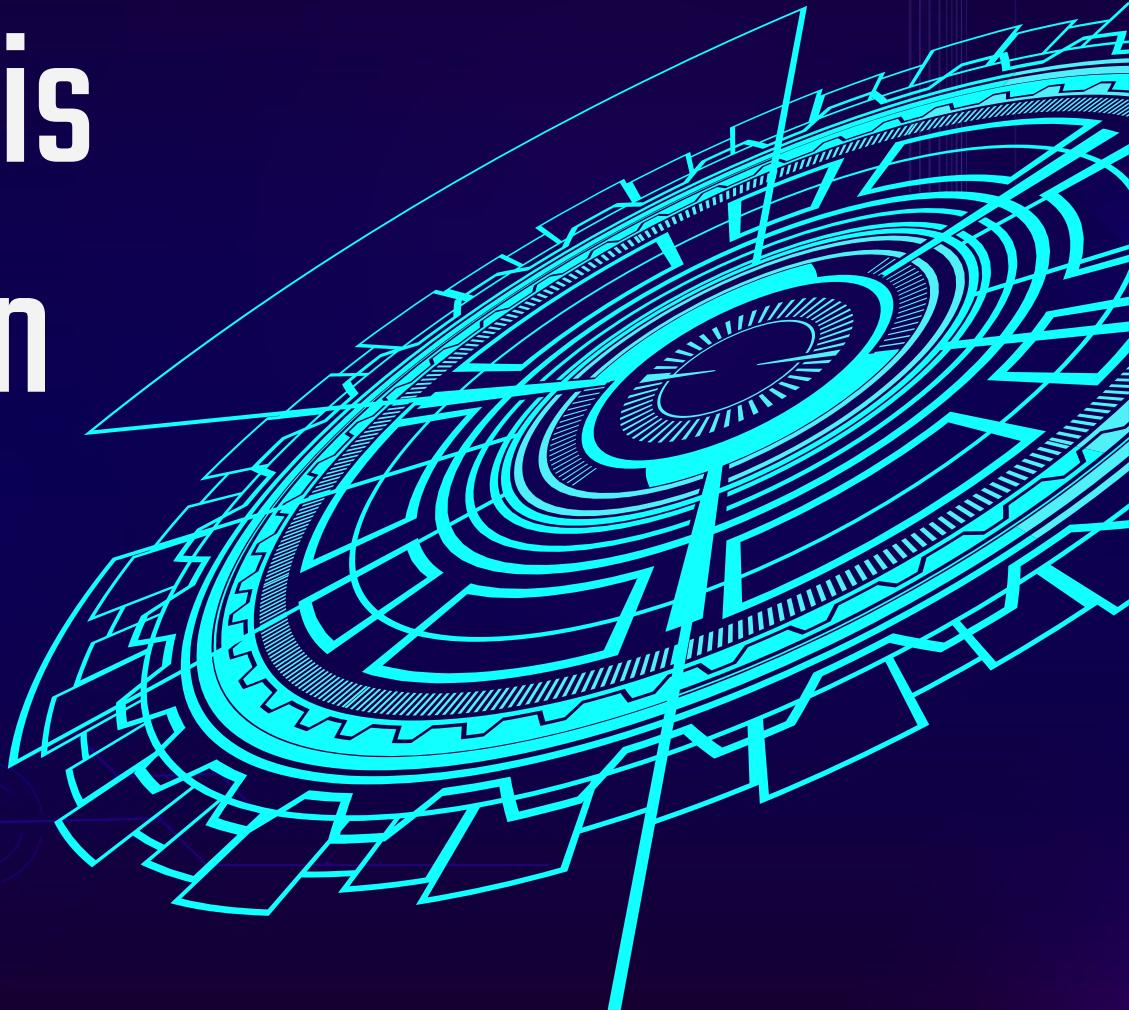
