Analysis of the SkillCraft dataset

What is starcraft ?

Starcraft is a competitive game developed by Blizzard. There is a ranked system within the game : (Bronze, Silver, Gold, Platinum, Diamond, Master, GrandMaster, and Professional leagues coded 1-7 (Ordinal) because the Professionnal leagues is not in the game). Players are placed in a league after having completed 5 placement matches.
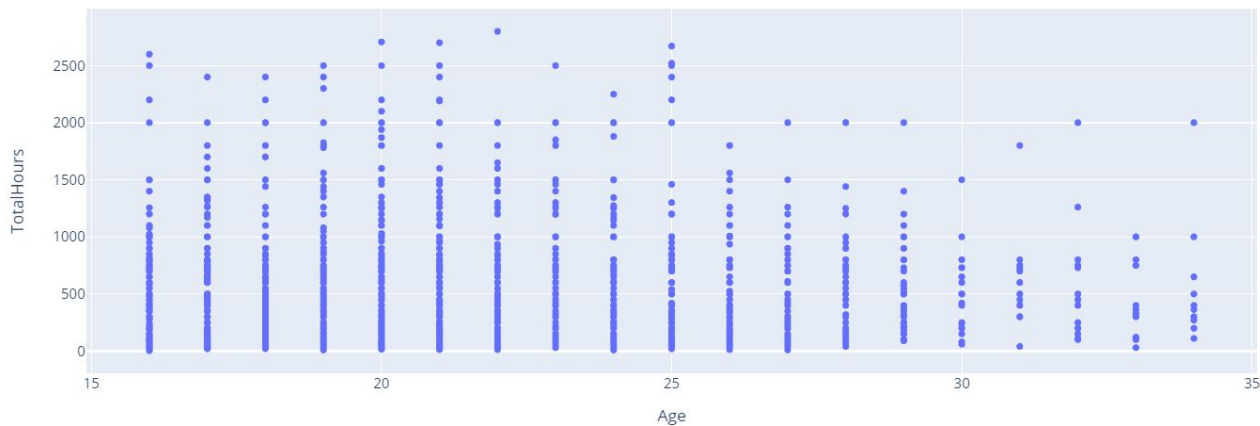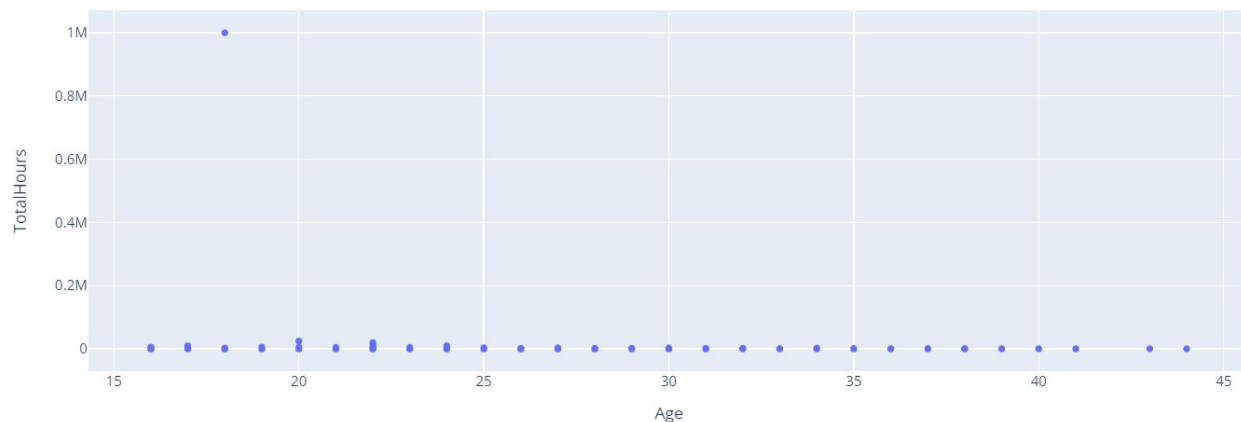
Our problematic

