****

**Assessment Report**

on

**“Movie Watch Pattern Clustering”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

In

**CSE(AI)**

By

Name : Nikhil Kumar Singh

Roll Number : 202401100300161

Section: C

**1. Introduction**

This project analyzes user behavior in movie-watching habits by clustering viewers based on three key features:

1. **Time of Watching** – The hour of the day when users watch movies (0-23).
2. **Genre Preference** – The most frequently watched genre (action, comedy, drama, thriller).
3. **Rating Behavior** – The average rating given by users (1-5 scale

**2. Problem Statement**

**Movie Watch Pattern Clustering**

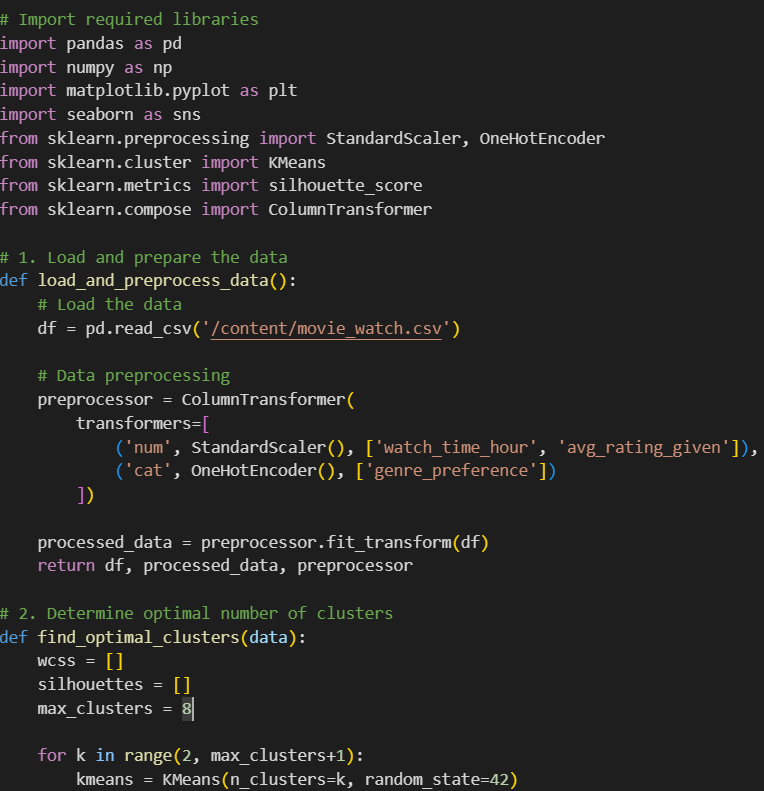
**Cluster users based on time of watching, genre preference, and rating behavior.**

**4. Methodoloy**

**The following steps were followed:**

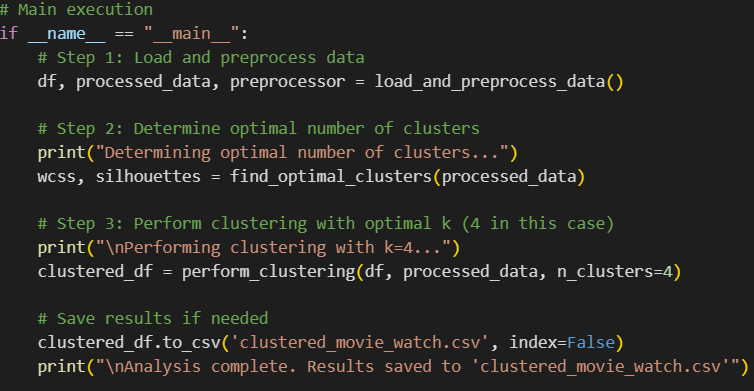
1. **Load and preprocess the dataset**
   * **Standardized numerical features (watch\_time\_hour, avg\_rating\_given)**
   * **Applied One-Hot Encoding to categorical feature (genre\_preference)**
2. **Determine the optimal number of clusters**
   * **Used Elbow Method and Silhouette Score for evaluation**
   * **Selected k = 4 as optimal number of clusters**
3. **Train the clustering model**
   * **Applied K-Means algorithm**
   * **Assigned each user to a cluster**
4. **Analyze clusters**
   * **Plotted distributions of clusters**
   * **Examined average values of features within each cluster**
   * **Visualized genre preferences by cluster**
5. **Save and export results**

