**ETL - Extracting, Transforming, and Loading Data**

ETL: The process of combining data from multiple sources, preparing the data, and loading the result into a destination.

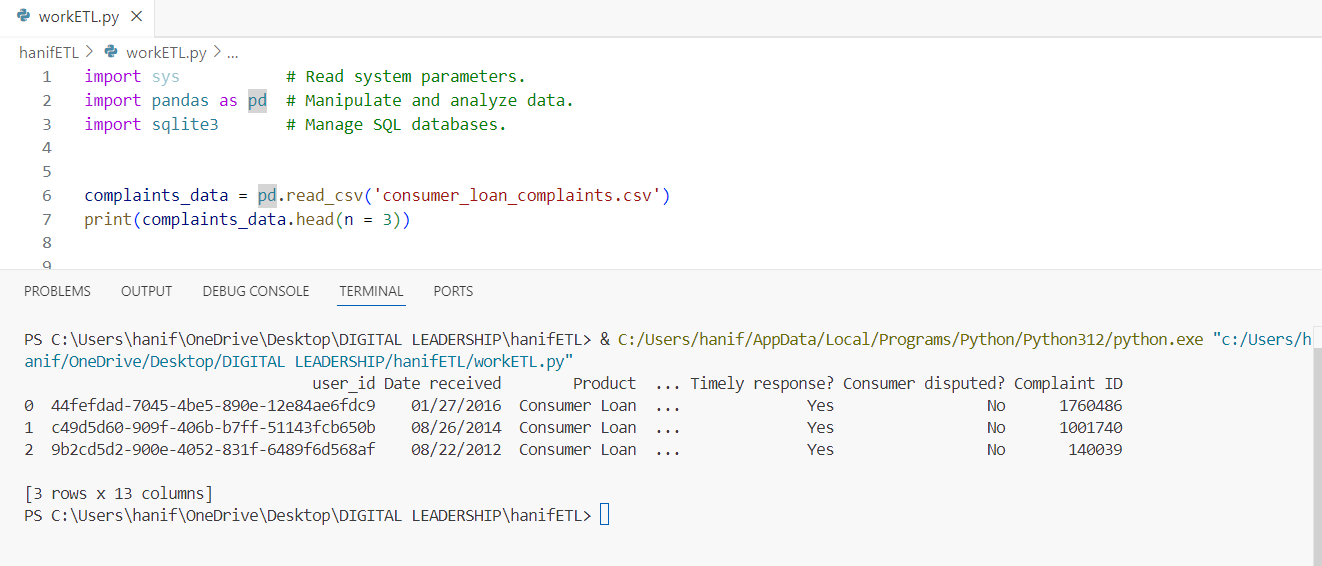
**Objective**

- learn to extract data from various data

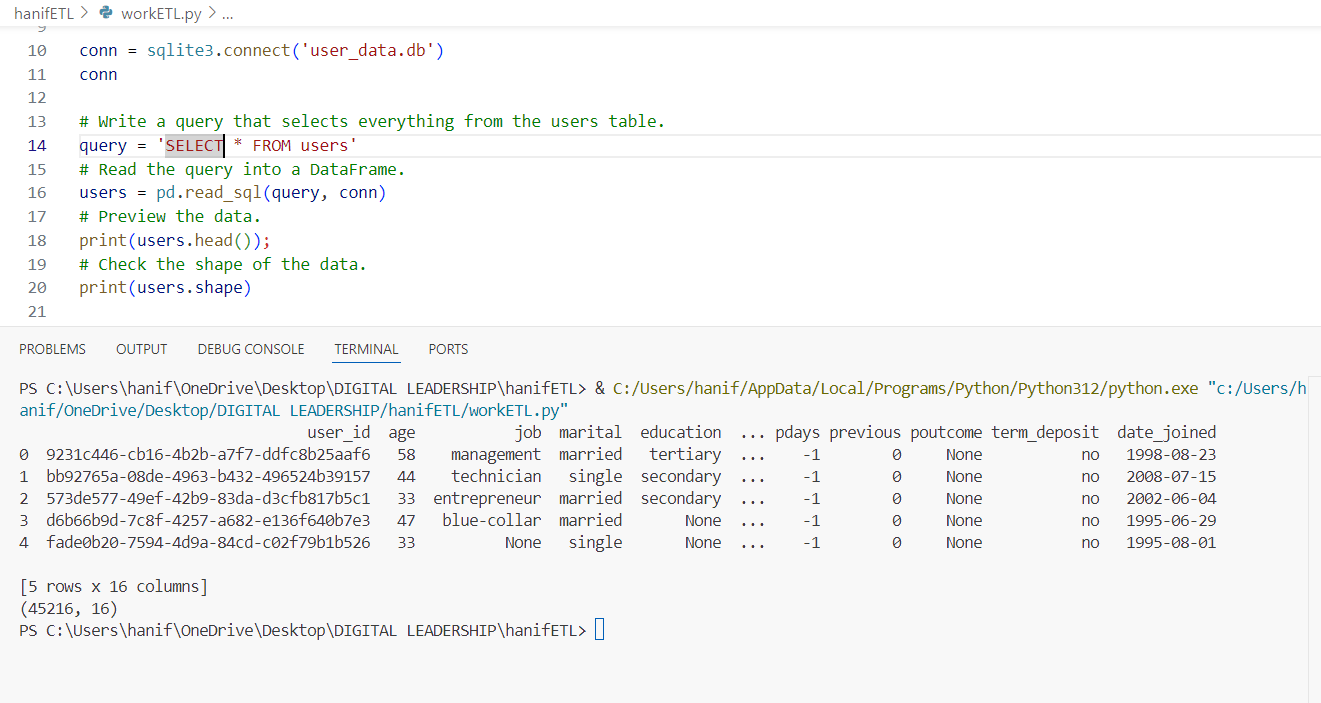
- learn to transform data into a proper format

