

Data Visualization & Analytics Challenge – Carlos José Viola – carlosjviola@gmail.com

Para realizar la tarea se ha construido un prototipo de software que consta de dos componentes:

1. Desarrollo de un cuaderno de Google Colab con un script de Python para analizar el conjunto de datos y generar un nuevo archivo csv con las transformaciones necesarias.
2. Desarrollo de un reporte de Power BI para visualizar claramente la información.

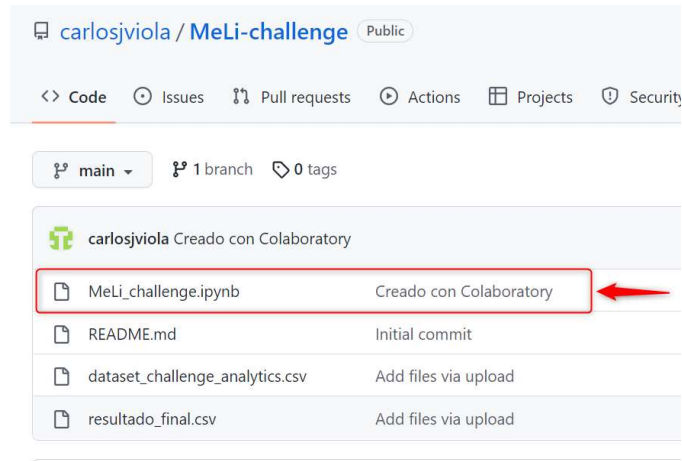
Consideraciones acerca del dataset “dataset_challenge_analytics.csv”

1. El archivo tiene registros correspondientes al mes de enero de 2022 para el campo *first_login_dt*
2. No hay especificaciones para el campo *week_year*. Se considera corresponde a la semana del año del campo *first_login_dt*
3. Haciendo una visualización rápida del archivo se detectan posibles inconsistencias o errores por lo cual se comienza un análisis más detallado.
4. Se va a suponer que *habito* significa que un usuario utiliza la aplicación un número determinado de veces en un período determinado.

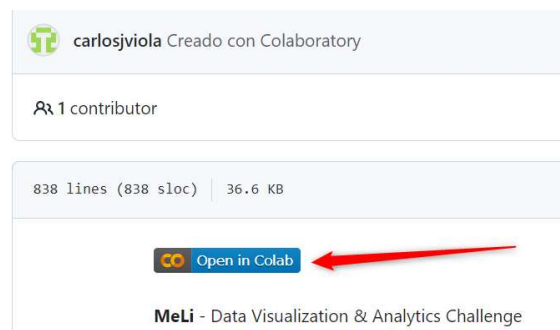
Se comparte el repositorio de GitHub que contiene los archivos del proyecto:

<https://github.com/carlosjviola/MeLi-challenge>

El archivo MeLi_challenge.ipynb contiene el script para realizar el análisis



Accedemos al archivo y tocamos el botón “Open in Colab” para realizar el análisis. Al finalizar el script genera un nuevo archivo “resultado_final.csv” con los cambios y transformaciones aplicadas.



Transformaciones y ajustes aplicados al dataset:

1. Se elimina la primera columna que contiene un *id* correlativo que se considera no es de utilidad.
2. Se eliminan los caracteres “-1” presentes al final de algunos registros en el campo *user_id*. Estos registros están duplicados.
3. Además, se localizan y eliminan otros *user_id* que también están duplicados.
4. La columna *habito* se transforma a número entero, de acuerdo a las especificaciones.
5. Las columnas de fecha que terminan en “_dt” no tienen información sobre horas. Estas columnas se convierten al formato de solo fecha o date.
6. La columna *week_year* tiene el valor 52 para los días 1 y 2 de enero. Se entiende que por error mantiene la semana del año anterior. Se decide reemplazar por 1 el campo en estos registros.
7. Se analizan los campos de fecha “_dt”. Se entiende que ninguna fecha puede ser anterior a la fecha en *first_login_dt*. Se localizan y eliminan los registros con este problema.
8. Se crean nuevas columnas “_dias” con el cálculo de la cantidad de días transcurridos entre la fecha de primer ingreso *first_login_dt* y las demás fechas para facilitar el desarrollo de los reportes.
9. Se genera un nuevo archivo *resultado_final.csv* con los cambios aplicados.

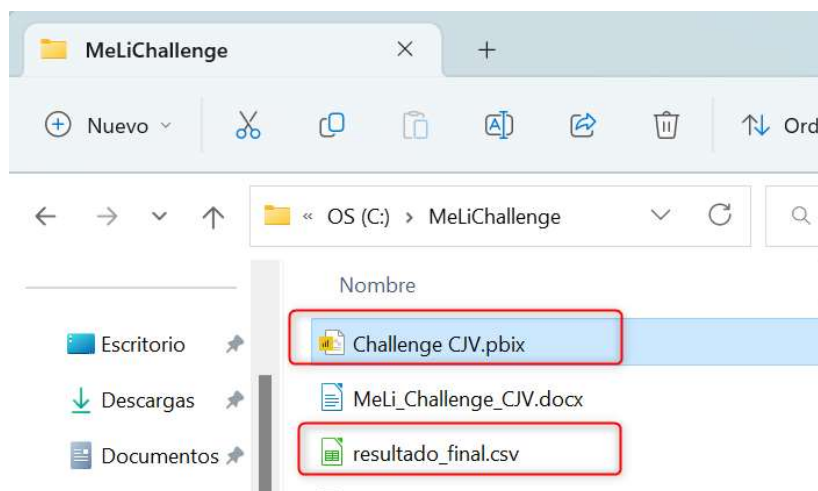
En el cuaderno de Colab se incluyen otros comentarios y aclaraciones del proceso.

Desarrollo del reporte

Se ha utilizado la herramienta Power BI. Necesitará tener instalada en su PC la aplicación Power BI Desktop.

Instalación:

Para poder visualizar el reporte deberá copiar los archivos “Challenge CJV.pbix” y “resultado_final.csv” en la carpeta “MeLiChallenge” del disco C: de su computador. Una vez realizada esta operación puede abrir el reporte con la aplicación Power BI desktop.



El objetivo del reporte es el siguiente:

Visualizar la cantidad de usuarios nuevos, que utilizan la aplicación por primera vez en un período determinado.



A partir de esta información conocer:

1. Cantidad de usuarios que realizan la activación, porcentaje que representan sobre el total de usuarios nuevos y el promedio en días que transcurren hasta este evento.



2. Cantidad de usuarios que realizan la configuración, porcentaje que representan sobre el total de usuarios nuevos y el promedio en días que transcurren hasta este evento.



3. Cantidad de usuarios que regresan después de realizar el primer ingreso, porcentaje que representan sobre el total de usuarios nuevos y el promedio en días que transcurren hasta este evento.



4. Cantidad de usuarios que se considera han adoptado la aplicación, porcentaje que representan sobre el total de usuarios nuevos y el promedio en días que transcurren hasta este evento.



También se podrá visualizar un comparativo de todos los indicadores en el período seleccionado.



El resultado final del reporte es el siguiente:

