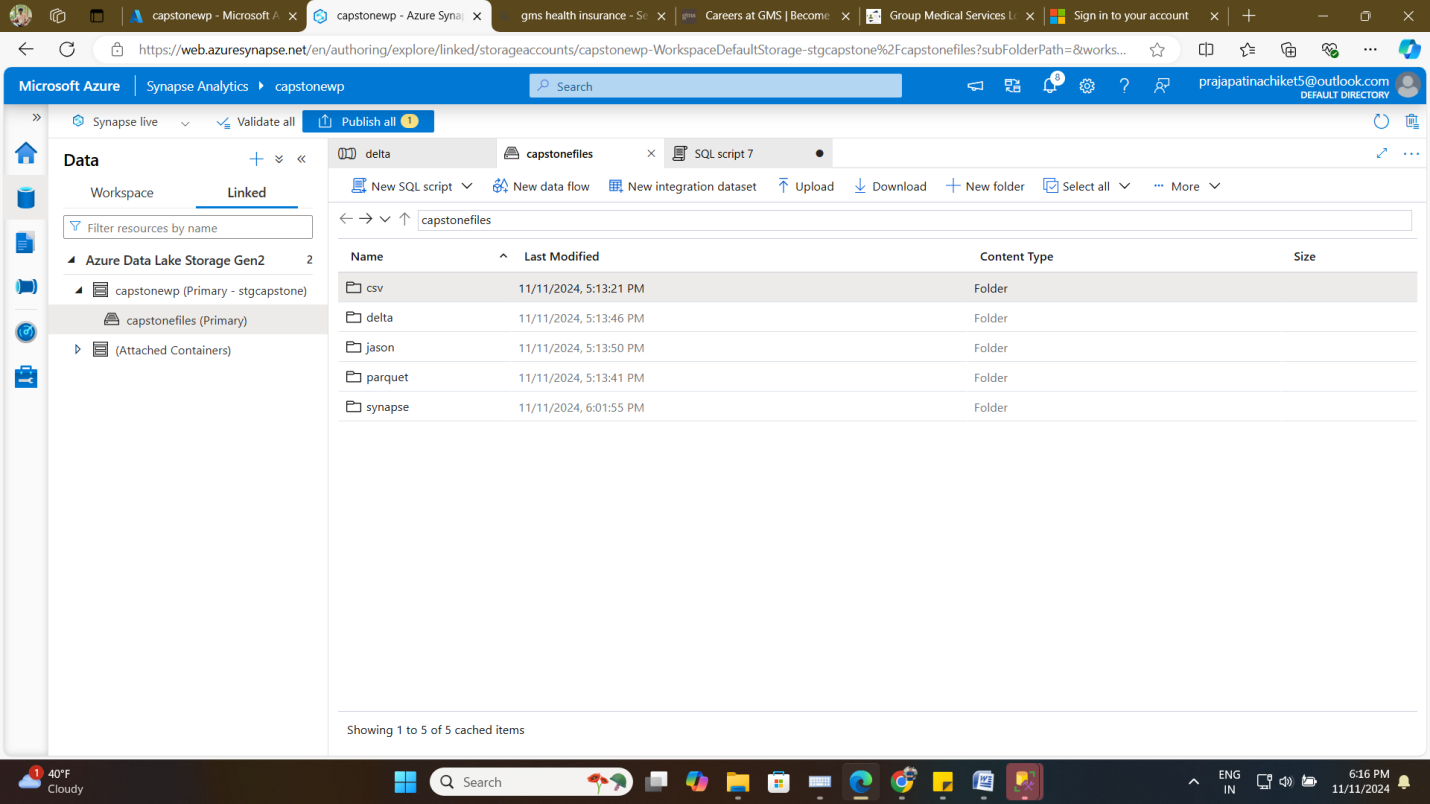
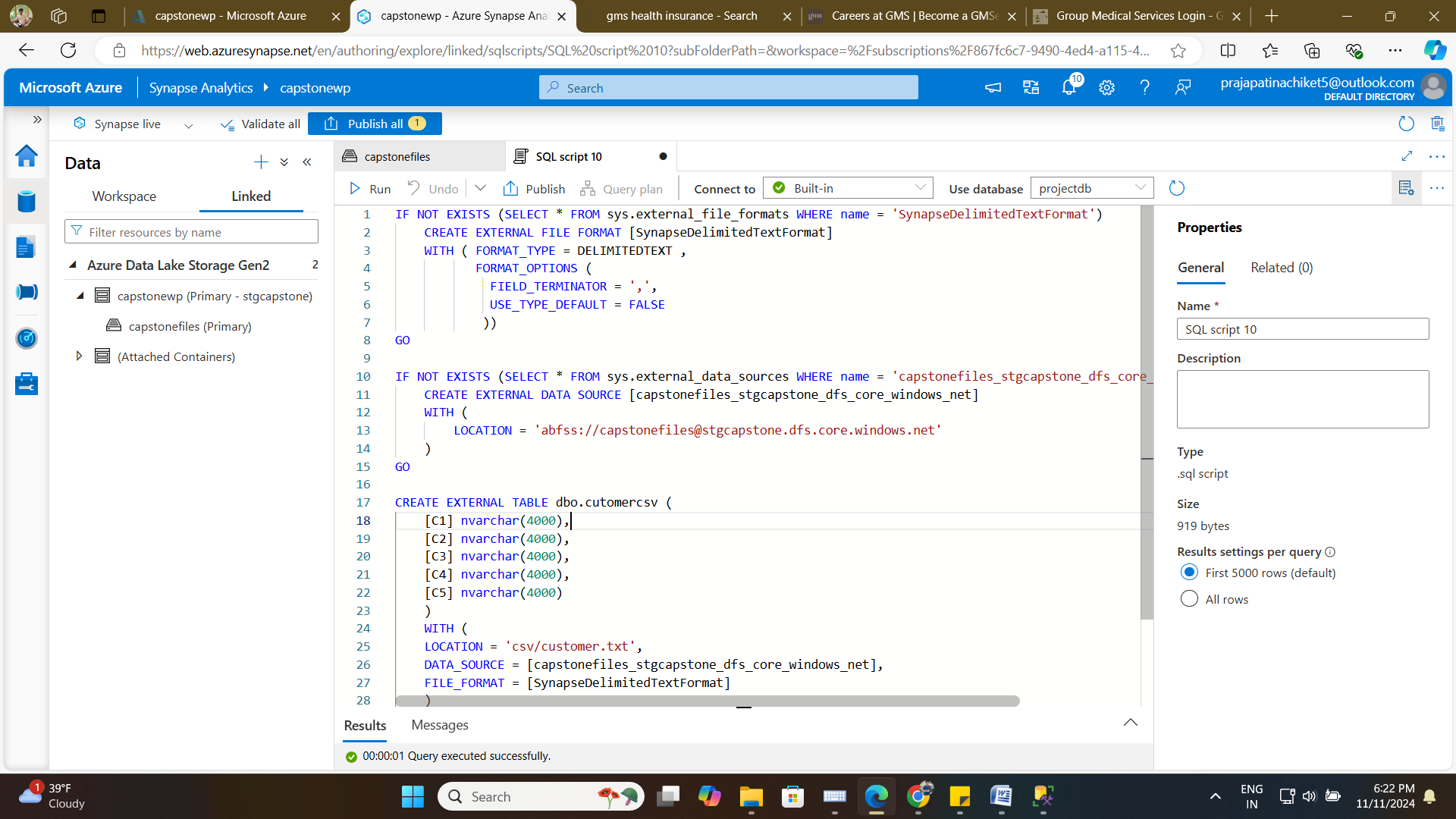
Name:- Nachiket Prajapati

