

Projet 705

Analyse de données d'un serveur Minecraft Événementiel

GARRAT Axelle
GAY Nicolas
COLLIGNON Alexandre

M1 Informatique
Parcours IA
2023-2024

Sommaire

I - Les données.....	2
II - La problématique.....	3
III - Choix du modèle et travail réalisé.....	3
Etape 1 - Chargement et transformation des données.....	3
Etape 2 - ACP.....	4
Etape 3 - AFDM.....	4
IV - Résultats de l'analyse et interprétation.....	5
Graphique des « squared loadings ».....	5
Projection des individus et des variables qualitatives.....	6
Cercle des corrélations.....	7
V - Conclusion.....	8
VI - Synthèses.....	8

I - Les données

Notre groupe a pris la décision de se pencher sur l'analyse des données relatives aux joueurs présents sur un serveur de jeu Minecraft.

En effet, nous avons récupéré des données d'une base de données relative à un serveur Minecraft dont le propriétaire est Alexandre, qui fait partie de notre groupe. Les données couvrent la période du 20 avril 2023 au 12 janvier 2024.

Pour le contexte, les joueurs sont majoritairement invités par d'anciens joueurs afin de participer à une saison inspirée du jeu télévisé Koh-Lanta et sa variante américaine Survivor, ici nommé "Survisland". Par conséquent, la plupart des joueurs ne jouent qu'une seule saison avant de partir.

Notre dataset contient pour chaque joueur les variables statistiques suivantes :

- **name** : Le pseudonyme du joueur. (Non pertinent dans notre analyse)
- **registered** : Timestamp* correspondant à la première connexion.
- **times_kicked** : Nombre d'expulsions du joueur.
- **nb_player_kills** : Nombre de joueurs tués.
- **nb_mob_kills** : Nombre de créatures tuées.
- **nb_deaths** : Nombre de morts.
- **nb_sessions** : Nombre de connexions au serveur.
- **longest_session** : Timestamp* indiquant le temps de la session la plus longue passée.
- **avg_session** : Timestamp indiquant la durée de session moyenne.
- **active_playtime** : Timestamp correspondant au temps d'activité.
- **afk_time** : Timestamp correspondant au temps d'inactivité.
- **favorite_world** : Nom du monde où le joueur passe le plus de temps.
- **geolocalisation** : Pays dans lequel le joueur se trouve.

Pour résumer et compléter les informations, nous avons un total de 13 variables statistiques dont 10 quantitatives et 2 qualitatives pour un total de 202 individus.

II - La problématique

Notre analyse vise à approfondir notre compréhension des interactions et des comportements des utilisateurs au sein de cet environnement virtuel particulier. En examinant de près les données générées par les joueurs, nous cherchons à identifier des tendances, des modèles de jeu et d'autres aspects significatifs, comme par exemple quelles variables influencent l'intérêt d'un joueur pour le serveur.

III - Choix du modèle et travail réalisé

Nos données contenant des variables qualitatives et des colonnes quantitatives, nous avons décidé d'utiliser l'analyse Factorielle de Données Mixtes (AFDM).

Pour cela, notre groupe a décidé d'adapter le script R pour un AFDM, fourni lors du cours, en se servant du logiciel RStudio.

Les étapes de notre analyse sont donc les suivantes.

Etape 1 - Chargement et transformation des données

La première étape consiste à charger les données contenues dans notre fichier CSV, préalablement collecté grâce à une requête SQL.

Il faut ensuite réaliser le centrage-réduction des variables quantitatives afin de standardiser les variables en les ramenant à une échelle commune.

En ce qui concerne les variables qualitatives, on applique un codage disjonctif complet pour les transformer en variables indicatrices, ceci est une méthode couramment utilisée pour représenter les variables qualitatives nominales dans les analyses multivariées.

Par la suite, on réalise une pondération des indicatrices permettant ainsi de garantir une contribution équilibrée des variables pour des réponses avec davantage de précision.

Etape 2 - ACP

La deuxième étape est l'Analyse en Composantes Principales (ACP). Elle permet de représenter les variables quantitatives et qualitatives codées en indicatrices sous forme de projections sur un espace de dimensions inférieures.

Afin de réaliser cette analyse, on concatène les différentes variables préparées à l'étape précédente. Puis, on les utilise avec la fonction **dudi.pca** de la librairie **ade4**, qui effectue une ACP sur un tableau de données.

Nous obtenons ainsi les coordonnées ACP des variables et des individus.

