

AFKAdventures: Recommendation System for Games on Steam

DATA 240: DATA MINING

Guided by:

Prof. Shayan Shams

Presented by: (Group 8)

Abdul Sohail Ahmed

Bhavik Patel

Poojan Gagrani

Smeet Sheth



TABLE OF CONTENT

- Introduction
- Motivation
- Background
- Literature Review
- Methodology
- Experiment and Results
- Future Scope
- Contribution
- References

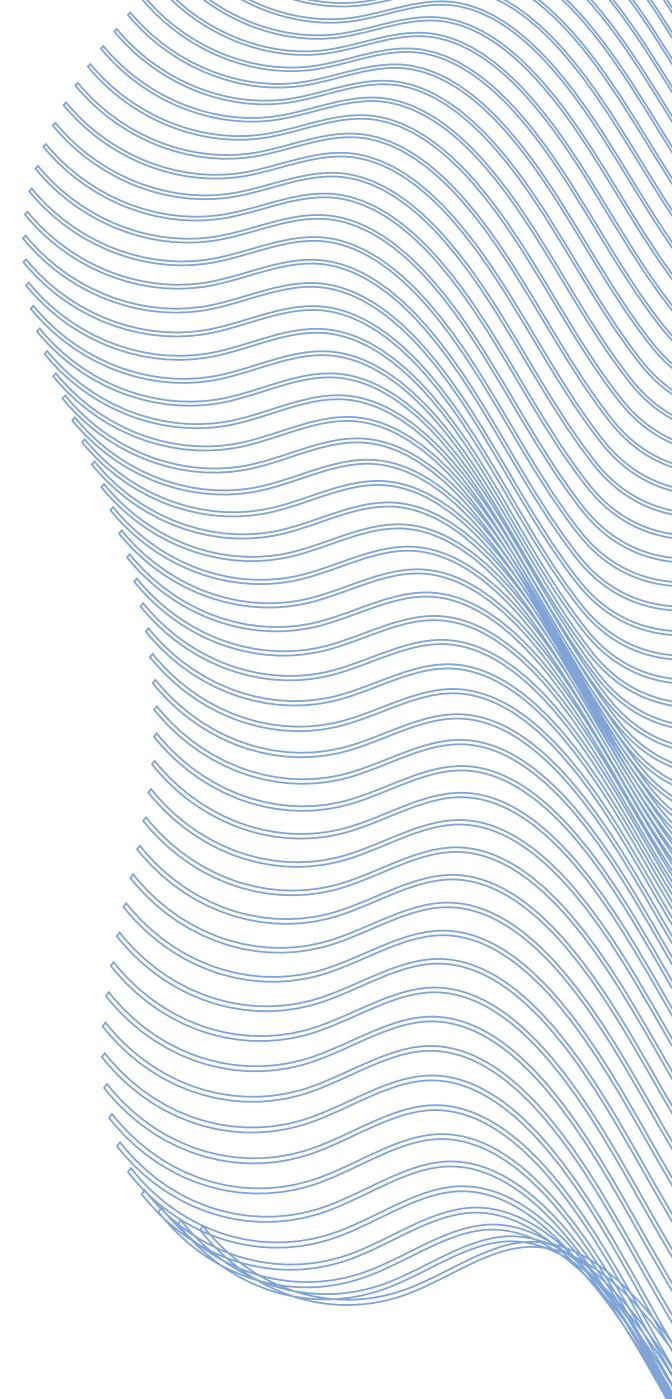


LET'S
PLAY

MENU

INTRODUCTION

- **About Platform:** Steam is a platform developed by Valve Corporation in 2003. Games range from AAA titles to Indie.
- **Market Leadership:** Steam dominates the market of PC gaming.
- **Reason for Popularity:** Vast game offering, user-friendly interface, and periodic sales like Winter Sales, Black Friday Sales, etc.
- **Cross-Platform Functionality:** Available in Operating Systems like Windows, Linux, and macOS.



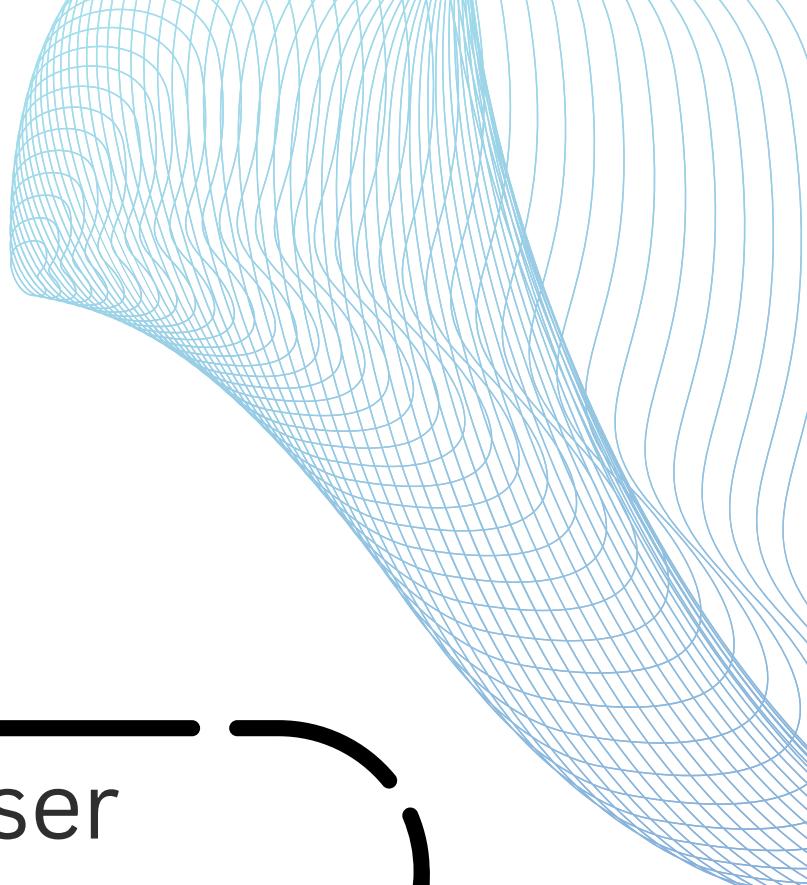
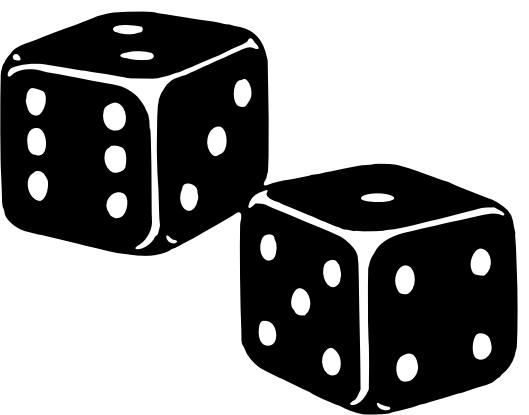
MOTIVATION

