Date: 27-12-23

Spark Assessment

Solution:

Processing CSV and JSON file:

Accessing CSV file using pyspark with the help of creating a spark session and viewing the data through. show() of keyword

A computer screen shot of a computer screen

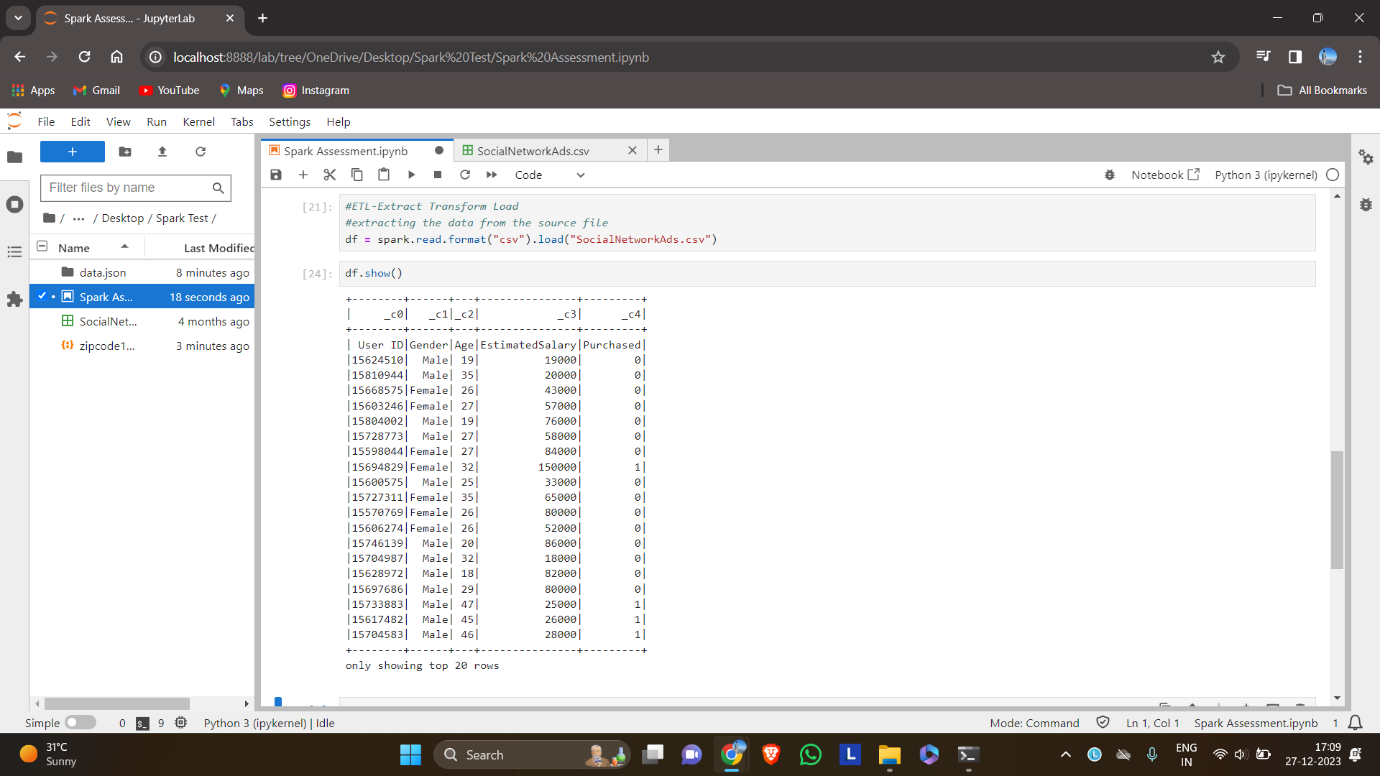
Description automatically generated

Accessing CSV file using pyspark with the help of creating a spark session and viewing the data through. show() of keywordA screenshot of a computer

