

# Analyse statistique des tirages du LOTO (FDJ) : étude des fréquences et des tendances

Yohan POSÉ

22 mai 2025

## Résumé

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Présentation du jeu : le LOTO . . . . .	1
1.2	Origine et structure des données analysées . . . . .	2
<b>2</b>	<b>Analyse descriptive univariée</b>	<b>4</b>
2.1	Distributions des tirages . . . . .	4
2.2	Montants des gains . . . . .	5
2.3	Jours et codes gagnants . . . . .	6
<b>3</b>	<b>Exploration multivariée</b>	<b>7</b>
3.1	Corrélation entre rang et montant gagné . . . . .	7
3.2	Évolution des cagnottes dans le temps . . . . .	7
3.3	Lien entre nombre de gagnants et montant du gain . . . . .	8
3.4	Saisonnalité des cagnottes . . . . .	9
3.5	Comparaison entre tirage principal et second tirage . . . . .	9
<b>4</b>	<b>Modélisations et inférences statistiques</b>	<b>10</b>
<b>5</b>	<b>Conclusion et perspectives</b>	<b>10</b>

## 1 Introduction

### 1.1 Présentation du jeu : le LOTO

Le LOTO est un jeu de hasard emblématique en France, organisé par la Française des Jeux (FDJ) depuis 1976. Il consiste à tirer aléatoirement une combinaison de **5 numéros distincts** compris entre 1 et 49, complétée par un numéro complémentaire, appelé « numéro

chance » dans les formules actuelles – compris entre 1 et 10. Ce jeu repose sur le principe d’une sélection sans remise, dans laquelle chaque tirage est censé être indépendant et équiprobable.

Les tirages ont lieu plusieurs fois par semaine, généralement le lundi, le mercredi et le samedi. Chaque combinaison jouée peut donner lieu à un gain, dont le montant dépend du nombre de bons numéros trouvés parmi ceux tirés.

Les gains sont répartis selon 9 « rangs » de gain, définis en fonction du nombre de numéros correctement prédits, avec ou sans le numéro *chance*. Ce système permet d’encourager une large gamme de participations, des plus ambitieuses (jackpot) aux plus accessibles (petits gains fréquents). Les rangs sont définis comme suit :

Rang	Nombre de bons numéros	Numéro chance
1	5	Oui
2	5	Non
3	4	Oui
4	4	Non
5	3	Oui
6	3	Non
7	2	Oui
8	2	Non
9	1	Oui

TABLE 1 – Définition des rangs

Les rangs supérieurs (1 à 3) offrent des gains importants mais peu fréquents, souvent variables car dépendants du nombre total de gagnants et du montant de la cagnotte en jeu. À l’inverse, les rangs inférieurs (4 à 9) proposent des gains fixes ou semi-fixes, statistiquement plus accessibles.

Ce mécanisme de répartition des gains cherche à établir un équilibre entre l’attractivité des gains exceptionnels et la satisfaction des joueurs réguliers par des récompenses plus fréquentes.

Compte tenu de sa popularité, de sa régularité et du volume important de données disponibles, le LOTO constitue un terrain d’étude privilégié pour une analyse statistique approfondie. Dans le cadre de ce projet, nous nous intéressons à une base de données regroupant les résultats historiques de ces tirages. L’objectif est d’examiner leur comportement statistique, de détecter d’éventuelles régularités ou anomalies, et d’évaluer leur conformité avec les lois théoriques de l’aléatoire.

## 1.2 Origine et structure des données analysées

Les données utilisées dans cette étude proviennent d’une compilation des tirages officiels du LOTO, publiés par la Française des Jeux (FDJ) [1]. Elles couvrent une période allant de 2019 à 2025, et ont été extraites à partir des archives mises à disposition en ligne, ainsi que de certaines sources secondaires (agrégateurs de données, sites d’analyse statistique).

Le fichier brut, au format `.csv`, contient l’ensemble des tirages réguliers effectués pendant cette période, incluant les tirages principaux, les seconds tirages (introduits à certaines dates), ainsi que les numéros « Joker+ » associés.

