Windsor Miners  
March 21st 2022



**PROJECT PROPOSAL**

Counter Strike: GO

Phase 1: EDA Analysis

Suhail Riyaz Ahmed

Raj Bharatkumar Bhalodwala

Bhavesh Trigunanand Mishra

Luis Palacios Melendez

Dhruvalbhai Sunilkumar Patel

Amit Sharma – **Group Leader**

Content

[1. Introduction 3](#_Toc98935539)

[2. Need/Problem 3](#_Toc98935540)

[3. Goals/Proposal 4](#_Toc98935541)

[4. Timetable 5](#_Toc98935542)

[5. EDA Analysis 6](#_Toc98935543)

[6. GIT Setup 9](#_Toc98935544)

[7. Appendix 10](#_Toc98935545)

# Introduction

Counterstrike is a FPS (First-Person Shooter) game in which two teams of 5 players face each other in a matchup. The game includes a Terrorist side (T) that is tasked to plant a bomb and have it detonated, and a Counter-Terrorist side (CT) that is tasked to defuse the bomb or prevent it from being planted.

The game of Counterstrike is a best of 30 rounds, the winning team being the first to win 16 rounds. The 30 rounds are played in two halves of 15 on each side of the map. There are 7 maps in the map pool that are available to be played competitively at any given time. Matches are normally played as a 'bo3' (Best of 3) maps, with finals often played as 'bo5’s. Counterstrike also has an economic system that governs the acquisitions of armour, weapons and grenades by the players.

**Motivation:**

Online gaming world is full of First-Person Shooter games, with CS: GO being one of the most sorted out game by online players/gamers. Most players (old or new) try to get along on the platform and try to hone their skills (either offensive and defensive) by playing competitive matches.

But the **other side of the story** is that CS: GO platform has the highest dropout rate for new/beginner online gamers. With the game and maps becoming more and more competitive and complex it becomes extremely difficult for a new gamer to keep on going on the platform after certain time on the platform. Beginners often lose interest as they face tough teams, maps and game styles which sometimes don’t suit either the skill level or the playing strategy of the beginner.

# Need/Problem

Beginner level Gamers starts enthusiastically on the Gaming platform but soon found wanting on the skills and the strategy level. The game has complex maps and mechanics, which are need thorough knowledge and tactics to master the area, on top there is style of play (either offensive or defensive) which drives the success rate for a player or team.

Often new gamers play for few days and as the game is tough, they lose a lot of initial matches, which in turn triggers the DROP-OUT tendency for the new gamer. Who likes Losing? – and thus the interest levels go down for the new players. Also, players don’t get a hang of game styles for a specific map. There is always a missing piece of the puzzle where the new gamer misses out on success (Map, Style, Economics, Skills etc.)

They need upfront guidance and insights for starting their online game journey. These insights can help the gamers to choose better maps, learn from top players/teams and use strategies (economic) that can raise their own interest in this complex gameplay. Eventually the gamer can perfect his own style of gameplay (offensive or defensive) by learning or copying from the very best around on the circuit.

# Goals/Proposal

The team will take a descriptive analysis approach in which team will try to develop an analytic product to examine the data from a dataset which is pulled from a specific timeline and containing professional match data for players, teams and maps for Counter Strike: Global Offensive online game/tournament.

This data can help Beginner level online gamers to gain insights related to the past trends for the online tournament games which can help them to gain strategic edge while starting their new online journey.

**Beneficiaries**

New gamers will be direct beneficiaries of the analytic trend data to understand game, maps, styles and economics.

Game Creators will be benefitted indirectly by having better retention percentage for beginner level players – hence retaining new customers.

The analysis will try to give answers to some of the below few questions: -

Which maps get most picked with high number of kills per round?

Which are the top players in an event and can decided awards for MVP?

