

Data Storytelling Project Analysis

Problems Faced

After attempting to create useful aggregates and graphs using my cleaned dataset, I was able to find the flaws in the dataframes created from my previous cleaning process. I found modified columns names with typos, creating an added altered column. I discovered a column from one out three of the dataframes containing 0 for every entry. It took many attempts getting information out of this column before realizing every entry was 0.

My initial data frames were too large for Jupyter to handle, creating long wait times for simply computations. I realized the dataframe needed to be reduced in size. I learned I needed to reduce the dataframe in a separate jupyter notebook, export it to a file, and then read that reduced file in a new journal. If I attempt to import the large file, then reduce it in the same notebook, I learned working the with the same dataframe is still too memory intensive for even a fast pc.

Insights

The main insights I discovered from this project was that players that make it to the top 5, especially first place, will end up with the most kills most likely. This challenges some assumptions on strategy that often people think of when playing PUBG; such as hiding and not engaging enemies will keep you safe and keep you alive. Since players that win first place tend to have high kills, It's safe to say their aggressive play style keeps them alive and able to win.

I'd like to better analyze the correlation between winning a match and distance traveled. With my current graphs making an observation is tough to make even a broad conclusion. Focusing the data and graph on this might better inspect this. For example for the column showing distance traveled by vehicle, majority of players don't get in the vehicle at all in a match, whether they win or not, it's hard to see if there's any correlating results.