**PYTHON AVANZADO (PYTHON CON LIBRERÍAS DE ANÁLISIS DE DATOS)**

**Fernanfrain Montoya Arbelaez**

**CC:1060652348**

**TALENTO TICS 2023**

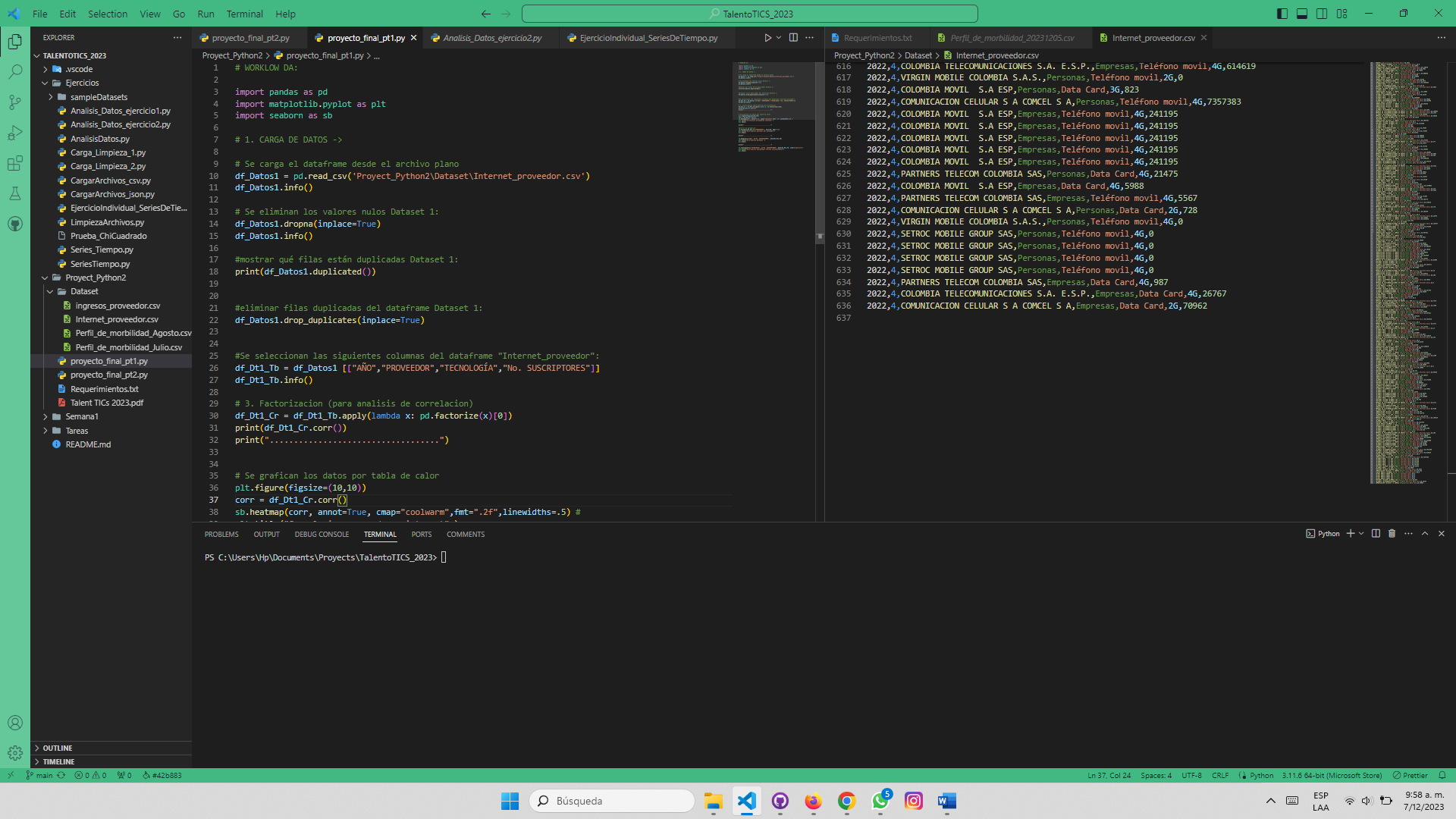
**OBEJTIVOS**

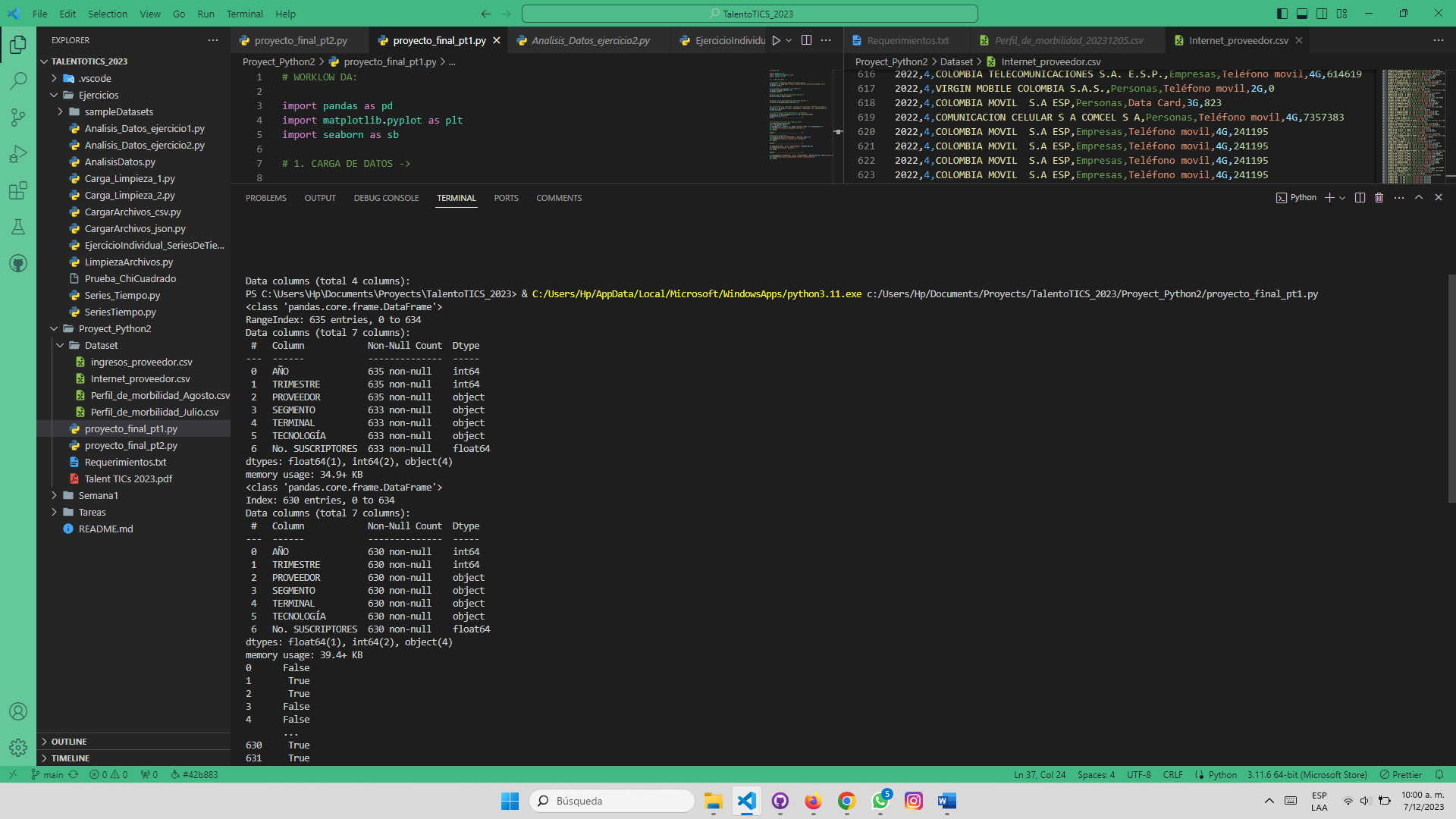
* Hacer uso de funcionalidades avanzadas de Python aplicado a tareas básicas de análisis de datos.
* Usar librerías como pandas, matplolib y demás en tareas básicas de carga, limpieza, análisis y visualización de datos.
* Aplicar de forma correcta las capacidades de análisis necesarias para hacer un uso adecuado de las librerías en tareas de análisis de datos.

