**INFORME EXPLICATIVO**

Transformación y limpieza de los datos:

* Descargar los archivos .csv e importarlos en un Jupiter notebook para su transformación y limpieza.
* Realizar un primer contacto visual de los dataframes.
* Recoger información de los dataframes, tipos de datos de las columnas, valores nulos, duplicados, filas vacías.
* En el dataframe de matriculaciones, Aparecen 1787 duplicados, he decidido analizarlos de forma visual generando un dataframe 'duplicados\_completos' que nos muestra todas las filas repetidas incluyendo su primera aparición, he llegado a la conclusión que no son duplicados como tal, sino que al no haber un identificador único hay probabilidades de que hayan filas duplicadas, sobre todo cuando el valor de [Model, Marca] es ND, que son prácticamente todas menos unas pocas al final, las cuales las columnas [Model, Marca] sí tienen valor y es más dificil que coincidan, de hecho el índice que se genera en este dataframe no es correlativo en estas filas, por lo que yo creo que ha medida que hay matriculaciones se registran los vehículos, en cambio en las de valor ND al ser más probable la coincidencia si que el índice generalmente es correlativo. Es mi apreciación, que seguro tampoco he de estar en lo cierto. Tomo la decisión de no eliminar los duplicados.
* Eliminar columnas, modificar nombres de columnas y categorías de éstas.
* Realizar en algunas columnas una agrupación de varias categorías en una sola.
* Visualizar outliers, en la columna ‘Unidades’ de dataframe matriculaciones hay paquetes de matriculaciones de muchas unidades que pueden parecer outliers, pero después de una exploración visual se aprecia que son motos y pertenecen al 'Cliente' Rac. No he encontrado información exacta, pero parece ser un código administrativo usado en el registro de vehículos. Posiblemente estas matriculaciones sean de empresas de alquiler, ayuntamientos...
* En el dataframe cargadores y su columna ’Tipo Carga’ hay 154 cargadores con una potencia de 0 kW y algún cargador de 400 kW.
* Modificar la columna ‘Tipo Carga’ clasificando las potencias de carga en categorias, ‘Caga Lenta’, ‘Semirrápida’…
* Creación de nuevas columnas para el posterior dashboard.
* Visualizar datos estadísticos.
* Unir dataframes mediante la columna común ‘Provincia’, ha sido imposible al aparecer el siguiente error:

**MemoryError: Unable to allocate 2.06 GiB for an array with shape (276175299, 1) and data type object *Output is truncated. View as a*** [***scrollable element***](command:cellOutput.enableScrolling?0089c693-b998-4930-805e-05d17c9e41a5) ***or open in a*** [***text editor***](command:workbench.action.openLargeOutput?0089c693-b998-4930-805e-05d17c9e41a5)***. Adjust cell output*** [***settings***](command:workbench.action.openSettings?%5B%22%40tag%3AnotebookOutputLayout%22%5D)***...***

* Sinceramente aunque aparezca el error por la enorme cantidad de filas generada, no veo sentido la unión de ambos dataframe.
* Exportar los dataframe a .csv.
* Importar los archivos en dos hojas diferentes del mismo libro excel, Dashboard\_matriculaciones\_cargadores.
* En matriculaciones se crea la columna ‘Fecha’, de la unión de las otras columnas ‘Mes’ y ‘Año’.
* Después de una exploración visual me doy cuenta que en la columna ‘Marca’ (matriculaciones), existen dos valores para la misma marca, LYNK&CO y Lynk & Co, los reemplazo y solo dejo el valor Lynk & Co.

Análisis descriptivo:

* Realización de KPI´s:
* Año / Mes con más matriculaciones.
* Provincia con más matriculaciones
* Tipo de cargador más instalado.
* Provincia con más cargadores.
* Cantidad de matriculaciones por ‘Tipo de Vehículo’. Al sorprenderme que la categoría con más matriculaciones es ‘Todoterreno’, examino sus ‘Marcas’ y ‘Modelos’ y me doy cuenta que los vehículos de tipo SUV los considera así. A mi manera de verlo los debería considerar como ‘Vehículos’. Para la categoría ‘Derivados’ incluye sobre todo camiones y furgonetas.
* Cantidad de matriculaciones por ‘Tecnología’.
* Top 3 de marcas más matriculadas por ‘Tecnología’, incluye segmentador para la selección.
* Gráficos:
* Evolución Temporal. Evolución del número de matriculaciones a lo largo de los años del estudio.Mediante macros se puede filtrar por provincias o la comunidad autónoma al completo, por tipo de ‘Tecnología’ o todos los vehículos.
* Matriculaciones Tipos de Vehículos. Cantidad de matriculaciones de las distintas categorías o por su conjunto según la elección del segmentador. Filtrado temporal mediante segmentadores de tiempo.
* Vehículos enchufables. En el primer gráfico se muestra la cantidad de vehículos Eléctricos o de Combustión/ Enchufables por provincia. En el segundo se muestra la cantidad por Tipo de Cargadores que hay por provincia. En la parte inferior se muestra la proporción de cargadores, la proporción de vehículos enchufables y la cantidad de vehículos que hay por cargador, según la macro de la provincia que se seleccione.

Informe explicativo del análisis:

Se puede apreciar el aumento de matriculaciones en el paso del tiempo, también es verdad que las normativas nos llevan a ello, pero si observamos las de solo vehículos eléctricos también van en un aumento considerable. Como anécdota si filtramos el primer semestre del año 2020 se aprecia la bajada de matriculaciones por el Covid. Barcelona, evidentemente, es la provincia más electrificada, pero no por ello la mejor preparada infrastructuramente en la carga de vehículos, ya que tiene un número de vehículos por cargador bastante superior al resto.