



BANKBOOSTX

EMPOWERING BANKING SUCCESS

AGENDA

01

OVERVIEW

02

BANKBOOSTX

03

Objective &
Predictions

04

Summary &
Features

05

hierarchical clustering

06

Unsupervised
Clustering

07

Supervised
Clustering

08

CONCLUSION

OBJECTIVE

Project Objective

- Study aims to predict customer subscription to term deposits in Portuguese bank's marketing campaigns.
- Dataset covers May 2008 to November 2010 with 41,188 examples and 20 variables.
- BankBoostX, running the campaigns, seeks insights for better marketing strategies.
- The study benefits marketers and researchers in understanding customer behavior and improving marketing tactics.

Stakeholder Selection:

- BankBoostX is the primary stakeholder aiming to optimize the bank's marketing strategies.
- Understanding customer decision factors can boost engagement, order values, and resource allocation.
- BankBoostX's collaboration highlights its dedication to marketing effectiveness.
- The study aligns with shared goals for success.

Data Exploration

- Explored dataset with 41,188 examples and 20 input variables.
- Investigated customer behavior and preferences.
- Identified patterns influencing term deposit subscription.
- Evaluated data to inform marketing strategies.

FEATURES

- **job**

- **education**

- **loan**

- **day_of_week**

- **pdays**

- **emp.var.rate**

- **poutcome**

- **euribor3m**

- **nr.employed**

- **marital**

- **housing**

- **month**

- **campaign**

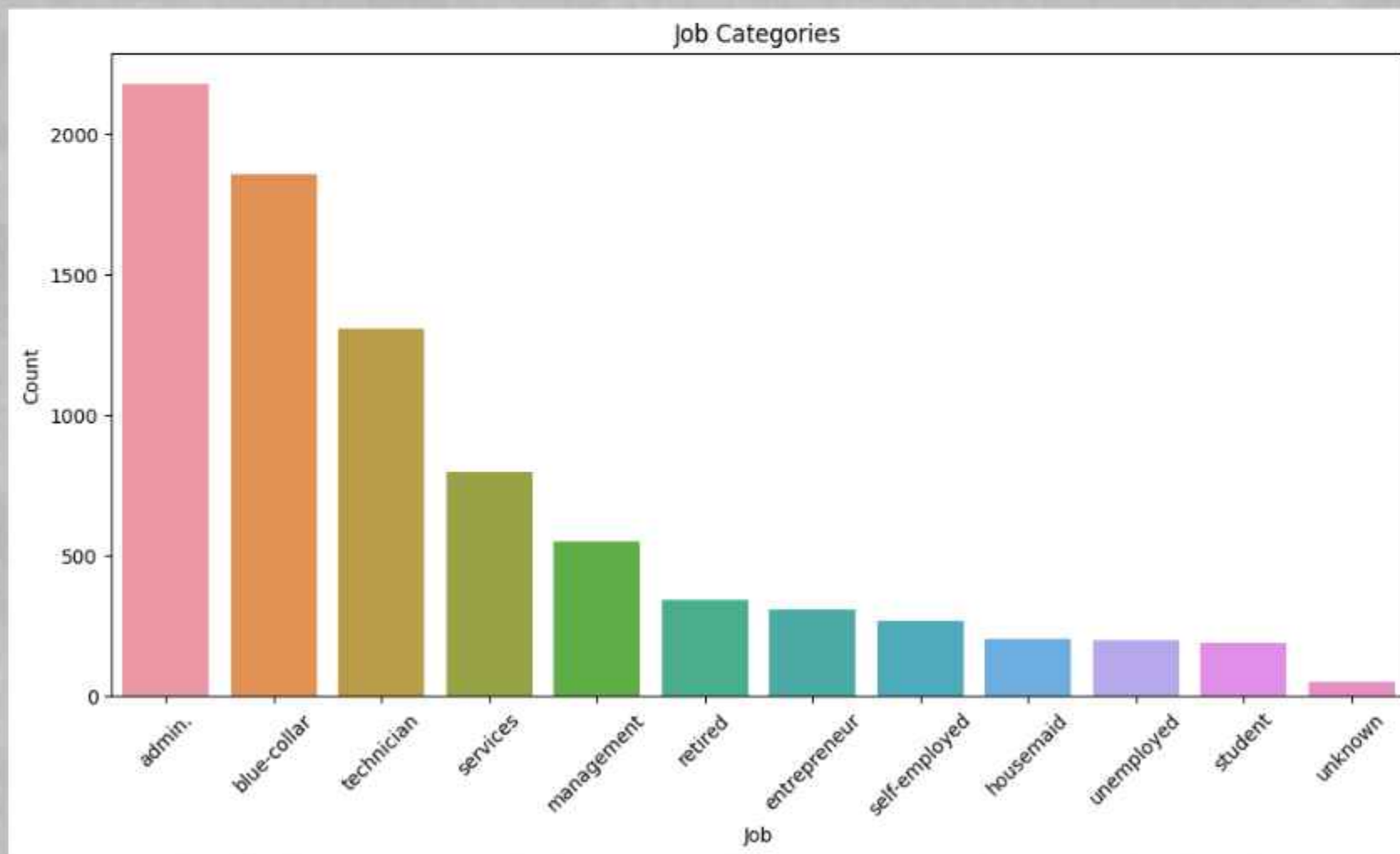
