



DataCorp.™

Steam Games Analysis

John Paul Hernandez Alcala



Problem Statement

01

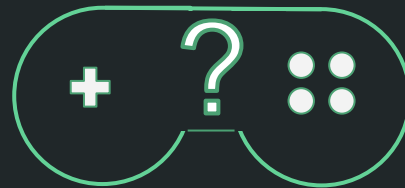
Investigate how some game features affect owner count outcome

02

Derive models that can predict above normal game ownership

03

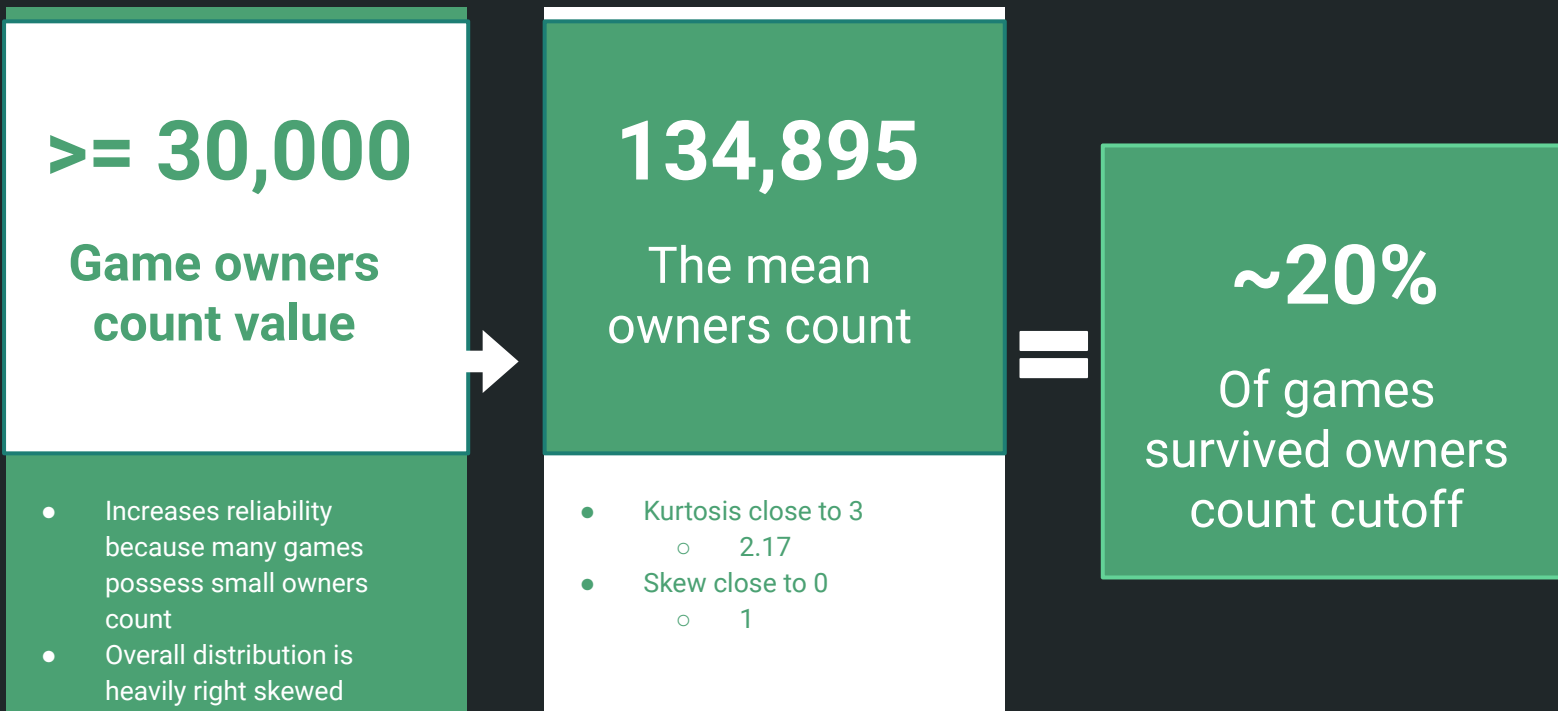
Propose 3 business recommendations for game design