Etape 3 - AFDM

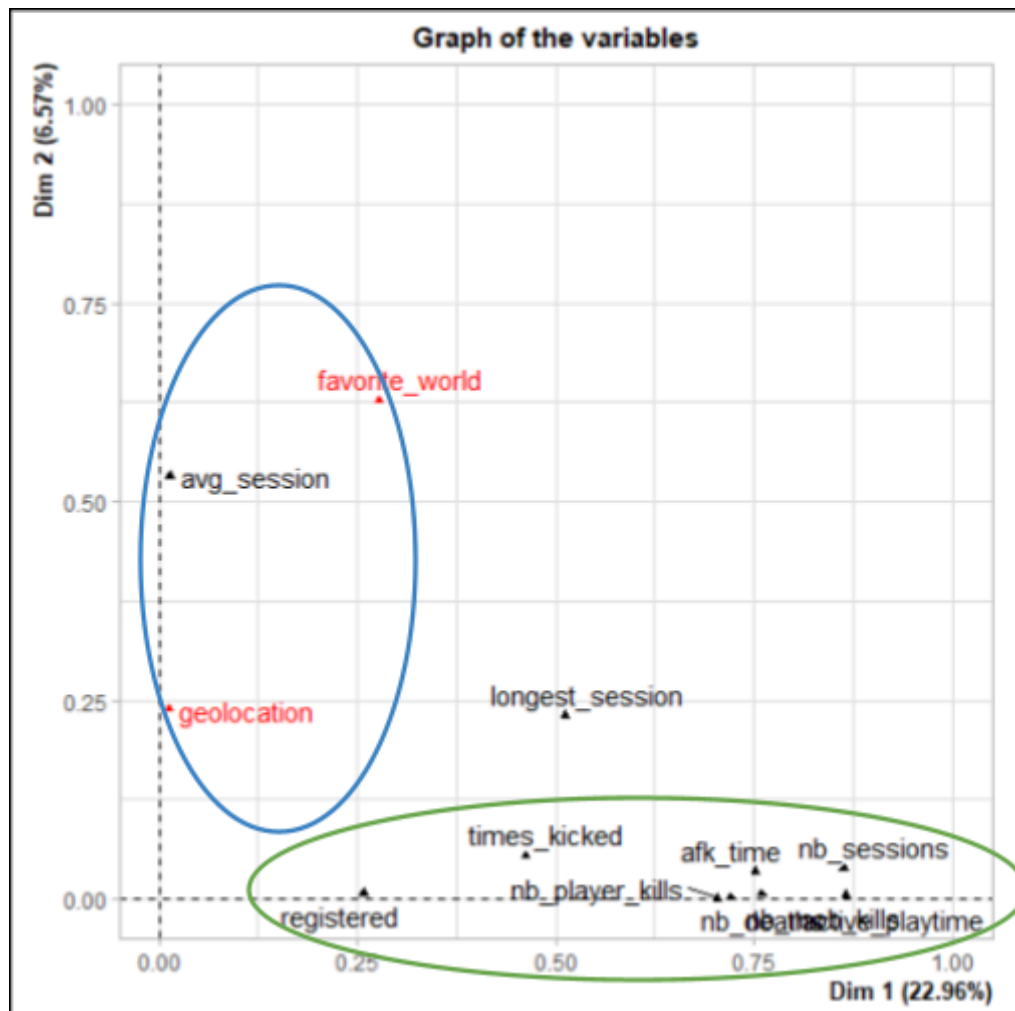
Cette troisième étape consiste en la réalisation de l'analyse factorielle de données mixtes (AFDM) permettant d'analyser un ensemble de données de variables quantitatives et qualitatives. Fonctionnant de la manière d'un ACP pour les variables quantitatives et comme un ACM pour les variables qualitatives.

Pour ce faire on peut utiliser plusieurs librairies comme **FactoMineR**, **ad4** ou encore **PCAmixdata**. Nous avons décidé d'utiliser FactoMineR car nous sommes plus familier avec cette librairie que les autres et avons trouvés plus pertinent les résultats obtenus dans la section suivante de ce rapport.

IV - Résultats de l'analyse et interprétation

Après exécution du code, nous obtenons quatres graphiques qui sont interprétés ci dessous.

Graphique des « squared loadings »



Ce graphique permet de visualiser l'impact des variables sur les deux facteurs, dim1 et dim2.

