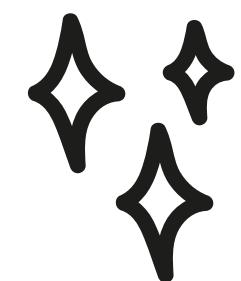
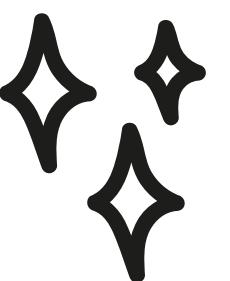


PROYECTO EVALUACIÓN DE AUTOMÓVILES



MINERIA DE DATOS

NICOLAS TELLO MENDEZ
JERSSON FABIAN BUITRAGO



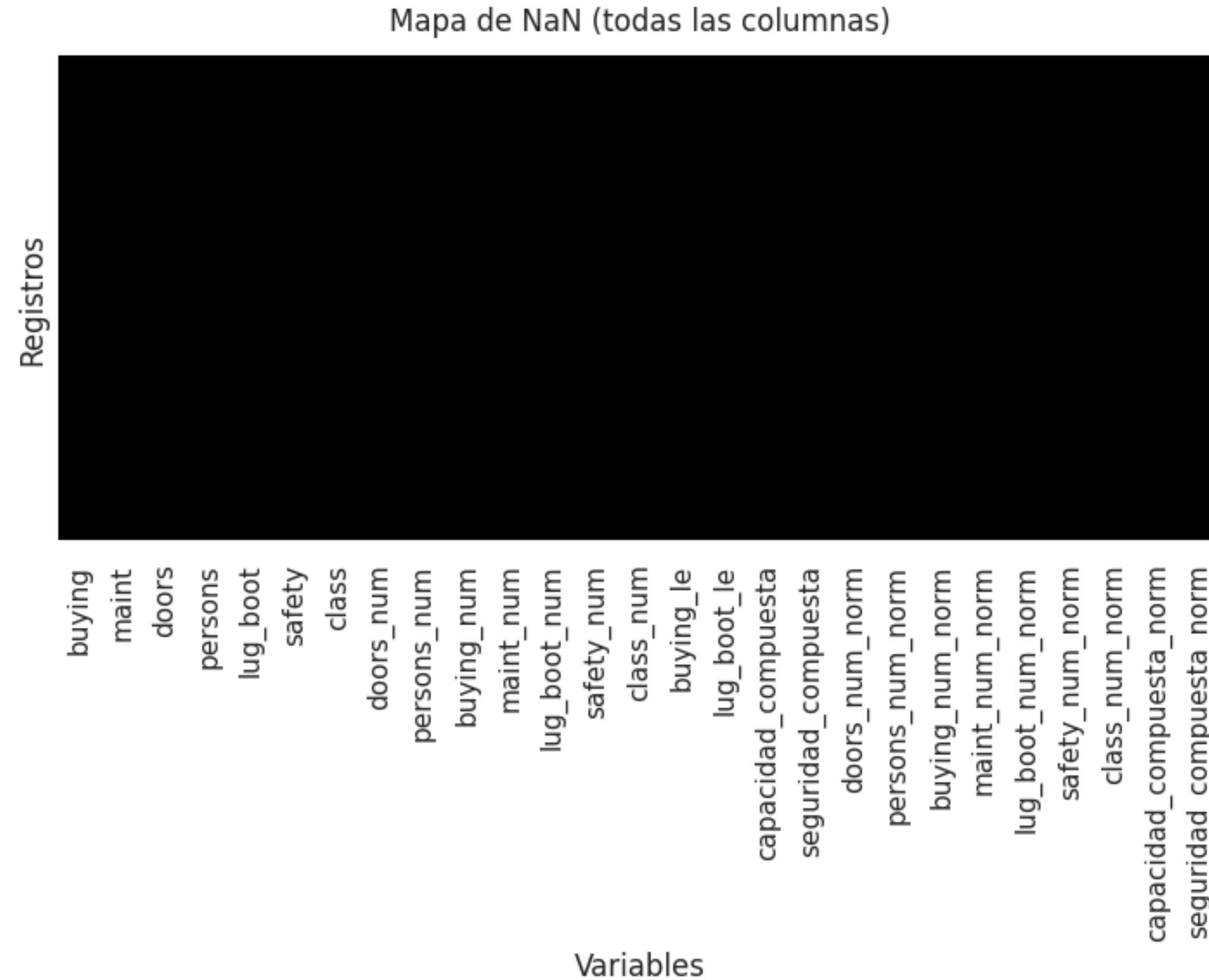


OBJETIVO

Identificar patrones y tendencias en la aceptabilidad de los automóviles.

¿Es posible predecir la calidad de un auto usando sus características técnicas y de seguridad?