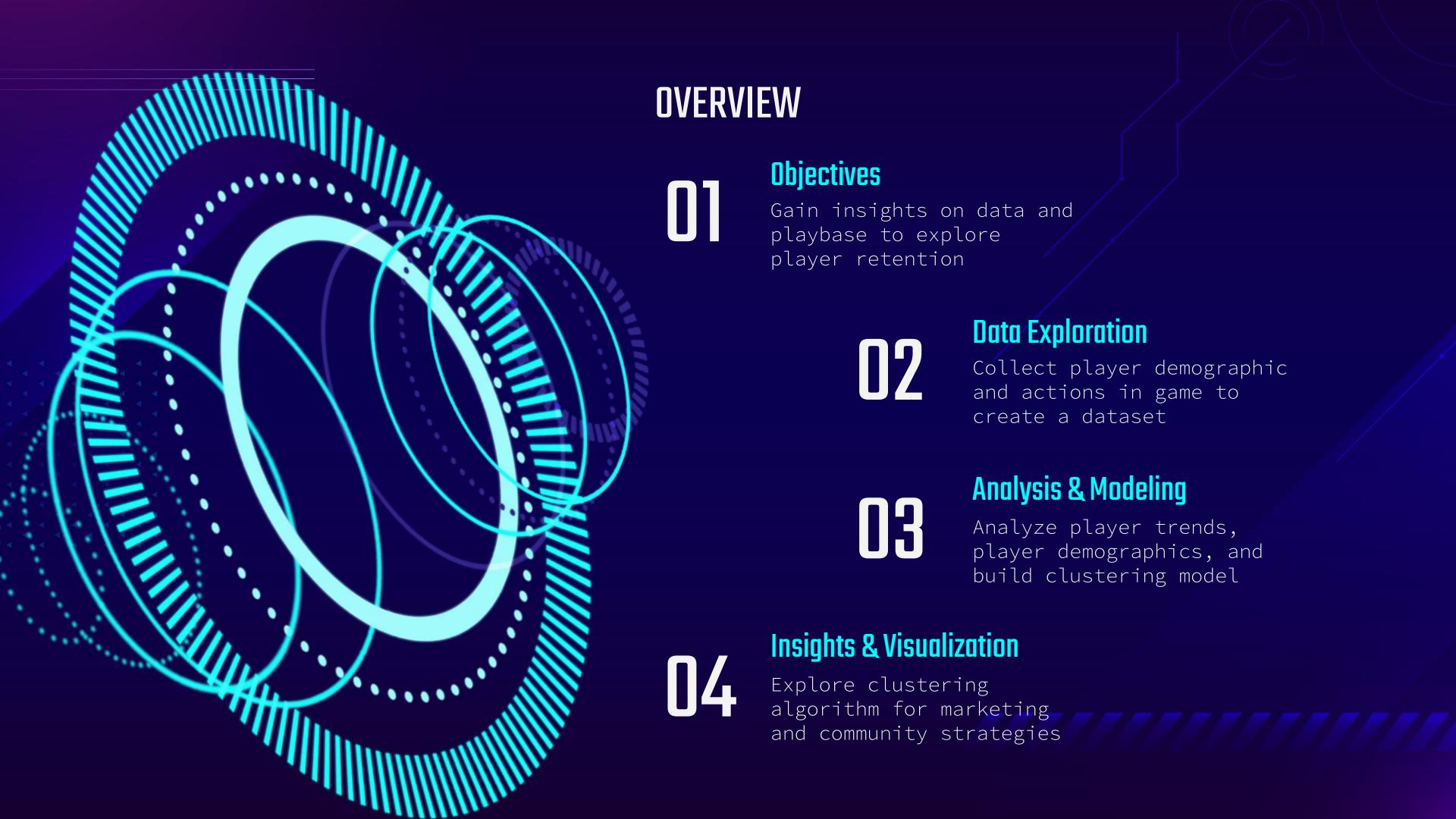