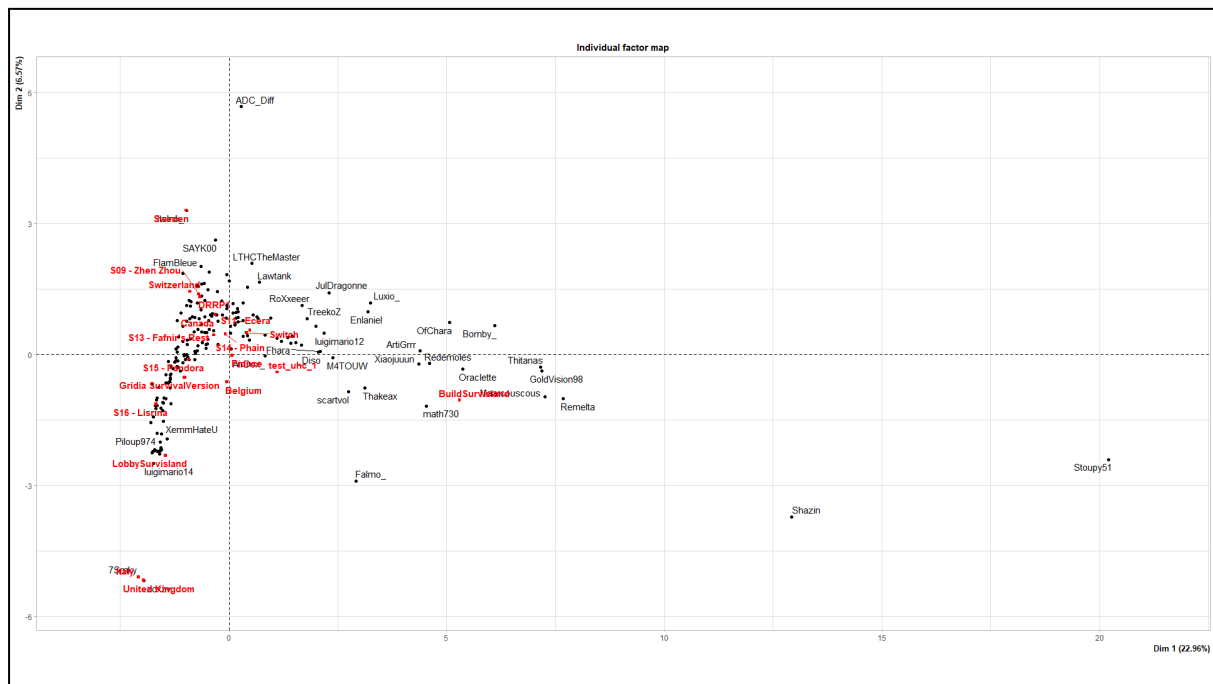
On remarque que les variables forment des groupes distincts.

Les variables du monde favori, le temps de session moyen, et la geolocation sont rattachées au deuxième facteur (dim2).

Tandis que le reste des variables influent sur le premier facteur (dim1).

Cependant, la variable quantitative de la plus longue session semble rattachée aux deux facteurs.

Projection des individus et des variables qualitatives



Ce graphique permet de visualiser les joueurs dans un espace de dimension 2 en fonction de leur corrélation avec les différents facteurs (dim1 et dim2). Nous pouvons remarquer, grâce aux variables qualitatives en rouge, que le monde favori des joueurs est axé majoritairement sur l'axe vertical.

Ensuite, on note une séparation en groupe des joueurs.

Vers la gauche du graphique, on retrouve les joueurs seulement présents pour participer à des saisons Survisland, une fois la saison terminée ils s'en vont.

En haut à droite, on a les joueurs qui reviennent sur le serveur pour jouer à des mini-jeux sur le monde "Switch".