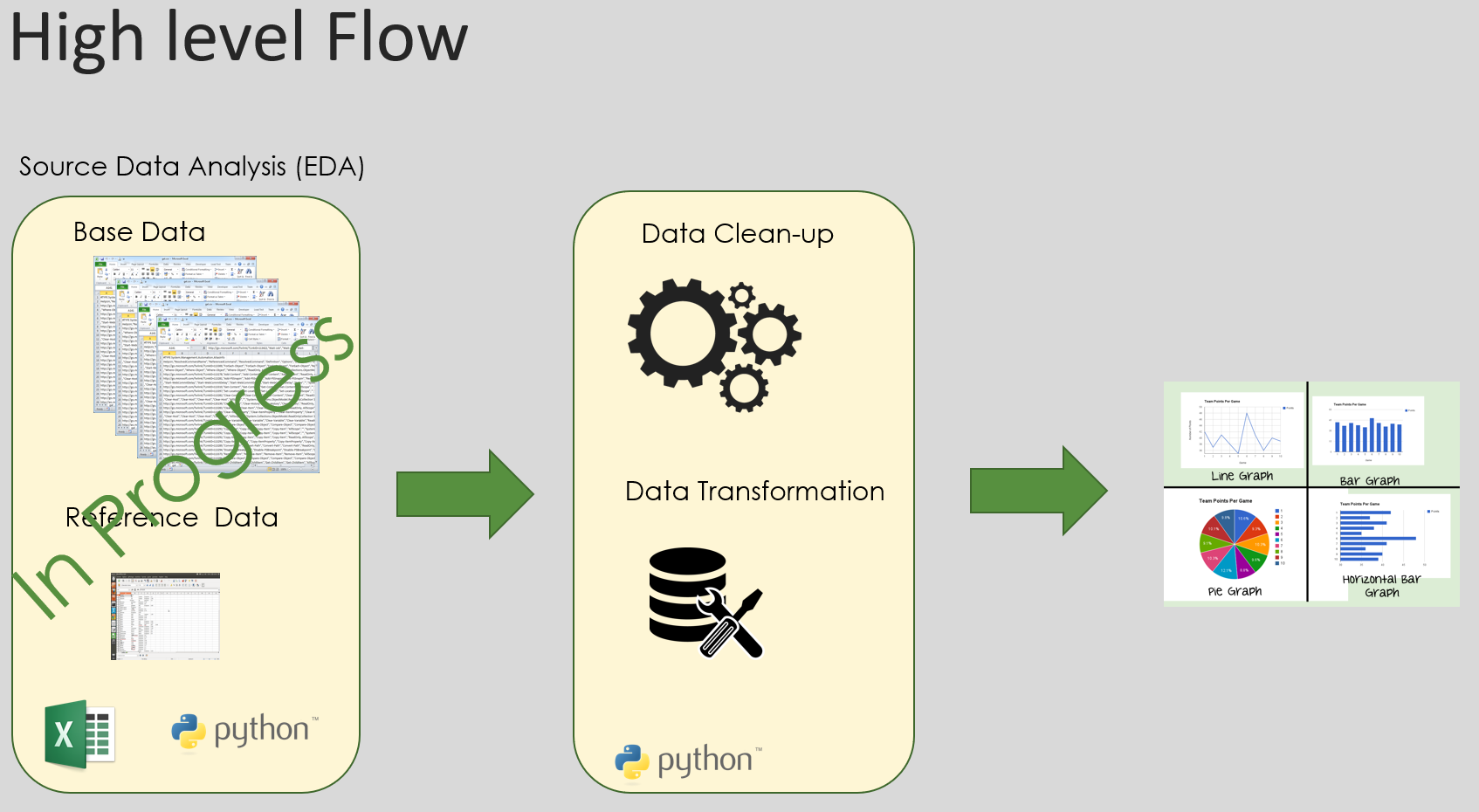
How to save economy in game to maintain a good inventory?

Where to start and what approach to take on a certain map?

# Timetable

Provide detailed information on the expected timetable for the project. Break the project into phases and provide a schedule for each phase.

|  |  |  |
| --- | --- | --- |
|  | **Description of Work** | **Tentative Delivery Dates** |
| **Phase One** | Proposal and EDA Analysis | March 25th, 2022 |
| **Phase Two** | Data – Cleanup and Transformation | Mid-April 2022 |
| **Phase Three** | Project Delivery - Visualizations | April 2022 - End |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ACTIVITY** | **IMPLEMENTATION TIME** | | | **RESPONSIBILITY** |
| 1.1 Discuss and implement the EDA approach and Proposal | Phase 1 | Phase 2 | Phase 3 |  |
| March 2022 |  |  | Full Team |
| 1.2. EDA for existing and New Data | March 2022 |  |  | Full Team |
| 1.3. Final Project Proposal | March 2022 |  |  | Full Team |
| 1.4 Phase 1 Submission | 25th March 2022 |  |  | Team Leader |
| 2.1 Python Training |  | April 2022 |  | Full Team |
| 2.2 Data Transformation and Clean up activity |  | April 2022 |  | Full Team |
| 2.3 Phase 2 Submission |  | Mid-April 2022 |  | Team Leader |
| 3.1 Tableau Training |  |  | Mid-April 2022 | Full Team |
| 3.2 Data Visualization – Draft version |  |  | Mid-April 2022 | Full Team |
| 3.3 Phase 3 Submission |  |  | April End 2022 | Team Leader |