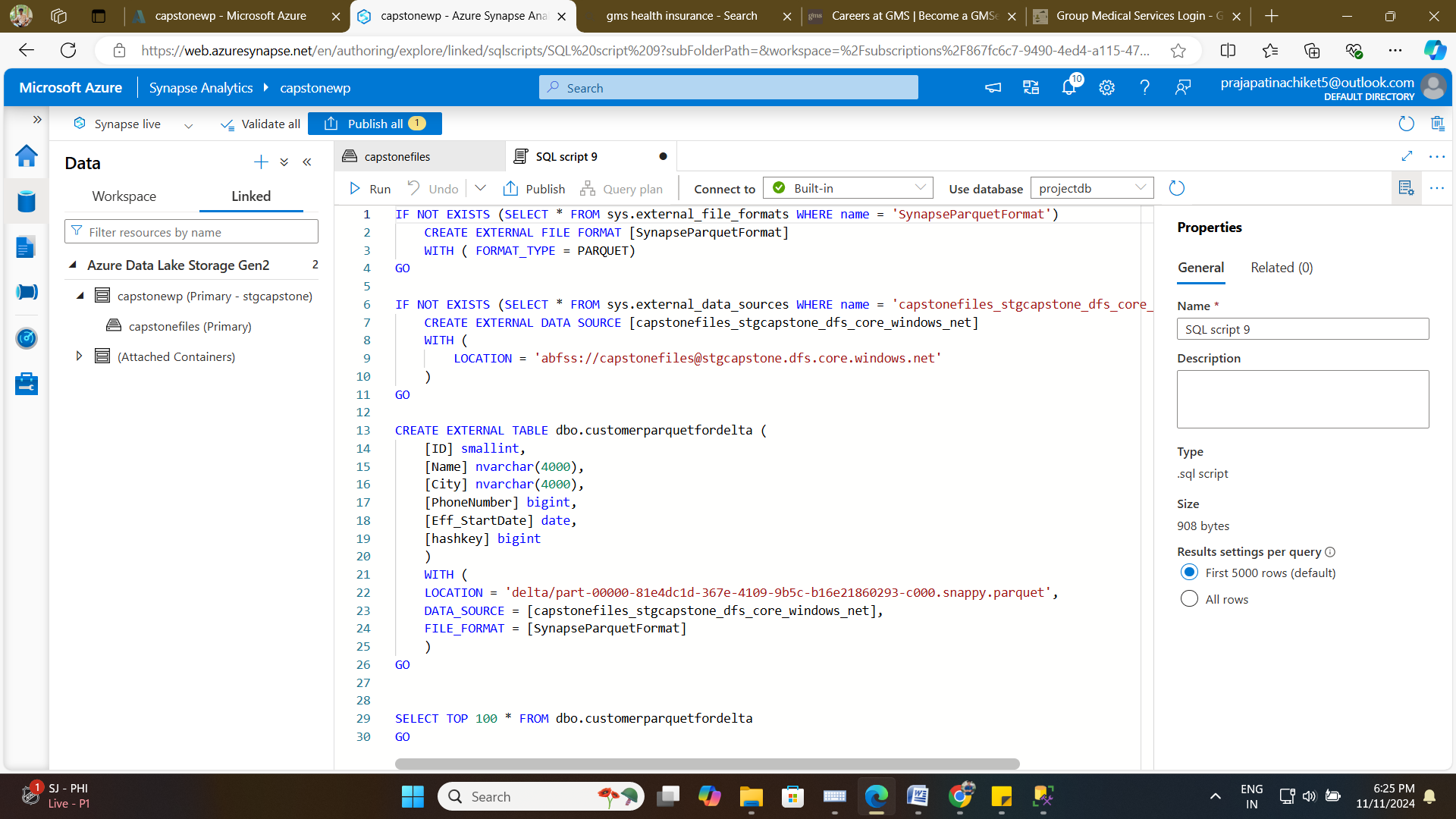
Project Objective:- This project aims to optimize data ingestion and query performance on Azure Synapse Serverless SQL Pool by effectively handling various file formats (CSV, JSON, Parquet, and Delta) from Azure Data Lake Storage Gen2. By implementing strategies like partitioning, indexing, and query tuning, we aim to improve data processing efficiency and reduce query execution times.

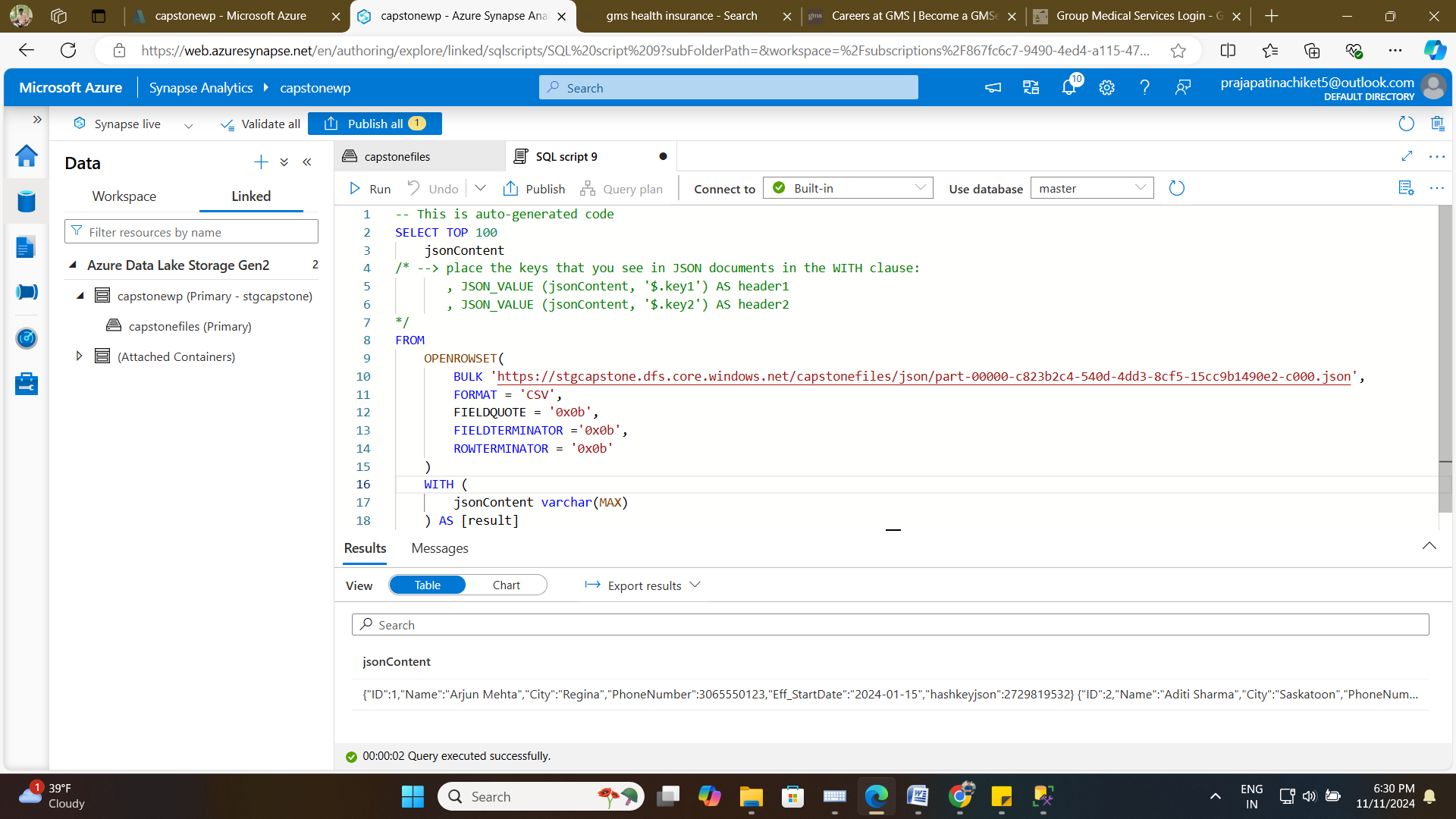
1. **Create External Data Source:** Define the location of the data in Azure Data Lake Storage Gen2.