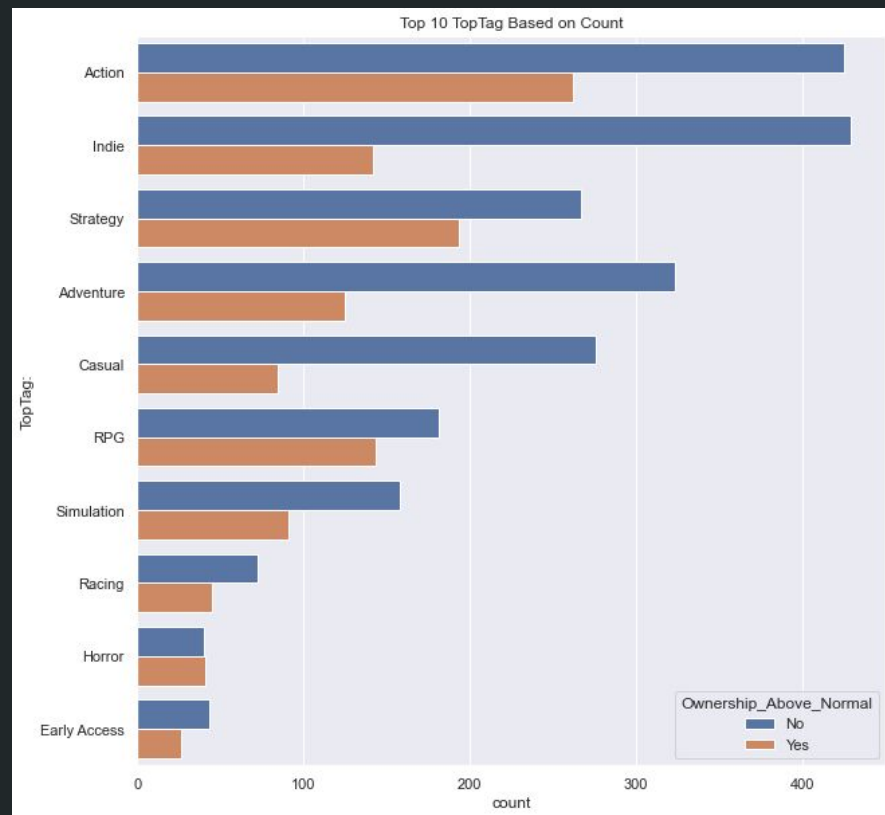
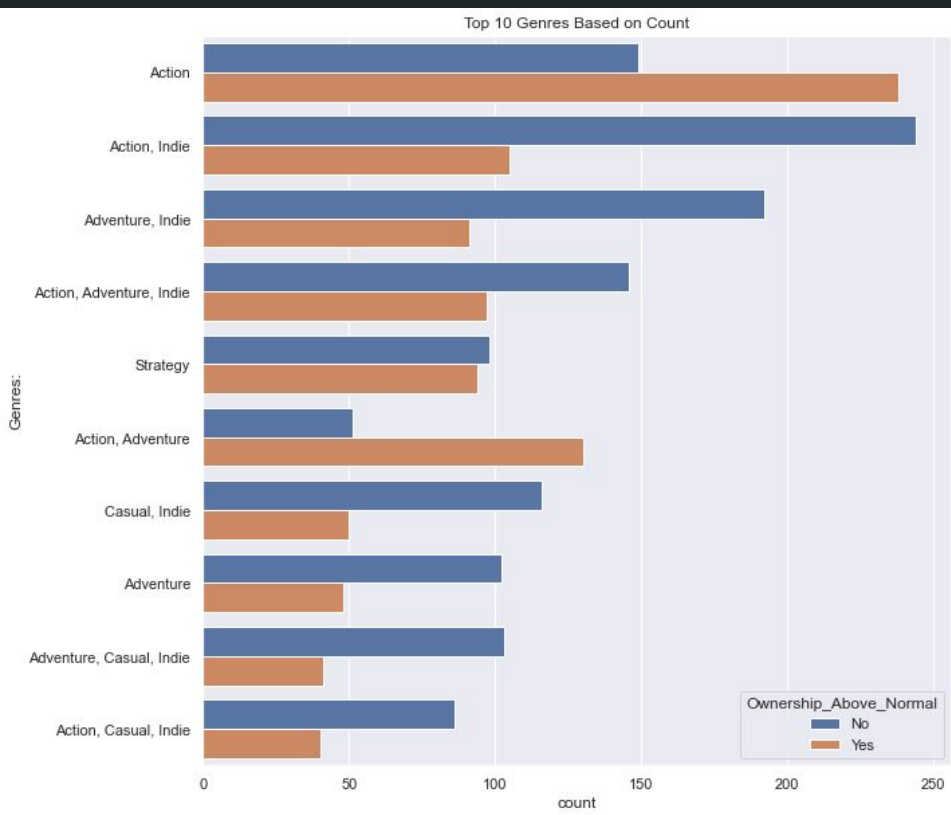
Methodology



