Summary of Michael’s Contribution

**Introduction**

In the introduction, I provided an overview of the project, including the problem statement, objectives, and the significance of the study. This sets the stage for the rest of the analysis.

**Load Data**

In this section, I loaded the dataset(s) and investigated the source of data we would be working on. These included importing necessary libraries (e.g., pandas, numpy, matplotlib) and reading the data from files (e.g., CSV).

**Exploratory Data Analysis (EDA)**

EDA involves examining the dataset to understand its structure, patterns, and anomalies. These included:

Descriptive statistics: Summarizing the main characteristics of the data.

Data visualization: created plots (e.g., histograms, scatter plots) to visualize distributions and relationships.

contributed to why so few quantity of rating in some movies are not representative to the quality of the movie

**Data Preprocessing**

Encoding categorical (genre) variables: converted categorical data into Boolean format for user genre (content-based filtering).

Feature scaling: normalized the user-item matrix to account for user bias and movie rating bias.

**Work with hector on choosing the models**

Worked with hector on building the recommendation systems.

**Notebook code**

Contributed to many preliminary codes for the recommendation system. Wrote many comments to the code blocks.

**PowerPoint**

Worked on the layout of the PowerPoint and provided necessary figures into the slides to complement the word context.