Unsupervised Learning on Bank Marketing Data

Elena Conderana & Sergio Cuenca

Table of contents

01

Introduction

About the dataset

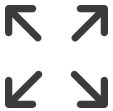
Demographic data, past marketing campaign responses

Complex relationships between consumer preferences and behavior

Influences in consumer engagement and loyalty

Problem vs Solution

Improving the efficacy of bank marketing initiatives



One-size-fits-all

tactics

VS

More sophisticated,

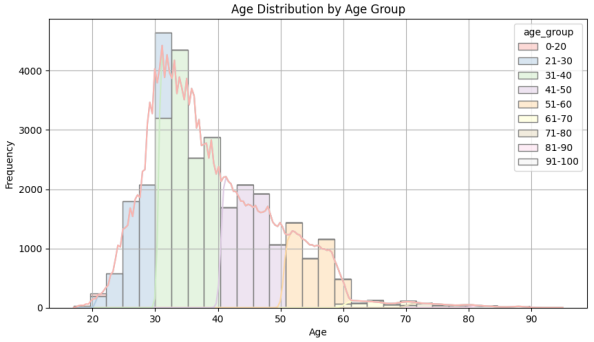
data-driven approaches

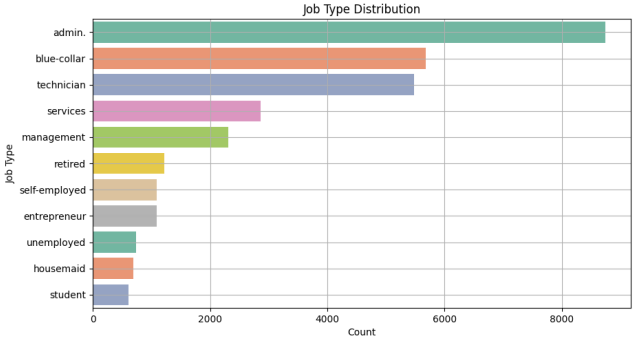
Table of contents

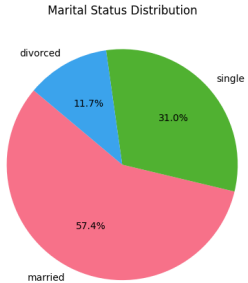
02

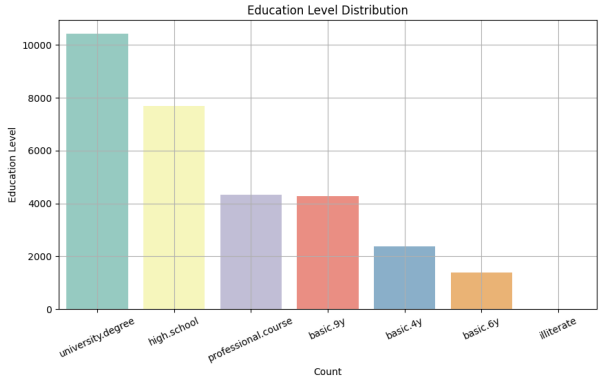
Exploratory Data Analysis (EDA)