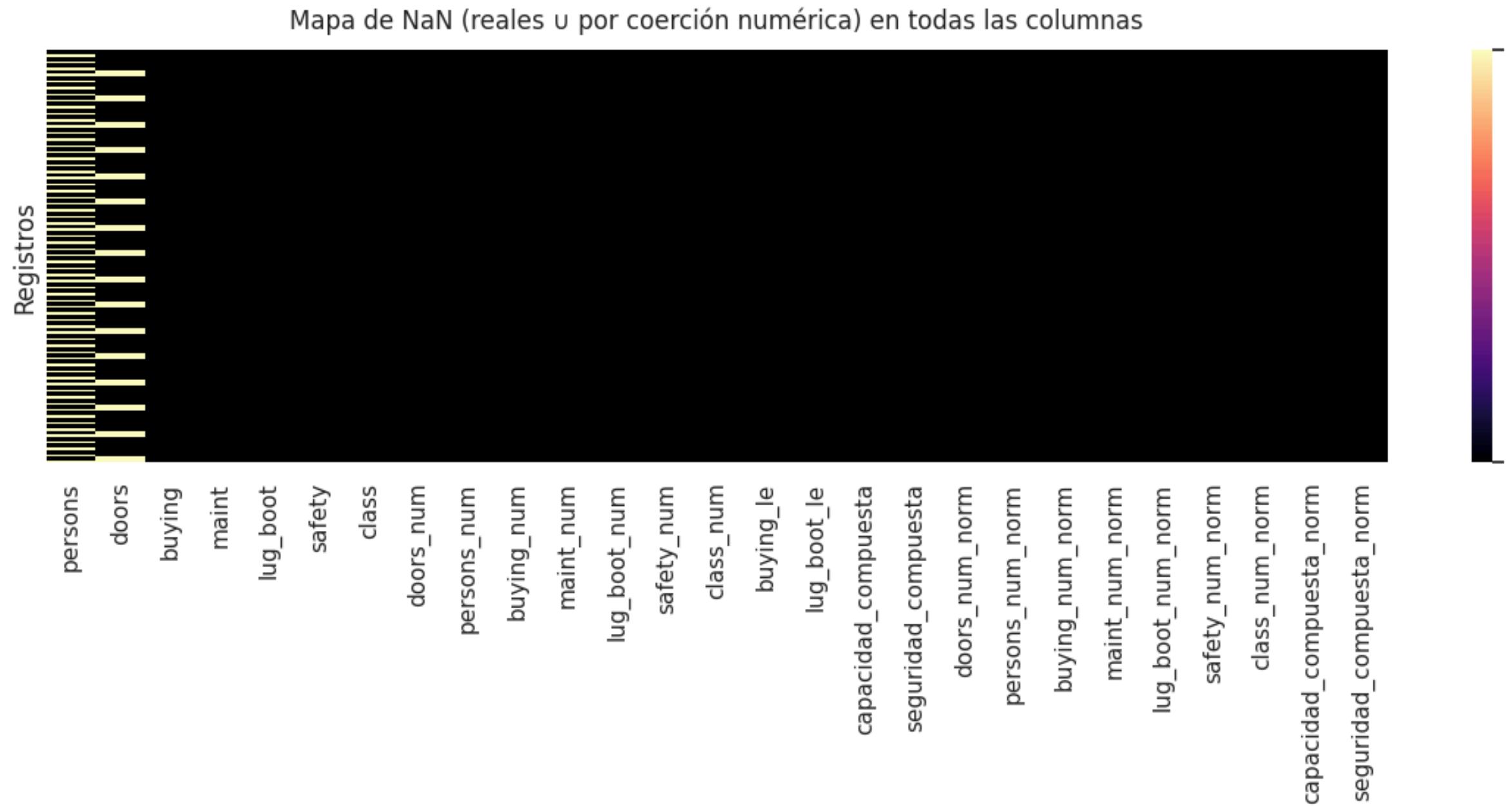
MAPA DE CALOR NAN



Se evidencia que no hay datos faltantes en el database.

