CS210 Project

My Valorant Analysis Report

Bülay Ecem Çakallı
30788

Summary:

Valorant is a team based (5v5) first-person tactical hero shooter game whoever wins the 13 rounds first, also wins the game. Each player chooses and agent before the game starts, and each agent has its' own special abilities. In my project, I'll be focusing on whether my received damage rate is higher than my dealt damage as the in-game playtime increases. With that I'll be able to find whether I was a burden to team or was a player who carried the team.

Motivation:

I used to play Valorant between the years 2020-2021 but for some reason I had to stop playing. However, nowadays I was wondering whether I should start again. But I did not want to play it alone this time, so asked some of my friends to play with me and told them that I would carry them in the game. Then, I questioned myself "Can I really carry them?". So, to find an answer I decided to go with that idea in my project.

Data Source:

First, I'd like to explain my dataset a bit. I only used competetive matches in my dataset. Unlike, unranked matches, competetive matches make it possible to match players with similar skill level. Therefore, it is possible to get more consisent data. Secondly, the dataset only contains the years between 2020-2021. The reason for that is because I stopped playing competetive matches specifically after 2021-10-15(can be seen in the data frame below) for various reasons. In other words, the data you see here contains all my competetive matches.

It is hard to get the Riot API, therefore I get the data from a website which tracks all my Valorant stats. The name of that website is tracker gg and I get all those stats in JSON format (It can be found inside Data Scraping folder's JSONS folder). Then I convert the JSONs into a CSV file by using CS210_MakeCSV.py (CSVs can be found inside Data Scraping folder's csv files folder) and then I combined the CSV files by using CS210_project_combineCSVs.py and in the end a combined CSV file named CS210_project_combineCSVs.csv was created.

Data Analysis:

My motivation was to find whether I carried the team or not, and to find that first I decided to go with Total Damage/Total Received Damage. However, some matches are more competitive than the other, meaning those matches would take more time to finish. Therefore, I decided to add playtime, and formed this hypothesis:

As my playtime increases, a corresponding rise in my damage is higher than the damage I receive.

If that hypothesis was correct that would mean overall, I carried my team in my matches. However, I had to change the hypothesis because it was the opposite…
Therefore, my hypothesis now is:

As my playtime increases, a corresponding rise in my damage is lower than the damage I receive.

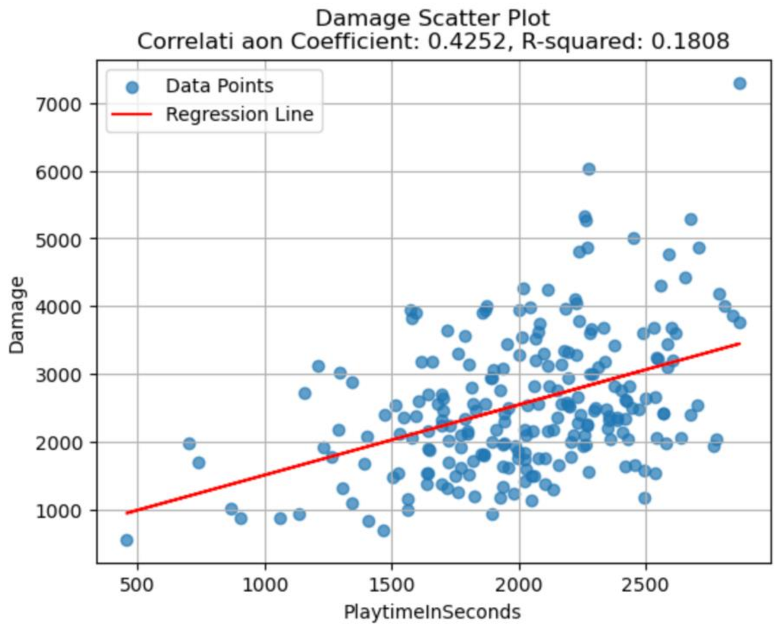
However, I had to obtain damage and received damage's relation with time separately then compare damage-playtime and received damage-playtime relations. Therefore, for hypothesis testing my hypothesis were:

For damage:
As my playtime increases, there is a corresponding rise in my damage.
For received damage:
As my playtime increases, there is a corresponding rise in my received damage.

I created a Data Frame from the CS210_project_combineCSVs. Then, I had to add a new column named PlaytimeInSeconds because the format of the playtime was '34m28s' like this. In addition, I converted the Damage and Received Damage into numeric. After that, I created a Scatter Plot.

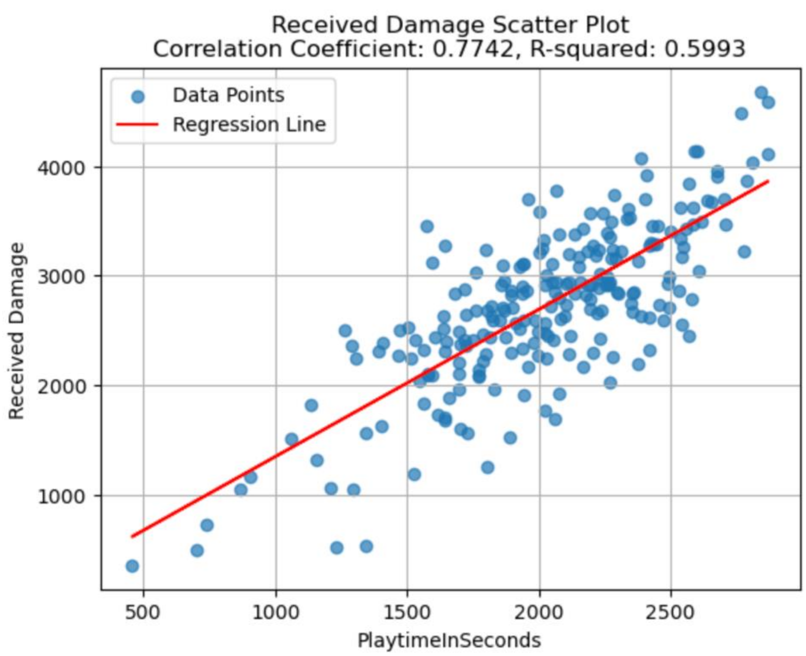
The Damage-Playtime and Received Damage-Playtime Scatter Plot

The purpose of this is to show whether damage and received damage had a positive correlation with the playtime. In the end, both damage and received damage did, since their correlation coefficient was bigger than zero. However, damage (it's R value was 0.18) had a weak relation with the playtime, whereas received damage (it's R value was 0.59) had a strong relation with the playtime. That meant I received more damage than I dealt.




```
Correlation Coefficient: 0.4252
R-squared: 0.1808
```

The correlation coefficient of 0.4252 indicates a moderate positive correlation between playtime and damage. However, the R-squared value of 0.1808 suggests that only approximately 18.08% of the variance in damage can be explained by playtime. In the end, there is a weak positive correlation. Moreover, Normal Distribution is seen.

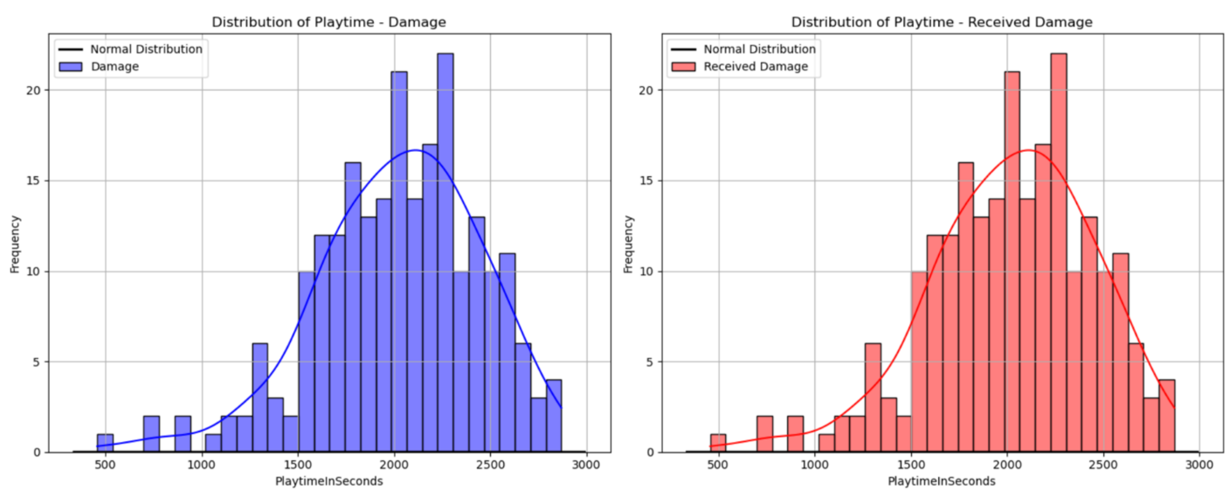



```
Correlation Coefficient: 0.7742
R-squared: 0.5993
```

A correlation coefficient of 0.7742 indicates a strong positive relationship between the playtime and received damage. The R-squared value of 0.5993 suggests that approximately 60% of the variability in the dependent received damage can be explained by changes in the independent playtime. In the end, there is a strong positive correlation. Moreover, Normal Distribution is seen.

The Normal Distribution Histogram

The purpose is to find whether my damage and received damage varies across different levels of playtime. In addition, there was a normal distribution.



.

Pearson Testing

Hypothesis for damage:
As my playtime increases, there is a corresponding rise in my damage.
Since my values were linear and normally distributed, I decided to use Pearson Testing. For damage, correlation coefficient is 0.42522309438843275. That means as playtime increase the

damage increase as well. In other words, damage has a weak positive correlation with playtime. The P-Value is 1.8009440673632214e-11, meaning the correlation was not happened randomly and the conventional significance is less than 0.05 which means there is strong evidence against my null hypothesis (There is no significant correlation between playtime and damage)

For received damage my hypothesis is:
As my playtime increases, there is a corresponding rise in my received damage.

Correlation coefficient is 0.7741574956181136. That means as playtime increase the damage increase as well. In other words, received damage has a strong positive correlation with playtime. The P-Value is 5.635296636492636e-47, meaning the correlation was not happened randomly. And the conventional significance is less than 0.05 which means there is strong evidence against my null hypothesis (There is no significant correlation between playtime and received damage)

Factor Analysis

There were some factors that I believed had correlation with my hypothesis. Factors:

1) Agent
2) Map
3) Rank

And I analyzed each one whether they were supporting (or not) or were they the causation for my hypothesis (or not).

1) Agent Analysis

Agents are the characters you play with in Valorant. Each of them has unique abilities,and those abilities appoints the agent's role. These are the roles:

Controllers: These agents are the agents that have the abilitiy to use smokes. With their smokes, agents are able to block off certain areas of the map to make offensive or defense easier.

Agents: Brimstone, Viper, Omen, Astra

Duelists: These agents excel in one-vs-one fights and their goal is win their duels. These are offensive agents.

Agents: Phoenix, Jett, Reyna, Raze, Yoru, Neon

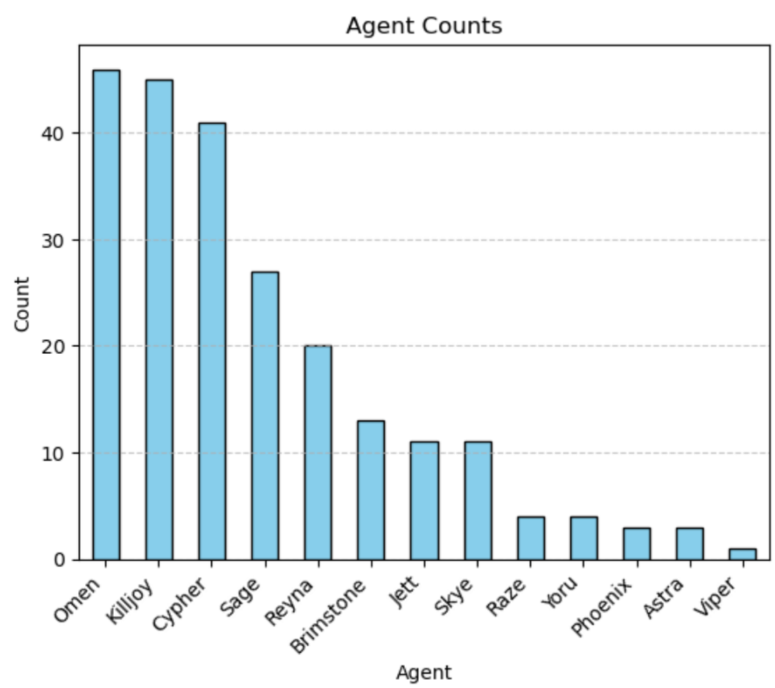
Initiators: These agents are the supportive class of agents in Valorant who have abilities that allow them to provide information to the team.

