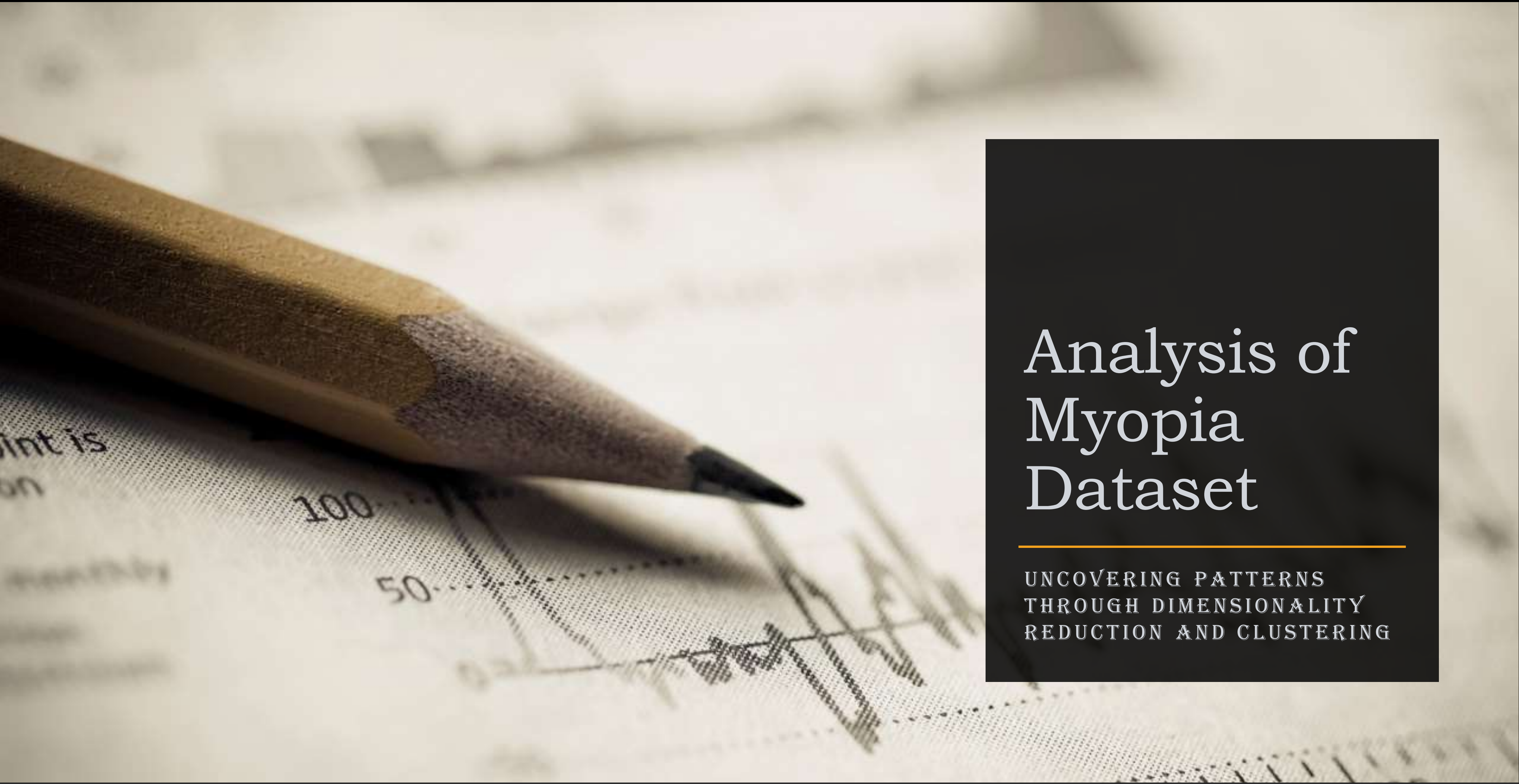
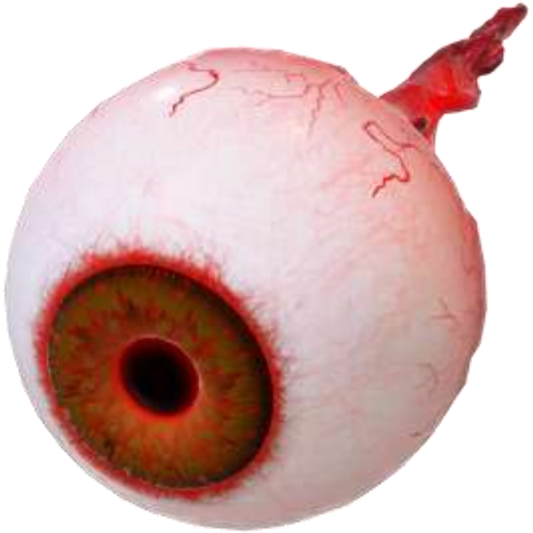


