**Project Overview:**

This project aims to implement an end-to-end ETL (Extract, Transform, Load) pipeline on the Azure platform to handle data from multiple sources, including an e-commerce website, a mobile app, and point-of-sale (POS) systems. The objective is to orchestrate data from these platforms, perform transformations, clean the data, and store it in a data warehouse for reporting and analytics via Power BI.

**Problem Statement:**

The project involves integrating transactional data from various sources (web, mobile app, and POS), performing data cleaning, validation, transformation, and loading it into a centralized data warehouse. This enables comprehensive reporting and analytics, including customer behavior analysis, product sales performance, and customer feedback insights, using Power BI.

Challenges include:

* Handling diverse datasets with varying structures.
* Transforming raw data into meaningful insights.
* Building reports on customer behavior, sales patterns, and feedback analysis.

**Data Sources:**

1. E-commerce website transaction data
2. E-commerce mobile app transaction data
3. Point-of-Sale (POS) store transaction data
4. Product details, customer details, and customer feedback

**Key Azure Services:**

* **Azure Data Factory** (ADF) – Orchestrating the data pipelines.
* **Azure Blob Storage** – Storing raw files and staging data.
* **Azure Data Lake Storage Gen2** – Centralized data repository for structured, semi-structured, and unstructured data.
* **Azure SQL Database** – Storing transformed, cleaned data.
* **Azure Synapse Analytics** – Centralized data warehouse for reporting.
* **Power BI** – Data visualization and reporting.

**ETL Pipeline:**

* **Extract**: Data is ingested from different sources into Azure Data Lake Storage. Data from the e-commerce website, mobile app, and POS is uploaded as CSV files.
* **Transform**: Azure Data Factory and Azure Databricks handle data cleaning, deduplication, and transformation.
  + Data from different platforms (web, app, POS) is merged based on customer and transaction details.
  + Data is normalized and cleaned.
  + Missing data is handled through imputation.
* **Load**: Transformed data is loaded into Azure SQL Database and Azure Synapse Analytics for advanced analytics.

**Data Flow:**

1. **Source data ingestion**: Data is ingested from web and app-based transaction systems and POS stores into Azure Blob Storage.
2. **Data orchestration**: Azure Data Factory manages the flow of data from Blob Storage to Azure SQL Database for transformations.
3. **Data transformation**: Cleaned and transformed data is loaded into Azure Synapse Analytics for reporting.
4. **Data visualization**: Power BI is used to visualize key metrics and build interactive dashboards.