**e-Commerce Use Case Document - Handbook**

**Data Pipeline Test:**

This script contains unit tests for a data pipeline designed to process various datasets related to an e-commerce business. The pipeline performs the following tasks:

* Read raw datasets from DBFS.
* Clean and transform the raw data.
* Create temporary views for analysis.
* Generate enriched tables by joining and transforming data.
* Create aggregate tables for analysis.
* Execute SQL queries to analyze the data.

**Dependencies:**

* unittest: A unit testing framework in Python.
* pandas
* pyspark
* com.crealytics:spark-excel\_2.12:0.13.4: A Spark package for reading Excel files.

**Class: DataPipelineTest**

**Methods:**

setUpClass(cls): This class method sets up the SparkSession and mounts Azure Data Lake Storage Gen2 using a service principal. It also defines the paths and schema for the datasets.

test\_raw\_tables\_creation(self): This method tests the creation of raw tables by reading datasets, cleaning the data, creating Spark DataFrames, and creating temporary views for analysis.

test\_enriched\_tables\_creation(self): This method tests the creation of enriched tables by reading raw tables, cleaning and transforming the data, and creating temporary views for analysis.

test\_enriched\_orders\_table\_creation(self): This method tests the creation of enriched orders table by joining raw and enriched tables, cleaning and transforming the data, and creating a temporary view for analysis.

test\_aggregate\_table\_creation(self): This method tests the creation of aggregate tables by reading enriched orders table, aggregating the data, and creating a temporary view for analysis.

test\_sql\_queries(self): This method tests the execution of SQL queries on the aggregated data to analyze profits by year, category, customer, and year.