# Data Analysis

# Presentation

Christopher Angeles





# OVERVIEW

01

## Objectives

Gain insights on data and playbase to explore player retention

02

## Data Exploration

Collect player demographic and actions in game to create a dataset

03

## Analysis & Modeling

Analyze player trends, player demographics, and build clustering model

04

## Insights & Visualization

Explore clustering algorithm for marketing and community strategies

# Life Beyond



DAREWISE  
ENTERTAINMENT

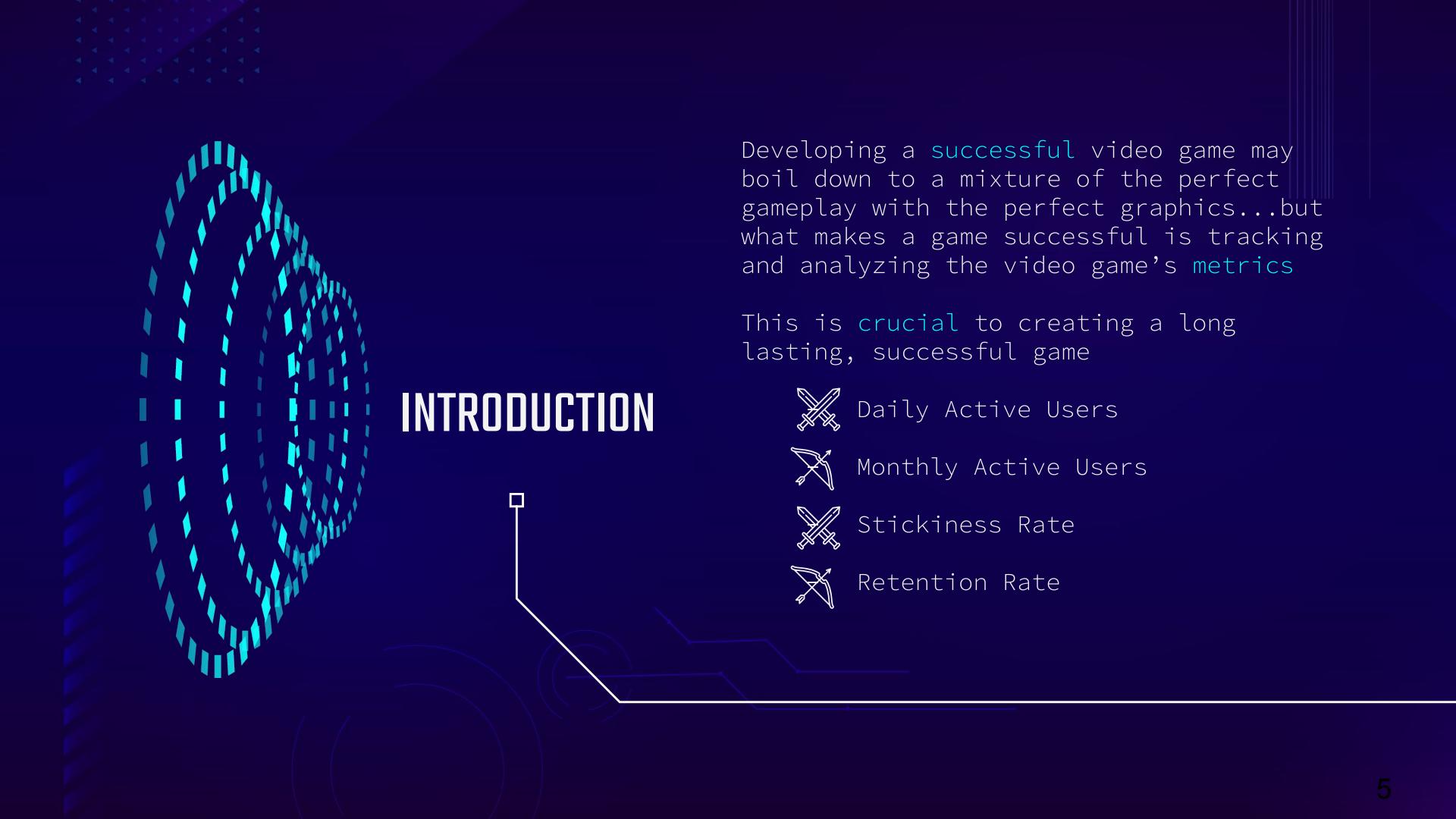
Darewise is a company looking to create long-term and meaningful social experiences for players, changing how the world views online games

They are the developers behind Life Beyond, a currently in development, open world and persistent game where teamwork and player choice unlock rich social adventures

## DISCLAIMER

- ◀ ALL DATA IS THE RIGHT OF DAREWISE AND THIS INFORMATION IS UNDER NDA
- ◀ GRAPHS AND FIGURES WILL SHOW TRENDS, BUT NO SPECIFIC NUMBERS

THANK YOU FOR  
UNDERSTANDING



# INTRODUCTION

Developing a successful video game may boil down to a mixture of the perfect gameplay with the perfect graphics...but what makes a game successful is tracking and analyzing the video game's metrics

This is crucial to creating a long lasting, successful game



Daily Active Users



Monthly Active Users



Stickiness Rate



Retention Rate



# PROBLEM

Darewise is looking to explore their player base and bring insights to player retention

By exploring player actions and building two models, we will present a strategy to increase player retention

# PROCESS & TOOLS

## DATA ANALYSIS

Extract and query data from Google Cloud Storage, and utilize python and pandas to create dataframe