Chaque ligne du jeu de données correspond à un tirage unique. Les colonnes renseignent un ensemble varié d'informations regroupables en plusieurs catégories, avec une typologie des variables précisée ci-dessous :

Variable(s)	Type(s)	Description(s)	Conservée(s)
annee_numero_de_tirage	str	identifiant concaténé du tirage	Non
jour_de_tirage	str	jour de la semaine (LUNDI,...,DIMANCHE)	Non
date_de_tirage	str	dates du tirage (JJ/MM/AAAA)	Oui
date_de_forclusion	str	dates de forclusion <sup>1</sup> (JJ/MM/AAAA)	Non
boule_i	int	numéros principaux du tirage de 1 à 49	Oui
numero_chance	int	numéros principaux du tirage de 1 à 10	Oui
combinaison_gagnante_en_ordre_croissant	str	combinaisons formatées (b1-b2-b3-b4-b5+c)	Oui
nombre_de_gagnant_au_rangi	int	nombre de gagnants par rang	Oui
rapport_du_rangi	float	gains associés par rang	Oui
nombre_de_codes_gagnants	int	nombre total de codes gagnants	Oui
rapport_codes_gagnants	float	montant fixe associé aux codes gagnants	Oui
codes_gagnants	str	identifiants des codes gagnants	Oui
boule_i_second_tirage	int	numéros du second tirage si présent de 1 à 49	Oui
combinaison_gagnant_second_tirage_en_ordre_croissant	str	combinaisons formatées (b1-b2-b3-b4-b5+c)	Oui
nombre_de_gagnant_au_rangi_second_tirage	int	nombre de gagnants par rang pour le second tirage	Oui
rapport_du_rangi_second_tirage	float	gains associés par rang pour le second tirage	Oui
numero_jokerplus	int	numéro associé au tirage Joker+	Oui
devise	str	devise utilisée pour les montants ici « eur »	Non
promotion_second_tirage	str	-	Non
Unnamed: 49	str	-	Non

TABLE 2 – Variables brutes de "not\_cleaned\_df\_loto\_19-25.csv"

Avant toute analyse, un travail rigoureux de nettoyage a été réalisé : correction des formats (notamment pour les dates et les montants), traitement des valeurs manquantes, suppression des colonnes incomplètes ou redondantes, et harmonisation des noms de variables. Les données issues des tirages (numéros, gains, dates) ont également été restructurées pour faciliter l'analyse statistique.

Variable(s)	Type(s)	Description(s)
date_de_tirage	datetime	dates du tirage (AAAA-MM-JJ)
jour_de_la_semaine_de_tirage	str	jour de la semaine (Lundi,...,Dimanche)
jour_de_tirage	int	jour compris entre 1 et 31
mois_de_tirage	int	mois compris entre 1 et 12
annee_de_tirage	int	annee compris entre 2019 et 2025
boule_i	int	numéros principaux du tirage de 1 à 49
numero_chance	int	numéros principaux du tirage de 1 à 10
combinaison_gagnante_en_ordre_croissant	str	combinaisons formatées (b1-b2-b3-b4-b5+c)
nombre_de_gagnant_au_rangi	int	nombre de gagnants par rang
nombre_de_codes_gagnants	int	nombre total de codes gagnants
rapport_codes_gagnants	float	montant fixe associé aux codes gagnants
codes_gagnants	str	identifiants des codes gagnants
boule_i_second_tirage	int	numéros du second tirage si présent de 1 à 49
combinaison_gagnant_second_tirage_en_ordre_croissant	str	combinaisons formatées (b1-b2-b3-b4-b5+c)
nombre_de_gagnant_au_rangi_second_tirage	int	nombre de gagnants par rang pour le second tirage
rapport_du_rangi_second_tirage	float	gains associés par rang pour le second tirage
numero_jokerplus	int	numéro associé au tirage Joker+

TABLE 3 – Variables nettoyées de "df\_loto\_19-25.csv"

Ce jeu de données ainsi préparé constitue une base solide pour explorer les caractéristiques empiriques des tirages du LOTO et mettre à l'épreuve les hypothèses de tendance et d'indépendance attendues d'un jeu de hasard uniforme.

## 2 Analyse descriptive univariée

Cette section vise à explorer la distribution individuelle des variables les plus significatives du jeu de données. L'analyse descriptive univariée permet de dégager des tendances centrales, de repérer d'éventuelles valeurs extrêmes et d'identifier des motifs fréquents, sans considérer de relations entre variables.

### 2.1 Distributions des tirages

La fréquence d'apparition de chaque boule (`boule_i`,  $i \in \llbracket 1, 5 \rrbracket$ ) a été agrégée pour former une distribution globale des numéros tirés. L'analyse montre une répartition relativement homogène, ce qui est compatible avec l'hypothèse d'équiprobabilité des tirages. Toutefois, quelques numéros se distinguent par une fréquence légèrement supérieure à la moyenne.