We want to help the constructor. Why ? Imagine a player has play starcraft for a long time but as never played in a rank situation (game). We can make the player play 5 games with other player in rank (incremental if he win) and estimate based on his stats what rank he deserve.
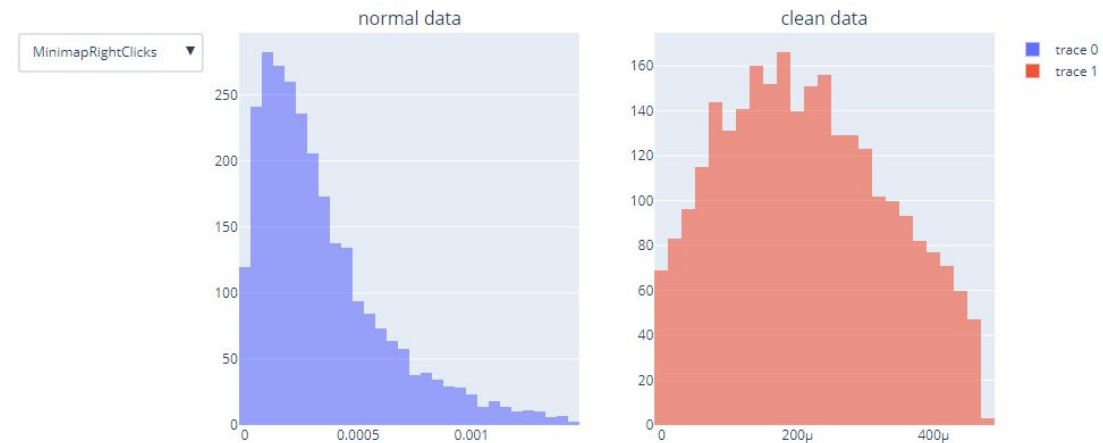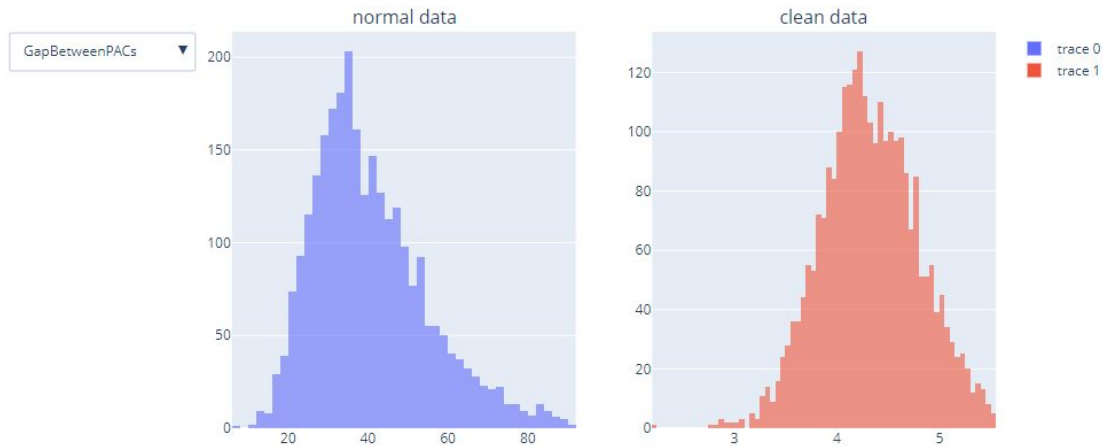
Data cleaning and preprocessing

