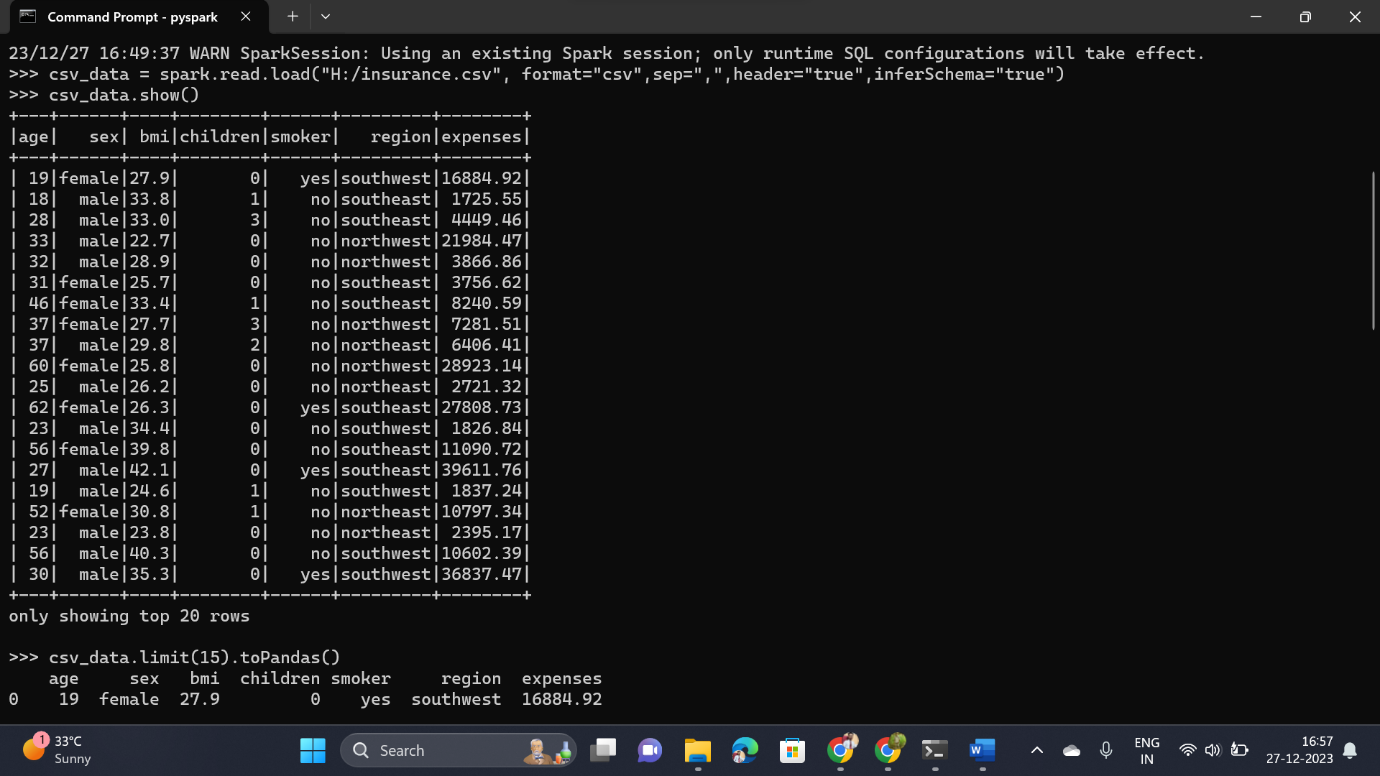
Spark Assessment [27-12-2023]

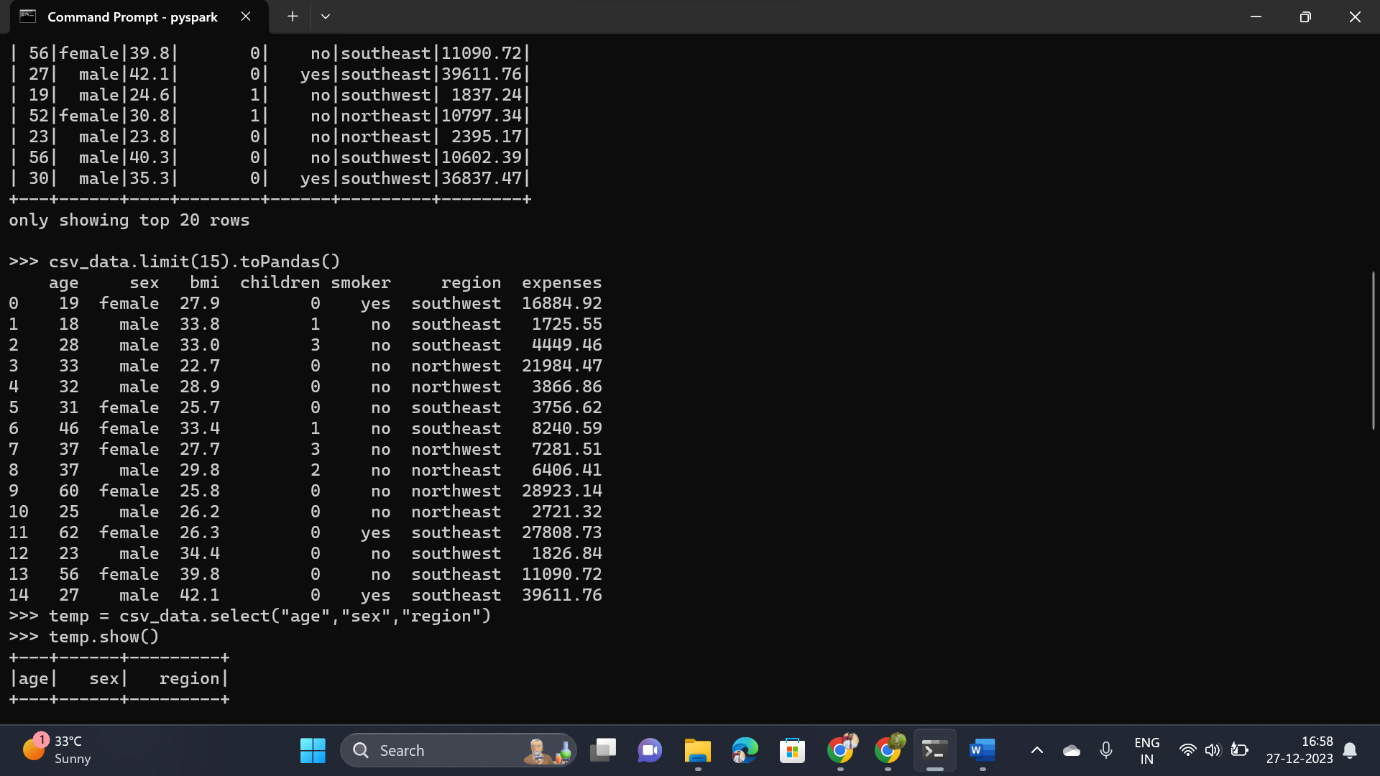
By Karthikeyan K R