CORRELACIÓN NUMÉRICA

Variables

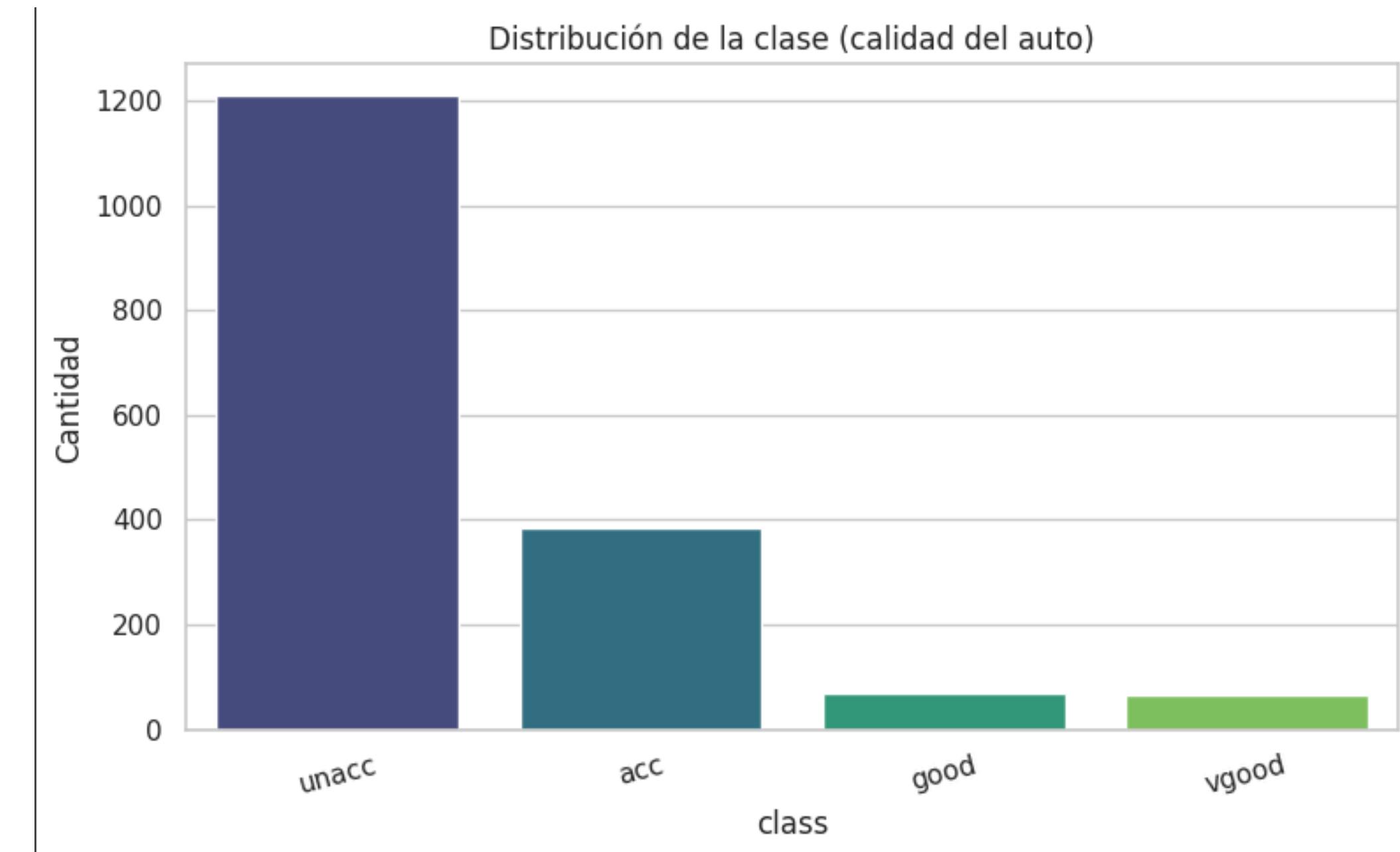


Datos faltantes NAN:
persons = 33.33%
doors = 25.0%

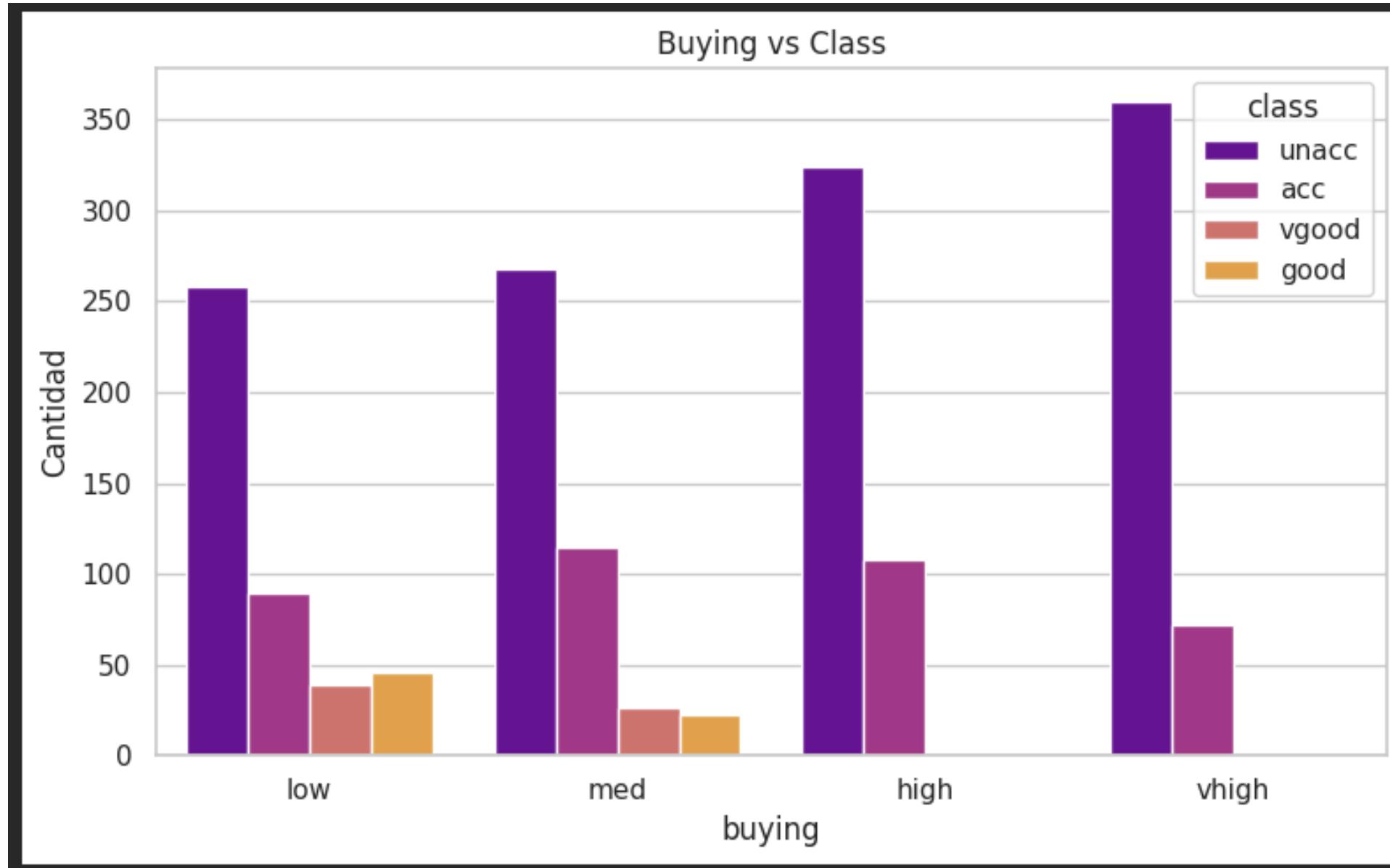
DISTRIBUCIÓN

Mayoria inaceptable.
El resto tiene menor