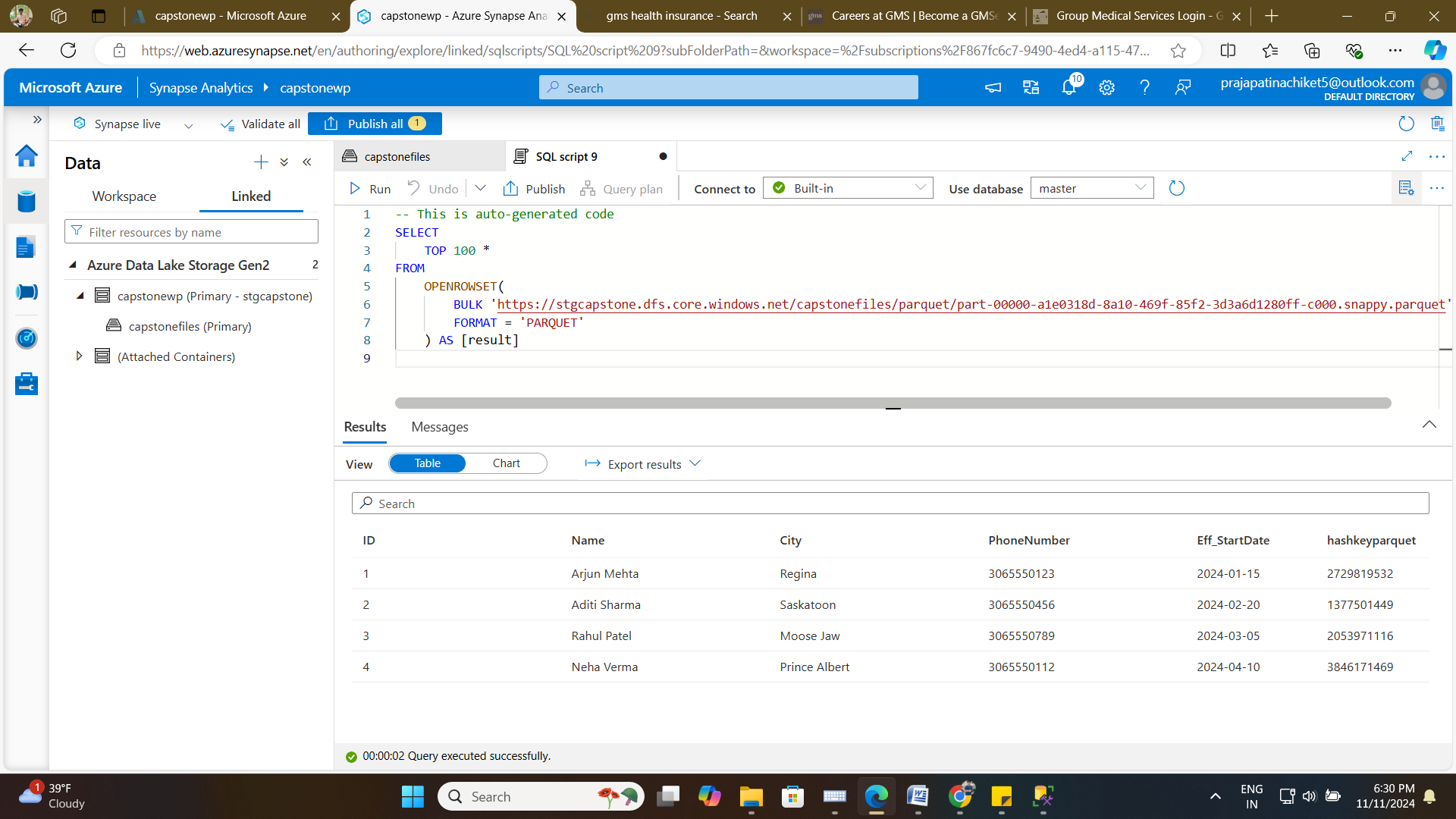


1. **Create External File Formats:** Specify the format for CSV, JSON, Parquet, and Delta files.



