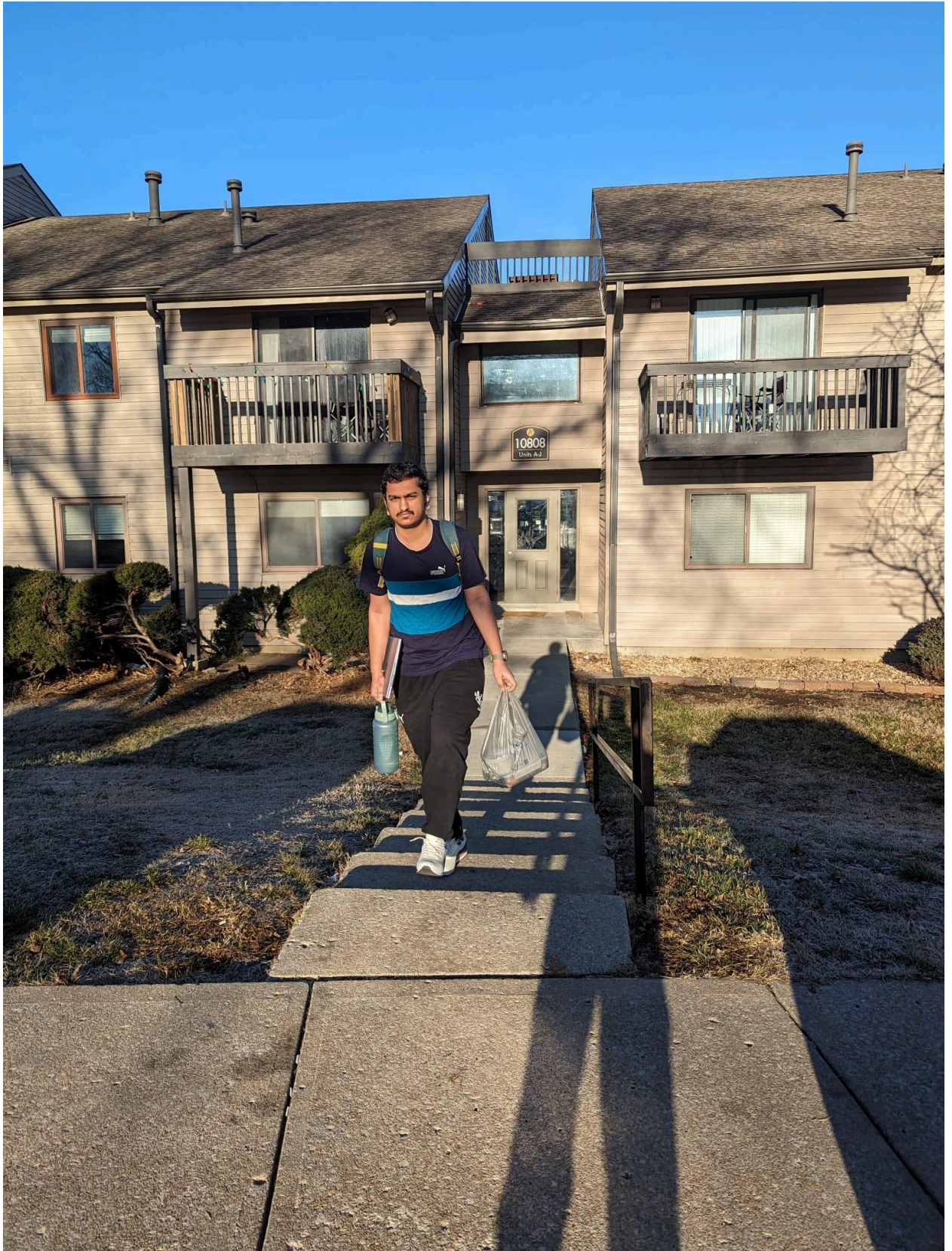


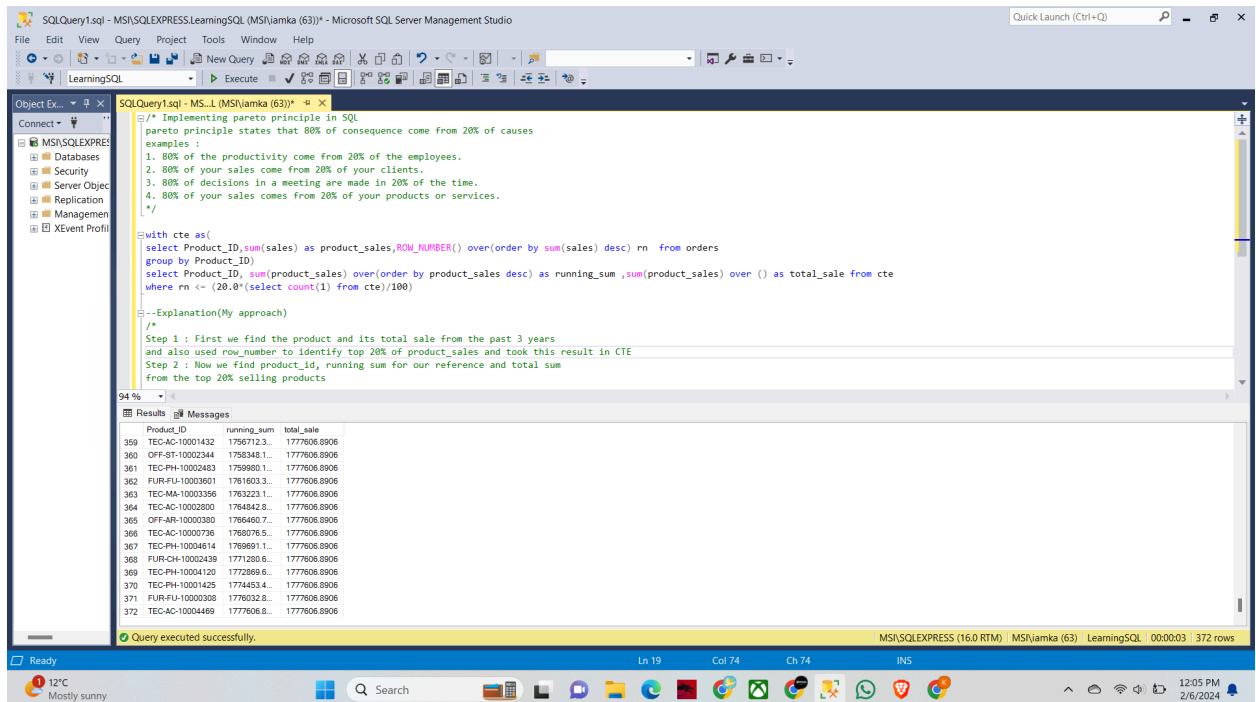
Day1 - Feb 6th 2024

1. Woke up at 5:45 am and completed cold water bath
2. Cooked breakfast and lunch

3. 8:20am headed to library



- Marketed my profile for Data Engineer role and applied few jobs and connected with professional data engineer's on linkedin to expand my network
- Later solved 2 complex SQL Queries(from Ankit Bansal youtube) and added explanation for future reference.

