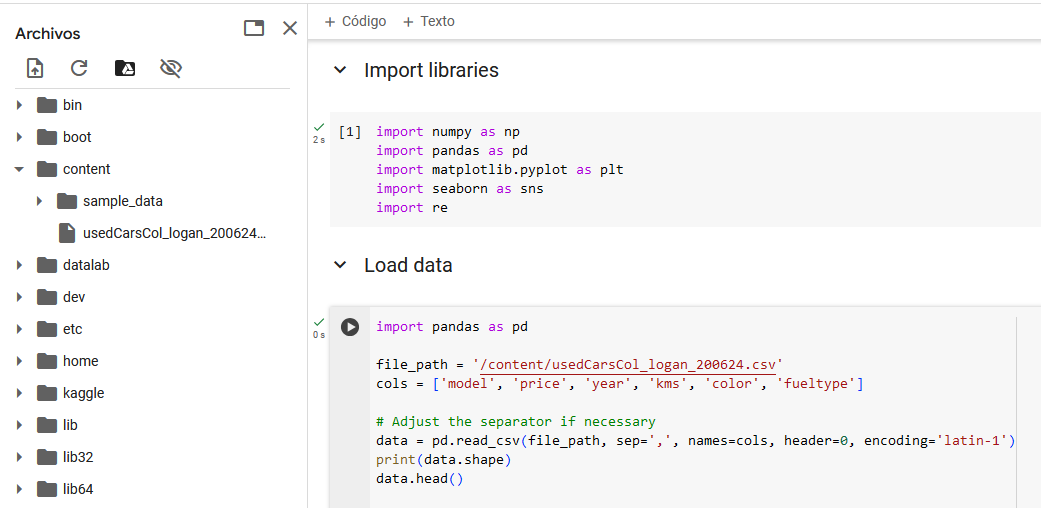
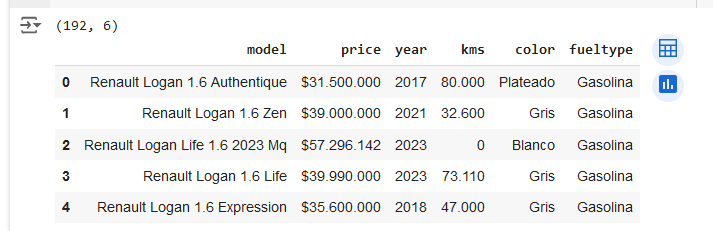
Roger Alberto Arias Triana - Taller 3

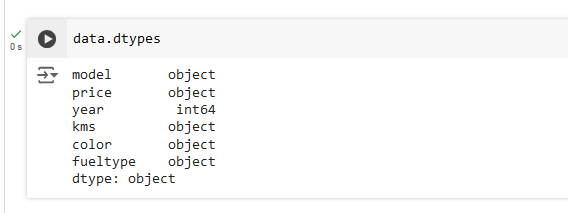
Para este ejercicio se trabajó con el archivo de vehiculos logan generados en el taller de web scraping, se sube el dataset a la carpeta sample data y se cambia el nombre en la ruta de la variable file\_path:



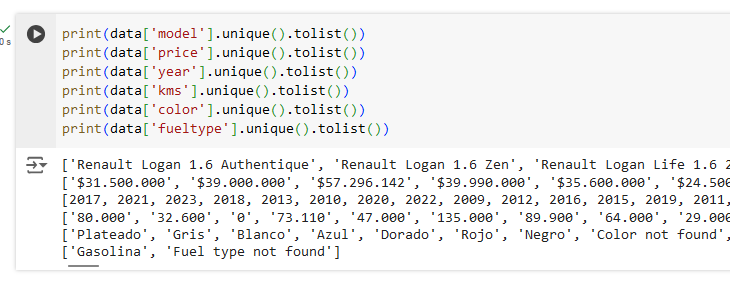
Visualizan los datos de los primeros registros, así:



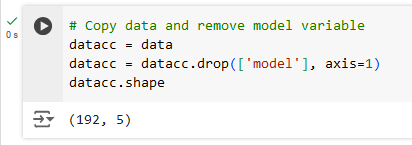
Se obtienen los tipos de datos



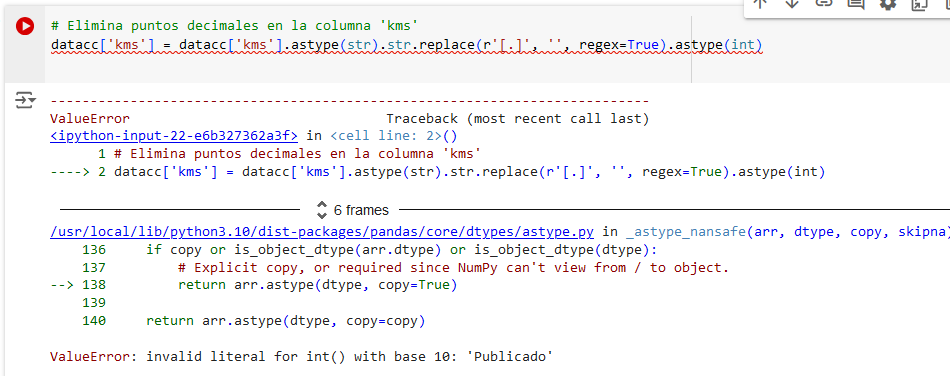
Se imprimen los datos



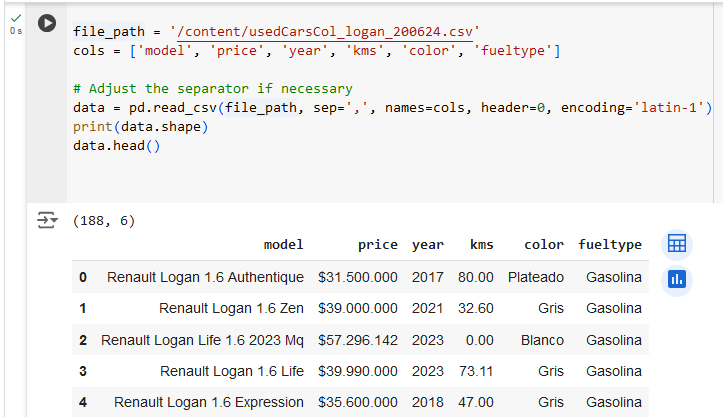
Se elimina el campo ‘model’ y se muestran las dimensiones del dataset:



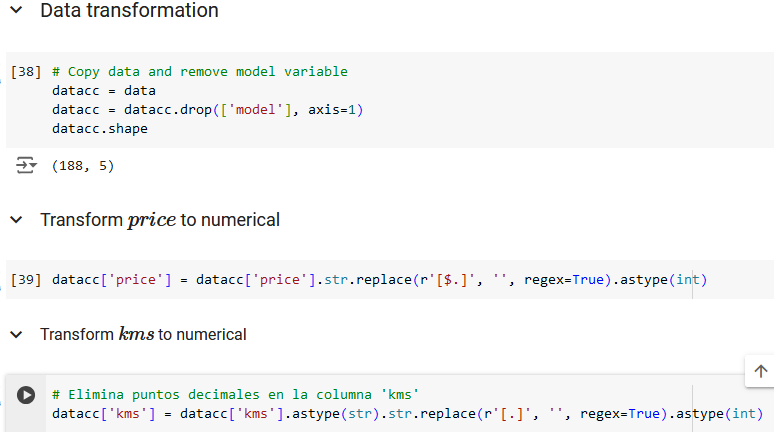
Se observa error en el campo kms por contar con un valor en string