Se realiza el análisis de datos de los siguientes Dataset los cuales se encuentran en el repositorio de GitHub al igual que el workflow del proyecto en Python.

* Repositorio GitHub: <https://github.com/Efrain06/TalentoTICS_2023/tree/799a72d9b0f314755075919961ecb41c8544b49f/Proyect_Python2>
* Carga del Dataframe

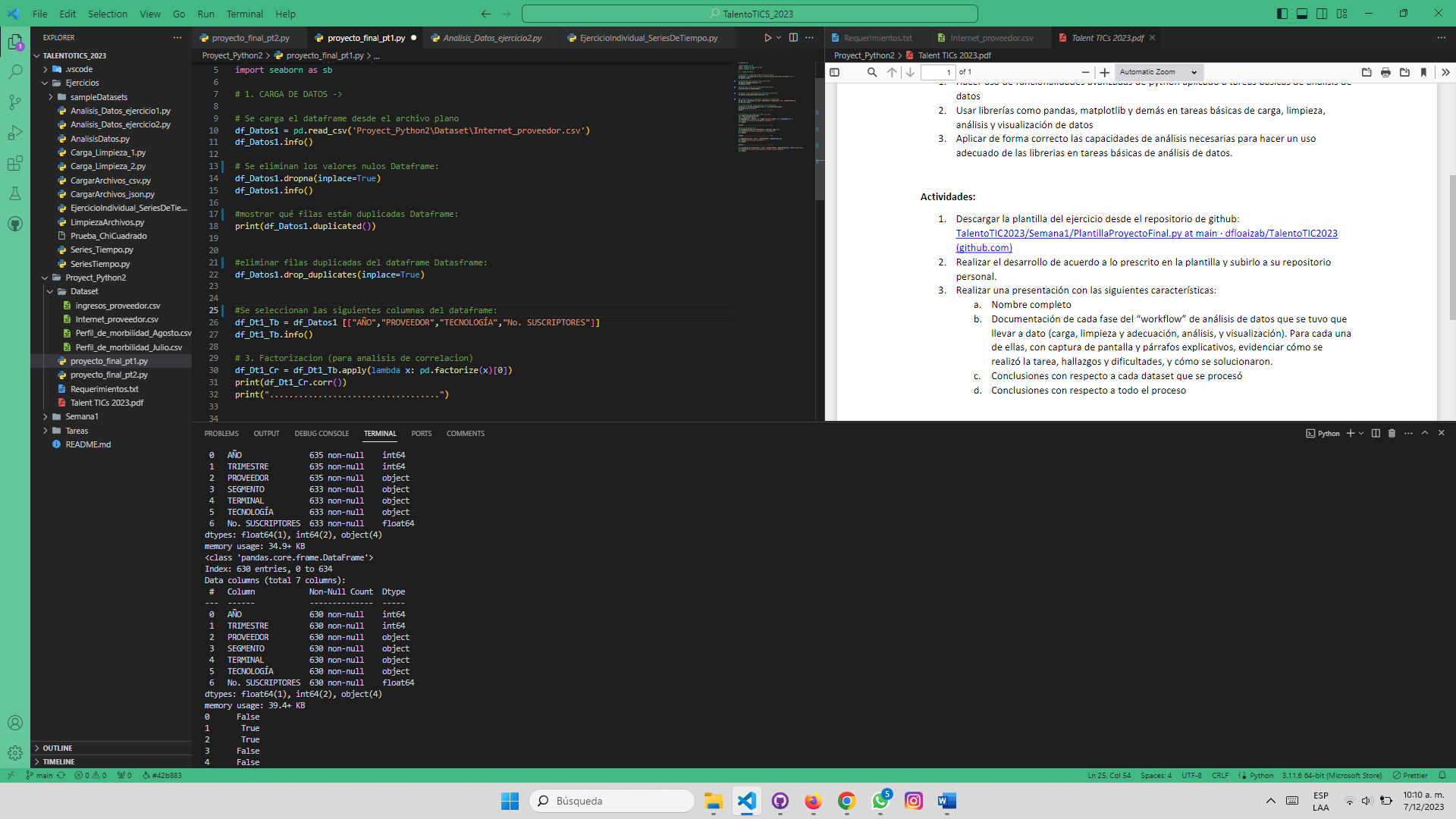
Se utiliza el comando read\_csv de la librería de Pandas, el cual carga con éxito el Dataframe mostrando como resultado 7 columnas; las cuales podemos visualizar que en la columnas “Segmento, terminal, tecnología y No. Suscriptores” poseen valores nulos.