Analysis of Myopia Dataset

UNCOVERING PATTERNS THROUGH DIMENSIONALITY REDUCTION AND CLUSTERING

Background

- The Myopia dataset is focused on improving the prediction of myopia, or nearsightedness. There have been previous attempts to enhance classification models on the dataset, but that has proven futile.
- Nonetheless, we believe that distinct groups of patients may exist, therefore warranting our separate analysis.
- We'll explore this possibility using:



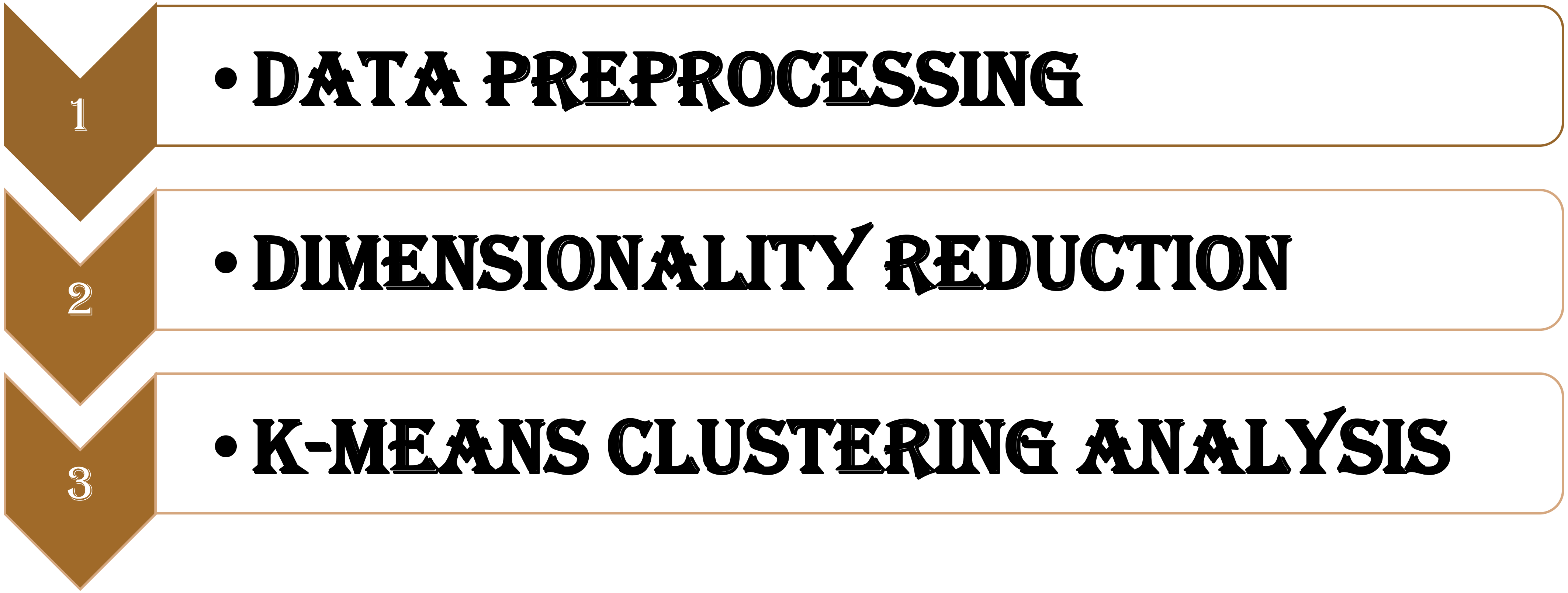
UNSUPERVISED MACHINE LEARNING TECHNIQUE

K – MEANS CLUSTERING

Objective

The primary goal is to identify if there are distinct patient groups within the myopia dataset. This exploration serves as the foundation for our aim to refine the analysis by focusing on subgroups that may exhibit different patterns or characteristics.

