***Data Analysis and Visualization***

***Individual Project Weighting: 50% Due: 4-5-2025***

Evan Liu - D00244247

**Theoretical framework**

I created a Valorant Player performance index (VPPI) for evaluating players across multiple teams and roles. Not like simple K/D ratio, VPPI uses a mixture of (ACS, K/D, headshot %, clutch rate) which categorized into

* **Aggression** (ACS, first kills, headshot %) The frequency of the player's first kill, their ACS, and their headshot shooting accuracy.
* **Survivability** (inverse first deaths, K/D) The frequency with which they do not die first and the quality of their K/D ratio.
* **Impact** (assists, clutch %, KAST) How frequently they assist teammates, win clutches, and participate in trades (KAST).

**Data Selection**

I picked up vct-challengers.json from <https://www.kaggle.com/datasets/sauurabhkr/valorant-champions-tour-2024/data>

with more than 2000 VCT-challengers’ players during 2024. I have only picked stats relevant to their individual performance.

**average\_combat\_score** — Overall match performance score combining damage, kills, and objective play.

**kill\_deaths** — Kill-to-death ratio (K/D).

**kill\_assists\_survived\_traded** — KAST (%) — how often a player contributes to a round.

**average\_damage\_per\_round** — How much damage the player deals per round.

**kills\_per\_round** — Average kills per round.

**assists\_per\_round** — Average assists per round.

**first\_kills\_per\_round** — Entry kills — how often the player secures the first kill.

**first\_deaths\_per\_round** — How often the player dies first in a round.

**headshot\_percentage** — Percent of shots that are headshots.

**clutch\_success\_percentage** — Win rate in clutch situations (1vX scenarios).

A screenshot of a computer screen

AI-generated content may be incorrect.

**Imputation of missing data**

I scanned the dataset for missing data across all performance metrics.

The following metrics were found to have missing data

A screenshot of a computer

AI-generated content may be incorrect.

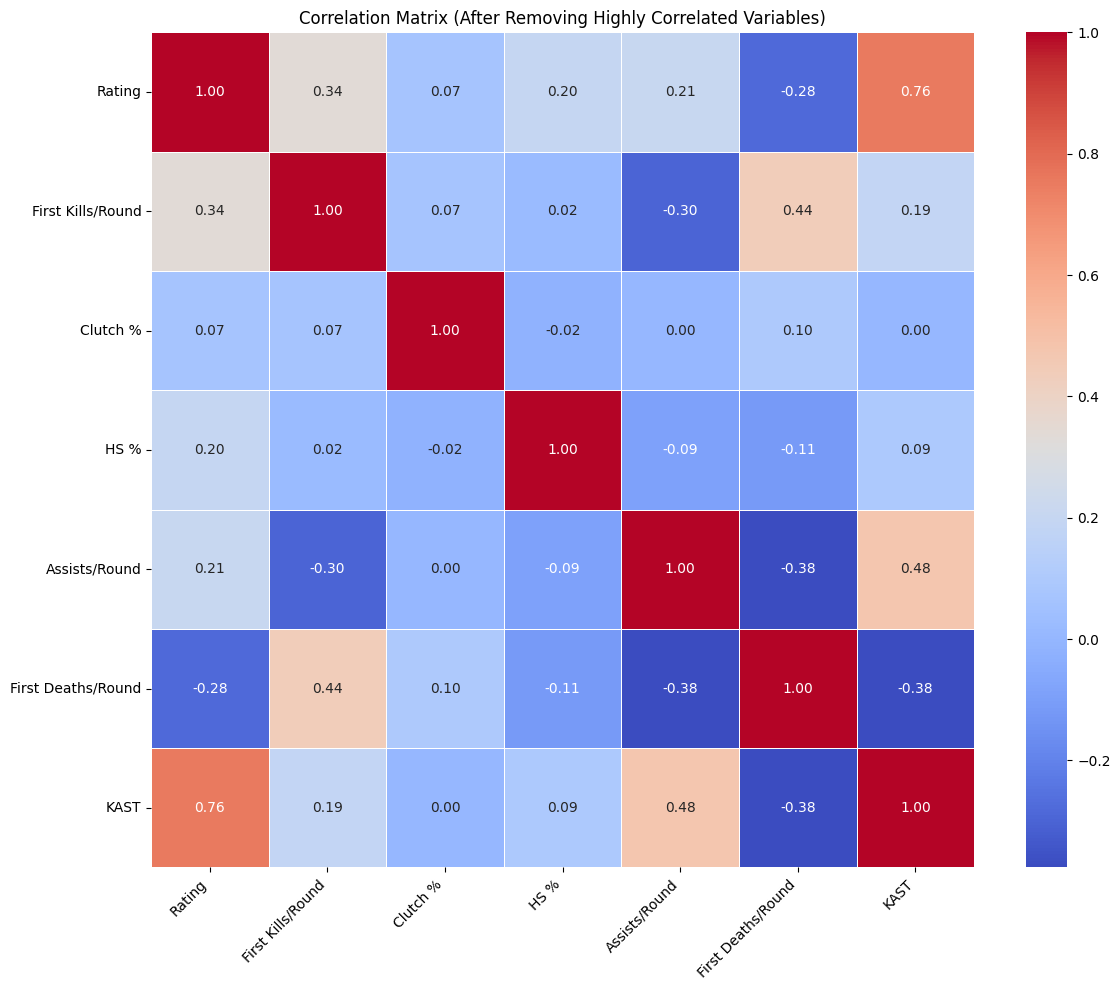
I didn’t want to exclude these rows , which would reduce the sample size and could skew the dataset, so I filled the missing values to preserve the players stats.

