Steaming Hot

A Recommendation Engine for Game Developers

Alex, Eduardo, Joseph, Kanishk, Ling, Rohit



Mentor: James Bramante

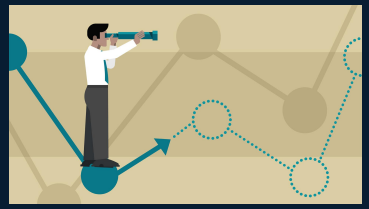
Problems and Goals



Development:
What features should game developers focus on?

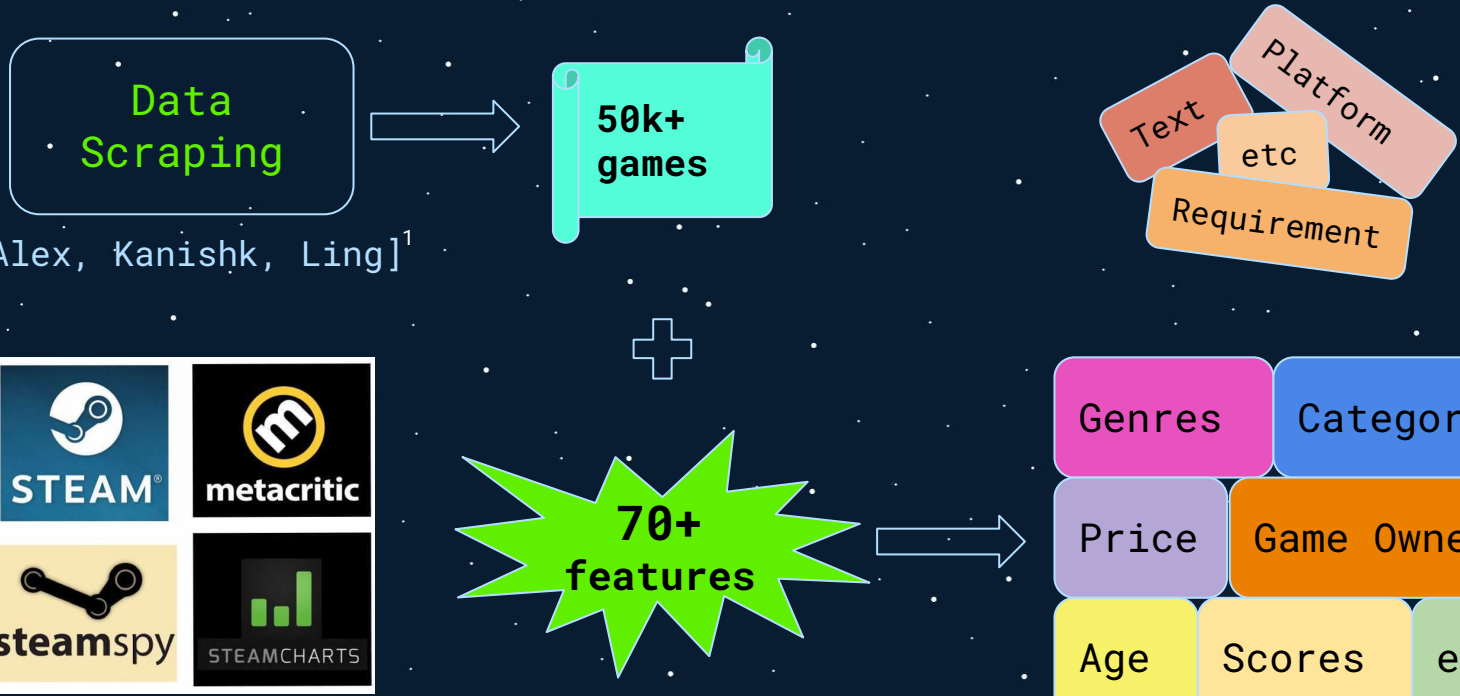


Deployment:
How to present a game in a Steam library store? What is the expected revenue?



Down The Road
How will the game perform in the future? What is the next decision?

Approach and Summary



[1]: files from https://github.com/CraigKelly/steam-data has been used.

Approach and Summary

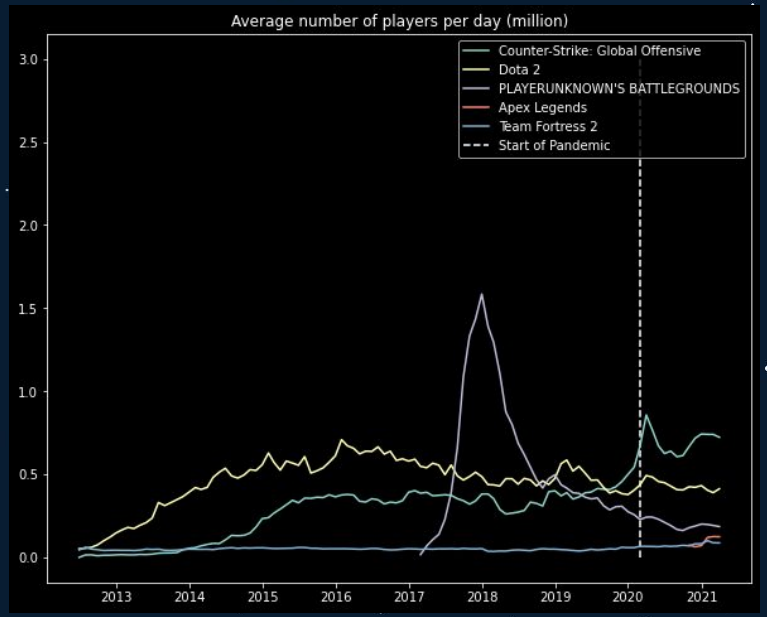
Data Scraping

[Alex, Kanishk, Ling]

