**Documento Descriptivo.**

1. Se construye flujo CuracionPreparacionDataSet, donde se establecen 5 contenedores:

* Carga DataSet
* EDA\_1
* PreparacionDataSet
* EDA\_2
* EDA\_3

1. A continuación, se describe la función de cada contenedor y su contenido:

* Carga DataSet: Se carga información de ImportacionesAutomoviles1985.csv, utilizando el objeto “Datos de Entrada”.
* EDA\_1: Se realiza análisis exploratorio, donde no se obtiene información relevante debido a que todos los campos son string. Por medio del objeto “examinar”.
* PreparacionDataSet:
* Seleccionamos las columnas a tener en cuenta para el análisis y renombramos las columnas con “Seleccionar”.
* Luego se asignan los tipos de dato con “campo automatico”.
* Se verifican cambios con “seleccionar”.
* Se retiran espacios “limpieza de datos profesional”.
* Se reemplaza los valores de numero\_puertas (four y two), por 4 y 2 respectivamente “formula”.
* Se cambia, numero\_puertas de string a Entero “Seleccionar”,
* Se imputan datos nulos por el valor promedio “imputar”, los campos imputados fueron (numero\_puertas (2 nulos), caballos\_fuerza (2 nulos), rpm\_maximas (2 nulos) y precio (4nulos)).
* EDA\_2: Obtenemos los datos estadísticos de los campos tipo “numero” del DataSet cargado, por medio del objeto “Resumen de Campos” y “Examinar” respectivamente.
* EDA\_3: Se genera información sobre los valores de las columnas del DataSet, por medio del objeto “Examinar”.

1. Observaciones, sobre la ejecución del flujo:

* Los campos “símbolos”, “perdidas normalizadas”, “diámetro” y “carrera”, no dan información relevante sobre los vehículos importados, por lo que son retirados del análisis exploratorio.
* Los datos son 100% aceptables para cada columna.
* Las siguientes columnas presentan nulos, numero\_puertas (2 nulos), caballos\_fuerza (2 nulos), rpm\_maximas (2 nulos) y precio (4 nulos).
* Los datos no presentan caracteres especiales.
* Los datos tipo numero tienen formatos validos para su transformación de String (como se cargaron desde el archivo csv), a float, int, y demás en el proceso de Preparación. Donde no se generó error de conversión.
* No se encontraron datos outliers.
* Las marcas con mayor cantidad de importaciones son Toyota, Nissan y Mazda.
* El tipo de combustible mas utilizado es el gas (gasolina).
* Hay muy pocas importaciones de vehículos con turbo.
* La mayor cantidad de vehículos importados son de 4 puertas.
* La carrocería mas importada es el Sedan seguido del hatchback.
* Los vehículos importados en su mayoría son de tracción delantera, seguidos de los vehículos de tracción trasera y muy pocas unidades con tracción en las 4 ruedas.
* La distribución de distancia entre ejes, no se presentan muchas diferencias. Al igual, que en medidas de alto, largo y ancho
* Respecto al peso en vacío (sin aceite, liquido refrigerante, gasolina, entre otros), los vehículos más pesados son de 4 toneladas aproximadamente y los más livianos de 1.4 toneladas.
* Los vehículos con mayor exportación ubican su árbol de levas en la cabecera de motor y implementan uno solo por motor, junto con configuración de 4 cilindros e inyección electrónica multipunto. Los cuales manejan muy bajos consumos de combustible.
* En cambio, los vehículos con especificaciones mas altas, en arboles de levas, sistemas de alimentación, cantidad de pistones y relaciones alta (como es el caso de vehículos deportivos, roadster, entre otros), presentan números muy bajos de importaciones.
* Los vehículos importados más económicos están entre 5.000 y 9.000 dólares. En cambio, los mas costosos superan los 20.000 dólares.