## MODEL BUILDING & METRICS

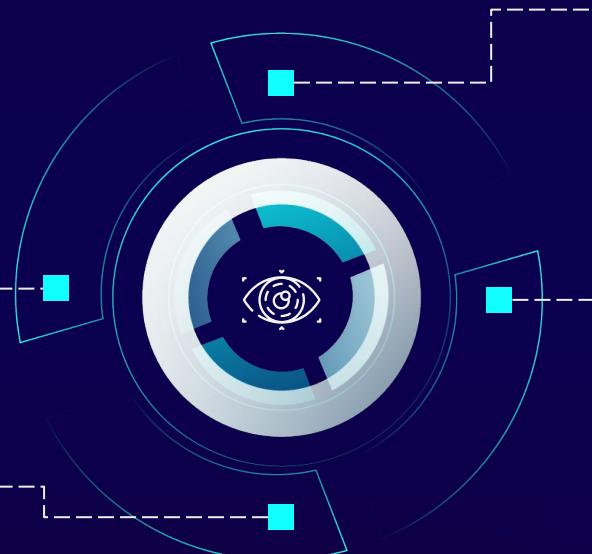
Build a model with scikit-learn library and pandas

## DATA PROCESSING

Explore and beautify data via reconfiguration, scaling and data manipulation

## STRATEGY & INSIGHTS

Present strategic insights to increase player retention



Identifying what your players are doing in game is just as important as knowing your community outside your game

## DATA ACQUISITION

Data was not only collected through players in game actions but also gamers that expressed interest in Life Beyond

# POTENTIAL & TARGET PLAYER METRICS

## GENDER

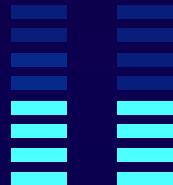


%

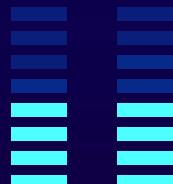


%

## AGE



15-25      26-35



35-45      45+

## AVERAGE TIME PLAYED



TOTAL DAYS PLAYED



## PREFERRED GENRE



## PREFERRED GAMEPLAY



# DATA EXPLORATION

After collecting our data and creating a dataframe, we asked what could this **information** tell us

DAU, MAU, & STICKINESS

How **many** players do we have a daily or monthly and which daily gamers are also monthly gamers?

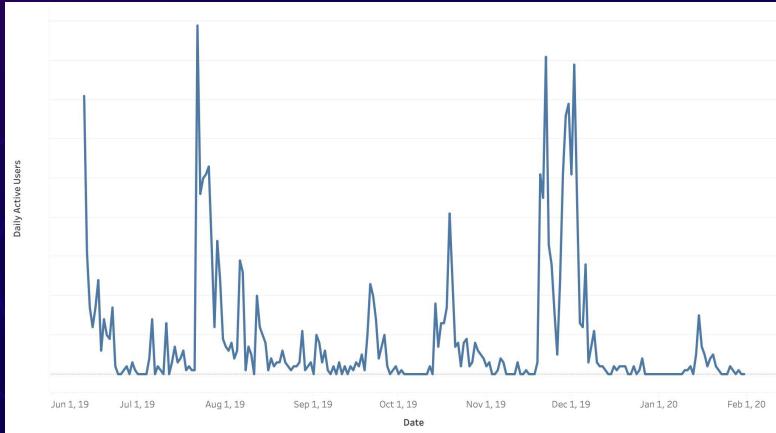
GAMEPLAY PREFERENCE

As an MMORPG, what **types** of gamers are playing Life Beyond

GAME ACTIVITY

What type of gamer is doing which **activity** in Life Beyond?