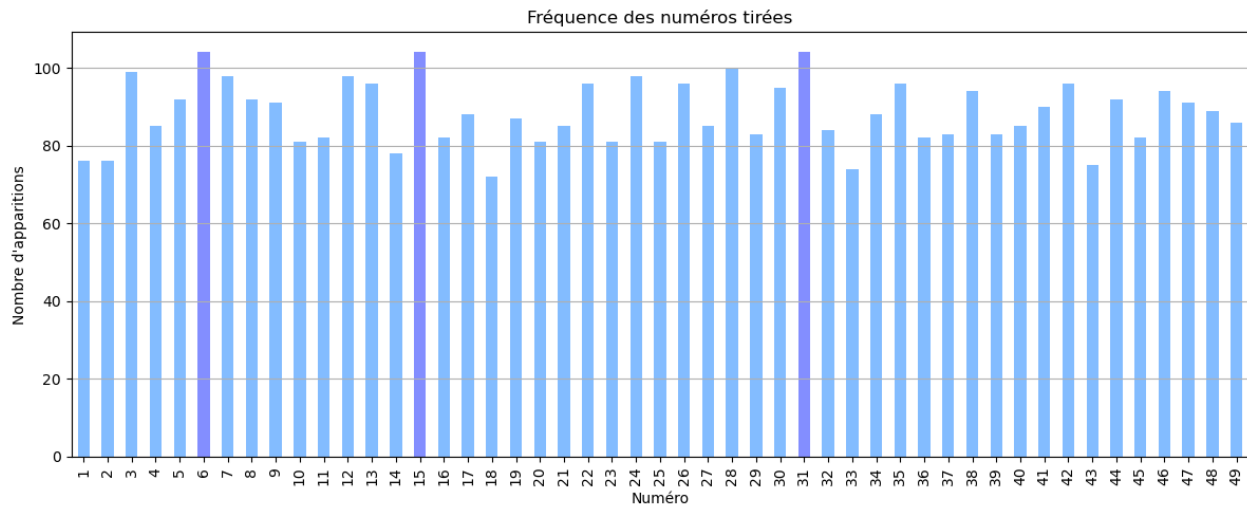


FIGURE 1 –

De même, la distribution du `numero_chance` est étudiée. La majorité des numéros de 1 à 10 apparaissent avec des fréquences proches, sans déséquilibre notable.

