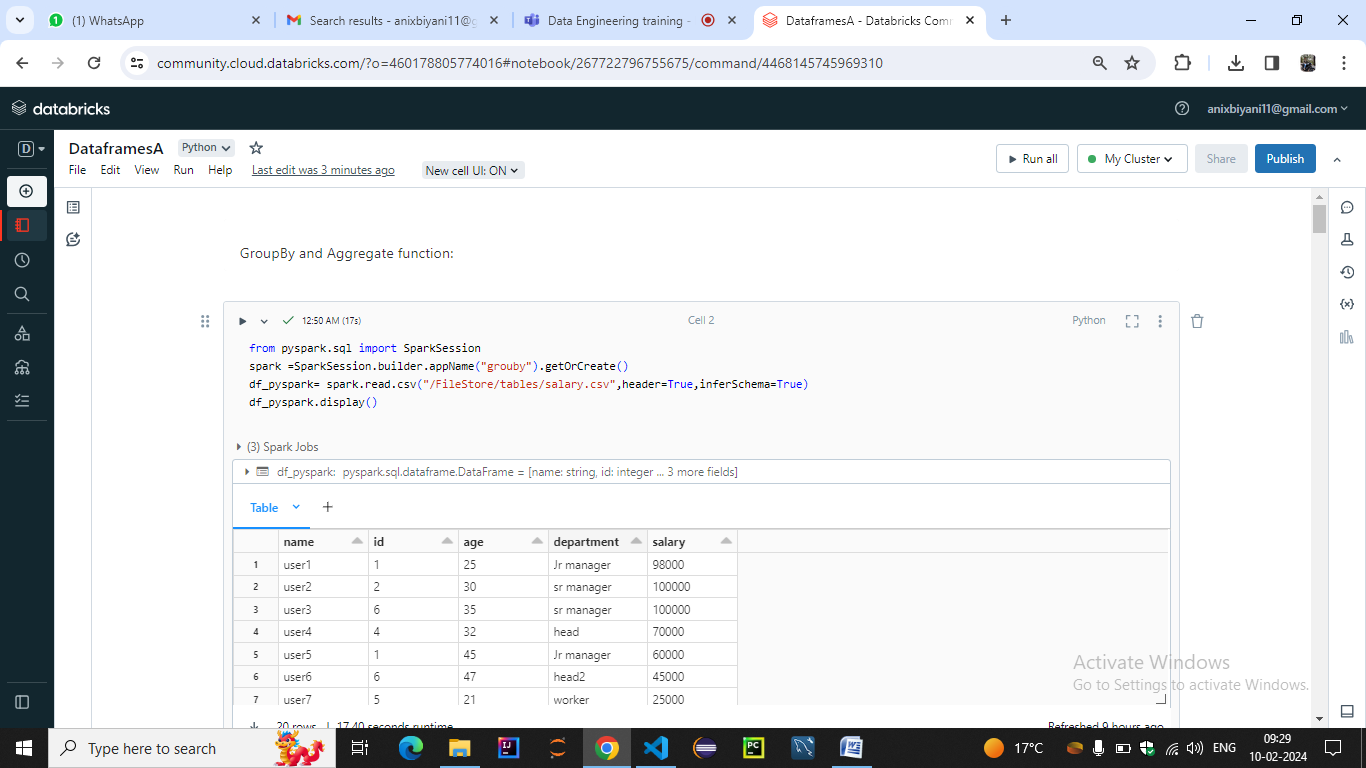
**NAME :- ANIKET SANJAYKUMAR BIYANI**

**DATA ENGINEERING BATCH -1**

**PYTHON DAY-14 ASSIGNMENT**

**GroupBy() and Aggregate Function:-**

* **In PySpark, the groupBy operation is used to group data based on one or more columns and perform aggregate operations on each group. \**

