David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

# CS105 Final Project Report - Rocket League Championship Data

**Project Overview**

The two datasets that we're using (games_by_teams.csv, games_by_players.csv) contain data from Championship Series 2021-2022 that is full of Rocket League player and team data. The data consists of a team's or players number of goals scored, number of assists, time spent on offensive positioning, time spent on defensive positioning, speed boost collected, amount of air time, and amount of ground time. The player's dataset contains exclusive features such as mvp rating, goal participation, and steering sensitivity. This dataset is provided by octane.gg and ballchasing.com to see if there is a pattern or strategy to winning games in the championship.

The purpose of this project is to use the RLCS '21-'22 datasets to draw correlations between winning matches alongside various game statistics of players that perform in the Rocket League Championship Series 2021-2022. The goal is to relate winning matches to data points like goals scored, speed boost, movement, positioning, steering sensitivity, and mvp status with respect to wins. To find correlations, we plan to do K-Means Clustering to find clustering for ball possession time and core score, and do Chi-Square Analysis for MVP status, ground movement time, steering sensitivity, and performance.

**Data Collection & Cleaning**

Our Rocket League Championship Series 2021-2022 Data from Kaggle:
https://www.kaggle.com/datasets/dylanmonfret/rlcs-202122?select=games_by_players.csv

The dataset we have is a Rocket League Championship Series dataset that includes game data for 34,000 teams and 101,000 players. The team dataset(games_by_teams.csv) includes a team's number of goals, assists, saves, and assists, along with a team's boost amount collected, total distance of movement, amount of time spent on offensive and defensive positioning, and the number of cars demolished. The player dataset(games_by_players.csv) contains the same features as the team dataset, with differences only in the amount of goal participation, MVP, and steering sensitivity.

David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

This is a preview of the games_by_team dataset before cleaning:

| | game_id | color | team_id | team_slug | team_name | team_region | ball_possession_time | ball_time_in_side | core_shots | core_goals | ... | positioning_time_defensive_third |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 616004f3143c37878b238690 | blue | 6020bc8ef1e4807cc700391a | https://octane.gg/teams/391a-ground-zero-gaming | GROUND ZERO GAMING | Oceania | 136.54 | 170.84 | 13 | 2 | ... | 513.73 |
| | 616004f3143c37878b238690 | orange | 614c8930f8090ec745286474 | https://octane.gg/teams/6474-ranga-roundup | RANGA ROUNDUP | Oceania | 153.60 | 148.51 | 4 | 1 | ... | 457.04 |
| | 616004f7143c37878b238697 | blue | 6020bc8ef1e4807cc700391a | https://octane.gg/teams/391a-ground-zero-gaming | GROUND ZERO GAMING | Oceania | 155.79 | 145.60 | 10 | 4 | ... | 507.95 |
| | 616004f7143c37878b238697 | orange | 614c8930f8090ec745286474 | https://octane.gg/teams/6474-ranga-roundup | RANGA ROUNDUP | Oceania | 126.81 | 185.78 | 6 | 2 | ... | 561.51 |
| | 616004fc143c37878b23869e | blue | 6020bc8ef1e4807cc700391a | https://octane.gg/teams/391a-ground-zero-gaming | GROUND ZERO GAMING | Oceania | 165.76 | 136.11 | 11 | 3 | ... | 476.66 |
| | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| [illegible] | 6295db30da9d7ca1c7bb1618 | orange | 607d7d1f473f172e4c0d12c2 | https://octane.gg/teams/12c2-maycam-evolve | MAYCAM EVOLVE | South America | 163.60 | 167.88 | 6 | 1 | ... | 459.75 |
| [illegible] | 6295db34c437fde7e02d8d91 | blue | 6020bda0f1e4807cc700e078 | https://octane.gg/teams/e078-true-neutral | TRUE NEUTRAL | South America | 158.32 | 130.84 | 11 | 4 | ... | 475.72 |
| [illegible] | 6295db34c437fde7e02d8d91 | orange | 607d7d1f473f172e4c0d12c2 | https://octane.gg/teams/12c2-maycam-evolve | MAYCAM EVOLVE | South America | 129.38 | 197.09 | 7 | 2 | ... | 546.21 |
| [illegible] | 6295db38c437fde7e02d8d98 | blue | 6020bda0f1e4807cc700e078 | https://octane.gg/teams/e078-true-neutral | TRUE NEUTRAL | South America | 131.27 | 135.18 | 14 | 4 | ... | 462.51 |
| [illegible] | 6295db38c437fde7e02d8d98 | orange | 607d7d1f473f172e4c0d12c2 | https://octane.gg/teams/12c2-maycam-evolve | MAYCAM EVOLVE | South America | 149.65 | 197.93 | 7 | 2 | ... | 603.22 |

