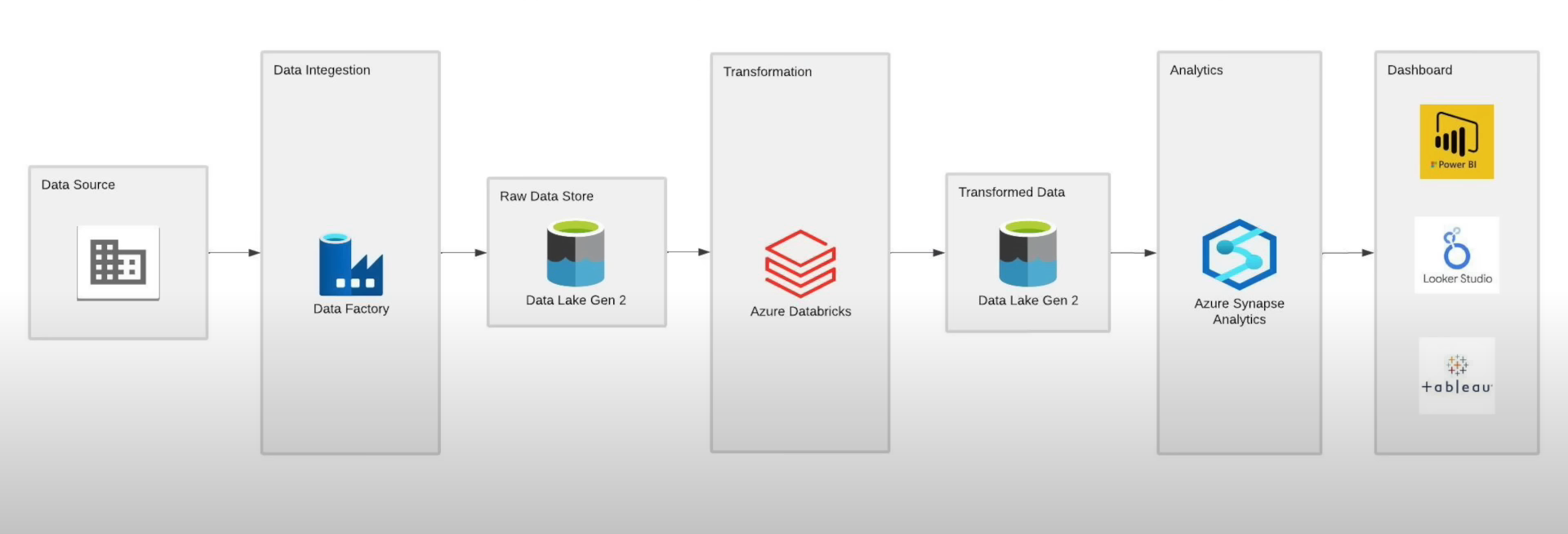
**End To End Sales Data Analytics Project on Azure**

**I. Introduction**

**A. Project Overview**

This sales data analytics project aims to analyse and derive insights from a dataset obtained from Kaggle. The project involves various stages, including data extraction, transformation, loading (ETL), and analysis using Azure services. The primary goal is to create a comprehensive sales analysis dashboard using Power BI.

**Project Architecture**

****

**II. Data Acquisition**

**A. Dataset Selection**

Source: [Kaggle](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce)

Contents: 9 CSV files containing diverse sales-related data.