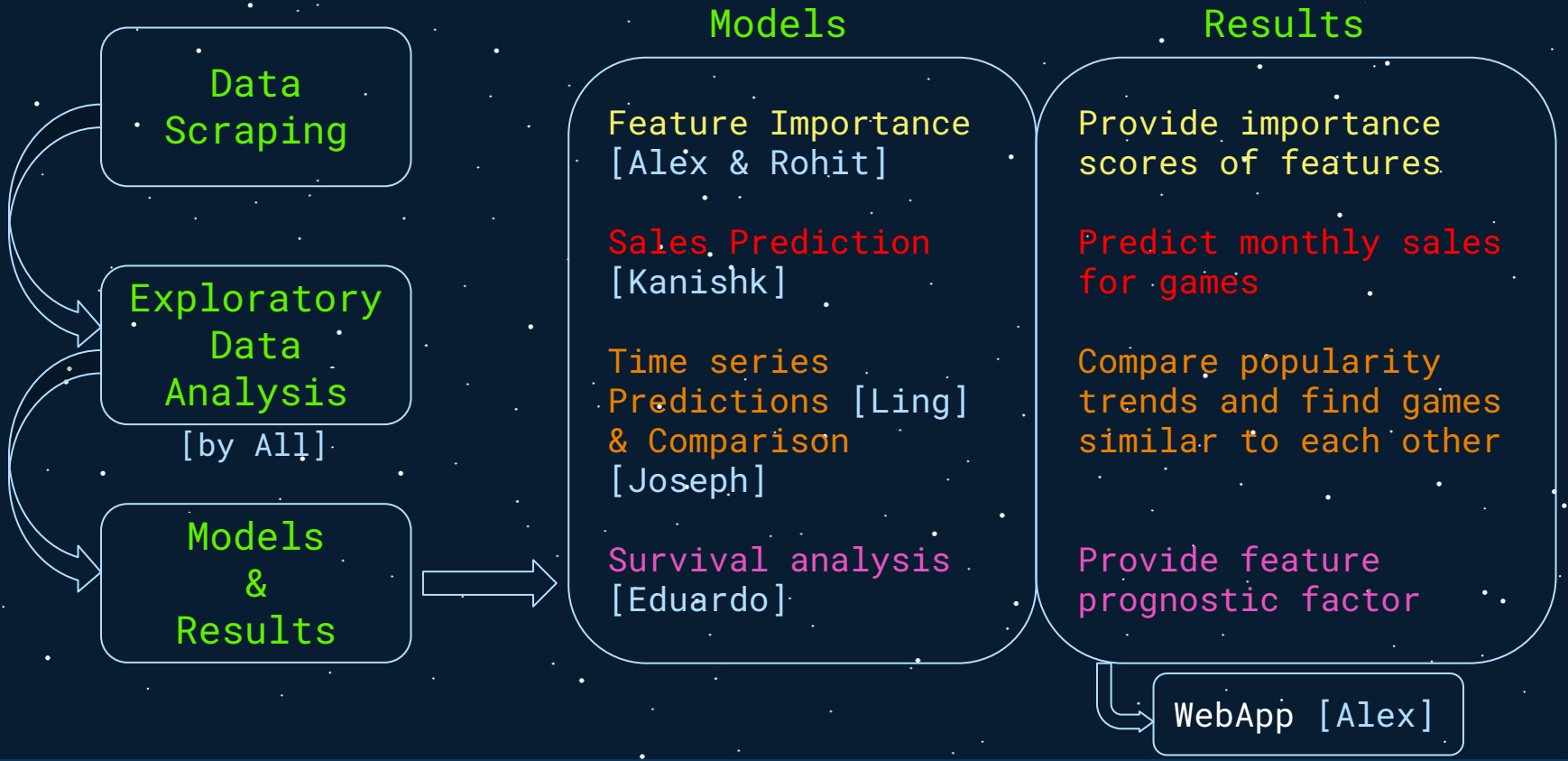
10k+ games



106 months

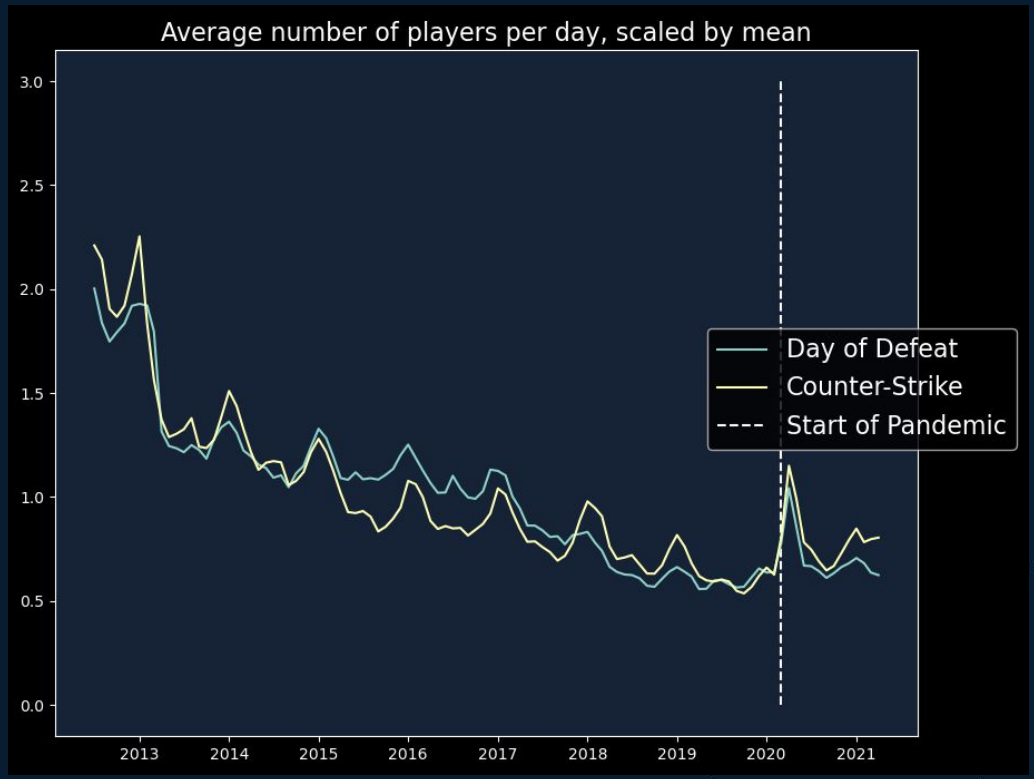


Approach and Summary



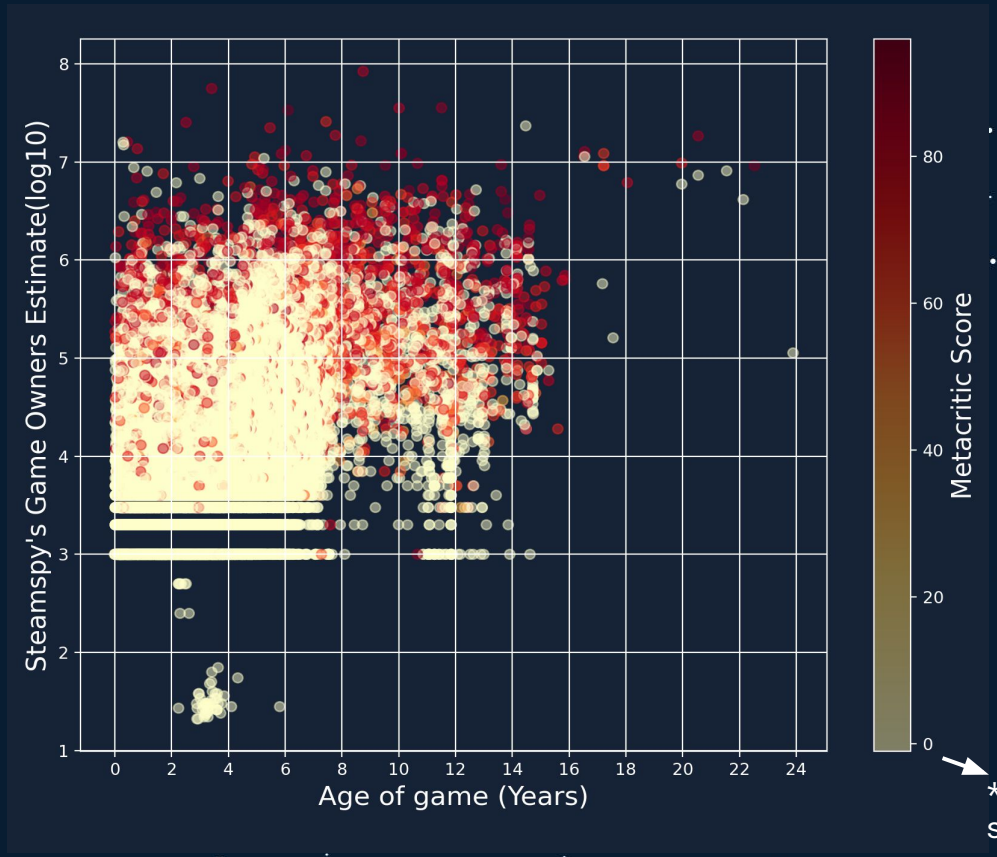
Exploratory Data Analysis



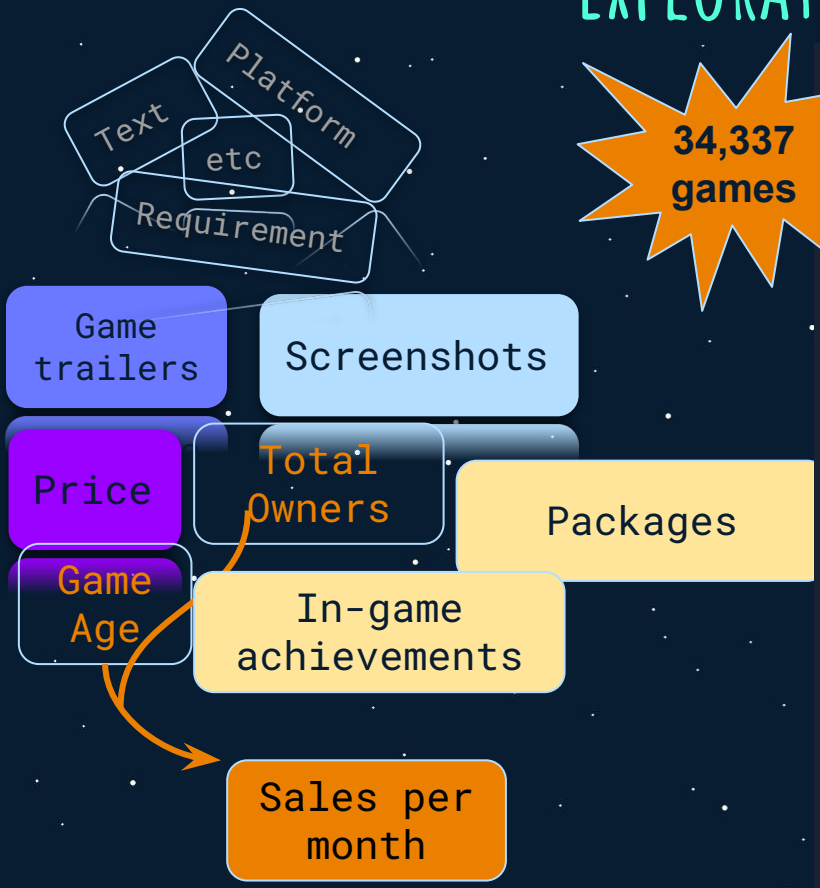


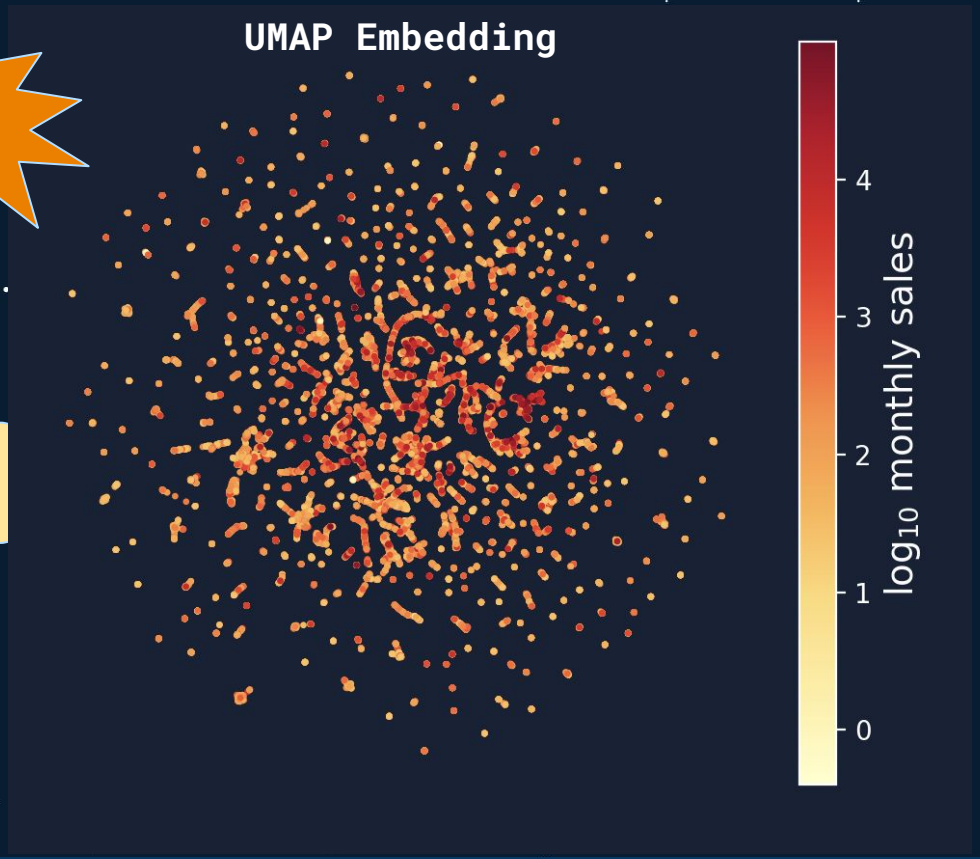
Exploratory Data Analysis



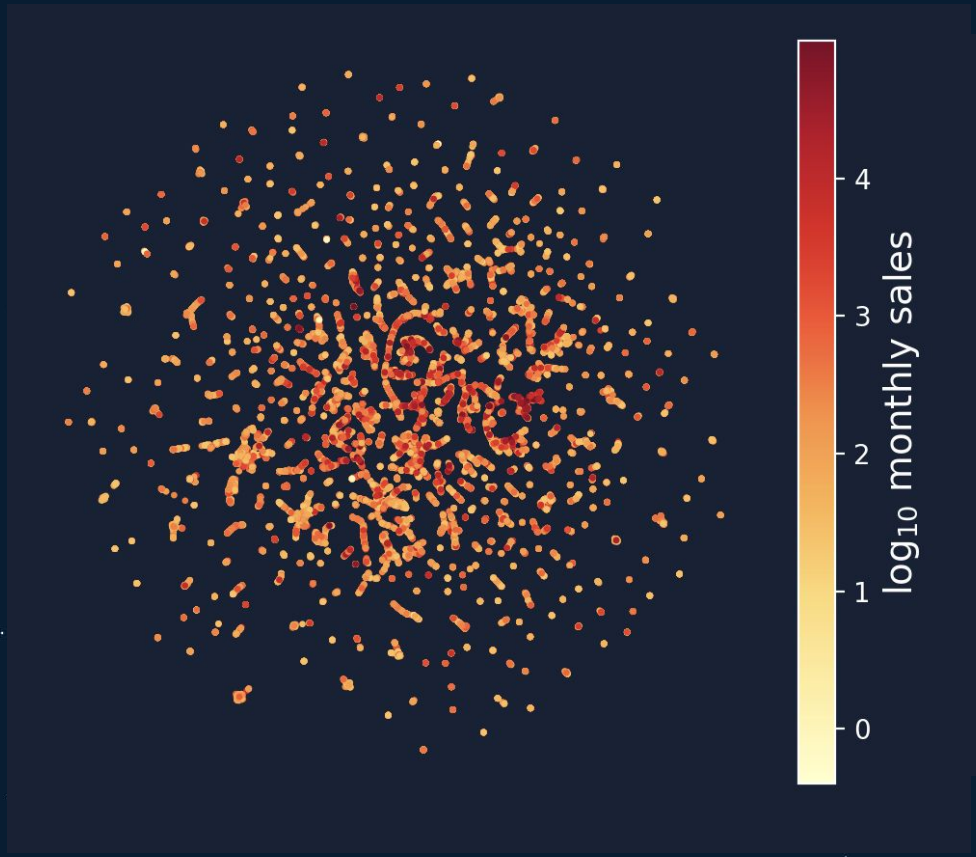


Exploratory Data Analysis





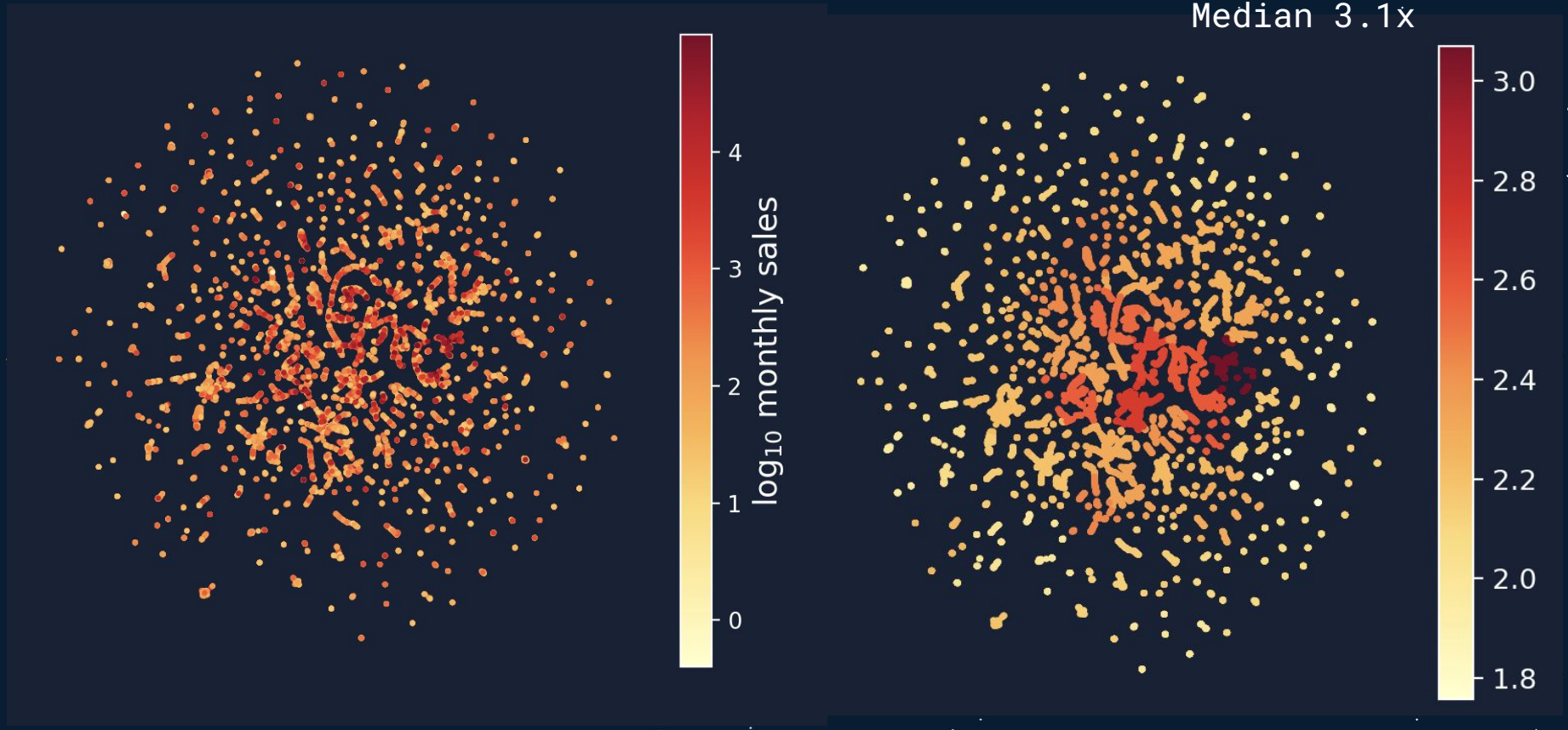
Sales prediction Model





Sales prediction Model

Errors
Mean 3.8x
Median 3.1x



Categorical features!



How To Select Features of a New Game?

- Heuristic Approach

35 categorical features

Target variable:“Popularity Rating”

IsFree

GenreIsAction

PlatformMac

CategoryMMO

ControllerSupport

Most Popular (5.0)
Popular (4.0)
Somewhat Popular (3.0)
Unpopular (2.0)
Most Unpopular (1.0)

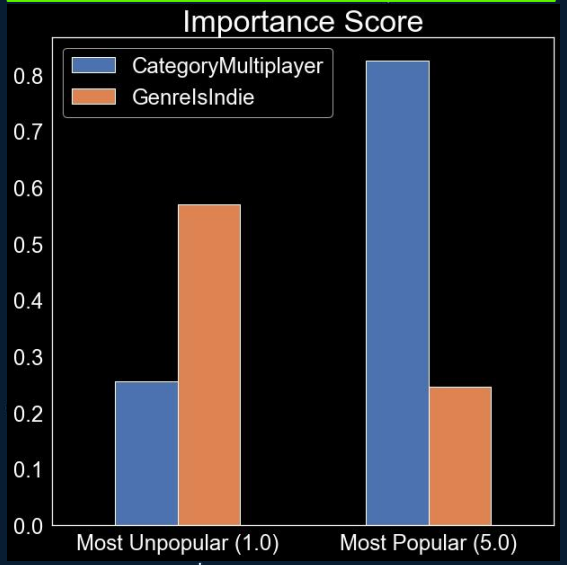
- Machine Learning
  - Classification
  - Survival Analysis

See our Github page for more!