This is a preview of the games_by_players dataset before cleaning

| | game_id | color | team_id | team_region | player_id | player_tag | core_shots | core_goals | core_saves | core_assists | ... | car_id | car_name | steering_sensitivity | camera_fov |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 616004f3143c37878b238690 | blue | 6020bc8ef1e4807cc700391a | Oceania | 5f3d8fdd95f40596eae2412e | Amphis | 4 | 1 | 1 | 0 | ... | 4284.0 | Fennec | 3.05 | 110.0 |
| 1 | 616004f3143c37878b238690 | blue | 6020bc8ef1e4807cc700391a | Oceania | 5f3d8fdd95f40596eae23e01 | Torsos | 5 | 0 | 2 | 1 | ... | 403.0 | Dominus | 1.00 | 110.0 |
| 2 | 616004f3143c37878b238690 | blue | 6020bc8ef1e4807cc700391a | Oceania | 5f3d8fdd95f40596eae23e53 | Express | 4 | 1 | 0 | 0 | ... | 4284.0 | Fennec | 1.30 | 110.0 |
| 3 | 616004f3143c37878b238690 | orange | 614c8930f8090ec745286474 | Oceania | 604e562901d675f81a96b270 | mel kin | 2 | 0 | 6 | 0 | ... | 403.0 | Dominus | 1.51 | 110.0 |
| 4 | 616004f3143c37878b238690 | orange | 614c8930f8090ec745286474 | Oceania | 5f7ca648ea8a0f0714fb9a20 | Laxin | 1 | 0 | 3 | 0 | ... | 23.0 | Octane | 1.75 | 110.0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 100730 | 6295db38c437fde7e02d8d98 | blue | 6020bda0f1e4807cc700e078 | South America | 601a30c6d35e6ea972756329 | Seck | 5 | 2 | 1 | 0 | ... | 4284.0 | Fennec | 2.00 | 110.0 |
| 100731 | 6295db38c437fde7e02d8d98 | blue | 6020bda0f1e4807cc700e078 | South America | 5fbda6d716fbe0e46966afec | Navarro | 4 | 1 | 1 | 0 | ... | 23.0 | Octane | 1.50 | 110.0 |
| 100732 | 6295db38c437fde7e02d8d98 | orange | 607d7d1f473f172e4c0d12c2 | South America | 6074ce3000af93009ae4f7d6 | LagLy | 3 | 0 | 4 | 0 | ... | 4284.0 | Fennec | 2.30 | 108.0 |
| 100733 | 6295db38c437fde7e02d8d98 | orange | 607d7d1f473f172e4c0d12c2 | South America | 5f3d8fdd95f40596eae23f4c | Wais | 0 | 0 | 3 | 1 | ... | 2269.0 | '99 Nissan Skyline GT-R R34 | 1.00 | 110.0 |
| 100734 | 6295db38c437fde7e02d8d98 | orange | 607d7d1f473f172e4c0d12c2 | South America | 60597e85426b95b9dd517d74 | Rmnn | 4 | 2 | 0 | 0 | ... | 23.0 | Octane | 7.00 | 110.0 |

100735 rows × 103 columns

To start our project, We cleaned the data in order to only include relevant data points in our dataset that we used for our analysis. We checked if there are missing values, and it turns out there isn't any missing data.

This is a preview of the games_by_team dataset after cleaning:

| | ball_possession_time | core_shots | core_goals | core_saves | core_assists | core_score | boost_count_collected_big | boost_count_collected_small | movement_total_distance | positioning_time_defensive_third | positioning_ti |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 136.54 | 13 | 2 | 3 | 1 | 921 | 48.0 | 208.0 | 1433406.0 | 513.73 | |
| 2 | 155.79 | 10 | 4 | 3 | 3 | 1283 | 60.0 | 202.0 | 1581436.0 | 507.95 | |
| 4 | 165.76 | 11 | 3 | 2 | 3 | 1201 | 71.0 | 219.0 | 1679856.0 | 476.66 | |
| 6 | 140.93 | 14 | 8 | 2 | 8 | 1938 | 78.0 | 203.0 | 1686848.0 | 395.84 | |
| 8 | 142.42 | 14 | 6 | 2 | 6 | 1567 | 71.0 | 243.0 | 1742004.0 | 405.76 | |

This is a preview of the games_by players dataset after cleaning

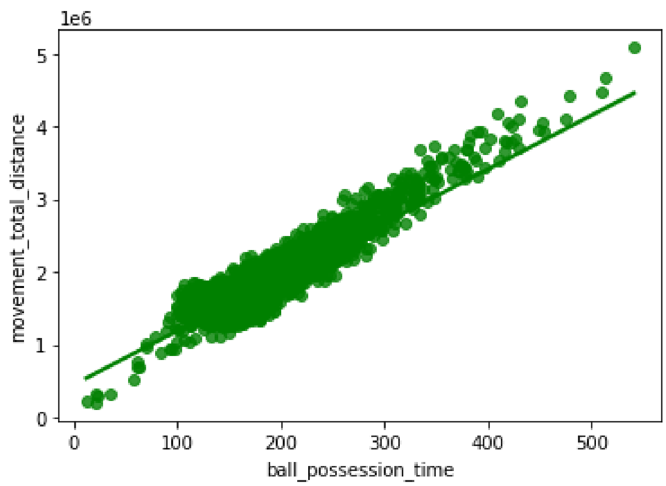
| | ball_possession_time | core_shots | core_goals | core_saves | core_assists | core_score | boost_count_collected_big | boost_count_collected_small | movement_total_distance | positioning_time_defensive_third | positioning_ti |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 136.54 | 13 | 2 | 3 | 1 | 921 | 48.0 | 208.0 | 1433406.0 | 513.73 | |
| 2 | 155.79 | 10 | 4 | 3 | 3 | 1283 | 60.0 | 202.0 | 1581436.0 | 507.95 | |
| 4 | 165.76 | 11 | 3 | 2 | 3 | 1201 | 71.0 | 219.0 | 1679856.0 | 476.66 | |
| 6 | 140.93 | 14 | 8 | 2 | 8 | 1938 | 78.0 | 203.0 | 1686848.0 | 395.84 | |
| 8 | 142.42 | 14 | 6 | 2 | 6 | 1567 | 71.0 | 243.0 | 1742004.0 | 405.76 | |

Now that we have a clean dataset, with no NaN values, it is time to move on to exploratory data analysis.