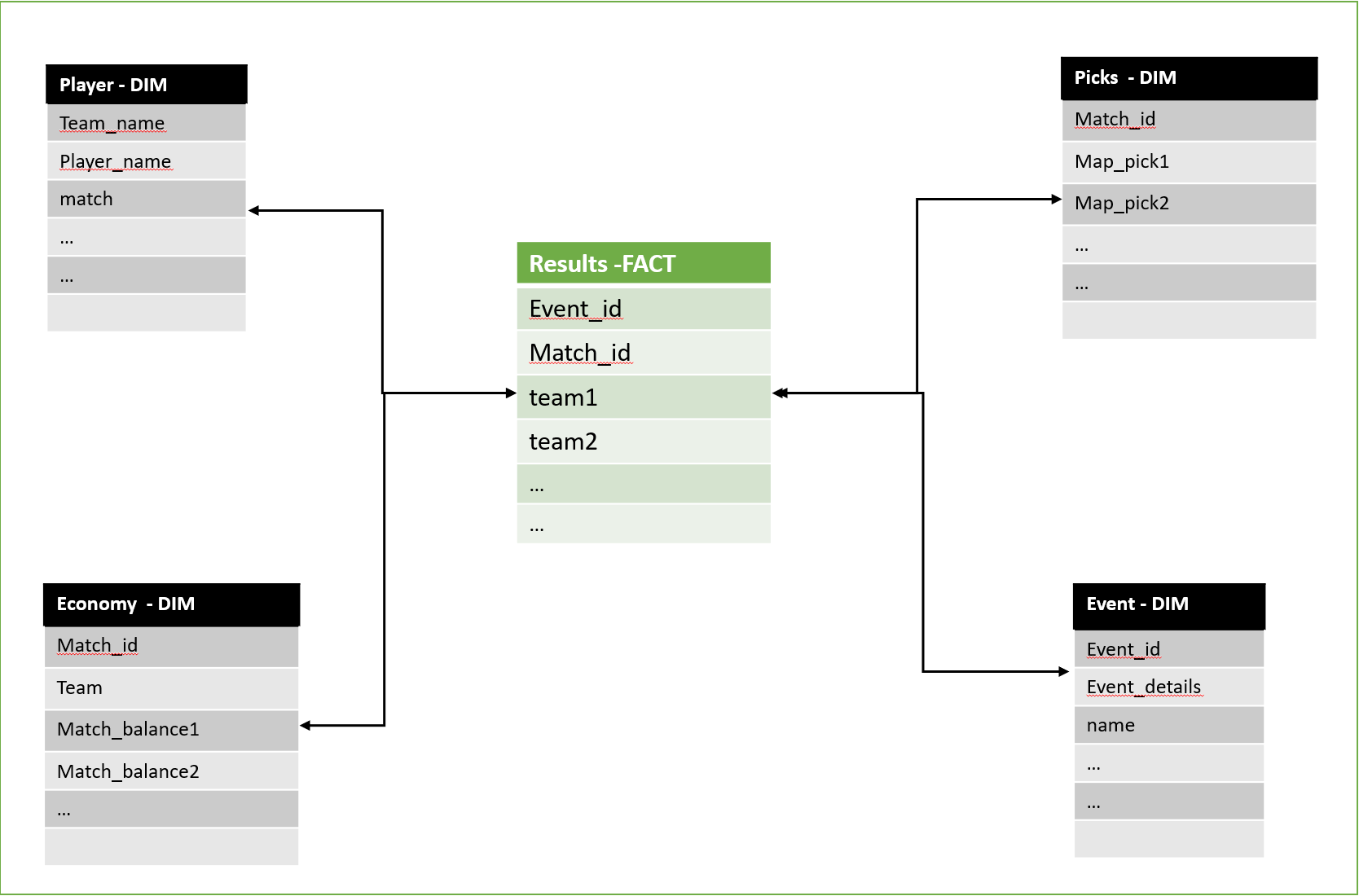
# EDA Analysis

About the Data

The data is split into 5 tables that store data related to:

* Results.csv: map scores and team rankings
* Picks.csv: order of map picks and vetoes in the map selection process.
* Economy.csv: round start equipment value for all rounds played
* Players.csv: individual performances of players on each map.
* Event.csv: Details of the events against event id

Values stored in 'eventid' and 'matchid' columns are unique for each match and event and shared between tables, so these columns can be used as keys to merge data between tables. It is necessary to note that the rows in the 'results' and 'economy' tables store data for each map played in a match, while the rows in 'picks' and 'players' table store data for the entire match. The dates of the matches range from 11/2015 to 03/2020.



**About Data Set**

**Economy** - contains record or every match game by date and event. Data table recollect records of match economy scored by team 1 and team 2 at the end of every round.

**Picks** – contains record of order of map picks and vetoes in the map selection process. Data table recollect records of maps eliminated by team 1 and team 2 in each game to select playing map.

**Players** - information of individual performances of players on each map. Data table recollect records of performance in terms of kills, assist, deaths, headshots, Kast (Percent of round the player killed, assisted, survived or was traded), kills-death difference, average damage per round, firstkills - firstdeaths, and rating in match in three different maps.

**Events** – contains data related to esports events held worldwide along with the type of event and country

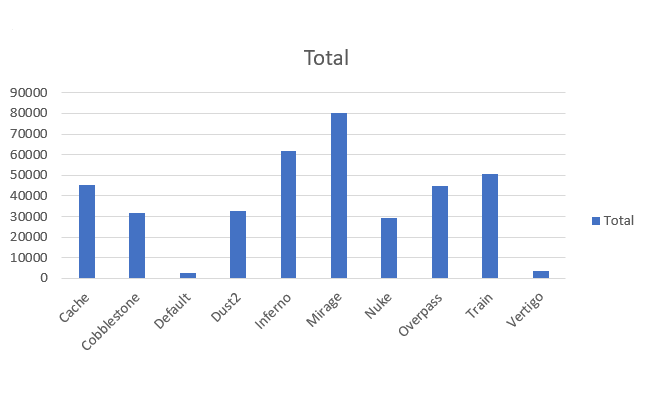
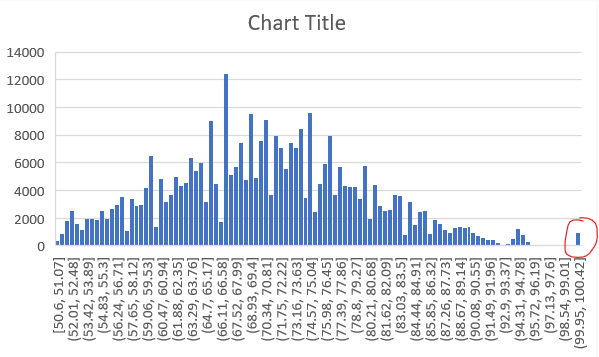
**Results** – contains the results of each game/match played by the teams, its contains full details on styles, strategy, map and winner details