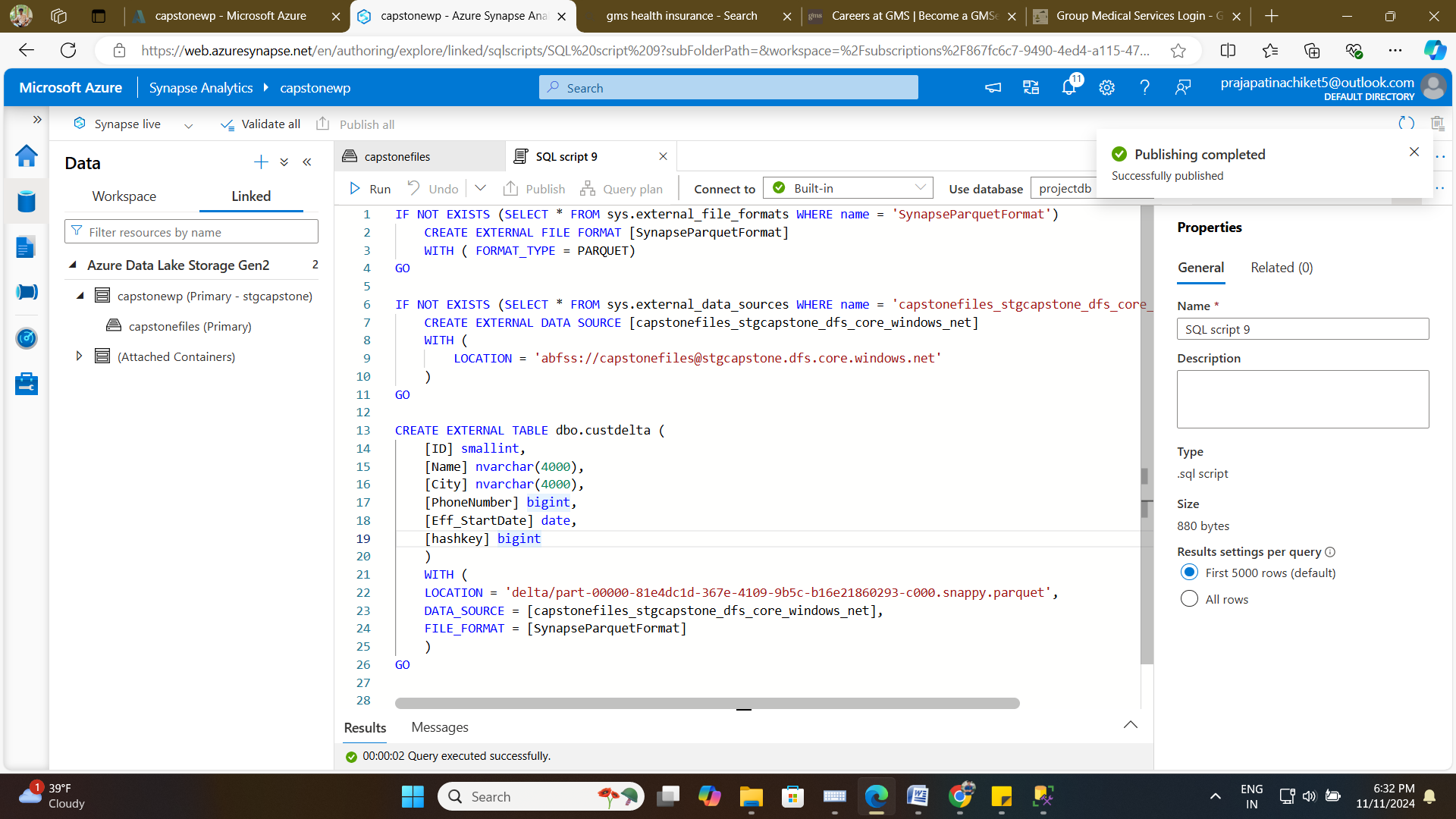




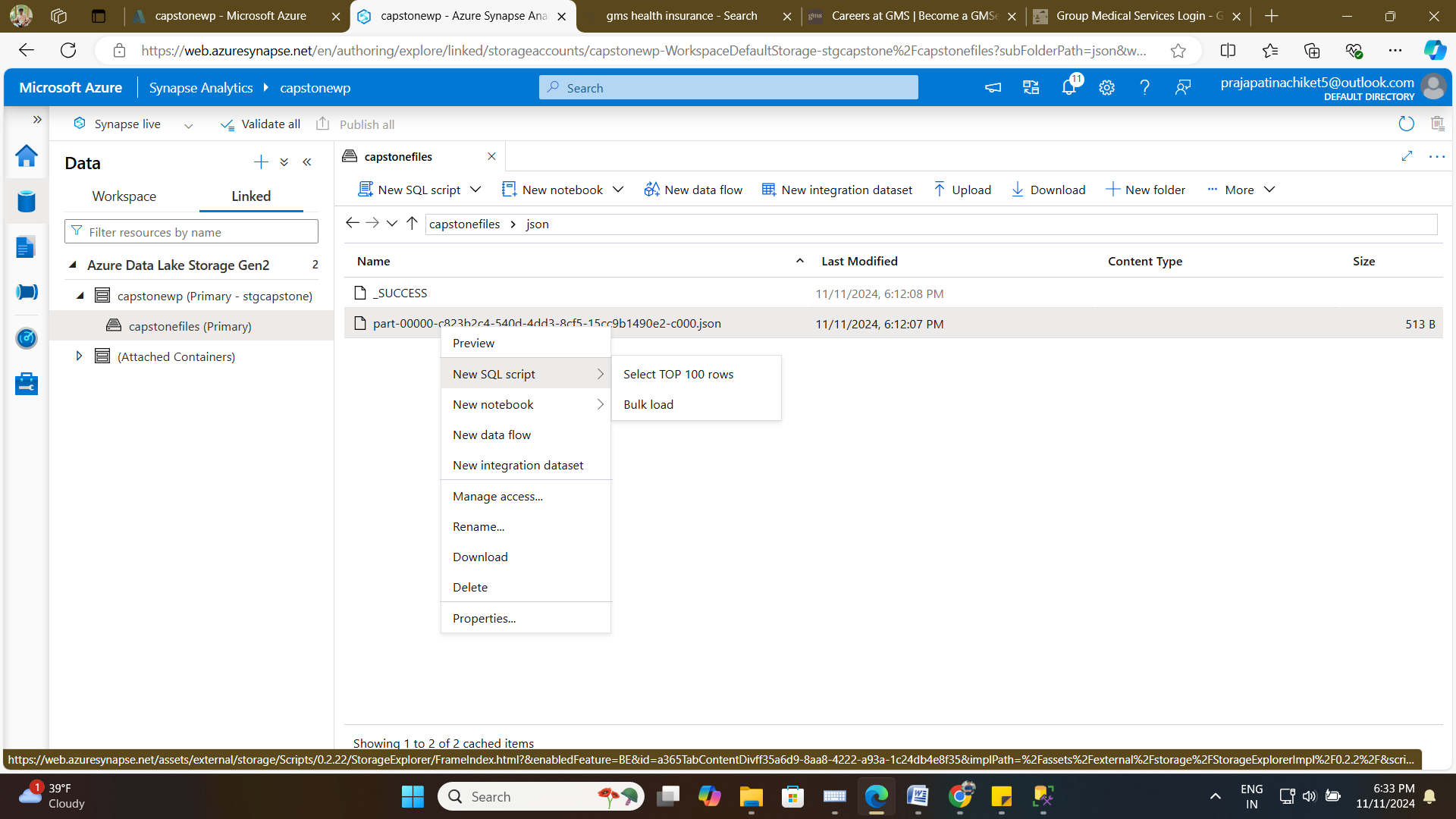


**Created external tables for Delta, Json, Parquet**

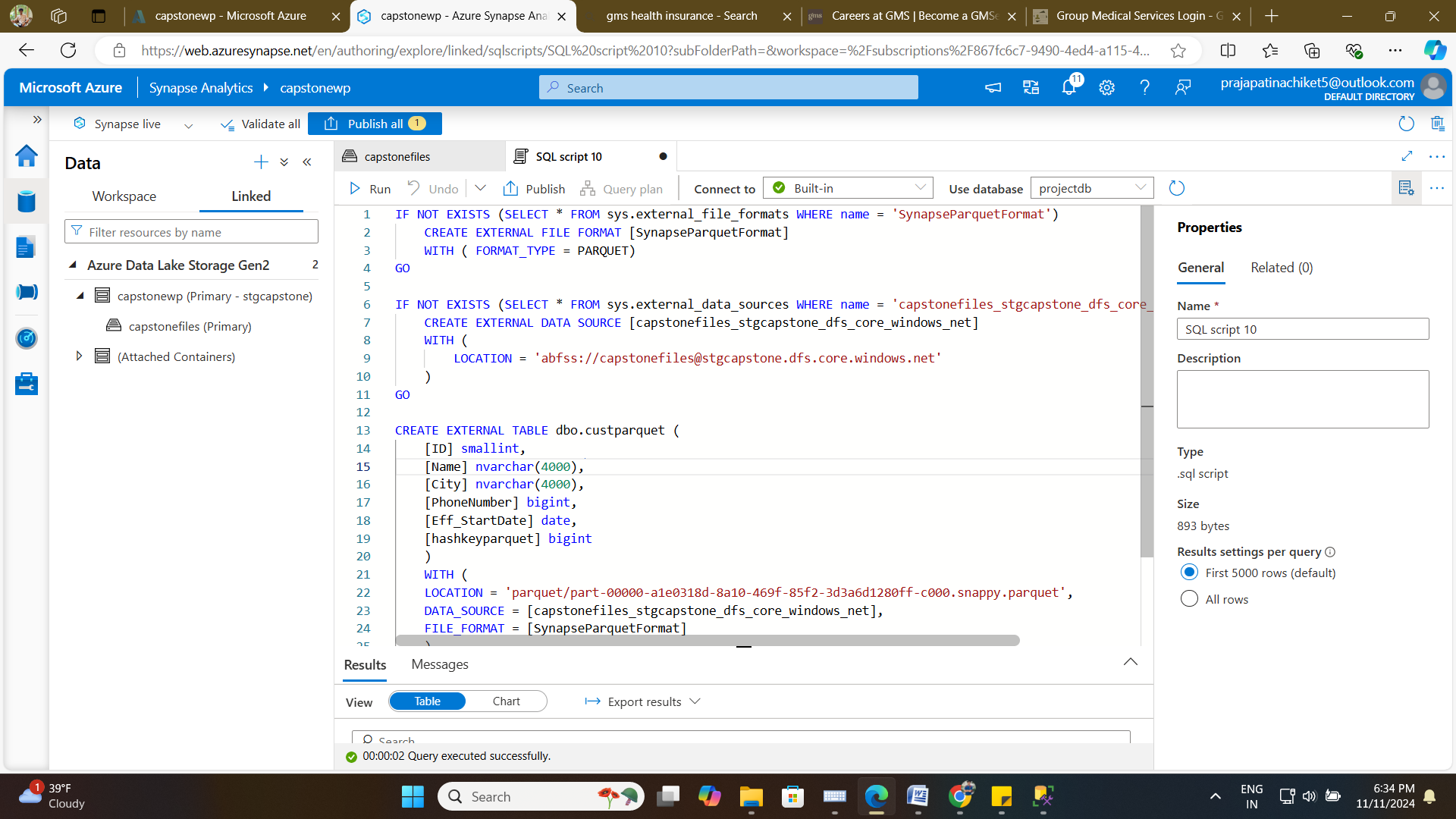