**B. Data Download and Storage**

**Download Process**: direct download.

**Storage**: Local storage for initial processing.

**III. Data Extraction and On-Premise Database**

**A. ETL Pipeline in Python**

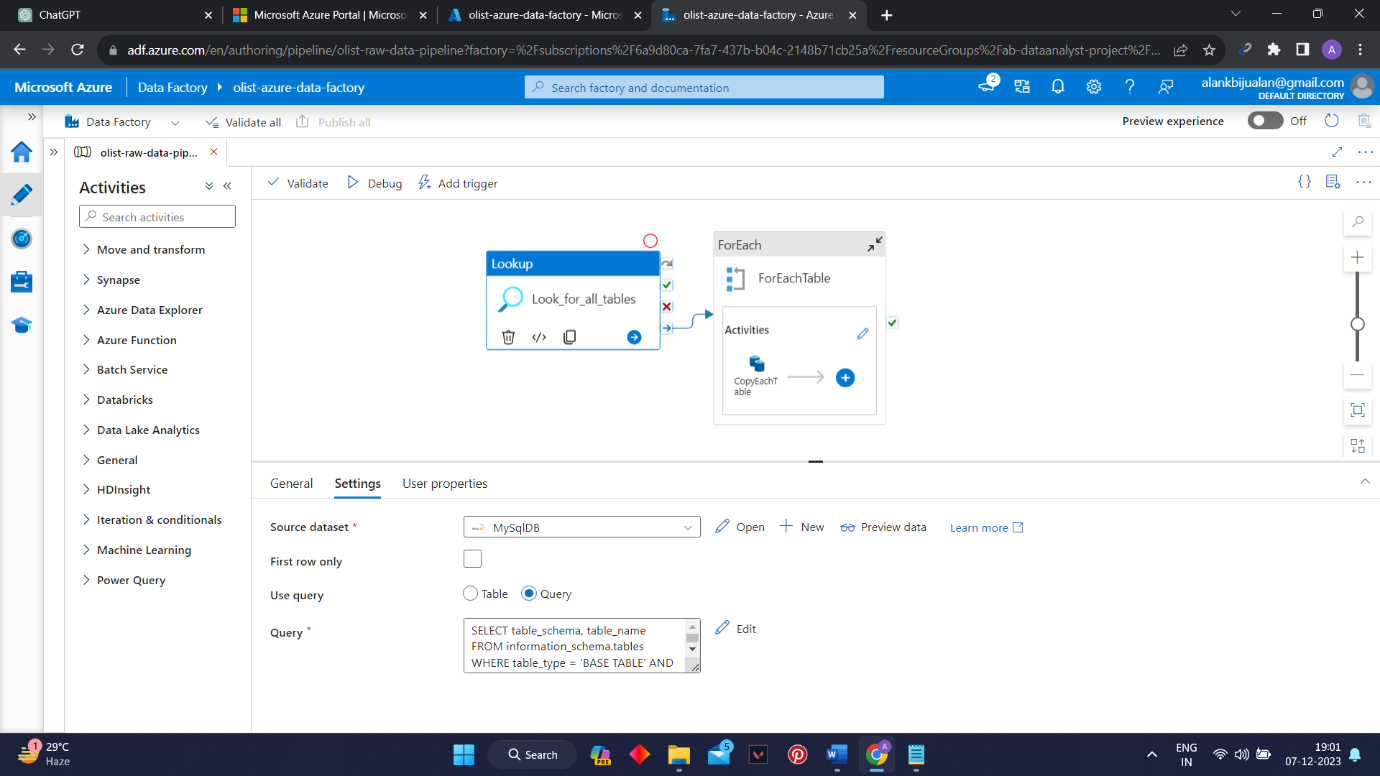
**Pipeline Development**: Developed a Python ETL pipeline for data extraction from CSV files. (Click here)

**MySQL Database**: Setup an on-premise MySQL database to store extracted data.

**IV. Cloud Migration with Azure Data Factory**

**A. Data Movement to Azure**

**Azure Data Factory**: Configured ADF for ETL pipeline orchestration.

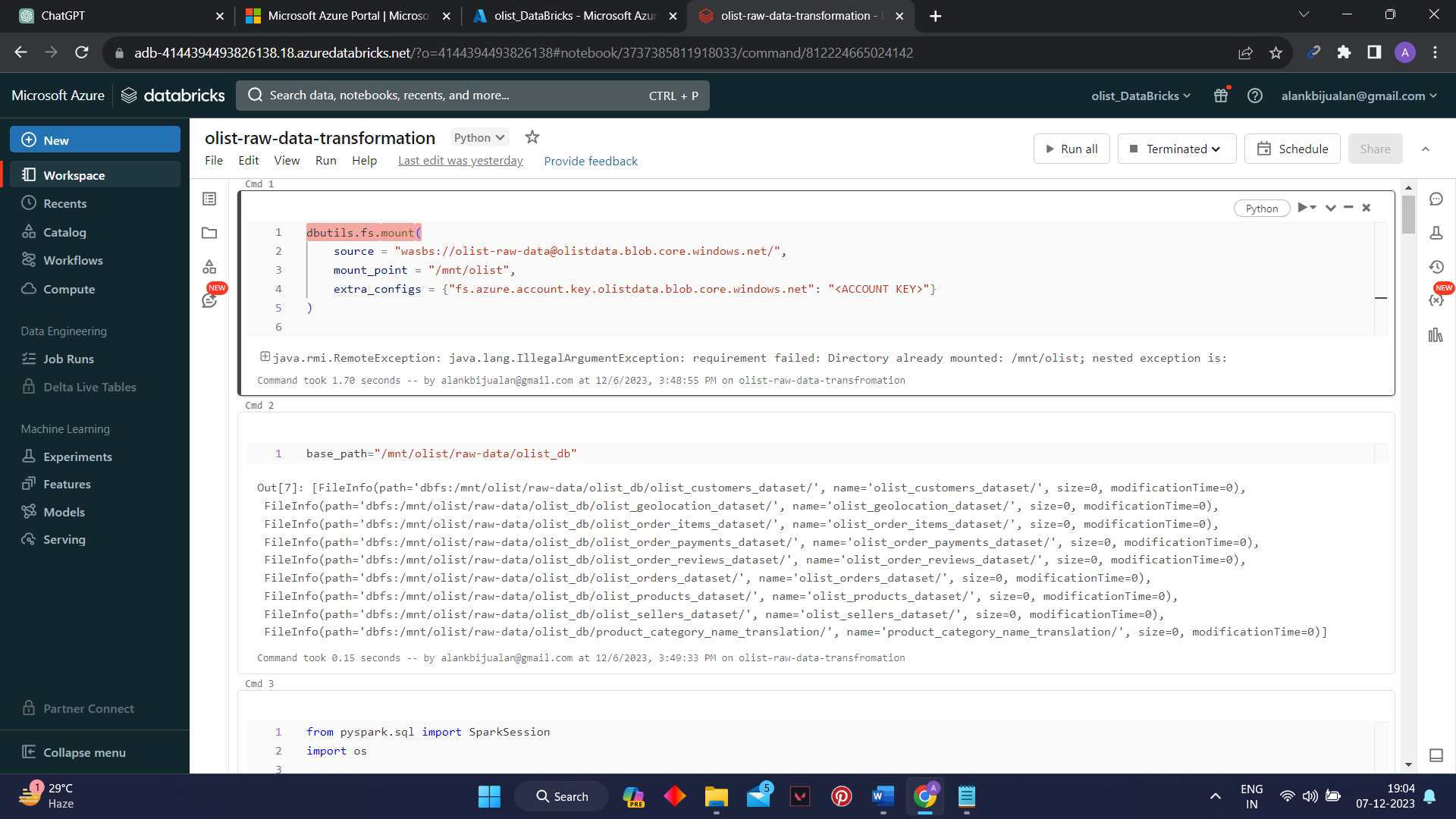


**MySQL to Azure Data Lake Gen2**: Extracted data from MySQL and stored it in Azure Data Lake Gen2

**V. Data Transformation with Databricks**

**A. Cleaning and Transformation**

**Azure Databricks**: Leveraged Databricks for data cleaning and transformation.



**(For the transformation code you can check ‘raw data transformation.py’ file)**

**Output Storage**: Cleaned data saved in Azure Data Lake Gen2 under 'cleaned data' directory.

**VI. Synapse Analytics and Data Warehousing**

**A. Serverless SQL Database**

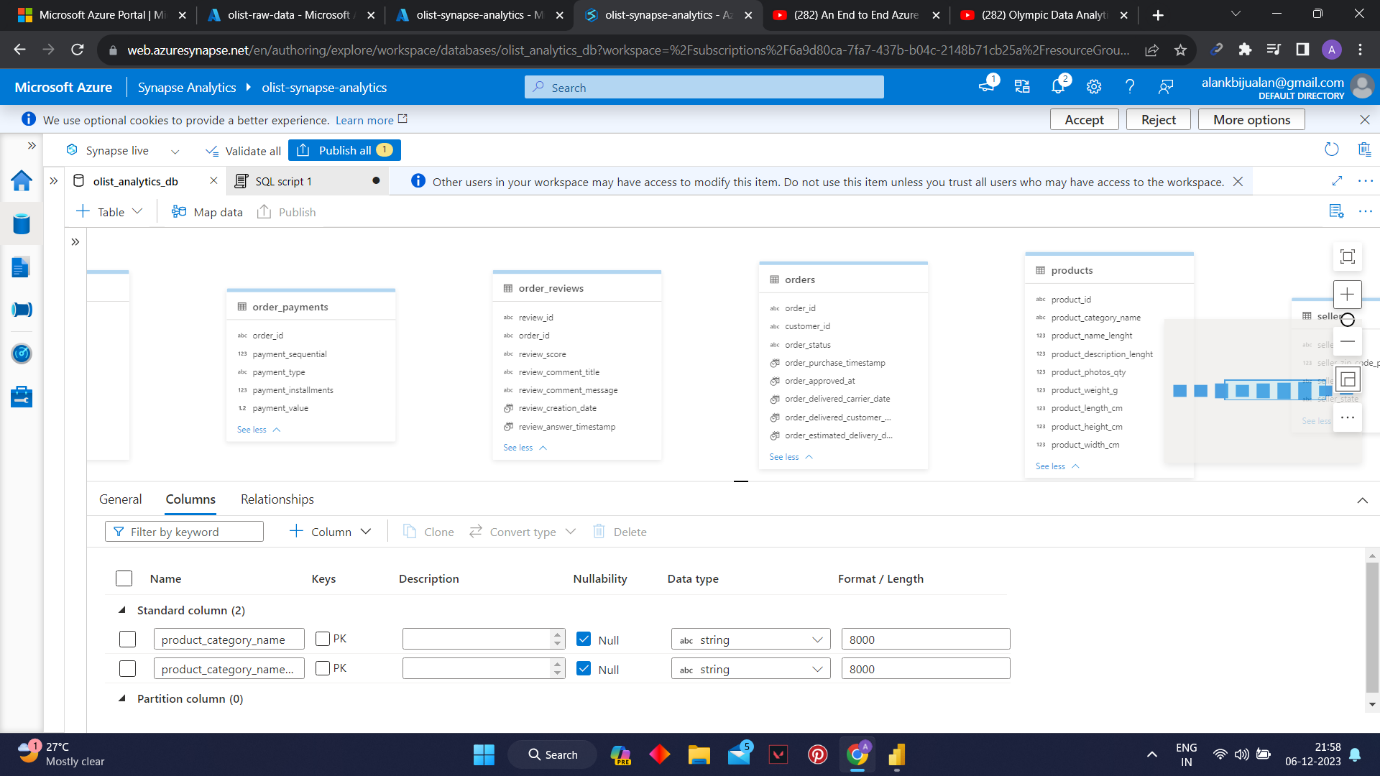
**Azure Synapse Analytics**: Established a serverless SQL database on Azure Data Lake.