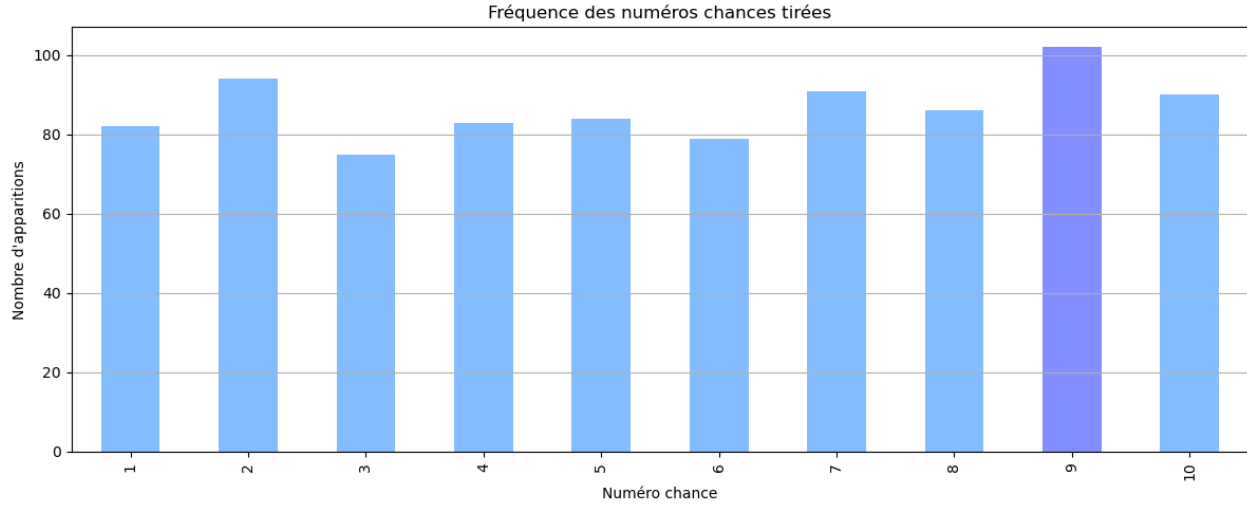


FIGURE 2 –

## 2.2 Montants des gains

Les variables `rapport_du_rangi` ( $i \in \llbracket 1, 9 \rrbracket$ ) présentent une forte variabilité. Le rang 1, associé à la cagnotte principale, se caractérise par des valeurs extrêmes, parfois très élevées lorsque les gains ne sont pas remportés sur plusieurs tirages. Les rangs inférieurs montrent des montants plus stables. Les premières mesures statistiques (moyenne, médiane, écart-type) ont été calculées pour chaque rang, offrant une vue synthétique des écarts entre niveaux de gain.

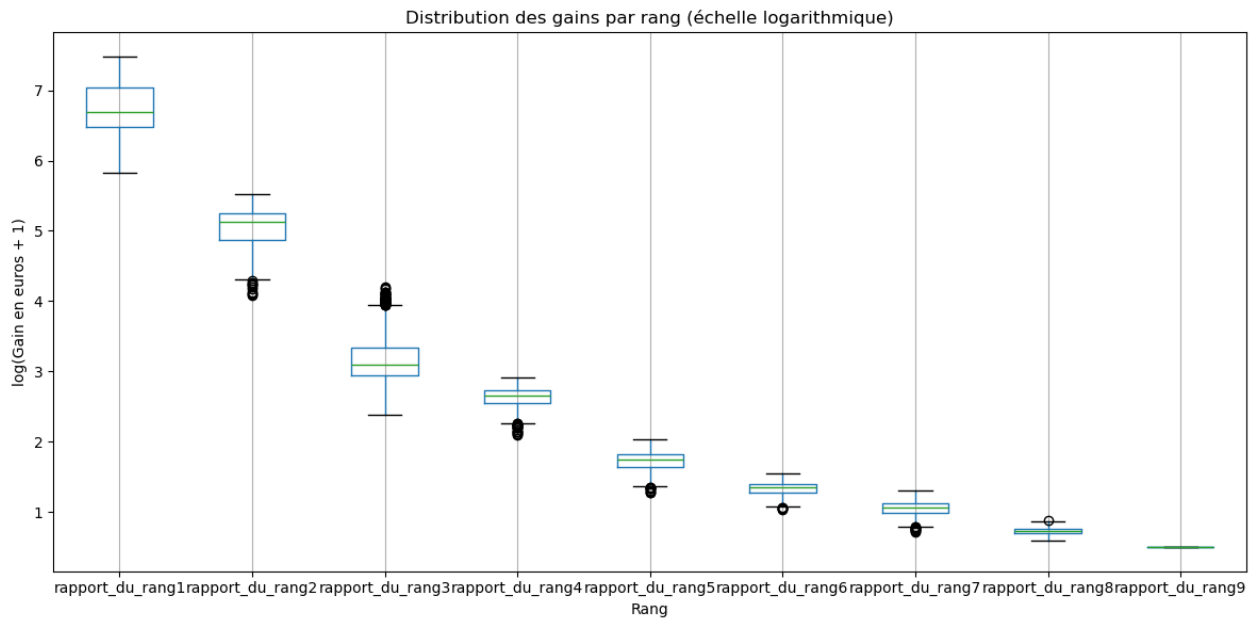


FIGURE 3 –

## 2.3 Jours et codes gagnants

Les jours de tirage (`jour_de_tirage`) sont dominés par trois modalités principales : Lundi, Mercredi et Samedi. Cela reflète la programmation habituelle des tirages par la FDJ.

