



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 . . . . .	7
3.3 Visualització Gràfica . . . . .	13
3.4 Discussió possibles valors atípics . . . . .	24
<b>4 Aplicació d'algorismes per la modelització</b> . . . . .	24
4.1 Metodologia . . . . .	24
4.2 Entranament del model amb el conjunt de dades . . . . .	25
<b>5 Resultats</b> . . . . .	27
5.1 Presentació dels resultats . . . . .	27
5.2 Mètriques d'avaluació . . . . .	27
<b>6 Conclusions</b> . . . . .	28
<b>7 Annexos</b> . . . . .	29
<b>8 References</b> . . . . .	30

# 1 Resum

El Sorteig Extraordinari de Nadal, conegut popularment com la Loteria de Nadal, és un dels esdeveniments més populars a Espanya. El sorteig es celebra 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 encarregat de la lectura i el processament de les dades s'implementarà en un script extern, que importarem amb la funció `source()` d'R. A continuació, mostrarem alguns exemples dels premis per tal de familiaritzar-nos amb el tipus de dades amb què treballarem en l'estudi.

```
library(readxl)
```

## Warning: package 'readxl' was built under R version 4.4.3

```
library(dplyr)

## 
## Adjuntando el paquete: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag

## The following objects are masked from 'package:base':
## 
##     intersect, setdiff, setequal, union

library(knitr)
library(kableExtra)

## Warning: package 'kableExtra' was built under R version 4.4.2

## 
## Adjuntando el paquete: 'kableExtra'

## The following object is masked from 'package:dplyr':
## 
##     group_rows

library(stringr)

## Warning: package 'stringr' was built under R version 4.4.3

library(dplyr)
library(tibble)

## Warning: package 'tibble' was built under R version 4.4.3
```

```

library(readxl)
library(readr)

datos <- read_excel("loteria_2015_2024.xlsx")
exemple <- head(datos[datos$any==2024,])

exemple %>%
  kable(caption = "Primers premis de la Loteria de Nadal 2024") %>%
  kable_styling(full_width = FALSE, bootstrap_options = c("striped", "hover"))

```

Table 2: Primers premis de la Loteria de Nadal 2024

numero	letra	premio	categoria	moneda	any
00014	t	1000	Decena	EUR	2024
00017	NA	1000	Decena	EUR	2024
00040	t	1000	Decena	EUR	2024
00055	NA	1000	Decena	EUR	2024
00059	NA	1000	Decena	EUR	2024
00080	NA	1000	Decena	EUR	2024

La taula mostra un extracte dels números premiats corresponents al sorteig del 2024. Cada fila representa un número premiat, 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**:

A més, també tindrem l'import del premi, la seva categoria i l'any.

### 3.2 Anàlisi descriptiva

```
source("nateja_dades.R")
options(scipen = 999)
df <- nateja(datos)
df %>%
  summarise(
    `Años analizados` = n_distinct(any),
    `Total de números` = n(),
    `Premio mínimo (€)` = min(premio, na.rm = TRUE),
    `Premio máximo (€)` = max(premio, na.rm = TRUE),
    `Premio medio (€)` = round(mean(premio, na.rm = TRUE), 2)
  ) %>%
  kable(
    caption = "Resumen global de la Lotería de Navidad",
    align = "c"
  ) %>%
  kable_styling(full_width = FALSE)
```

Table 3: Resumen global de la Lotería de Navidad

Años analizados	Total de números	Premio mínimo (€)	Premio máximo (€)	Premio medio (€)
10	53050	1000	4000000	2262.05

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.

```
df %>%
  count(premio, name = "Frecuencia") %>%
  arrange(desc(Frecuencia)) %>%
  head(10) %>%
  kable(
    caption = "Premios más frecuentes en la Lotería de Navidad",
```

```

col.names = c("Premio (€)", "Frecuencia"),
align = "c"
) %>%
kable_styling(full_width = FALSE)

```

Table 4: Premios más frecuentes en la Lotería de Navidad

Premio (€)	Frecuencia
1000	52860
60000	80
9600	20
12500	20
20000	20
200000	20
500000	10
1250000	10
4000000	10

```

df %>%
group_by(categoría) %>%
summarise(
`Nº observaciones` = n(),
`Premio mínimo (€)` = min(premio, na.rm = TRUE),
`Premio máximo (€)` = max(premio, na.rm = TRUE),
`Premio medio (€)` = round(mean(premio, na.rm = TRUE), 2)
) %>%
arrange(desc(`Nº observaciones`)) %>%
kable(
caption = "Análisis de premios por categoría",
align = "c"
) %>%
kable_styling(full_width = FALSE)

```

Table 5: Análisis de premios por categoría

categoria	Nº observaciones	Premio mínimo (€)	Premio máximo (€)	Premio medio (€)
Setenta y dos mil	787	1000	4000000	12950.44
Cuatro mil	777	1000	1250000	4938.48
Seis mil	697	1000	1250000	3650.22
Cuarenta y dos mil	696	1000	200000	1571.84
Cinco mil	686	1000	4000000	7723.32
Cincuenta y dos mil	683	1000	500000	2047.14
Cuarenta y un mil	682	1000	200000	1583.58
Setenta y un mil	680	1000	4000000	7229.41
Cincuenta y un mil	614	1000	1250000	3167.75
Sesenta y seis mil	613	1000	4000000	7681.89
Setenta y siete mil	605	1000	200000	1328.93
Setenta y ocho mil	599	1000	500000	1861.77
Veintiún mil	599	1000	1250000	3123.54
Cincuenta y cuatro mil	596	1000	200000	1531.88
Sesenta y un mil	596	1000	200000	1333.89
Setenta y nueve mil	594	1000	4000000	7895.62
Doce mil	593	1000	1250000	3145.03
Setenta y cinco mil	593	1000	200000	1534.57
Diez mil	591	1000	1250000	3152.28
Treinta y un mil	591	1000	500000	1973.27
Cuarenta y ocho mil	588	1000	200000	1338.44
Diecinueve mil	587	1000	500000	2080.41
Noventa y tres mil	585	1000	200000	1340.17
Cincuenta y ocho mil	583	1000	1250000	3283.02
Siete mil	582	1000	200000	1443.30
Ochenta y ocho mil	581	1000	4000000	8151.46
Cuarenta y cinco mil	580	1000	500000	2195.17
Tres mil	580	1000	4000000	8163.79