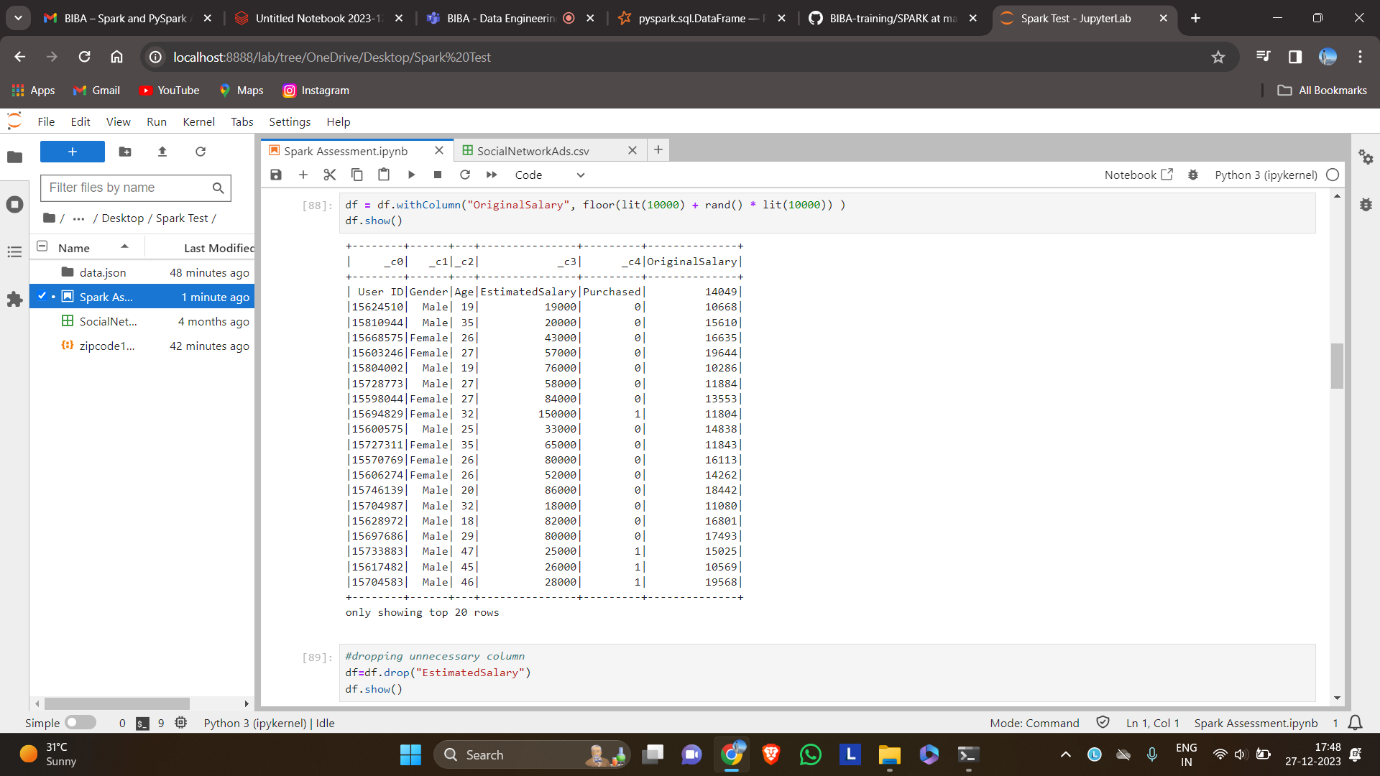
Description automatically generated

**ETL:**

**Extract:** Retrieve data from various sources like databases, files, or APIs.



Transform: Clean, aggregate, and manipulate data to fit your analysis needs.



