-NN (k-nearest neighbors) are distinct machine learning algorithms with different purposes and methodologies:

K-Means (k-means clustering):

Objective: The primary goal of K-Means is to partition data points into distinct groups (clusters) based on their features. Unsupervised Learning: K-Means is an unsupervised learning algorithm, meaning it doesn't require labeled data. Clustering: It is used for clustering similar data points together without any predefined class labels. Centroid-Based: It works by iteratively assigning data points to clusters based on the similarity of features and updating the cluster centroids.

K-NN (k-nearest neighbors):

Objective: K-NN is used for classification or regression tasks. It predicts the label or value of a data point based on the majority class or mean value of its k-nearest neighbors.

Supervised Learning: K-NN is a supervised learning algorithm, meaning it requires labeled training data.

Classification and Regression: It can be used for both classification and regression tasks, depending on the nature of the target variable.

Instance-Based: It classifies a new data point by considering the labels of its k-nearest neighbors in the training set.

Interference:

K-Means is designed for unsupervised clustering, where the goal is to group similar data points together without predefined labels. K-NN, on the other hand, is a supervised algorithm used for making predictions based on the similarity of a data point to its labeled neighbors. The code you provided is related to K-Means clustering, where you are partitioning the data into clusters and analyzing the results.

In summary, K-Means and K-NN serve different purposes: K-Means is for clustering, while K-NN is for classification or regression tasks based on labeled training data.