For delta



Unable to make external table for Json format

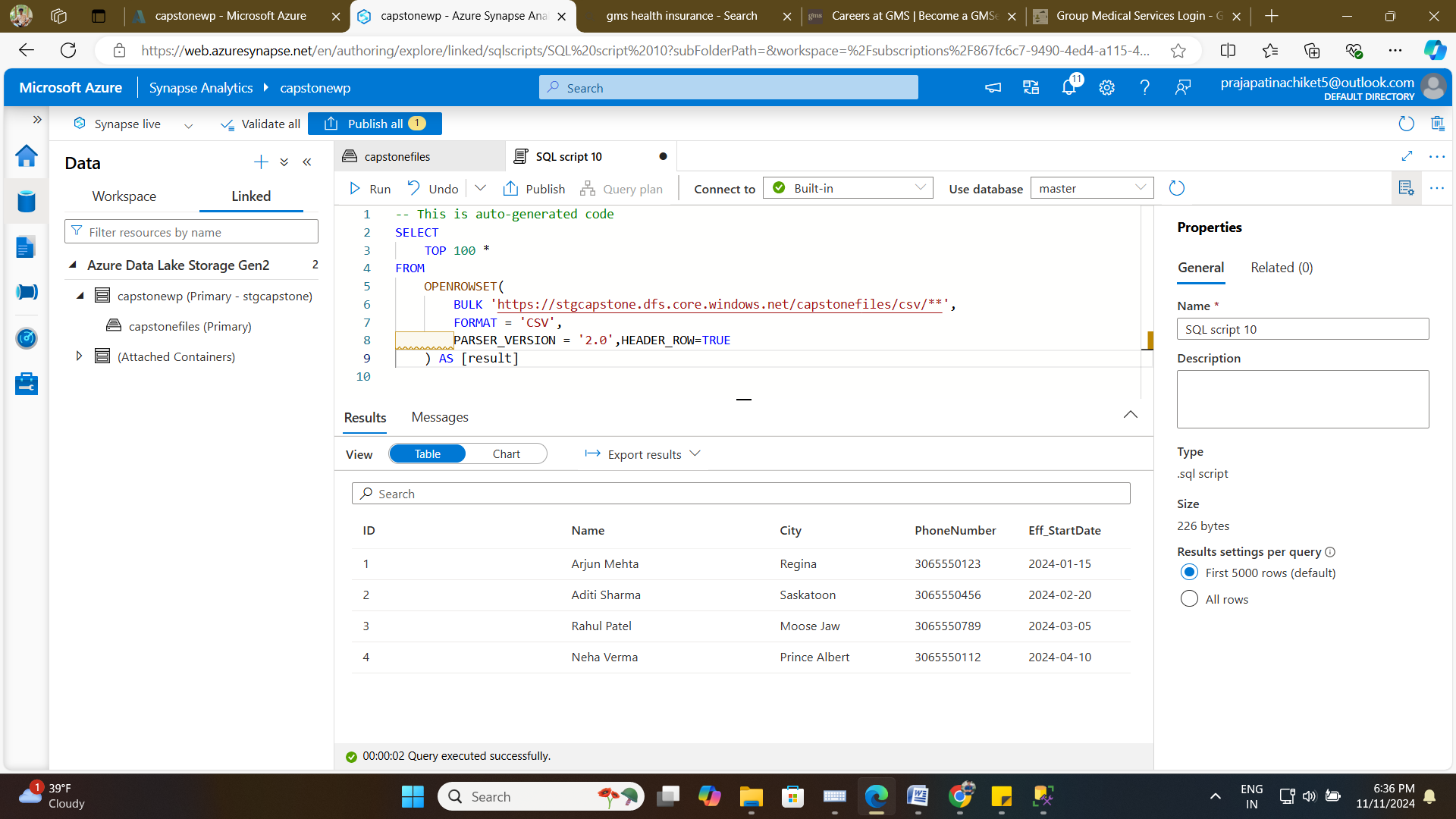


For Parquet

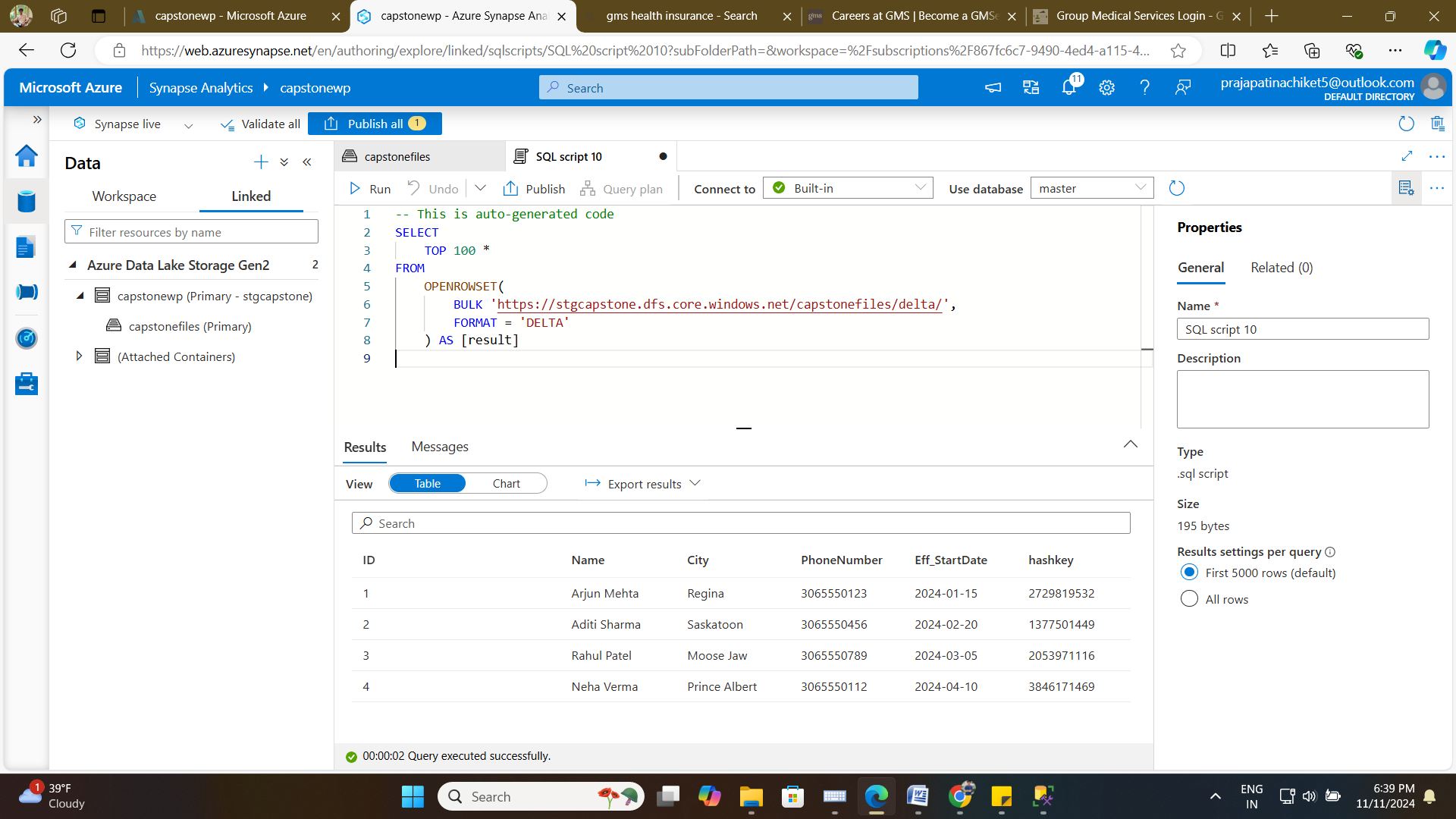