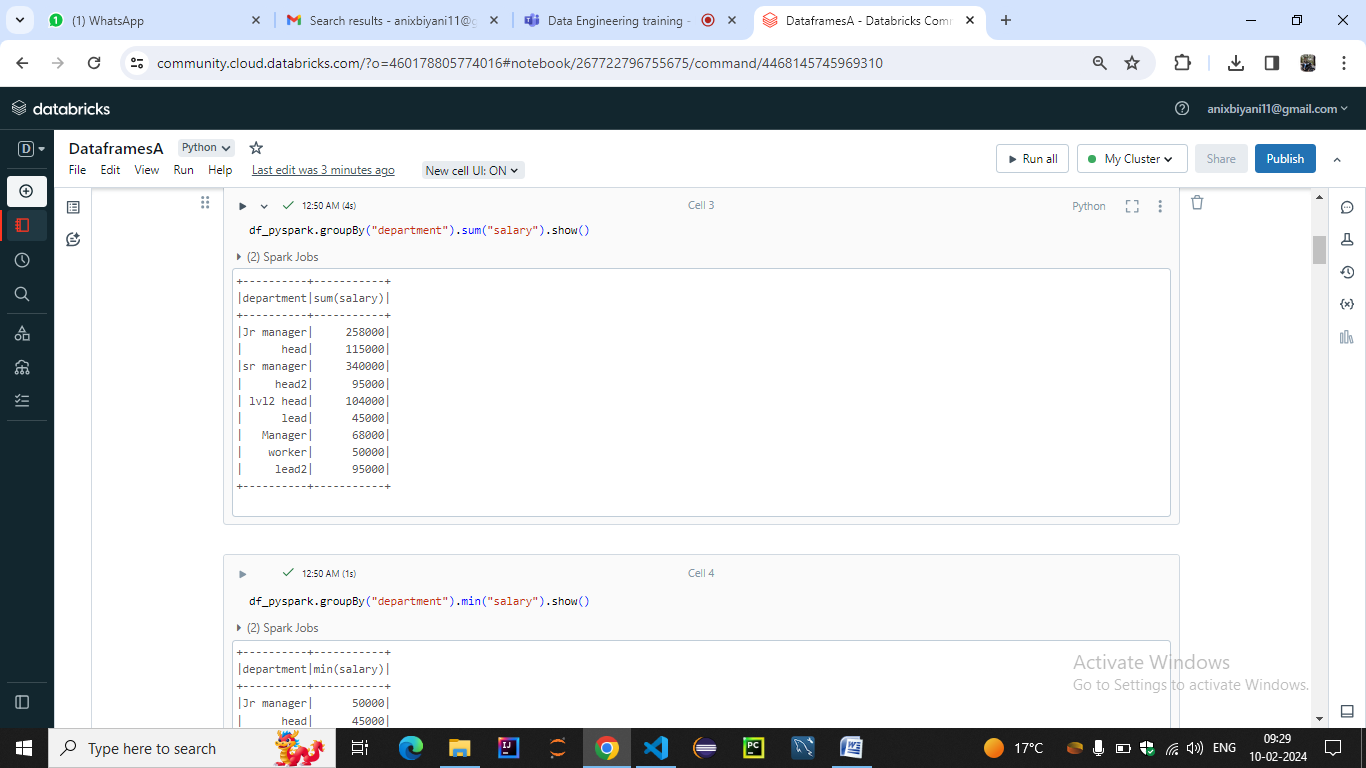
****

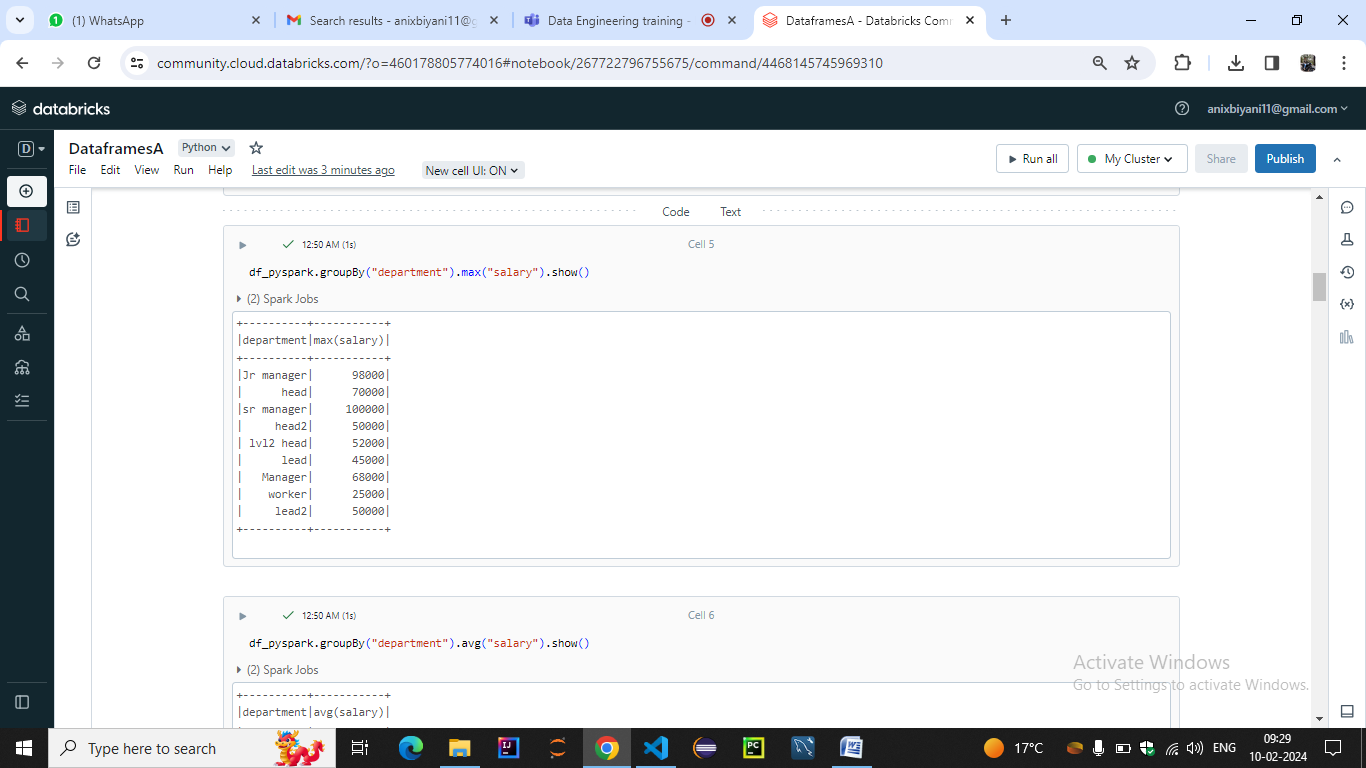
**SUM**

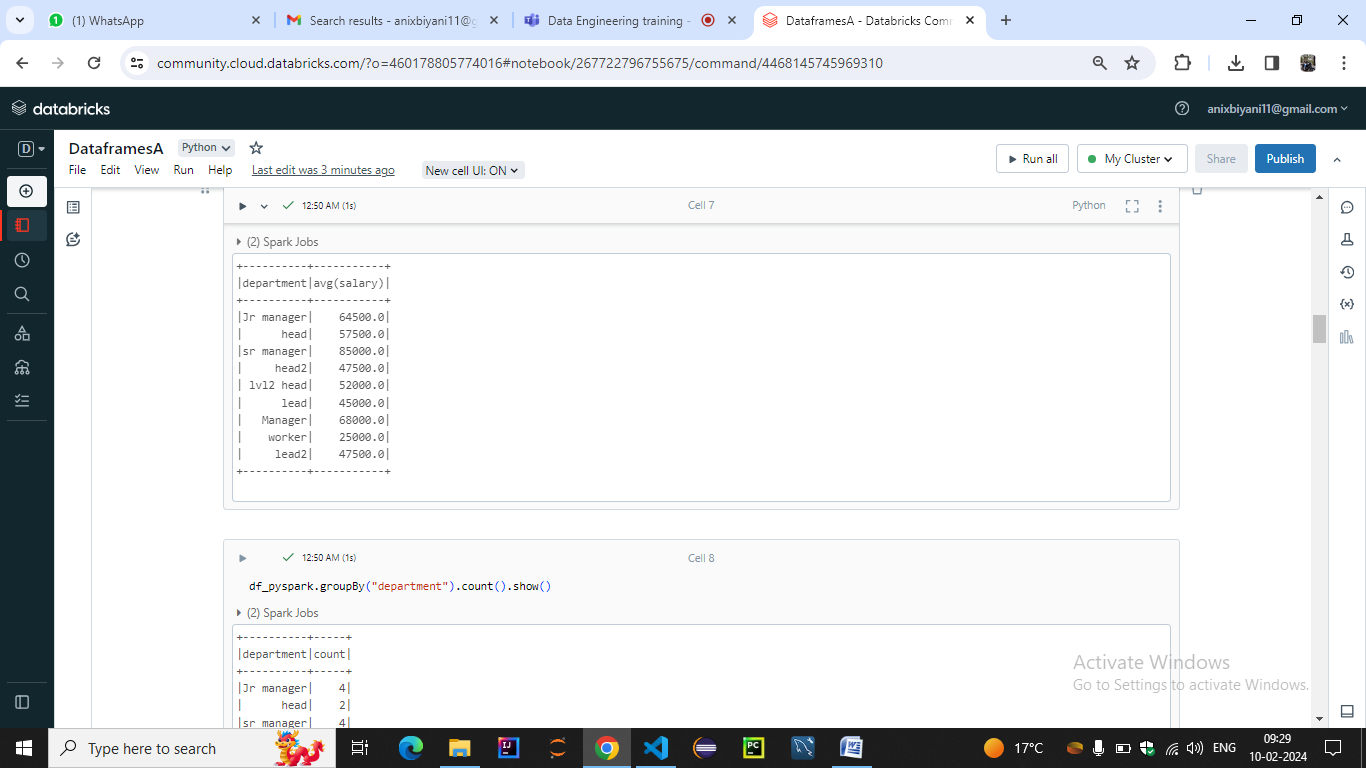
**MIN**

**MAX**

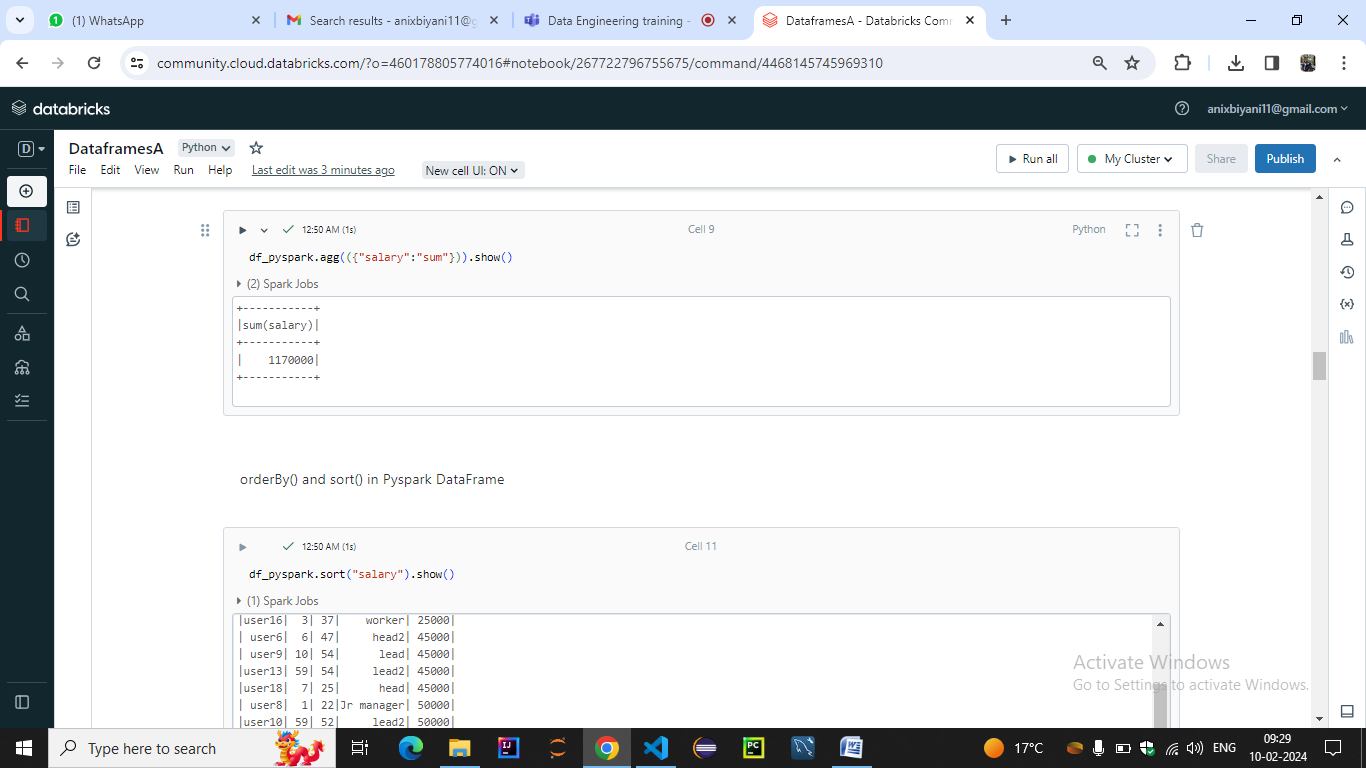
