

DAVID GÓMEZ, LAURA RINCÓN

---

TALLER 1  
PRYE

---

MATEMÁTICAS

2024

# Índice

<b>1. Enunciado</b>	<b>2</b>
<b>2. Desarrollo</b>	<b>2</b>
2.1. Leer el archivo csv . . . . .	2
2.2. Extraer las empresas . . . . .	2
2.3. Encontrar las mayores ventas . . . . .	2
2.4. Resultado . . . . .	3
2.5. Gráfica . . . . .	3
2.6. Tabla . . . . .	3
<b>3. Conclusiones</b>	<b>4</b>

## 1. Enunciado

El taller consistía en manejar una tabla de datos correspondientes a videojuegos. La información relacionada a cada título, en este taller, es la empresa la cual lo vende o vendió y sus ventas globales.

## 2. Desarrollo

### 2.1. Leer el archivo csv

Para lograr esto, se extrajo el archivo `xlsx` a una extensión `csv`, con el fin de poder procesarlo con Python. El procedimiento en Python fue leer el archivo usando la librería `pandas`. Un primer inconveniente fue la codificación en la cual era posible leer dicho archivo, con lo que fue necesario buscar la codificación que permitiera leerlo extensivamente. La función definida para esto es `CheckEncoding`, la cual toma como argumento la ruta del archivo a leer. Lo que se hace es tomar un arreglo con todas las codificaciones posibles e intentar una a una hasta encontrar la que no genere error. En ese momento, se imprime en consola el nombre de la codificación que no generó error. Para este caso, funcionó `'cp437'`.

### 2.2. Extraer las empresas

Posteriormente, se genera un `DataFrame` de `pandas` con el archivo leído. Para extraer los nombres de todas las empresas, se tomó la columna “Publisher”, se pasó a la estructura de conjunto, para así eliminar elementos repetidos, y se entregó como un arreglo.

### 2.3. Encontrar las mayores ventas

Para esto se generó un diccionario con los nombres de las empresas como índices y un arreglo el cual sería destinado a guardar los valores de las ventas totales de cada empresa. Luego se iteró sobre los nombres de las empresas y se filtró el `DataFrame` para obtener una tabla más pequeña en la que la columna “Publisher” coincidiera con el nombre de la iteración. Con esta tabla, se sumaron todas las entradas de la columna de las ventas globales, obteniendo así, las ventas totales de dicha empresa. Durante esta misma iteración, se guardaron en el arreglo mencionado, los valores de las ventas calculadas.

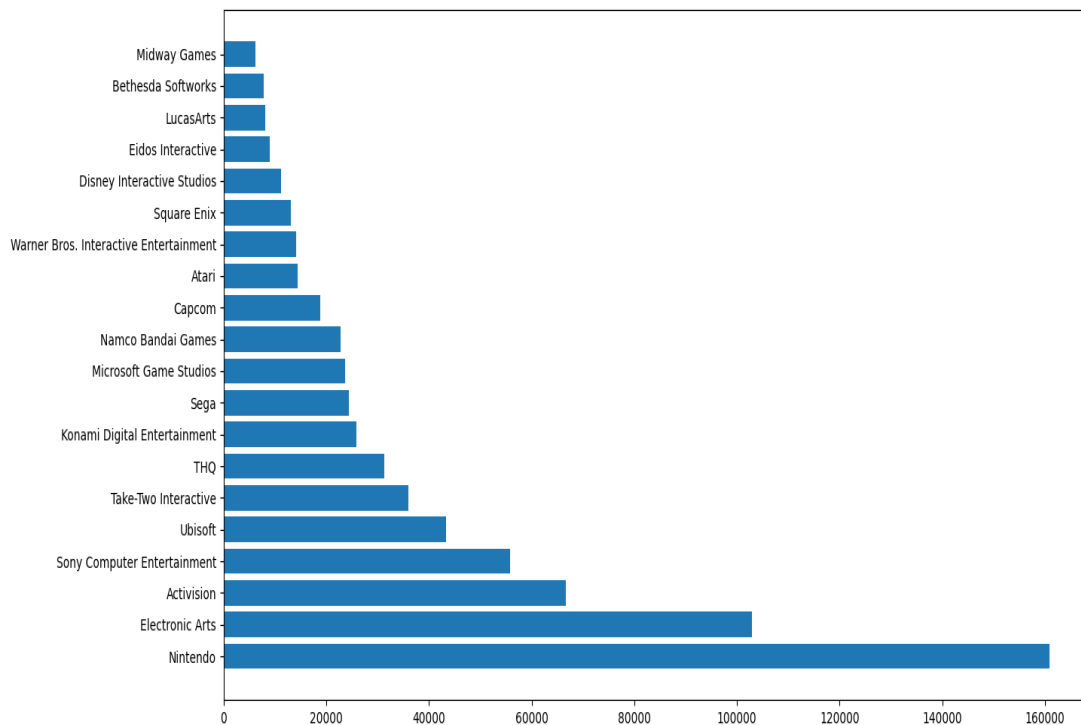
Posteriormente se realizó un diccionario el cual funcionaría como el inverso del paso anterior. Dado que existía la posibilidad de que varias empresas coincidieran en el total de ventas, las entradas para cada valor debían ser una lista con los nombres que coincidieran en el número de ventas.

Por último, se ordenó de forma decreciente el arreglo de los valores de las ventas totales y se tomaron únicamente los primeros 20 valores. Al tomar estos valores en el diccionario “inverso”, se vió que había una única empresa por cada valor en esta lista, con lo que se podía generar otro **DataFrame** en el cual se tuvieran los valores del nombre de la empresa y sus ventas totales (esto hubiera sido un poco más complicado si se tuvieran más de una empresa por valor, pues la librería no admite que un **DataFrame** tenga columnas desiguales).

Ya con estos datos, usando la librería `matplotlib.pyplot`, se generó la gráfica de barras de estas 20 empresas y sus ventas totales.

## 2.4. Resultado

## 2.5. Gráfica



## 2.6. Tabla

Puesto	Empresa	Ventas
1	Nintendo	160879
2	Electronic Arts	102969
3	Activision	66636

4	Sony Computer Entertainment	55790
5	Ubisoft	43320
6	Take-Two Interactive	36010
7	THQ	31329
8	Konami Digital Entertainment	25835
9	Sega	24424
10	Microsoft Game Studios	23570
11	Namco Bandai Games	22799
12	Capcom	18858
13	Atari	14365
14	Warner Bros. Interactive Entertainment	14031
15	Square Enix	13048
16	Disney Interactive Studios	11126
17	Eidos Interactive	8896
18	LucasArts	8100
19	Bethesda Softworks	7774
20	Midway Games	6232

### 3. Conclusiones

En esta table se evidencia una posible relación entre las ventas totales de las empresas y su popularidad. Siendo Nintendo una empresa la cual casi cualquier persona conoce.