representación.

Desbalance significativo
en los datos.



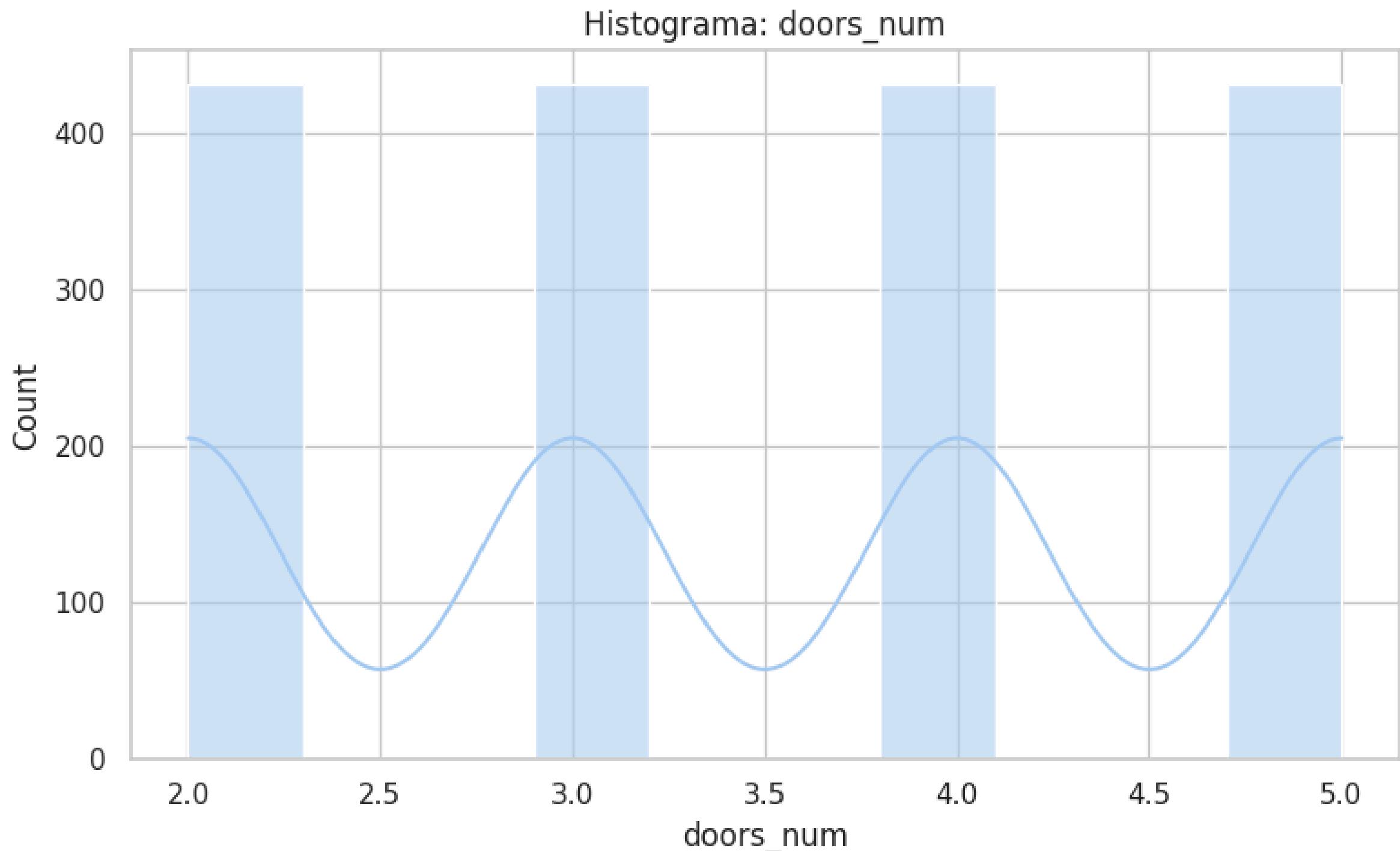
RELACIÓN



La mayoría de los autos,
sin importar el precio de
compra, son clasificados
como calidad inaceptable