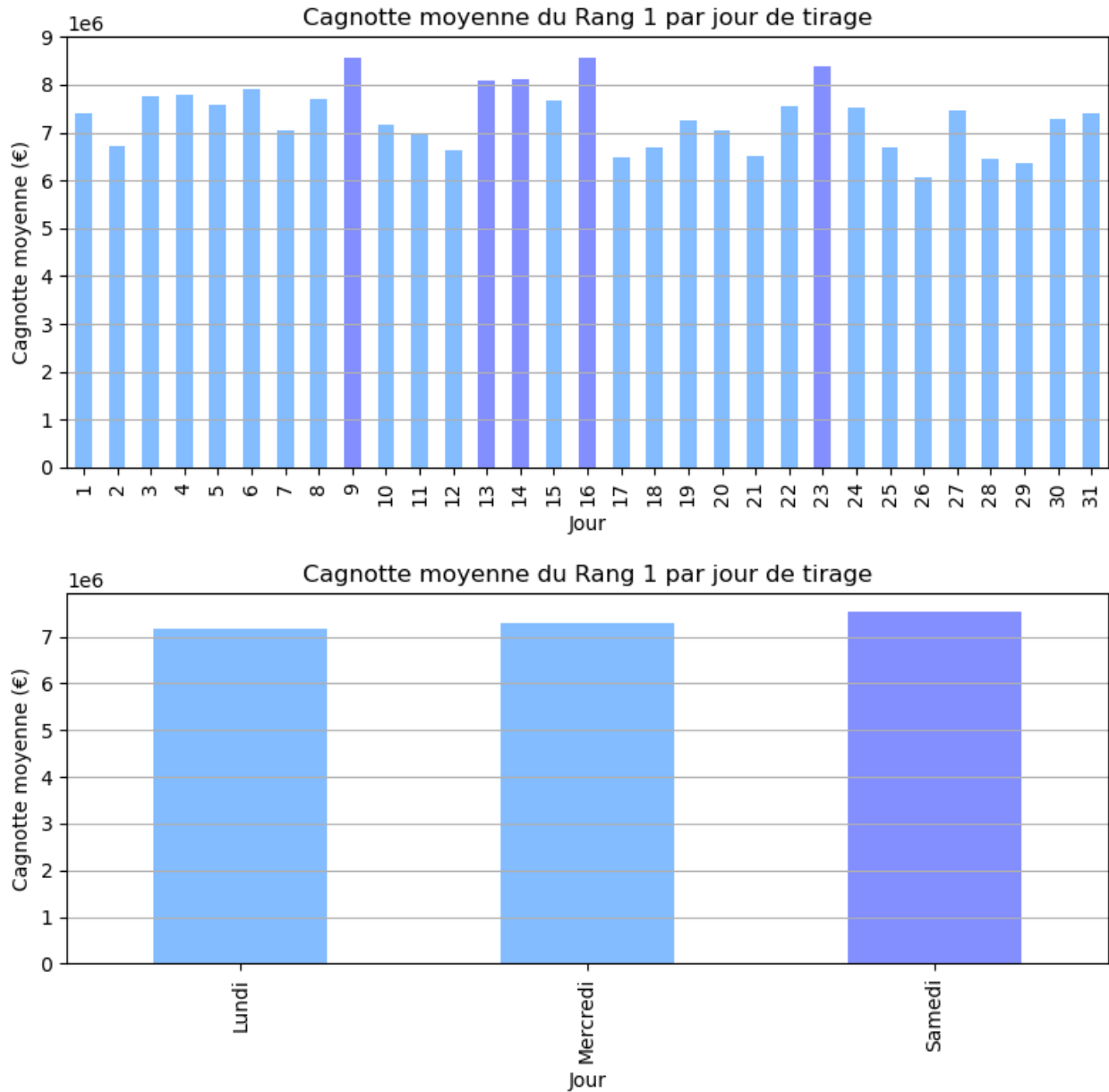


FIGURE 4 –

Concernant les `codes_gagnants`, leur analyse se limite à une inspection de la structure (chaîne de caractères formalisée). Ces codes ne présentent pas d'intérêt statistique direct en dehors de leur fréquence de distribution.

Enfin, les variables temporelles ont été converties en années et mois pour permettre des

regroupements. L'analyse de la fréquence des tirages par mois ou année montre une répartition stable dans le temps, sauf interruptions exceptionnelles.

### 3 Exploration multivariée

L'exploration multivariée vise à mettre en évidence les éventuelles relations statistiques entre plusieurs variables. Cette étape permet notamment d'interroger l'aléa des tirages, la stabilité des gains dans le temps, ou encore les interactions entre date, fréquence et montant des lots distribués.

#### 3.1 Corrélation entre rang et montant gagné

Nous avons analysé les corrélations entre les différentes colonnes `rapport_du_rangi` pour  $i \in \llbracket 1, 9 \rrbracket$  afin d'évaluer s'il existe des liens linéaires entre les gains aux différents rangs.

