

# Final project - Mid-Project update

## MOVIE RECOMMENDER SYSTEM

### TEAM MEMBERS

S.No	NAME OF THE MEMBER
1.	Navya Sai Reddy Gorre
2.	Chetana Alekhya Reddy Gorre
3.	Maheshwar Reddy Peram

### PROGRESS

#### BUSINESS PURPOSE OF OUR PROJECT

Big Business organizations constantly try to know more about their customers. They make money by analyzing the past data they collected. It would be a difficult task for a human to use this large data. This is where Machine Learning comes into view. There are many machine learning concepts that are being implemented in Business sector. One of the advanced Machine Learning concepts is the Recommender System.

Recommender Systems are helping organizations to know their customer's interests and habits simply through their data. This basically helps them to know what their users are expecting. So, these Recommender Systems make recommendations to the user for what they would purchase or what they would like more.

These are the various use cases of Recommender Systems:

- Increasing Customer Loyalty
- Increasing Sales and Profits
- Improving User Experience
- Knowing the Market

We gathered this information and decided to build a Movie Recommender System.

## **DATA PREPARATION**

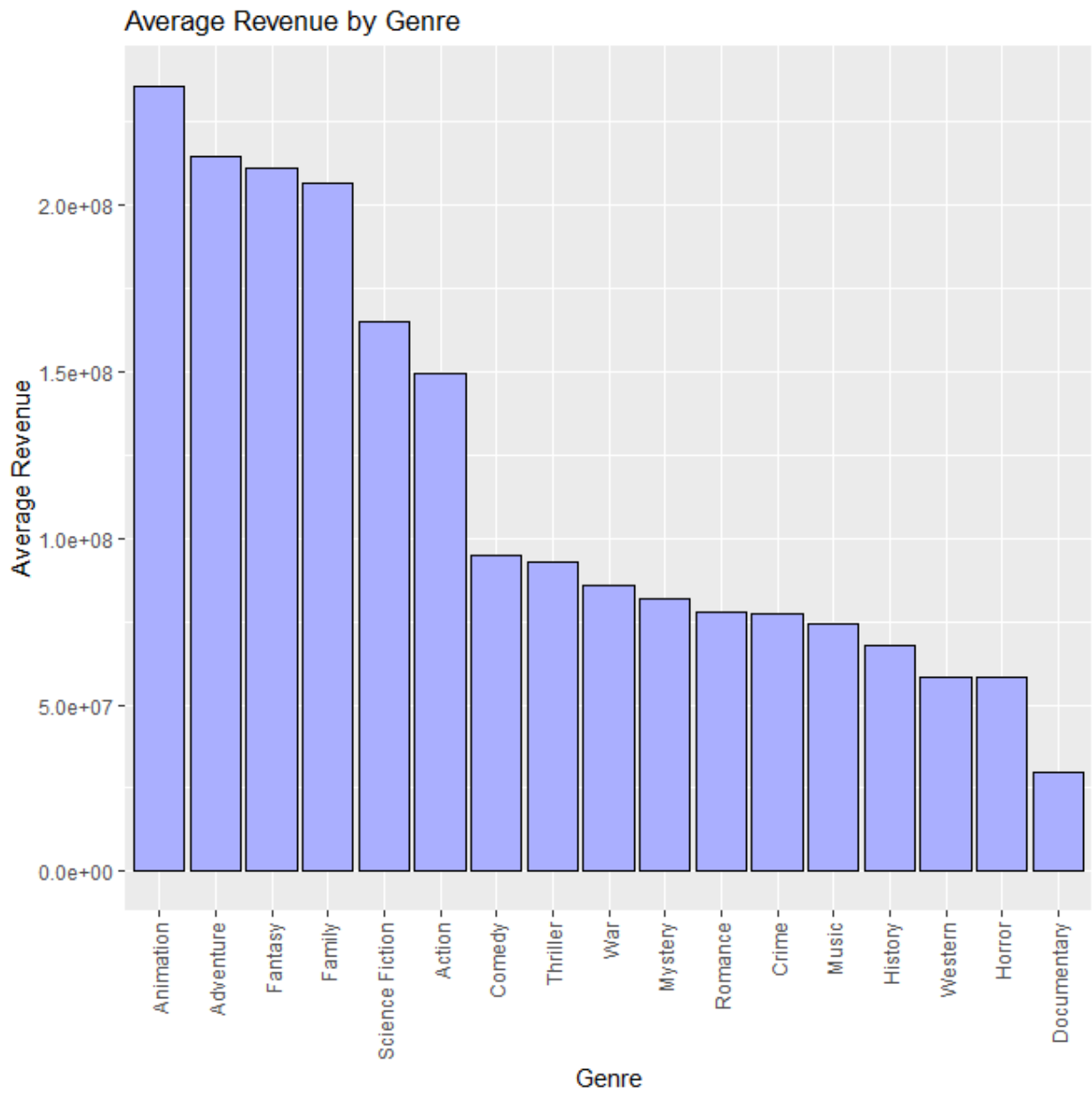
We collected the data from the Kaggle website and MovieLens website to build Recommender Systems using three different approaches.