Utilizing Data
Effectively

Improving User
Experience

Enhancing Platform
Engagement

Boosting Sales of the
Games





BACKGROUND

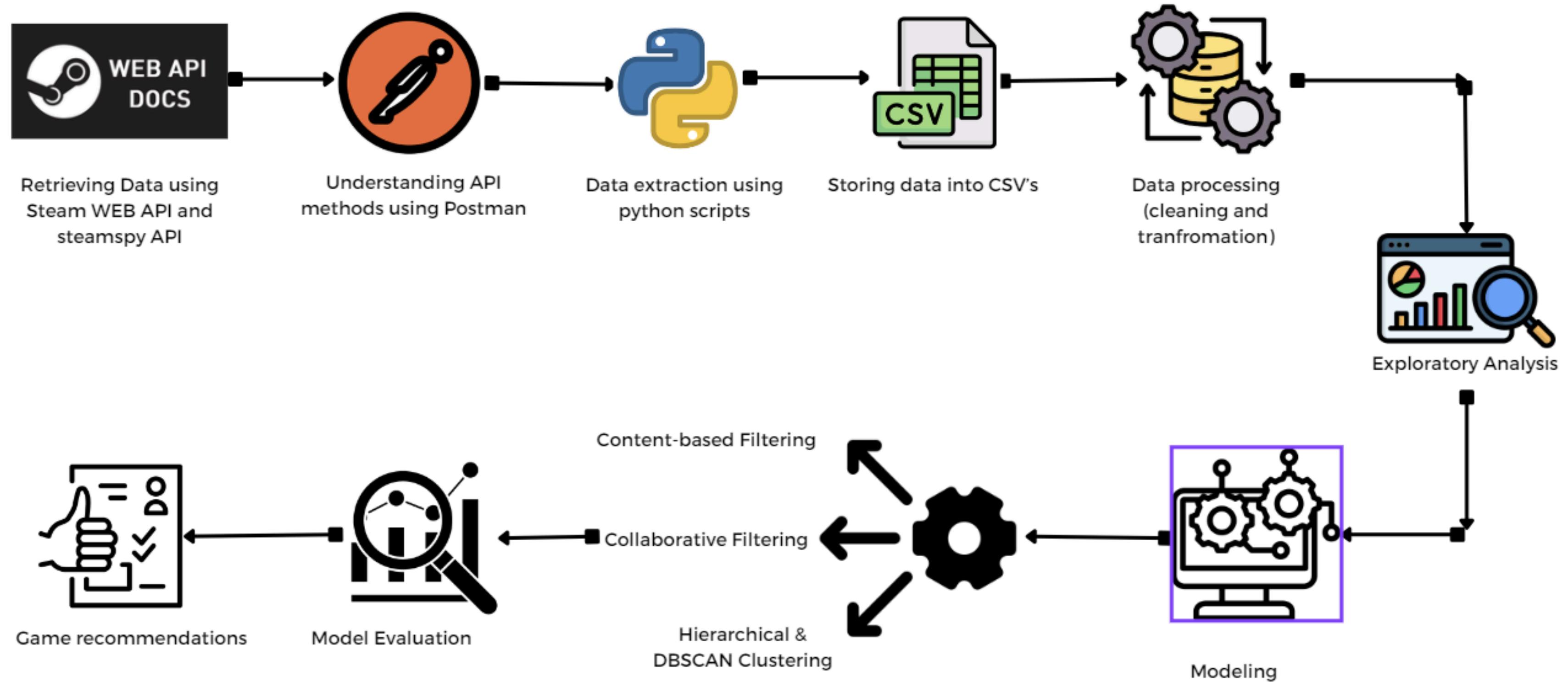
- **Overview:** In 2022, the gaming industry's worth reached approximately \$220 billion, surpassing the combined value of the film and music sectors.
- **Popular Game Platform:** Steam is one of the most comprehensive platforms, hosting over 50,000 games across myriad genres.
- **Challenges of Game Discovery:** Options are abundant but choosing one game over another is challenging.
- **Early Recommendation Systems:** Most of the initial systems focused on the popularity and sales of the games, but didn't focus much on user preference.
- **Limitations of Earlier Systems:** Generic game recommendation which lacks the personalization that leads to missing interesting games preferred by users.
- **Improving Recommendation Systems:** To generate recommendations based on user game playing history that helps to gain an idea about the preferences and also provides the facility to select category and genre and provide appropriate games to them.



LITERATURE REVIEW

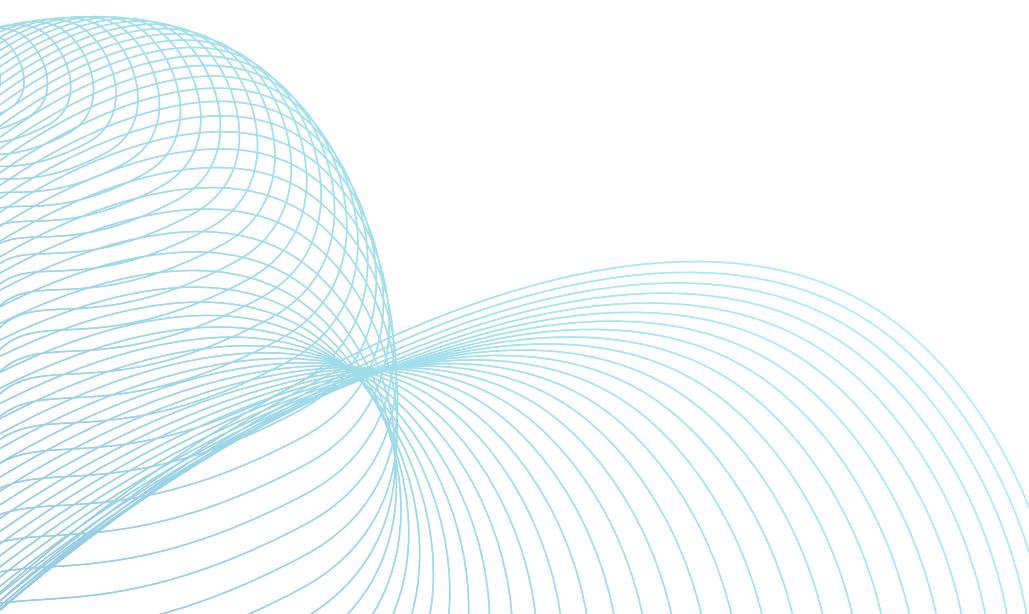
Author	Title	Dataset Used	Models / Methodology	Results
Kang, H. N., Yong, H. R., & Hwang, H. S.	A study of analyzing on online game reviews using a data mining approach: Steam community data	Open Source Game Review Data (https://github.com/mulhod/STEAM_reviews) extracted from Steam	CART (Classification and Regression Trees) ANN	CART : MAE (0.112E+6) SSE (1213.269) ANN: MAE (0.188E+3) SSE (2691.165)
Cheuque, G., Guzmán, J. L., & Parra, D.	Recommender Systems for Online Video Game Platforms: the Case of STEAM	Three Datasets (provided by Julian McAuley) First: Purchase history of Australian Users of Steam Second: Opinion of users on different items in platform Third: Detailed features of available games	ALS (Baseline) FM (with/without) DeepNN (with/without) DeepFM (with/without)	Best Model: MAP @10 (0.897) NDCG (0.948) Novelty (0.197)
Lomanto, W.E., Andrain, V., Achmad, S., & Sutoyo, R.	Collaborative Filtering for Steam Games Recommendation	Steam Video Games Dataset (Kaggle) (12,393 users and 185,673 games)	SVD Pearson Correlation	RMSE (PC): 1.90 MAE (PC): 1.80 RMSE (SVD): 1.00 MAE (SVD): 0.90
Batra, S., Sharma, V., Sun, Y., Wang, X., & Wang, Y.	Steam Recommendation System	Steam Video Game and Bundle Data provided by Julian McAuley	ALS (playtime) ALS (playtime, sentiments) ALS (playtime, recommendations)	RMSE (3.31) RMSE (2.64) RMSE (3.32)

