Détection d'un casino en ligne frauduleux

Table des matières

Table des figures

Liste des tableaux

1 Mise en contexte

On considère ici un jeu de données provenant d'un casino en ligne dans un pays qui a des restrictions sur les jeux d'argent en ligne (non spécifié). Le principe est le suivant : les joueurs peuvent parier des objets virtuels d'un jeu vidéo (des skins), ces skins ont une certaine valeur en euro (c'est donc comme s'ils parient de l'argent), et choisir un nombre entre 1 et $+\infty$. On tire alors une roulette fictive (la distribution des valeurs est inconnue) et il y a deux issues, si on est en dessous du nombre choisi on gagne sa mise mutlipliée par ce nombre et sinon on perd sa mise. On a un jeu de données de 55 575 parties, voici les 10 premières :

	Joueurs	Argent	Pari	Victoire	Defaite	Gains	Moderateur
2091103	144	283.57	14.30	125.46	0.25	408.78	0
2091104	134	279.30	1.14	5.81	177.04	108.07	0
2091105	139	282.87	3.91	125.01	5.96	401.92	0
2091106	139	271.44	1.15	6.90	181.85	96.49	0
2091107	142	304.88	1.00	0.00	304.88	0.00	0
2091108	161	463.81	2.46	118.96	15.45	567.32	0
2091109	157	342.79	1.45	36.50	119.43	259.86	0
2091110	160	302.06	2.73	93.70	25.94	369.82	0
2091111	155	1218.63	5.37	184.01	3.16	1399.48	0
2091112	129	236.01	2.09	64.04	12.02	288.03	0

Table 1 – *Jeu de données*

La colonne joueurs représente le nombre de joueurs durant une partie, la colonne argent représente l'argent misé dans cette partie, la colonne pari représente le résultat de la roulette durant cette partie, la colonne victoire représente la quantité de bénéfice durant cette partie (en ayant retiré la mise) , la colonne défaite représente la quantité d'argent perdu durant cette partie, la colonne gains représente la quantité d'argent gagné durant cette partie (sans avoir retiré la mise), la colonne modérateur représente s'il y a ou non un modérateur durant cette partie.

Tout d'abord, on peut tracer la distribution des valeurs de la colonne pari :