- **previous**

- **previous**

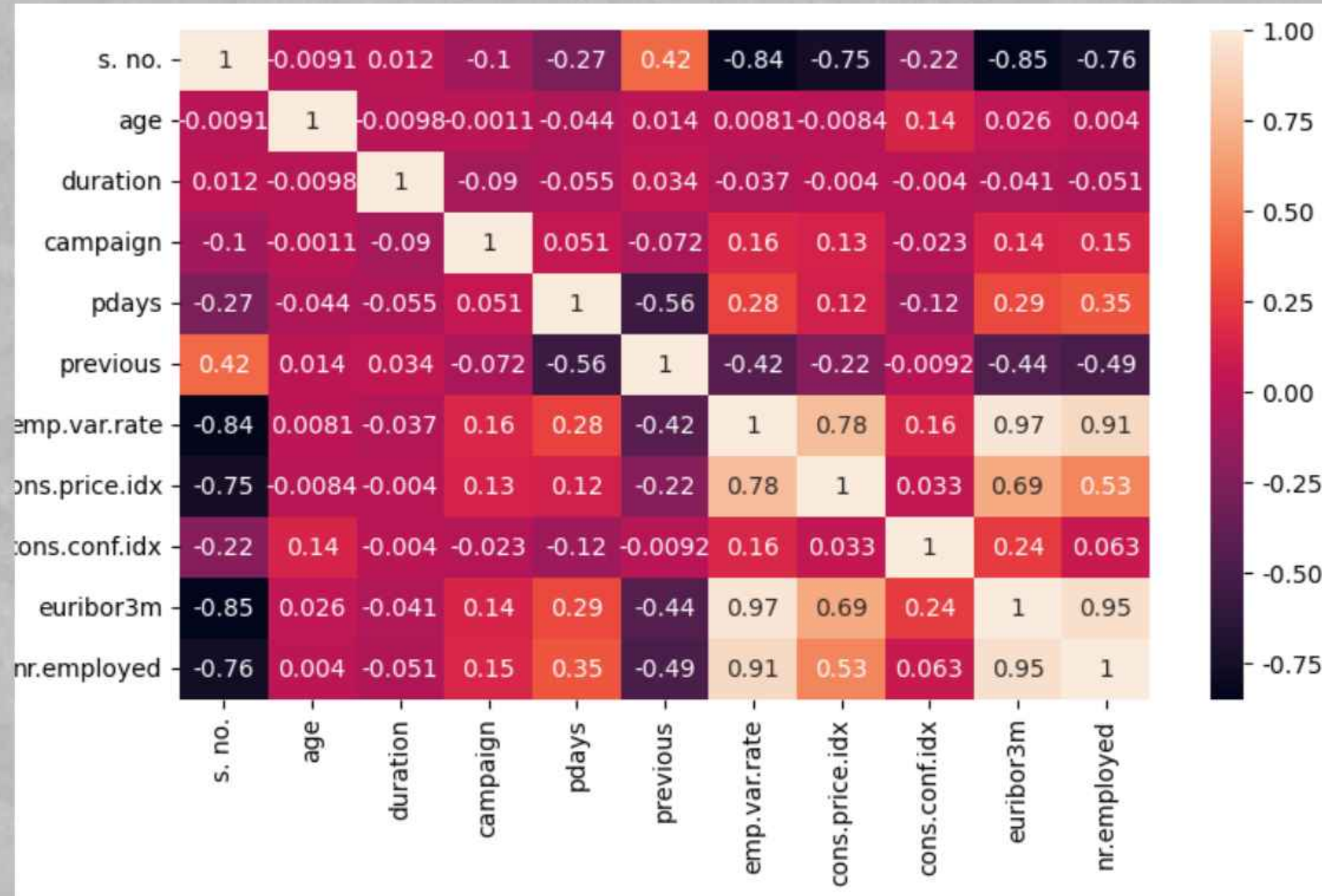
- **cons.price.idx**

- **cons.conf.idx:**

- **target**



DATA EXPLORATION



- Age positively correlates with consumer confidence but negatively with employment variation, impacting older individuals more.
- Previous campaign contacts negatively affect current campaign contacts, reflecting changing engagement.
- Employment variations correlate strongly with consumer prices, suggesting interdependence.
- Euribor 3-month rate and employee count are closely connected.
- Employment variations link to consumer confidence via economic factors.

CLUSTERING

hierarchical clustering

- Hierarchical clustering highlights distinct patterns within the dataset, aiding in the identification of underlying structures.
- It assists in determining the optimal number of clusters, as demonstrated by the dendrogram and agglomeration method's agreement on four clusters.
- The relationships between features and clusters are unveiled, such as the influence of longer call durations on subscription choices.
- Hierarchical clustering offers both manual and automated methods, providing flexibility in data analysis.

K-means clustering

- Individuals within the same cluster are considered similar based on the specified features (pdays, duration, y_binary).
- The clusters are created in such a way that individuals within a cluster are more similar to each other than to individuals in other clusters.

