

Predicting Online Game Player Count through SARIMAX

Ryan McArthur, Flatiron School,
Data Science



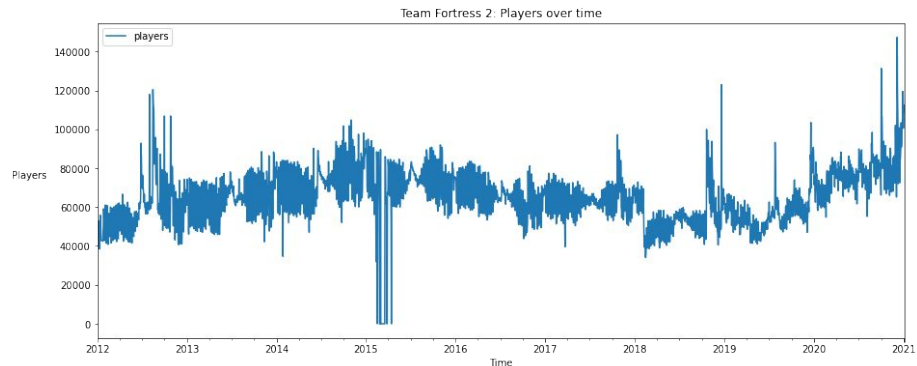
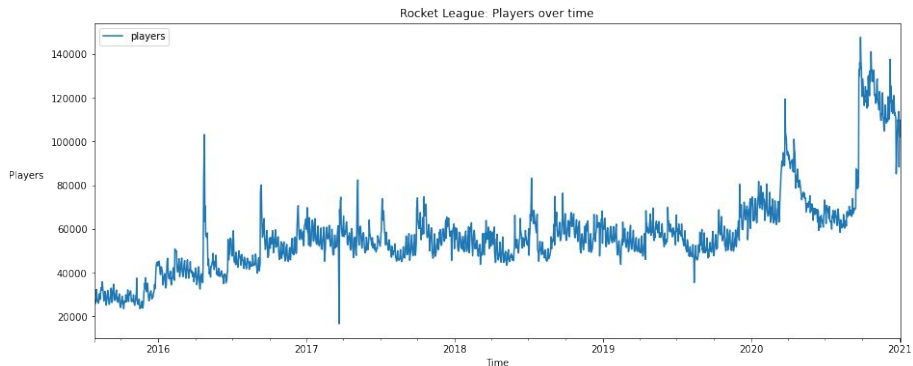
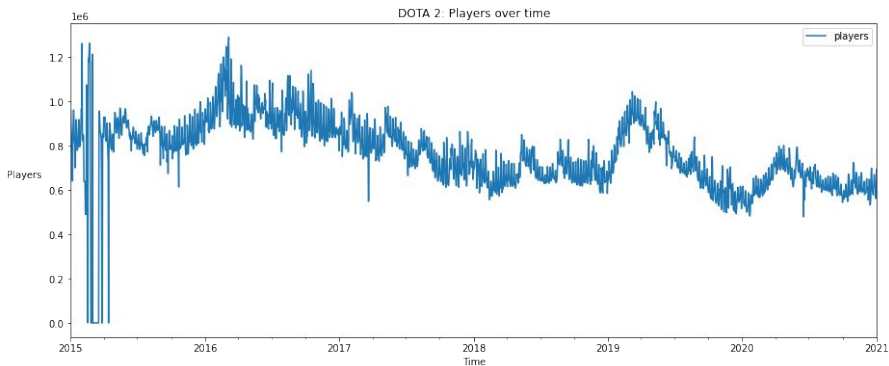
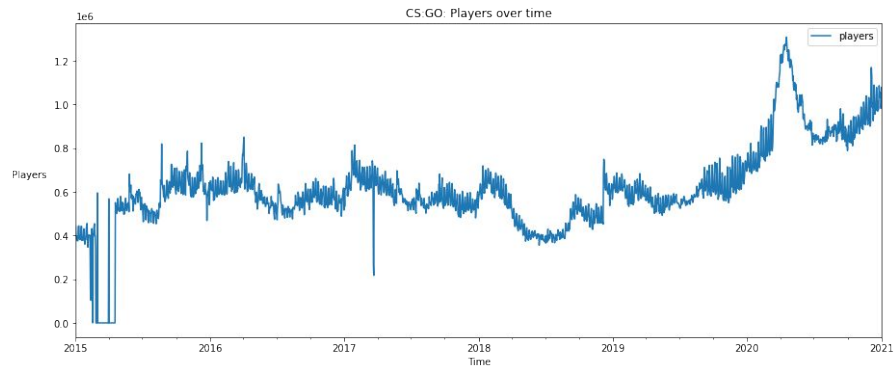
Introduction

