

GRÁFICOS CHEAT SHEETS

"Un gráfico dice un poco más que mil palabras"

I Bootcamp Business Intelligence

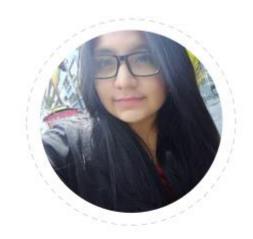


Preparado para:

MAYUMY CH

Lic. Scientific Computing | Data Scientist in training | Datalover

"Nunca te rindas, lo que es hoy es difícil, el mañana es una conquista"



- Mayumy CH
- mayumych
- heydy-mayumy-carrasco-huaccha





Caso Estudio



Se tiene una data proporcionada por Kaggle el cual corresponde a ventas de videojuegos y los scores obtenidos; Todos estos registros pertenecen al género de Acción.

Fuente: Video Game Sales with Ratings | Kaggle



Etapas de un proyecto - Ciencia de datos -



- Comprensión del Negocio
- Extracción y comprensión de los datos
- Preparación de Datos
- Modelamiento y evaluación
- Despliegue

Comprensión del Negocio

Comprensión del Negocio



- Introducción
- Antecedentes
- Objetivo

Fuente: Video Game Sales with Ratings | Kaggle

Extracción y comprensión de los datos

Datos de los juegos de Género Acción registrados

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating	GPlatforms	Has_Score
0	Assassin's Creed III	PS3	2012.0	Action	Ubisoft	2.64	2.52	0.16	1.12	6.45	85.0	41.0	6.9	1307.0	Ubisoft	М	Playstation	True
1	Assassin's Creed II	PS3	2009.0	Action	Ubisoft	2.54	1.93	0.21	0.86	5.55	91.0	70.0	8.6	1202.0	Ubisoft Montreal	М	Playstation	True
2	Assassin's Creed III	X360	2012.0	Action	Ubisoft	3.13	1.69	0.03	0.44	5.29	84.0	61.0	6.7	1196.0	Ubisoft	М	Xbox	True
3	Assassin's Creed II	X360	2009.0	Action	Ubisoft	3.11	1.55	0.08	0.51	5.26	90.0	82.0	8.8	1305.0	Ubisoft Montreal	М	Xbox	True
4	Resident Evil 5	PS3	2009.0	Action	Capcom	1.96	1.41	1.08	0.64	5.09	84.0	76.0	7.5	667.0	Capcom	М	Playstation	True