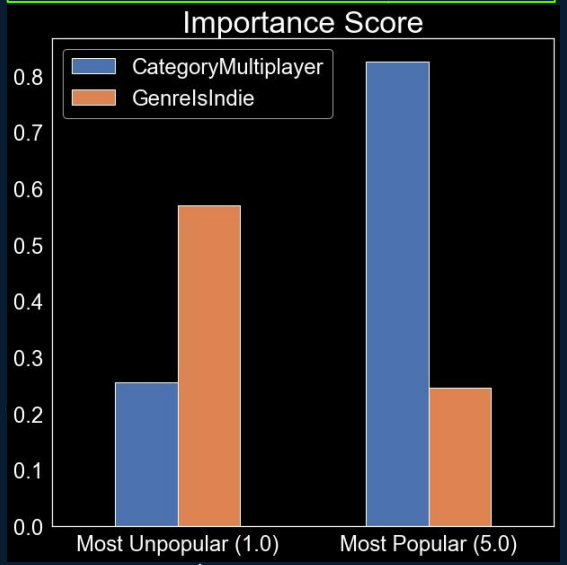
Heuristic Scores For Feature Selection

$I_i(F)$ = Fraction of games in class C_i with feature F

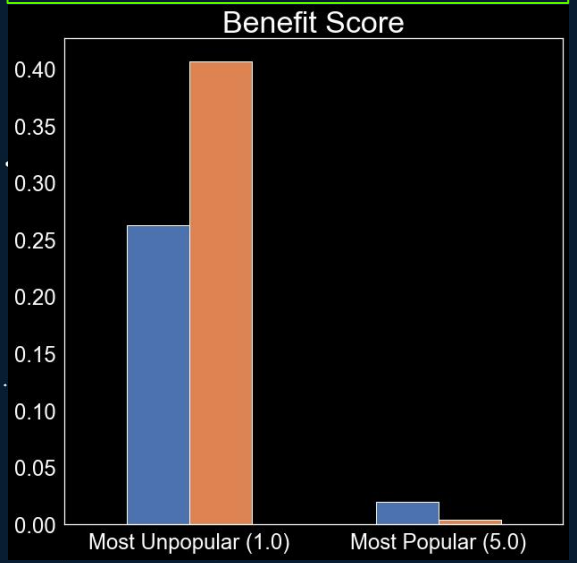


Heuristic Scores For Feature Selection

$I_i(F)$ = Fraction of games in class C_i with feature F

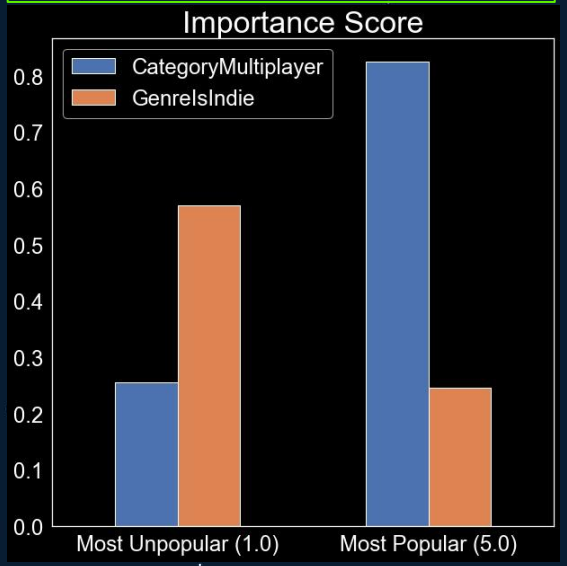


$B_i(F)$ = Fraction of games with feature F in class C_i

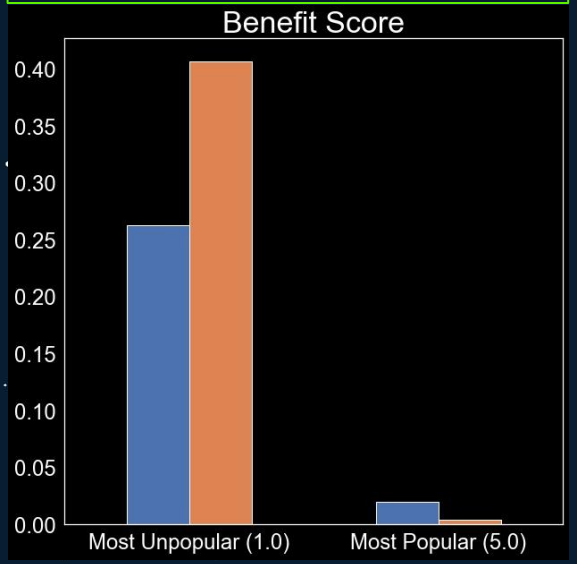


Heuristic Scores For Feature Selection

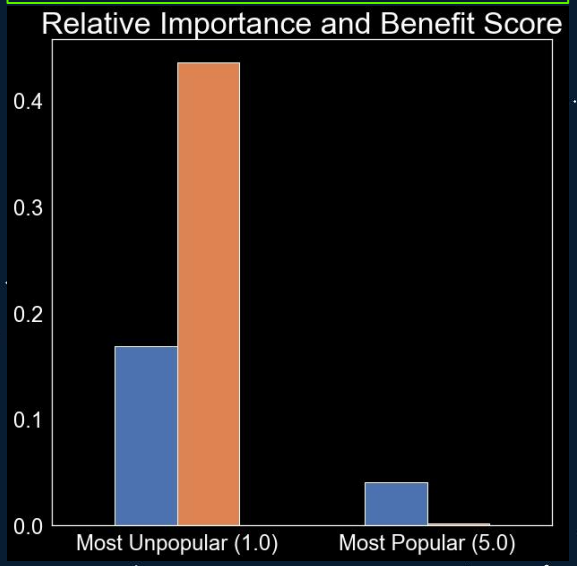
$I_i(F)$ = Fraction of games in class C_i with feature F



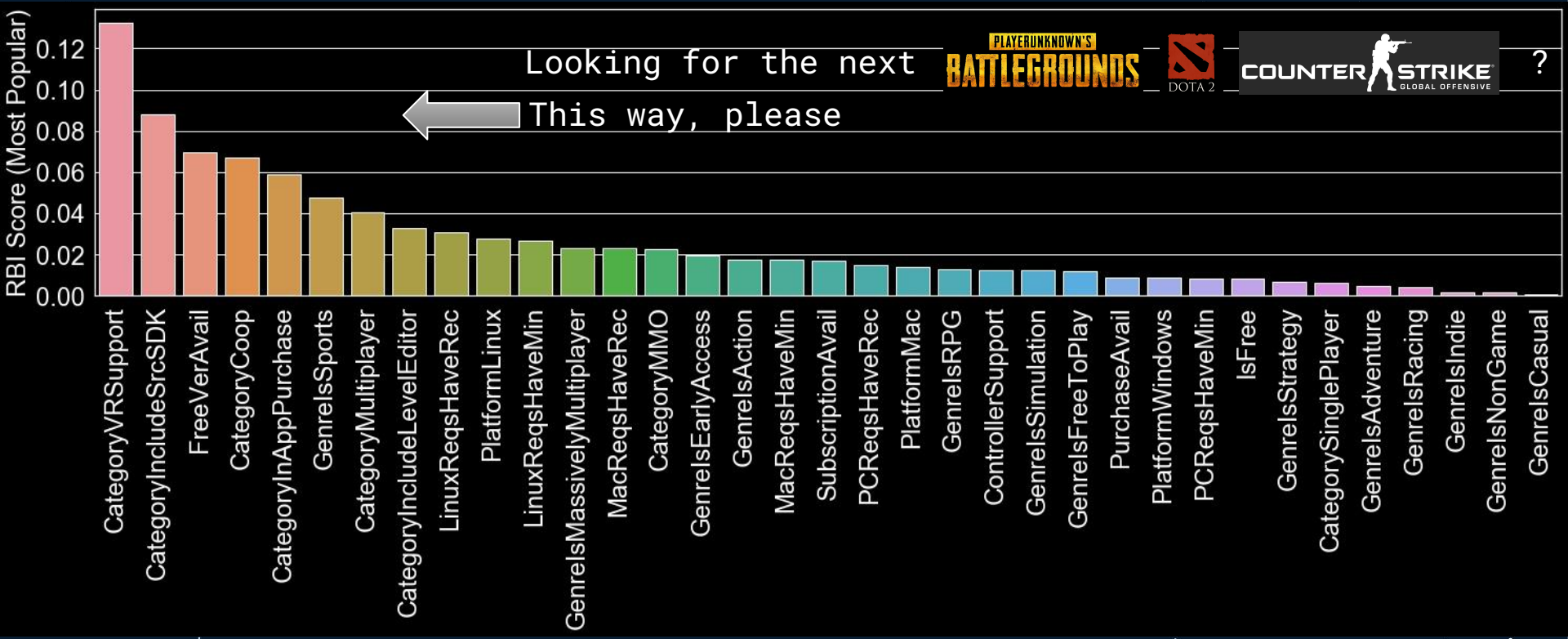
$B_i(F)$ = Fraction of games with feature F in class C_i



$RBI_i(F) = I_i(F)*B_i(F)$

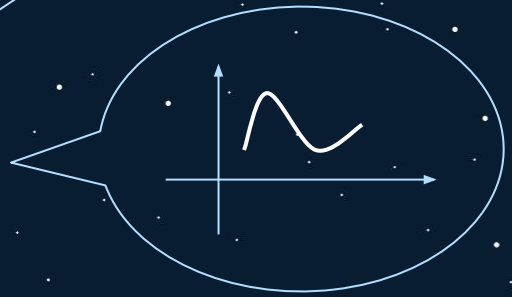


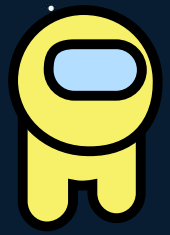
Heuristic Feature Selection



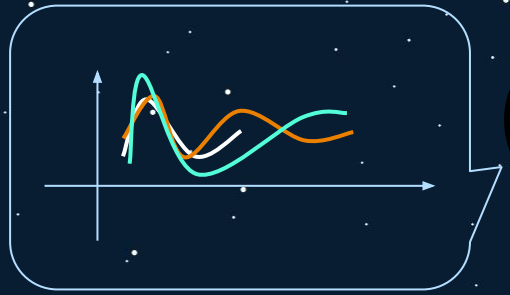
Time Comparison

My game is 1 year old, what to do next?





Learn from the past you must, find guidance you will.

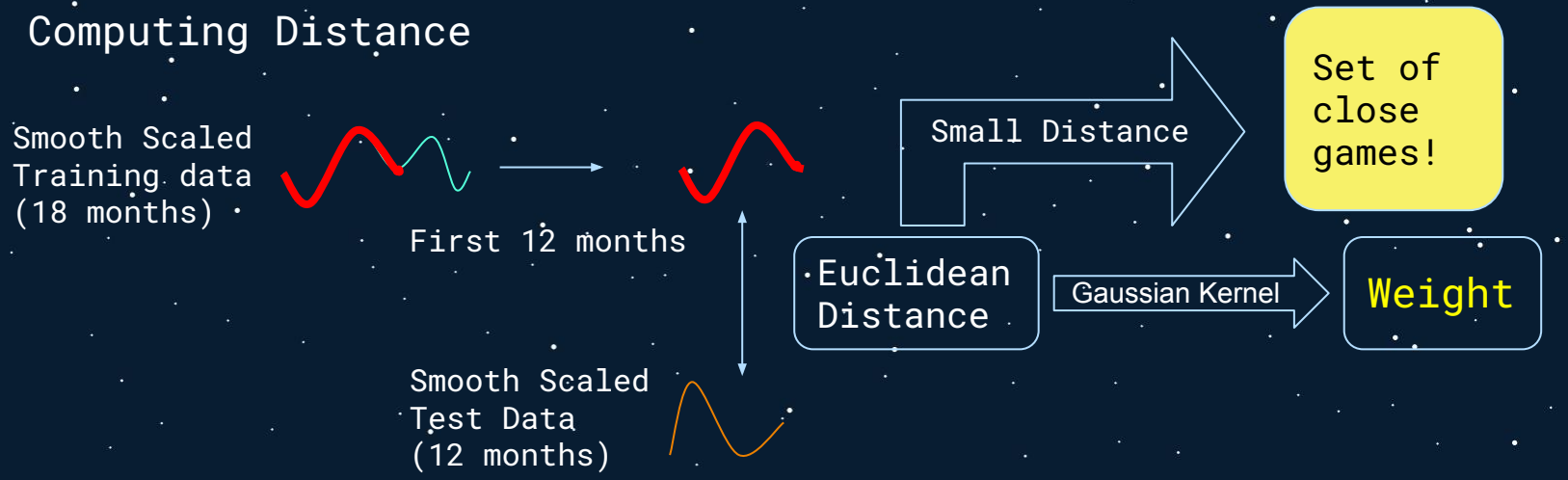




Time Comparison

Objective: Given first 12-month data of average no. of players, find close games in the past and compare.

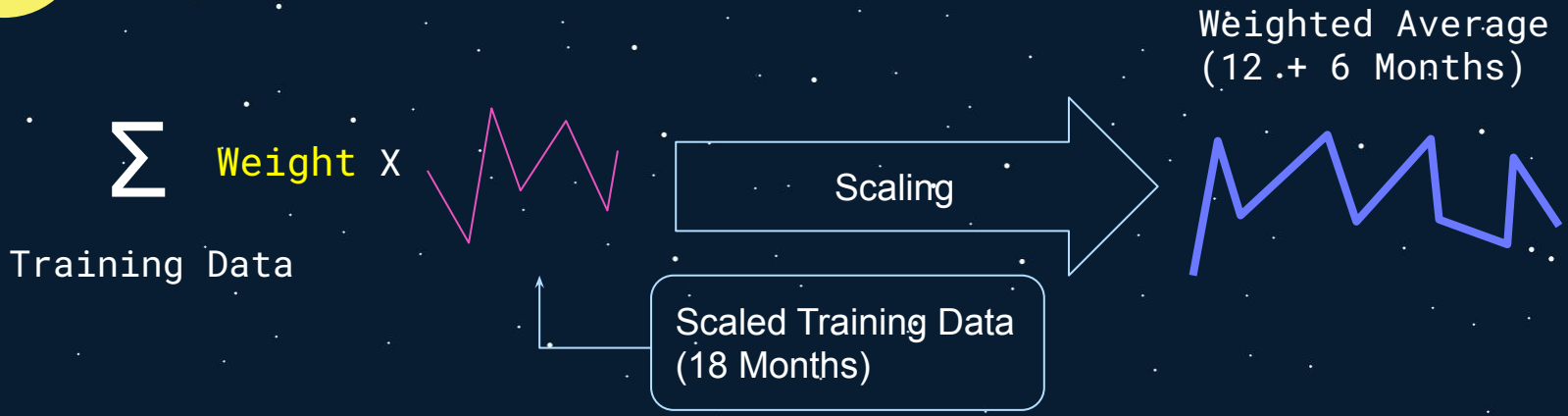
01 Computing Distance



Time Comparison

Objective: Given first 12-month data of average no. of players, find close games in the past and compare.

02 Compute Weighted Average

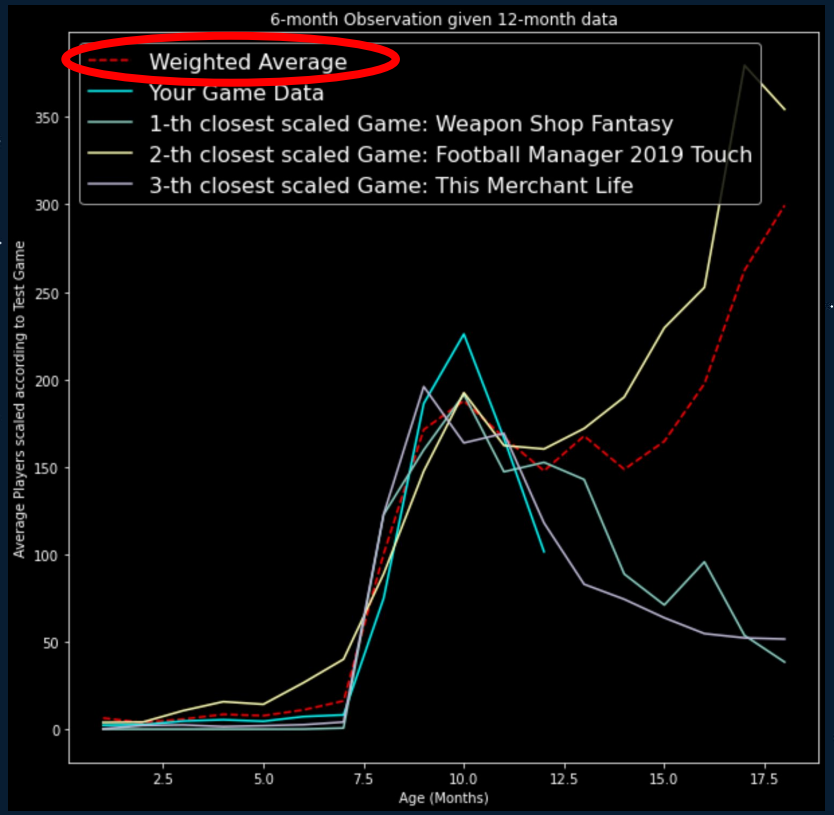


Time Comparison

What we achieved:

- Time series data from **ALL** past games as guidance
- **Games with similar history** obtained





Time Comparison



	Naive/SES/ARIMA	Weighted Average
Average % Error	~ 50%	~80% (Can adjust model to ~60%, but loses average meaning)

- Provide index of success: weighted average
- Case study
  - Timed-Features selection
    - e.g. DLC, Updates, Price change
  - Mimic events in past successful games with similar starting history
    - e.g. Tournaments, Conventions, Promotions

Web App

A user friendly web app that allows you to try three of our algorithms:

- Time Comparison
- Feature Score
- Owners Prediction



https://share.streamlit.io/alexcasella/streamlit/main/main.py

Challenges...

