

PROYECTO FINAL
ANÁLISIS DE DATOS

Integrantes:

Jean Tituaña

Jhael Nicolalde

Andy Loor

Índice de Contenido y Figuras

Contenido

Índice de Contenido y Figuras.....	2
Índice de Ilustraciones	3
1. OBJETIVOS.....	4
1.1 OBJETIVOS GENERALES	4
1.2 OBJETIVOS ESPECIFICOS.....	4
2. HERRAMIENTAS UTILIZADAS.....	4
3. PROCESOS DE ANÁLISIS DE DATOS	5
3.1 Extracción de datos.....	5
3.2 Limpieza de datos	5
3.3 Visualización e Interpretación de Datos	5
4. ILUSTRACIONES	6
4.1 Extracción de Datos	6
4.2 Limpieza de datos	7
4.3 Visualización de Datos	7
5. CONCLUSIONES	9
6. RECOMENDACIONES.....	9
7. BIBLIOGRAFÍA.....	9

Índice de Ilustraciones

ILUSTRACIÓN 1 INSTALACIÓN DE LIBRERÍAS DE MANIPULACIÓN DE DATOS.....	6
ILUSTRACIÓN 2 INSTALACIÓN DE LIBRERÍAS DE MANIPULACIÓN DE DATOS.....	6
ILUSTRACIÓN 3 CODIFICACIÓN PARA LA EXTRACCIÓN Y PRESENTACIÓN DE DATOS	6
ILUSTRACIÓN 4 LIMPIEZA DE COLUMNAS DE DATOS.....	7
ILUSTRACIÓN 5 LIMPIEZA DE COLUMNAS DE DATOS.....	7
ILUSTRACIÓN 6 LIMPIEZA DE COLUMNAS DE DATOS.....	8
ILUSTRACIÓN 7 VENTAS POR PLATAFORMA	8

1. OBJETIVOS

1.1 OBJETIVOS GENERALES

Realizar el análisis de cualquier tipo de documento o enlace (Xml, Txt, Pdf, Web) mediante la manipulación de datos para su extracción e interpretación de estos.

1.2 OBJETIVOS ESPECIFICOS

- [1] Obtención de datos
- [2] Limpieza de datos
- [3] Visualización de datos
- [4] Interpretación de resultados

2. HERRAMIENTAS UTILIZADAS

Colaboratory

Colaboratory, o "Colab" para abreviar, es un producto de Google Research. Permite a cualquier usuario escribir y ejecutar código arbitrario de Python en el navegador. Es especialmente adecuado para tareas de aprendizaje automático, análisis de datos y educación. Desde un punto de vista más técnico, Colab es un servicio de cuaderno alojado de Jupyter que no requiere configuración y que ofrece acceso sin coste adicional a recursos informáticos, como GPUs.

Python

Un lenguaje sencillo, legible y elegante que atiende a un conjunto de reglas que hacen muy corta su curva de aprendizaje. Si ya tienes unas nociones de programación o vienes de programar en otros lenguajes como Java no te será difícil comenzar a leer y entender el código desarrollado en Python.

Numpy

Librería de Python para crear vectores (arrays de una dimensión) y matrices (arrays multidimensionales) grandes multidimensionales, junto con una gran colección de funciones matemáticas para operar con ellas.

Pandas

Es una librería considerada como extensión de NumPy para la manipulación y el análisis de datos. Permite el análisis, limpieza y preparación de datos.

Matplotlib

Permite crear y personalizarlos tipos de gráficos más comunes y las combinaciones entre ellos. El código para su instalación es ***conda install matplotlib/pip install matplotlib***

Seaborn

Seaborn está basada en matplotlib y proporciona una interfaz de alto nivel para realizar gráficos estadísticos de forma vistosa (requiere de las librerías numpy, pandas, matplotlib). El código para su instalación es ***conda install seaborn/pip install seaborn***

3. PROCESOS DE ANÁLISIS DE DATOS

3.1 Extracción de datos

1. Se instalaron las varias librerías para la extracción y manipulación de datos (Numpy, Pandas).
2. Se importaron las diferentes librerías para la extracción y manipulación de datos (Numpy, Pandas).
3. Se codificó secuencialmente para extraer los datos de un documento .csv con el lenguaje Python.
 - I. Se creó una variable para leer el documento .csv con la extensión de Pandas.
 - II. Se mostró la variable misma que refleja los resultados de los datos del documento.