DOORS

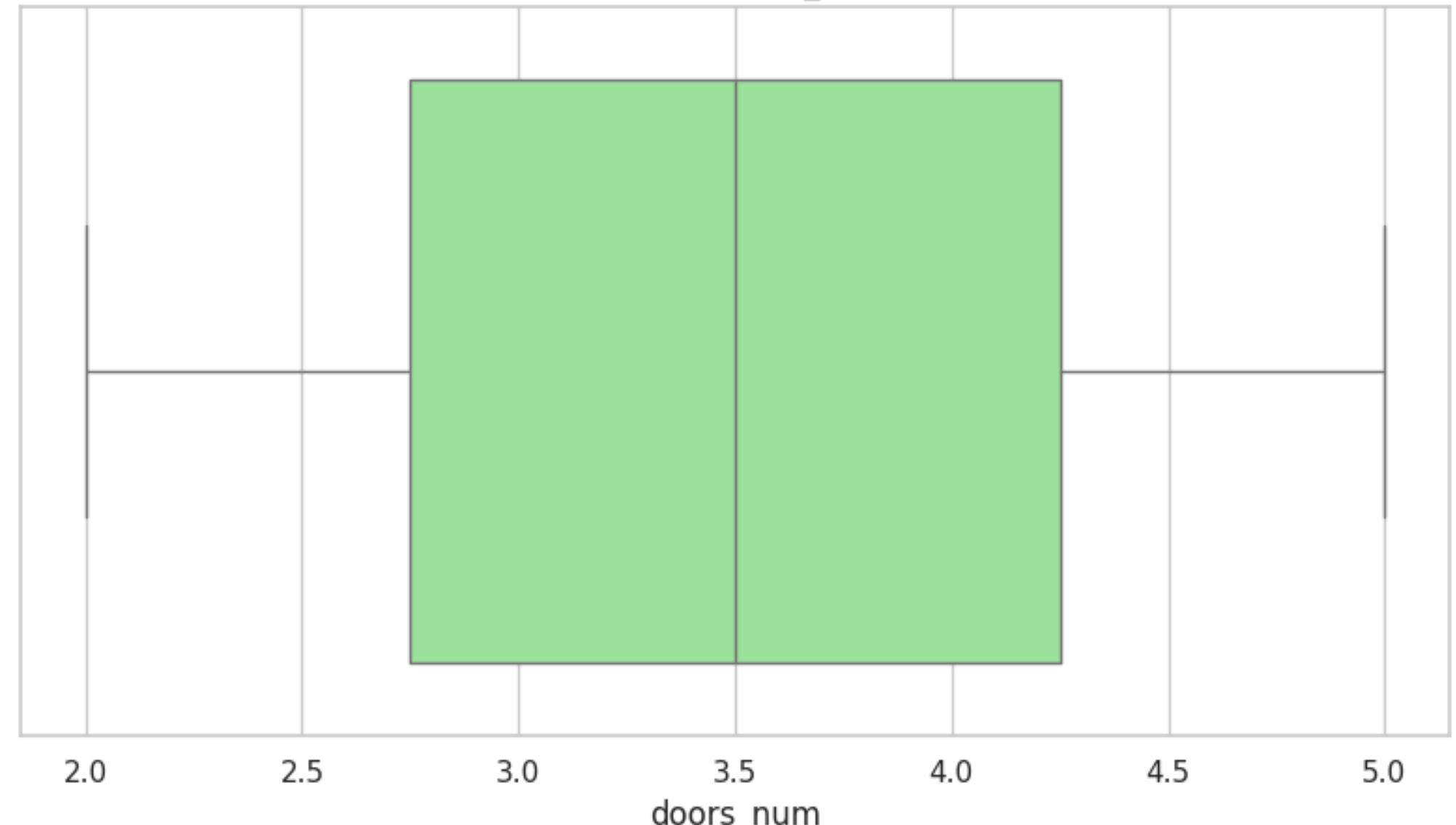
La mayoria de los autos
tienen 2, 3, 4 o 5
puertas, con frecuencias
similares en cada grupo