* Limpieza del Dataframe

Se utiliza el comando dropna de la librería de Pandas, El cual elimina con éxito los valores nulos del Dataframe para continuar con un analisis de datos mas limpio y organziado; verificamos tambien si posee filas duplicadas con el comando “duplicated” y “drop.duplicates”.

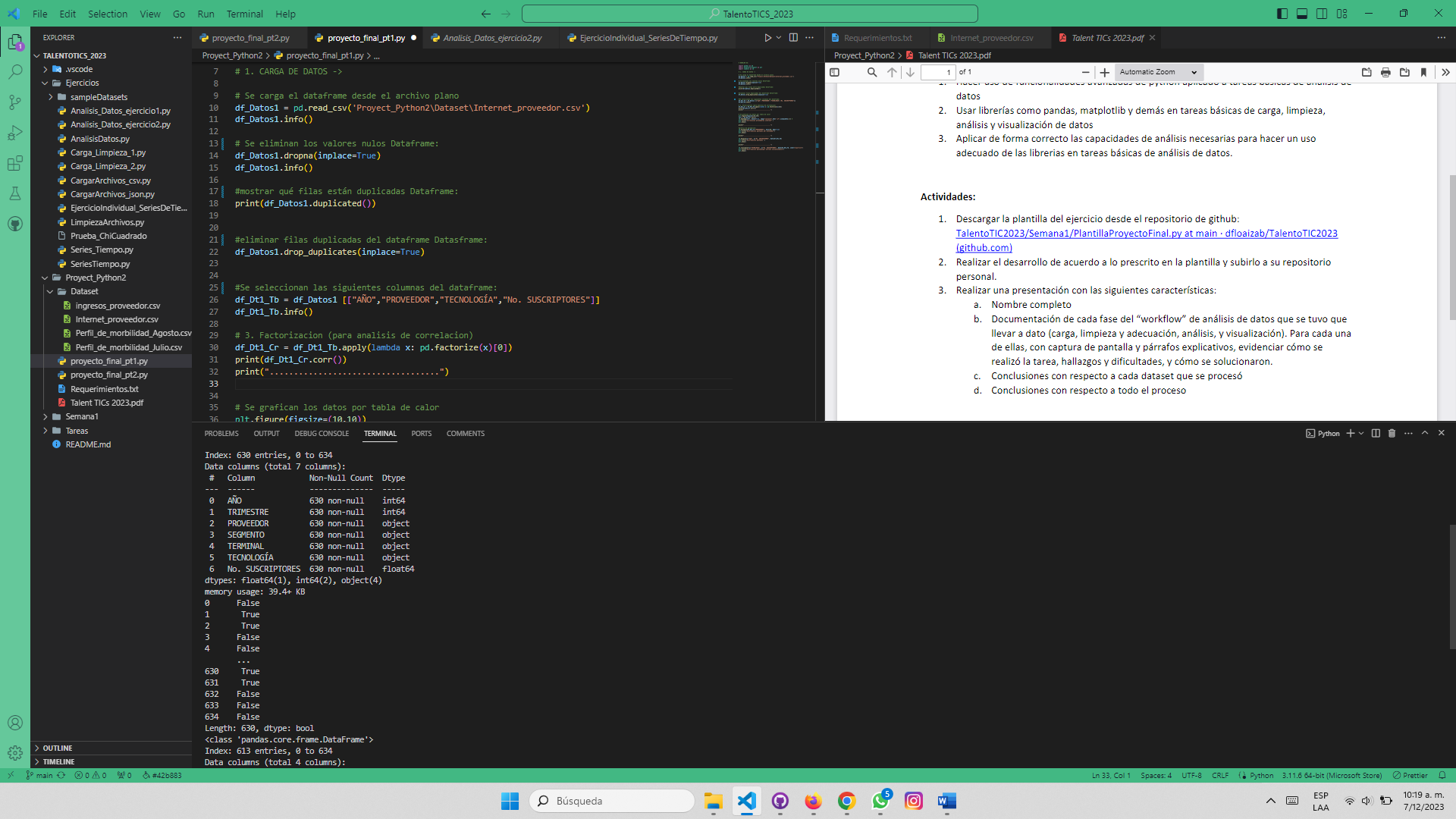


Texto

Descripción generada automáticamente

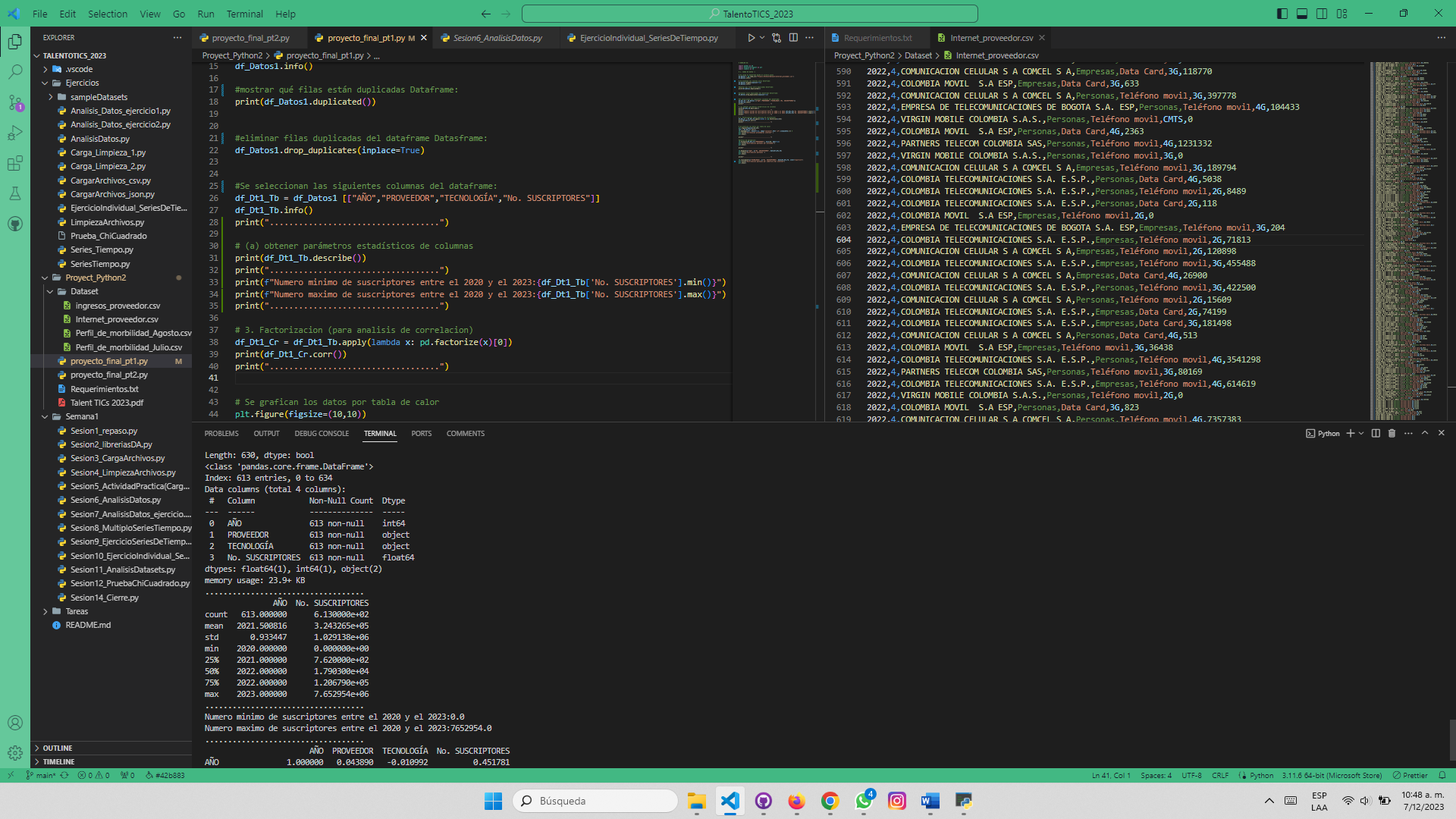
Texto

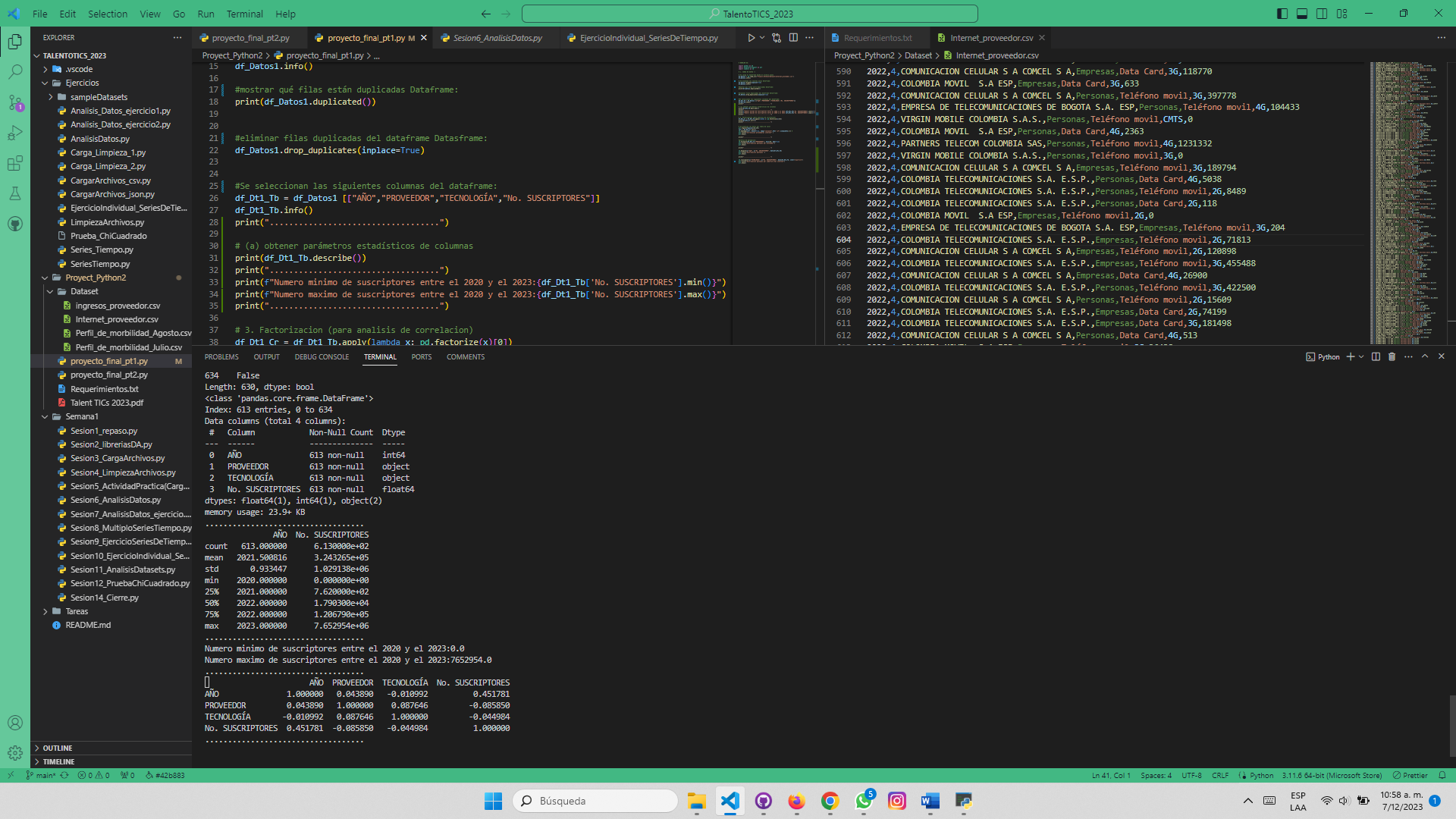
Descripción generada automáticamente



* Tareas de análisis básicas del dataset:

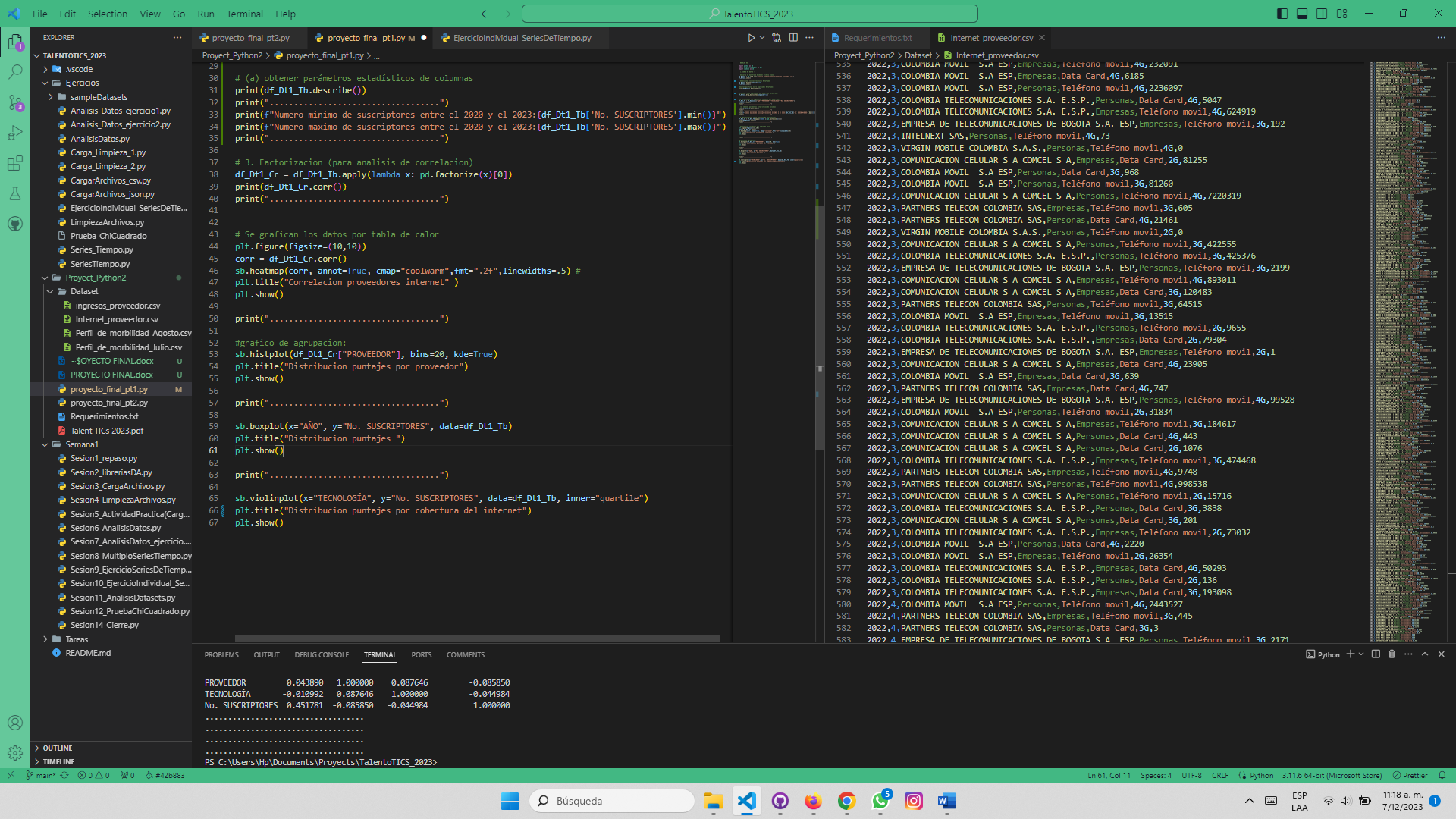
Se seleccionan las columnas que tomaremos como prioridad para iniciar el analisis de datos del Dataframe para lo cual iniciamos con datos estadisticos como el comando “describe” de la librería Pandas el cual nos arroja valores estadisticos del Dataframe ya depurado “df\_Dt1\_Tb; tambien utilizamos la linea de codigo “factorize” y “corr” para ver la correlacion del Dataframe.

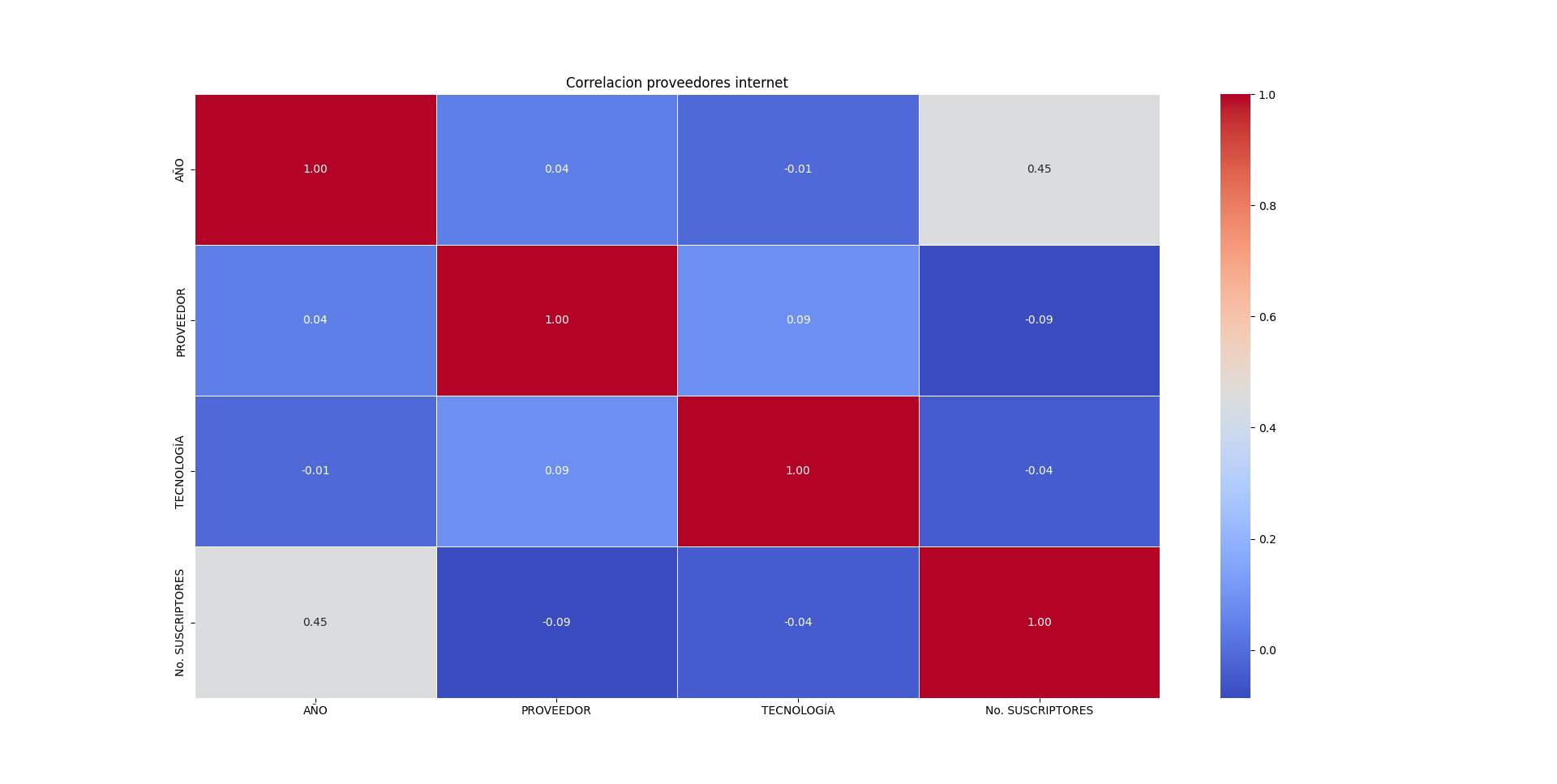


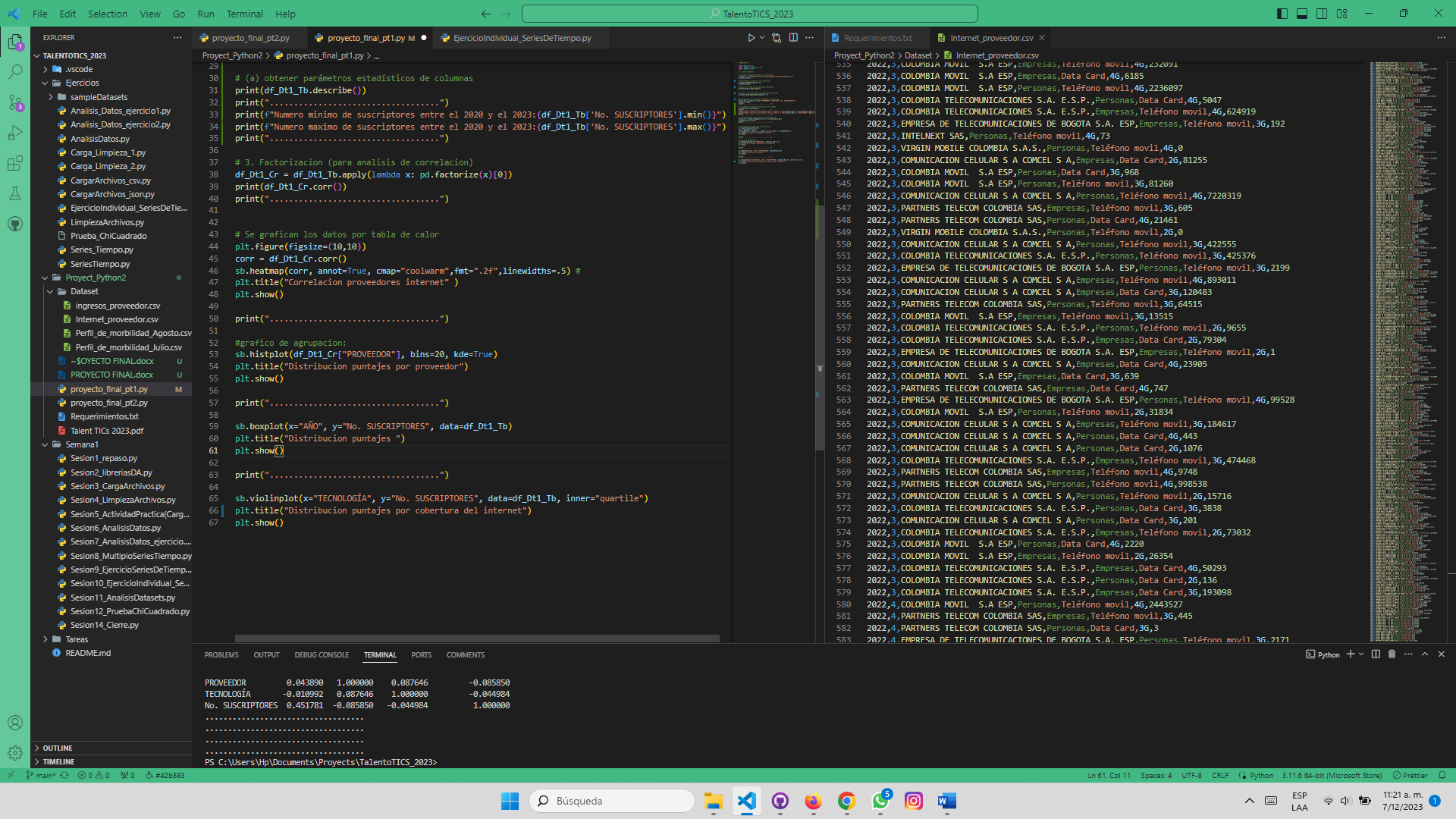


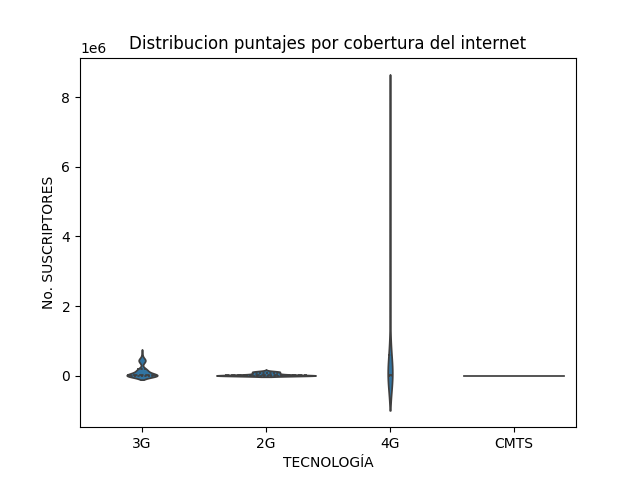
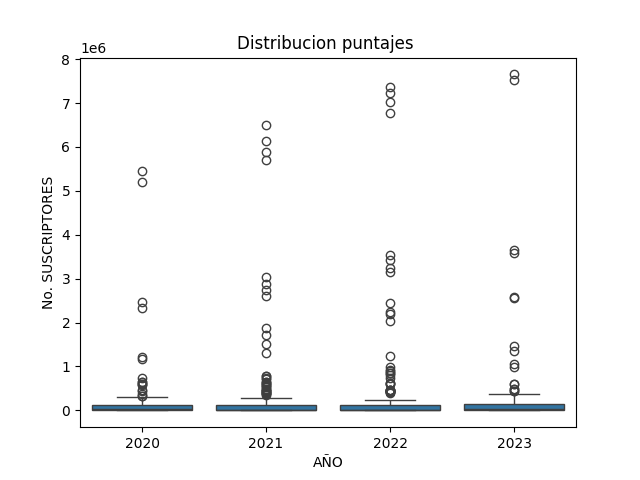
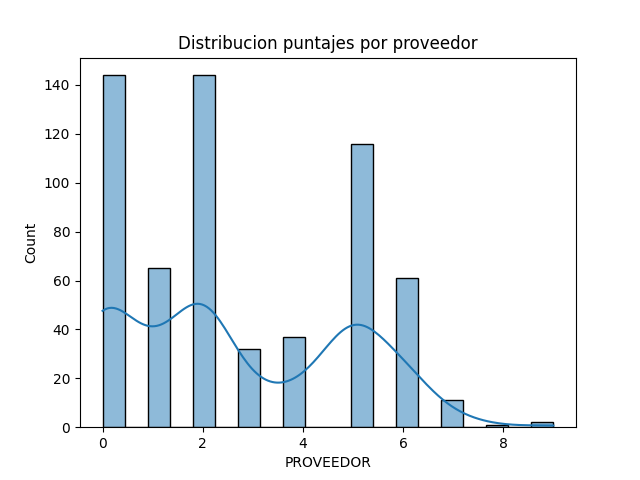
Del analisis estadistico podemos ver que los datos suministrados en el Dataframe no cuentan con alta correlacion.

A continuacion se realiza el analisis de datos con diferentes tipos de graficos para tener la informacion mas concisa posible.



Nota: No posee correlación entre los datos del Dataframe.



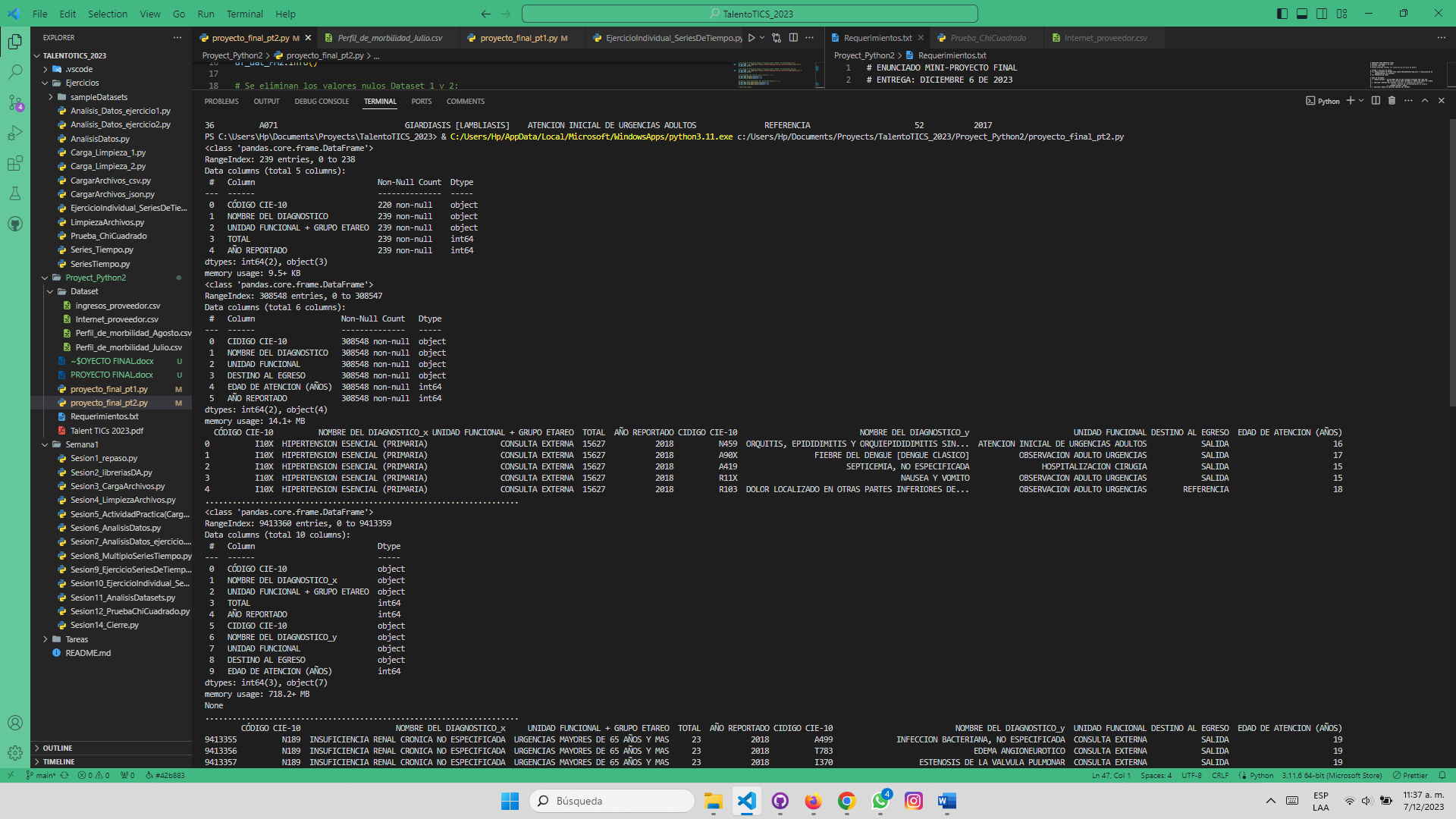


Nota: El internet con mayor cobertura en tecnología es el 4G.

Para el siguiente punto del proyecto “Series de tiempo” seleccionamos los siguientes Dataset los cuales se encuentran en el repositorio y también el workflow.

Se realiza la limpieza del Dataframe eliminando los valores nulos y las filas duplicadas de los dos Dataframe; iniciando el análisis de datos utilizando “merge”, “to\_numeric” de la librería de Pandas.



Nota: Se visualiza la informacion de cada uno de los Dataframe

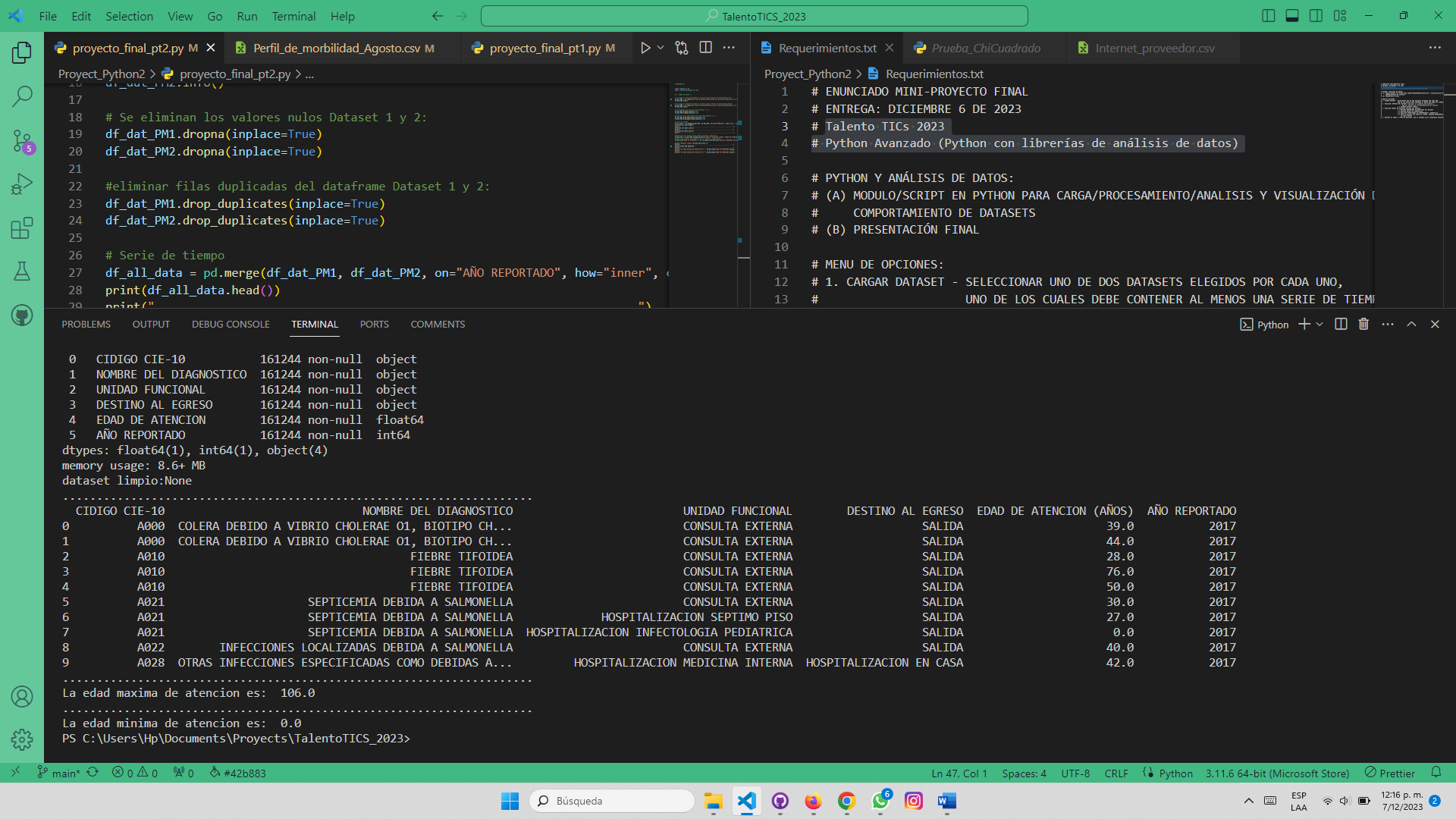
Una captura de pantalla de una computadora

Descripción generada automáticamente

Nota: Podemos visualizar con el método head las primeras 5 filas del Dataframe (“df\_all\_data”), resultado de la unión de los dos Dataframe con el método “merger”; también visualizamos la información de Dataframe unido.

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza mediaNota: Con el metodo “tail”, visualizamos las 5 ultimas filas del Dataframe; tambien visualizamos la informacion del Dataframe con la modificacion del nombre de la columna “Año reportado”.



Nota: Con el metodo “to\_numeric” convierte los valores de “edad de atenion” en numeros enteros.

**CONCLUSIONES**

Python es uno de los lenguajes mas utilizados hoy en día en la analítica de datos, también es utilizado en compañía de lenguajes funcionales que permiten una mejor comprensión y visualización de los datos por la variedad de librerías que este maneja.

Este curso de Python 2 enfocado a la analítica de datos es un gran paso al campo laboral que este abarca, hoy en día el análisis los podemos analizar mediante gráficos, tablas, esquemas, con tal de conocer las necesidades por la que estamos realizando el análisis a un Dataset