- Clean the data: *only relevant informations!*
- Interpret the data: *are we answering the questions?*
- Determine the correct methods: *adapt and create to better fit our context.*

...and future work!

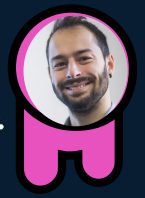
- Better predictions with more and new data.
- Link methods to improve each algorithm.
- Update the Web App to create the **Ultimate Weapon** for game developers.

Thank you!

Team 45: Steaming Hot

github.com/steaming-hot/steaming-hot

Alex Casella
acasella@fsu.edu
FSU



Eduardo Medina
medin215@umn.edu
UMN



Wing Hong Leung
leung.179@osu.edu
OSU



Kanishk Jain
kanishkbjain@gmail.com
Emory University



Ling Zhou
zhou.2568@osu.edu
OSU



Rohit Satija
rsatija@berkeley.edu
UC Berkeley

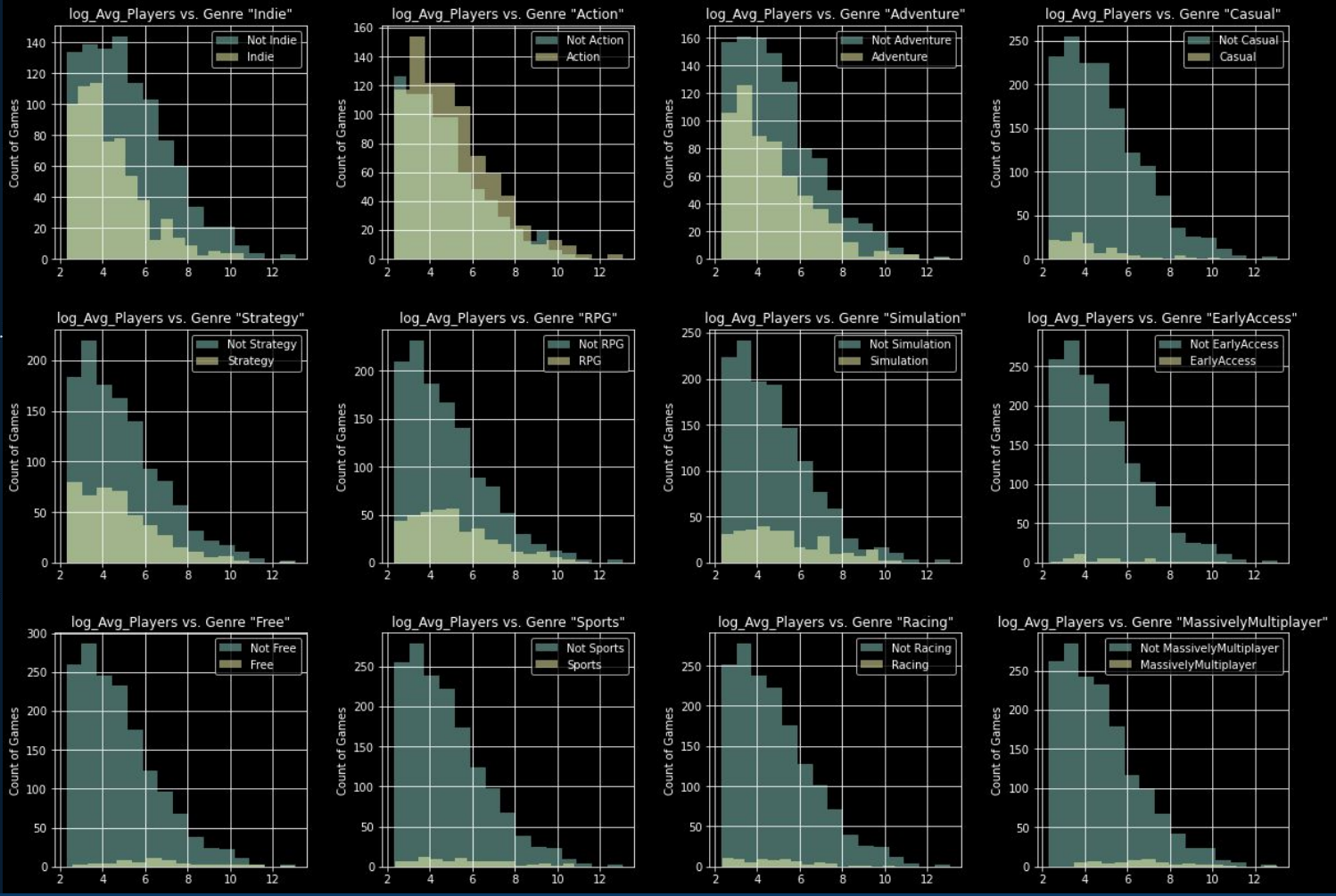


Backup slides

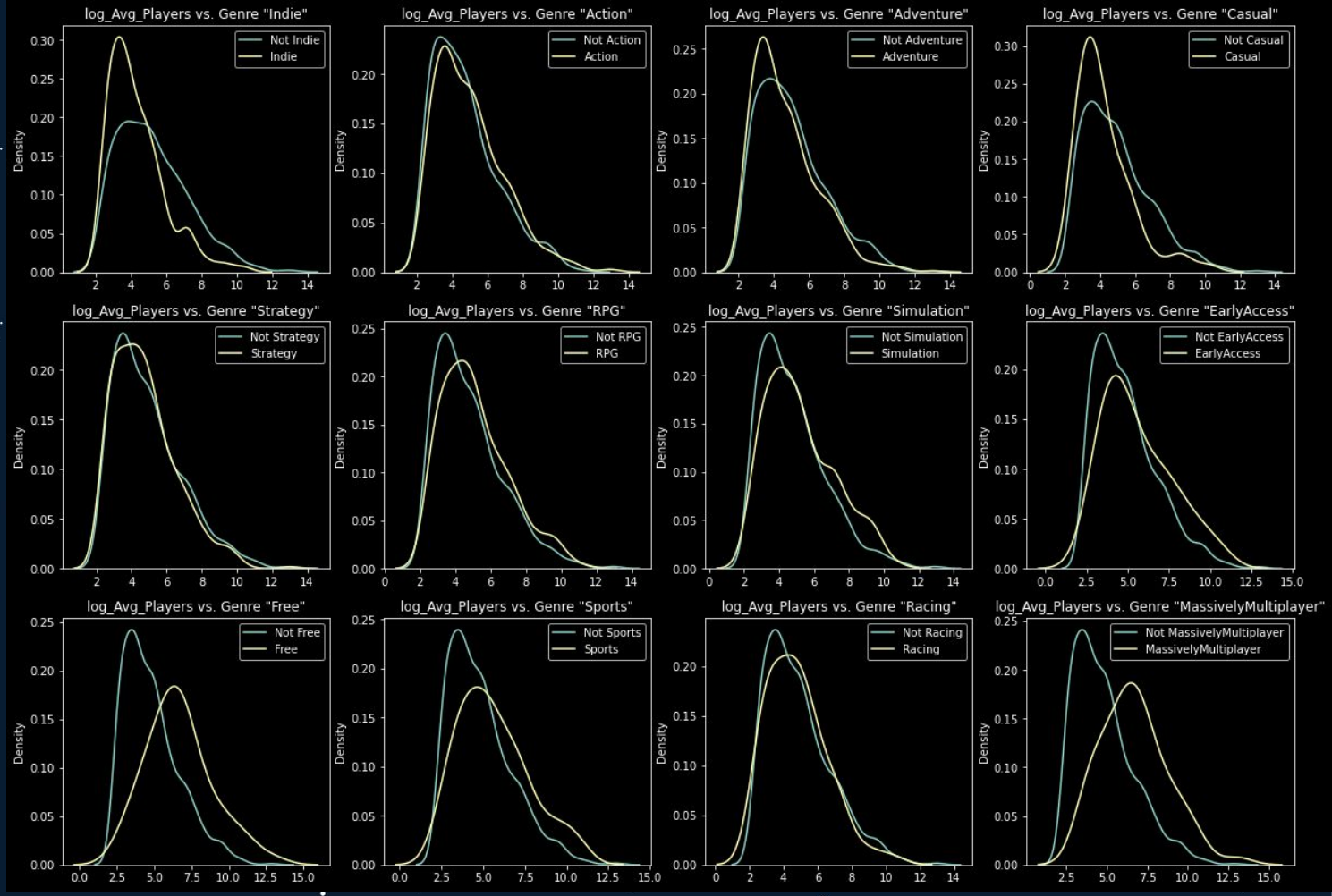
Popularity Rating

Most Popular (5.0)	mean_players>10,000
Popular (4.0)	10,000>mean_players>1,000
Somewhat Popular (3.0)	1,000>mean_players>100
Unpopular (2.0)	100>mean_players>10
Most Unpopular (1.0)	10>mean_players>1

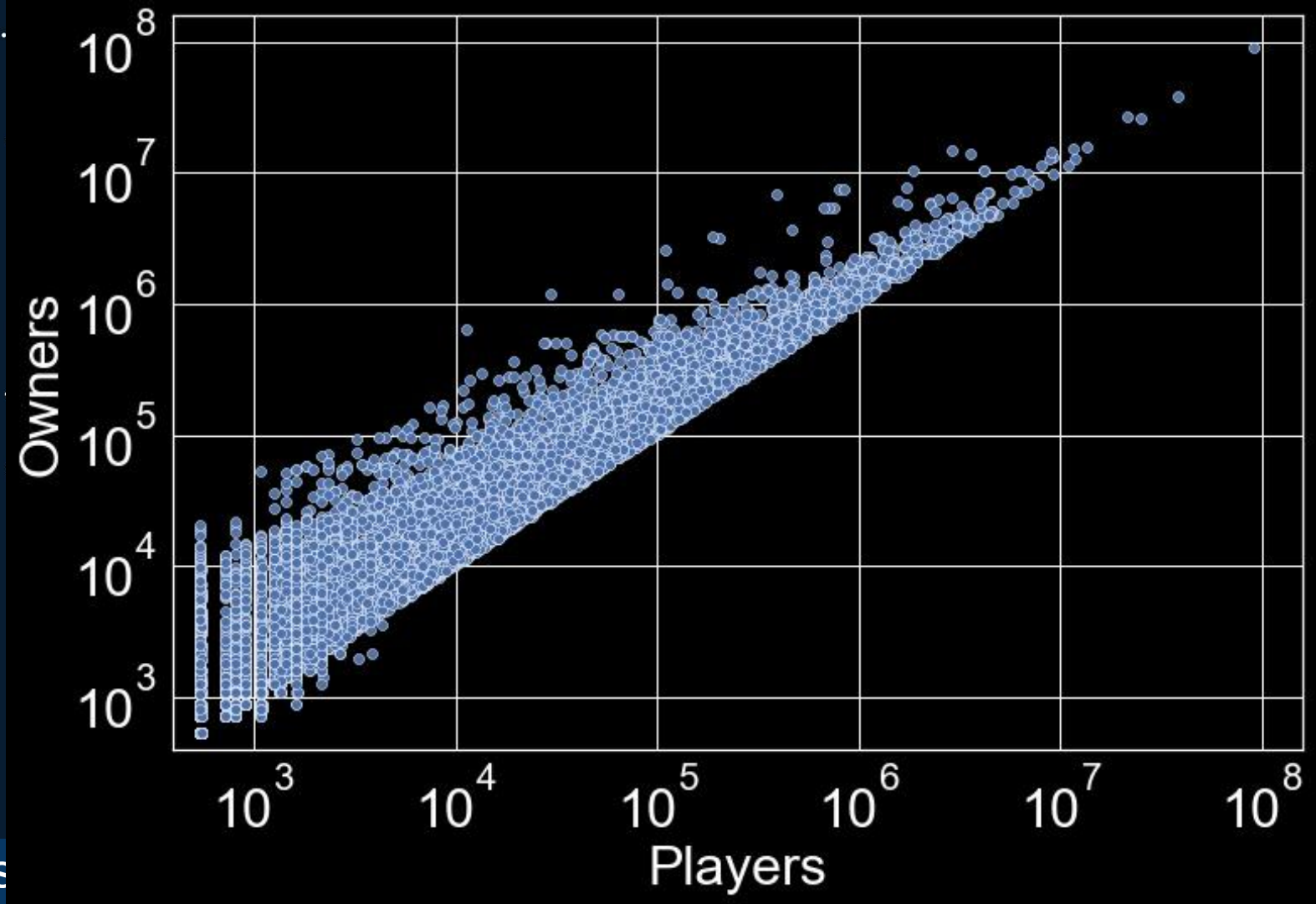
More EDA


log_Avg_Players vs. Genre "Indie"
Not Indie
Indie
log_Avg_Players vs. Genre "Action"
Not Action
Action
log_Avg_Players vs. Genre "Adventure"
Not Adventure
Adventure
log_Avg_Players vs. Genre "Casual"
Not Casual
Casual
log_Avg_Players vs. Genre "Strategy"
Not Strategy
Strategy
log_Avg_Players vs. Genre "RPG"
Not RPG
RPG
log_Avg_Players vs. Genre "Simulation"
Not Simulation
Simulation
log_Avg_Players vs. Genre "EarlyAccess"
Not EarlyAccess
EarlyAccess
log_Avg_Players vs. Genre "Free"
Not Free
Free
log_Avg_Players vs. Genre "Sports"
Not Sports
Sports
log_Avg_Players vs. Genre "Racing"
Not Racing
Racing
log_Avg_Players vs. Genre "MassivelyMultiplayer"
Not MassivelyMultiplayer
MassivelyMultiplayer
Count of Games

More EDA

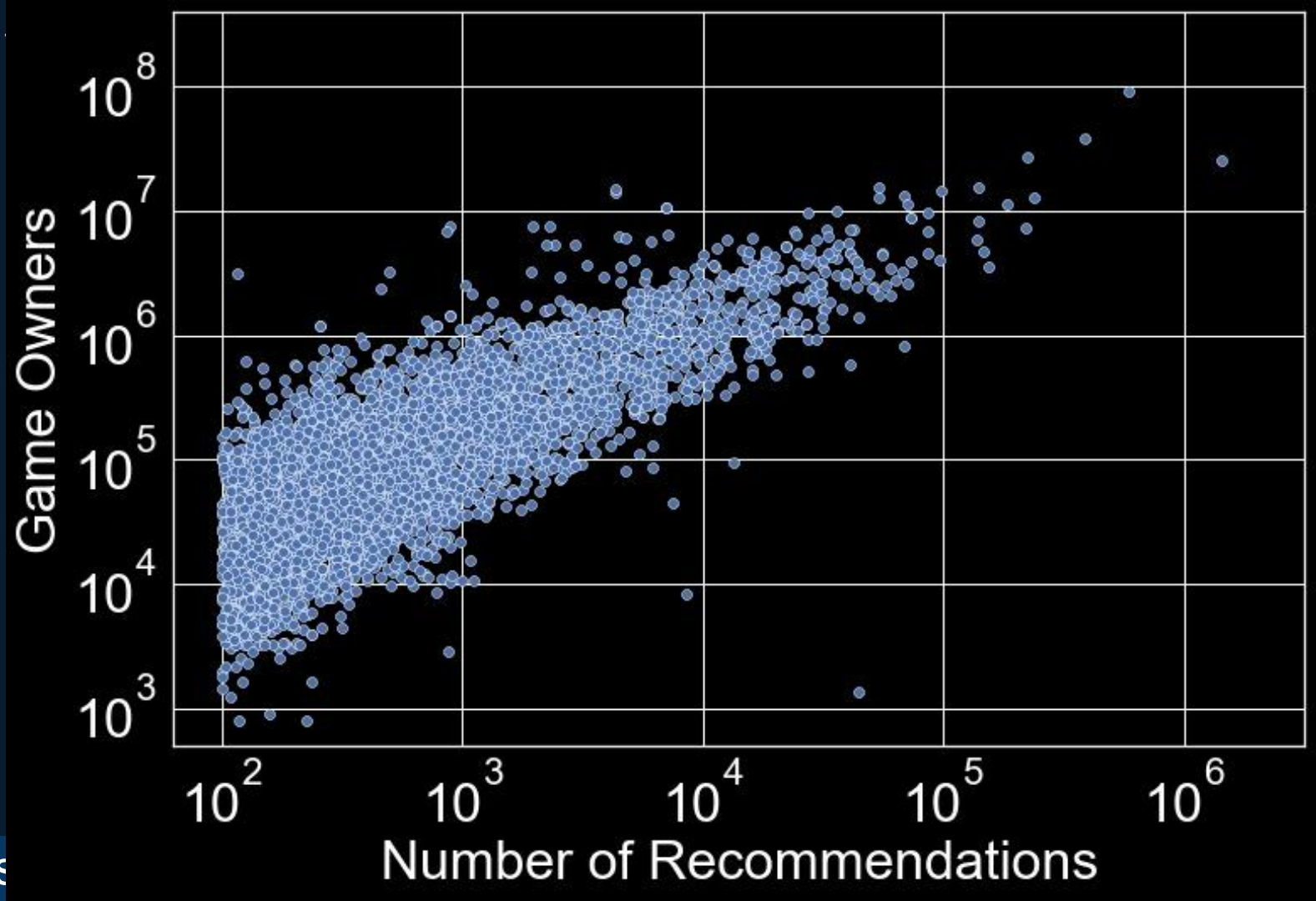

log_Avg_Players vs. Genre "Indie"
Not Indie
Indie
log_Avg_Players vs. Genre "Action"
Not Action
Action
log_Avg_Players vs. Genre "Adventure"
Not Adventure
Adventure
log_Avg_Players vs. Genre "Casual"
Not Casual
Casual
log_Avg_Players vs. Genre "Strategy"
Not Strategy
Strategy
log_Avg_Players vs. Genre "RPG"
Not RPG
RPG
log_Avg_Players vs. Genre "Simulation"
Not Simulation
Simulation
log_Avg_Players vs. Genre "EarlyAccess"
Not EarlyAccess
EarlyAccess
log_Avg_Players vs. Genre "Free"
Not Free
Free
log_Avg_Players vs. Genre "Sports"
Not Sports
Sports
log_Avg_Players vs. Genre "Racing"
Not Racing
Racing
log_Avg_Players vs. Genre "MassivelyMultiplayer"
Not MassivelyMultiplayer
MassivelyMultiplayer
Density