# GAMEPLAY OVERVIEW

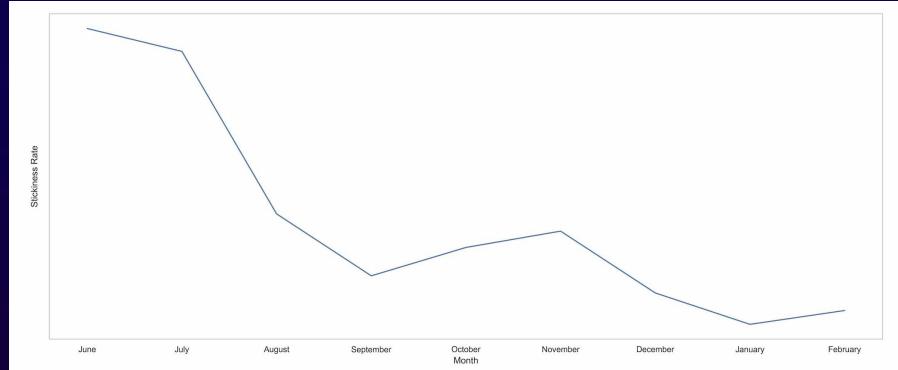


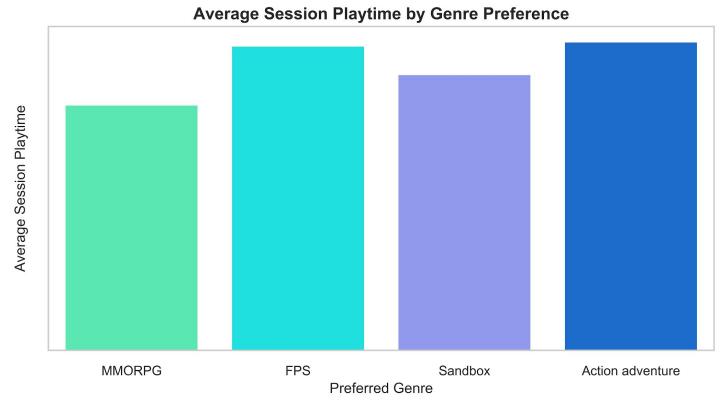
## DAILY ACTIVE USERS

The DAU trends peak and fall, most likely due to new content releases

## STICKINESS RATE

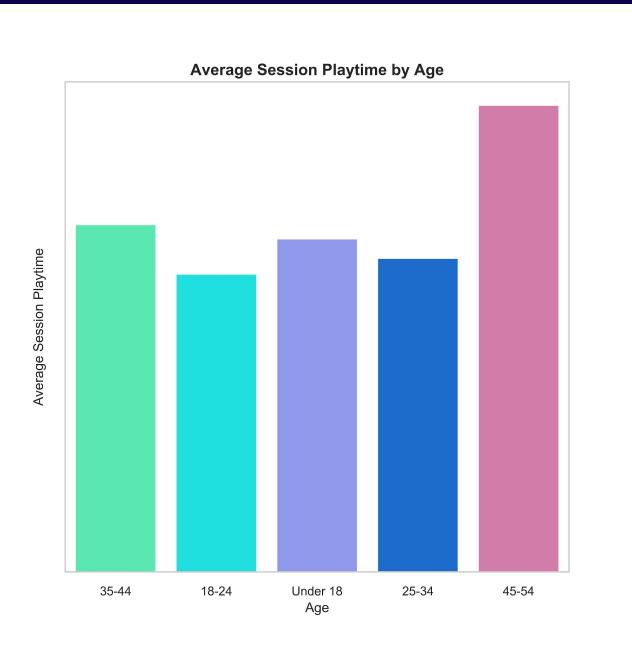
The ratio of monthly users that are also daily users gradually decline, meaning a decrease in retention of players





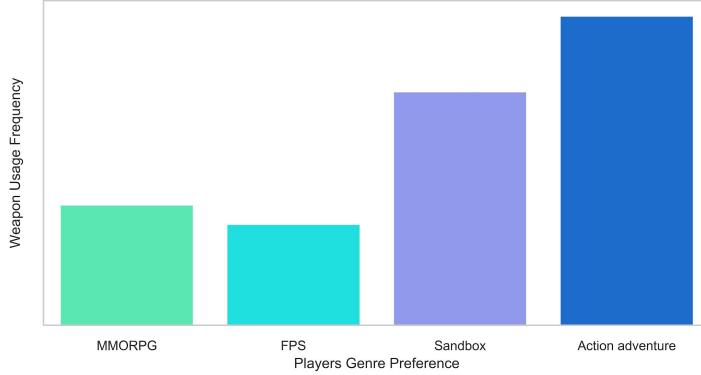
The oldest age group tends to have [longer](#) play sessions as compared to the younger age group

Although Life Beyond is an [MMORPG](#), players who prefer FPS, Sandbox, and Action games are playing longer