3.2 Limpieza de datos

1. Una vez obtenido los datos, se procedió a limpiar las columnas que contenían datos NaN (Ver ilustración 4)
2. Se realizó el cambio de nombres

3.3 Visualización e Interpretación de Datos

1. Las ventas de Konami, Sega y Namco no superan las 500 ventas, por otro lado, la empresa Nintendo logra superar las 1750 ventas, que son más del triple que las ventas de Sony esto durante el año 2016. (ver ilustración 5)
2. Los juegos de género de acción y deportes son las ventas más altas. Mientras los juegos para los géneros de Estrategia y Puzzle son los más bajos, los cuales se encuentran dentro de un rango de entre 500 y 1000 ventas durante el año 2016 (ver ilustración 6).
3. La PS2, DS tienen las ventas globales totales más altas. La XB, GBA tienen las ventas globales totales más bajas (ver ilustración 7).

4. ILUSTRACIONES

4.1 Extracción de Datos

```
#Instalación de librerías para la extracción de datos
!pip install numpy
!pip install pandas
!pip install matplotlib
!pip install seaborn

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: numpy in /usr/local/lib/python3.7/dist-packages (1.21.6)
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (1.3.5)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (from pandas) (2022.2.1)
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.7/dist-packages (from pandas) (1.21.6)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from python-dateutil>=2.7.3->pandas) (1.15.0)
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (3.2.2)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib) (2.8.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from matplotlib) (0.11.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (from matplotlib) (1.4.4)
```

Ilustración 1 Instalación de librerías de manipulación de datos

```
#Importación de librerías para la extracción de datos
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Ilustración 2 Instalación de librerías de manipulación de datos

