

Documentación del Análisis de Regresión Logística en Videojuegos

1. Librerías Utilizadas

```
library(tidyverse)
library(broom)
```

- **tidyverse**: Conjunto de paquetes para manipulación de datos, visualización y modelado estadístico.
- **broom**: Convierte los resultados de modelos estadísticos en data frames ordenados.

2. Carga y Limpieza de Datos

```
dates <- read_csv("Dataset/video games sales.csv", show_col_types = FALSE) %>%
select(-Name, -Rank) %>%
distinct() %>%
mutate(
Year = as.numeric(Year),
Altas_Ventas = ifelse(Global_Sales > 5, 1, 0),
Genre = as.factor(Genre),
Platform = as.factor(Platform),
Publisher = as.factor(Publisher)
) %>%
drop_na(Year, Genre, Platform, Global_Sales)
```

- Carga del archivo CSV con ventas de videojuegos.
- Eliminación de columnas irrelevantes: **Name**, **Rank**.
- Eliminación de duplicados.
- Conversión de **Year** a numérico y transformación de variables categóricas.
- Creación de variable binaria **Altas_Vventas(ventas mayores a 5 millones)**. Eliminación de valores faltantes.

3. Regresión Logística

- `model <- glm(Altas_Ventas ~ Genre + Platform + Year,
data = dates,
family = binomial)`

- Modelo de regresión logística para predecir *Altas_Ventas*. *Variables predictoras* : Genre, Platform, Year
- Familia binomial: regresión logística.

4. Evaluación del Modelo

```
probabilities <- predict(model, type = "response")
predictions <- ifelse(probabilities > 0.5, 1, 0)
real <- dates$Altas_Ventas

conf_matrix <- table(Predicho = predictions, Real = real)
accuracy <- mean(predictions == real)

print(conf_matrix)
cat("Precisión del modelo:", round(accuracy, 3), "\n")
```

- Cálculo de probabilidades y clasificación binaria.
- Comparación con valores reales y construcción de matriz de confusión.
- Cálculo de precisión del modelo.

5. Visualización de Importancia de Variables

```
tidy(model) %>%
filter(term != "(Intercept)") %>%
ggplot(aes(x = reorder(term, estimate), y = estimate)) +
geom_col(fill = "steelblue") +
coord_flip() +
labs(title = "Importancia de variables", x = "Variable", y = "Coeficiente")
```

- Visualización de coeficientes estimados.
- Análisis del impacto de cada variable en la predicción.
- Gráfico de barras horizontales para facilitar la interpretación.

6. Observaciones Finales

- El script está optimizado para ejecutarse en Visual Studio Code con soporte gráfico.
- Puede compartirse como archivo `.R` o `.Rmd`.
- Se recomienda simplificar niveles de factores si son demasiados para mejorar interpretabilidad.