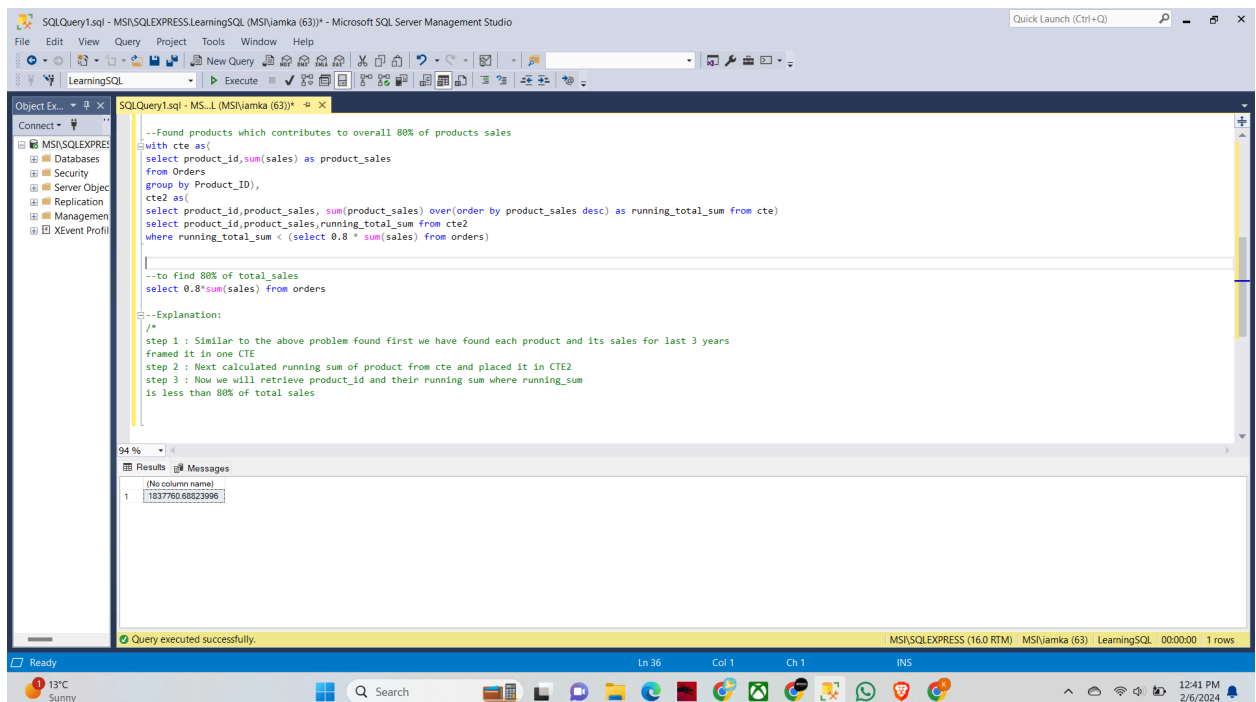


```
/* Implementing pareto principle in SQL
pareto principle states that 80% of consequence come from 20% of causes
examples :
1. 80% of the productivity come from 20% of the employees.
2. 80% of your sales come from 20% of your clients.
3. 80% of decisions in a meeting are made in 20% of the time.
4. 80% of your sales comes from 20% of your products or services.
*/

with cte as(
select Product_ID,sum(sales) as product_sales,row_number() over(order by sum(sales) desc) rn from orders
group by Product_ID)
select Product_ID, sum(product_sales) over(order by product_sales desc) as running_sum ,sum(product_sales) over () as total_sale from cte
where rn <= (20.0*(select count(1) from cte)/100)

--Explanation(My approach)
/*
Step 1 : First we find the product and its total sale from the past 3 years
and also used row_number to identify top 20% of product_sales and took this result in CTE
Step 2 : Now we find product_id, running sum for our reference and total sum
from the top 20% selling products
*/
```

Product_ID	running_sum	total_sale
359 TEC-AC-10001432	17967123..	1777606.6906
360 OFF-ST-10002344	17563481..	1777606.6906
361 TEC-PH-10002483	17599801..	1777606.6906
362 FUR-FU-10003601	17618033..	1777606.6906
363 TEC-MA-10003356	17632231..	1777606.6906
364 TEC-AC-10002800	17648428..	1777606.6906
365 OFF-AR-10000380	17664607..	1777606.6906
366 TEC-AC-10000736	17680176..	1777606.6906
367 TEC-PH-10004614	17696911..	1777606.6906
368 FUR-CH-10002439	17712801..	1777606.6906
369 TEC-PH-10004120	17728696..	1777606.6906
370 TEC-PH-10001425	17744534..	1777606.6906
371 FUR-FU-10000308	17760328..	1777606.6906
372 TEC-AC-10004469	17776068..	1777606.6906



```
--Found products which contributes to overall 80% of products sales
with cte as(
select product_id,sum(sales) as product_sales
from Orders
group by Product_ID),
cte2 as(
select product_id,product_sales, sum(product_sales) over(order by product_sales desc) as running_total_sum from cte)
select product_id,product_sales,running_total_sum from cte2
where running_total_sum < (select 0.8 * sum(sales) from orders)

--to find 80% of total_sales
select 0.8*sum(sales) from orders

--Explanation:
/*
step 1 : Similar to the above problem found first we have found each product and its sales for last 3 years
framed it in one CTE
step 2 : Next calculated running sum of product from cte and placed it in CTE2
step 3 : Now we will retrieve product_id and their running sum where running_sum
is less than 80% of total sales
*/
```

(No column name)
1837760.68823996

-
-
-
-
-
-
- Started Learning BigData on GreatLearning platform.
- Course Link : <https://olympus.mygreatlearning.com/courses/10977>

9. Completed Learning Hadoop part 1 on feb 3rdth and feb 4th :
https://docs.google.com/document/d/1V14CVJkM8RBhT9tZH2iKh3mj-tiBoF7qXXSN_NvWub8/edit?usp=sharing (47 Pages)
10. Today Completed till Hadoop 3.X ..next should run hadoop on AWS EC2 Ubuntu instance
11. Completed till 27 pages on feb 6th 2024 :
<https://docs.google.com/document/d/1CPiJa17UKwzBhtvngG62Hnjv6xBYkkhxFFQA0dcZ86Y/edit?usp=sharing>
12. Today my time was wasted a lot..but I can't avoid going out with friends.
13. Will work even harder tomorrow and should complete more chapters.