Create Customer Segmentation Report for Arvato Financial Services

Project Overview

First of all creating customer segmentation and comparing to general population, then find which more likely to be customer. after that I used different method to preprocess the data, and then used unsupervised learning techniques Principle Components Analysis (PCA) to perform customer segmentation and identify customer of the company and use models to predict which individuals convert to be customers.

missing data in each column

A screenshot of a cell phone

Description automatically generated

Missing data in each row

A screenshot of a cell phone

Description automatically generated

Compare distribution

A screenshot of a cell phone

Description automatically generated

Apply PCA

Variance explained by each component

A screenshot of a cell phone

Description automatically generated

Cumulative Variance Explained

A screenshot of a cell phone

Description automatically generated

Apply k-mean

Compute the average within-cluster distances

A close up of a map

Description automatically generated

Compare customer vs General clusters

A screenshot of a cell phone

Description automatically generated

Conclusion:

What we find is PCA and Cluster can help to reduce the number of columns when we use huge dataset. It helps to ease the process of supervised and unsupervised learning. Data preprocessing is very important to ease the analysis.