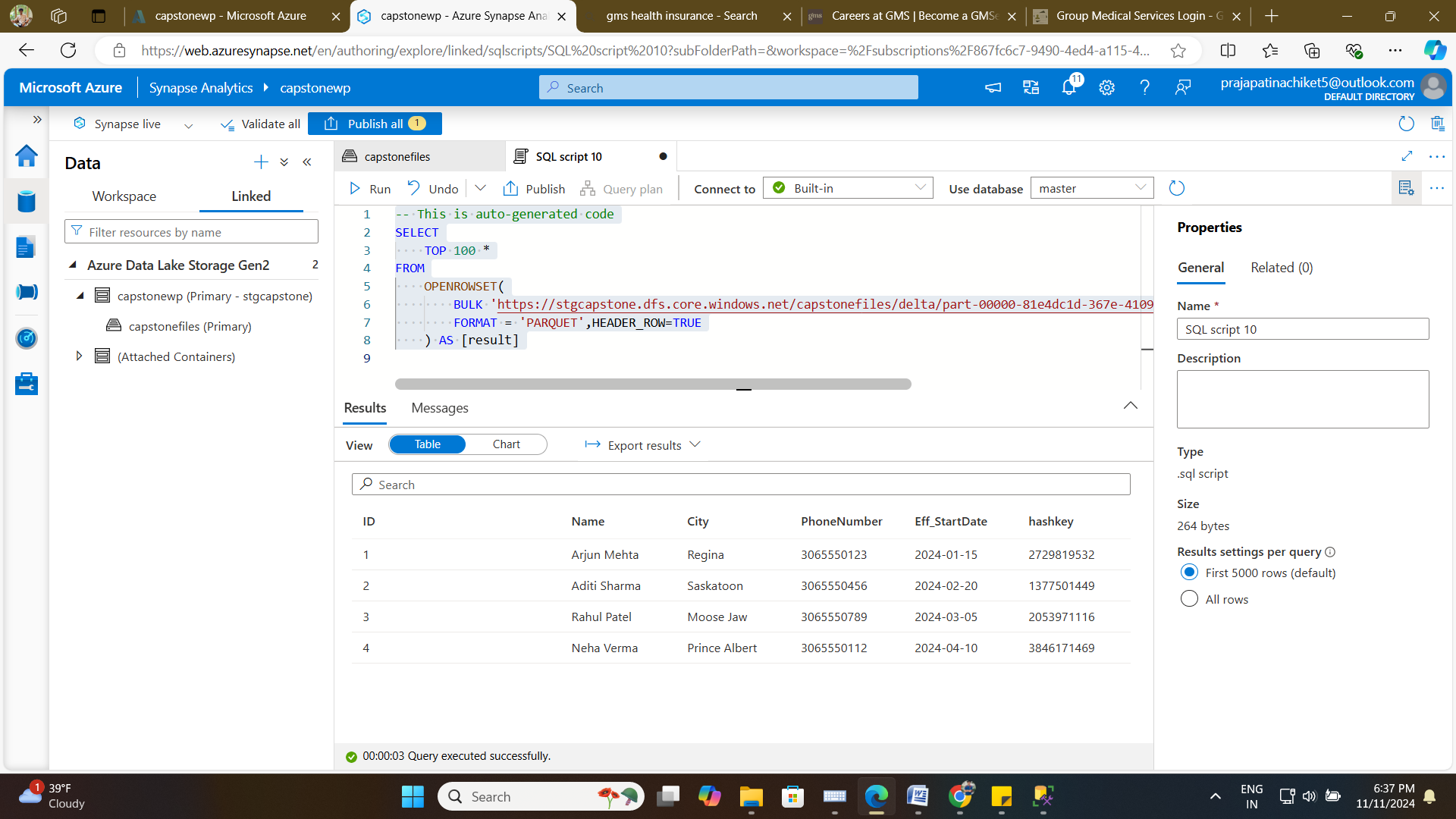
1. **Query Data with Wildcards:** Use wildcards to read data from multiple files in a folder or subfolder.

