

RAPPORT DE PROJET ANALYSE EXPLORATOIRE DES DONNÉES

Sujet : Découverte du monde de StarCraft

Le 6 Décembre, 2019



Réalisé par : TRAN Trong Hieu - TRAN Le Minh

Group A1 - 5 SDBD - Département GEI - INSA Toulouse, FRANCE

Sommaire

1. *Introduction du jeux de données*
 2. *Phase d'exploration des données*
 3. *Conclusion*
 4. *Annexe*
-

I. Introduction

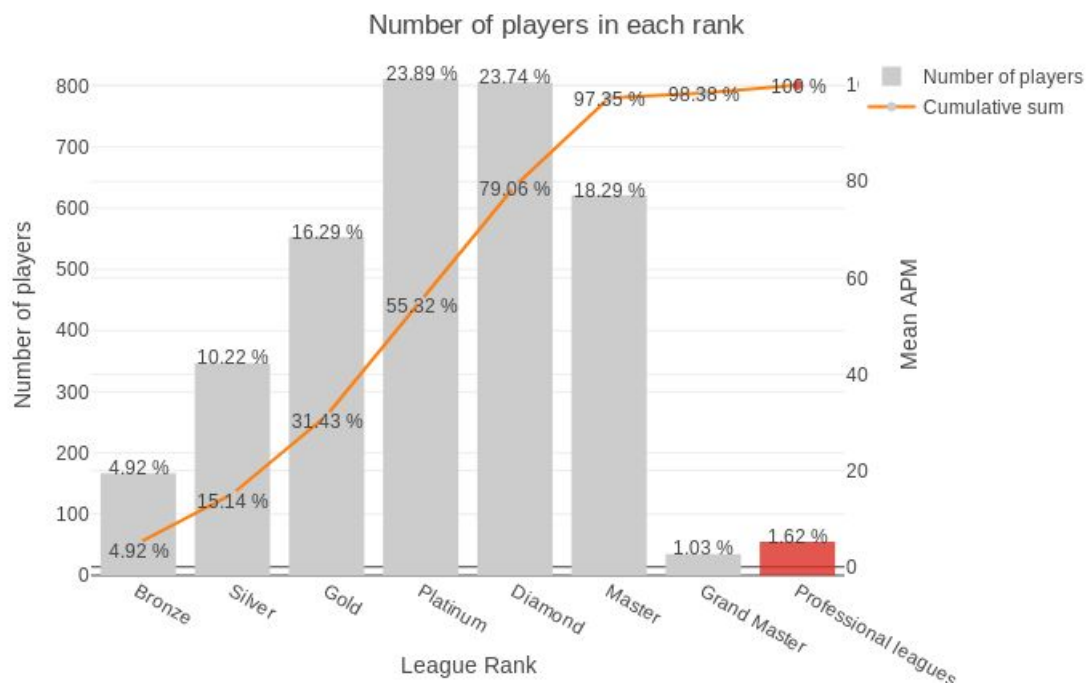
Dans l'univers du eSport, il existe un jeu qui non seulement affirme sa classe depuis longtemps en recueillant le plus de réussite, mais encore incite l'équipe DeepMind de Google, après le lancement de l'agent IA AlphaGo qui a vaincu le championnat mondial du jeu de Go, à enchaîner leur succès en créant un autre agent baptisé AlphaStar pour conquérir ce jeu. Il s'appelle : **Starcraft 2**.

Dans le cadre du TP Analyse Exploratoire, nous allons explorer le jeu de données collecté par *Joe Thompson*, *Andrew Henrey* et *Bill Chen* en Septembre 2013, contenant des données sur 3395 joueurs de différents niveaux de compétences, pour découvrir le monde de ce chef-d'oeuvre stratégique en temps réel. Ce jeu de données est disponible en open-source sur le site **UCI Machine Learning Repertory**¹. Les différents attributs de chaque joueurs, correspondant à des colonnes du jeu de données, seront éventuellement présentés au fil de l'exploration des données. Il y aura des données mesurées par timestamp, avec timestamp l'unité de temps du jeu correspondant à 1/88 seconde en temps réel.

II. Phase d'exploration des données

1. Comment se distribuent les joueurs en fonction du niveau de compétence ?

Dans l'univers de Starcraft 2, le classement des joueurs se divise en 8 ligues qui se rangent de "Bronze" pour les joueurs du classement le plus bas à "Professional Leagues" pour les joueurs professionnels.