DOORS

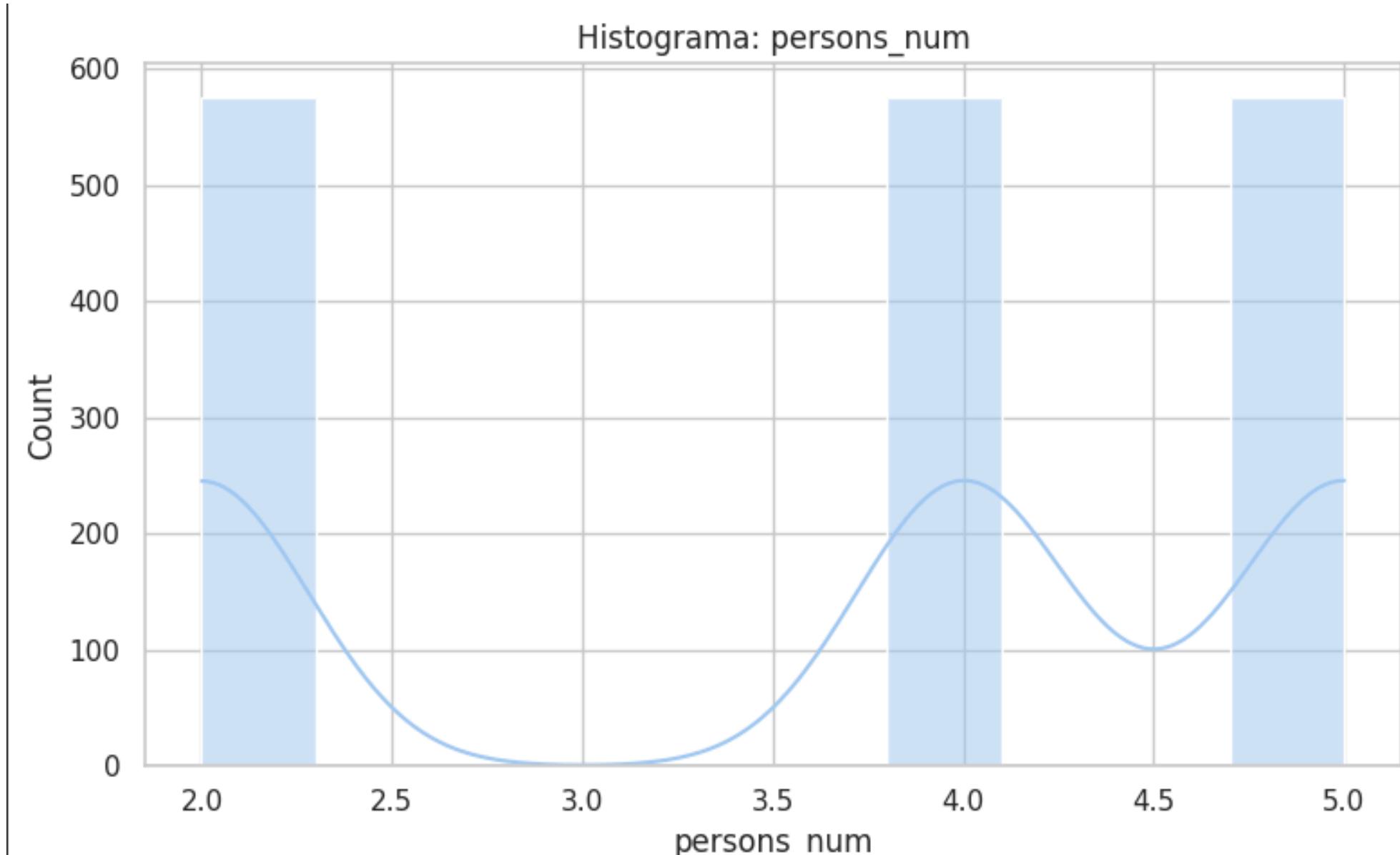
La cantidad de puertas en los autos varia entre 2 y 5, y no se presentan valores extremos ni atípicos.