Ochenta y seis mil	577	1000	4000000	8201.04
Veinticinco mil	574	1000	200000	1346.69
Once mil	573	1000	500000	1900.87
Trece mil	571	1000	200000	1348.51
Cincuenta y nueve mil	567	1000	200000	1455.03
Sesenta y siete mil	566	1000	200000	1351.59
Noventa y un mil	564	1000	200000	1457.45
Treinta y ocho mil	564	1000	200000	1457.45
Cuarenta y nueve mil	563	1000	200000	1458.26
Cuarenta mil	561	1000	1250000	3267.38
Veintiséis mil	561	1000	4000000	8301.25
Centena	533	1000	500000	2189.87
Cuarenta y cuatro mil	514	1000	1000	1000.00
Setenta y tres mil	508	1000	60000	1116.14
Treinta mil	504	1000	1000	1000.00
Veintidós mil	504	1000	60000	1234.13
Ocho mil	501	1000	1000	1000.00
Noventa y nueve mil	499	1000	1000	1000.00
Catorce mil	494	1000	1000	1000.00
Cincuenta y cinco mil	491	1000	60000	1120.16
Cincuenta y tres mil	490	1000	1000	1000.00
Noventa y cuatro mil	490	1000	1000	1000.00
Treinta y dos mil	490	1000	1000	1000.00
Treinta y seis mil	488	1000	60000	1120.90
Setenta mil	487	1000	60000	1363.45
Veinticuatro mil	487	1000	60000	1363.45
Cuarenta y seis mil	486	1000	1000	1000.00
Noventa y cinco mil	486	1000	1000	1000.00
Ochenta y cinco mil	486	1000	1000	1000.00
Quince mil	486	1000	1000	1000.00
Treinta y cinco mil	486	1000	1000	1000.00

Veintiocho mil	486	1000	60000	1121.40
Cuarenta y siete mil	485	1000	60000	1121.65
Noventa y dos mil	485	1000	60000	1243.30
Noventa y ocho mil	484	1000	1000	1000.00
Sesenta y cinco mil	484	1000	1000	1000.00
Treinta y tres mil	484	1000	1000	1000.00
Cuarenta y tres mil	483	1000	60000	1366.46
Sesenta y nueve mil	483	1000	60000	1244.31
Treinta y nueve mil	482	1000	60000	1122.41
Treinta y siete mil	481	1000	60000	1490.64
Ochenta y nueve mil	480	1000	60000	1368.75
Setenta y seis mil	480	1000	1000	1000.00
Ochenta y dos mil	478	1000	1000	1000.00
Noventa y siete mil	477	1000	60000	1123.69
Sesenta y ocho mil	477	1000	60000	1247.38
Veintisiete mil	475	1000	1000	1000.00
Ochenta y cuatro mil	474	1000	1000	1000.00
Sesenta y tres mil	474	1000	60000	1124.47
Cincuenta y seis mil	473	1000	1000	1000.00
Cincuenta mil	471	1000	1000	1000.00
Dieciséis mil	470	1000	1000	1000.00
Ochenta y siete mil	470	1000	60000	1125.53
Ochenta y un mil	469	1000	60000	1125.80
Veinte mil	469	1000	60000	1125.80
Dieciocho mil	468	1000	60000	1378.21
Noventa y seis mil	468	1000	1000	1000.00
Ochenta mil	468	1000	1000	1000.00
Ochenta y tres mil	468	1000	1000	1000.00
Dos mil	467	1000	60000	1126.34
Sesenta y dos mil	467	1000	60000	1126.34
Cincuenta y siete mil	466	1000	60000	1126.61

Noventa mil	466	1000	1000	1000.00
Setenta y cuatro mil	463	1000	60000	1254.86
Veintitrés mil	462	1000	60000	1127.71
Nueve mil	461	1000	1000	1000.00
Veintinueve mil	457	1000	60000	1129.10
Diecisiete mil	456	1000	1000	1000.00
Mil	454	1000	60000	1129.96
Sesenta mil	454	1000	60000	1259.91
Treinta y cuatro mil	451	1000	60000	1130.82
Sesenta y cuatro mil	444	1000	1000	1000.00
Decena	44	1000	1000	1000.00
Unidad	3	1000	1000	1000.00

```
df %>%
  count(letra)
```

```
## # A tibble: 4 x 2
##   letra     n
##   <chr> <int>
## 1 a        60
## 2 c       4950
## 3 t      29970
## 4 <NA>    18070
```

```
df %>%
  count(numero, name = "Veces premiado") %>%
  arrange(desc(`Veces premiado`)) %>%
  head(10) %>%
  kable(
    caption = "Números más premiados históricamente",
    align = "c"
  ) %>%
```

```
kable_styling(full_width = FALSE)
```

Table 6: Números más premiados históricamente

numero	Veces premiado
03348	5
06050	5
06948	5
13948	5
37240	5
38350	5
54450	5
71168	5
86713	5
00219	4