- learn to load data into a destination

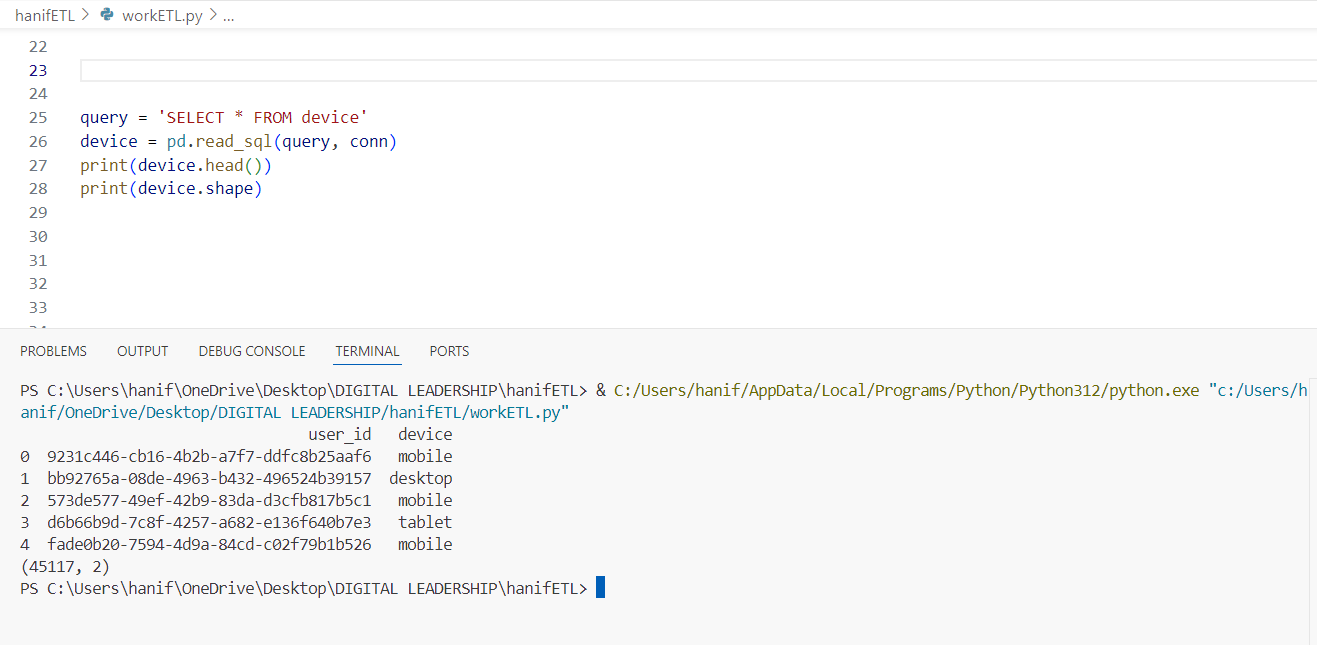
**1)Extract Data**



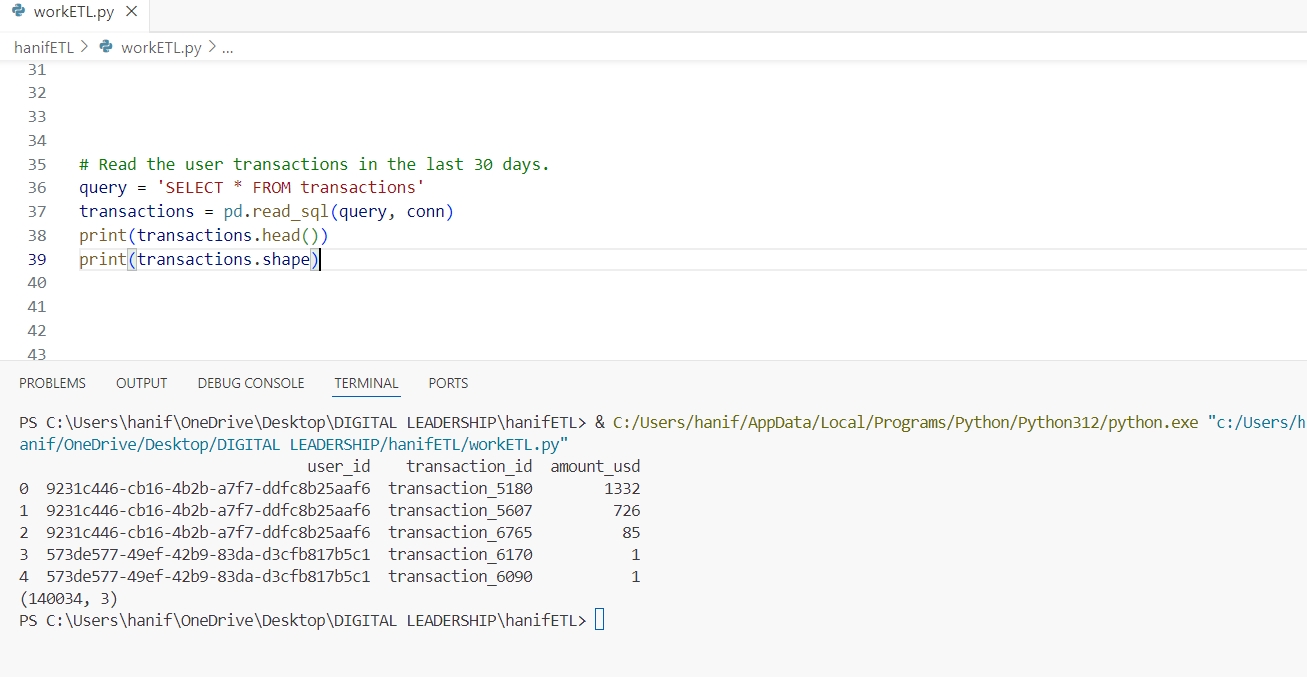
* Load the data from csv file
* Display the data and showing the 3 data first
* The column or feature is 13