¹ Source de données :

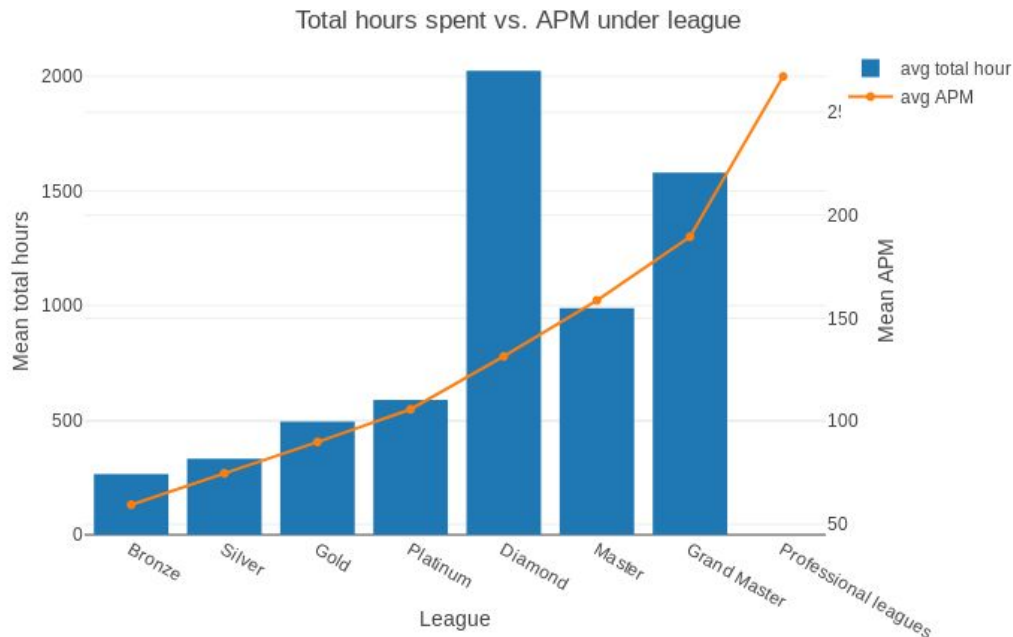
<https://archive.ics.uci.edu/ml/datasets/SkillCraft1+Master+Table+Dataset>

Graphique 1: Distribution des joueurs en fonction de ligue

Donc le premier problème auquel nous nous intéressons est de savoir où se situe l'ensemble des joueurs dans ce système de classement. Grâce à **Graphique 1**, on peut déduire que le classement moyen des joueurs est la ligue Platinum en se basant sur le fait que la majorité des joueurs se trouvent dans cette ligue, et de plus elle représente aussi la centile 50% sur l'ensemble joueur. D'ailleurs, le taux des joueurs qui atteignent "Grandmaster" et qui se lancent dans la ligue des pros est très faible, juste respectivement 1.03% et 1.62%. C'est en effet le chiffre prévu concernant la ligue d'élite.

2. Quel est la variation de la performance des joueurs selon chaque niveau de compétence ?

Pour des jeux stratégiques en temps réel tels que *Starcraft II*, il existe un indice crucial qui peut révéler la performance d'un joueur, c'est "Action Per Minute" (APM), ce qui signifie le nombre d'action qu'une personne peut effectuer sur le clavier et la souris pendant une minute. Comme ce jeu concerne une forte multitâche : de la gestion d'économie au contrôle de l'armée, donc une bonne maîtrise de ce jeu équivaut à un haut niveau d'APM.