**CODE:**

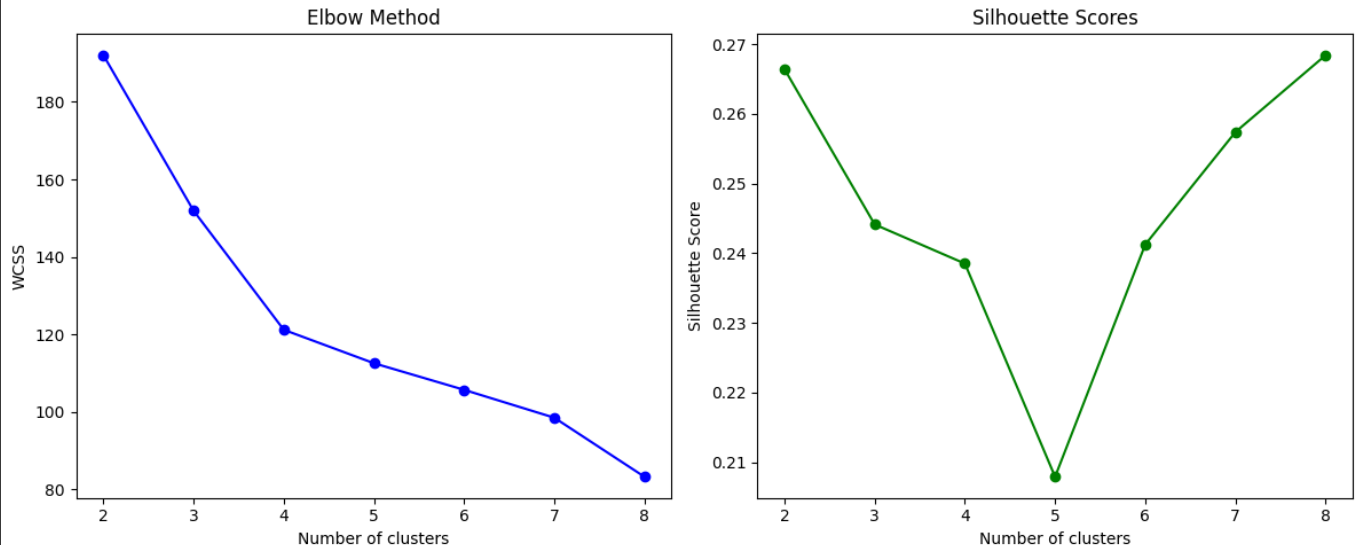


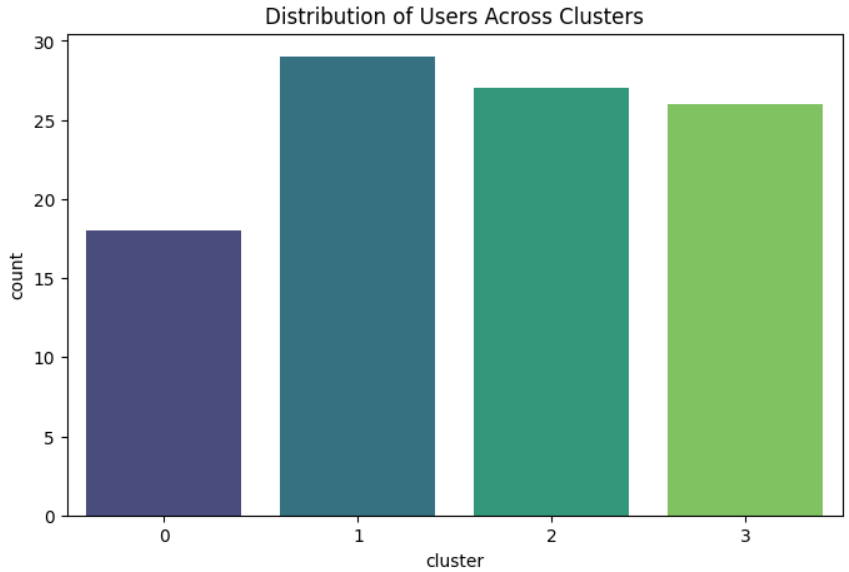


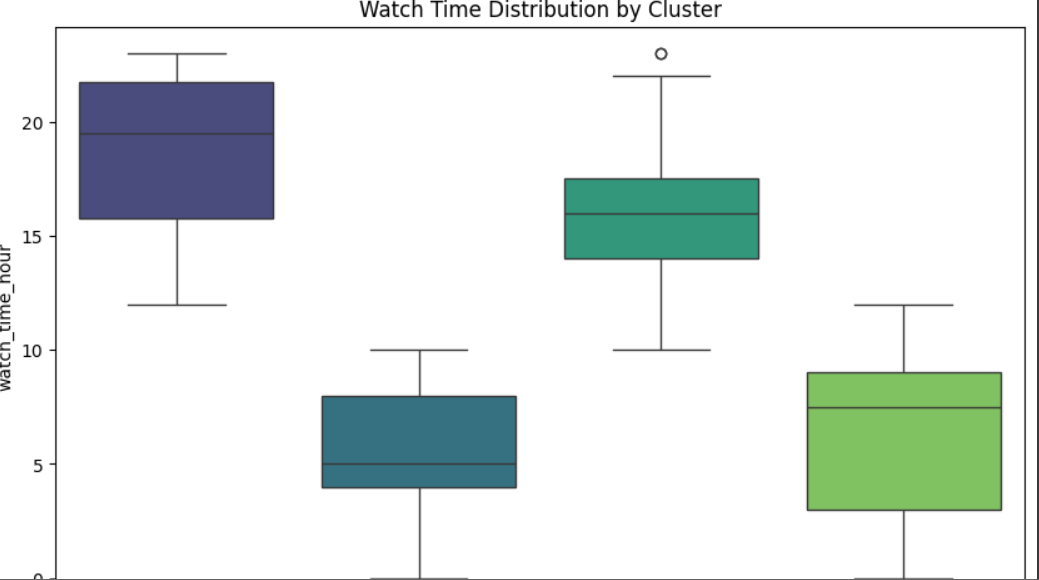


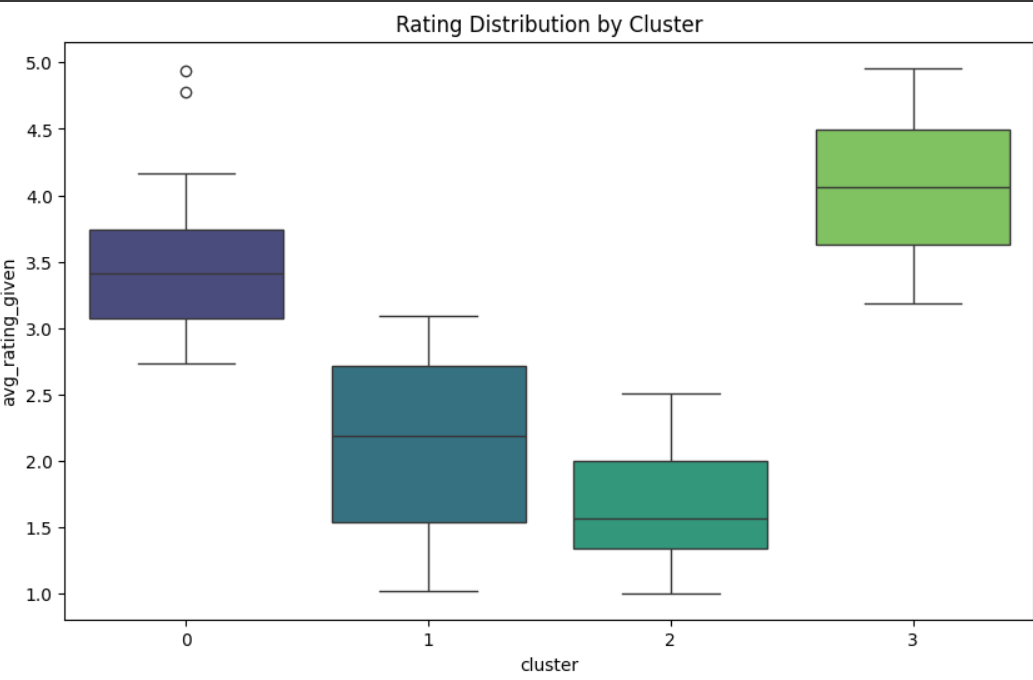


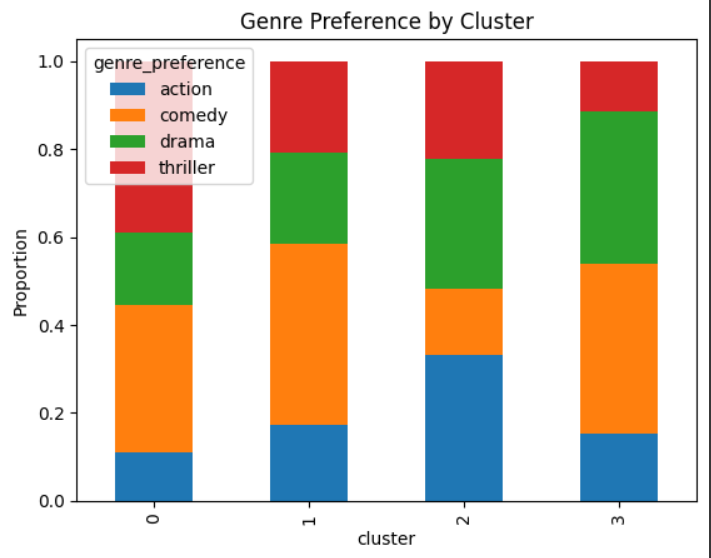
**OUTPUT:**











**REFRENCES:**

**1. Dataset provided for the exam**

**2. Scikit-learn documentation:** [**https://scikit-learn.org**](https://scikit-learn.org)

**3. Matplotlib and Seaborn for visualization**