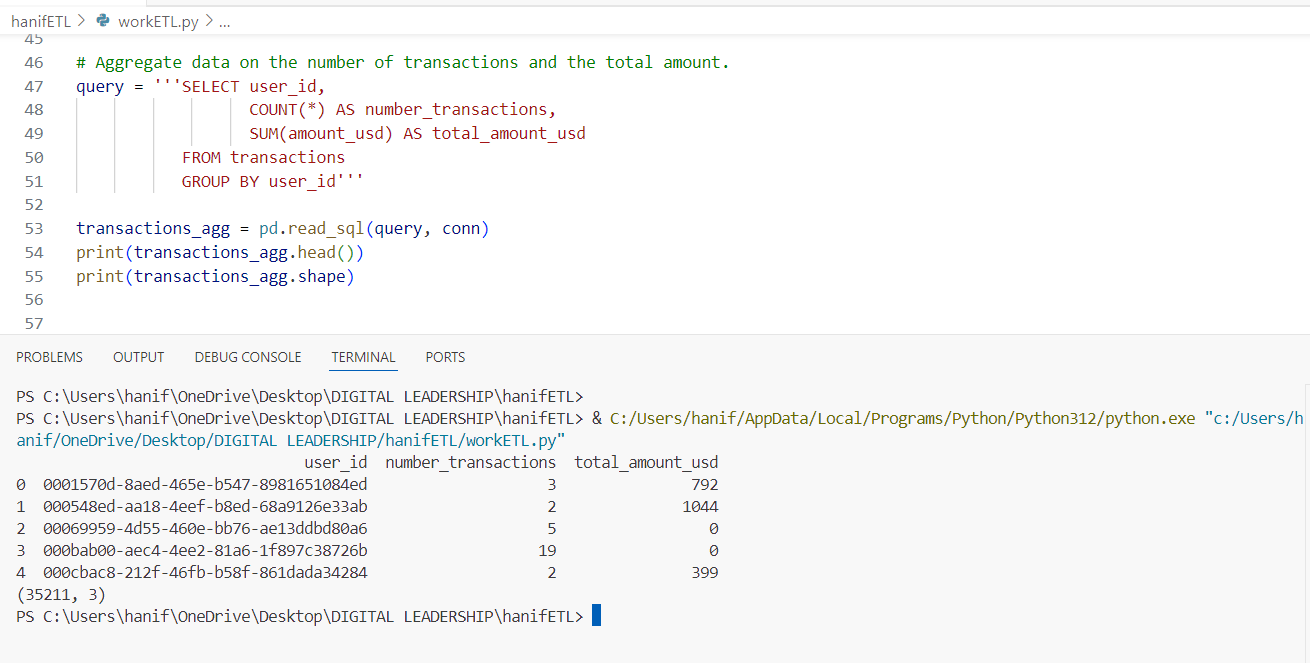
* Load the data from database user\_data.db
* select the users data
* Display data and showing the data by default 5 row only
* The column or feature is 16
* The total all data or rows is 45216



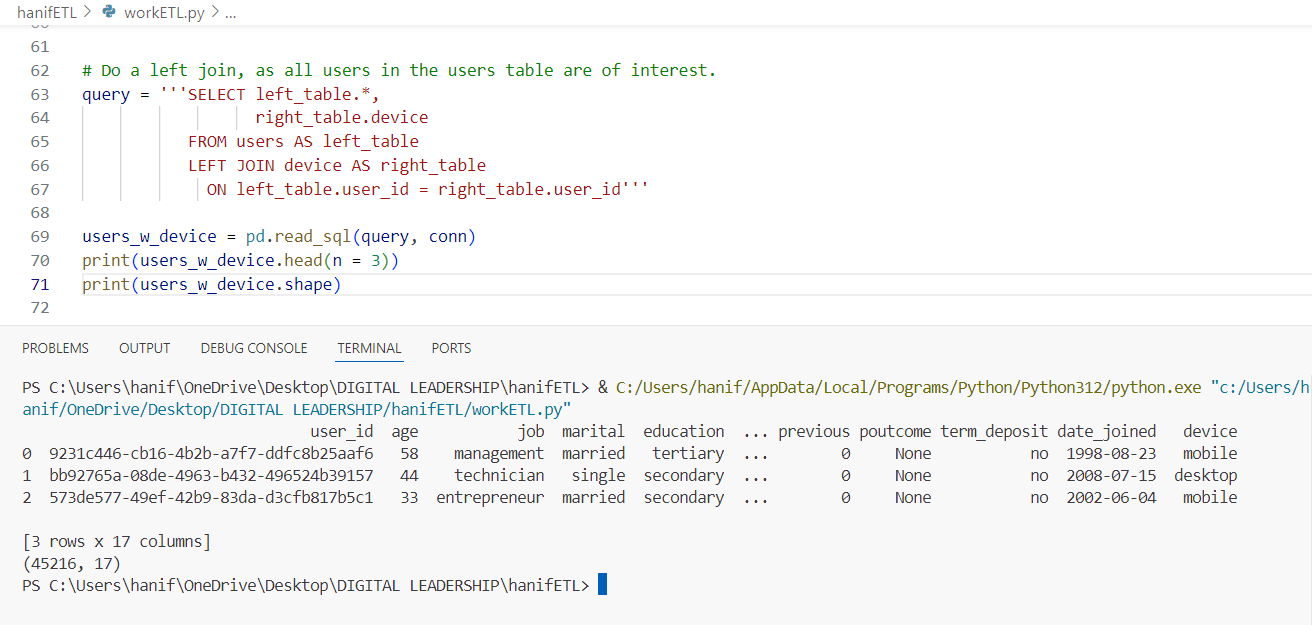
* select the device data
* Display data and showing the data by default 5 row only
* The column or feature is 2
* The total all data or rows is 45117



* select the transaction data
* Display data and showing the data by default 5 row only
* The column or feature is 3
* The total all data or rows is 140034



* Select data user with total transaction and total amount usd
* Display the data and showing the 3 data first
* Display data and showing the data by default 5 row only
* The total all data or rows is 35211



* Join 2 table user and device using sql query
* Display the data and showing the 3 data first
* The column or feature is 17
* The total all data or rows is 45216



* Merge data user with device and transaction
* Display the data and showing the data by default
* The column or feature is 19
* The total all data or rows is 45216