A screenshot of a computer

Description automatically generated

Load: Store the transformed data into a database or data warehouse for analysis

A screenshot of a computer

Description automatically generated

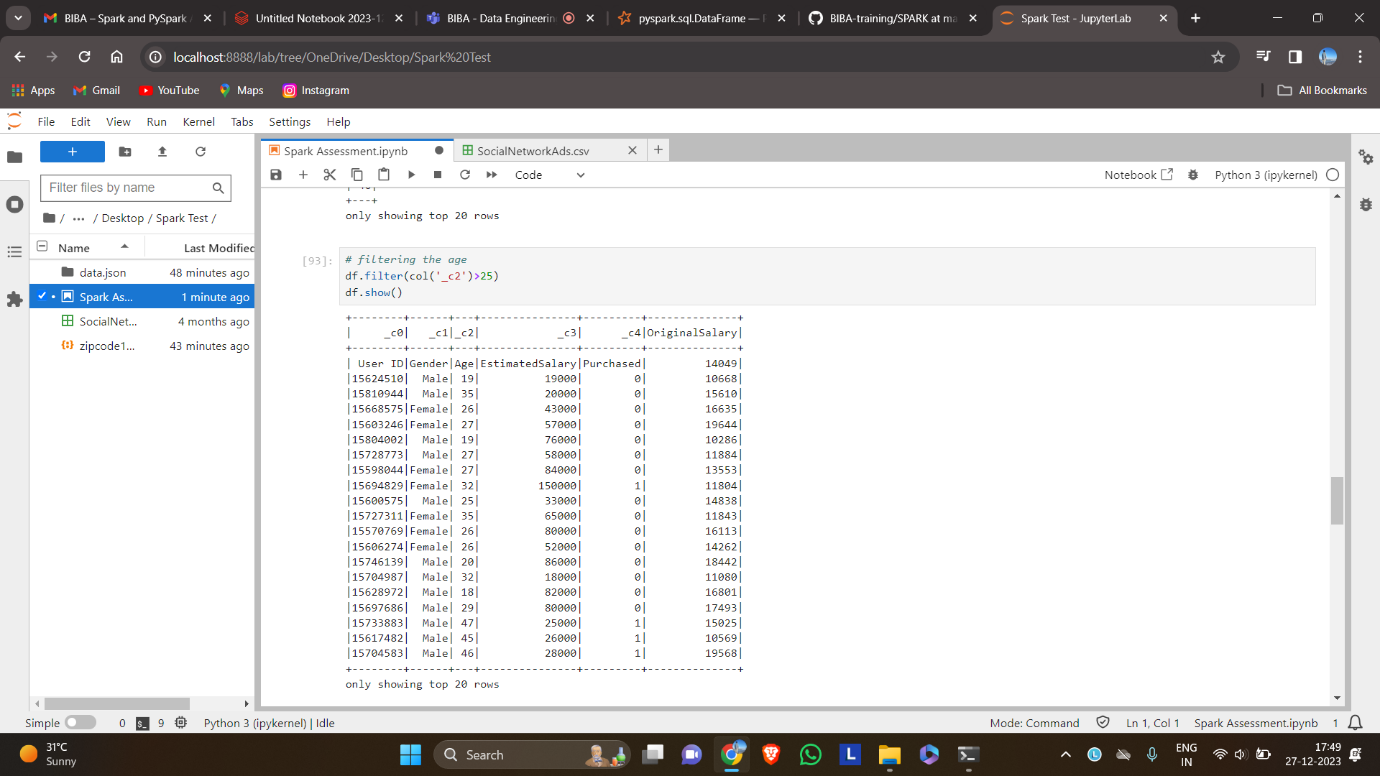
Transformation using filter, count, group by etc

Viewing a single column:

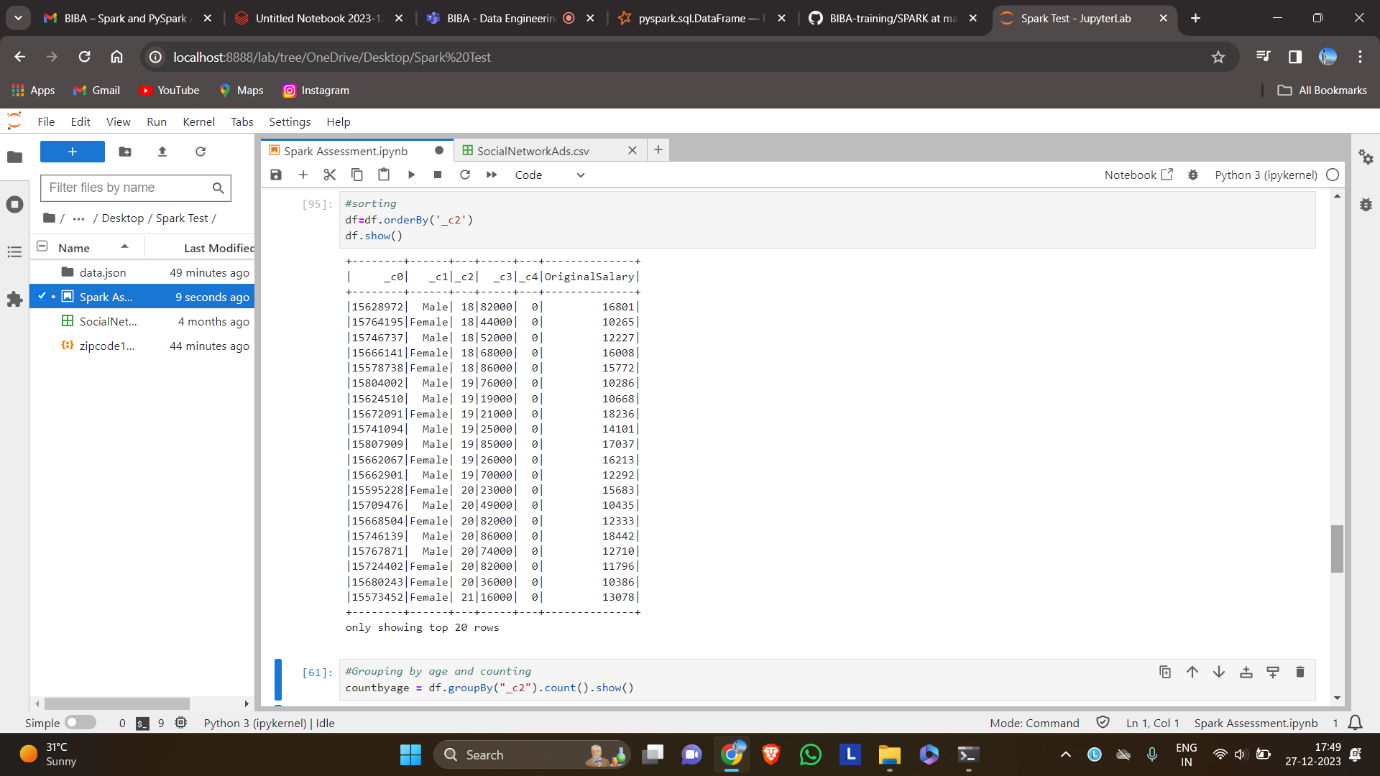
A screenshot of a computer

Description automatically generated

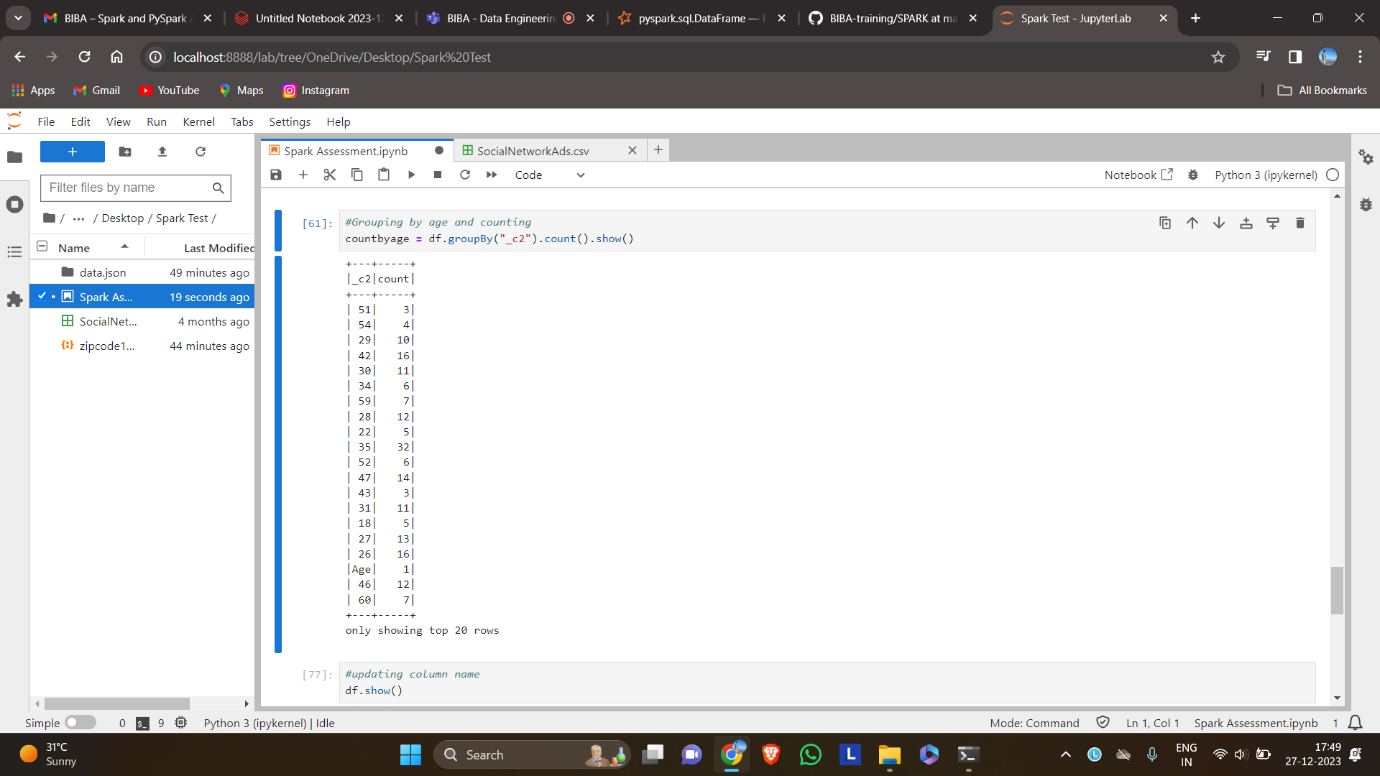
Using filtering function to remove people age<25



Using Sorting function:



Using Groupby and count for finding number of people in each age



Creating a data frame:

