

ANÁLISIS DEL PROFILING

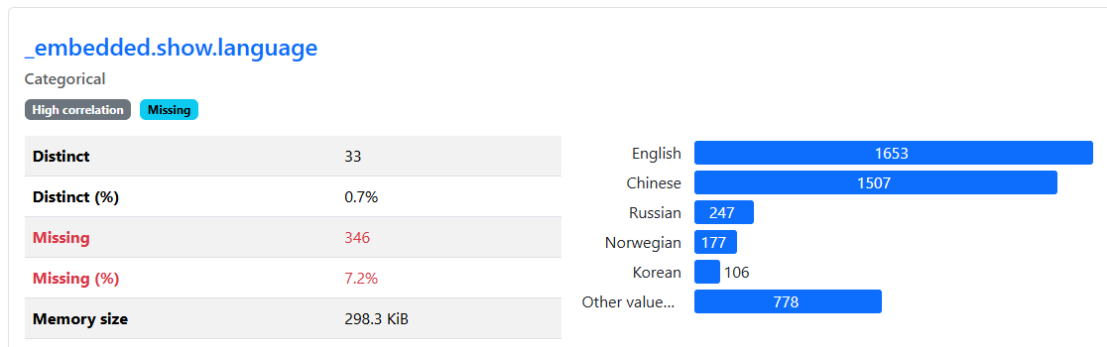
Se realiza el profiling con el fin de analizar y examinar la estructura, calidad, distribución y tendencias del conjunto de datos, y por lo tanto, obtener una visión más amplia de manera previa a la limpieza o transformación de estos.

Una vez normalizado y consolidado el archivo JSON obtenido desde la API, se observa que este cuenta con **4.814** registros y **66** columnas correspondientes a las series emitidas durante el mes de enero del año 2024. Como campo de diferenciación para las fechas se tiene **airdate**.

Para todos los campos es posible tener información sobre la cantidad y porcentaje de valores distintos y vacíos, así como medidas estadísticas cuando se trata de variables no categóricas, tales como el promedio, el mínimo, el máximo, entre otros.

Para el ejercicio en cuestión, se tienen en cuenta las siguientes consideraciones en la limpieza de los datos:

- ✓ No realizar limpieza de caracteres especiales tales como tildes dado que la base está compuesta por cadenas de texto descritas en idiomas distintos al español tal como lo muestra el profiling para el campo **_embedded.show.language**:



- ✓ No realizar transformaciones de tipo de datos sobre los campos que representan fechas dado que se observa que para algunos casos estas traen únicamente el formato de año, mes y día, mientras que otras tienen también la hora. Lo anterior con el fin de no afectar la integridad de los datos.
- ✓ Realizar limpieza en los nombres de las columnas realizando cambio del carácter punto (.) por guión bajo (_).

- ✓ Para los campos que contienen listas luego de la normalización del JSON, convertir este tipo de dato a string con el fin de no afectar el almacenamiento del dataframe como base de datos sqlite. Asimismo, se reemplaza en todo el dataframe el caracter apóstrofe ('), dado que este puede actuar como delimitador de texto en el lenguaje SQL y afectar la integridad de la información al crear e insertar los datos en la tabla de la base de datos.
- ✓ Conservar la estructura para los registros con valores nulos dado que la información de missings puede aportar apreciaciones significativas sobre los análisis estadísticos que se realicen sobre la data.