PROJECT-1

Sleep, Health and Life Style.

Steps followed:

1. Upload data to Storage account - Raw Container

2. Setup Databricks

a. Create data bricks workspace

b. Create Azure Service Principal

c. Grant access to ADLS Gen2

d. Mount ADLS Gen2 in Databricks

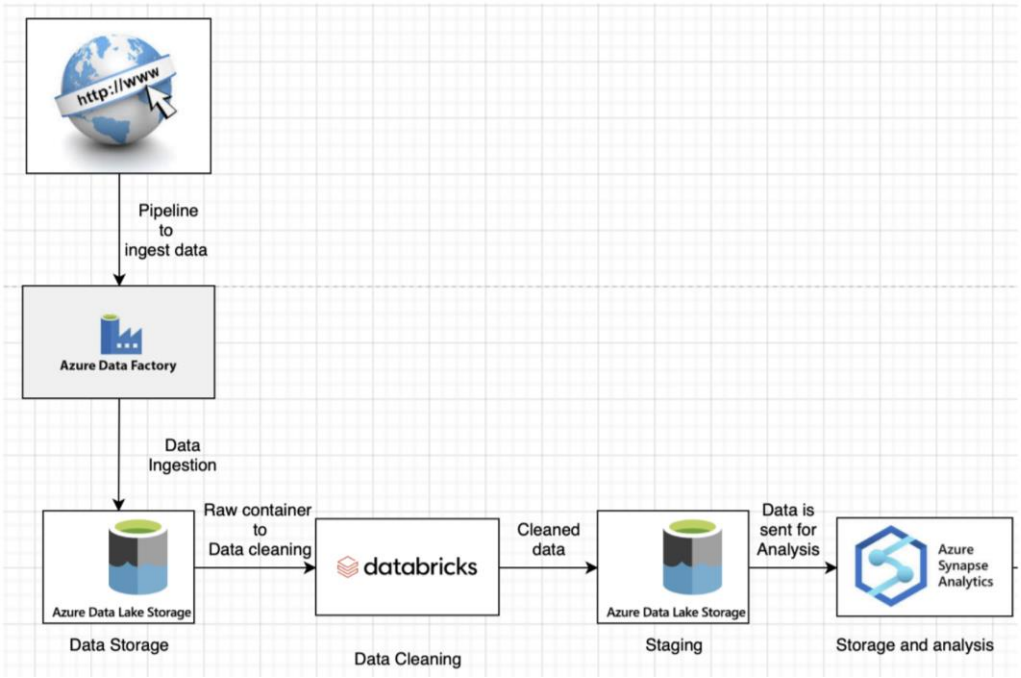
3. Read the file to data bricks and perform any transformation

4. Load the data to Synapse Analytics

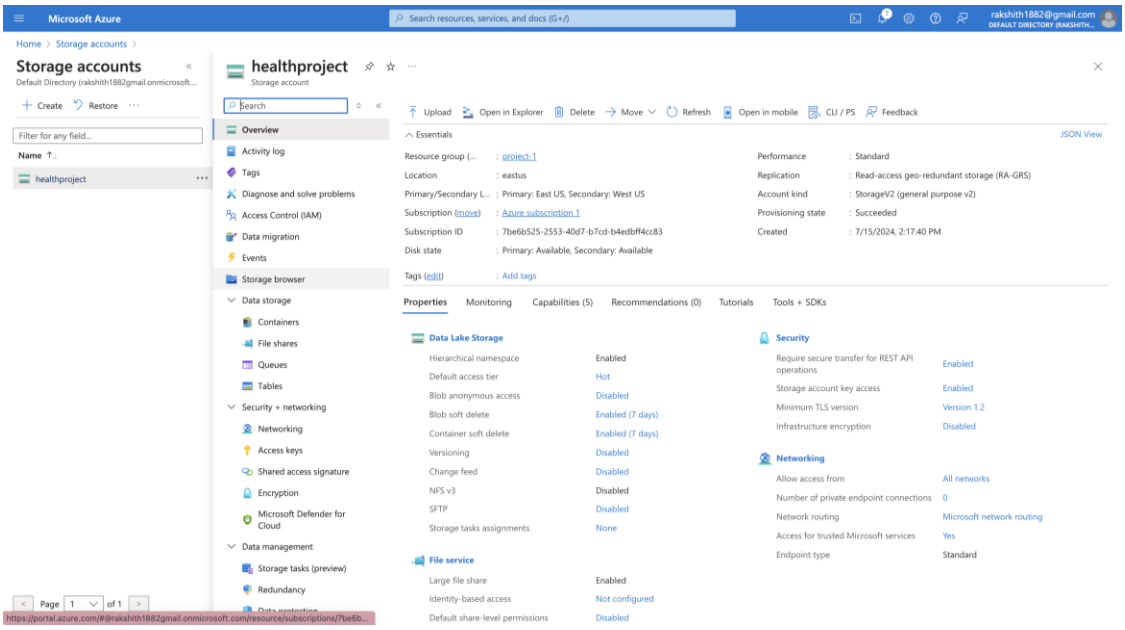
5. Perform Analysis using SQL in Synapse Analytics

6. Visualize using Power BI

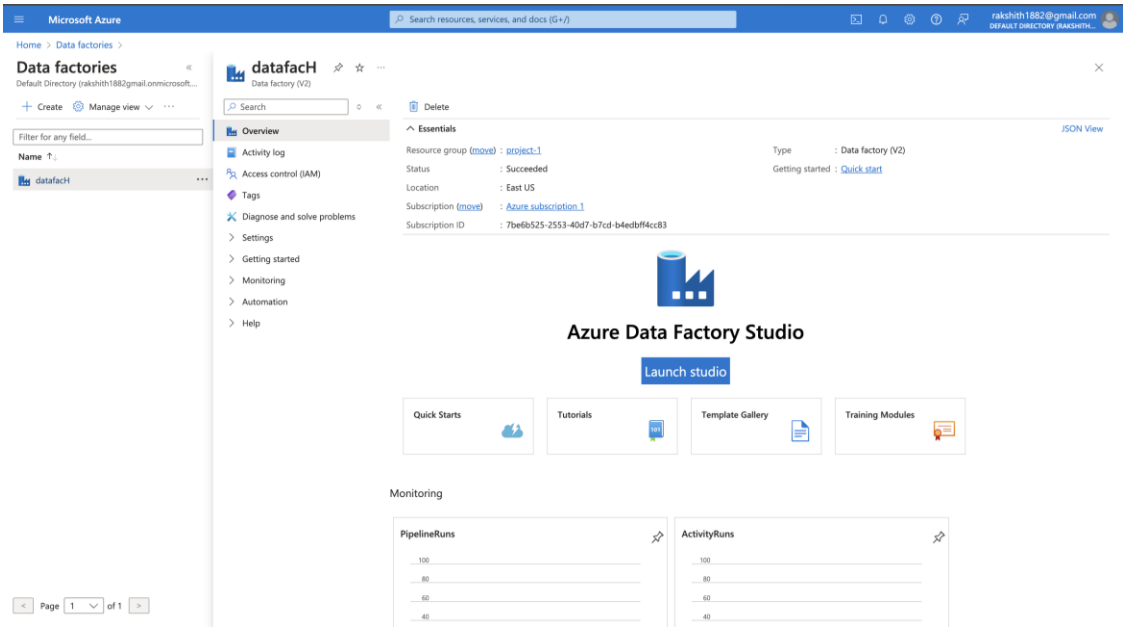
Architectural Diagram:



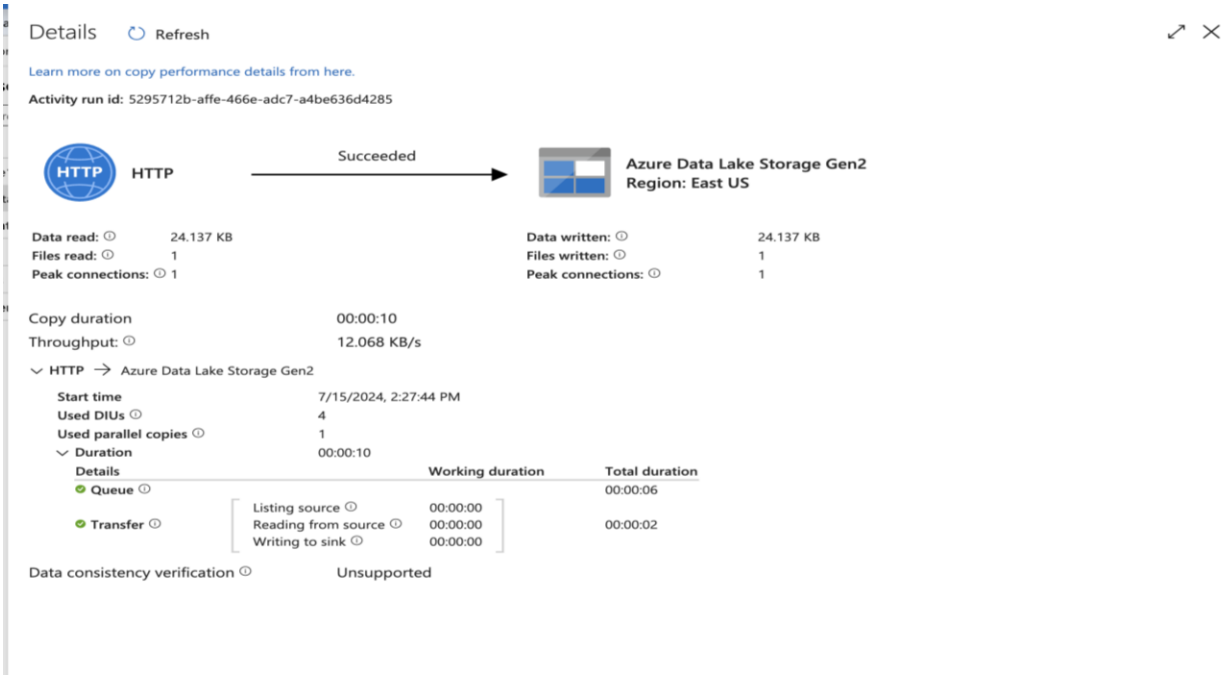
Create ADLS account:



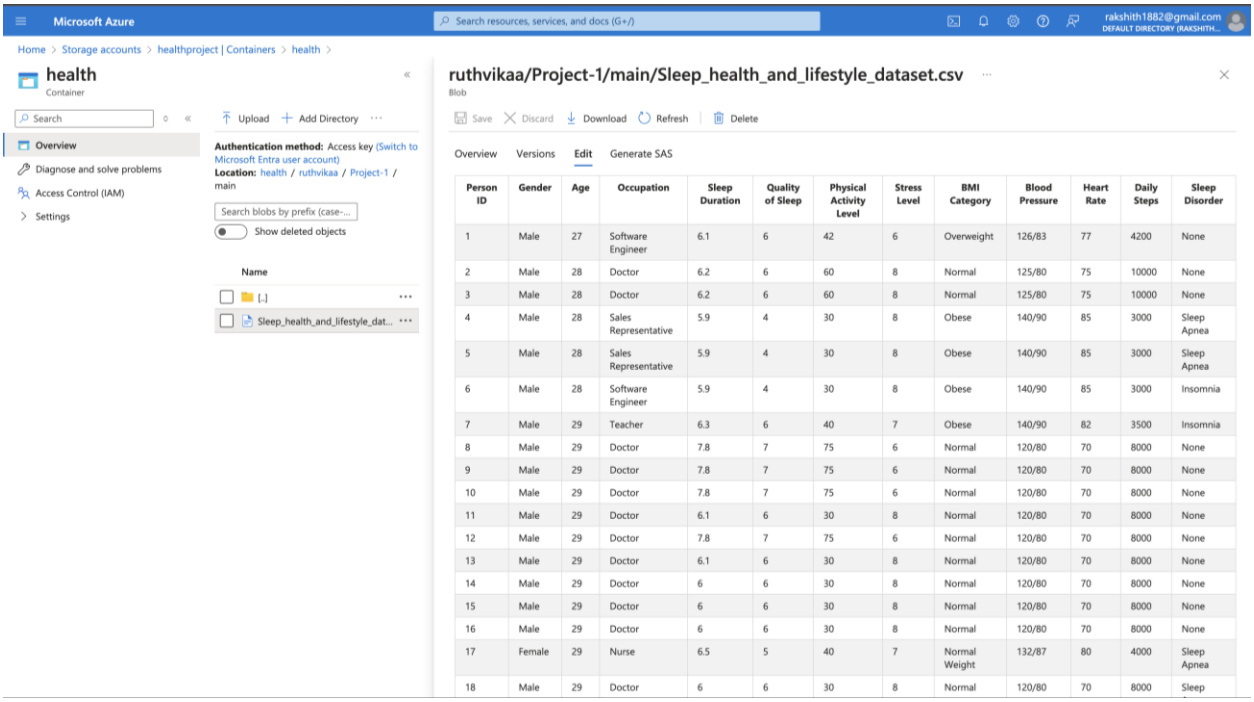
Create Data factory Account:



Creating a pipeline to copy data from GIT to ADLS:

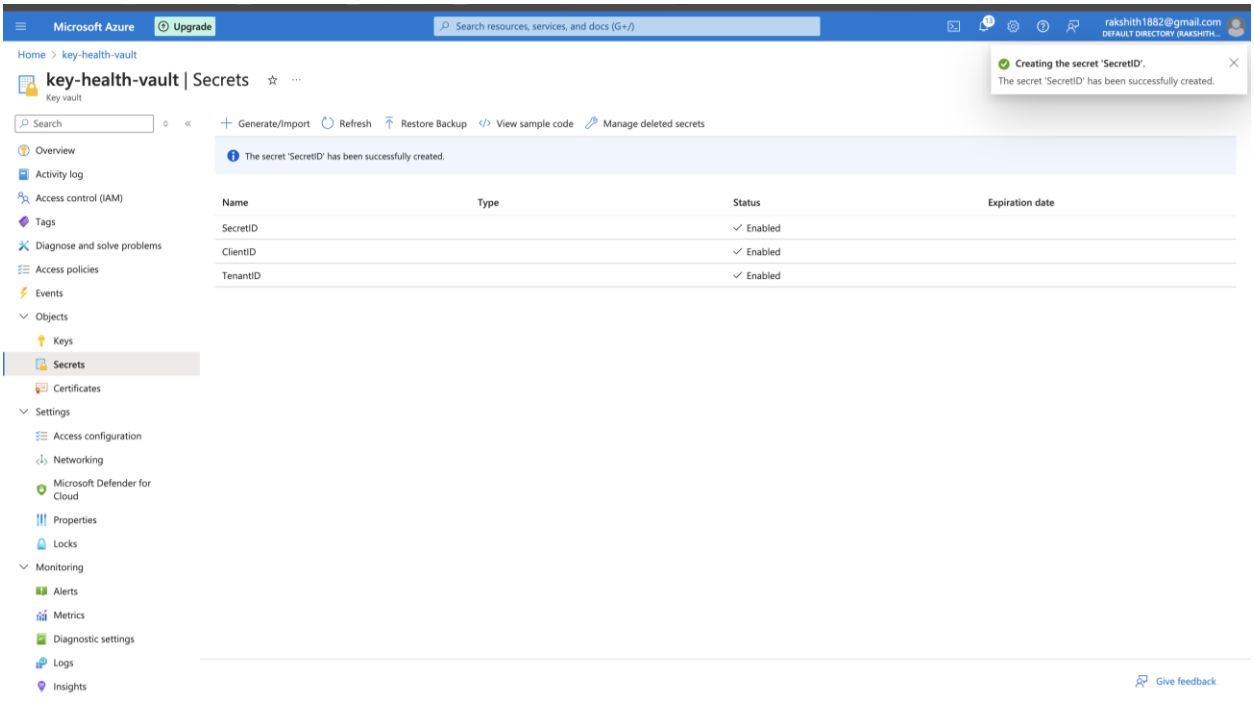


Ingested data:

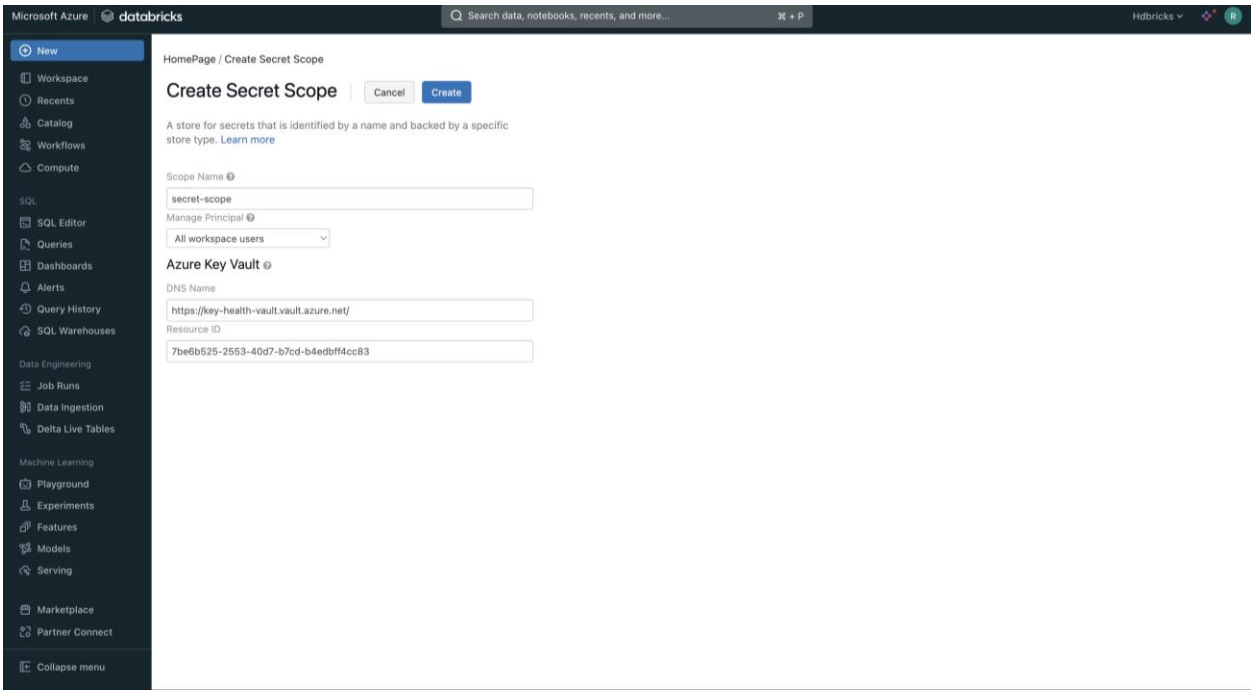


Person ID	Gender	Age	Occupation	Sleep Duration	Quality of Sleep	Physical Activity Level	Stress Level	BMI Category	Blood Pressure	Heart Rate	Daily Steps	Sleep Disorder
1	Male	27	Software Engineer	6.1	6	42	6	Overweight	126/83	77	4200	None
2	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
3	Male	28	Doctor	6.2	6	60	8	Normal	125/80	75	10000	None
4	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
5	Male	28	Sales Representative	5.9	4	30	8	Obese	140/90	85	3000	Sleep Apnea
6	Male	28	Software Engineer	5.9	4	30	8	Obese	140/90	85	3000	Insomnia
7	Male	29	Teacher	6.3	6	40	7	Obese	140/90	82	3500	Insomnia
8	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
9	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
10	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
11	Male	29	Doctor	6.1	6	30	8	Normal	120/80	70	8000	None
12	Male	29	Doctor	7.8	7	75	6	Normal	120/80	70	8000	None
13	Male	29	Doctor	6.1	6	30	8	Normal	120/80	70	8000	None
14	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
15	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
16	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	None
17	Female	29	Nurse	6.5	5	40	7	Normal Weight	132/87	80	4000	Sleep Apnea
18	Male	29	Doctor	6	6	30	8	Normal	120/80	70	8000	Sleep

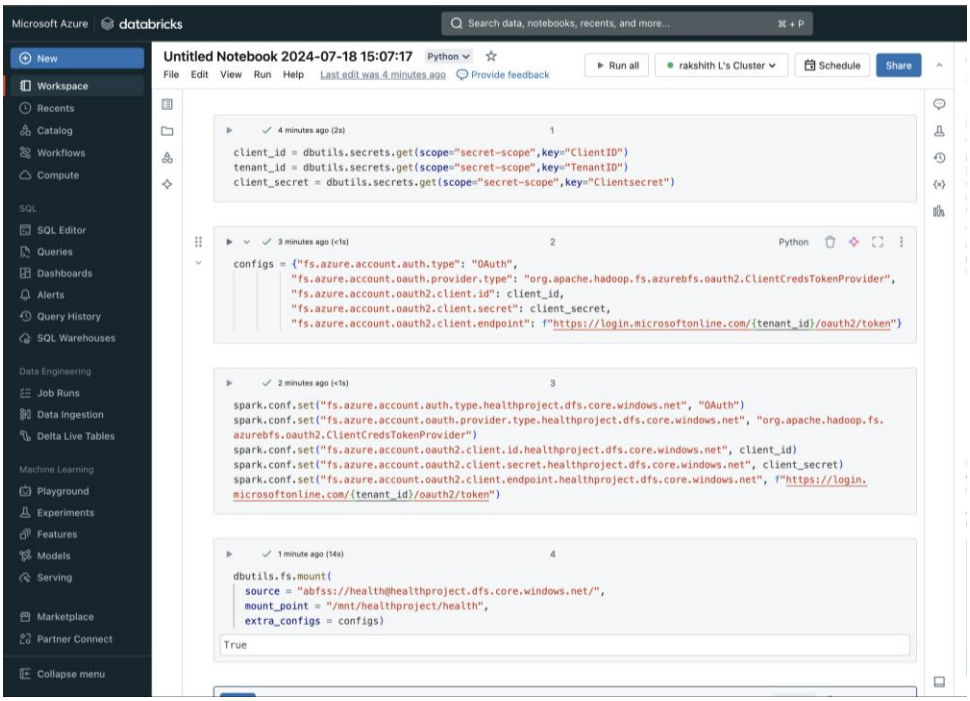
Creating Secrets in Key vault:



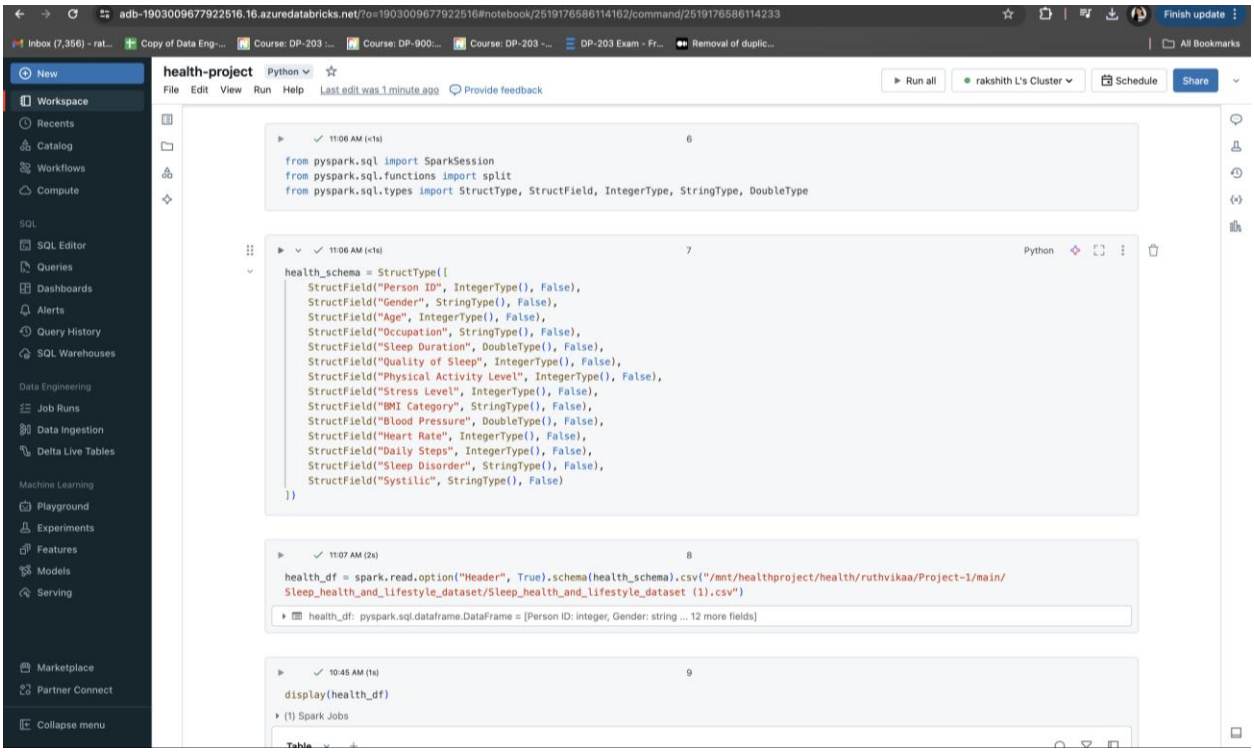
Creating secret scope:



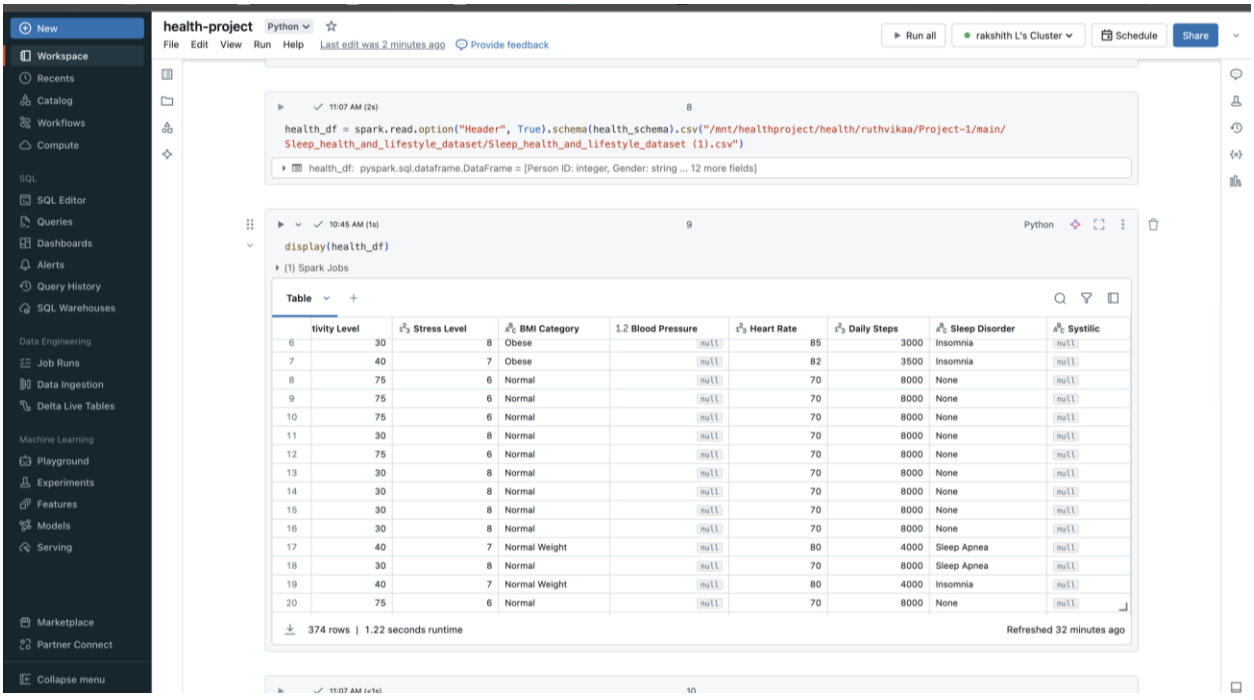
Configuration:



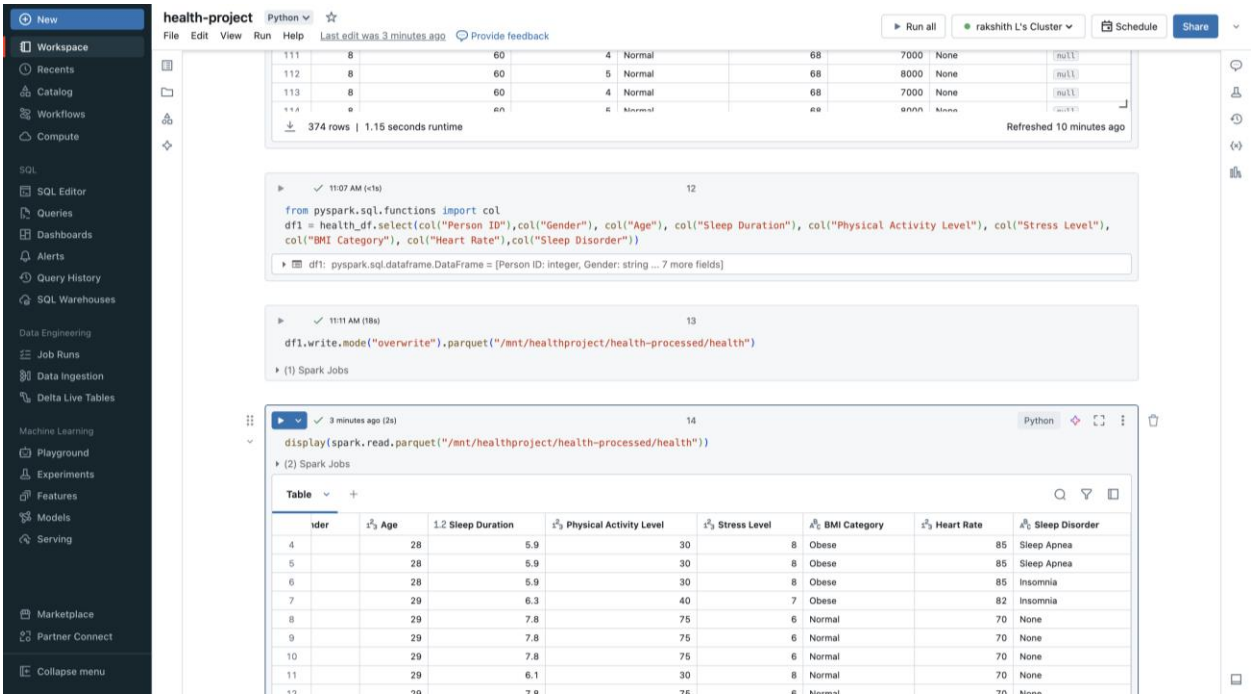
Define schema



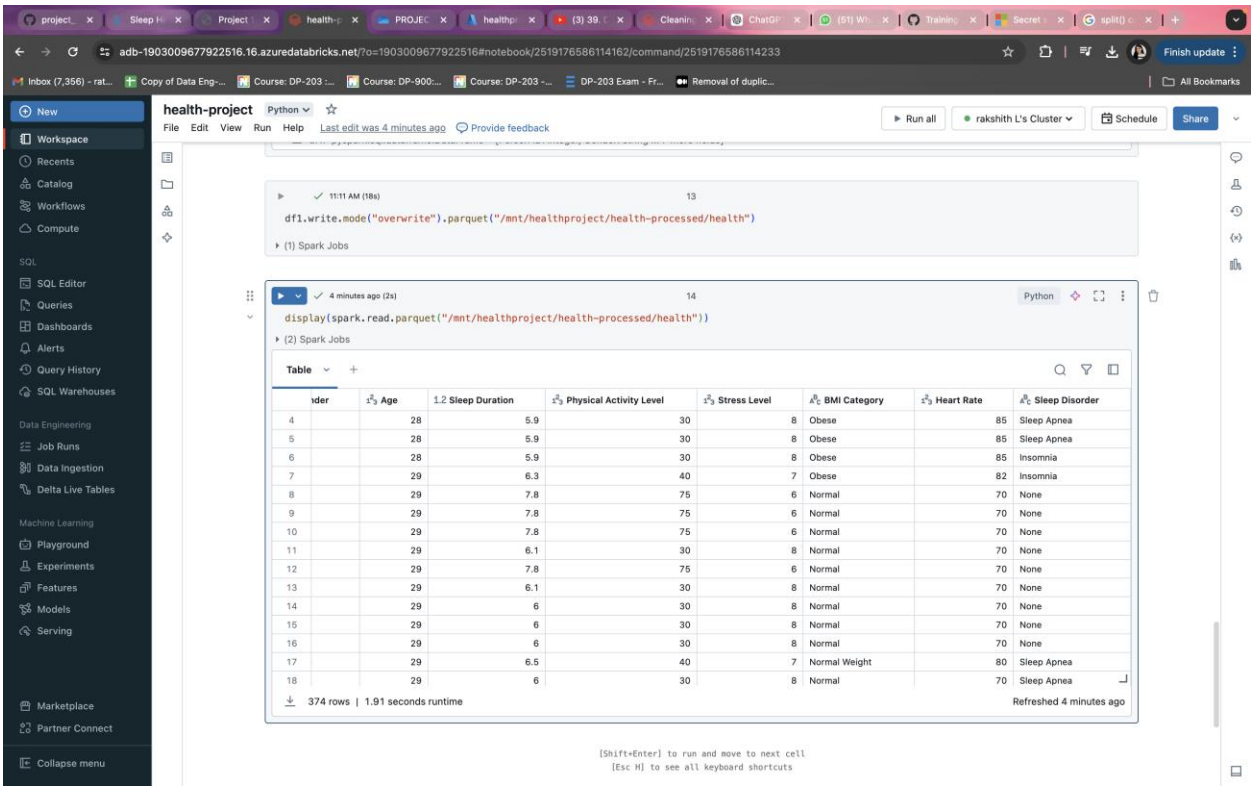
Read CSV File to Data bricks :



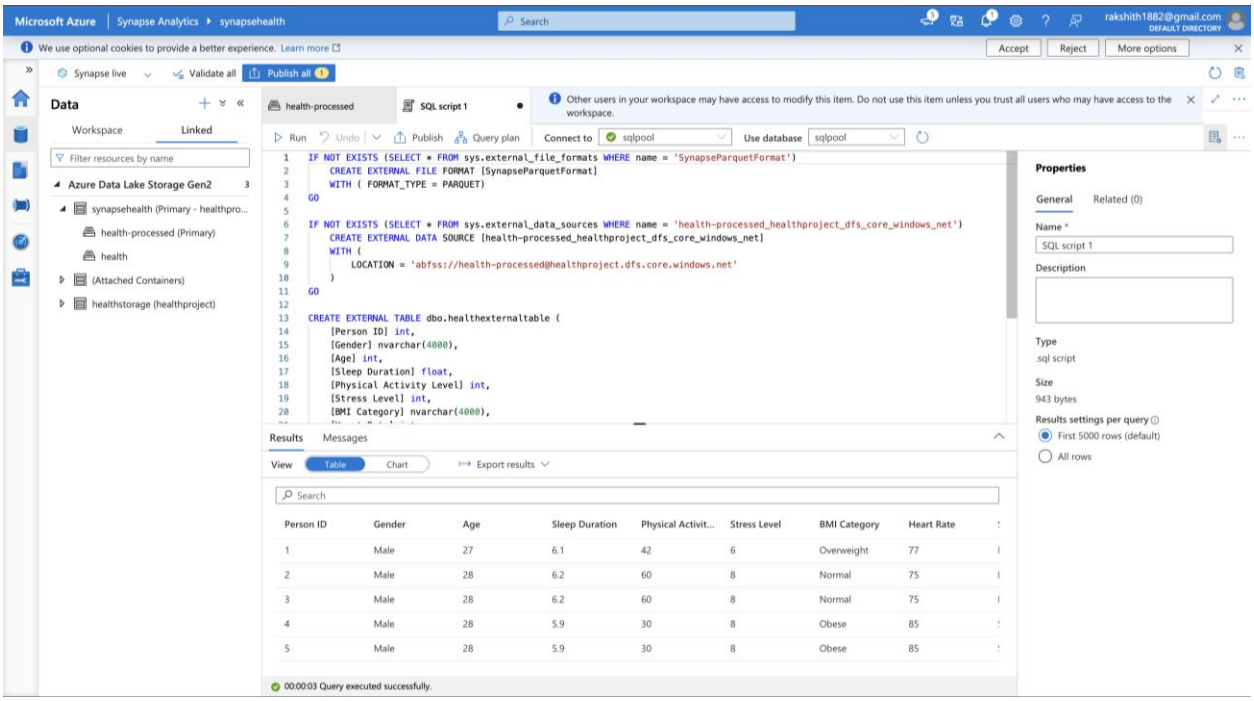
Selected Required Columns:



Wrote the file to adls processed container in parquet format :



Created a table in synapse



AGGREGATIONS: