**Jawwy TV**

Diversifying Kingdom’s economy is vital for its sustainability. Even though oil and gas are essential pillars of the economy, the Kingdom has expanded its investments into new sectors. Moreover, one of those sectors is entertainment. The cinema industry is one of the most critical parts of the entertainment sector, but in Saudi Arabia, it is tiny because it considers a new industry. In this project, we will try to understand the audience in Saudi Arabia to help grow this field until the Saudi cinema industry reaches internationality stardom. Analyzing the dataset related to movies and tv shows is part of this project. This dataset provides by Jawwy TV, which is a Saudi application. After analyzing this dataset, we have found many outcomes from Jawwy TV.

The first analysis was for a bar chart for average rating and genre, we tried to find the most popular types of movies and TV shows that Saudi subscribers like to watch on Jawwy TV. It is clear from the graph that Saudis like to watch drama and Horror movies. On the other hand, we can see clearly that Jawwy TV audiences do not like to watch documentaries and adventure programs.

The bar chart shows that number of votes during months is almost the same except for August and June. The votes were higher than others, reaching up to 450 votes in June and August. Also, on the bar chart of the distribution of program genre based on the duration of seconds, we noticed that the last three months of the year have the highest numbers of views. November views go above 10 million seconds. Furthermore, it decreases in the first two months of the year where the duration of seconds went below 3 million.

According to the pie chart of the HD quality, 37% of the movies and programs do not have an HD option. At the same time, just 63% have an HD option. Also, it is clear from the number of viewers that Saudis like to watch series more than movies. The difference is 71.773 viewers, which is a slightly large number. Animation movies have the highest number of movies with HD quality with almost 150 thousand protections and the highest who have not HD moves based on their large number of production where it went over 250 thousand.

Moreover, we found out the top five types of movie subscribers like on Jawwy TV, which were drama, horror, comedy, action and the last one of the top five is a thriller. The ratings of those types of movies are from 20.000 to 30.000 ratings. For example, all drama movies rating is equal to 30K. However, the adventure movies rate equals 10K. The line chart shows that views increase above 150 thousand seconds by month between February and Jun. Moreover, decrease below 100 thousand seconds between July and January.

After analyzing the data and getting some outcomes from the analysis. We have looked and searched for an appropriate way to build a recommendation system. There are three types of the recommendation system collaborative filtering, content-Based Filtering, and hybrid Recommendation systems. The most appropriate one in our case is the collaborative filtering because we do not have a description of the movies and identify preferences and information from many users. We have started building a clustering model first after choosing the type of the recommendation.

The first model is a clustering model for clustering users and movies, which is useful to build a recommendation system. In this model, we have clustered the data into 8 clusters based on movie id, user id, and the average rating. Then we have built a recommendation system to get different movies id, and titles suggested for the users.

The second model is the apriori which can work for frequent item set mining and association rule learning over relational databases. Here we have use it to help in generating movies recommendations based on user id and views in the dataset. The results of the recommendation system of apriori model were more sufficient than in the clustering model. It can suggest movies to various users based on the Jawwy TV data that they have collected for that user. A future plan to improve this project is using the matrix factorization which can improve our machine learning.