Removing outliers

# Data preprocessing

Standardisation

for each features

Most relevant features

	Importance	Chi2	Corr
ActionLatency	0.081724	18.883694	0.630793
APM	0.072595	149.586781	0.616659
TotalHours	0.070703	1287.631017	0.478378
NumberOfPACs	0.065927	0.068338	0.553018
GapBetweenPACs	0.061926	32.669800	0.483009
SelectByHotkeys	0.059685	0.478959	0.523275
AssignToHotkeys	0.057333	0.031468	0.434344
MinimapAttacks	0.053809	0.008517	0.311913
WorkersMade	0.052969	0.012709	0.332405
HoursPerWeek	0.051774	31.702385	0.176768
UniqueHotkeys	0.051261	197.669436	0.287482
ActionsInPAC	0.051202	1.920662	0.165634
Age	0.050066	0.002262	0.065518
TotalMapExplored	0.049154	26.636277	0.195719
MinimapRightClicks	0.047928	0.007278	0.194399
UniqueUnitsMade	0.047165	11.460188	0.120097
ComplexAbilitiesUsed	0.043735	0.005919	0.173853
ComplexUnitsMade	0.031043	0.003081	0.157519

# Amount of player in each rank

It means that it's possible that we lack data on Grandmaster players and possibly would be less accurate on their predictions.
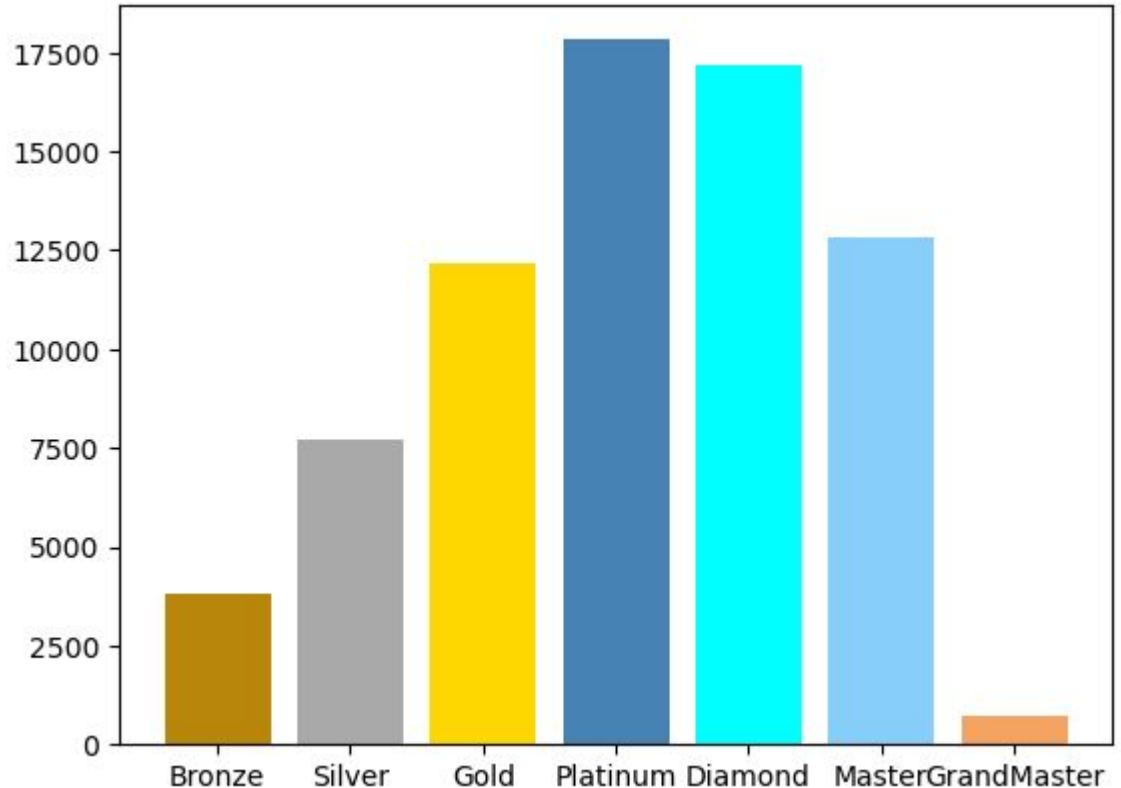
Comparison between Leagues

