



Predicting Steam Games User Count

By: Myrna Grace Calip

Table of Contents

01

Overview

04

Models

02

Business Value

05

Conclusion

03

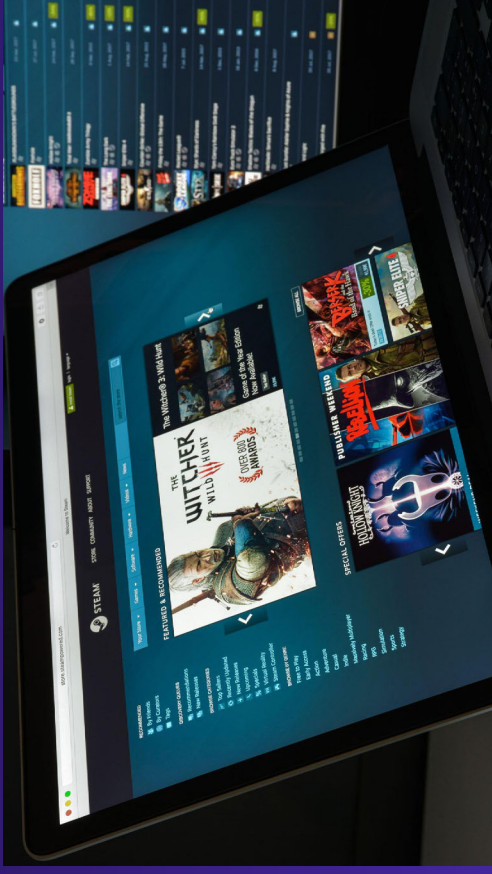
Data

06

Future Work



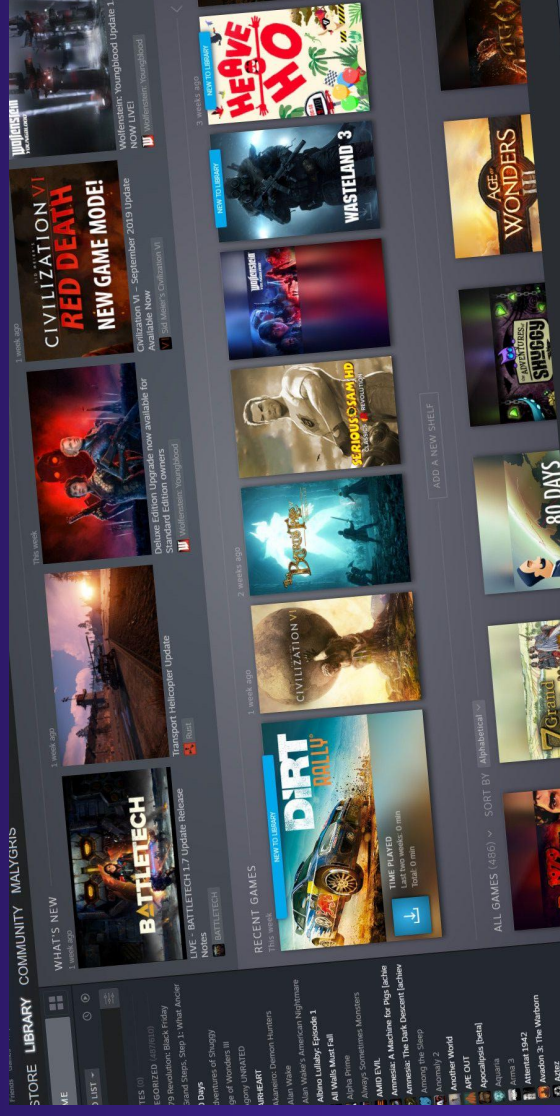
- Game Distribution Platform
- 75% of global market share
- Released in 2003 by Valve
- Offers 50,000+ games (May 2022)
- Games varies from AAA to Indie



Source: comparecamp.com

Business Value

Goal: Provide game developers and publishers insights on which game features influences high user count on Steam.



Data Acquisition

6000 Games Data Extracted from:

- steamspy
- steamworks
- NikDavid Kaggle Dataset

How?

