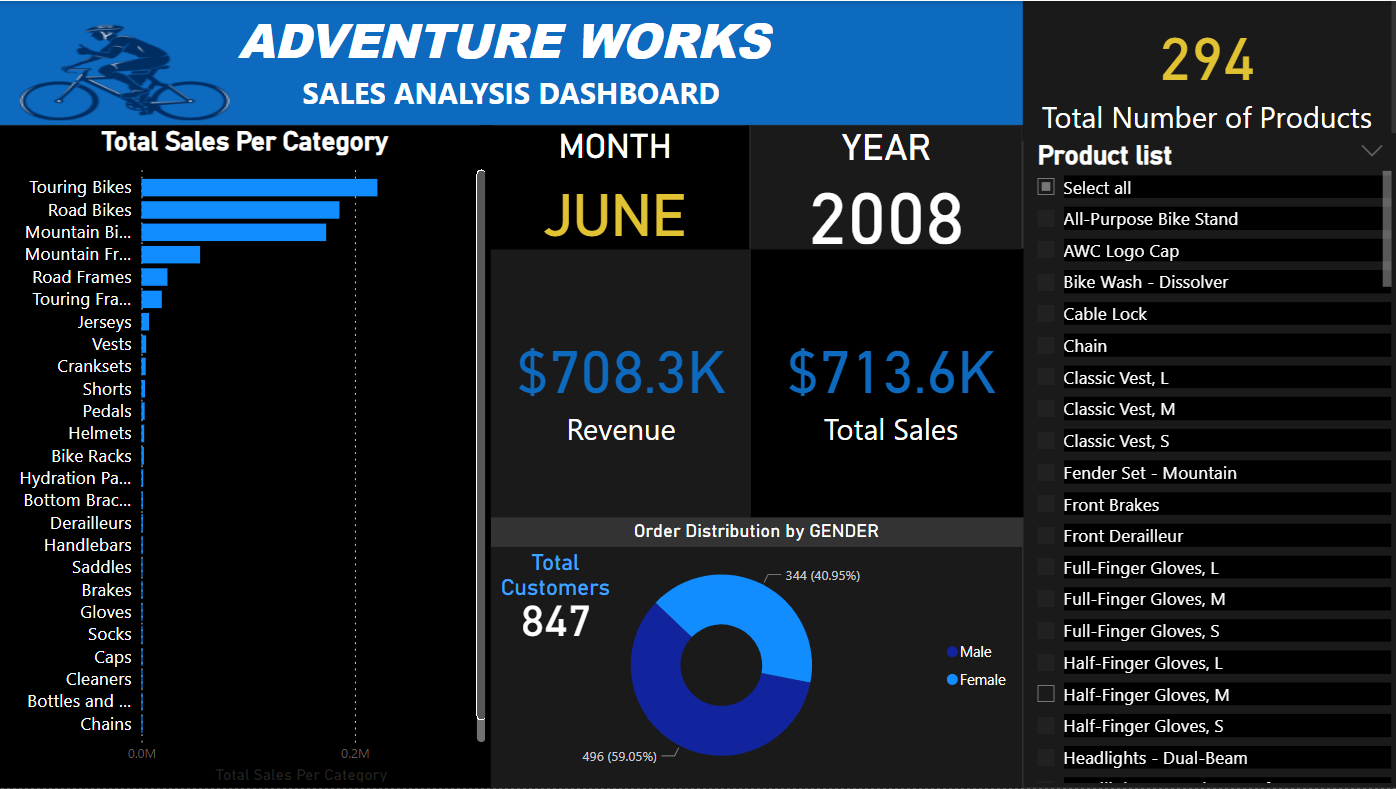
# Adventure Works Sales Analysis



On-prem DB to Azure Cloud Pipeline with Data Factory, Lake Storage, Spark, Databricks, Synapse, PowerBI

📝 Table of Contents

Project Overview

Key Insights

Project Architecture

3.1. Data Ingestion

3.2. Data Transformation

3.3. Data Loading

3.4. Data Reporting

Credits

Contact

🔬 Project Overview

This is an end-to-end data engineering project on the Azure cloud. Where I did data ingestion from an on-premise SQL Server to Azure Data Lake using Data Factory to transformation using Databricks and Spark, loading to Synapse, and reporting using PowerBI. Also, I used Azure Active Directory (AAD) and Azure Key Vault for data monitoring and governance purposes.

💾 Dataset

AdventureWorks is a database provided by Microsoft for free on online platforms. It is a product sample database originally published by Microsoft to demonstrate the supposed design of an SQL server database using SQL server.

- AdventureWorks database supports a manufacturing MNC named Adventure Works Cycles.

- It is a sample Online Transaction Processing (or OLTP) database, which is a type of data processing where multiple transactions occur concurrently. These are shipped by Microsoft with all of their SQL server products.

🎯 Project Goals

- Establish a connection between on-premise SQL server and Azure cloud.

- Ingest tables into the Azure Data Lake.

- Apply data cleaning and transformation using Azure Databricks.

- Utilize Azure Synapse Analytics for loading clean data.

- Create interactive data visualizations and reports with Microsoft Power BI.

- Implement Azure Active Directory (AAD) and Azure Key Vault for monitoring and governance.

