Data Analysis Project
P0: Project Idea
(Group 10)

Mathieu Groenen and Jiri Käser

University of Basel
Databases (CS244) course
Autumn Semester 2023

1 Introduction

Counter Strike - Global Offensive (CSGO) is one of the biggest Esport games to date, with usually about 1 million players across multiple continents that are active simultaneously. With numerous tournaments all across the globe CSGO, is an internationally well-known tactical game.

For our project, we will be looking at professional tournament demos from HLTV with positional data of various players and game metadata from over 1500 different games. With this large amount of information, we would like to analyse common behaviour and figure out frequent reactions to different in-game events in order to understand specific actions as well as their implications on the outcome of a round and have a more precise overview over how the game is played by professionals.

On the other hand, we also want to have a look into the Steam Community Market for CSGO. It is a large marketplace where players can trade their in-game items for real money. Single items can easily be sold for thousands of CHF. This marketplace is rather reactive. For example, with the release of the new counter-strike CS2 announcement, prices for many skins skyrocketed. We want to analyse such behaviour of the marketplace in response to public events related to CSGO or even unrelated events such as Christmas or Easter. Every CSGO player that has played in a major tournament (the events hosted by Valve directly) gets his own in-game sticker. The same goes for their team. Thus, performance of individual players and teams may impact the prices of these stickers.

2 Datasets

Our project consists of two datasets, both related to CSGO. The esta dataset contains data from online tournaments between January 2021 and May 2022 and from LAN tournaments between July 2021 and May 2022. The Steam Community Market contains the prices of different cosmetics for the game CSGO.

The datasets are connected in that the esta datasets contains the usage of different weapons and the Steam Community Market contains the prices of different cosmetics for these weapons. Also the esta dataset consists of tournament

demos with data about the different teams and players which is connected to cosmetics of tournaments in the marketplace.

After complete integration we estimate the Database to be around 30 GB.

2.1 Esports Trajectories & Actions (Esta) Dataset

Esta[1] is a CSGO dataset consisting of player actions, movement, game events, and game metadata from tournament demos from HLTV. It is 99.1 GB large. It contains 1558 games which are each stored in a JSON file, capturing all the information of every player's action and game metadata per tick (frequency of game state update to server per second). In all the games combined, there are 7.9 million frames.
A detailed setup of the data can be found in the awpy docs[2].

2.2 Steam Community Market API

The Steam Community Market is a marketplace for virtual in-game items from different Steam games. The Steam market can be accessed through a public API[3]. Our field of interest only requires the items of CSGO. There exist around 22'000 in-game items for CSGO on the steam market. The total size of the data important for our analysis is roughly 1.8 GB in JSON format. Through this API, almost all items that have ever existed on the steam market can be retrieved. The API provides each item with a lot of additional information such as full name, full price history with timestamp per hour, price and amount sold.

3 Analysis Goals

3.1 Goals

With the information provided in the esta dataset, we will be looking at common player behaviour related to positions, strategies, and actions on different maps and for different teams, as well as some overarching questions related to gameplay.
Our main questions are:

- What are some common player positions and what is their effectiveness in the outcome of a round?
- Are there deciding factors or actions to winning a round, and if yes, what are they?

With the Steam Community marketplace data, we are going to focus more on how the market reacts in relation to events, i.e. tournaments etc.
We want to know:

[1] `https://github.com/pnxenopoulos/esta`
[2] `https://awpy.readthedocs.io/en/latest/parser_output.html`
[3] `https://github.com/Revadike/InternalSteamWebAPI`

– How do official Esport tournaments affect the price of skins?
– Do global events outside of the Esport world have an impact on the steam market?

Looking at both datasets, we would like to combine the data of player and team activity during the games of the esta dataset and relate them to how the prices of items (especially the player and team stickers) on the steam market fluctuate with this activity.
The questions and themes that we would like to pursue here are:

– How does in-game usability of a weapon affect the price of skins on the marketplace?
– Does player or team success in tournaments affect the price of team/player stickers and if yes, how?

3.2 Methods

Using the positional data in coordinates per tick from the esta dataset, we will display the positions of each player onto the corresponding map by using a mapping function. Utilizing this, we can then visualize the movement trajectories of the players, show specific events such as kills, deaths, etc. at their respective location of occurrence and use heat-maps to indicate common attacking and defending player positions. With this amount of data, we can look at specific actions at certain locations and compare the outcome of their respective rounds to see if there is a correlation between the two. The trajectories allow us to visualize player movement, which again can be coupled with action events to describe player behaviour. Heatmaps allow us to understand player positioning and effectiveness in defending or attacking.

For the market data, we will plot the prices for individual items over time. After that, we can analyse how the prices are affected short-term and long-term by tournament match outcomes and player performance gathered from the esta dataset. Tournaments taking place between 2021 and 2022 are going to be the main focus, as the esta dataset contains matches starting from January 2021 and ending at May 2022.

4 Tools

A PostgreSQL[4] Database running in a Docker container[5] using the `postgres:x` container image[6] for ease of setup will be used for this project. Due to its easy data manipulation capabilities and extensive libraries, Python is our programming language of choice for the data integration together with the Psycopg[7]

[4] `https://www.postgresql.org/`
[5] `https://www.docker.com/`
[6] `https://hub.docker.com/_/postgres`
[7] `https://www.psycopg.org/`

library to be able to connect to the Database via our Python script. Our Rest API will be implemented in Python as well, using the Flask Web Framework[8], which will also be able to run in a separate Docker container.
For the analysis part, we will again be using Python for its flexibility and mostly the libraries Numpy[9], Pandas[10], Matplotlib[11] and Seaborn[12] to help with the data manipulation and visualization.

All work will be done in the JetBrains PyCharm IDE[13] and DataGrip[14] because of its integrated tools related to Databases and GitHub. Python and its libraries will be managed through the Anaconda Enviroment[15].

All relevant files will be uploaded to our University Basel GitLab repository.

[8] https://flask.palletsprojects.com/en/3.0.x/
[9] https://numpy.org/
[10] https://pandas.pydata.org/
[11] https://matplotlib.org/
[12] https://seaborn.pydata.org/
[13] https://www.jetbrains.com/pycharm/
[14] https://www.jetbrains.com/datagrip/
[15] https://www.anaconda.com/