N=10,364



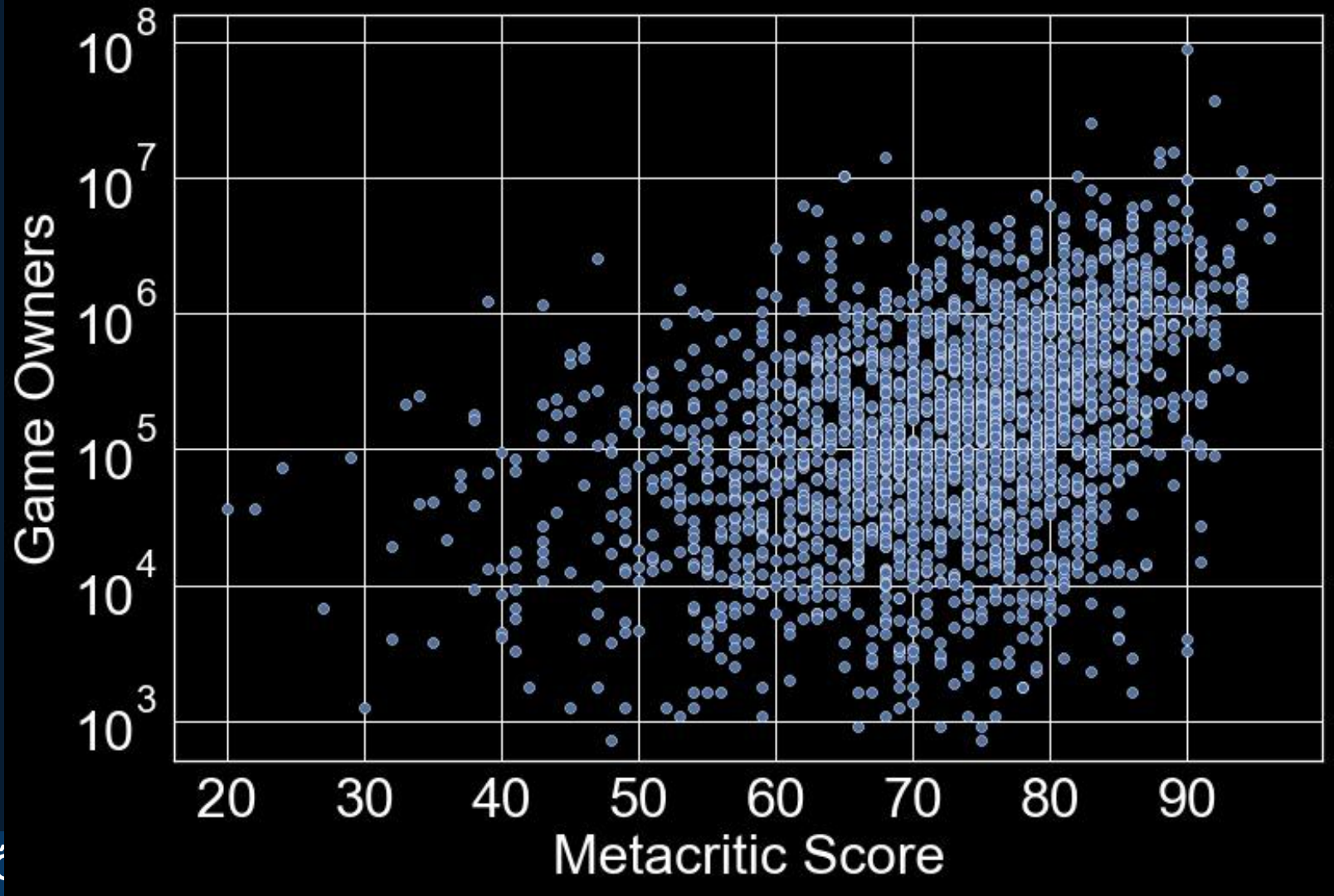
ő The Erdős

N=4,840

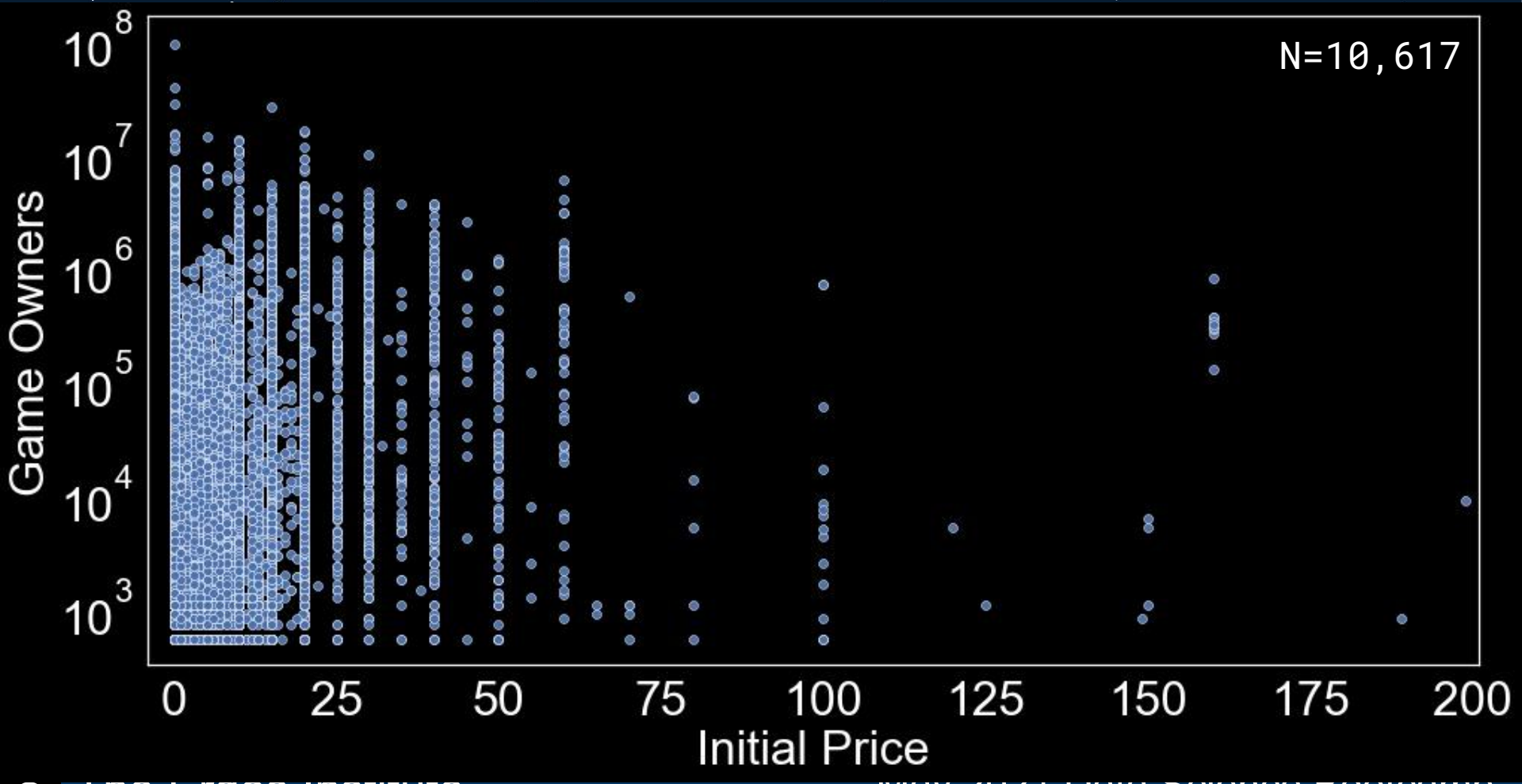


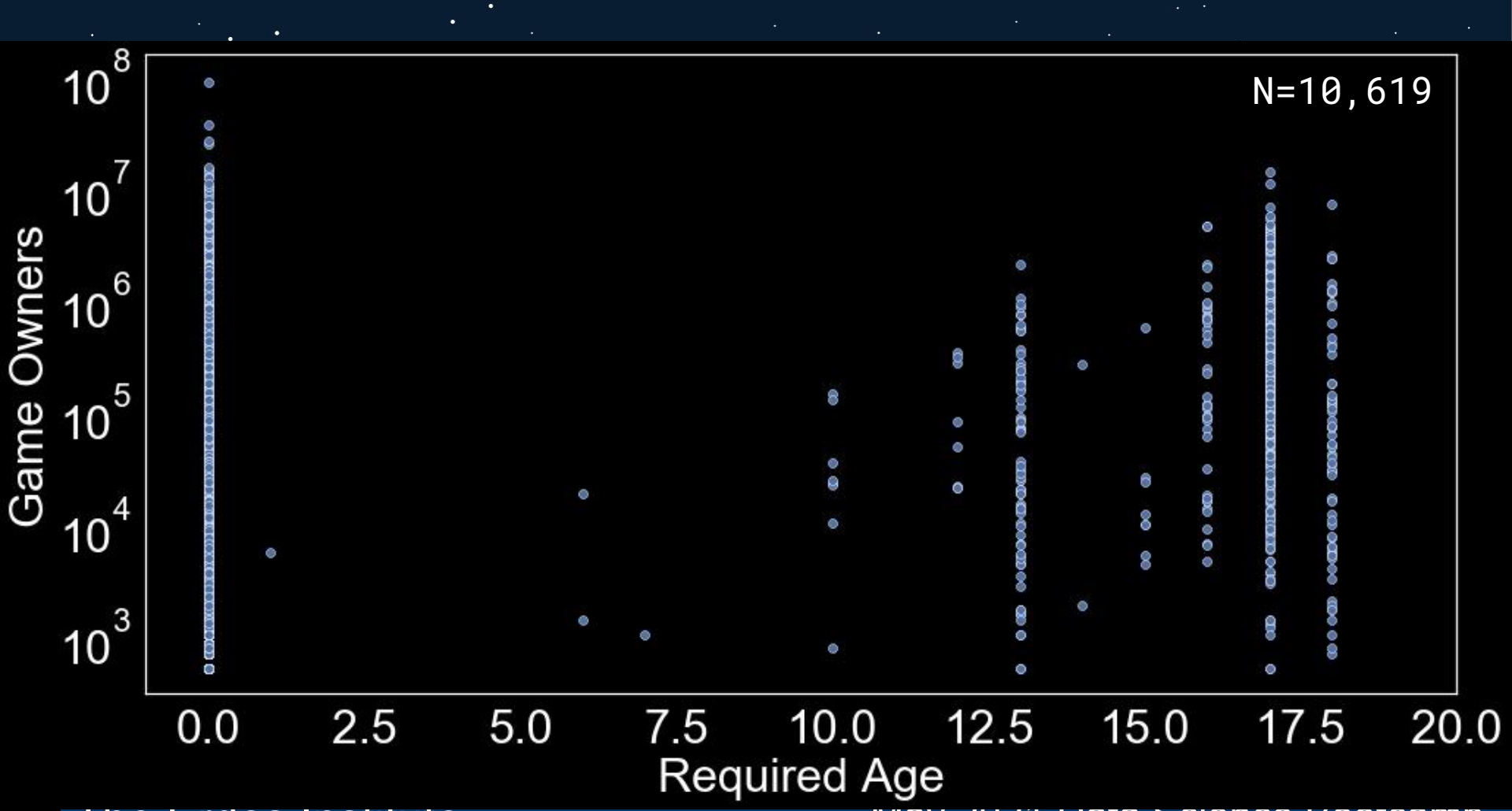
ő The Erdős

N=2,224



ő The Erdő

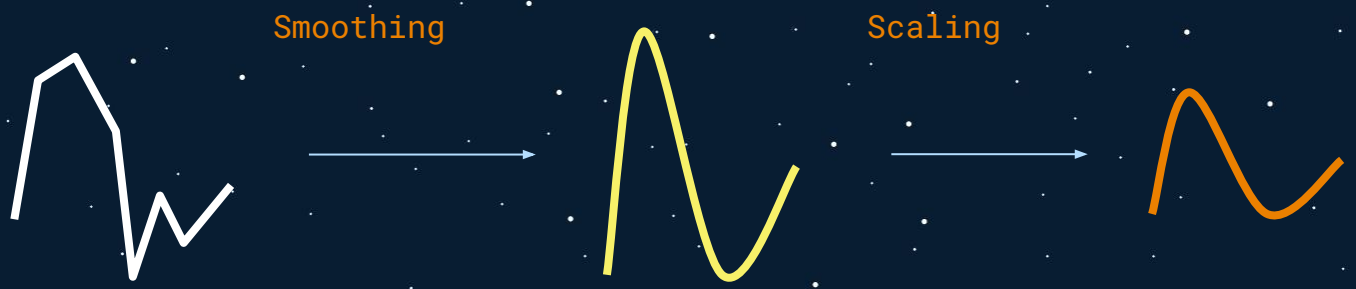

N=10,617
Game Owners
10^8
10^7
10^6
10^5
10^4
10^3
0
25
50
75
100
125
150
175
200
Initial Price


N=10,619
Game Owners
10^8
10^7
10^6
10^5
10^4
10^3
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Required Age

Time Comparison

Objective: Given first 12-month data of average no. of players, find close games in the past and compare.

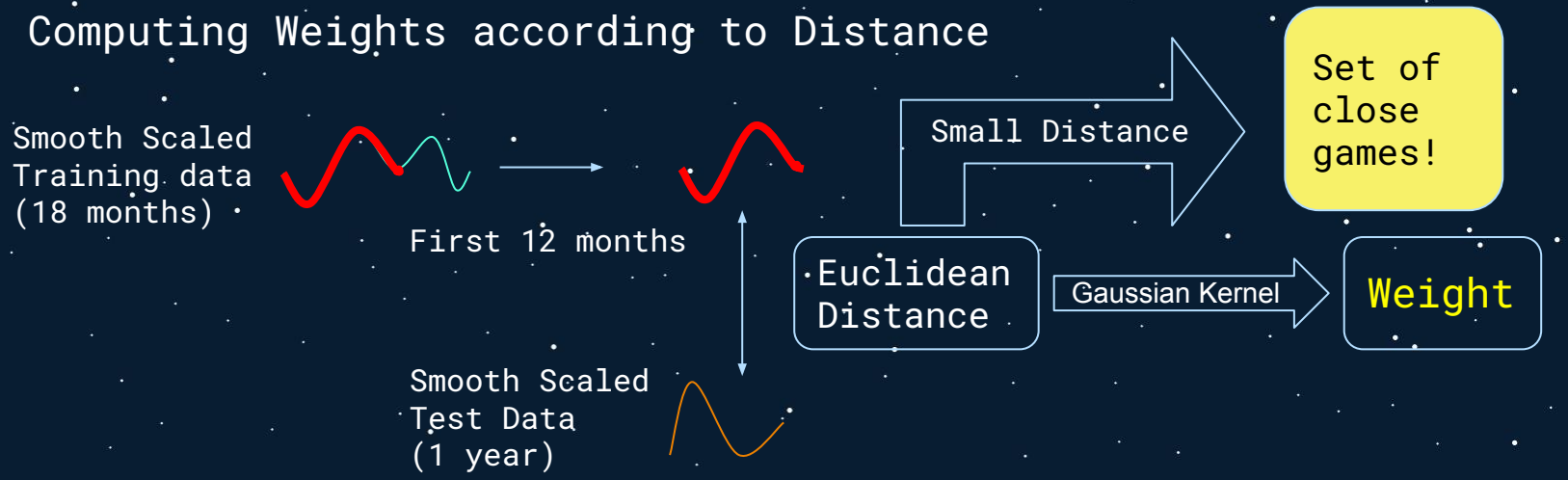
01 Scaling and Smoothing



Time Comparison

Objective: Given first 12-month data of average no. of players, find close games in the past and compare.

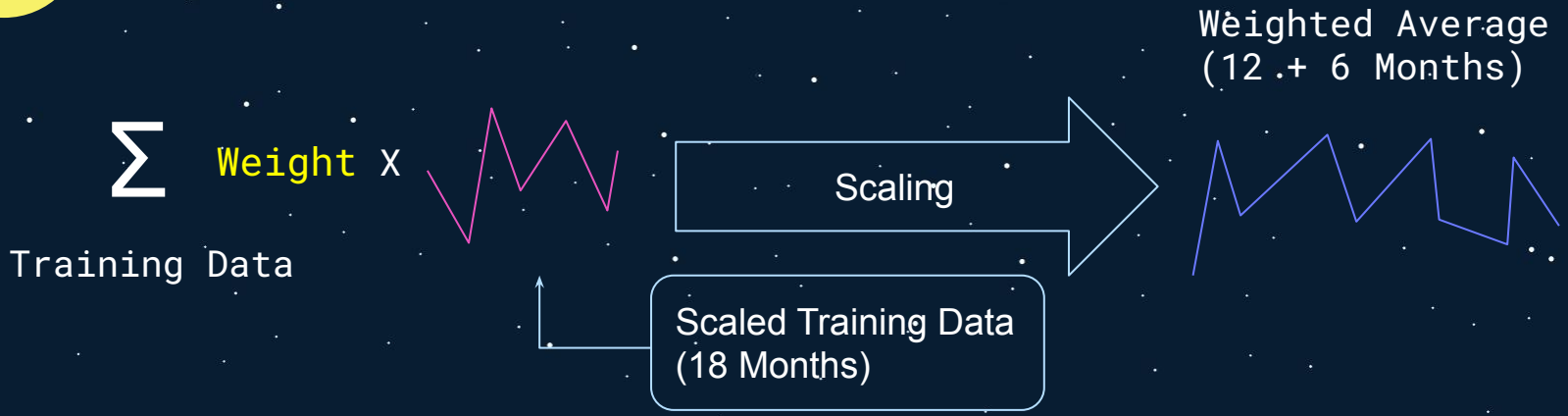
02 Computing Weights according to Distance



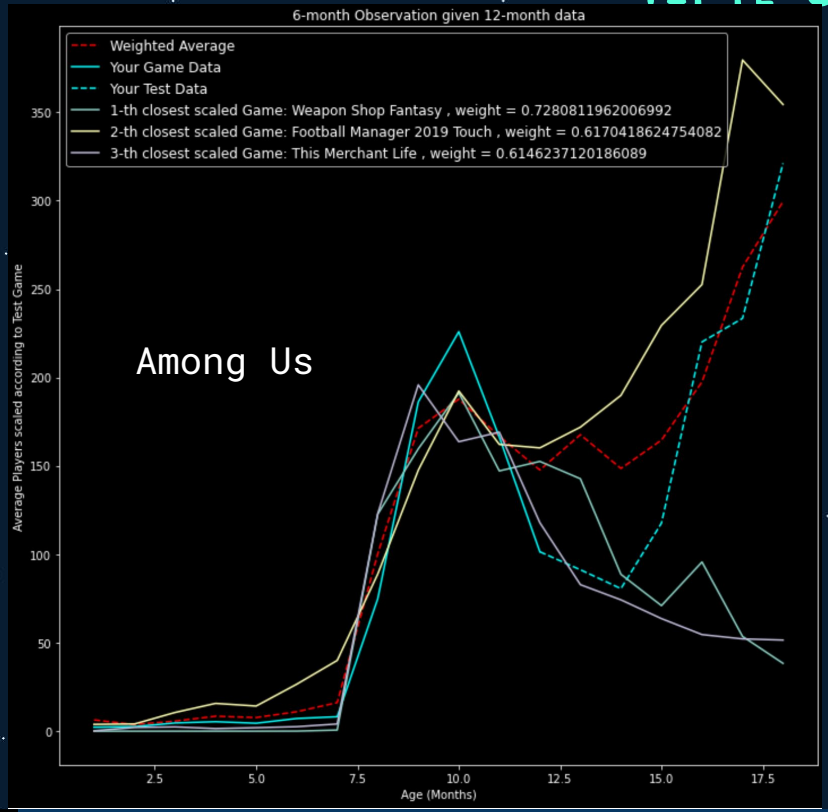
Time Comparison

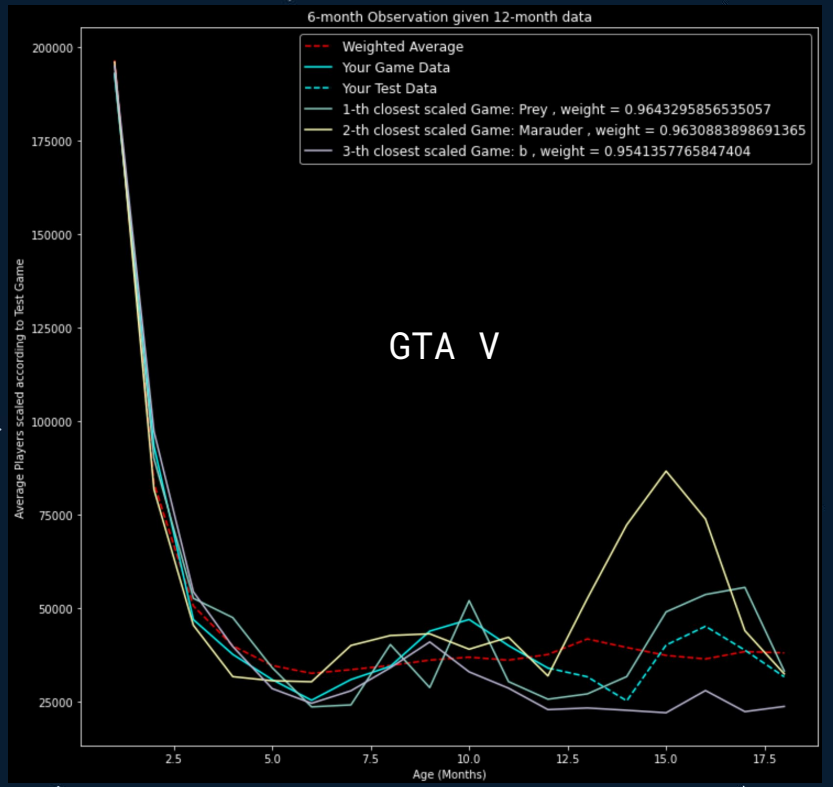
Objective: Given first 12-month data of average no. of players, find close games in the past and compare.

03 Compute Weighted Average



Time Comparison





Machine Learning: Classification

	ControllerSupport	IsFree	PurchaseAvail	PlatformLinux	PlatformMac	PCReqsHaveRec	LinuxReqsHaveMin	LinuxReqsHaveRec	MacReqsHaveMin	MacReqsHaveRec	CategorySinglePlayer	CategoryMultiplayer	CategoryCoop
0	0	0	1	0	0	1	1	1	1	1	0	1	0
1	1	0	1	1	1	0	1	0	1	0	0	1	0
2	0	1	0	1	1	0	1	0	1	0	0	1	1
3	0	0	1	0	0	1	1	1	1	1	1	0	0
4	1	0	1	0	0	1	0	0	0	0	1	0	0

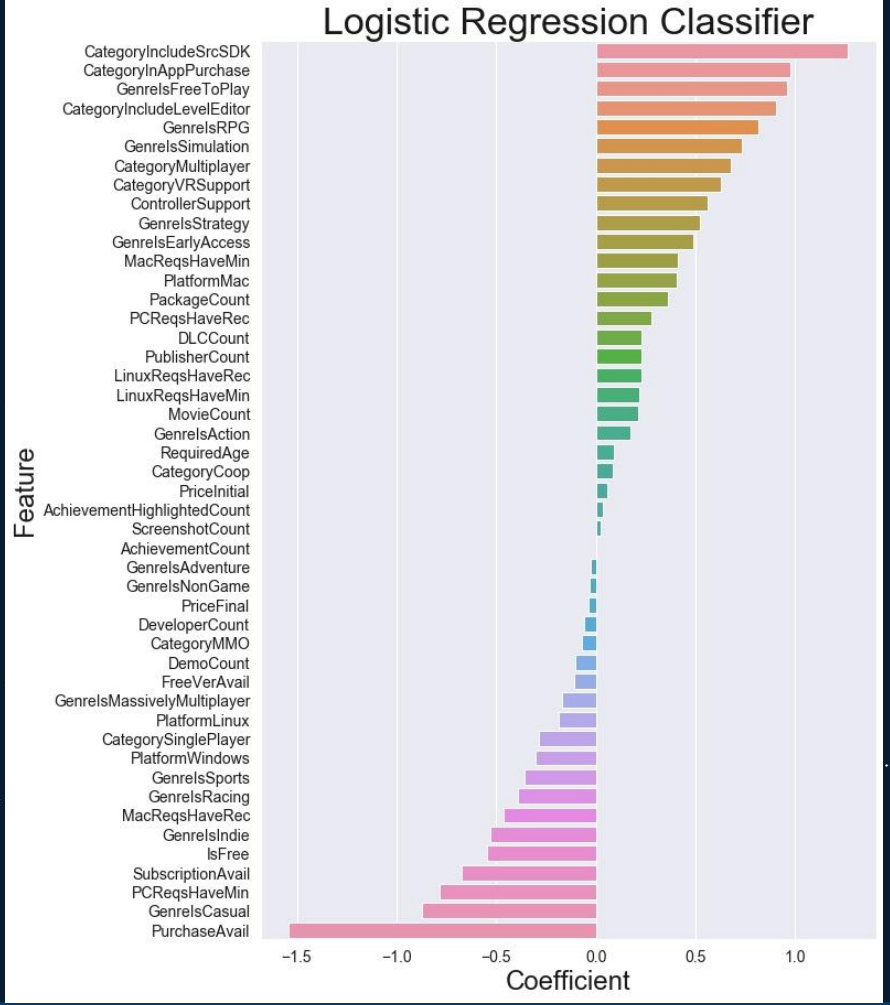
	CategoryMMO	CategoryInAppPurchase	GenreIsIndie	GenreIsAction	GenreIsAdventure	GenreIsCasual	GenreIsStrategy	GenreIsRPG	GenreIsSimulation	GenreIsEarlyAccess	GenreIsFreeToPlay	GenreIsSports	GenreIsRacing	GenreIsMassivelyMultiplayer
0	0	0	0	1	1	0	0	0	0	0	0	0	0	1
1	0	1	0	1	0	0	0	0	0	0	0	0	0	0
2	0	1	0	1	0	0	1	0	0	0	1	0	0	0
3	0	0	0	0	0	0	0	1	0	0	0	0	0	0
4	0	0	0	0	0	0	0	1	0	0	0	0	0	0