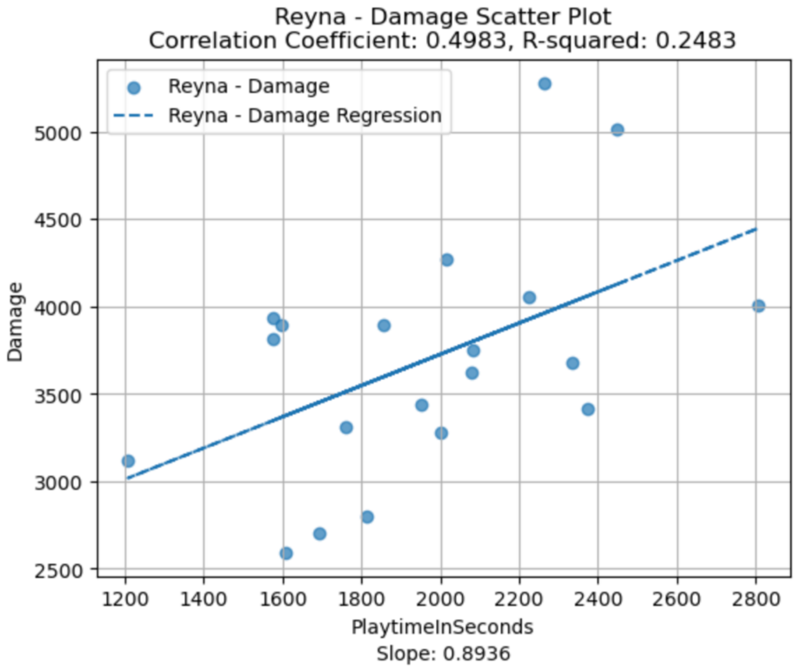
Agents: Sova, Breach, Skye

Sentinels: These agents are the defensive class of agents and most of the time they slow down or sometimes outright stop enemy pushes on certain areas of the map.

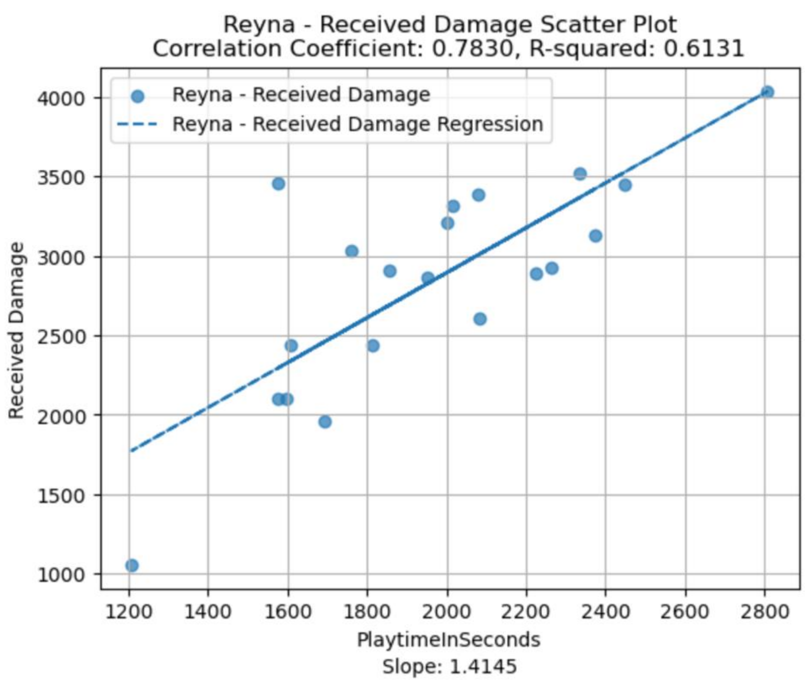
Agents: Killjoy, Cypher, Sage, Chamber

The expected damage from a duelist is higher than a sentinel. However, it still depends on the player's skills as well. I looked at each agent's damage and received damage as playtime gets a higher value, and to obtain a more consistent data set, I got rid of the agents which was not chosen more than ten matches. I showed the relations in a scatter plot. Then, showed each agent's slope value on a bar chart. That way I get to compare received damage-playtime and damage-playtime. From 8 agents 5 of them had a higher received damage-playtime. Therefore, my hypothesis was supported, and agent was one of the causations for my hypothesis.

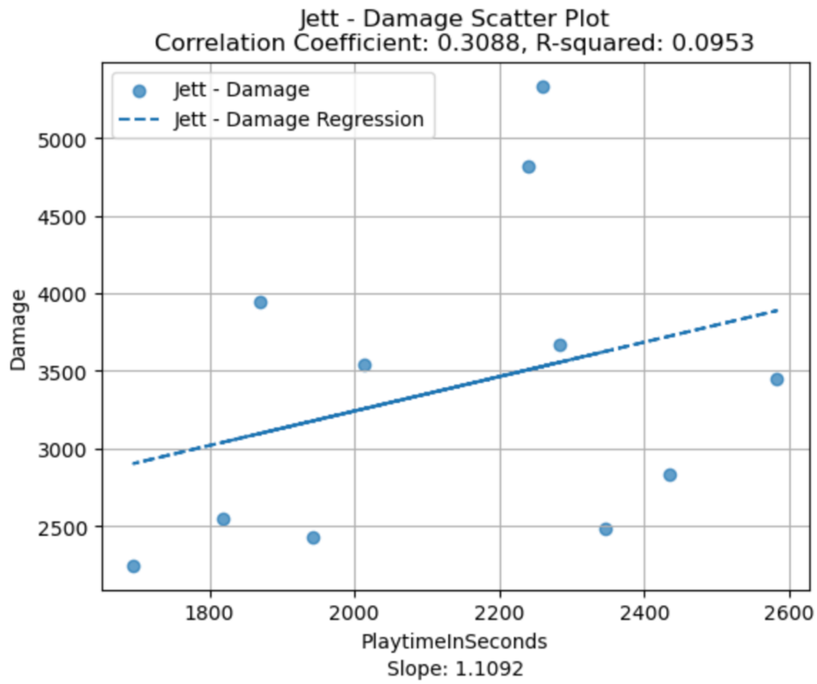




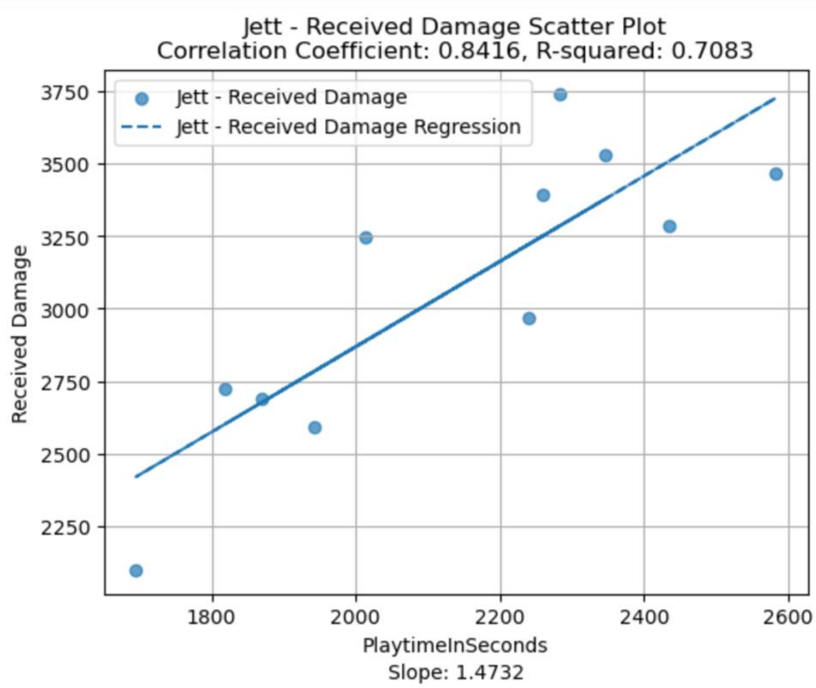

```
Reyna - Damage:
Correlation Coefficient: 0.4983
R-squared: 0.2483
```



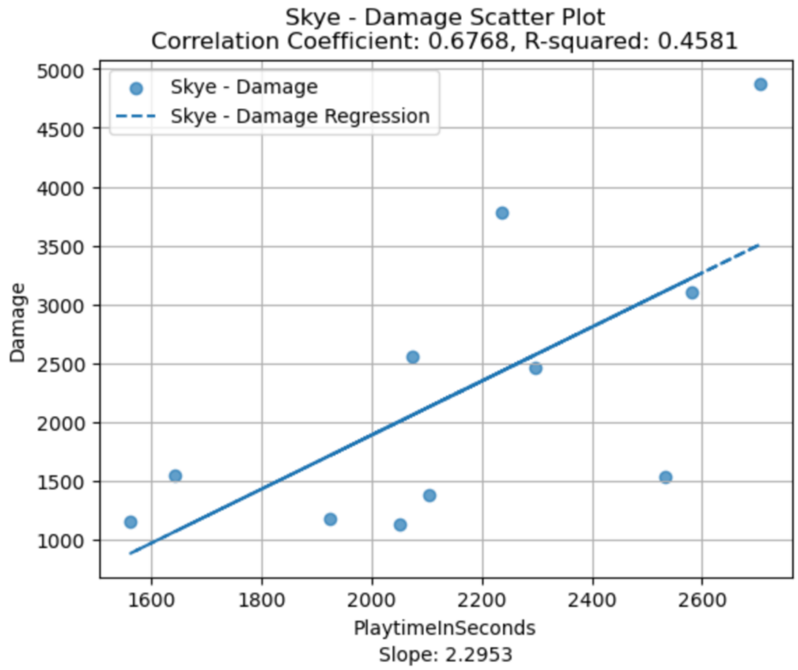

```
Reyna - Received Damage:
Correlation Coefficient: 0.7830
R-squared: 0.6131
```



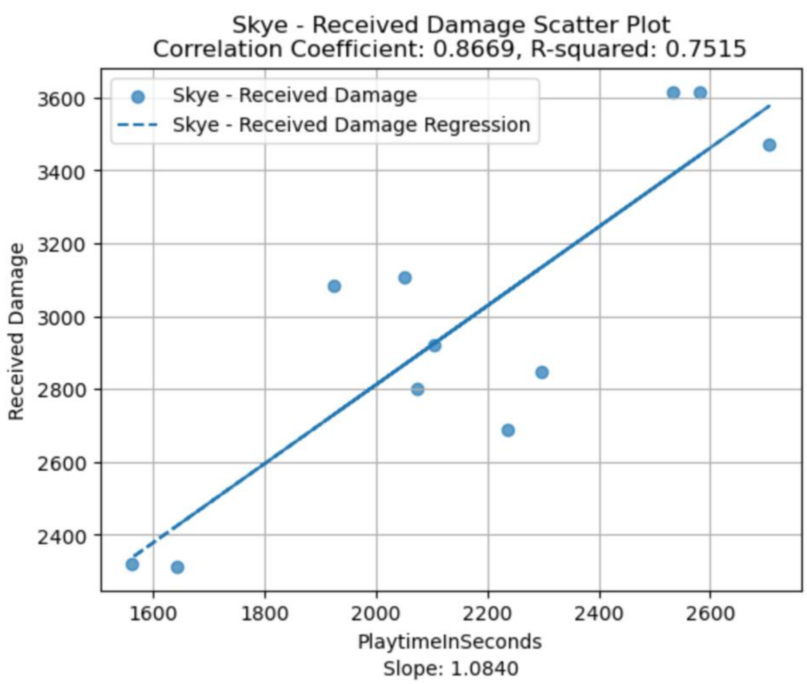

```
Jett - Damage:
Correlation Coefficient: 0.3088
R-squared: 0.0953
```