For rating , kast and headshot\_percentage have a relatively low missing entries, so for the data not be significantly skewed I used the mean of the column.

Clutch\_success\_percentage has a over 25% missing value which indicate it need careful action. I picked median which is more robust to outliers and suited for data that could be skewed. ( some players never encounter clutch scenarios, while others are clutch gods)

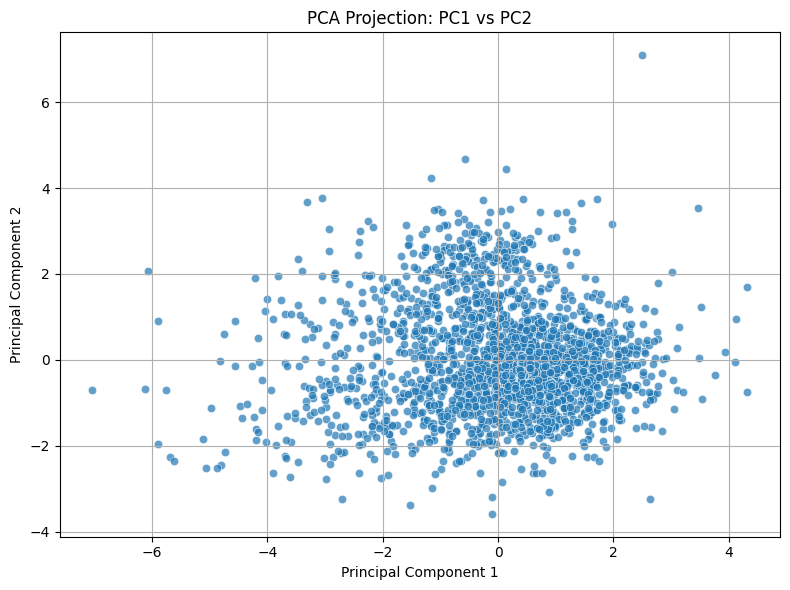
**Multivariate Analysis**

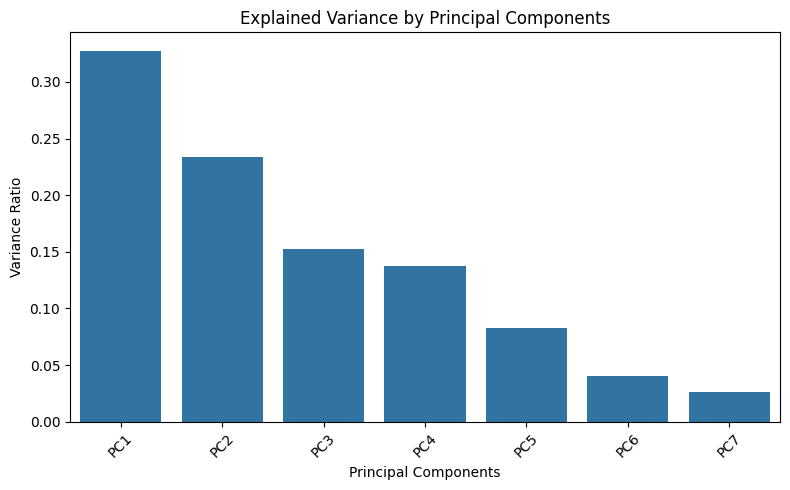
Correlation Matrix has some variable closely related. There were significant overlaps (above 0.85 correlation) between kills per round, K/D, ADR, and ACS.  
This makes sense because a player with a high kill total is most likely dealing a lot of damage and earning a high combat score.



KAST and Rating had a noteworthy correlation of 0.76, indicating that overall performance is strongly influenced by surviving and participating in transactions.

**PCA – Principal Component Analysis**

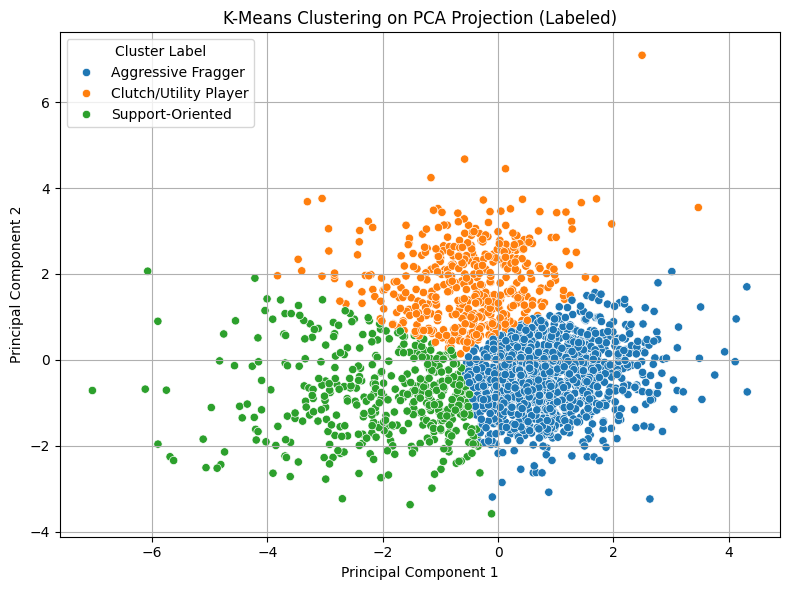
We further simplified the dataset using PCA after lowering the variables. The goal was to preserve the majority of the information while reducing it to a small number of dimensions.  
  
The scatter plot displays the participants' two-dimensional distribution (PC1 and PC2).  
How much each "Principal Component" contributes is shown in the bar chart:  
About 33% of the data may be explained by the major axis of variation, or PC1.  
About 23% can be explained by PC2.  
The first four PCs collectively account for more than 85% of the variance, which is a really good explanation.  
  
This demonstrates that we can comprehend the majority of player differences using only a few parameters, which improves grouping and visualisation. 



**Clustering (K-Means)**

I grouped players according to performance trends using K-Means Clustering (k=3).  
  
Aggressive Fraggers (blue): These players are known as "entry fraggers" or duellists since they lead the charts in headshots and first kills.

Support-Oriented Players (green): These players dominate in KAST and assists. In addition to fragging, they excel in supporting teammates.  
  
Clutch/Utility Players (orange): They frequently carry in difficult rounds and, despite their lack of flash, are reliable in clutches and exchanges.  
  
Coaches or analysts can use this grouping to identify who contributes what kind of value to a squad and see beyond just "top fragger."



This step was very beneficial due to:

It cleaned our data, retaining just variables that were useful.  
PCA assisted in simplifying the data while preserving its narrative.  
In addition to rankings, clustering provided us with distinct classifications of player types.  
Because of this, the VPPI is more than just a figure; it also explains a player's performance, not just how much they do.

**Normalisation**

We had to normalise the values to ensure that each performance parameter contributed equitably to the final VPPI score. Because the statistics I utilise (such as kills, assists, headshot percentage, etc.) all fall on rather various scales, this phase is particularly crucial.

I applied Z-score normalization to the reduced dataset. The variable would have a mean of 0 and standard deviation of 1.

A screenshot of a computer

AI-generated content may be incorrect.

A player with a rating of 3.32 is performing more than three standard deviations above average, which is considered to be very good.  
  
A score of -1.40 for first deaths per round indicates that the player rarely dies first, which is positive for survivability.

A clutch of 2.87 indicates that the player is strongly active in clutch situation.

**Weighting and Aggregation**

Now that Z-scores have been used to normalise all of our statistics, I can aggregate them into a single number called the Valorant Player Performance Index (VPPI), which indicates a player's overall performance.

A screenshot of a graph

AI-generated content may be incorrect.

The objective is to:  
Sort comparable stats into groups (such as support or aggressiveness).  
Determine each category's subscore.  
Created a single final VPPI score by combining all of the sub-scores.  
This provides us with a balanced, detailed picture of a player's performance across several gameplay factors rather than just one.

First Deaths/Round was flipped since lower values are preferable. We just flipped the Z-score since we believe that a player who dies less early in the round should be rewarded.

I used equal weighting across variables to maintain fairness and ease of interpretation:  
  
This indicates that none of the categories are viewed as being more significant than the others. Aggression, support, and impact all make equal contributions to the final index.

Links to other data

<https://tracker.gg/valorant/leaderboards?platform=pc&region=na&act=aef237a0-494d-3a14-a1c8-ec8de84e309c&page=1>

https://www.vlr.gg/stats