- 85% of all video game revenue is generated from free-to-play games, which rely on advertisements to generate revenue.
- Game production companies must be able to show that their games reach a large number of players
- By using time-series data on number of concurrent players, we will be able to construct a model that production companies may use to prepare for the future.

Data

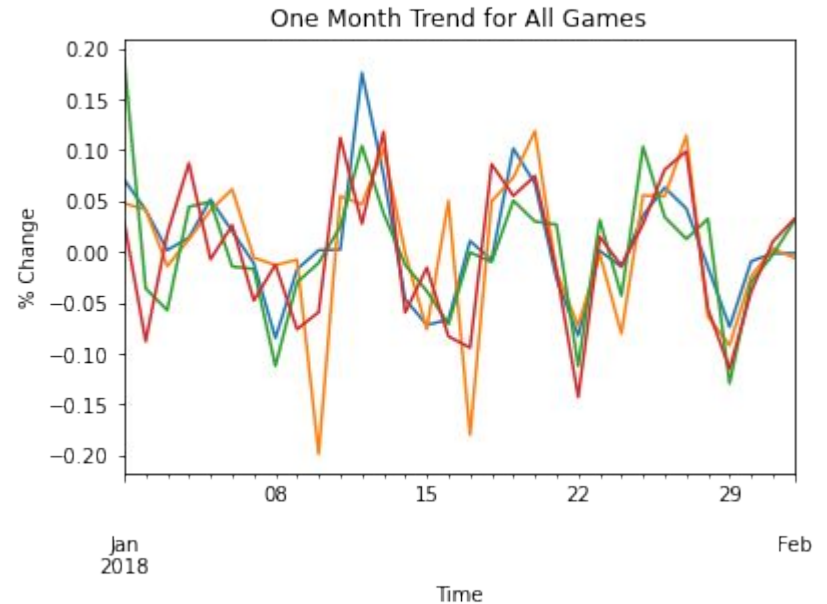
- Our data comes from SteamDB, or Steam Database.
- Data provides daily statistics on the number of concurrent players, number of concurrent Twitch viewers, and whether an event is ongoing.
- We used 4 games:
 - CS:GO
 - DOTA 2
 - Rocket League
 - Team Fortress 2

Visualizing our Raw Data



All our games share the same Trend

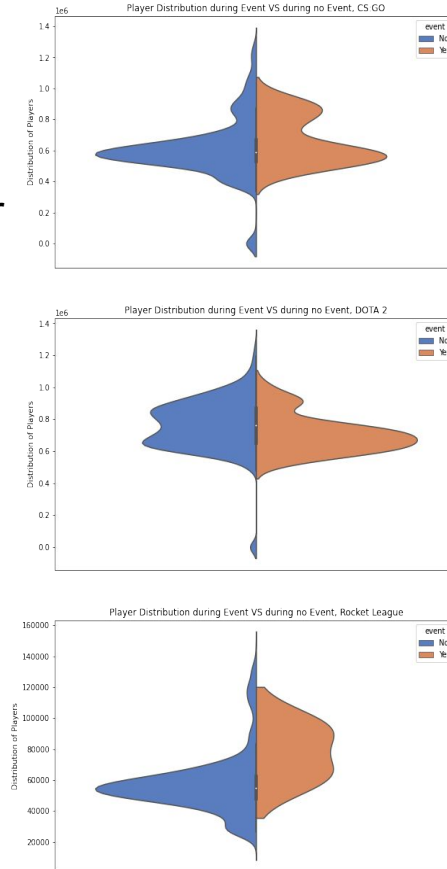
- We found that when visualizing trends on a monthly level, it was plain to see that all of our games follow the same trends in changes in player counts.
- Player counts are higher during weekends.



