**D212 PA**

Jon Isak

D212: Data Mining II

Contents

[**D212 Task 2 3**](#_3xni6yjcaxld)

[**Part I: Research Question 3**](#_rc29ops3t4ge)

[**Part II: Method Justification 3**](#_wqk7iynz2hav)

[**Part III: Data Preparation 4**](#_nncaoeoy6o4n)

[**Part IV: Analysis 4**](#_5vl7hmvmrk07)

[**Part V: Attachments 4**](#_qiqrs0a7mir8)

## 

## D212 Task 2

## Part I: Research Question

A1. This data analysis will investigate the question, can we reduce the dimensionality of the dataset by identifying principal components that explain the most variance?

A2. The objective of this data analysis is to utilize PCA to uncover underlying patterns and extract principal components from the specific set of continuous variables within the churn dataset. The focus is on identifying those principal components that significantly contribute to understanding customer behavior, preferences, and demographics.

## Part II: Method Justification

B1. Principal Component Analysis is chosen for analyzing the selected dataset due to its ability to manage high-dimensional data by reducing its dimensionality while retaining most of the variability present in the original data. PCA works by transforming the original variables into a new set of uncorrelated variables called principal components. The expected outcomes of PCA are a reduction in variables while retaining most of the information in the data. Identification of principal components that represent patterns in the data and which variables contribute the most variability (Analyticsvidhya, 2023)

B2. One assumption of PCA is that it assumes a linear relationship among variables (Laerd, n.d.).

## Part III: Data Preparation

C1. My continuous data set variables are "Outage\_sec\_perweek", "Yearly\_equip\_failure", "Tenure", "MonthlyCharge", and "Bandwidth\_GB\_Year".

C2. The file is attached.

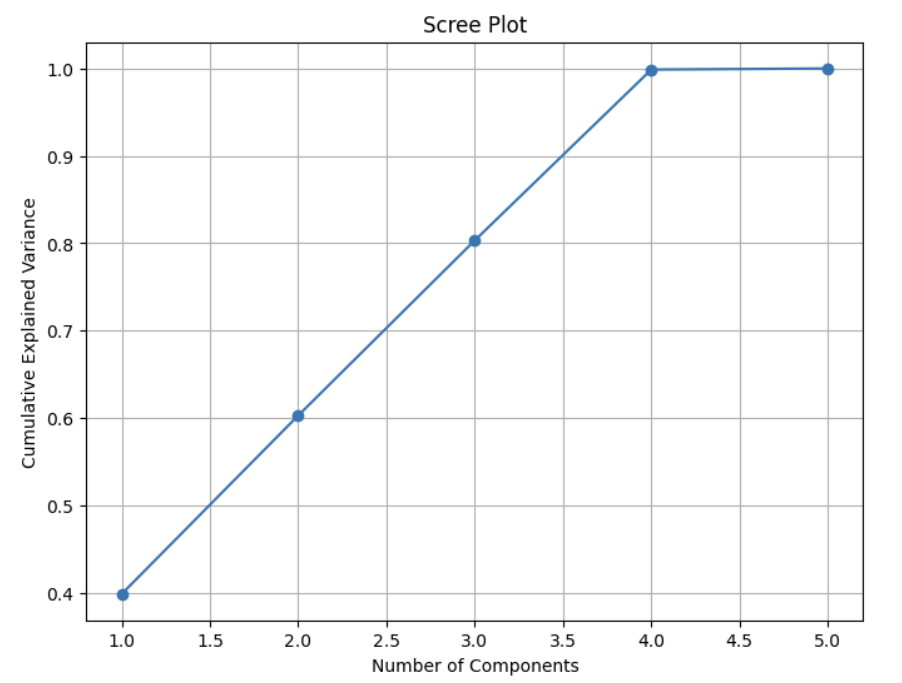
## Part IV: Analysis

D1. The screenshot includes the matrix.

A screenshot of a computer

Description automatically generated

D2. Based on kaiser criterion, the total number of principal components is three.



D3. There were three principal components with a variance. PC 1 has a variance of 1.99. PC 2 has a variance of 1.02, PC 3 has 1.001.

D4. The total variance captured by principal components is 4.017.

D5. Using the kaiser criterion, the number of principal components was identified as three. These three components should explain the majority of the variability in the data and represent the relationships within the data. The total variance captured by principal components is 4.017. The three components identified from the analysis can be used to group customers into segments. Through further analysis, the company can develop targeted marketing strategies for these segments.

## Part V: Attachments

E.   
 Jakevdp. (n.d.). In Depth: Principal Component Analysis | Python Data Science Handbook. Retrieved December 12, 2023, from https://jakevdp.github.io/PythonDataScienceHandbook/05.09-principal-component-analysis.html

Pani, A. (n.d.). *Clustering with PCA*. Kaggle. Retrieved December 12, 2023, from https://www.kaggle.com/code/aninditapani/clustering-with-pca

F.

Analyticsvidhya. (2023, September 13). Principal Component Analysis | PCA Explained. Analytics Vidhya. Retrieved December 12, 2023, from https://www.analyticsvidhya.com/blog/2016/03/pca-practical-guide-principal-component-analysis-python/

Laerd. (n.d.). *How to perform a principal components analysis (PCA) in SPSS Statistics*. Laerd Statistics. Retrieved December 12, 2023, from https://statistics.laerd.com/spss-tutorials/principal-components-analysis-pca-using-spss-statistics.php