

---

## Segundo Examen Parcial

---

### I. Predecir los precios de los automóviles usados

El archivo ToyotaCorolla.csv contiene datos sobre automóviles usados (Toyota Corolla) a la venta a finales del verano de 2004 en los Países Bajos. Tiene 1436 registros que contienen detalles sobre 38 atributos, incluidos Precio (Price), Edad (Age), Kilómetros (Kilometers), HP y otras especificaciones. El objetivo es predecir el precio de un Toyota Corolla usado según sus especificaciones.

Divide los datos en conjuntos de datos de entrenamiento (50%), validación (30%) y prueba (20%).

- a. ¿Cuáles parecen ser las tres o cuatro especificaciones de automóvil más importantes para predecir el precio del automóvil?
- b. Utilizando métricas que consideres útiles, evalúa el rendimiento del modelo en la predicción de precios.

### II. Ventas de automóviles Toyota Corolla

El archivo ToyotaCorolla.csv contiene datos sobre coches usados (Toyota Corolla) a la venta a finales del verano de 2004 en los Países Bajos. Tiene 1436 registros que contienen detalles sobre 38 atributos, incluidos Precio, Edad, Kilómetros, HP y otras especificaciones. El objetivo será predecir el precio de un Toyota Corolla usado en función de sus especificaciones.

- a. Identifica las variables categóricas.
- b. Explica la relación entre una variable categórica y la serie de variables ficticias binarias derivadas de ella.
- c. ¿Cuántas variables binarias ficticias se requieren para capturar la información en una variable categórica con  $N$  categorías?
- d. Utiliza R para convertir las variables categóricas de este conjunto de datos en variables ficticias y explica en palabras, para un registro, los valores de las variables ficticias binarias derivadas.
- e. Utiliza R para producir una matriz de correlación y una gráfica de matriz. Comenta sobre las relaciones entre variables.