- Data Cleaning:
  - We removed the rows which contain Null values.
  - We removed all the duplicate values present in the data.
  - We changed data-types of particular columns to do the computation according to our requirement.
- Data Filtering:
  - We removed the unwanted columns in our dataset. This will also reduce the burden to the system.
  - We added a few extra columns from existing columns which helped us to do Analysis.
- Outliers Removal:
  - We removed only the one outlier present in our data.

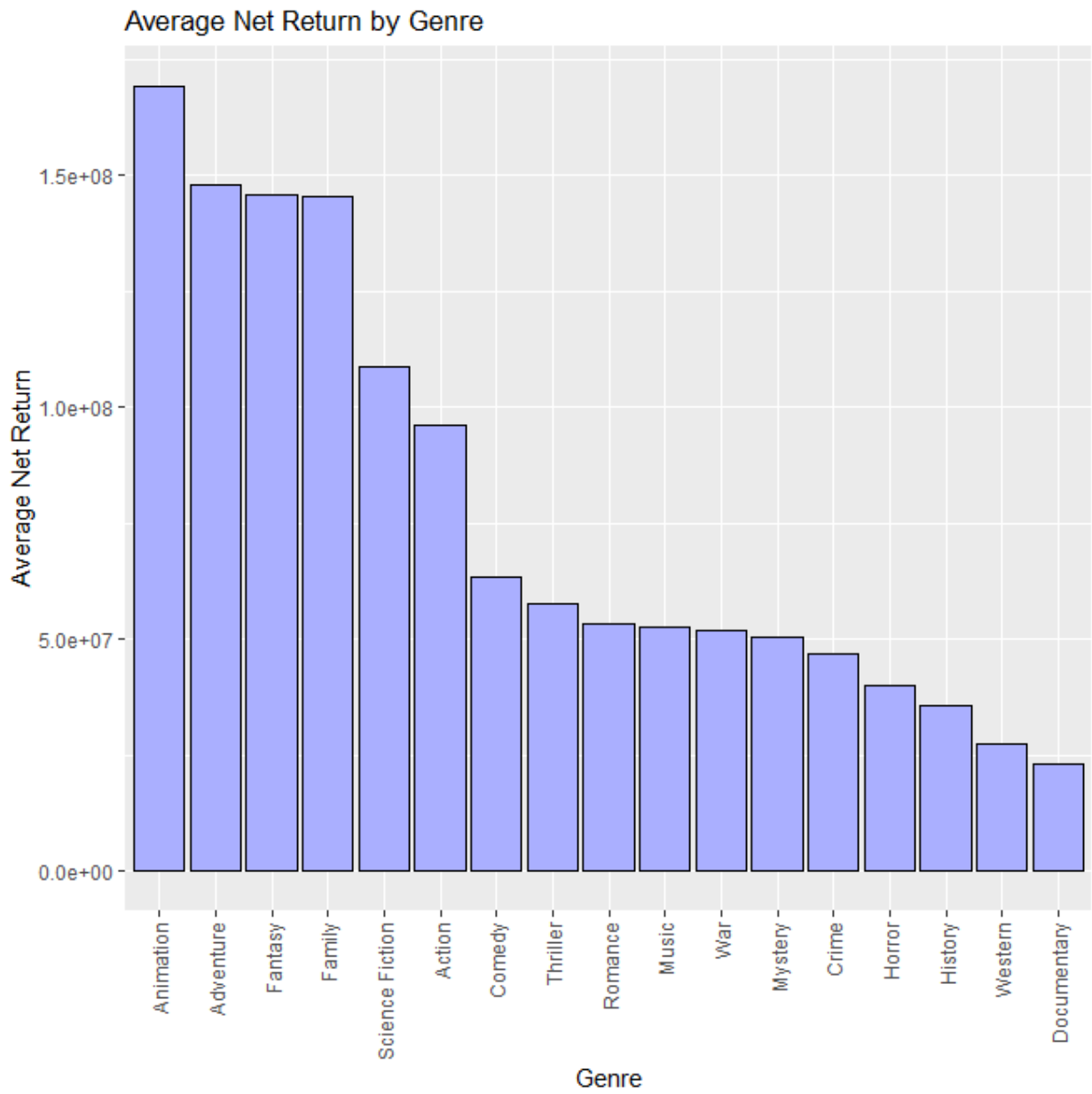
## **EXPLORATORY DATA ANALYSIS**

We divided our data into Genre-wise distribution to see how genres are affecting the financial matters of the movies and we analyzed how the runtime of the movies are affecting the revenue of the movies and also analyzed how the Average ratings of the users are affecting revenue.

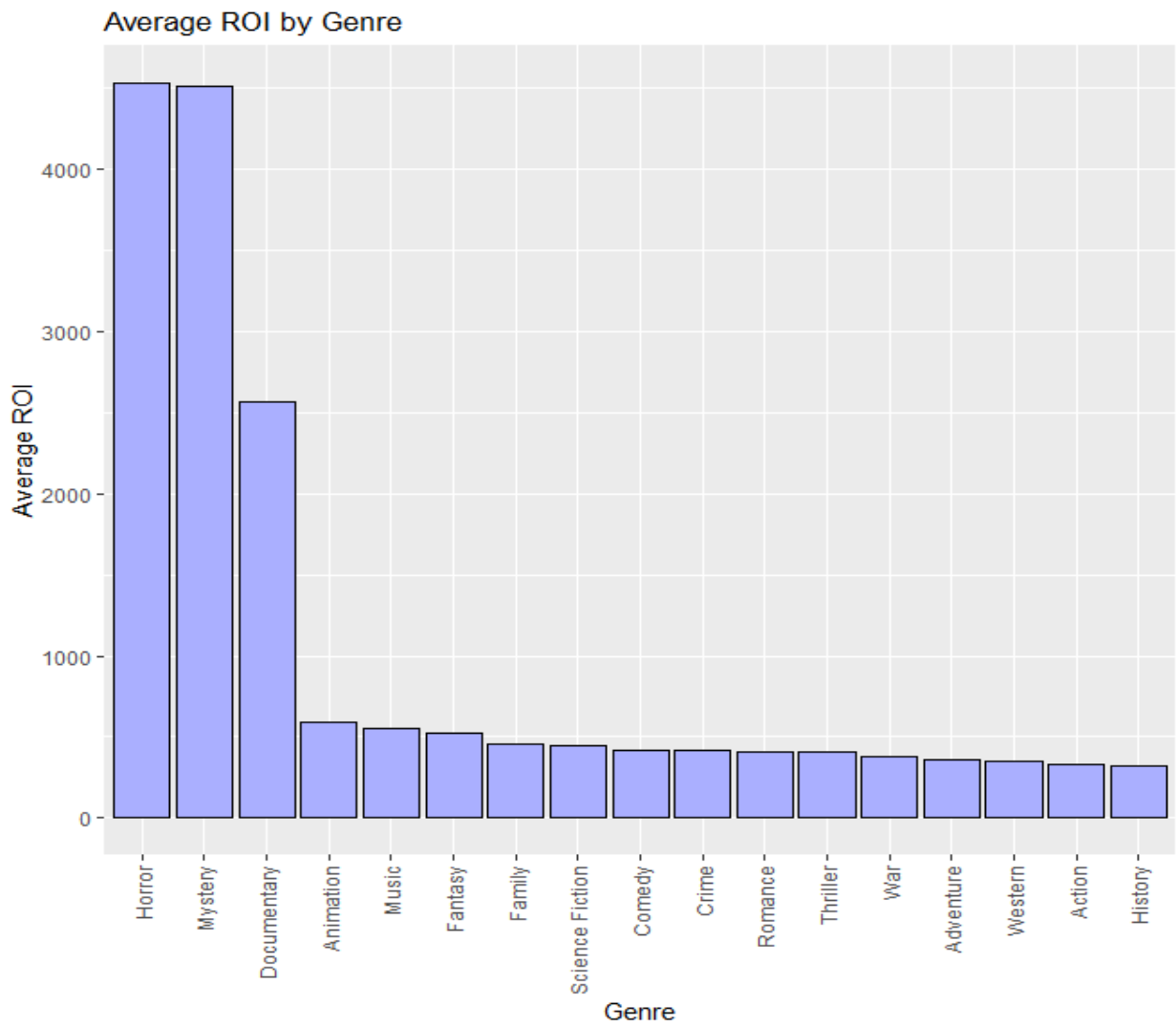
Here are some plots which gives us detailed results :