**COUNT AGG FUNCTIONS WITH GROUP BY**

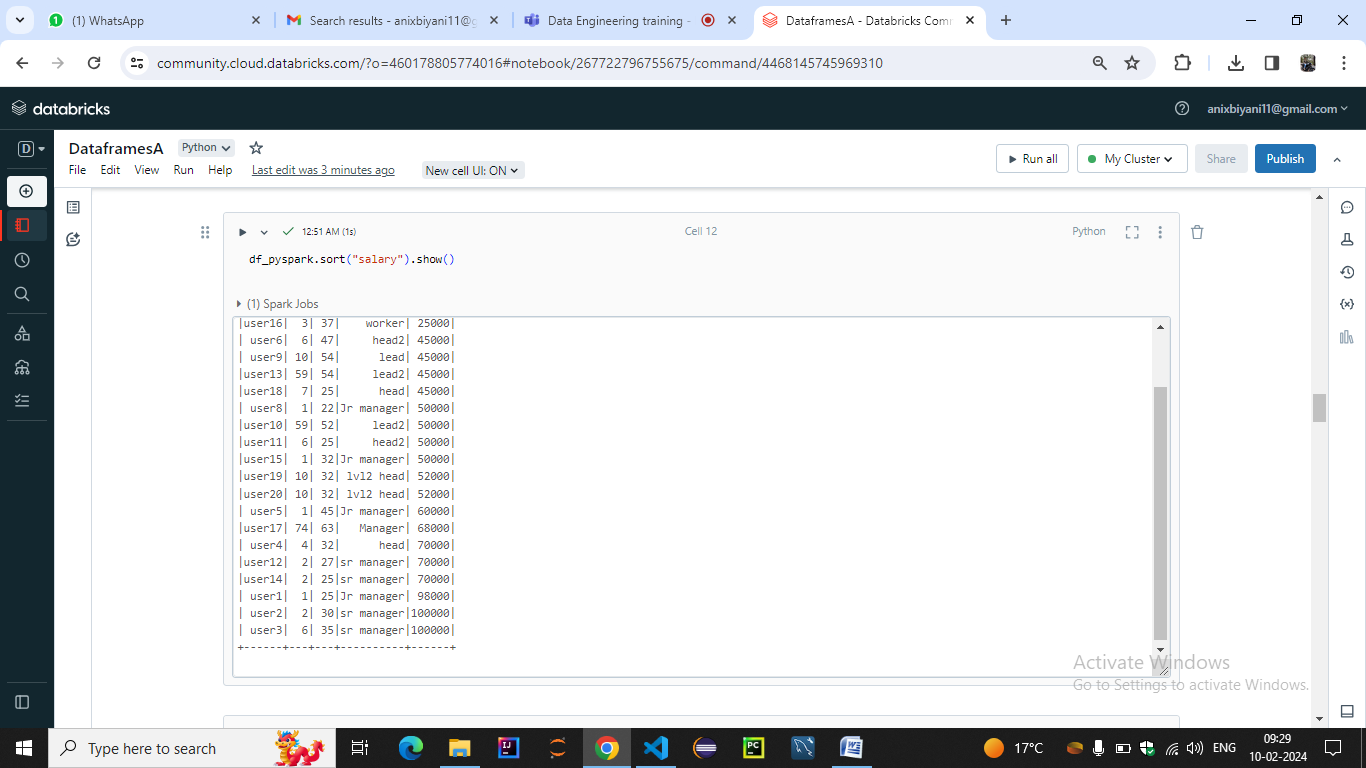
****

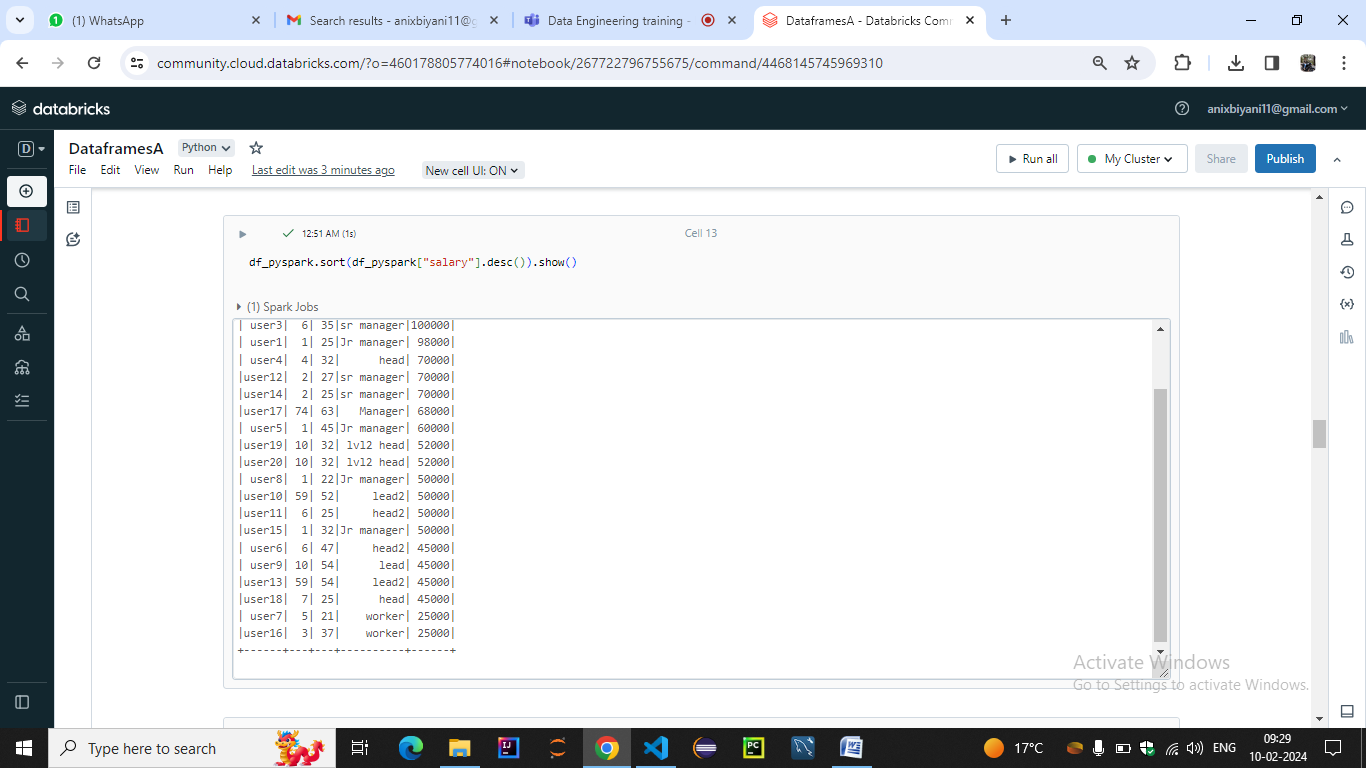


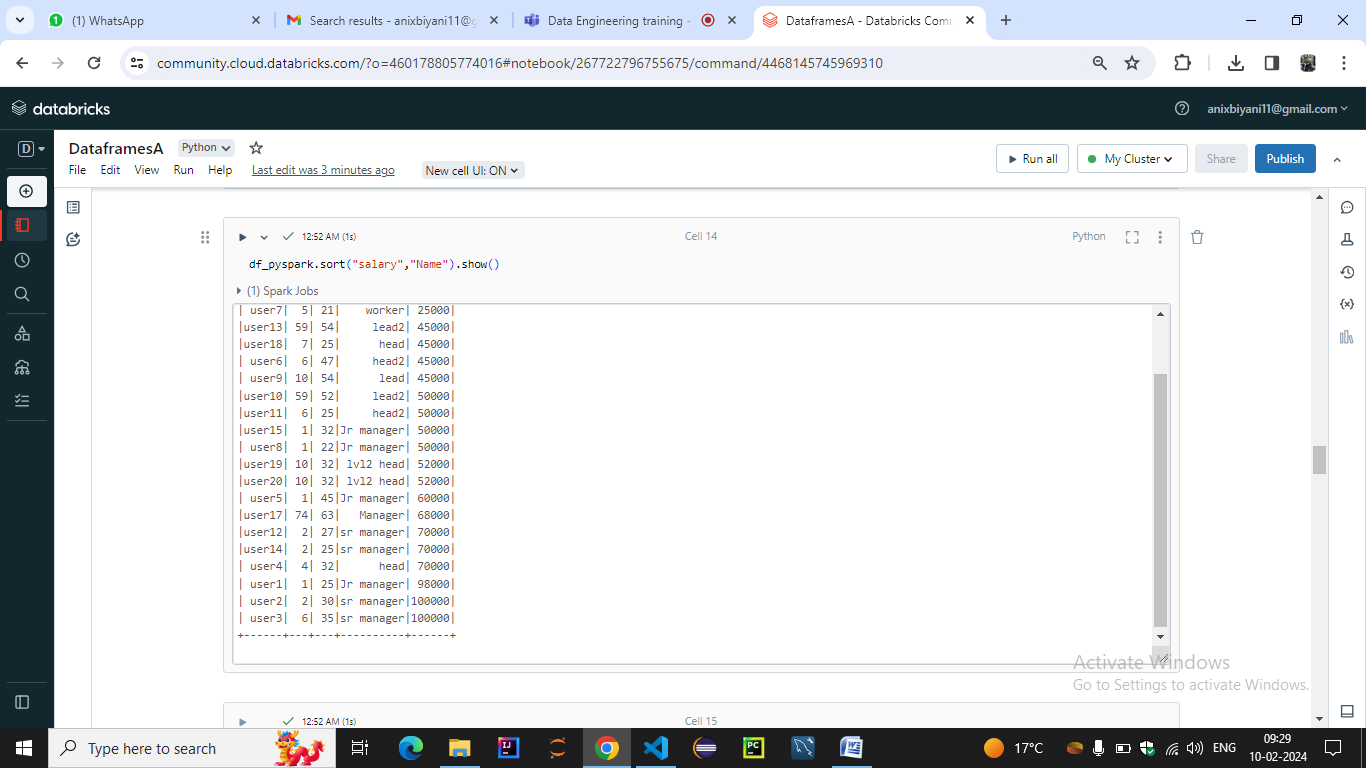


ORDER BY SORT