- API
- All request
- Iterate through Steam App list



Data Preparation

Clean:

- Extracted: 6000 games
- Remove missing values
- Change columns values to simple delimited list
- Final Clean Data: 5489

Preprocessing & Feature Engineer:

- Create categorical columns
- One Hot Encode
- Create Target Column
- # of columns: 280

Model

Target: Usercount

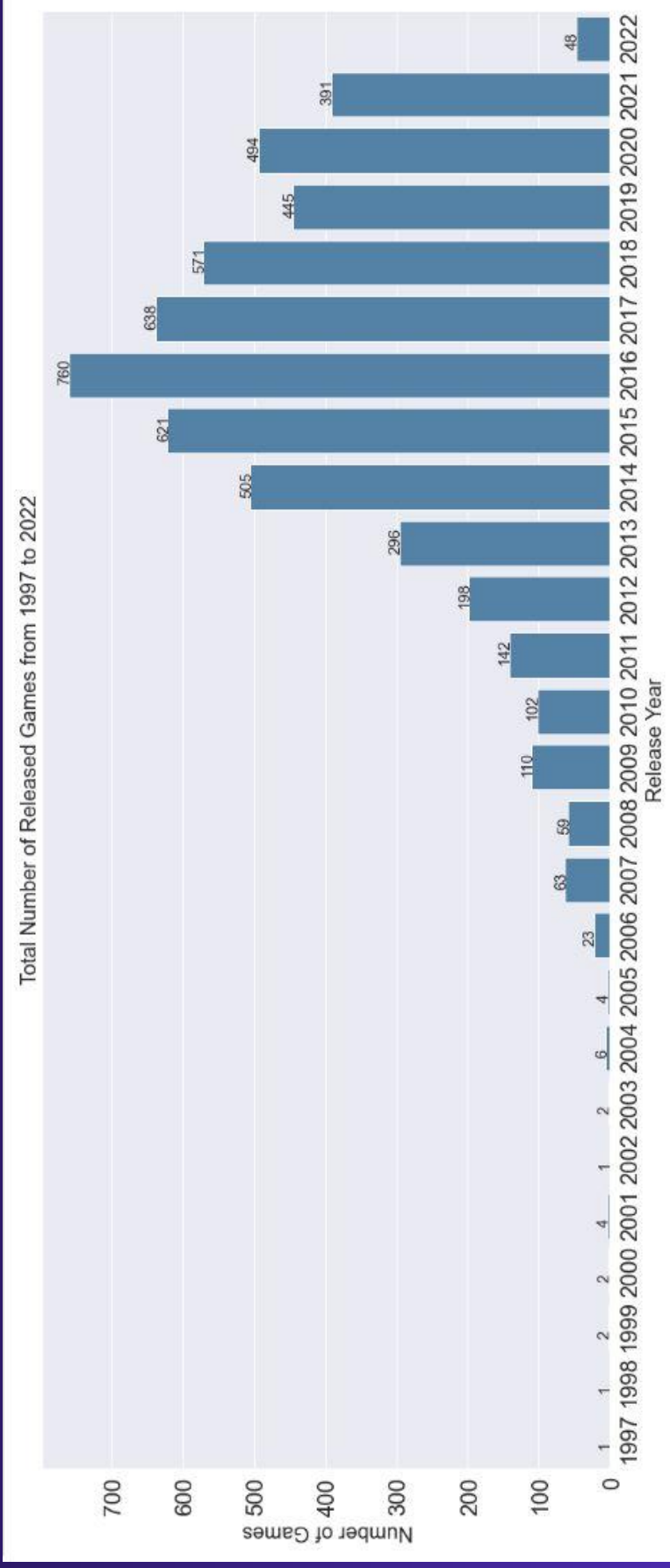
Features:

- Tags,
- Categories,
- Discount,
- Release
- Year,
- Rating,
- Max Revenue
- Genre

