**PRACTICA SEMANA 2**

**GRUPO: # 2**

**INTEGRANTES:**

* **Ángel Gaona,**
* **Daniela Álvarez**
* **Katherine Vera**
* **Christian Rodríguez**

**ANTECEDENTES**

En este estudio, nos proponemos investigar la relación entre el uso de redes sociales y el bienestar emocional. Para ello, utilizaremos un conjunto de datos exhaustivo recopilado por Emirhan Bulut. Este conjunto, dividido en tres tablas (test.csv, train.csv y val.csv), contiene información detallada sobre el comportamiento de los usuarios en diversas plataformas, así como indicadores de su estado emocional.

Buscaremos identificar patrones y correlaciones que nos permitan comprender cómo diferentes aspectos del uso de redes sociales (tiempo de uso, tipo de contenido consumido, interacción social) se relacionan con variables como la satisfacción con la vida, la ansiedad y la depresión.

**Características:**

**User\_ID**: Identificador único del usuario.

**Age:** Edad del usuario.

**Gender:** Género del usuario (Mujer, Hombre, No binario).

**Platform:** Plataforma de redes sociales utilizada (por ejemplo, Instagram, Twitter, Facebook, LinkedIn, Snapchat, Whatsapp, Telegram).

**Daily\_Usage\_Time:** (minutos): Tiempo diario de permanencia en la plataforma en minutos.

**Posts\_Per\_Day:** Número de publicaciones realizadas por día.

**Likes\_Received\_Per\_Day:** Número de Me gusta recibidos por día.

**Comments\_Received\_Per\_Day:** Número de comentarios recibidos por día.

**Messages\_Sent\_Per\_Day:** Número de mensajes enviados por día.

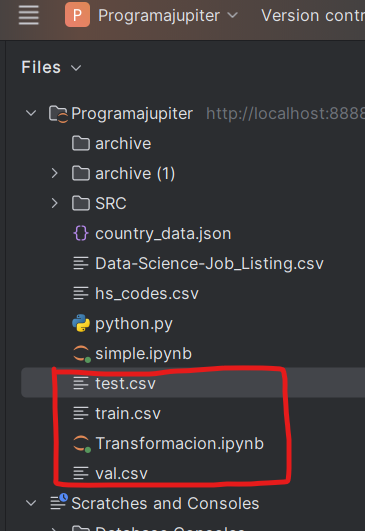
**Dominant\_Emotion:** estado emocional dominante del usuario durante el día (por ejemplo, felicidad, tristeza, ira, ansiedad, aburrimiento, neutral).

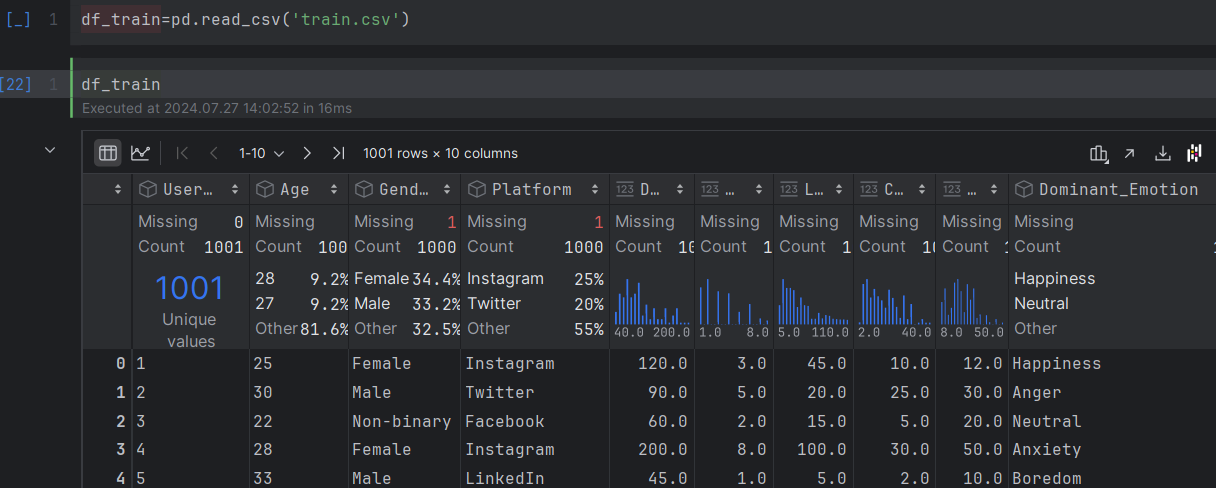
Este conjunto de datos se puede utilizar para varios análisis, incluidos, entre otros:

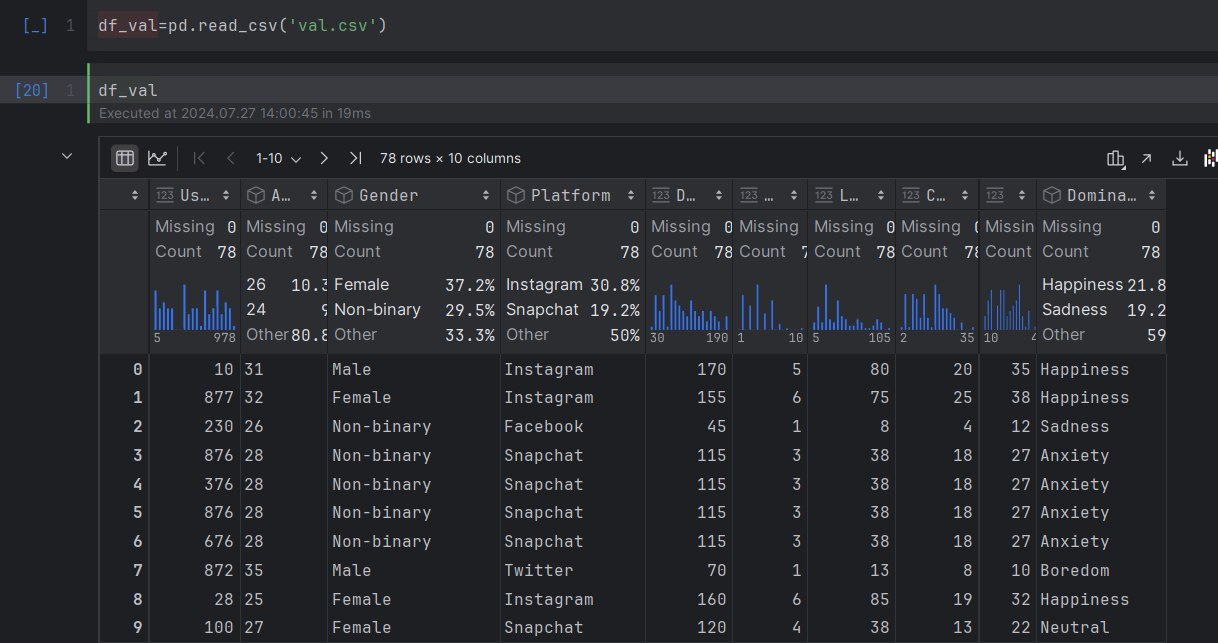
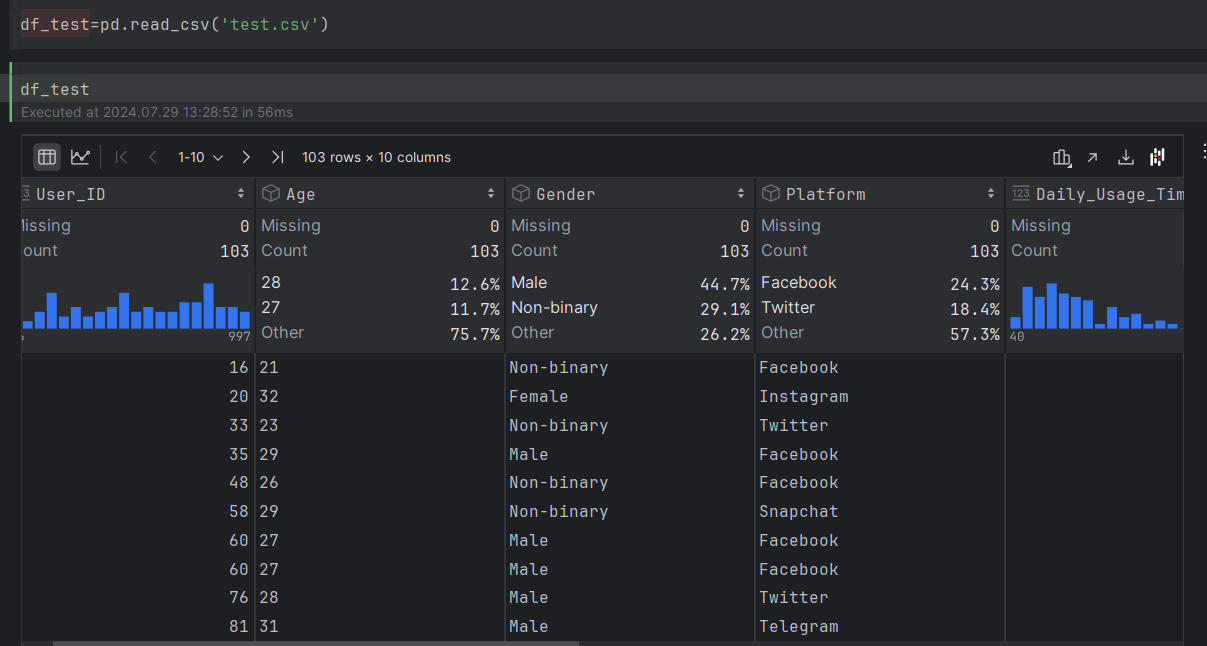
* Predecir el bienestar emocional de los usuarios en función de su uso de las redes sociales.
* Agrupar usuarios según patrones de uso.
* Examinar el impacto de diferentes plataformas en las emociones de los usuarios.

**DESARROLLO**

**Selección de la Data**



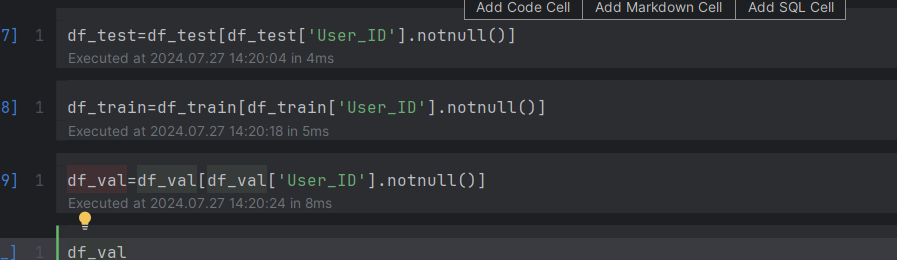


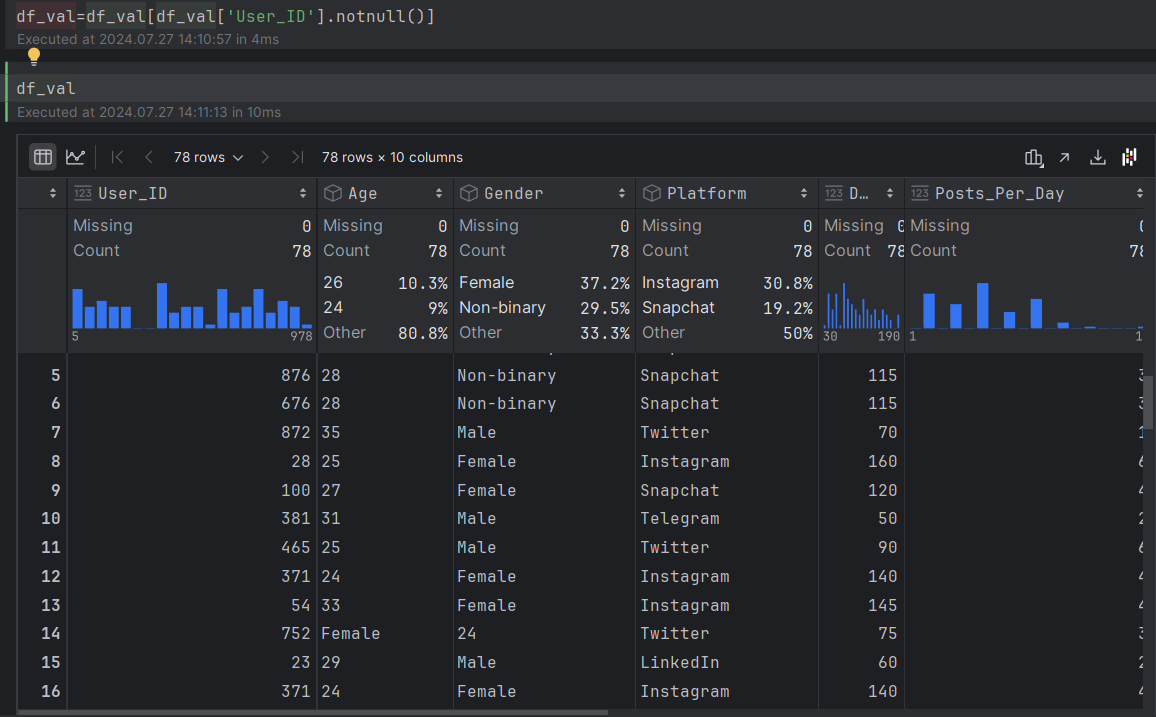


**TRANSFORMACION**

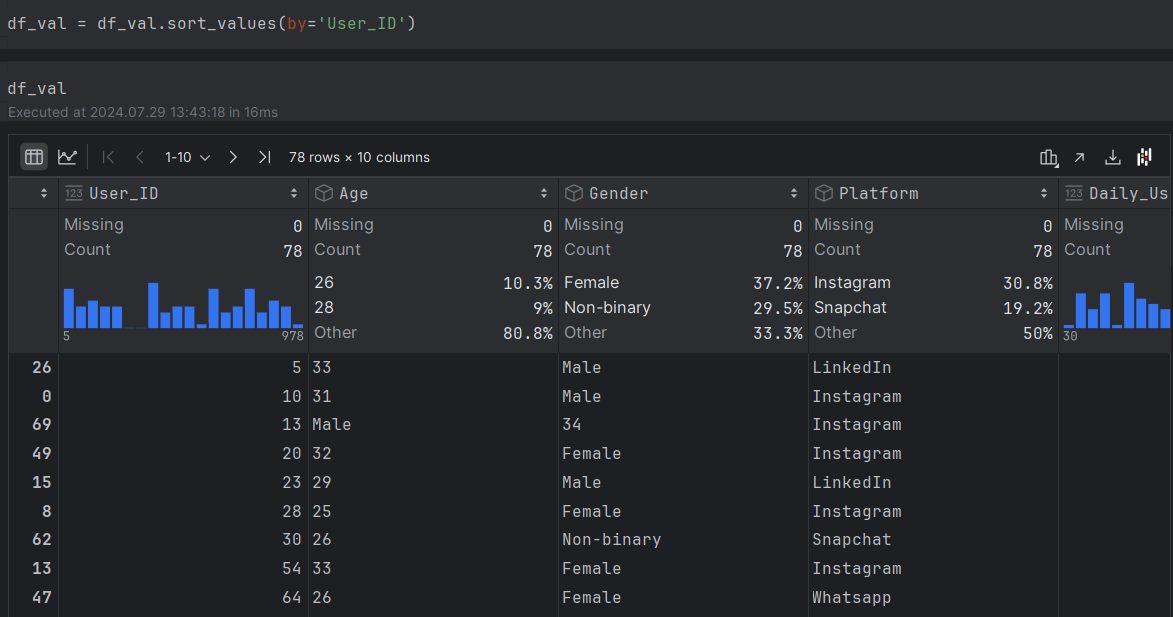
**a) Limpieza de Datos**

Para garantizar la integridad de los datos, procedimos a eliminar las filas que contenían valores nulos en al menos una de las columnas."

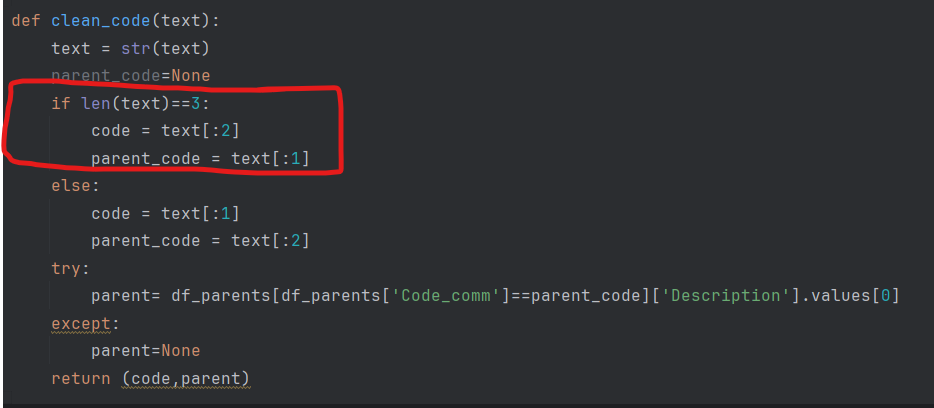




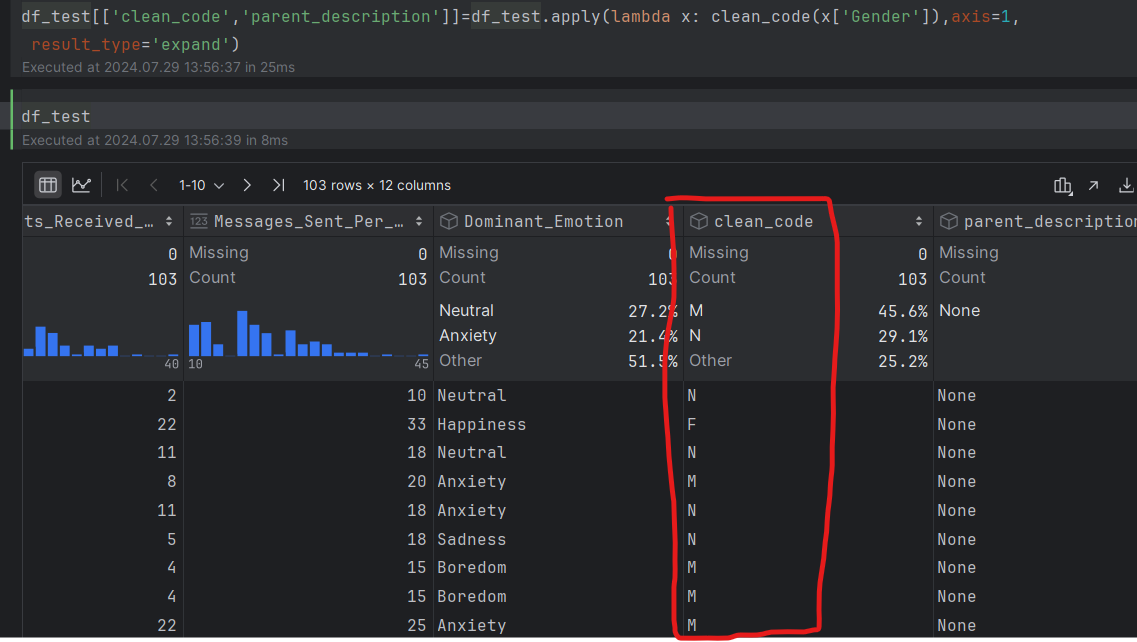
Para facilitar el análisis, se ordenó los datos por el identificador único de usuario (USER\_ID), con la función sort\_values



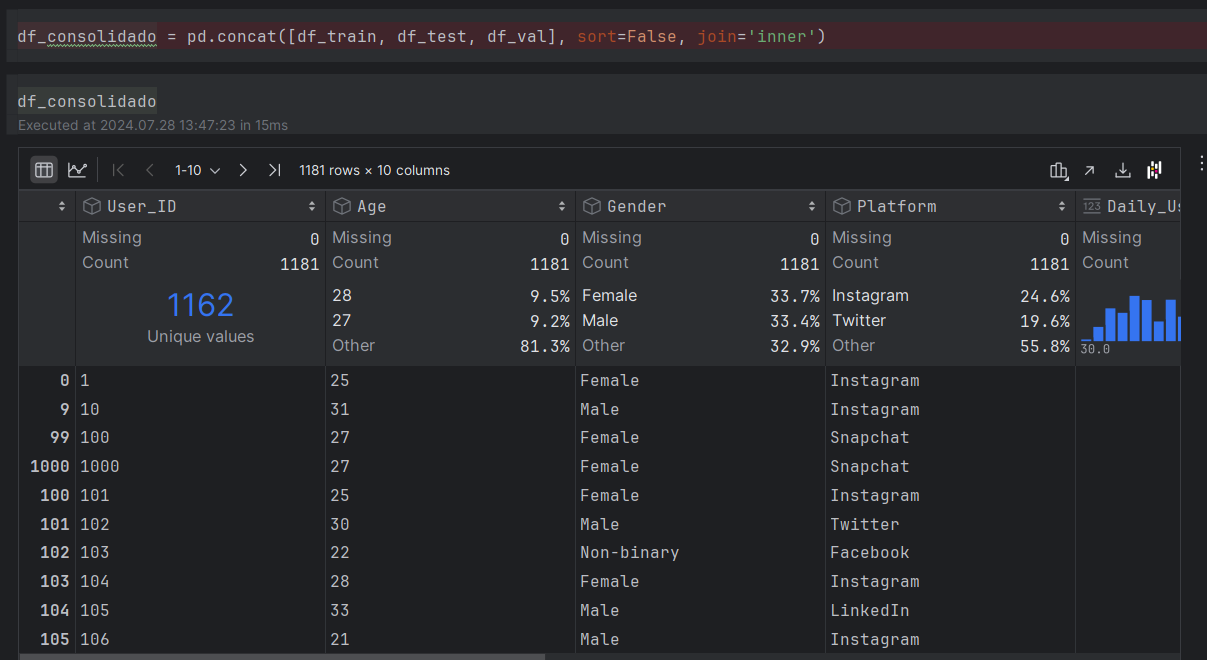
Para fines didácticos, se desarrolló una función llamada Clean Code con el objetivo de agregar una nueva columna a los registros que cumplieran con una determinada longitud. En un primer momento, esta función estaba configurada para buscar códigos de 11 caracteres. Sin embargo, dado que los códigos en nuestro conjunto de datos eran más cortos, se modificó la función para que se adaptara a esta característica particular

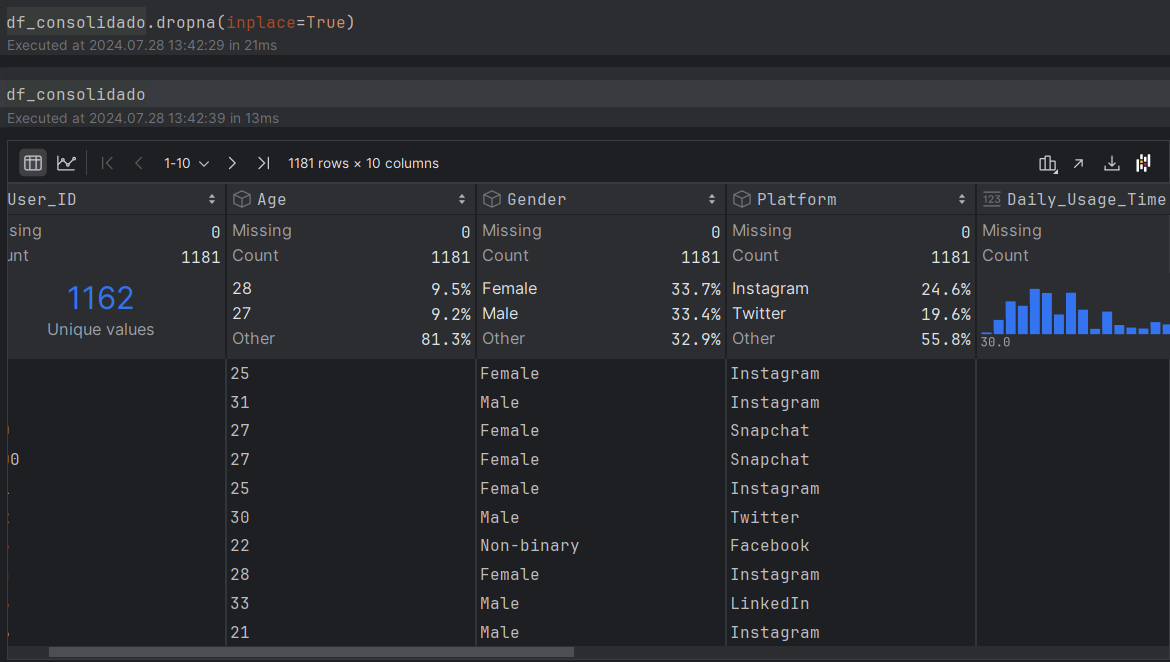


Para obtener la inicial de género de cada registro, se emplea una función lambda que llama a clean\_code, pasando como parámetro la columna 'Gender'. La función clean\_code, a su vez, extrae el primer carácter de cada valor en esta columna y lo devuelve.



Con el fin de simplificar el proceso de análisis, se unieron las tres tablas, que compartían la misma estructura, en una única tabla de hechos denominada df\_consolidado. Esta acción se llevó a cabo mediante la concatenación vertical de las tres tablas, lo que permitió obtener un conjunto de datos más manejable y coherente



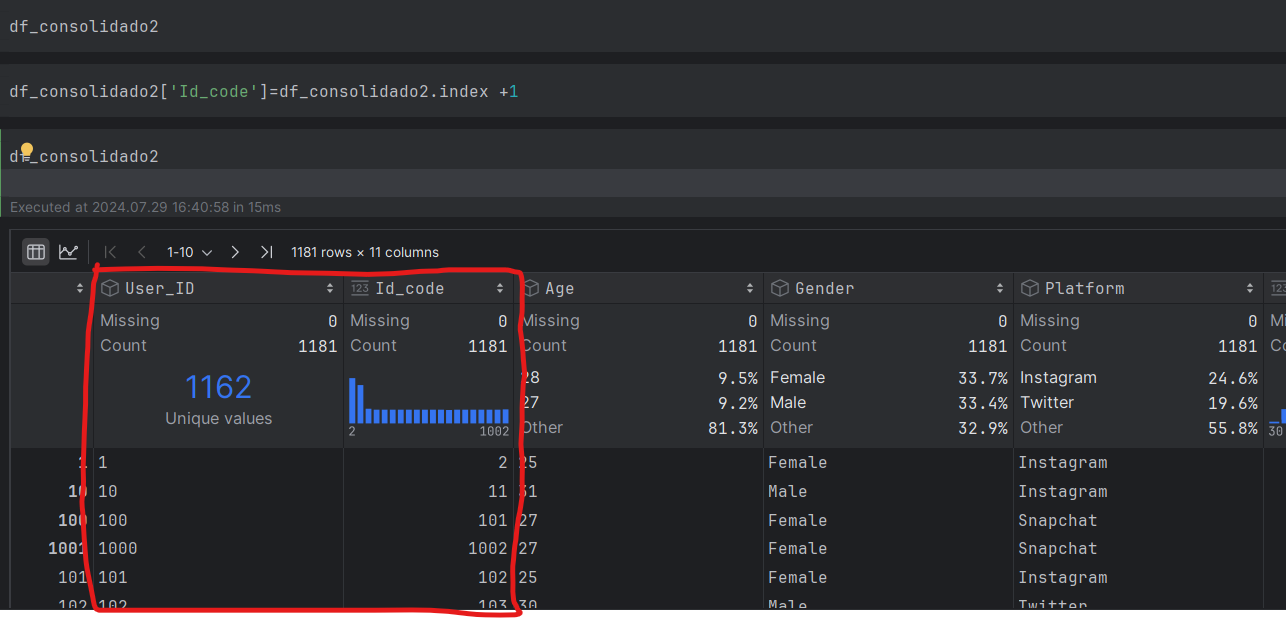
A fin de garantizar la calidad de los datos para el análisis subsiguiente, se aplicó la función dropna() al DataFrame consolidado. Esta función permitió eliminar las filas que contenían al menos un valor nulo o vacío, asegurando así la integridad de los datos

Para proteger los datos originales y facilitar las transformaciones, se crea una copia del DataFrame consolidado, denominada consolidad2. Esta copia servirá como punto de partida para aplicar las operaciones de dimensionamiento necesarias

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

A fin de garantizar la singularidad de cada registro, se incrementa el valor del campo USER\_ID en 1 y se convierte el resultado a un formato numérico, generando así un nuevo identificador único

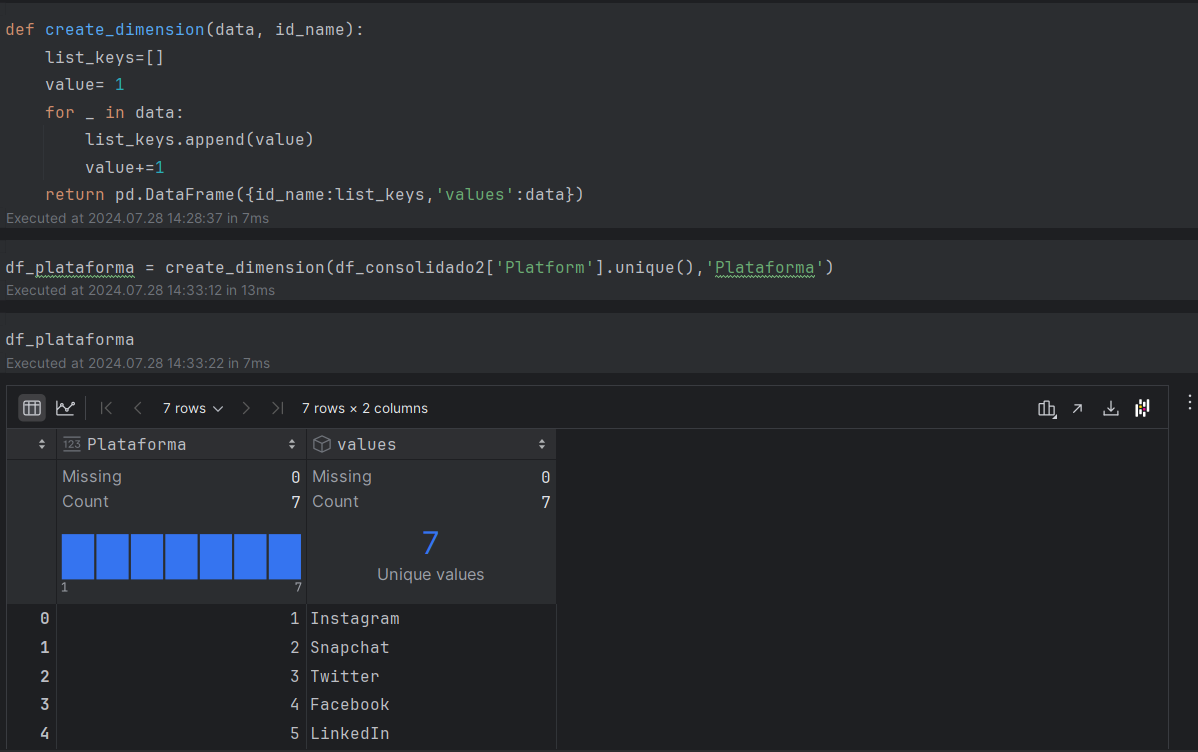


**DIMENCIONES**

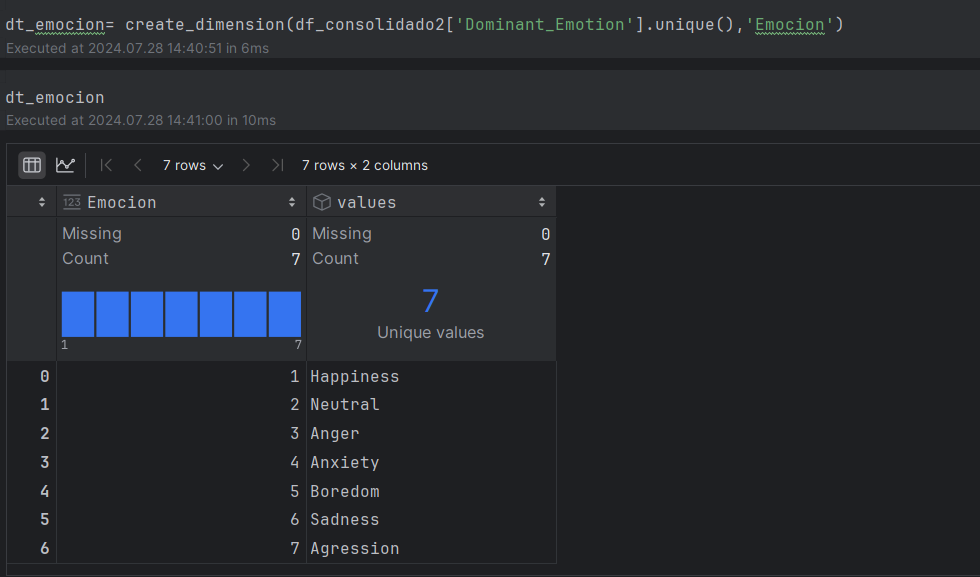
Tras un exhaustivo análisis de los datos, se identificó la necesidad de las siguientes dimensiones para obtener una comprensión integral del fenómeno en estudio:

df\_plataforma , df\_genero, df\_emocion

Creación de dimensión plataforma



Creación de dimensión emoción



Creación de dimensión Genero

Captura de pantalla de un celular

Descripción generada automáticamente