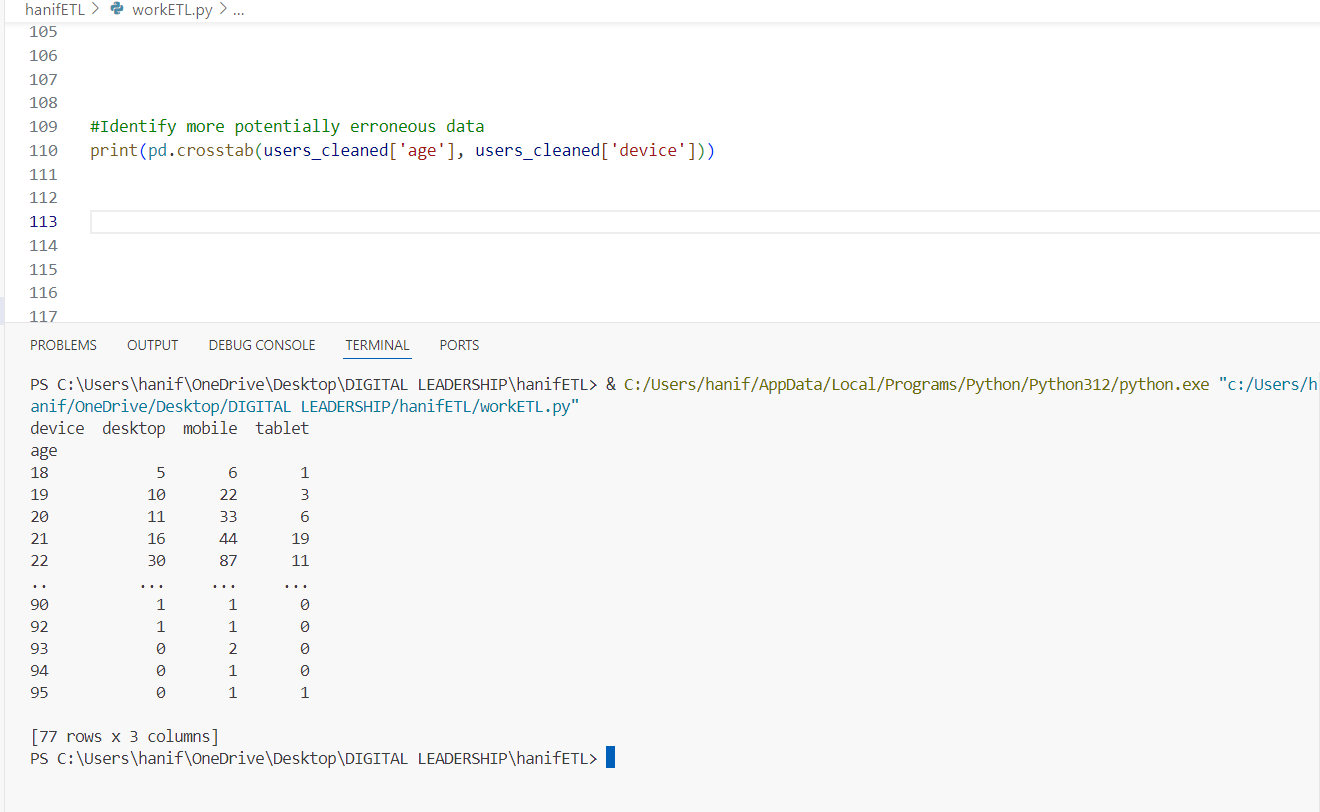
**2)Transform Data**



* Check the data user age is more than 150
* The column or feature is 19
* The total all data or rows is 2



* Remove data which user age is more 150 or keep data user age below 150
* The column or feature is 19
* The original total all data is 45216 to 45214



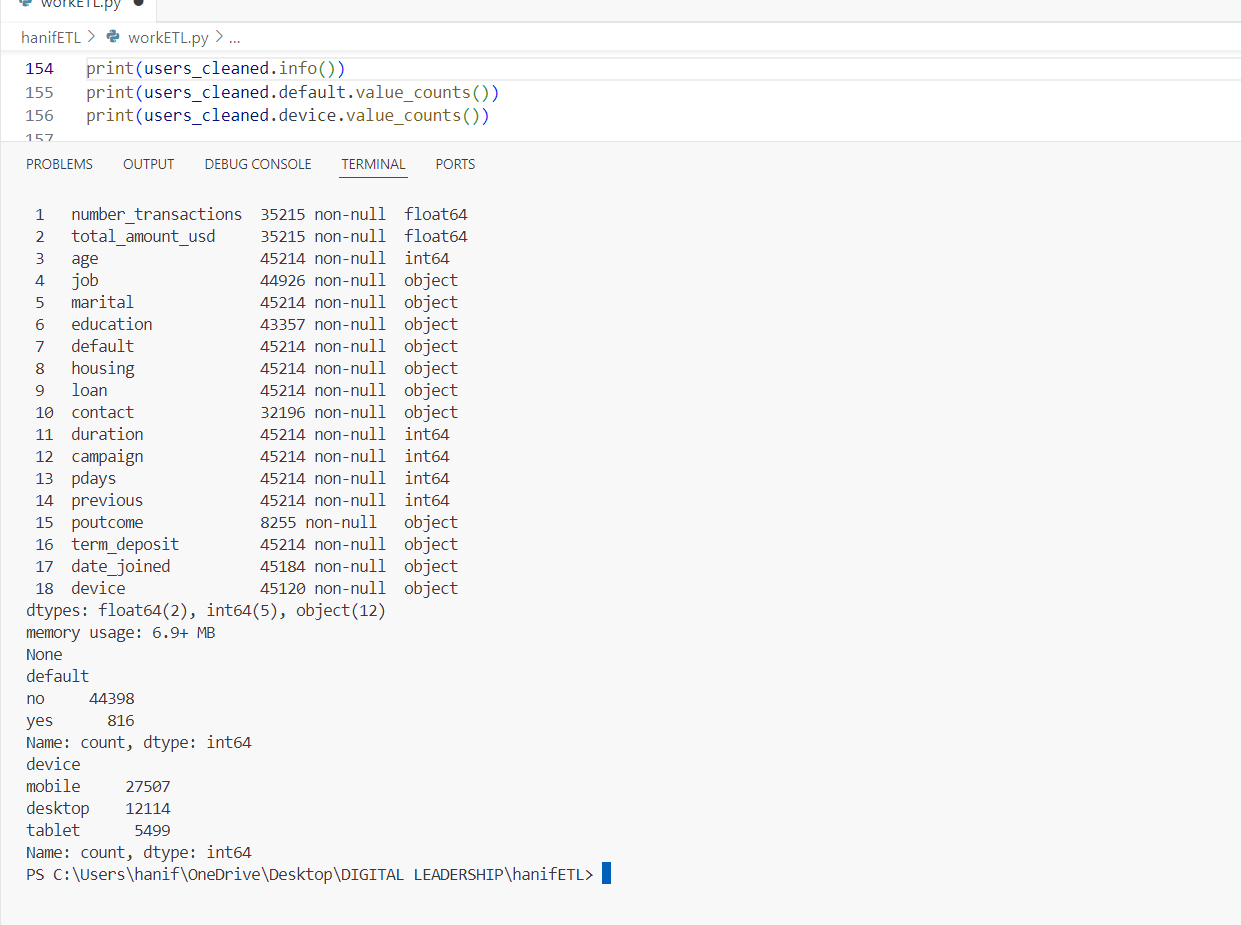
* Do crosstab data from row to column/feature such device [desktop, mobile, tablet]
* Display the total device used based on age data
* the age, desktop, mobile and tablet as feature
* The column or feature is 3
* The total all data or rows is 77