Insights Gained from Exploration



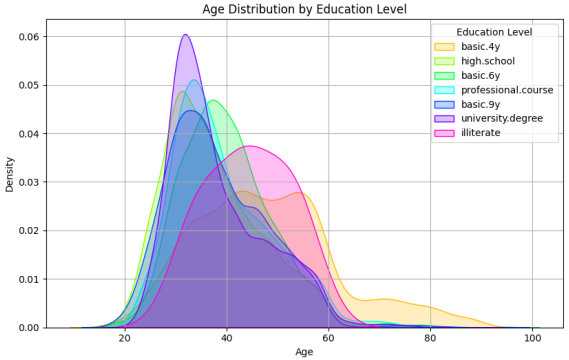


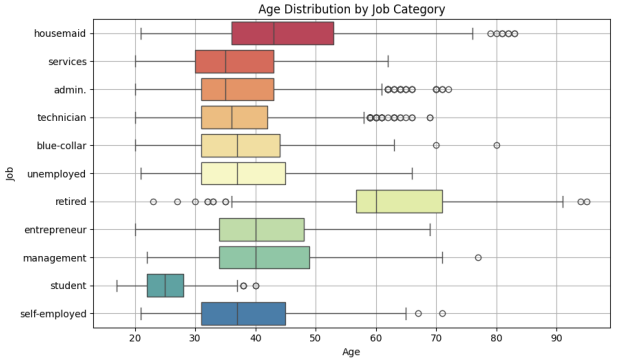


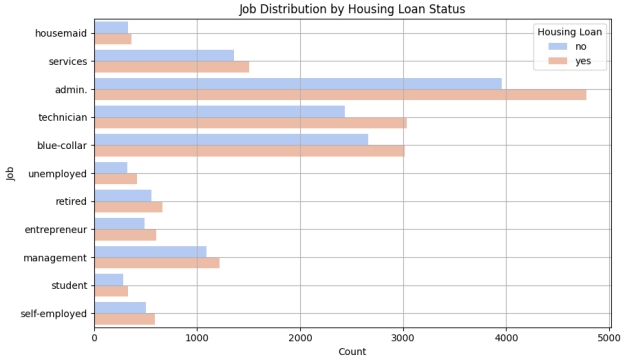


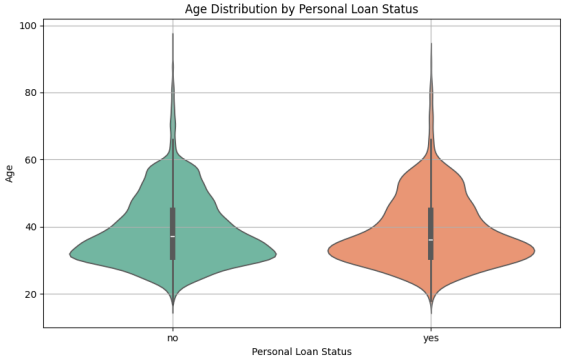
- The prevailing age group ranges from 30 to 40 years of age
- There are three types of job considerably more frecuent among the targeted audience (administrators, blue-collar workers and technicians)
- The 25% and 75% percentile of the three jobs encaptures the 30 to 40 age group wholly. The same situation occurs with the education degree
- The majority of the audience is in possesion of either a university degree or have finished high school

Insights Gained from Exploration



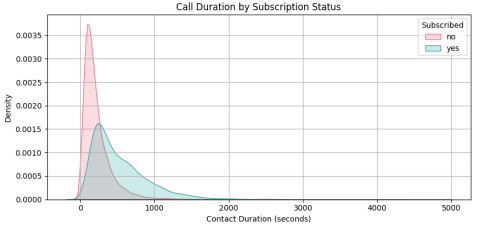


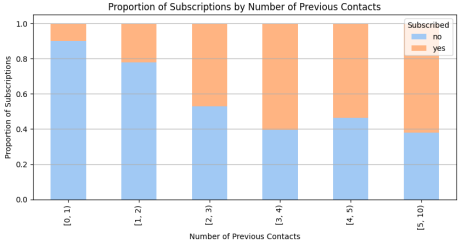


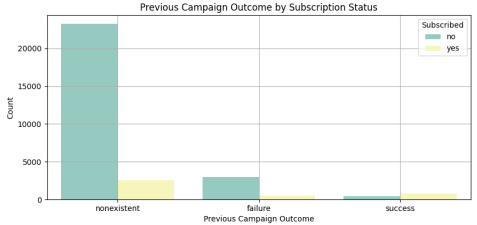


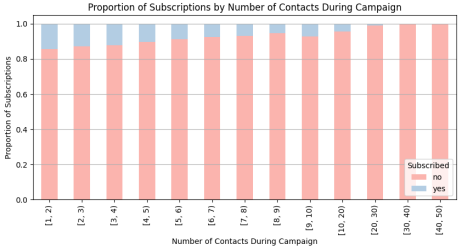
- Either house or personal loans, the distributions even out. There is a very similar amount of people that have taken a personal loan and who haven't.
- House loans are equally frequent independent of the employment of each customer. There is evidently a higher density of house loans among technicians, administrators and blue-collar workers.
- The prevailing trend is that more people do have a house loan.

