PYTHON PROJECT STAR CRAFT 2

SILVA ANAELLE AND VACHER QUENTIN

OVERVIEW

1. The data set
2. Statistic description
3. Modeling
4. Conclusion

THE DATASET

Dataset by Mark Blair, Joe Thompson, Andrew Henrey and Bille Chen created in September 2013.

This dataset concerns the game StarCraft 2. It lists 20 metrics for around 3400 players about different informations about the way they play.

StarCraft is a game used a lot in reinforcement learning because it require a lot of micromanagement processing. This dataset could be interesting in this way to train different reinforcement learning algorithm (which we will not see here but this data set probably has been create in this way).

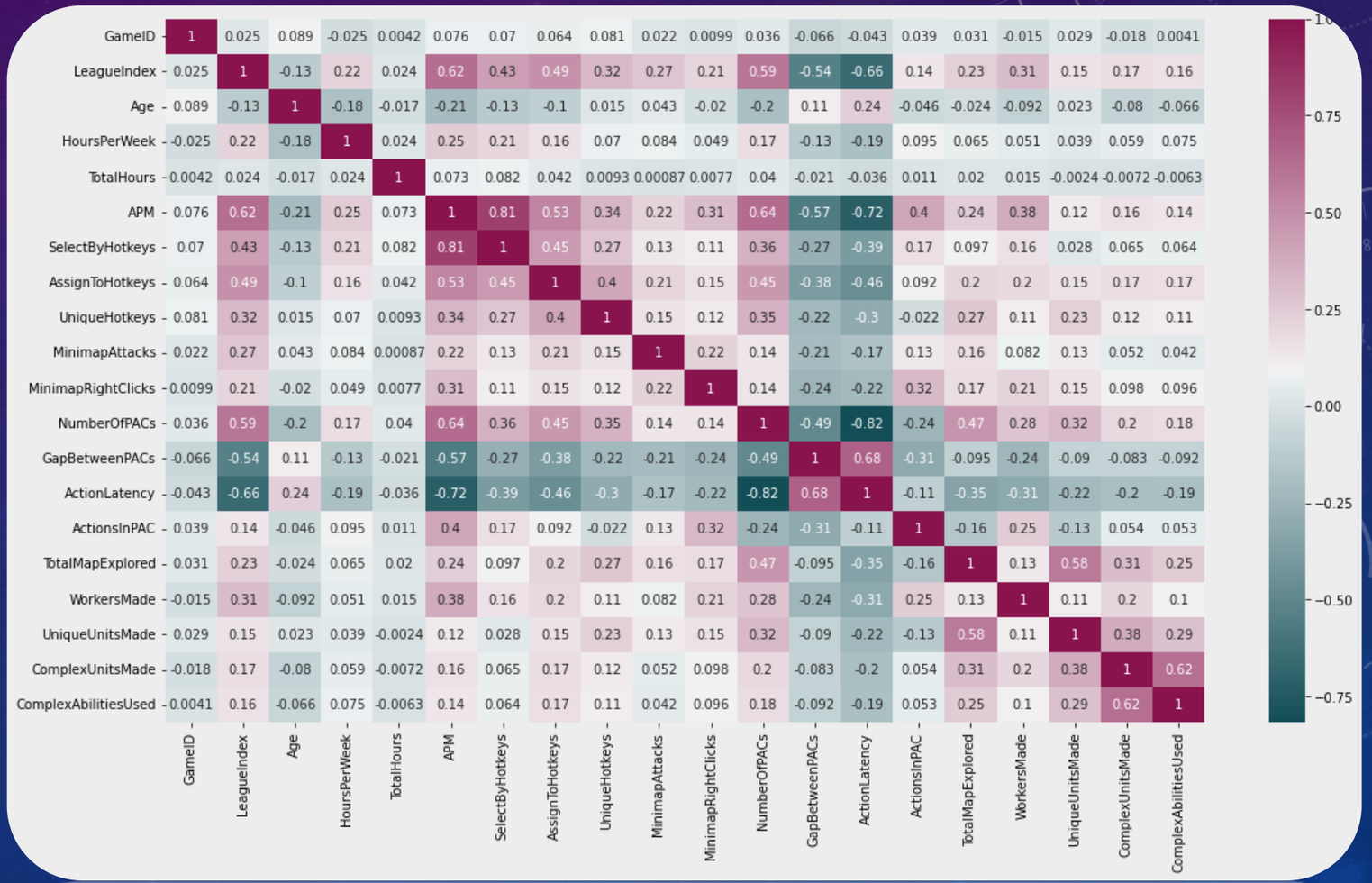
The dataset contains the LeagueIndex of the players, this will be our research value. Depending on the other attribute the objective will be to find the LeagueIndex of the player.