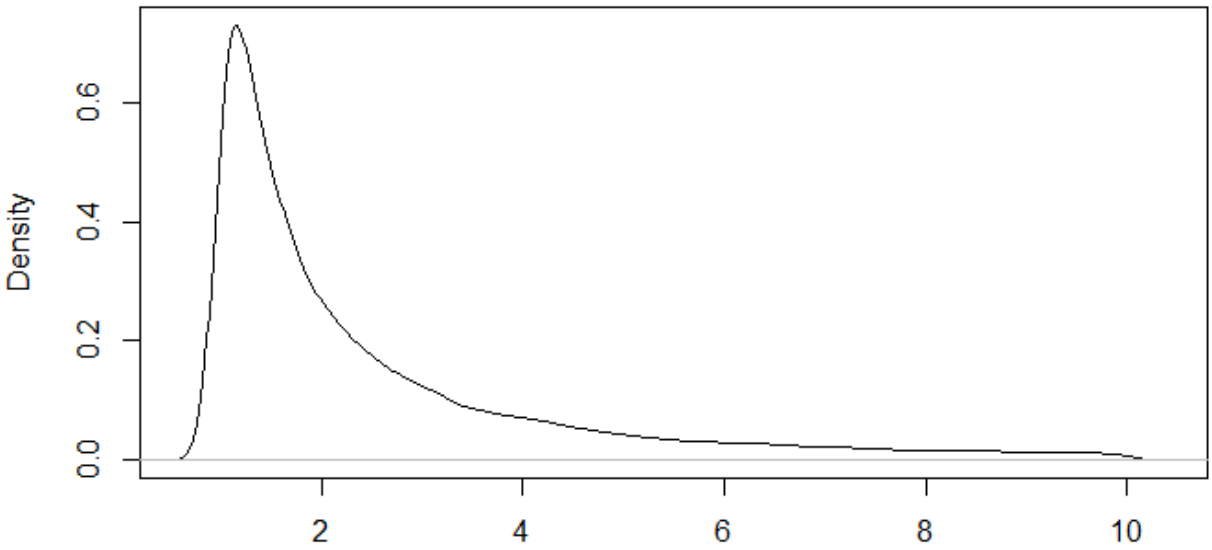

Figure 1 – *Distribution des valeurs de la colonne pari*

A première vue, on peut voir que la distribution est régulière, on peut donc penser naïvement que le casino n'est pas truqué. Cependant, nous allons voir au travers différentes analyses que c'est tout à fait l'inverse.

2 Analyse de données

2.1 Analyse en composantes principales

Tout d'abord, nous avons réalisé une analyse en composantes principales. Voici en premier lieu le graphe d'inertie des valeurs propres :

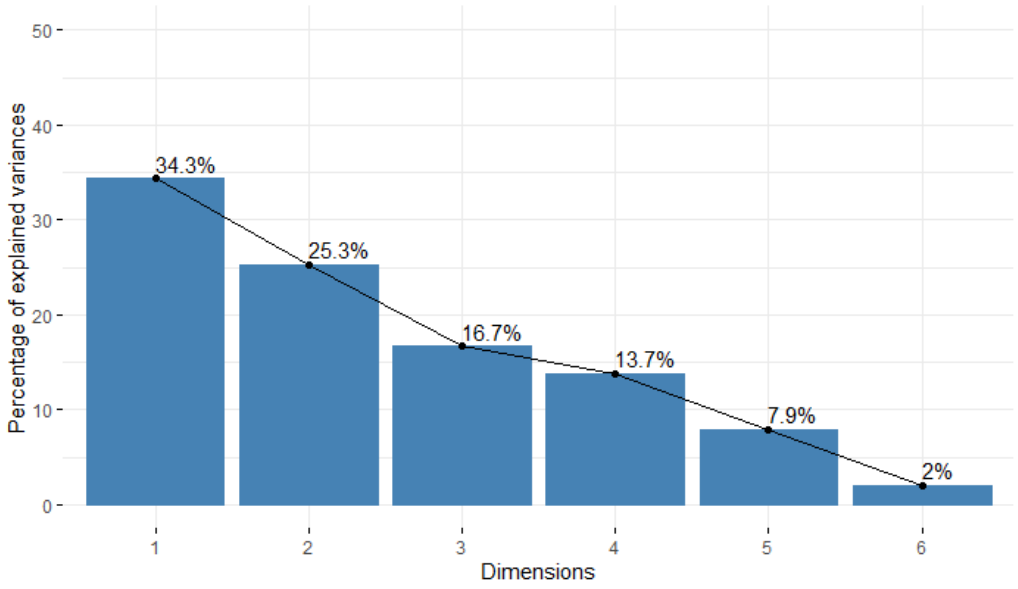

Figure 2 – *Graphe d'inertie des valeurs propres*

On peut voir que les deux premières composantes principales représentent à elles seules 60% de l'inertie totale, on va donc se contenter de les étudier. Voici à présent la carte factorielle couplé du cercle des corrélations :