**Processing the CSV File data:**

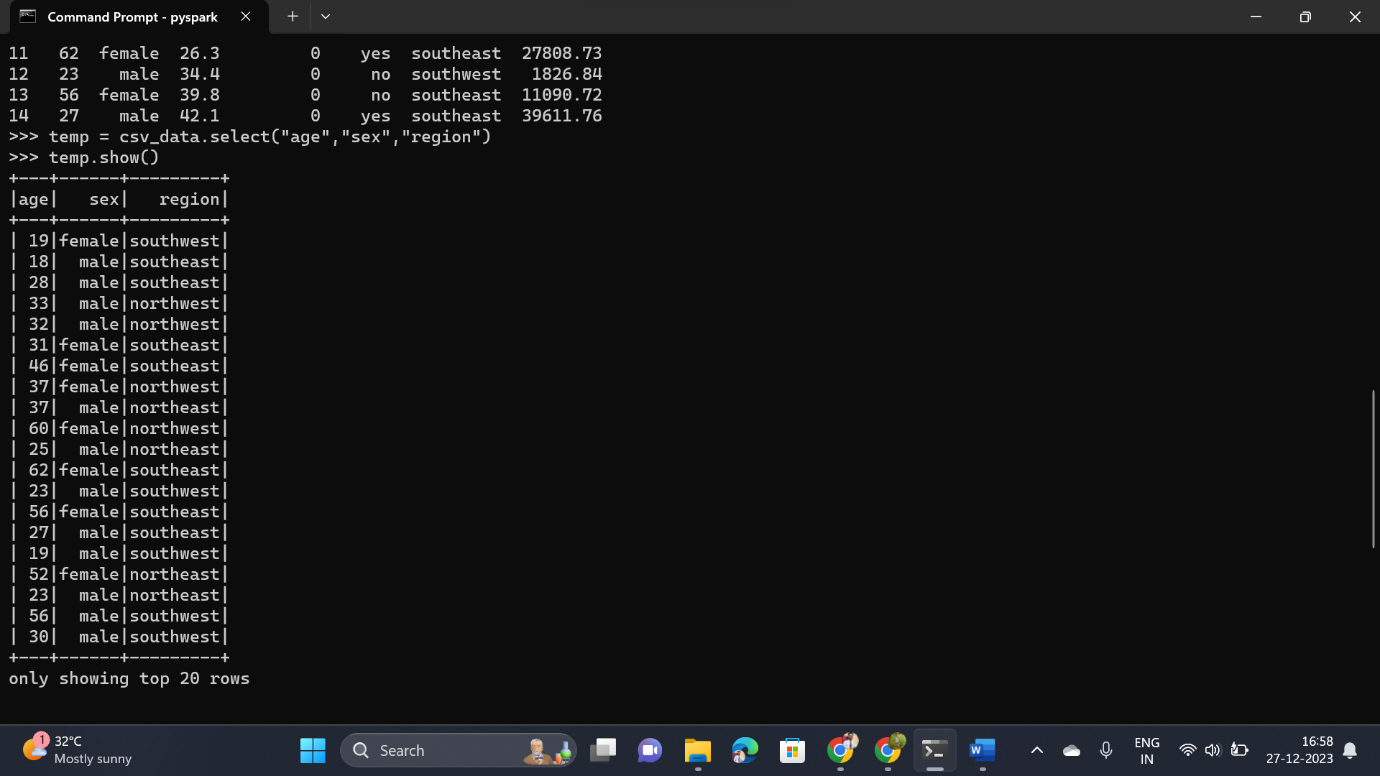
* Reading and displaying the data from csv file



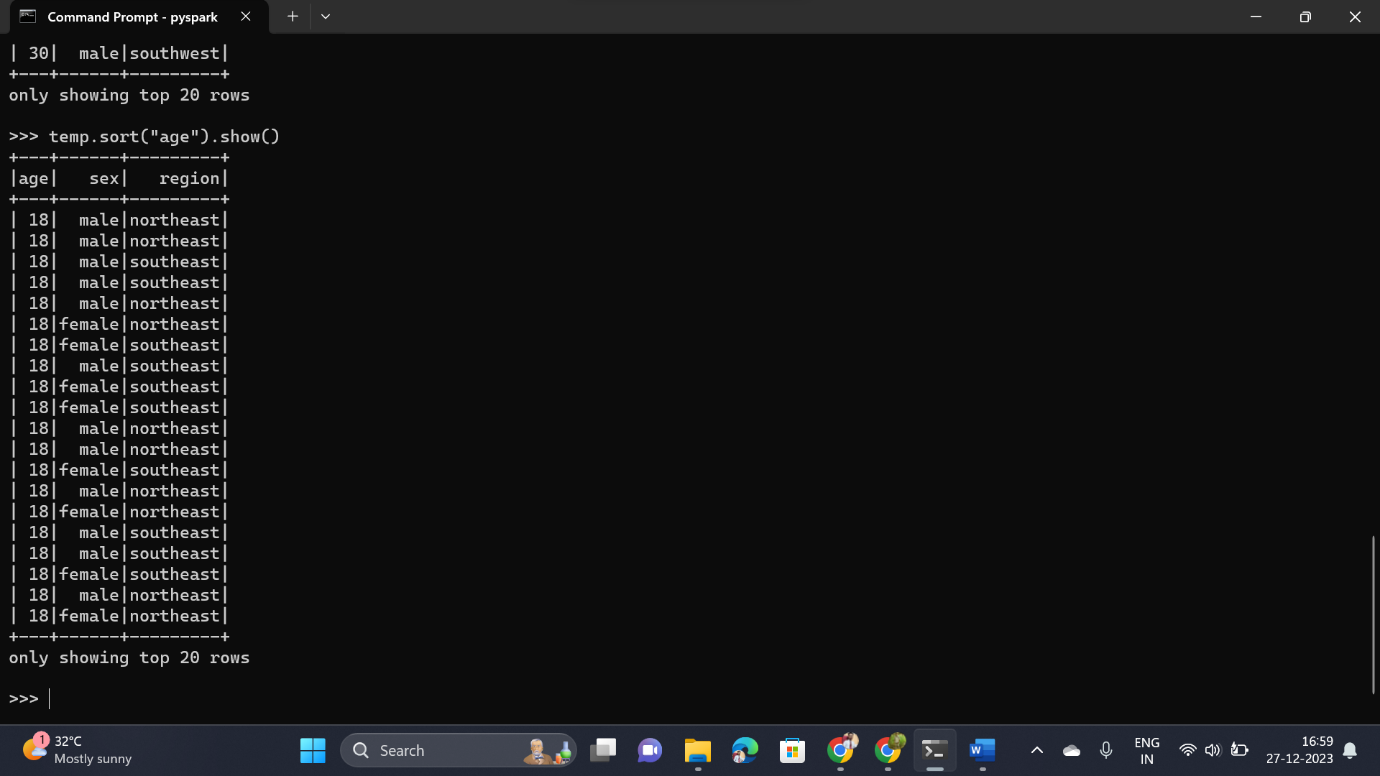