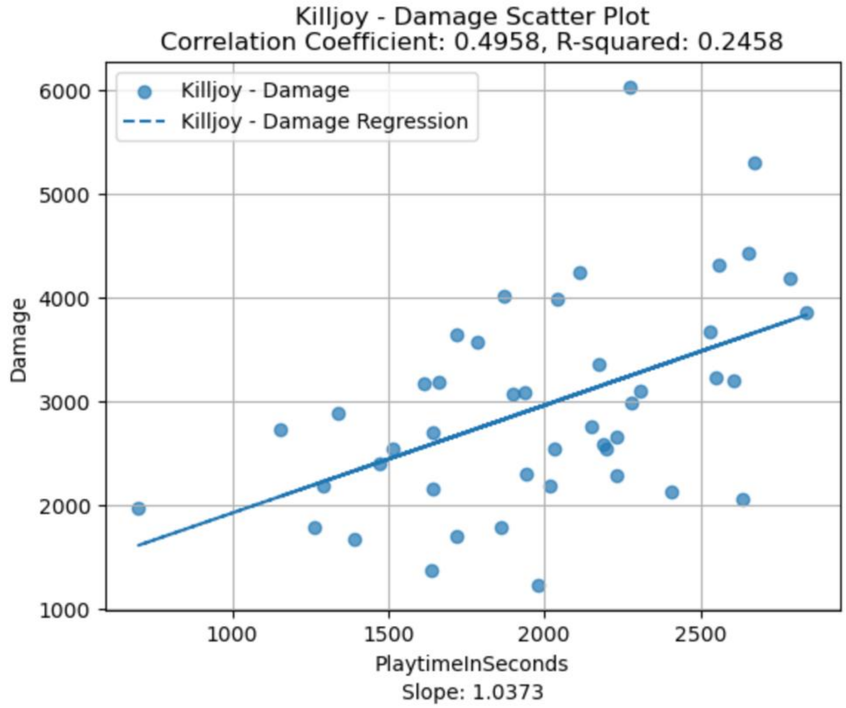

```
Jett - Received Damage:
Correlation Coefficient: 0.8416
R-squared: 0.7083
```



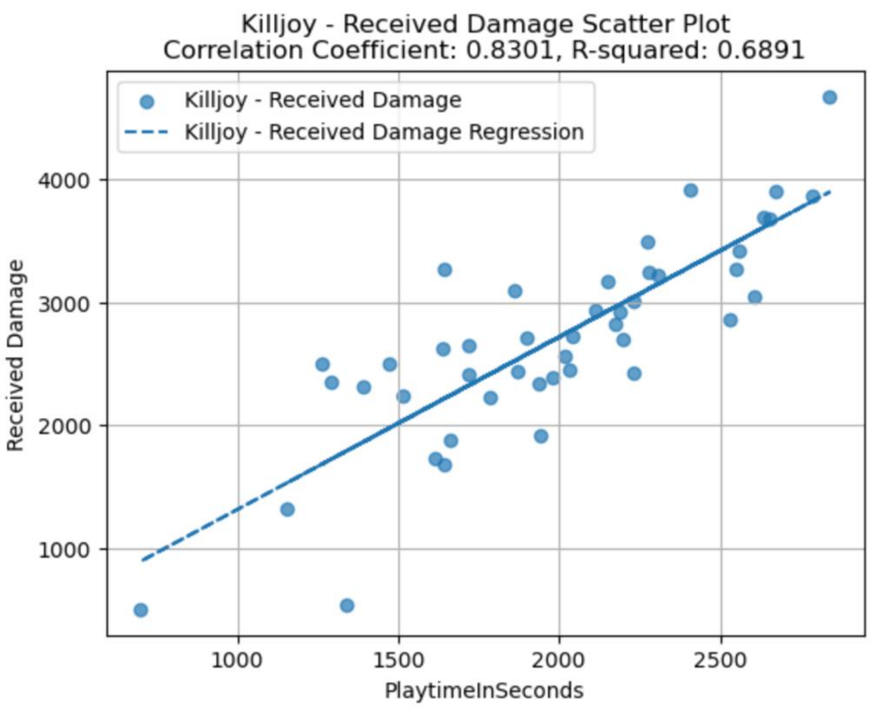
Skye - Damage:
Correlation Coefficient: 0.6768
R-squared: 0.4581



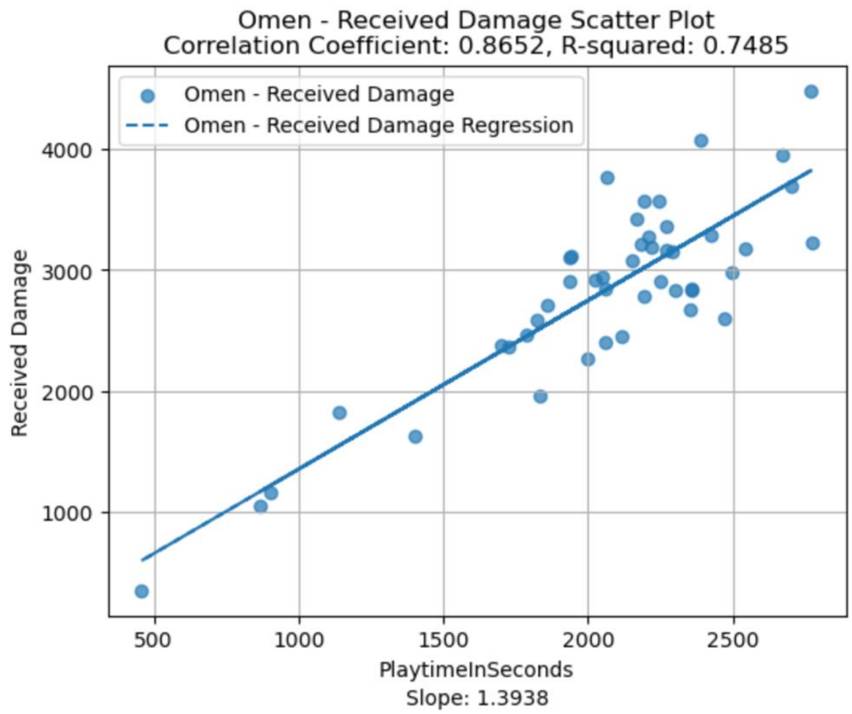
Skye - Received Damage:
Correlation Coefficient: 0.8669
R-squared: 0.7515



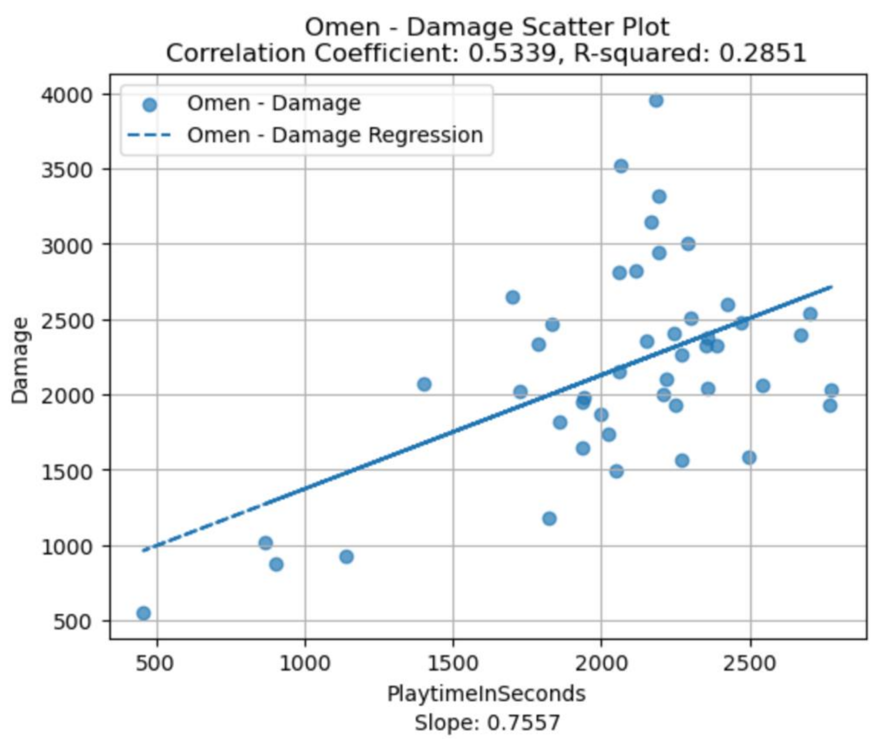
Killjoy – Damage:
Correlation Coefficient: 0.4958
R–squared: 0.2458



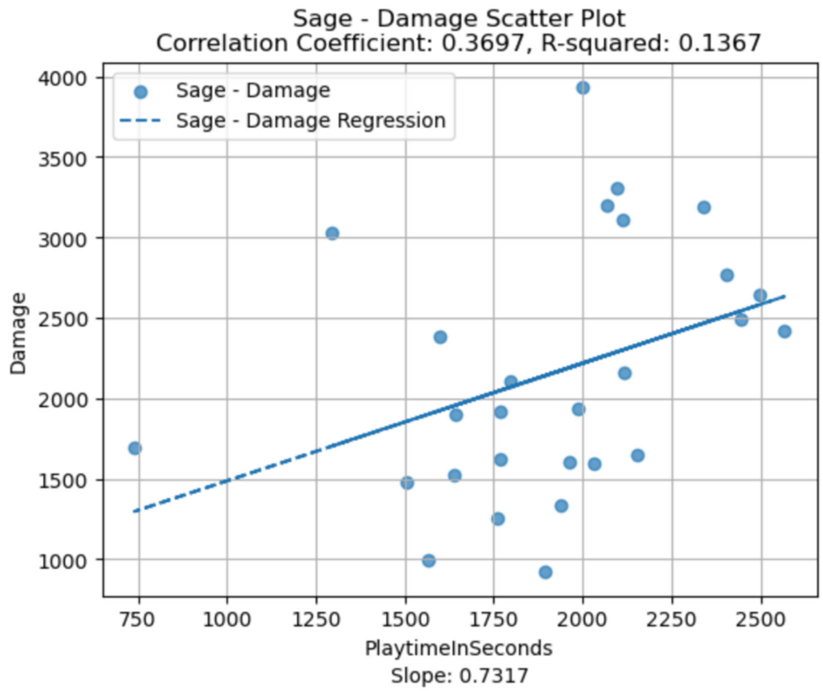
Killjoy – Received Damage:
Correlation Coefficient: 0.8301
R–squared: 0.6891



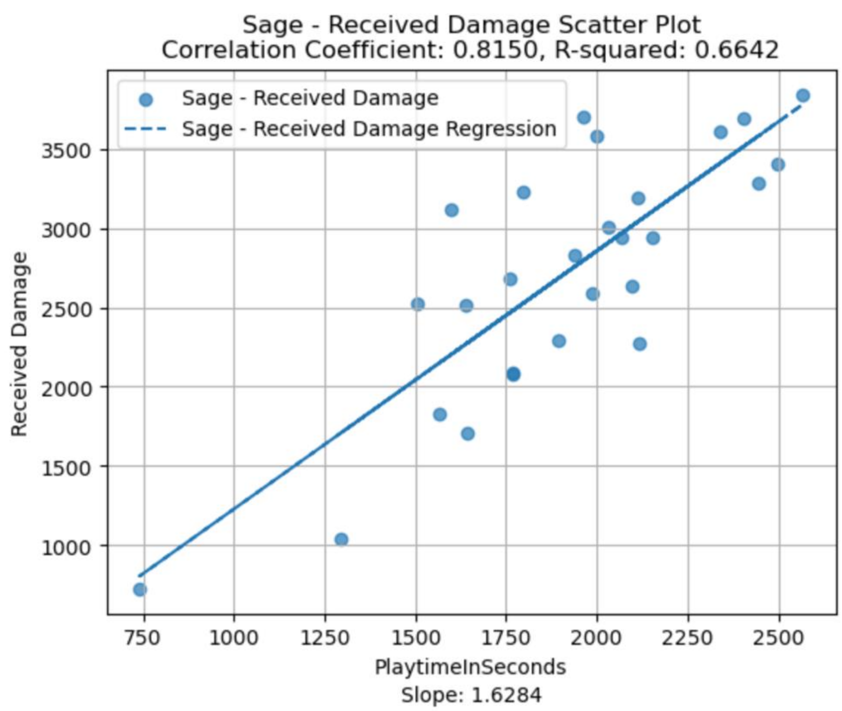

```
Omen - Received Damage:
Correlation Coefficient: 0.8652
R-squared: 0.7485
```