Identificar números que no han sortit mai

```
n <- tibble(numero = 00000:99999)

freq <- df %>%
  count(numero, name = "vegades_premiat") %>%
  arrange(desc(vegades_premiat))

freq <- freq %>% mutate(numero = as.integer(numero))

mai <- n %>%
  anti_join(freq, by = "numero");
nrow(mai)

## [1] 59495
```

### 3.3 Visualització Gràfica

A continuació visualitzarem diverses representacions gràfiques.

```

library(dplyr)
library(ggplot2)

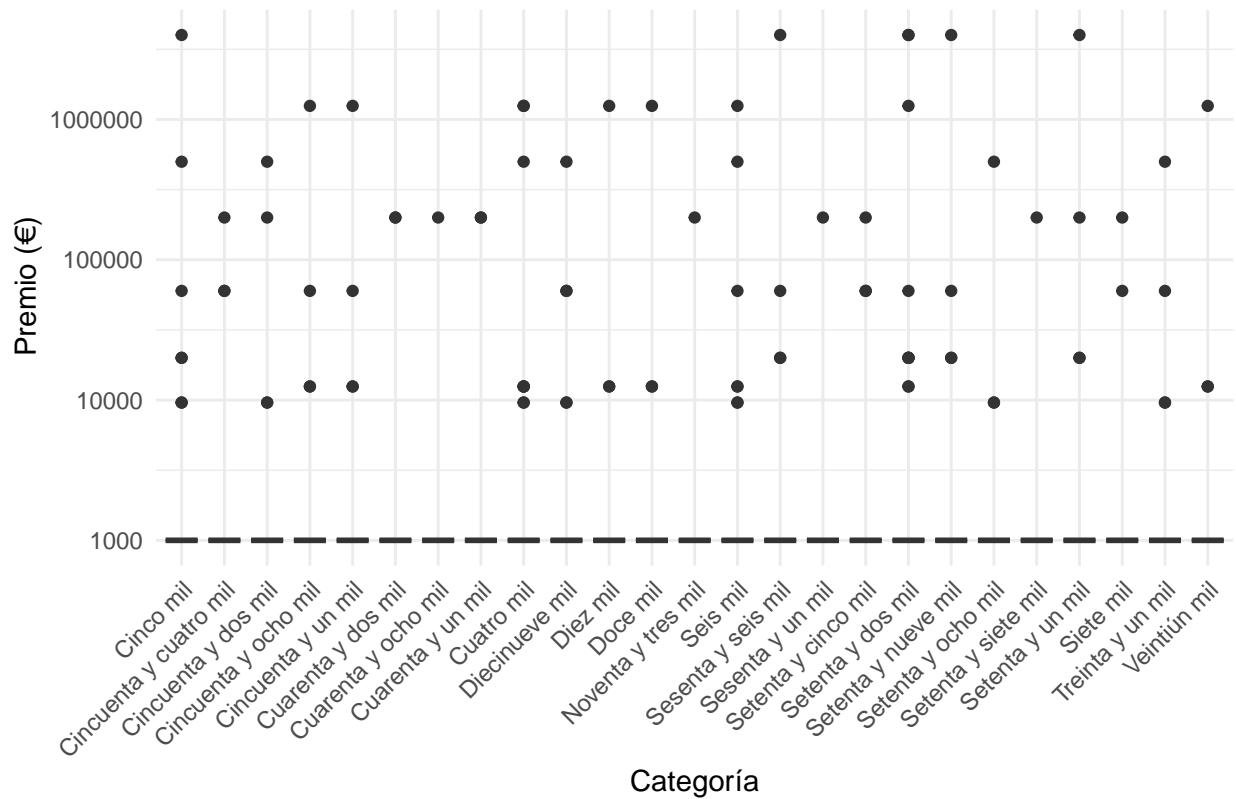
## Warning: package 'ggplot2' was built under R version 4.4.3

top_categorias <- df %>%
  count(categoría, sort = TRUE) %>%
  slice_head(n = 25) %>%
  pull(categoría)

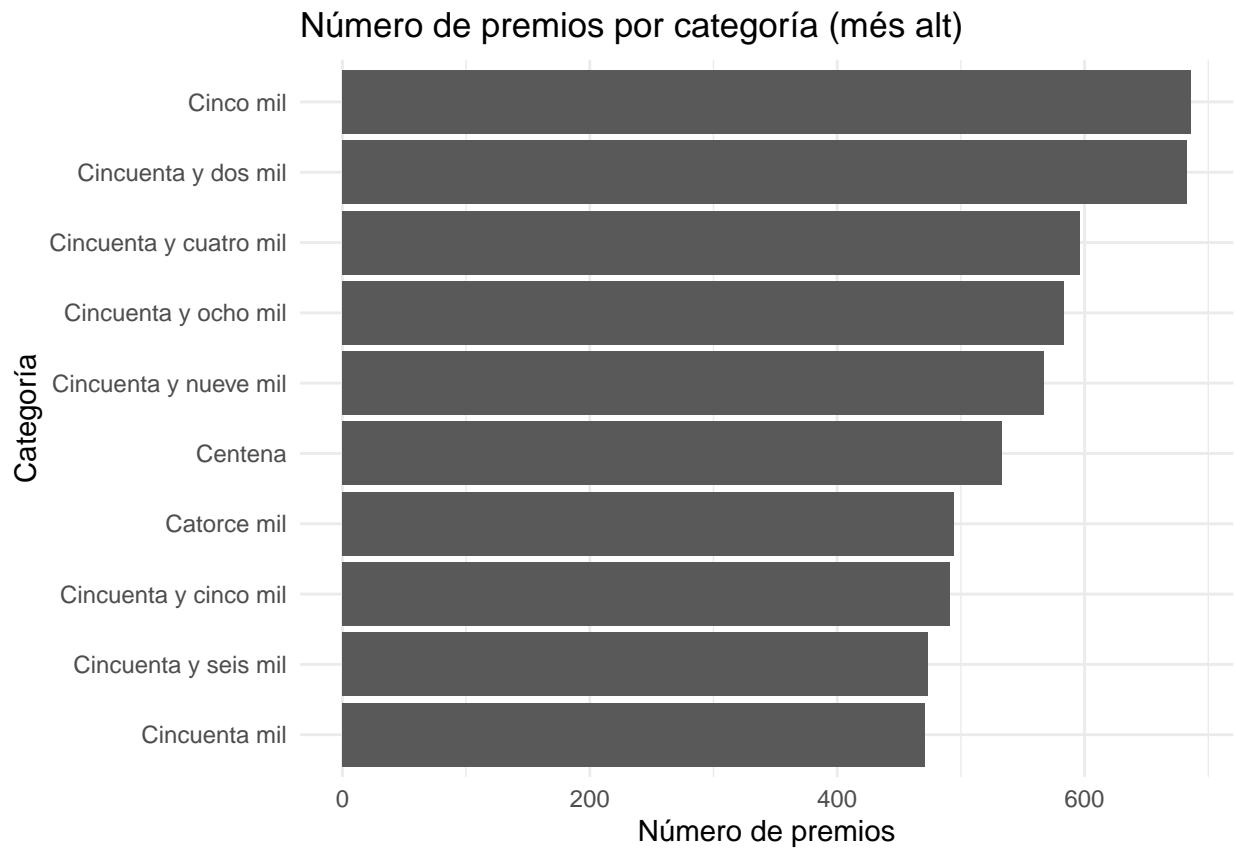
df %>%
  filter(!is.na(premio), categoría %in% top_categorias) %>%
  ggplot(aes(x = categoría, y = premio)) +
  geom_boxplot() +
  scale_y_log10() +
  labs(
    title = "Distribución del premio por categoría (Top 100 más usadas)",
    x = "Categoría",
    y = "Premio (€)"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

