Game Data Analysis

By: Mateo Fernandez, Jason Pham, Jake Preston

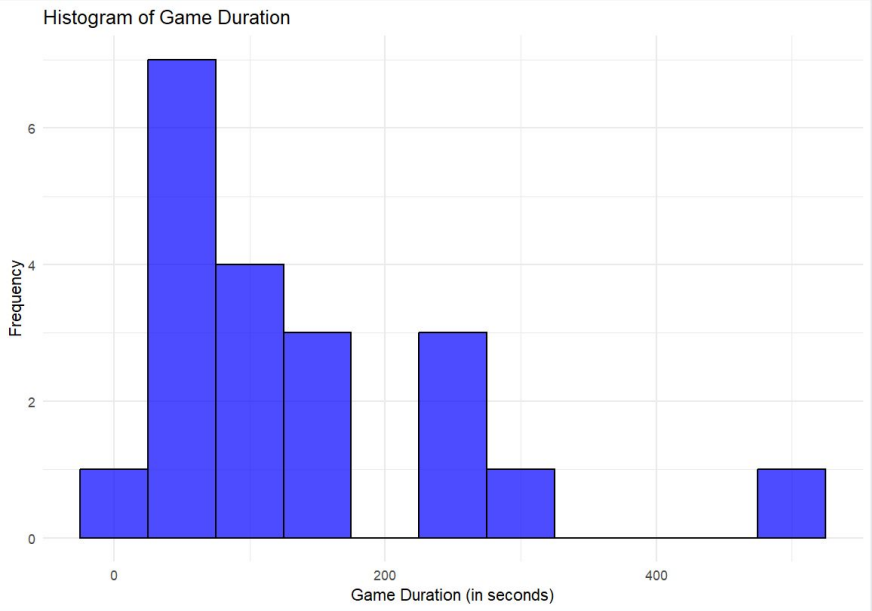
March 13, 2025

Overview

After going through the data hooks and raw data collected we determined that there was no significant relationships that we could detect between meaningful variables. We attempted to find some sort of correlation between any two variables but nothing meaningful could be found. The only slightly meaningful correlation was that only players who accused the advisor had a value for $TownSquareTime and no value for $LilyRoomTime. Players who accused the other 2 options had the opposite, a value for $LilyRoomTime and not $TownSquareTime. Upon further inspection, however, this is due to the fact that there is no way to go back in there game without pressing the back arrow and doing so deletes the data collected by data hooks because you are rewinding the game. Looking deeper we found some

Question: What does overall player engagement with the game look like?

Player Engagement (Methodology)



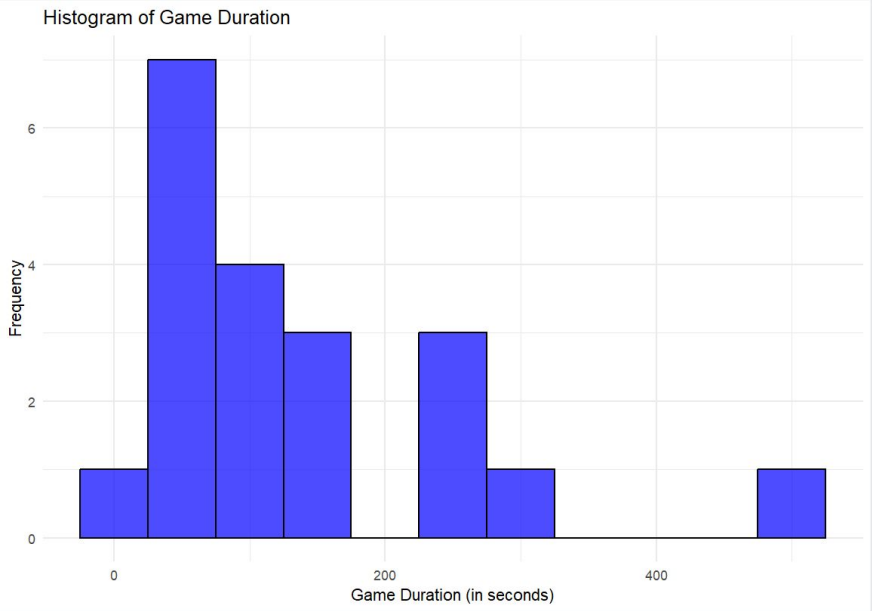
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16.22	50.85	109.03	139.40	190.25	479.87

Values in seconds

In order to find what engagement looked like for the audience, we decided to plot out a histogram of the distribution of the gameplay duration in seconds ($GameDuration), along with some numbers from the distribution (min, max, mean, etc.)

Std. dev: 116.97 seconds
Variance: 13,682.39

Player Engagement (Findings)



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16.22	50.85	109.03	139.40	190.25	479.87

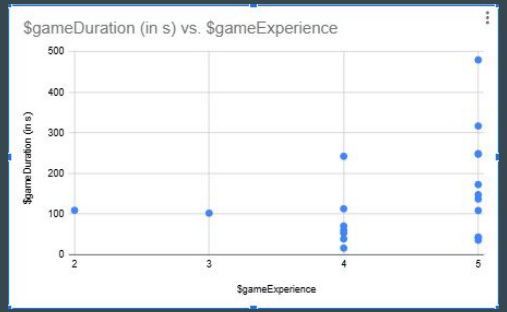
Values in seconds

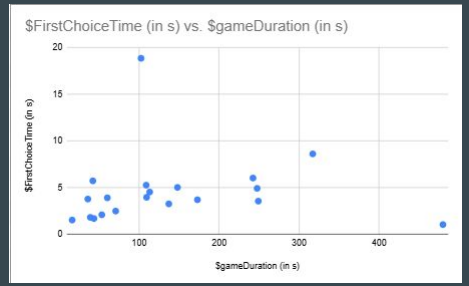
Std. dev: 116.97 seconds
Variance: 13,682.39

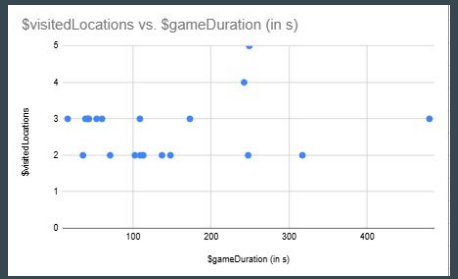
- The average engagement time with the game for 139.4 seconds (mean).
- Right-skewness of the histogram suggests that most players play shorter sessions, there is are smaller amounts of players that play for longer sessions, which increased the mean.
- With the skewness, the median playtime was 109.03 seconds.
- This notion of variability in playtime is also supported by a large standard dev. (116.97s)

Looking further into Game Duration (Methods)

Looking further in, we wanted to see if any variables had a meaningful relationship with the $GameDuration variable in order to determine if certain decisions kept players playing for longer periods. To do this we compared other variables like $FirstChoiceTime, $VisitedLocations, and $GameExperiences. We compared each variable against $GameDuration by creating scatterplots that would hopefully allow us to see any type of relationship between the variables.

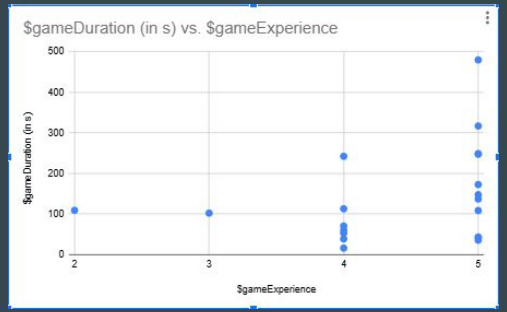


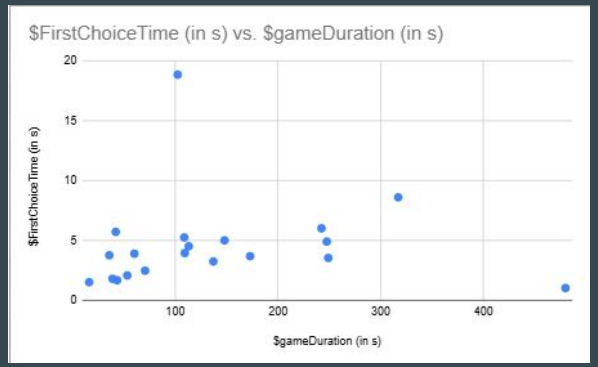


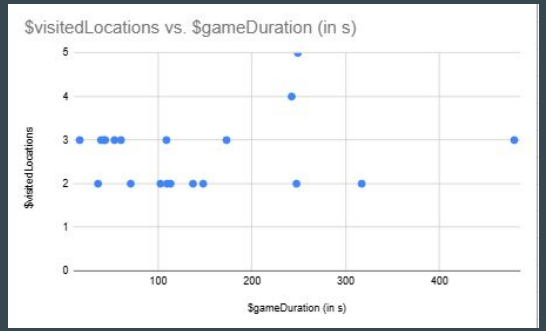


Looking further into Game Duration (Findings)

At first we attempted to find correlations between variables and the $GameDuration variable, to see if certain decisions kept players engaged for longer. We could find no correlation, however. $VisitedLocations and $FirstChoiceTime had no correlation. $FirstChoiceTime does demonstrate a slight positive correlation but only if you ignore the outlier where the player with the highest $GameDuration also had the quickest $FirstChoiceTime. Even then, it makes sense that a higher First Choice Time would lead to a higher Game Duration, so this does not tell us much. Even $GameExperience had little to no relationship with the $GameDuration Variable, with players from all experiences having different durations.