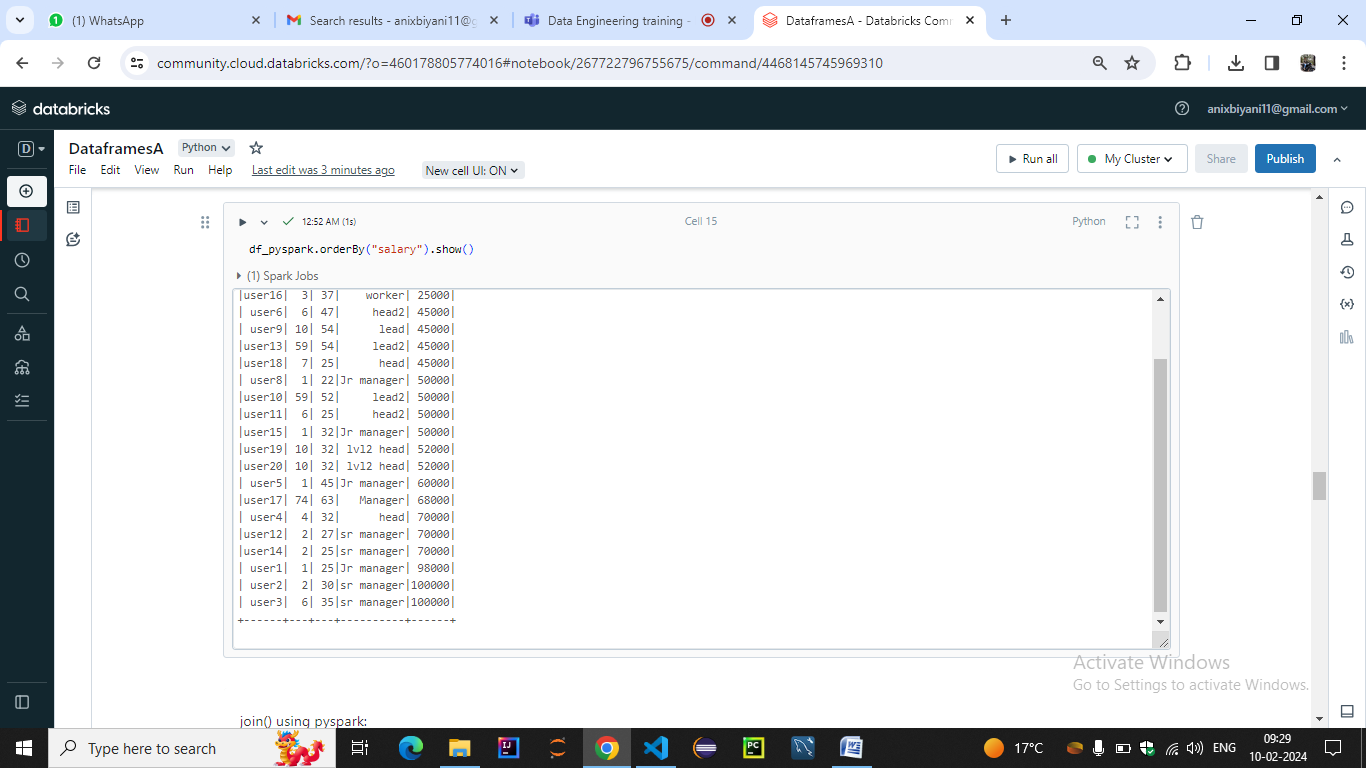




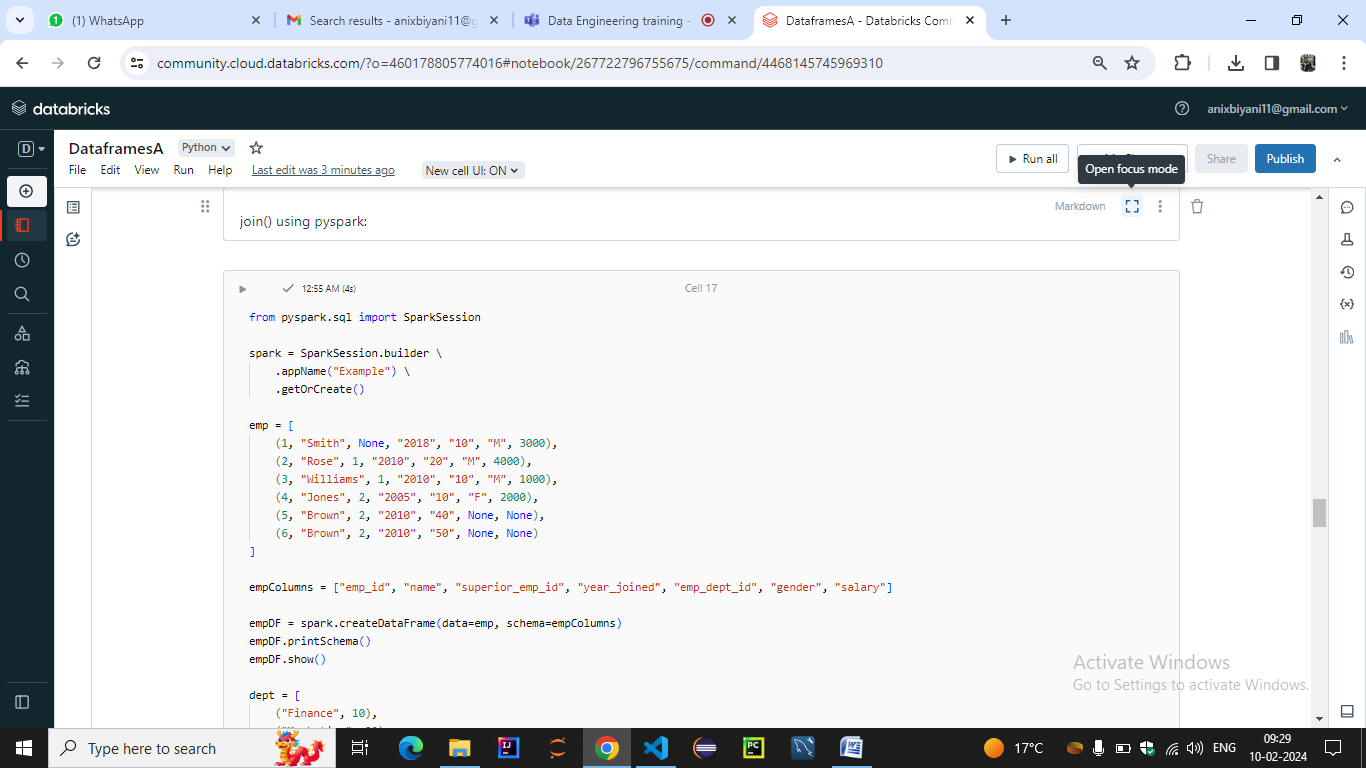


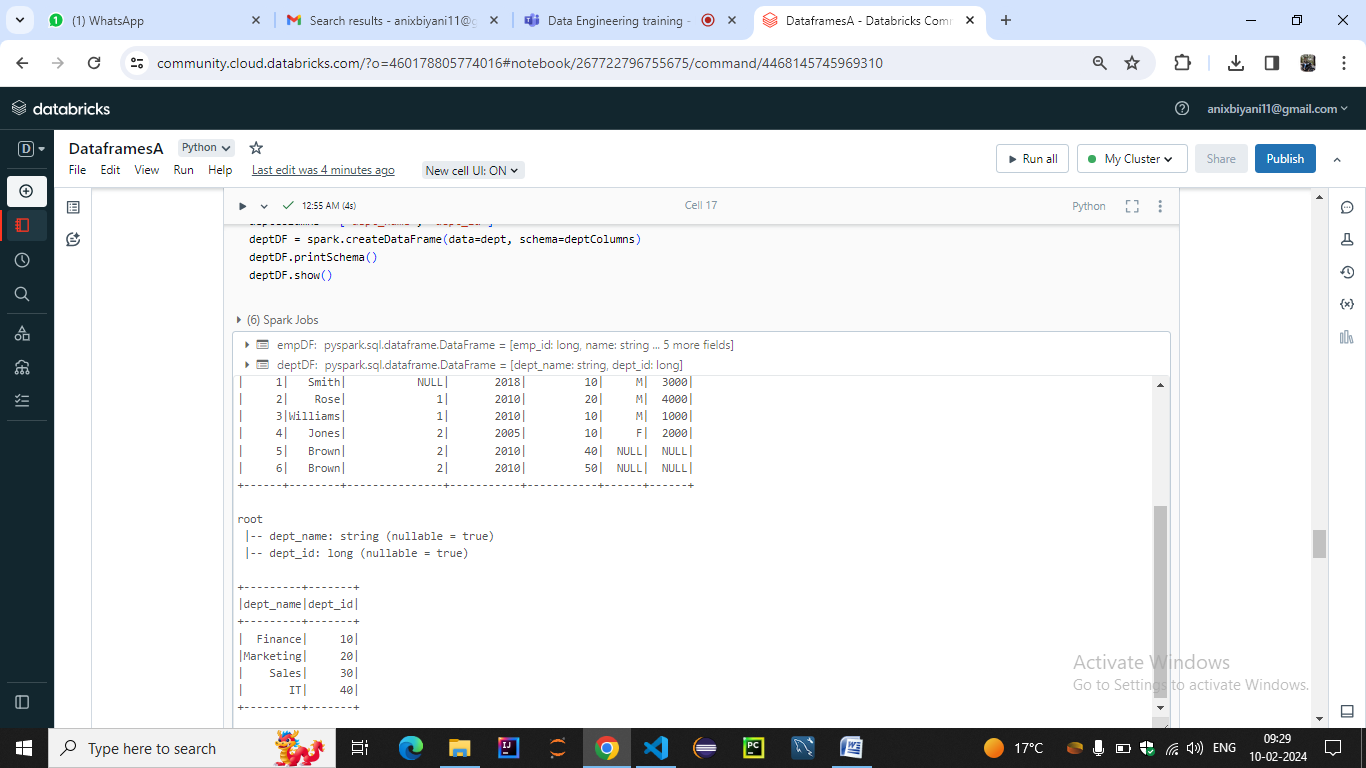


ORDER BY

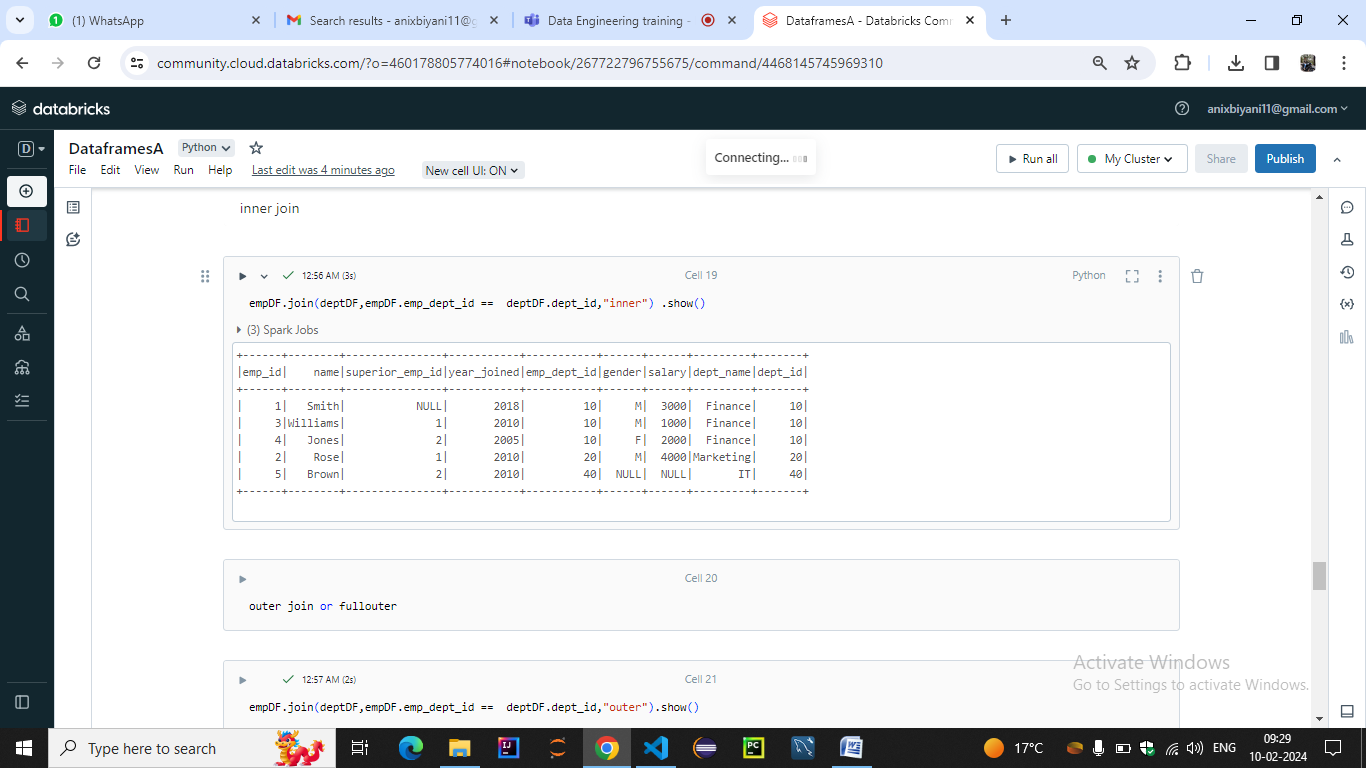


JOINS :