Whole Dataset contains a total of 236 columns across sheets, with maximum columns being **Decimal and Integer values (196 in total)**, these variables are more of continuous in nature with very few of categorical values. Of the **27 String columns** many are repeated across sheets to maintain a connectivity in them. The last category contains **Date fields and Boolean fields** to maintain the timeline and yes/no type answers

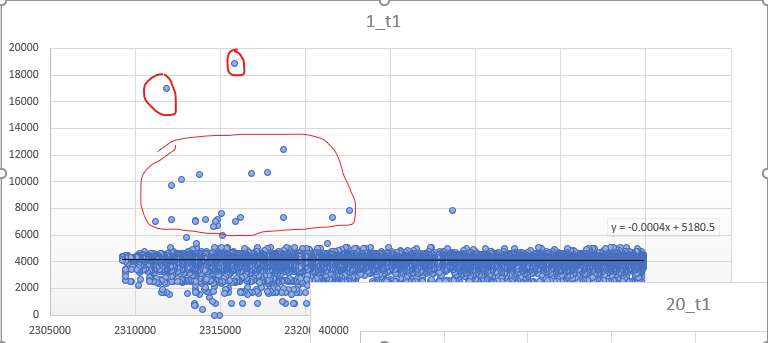
**Invalid and Missing values**

Most of the fields have valid records, there's some fields of t1\_removed\_3, t2\_removed\_3, t1\_picked\_1 and t2\_picked\_1 with some missing values, because the game consists of best of 3 rounds. Player names and some text fields have **invalid records (symbols, nulls and blanks).** Some columns in specific maps have missing values, because data is collected based on performance in three maps

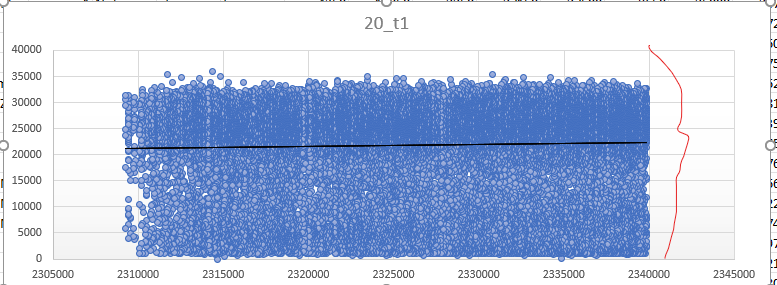
**Outliers**



Economy Table:



As we can see some outliers from the very first games, but data very centralizaed around 5,000 pts



While games are going forward, we see no outliers, but data is very dispersed from 0 to 30,000 pts

Is the data segmented in groups?

Is the data imbalanced? – common theme

Some a few records of points scored on game could have lower results depending of the stage

Some a few records of map 3 in game could have lower results. So, it shows 0 in the sheet as the result.

Some a few records of points scored on map could have lower results depending of the stage

