Final project

- Python for data analysis – DIA4
- GALLET Guillaume - ERIC Thessa - NELSON Théophile

Definition of the problem

StarCraft 2 is a real-time strategy video game developed by Blizzard Entertainment. This game had a solid competitive community since it was released in 2010.

We want to predict the level of the players, represented by their league (from 1 to 8). To do that, we have a dataset of several characteristics for each player, such as his age, number of hours played or their number of actions per minute.

What are the following steps to do:

- Data pre processing : We are going to use several methods to detect and delete/replace the outliers in our dataset, in order to apply the machine learning models in a clean dataset.
- Data visualization : The importance of data visualization is simple, it helps people see, interact with, and better understand data. Whether simple or complex, the right visualization can bring everyone on the same page, regardless of their level of expertise.
- Modeling : We can use whether regression or classification. Depends of the target. Furthermore it is possible play with the hyperparameters to improve our model.
- Django : The django framework is a dynamic application used to create a showcase for analyzing and visualizing the data studied in the notebook.

Your thoughts on the asked question

Our experience in video Games lets us think that the most important values could be the number of Hour played, the actions per minute and the action latency.

Furthermore, we think that it would be interesting to study League Index feature.

Project organisation

- After having appropriated the data using data processing, we decided to display the values in the form of graphs, histograms, pie charts, and maps using libraries such as Matplotlib.pyplot , plotly.express, seaborn during the data visualization then we ended up modeling to predict the level of the player.
- Each person in the group participated in the development of the different parts of the notebook by trying to be as precise and exhaustive as possible, and by reviewing all the attributes that were available to us.
- Then we looked at Django, we chose to represent 2 different graphs with a mini widget allowing the user to choose the category concerned and the Framework executes the code necessary for the visualization requested.

Encountered difficulties

The difficulties encountered on this project were few, the most delicate part to deal with was the modeling because the results are unsatisfactory, with confusion matrices which have rather low precision

Division of labor :

The distribution of work was globally balanced, we worked at home and we saw each other several times to move forward together, help each other and deepen the work of one or the other.

- Notebook distribution: 33% Guillaume, 33% Thessa and 33% Théophile
- Django : 33% Guillaume, 33% Thessa and 33% Théophile
- Distribution Ratio: 33% Guillaume, 33% Thaïs and 33% Swann