```

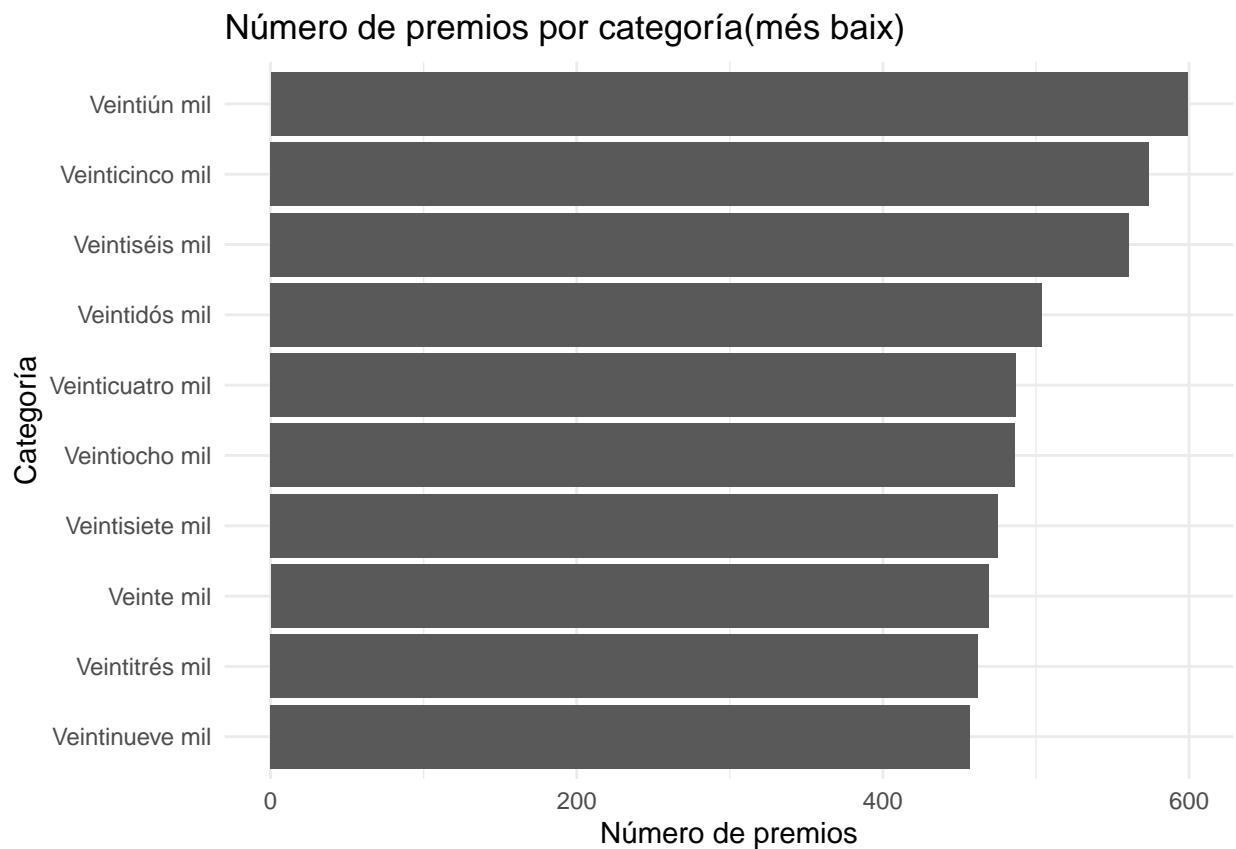
## Distribución del premio por categoría (Top 100 más usadas)



```
df %>%
  count(categoría) %>%
  slice_head(n=10) %>%
  ggplot(aes(x = reorder(categoría, n), y = n)) +
  geom_col() +
  coord_flip() +
  labs(
    title = "Número de premios por categoría (más alt)",
    x = "Categoría",
    y = "Número de premios"
  ) +
  theme_minimal()
```



```
df %>%
  count(categoría) %>%
  slice_tail(n=10) %>%
  ggplot(aes(x = reorder(categoría, n), y = n)) +
  geom_col() +
  coord_flip() +
  labs(
    title = "Número de premios por categoría(més baix)",
    x = "Categoría",
    y = "Número de premios"
  ) +
  theme_minimal()
```



```

library(dplyr)
library(tidyr)

# Convertir a character para separar dígitos
df_digitos <- df %>%
  filter(!is.na(numero)) %>%
  mutate(numero = as.character(numero)) %>%
  # Separar cada dígito en columnas
  mutate(
    d1 = substr(numero, 1, 1),
    d2 = substr(numero, 2, 2),
    d3 = substr(numero, 3, 3),
    d4 = substr(numero, 4, 4),
    d5 = substr(numero, 5, 5)
  ) %>%