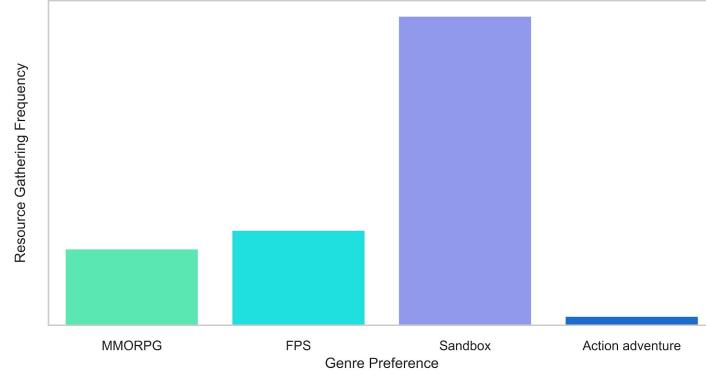


# GAME ACTIVITY

Weapon Usage Frequency by Genre Preference

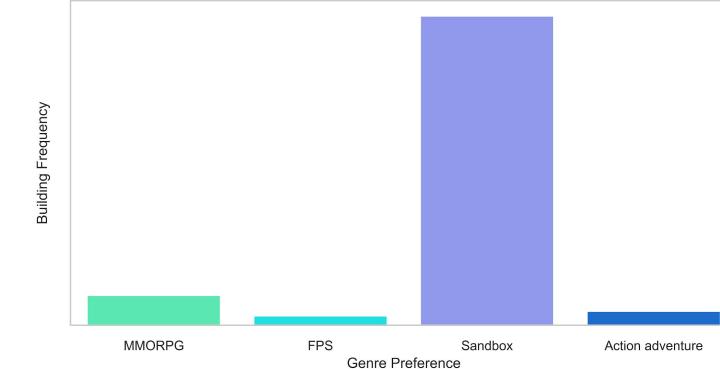


Resource Gathering Frequency by Genre Preference



Players whose favorite type of game are sandboxes tend to gather and build in game **more** frequently

Building Frequency by Genre Preference



Action adventure gamers **prefer** to combat and shooting, rather than resource collecting or building

# MODEL OBJECTIVES



## CLUSTERING

A clustering model can help understand not only what players are doing, but also group them to enhance retention strategies



## PREDICTION

A prediction model can use our players activities to see if new players will play shorter or longer than the average current player

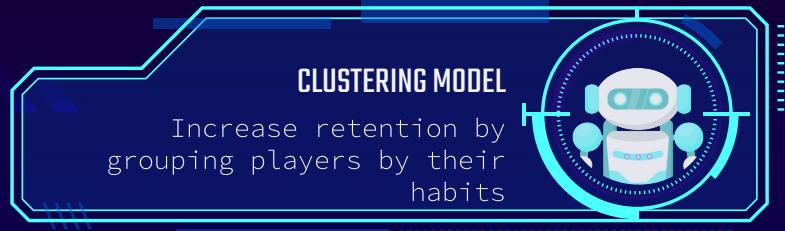


## SOLUTION

Using these models, we can help increase retention by selecting new players that are most likely going to play our game for longer periods of time

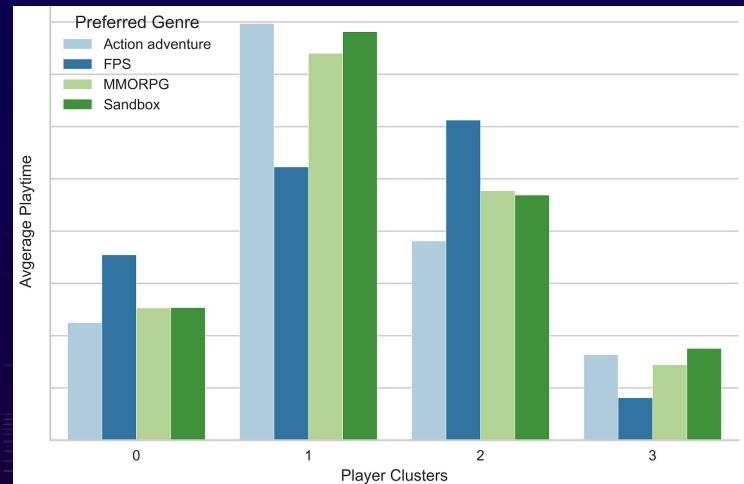
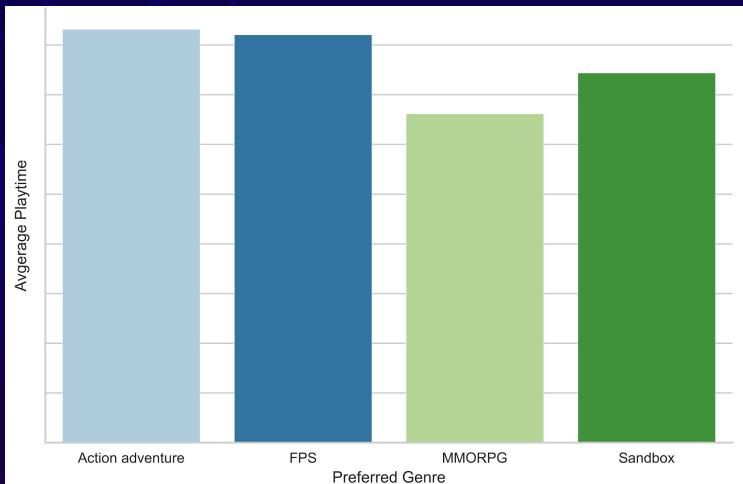
# CLUSTERING ALGORITHM