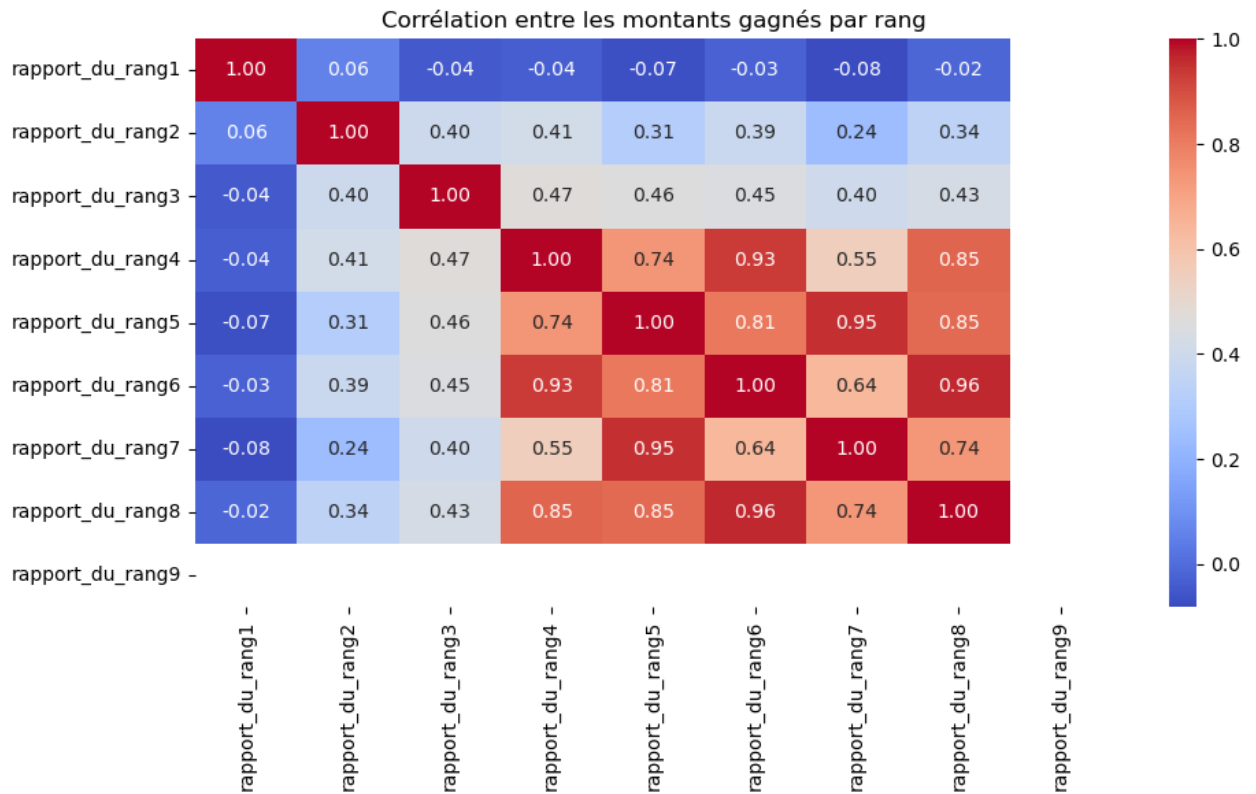


FIGURE 5 –

#### 3.2 Évolution des cagnottes dans le temps

Une visualisation de la variable `rapport_du_rang1` en fonction du temps permet d'illustrer les hausses brutales liées à l'accumulation des jackpots non remportés.