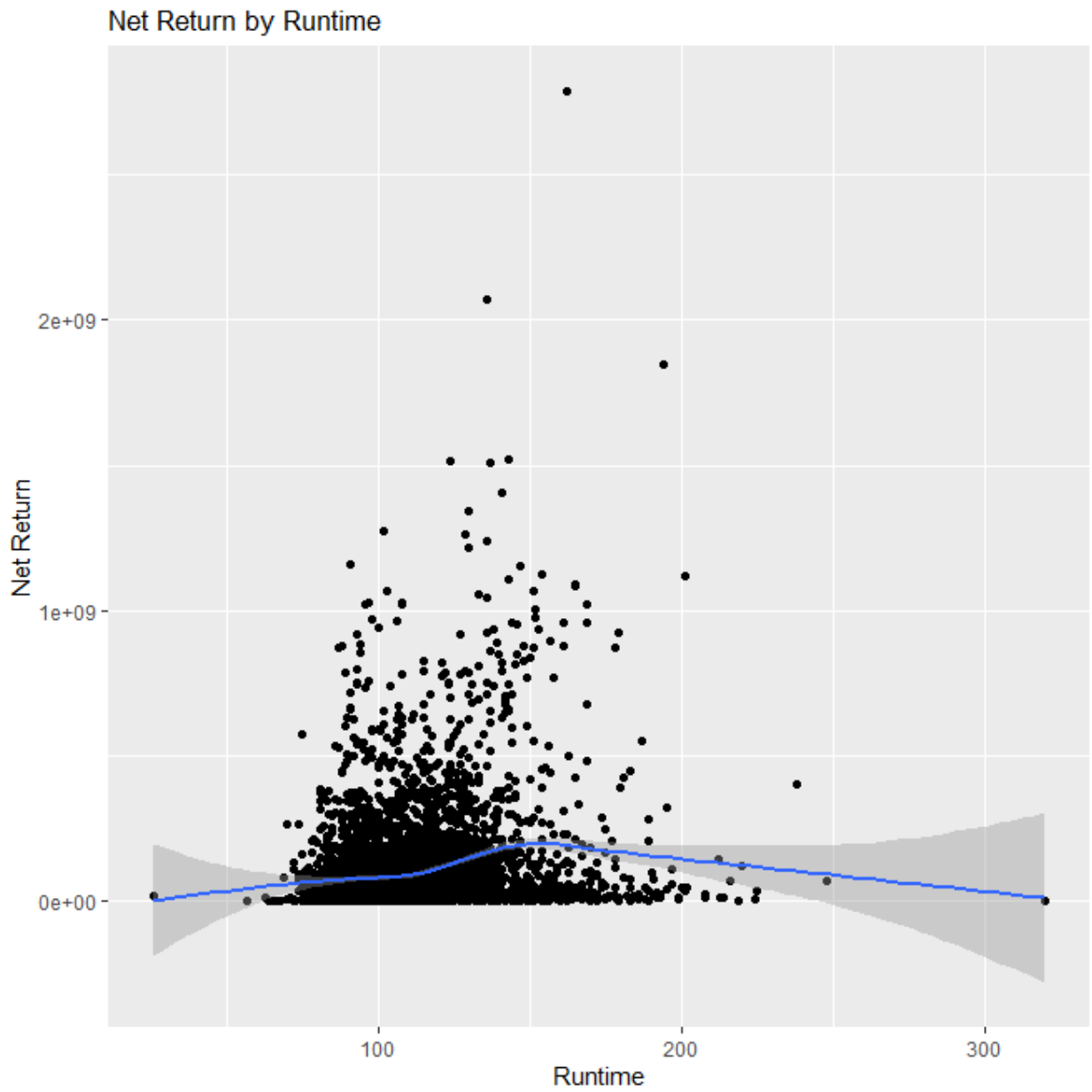
(Average Revenue vs Genre)