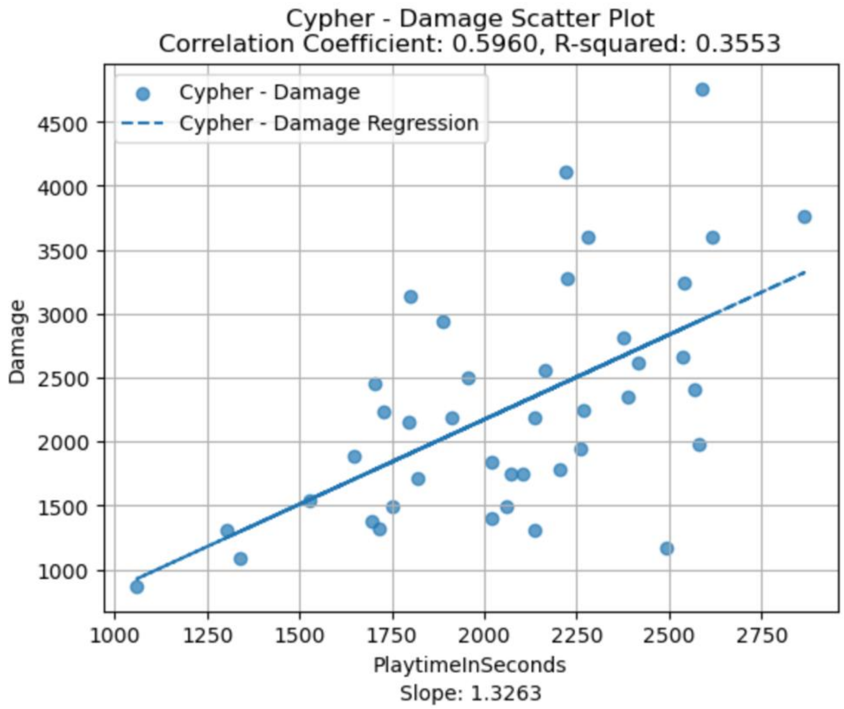

```
Omen - Damage:
Correlation Coefficient: 0.5339
R-squared: 0.2851
```



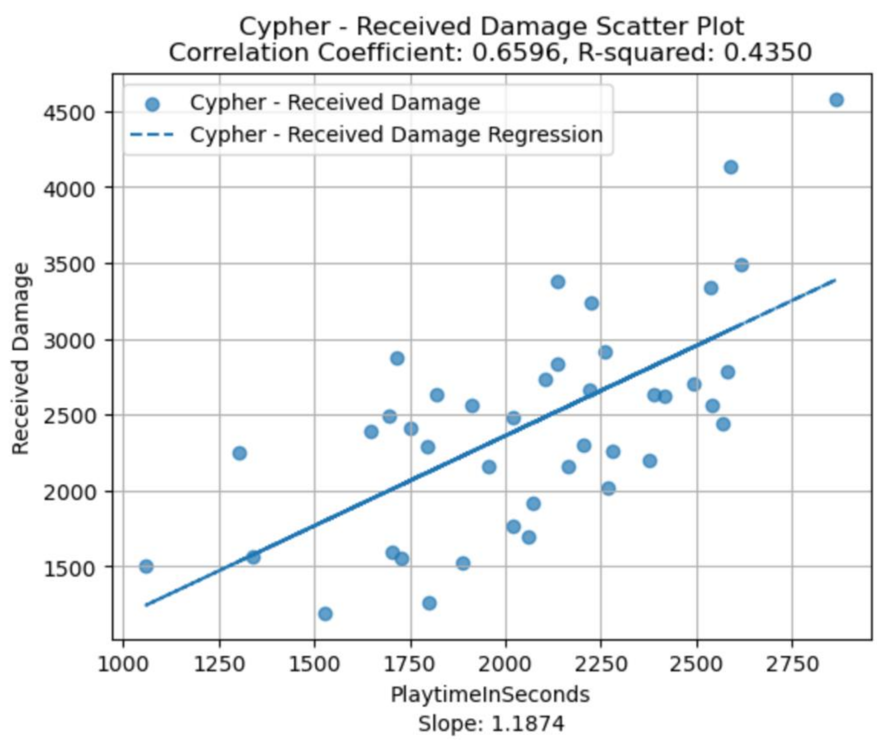

```
Sage - Damage:
Correlation Coefficient: 0.3697
R-squared: 0.1367
```



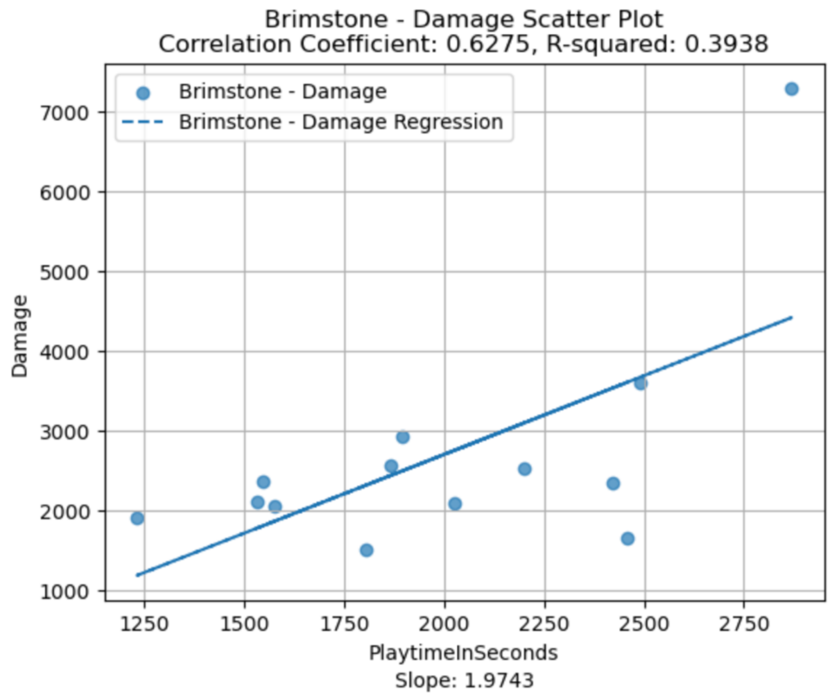

```
Sage - Received Damage:
Correlation Coefficient: 0.8150
R-squared: 0.6642
```



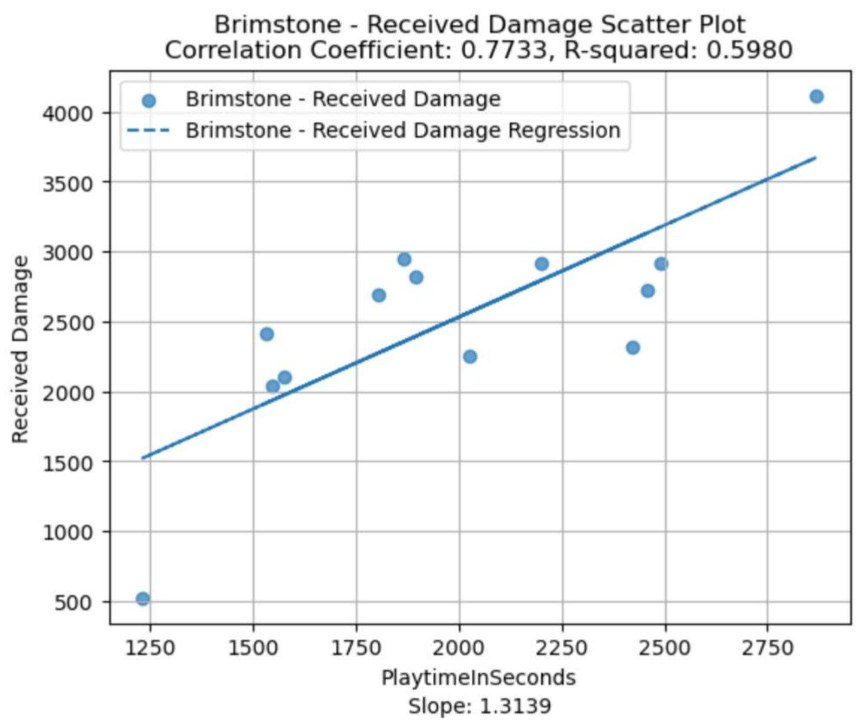

```
Cypher - Damage:
Correlation Coefficient: 0.5960
R-squared: 0.3553
```



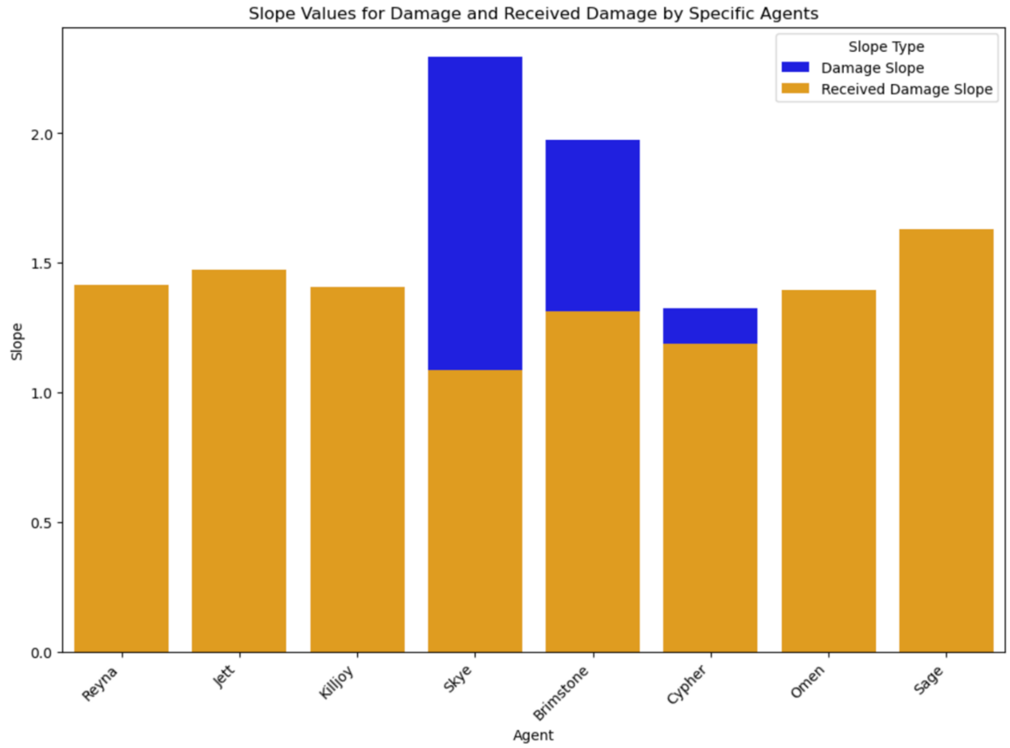

```
Cypher - Received Damage:
Correlation Coefficient: 0.6596
R-squared: 0.4350
```




```
Brimstone - Damage:
Correlation Coefficient: 0.6275
R-squared: 0.3938
```

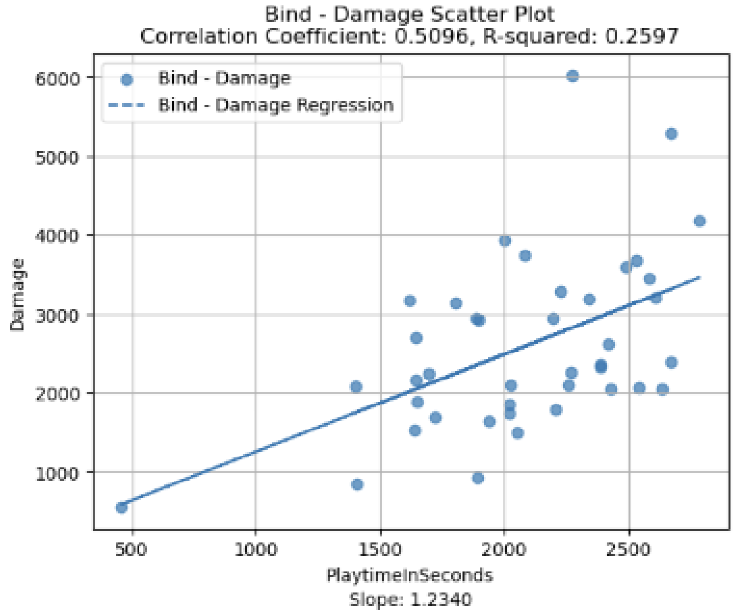



```
Brimstone - Received Damage:
Correlation Coefficient: 0.7733
R-squared: 0.5980
```

