

Procesamiento de datos con reservas de viajes (NCR Ride Bookings)

Contexto

Tienes un registro de 150.000 reservas de viajes con información de cliente, estado de la reserva, tipo de vehículo, ubicaciones de origen/destino, tiempos (VTAT/CTAT), valor del viaje, distancia, calificaciones y método de pago.

Objetivos de aprendizaje

1. Limpiar y estandarizar datos (nombres de columnas, tipos, duplicados, faltantes, outliers).
2. Enriquecer con variables temporales a partir de fecha y hora.
3. Analizar y resumir métricas clave del negocio.

Dataset (columnas principales)

- Date, Time, Booking ID, Booking Status, Customer ID
- Vehicle Type, Pickup Location, Drop Location
- Avg VTAT, Avg CTAT
- Cancelled Rides by Customer / Driver + razones
- Incomplete Rides + razón
- Booking Value, Ride Distance
- Driver Ratings, Customer Rating
- Payment Method

Tareas

1. Carga & perfilado rápido
 - Carga el CSV
2. Estandarización
 - Normaliza nombres de columnas a snake_case.
 - Convierte tipos (numéricos/categoricos) y normaliza categorías (minúsculas, sin espacios).
3. Tiempo & features
 - Une Date + Time → datetime.
 - Detecta y elimina duplicados por booking_id.

4. Limpieza

- ¿Cómo crees que se pueden identificar outliers?
- Guarda un CSV limpio.

5. Métricas del negocio

- Total de ingresos.
- Distancia promedio.
- Tasa aparente de cancelación (filas con "cancel" en `booking_status`).