

Carrier Data Managment: Prueba de habilidades

Instrucciones

Tomando los datos del set **Car Insurance Claim** contesta los siguientes puntos utilizando como herramienta cualquier lenguaje para análisis de datos (Python, R, SQL, Julia, etc). El set de datos incluye las siguientes variables:

1. **policy_id**: Identificador único del titular de la póliza.
2. **policy_tenure**: Período de tiempo de la póliza.
3. **age_of_car**: Edad normalizada del coche en años.
4. **age_of_policyholder**: Edad normalizada del titular de la póliza en años.
5. **area_cluster**: Cluster o zona geográfica del titular de la póliza.
6. **population_density**: Densidad poblacional de la ciudad del titular.
7. **make**: Fabricante del coche (codificado).
8. **segment**: Segmento del coche (A/B1/B2/C1/C2).
9. **model**: Nombre del modelo del coche (codificado).
10. **fuel_type**: Tipo de combustible usado por el coche.
11. **max_torque**: Máximo torque generado por el coche (Nm @ rpm).
12. **max_power**: Potencia máxima generada por el coche (bhp @ rpm).
13. **engine_type**: Tipo de motor usado en el coche.
14. **airbags**: Número de airbags instalados en el coche.
15. **transmission_type**: Tipo de transmisión del coche.
16. **is_claim**: Compensación por daños o pérdidas en el coche.

Para todos los puntos deben agregar el código o pasos con los que obtuvieron el resultado. Pueden pegarlos en este file y además enviar su script (.py), jupyter notebook (.ipynb), quarto (.qmd), etc.

Además, se debe enviar el nuevo data set con lo solicitado en el punto 1.c.

1. Limpieza de Datos

- a. ¿Existen valores faltantes en alguna de las variables?
- b. En caso de existir, ¿Podemos hacer algo para resolverlo?
- c. Para la variable *max power* (potencia máxima del motor), divide el campo en dos: caballos de fuerza (bhp) y revoluciones por minuto (rpm).

Hint: Sepáralos por el patrón @

2. Análisis de Datos Exploratorio (EDA):

Para este análisis se debe tomar en cuenta que la variable de interés es **is_claim**. Esta variable representa si el asegurado solicita a la compañía de seguros un pago o compensación por los daños o pérdidas cubiertas por la póliza de seguro.

Valor de **1** → el asegurado presentó reclamación.

Valor de **0** → no ha presentado reclamación.

- a. ¿Cuáles son los principales patrones que podemos observar entre los asegurados que presentaron alguna reclamación y los que no presentaron?
- b. ¿Cómo se asocian las características del vehículo como la antigüedad, transmisión automática, bolsas de aire, cilindros en la probabilidad de presentar un *claim*?

(Presentar visualizaciones)