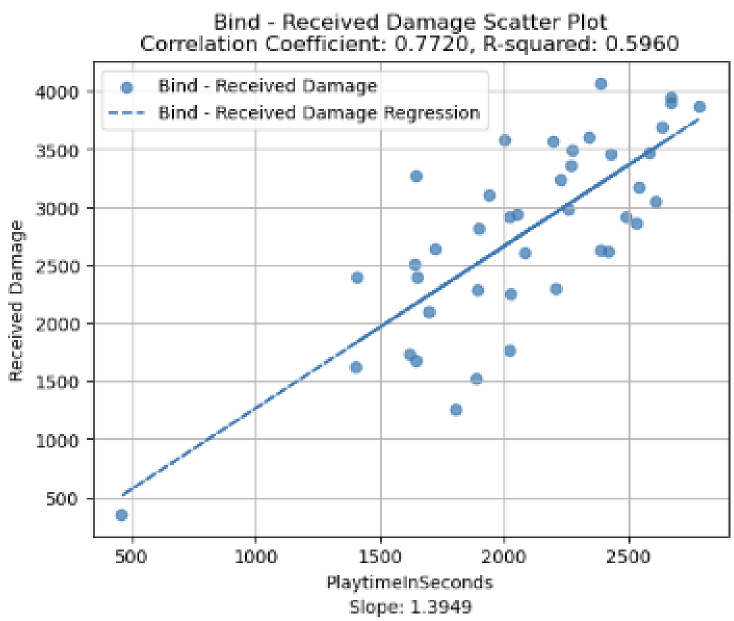


2)Map Analysis

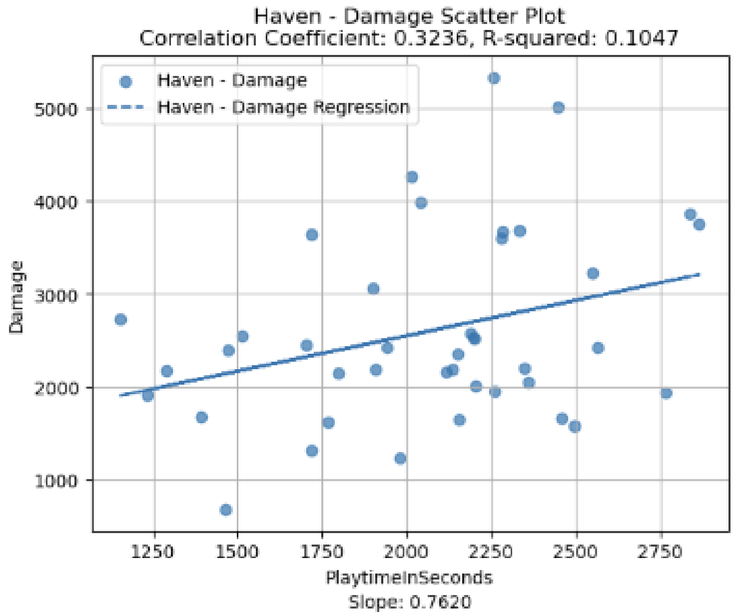
In my dataset, maps have 2 sites (site A and site B) except Haven (it has also site C). These sites are areas where you can plant the spike. Planting the spike is offense team's objective and if the spike explodes than offense team wins the round. In Valorant, each map is different from each other. Therefore, like agents it affects the game. First, I'll excluded some maps that does not exceed 10 matches. Then like I did in agent analysis part; I showed the relations in a scatter plot and each map's slope value on a bar chart. That way I get to compare received damage-playtime and damage-playtime. Overall, the number of maps had stronger received damage-playtime relation than damage-playtime. Therefore, Map is a supporting factor for my hypothesis and one of the causations.



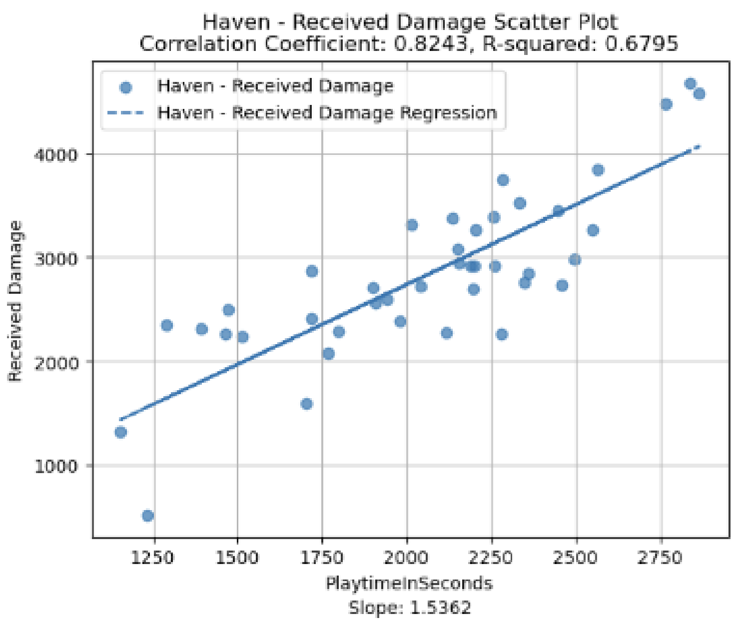

```
Bind - Damage:
Correlation Coefficient: 0.5096
R-squared: 0.2597
```



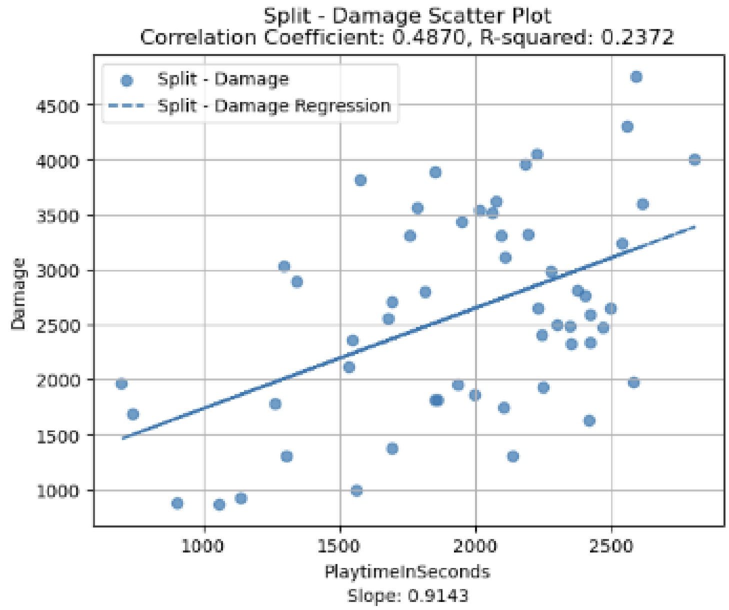

```
Bind - Received Damage:
Correlation Coefficient: 0.7720
R-squared: 0.5960
```



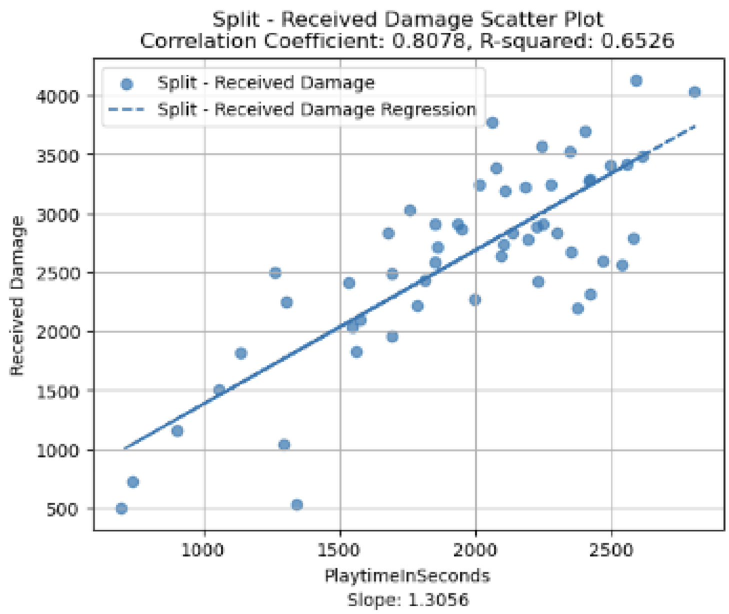

```
Haven - Damage:
Correlation Coefficient: 0.3236
R-squared: 0.1047
```



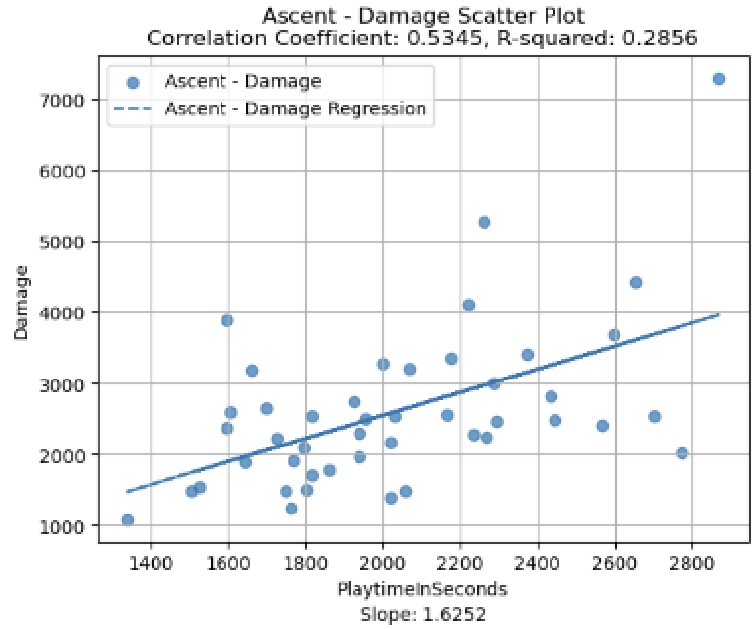

```
Haven - Received Damage:
Correlation Coefficient: 0.8243
R-squared: 0.6795
```



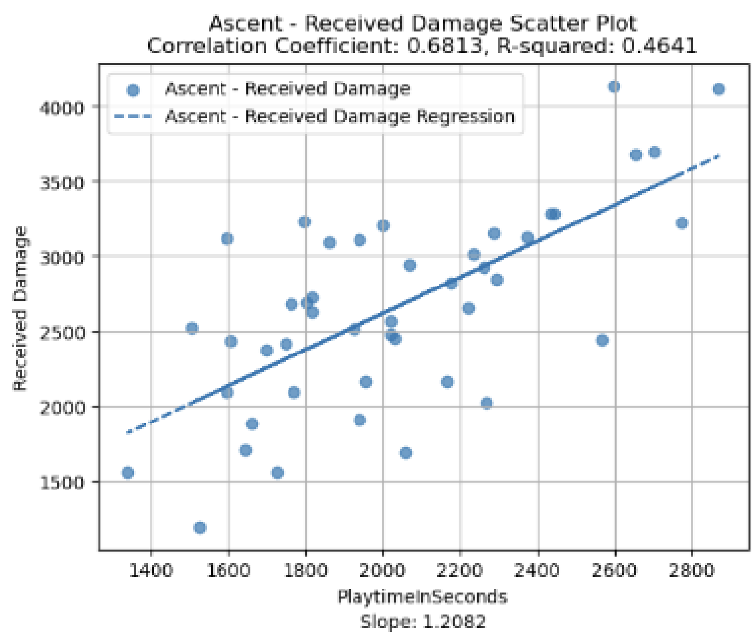

```
Split - Damage:
Correlation Coefficient: 0.4870
R-squared: 0.2372
```



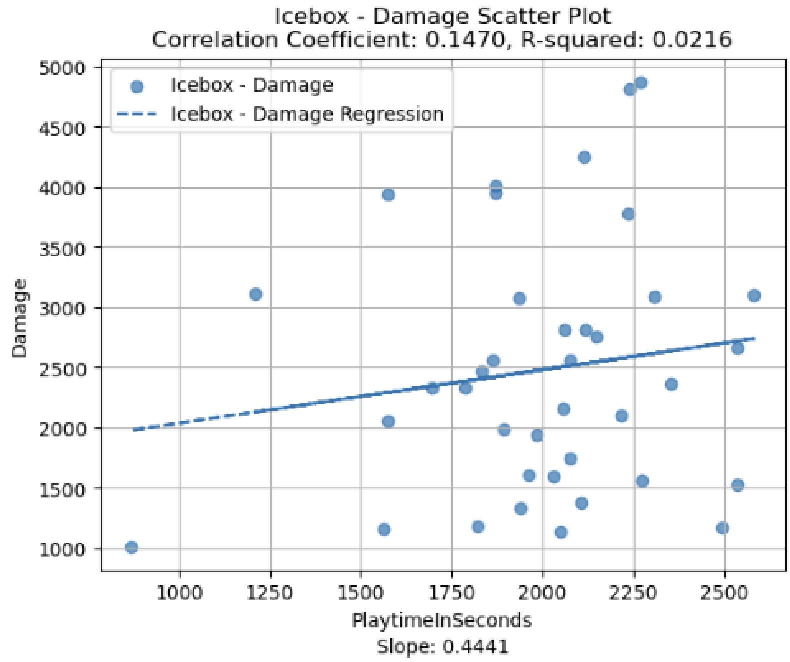

```
Split - Received Damage:
Correlation Coefficient: 0.8078
R-squared: 0.6526
```