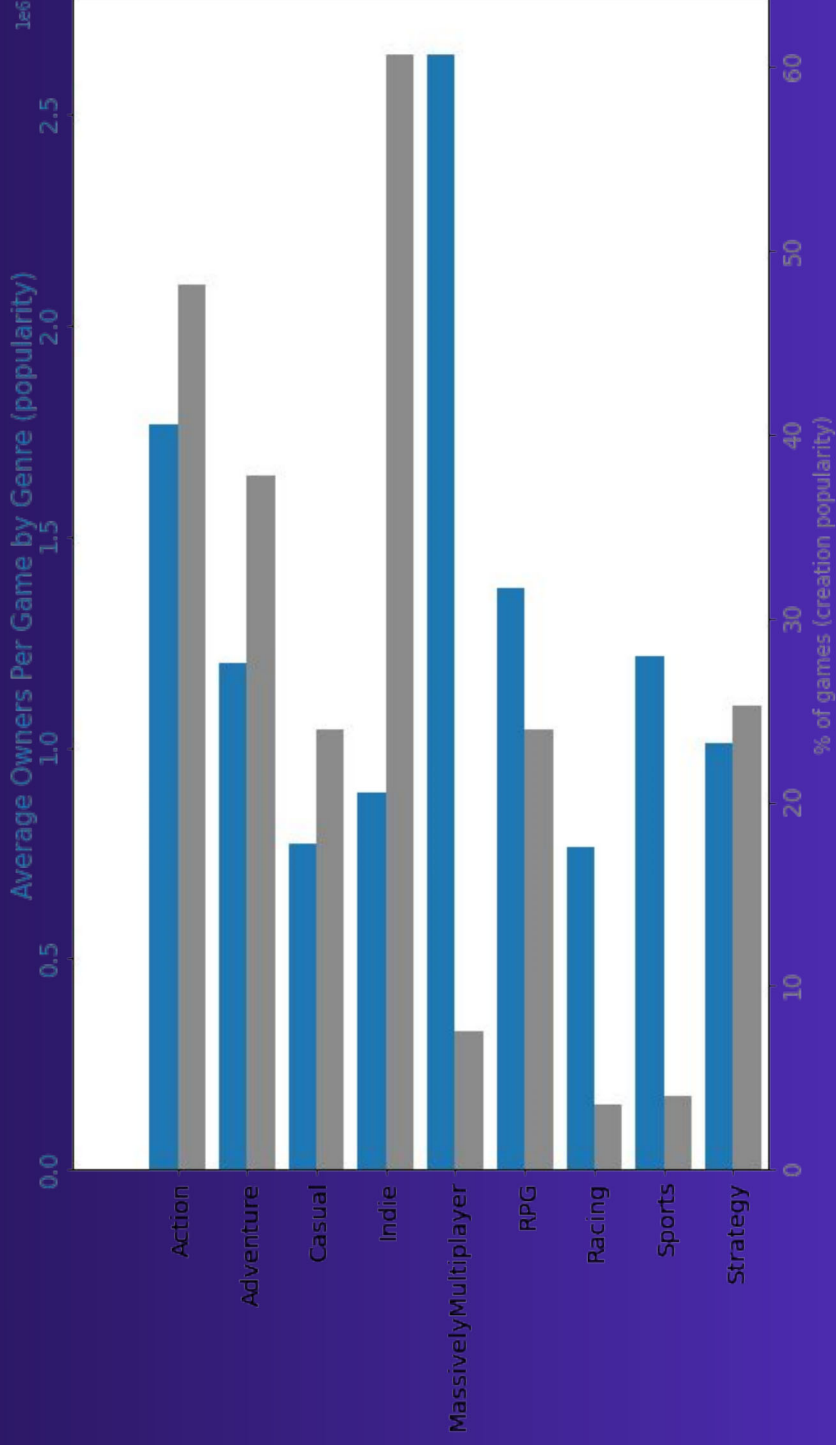
Models Tested

- Binary Classification Threshold: 1 Million Users
- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost

Total Number of Games Released



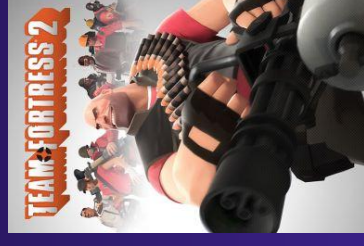
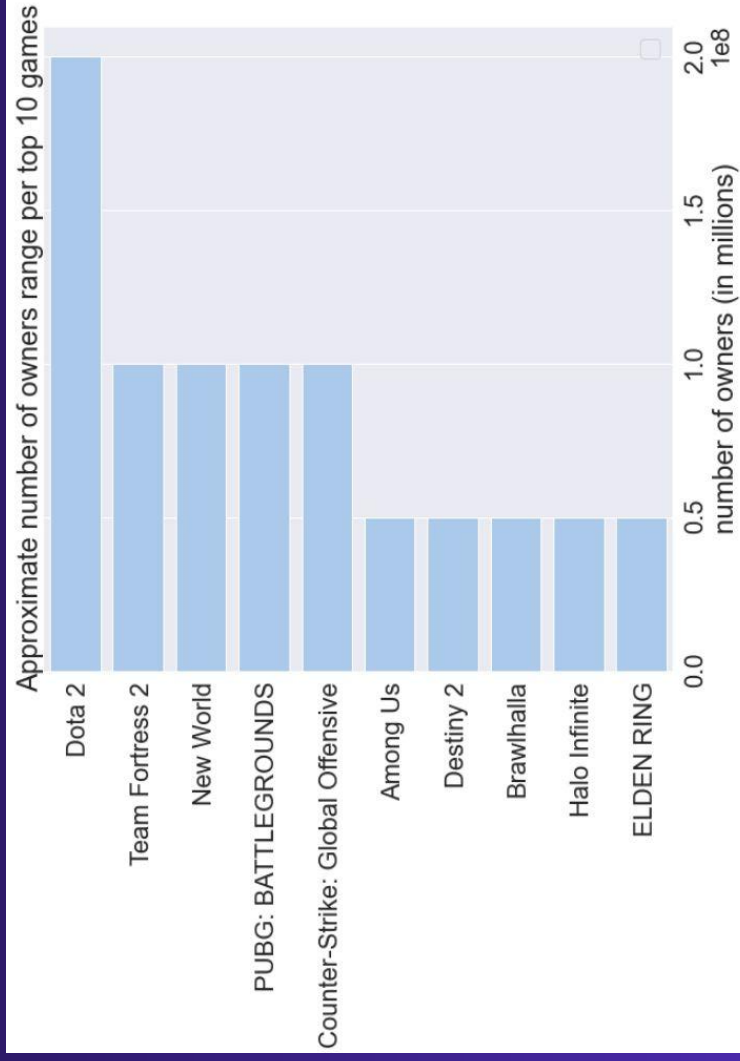
Genre Popularity by Average Owners



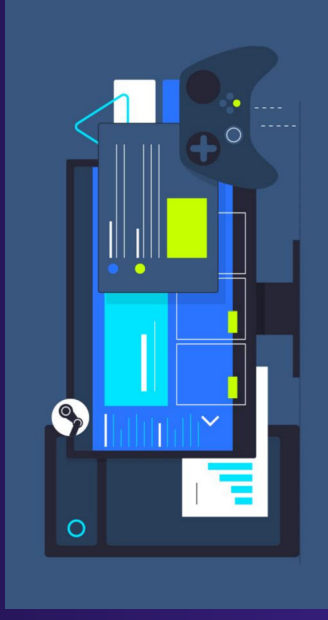
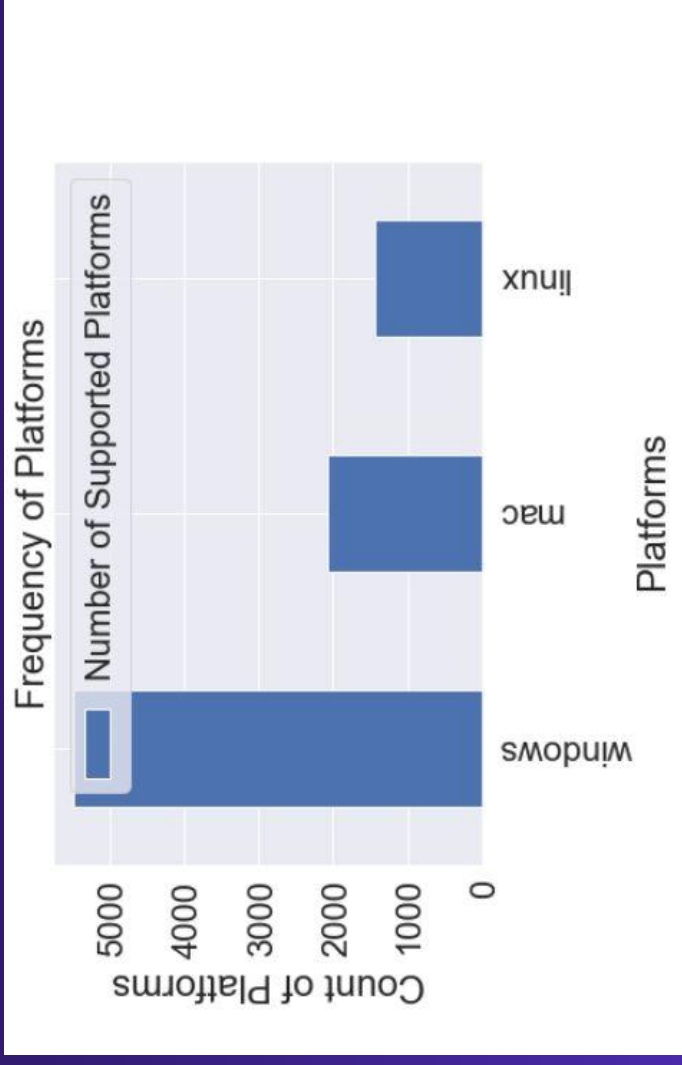
**Which game currently has the highest
user count?**



