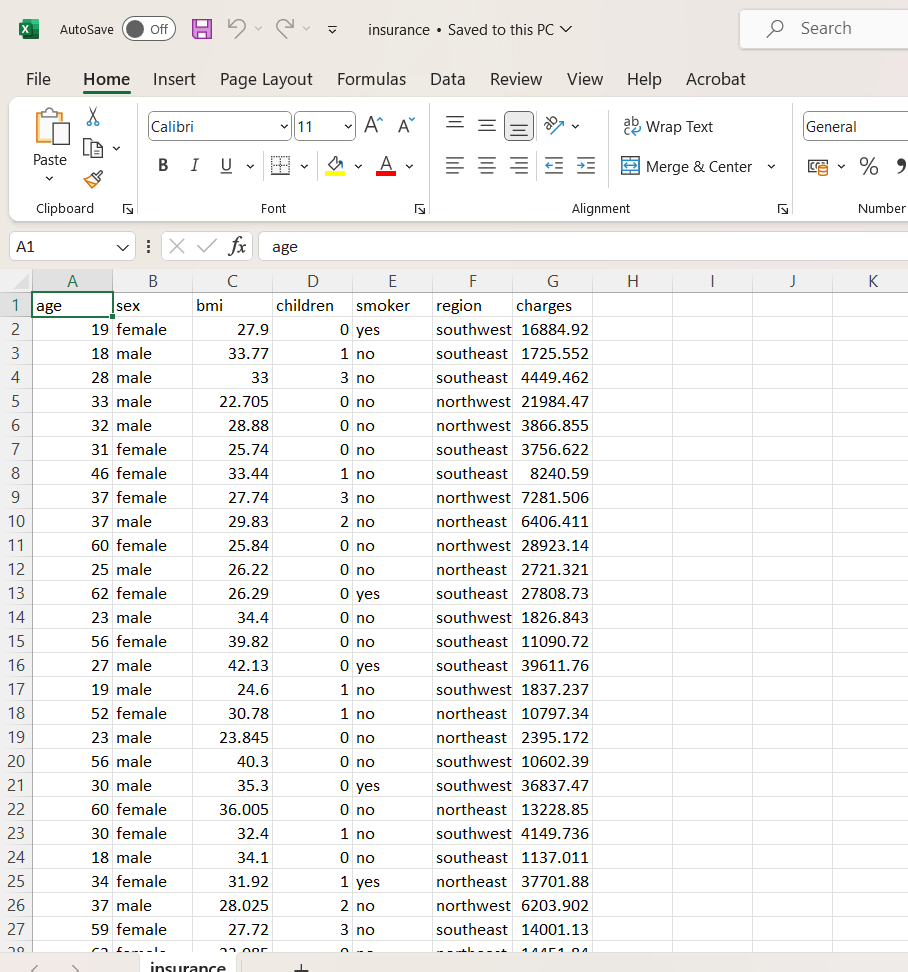
**Project Title: Insurance Analysis**

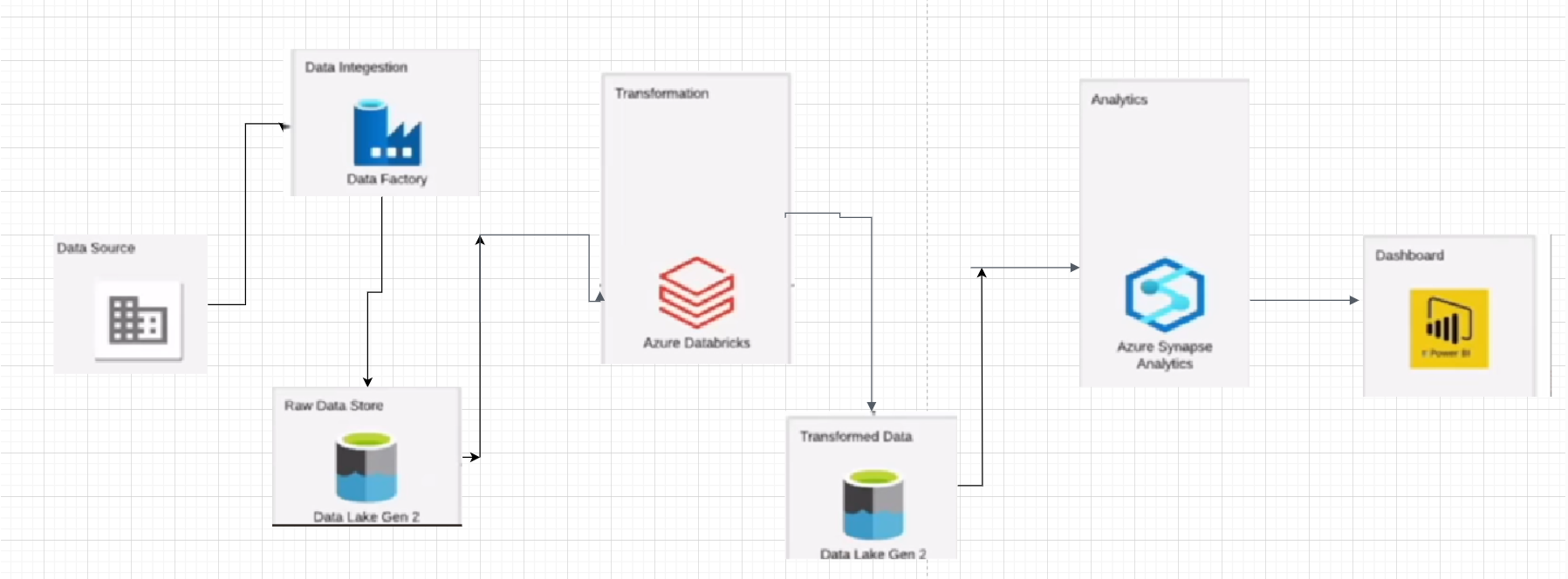
1. **Finding and Analyzing Data:**

Analysis is done on Insurance.CSV file which as data like Age,BMI,No of children,Smoker