STATISTIC DESCRIPTION

Our goal is to found variables correlate with the League Index

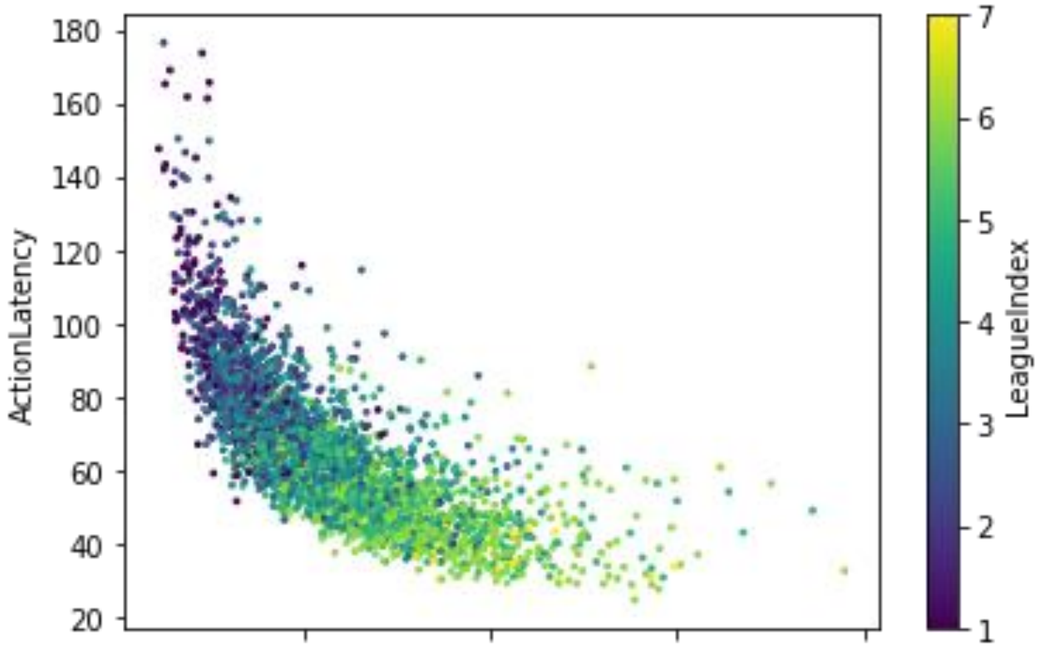
Depending of this correlation matrix, we will chose only several variables :

- APM
- SelectedByHotkeys
- AssignToHotkeys
- NumberOfPACs
- GapBetweenPACs
- ActionLatency