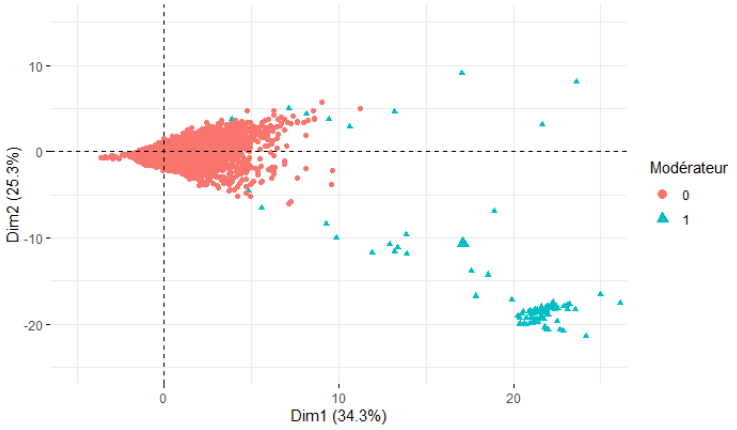

Figure 3 – *Carte factorielle*

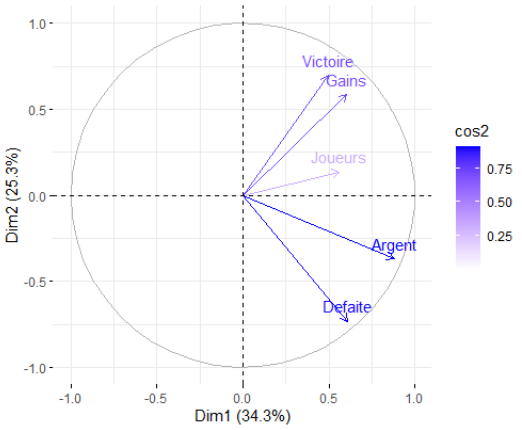

Figure 4 – *Cercle des corrélations*

Lorsqu'on représente les parties en fonction de si il y a un modérateur ou non, on remarque que les parties avec modérateur sont très proches les unes des autres, ce qui signifie que les parties avec modérateur sont très similaires entre elles. On remarque également que les parties avec modérateurs sont très corrélées aux colonnes argent et défaite, ce qui signifie que les parties avec modérateur sont très corrélées à la mise et à la perte. On peut donc penser que les parties avec modérateur sont des parties où les joueurs misent beaucoup d'argent et perdent beaucoup d'argent. Sûrement que les modérateurs misent beaucoup d'argent en étant sûrs de gagner et avec un score de roulette très faible ce qui expliquerait la corrélation avec la colonne défaite.

2.2 Analyse en clusters

Lorsqu'on réalise une analyse en clusters sur les deux premières composantes principales, on obtient la figure suivante :