1098 registros / 18 columnas

Información

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1098 entries, 0 to 1097
Data columns (total 18 columns):
     Column
                      Non-Null Count Dtype
                      1098 non-null
                                      object
     Name
     Platform
                      1098 non-null
                                      object
     Year of Release 1093 non-null
                                      float64
                                      object
     Genre
                      1098 non-null
     Publisher
                     1098 non-null
                                      object
     NA Sales
                     1098 non-null
                                      float64
     EU Sales
                     1098 non-null
                                      float64
     JP Sales
                                     float64
                     1098 non-null
                                     float64
     Other Sales
                      1098 non-null
     Global Sales
                      1098 non-null
                                     float64
                                     float64
    Critic Score
                      580 non-null
    Critic Count
                      580 non-null
                                     float64
    User Score
                      670 non-null
                                      object
                                      float64
     User Count
                      556 non-null
     Developer
                      677 non-null
                                      object
     Rating
                      665 non-null
                                      object
16 GPlatforms
                      1098 non-null
                                      object
17 Has Score
                      1098 non-null
                                      bool
dtypes: bool(1), float64(9), object(8)
memory usage: 147.0+ KB
```

Datos Nulos

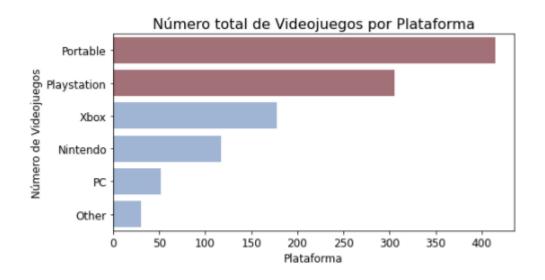
User_Count	542
Critic_Count	518
Critic_Score	518
Rating	433
User_Score	428
Developer	421
Year_of_Release	5
Has_Score	0
NA_Sales	0
Platform	0
Genre	0
Publisher	0
Other_Sales	0
EU_Sales	0
JP_Sales	0
GPlatforms	0
Global_Sales	0
Name	0
dtype: int64	

Análisis Univariante

Variable "plataformas"

65.7%

De los juegos fueron de playstation y portable

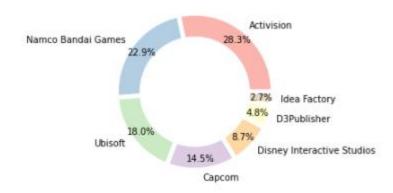


Variable "Publisher"

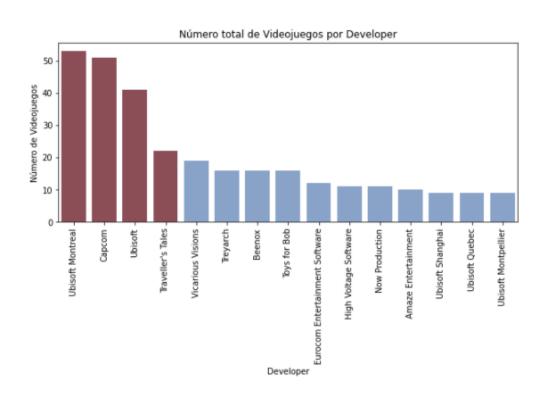
7 Estudios evaluados

2 son de EEUU y 2 de Japón y estos representan un 74.4% de total

Número total de Videojuegos por Publisher



Variable "Developer"



Los 15 developer

Que mas registros de videojuegos tienen desde 1980 -2016

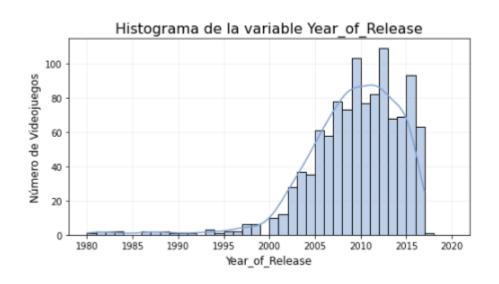
Variable "Name"



Spider, Man, Resident y Evil

son las palabras que mas aparecen en los nombres de los registros que se tienen.

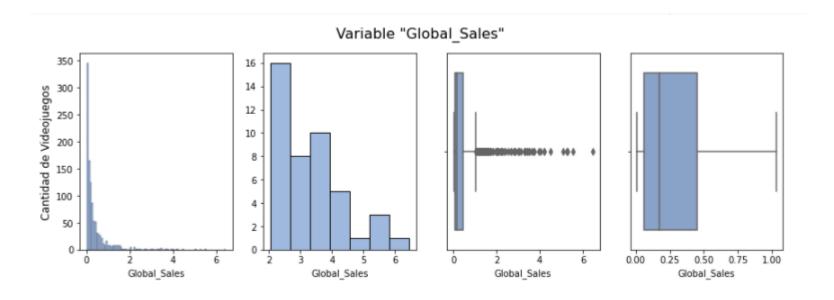
Variable "Year_of_Release"



1995 - 2016

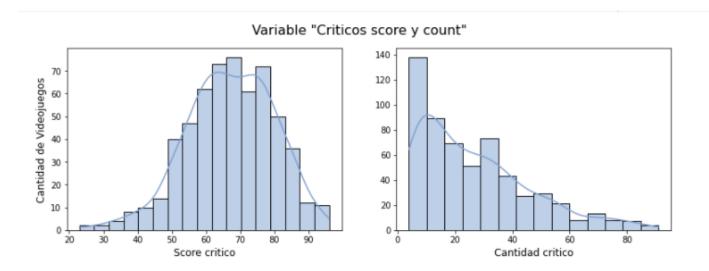
Se observa que hubo un CRECIMIENTO en la CANTIDAD de VIDEOJUEGOS.