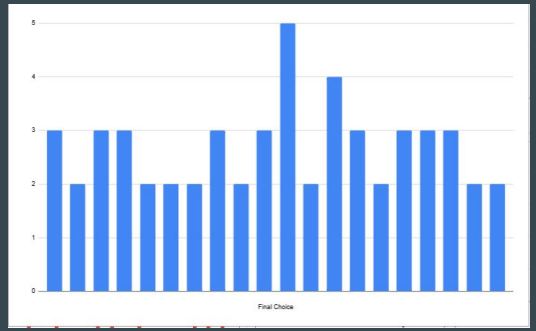




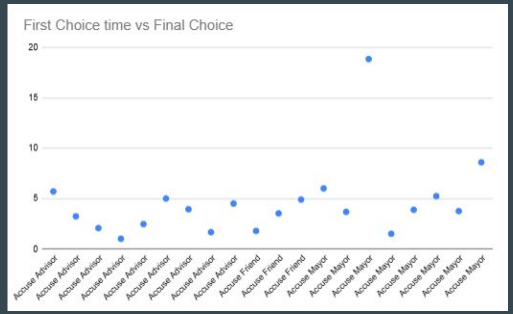
Question: What influences the final ending players got?

Methods ($FinalChoice)

We wanted to see if certain decisions and other variables influences the players Final Choice. To do this we compared certain variables against the $FinalChoice variable. Ultimately most of the variables had no correlation but after sorting we determined that there may be some correlation between the $FirstChoiceTime and $FinalChoice variables. We took the averages of the $FirstChoiceTime of each player sorted by $FinalChoice in order to determine this. We then did the same thing for the $GameDuration variable too.

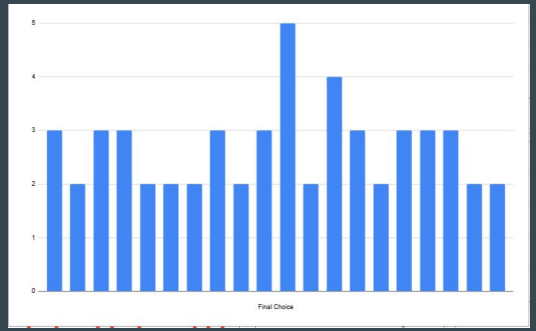


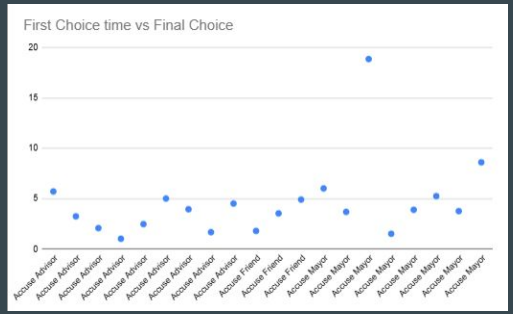
3.303888889	132.9874444
3.421666667	178.4746667
6.458125	131.96675



Analysis ($FinalChoice)

Then we tried to see if the final choice was influenced by any of the other variables. Here we were able to find something interesting. We found that players whose Final Choice was Accuse Mayor had a significantly higher average $FirstChoiceTime. While the other 2 options have averages of 3.3 and 3.4 respectively, Accusing the Mayor had an average of 6.4, almost double. Removing the large outlier of 18 seconds still leaves the average at 4.6, which we believe to be significantly higher than the other 2. This could mean that players who were more hesitant or calculating tended to accuse the mayor, since them taking a longer time to make the first decision could mean they are more hesitant or apprehensive as players. This however, is not backed up by the $GameDuration variable. For this variable all Final choices had very similar results (Advisor: 133, Friend: 178, Mayor: 132). While the friend option is higher, we believe that this is due to the fact that only 3 players chose this option, and if more players played the game it could very easily level out to a similar value to the other options.





Conclusion

The average player was not as engaged with the game as the outlier players were. The graphs of player game duration, or engagement with the game, were very right skewed. This was also backed up by the fact that we compared multiple variables against game duration and found that playstyle did not impact player engagement in a meaningful way, demonstrating that players tended to not play for very long no matter how they played the game, meaning the game may not have strong features that keep players playing. This could also be due to the fact that playstyle did not seem to impact the ending players got, and even playing the game it seemed like I could not impact the game much because it railroaded me to whatever end I started on and I could not go back with game mechanics.