

15 DE SEPTIEMBRE DE 2021



**Tecnológico
de Monterrey**

OBTENCIÓN DE ESTADÍSTICAS DESCRIPTIVAS

HERRAMIENTAS COMPUTACIONALES: EL ARTE DE LA ANALÍTICA

Profesor: Pedro Oscar Pérez Murueta

Julio César Gómez González - A01274966

Julio César Pérez Rodríguez - A01705763

Flavio Ruvalcaba Leija - A01367631

Antes de empezar realizamos un análisis de los datos desde Excel y nos dimos cuenta de que era mucha información la que teníamos en la database, por lo que decidimos reducir la información a solo 100 juegos.

Ya con menos cantidad de información buscamos las variables que teníamos para enfocarnos en unas pocas para poder analizar mejor los datos y buscar una relación entre ellos, por lo que nos enfocamos en la cantidad de copias vendidas y en las empresas las cuales desarrollaban los videojuegos para encontrar la que mas ventas tuviera.

Primero sacamos la media, mediana y desviación estándar para conocer cuáles son las estadísticas generales de las empresas que están dentro del top 100 de ventas. (Las ventas son equivalentes a 1 = 100,000)

```
gs = df['Global_Sales']  
print('Media:', gs.mean())  
print('Mediana:', gs.median())  
print('Desviación estándar:', gs.std())
```

```
Media: 14.619799999999989  
Mediana: 11.254999999999999  
Desviación estándar: 9.936065109417603
```

Lo que obtuvimos es que la media de ventas de un video juego en el top 100 son de 1,461,979, la mediana de las ventas equivale a 1,125,499, y la desviación estándar es de 993,606 aproximadamente.

Después buscamos cuales eran las diferentes empresas que estaban en el top 100.

```
pd.unique(df['Publisher'])
```

```
array(['Nintendo', 'Microsoft Game Studios', 'Take-Two Interactive',  
      'Sony Computer Entertainment', 'Activision', 'Ubisoft',  
      'Bethesda Softworks', 'Electronic Arts', 'Sega', 'SquareSoft',  
      'Atari'], dtype=object)
```

Realizamos las mismas operaciones con cada una de las empresas para así obtener los siguientes datos:

```
dfN = df[p == 'Nintendo']
gsN = dfN['Global_Sales']
print('Media Nintendo:', gsN.mean())
print('Mediana Nintendo:', gsN.median())
print('Desviación estándar Nintendo:', gsN.std())
```

Media Nintendo: 17.69769230769231
 Mediana Nintendo: 12.655
 Desviación estándar Nintendo: 12.618799019232487

```
dfM = df[p == 'Microsoft Game Studios']
gsM = dfM['Global_Sales']
print('Media Microsoft Game Studios:', gsM.mean())
print('Mediana Microsoft Game Studios:', gsM.median())
print('Desviación estándar Microsoft Game Studios:', gsM.std())
```

Media Microsoft Game Studios: 11.881666666666666
 Mediana Microsoft Game Studios: 9.82
 Desviación estándar Microsoft Game Studios: 5.020758574823795

```
dfTTI = df[p == 'Take-Two Interactive']
gsTTI = dfTTI['Global_Sales']
print('Media Take-Two Interactive:', gsTTI.mean())
print('Mediana Take-Two Interactive:', gsTTI.median())
print('Desviación estándar Take-Two Interactive:', gsTTI.std())
```

Media Take-Two Interactive: 14.347777777777774
 Mediana Take-Two Interactive: 13.1
 Desviación estándar Take-Two Interactive: 4.681193164615666

```
dfSCE = df[p == 'Sony Computer Entertainment']
gsSCE = dfSCE['Global_Sales']
print('Media Sony Computer Entertainment:', gsSCE.mean())
print('Mediana Sony Computer Entertainment:', gsSCE.median())
print('Desviación estándar Sony Computer Entertainment:', gsSCE.std())
```

Media Sony Computer Entertainment: 10.399999999999999
 Mediana Sony Computer Entertainment: 10.245000000000001
 Desviación estándar Sony Computer Entertainment: 2.3228307853011714

```
dfA = df[p == 'Activision']
gsA = dfA['Global_Sales']
print('Media Activision:', gsA.mean())
print('Mediana Activision:', gsA.median())
print('Desviación estándar Activision:', gsA.std())
```