Approach



1

- DATA PREPROCESSING

2

- DIMENSIONALITY REDUCTION

3

- K-MEANS CLUSTERING ANALYSIS

Data Preprocessing

FEATURE SCALING

- Ensure that all features in the myopia dataset are on a consistent scale. This is crucial for the proper functioning of machine learning algorithm.

STANDARD SCALER

- **Consistent Scale:** Standard Scaler standardizes features by removing the mean and scaling to unit variance.
- Enhanced model performance

DIMENSIONALITY REDUCTION

PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA efficiently reduced the dimensionality of the myopia dataset.

Retained essential patterns and relationships within the data.

T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (T-SNE)

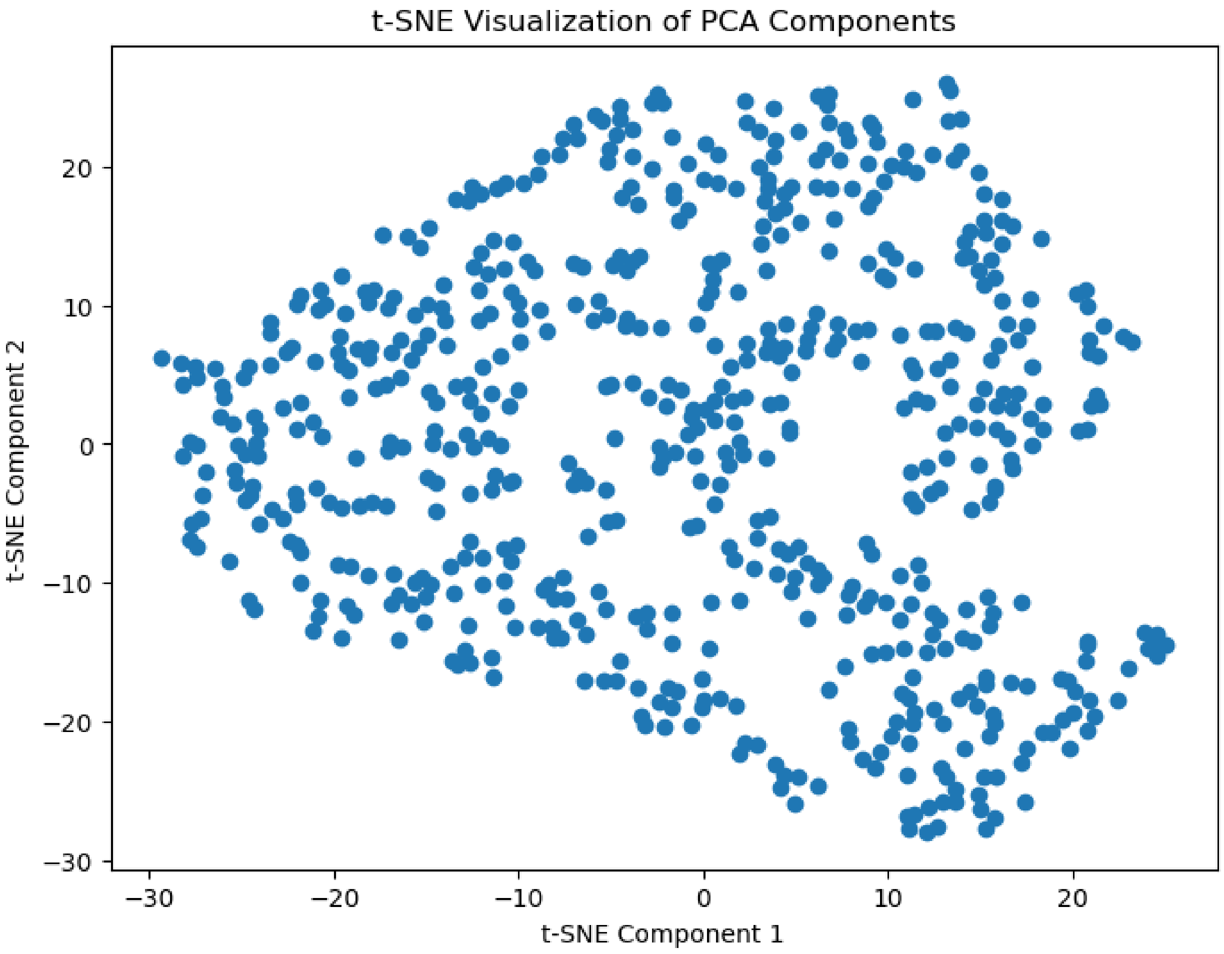
Continuation from PCA: Building upon PCA results, t-SNE provided a deeper exploration of the myopia dataset's inherent structures.