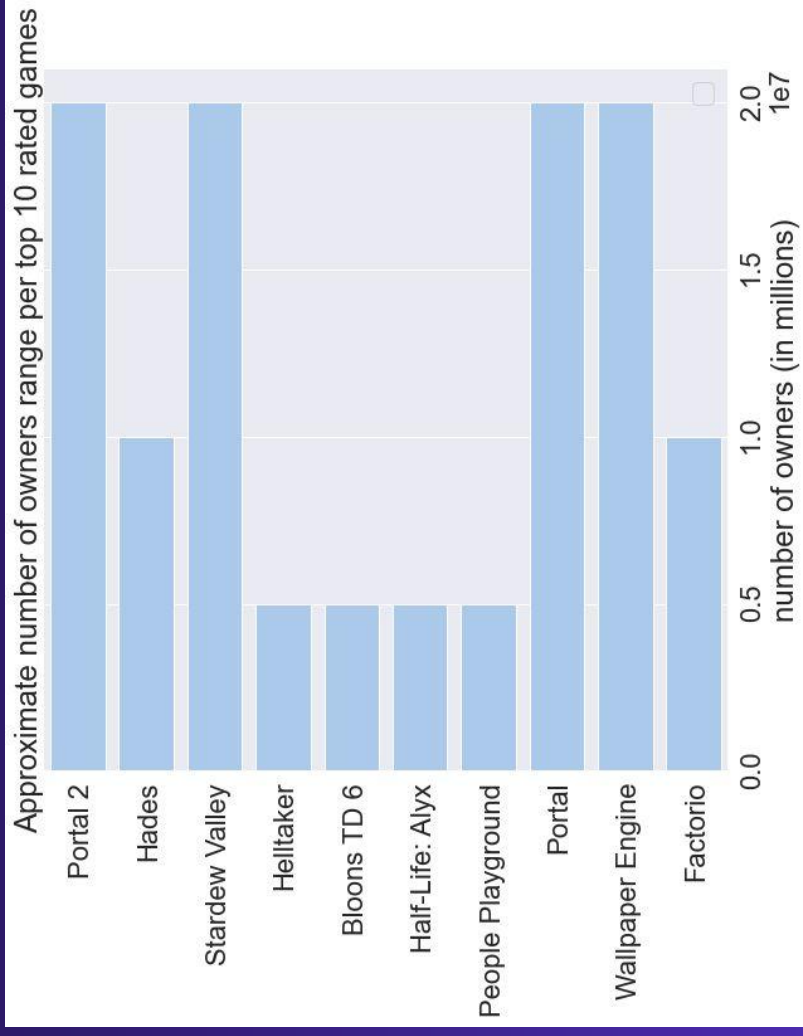
Answer: Dota 2!



