

Proyecto de pruebas A/B

Ejercicio

Has recibido una tarea analítica de una tienda en línea internacional. Tus predecesores no consiguieron completarla: lanzaron una prueba A/B y luego abandonaron (para iniciar una granja de sandías en Brasil). Solo dejaron las especificaciones técnicas y los resultados de las pruebas.

Descripción técnica

- Nombre de la prueba: `recommender_system_test`
- Grupos: A (control), B (nuevo embudo de pago)
- Fecha de lanzamiento: 2020-12-07
- Fecha en la que dejaron de aceptar nuevos usuarios: 2020-12-21
- Fecha de finalización: 2021-01-01
- Audiencia: 15% de los nuevos usuarios de la región de la UE
- Propósito de la prueba: probar cambios relacionados con la introducción de un sistema de recomendaciones mejorado
- Resultado esperado: dentro de los 14 días posteriores a la inscripción, los usuarios mostrarán una mejor conversión en vistas de la página del producto (el evento `product_page`), instancias de agregar artículos al carrito de compras (`product_cart`) y compras (`purchase`). En cada etapa del embudo `product_page` → `product_cart` → `purchase`, habrá al menos un 10% de aumento.
- Número previsto de participantes de la prueba: 6 000

Descarga los datos de la prueba, comprueba si se ha realizado correctamente y analiza los resultados.

Descripción de los datos

Descarga los datasets de [Notion](#).

Para acceder a los datasets de la plataforma, agrega `/datasets/` al principio de la ruta del archivo (por ejemplo, `/datasets/ab_project_marketing_events_us.csv`).

- `ab_project_marketing_events_us.csv` — el calendario de eventos de marketing para 2020
- `final_ab_new_users_upd_us.csv` — todos los usuarios que se registraron en la tienda en línea desde el 7 hasta el 21 de diciembre de 2020
- `final_ab_events_upd_us.csv` — todos los eventos de los nuevos usuarios en el período comprendido entre el 7 de diciembre de 2020 y el 1 de enero de 2021
- `final_ab_participants_upd_us.csv` — tabla con los datos de los participantes de la prueba

Estructura ab_project__marketing_events_us.csv:

- name — el nombre del evento de marketing
- regions — regiones donde se llevará a cabo la campaña publicitaria
- start_dt — fecha de inicio de la campaña
- finish_dt — fecha de finalización de la campaña

Estructura final_ab_new_users_upd_us.csv:

- user_id
- first_date — fecha de inscripción
- region
- device — dispositivo utilizado para la inscripción

Estructura final_ab_events_upd_us.csv:

- user_id
- event_dt — fecha y hora del evento
- event_name — nombre del tipo de evento
- details — datos adicionales sobre el evento (por ejemplo, el pedido total en USD para los eventos purchase)

Estructura final_ab_participants_upd_us.csv:

- user_id
- ab_test — nombre de la prueba
- group — el grupo de prueba al que pertenecía el usuario

Instrucciones para completar la tarea

- Describe los objetivos del estudio.
- Explora los datos:
 - ¿Es necesario convertir los tipos?
 - ¿Hay valores ausentes o duplicados? Si es así, ¿cómo los caracterizarías?
- Lleva a cabo el análisis exploratorio de datos:
 - Estudia la conversión en las diferentes etapas del embudo.
 - ¿El número de eventos por usuario está distribuido equitativamente entre las muestras?
 - ¿Hay usuarios que están presentes en ambas muestras?
 - ¿Cómo se distribuye el número de eventos entre los días?

- ¿Hay alguna peculiaridad en los datos que hay que tener en cuenta antes de iniciar la prueba A/B?
- Evaluar los resultados de la prueba A/B:
 - ¿Qué puedes decir sobre los resultados de la prueba A/B?
 - Utiliza una prueba z para comprobar la diferencia estadística entre las proporciones.
- Describe tus conclusiones con respecto a la etapa EDA y los resultados de la prueba A/B.