Graphique 2: Temps de jeux moyen & AMP moyen par ligue

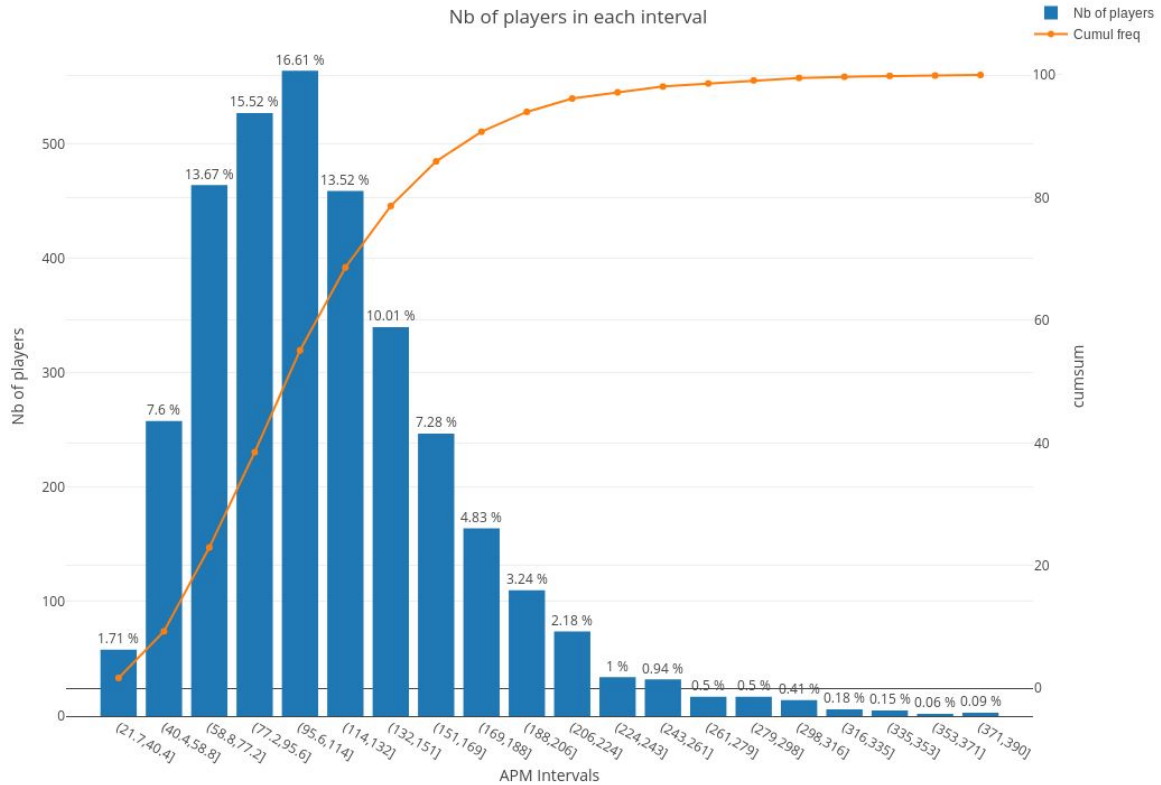
En effet, **Graphique 2** nous montre tout d'abord que l'APM augmente proportionnellement en fonction des ligues, où on observe un saut considérable de l'APM entre la ligue 'Grandmaster' et la ligue professionnelle, ce qui s'explique par le fait que les tournois sont beaucoup plus exigeants au niveau de compétence personnelle que des matchs occasionnels.

Un autre indice qui capte notre intérêt dans **Graphique 2**, c'est le temps total de jeu moyen par ligue, car il peut montrer en général il exige en moyenne combien de temps pour atteindre la prochaine ligue. Donc selon ce graphique, tout d'abord il ne prend pas beaucoup de temps pour aller de la ligue "Bronze" à la ligue "Diamond", mais il ne semble pas facile de monter à la ligue "Master" car il exige jusqu'à plus de 2000 heures de jeu en moyen, ce qui est beaucoup plus large par rapport au temps de jeu moyen des autres

ligues. Ceci suggère que cette ligue pourrait être une frontière qui sépare les joueurs plutôt occasionnels et les joueurs élités dont un certain ont même envie de s'orienter vers la scène professionnelle. Comme on ne compte pas le temps de jeu des joueurs pro, "Grandmaster" étant la ligue la plus élitée, son temps de jeu moyen reste haut et continue à augmenter.

3. Comment se trouvent les joueurs en fonction de l'APM ?

Maintenant que l'on a identifié l'APM comme un indice crucial concernant la performance d'un joueur, on va s'intéresser à la distribution des joueurs en fonction de l'APM.