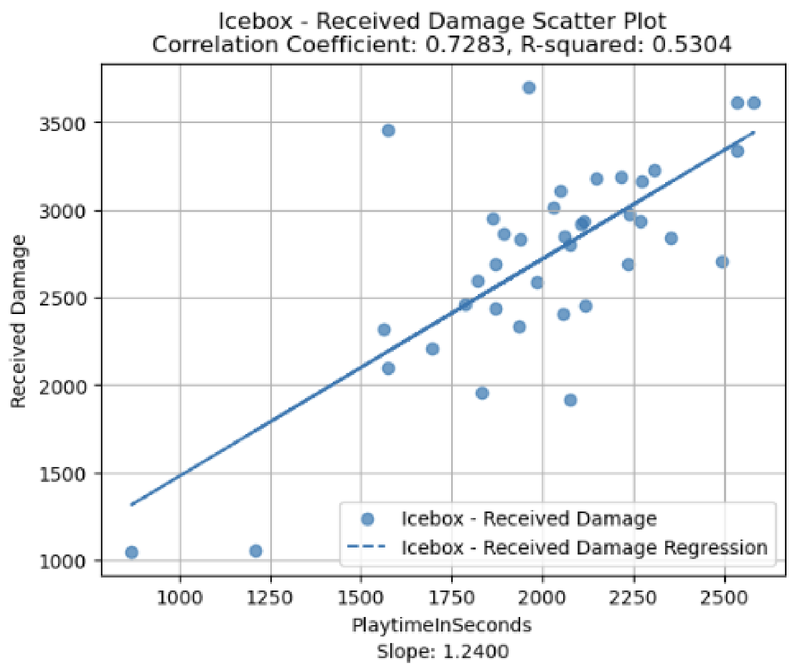

```
Ascent - Damage:
Correlation Coefficient: 0.5345
R-squared: 0.2856
```



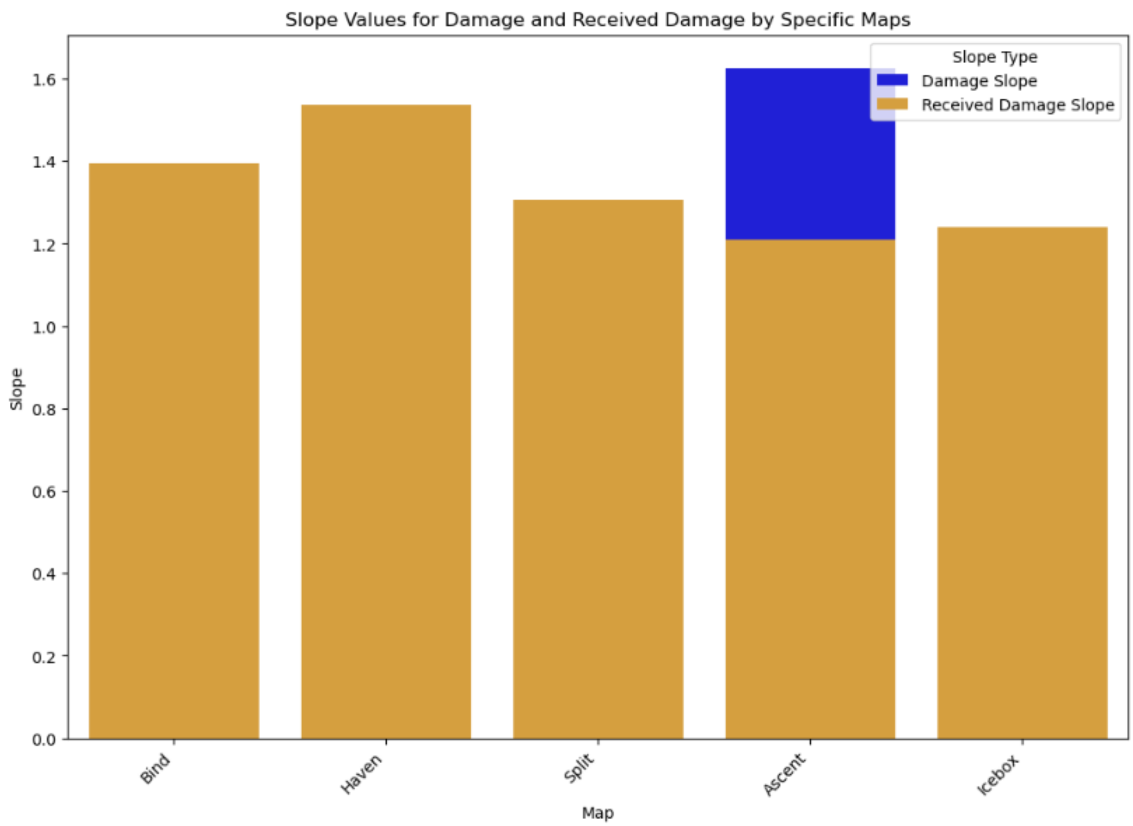

```
Ascent - Received Damage:
Correlation Coefficient: 0.6813
R-squared: 0.4641
```




```
Icebox - Damage:
Correlation Coefficient: 0.1470
R-squared: 0.0216
```




```
Icebox - Received Damage:
Correlation Coefficient: 0.7283
R-squared: 0.5304
```

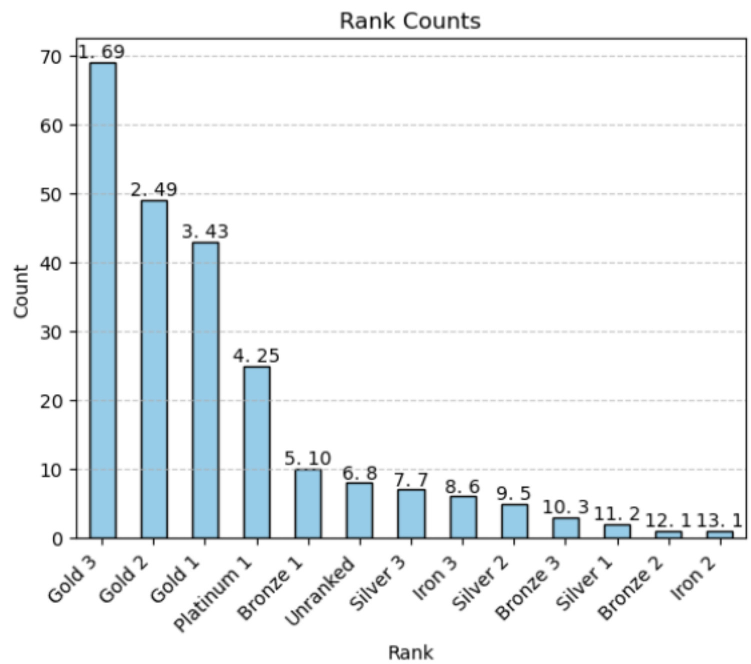


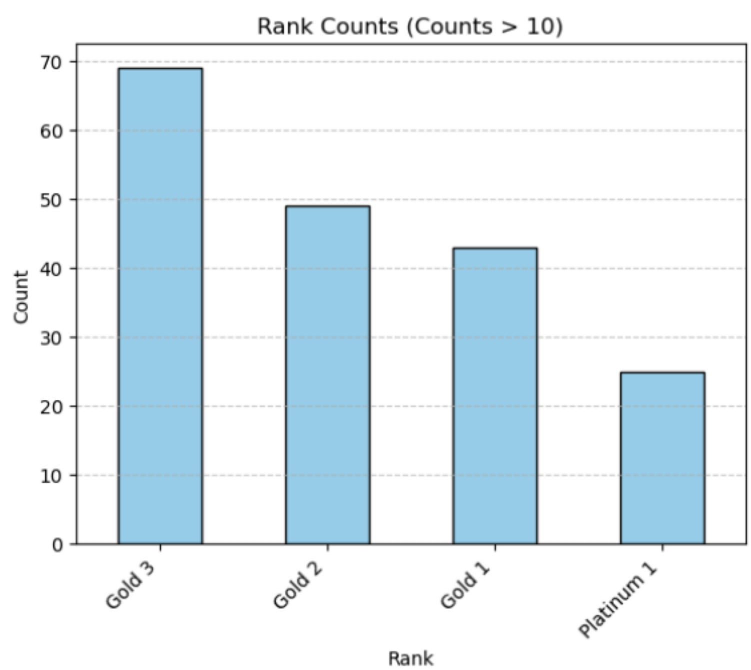
3) Rank Analysis

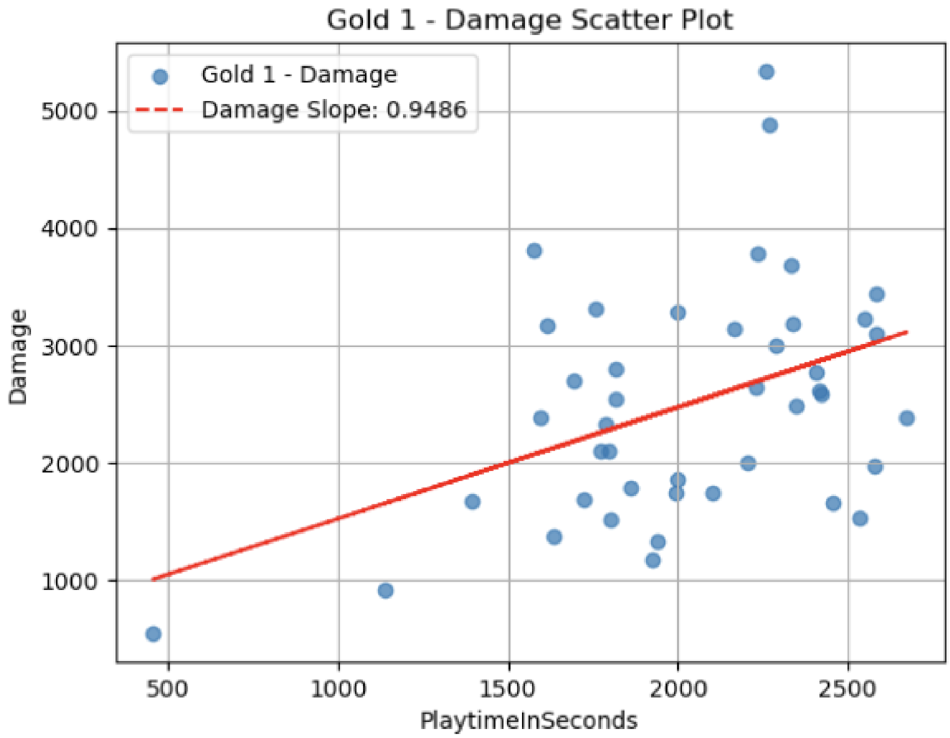
In Valorant, by the result of your comptetive matches a rank is assigned to the player. The more you win the higher your rank is and the more you lose, your rank gets lower. Now I'd like to see whether my received damage-playime slope was higher than the damage-playtime slope as my rank got higher.
The ranks I was assigned and their order:

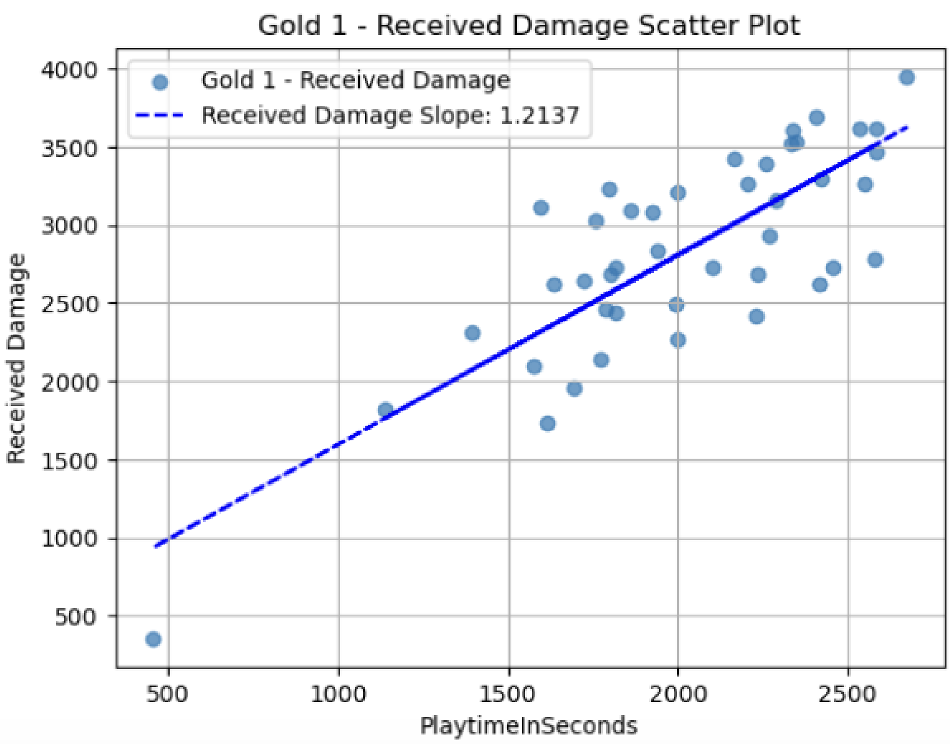
Iron 2<Iron 3<Bronze 1<Bronze 2<Bronze 3<Silver 1<Silver 2<Silver 3<Gold 1<Gold 2<Gold 3<Platinum 1