Media Activision: 11.848571428571429
 Mediana Activision: 13.095
 Desviación estándar Activision: 2.631901497872877

```
dfU = df[p == 'Ubisoft']
gsU = dfU['Global_Sales']
print('Media Ubisoft:', gsU.mean())
print('Mediana Ubisoft:', gsU.median())
print('Desviación estándar Ubisoft:', gsU.std())
```

Media Ubisoft: 9.89
Mediana Ubisoft: 9.89
Desviación estándar Ubisoft: 0.5232590180780453

```
dfBS = df[p == 'Bethesda Softworks']
gsBS = dfBS['Global_Sales']
print('Media Bethesda Softworks:', gsBS.mean())
print('Mediana Bethesda Softworks:', gsBS.median())
print('Desviación estándar Bethesda Softworks:', gsBS.std())
```

Media Bethesda Softworks: 8.84
Mediana Bethesda Softworks: 8.84
Desviación estándar Bethesda Softworks: nan

```
dfEA = df[p == 'Electronic Arts']
gsEA = dfEA['Global_Sales']
print('Media Electronic Arts:', gsEA.mean())
print('Mediana Electronic Arts:', gsEA.median())
print('Desviación estándar Electronic Arts:', gsEA.std())
```

Media Electronic Arts: 7.969999999999999
Mediana Electronic Arts: 8.11
Desviación estándar Electronic Arts: 0.46092298705966067

```
dfs = df[p == 'Sega']
gsS = dfs['Global_Sales']
print('Media Sega:', gsS.mean())
print('Mediana Sega:', gsS.median())
print('Desviación estándar Sega:', gsS.std())
```

Media Sega: 8.06
Mediana Sega: 8.06
Desviación estándar Sega: nan

```
dfSq = df[p == 'SquareSoft']
gsSq = dfSq['Global_Sales']
print('Media SquareSoft:', gsSq.mean())
print('Mediana SquareSoft:', gsSq.median())
print('Desviación estándar SquareSoft:', gsSq.std())
```

Media SquareSoft: 7.86
Mediana SquareSoft: 7.86
Desviación estándar SquareSoft: nan

```
dfAt = df[p == 'Atari']
gsAt = dfAt['Global_Sales']
print('Media Atari:', gsAt.mean())
print('Mediana Atari:', gsAt.median())
print('Desviación estándar Atari:', gsAt.std())
```

```
Media Atari: 7.81
Mediana Atari: 7.81
Desviación estándar Atari: nan
```

Al ver los resultados nos damos cuenta de que algunos resultados no nos aportan tanta información de la empresa, debido a que solo cuentan con un videojuego en el top 100 por lo cual no podemos obtener su desviación estándar, además que su media y mediana no tienen con que compararse, este es el caso de Atari, SquareSoft, Sega y Bethesda Softworks por lo cual no son relevantes.

Otra información que obtenemos al realizar estas estadísticas es que Nintendo es la empresa la cual tiene mas ventas de videojuegos, ya que tiene una media de 1,769,769 copias vendidas por videojuego, lo cual la convierte en la empresa que mas copias de videojuegos vende. A pesar de esto tiene una desviación estándar muy alta también, debido a que tiene juegos muy exitosos que están hasta arriba del puesto y otros que están muy abajo.

Por el otro lado tenemos a Electronic Arts la cual es la empresa del top 100 que menos copias vende por videojuego, esto sin contar a las otras empresas que solo tienen un videojuego en el top. Electronic Arts tiene una media de venta de 796,999 por videojuego. También tiene la desviación más pequeña de todas, ya que sus videojuegos se encuentran en lugares muy cercanos en el top.

Después de los análisis podemos llegar a concluir que existe una relación entre la media de ventas con la desviación estándar, debido a que, si una aumenta, la otra también, de la misma manera si una disminuye, la otra también pero esto no siempre se va a cumplir como lo vemos en Microsoft Game Studios debido a que sus media es baja pero su desviación es mayor a comparación de Take-Two Interactive. Por lo cual, no es causa y efecto, debido a que una no afecta a la otra pero tiende a cumplir esta norma, ya que los juegos pueden llegar a tener mucha diferencia en el top, como no lo muestra Nintendo, ya que tiene muchos juegos exitosos pero al tener otros que no llegan a ser tan exitosos producen una gran variación de copias vendidas pero de lo que podemos estar seguros es que Nintendo vende muchas copias de videojuegos debido a que la empresa de Nintendo ocupa los primeros 15 puestos del top 100 de juegos mas vendidos.

---Repositorio: https://github.com/JulioGmz89/RetoAnalitica_Equipo8---