```

```

# Unir todos los dígitos en una sola columna
pivot_longer(cols = d1:d5, names_to = "posicion", values_to = "digito")

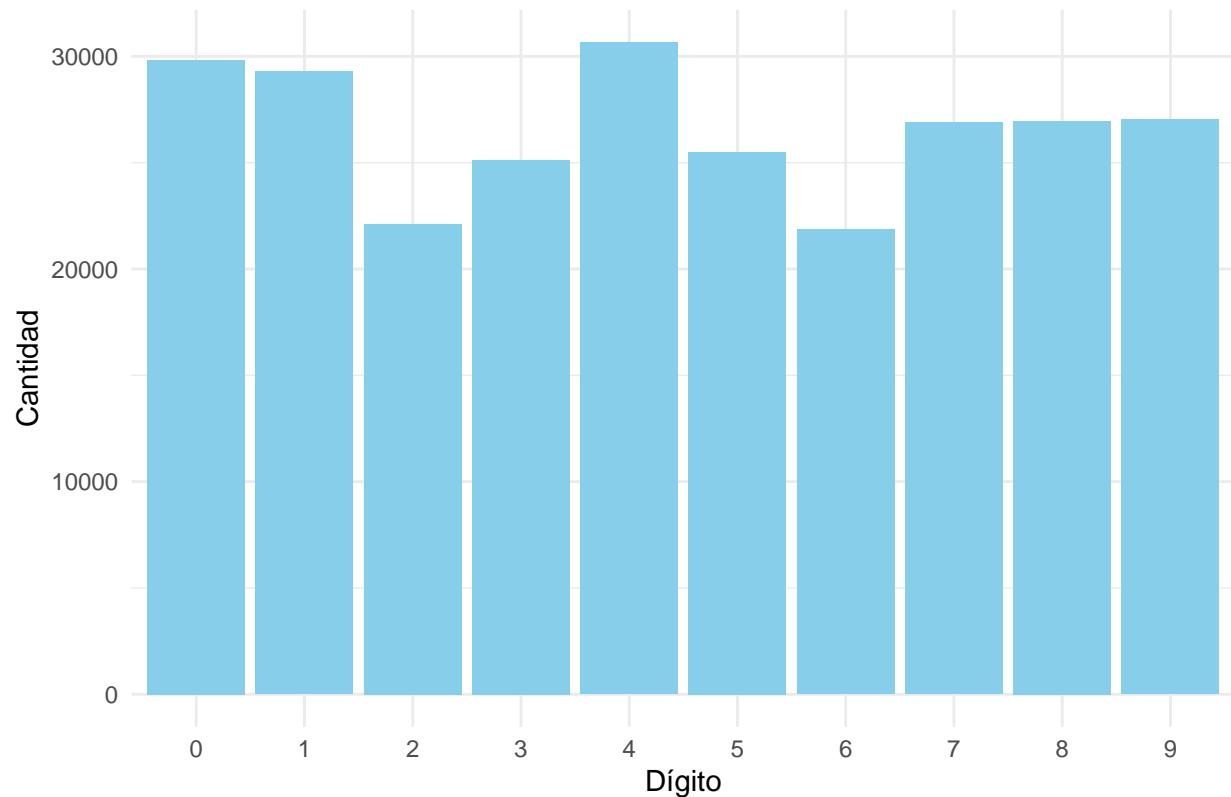
library(dplyr)
library(tidyr)

# Convertir a character para separar dígitos
df_digitos <- df %>%
  filter(!is.na(numero)) %>%
  mutate(numero = as.character(numero)) %>%
# Separar cada dígito en columnas
  mutate(
    d1 = substr(numero, 1, 1),
    d2 = substr(numero, 2, 2),
    d3 = substr(numero, 3, 3),
    d4 = substr(numero, 4, 4),
    d5 = substr(numero, 5, 5)
  ) %>%
# Unir todos los dígitos en una sola columna
pivot_longer(cols = d1:d5, names_to = "posicion", values_to = "digito")
library(ggplot2)

df_digitos %>%
  count(digito) %>%
  ggplot(aes(x = digito, y = n)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(
    title = "Frecuencia de los dígitos en los números de lotería",
    x = "Dígito",
    y = "Cantidad"
  ) +
  theme_minimal()

