

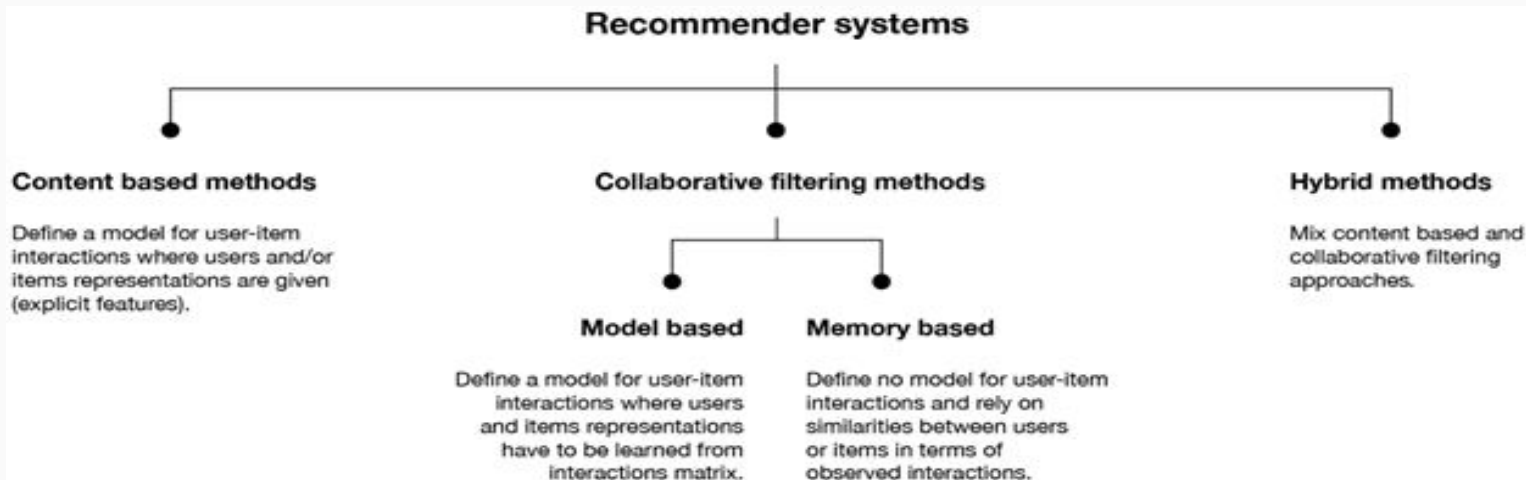
# DSI-13 Capstone Project Movie Recommender System



By Davis Hong

# Introduction

The below are the various recommender systems currently used by content providers, online retailers, etc.



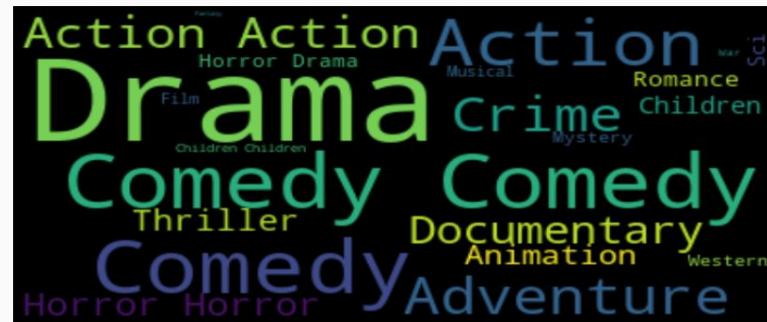
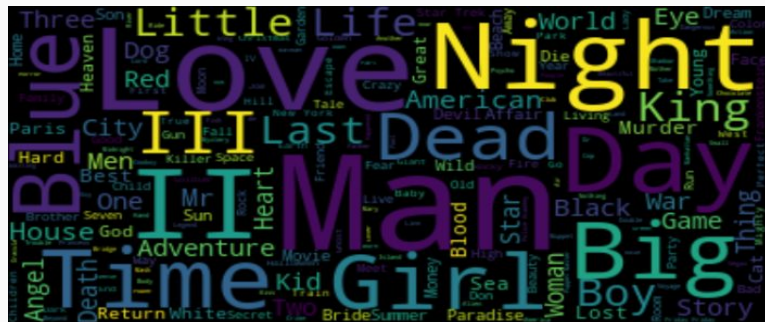
# Introduction (con't...)