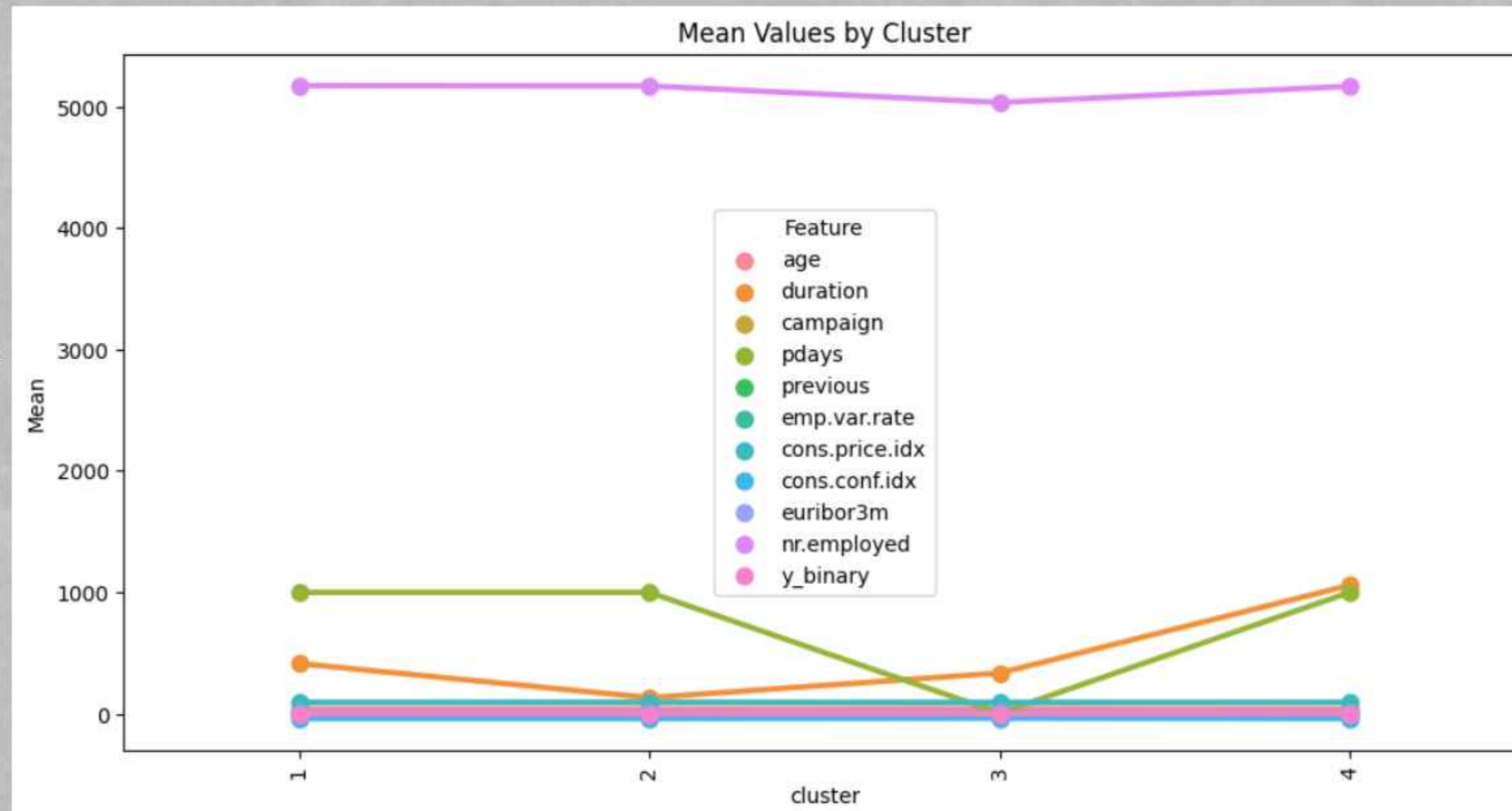
kNN Function

- The KNN model consistently predicts a class label of 1 for the given data point. This shows that the consumer subscribed to the program, data point is closest to instances in the dataset with a class label of 1.
- According to the KNN model, the given data point is more likely to be similar to dataset instances where users have paid for program subscriptions. This suggests that the given data item demonstrates traits or features that are more important among clients whose decision to sign up for the service. It's important to remember that the choice of k and the particular features of the dataset can affect how accurate the predictions are. To get more accurate forecasts, the model may need to be further examined and adjusted.

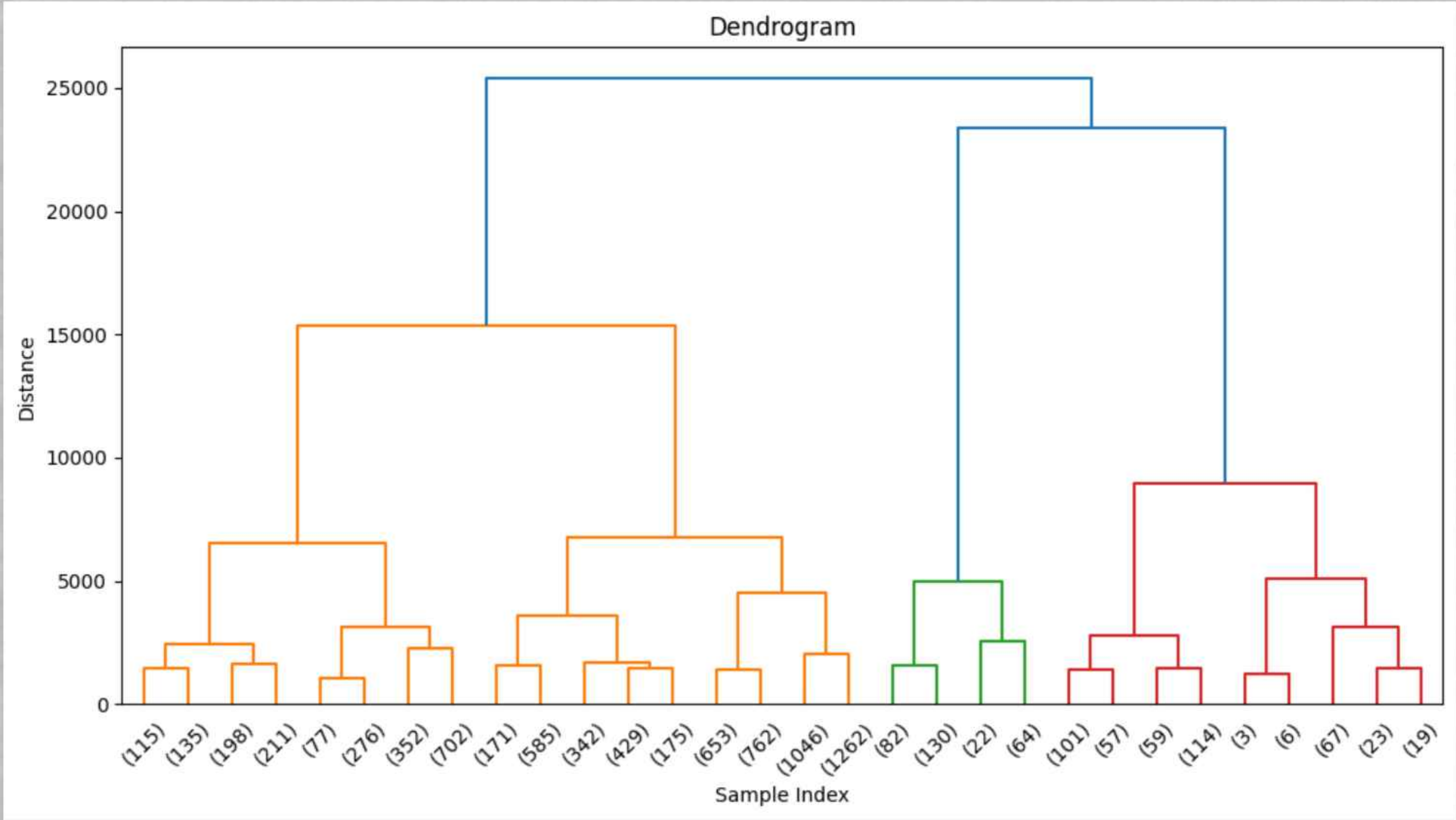
HIERARCHICAL CLUSTERING

Key Points

- Hierarchical clustering identified distinct customer segments within the dataset based on similarities in their characteristics and behavior.
- The results demonstrated that customers can be grouped into four main clusters, with each cluster exhibiting unique features related to call duration and subscription behavior.
- This clustering analysis provides valuable insights for targeted marketing efforts, allowing the bank to tailor its strategies to different customer segments and increase the effectiveness of marketing campaigns.



DENDROGRAM



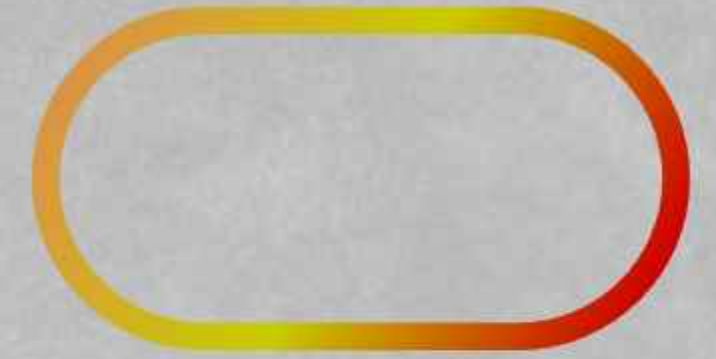


K-MEANS CLUSTERING

-
- Individuals in cluster 0 have relatively higher values for 'duration' and 'y_binary' and have been contacted recently (pdays is not 999).
 - Individuals in cluster 1 have higher values for 'pdays', indicating they have not been contacted recently.

-
- Individuals in cluster 2 have lower 'duration' and 'pdays' values.
 - Individuals in cluster 3 have moderate values for 'duration' and 'pdays'.

KNN FUNCTION



-
- The K-Nearest Neighbors (KNN) model consistently predicts a class label of 1 for different values of k for the provided data point.
 - This suggests that the data point shares similarities with instances in the dataset where customers subscribed to the program.

-
- The choice of k can influence prediction accuracy and should be selected based on dataset characteristics and model performance.
 - Further model tuning and analysis are required to improve prediction accuracy and gain more insights from the data.



CONCLUSION

- Clustering techniques revealed distinct groups with varying characteristics, aiding in the identification of potential target segments. Descriptive visualizations and correlation analyses illuminated patterns and relationships among features, guiding marketing strategies.
- The implementation of hierarchical clustering, K-means, and KNN demonstrated the versatility of clustering and classification methods in understanding and predicting customer responses.
- The study's outcomes can empower the banking institution, represented by BankBoostX, to optimize marketing strategies, allocate resources effectively, and enhance customer engagement.