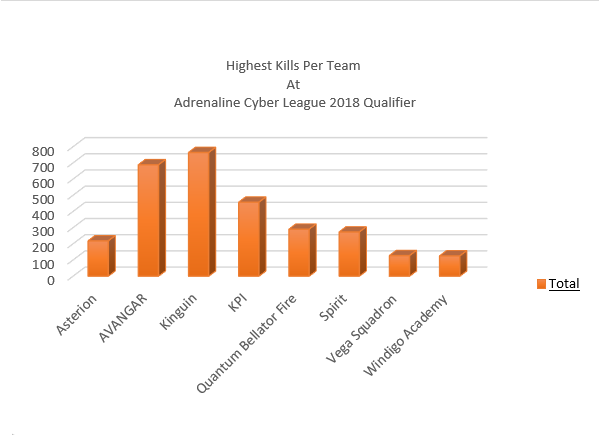
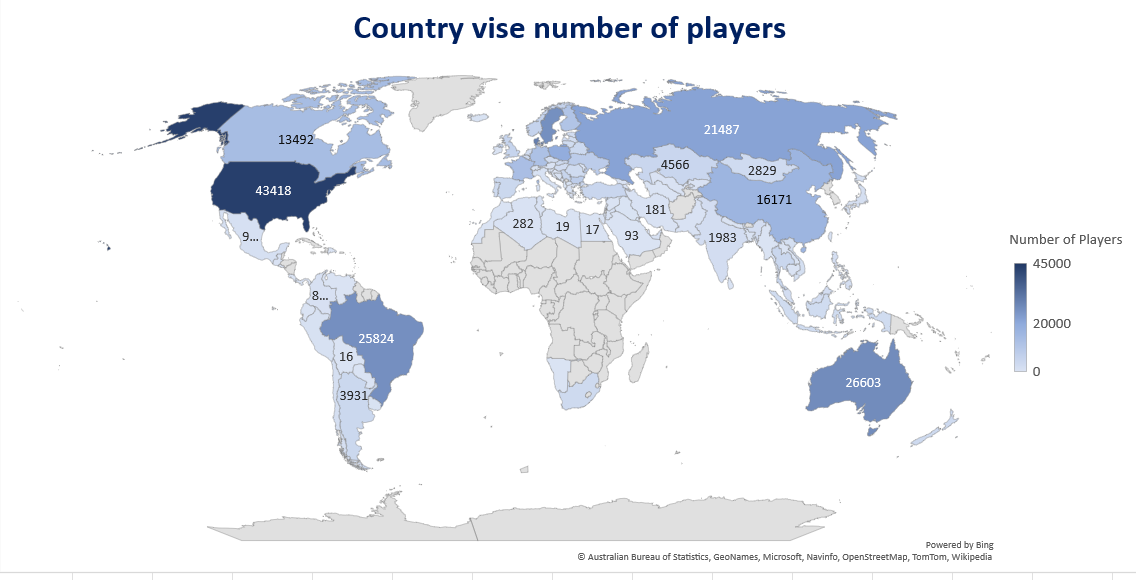
Some element correlated?

Matches were recorded in a range from 11/2015 to 03/2020. This data was scraped off [**https://www.hltv.org/**](https://www.hltv.org/)and from that they collected individual performance of player.

Can you generate preliminary visualizations for individual features? – Excel graphs – maybe 2/3

Yes, this table can analyze tendency of every team along the match, and combine the map of each game.

Yes, this table can analyze tendency of every team along the match and combine the map of each game.



# GIT Setup

Team has tried using Version control for initial documentation level. A shared GIT Repo for is created for the team members to add and update documents. This repo will be further used to add code and visualization part.

• <https://github.com/miners-work/DAB-103-Project>

# Appendix

Provide supporting material for your proposal here. It may be:

**Data Link:**

<https://www.kaggle.com/datasets/mateusdmachado/csgo-professional-matches?select=results.csv>

**Team** Logo:

<https://www.canva.com/q/pro/?utm_source=google_sem&utm_medium=cpc&utm_campaign=REV_CA_EN_CanvaPro_Branded_Tier1_Core_EM&utm_term=REV_CA_EN_CanvaPro_Branded_Tier1_Canva_EM&utm_content=358_control&gclid=CjwKCAjwxOCRBhA8EiwA0X8hiwLGbNQbY7mfXqwBIPHJpf8nsWgp0jsCEJ6iQTRehEpI1wci99YBpRoCjzUQAvD_BwE&gclsrc=aw.ds>

**Python:**

**Git:**

<https://github.com/>