**DAMG 7275 – P3 Assignment**

**Analyzing Layoff Data and Stock**

**Performance: A SQL-Graph Hybrid Model**

**Team – 3**

**Data Preprocessing:**

We've obtained the necessary layoff dataset from the website [https://layoffs.fyi](https://layoffs.fyi/) for our project. To prepare the data for analysis, we're conducting preprocessing tasks using Python within a Jupyter notebook environment. Our dataset encompasses various facets of layoff information, and to organize it effectively, we've structured it into six distinct tables as delineated in the Entity-Relationship Diagram (ERD) for our relational database implementation.

As part of our preprocessing pipeline, we've carefully extracted relevant data points from the raw dataset, ensuring accuracy and consistency. These data points are then transformed and formatted to adhere to the specifications outlined in our project requirements. The preprocessing steps involve cleaning the data, handling missing values, and transforming data types as necessary to facilitate downstream analysis.

In addition to the Python notebook containing our preprocessing code, we've included the CSV files that we extracted from the original dataset. Furthermore, we've incorporated supplementary CSV and Excel files that augment our analysis, enriching our dataset with additional contextual information where applicable. These files serve as valuable resources in our endeavor to derive meaningful insights from the layoff data.

Screenshots:

Company Table:

A screenshot of a computer

Description automatically generated

Industry Table:

A screenshot of a computer

Description automatically generated

Region Table:

A screenshot of a computer

Description automatically generated

Company-Region Table:

A screenshot of a computer

Description automatically generated

Person Table:

A screenshot of a computer

Description automatically generated

Layoff event Table:

A screenshot of a computer

Description automatically generated

**Data Storage:**

After preprocessing the layoff dataset and obtaining the necessary CSV files, we proceeded with storing the data in Azure Blob Storage and creating a pipeline in Azure Data Factory for data ingestion.

* We first uploaded the CSV files containing the preprocessed layoff data and supplementary files to Azure Blob Storage, using terminal.

A screenshot of a computer

Description automatically generated

A black screen with white text

Description automatically generated

A screenshot of a computer program

Description automatically generated

* Two separate containers were created in Azure Blob Storage to organize the data: one named "layoffdata" and another named "stockdata".

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Data Ingestion:**

* In Azure Data Factory, linked services were created for Azure Blob Storage, Azure SQL Database, and Azure Cosmos DB to establish connections for data movement.
* Datasets were defined in Azure Data Factory for each CSV file stored in Azure Blob Storage. Two separate datasets were created for the layoff data and stock data, specifying the path to the respective CSV files in their corresponding containers.
* A pipeline was created in Azure Data Factory to orchestrate the data ingestion process.
* Within the pipeline, copy data activities were added:
* The first few copy data activities was configured to copy data from the "layoffdata" container in Azure Blob Storage to the Azure SQL Database.
* Another seprate data activity was configured to copy data from the "stockdata" container in Azure Blob Storage to Azure Cosmos DB.
* The pipeline was tested to ensure that data ingestion worked as expected.
* Parameters such as scheduling, data movement, and error handling were configured as per project requirements.

Some of the screenshots

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

By following these steps, the team successfully orchestrated the data ingestion process from Azure Blob Storage to Azure SQL Database for layoff data and to Azure Cosmos DB for stock data using Azure Data Factory. This setup allows for seamless movement of data across different Azure services, facilitating further analysis and insights derivation from the layoff dataset.