```
#Extracción de los datos
xml=pd.read_csv("/Videos/Games_Sales_es_at_22_Dec_2015.csv")
xml
```

	Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	Critic_Cou
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	6.45	82.53	76.0	5
1	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.56	6.81	0.77	40.24	NaN	N
2	Mario Kart Wii	Wii	2008.0	Racing	Nintendo	15.66	12.76	3.79	3.29	35.52	82.0	7
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77	80.0	7
4	Pokémon Red/Pokémon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37	NaN	N
...
16714	Samurai Warriors: Sanada Maru	PS3	2010.0	Action	Tecmo Koei	0.00	0.00	0.01	0.00	0.01	NaN	N

Ilustración 3 Codificación para la extracción y presentación de datos

4.2 Limpieza de datos

	Nombre	Plataforma	Año	Genero	Editor	Ventas_NA	Ventas_EU	Ventas_3P	Otras_Ventas	Ventas_Globales
0	Wii Sports	Wii	2006.0	Sports	Nintendo	41.36	28.96	3.77	8.45	82.53
1	Super Mario Bros.	NES	1985.0	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	Mario Kart Wii	Wii	2006.0	Racing	Nintendo	15.68	12.76	3.79	3.29	35.52
3	Wii Sports Resort	Wii	2009.0	Sports	Nintendo	15.61	10.93	3.28	2.95	32.77
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37
...
16714	Samurai Warriors: Sanada Maru	PS3	2016.0	Action	Tecmo Koei	0.00	0.00	0.01	0.00	0.01
16715	LMA Manager 2007	X360	2006.0	Sports	Codemasters	0.00	0.01	0.00	0.00	0.01
16716	Haitaka no Psychodelica	PSV	2016.0	Adventure	Idea Factory	0.00	0.00	0.01	0.00	0.01
16717	Spirits & Spells	GBA	2003.0	Platform	Wanadoo	0.01	0.00	0.00	0.00	0.01
16718	Winning Post 8 2016	PSV	2016.0	Simulation	Tecmo Koei	0.00	0.00	0.01	0.00	0.01

Ilustración 4 Limpieza de columnas de datos

4.3 Visualización de Datos

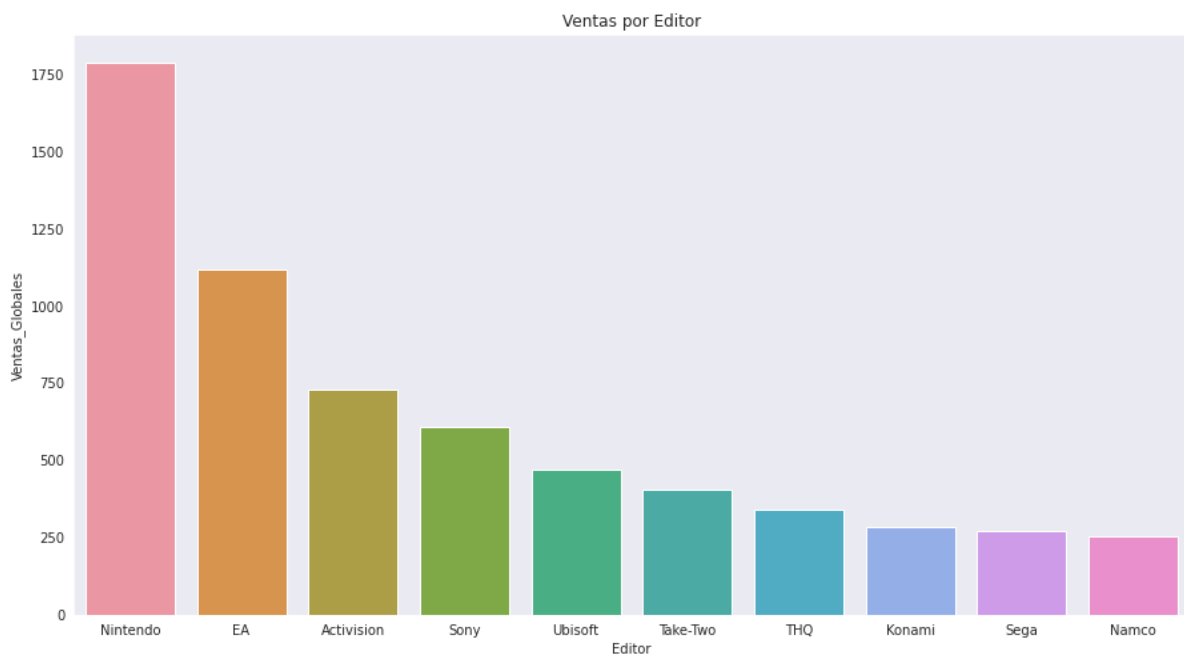


Ilustración 5 Limpieza de columnas de datos

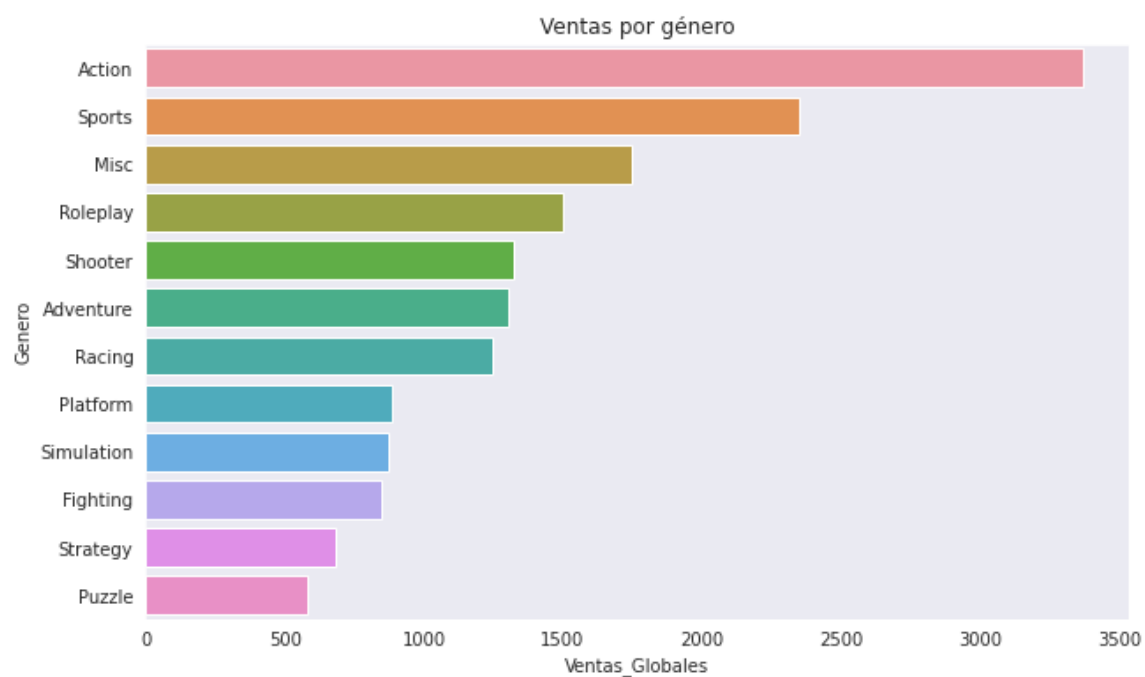


Ilustración 6 Limpieza de columnas de datos

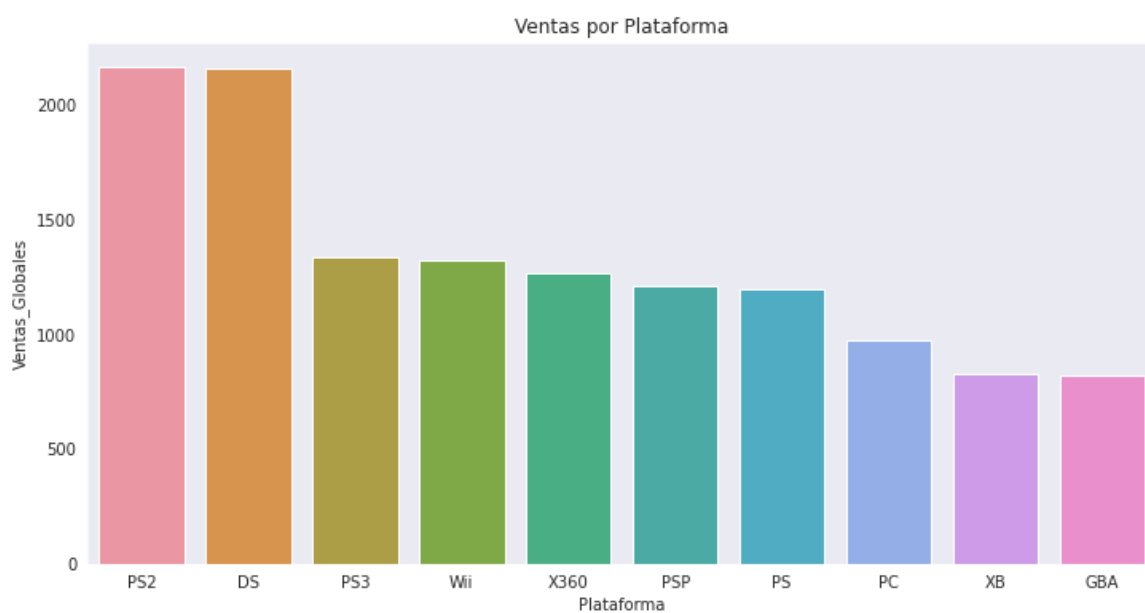


Ilustración 7 Ventas por Plataforma

5. CONCLUSIONES

1. El uso de un gráfico de barras nos permite visualizar los datos en diferentes categorías, como, por ejemplo, por editor, género y plataforma. De manera que, podremos comparar y mostrar los datos a través de barras tanto horizontales como verticales, las ventas de videojuegos que se realizaron en el año 2016
2. Por medio del ejemplo expuesto se puede determinar que es necesario eliminar los diferentes tipos de datos que se encuentran sin valor, ya que es importante para poder realizar su respectivo análisis y graficas estadísticas.
3. Finalmente, luego de un riguroso análisis a los diferentes tipos de datos obtenidos en la extracción de datos con las libreas de manipulación de datos, misma que dieron paso a concluir que el mayor consumo en cuanto a los géneros de juegos en el mundo gaming son de deporte y acción.

6. RECOMENDACIONES

- Se recomienda utilizar nombres sencillos para los títulos de las columnas, así se evitará errores y será más sencillo al momento de crear las gráficas utilizando estas columnas del dataframe.
- Se recomienda usar el método figure de la librería matplotlib para aumentar el tamaño de las gráficas para una mejor visualización.
- Se recomienda que, en el proceso de la limpieza de datos, además de limpiar datos no válidos, también se deberá eliminar columnas y/o filas que no tengan un contenido resaltable el cual deje un resultado con cifras de comparación ante cualquier tema después de su análisis.

7. BIBLIOGRAFÍA

- [1] Kirubi, R. (2016, 10 septiembre). Video Games Sales Dataset. Kaggle. Recuperado 10 de septiembre de 2022, de <https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings>
- [2] Álvarez, I. Preparación de datos. Recuperado 10 de septiembre de 2022, de <https://aulasvirtuales.epn.edu.ec/mod/resource/view.php?id=4795447>
- [3] Álvarez, I. Visualización de datos. Recuperado 10 de septiembre de 2022, de <https://aulasvirtuales.epn.edu.ec/mod/resource/view.php?id=4827217>
- [4] Google. (s. f.). Colaboratory. Google Research. Recuperado 11 de septiembre de 2022, de <https://research.google.com/colaboratory/intl/es/faq.html>
- [5] Robledano, A. (2019, 23 septiembre). Qué es Python: Características, evolución y futuro. Open Webinars. Recuperado 11 de septiembre de 2022, de <https://openwebinars.net/blog/que-es-python>