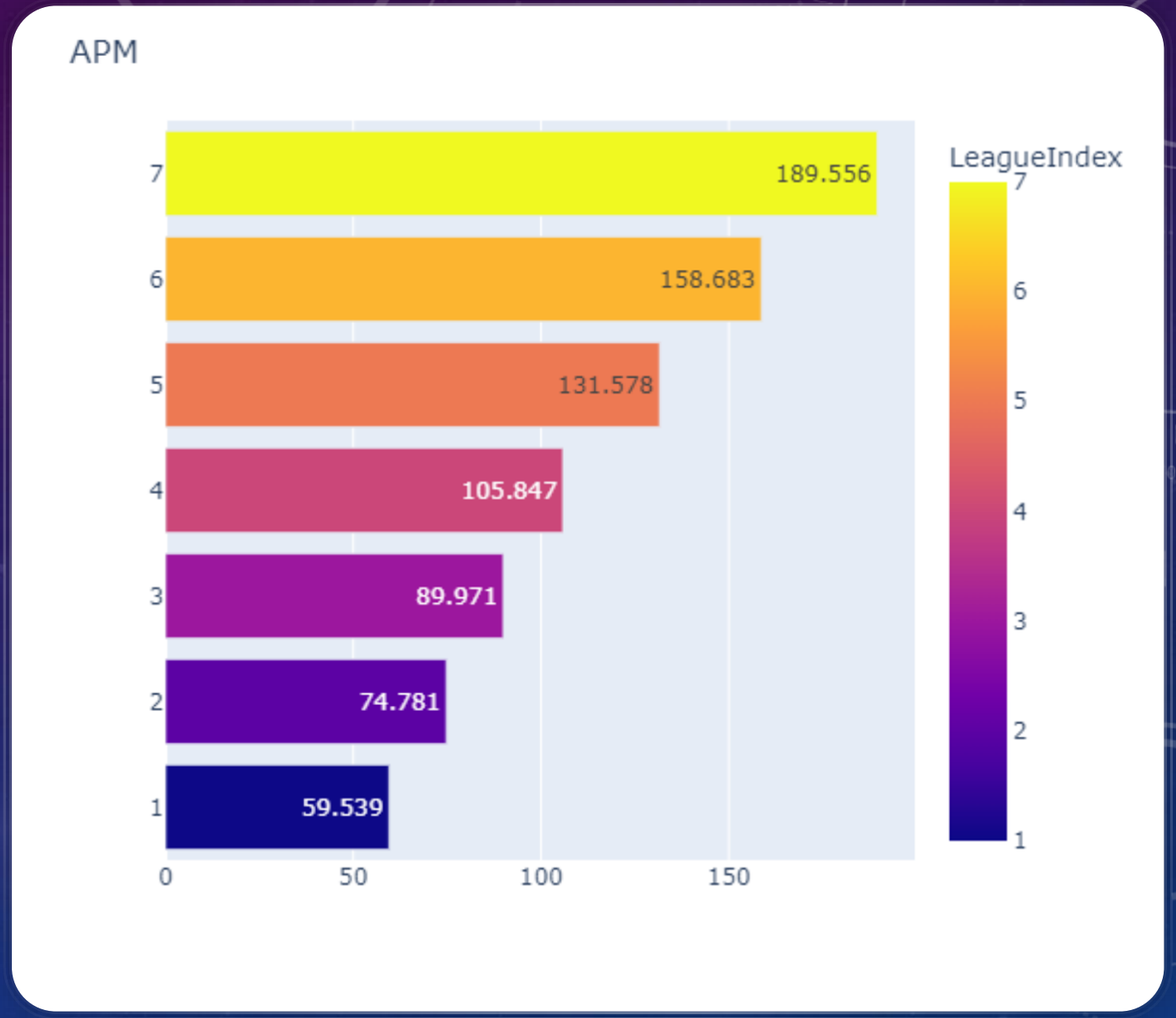
STATISTIC DESCRIPTION

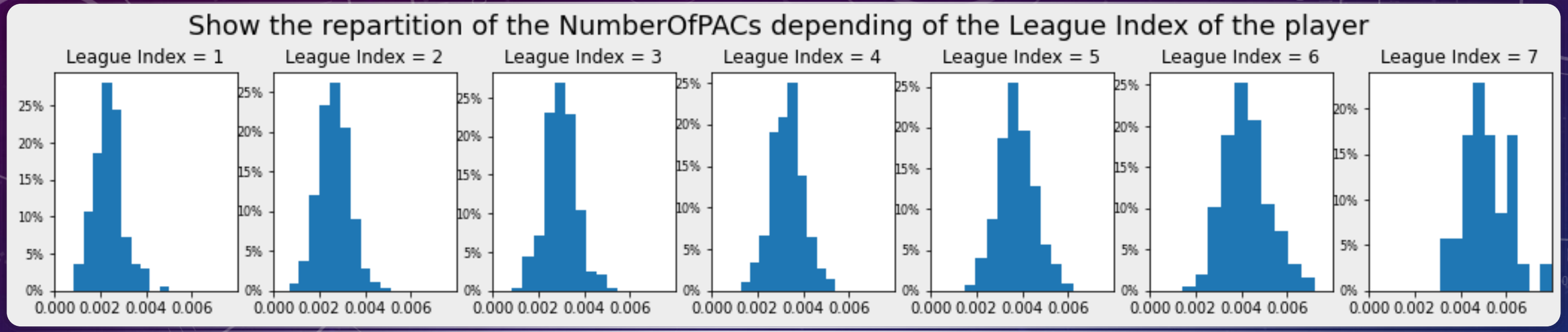
THE ACTIONLATENCY DEPENDING ON THE APM WHERE THE COLOR'S DOT DEPENDS ON THE LEAGUEINDEX



STATISTIC DESCRIPTION

THE APM DEPENDING ON THE LEAGUE INDEX





STATISTIC DESCRIPTION

EACH HISTOGRAM REPRESENT THE REPARTITION OF NUMBER OF PACS FOR ONE LEAGUE INDEX

MODELING

- Linear Support Vector Cassification (LSVC) : 33.83 %
- Support Vector Cassification (SVC) : 35.18 %
- Random Forest : 36.23 %
- Naive Bayes : 37.13 %
- Extra Trees Classifier with great search: 47,37 %

CONCLUSION

With a maximum accuracy of 47% we can not say that the research was a success. The great search allows us to find a way bigger accuracy.

One of the reasons is the poor quantity of data. There are only around 3 thousands data which is not a lot.

Another reason is that the different variables correlate with the LeagueIndex were also correlate a lot with each other.

However, this project was interesting because it allows us to test different models and algorithms on a true database.