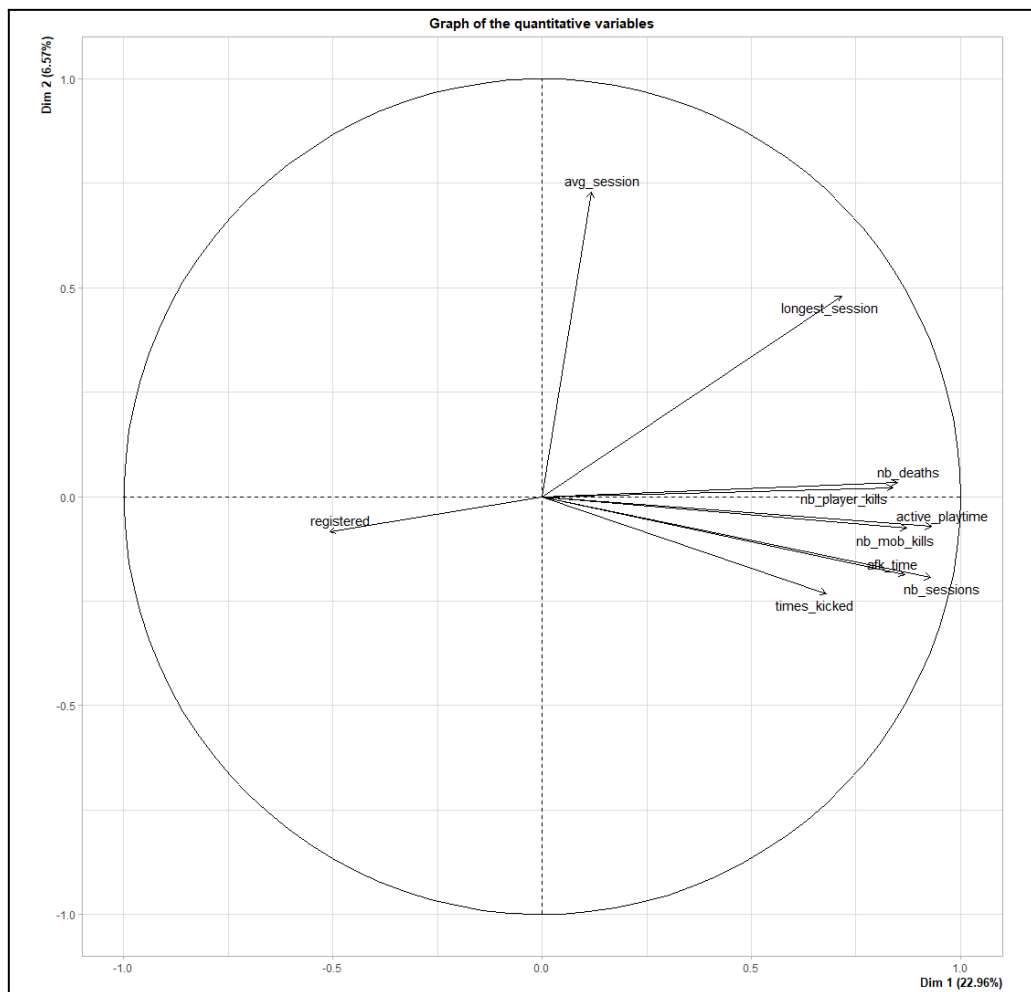
Ensuite, à droite vers le milieu et le bas du graphique, on repère les joueurs qui agissent en tant que staff.

Et enfin, en bas à gauche, on a des joueurs isolés, qui représentent des personnes ne s'étant connectées qu'une seule fois ou venant d'un pays peu représenté.

On remarque que ces groupes se différencient par leur niveau d'activité (active_playtime, nb_sessions, ...) et leur monde favori. Par exemple, les staffs sont plus actifs, donc vers la droite, et ont des mondes favoris spécifiques par rapport à d'autres joueurs, d'où leur regroupement.

Ces observations correspondent au graphique des « squared loadings ».

Cercle des corrélations



Nous pouvons voir dans ce graphique des variables quantitatives la corrélation entre la coordonnée dans chacune des dimensions pour les individus. On remarque que plusieurs variables sont intimement liées à l'activité tel que le nombre de morts, de sessions, et de monstres ou joueurs tués. Ces variables influent beaucoup sur le premier facteur (la coordonnée dans la 1ère dimension) et cela se vérifie avec le joueur le plus actif "Stoupy51" qui se trouve projeté au loin vers la droite comme on a pu le constater sur le graphique de projection des individus.

On remarque cependant que l'activité d'un joueur ne garantit pas que ses sessions de jeu soient longues, voilà pourquoi la variable du temps moyen de session agit beaucoup sur la deuxième dimension.

Pour rappel, la variable quantitative "registered" indique la date de la première connexion. Les nouveaux joueurs sont donc inversement proportionnels à l'activité étant donné que cette variable influe négativement sur le premier facteur.

V - Conclusion

En conclusion, notre analyse des données du serveur Minecraft a jeté une lumière intéressante sur les comportements des joueurs au sein de cet univers virtuel. En utilisant l'Analyse Factorielle de Données Mixtes (AFDM), nous avons pu explorer de manière approfondie à la fois les aspects quantitatifs et qualitatifs des données que nous possédions, dévoilant ainsi des tendances significatives.

Les différentes étapes de notre analyse, de la transformation des données à l'interprétation des résultats, ont permis de mettre en évidence des regroupements distincts parmi les joueurs, révélant des motifs d'engagement variés. Nous avons observé des corrélations intéressantes entre les différentes variables, mettant en lumière des facteurs clés tels que l'activité, les préférences de jeu et la fidélité des joueurs.

Cependant, il est important de souligner que notre analyse représente un instant T dans le temps et que chaque nouveauté pourrait faire évoluer les comportements des joueurs. Il serait donc bénéfique de poursuivre cette étude à des dates ultérieures, en examinant le changement des comportements des joueurs et en ajustant les analyses en conséquence si le besoin se faire ressentir.

Pour conclure, cette analyse nous a permis d'acquérir des statistiques intéressantes sur le serveur Minecraft, ainsi que de nous familiariser avec une technique d'analyse, qui est l'AFDM.

VI - Synthèses

Nous n'avons pas rencontré de problèmes particuliers, à part la difficulté de paralléliser les différentes étapes de l'analyse. Nous avons donc réalisé la majorité du travail en groupe, avec chacun un apport équivalent.