By grouping our players, based on their ingame habits, we are able to target players that have a high average playtime



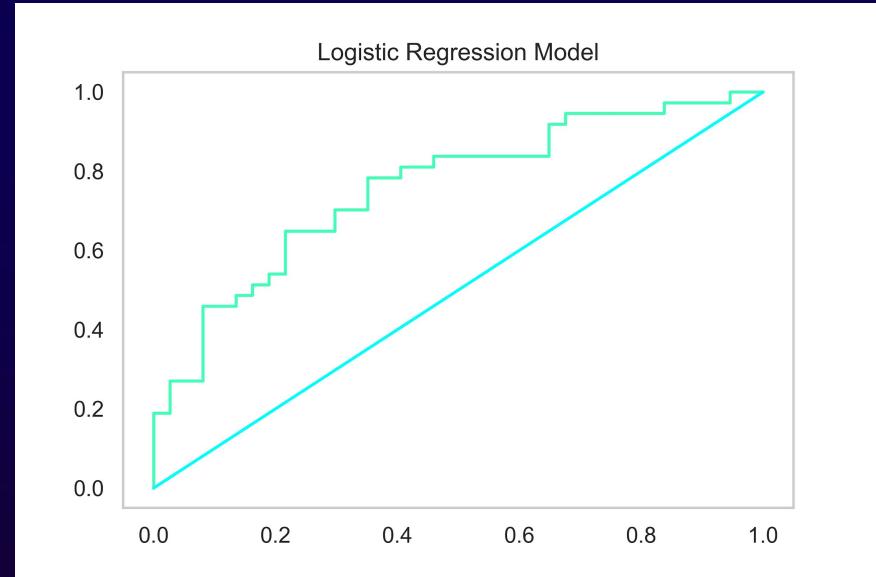
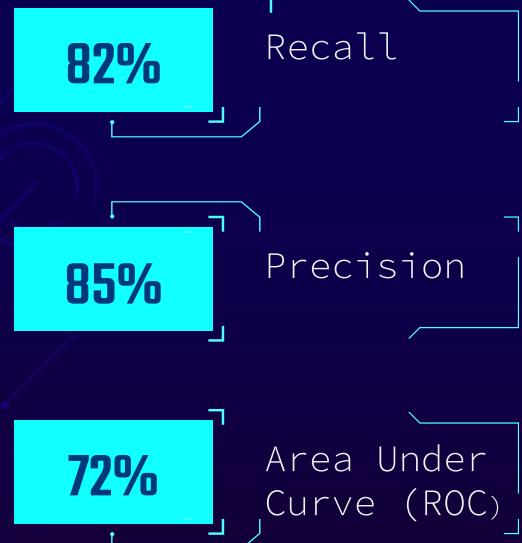
4

Clusters



# PREDICTION ALGORITHM

This model predicts players that have long play sessions



# MODEL OPTIMIZATION

## PARAMETERS

To optimize the model, we tested different models on their accuracy

		0	1
1	GaussianNB	0.702703	
0	KNeighborsClassifier	0.72973	
3	RandomForestClassifier	0.824324	
4	CatBoostClassifier	0.851351	
2	DecisionTreeClassifier	0.864865	

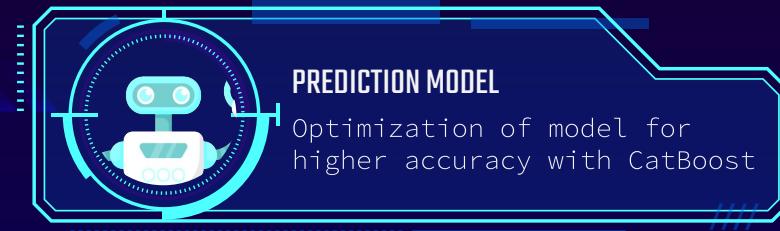
```
1 X_train, X_test, y_train, y_test = train_test_split(df2.drop('Cluster', axis=1),  
2 df2.Cluster, test_size=1/3)  
  
1 model_list=[KNeighborsClassifier(), GaussianNB(),  
2 DecisionTreeClassifier(),  
3 RandomForestClassifier(),  
4 CatBoostClassifier()]  
  
1 l_acc = []  
2 l_cm = []  
3 for model in model_list:  
4     model2=model.fit(X=X_train, y=y_train)  
5     y_pred2 = model2.predict(X_test)  
6     l_acc.append(accuracy_score(y_test,y_pred2))  
7     l_cm.append(confusion_matrix(y_test,y_pred2))  
8     print(type(model2).__name__, ' is done')
```