* Find the maximum and minimum total device used by user based on age



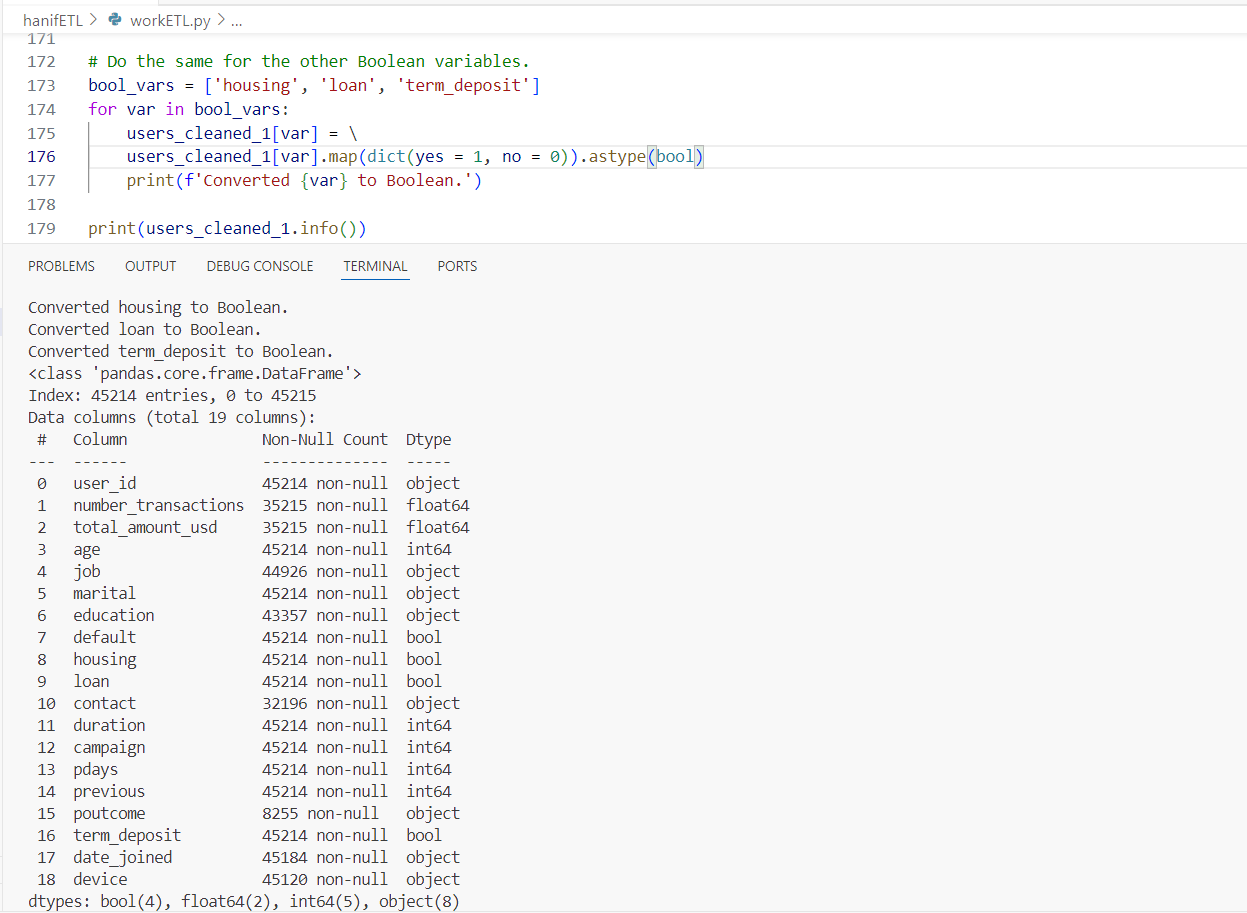
* Find the maximum and minimum total for each device used by user based on age



* Get the data info about structure user table like feature and type of data
* Total data of value for default feature
* Total data of value for device feature



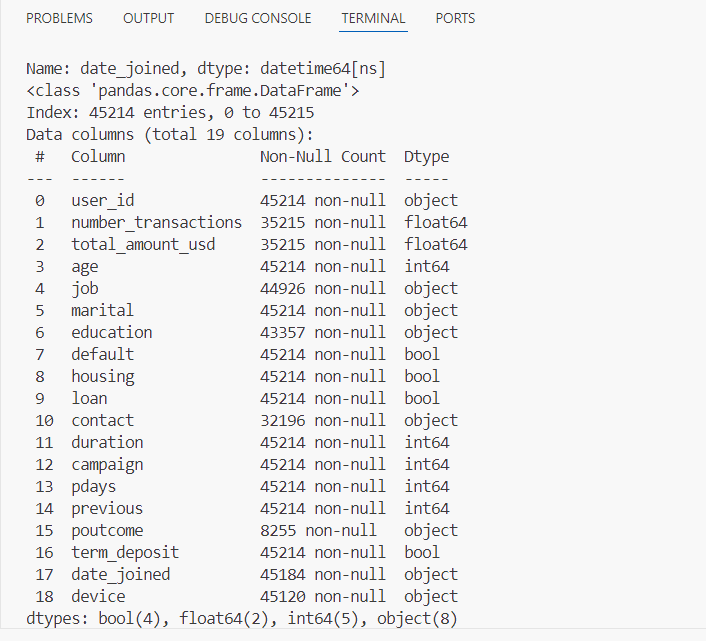
* Convert the default value to bool since only yes and no as value
* Display the data and showing the data by default
* Show feature default value