* Using toPandas()



* Using Select() Command



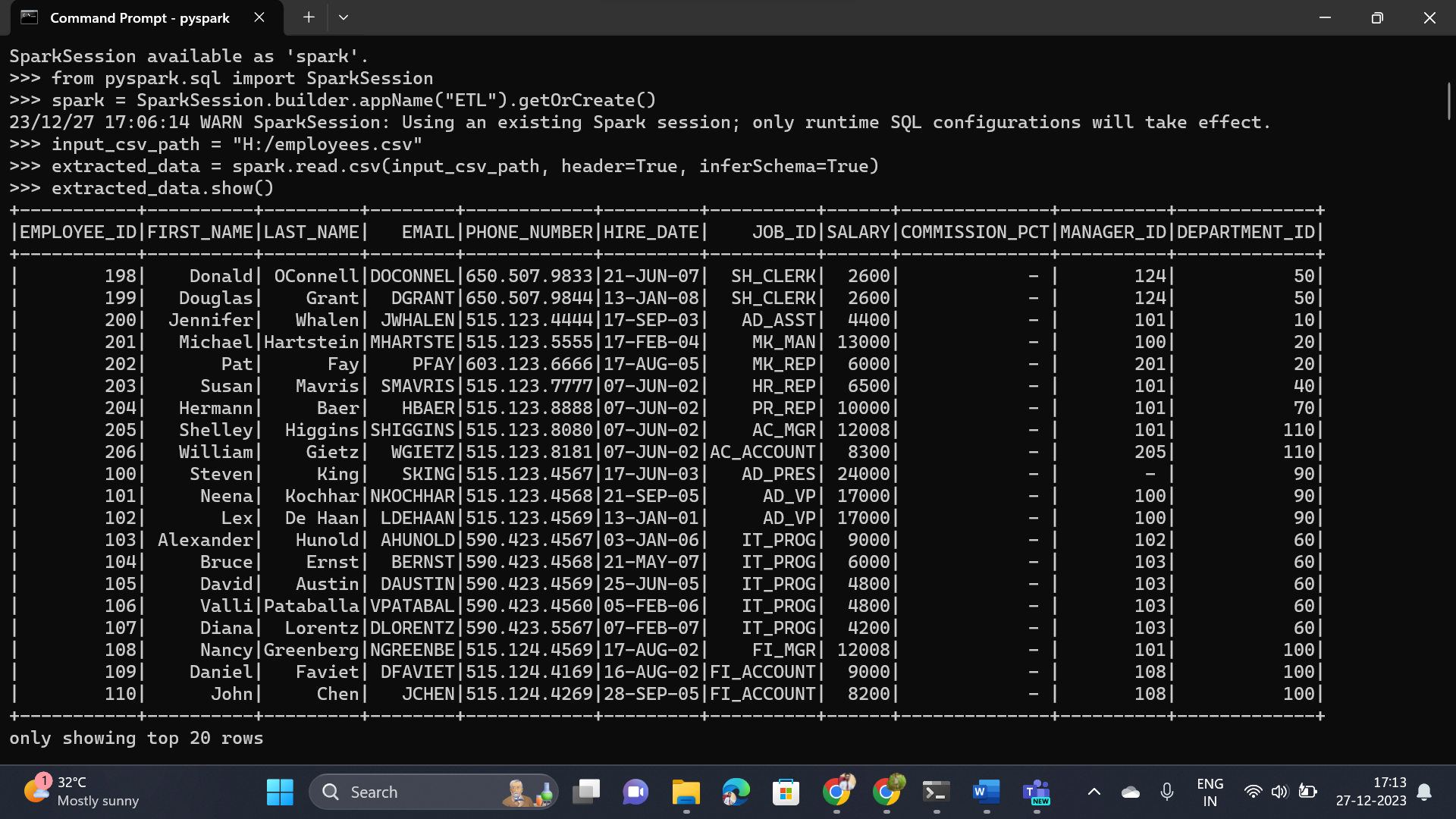
* Using Sort() command to sort the values



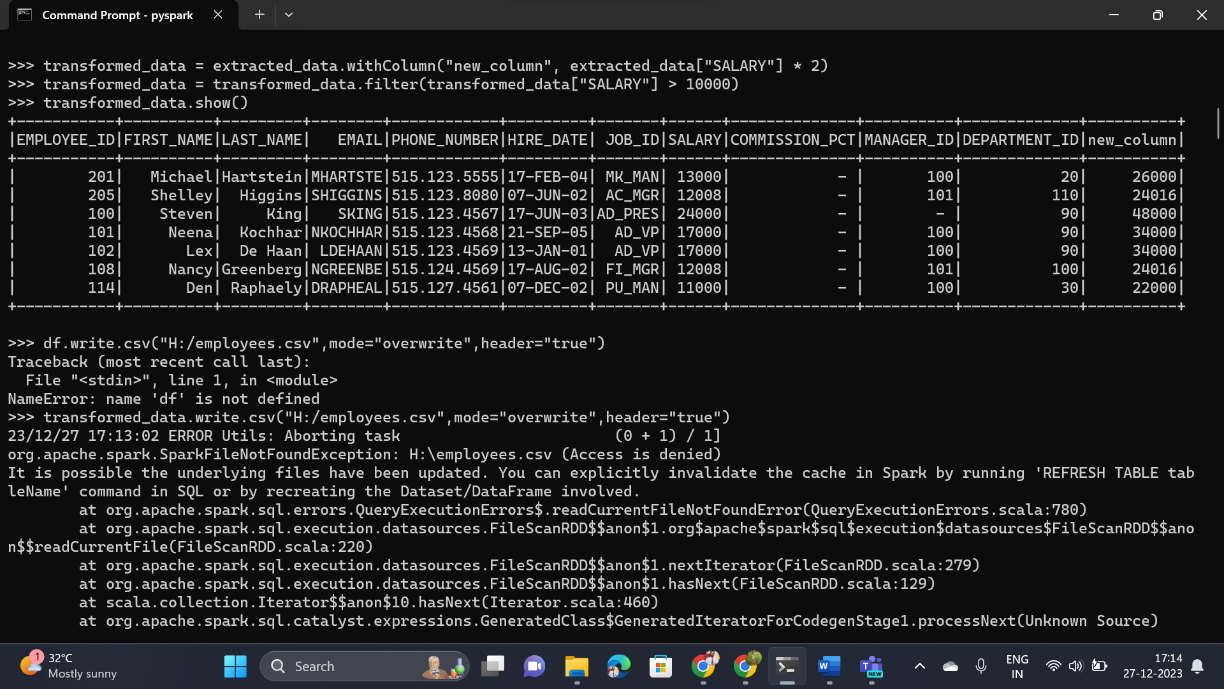
**ETL Process**

ETL, which stands for Extract, Transform, Load, is a process used in data integration and data warehousing to collect data from various sources, transform it into a suitable format, and then load it into a target database or data warehouse for analysis and reporting.