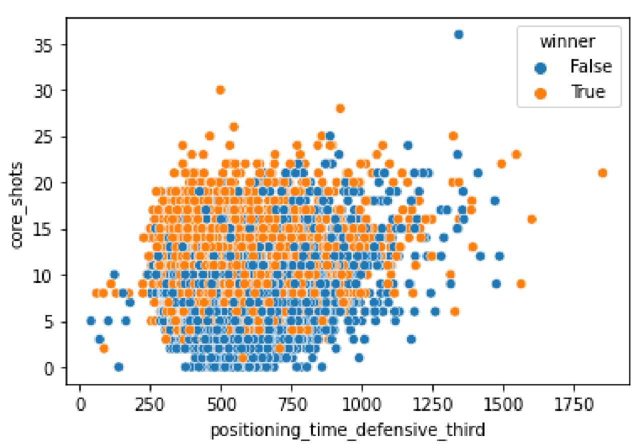
David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

**Exploratory Data Analysis**

David and Jonathan first used linear regression models to gauge whether correlations existed between a plethora of variable pairings which provided valuable insights into how Rocket League games often play out at a tournament level. Ball possession time and the total distance traveled over a game on a team-level turned out to be highly correlated demonstrating that the mechanics of the game are highly dynamic.

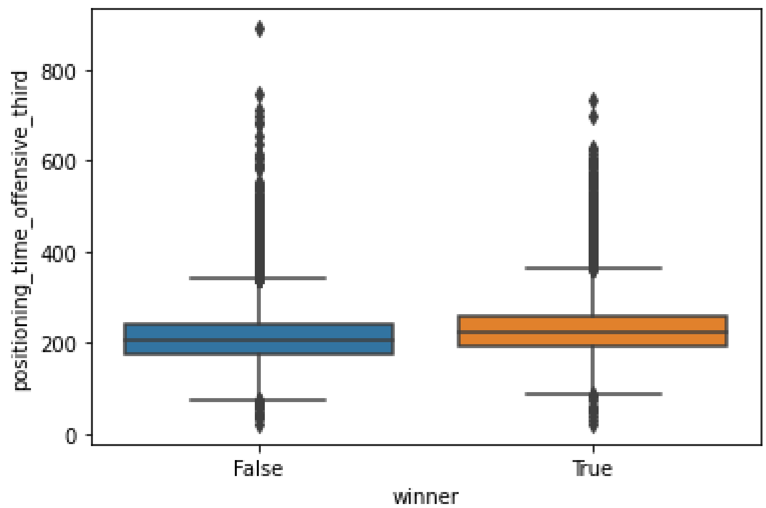



An interesting observation made in EDA that further gave insight into the value of ball possession within the context of scoring. David and Jonathan found that the magnitude of offensive positioning was positively correlated to goals attempted, independent of whether each team won or lost their match.

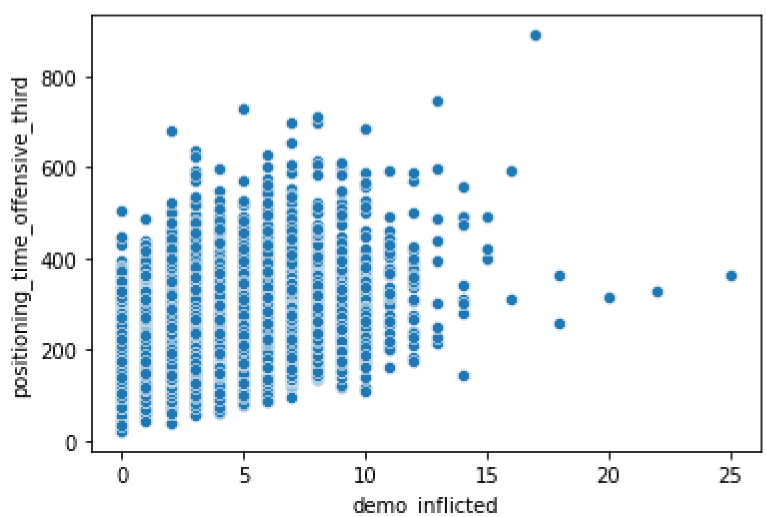


David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

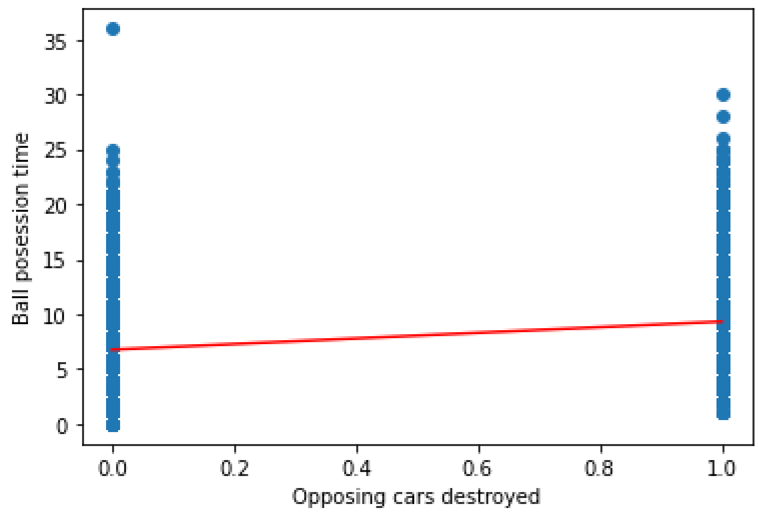
Additionally, it was realized that there was not much variability amongst winning and losing teams when it came to time spent on offense. We can therefore conclude that what sets high performing teams apart is not their ability to retain offensive positioning but creating and maximizing shot opportunities when on offense.