* Convert the housing, loan and term\_deposit value to bool format



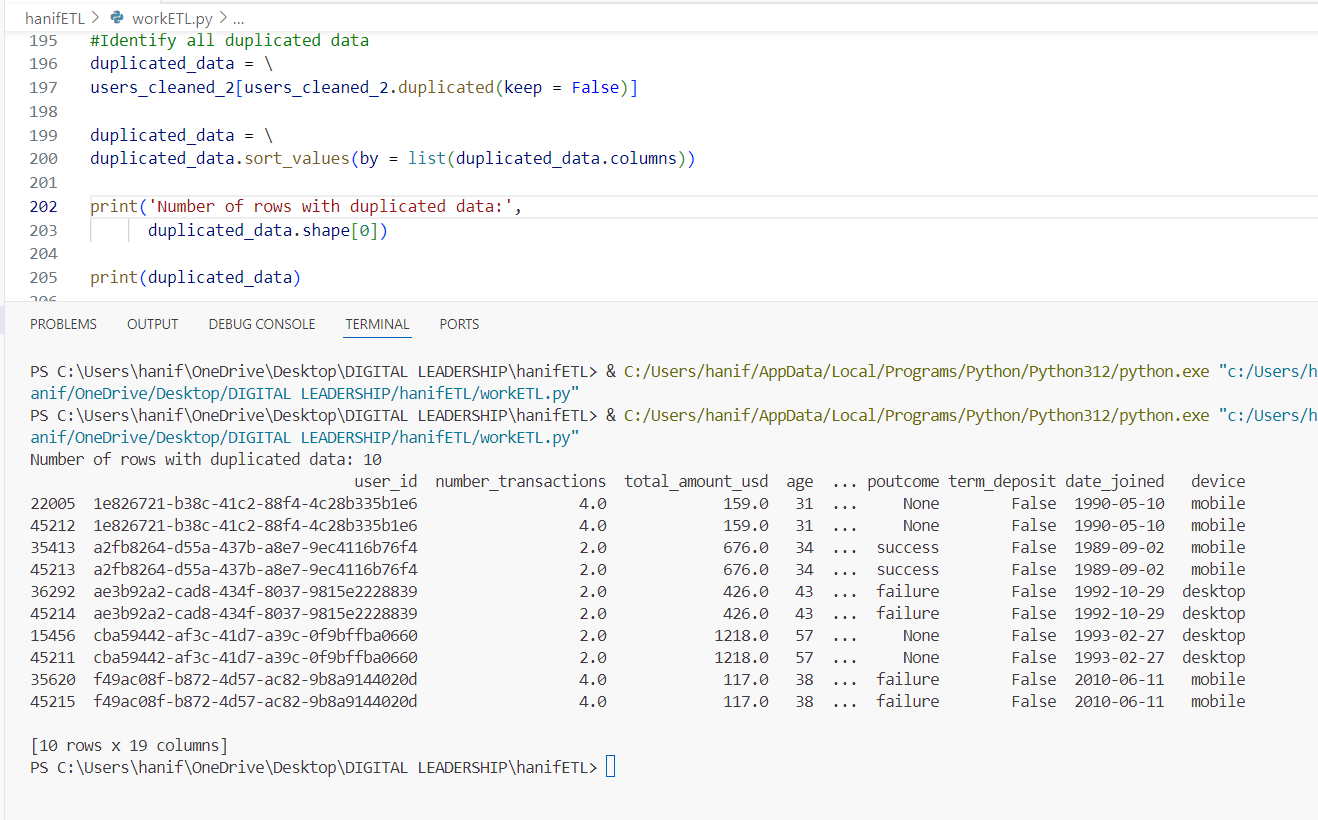
* Convert the date\_joined value to date format as Y-M-D
* Show feature date\_joined value



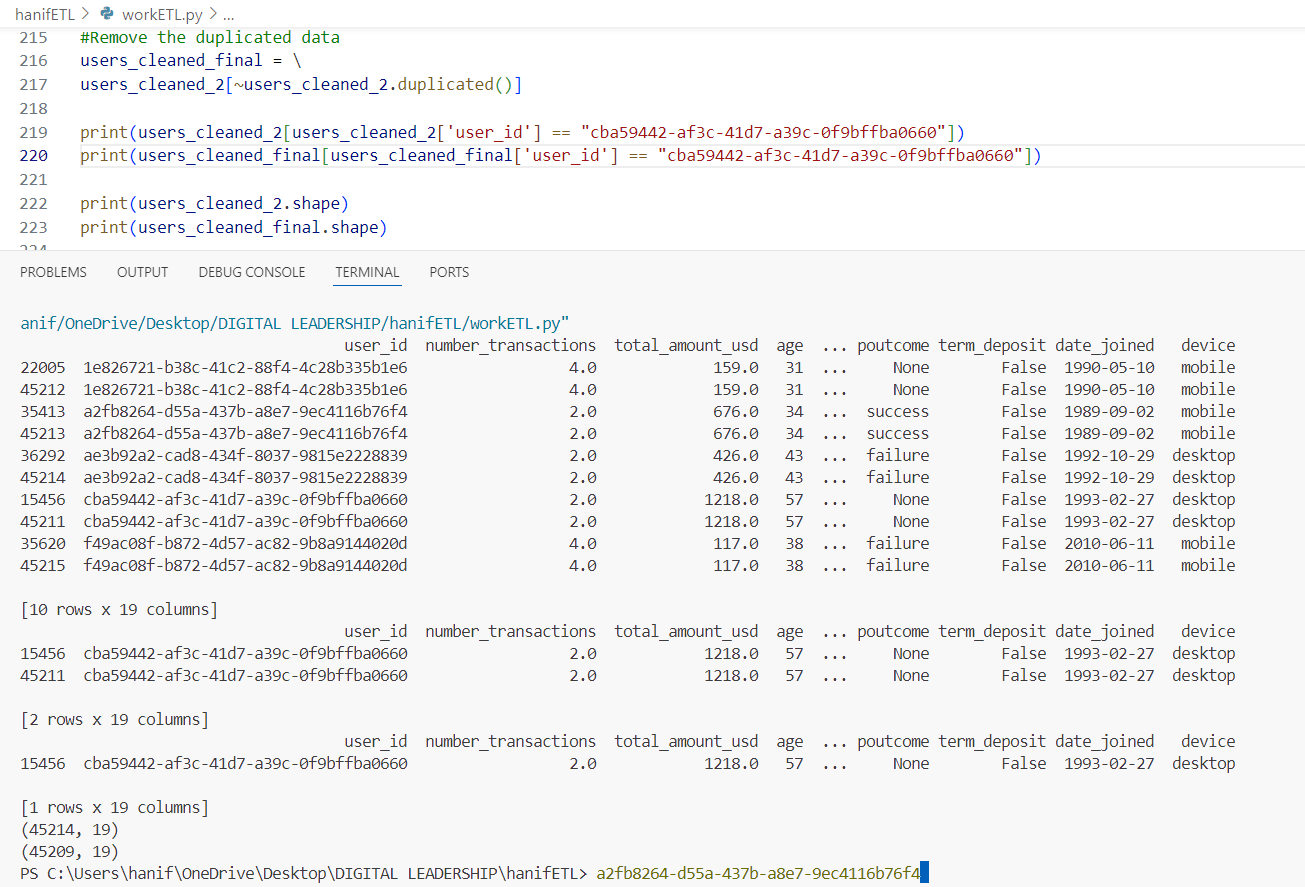
* Check data format for date\_joined before - object