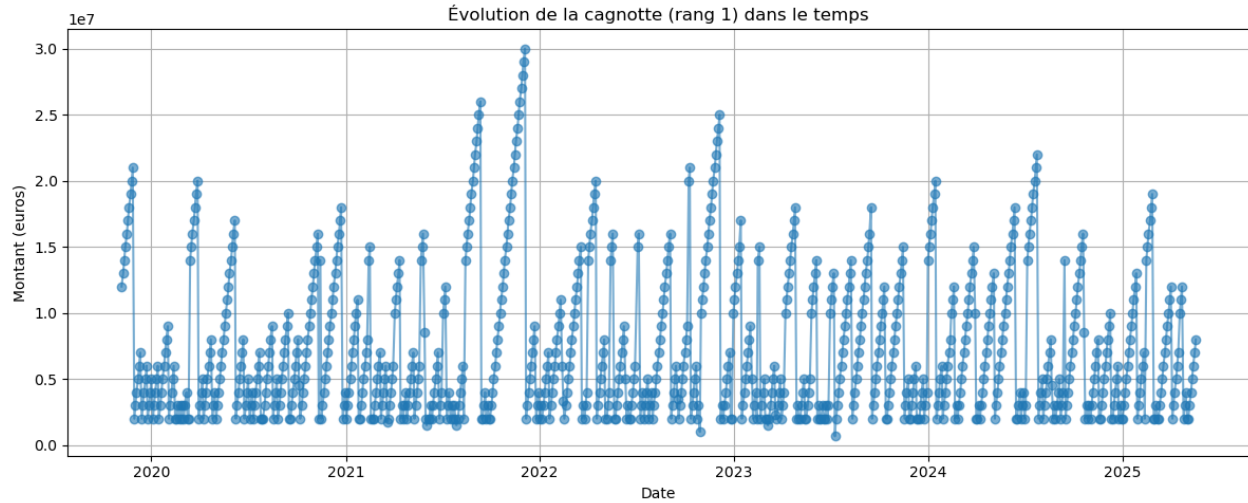


FIGURE 6 –

### 3.3 Lien entre nombre de gagnants et montant du gain

L'analyse de la relation entre le nombre de gagnants au rang 1 et le montant de la cagnotte montre une relation inverse, comme attendu. Lorsque plusieurs joueurs trouvent la combinaison gagnante, la cagnotte est répartie, ce qui réduit le gain individuel. Inversement, des cagnottes très élevées apparaissent en l'absence de gagnants, en raison de l'effet de report.



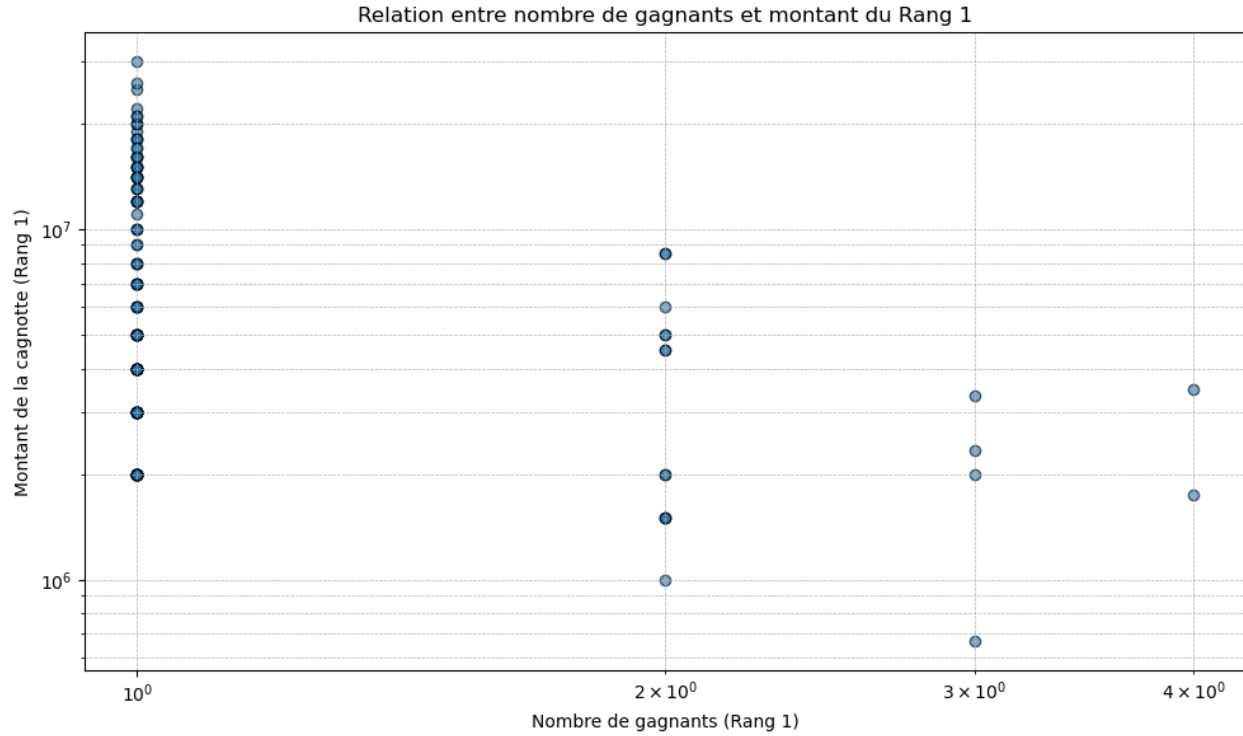


FIGURE 7 –

### 3.4 Saisonnalité des cagnottes

L'évolution moyenne des montants du rang 1 par mois et par année permet d'étudier une éventuelle saisonnalité. L'analyse ne montre pas de tendance mensuelle nette, mais met en évidence quelques années atypiques, notamment celles où le nombre de tirages a été affecté par des événements exceptionnels (périodes de confinement, modifications du règlement, etc.).

### 3.5 Comparaison entre tirage principal et second tirage

La mise en relation des gains du rang 1 entre le tirage principal et le second tirage révèle que les montants distribués lors du second tirage sont systématiquement inférieurs, conformément aux règles du jeu. Cette visualisation confirme l'existence d'un écart structurel entre les deux niveaux de tirage.

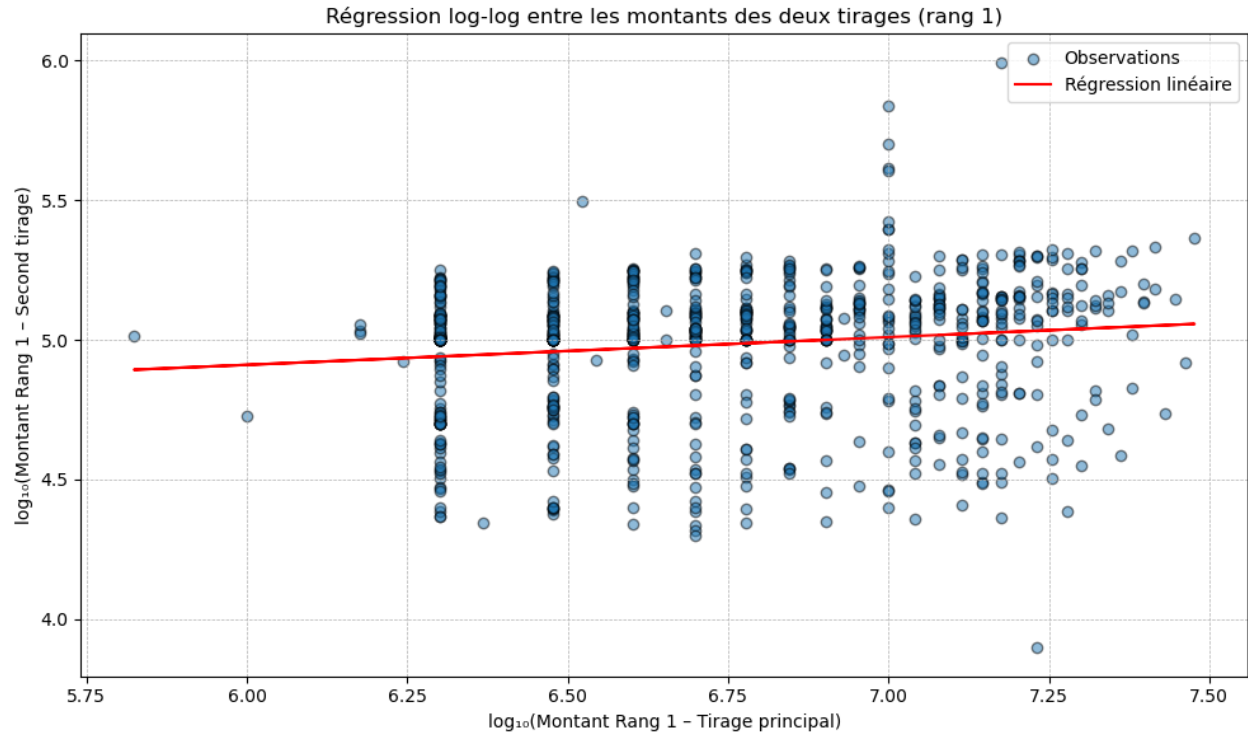


FIGURE 8 –

## 4 Modélisations et inférences statistiques

## 5 Conclusion et perspectives

## Références

- [1] Française des JEUX. *Résultats officiels du Loto*. Consulté le 22 mai 2025. 2025. URL : <https://www.fdj.fr/jeux-de-tirage/loto/resultats> (visité le 22/05/2025).