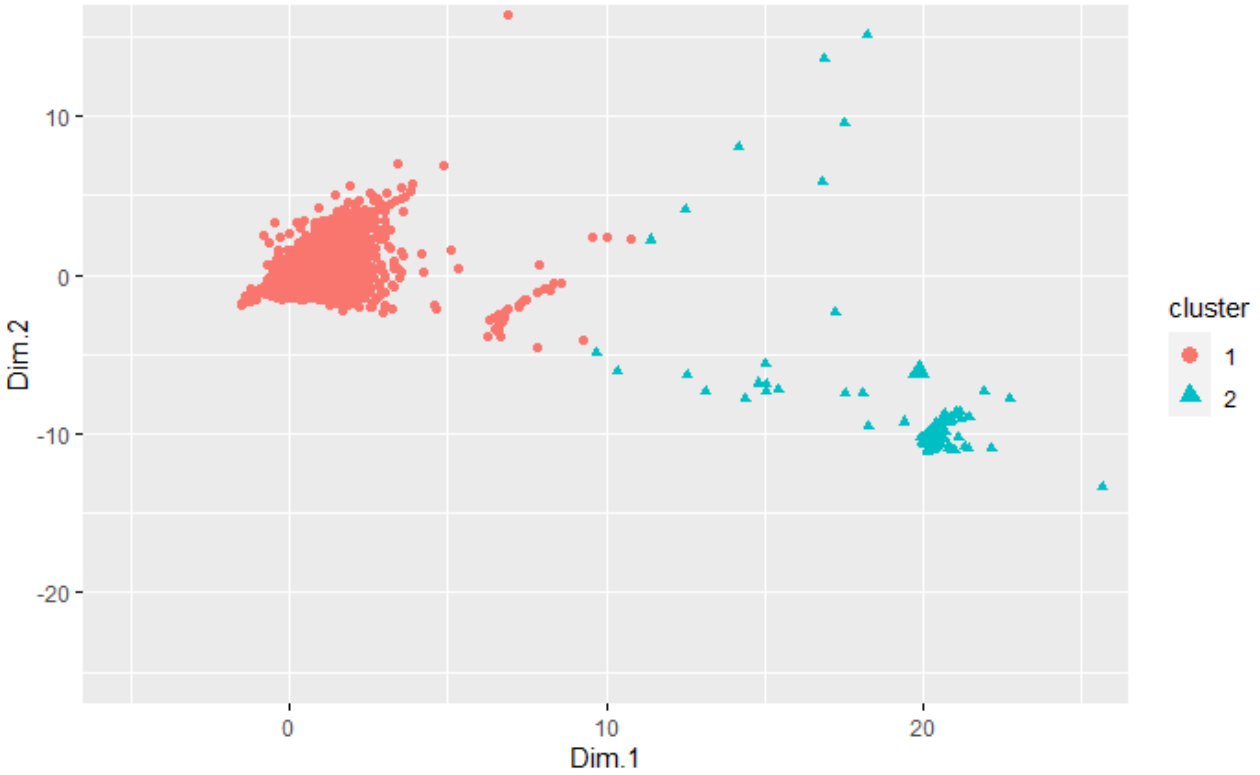

Figure 5 – *Analyse en clusters*

On remarque que les deux clusters trouvés coïncident avec la présence ou non d'un modérateur. On peut donc penser que les parties avec modérateur sont très différentes des parties sans modérateur.

2.3 Analyse discriminante

A présent, faisons une analyse discriminante sur le fait qu'il y ait ou non un modérateur. Voici le tableau des moyennes selon la présence d'un modérateur :

	Joueurs	Argent	Pari	Victoire	Defaite	Gains
0	123.01	277.69	12.05	65.80	72.76	271.07
1	131.90	3208.31	3.18	402.30	2692.85	784.45

Table 2 – *Moyennes selon la présence d'un modérateur*

On remarque que les moyennes sont très différentes selon la présence d'un modérateur. On confirme aussi nos suppositions quant à la partie 2.1, les parties avec modérateur sont des parties où les joueurs misent beaucoup d'argent et perdent beaucoup d'argent. On remarque également que les moyennes de la colonne pari sont très différentes, elle est beaucoup plus basse dans le cas d'une partie avec modérateur ce qui signifie que les modérateurs influencent la roulette pour qu'elle soit plus faible en misant peu et en faisant perdre beaucoup d'argent aux joueurs.

On peut aussi visualiser le résultat de l'analyse discriminante :

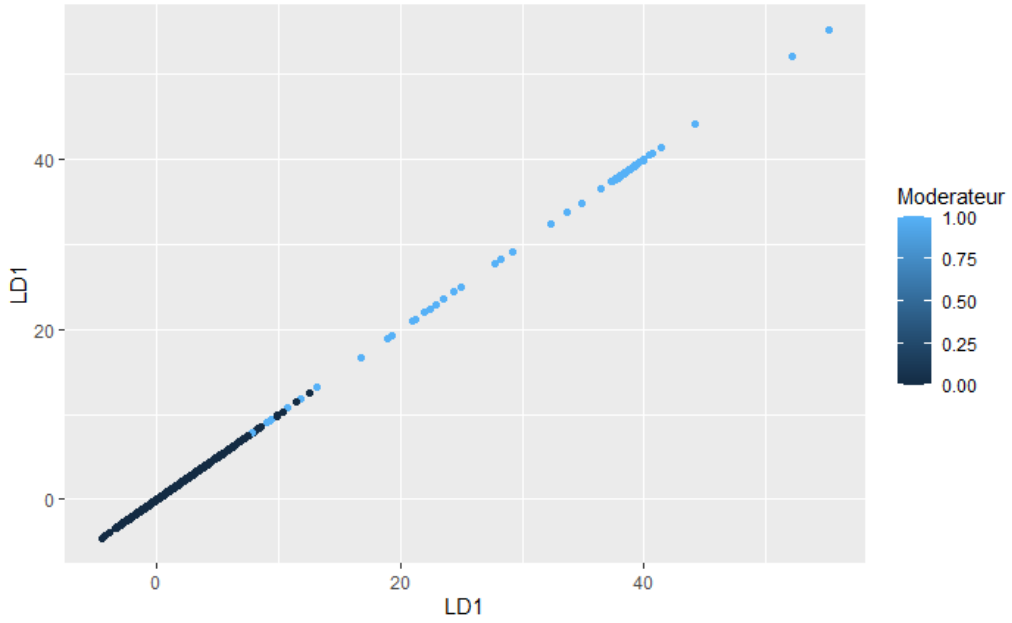

Figure 6 – *Résultat de l'analyse discriminante*

On remarque encore une fois que les deux groupes sont très bien séparés, ce qui confirme nos suppositions.

2.4 Test du χ^2

À présent on peut finalement réaliser un test du χ^2 entre la valeur de la roulette et la présence d'un modérateur. On obtient alors un χ^2 de 666880 et une p-value de 2.2×10^{-16}, on rejette donc l'hypothèse nulle et on peut affirmer que la valeur de la roulette est dépendante de la présence d'un modérateur. On peut donc affirmer que le casino est truqué.

2.5 ANOVA

En effectuant une ANOVA sur la valeur de la roulette en fonction de la présence d'un modérateur, on obtient la table suivante :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Moderateur	1	10911.00	10911.00	0.19	0.6641
Residuals	55573	3214320632.19	57839.61		

Table 3 – *ANOVA avec toutes les données*

Les résultats de l'ANOVA montrent une valeur F de 0.189 et une p-value de 0.664. La valeur F est relativement faible, ce qui suggère que la variabilité entre les groupes n'est pas particulièrement grande par rapport à la variabilité à l'intérieur des groupes. Plus important encore, la p-value, qui est bien au-dessus du seuil conventionnel de 0.05, indique qu'il n'y a pas de différence statistiquement significative entre les montants moyens des paris dans les situations avec et sans modérateur. Ce qui contredit les résultats de l'analyse discriminante et du test du χ^2, il faut noter que ces résultats ne sont pas complètement contradictoires, car la taille des deux échantillons est très différente.

On va alors recommencer l'ANOVA mais en choisissant aléatoirement 139 valeurs de la roulette sans modérateurs pour avoir un échantillon de même taille que celui avec modérateurs. On obtient alors la table suivante :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Moderateur	1	477.60	477.60	3.93	0.0483*
Residuals	276	33503.04	121.39		

Table 4 – *ANOVA avec des données réduites*

On obtient alors une valeur F de 3.93 et une p-value de 0.0483 significative, on peut donc affirmer que la valeur de la roulette est dépendante de la présence d'un modérateur.

2.6 Analyse factorielle des correspondances

A présent nous allons réaliser une analyse factorielle des correspondances sur la présence d'un modérateur. Voici le graphe d'inertie des valeurs propres :

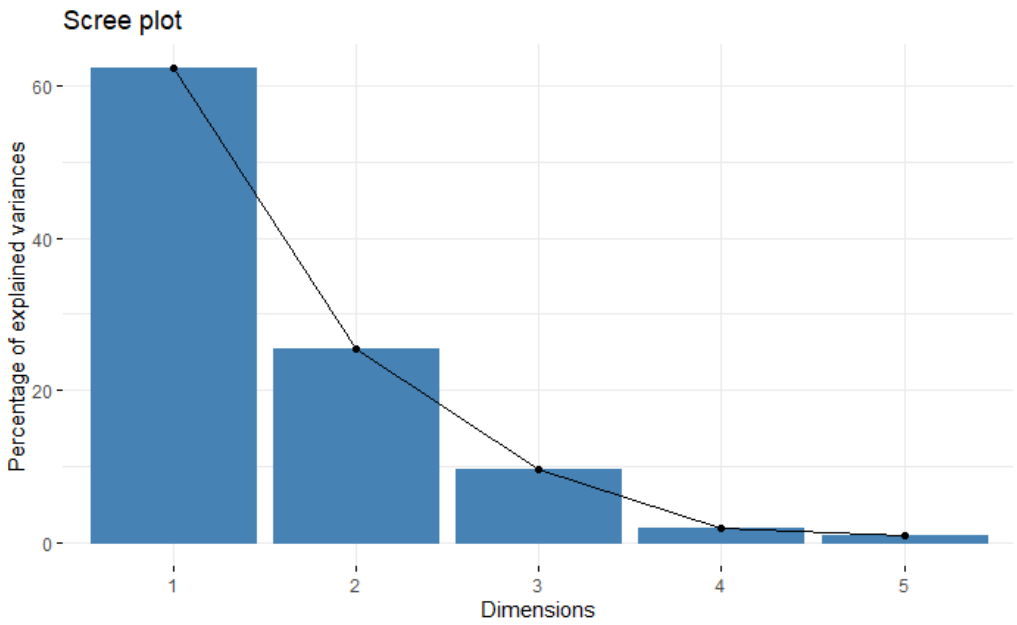

Figure 7 – *Graphe d'inertie des valeurs propres*

On peut voir que les deux premières composantes principales représentent à elles seules 87% de l'inertie totale, on va donc se contenter de les étudier. Voici à présent la carte factorielle où les point sont différenciés selon la présence d'un modérateur et le graphe des colonnes :

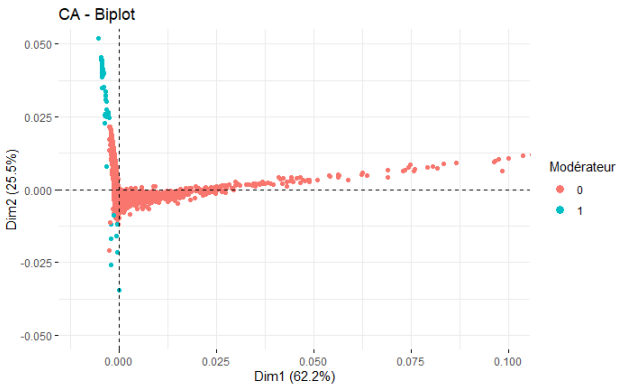

Figure 8 – *Carte factorielle*

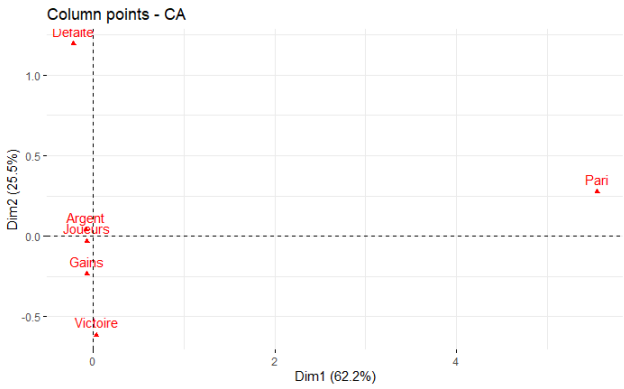

Figure 9 – *Graphe des colonnes*

Encore une fois, on remarque que les parties avec modérateur sont très corrélées avec la colonne défaite mais très peu avec la colonne Pari, ce qui est cohérent avec les différentes analyses faites précédemment.

3 Conclusion

Pour résumer, à travers toutes ces différentes analyses de données, on a pu démontrer que le casino est truqué. En effet, on a pu voir que les parties avec modérateur sont très différentes des parties sans modérateur, les joueurs perdent beaucoup plus d'argent lors de parties avec modérateurs car la valeur de la roulette est artificiellement plus faible. Paradoxalement les joueurs misent beaucoup plus d'argent lors de parties avec modérateurs, ce qui signifie que les modérateurs misent beaucoup d'argent en étant sûrs de gagner.