



Universitat Autònoma de Barcelona

**Consultoria Estadística**

## **LOTERIA DE NADAL**

Maria Marín Méndez (1668394)

Andrea Acuña Villagaray (1639232)

**29 Gener, 2026**

# Contents

<b>Contents</b> . . . . .	i
<b>1 Resum</b> . . . . .	1
<b>2 Descripció del conjunt de dades</b> . . . . .	2
2.1 Introducció de les dades . . . . .	2
<b>3 Exploració de les dades</b> . . . . .	4
3.1 Lectura de dades . . . . .	4
3.2 Anàlisi descriptiva . . . . .	5
3.3 Visualització Gràfica . . . . .	5
3.4 Discussió possibles valors atípics . . . . .	5
<b>4 Aplicació d'algorismes per la modelització</b> . . . . .	6
4.1 Metodologia . . . . .	6
4.2 Entranament del model amb el conjunt de dades . . . . .	6
<b>5 Resultats</b> . . . . .	6
5.1 Presentació dels resultats . . . . .	6
5.2 Mètriques d'avaluació . . . . .	6
<b>6 Conclusions</b> . . . . .	7
<b>7 Annexos</b> . . . . .	8
<b>8 References</b> . . . . .	9

# 1 Resum

El Sorteig Extraordinari de Nadal, també conegut com a Loteria de Nadal, és un dels sorteigs més populars que se celebra a Espanya cada desembre des de l'any 1812.

L'objectiu principal d'aquest estudi es centra en l'anàlisi exhaustiva dels resultats històrics del sorteig, així com el desenvolupament d'un model capaç de predir els números guanyadors de la Loteria de Nadal.

## 2 Descripció del conjunt de dades

### 2.1 Introducció de les dades

En aquest repositori es recull el desenvolupament del Treball Final de Consultoria Estadística 2025, centrat en l'anàlisi exhaustiva de la Loteria de Nadal des de la seva vessant més tècnica. L'objectiu no és sols descriure el sorteig, sinó avaluar amb rigor estadístic si la variabilitat dels resultats històrics (des del 2000 fins a l'any 2025) respon purament a l'atzar o si existeixen anomalies mesurables en l'homogeneïtat del sistemes,

A través de metodologies de web scraping, tests d'independència .....

La Loteria de Nadal no és sols un sorteig de boles; és l'unic moment de l'any en què un país sencer es posa d'acord per ignorar les lleis de l'estadística. Des d'un punt de vista matemàtic, és un "impost a l'esperança", però des del punt de vista de les dades, és un ecosistema fascinant.

#### L'arquitectura del "GORDO"

La loteria no ven números a l'atzar, sinó que segueix una jerarquia rígida que determina quants diners es mouen i quina probabilitat real tens de guanyar.

El sistema treballa amb un ventall de 100.000 números (des del 00000 fins al 99999). Això ens deixa una probabilitat de guanyar el primer premi amb un sol dècim de un 0.001%.

Per al 2025, SELAE (Societat Estatal Loteries i Apostes de l'Estat) va emetre 197 sèries de cada número. Això significa que de cada número n'hi ha 197 billets idèntic repartits per tot l'estat. On cada sèrie es dovodeox en 10 dècims, el que ens dona un total de 1970 dècims per cada número.

- Preu del dècim: 20€
- Recaptació potencila, si es ven tot, el sorteig mou 3940 milions d'euros. D'aquests, el 705 es destina a premis.

El que veiem a la tele és folklore, però el que hi ha dins dels bombos és física pura. com ens explica el professor Badiella, el sistema està dissenyat per ser homogeni.

Hi ha 100.000 boles al bombo gran, totes fetes de goma de boio, amb un diàmetre de 3cm i un pes unificat. On per evitar que el pes de la pintura alteri el camí de la bola cap a la trompeta, els números estan impresos amb l'àser. Així s'elimina la teoria que els números amb més "tinta", com el 88.888, pesen més que el 11.111.

Per a realitzar el sorteig es fan servir dos bombos que giren simultàniament: un per als números i un latre amb les 1807 boles de premis. fins que el bombo petit no queda buit, el sorteig no es dona per acabat.

## El repartiment del “Pastís”

Tot i que el focus està en el Gordo, la realitat és que el sorteig és una pluja fina de premis petits que serveixen per “alimentar l’esperança” de cara a l’anys següent.