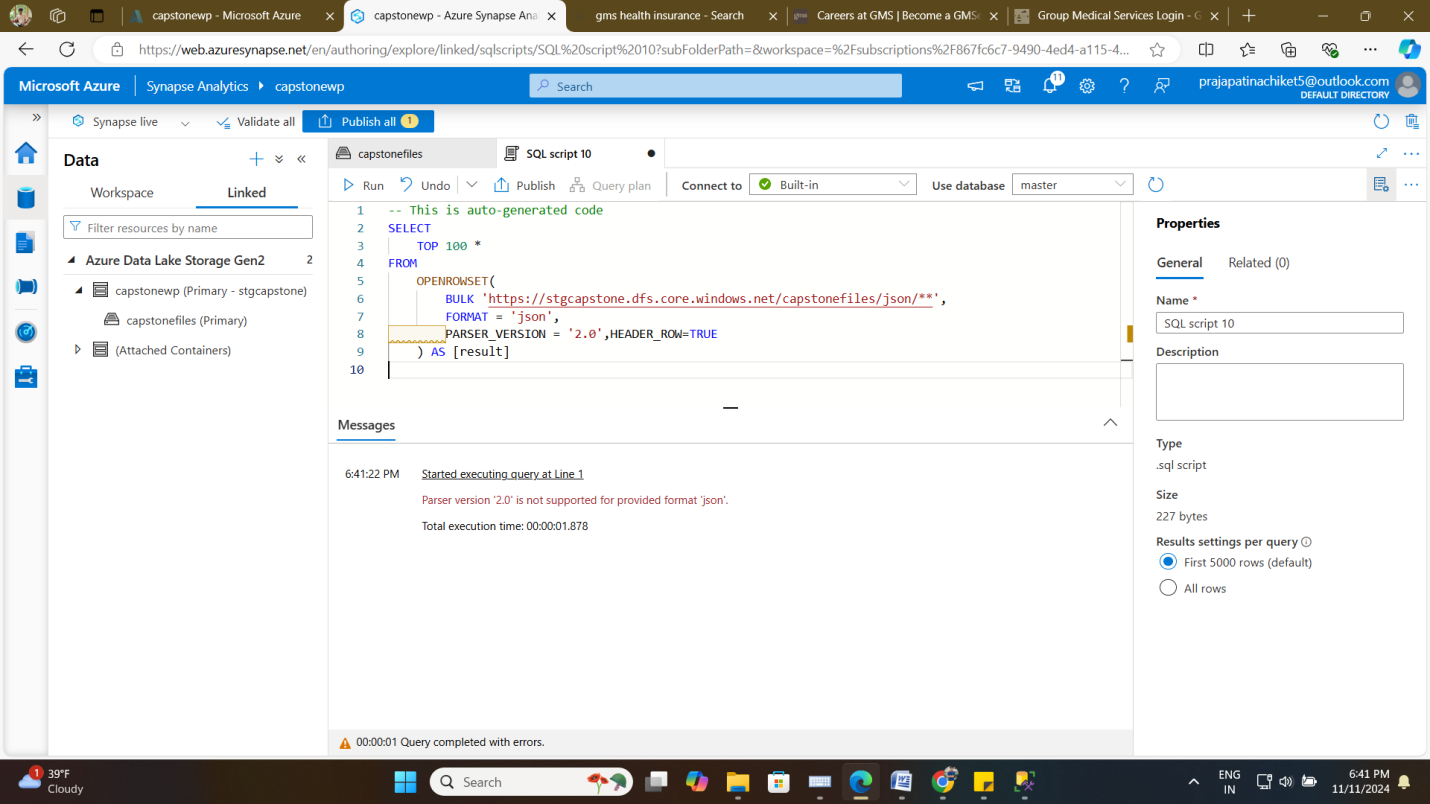
For csv:- used header row = true



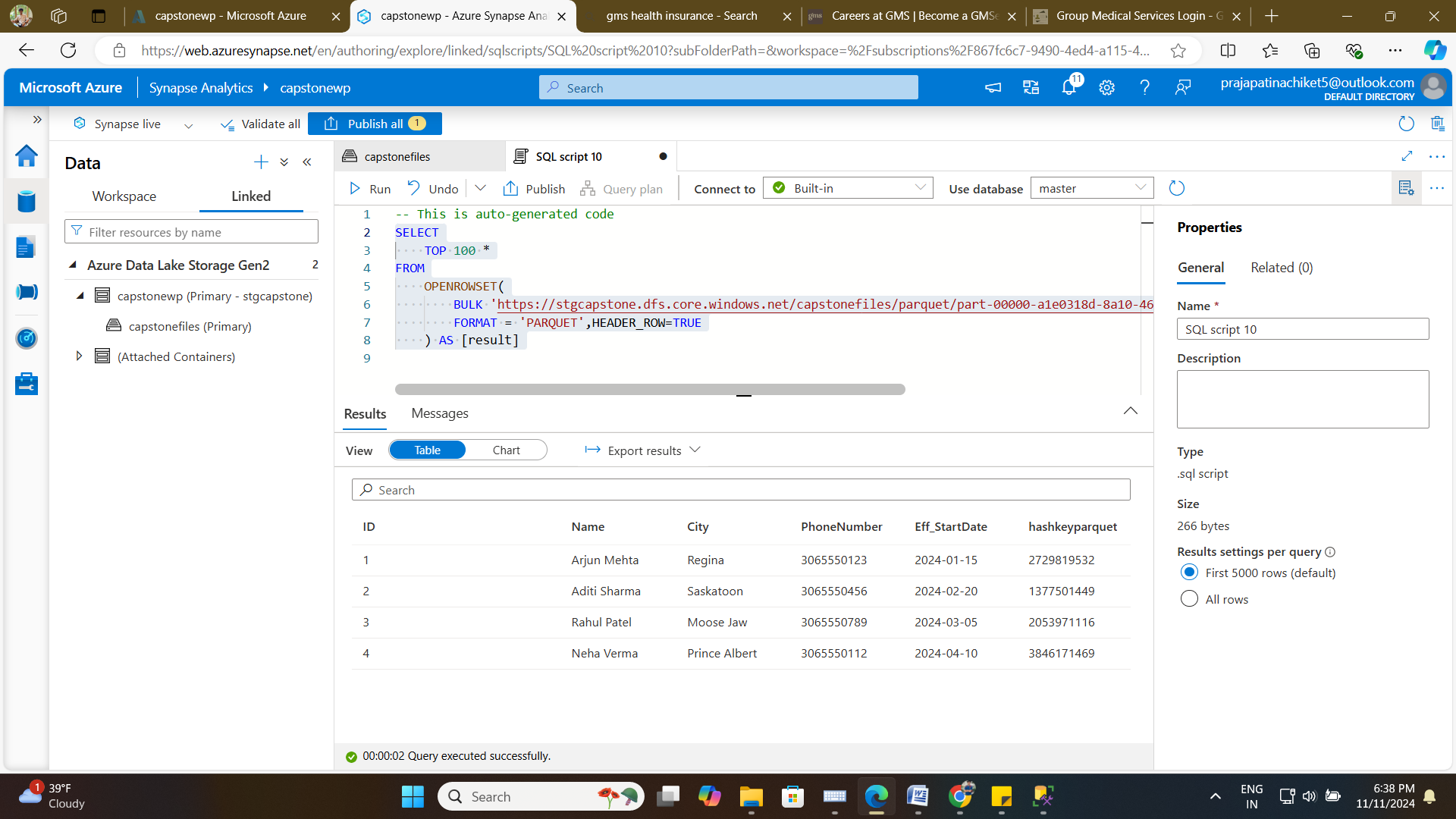
For delta:-

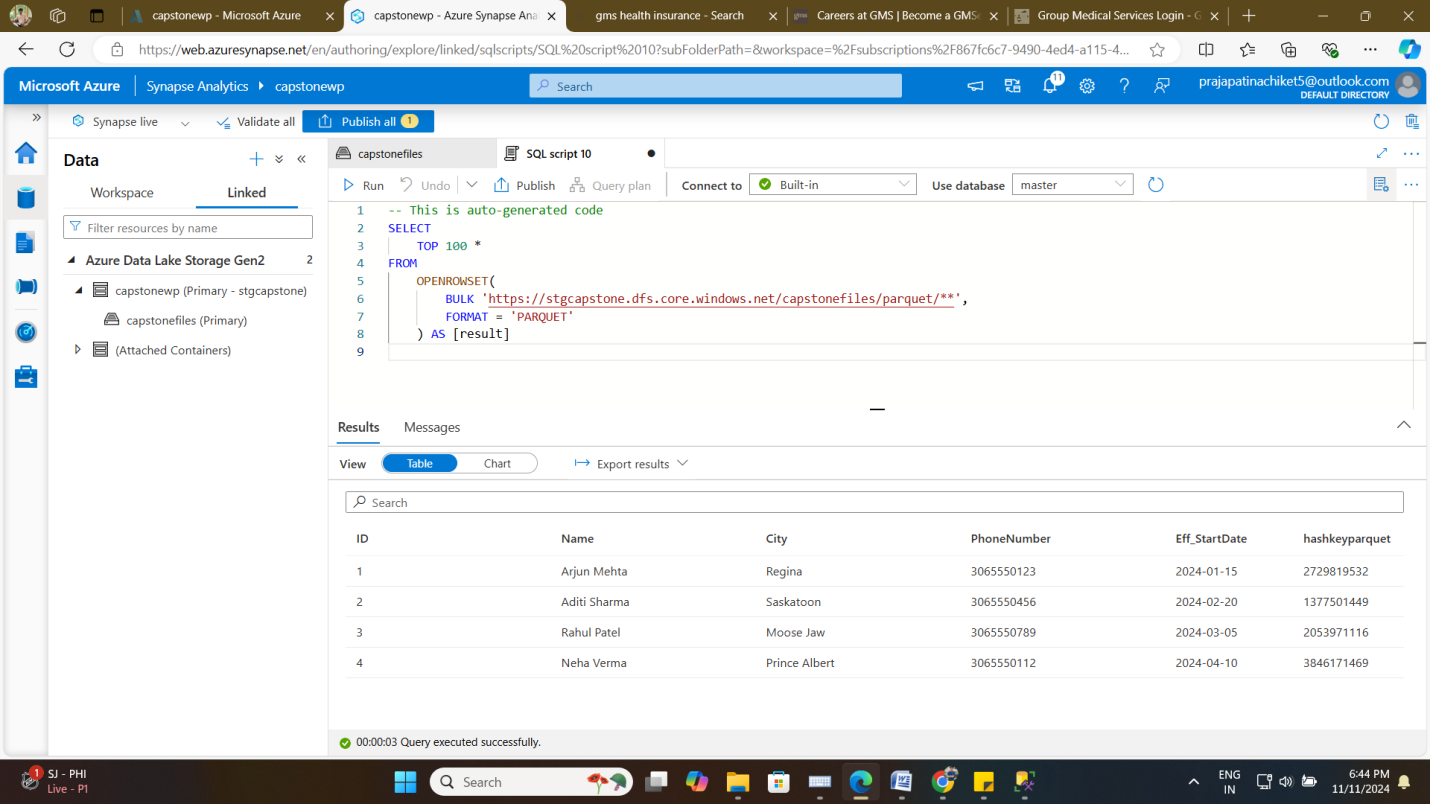


For Json:- 



For Parquet:-



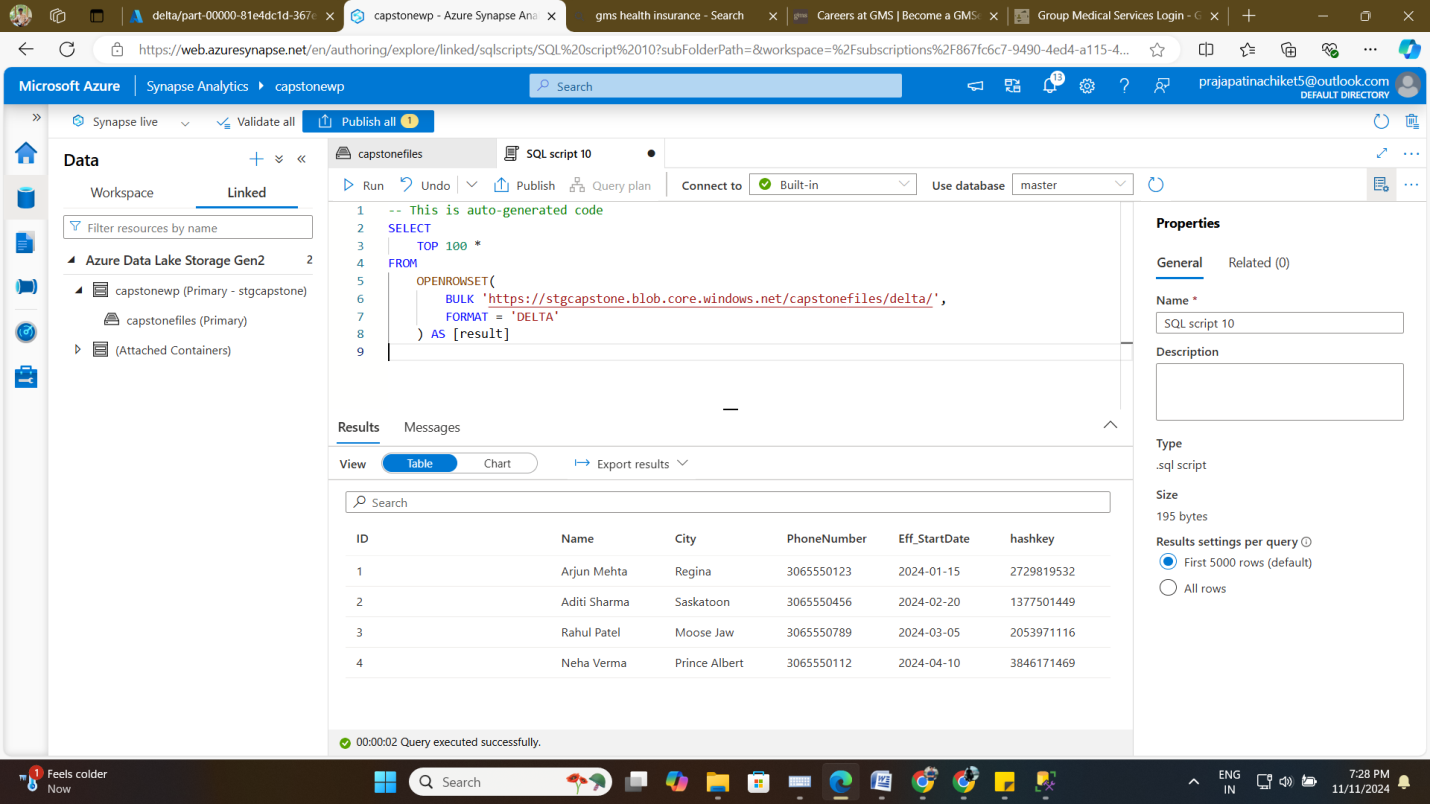