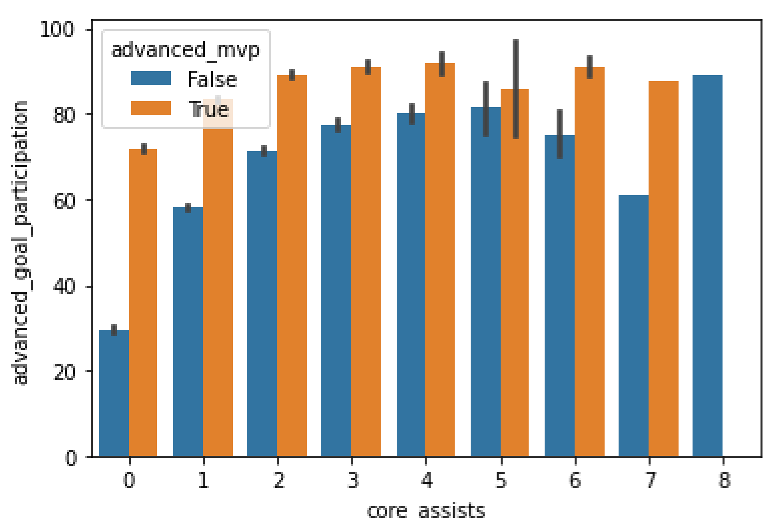

Acts of demolition or direct attacks on opposing players' vehicles proved to have variable relationships with time spent on offense and defense. Our initial assumption was that teams that inflicted more demolitions would spend a larger proportion of the game on offense as the mechanism of attack allows for the changing of ball possession. From the following visualizations, it is inconclusive as to whether that is the case in actuality.



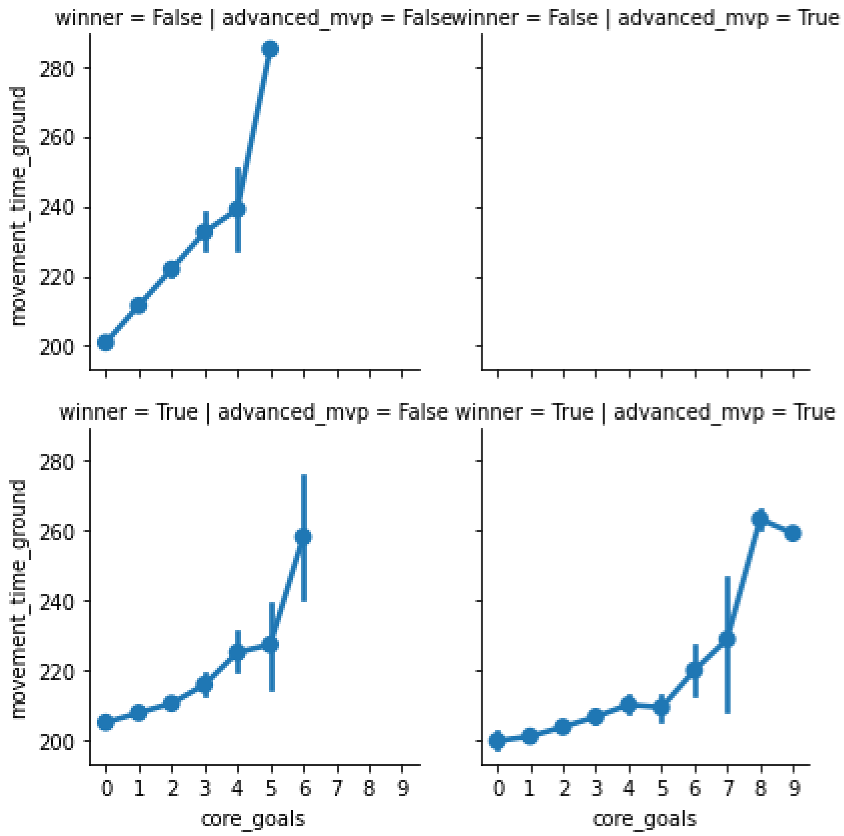
David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang



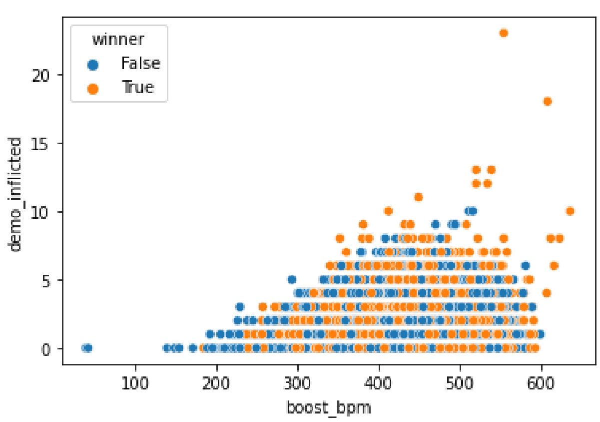

Kendrew and Nicholas made a barplot showing that winning MVP players tend to score and assist goals more than regular players. They both also made a pointplot comparing MVP players and regular players, and seeing the relationship between movement time spent on the ground and number of goals. Those who are winning players, regardless of MVP status tend to stay on the ground less when scoring early goals. MVP players, in fact, score early goals on the ground much less that regular winning players, meaning they score more often in the air, early in the game.

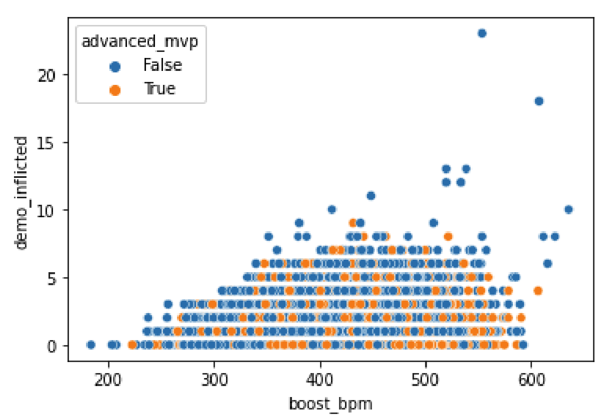


David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang



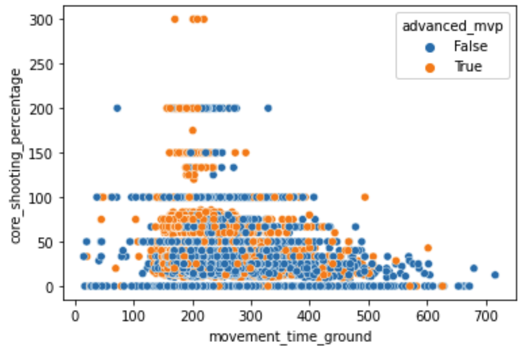

Kendrew and Nicholas made 2 scatterplots comparing the relationship between boost collected per minute and demolition of cars inflicted. What they found was that regardless of losing, winning, or gaining mvp status, the relationship between boost collected and demolition inflicted is the same.

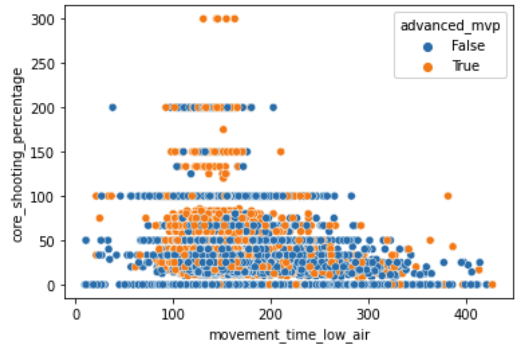


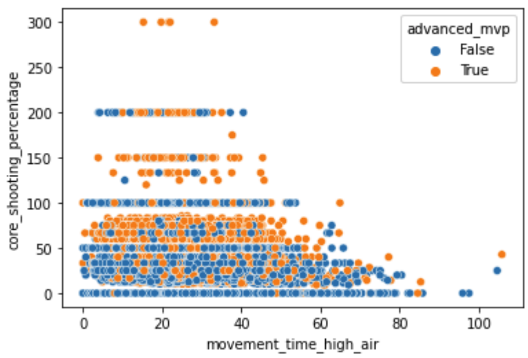



David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

Kendrew and Nicholas made 3 scatterplots showcasing relationships between movement time spent on ground vs. core shooting percentage, movement time spent on low air vs. core shooting percentage, and movement time spent on high air vs. core shooting percentage. What they found is that MVPs with a greater shooting percentage tend to spend more time high in the air, which is most likely where they take their shots from.



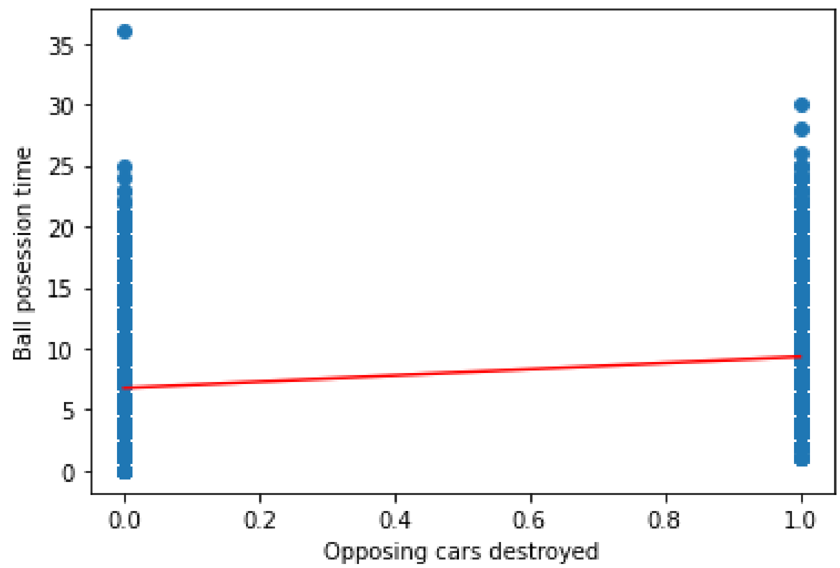



David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

**Linear Regression Analysis**

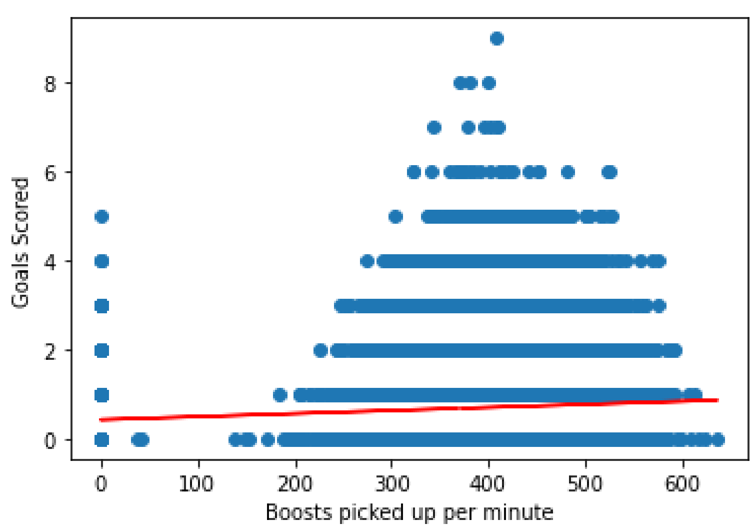
David and Jonathan made a Linear Regression model for predicting possession time based on how often destructions are inflicted (Does destroying opposing cars lead to more ball possession?)

We discovered that destruction infliction also did not work very well to predict possession time, despite the fact that destroying an enemy's car in thought does mean a greater chance for your team to be in control of the ball.




David and Jonathan made a Linear regression for predicting goals scored based on boost pickups per minute(Does picking up more boosts result in more goals being scored?)