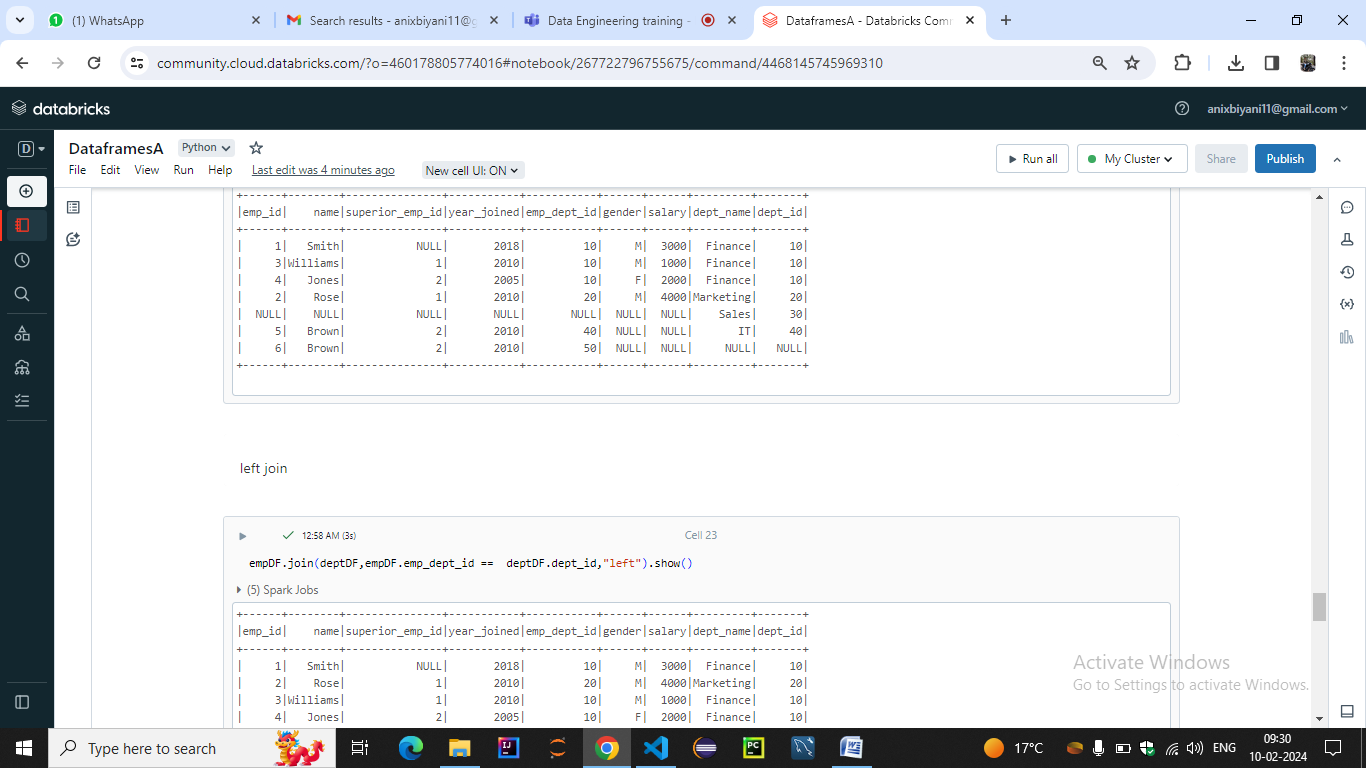


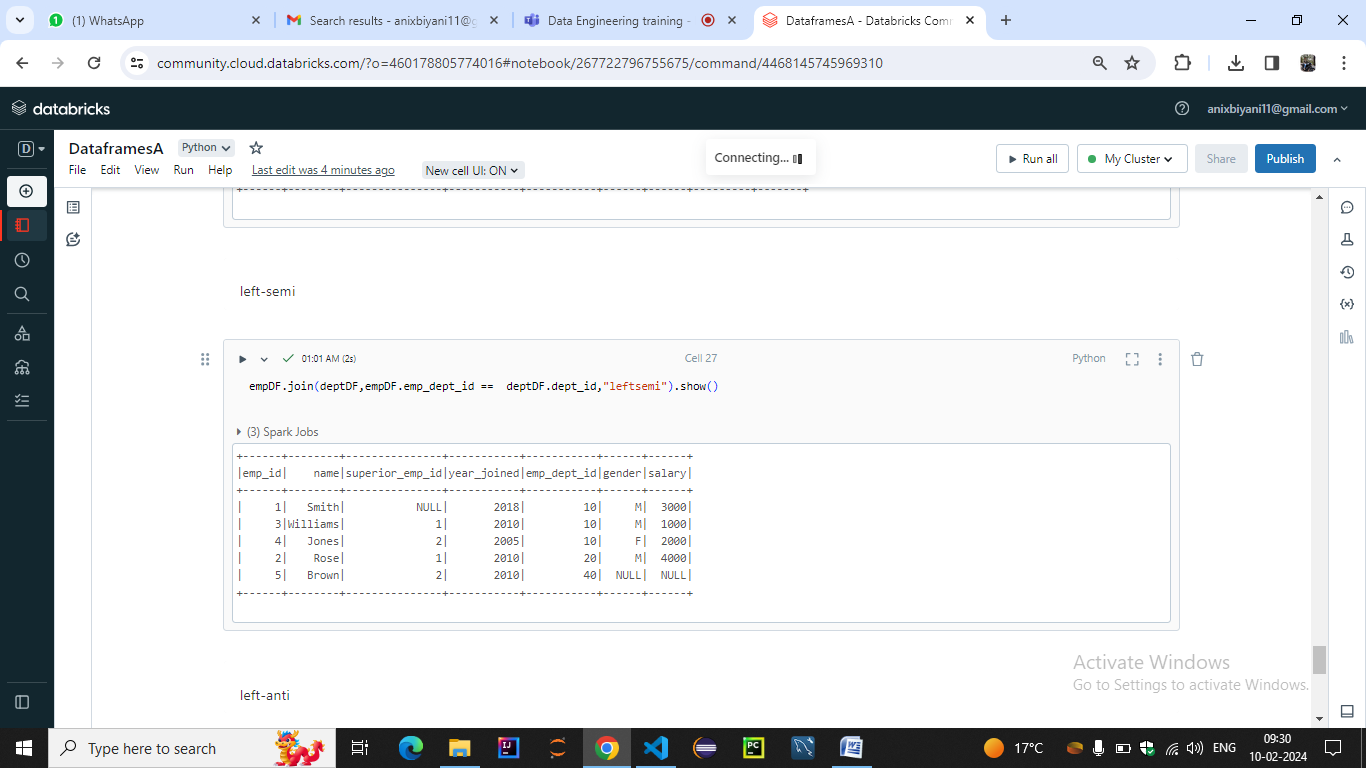


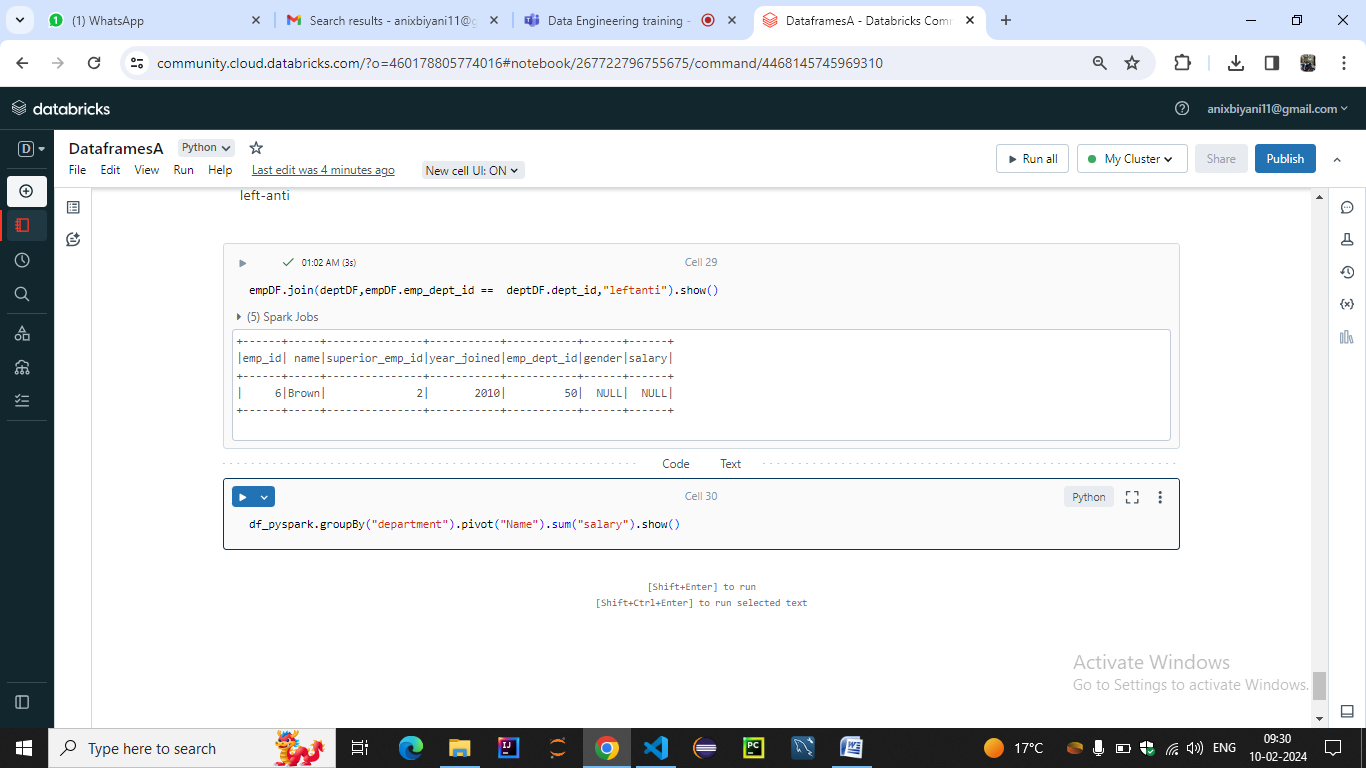
INNER JOIN



OUTER JOIN