Preparation for Clustering: Set the stage for subsequent clustering analyses by revealing potential clusters or subgroups.

Visualization

The t-SNE visualization of the myopia dataset contributed to a more nuanced understanding by emphasizing local structures and providing a visually insightful representation.

It however did not reveal any apparent clusters in the myopia dataset



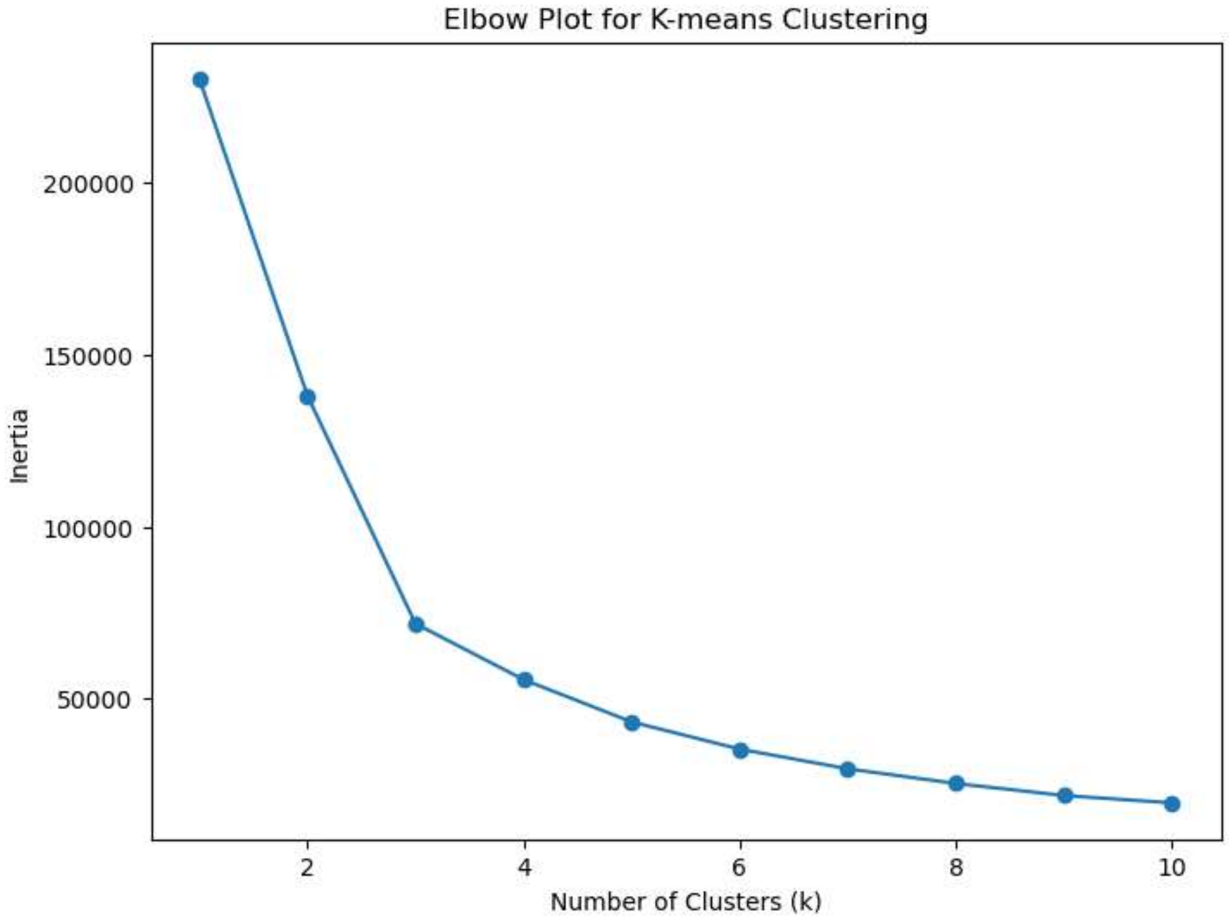
K-Means Clustering

- **Benefits:**
  - Identify natural groupings within the myopia dataset.
  - Enable the examination of distinct patient clusters, potentially revealing unique characteristics or patterns.
- **Elbow Method for Optimal k:**
  - Use the elbow method to determine the optimal number of clusters.
  - **Elbow Point:** Indicates a point where adding more clusters provides diminishing returns in explaining variance.
  - **Optimal k**: 3
- **Visualization of Clustered Data:**

A scatter plot was generated, displaying t-SNE features on the x- and y-axes.