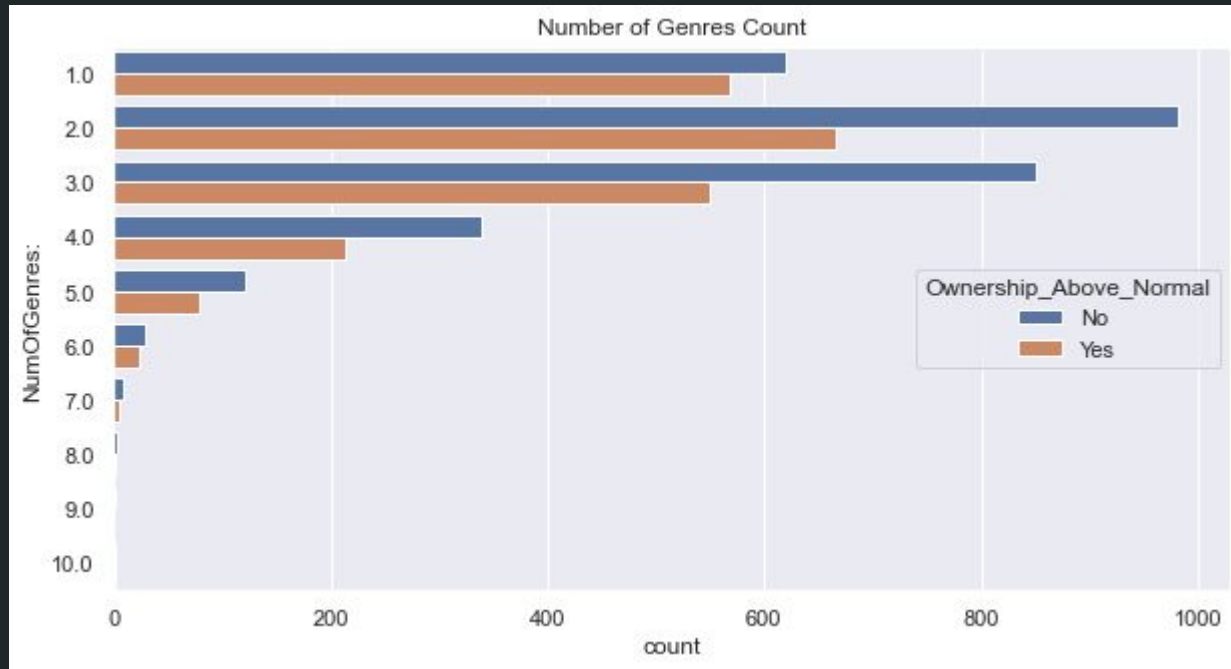
Metric Definition: Above Average Ownership