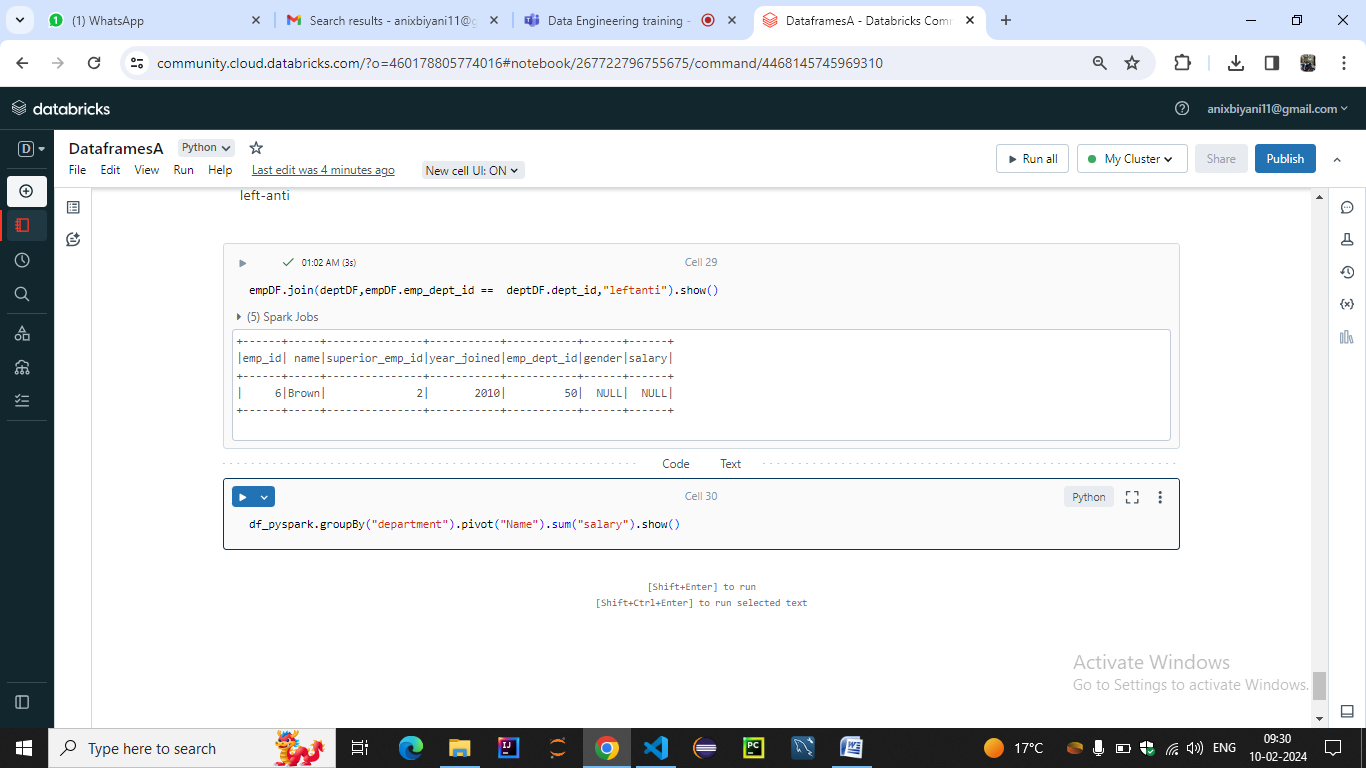
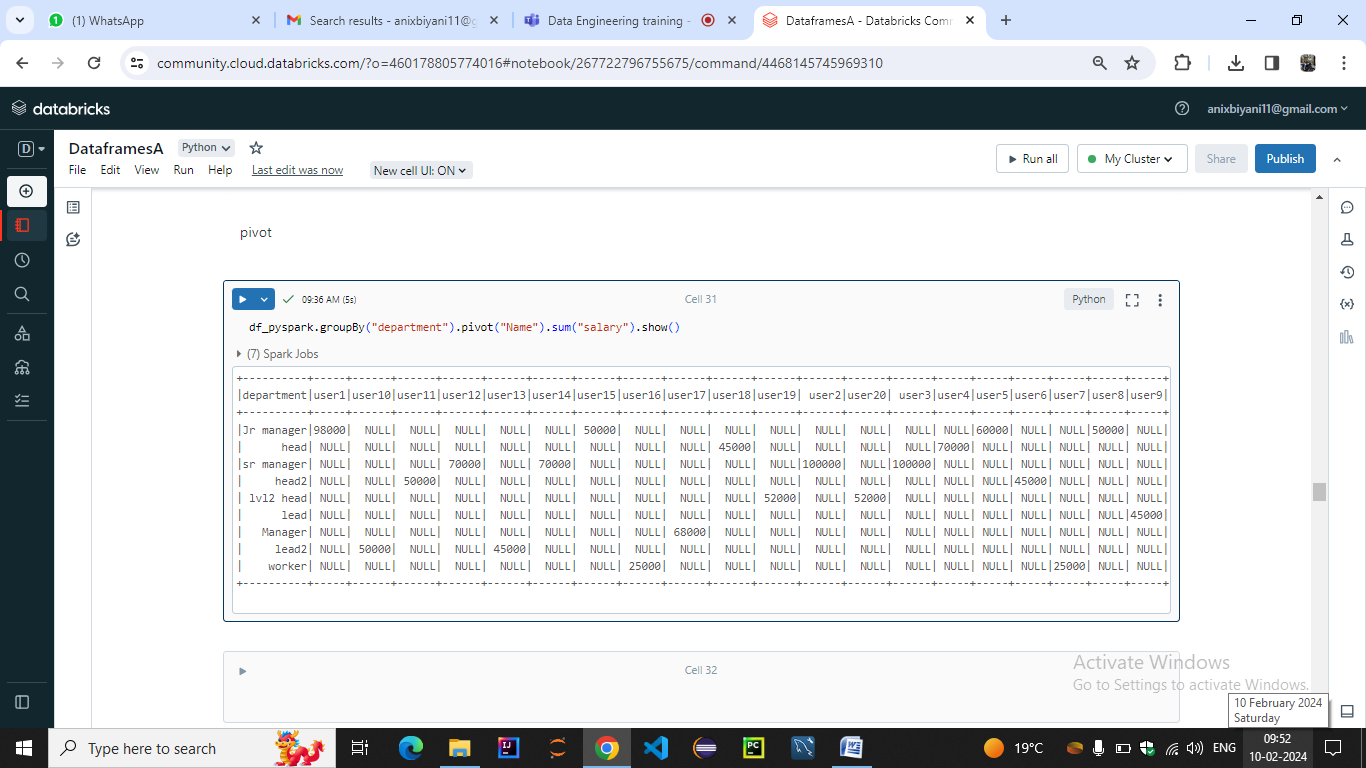






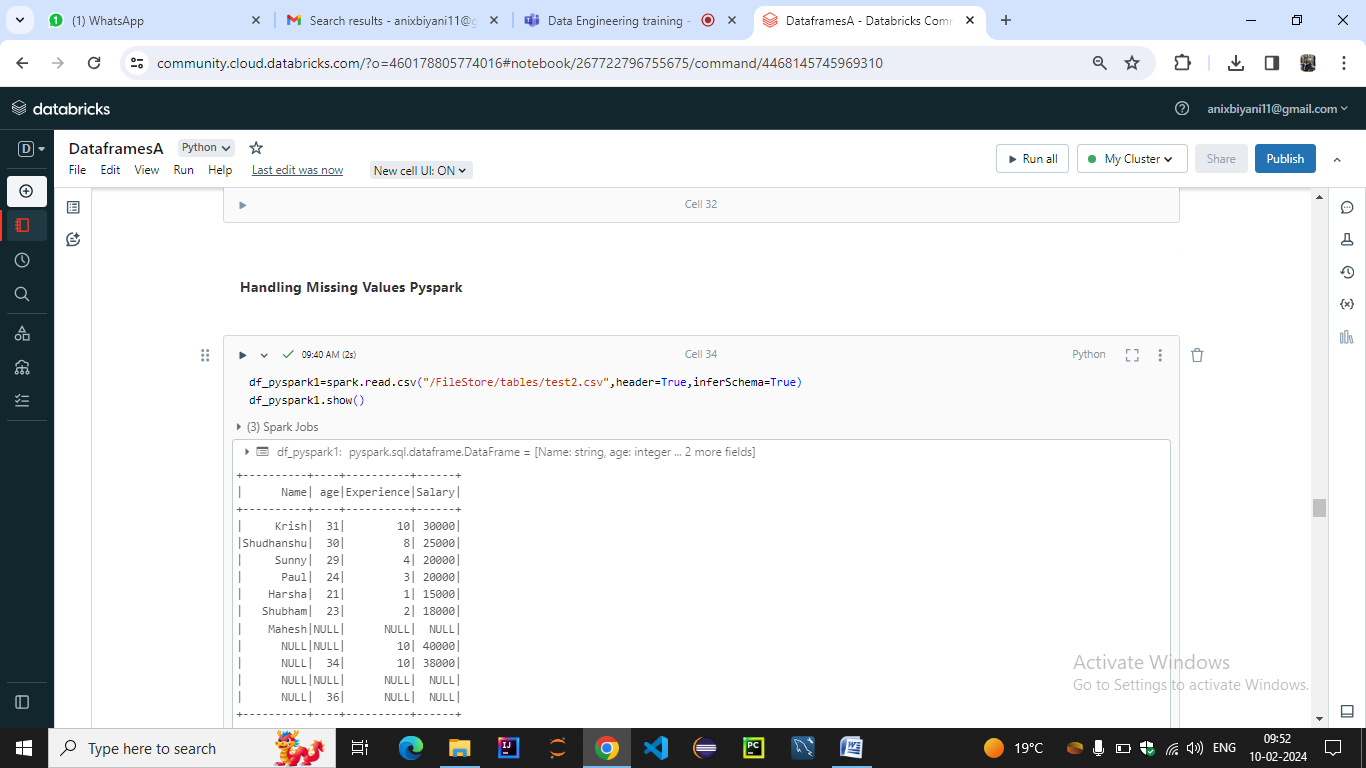
\*\*pivot(pivot\_col, values=None)\*\*

.