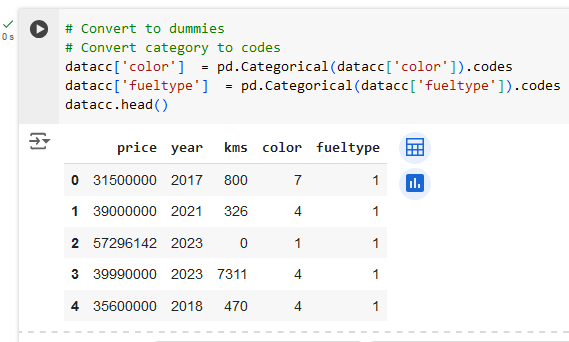
Se eliminan registros con datos incompletos y se vuelve a subir el archivo



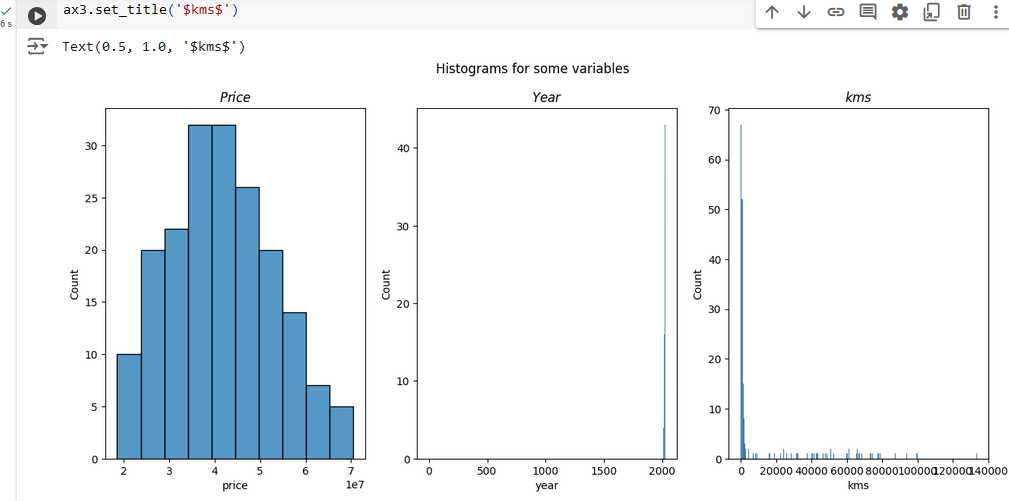
Se elimina el campo ‘modelo’ y se convierten a numéricos los campos ‘precio’ y ‘kms’



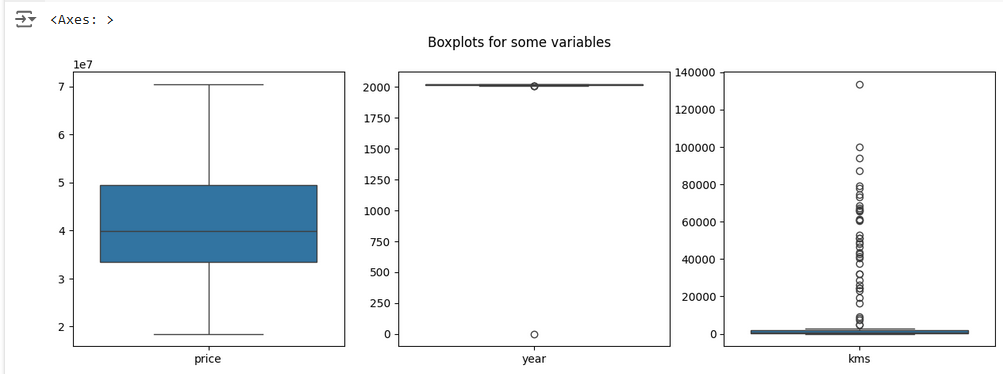
Se categorizan(asignan valores numéricos) al campo ‘fueltype’



Se grafican los variables ‘price’, ‘year’ y ‘kms’



Se generan gráficos boxplots en donde se identifica que los datos contienen demasiados datos atipicos por lo cual se debe generar nuevamente el dataset



Se deben eliminar demasiados registros del archivo generado en el ejercicio de webscraping para la marca Renault y el modelo Logan