```

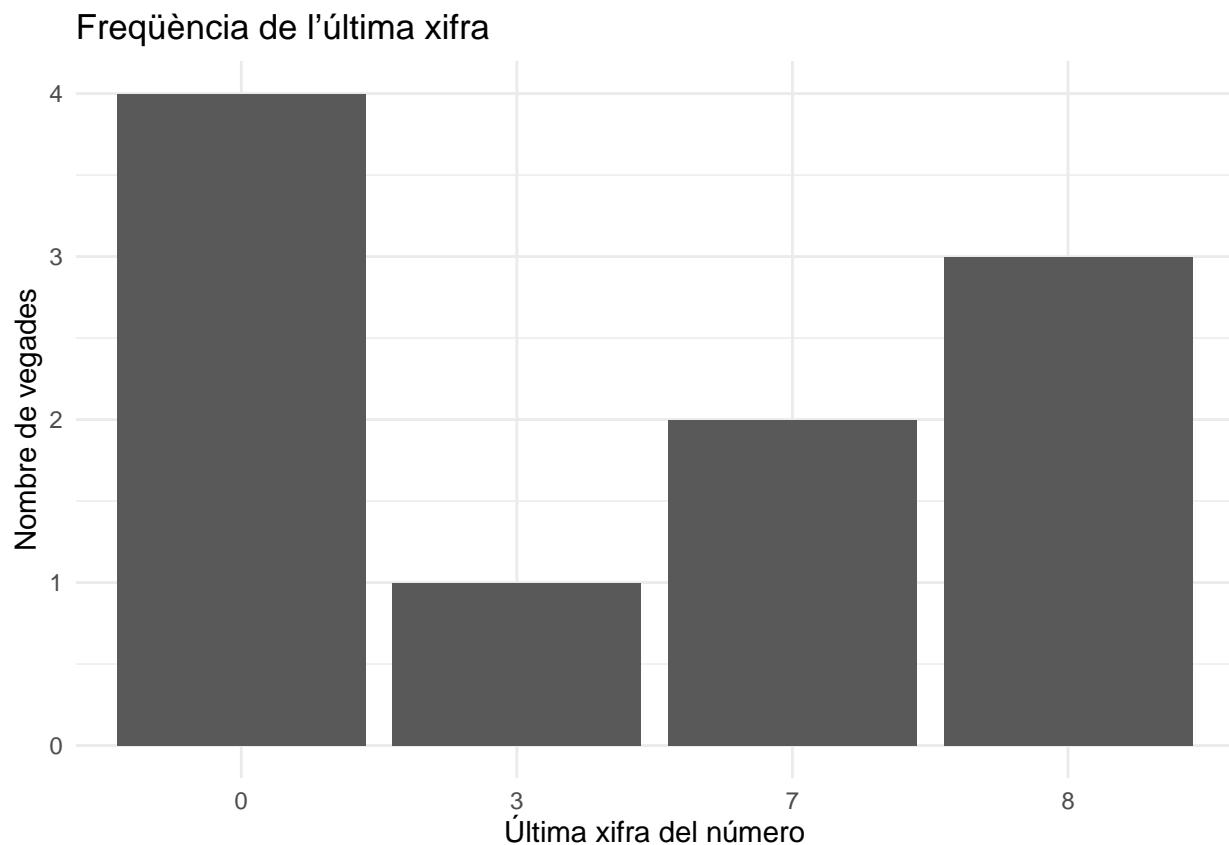
## Frecuencia de los dígitos en los números de lotería



```
a <- df %>%
  filter(!is.na(premio)) %>%
  filter(premio == max(premio)) %>%
  mutate(
    ultima_cifra = substr(numero, nchar(numero), nchar(numero)))
) %>%
count(ultima_cifra, sort=TRUE)

a %>%
ggplot(aes(x = ultima_cifra, y = n)) +
geom_col() +
labs(
  title = "Freqüència de l'última xifra",
  x = "Última xifra del número",
  y = "Nombre de vegades"
```

```
) +  
theme_minimal()
```



```
historial <- read_table2(  
  "historial_gordo.txt",  
  col_types = cols(  
    Any = col_integer(),  
    Numero = col_character(),  
    Terminació = col_integer()))  
  
## Warning: 'read_table2()' was deprecated in readr 2.0.0.  
## i Please use 'read_table()' instead.  
## This warning is displayed once every 8 hours.  
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was  
## generated.
```

```

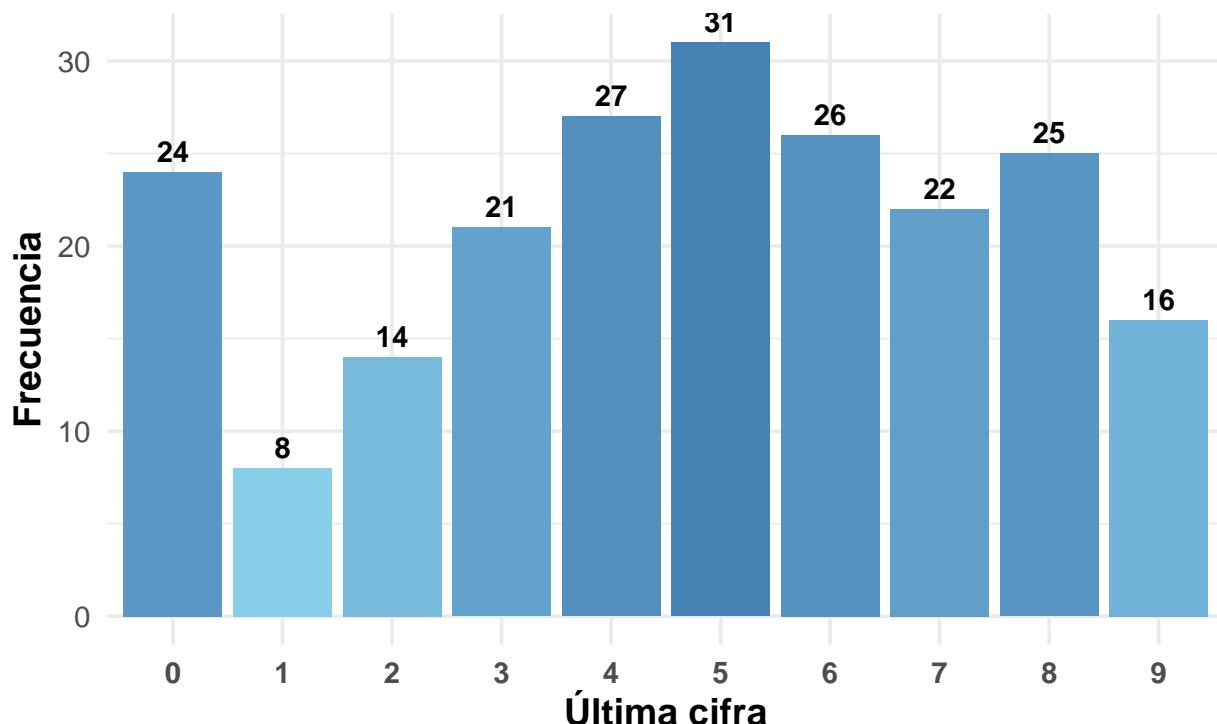
df_plot <- historial %>%
  count(Terminació) %>%
  mutate(Terminació = factor(Terminació, levels = 0:9))

ggplot(df_plot, aes(x = Terminació, y = n, fill = n)) +
  geom_col(show.legend = FALSE) +
  geom_text(aes(label = n), vjust = -0.5, size = 4, fontface = "bold") +
  scale_fill_gradient(low = "skyblue", high = "steelblue") +
  labs(title = "Distribució de la última xifra dels números premiats",
       subtitle = "Resultats Històrics de la Loteria de Nadal",
       x = "Última cifra",
       y = "Frecuencia") +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(face = "bold"),
        axis.title = element_text(face = "bold"),
        plot.title = element_text(face = "bold", size = 16),
        plot.subtitle = element_text(size = 12))