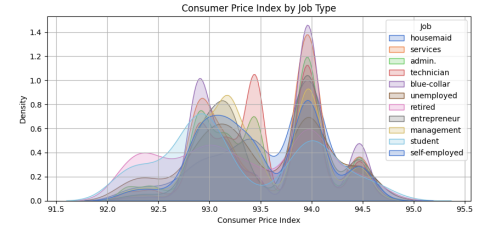
Insights Gained from Exploration

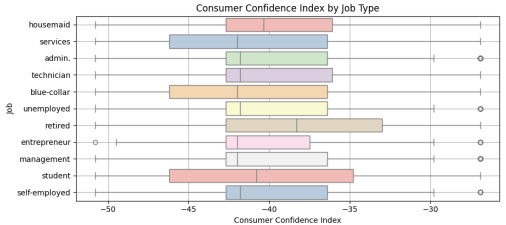












- During the time previous to the campaign the more times the audience was contacted the more subscriptions there tended to be. This trend utterly switches during the campaign. During that time the more calls a person received, the less likely he was to subscribe a new deposit
- Before the campaign 3 calls was an optimum amount to try to get a person to subscribe. During the campaign, however, if the second call had not succeeded it would be advisable to drop that customer and move towards the next
- Calls with subscribed members or people that ended up subscribing have a tendency to have a longer duration

Table of contents

03

Principal Component Analysis (PCA)

Encoding

All features must be in numeric format. In order to encode all columns three different methods are used:

Ordinal

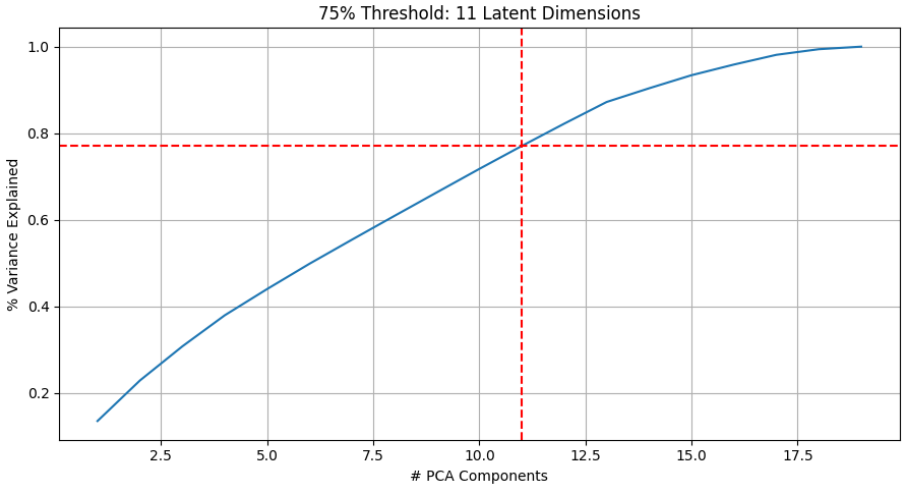
Binary

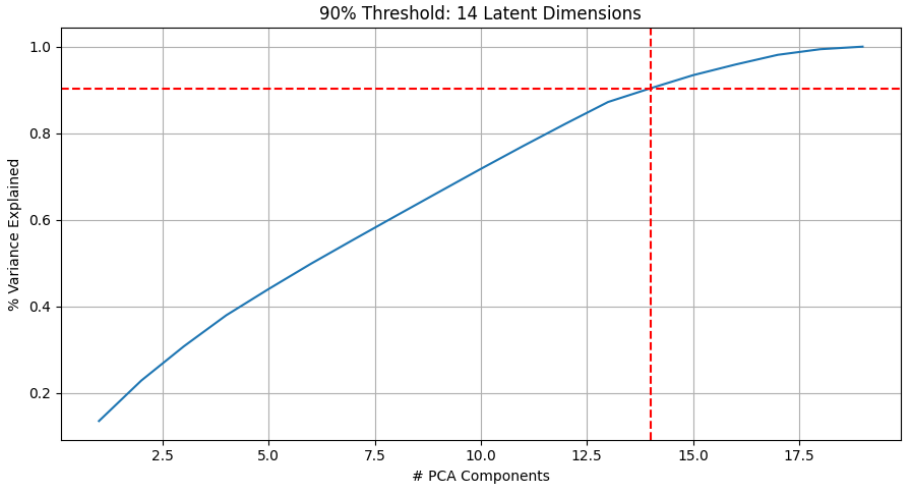
Nominal

Kaisser Rule vs Explained Variance

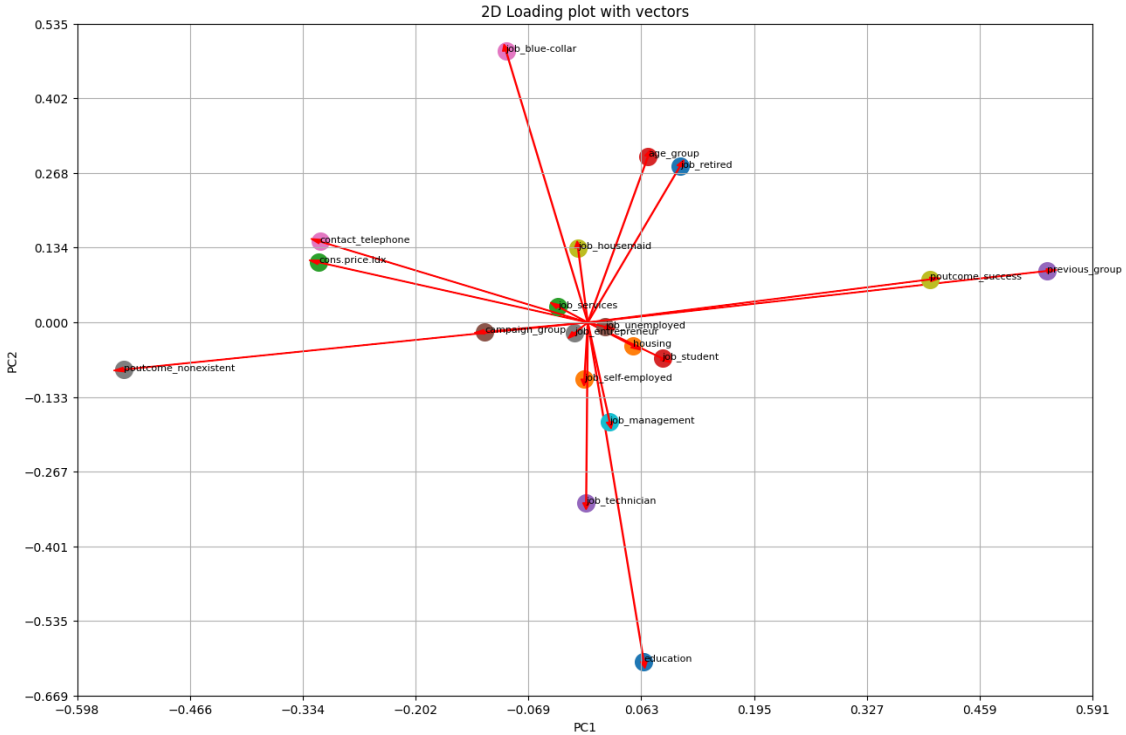
Kaisser Rule: 11 Latent Dimensions

VS





Important Variables



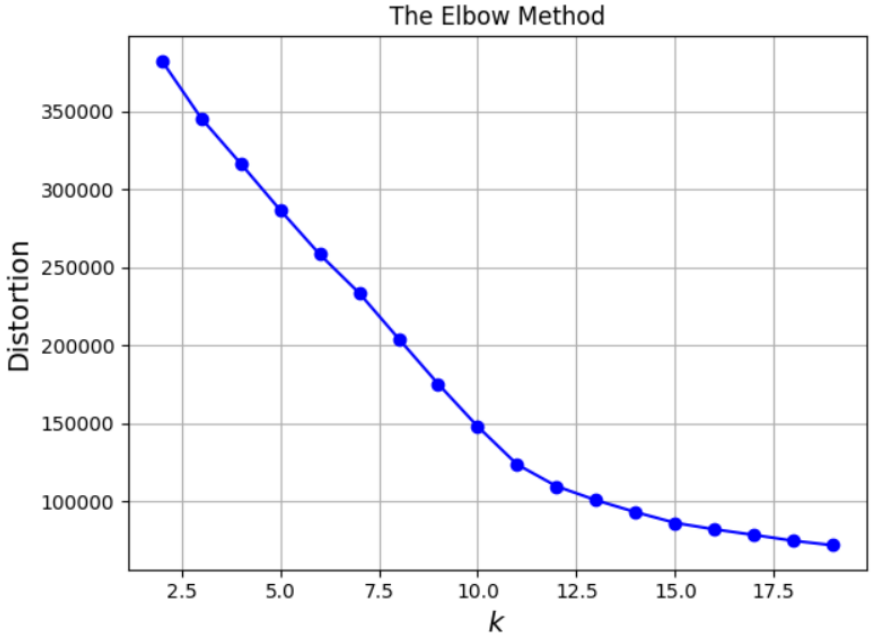
- Age_group, job_retired, job_technician, job_blue-collar, education poutcome_nonexistent, poutcome_success and previuous_group hold the highest variance of the information.
- In general, features are fairly orthogonal between each other and have large projections.