or Non-Smoker,Region,Sex,Charges.

Insurance.csv file I got through HTTPS server(backend team) .

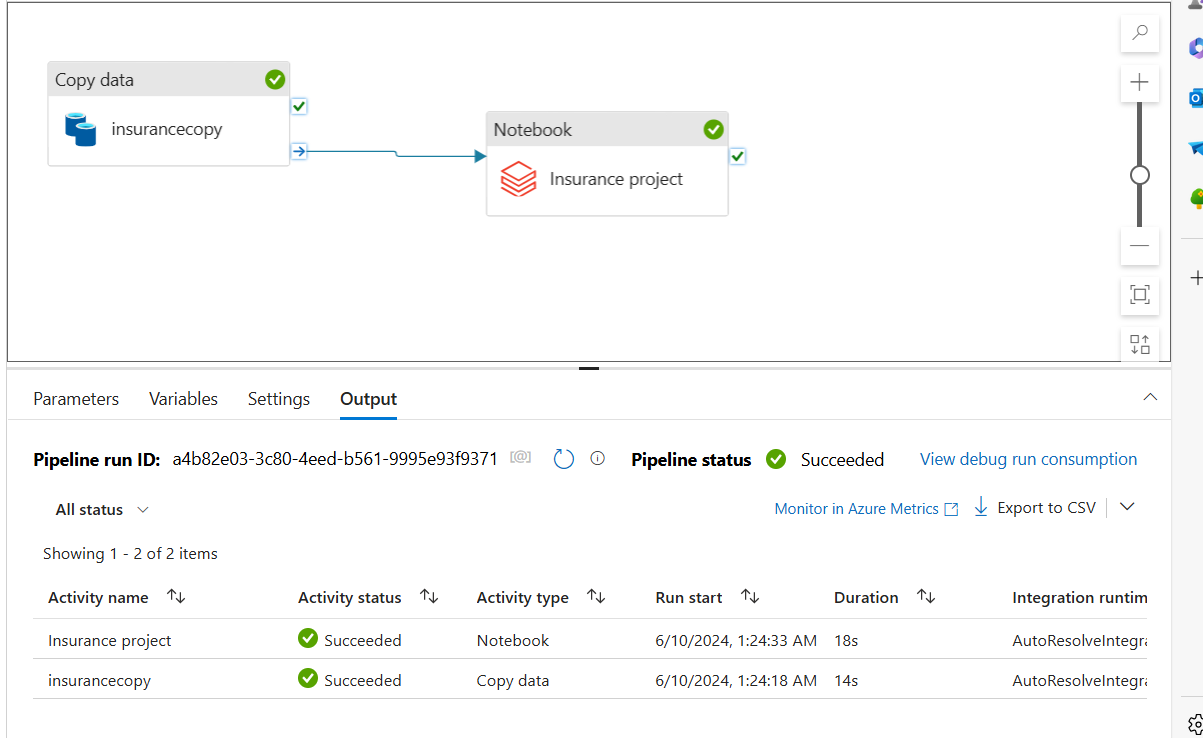


1. **Architectural Diagram: (Use tools like** [**https://draw.io/**](https://draw.io/)**)**



**3.Data Pipeline Creation:**

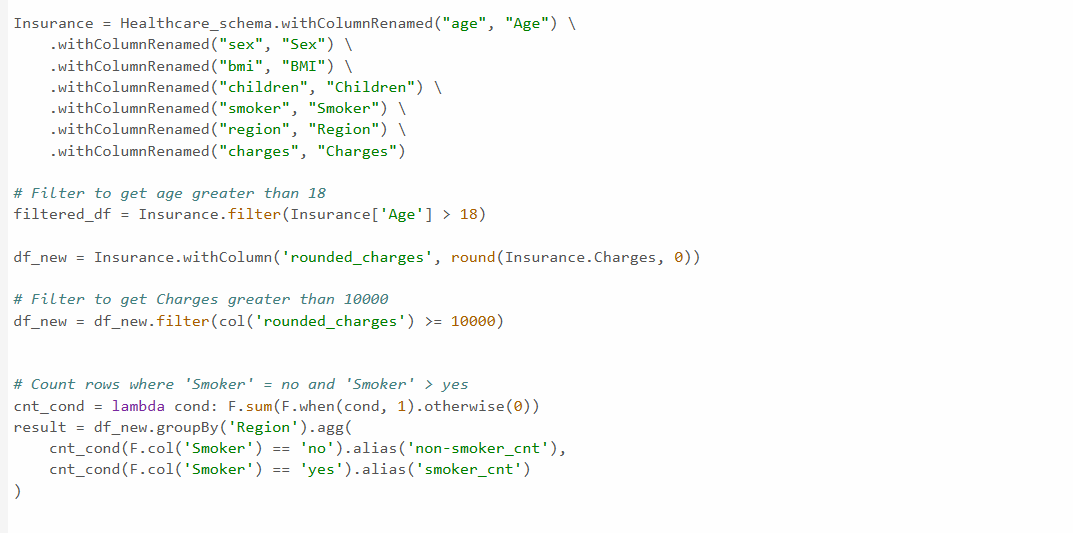
* From external datasource like HTTPS Server data is ingested into ADLS using copy activity in Data factory.
* Once the Copy activity is done , We perform transformations using Azure databricks notebook(pyspark Code)

