Project Overview

In this project you will apply unsupervised learning techniques on product spending data collected for customers of a wholesale distributor in Lisbon, Portugal to identify customer segments hidden in the data. You will first explore the data by selecting a small subset to sample and determine if any product categories highly correlate with one another. Afterwards, you will preprocess the data by scaling each product category and then identifying (and removing) unwanted outliers. With the good, clean customer spending data, you will apply PCA transformations to the data and implement clustering algorithms to segment the transformed customer data. Finally, you will compare the segmentation found with an additional labeling and consider ways this information could assist the wholesale distributor with future service changes.

Project Highlights

This project is designed to give you a hands-on experience with unsupervised learning and work towards developing conclusions for a potential client on a real-world dataset. Many companies today collect vast amounts of data on customers and clientele, and have a strong desire to understand the meaningful relationships hidden in their customer base. Being equipped with this information can assist a company engineer future products and services that best satisfy the demands or needs of their customers.

Things you will learn by completing this project:

- How to apply preprocessing techniques such as feature scaling and outlier detection.
- How to interpret data points that have been scaled, transformed, or reduced from PCA.
- How to analyze PCA dimensions and construct a new feature space.
- How to optimally cluster a set of data to find hidden patterns in a dataset.
- How to assess information given by cluster data and use it in a meaningful way.

NEXT

Overview