****

**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 : Praveer Singh

Roll NO. : 202401100300181

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**

**. Data Understanding**

We started with a dataset containing:

* watch\_time\_hour: The hour of the day when a movie was watched.
* genre\_preference: The user’s preferred genre (text).
* avg\_rating\_given: Average rating the user gives to movies.

**a. Convert watch\_time\_hour to Time Blocks**

To make analysis more meaningful, the 24-hour format is grouped into:

* **Morning** (5–11)
* **Afternoon** (12–17)
* **Evening** (18–22)
* **Night** (23–4)

This captures *viewing behavior* better than using raw hour values.

**b. Encode Categorical Features**

* genre\_preference and time\_block are categorical.
* Used **Label Encoding** to convert them into numeric format for clustering.

**c. Scale Features**

* Standardized the features (mean = 0, std = 1) using StandardScaler.
* KMeans is distance-based, so scaling ensures fair contribution from all features.

**3. Clustering (KMeans)**

* Applied **KMeans Clustering** with 3 clusters (n\_clusters=3) to group users.
* The algorithm tries to minimize intra-cluster variance and maximize inter-cluster separation.

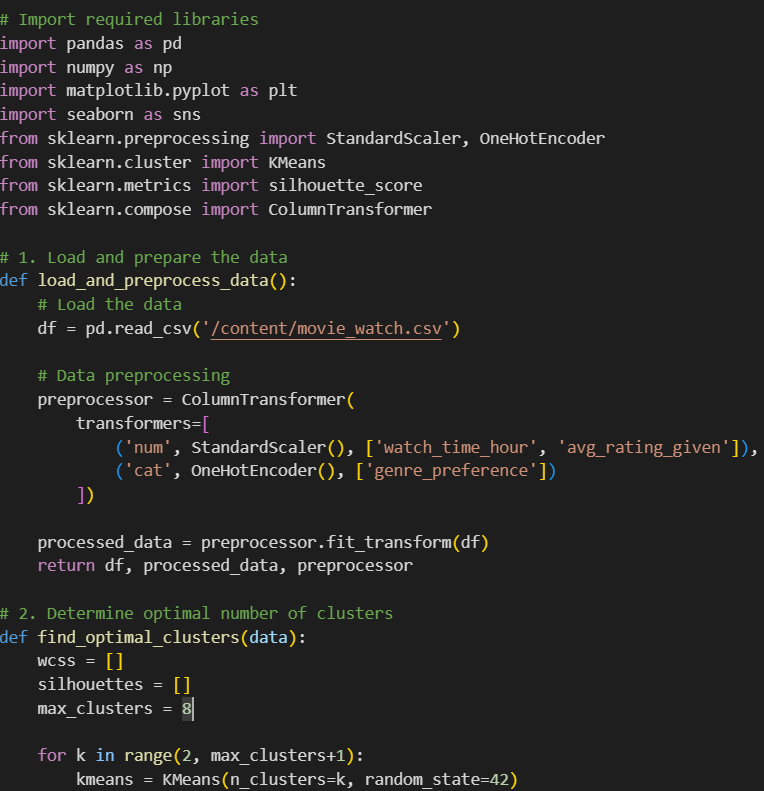
**4. Dimensionality Reduction for Visualization (PCA)**

* Since we had 3 features, we used **PCA (Principal Component Analysis)** to reduce them to 2D.
* This helps visualize the clusters clearly in a 2D scatter plot.

**5. Visualization**

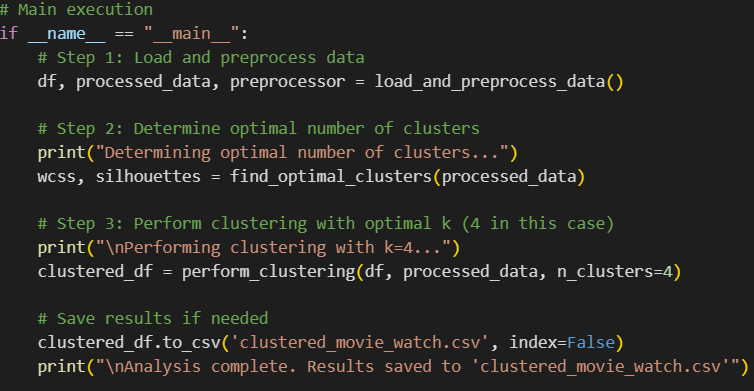
* Used **Seaborn** to create a colored scatter plot showing clusters.
* Each color represents a group of