N_features = 47, N_observations = 3,597

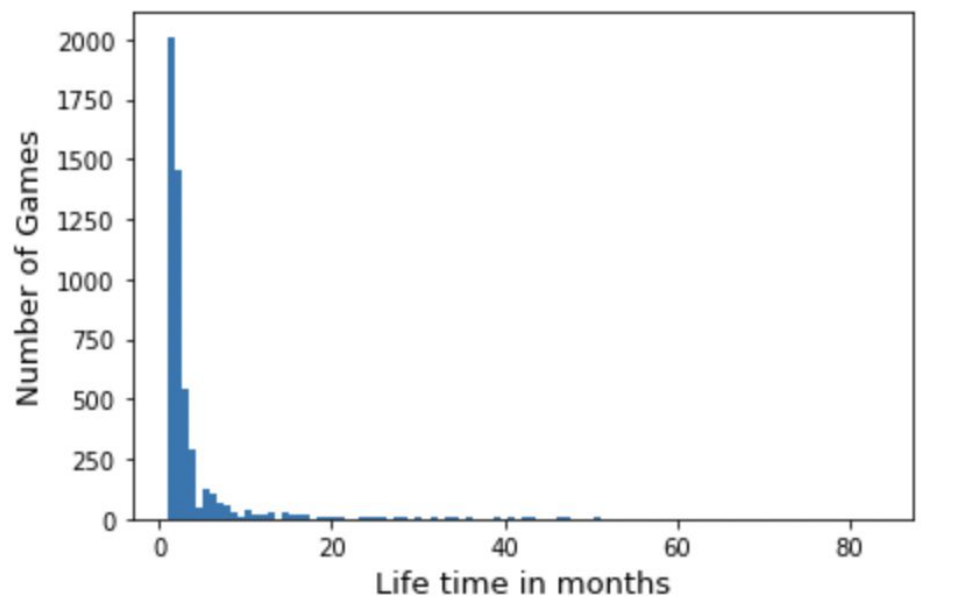
Target variable:“Popularity Grade”

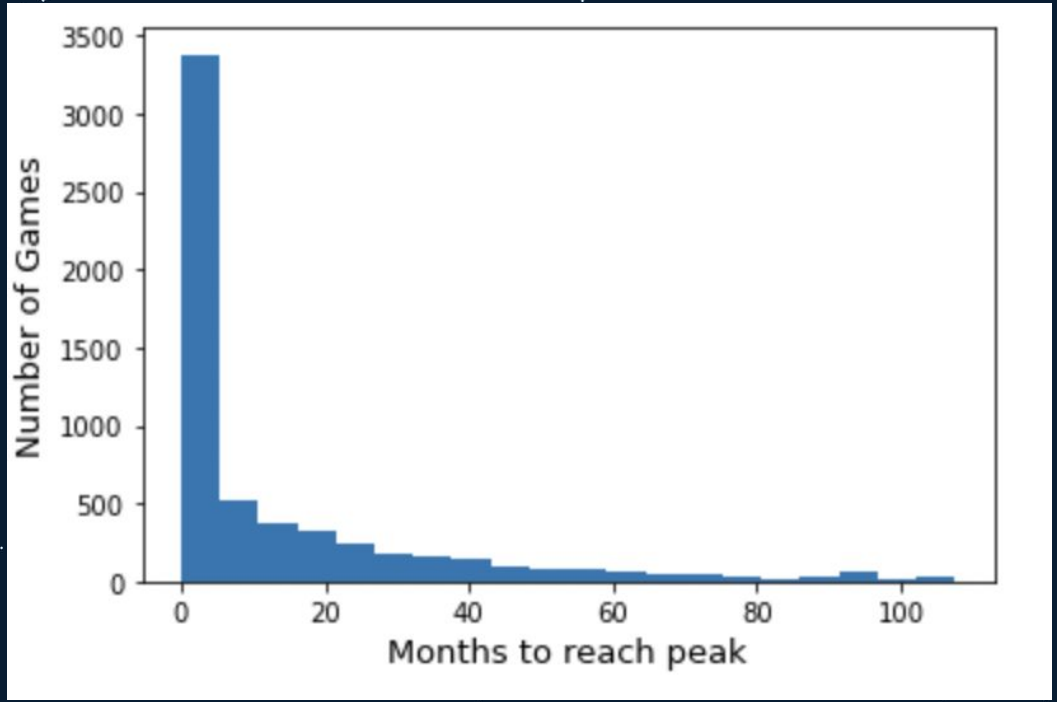
- mean_players>100:1.0
- 1<Avg_Players<5:0.0

Model	accuracy_score %	precision_score %	recall_score %	f1_score %	AUC_score %
K-Nearest Neighbors	73.78	74.94	71.96	73.42	81.86
Decision Tree	75.86	78.84	75.96	74.77	82.18
Support Vector Machine	75.78	73.45	81.24	77.15	83.1
Logistic Regression	79.11	80.05	77.92	78.97	87.41
AdaBoost	79.68	81.08	78.47	79.38	88.8
Random Forest	81.46	81.81	81.86	81.58	89.61

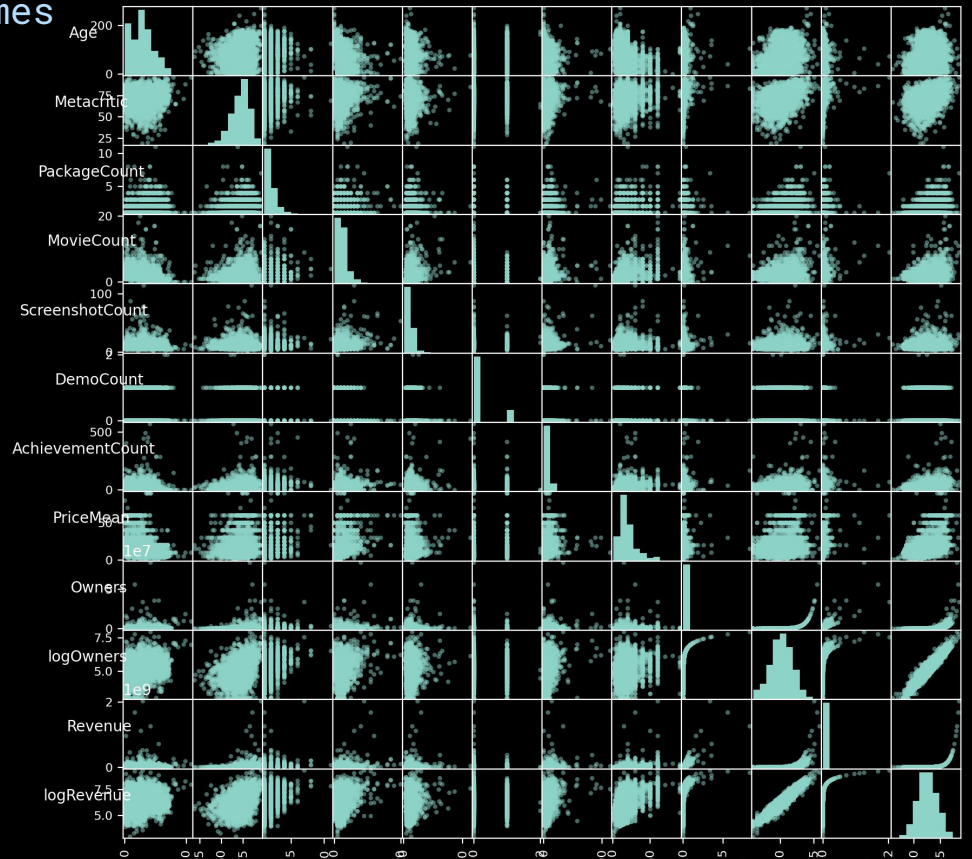

Logistic Regression Classifier
CategoryIncludeSrcSDK
CategoryInAppPurchase
GenreIsFreeToPlay
CategoryIncludeLevelEditor
GenreIsRPG
GenreIsSimulation
CategoryMultiplayer
CategoryVRSupport
ControllerSupport
GenreIsStrategy
GenreIsEarlyAccess
MacReqsHaveMin
PlatformMac
PackageCount
PCReqsHaveRec
DLCCount
PublisherCount
LinuxReqsHaveRec
LinuxReqsHaveMin
MovieCount
GenreIsAction
RequiredAge
CategoryCoop
PriceInitial
AchievementHighlightedCount
ScreenshotCount
AchievementCount
GenreIsAdventure
GenreIsNonGame
PriceFinal
DeveloperCount
CategoryMMO
DemoCount
FreeVerAvail
GenreIsMassivelyMultiplayer
PlatformLinux
CategorySinglePlayer
PlatformWindows
GenreIsSports
GenreIsRacing
MacReqsHaveRec
GenreIsIndie
IsFree
SubscriptionAvail
PCReqsHaveMin
GenreIsCasual
PurchaseAvail
Feature
−1.5
−1.0
−0.5
0.0
0.5
1.0
Coefficient

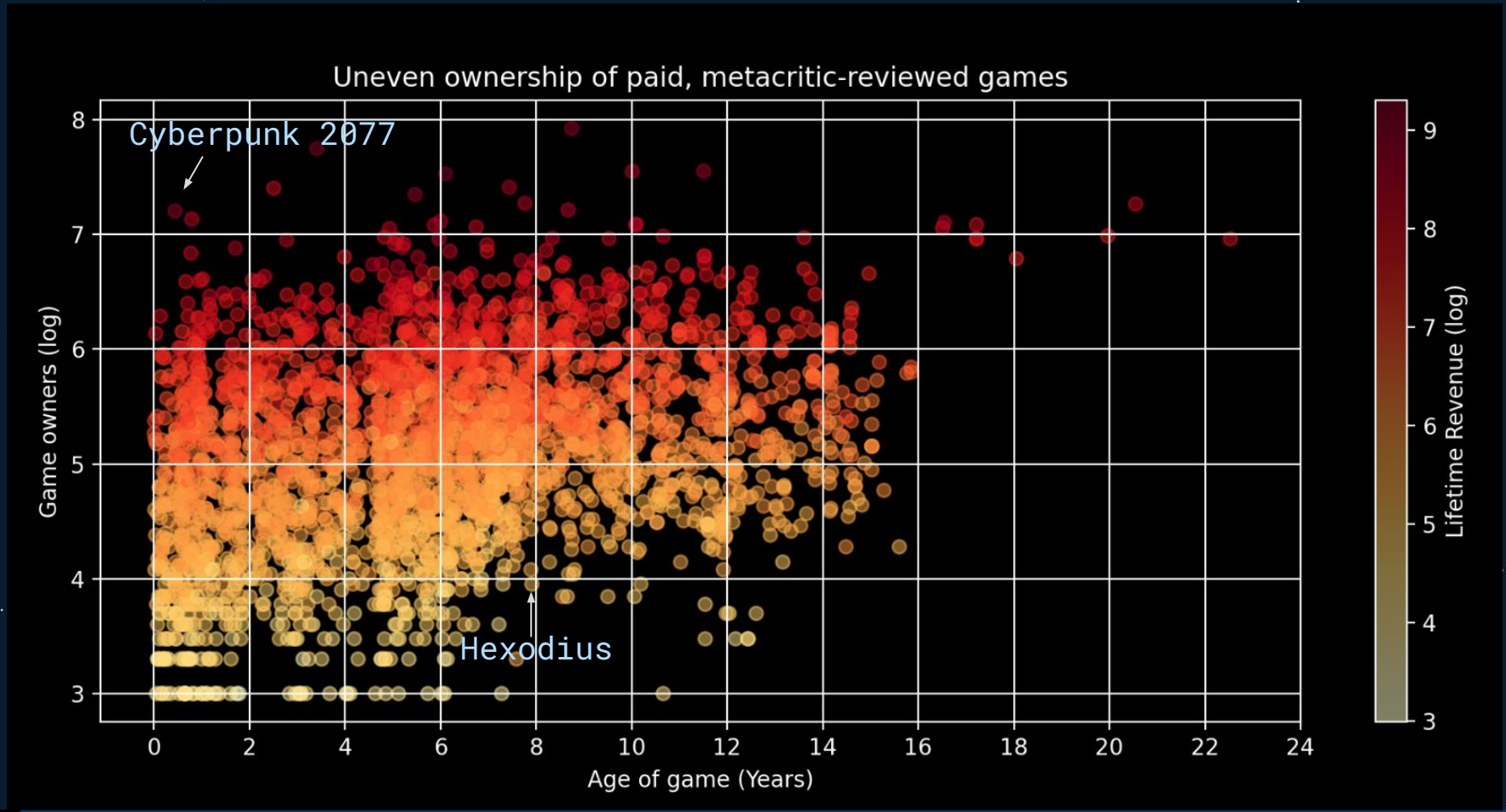
Total Games	Died Less Than 6 Months	Still Alive	Die in More than 6 Months
6059	4468	978	613
%100	%73.74	%16.14	%10.12