Variable "Global_Sales"



El 50% de las ventas globales están en el rango entre 60 mil y 450 mil dólares

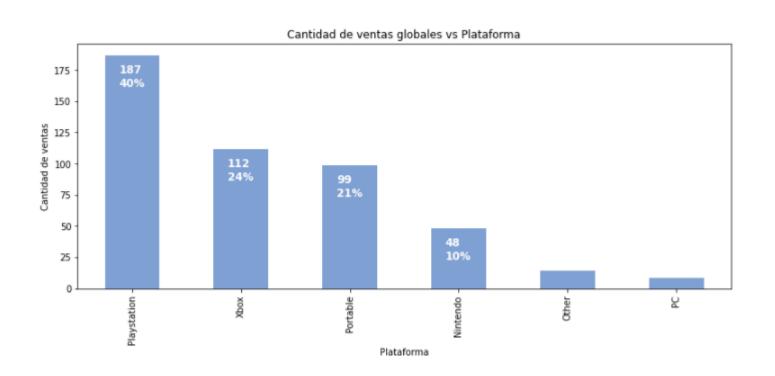
Variable "Criticos"



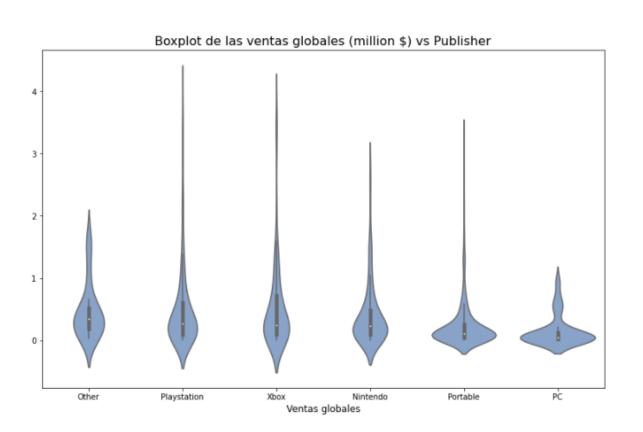
Solo el 25% de los videojuegos tienen más de 38 criticos; Siendo el máximo 91 criticos

Análisis Bivariante

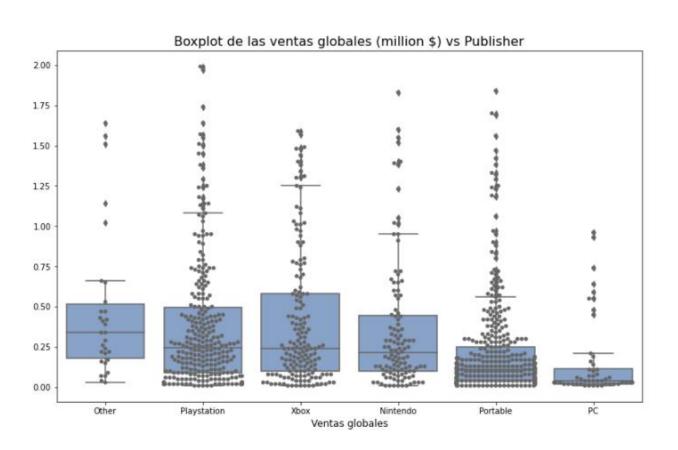
Variable "Global_Sales" vs "Platforms"



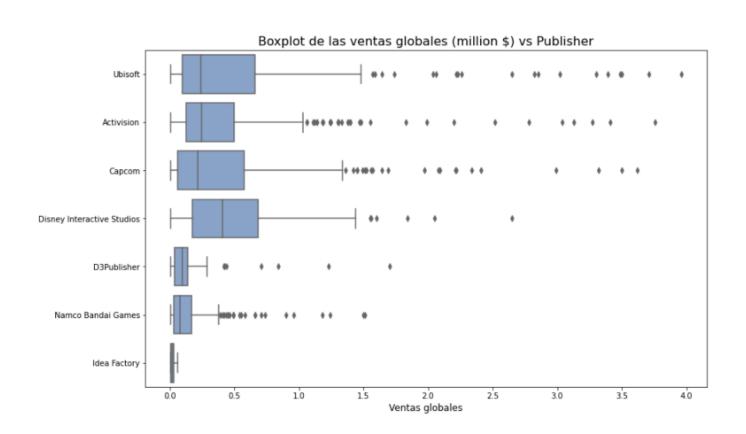
Variable "Global_Sales" vs "Platforms"



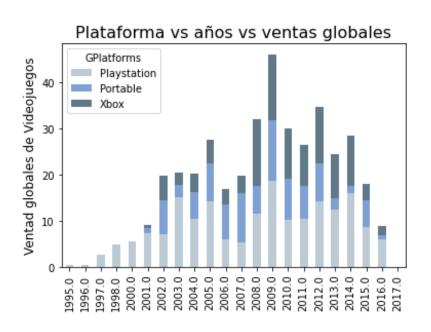
Variable "Global_Sales" vs "Platforms"



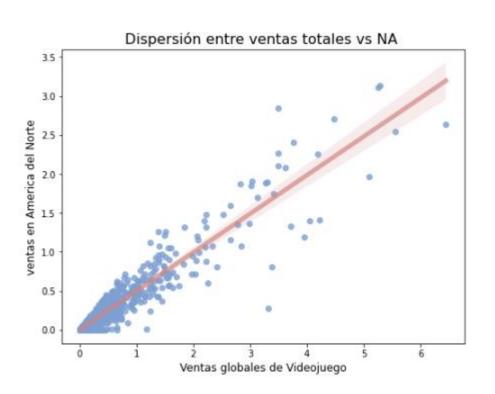
Variable "Global_Sales" vs "Publisher"



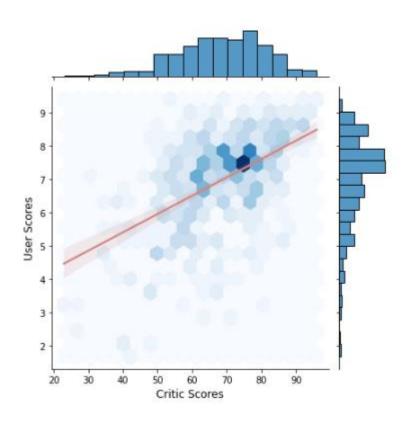
Variable "Global_Sales" vs "Publisher" vs "Year"



Variable "Global_Sales" vs "NA_Sales"

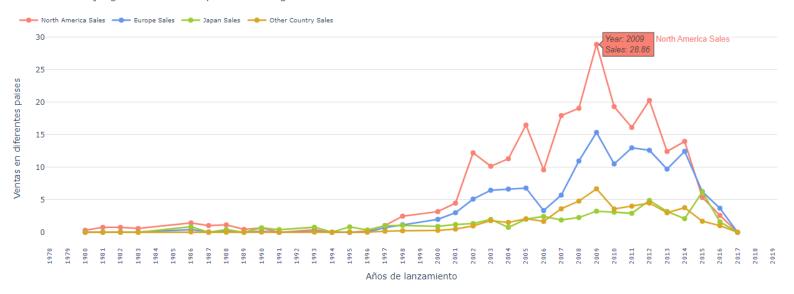


Variable "Critic_Scorevs "User_Scor"

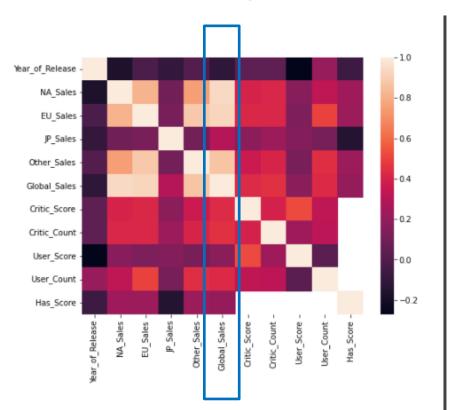


Variable "Sales" vs "Paises" vs "Year"





Correlaciones



Global Sales	1.000000
NA_Sales	0.939124
EU_Sales	0.929998
Other_Sales	0.866199
Critic_Count	0.451868
Critic_Score	0.432821
User_Count	0.428003
JP_Sales	0.289934
Has_Score	0.203509
User_Score	0.168544
Year_of_Release	-0.118113

"NA_Sales" y "EU_Sales " son las variables que estan mas correlacionadas con el target.

Pero debido a que entre ellas estan correlacionadas solo se escojera una de ellas.

"Atesora todo lo aprendido en el camino y reutiliza funciones"



Nuestras Redes Sociales

- @datasciencewomen
- @datasciencewomen
- in Data Science Women
- Data Science Women