## CATEGORICAL BOOST

Proceeded with Cat Boost since this model allows for further fine tuning

# PREDICTION ALGORITHM

By optimizing my model using Categorical Boost, we are able to increase accuracy



## PREDICTION MODEL

Optimization of model for higher accuracy with CatBoost

**92%**

Accuracy

**95%**

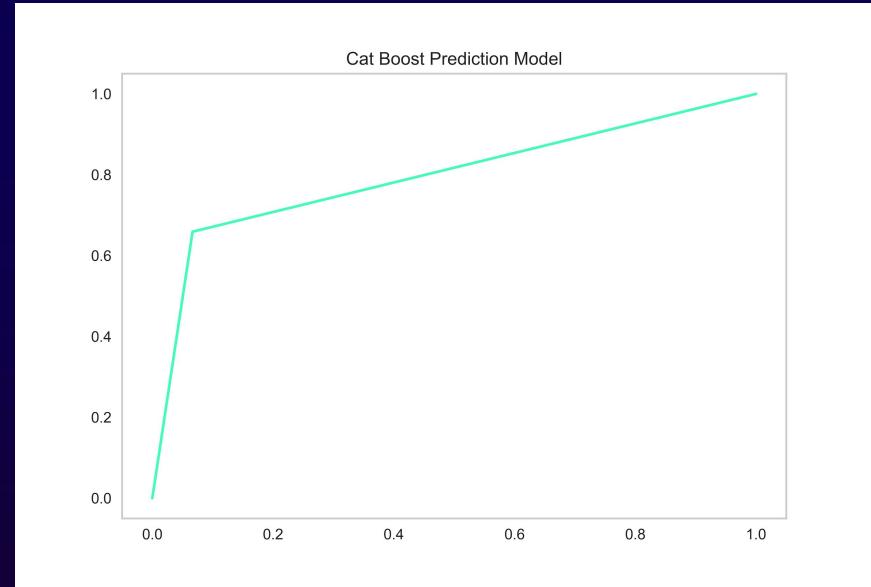
Recall

**91%**

Precision

**79%**

Area Under Curve (ROC)



# STRATEGY & SOLUTION

## PLAYER STATISTICS

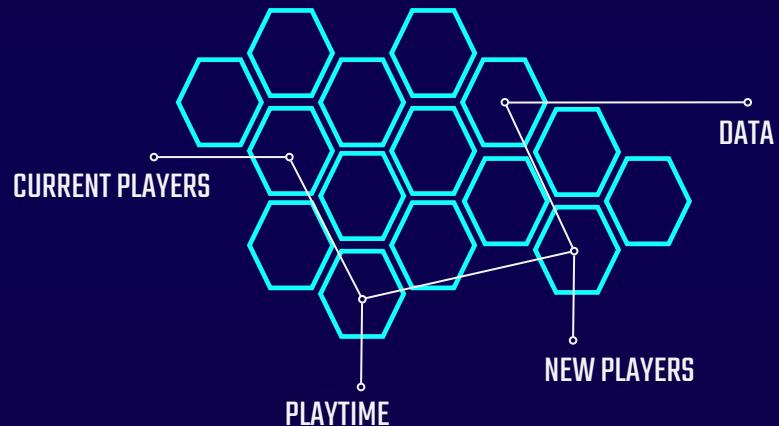
Since Life Beyond is still in development, only a select number of players are in game

## NEW PLAYERS

With this in mind, numerous players have shown interest in playing

## INVITE BEST PLAYERS

Using this prediction model, we can invite new players that will play Life Beyond the longest



## THANKS

Does anyone have any  
questions?  
[Christopher.Angeles@gmail.com](mailto:Christopher.Angeles@gmail.com)

## CREDITS

Special thanks to Darewise for  
offering me their data  
  
And to Ironhack for teaching  
me how to be a data analyst



[github.com/cangeles14](https://github.com/cangeles14)



[linkedin.com/in/chris-angeles](https://linkedin.com/in/chris-angeles)