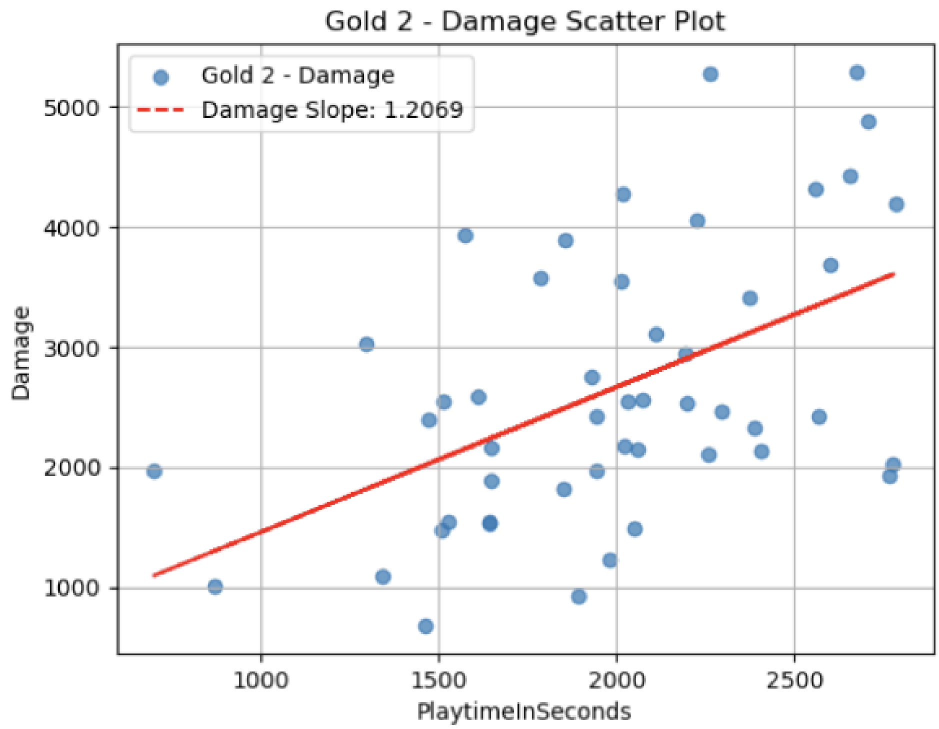
First, I got the rid of the ranks I played less than 10 matches. Then, I looked at each rank's damage-playtime and received damage-playtime. In the end, received damage-playtime's relationship was higher than damage-playtime in each rank. However, there was not a correlation between that relationship and the ranks. Becasue received damage-playtime did not have a linear increase as the ranks got higher. Therefore, rank was not a causation for my hypothesis.

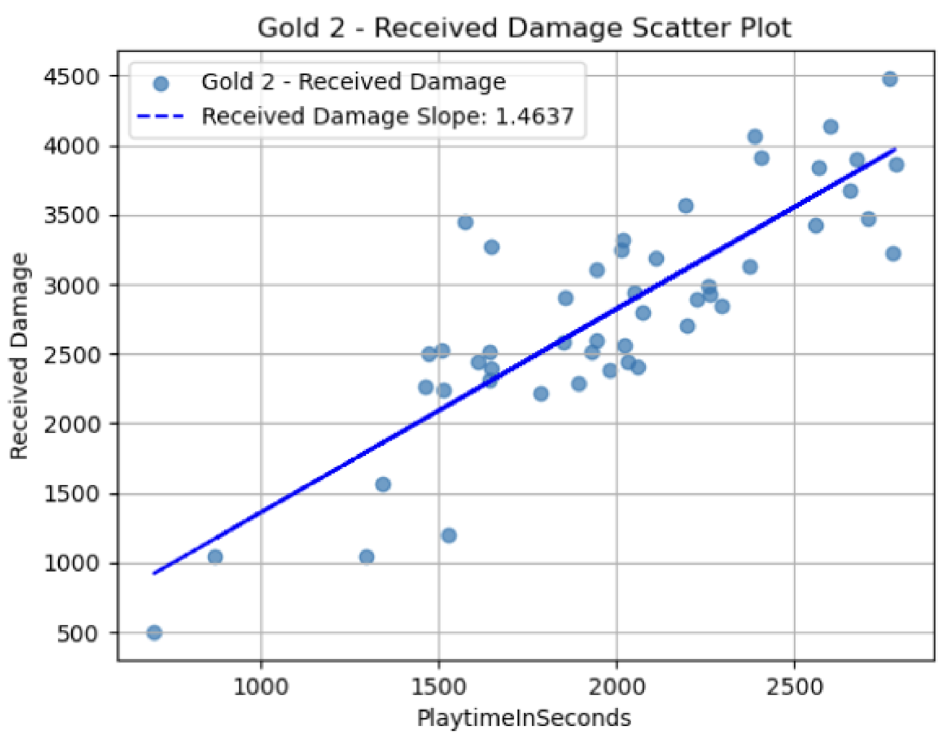


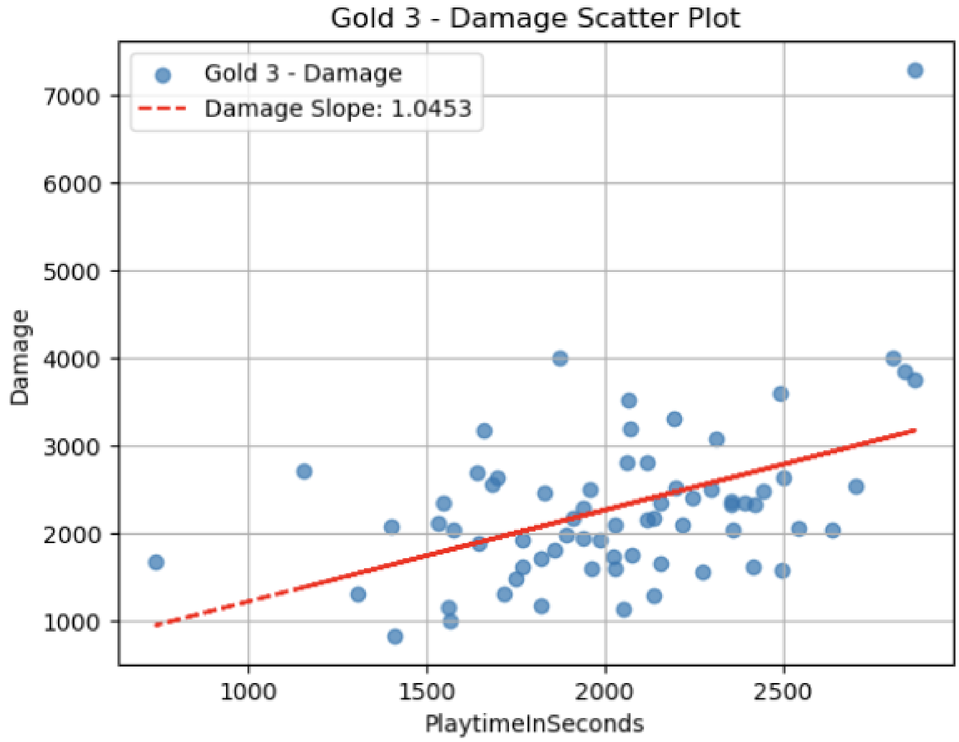


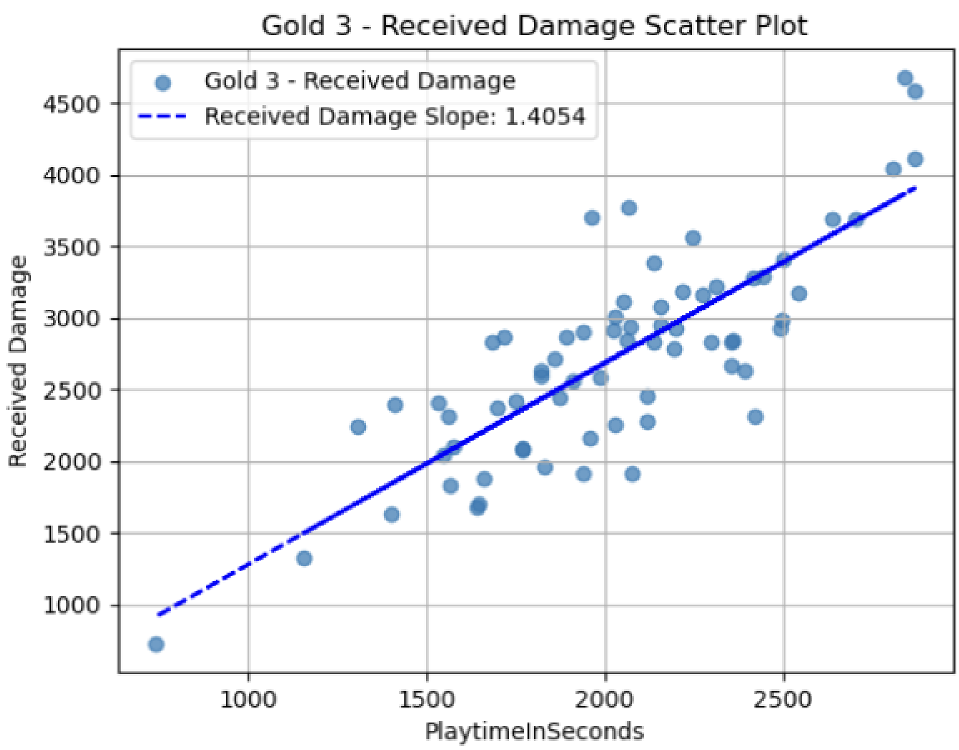

Gold 1 - Damage Scatter Plot
Gold 1 - Damage
Damage Slope: 0.9486
Damage
PlaytimeInSeconds
5000
4000
3000
2000
1000
500
1000
1500
2000
2500

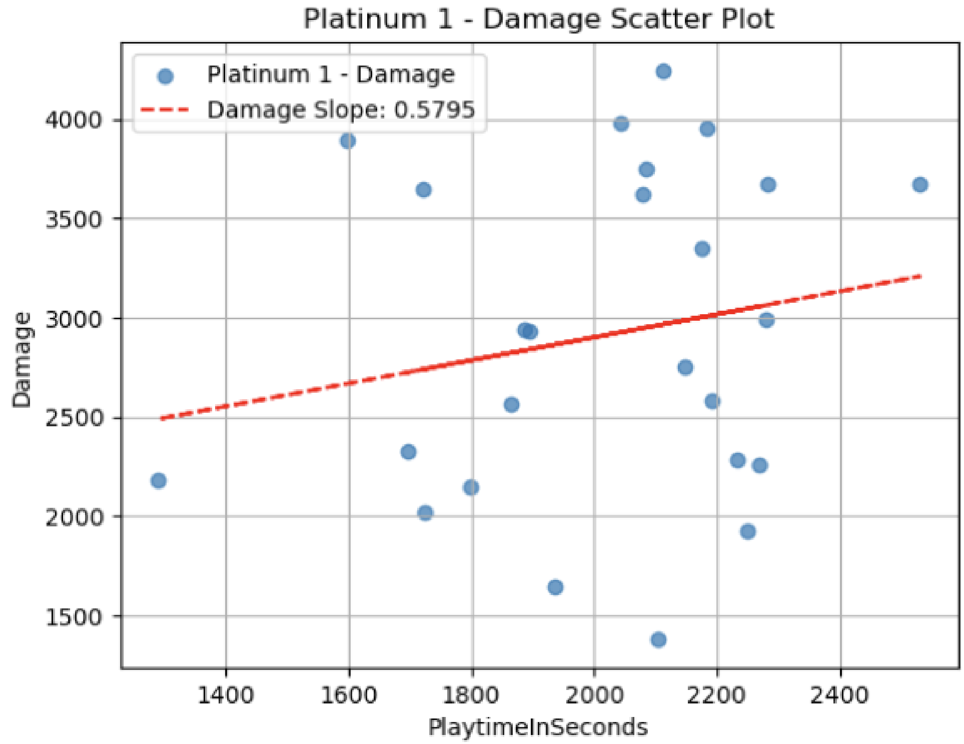

Gold 1 - Received Damage Scatter Plot
Gold 1 - Received Damage
Received Damage Slope: 1.2137
Received Damage
PlaytimeInSeconds
4000
3500
3000
2500
2000
1500
1000
500
500
1000
1500
2000
2500