METHODOLOGY

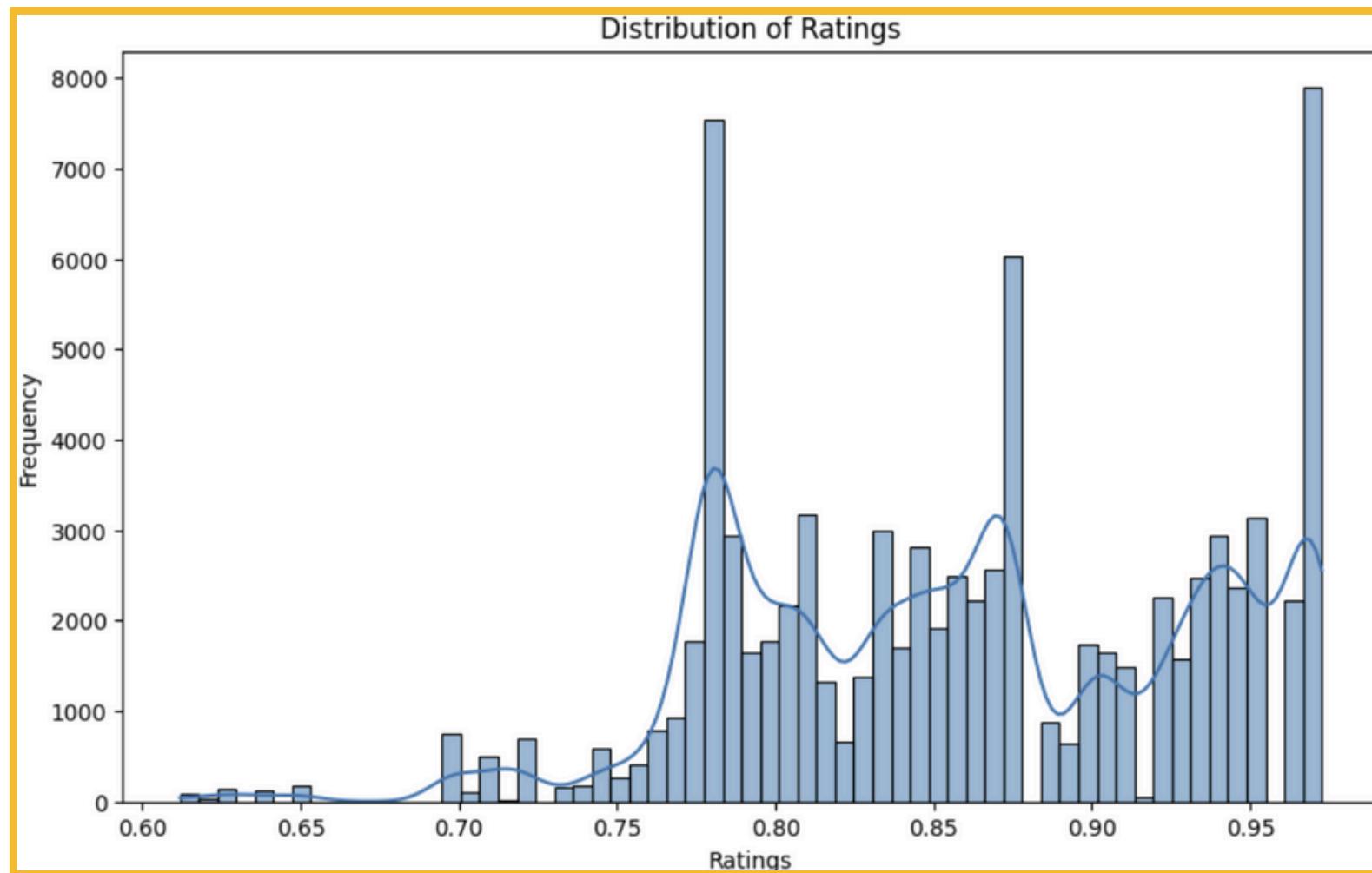


DATA STATISTICS AND PREPROCESSING (USER)