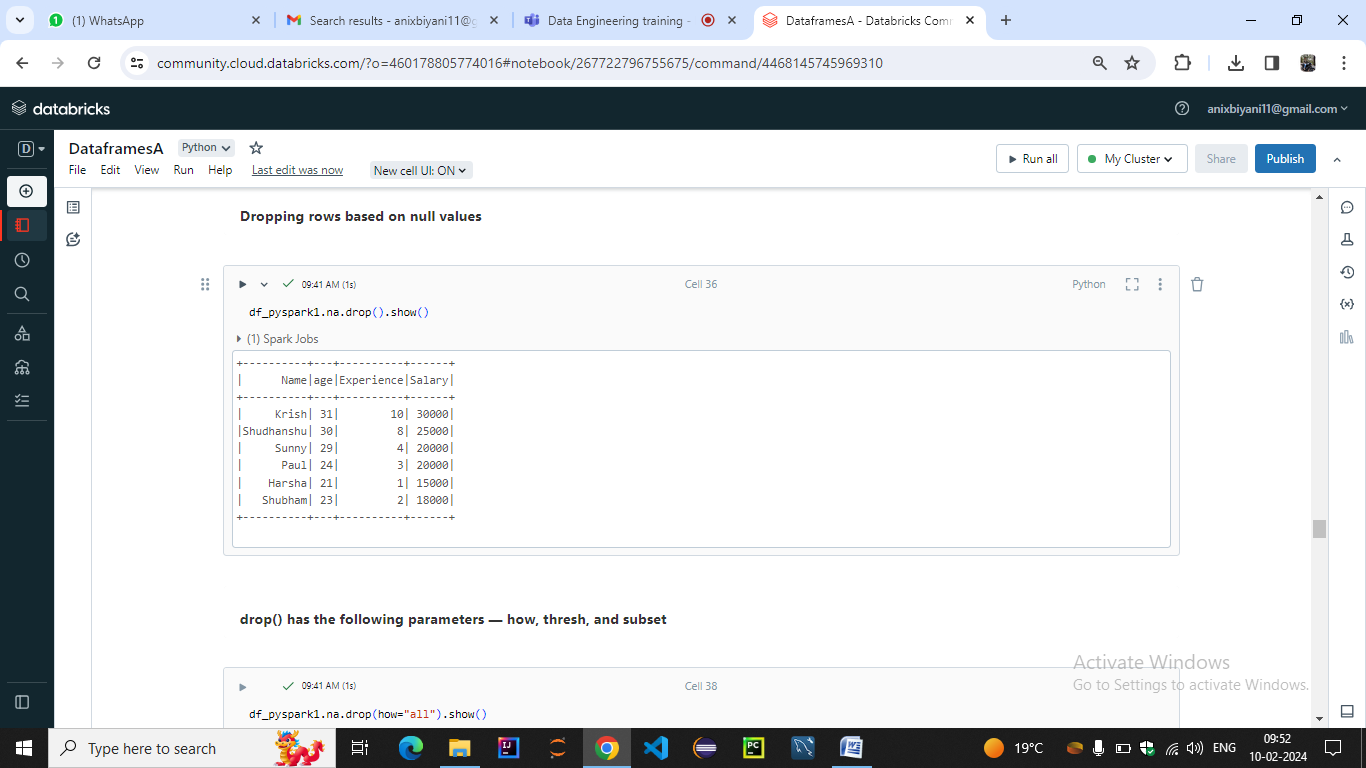
 

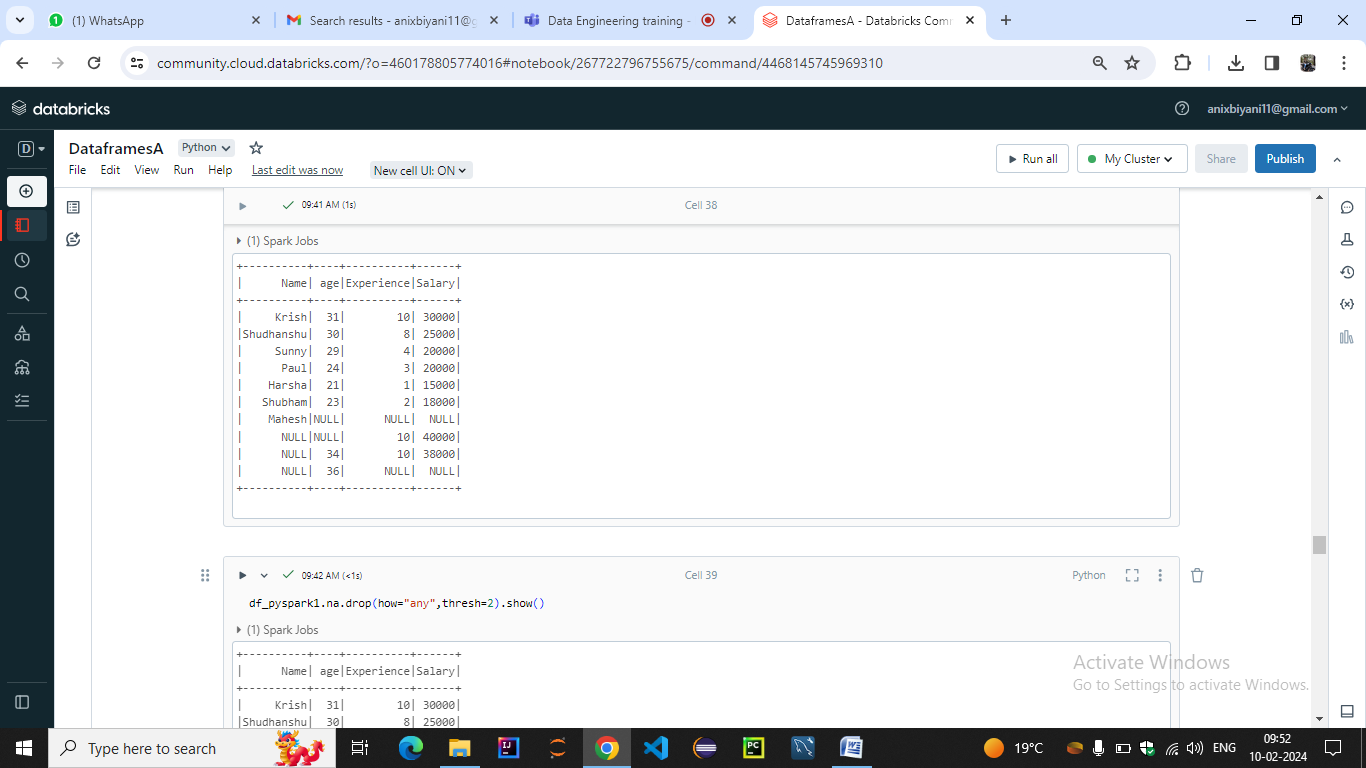
HAN

HANHAN\*\*HANDLING MISSING VALUES

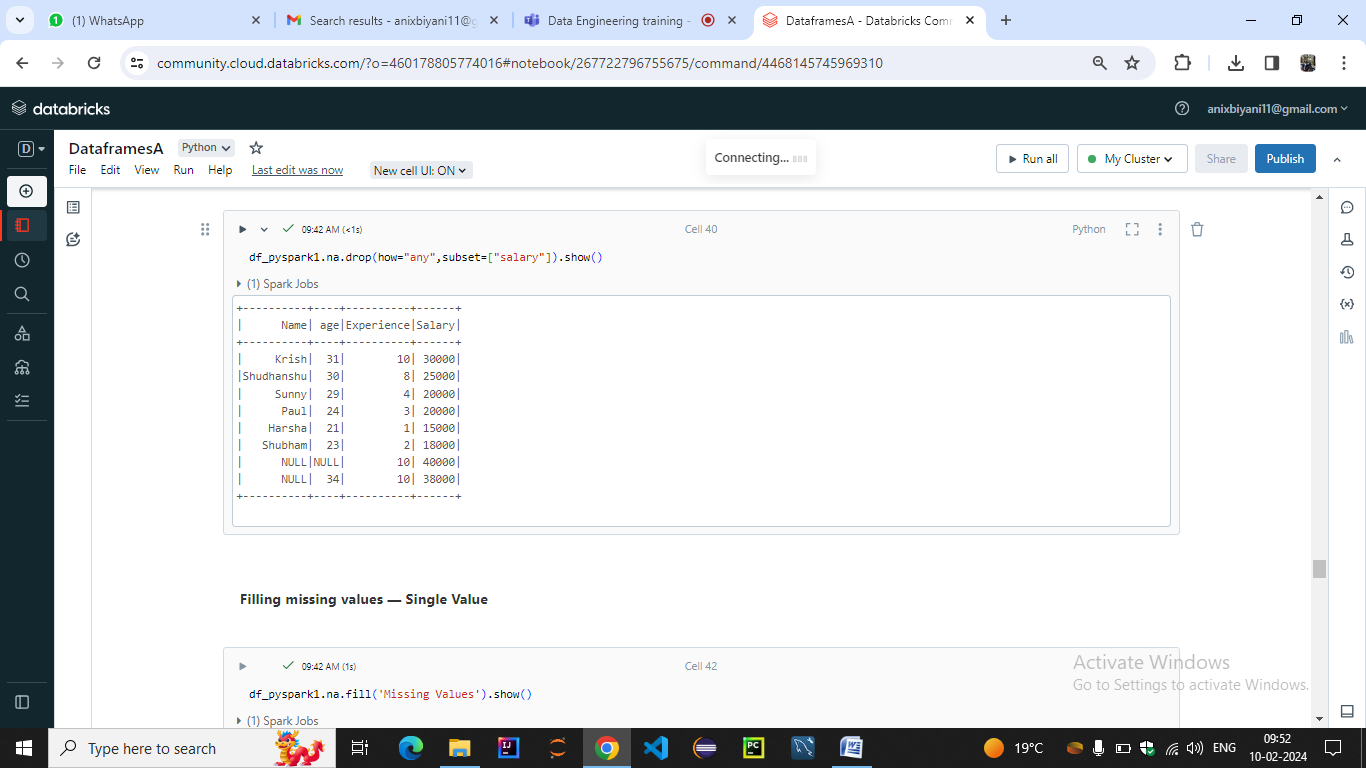
. 