```

# Distribució de la última xifra dels números premiats

Resultats Històrics de la Loteria de Nadal



```
library(dplyr)
library(ggplot2)
library(readr)
library(lubridate)

## Warning: package 'lubridate' was built under R version 4.4.2

##
## Adjuntando el paquete: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```

df <- df %>%
  mutate(ultima_cifra = as.integer(substr(numero, nchar(numero), nchar(numero))))

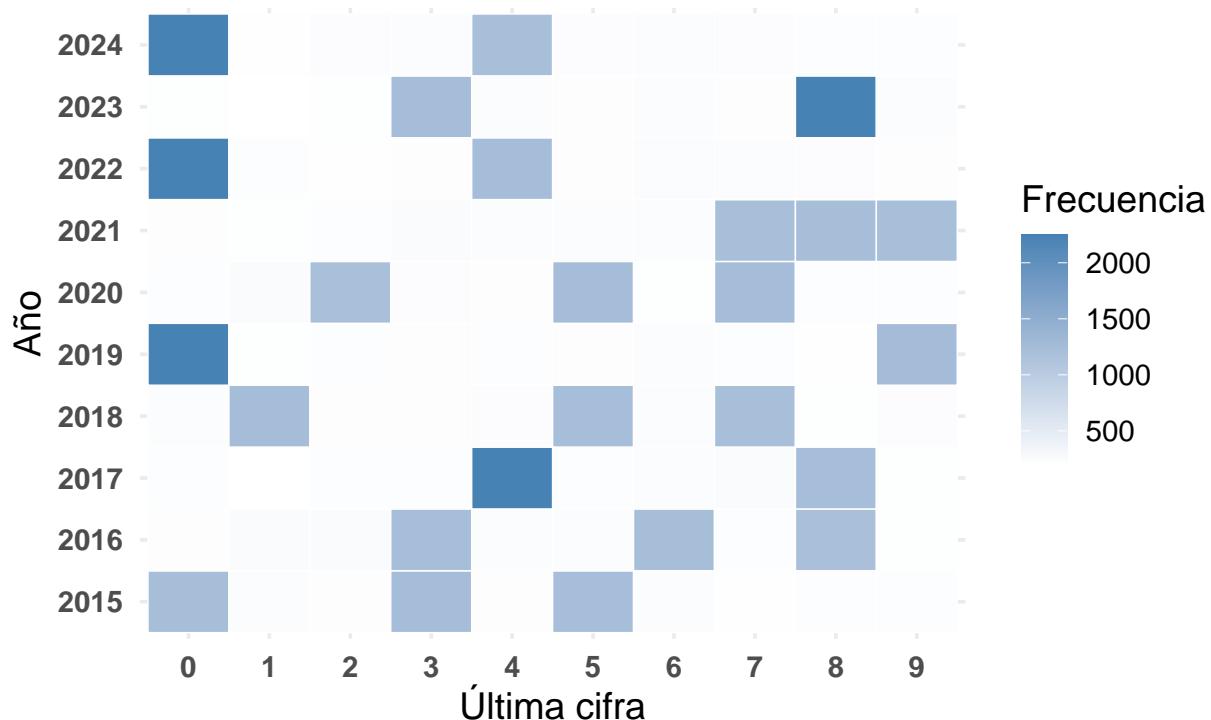

df_heat <- df %>%
  group_by(any, ultima_cifra) %>%
  summarise(frecuencia = n(), .groups = "drop")


# Heatmap solo colores
ggplot(df_heat, aes(x = factor(ultima_cifra), y = factor(any), fill = frecuencia)) +
  geom_tile(color = "white") +                               # cada celda
  scale_fill_gradient(low = "white", high = "steelblue") +  # degradado según frecuencia
  labs(
    title = "Mapa de calor de la última cifra de los números ganadores",
    subtitle = "Distribución por año",
    x = "Última cifra",
    y = "Año",
    fill = "Frecuencia"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(face = "bold"),
    axis.text.y = element_text(face = "bold"),
    plot.title = element_text(face = "bold", size = 16),
    plot.subtitle = element_text(size = 12)
  )