Premi	Import per dècim	Boles premiades	Probabilitat
1r premi (“el Gordo”)	400.000€	1	0,001%
2n premi	125.000€	1	0,001%
3r premi	50.000€	1	0,001%
4rt premi	20.000€	2	0,002%
5é premi	6.000€	8	0,008%
La Pedrea	100€	1.794	1,794%
Reintegrament	20€	1 de cada 10 xifres	10,00%

Aleshores amb aquesta informació la probabilitat de que et toqui el Gordo és la mateixa que la de qualsevol número, 0,001%. Gairebé el 99% dels números que reben premi directe del bombo són pedres de 100€ (1.794 de 1.807). Per altra banda, tenim un 10% de possibilitats de recuperar 20€ si l’última xifra del teu número coincideix amb la del Gordo.

Hem de recordar que per als tres primers premis, el premi realment és menor, ja que Hisenda entra al joc aplicant un 20% d’impost a qualsevol premi que superi els 40.000€. Això vol dir que, el guanyador del primer premi realment es de 328.000€, ja que els primers 40.000€ estan exempts i es paga 20% dels 360.000€ restant. El segon premi sseria de 108.000€ i del tercer 48.000€.

Per on podem començar l’anàlisi?

1. Històric de terminacions: quins números han sortir més.
2. Geografia de la sort: Analitzar si és veritat que el premi toca més a Madrid o Barcelona, o si és simplement perquè s’hi venene més dècims.
3. Simulació de Montecarlo (proposta gemini): crear un script en R que simuli 10.000 sortejos per veure queantes vegades guanyaries el Gordo si juguessis el mateix número durant 100 anys.

### 3 Exploració de les dades

A continuació presenterem diverses estadístiques descriptives així com visualitzacions gràfiques que permetràn poder observar el comportament de les dades.

### 3.1 Lectura de dades

En primer lloc, realitzarem una lectura inicial de tots els números premiats en el sorteig de la Loteria de Nadal corresponents al període entre 2000 i 2025. Les dades utilitzades provenen dels resultats oficials publicats un cop finalitzat cada sorteig.

A l'endemà del sorteig, es publica el següent arxiu per comprovar si el número ha sigut premiat:

Figure 1: Resultats Sorteig loteria de nadal 2023

El codi utilitzat per a la lectura i el processament de les dades s'implementarà a través d'un script extern, el qual importarem mitjançant la funció `d'R source()`.

A la taula següent mostrarem alguns exemples dels premis per veure amb quines dades treballarem:

La taula mostra un extracte dels números premiats corresponents al sorteig del 2024. Cada fila representa un número premia, identificat per la seva xifra, juntament amb informació addicional com la lletra associada.

Hi han tres tipus de lletra que ens podem trobar dins del nostre dataset:

1. Lletra **t**:

2. Lletra **a**:

3. Lletra **b**:

4. Lletra **c**:

Finalment, també obtindrem els diners premiats, la categoria del premi, l'any i en quina moneda s'ha repartit els premis. Ja que ens els primers anys tindrem la moneda com a pessetes espanyoles.

### 3.2 Anàlisi descriptiva

(Posar taula: Estadística descriptiva conjunt de dades)

Aquest conjunt de dades compta amb un dataset corresponent a cada any analitzat, amb una mida mitjana de 4000 números premiats per any. Així que tindrem una mostra suficient per dur a terme el nostra anàlisis.

### 3.3 Visualització Gràfica

A continuació visualitzarem diverses representacions gràfiques.

(Posar: Gràfic distribució números)

(Posar: Última xifra números premiats)

(Fer: Quantitat premis per categoria (decena, centena...), modo serie temporal amb punts)

(Probar Heatmap a veure si es veu bé)

(Fer: Interpretació)

### 3.4 Discussió possibles valors atípics

Explicar com tractarem lo de que van cambiant el sorteig + PESETAS / EUROS

Que fem???

## **4 Aplicació d'algorismes per la modelització**

### **4.1 Metodologia**

Hi ha diversos algorismes que podem aplicar per desenvolupar un model predictiu, en aquest treball utilitzarem l'algorisme anomenat x.

X és una tècnica on .... D'aquesta manera el model podrà...

x calcula .... i ....

### **4.2 Entranament del model amb el conjunt de dades**

Per entrenar el model amb ... hem .... D'aquesta manera, podrem analitzar....

Hem dividir el dataset en conjunts d'entrenament i validació, amb les dades de l'any 2000 al 2024 dels resultats per entrenar el model i l'últim concurs realitzat al 2025 per a fer la validació. Així podrem observar el rendiment del model.

Un cop preparada la mostra d'entrenament,... Ja realitzat tot el prepossessing, ja podem definir i entrenar el model.

L'estructura d'aquest es veu tal que:

## **5 Resultats**

### **5.1 Presentació dels resultats**

### **5.2 Mètriques d'avaluació**

## 6 Conclusions

## **7 Annexos**

A continuació, introduirem el directori de Git-Hub on podràs trobar tot el codi d'R utilitzat:

<https://github.com/MariaMM2204/Consultoria-2025>

## **8 References**