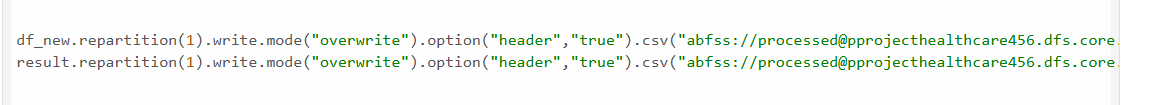


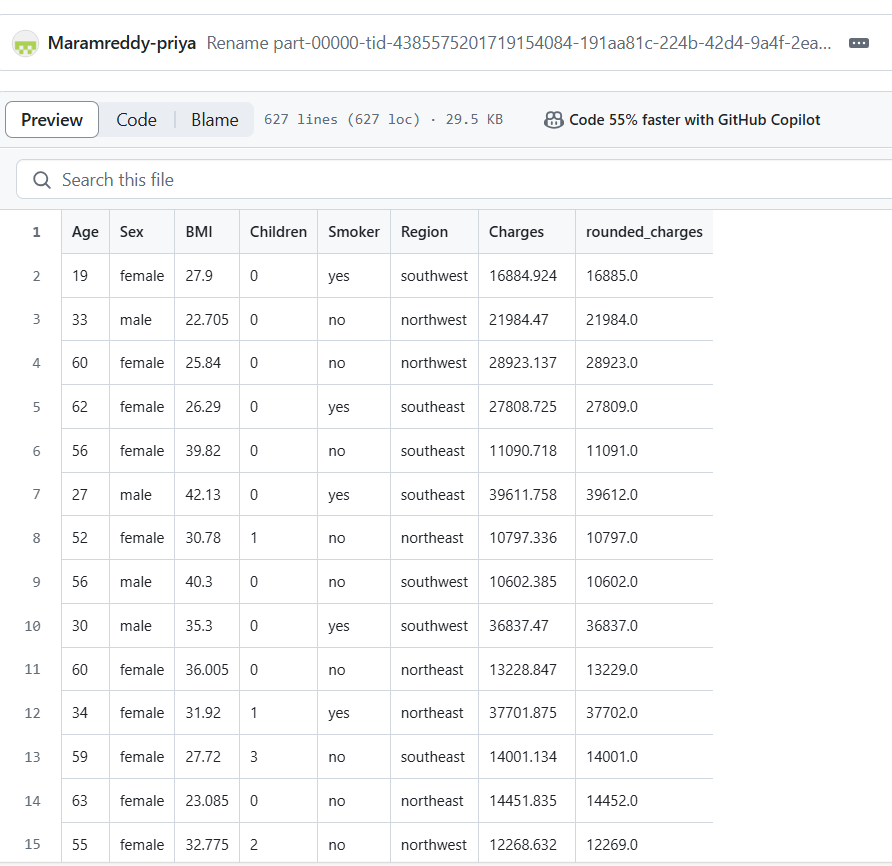
**4.Transformation and Analytics:**

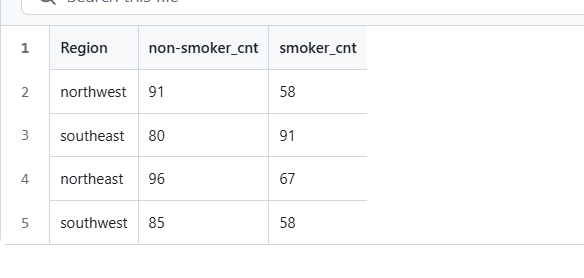
* Creating an Azure Data Lake Storage Gen2 storage account and containers named “Raw” and “processed” in ADLS Gen 2 storage.
* Creating an Azure Databricks workspace from Azure portal and creating a Spark cluster within that Azure Databricks workspace.
* Open a new notebook within Azure databricks for data cleaning and transformation.
* connecting one storage location (ADLS Gen2) to another within Azure Databricks using Access keys. This makes data stored in ADLS Gen2 accessible within the Databricks environment.
* Using PySpark , I performed data cleaning techniques which were removal of unnecessary columns and ensuring that all the columns have the correct datatype .
* Next step , I performed aggregation on the cleaned data using Pyspark .And I uploaded structured data to processed container. This Processed container holds structured format data which is suitable for reporting, visualization.
* Created Azure Synapse workspace , From there I created Lake database to access transformed data from processed container.
* Created DB,Table on that data performed analysis using SQL Queries.



