

Descripción del conjunto de datos: Transporte masivo

Introducción

Este documento describe el conjunto de datos recopilado sobre el uso de transporte masivo en diferentes estaciones, horas del día y condiciones climáticas. El objetivo es analizar los patrones de demora en el transporte con base en diversas características.

Descripción de las columnas

1. ID_usuario:

- **Descripción:** Identificador único asignado a cada usuario que utilizó el transporte.
- **Tipo de dato:** Numérico (entero).
- **Ejemplo:** 1000, 1001, 1002, etc.
- **Observaciones:** Este campo no tiene un impacto directo en el análisis de las demoras, pero es útil para identificar cada registro de manera única.

2. Hora:

- **Descripción:** Hora del día en la que el usuario realizó el viaje.
- **Tipo de dato:** Numérico (entero).
- **Rango:** 0 - 23 (en formato de 24 horas).
- **Ejemplo:** 7, 10, 19, 22.
- **Observaciones:** La hora es importante para identificar patrones de tráfico o congestión en diferentes momentos del día.

3. Origen:

- **Descripción:** Estación desde la cual el usuario inició su viaje.
- **Tipo de dato:** Categórico (cadena de texto).
- **Ejemplo:** Estacion_A, Estacion_B, Estacion_C, Estacion_D.
- **Observaciones:** Este campo permite analizar si algunas estaciones son más propensas a demoras en comparación con otras.

4. Destino:

- **Descripción:** Estación a la cual el usuario se dirigía.
- **Tipo de dato:** Categórico (cadena de texto).
- **Ejemplo:** Estacion_A, Estacion_B, Estacion_C, Estacion_D.
- **Observaciones:** Similar al campo "Origen", este campo es útil para identificar rutas que podrían tener más o menos demoras.

5. Transporte:

- **Descripción:** Tipo de transporte utilizado por el usuario.
- **Tipo de dato:** Categórico (cadena de texto).
- **Ejemplo:** Metro, Autobus, Tren.
- **Observaciones:** Se pueden realizar comparaciones entre diferentes tipos de transporte para evaluar cuál es más eficiente o propenso a demoras.

6. Demora:

- **Descripción:** Tiempo de demora en minutos que experimentó el usuario

durante su viaje.

- **Tipo de dato:** Numérico (entero).
- **Ejemplo:** 0, 2, 5, 7.
- **Observaciones:** Esta es la variable objetivo de análisis. Se busca predecir este valor en función de las demás características.

7. Clima:

- **Descripción:** Condición climática durante el viaje del usuario.
- **Tipo de dato:** Categórico (cadena de texto).
- **Ejemplo:** Soleado, Nublado, Lluvia, Nieve.
- **Observaciones:** El clima puede tener un impacto considerable en el tiempo de viaje, con algunas condiciones climáticas aumentando las demoras.

Ejemplo de datos

ID_usuario	Hora	Origen	Destino	Transporte	Demora	Clima
1000	22	Estacion_A	Estacion_C	Metro	7	Nublado
1001	7	Estacion_A	Estacion_D	Metro	4	Nublado
1002	14	Estacion_A	Estacion_A	Metro	8	Lluvia
1003	5	Estacion_A	Estacion_C	Metro	4	Lluvia
1004	5	Estacion_B	Estacion_A	Metro	5	Nublado

Observaciones

- **Hora:** Se puede analizar cómo las demoras varían en horas punta (ej. 7:00 - 9:00 y 17:00 - 19:00).
- **Origen/Destino:** Es posible investigar si ciertos trayectos, como entre las estaciones A y C, tienen más demoras.
- **Transporte:** La comparación entre el Metro, Autobus y Tren puede mostrar cuál es el transporte más eficiente.
- **Clima:** Las condiciones climáticas adversas como lluvia y nieve pueden correlacionarse con un aumento en las demoras.

Conclusiones

Este conjunto de datos ofrece una base sólida para analizar el comportamiento del sistema de transporte masivo en función de diversos factores, como la hora del día, el clima y el tipo de transporte. A partir de esta información, se pueden desarrollar modelos predictivos para mejorar la gestión de los sistemas de transporte y reducir las demoras.