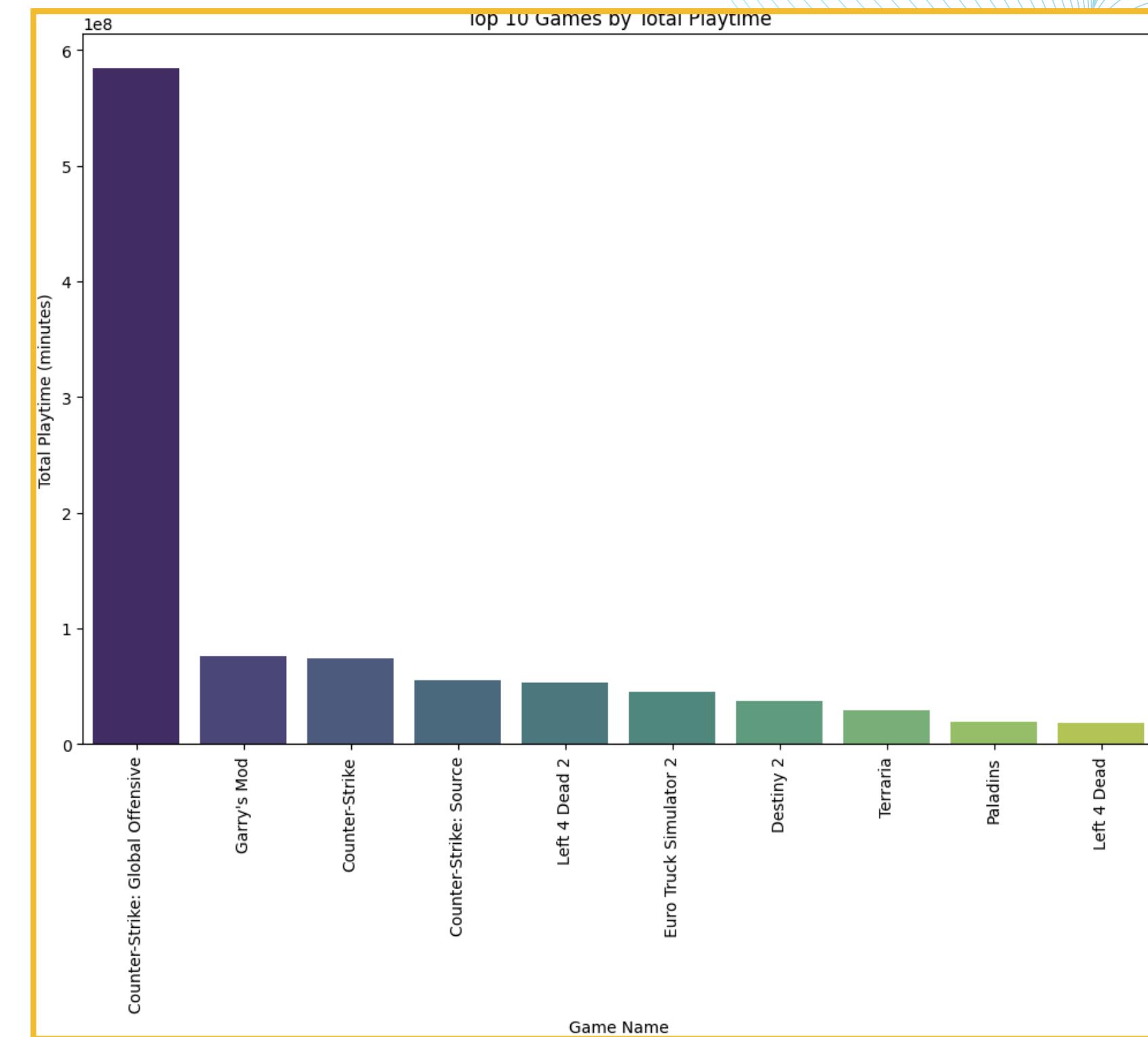
	Number of Features	Number of Unique Games	Number of Unique Users	Null Values	Data Cleaning Operations
Raw Data	12	837	5505	16,271 (Country Code) 368 (Developer) 14,809 (Currency)	Column renaming Dropping irrelevant columns Dropping values where game name and genre is missing Filling missing value in country name
Cleaned Data	11 (Dropped Currency and Country Code) (Added Log Playtime)	817	5504	0	Correcting misspelled country names



EDA (USER)