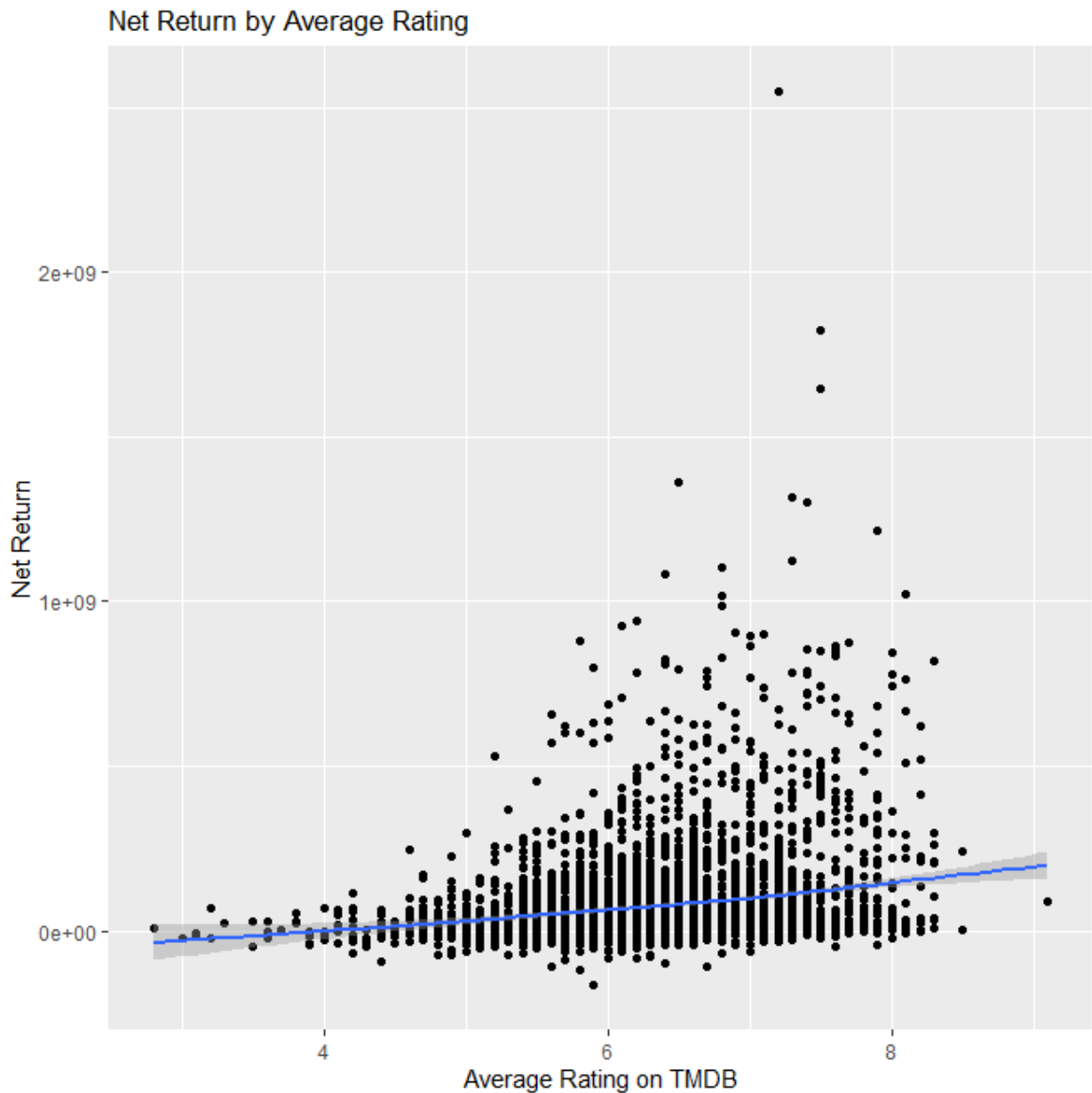
*(Average Net Return vs Genre)*



(Average ROI vs Genre)  
\*ROI - Return of Investment



*(Net Return vs Runtime)*



## RECOMMENDER SYSTEMS

### Ratings-based Recommender System

We built this model considering that the users will most likely enjoy movies that are of high budget and also movies which gain high profits. So, we can say this model is not able to give personalized recommendations based on the user.

	top_10.original_title
1	Avatar
2	Star Wars: The Force Awakens
3	Titanic
4	The Avengers
5	Jurassic World
6	Furious 7
7	Avengers: Age of Ultron
8	Harry Potter and the Deathly Hallows: Part 2
9	Frozen
10	Beauty and the Beast

*(Top 10 recommendations using ratings-based model)*

We can see our model recommended movies that gained high profits. All the movies that are displayed in the above image are highly acclaimed by the people.

## **PROGRESS LINK :**

[https://github.com/maheswarreddy01/movie\\_recommendation\\_system](https://github.com/maheswarreddy01/movie_recommendation_system)

## **CHALLENGES**

Detecting outliers has become a challenge for us when we are preprocessing the data. The results varied a lot when we did analysis without removing outliers. There are several attributes like budget and revenue. We used these attributes to find outliers.

The only outlier present in the data after filtering has a budget as 5000 dollars and revenue as 11083449 dollars which is highly impossible. So, we removed it successfully.

While coming to the code designing part, there are sometimes we are not able to meet in person. So, we used the Discord application to collaborate and work together.

We are still gathering information to complete the other two models. As we saw earlier, Ratings based Recommender is a simple recommender without any personalization.



The upcoming models require better understanding of the data. So, this remained a challenge for us.

## **COLLABORATION**

We are a group of three members. We stay at the same place and discuss frequently on this project. We have divided the development work among ourselves equally. We divided data collection, data cleaning, data filtering and outliers removal among 3 of us and worked on rating based recommender system. We helped out each other when stuck at any point in the individual work. We are still working on the other 2 models. So, this helped us to make our progress so far.

## **NEXT STEPS**

We only did a Ratings-based Recommender System. We are yet to complete the Content-based Recommender System and Collaborative-based Recommender System.

Then we have to evaluate every model taking the parameters like RMSE and Silhouette score.

We need to refer to more research papers to understand the remaining models. This will lead us to complete our models efficiently. Working on the other two remaining models might be a challenge for us. In order to overcome this problem, we need to explore more libraries in R language in order to complete the remaining steps.