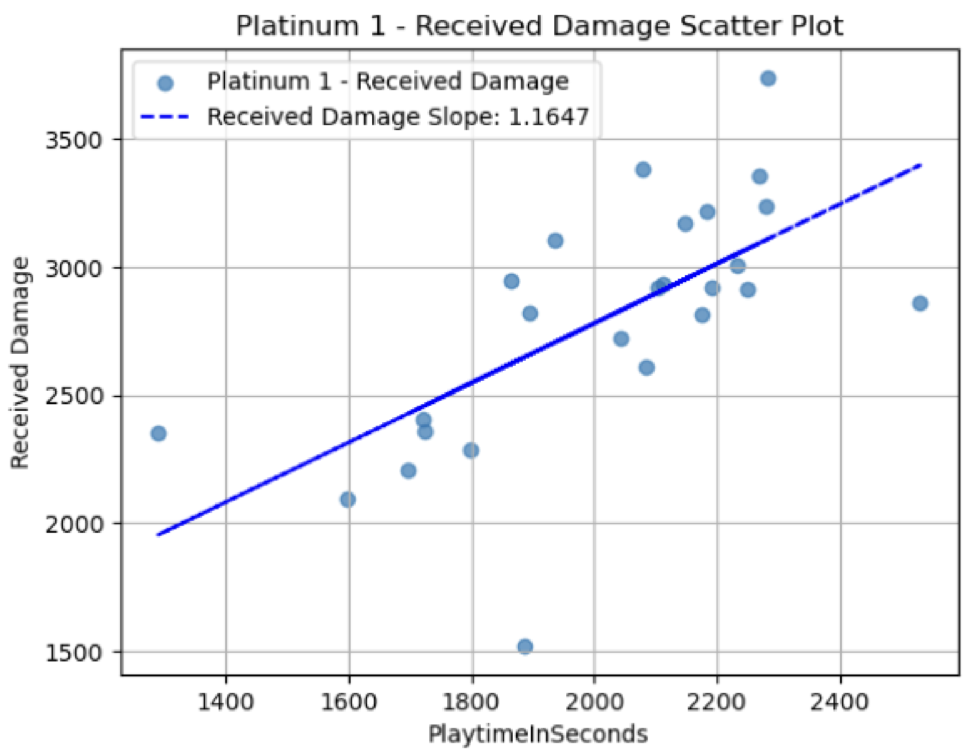

Gold 2 - Damage Scatter Plot
Gold 2 - Damage
Damage Slope: 1.2069
Damage
PlaytimeInSeconds
5000
4000
3000
2000
1000
1000
1500
2000
2500

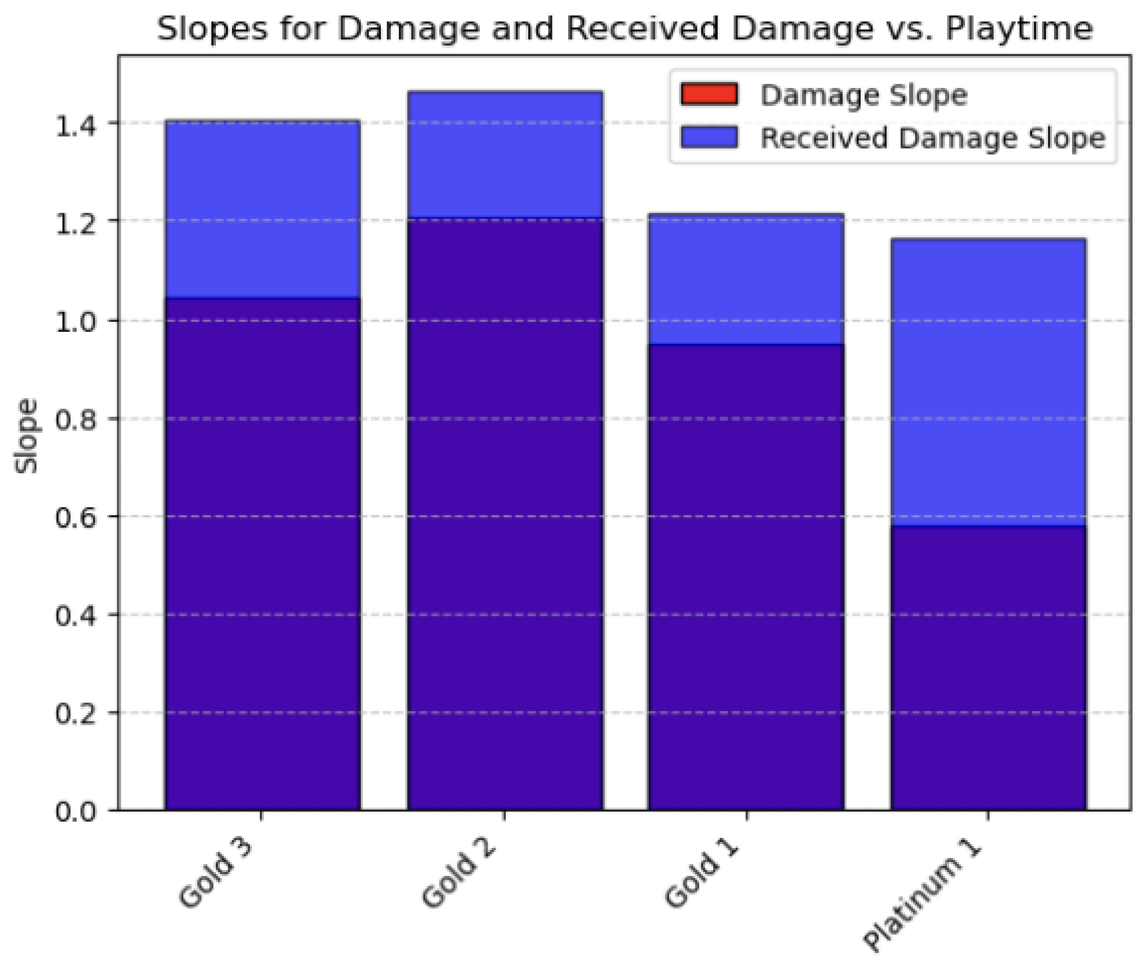

Gold 2 - Received Damage Scatter Plot
Gold 2 - Received Damage
Received Damage Slope: 1.4637
Received Damage
PlaytimeInSeconds
4500
4000
3500
3000
2500
2000
1500
1000
500
1000
1500
2000
2500


Gold 3 - Damage Scatter Plot
Gold 3 - Damage
Damage Slope: 1.0453
Damage
7000
6000
5000
4000
3000
2000
1000
1000
1500
2000
2500
PlaytimeInSeconds


Gold 3 - Received Damage Scatter Plot
Gold 3 - Received Damage
Received Damage Slope: 1.4054
Received Damage
4500
4000
3500
3000
2500
2000
1500
1000
1000
1500
2000
2500
PlaytimeInSeconds


Platinum 1 - Damage Scatter Plot
Platinum 1 - Damage
Damage Slope: 0.5795
Damage
PlaytimeInSeconds
4000
3500
3000
2500
2000
1500
1400
1600
1800
2000
2200
2400

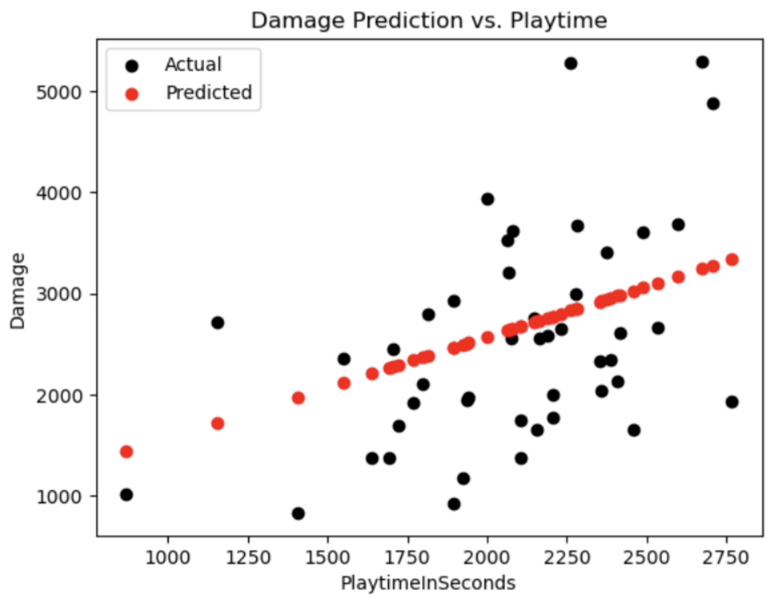

Platinum 1 - Received Damage Scatter Plot
Platinum 1 - Received Damage
Received Damage Slope: 1.1647
Received Damage
PlaytimeInSeconds
3500
3000
2500
2000
1500
1400
1600
1800
2000
2200
2400



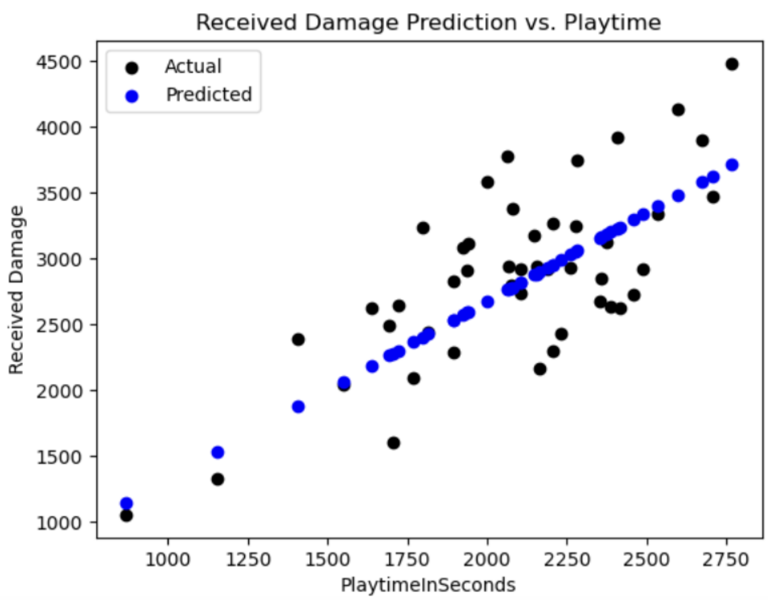
ML: Linear Regression

I created an ML (Machine Learning). It aims to find relationship between playtime with received damage and the relationship with playtime with damage. It evaluates the models' performance by calculating the Mean Squared Error. The scatter plots visually represent how well the models predict damage and received damage based on playtime. MSE for damage was 862149.1369, and MSE for received damage was 219514,2050. Lower MSE is preferred since it indicates that the predictions are closer to the actual value, however if the range is wide these values are also acceptable. Therefore, I looked for the max and min values. The range was wide, therefore MSEs were acceptable.

```
Mean Squared Error for Damage: 862149.1369
```




```
Mean Squared Error for Received Damage: 219514.2050
```




```
Minimum Damage: 552
Maximum Damage: 7299
Minimum Received Damage: 354
Maximum Received Damage: 4677
```

Findings & Future Work

My hypothesis: "As my Playtime increases, a corresponding rise in my damage is lower than the damage I receive." is proved to be correct, meaning I was a burden to my team… However, I decided to start playing again and thanks to this project now I know I'll be playing with Skye in competitive and try to get better at playing with other agents. In addition, I'll explore each map especially Haven since it had the worst results but aside from that I cannot do much about maps since it is chosen by the game, not me. The ranks did not support my hypothesis since they did not have a linear correlation, however the received damage-playtime is higher in all ranks than the damage-playtime, therefore I might play unranked until I get used to the maps and the agents. For this project, I plan to add more data to improve my ML. Moreover, I'll get the results of my matches when I start playing again and compare them with the ML's predictions. In addition, I might make a new project where I compare these results and my new results.