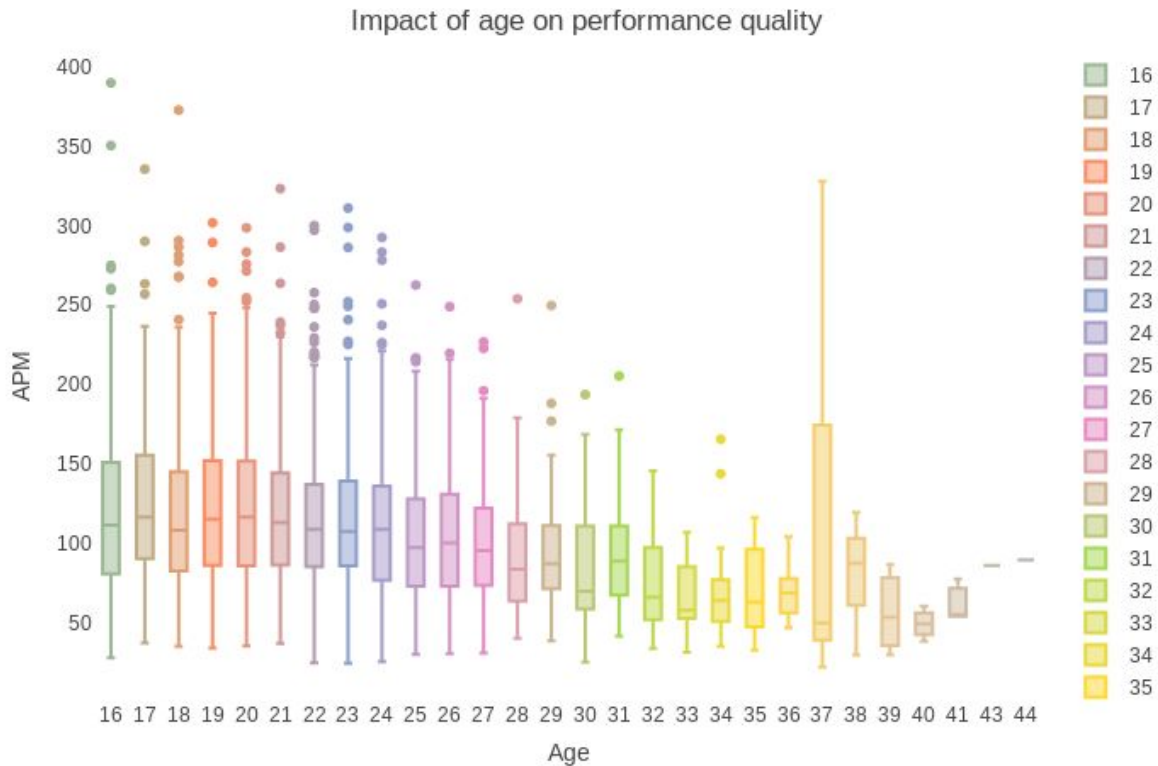


Graphique 3: Distribution des joueurs en fonction de l'APM

Graphique 3 montre que cette distribution se concentre plutôt à gauche. La plupart des joueurs ont un APM entre 58.8 à 151, comprenant 69.33% des joueurs. Avec le centile 50% qui se trouve au niveau d'APM entre 95.6 et 114, il s'agit donc la capacité personnelle moyenne de l'ensemble des joueurs. Comme prévu, il y a très peu de joueurs qui possèdent un extrêmement haut APM : les joueurs qui peuvent effectuer au moins 4 actions chaque secondes se trouvent dans le top 1.88% des joueurs avec le plus grand APM. Conseils pour les joueurs : pour avoir une différence dans la performance par rapport aux joueurs moyens, ie. pour atteindre le top 20% APM, entraînez-vous à effectuer au moins 3 actions par seconde !

4. Est-ce que l'âge a un impact sur la performance des joueurs ?

Concernant les données des joueurs, leurs âges sont aussi collectés. Donc dans cette partie, nous allons étudier l'impact de l'âge sur la performance en comparant l'APM en fonction de l'âge des joueurs.