- GM player mean hour per week is 2 times bigger than casual players
- It seems there's a big gap between diamond and master+ player because Diamond players may have a lot of total hours with progressing or ancient master+ players who no longer play as often are demoted to diamond
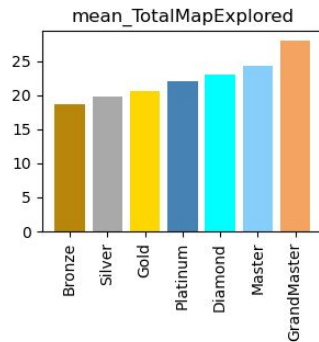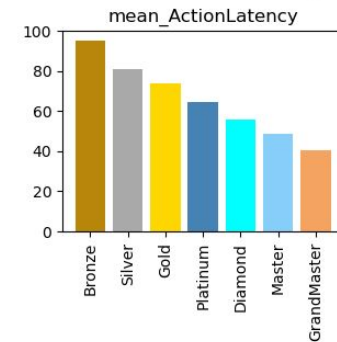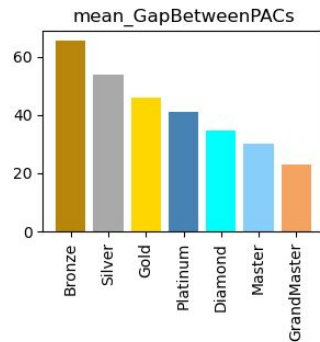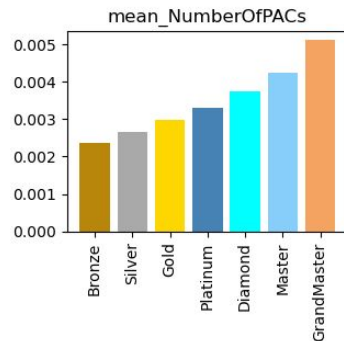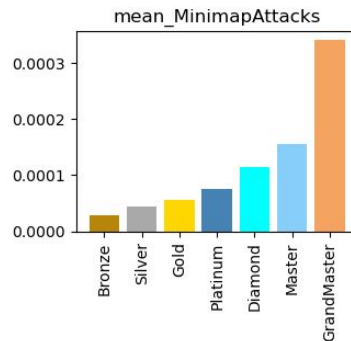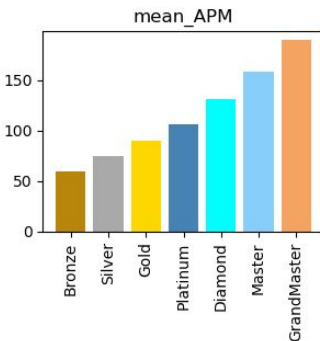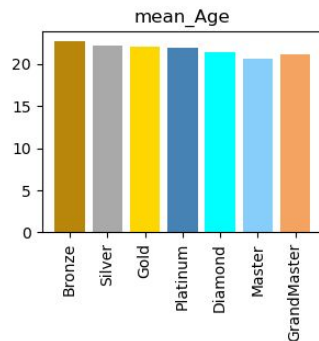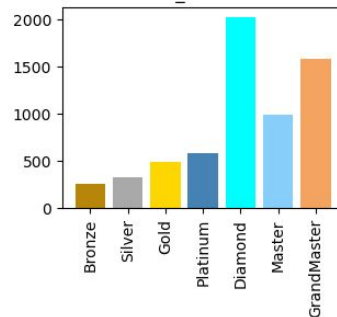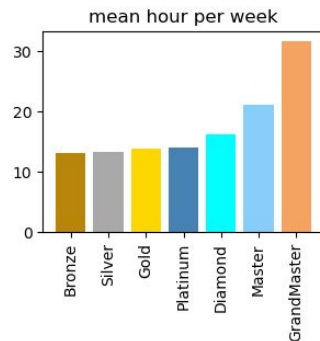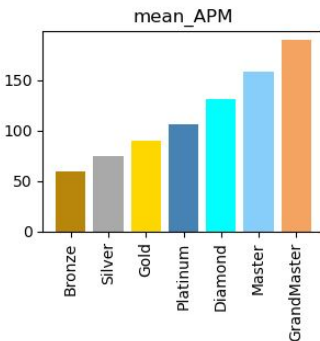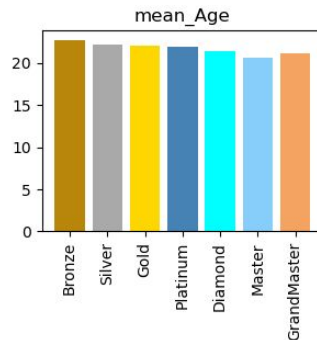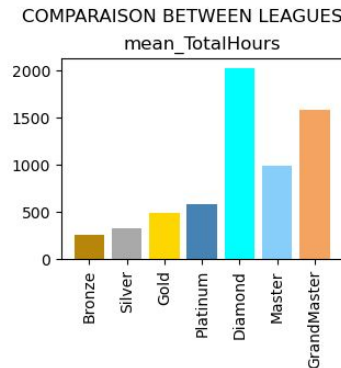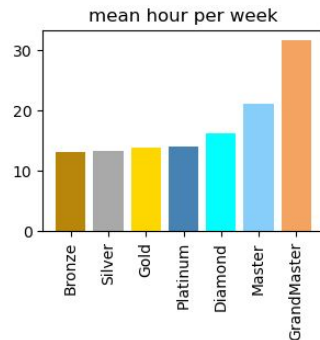
Comparison between Leagues