Most used Platform for Steam Games



Owner Count for Top 10 Rated Games



Model Metrics

- Highest F1 Score for user count ≥ 1 million to predict % of positive predictions
- Highest AUC Score = measures All Positive and Negative Correctly
- Scores are from final tuning and tweaking

Logistic Regression

- Test AUC Score: 92%
- F1 Score: 84%



Tuned XGBoost

- Test AUC Score: 97%
- F1-Score: 87%

Tuned Random Forest

- Test AUC Score: 85%
- F1 Score: 74%



Tuned Decision Tree

- Test AUC Score: 81%
- F1 Score: 72%

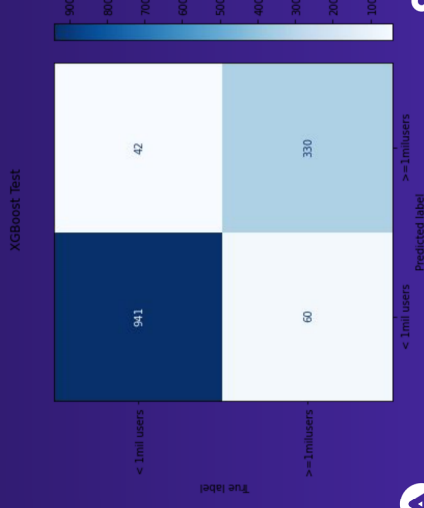
Conclusion

- Our chosen model to follow our results with is our Tuned XGBoost Model
- Each model improved with every small tweak and tuning
- All our Models followed a pattern of being able to predict less than 1 million game owners better than over 1 million owners



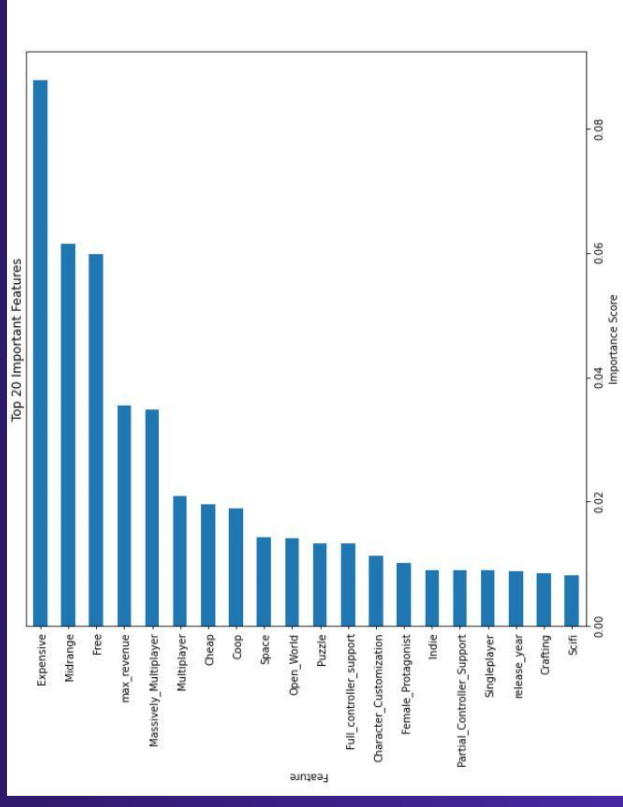
	precision	recall	f1-score	support
0	0.94	0.96	0.95	983
1	0.89	0.85	0.87	390
accuracy				
macro avg	0.91	0.90	0.91	1373
weighted avg	0.93	0.93	0.93	1373

Accuracy Score = 0.92571
Log Loss = 0.25640



Recommendations

- **Release a Massive Multiplayer that allows more users to play with other players**
- **MMO = More players owning your game**
- **Make the setting of the game in space that is open for players to explore that requires leveling up by solving puzzles**
- **Mixture of Among Us, Breath of The Wild, and Risk of Rain 2**



Future Work



- Use different targets to predict popular Steam Games
- Game Review Sentiment Analysis
- Linear Regression on Game Revenue
- NLP System Requirements

Thanks!



Do you have any questions?

Email: myrnagrace@gmail.com
GitHub: <https://github.com/myrnagrace>

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**