Distribution of Ratings across the Games

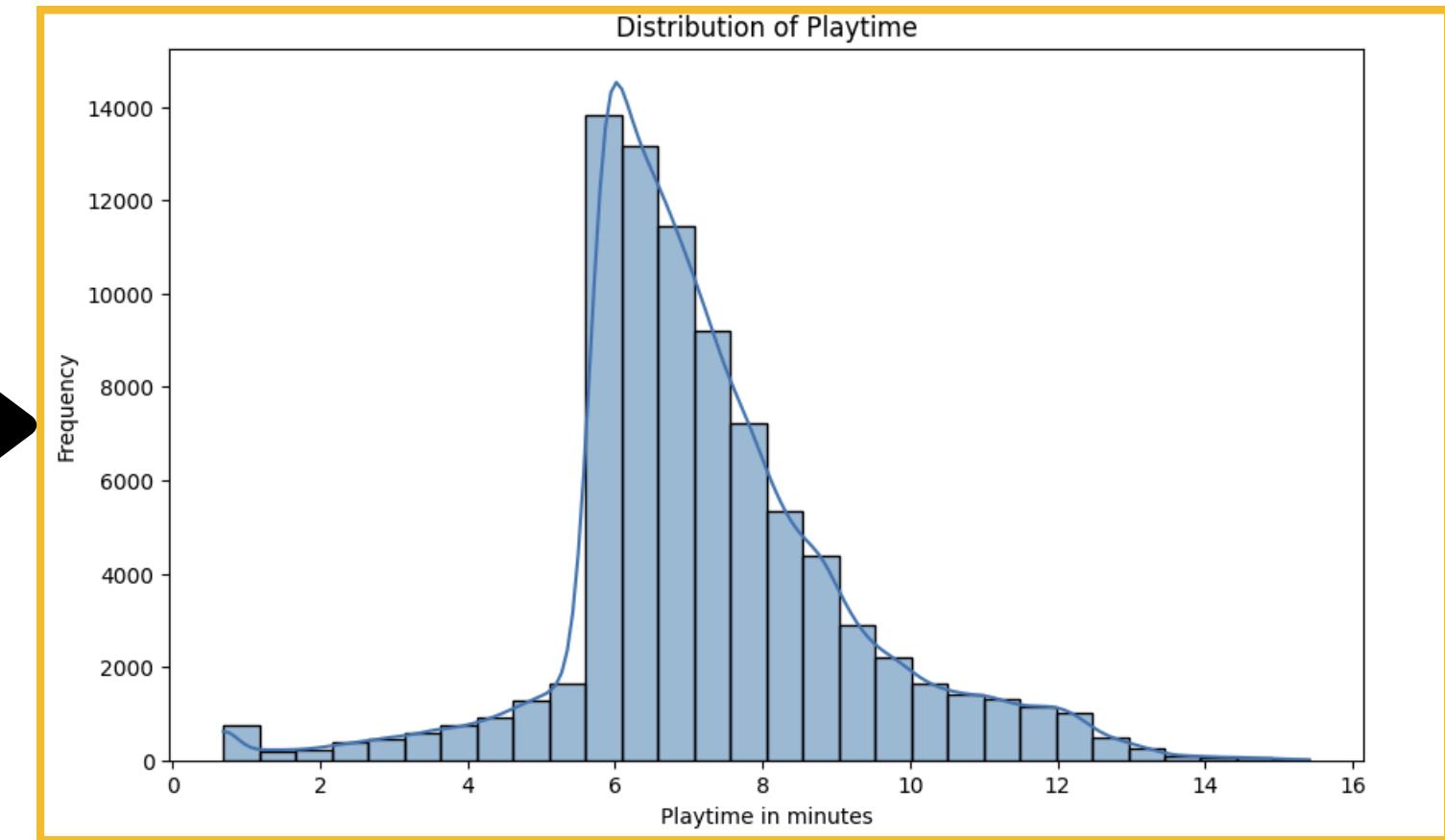
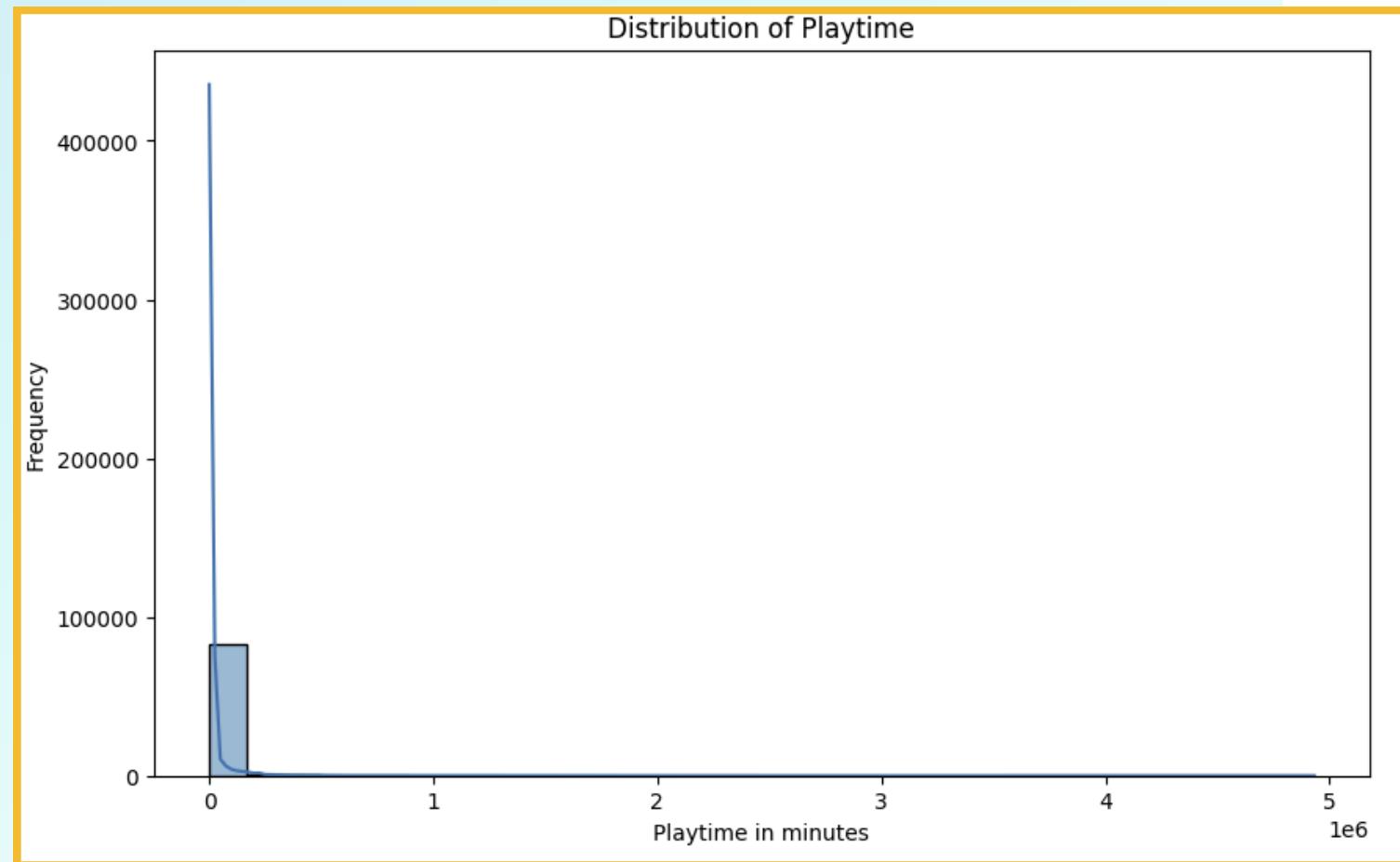


Top 10 Games based on Total Playtime



DATA TRANSFORMATION (USER)