- Age doesn't seem to have a big influence on performance
- For the last 6 statistics concerning in game mechanics it seems that rank is correlated to the capacity to perform in this mechanics
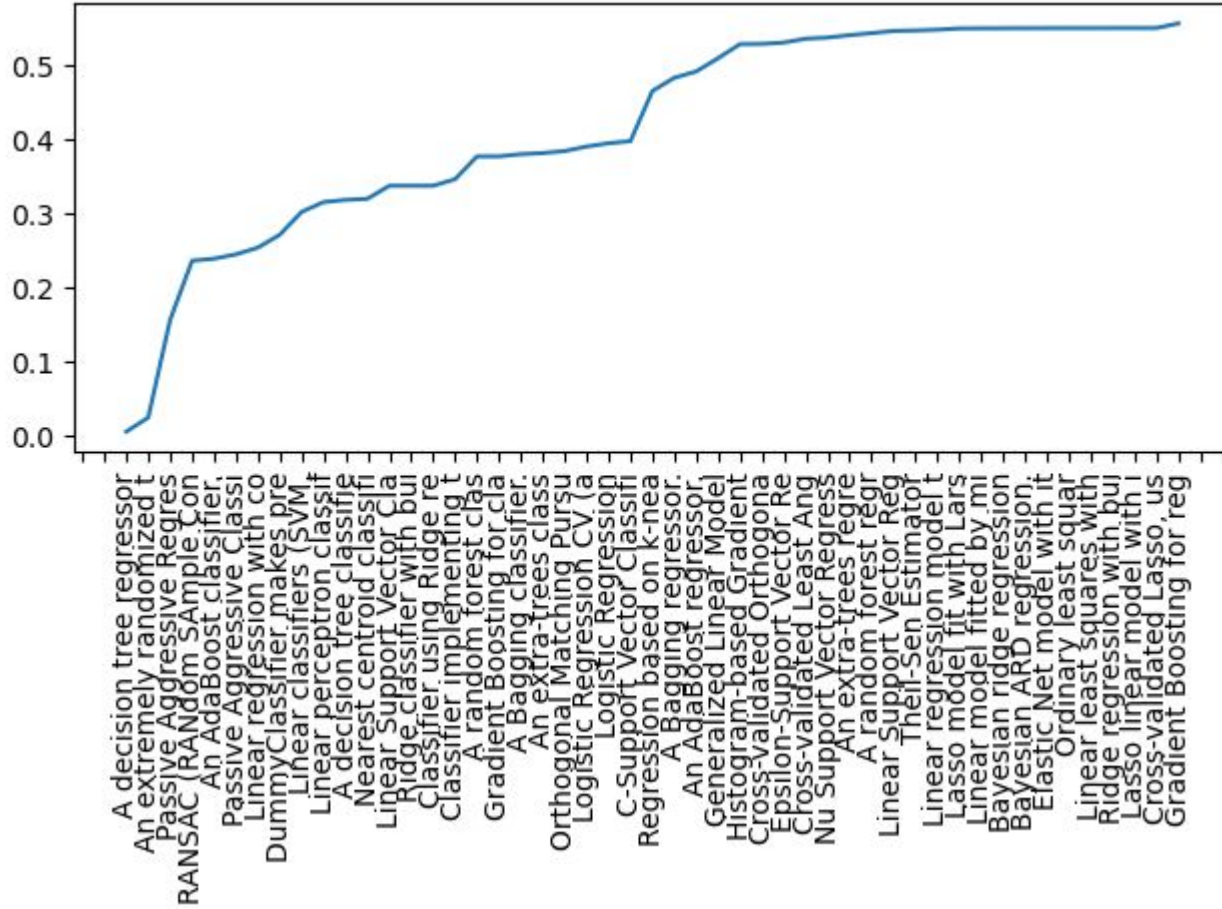

