Ride Wise-Smart Insights for Bike Sales

Introduction:

Ride Wise is a data engineering project aimed at analysing historical bike sales data using the AdventureWorks dataset. The project focuses on building a modern data pipeline using **Azure-based technologies** to enable insightful business reporting through **Power BI**. The solution is designed with the **Medallion Architecture** in mind to ensure scalability, data quality, and efficient data transformations.

Key Use Cases:

- Identify top-selling bike categories, subcategories, and products.
- Analyze customer buying behavior based on location and order history.
- Monitor sales performance over time.
- Generate region-wise and product-wise sales trends.
- Provide clean, curated data to business users via Power BI dashboards.

Objectives:

- Implement an end-to-end data pipeline from raw CSV ingestion to curated datasets.
- Apply Medallion Architecture principles (Bronze → Silver → Gold layers).
- Enable efficient data transformations using **Databricks and PySpark**.
- Orchestrate and automate workflows using **Azure Data Factory (ADF)**.
- Store and query structured data efficiently with **Azure Synapse Analytics**.
- Deliver interactive, self-service analytics through **Power BI**.

Tools and Technology:

- **Azure Data Factory (ADF)** – for data orchestration and pipeline automation
- **Azure Databricks (PySpark)** – for scalable transformations
- **Azure Data Lake Storage Gen2** – to store data across bronze, silver, and gold layers
- **Azure Synapse Analytics** – for optimized querying and data serving
- **Power BI** – for reporting and dashboarding
- **GitHub** – for version control and collaboration

Key Steps and Implementation:

1. Data Ingestion (ADF → Bronze)
2. Data Cleaning & Transformation (Databricks → Silver)
3. Business-Level Aggregations (Databricks → Gold)
4. Data Serving Layer (Synapse)
5. Reporting (Power BI)

Data Flow:

Raw CSVs (GitHub/Local)

   ↓

ADF Ingestion → Data Lake (Bronze)

   ↓

Databricks Cleaning → Data Lake (Silver)

   ↓

Business Aggregations → Data Lake (Gold)

   ↓

Synapse SQL Views & Tables

   ↓

Power BI Dashboards

Outputs:

- Clean, transformed, and analytics-ready datasets in Synapse
- Power BI dashboards