🕵️ Key Insights:

💸 Total Revenue: $ 708,300.00

A cartoon character holding a stack of red cubes

Description automatically generated Total Sales: $ 713,600.00

A colorful bike with black background

Description automatically generatedTouring Bikes is the top 1 category generating revenue with 32% followed by Road Bikes with 26% and Mountain Bikes with 24%.

🌍 Total Number of customers are 847

🚻 Revenue by Gender

59% of the revenue is generated by Male customers and 41% of the revenue is generated by Female customers.

This can be explained by males have more interest in doing outdoor activities with the different categories of Bikes than females.

📝 Project Architecture

You can find the detailed information on the diagram below:

A diagram of a software company

Description automatically generated with medium confidence

📤 Data Ingestion

Connected the on-premise SQL Server with Azure using Microsoft Integration Runtime.

A screenshot of a computer

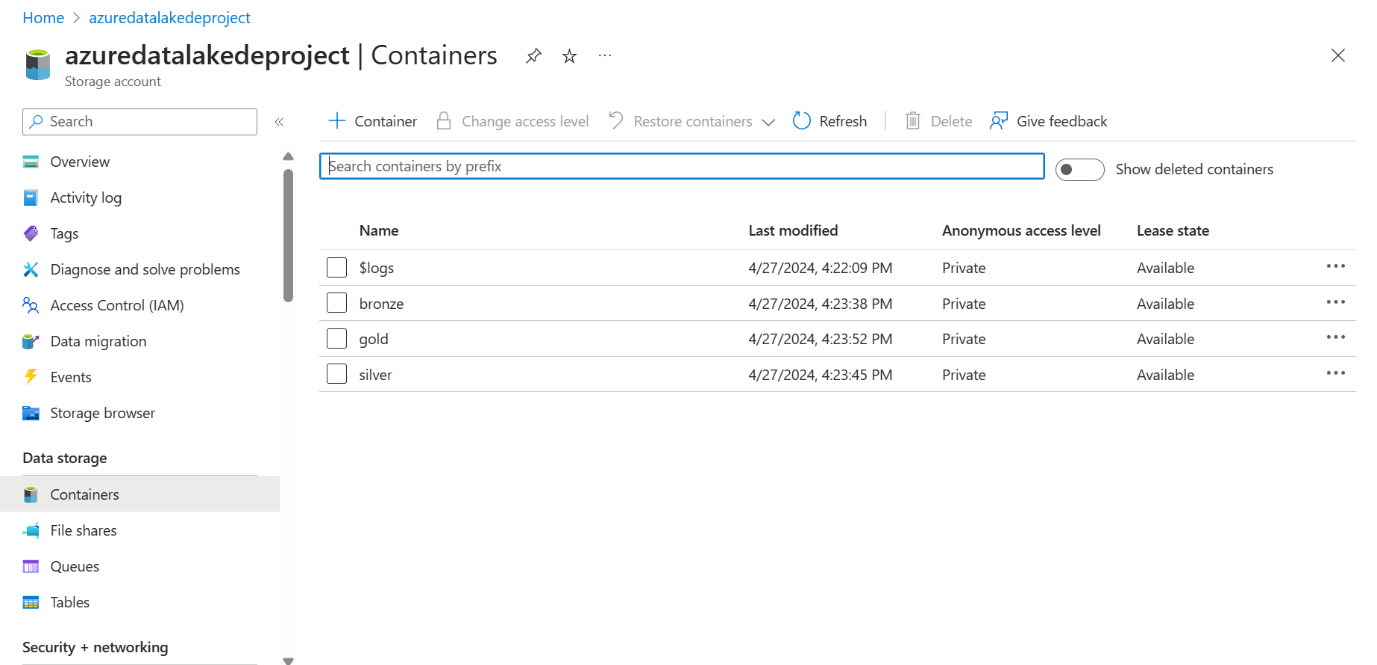
Description automatically generated

Setup the Resource group with needed services (Key Vault, Storage Account, Data Factory, Databricks, Synapse Analytics)

A screenshot of a computer

Description automatically generated

Migrated the tables from on-premise SQL Server to Azure Data Lake Storage Gen2.



⚙️ Data Transformation

Mounted Azure Blob Storage to Databricks to retrieve raw data from the Data Lake.

Used Spark Cluster in Azure Databricks to clean and refine the raw data.

Saved the cleaned data in a Delta format; optimized for further analysis.

A screenshot of a computer code

Description automatically generated

End to End data Pipeline

A screenshot of a computer

Description automatically generated

📥 Data Loading

Used Azure Synapse Analytics to load the refined data efficiently.

Created SQL database and connected it to the data lake.

synapse-pipeline db-synapse

A screenshot of a computer

Description automatically generated

📊 Data Reporting

Connected Microsoft Power BI to Azure Synapse and used the Views of the DB to create interactive and insightful data visualizations.

PowerBI-dashboard

A screenshot of a computer

Description automatically generated

🛠️ Technologies Used

Data Source: SQL Server

Orchestration: Azure Data Factory

Ingestion: Azure Data Lake Gen2

Storage: Azure Synapse Analytics

Authentication and Secrets Management: Azure Active Directory and Azure Key Vault

Data Visualization: PowerBI

📋 Credits

This Project is inspired by the video of the YouTube Channel "Mr. K Talks Tech"

📨 Contact Me

A black square with white letters in it

Description automatically generated : [www.linkedin.com/in/rahulmogal](http://www.linkedin.com/in/rahulmogal)

 : [rahulmogal@outlook.com](mailto:rahulmogal@outlook.com%20)