

Sprint 6 - Proyecto

Proyecto integrado

¡Enhorabuena! Has completado oficialmente la primera parte del curso en la plataforma interactiva. Llegó la hora de reunir todo lo que has aprendido hasta el momento en tu primer proyecto integrado, un estudio de caso analítico de la vida real.

Cuando finalices el proyecto envía tu trabajo al revisor de proyecto para su evaluación. Te dará su opinión en 48 horas. Utiliza los comentarios para realizar cambios y luego envía la nueva versión al revisor del proyecto.

Puede que recibas aún más feedback en la nueva versión. Esto es totalmente normal. Es común pasar por varios ciclos de comentarios y revisiones.

Tu proyecto se considerará completado una vez que el revisor del proyecto lo apruebe.

Descripción del proyecto

Trabajas para la tienda online Ice que vende videojuegos por todo el mundo. Las reseñas de usuarios y expertos, los géneros, las plataformas (por ejemplo, Xbox o PlayStation) y los datos históricos sobre las ventas de juegos están disponibles en fuentes abiertas. Tienes que identificar

patrones que determinen si un juego tiene éxito o no. Esto te permitirá detectar proyectos prometedores y planificar campañas publicitarias.

Delante de ti hay datos que se remontan a 2016. Imaginemos que es diciembre de 2016 y estás planeando una campaña para 2017.

Lo importante es adquirir experiencia de trabajo con datos. Realmente no importa si estás pronosticando las ventas de 2017 en función de los datos de 2016 o las ventas de 2027 en función de los datos de 2026.

El dataset contiene una columna "rating" que almacena la clasificación ESRB de cada juego. El Entertainment Software Rating Board (la Junta de clasificación de software de entretenimiento) evalúa el contenido de un juego y asigna una clasificación de edad como Adolescente o Adulto.

Instrucciones para completar el proyecto

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

Ruta de archivo:

`/datasets/games.csv` . [Descarga el dataset](#)

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.

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

Establece tu 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.

¿Cómo será evaluado mi proyecto?

Lee atentamente estos criterios de evaluación de proyectos antes de empezar a trabajar.

Esto es lo que buscan los revisores de proyecto cuando evalúan tu proyecto:

- ¿Cómo describirías los problemas identificados en los datos?
- ¿Cómo se prepara un dataset para el análisis?
- ¿Cómo creas gráficos de distribución y cómo los explicas?
- ¿Cómo calculas la desviación estándar y varianza?
- ¿Formulas las hipótesis alternativas y nulas?
- ¿Qué métodos aplicas a la hora de probarlos?
- ¿Explicas los resultados de tus pruebas de hipótesis?
- ¿Sigues la estructura del proyecto y mantienes tu código ordenado y comprensible?
- ¿A qué conclusiones llegas?
- ¿Has dejado comentarios claros y relevantes en cada paso?