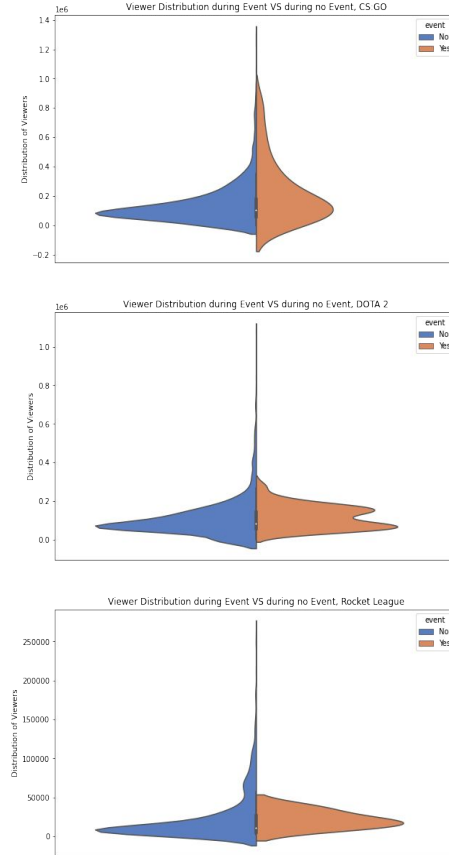
Events

- We found that while events do have a positive impact on number of players, it was not present enough to merit further investigation.
- The same is true for number of viewers, suggesting that events do not have a strong impact on these counts.

Players

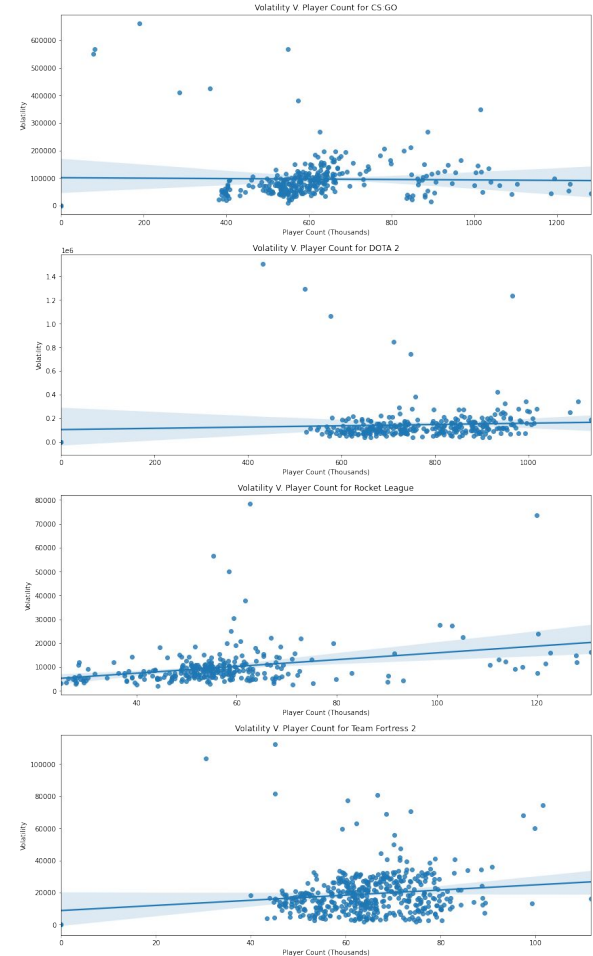


Viewers



Volatility trends

- We found that volatility tends to increase as number of players increases
- This suggests that a games maximum number of players increases at a rate greater than average number of players



