**📅 Project 1: BNPL Customer Insights & Risk Analysis**

**🎯 Objective**

Analyze Buy Now, Pay Later (BNPL) customer data to uncover trends, assess repayment behavior, and identify high-risk patterns.

**✅ Phase 0: Python Data Pipeline (ETL Setup)**

**📀 Purpose:**

Build a modular data pipeline to manage the entire flow of data: from raw generation or ingestion to transformation and cleaning.

**📦 Pipeline Components:**

1. **extract\_data()** – Loads or generates raw data.
2. **clean\_data(df)** – Handles nulls, duplicates, and fixes data types.
3. **transform\_data(df)** – Creates new features, e.g., total repayment.
4. **validate\_data(df)** – Ensures data quality (valid credit scores, no negatives).
5. **save\_data(df, filename)** – Exports the clean file for later use.

**✅ Phase 1: Interactive Data Exploration (Optional)**

If you want students to understand each transformation step more clearly:

* Use pandas to manually inspect nulls, types, and outliers.
* Visualize distributions (e.g., income, purchase amount).
* Prepare them for abstraction into pipelines.

**🔢 Phase 2: SQL Analysis**

Students will import the cleaned CSV into a database and write SQL queries such as:

1. Total purchases by customer
2. Missed payments by merchant
3. Avg. income vs. credit score
4. Most common product categories per location
5. Risk classification based on late payments

**📊 Phase 3: Tableau / Power BI Visualization**

**Suggested Visuals:**

* KPIs: Total Users, Active Users, Default Rate
* Bar Chart: Product Categories by Revenue
* Heatmap: Risk by Region
* Line Chart: Monthly Sales Volume
* Pie Chart: Gender Distribution

**📁 Dataset Details (20,000 Rows)**

Columns include:

* Transaction data (amount, date, merchant)
* Customer info (age, gender, income, credit score)
* Payment plan details (down payment, installments)
* Behavioral indicators (late payments, fraud flag)