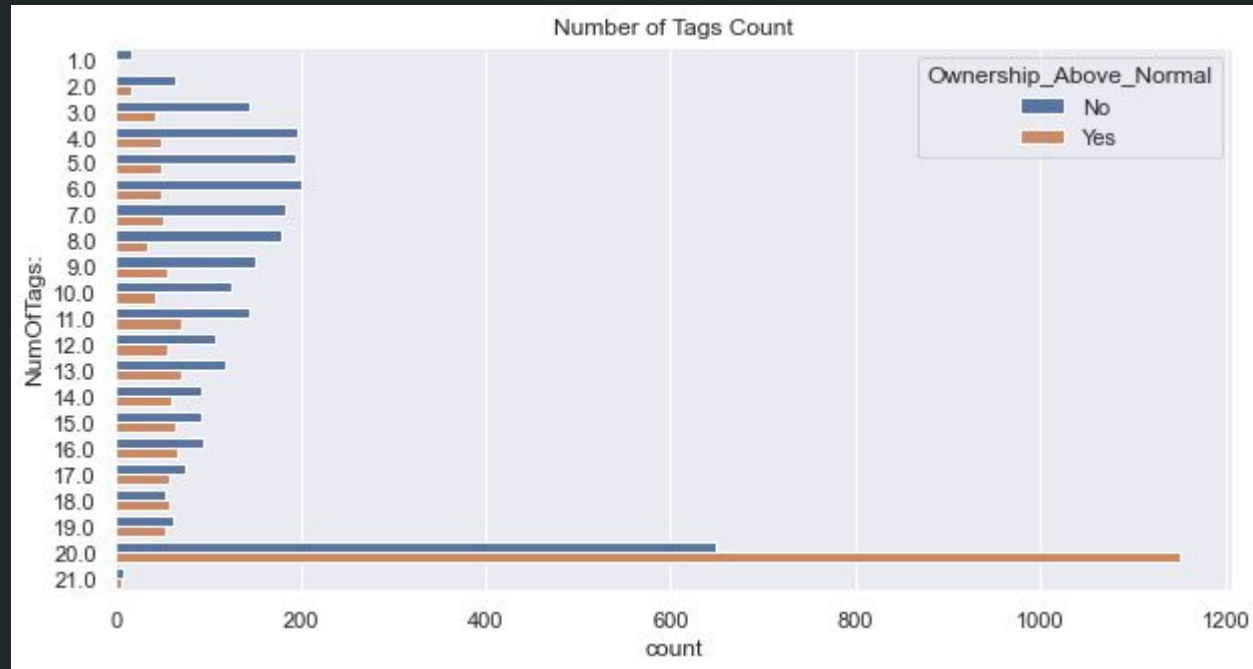
Bar Graph of 'Genres', and 'TopTag' Bar Graphs



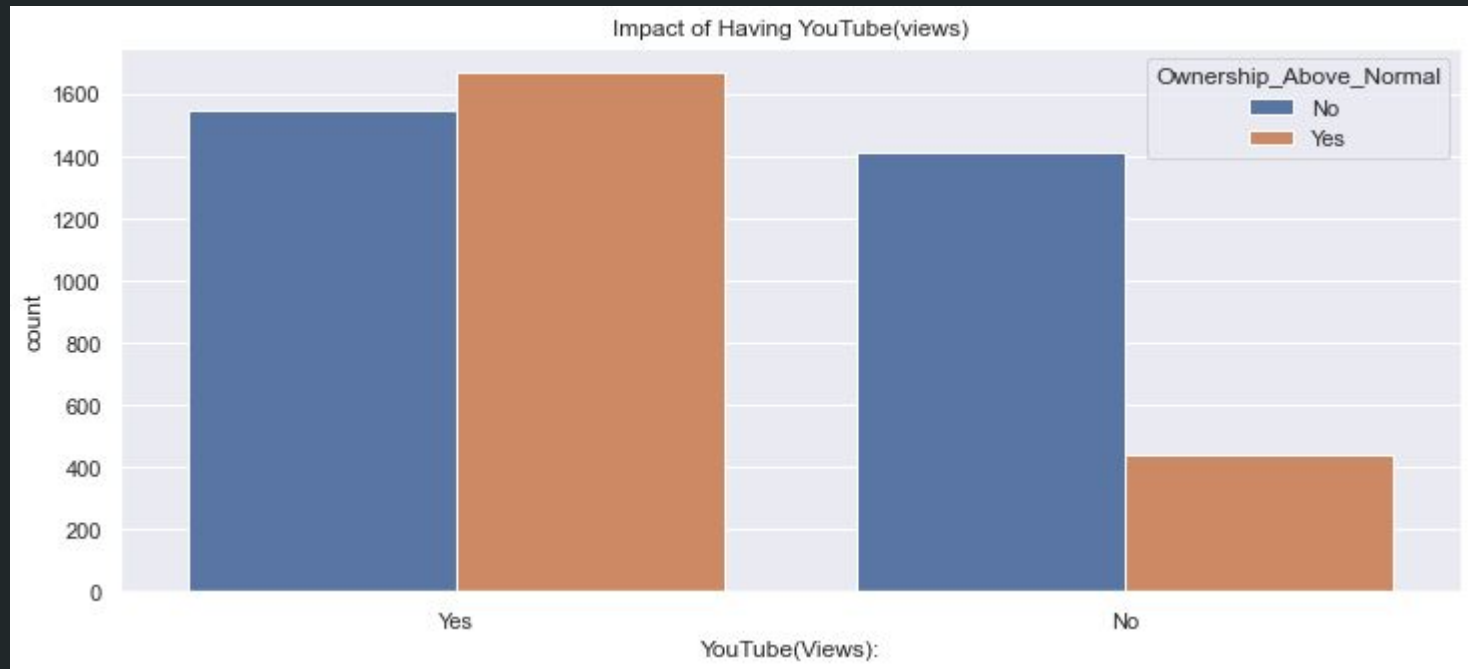
Bar Graph of 'Genres', and 'TopTag' Bar Graphs