- All 3 modeling approaches will be used for this project.
  - Content-Based Filtering makes recommendation based on user's shopping and browsing habits
  - Collaborative Filtering based on user-to-user or item-to-item to make recommendations
  - Hybrid Filtering combines Content-Based and Collaborative Filtering for better results

# Problem Statement

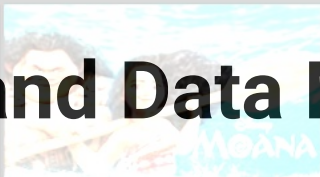
- Nowadays, users have very short attention span. With more and more players coming into the online video contents space, a recommender system is a must.
- Objective: Develop and identify the best recommender system that is;
  - fast and responsive
  - able to predict users' preference fairly accurately and make recommendations

- MovieLens datasets will be used
- Cleaning is required as the datasets are clean
- EDA shows that movie title on Night Love triumphs it all
- Drama, comedy and action genre are the hot picks

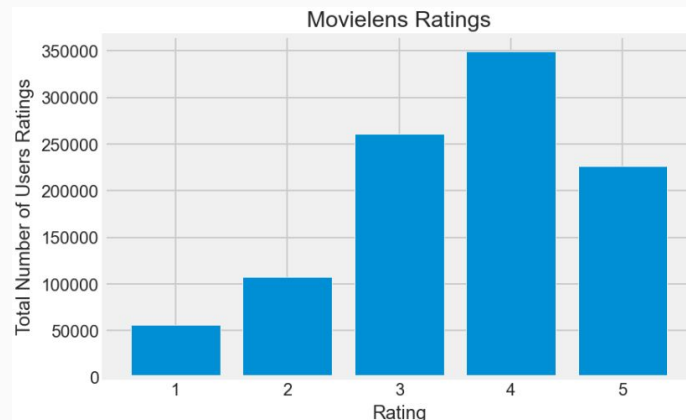


STREET  
THINGS

# EDA and Data Munching (con't...)



- Majority of the ratings fall in the 4-star category
- It is a negatively skewed distribution (Mode > Median > Mean)
- The ratings may not be a good indicator whether the movie is good or not as different people have different rating style







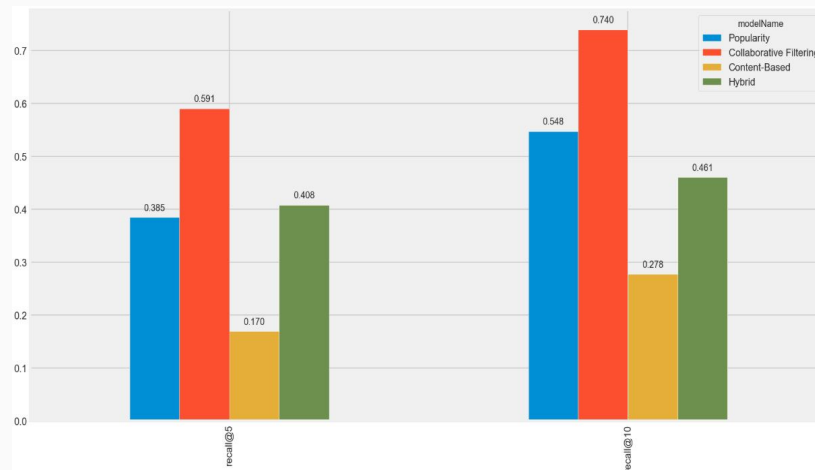
# Modeling

- A simple model using cosine similarity with TF/ IDF added subsequently to see recommender in action
- The below models were developed;
  - Popularity
  - Collaborative Filtering
  - Content-Based Filtering
  - Hybrid (a combination of Collaborative and Content based filtering)

# Evaluation of models

Top-N accuracy Recall@N metrics was used to evaluate the 4 models;

	recall@5	recall@10
modelName		
Popularity	0.384924	0.547870
Collaborative Filtering	0.590706	0.739545
Content-Based	0.169519	0.277652
Hybrid	0.408239	0.460618







# Conclusions and Recommendations

- X