**Data Warehousing**: Created a data warehouse for efficient analysis.

**B. Data Modelling in Synapse Analytics Studio**

**Table Creation**: Defined tables in Synapse Analytics Studio for each dataset.

**Schema Design**: Ensured optimal schema design for analytical queries.

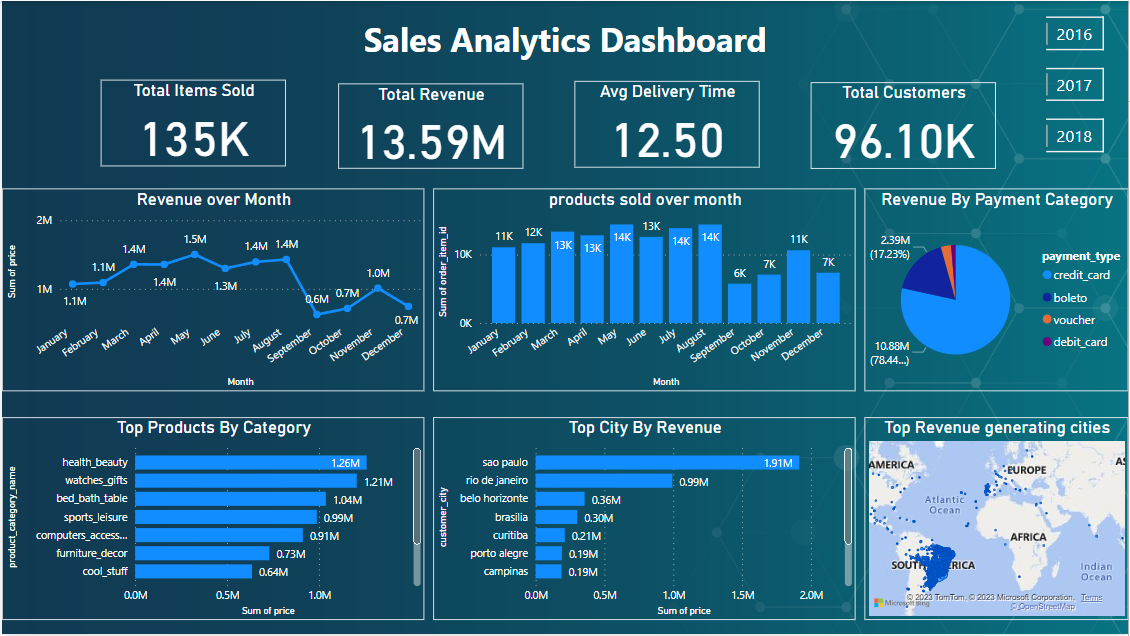


**VII. Power BI Visualization**

**A. Dashboard Creation**

**Connection Setup**: Connected Power BI to the Azure Synapse Analytics data.

**Data Visualization**: Developed interactive dashboards for sales analysis.



**(For detailed visualisation report, you can check the Power BI file)**

**VIII. Conclusion**

**A. Project Achievements**

* Successfully migrated and transformed data from Kaggle to Azure.
* Established an end-to-end analytics pipeline, incorporating various Azure services.
* Developed an insightful sales analysis dashboard using Power BI.