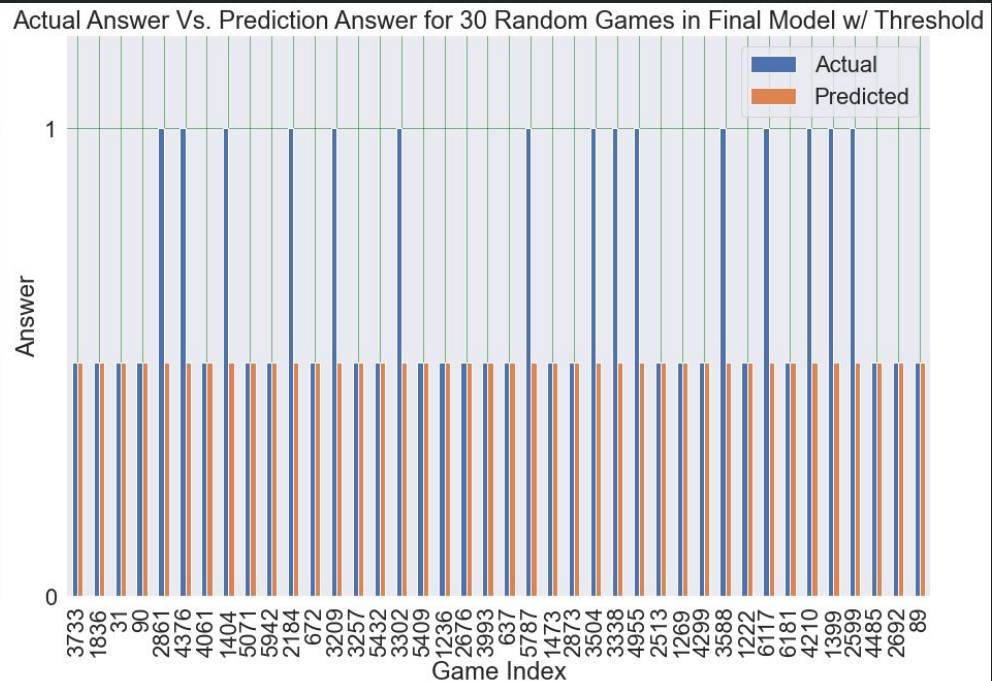
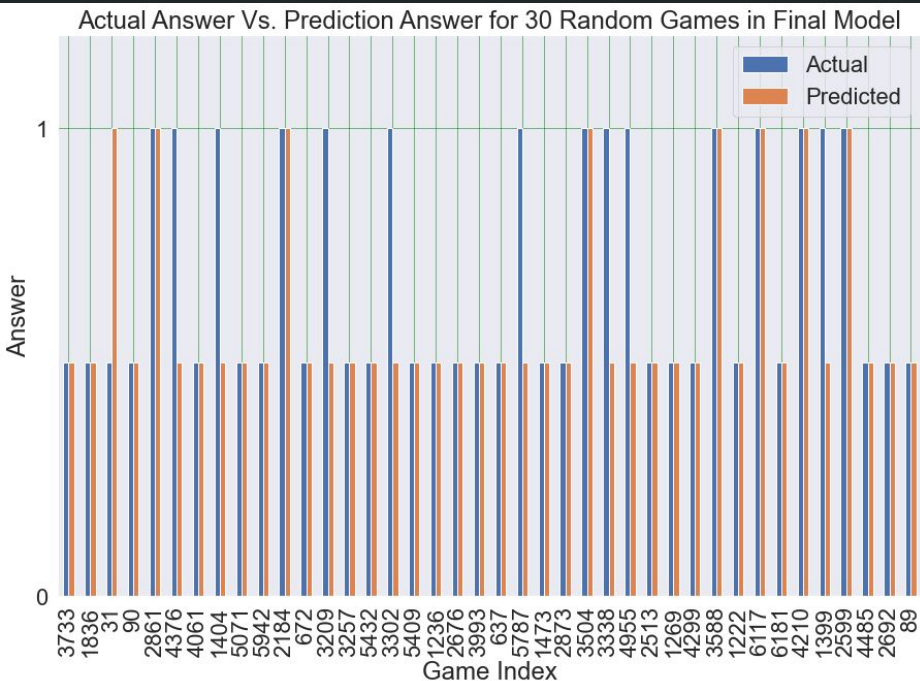
Bar Graph of 'Genres', and 'TopTag' Bar Graphs