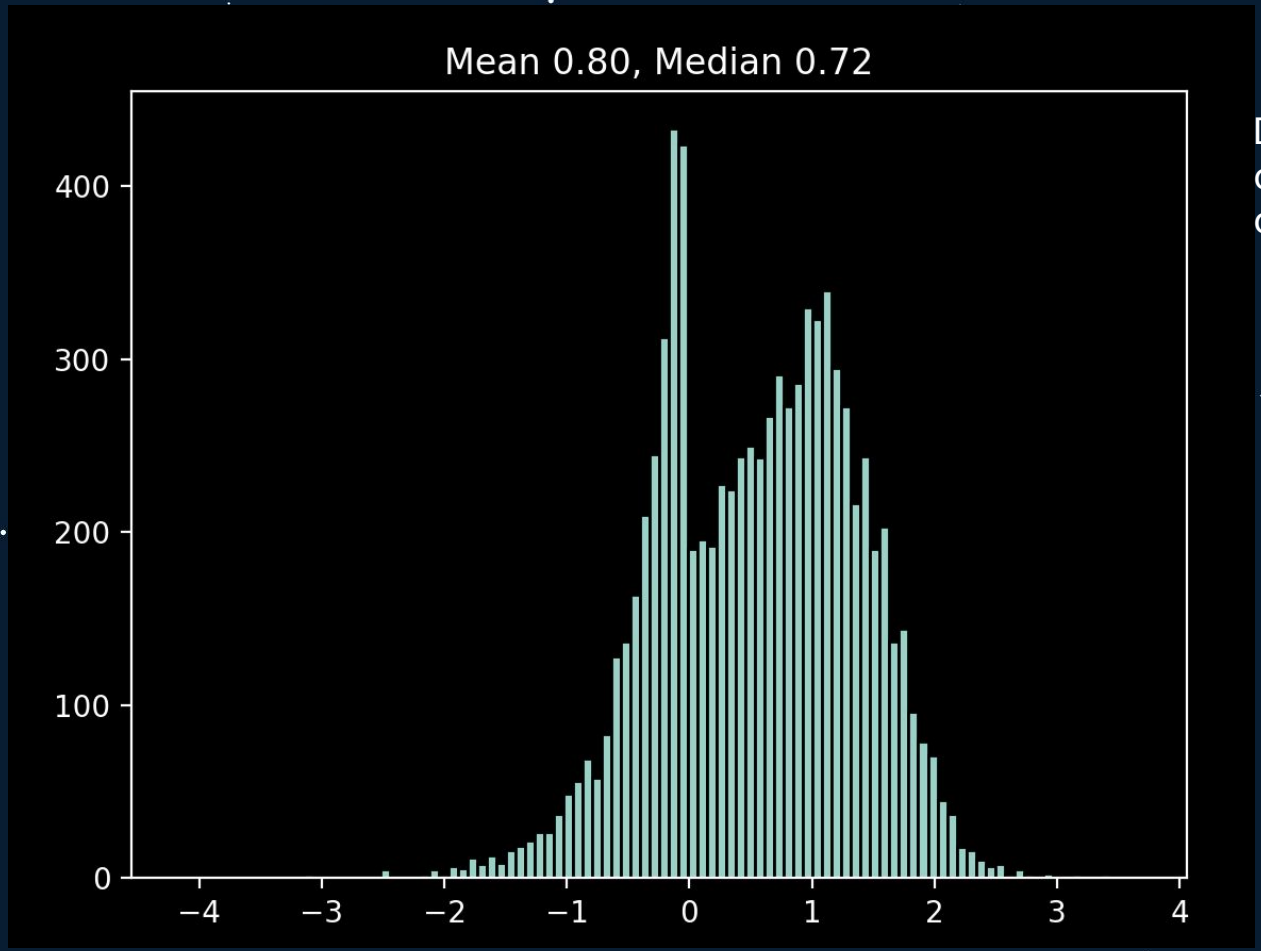



3500
3000
2500
2000
1500
1000
500
0
Number of Games
0
20
40
60
80
100
Months to reach peak

paid, metacritic-reviewed games



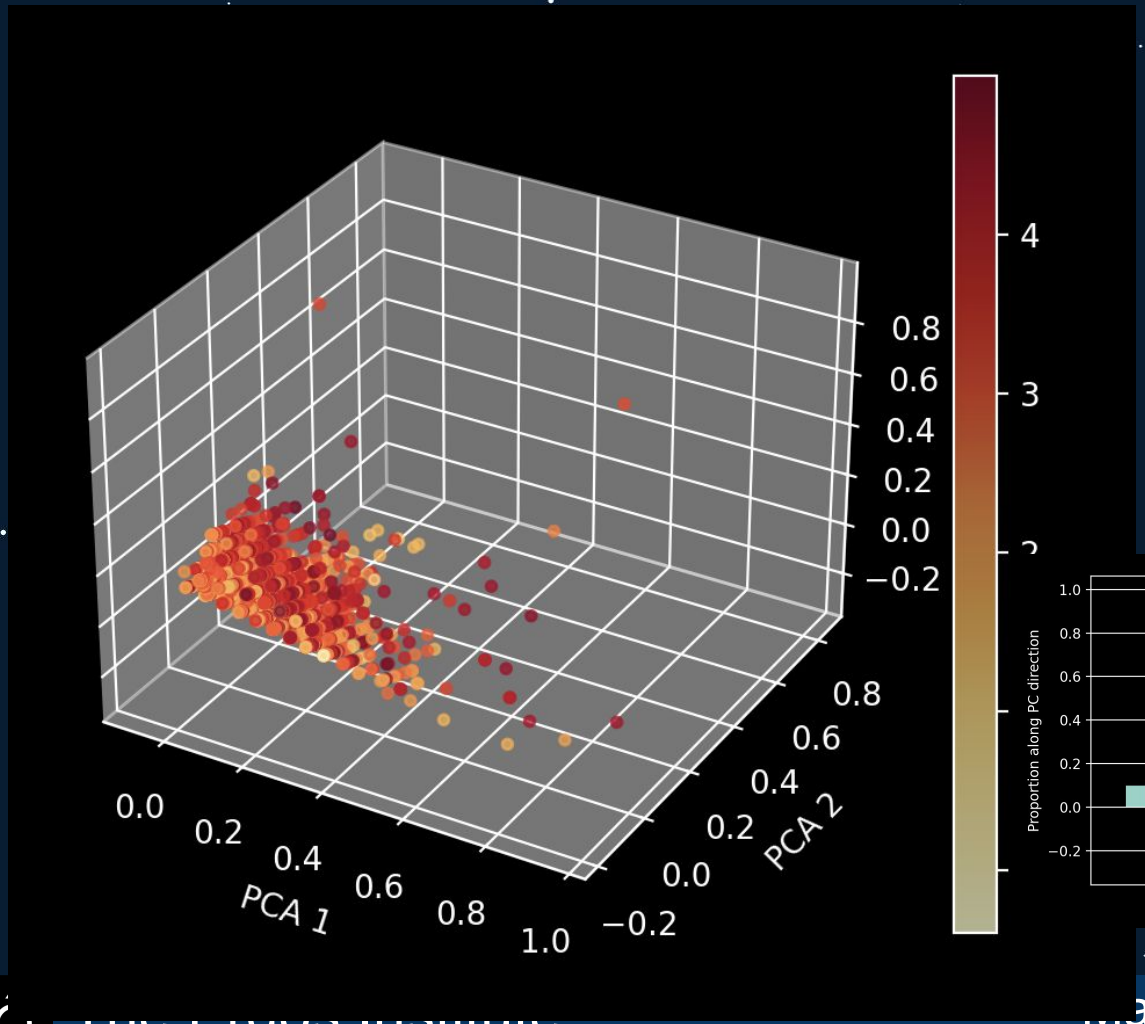

Uneven ownership of paid, metacritic-reviewed games
Cyberpunk 2077
Hexodius
Game owners (log)
Age of game (Years)
Lifetime Revenue (log)
8
7
6
5
4
3
0
2
4
6
8
10
12
14
16
18
20
22
24
9

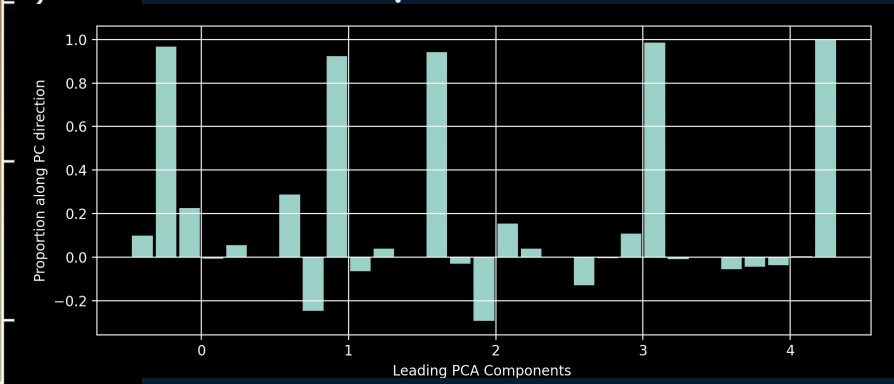


Decision tree regressor on games with monthly owners less than 100,000

PCA Monthly owners colorbar

'PackageCount',
'MovieCount',
'ScreenshotCount',
'AchievementCount',
'PriceMean'





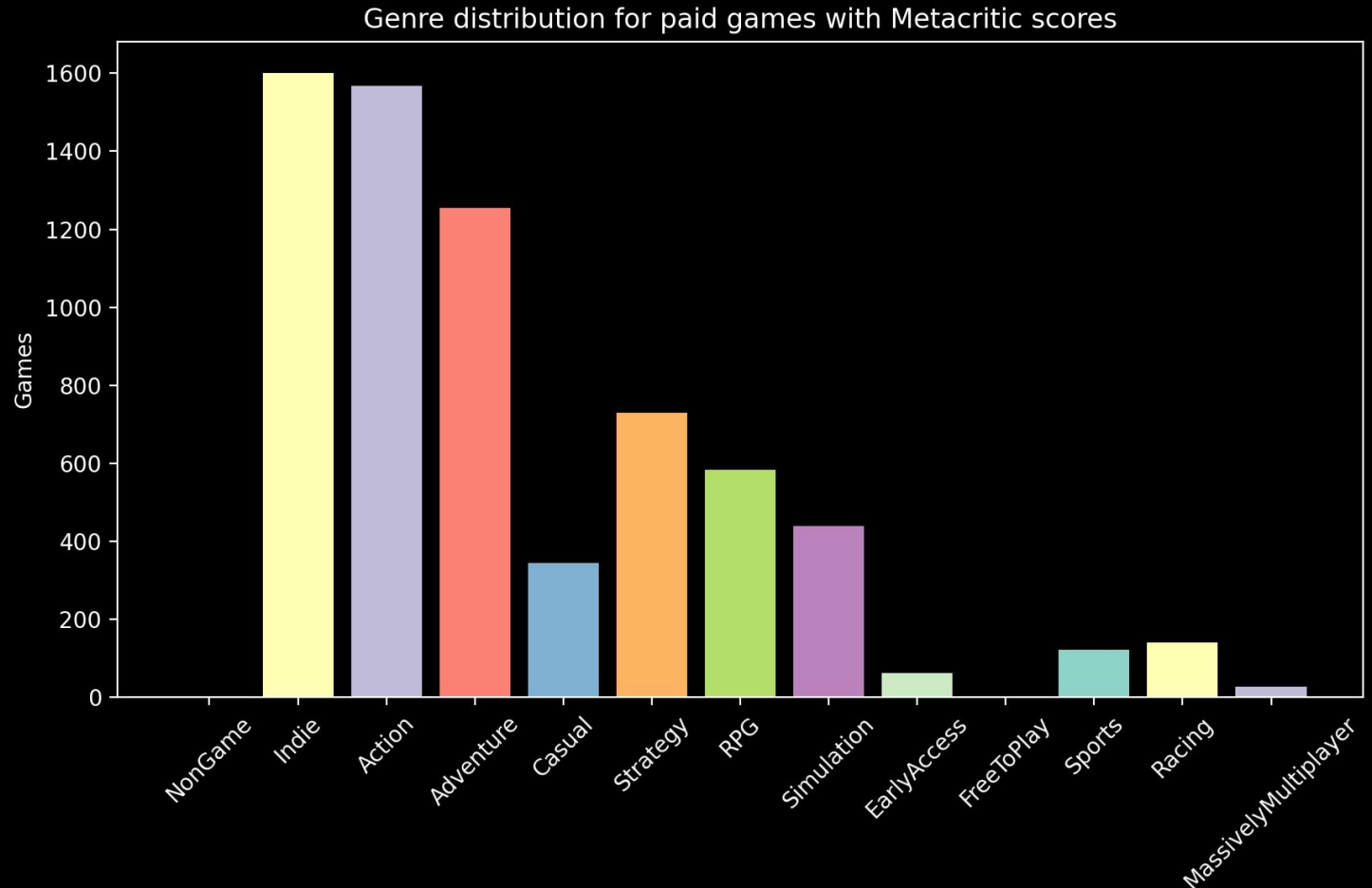

Genre distribution for paid games with Metacritic scores
Games
0
200
400
600
800
1000
1200
1400
1600
NonGame
Indie
Action
Adventure
Casual
Strategy
RPG
Simulation
EarlyAccess
FreeToPlay
Sports
Racing
MassivelyMultiplayer

Image sources:

https://www.lynda.com/Marketing-tutorials/Online-Marketing-Foundations/693114-2.html

https://gameanalytics.com/blog/8-areas-you-should-focus-your-game-development-in-2018/

https://medium.com/@Momentum_London/sales-forecasting-8c5a0b61ac5a

https://www.lynda.com/Business-Skills-tutorials/Select-forecasting-technique/548711/601858-4.html

https://learningsolutionsmag.com/articles/metafocus-choosing-a-serious-game-development-strategy

RequiredAge	-0.008818
DemoCount	0.025099
DeveloperCount	0.127606
DLCCount	-0.002965
Metacritic	0.000101
MovieCount	0.065993
PackageCount	0.073392
RecommendationCount	-0.000005
PublisherCount	-0.372406
ScreenshotCount	-0.001193
AchievementCount	0.000083
AchievementHighlightedCount	0.013701
ControllerSupport	0.177140
IsFree	0.144621
FreeVerAvail	1.049263
PurchaseAvail	-0.265359
SubscriptionAvail	-0.478384

CReqsHaveMin	-0.455764
PCReqsHaveRec	0.057429
LinuxReqsHaveMin	0.065472
LinuxReqsHaveRec	-0.033339
MacReqsHaveMin	-0.028826
MacReqsHaveRec	-0.116675
CategorySinglePlayer	-0.325486
CategoryMultiplayer	-0.112083
CategoryCoop	-0.011081
CategoryMMO	-0.898058
CategoryInAppPurchase	0.010689
CategoryIncludeSrcSDK	-0.338294
CategoryIncludeLevelEditor	-0.393233
CategoryVRSupport	0.013538
GenreIsNonGame	-0.197014
GenreIsIndie	0.353426
GenreIsAction	0.304676
GenreIsAdventure	0.025963
GenreIsCasual	0.362374
GenreIsStrategy	0.003225
GenreIsRPG	0.068681
GenreIsSimulation	-0.180593

GenreIsEarlyAccess	0.226580
GenreIsFreeToPlay	-0.356964
GenreIsSports	0.382582
GenreIsRacing	0.061259
GenreIsMassivelyMultiplayer	0.222382
PriceInitial	-0.014243
PriceFinal	0.010934