**CODE:**

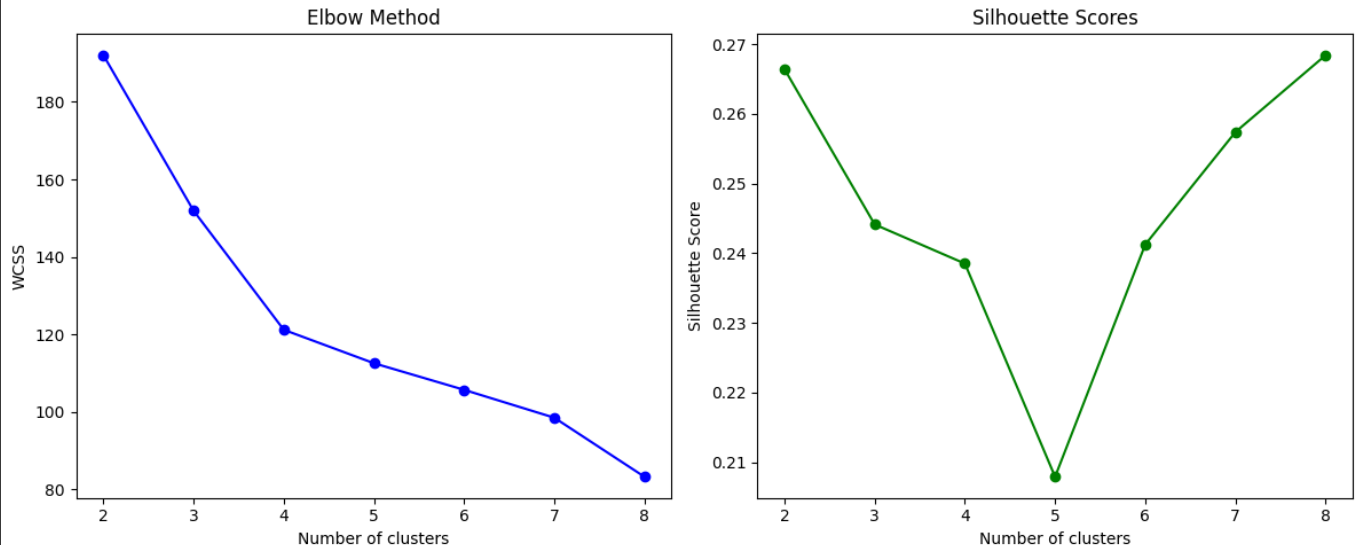


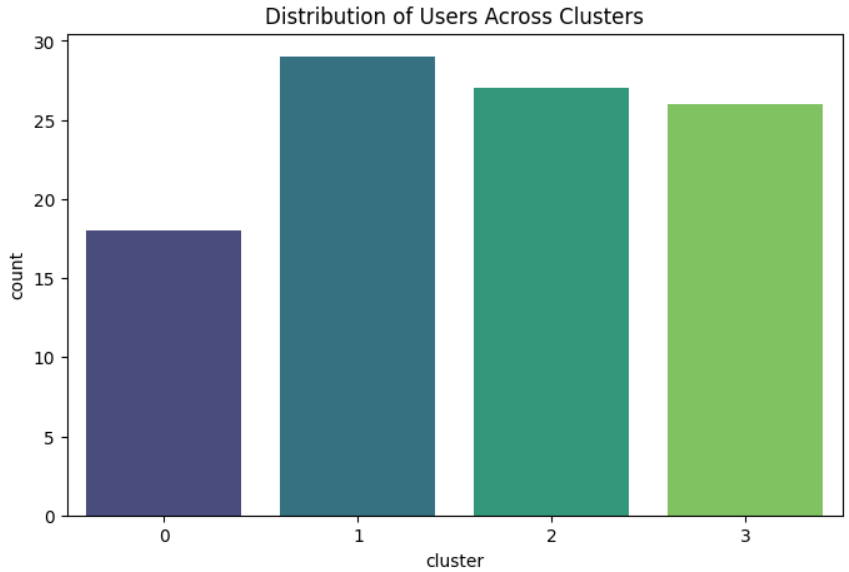


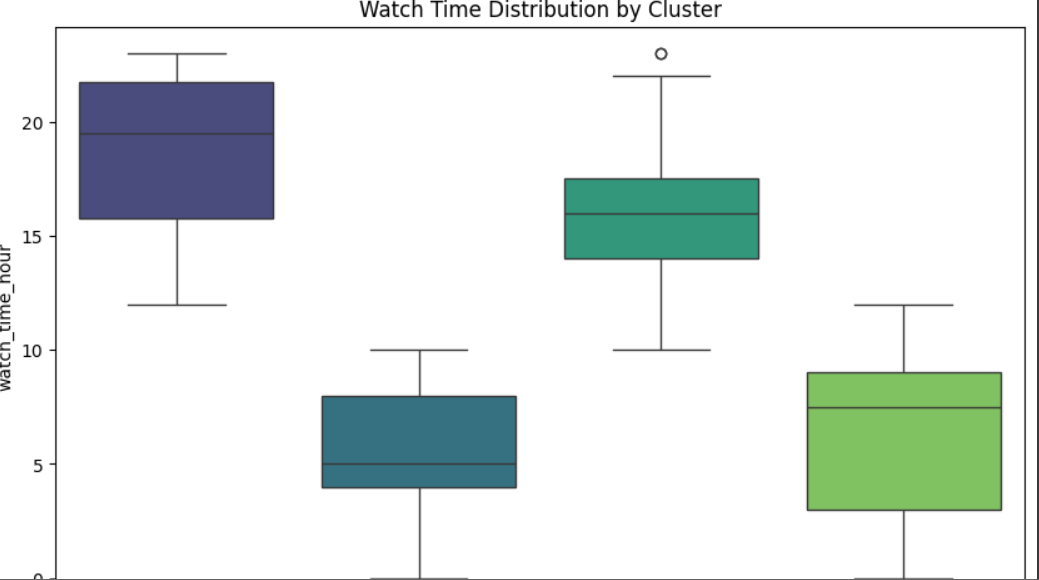


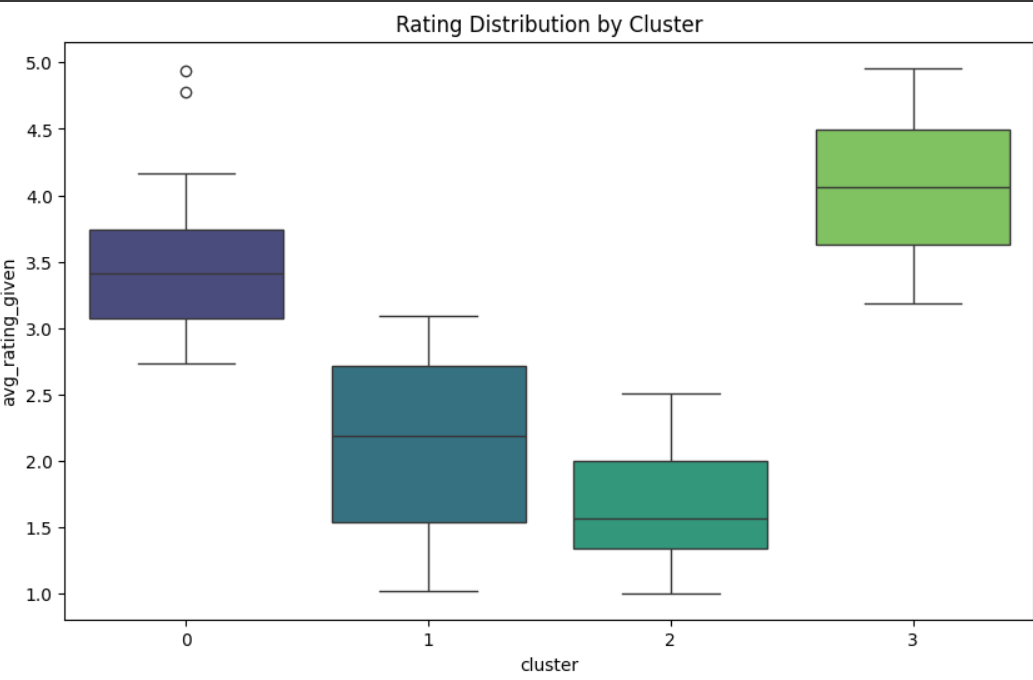


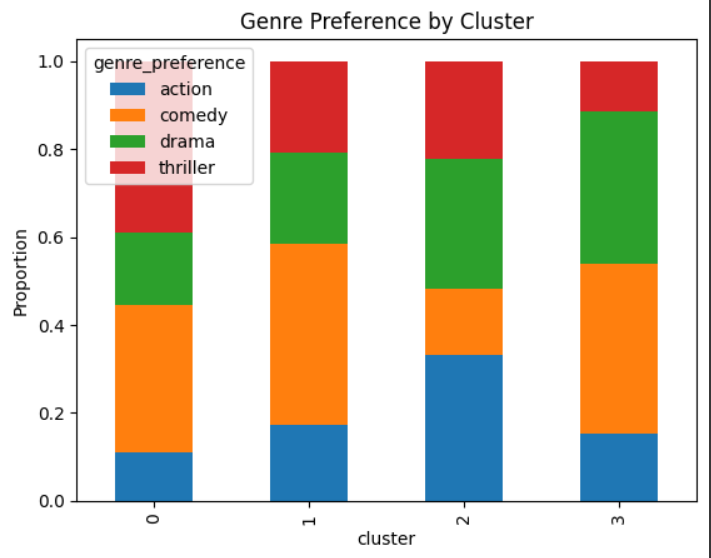
**OUTPUT:**











**REFRENCES:**

* 1. **Dataset  
     Custom dataset named movie\_watch.csv containing user watch patterns (watch time, genre preference, and rating behavior). If sourced or simulated, specify origin or tool used (e.g., self-generated,etc.).**
  2. **Scikit-learn: Machine Learning in Python**[**https://scikit-learn.org/**](https://scikit-learn.org/) **Used for KMeans clustering, PCA, data preprocessing, and evaluation metrics.**