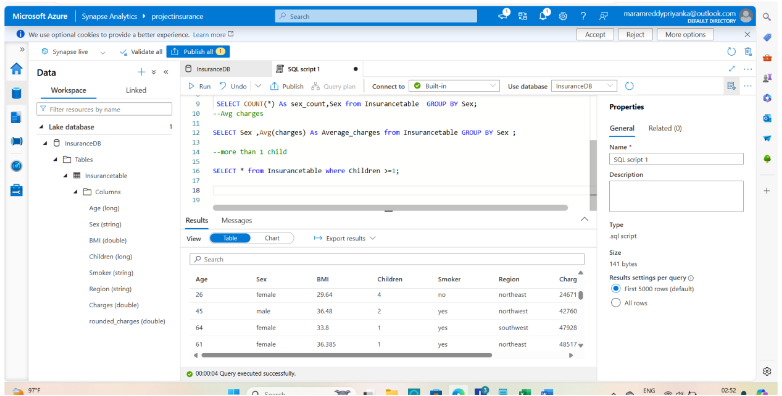


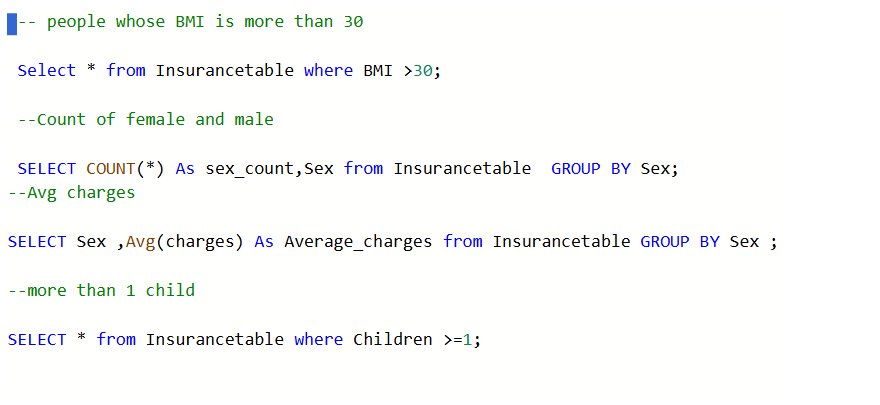




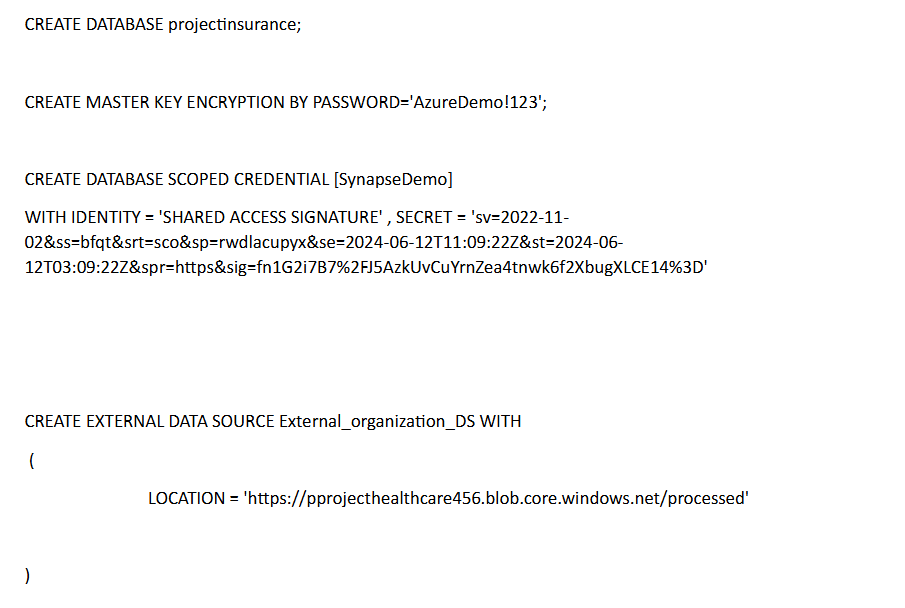


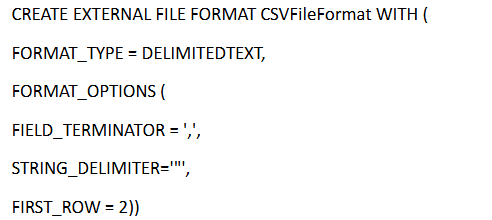
Azure synapse Analytics

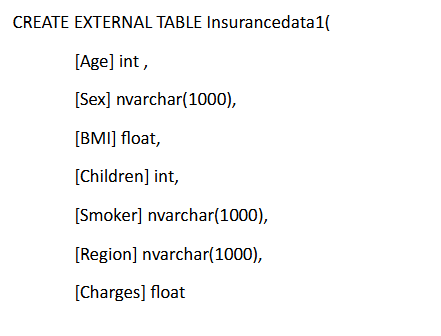


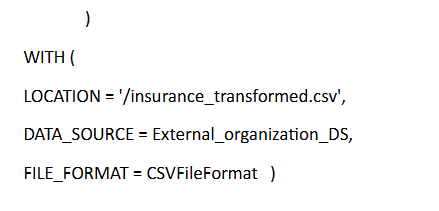


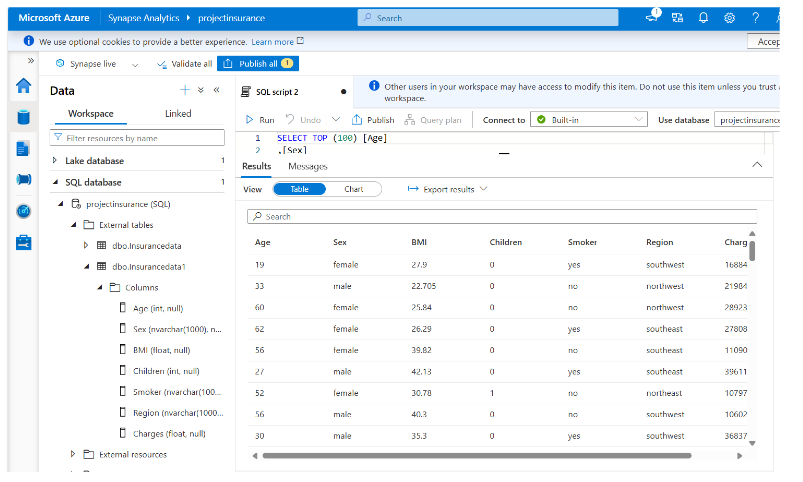
Azure Synapse Analytics(External Table)





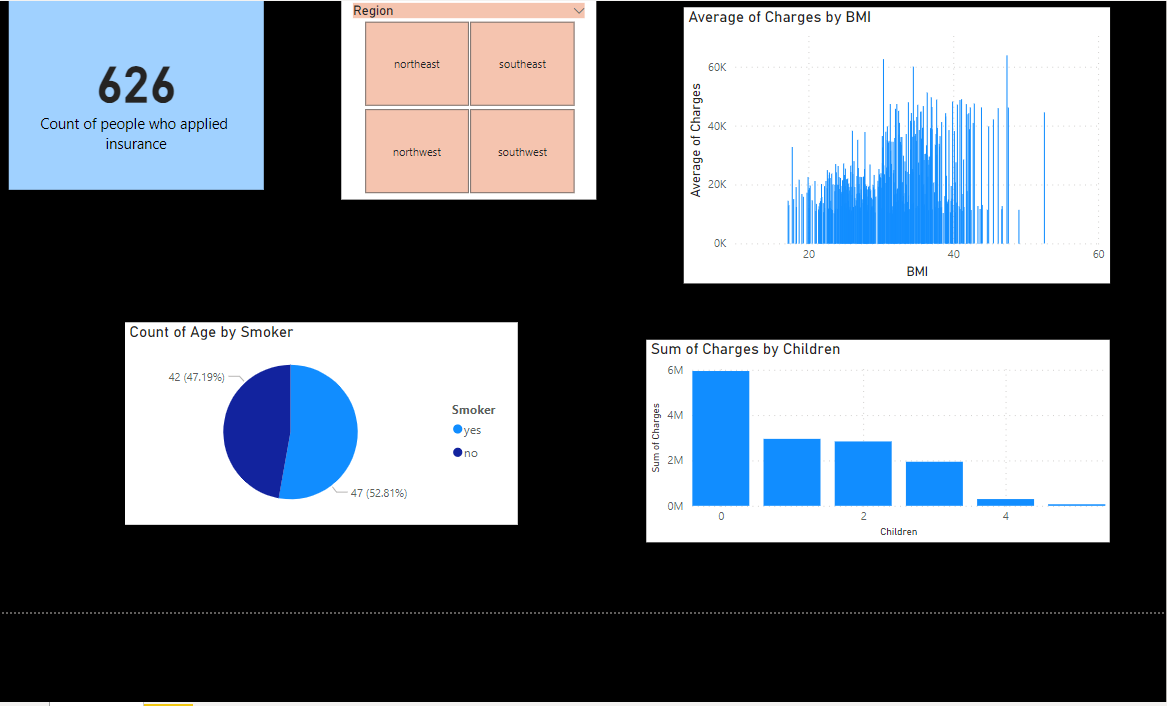






**5.Visualizations:**

* I Used Power BI to create bar charts and line graphs,cards,slicer depicting trends by connecting to azure Synapse analytics using Serverless end point URL.
* These visualizations offer insightful representations of the data.



**6.Conclusion:**

By following these steps, a robust ETL pipeline is built using Azure Databricks, Azure Data Factory and ADLS Gen2. This setup allows efficient data processing, transformation, and storage, facilitating advanced analytics and reporting using power BI and Azure Synapse.