Log Transformation

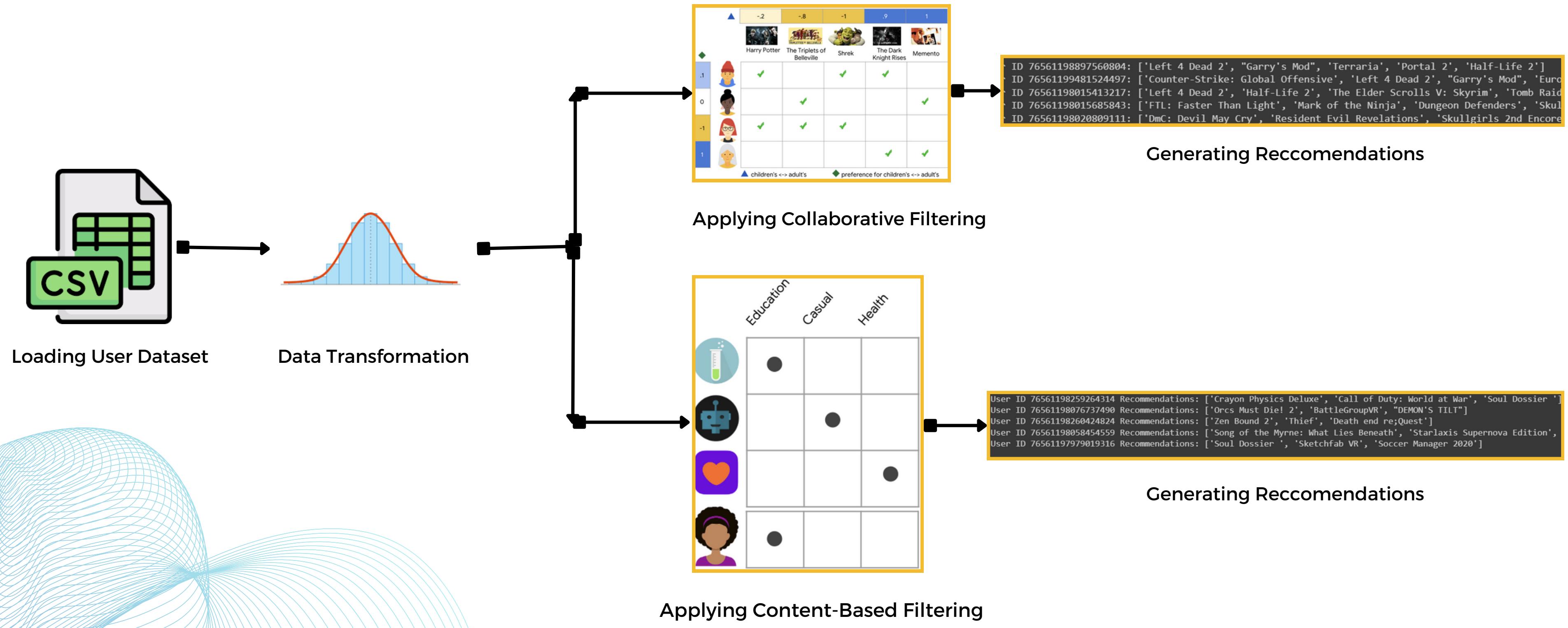


Before

After



MODEL WORKFLOW (USER PREFERENCE)



EXPERIMENT AND RESULTS

Collaborative Filtering (Bayesian Personalized Ranking (BPR))		
Factors	10-100	100
Learning Rate	0.001-0.1	0.1
Regularization	0.01-1.0	0.01
Iterations	50-500	50
Precision@K (1-10)	0.26 - 0.9	0.2
Recall@K (1-10)	0.11 - 0.33	0.16
AUC	0 - 1	0.18