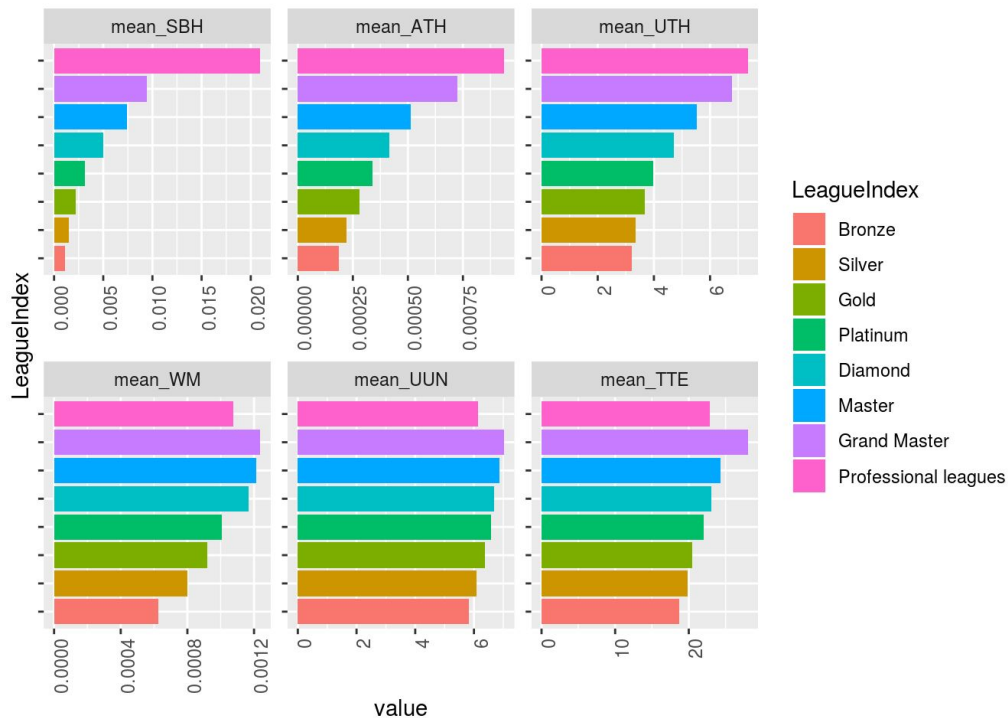


Graphique 4: Distribution des l'APM par rapport à l'âge

A travers **Graphique 4**, on trouve tout d'abord que généralement l'APM maximal et l'APM médiane ont tendance à diminuer quand les joueurs se vieillissent. De plus, à chaque âge on trouve qu'il existe des "anomalies" - les points au dessus de chaque "BoxPlot" qui montre la performance extraordinaire et qui apparaissent de manière remarquable chez les jeunes joueurs. Avec une distribution de l'APM assez similaires, cette phénomène chez les jeunes joueurs de 16 à 25 ans est expliquée par le fait que c'est l'âge où une personne exprime sa meilleure santé physique et mentale, ainsi que capacité cognitive. Par ailleurs, chez les joueurs plus âgés, ils montrent sur ce graphique les informations qui semble non conforme à la tendance mentionnée précédemment. En réalité, dans notre espace d'échantillon, le nombre de joueurs âgés ne sont pas très nombreux, et l'existence des joueurs exceptionnellement compétents a influé sur la distribution générale de l'APM dans cette tranche d'âge.

5. Quelles sont les différences dans le style de jeu de différentes ligues ?

Pour que les joueurs puissent atteindre le meilleur classement, l'amélioration de la capacité personnelle telle que l'APM est nécessaire mais pas indispensable, car il y a aussi des autres techniques à prendre en compte. Dans cette partie, on va comparer, à travers 8 ligues, ces trois techniques suivantes concernant la maîtrise du clavier : Nombre d'unité choisis par raccourci clavier par timestamp (SBH), nombre d'unité affectés à un raccourci clavier par timestamp (ATH) et nombre de raccourcis clavier uniques utilisés par timestamp (UTH); et ces trois aspects stratégiques du jeu : la superficie de la carte du jeu explorée par timestamp (TTE), nombre de travailleurs créés par timestamp (WE) et nombre d'unité unique créées par timestamp (UUT).



Graphique 5: Comparaison du style de jeu en fonction de ligue

Graphique 5 montre que la valeur de chaque indice augmente proportionnellement en fonction du niveau de classement. Les joueurs professionnels, sans aucun doute, possèdent des compétences incomparables par rapport au reste du monde au niveau de la maîtrise du clavier. Par contre, il semble qu'ils n'atteignent pas le même niveau stratégique vis-à-vis des autres joueurs de haut classement. En réalité, dans le cadre des compétitions professionnelles, on porte une forte priorité sur les stratégies optimales qui consistent en une durée d'un match la plus courte possible (donc non nécessité d'explorer excessivement la carte), en création suffisante des travailleurs responsables des constructions et des récupérations des ressources économiques et en concentration sur un certain type d'unités spécifiques convenable à la stratégie choisie.

En bref, à travers ce graphique, nous avons observé la différence de style de jeu entre d'une part les joueurs de différentes ligues et d'autre part les joueurs amateurs et les joueurs professionnels. Un conseil pour les joueurs de bas classement : il vaut mieux renseigner le style de jeu des meilleurs joueurs si vous voulez améliorer votre rang !

III. Conclusion

A travers l'exploration du jeu des données sur **Starcraft 2**, nous avons découvert plusieurs aspects du monde des joueurs de cette discipline eSport stratégique en temps réel. D'une part, nous avons renseigné sur la différence en terme de performance et de style de jeu entre les joueurs de différents classements. D'autre part, nous en avons tiré aussi des conseils, que nous espérons bien qu'ils seront utiles pour les joueurs qui souhaitent améliorer leur qualité de jeu !

Annexe :

Vous trouverez dans notre répertoire *Github* dont le fichier *Starcraft EDA.html* qui contient nos graphiques en mode interactif en utilisant la librairie **Plotly en R**, ainsi que les autres graphiques intéressants que nous avons exploré pendant notre exploration du jeu de données.

Lien vers notre github : <https://github.com/kuro10/Exploratory-Data-Analysis>,