

Módulo 3 – Análisis Exploratorio y Programación Estadística

Actividad: Millas por Galón (Múltiple)

Ciencia de Datos

Actividad: Millas por Galón (Modelo Regresivo Multivariado)



Problema

Se está haciendo una investigación en la industria automotriz para mejorar el rendimiento de los motores a combustible. Para esto, se recolectó un set de datos que contiene el rendimiento (millas por galón) de distintos modelos de automóviles, así como también de los factores que se piensa influyen.



Problema

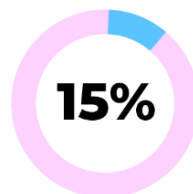
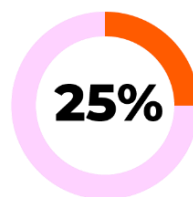
Rendimiento
(millas por galón)

	mpg	car name	cylinders	displacement	horsepower	weight	acceleration	model year	origin
1	18.0	chevrolet ch...	8	307.0	130	3504	12.0	70	1
2	15.0	buick skylark...	8	350.0	165	3693	11.5	70	1
3	18.0	plymouth sat...	8	318.0	150	3436	11.0	70	1
4	16.0	amc rebel sst	8	304.0	150	3433	12.0	70	1
5	17.0	ford torino	8	302.0	140	3449	10.5	70	1
6	15.0	ford galaxie ...	8	429.0	198	4341	10.0	70	1
7	14.0	chevrolet im...	8	454.0	220	4354	9.0	70	1
8	14.0	plymouth fur...	8	440.0	215	4312	8.5	70	1
9	14.0	pontiac catal...	8	455.0	225	4425	10.0	70	1
10	15.0	amc ambass...	8	390.0	190	3850	8.5	70	1
11	15.0	dodge challe...	8	383.0	170	3563	10.0	70	1
12	14.0	plymouth 'cu...	8	340.0	160	3609	8.0	70	1
13	15.0	chevrolet mo...	8	400.0	150	3761	9.5	70	1

Problema

Elabore un modelo multivariado utilizando todas las variables disponibles para explicar el consumo en millas por galón.

1. Cree un Notebook **ordenado, documentado y reproducible** con su análisis.
2. Haga un breve **análisis exploratorio** de los datos, ¿es un problema lineal?.
3. Elabore un **modelo regresivo multivariado**.
4. Realice el procedimiento de **selección de variables del modelo**.
5. Calcular las **métricas de error** para cada modelo.
6. Calcular el **coeficiente de determinación**.
7. Hacer **análisis de residuales**.
8. ¿Cuál es el factor que mayor incidencia tiene en el rendimiento?, ¿Por qué?
9. ¿Es un mejor modelo que el simple? ¿Cómo se evidencia aquello?



¡Éxito!