```

## Mapa de calor de la última cifra de los números ganados

Distribución por año



### 3.4 Discussió possibles valors atípics

Explicar com tractarem lo de que van canviant 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 Entrament del model amb el conjunt de dades

```
library(dplyr)
library(ggplot2)
library(forecast)    # para ARIMA

## Warning: package 'forecast' was built under R version 4.4.3

## Registered S3 method overwritten by 'quantmod':
##   method           from
##   as.zoo.data.frame zoo

library(tidyr)

# Suponemos que df ya tiene columnas: año, premio_num (premio numérico)
df <- df %>%
  mutate(premio_num = as.numeric(as.character(premio))) %>%
  filter(!is.na(premio_num))

# Extraer la grossa (premio más alto) por año
grossa_por_any <- df %>%
  group_by(any) %>%
  slice_max(premio_num, n = 1) %>%
  ungroup()

# Crear serie temporal
ts_grossa <- ts(grossa_por_any$premio_num, start = min(grossa_por_any$any), frequency = 1)

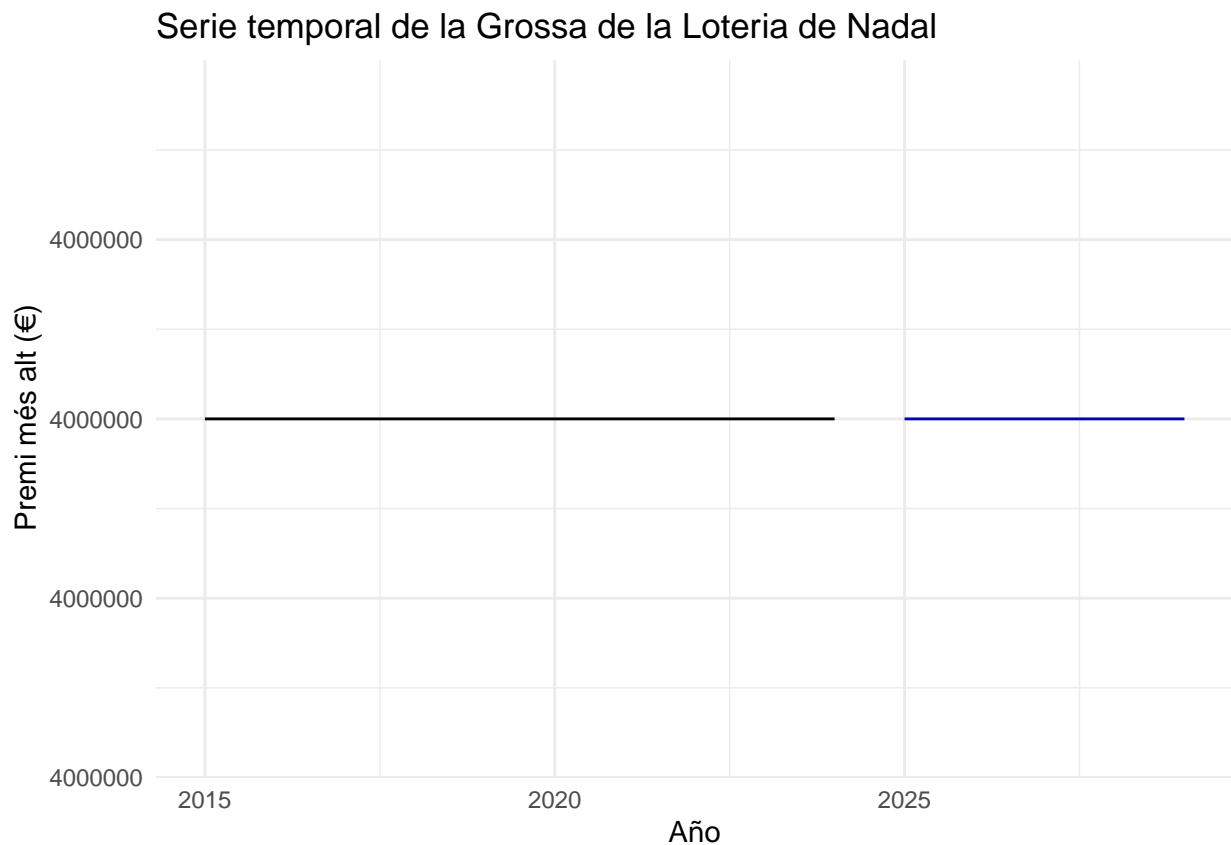
# Ajustar modelo ARIMA automáticamente
modelo_arima <- auto.arima(ts_grossa)

# Predicción a 5 años
prediccion <- forecast(modelo_arima, h = 5)
```

```

# Graficar la serie histórica + predicción
autoplot(prediccion) +
  labs(
    title = "Serie temporal de la Grossa de la Loteria de Nadal",
    x = "Año",
    y = "Premi més alt (€)"
  ) +
  theme_minimal()

```



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**