Table of contents

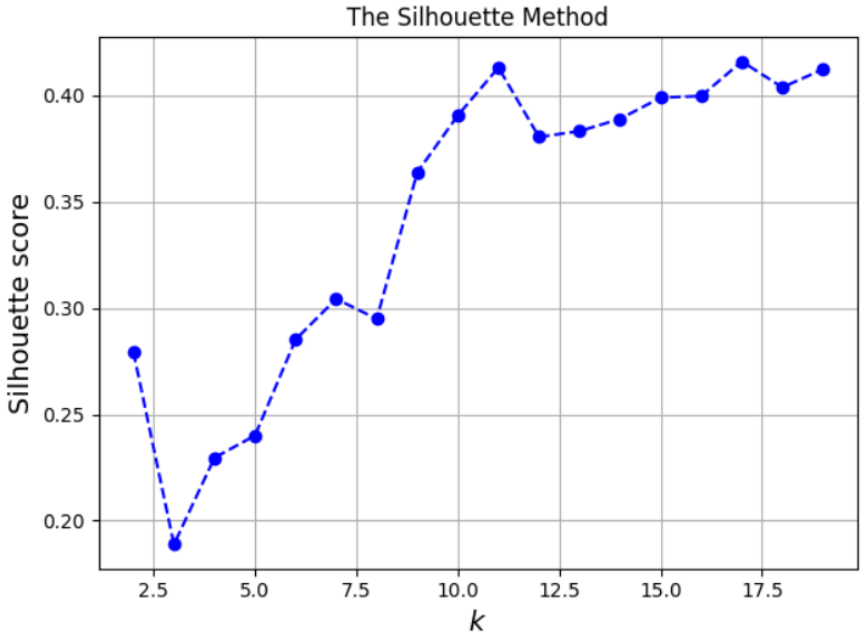
04

Clustering Analysis

K-Means

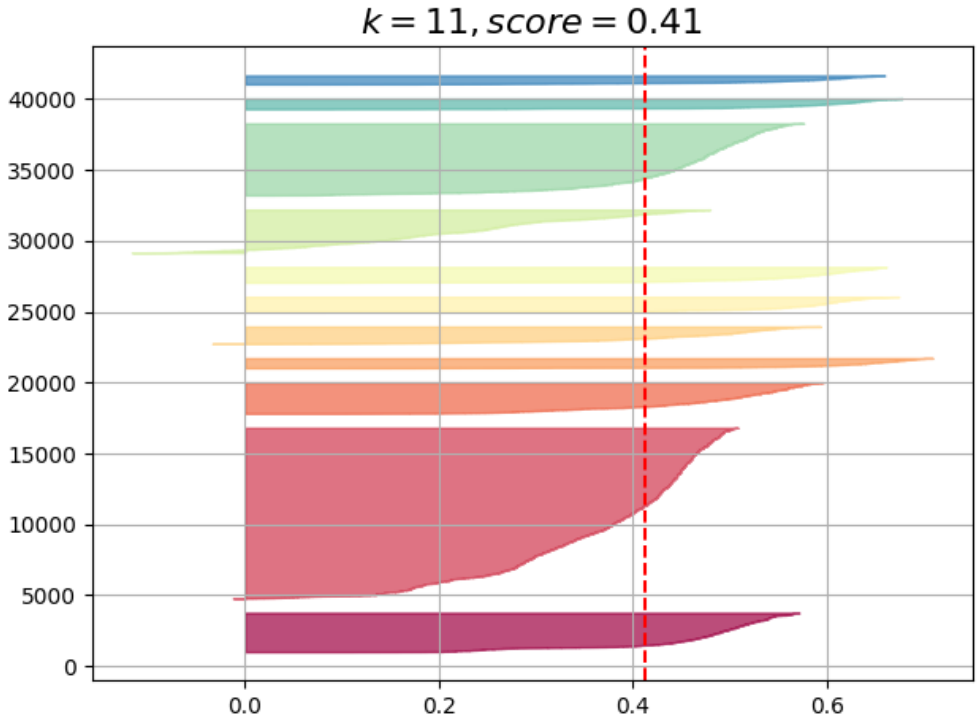


Change in slope to indicate a potentially interesting number of clusters (11)



The higher the final coefficient, the more optimal the number of clusters (11)

K-Means Silhouette Diagram



Measure how close each point in one cluster is to points in the neighboring clusters.