* Check data format for date\_joined after – datetime64[ns]

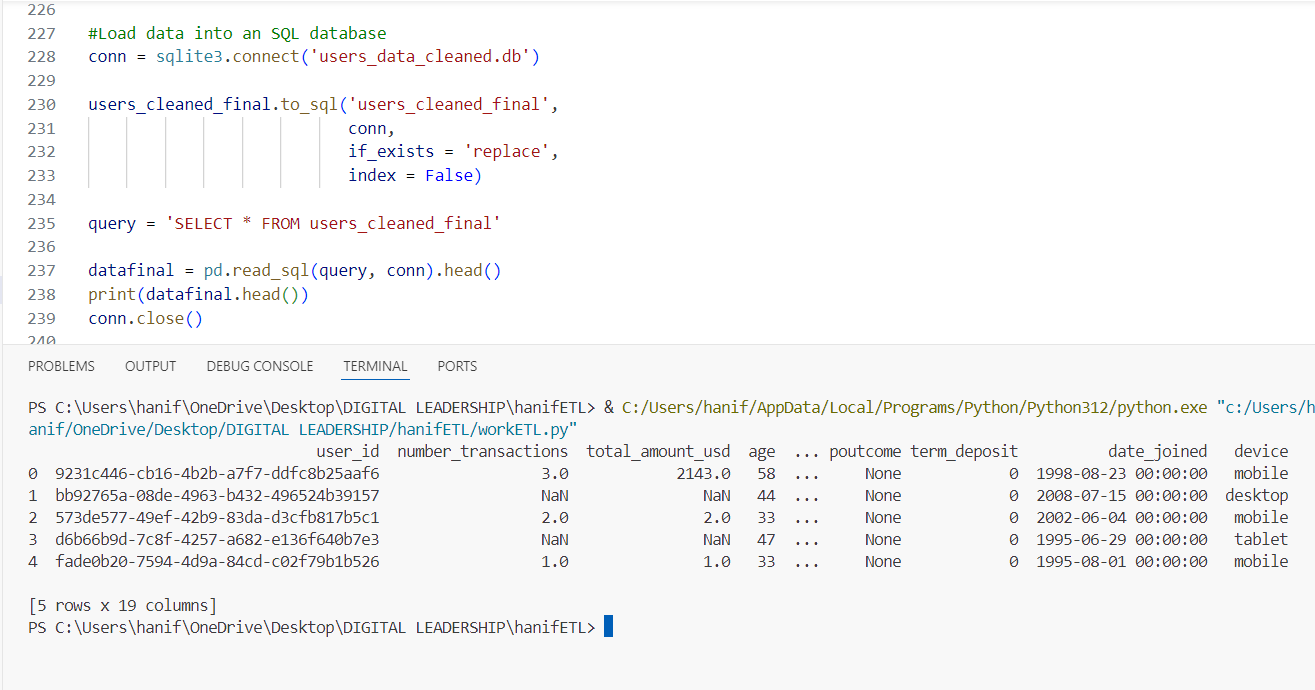


* Check duplicate data
* The total all data or rows is 10

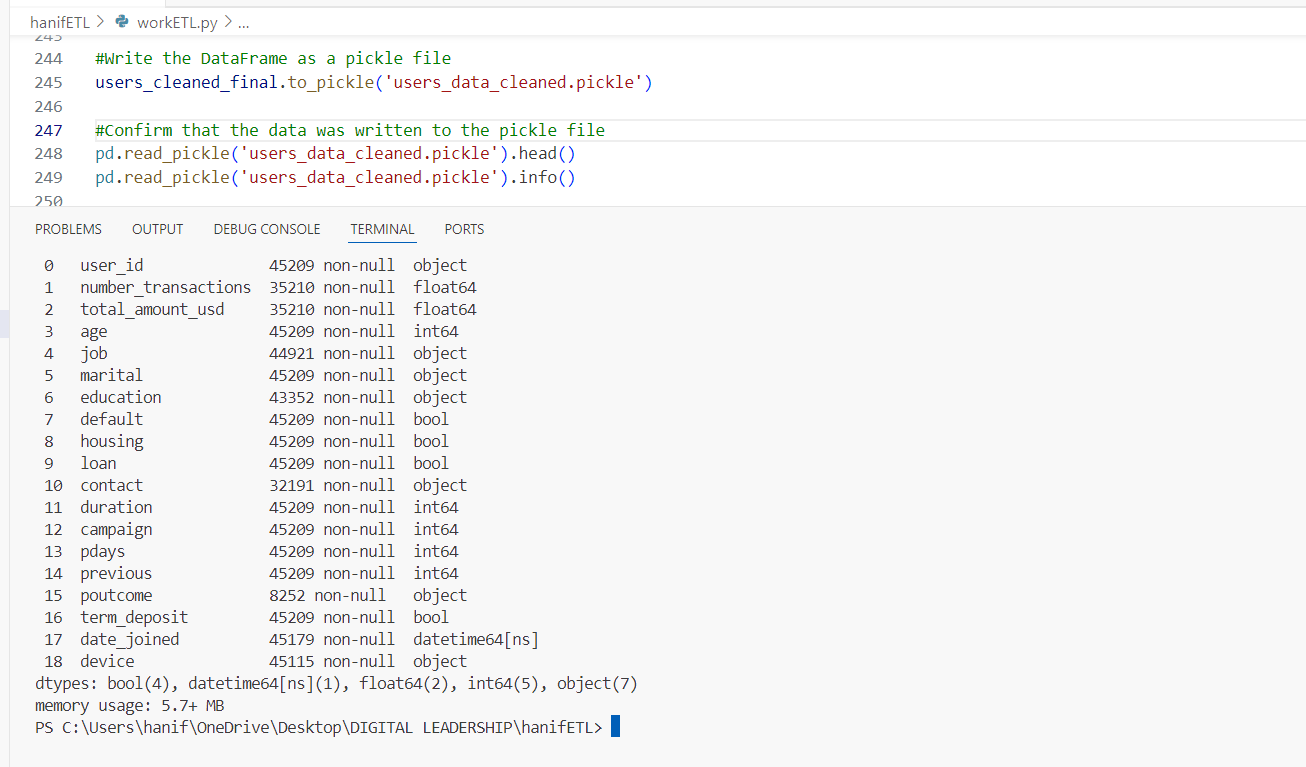


* Remove duplicate data
* Display the data and showing the data based on user id
* Check total data before and after remove duplicate
* Before – 45214, after - 45209

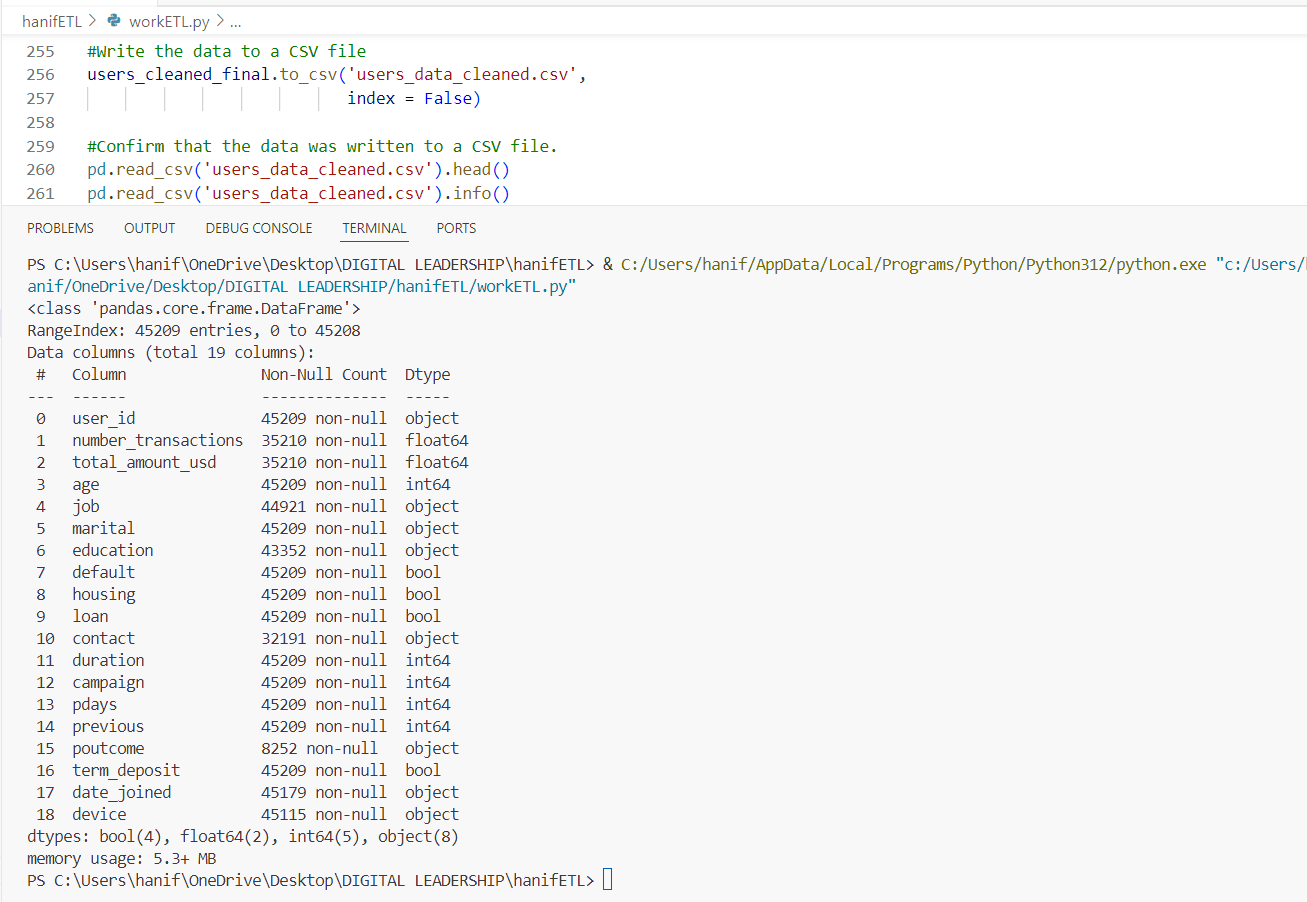
**3)Load Data**



* Create db file for user data that has been cleaned
* Display the data and showing the data by default
* The column or feature is 19



* Create a pickle file
* Saves complex objects like **lists, dictionaries, tuples, or classes**
* Simplifies **data transfer** between systems
* Used in **Machine Learning** to save trained models
* Store the data that has been cleaned



* Create the csv file
* Store the data that has been cleaned