Boxplot: doors_num



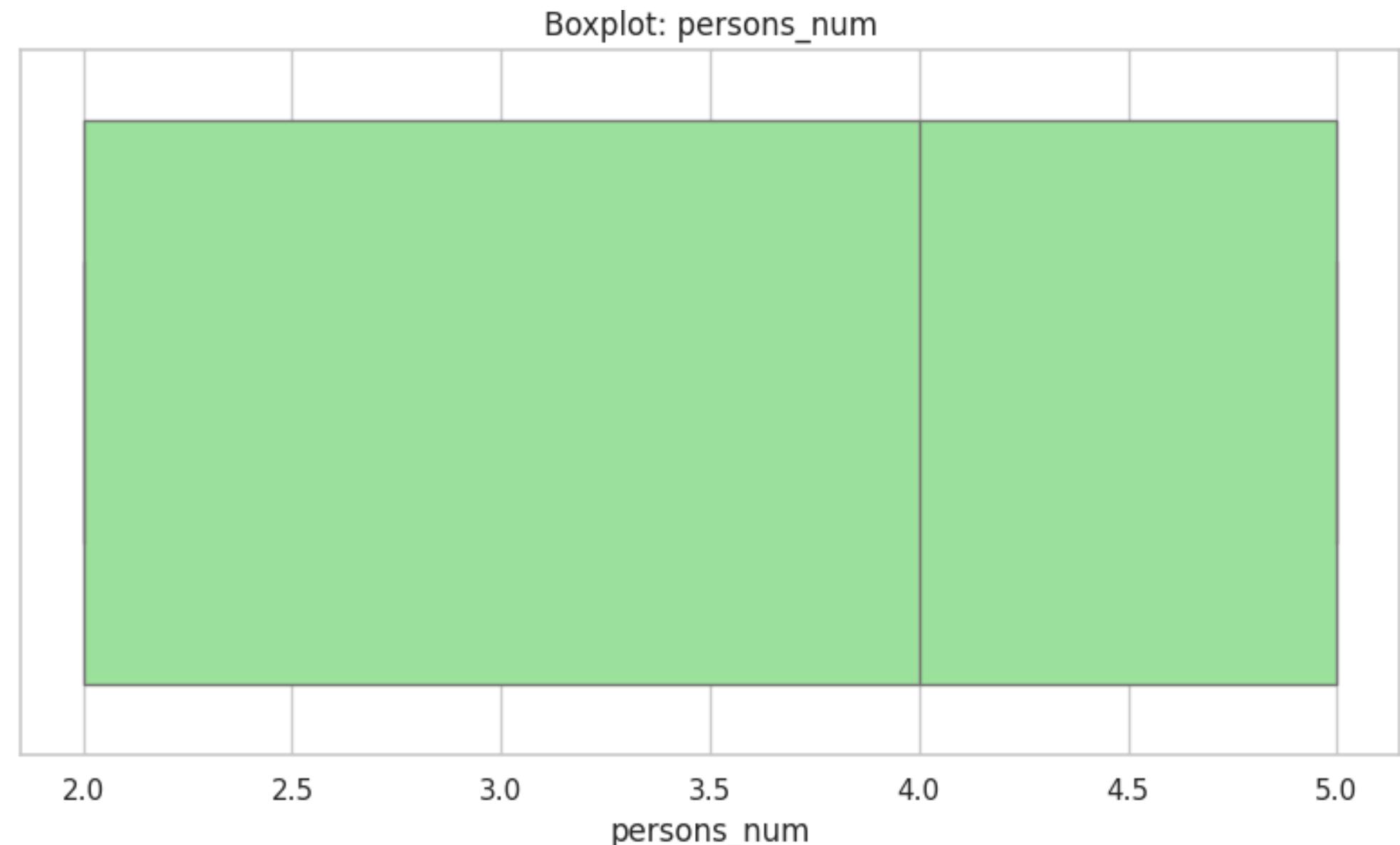
PERSONS

La mayoría de los autos
pueden transportar 2, 4
o 5 personas, siendo muy
pocos los de 3 plazas.



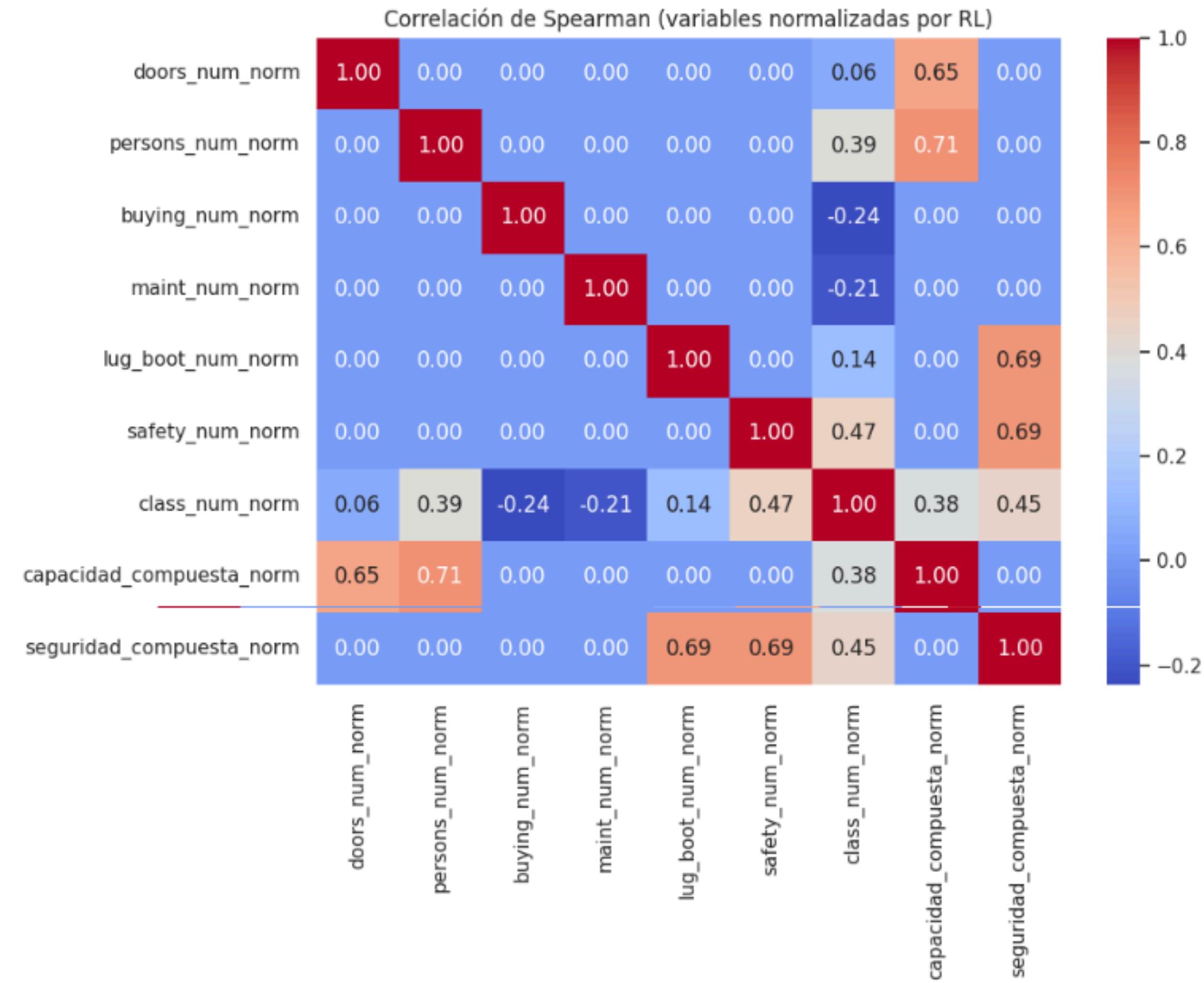
PERSONS

El numero de plazas varia entre 2, 4 y 5 personas por auto, y no se observan valores atípicos en la distribucion.