- Clusters 2 and 9 exhibit the highest concentration of clients, thereby constituting the largest clusters. Clusters 1, 3, and 10 also manifest a significant aggregation of clients, whereas the remaining clusters are comparatively smaller and exhibit a degree of homogeneity.
- The proximity measure for all clusters is approximately greater than 0.5, suggesting that the data points are substantially distant from those in neighboring clusters, a condition that is deemed favorable.
- Cluster 3 may be considered the least favorable, as it demonstrates a negative skew and possesses the lowest silhouette score. Conversely, Cluster 7 is potentially the most favorable, denoted by its attainment of the highest silhouette score.

Cluster Distribution

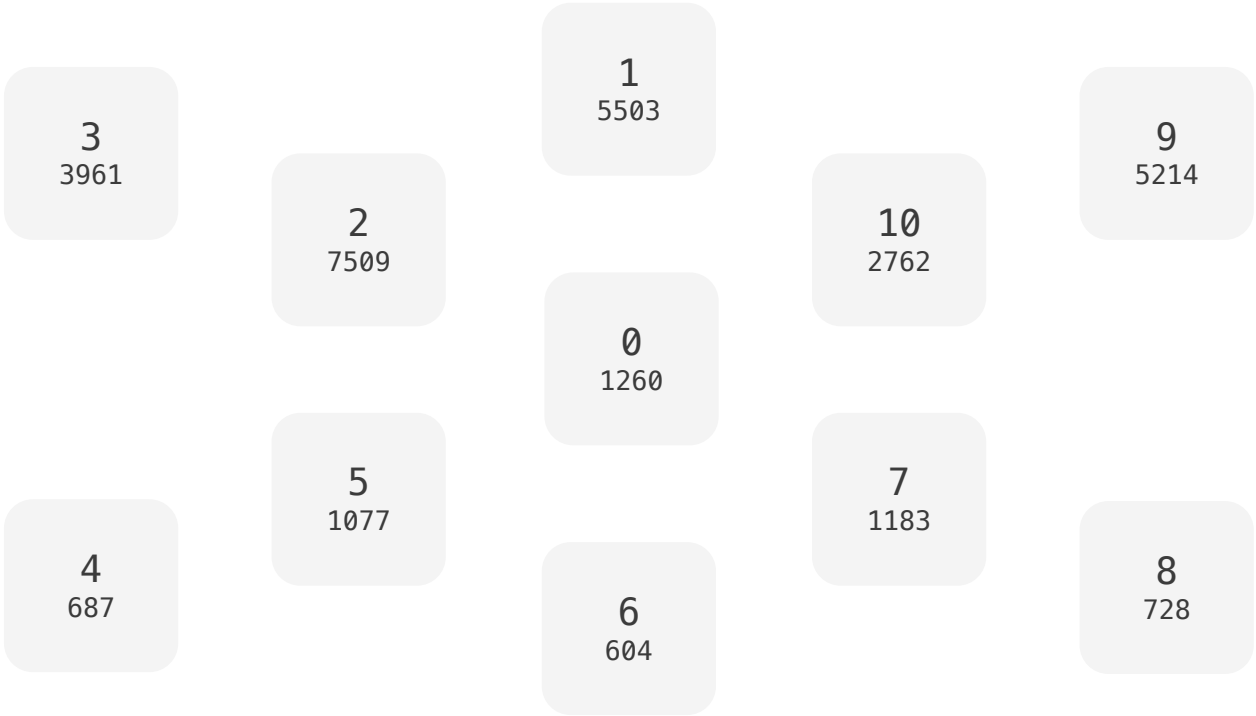


Table of contents

05

Conclusions

Conclusions

PCA

This technique minimized the dimensionality of the dataset while preserving the majority of its variance. As a result, the dataset became more straightforward to examine and visualize

- Set of features containing majority of information variance
- Less dimensions conclude in better results

Clustering

This technique enabled to divide the bank's customer base into discrete clusters. The bank might potentially increase the campaign's success rate by customizing its marketing techniques

- Latent common features among observations
- Imbalanced distribution of observations

“Unsupervised learning is where you only have input data and no corresponding output labels. The system tries to learn without a teacher.”

Yann LeCun