* Extracting the data

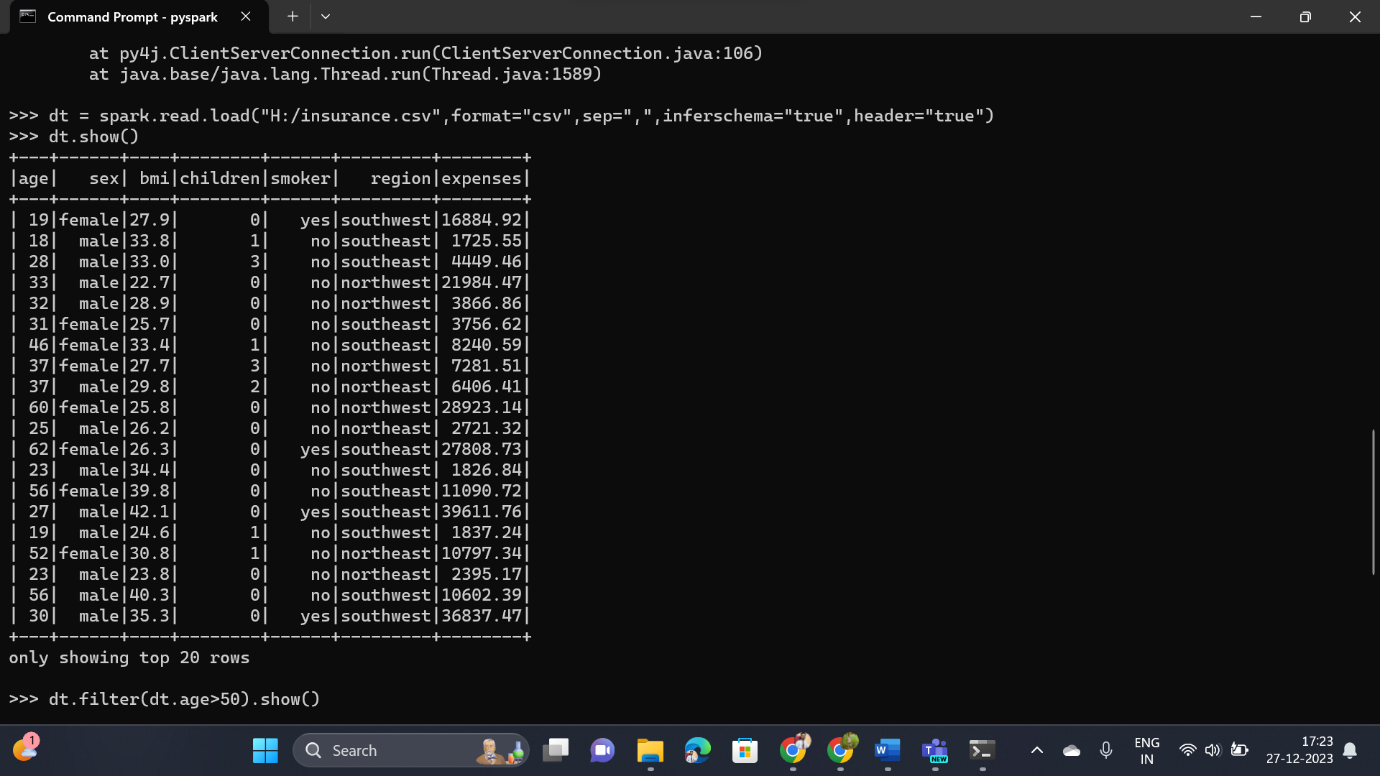


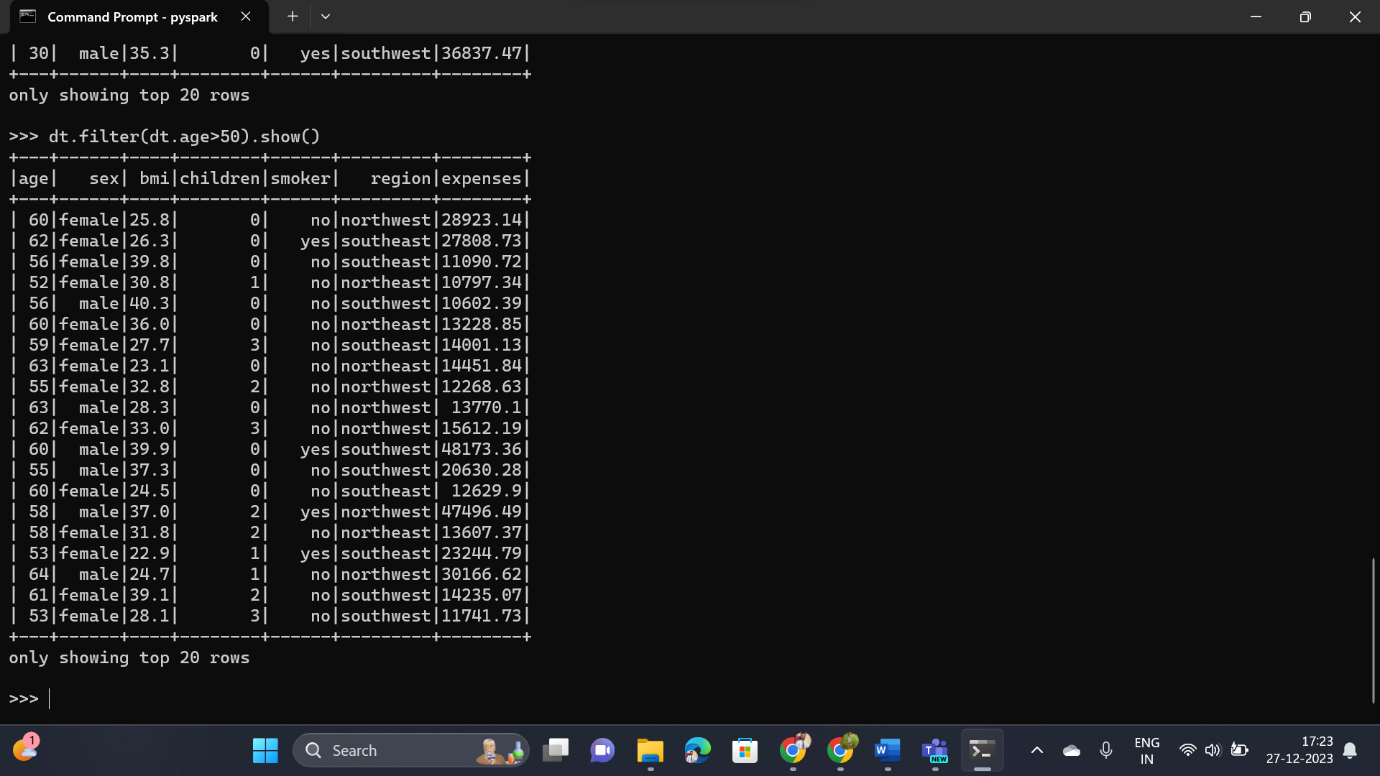
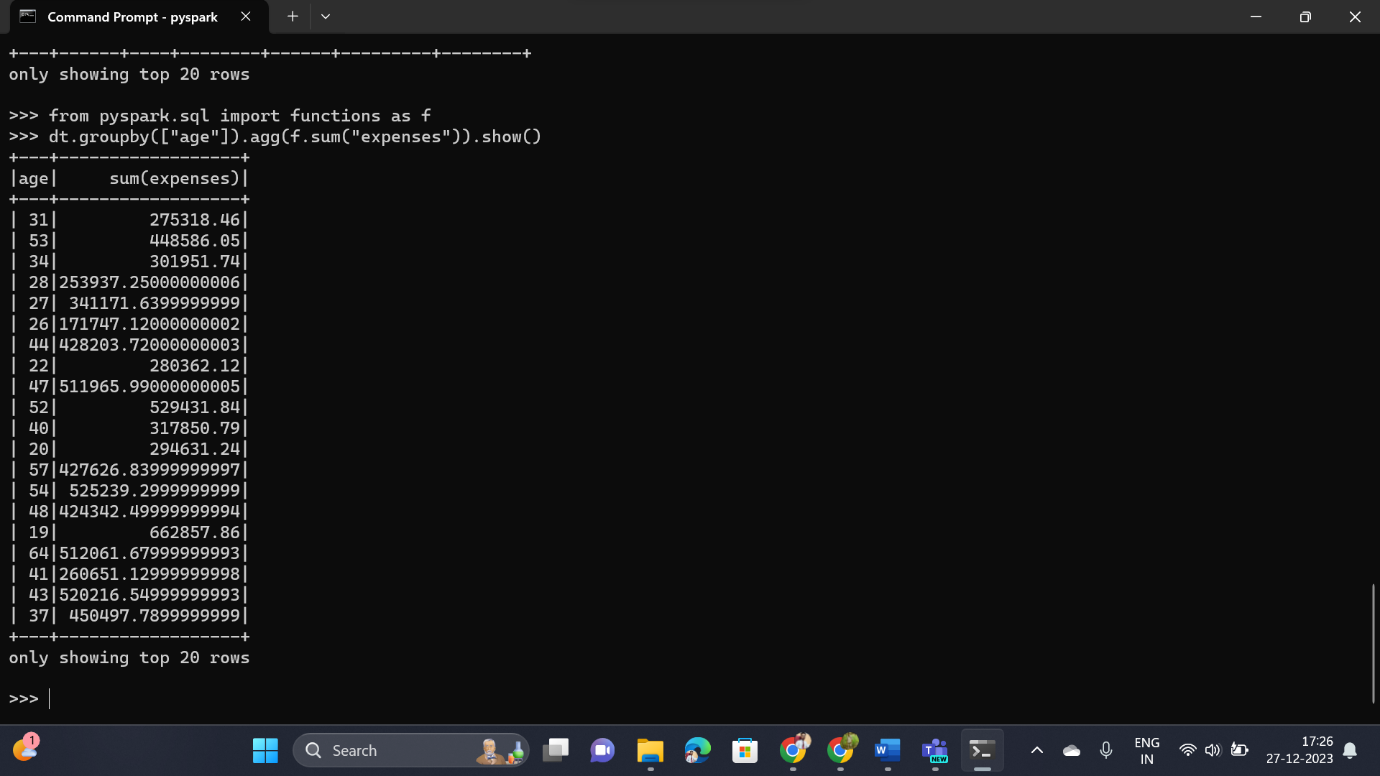