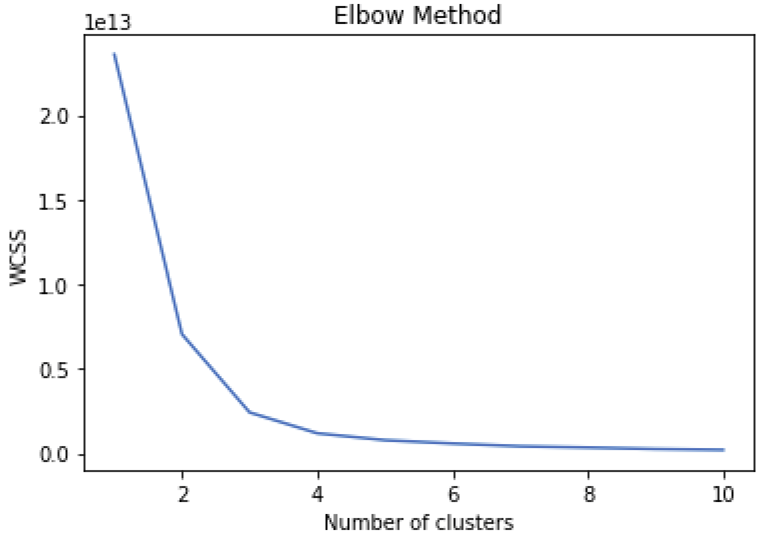
We discovered that boost pickups per minute were not a good predictor of goals, despite the fact that boosts are very important in the game as they allow quicker movement which is harder to defend against.



David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

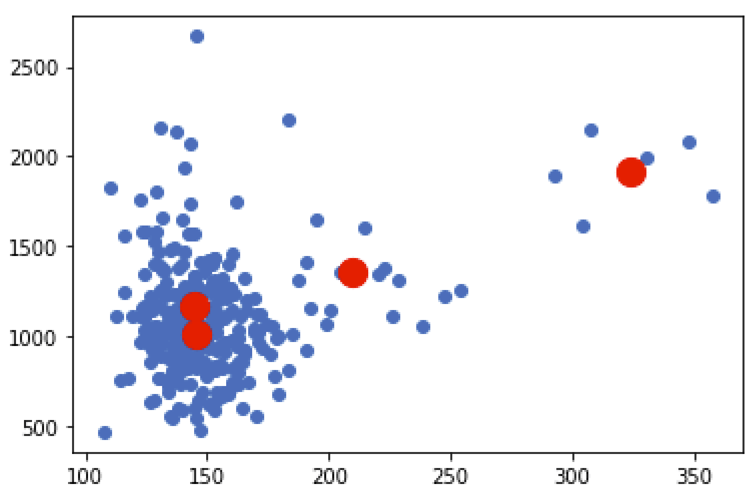
**K-Means Clustering**

Nicholas made a K-Means Clustering model for the team's dataset, with a focus on ball possession time and core score. First K-Means used to find clustering for ball possession time and the core score. Then we find the optimal number of clusters using the elbow method. We picked the number at the elbow which is 4 in this case.



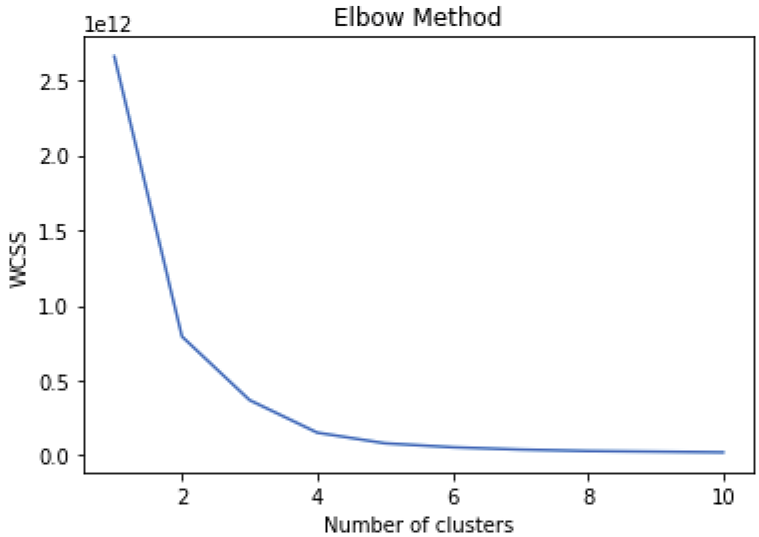

Using the optimal number of clusters we found we plug them. The points put into the graph which are called centroids We calculate the distance of each point to the centroid and find the average Those steps are repeated until clusters are found.

The groups are formed where the clusters are. This means that there are clusters of data when looking at a ball possession time and core score graph.

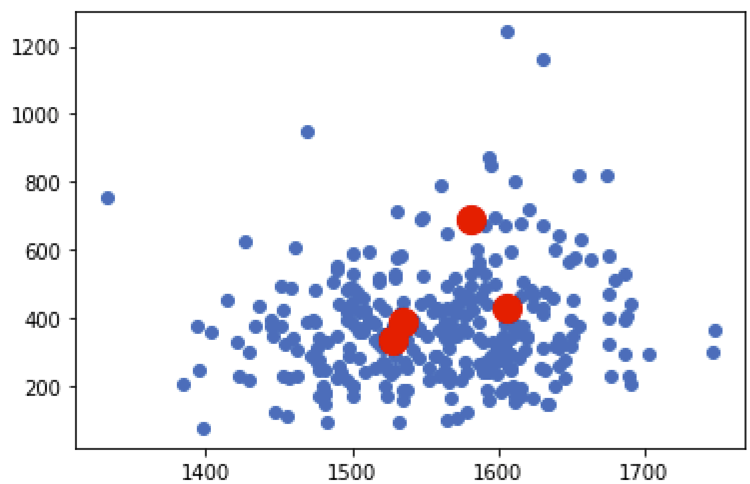



The same is done for the other one for average movement speed and core score. We find the optimal number of clusters using the elbow method which is 4.

David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang




And we find the groups using the clusters.



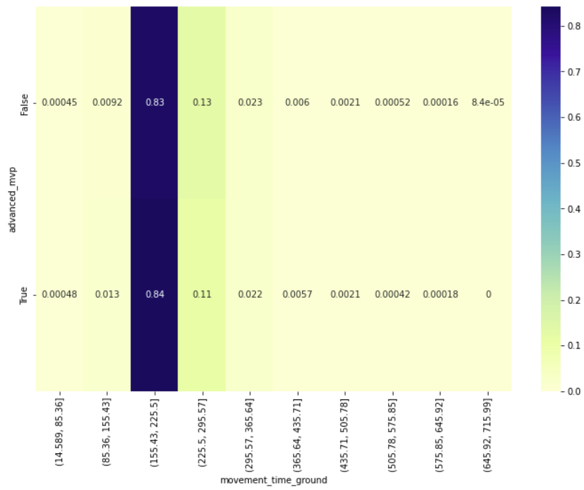

**Chi Squared Test [Correlation Testing]**

Kendrew produced a contingency table and ran a Chi-squared test for movement time_ground, advanced mvp, core_shooting_percentage, core_goals, steering_sensitivity.

David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

| movement_time_ground | (14.589, 85.36] | (85.36, 155.43] | (155.43, 225.5] | (225.5, 295.57] | (295.57, 365.64] | (365.64, 435.71] | (435.71, 505.78] | (505.78, 575.85] | (575.85, 645.92] | (645.92, 715.99] |
|---|---|---|---|---|---|---|---|---|---|---|
| advanced_mvp | | | | | | | | | | |
| **False** | 37 | 763 | 68944 | 10632 | 1951 | 500 | 175 | 43 | 13 | 7 |
| **True** | 8 | 221 | 13998 | 1884 | 364 | 94 | 35 | 7 | 3 | 0 |

P-Value:  2.4394338384981347e-08




Given the results of our chi square test, we do not reject our null hypothesis that a player's individual MVP status and time spent on the ground are independent of one another. Given how low our p-value is, we would still maintain our null hypothesis even at a .001% confidence level.

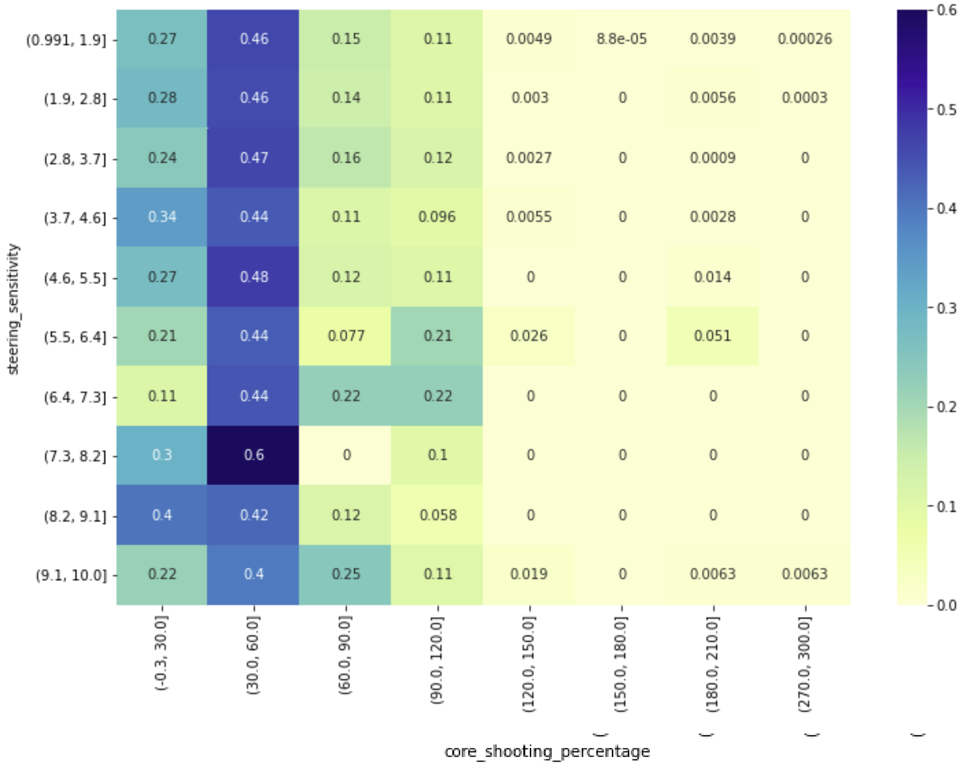
| core_shooting_percentage | (-0.3, 30.0] | (30.0, 60.0] | (60.0, 90.0] | (90.0, 120.0] | (120.0, 150.0] | (150.0, 180.0] | (180.0, 210.0] | (270.0, 300.0] |
|---|---|---|---|---|---|---|---|---|
| steering_sensitivity | | | | | | | | |
| **(0.991, 1.9]** | 3057 | 5273 | 1725 | 1242 | 56 | 1 | 44 | 3 |
| **(1.9, 2.8]** | 955 | 1538 | 486 | 371 | 10 | 0 | 19 | 1 |
| **(2.8, 3.7]** | 272 | 523 | 184 | 133 | 3 | 0 | 1 | 0 |
| **(3.7, 4.6]** | 125 | 159 | 41 | 35 | 2 | 0 | 1 | 0 |
| **(4.6, 5.5]** | 20 | 35 | 9 | 8 | 0 | 0 | 1 | 0 |
| **(5.5, 6.4]** | 8 | 17 | 3 | 8 | 1 | 0 | 2 | 0 |
| **(6.4, 7.3]** | 1 | 4 | 2 | 2 | 0 | 0 | 0 | 0 |
| **(7.3, 8.2]** | 3 | 6 | 0 | 1 | 0 | 0 | 0 | 0 |
| **(8.2, 9.1]** | 21 | 22 | 6 | 3 | 0 | 0 | 0 | 0 |
| **(9.1, 10.0]** | 35 | 63 | 39 | 17 | 3 | 0 | 1 | 1 |