COMPARAISON BETWEEN LEAGUES
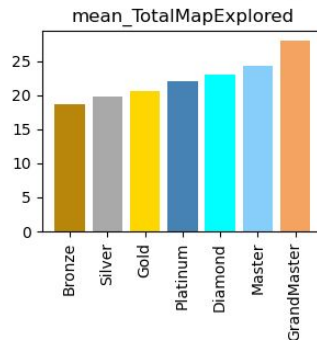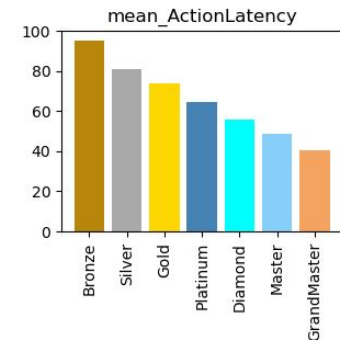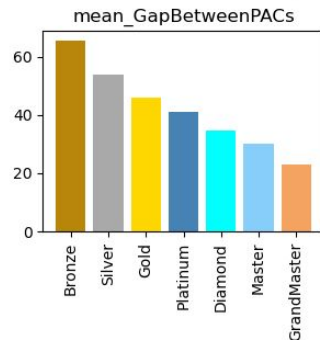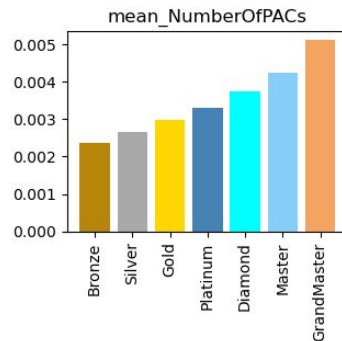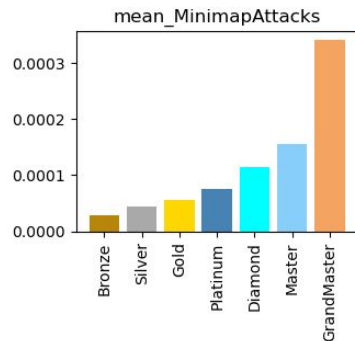
Best models for our use case

	performance
Gradient Boosting for reg	0.557474
Cross-validated Lasso, us	0.551294
Lasso linear model with i	0.551287
Ridge regression with bui	0.551254
Linear least squares with	0.551190
Ordinary least squar	0.551165
Elastic Net model with it	0.551115
Bayesian ARD regression.	0.551075
Bayesian ridge regression	0.550949
Linear model fitted by mi	0.550878
Lasso model fit with Lars	0.550677

Model Comparison of all sklearn models

Streamlit

Démo