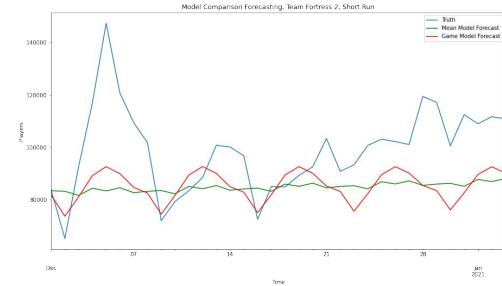
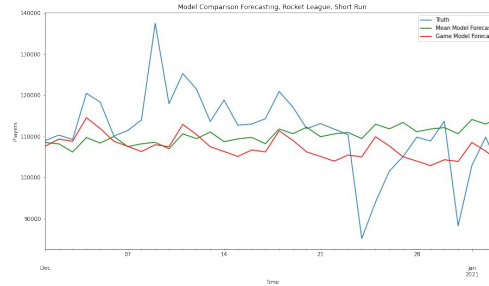
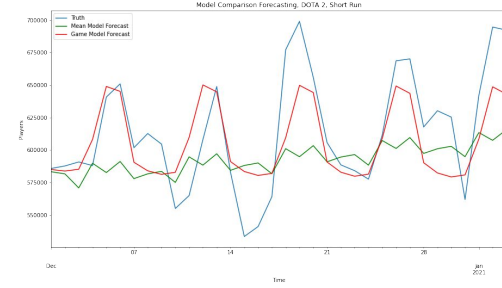
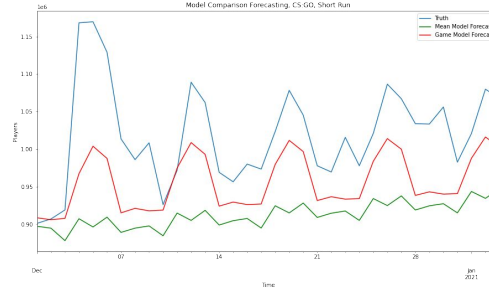
Modelling Process

- Two types of models:
 - First, we constructed basic models for each game, training on its own data.
 - Second, we constructed an averaged model, using average percent change in number of players across all games.

Modelling Results

Short Run Forecasting

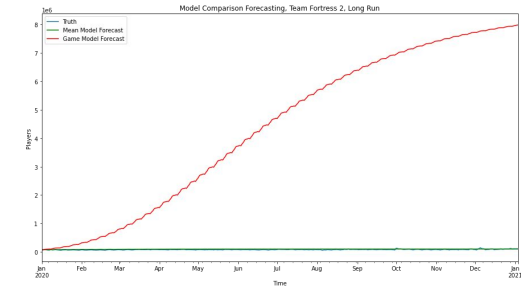
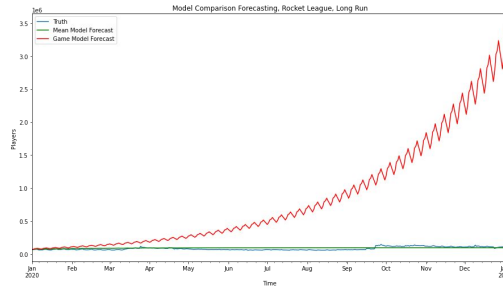
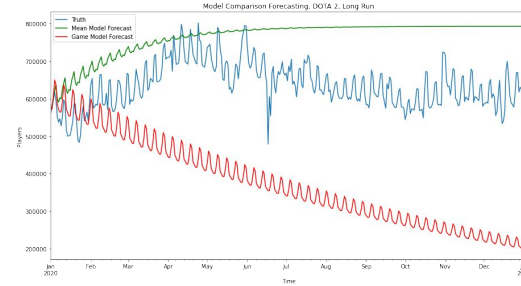
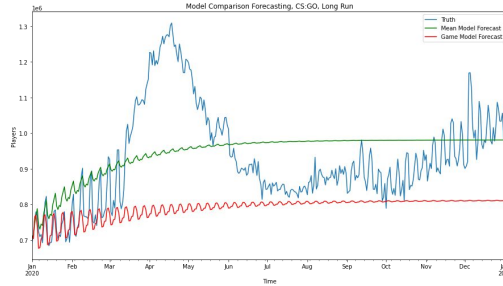
- Forecasting 1 month, basic models perform best.
- Mean Model fails to capture seasonality



Modelling Results

Long Run Forecasting

- Our basic models diverge, or do not predict trend well.
- Our averaged model performs best for long run predictions, predicting long run trends best.



Recommendations to Production Companies

1. Server Capacity must increase faster than player growth, since volatility is more of an issue at large player counts.
2. Events may not be the best way to increase social media presence, since number of viewers was not shown to increase during an event.
3. Long-term predictions should be made on market-wide analysis, while short-term predictions are best made by a games own historical data

Further Work

- Collecting hourly statistics instead of daily statistics
- Event Investigations
- Company investigations



Thank You!

Ryan McArthur, Flatiron School
rmcar522@gmail.com