1. **Query Data by Filename and Filepath:** Read specific files based on their name and path.

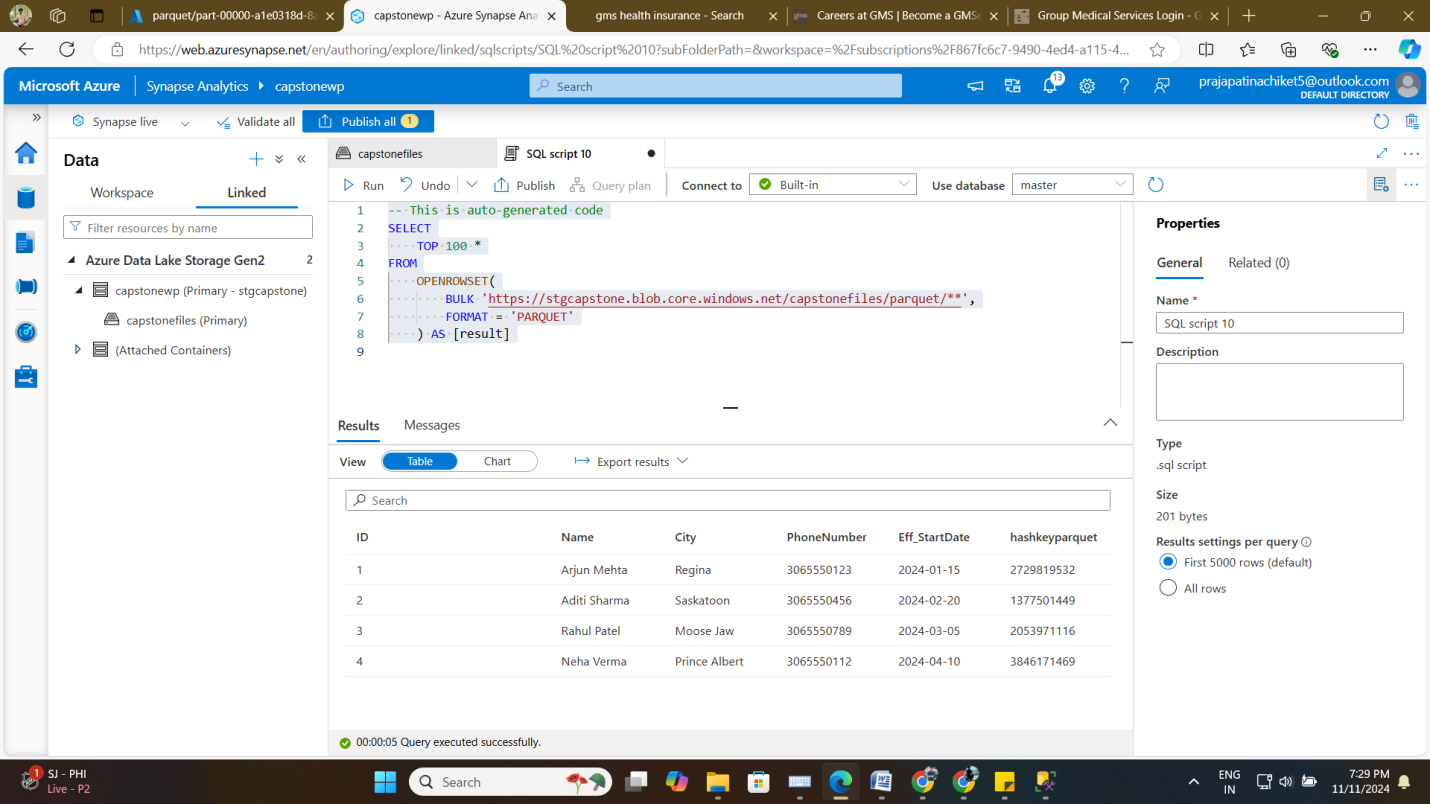
For csv file from storage account:-



For delta file format:-



For parquet filepath:-



Last SSMS view:-

