

Bertelsmann Arvato – Customer segmentation

Machine Learning Engineer Udacity Nanodegree - Capstone Project

This repository contains code and report for "Capstone Project - Arvato Customer Segmentation" done as part of my Udacity Machine Learning Engineer Nanodegree program.

Table of Contents

- [Project Overview](#)
 - [Data Description](#)
 - [Requirements](#)
 - [Results](#)
-

Project Overview

As part of this project, I am analyzing demographic data of German population and the customer data in order to perform Customer Segmentation and Customer Acquisition.

This project is to help a Mail-Order company to acquire new customers to sell its organic products. The goal of this project is to understand the customer demographics as compared to general population in order to decide whether to approach a person for future products.

There are three parts to project.

1. Data cleaning
2. Unsupervised learning based segmentation of data using PCA and K-means
3. Supervised model to fit the supplied training set

Data Description

The data has been provided by Udacity and Arvato Financial Solutions. The dataset contains 4 data files and 2 description files. The description files have information about the features and their explanation. The 4 data files include:

- Customer Segmentation
 - General Population demographics
 - Customer demographics
- Customer Acquisition
 - Training data
 - Test data

Please note that data files are not included in capstone submission because

- These are standard files provided by Udacity so no external data is used.
- Files are too big and including them violates project zip submission guidelines (500MB uncompressed).

Requirements

The project uses following utilities

- Numpy
- Panda
- Sklearn
- Seaborn data visualization

There are no other external software packages used for the project.

Results

The results have been clearly documented in the Jupyter Notebook. Please refer to associated "Project Workbook.ipynb" file.