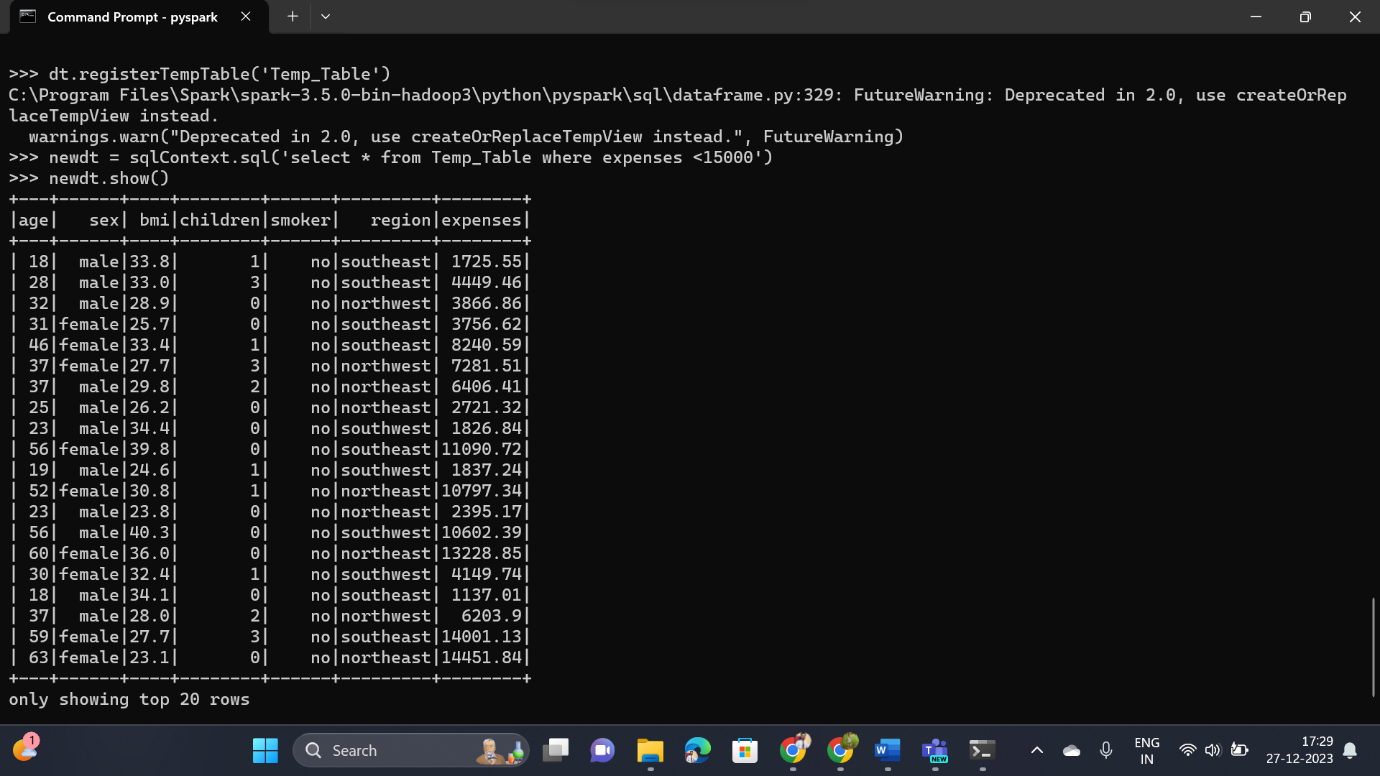
* Transforming the data



**Executing** **Filter, Simple Aggregations, and GroupBy**

* The Original data before processing



* Using Filter() Command
* Using GroupBy with Aggregation Function
* Writing SQL Query in PySpark