**Instrucciones para completar el proyecto**

Paso 1. Abre el archivo de datos y estudia la información general

Ruta de archivo:

Paso 2. Prepara los datos

* Reemplaza los nombres de las columnas (ponlos en minúsculas).
* Convierte los datos en los tipos necesarios.
* Describe las columnas en las que los tipos de datos han sido cambiados y explica por qué.
* Si es necesario, elige la manera de tratar los valores ausentes:
  + Explica por qué rellenaste los valores ausentes como lo hiciste o por qué decidiste dejarlos en blanco.
  + ¿Por qué crees que los valores están ausentes? Brinda explicaciones posibles.
  + Presta atención a la abreviatura TBD: significa "to be determined" (a determinar). Especifica cómo piensas manejar estos casos.
* Calcula las ventas totales (la suma de las ventas en todas las regiones) para cada juego y coloca estos valores en una columna separada.

Paso 3. Analiza los datos

* Mira cuántos juegos fueron lanzados en diferentes años. ¿Son significativos los datos de cada período?
* Observa cómo varían las ventas de una plataforma a otra. Elige las plataformas con las mayores ventas totales y construye una distribución basada en los datos de cada año. Busca las plataformas que solían ser populares pero que ahora no tienen ventas. ¿Cuánto tardan generalmente las nuevas plataformas en aparecer y las antiguas en desaparecer?
* Determina para qué período debes tomar datos. Para hacerlo mira tus respuestas a las preguntas anteriores. Los datos deberían permitirte construir un modelo para 2017.
* Trabaja solo con los datos que consideras relevantes. Ignora los datos de años anteriores.
* ¿Qué plataformas son líderes en ventas? ¿Cuáles crecen y cuáles se reducen? Elige varias plataformas potencialmente rentables.
* Crea un diagrama de caja para las ventas globales de todos los juegos, desglosados por plataforma. ¿Son significativas las diferencias en las ventas? ¿Qué sucede con las ventas promedio en varias plataformas? Describe tus hallazgos.
* Mira cómo las reseñas de usuarios y profesionales afectan las ventas de una plataforma popular (tu elección). Crea un gráfico de dispersión y calcula la correlación entre las reseñas y las ventas. Saca conclusiones.
* Teniendo en cuenta tus conclusiones compara las ventas de los mismos juegos en otras plataformas.
* Echa un vistazo a la distribución general de los juegos por género. ¿Qué se puede decir de los géneros más rentables? ¿Puedes generalizar acerca de los géneros con ventas altas y bajas?

Paso 4. Crea un perfil de usuario para cada región

Para cada región (NA, UE, JP) determina:

* Las cinco plataformas principales. Describe las variaciones en sus cuotas de mercado de una región a otra.
* Los cinco géneros principales. Explica la diferencia.
* Si las clasificaciones de ESRB afectan a las ventas en regiones individuales.

Paso 5. Prueba las siguientes hipótesis:

— Las calificaciones promedio de los usuarios para las plataformas Xbox One y PC son las mismas. Como el pvalue es mayor que el nivel de significancia no se puede rechazar la H nula.

— Las calificaciones promedio de los usuarios para los géneros de Acción y Deportes son diferentes.

Establece tú mismo el valor de umbral *alfa*.

Explica:

— Cómo formulaste las hipótesis nula y alternativa.

— Qué criterio utilizaste para probar las hipótesis y por qué.

Paso 6. Escribe una conclusión general

Formato: Completa la tarea en Jupyter Notebook. Inserta el código de programación en las celdas *code* y las explicaciones de texto en las celdas *markdown*. Aplica formato y agrega encabezados.

**Descripción de datos**

— *Name* (Nombre)

— *Platform* (Plataforma)

— *Year\_of\_Release* (Año de lanzamiento)

— *Genre* (Género)

— *NA\_sales* (ventas en Norteamérica en millones de dólares estadounidenses)

— *EU\_sales* (ventas en Europa en millones de dólares estadounidenses)

— *JP\_sales* (ventas en Japón en millones de dólares estadounidenses)

— *Other\_sales* (ventas en otros países en millones de dólares estadounidenses)

— *Critic\_Score* (máximo de 100)

— *User\_Score* (máximo de 10)

— *Rating* (ESRB)

Es posible que los datos de 2016 estén incompletos.

viceversa