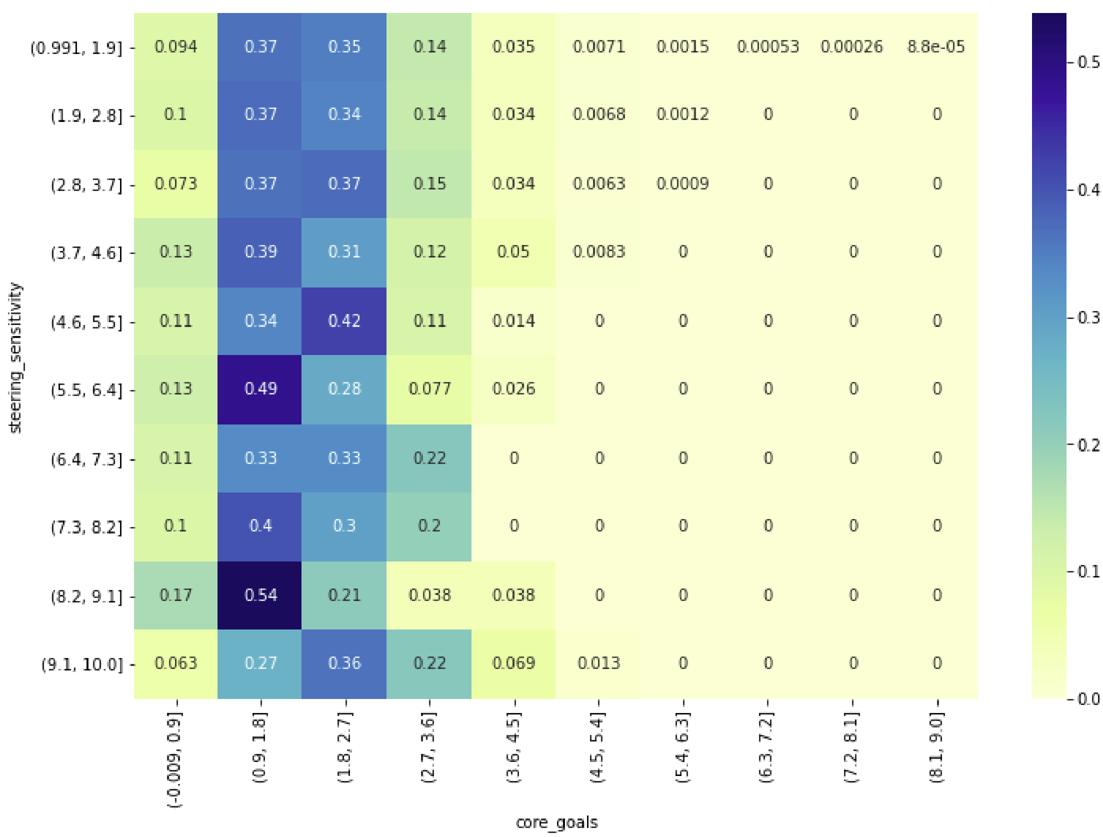
[Player Steering Sensitivity Setting to Shooting Percentage] P-Value:  0.000178140588134638

| core_goals | (-0.009, 0.9] | (0.9, 1.8] | (1.8, 2.7] | (2.7, 3.6] | (3.6, 4.5] | (4.5, 5.4] | (5.4, 6.3] | (6.3, 7.2] | (7.2, 8.1] | (8.1, 9.0] |
|---|---|---|---|---|---|---|---|---|---|---|
| steering_sensitivity | | | | | | | | | | |
| **(0.991, 1.9]** | 1071 | 4195 | 3985 | 1641 | 401 | 81 | 17 | 6 | 3 | 1 |
| **(1.9, 2.8]** | 339 | 1263 | 1160 | 476 | 115 | 23 | 4 | 0 | 0 | 0 |
| **(2.8, 3.7]** | 81 | 409 | 416 | 164 | 38 | 7 | 1 | 0 | 0 | 0 |
| **(3.7, 4.6]** | 49 | 140 | 111 | 42 | 18 | 3 | 0 | 0 | 0 | 0 |
| **(4.6, 5.5]** | 8 | 25 | 31 | 8 | 1 | 0 | 0 | 0 | 0 | 0 |
| **(5.5, 6.4]** | 5 | 19 | 11 | 3 | 1 | 0 | 0 | 0 | 0 | 0 |
| **(6.4, 7.3]** | 1 | 3 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| **(7.3, 8.2]** | 1 | 4 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| **(8.2, 9.1]** | 9 | 28 | 11 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| **(9.1, 10.0]** | 10 | 43 | 58 | 35 | 11 | 2 | 0 | 0 | 0 | 0 |

[Player Steering Sensitivity Setting to Goals Scored] P-Value:  0.7884866377610897

David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang



David Ryan
Kendrew Christanto
Jonathan Thai
Nicholas Chang

We can definitively conclude that there is no correlation between steering sensitivity and an individual player's shooting percentage given our computed p-value using the chi squared test.

Interestingly, there did appear to be a correlation between steering sensitivity and the pure number of goals scored for MVP players. Our computer p-value for this relation was 0.79 meaning that we can maintain the null hypothesis at a 95% confidence level but reject it at a 75% confidence level. We can conclude there is a decent probability that the number of goals scored for MVP players is a response function of their steering sensitivity.

**Contributions**:
We all contributed equally throughout each part of the project, but our specific contributions are listed above.