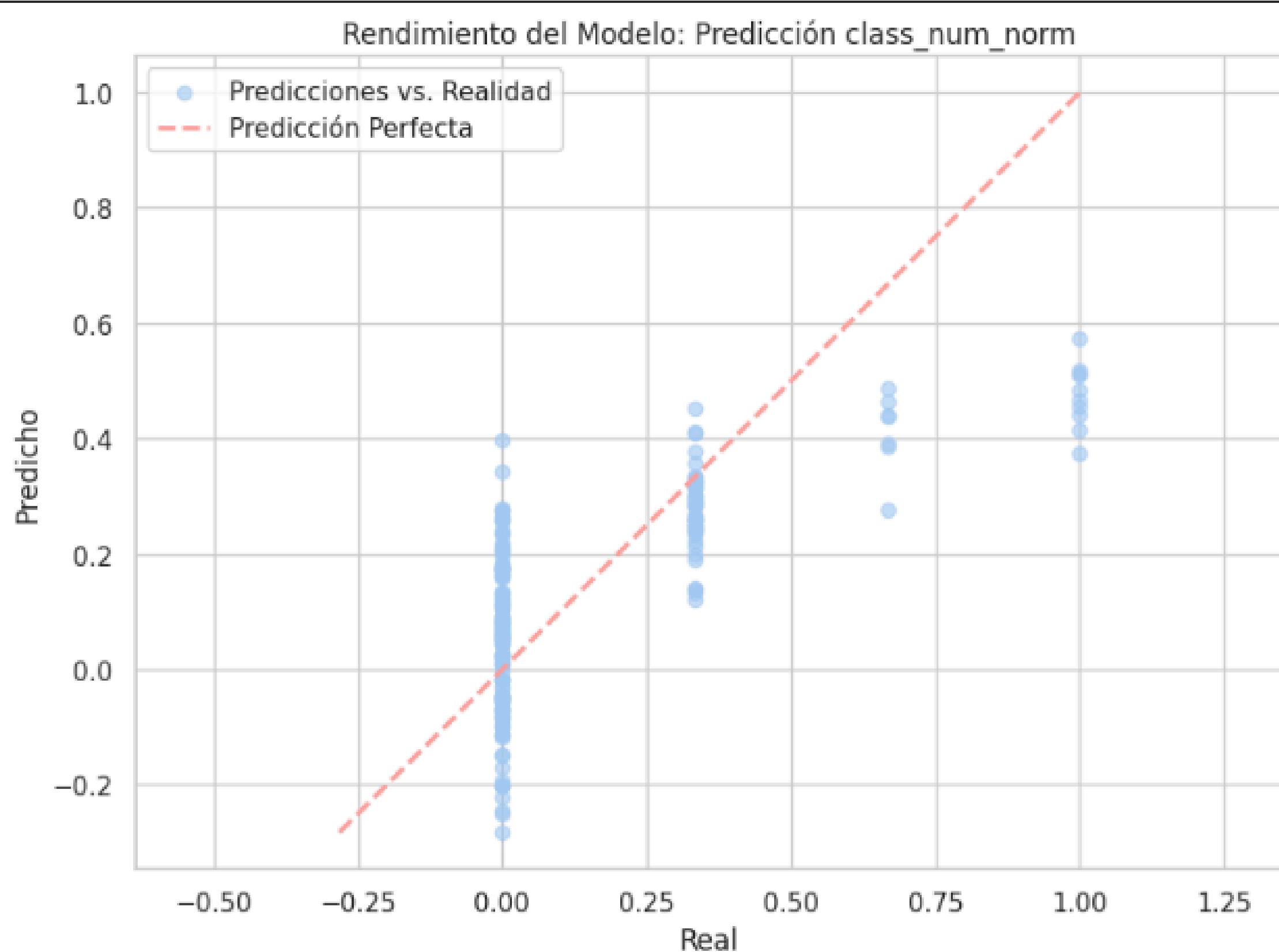
CORRELACIÓN

Las variables compuestas tienen fuerte relación entre si y con algunas variables principales.



PREDICCIÓN LINEAL

La grafica muestra que las predicciones del modelo se acercan a la realidad, pero existen algunas diferencias.





CONCLUSIONES

- La mayoría de los autos del dataset son de calidad “Inaceptable” y predominan los modelos con capacidad para 2, 4 o 5 personas.
- Algunas variables compuestas tienen fuertes correlaciones entre si y con variables principales, lo que puede simplificar los modelos.
- El modelo de regresion logra aproximar la calidad, pero no es perfecto, ya que existen diferencias notables entre lo predicho y lo real.