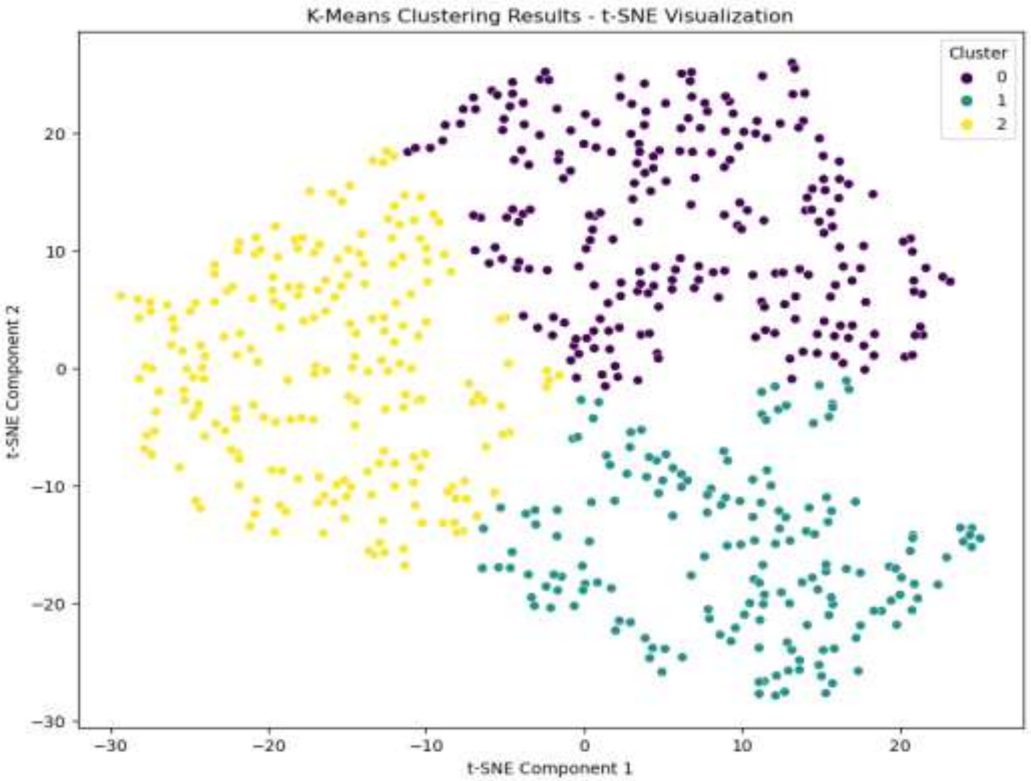
Elbow Plot

Optimal k = 3



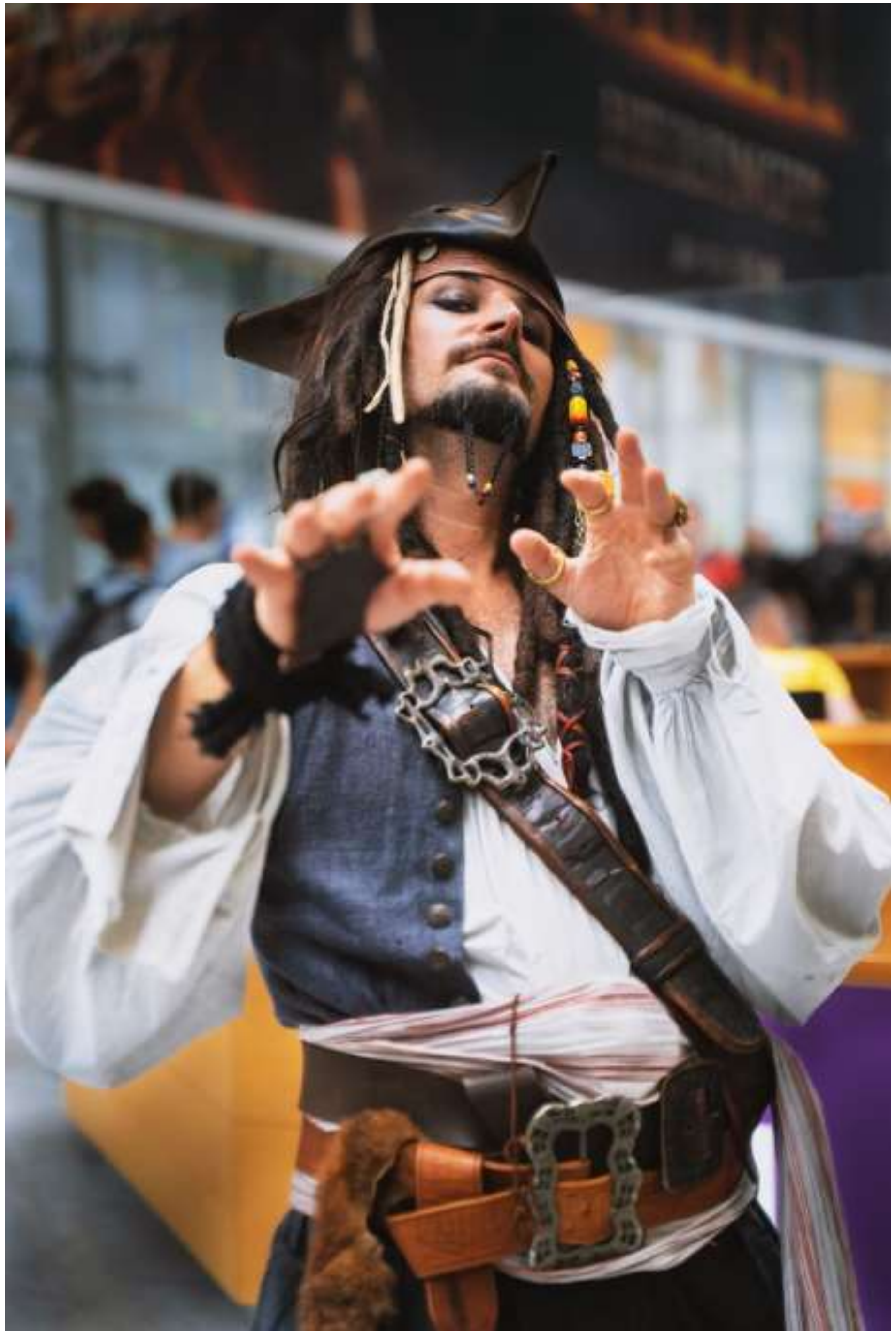
Visualization of the color-coded result

IMPLEMENT K-MEANS CLUSTERING WITH OPTIMAL K



Findings and Recommendations

- Successfully clustered the patients into three distinct groups.
- Based on the analysis, it is recommended to further explore and analyze the three identified clusters.
- Each cluster may represent a subgroup of patients with similar myopia-related characteristics.
- This refined understanding can lead to more targeted and effective strategies for predicting and managing myopia.



2024