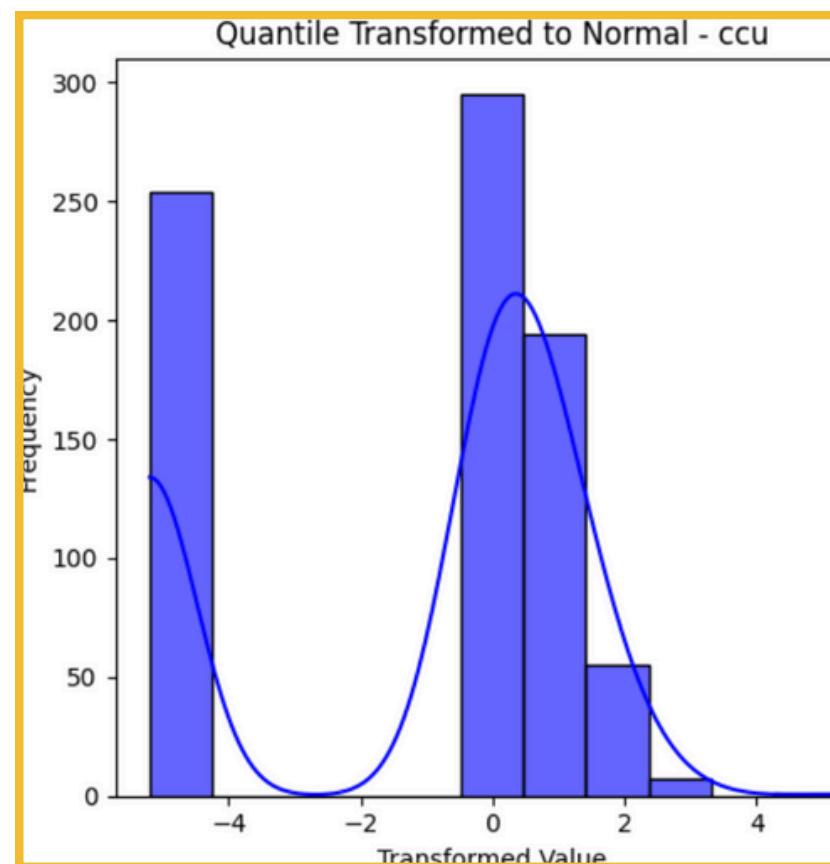
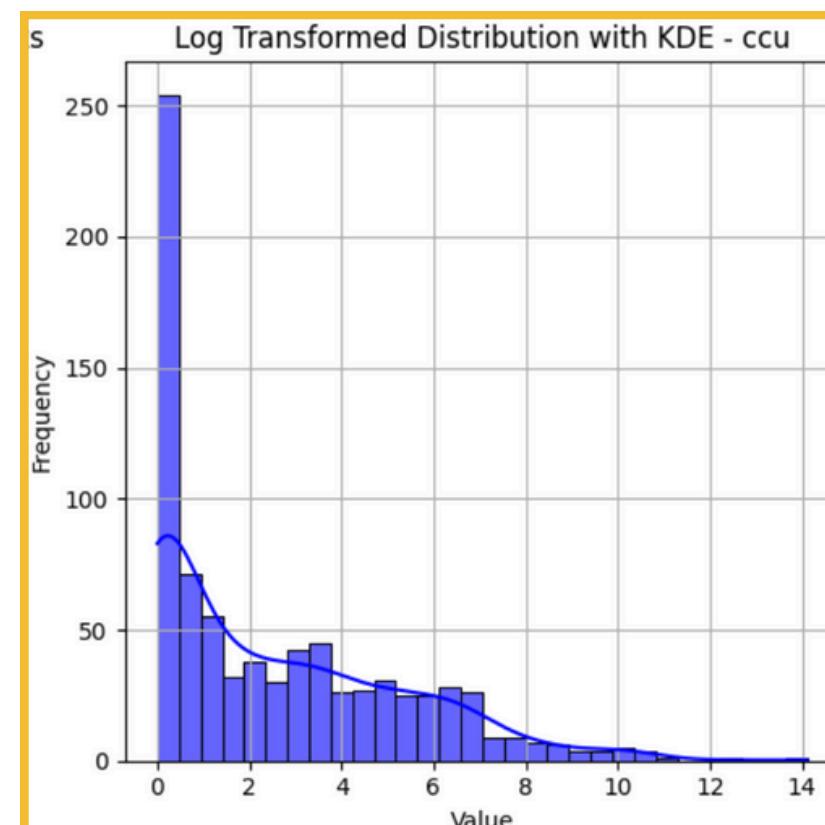
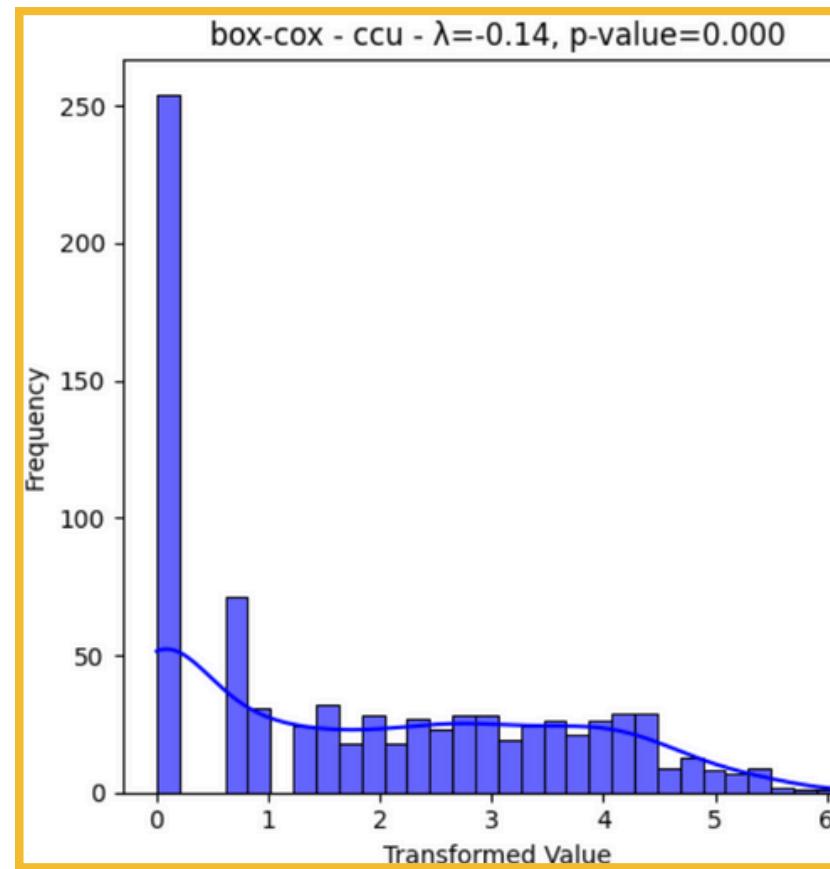
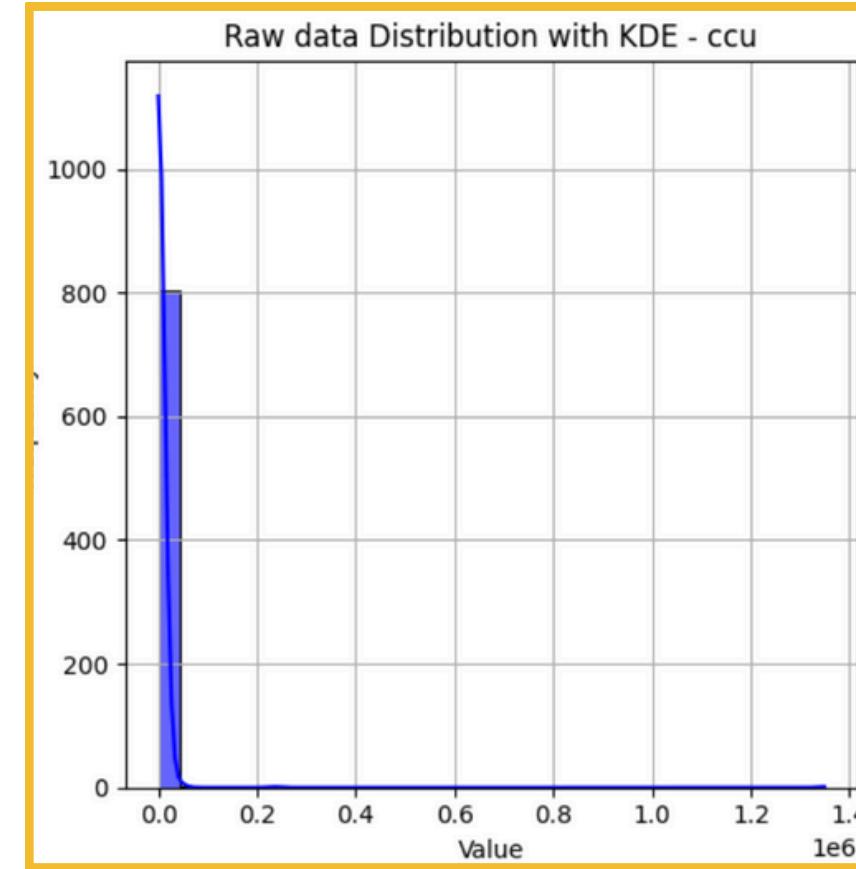
Content-Based Filtering	
Train-Test set	80:20
Random State	42
Feature Scaling	Min-Max Scaling
Categorical Encoding	One-Hot Encoding
Weighting Feature	Playtime (log_playtime)
Similarity Metric	Cosine
Precision@K (1-10)	0.09
Recall@K (1-10)	0.07
AUC	0.5