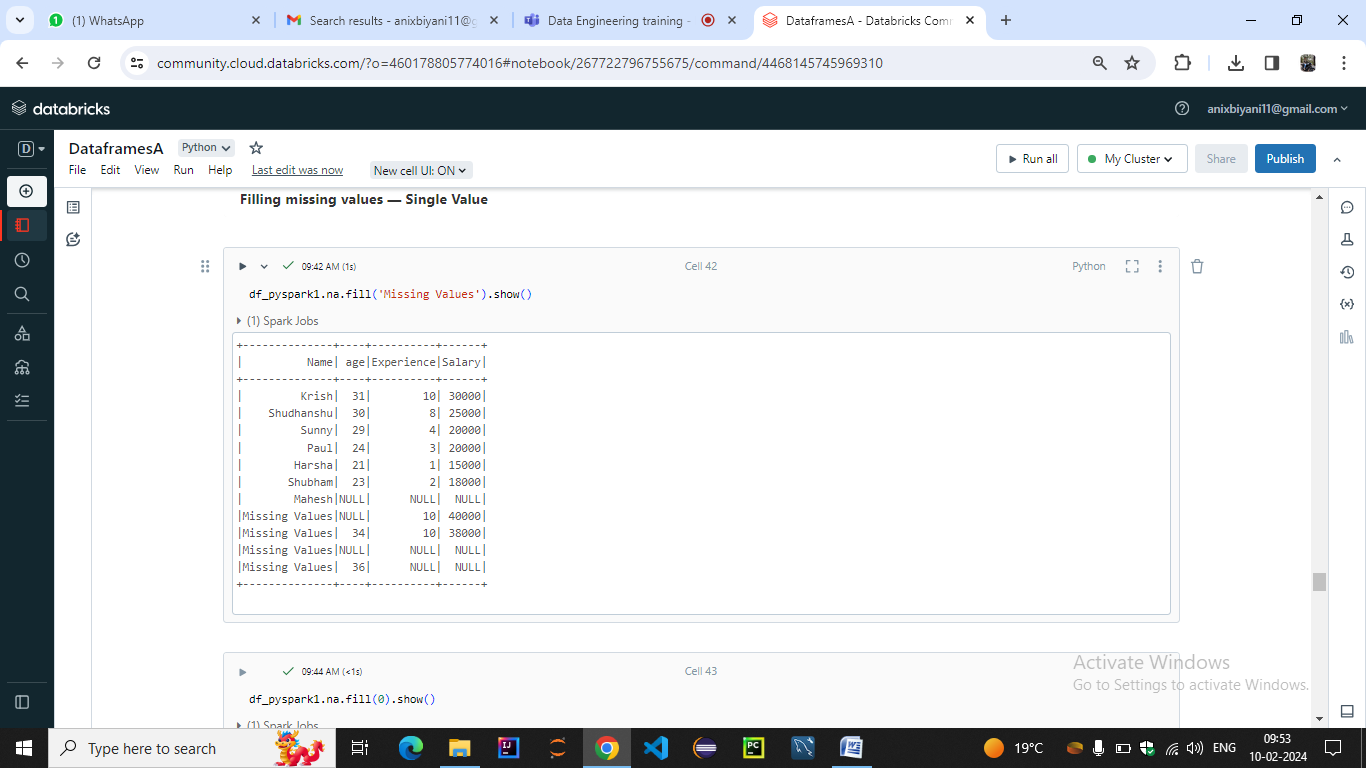
\*\*DROPPING ROWS

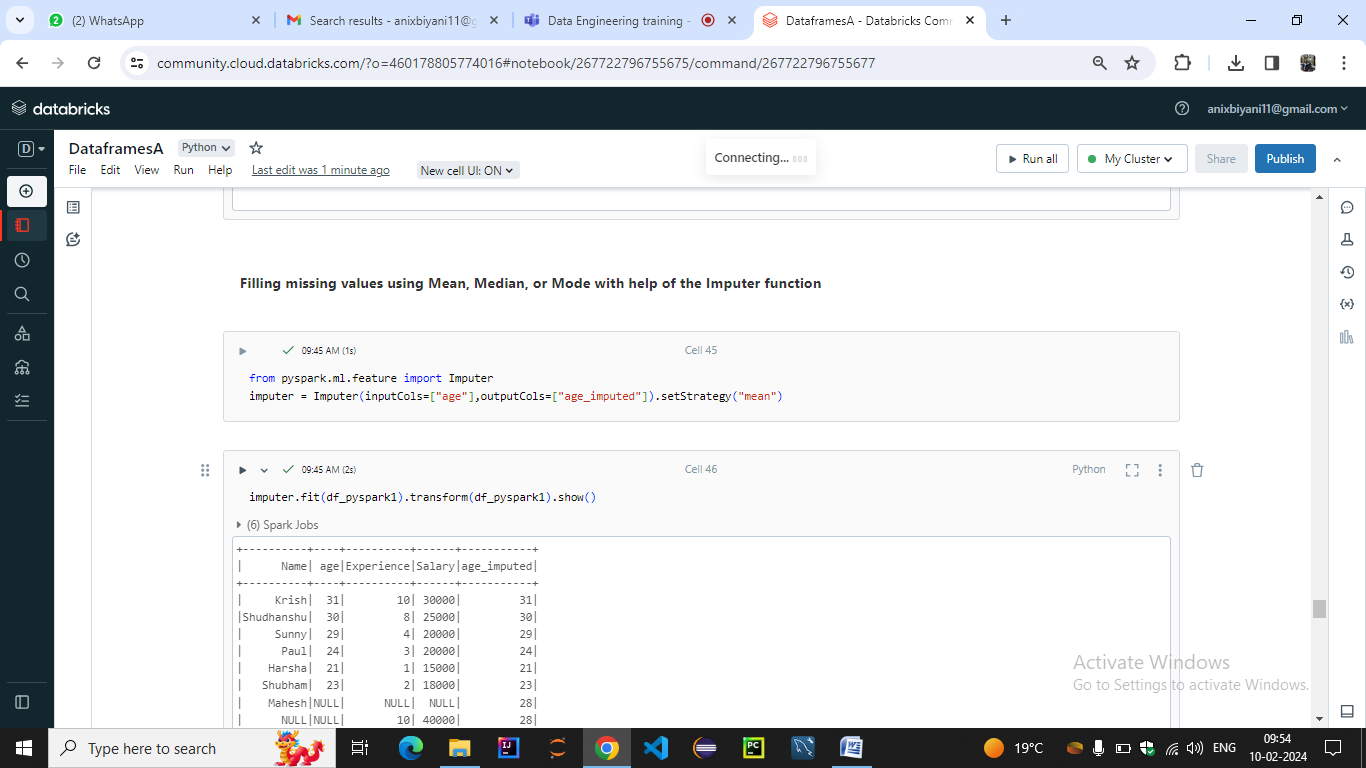




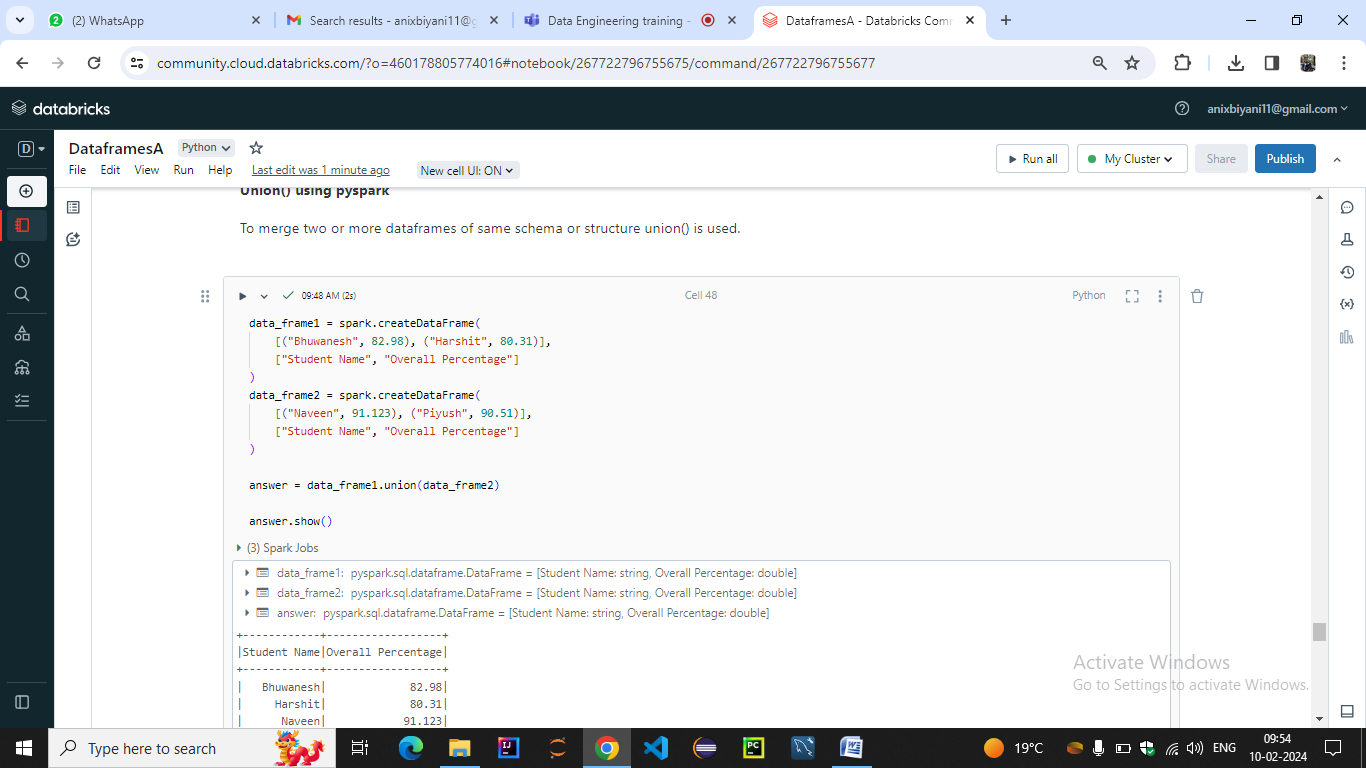
MISSING VALUES

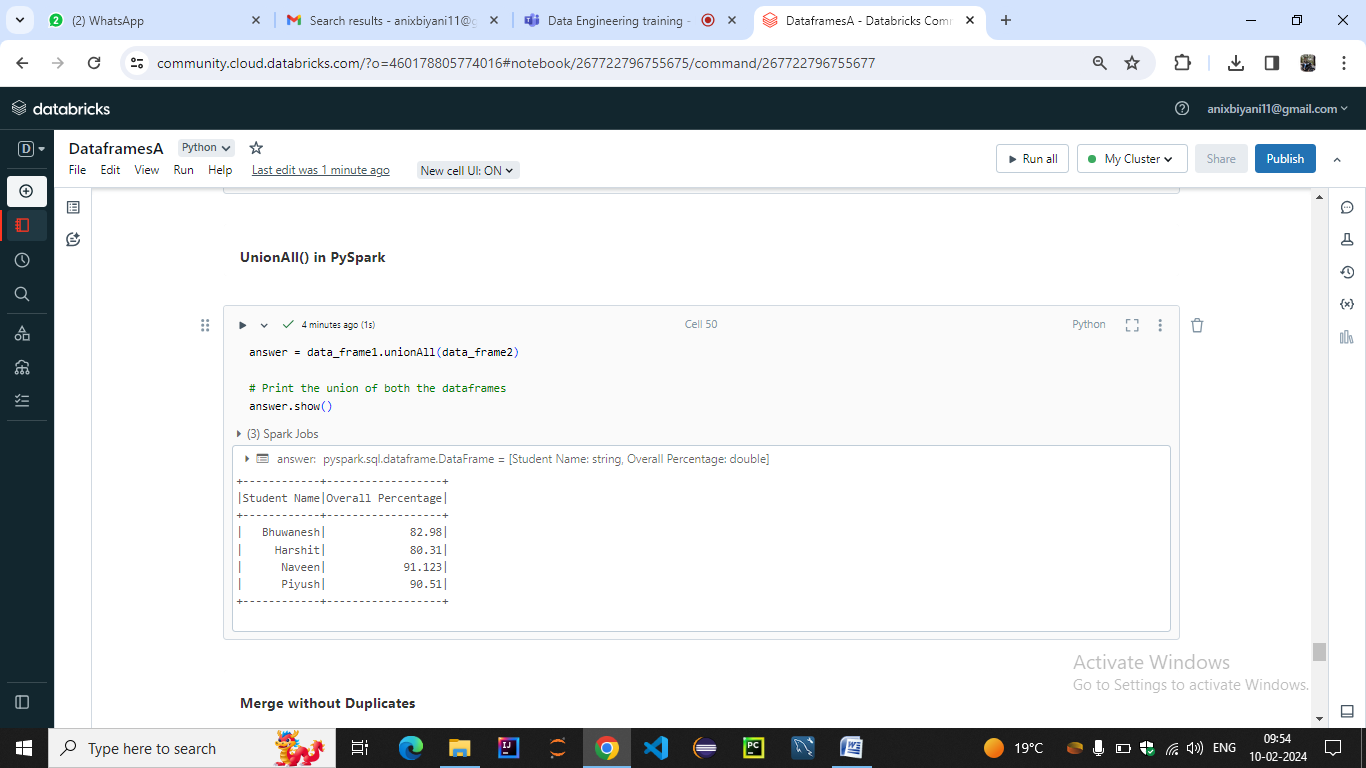






UNION AND UNION ALL





MERGW WITHOUT DUPLICATES