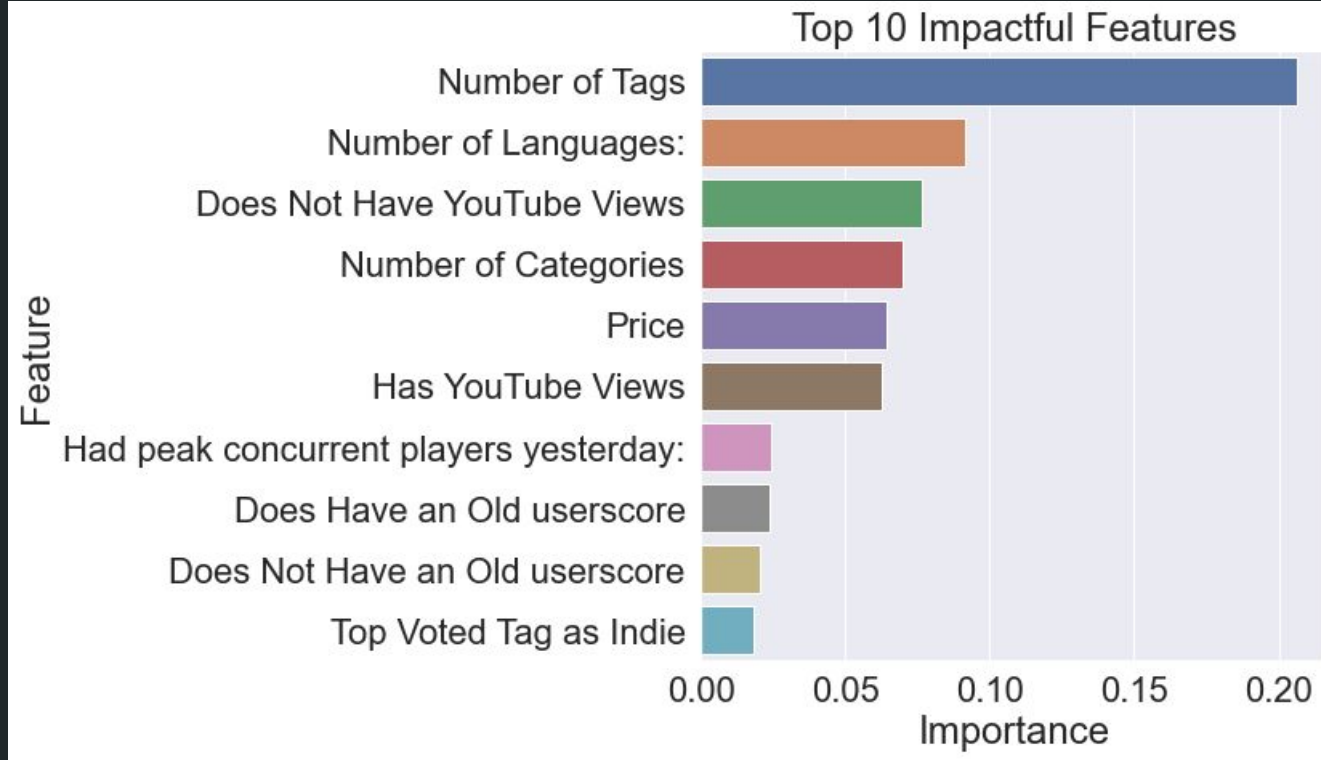
Bar Graph of 'YouTube(views)'






Final Model vs Final Model with Threshold



Top 10 Features



Recommendations Derived from Model and Analysis

-  1
 - Above average ownership is highly dependent on number of tags which should be greater than 6 in general
-  2
 - The probability of a game having above average ownership is higher when the game has YouTube views.
-  3
 - Action by itself as a genre has the highest above average ownership quotient and count



Future Work

01	More advanced model	<ul style="list-style-type: none">• Time-series based driven model with ownership over time to evaluate sustained ownership
02	Employ Other Imbalanced Data Techniques	<ul style="list-style-type: none">• Right evaluation metrics, resample the training set (under-sampling or over-sampling), cluster the abundant class, anomaly detection
03	More data	<ul style="list-style-type: none">• Obtain more data from other sources such as Steam Charts





Get to designing!

Thank You

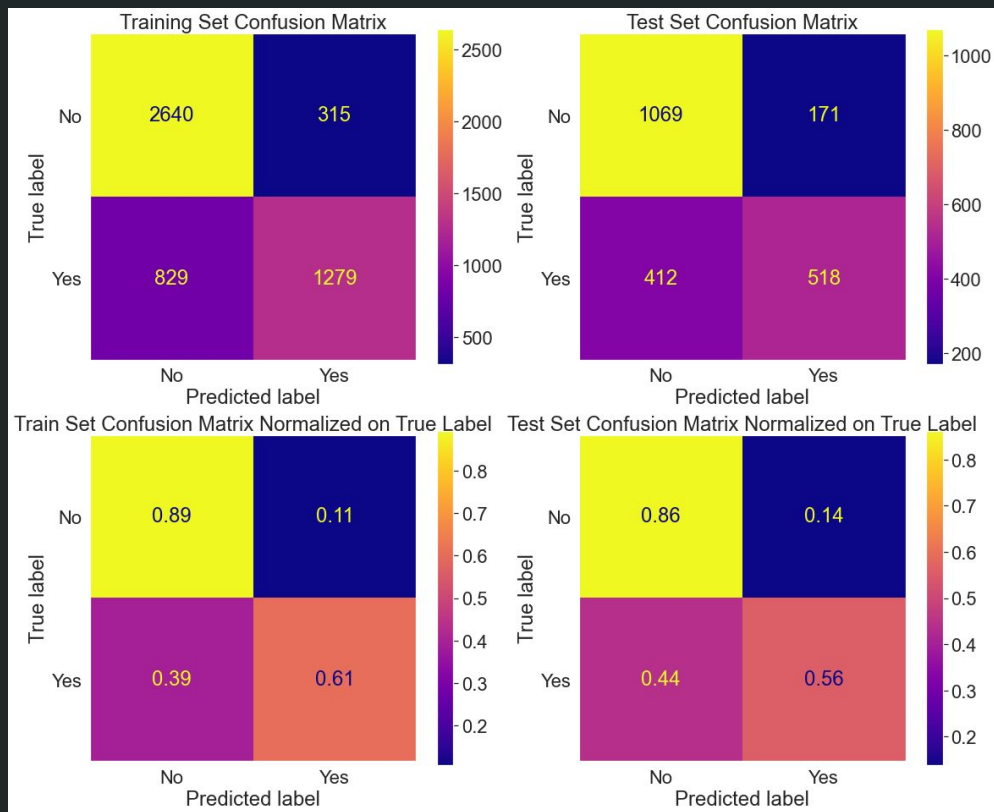
Appendix 1a. Features from Dataset

```
## Return format for an app: ##
```

```
* appid - Steam Application ID. If it's 999999, then data for this application is hidden on developer's request, sorry.  
* name - game's name  
* developer - comma separated list of the developers of the game  
* publisher - comma separated list of the publishers of the game  
* score_rank - score rank of the game based on user reviews  
* owners - owners of this application on Steam as a range.  
* average_forever - average playtime since March 2009. In minutes.  
* average_2weeks - average playtime in the last two weeks. In minutes.  
* median_forever - median playtime since March 2009. In minutes.  
* median_2weeks - median playtime in the last two weeks. In minutes.  
* ccu - peak CCU yesterday.  
* price - current US price in cents.  
* initialprice - original US price in cents.  
* discount - current discount in percents.  
* tags - game's tags with votes in JSON array.  
* languages - list of supported languages.  
* genre - list of genres.
```



Appendix 1b. Confusion Matrix for Final Model



Appendix 1b. Confusion Matrix for Final Model with custom threshold



Appendix 2. ROC-AUC Curve Graph