DATA STATISTICS & PREPROCESSING (GAMES)

	Number of Features	Number of Instances	Null Values
Raw Data	14	816	Name -> 2 Genre -> 20 Category -> 1
After Transformation	43	807	0

Preprocessing Steps
Data Cleaning
Feature Engineering
Encoding Categorical Features
Transforming Numerical Features
Scaling Numerical Features

EDA (GAMES)

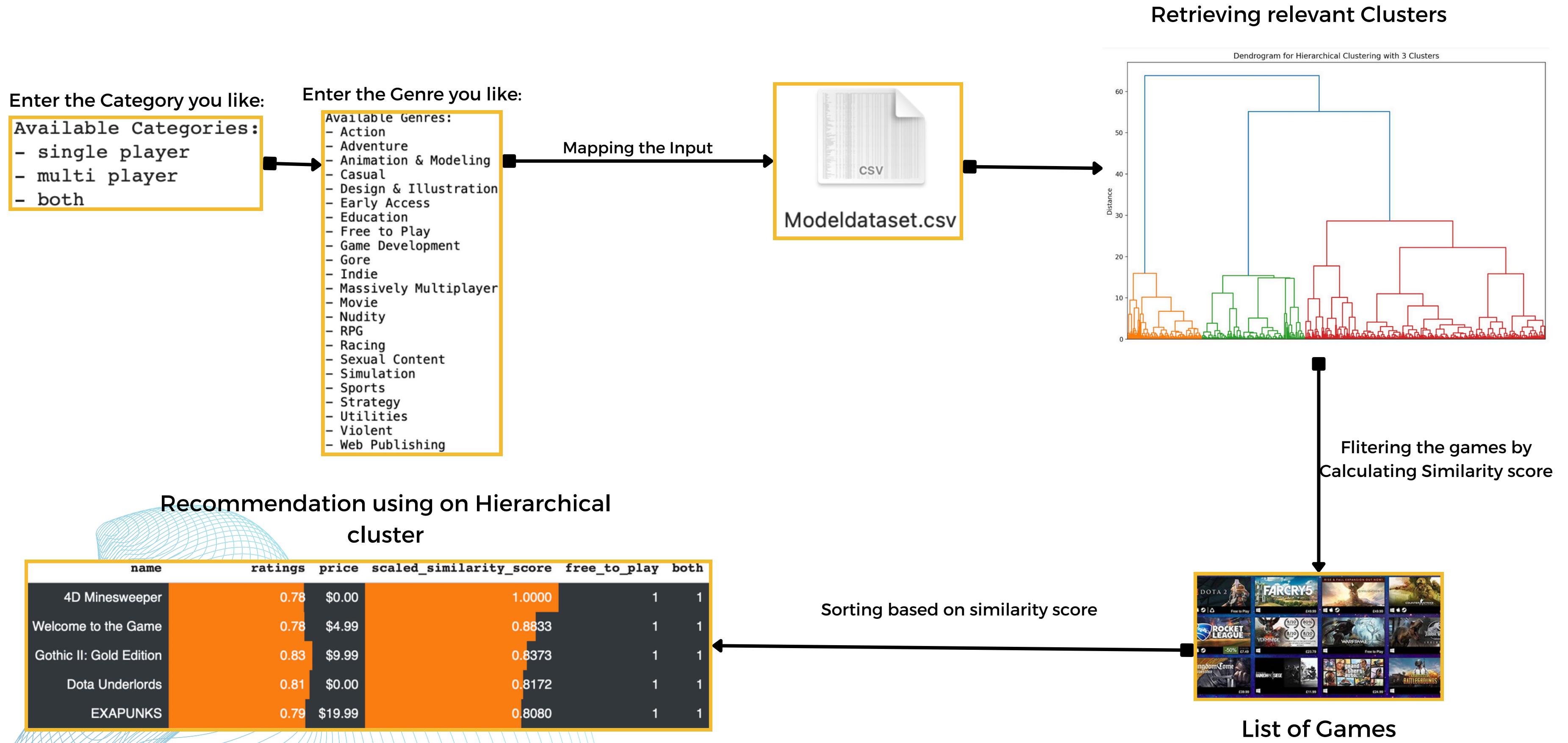


ccu: current concurrent user (as of yesterday)

Transformed columns name

'average_forever', 'average_2weeks', 'median_forever',
 'median_2weeks', 'ccu', 'price',
 'ratings', 'owners_average', 'MutualPlayerCount'

MODEL WORKFLOW (GAME PREFERENCE)



EXPERIMENT AND RESULTS

Hierarchical Clustering	
Cluster Range	2-15
Best Number of Clusters	2
Affinity	Euclidean
Linkage	Ward Method
Silhouette Score	0.45

DBSCAN	
Min Sample Range	3-10
Final Min Sample	4
Number of Cluster	12
Epsilon Range	0.5 - 0.95
Epsilon Value	0.9
Silhouette Score	0.39

DISCUSSION

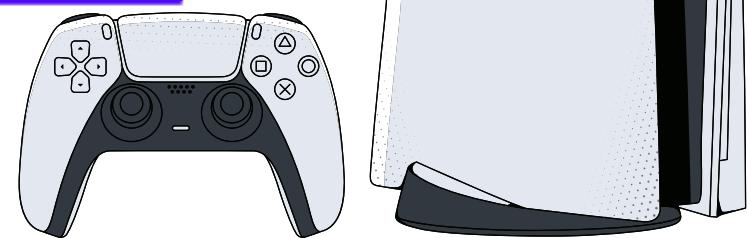
- Created 2 robust Recommender Systems generating user-centric recommendations.
- Existing user can explore new games based on their preferences.
- New users can scheme through a variety of options available and get the most trending games.
- Implemented Collaborative Filtering with Bayesian Personalized Ranking (BPR) and Content-Based Filtering to provide recommendations based on user preference.
- Implemented Agglomerative Clustering to create cohesive clusters for accurate recommendations.



FUTURE SCOPE



- **Integration of Advanced AI/ML:** Leveraging more sophisticated techniques like Neural Networks to improve game recommendations.
- **Community-based Recommendations:** Integrating recommendations based on the community favorites to enhance user experience.
- **Incorporating Targeted Marketing:** Based on user behavior campaigns or promotional offers can be provided to improve the preference of the user for that game.
- **Feedback mechanism:** Implementing a mechanism to incorporate feedback received from the user regarding recommendations and enabling proper methodology to improve them.





CONTRIBUTION

Team Member

Contribution

Abdul Sohail Ahmed

Data Exploration, Cleaning, Transformation, Feature Engineering and Modeling (User Preference)

Bhavik Patel

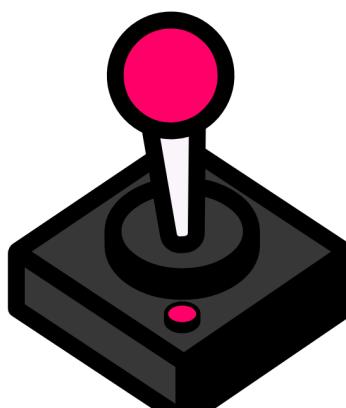
Data Exploration, Cleaning, Transformation, Feature Engineering and Modeling (Game Preference)

Poojan Gagrani

Data Extraction, Cleaning, Transformation, Feature Engineering and Modeling (User Preference)

Smeeth Sheth

Data Extraction, Cleaning, Transformation, Feature Engineering and Modeling (Game Preference)



REFERENCES

- [1] Kang, H. N., Yong, H. R., & Hwang, H. S. (2017). A study of analyzing on online game reviews using a data mining approach: Steam community data. *International Journal of Innovation, Management and Technology*, 8(2), 90-94. <https://doi.org/10.18178/ijimt.2017.8.2.709>
- [2] Cheuque, G., Guzmán, J. L., & Parra, D. (2019). Recommender Systems for Online Video Game Platforms: the Case of STEAM. *WWW '19: Companion Proceedings of the 2019 World Wide Web Conference*. <https://doi.org/10.1145/3308560.3316457>
- [3] Lomanto, W. E., Andrian, V., Achmad, S., & Sutoyo, R. (2023). Collaborative Filtering for Steam Games Recommendation. *International Conference on Cybernetics and Intelligent System*. <https://doi.org/10.1109/icoris60118.2023.10352243>
- [4] Batra, S., Sharma, V., Sun, Y., Wang, X., & Wang, Y. (2023). Steam Recommendation System. <https://doi.org/10.48550/arXiv.2305.04890>



Thank You!

Hanserkit Ghostron Yuna