Project Plan

Dataset: Customers, Products, Sales, Stores

Theme: Retail Sales Analysis

# Day 1 (4 Hours)

## 1. Dataset Creation – 4 Hours

- Use Python with Faker library to generate synthetic datasets.  
- Minimum 50,000 records each for customers, products, stores, and sales.  
- Ensure data variety (customer demographics, product categories, multiple stores, realistic sales transactions).  
- Save as customers.csv, products.csv, stores.csv, sales.csv.

Deliverable: Synthetic datasets (CSV files with ≥50k records each).

# Day 2 (8 Hours)

## 2. Data Preparation & ETL (Python + Pandas) – 4 Hours

- Load customers.csv, products.csv, stores.csv, sales.csv in Pandas.  
- Clean data: handle nulls, duplicates, inconsistent formats.  
- Generate dim\_date from sales\_date.  
- Save staging files (stg\_customers, stg\_products, etc.).  
- Write Python ETL script to:  
 • Transform data (calculate total\_amount = quantity \* price).  
 • Insert into dimension tables and fact\_sales (CSV/SQL inserts).

Deliverable: Python ETL script + cleaned CSVs.

## 3. Data Warehouse Design & SQL – 4 Hours

- Create schema in MySQL/Postgres/SQL Server.  
- Build dim\_customer, dim\_product, dim\_store, dim\_date, fact\_sales.  
- Load transformed CSVs into SQL tables.  
- Write SQL insights:  
 1. Top 5 products by revenue.  
 2. Monthly revenue trend.  
 3. Revenue by store & state.  
 4. Customer age group contribution.

Deliverable: SQL schema + insert scripts + insights queries.

# Day 3 (8 Hours)

## 4. Dashboard Building (Power BI / Tableau) – 4 Hours

- Connect Power BI to SQL Data Warehouse.  
- Create visuals:  
 • Revenue by Category  
 • Top 5 Products  
 • Monthly Trend (line chart)  
 • Sales by Store (map or bar)  
- Add slicers (Year, State, Category).  
- Format dashboard with KPIs (Total Revenue, Avg Order Value).

Deliverable: Power BI dashboard (.pbix).

## 5. Documentation & Presentation – 4 Hours

- Document:  
 • ETL workflow (Python → Staging → DW → BI).  
 • SQL schema design (ER diagram).  
 • Key SQL queries & insights.  
 • Dashboard screenshots + findings.  
- Prepare final PPT (5–6 slides).  
- Submit code + CSVs + SQL scripts + dashboard + documentation.

Deliverable: Project Report + PPT + all scripts.

# Final Outcome (Per Member)

Each of the 6 members will independently produce:  
- Synthetic Datasets (≥50k records each)  
- Cleaned CSVs (staging data)  
- Python ETL script  
- SQL schema + queries  
- Data Warehouse (populated with fact/dim tables)  
- Power BI Dashboard (.pbix)  
- Documentation + PPT  
  
This ensures all 6 practice the full stack (Python → SQL → DW → BI → Docs).

**DATA SET Samples**

**Customer**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| customer\_id | first\_name | last\_name | gender | age | city | state |
| 1 | Kathryn | Brown | Male | 50 | New Curtis | Illinois |
| 2 | Carlos | Hudson | Female | 61 | South Ronaldstad | Ohio |
| 3 | Mario | Watson | Male | 23 | North Jeffrey | Tennessee |
| 4 | Ashley | Owens | Female | 44 | Gordontown | New Mexico |
| 5 | Michael | Hill | Male | 18 | Jimmouth | Virginia |
| 6 | Nicholas | Matthews | Male | 30 | North Mary | South Carolina |
| 7 | Jennifer | Obrien | Female | 35 | Perezmouth | Washington |
| 8 | Jamie | Fuller | Female | 33 | New Dale | Maryland |
| 9 | Shannon | Webb | Male | 32 | Port Brianview | Kentucky |
| 10 | Mitchell | Williams | Male | 61 | North Sharonton | Maryland |
| 11 | Christine | Soto | Female | 19 | Port Matthewchester | Texas |
| 12 | Eric | Sanders | Male | 51 | Melissabury | Wyoming |

**Product**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| product\_id | product\_name | category | sub\_category | price |
| 1 | Use Mobile | Electronics | Mobile | 381.5 |
| 2 | Someone Mobile | Electronics | Mobile | 48.99 |
| 3 | Mr Mobile | Electronics | Mobile | 223.44 |
| 4 | Relate Mobile | Electronics | Mobile | 424.88 |
| 5 | Yourself Mobile | Electronics | Mobile | 863.54 |
| 6 | Hand Laptop | Electronics | Laptop | 171.11 |
| 7 | Analysis Laptop | Electronics | Laptop | 256.47 |
| 8 | American Laptop | Electronics | Laptop | 673.11 |
| 9 | Vote Laptop | Electronics | Laptop | 280.33 |
| 10 | Down Laptop | Electronics | Laptop | 609.31 |
| 11 | Report Tablet | Electronics | Tablet | 686.21 |
| 12 | Wear Tablet | Electronics | Tablet | 144.09 |

**Stores**

|  |  |  |  |
| --- | --- | --- | --- |
| store\_id | store\_name | city | state |
| 1 | Store\_1 | Contrerasville | Rhode Island |
| 2 | Store\_2 | Lake Gregoryport | Mississippi |
| 3 | Store\_3 | Tylerburgh | Vermont |
| 4 | Store\_4 | Arthurville | Georgia |
| 5 | Store\_5 | South Jasonstad | Missouri |

**Sales**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| sales\_id | customer\_id | product\_id | store\_id | quantity | sales\_date |
| 1 | 36 | 32 | 4 | 5 | 11-12-2024 |
| 2 | 24 | 21 | 1 | 1 | 05-09-2024 |
| 3 | 28 | 19 | 3 | 2 | 05-07-2025 |
| 4 | 36 | 45 | 5 | 2 | 12-01-2025 |
| 5 | 31 | 23 | 1 | 4 | 04-08-2025 |
| 6 | 43 | 24 | 1 | 5 | 22-06-2025 |
| 7 | 48 | 36 | 3 | 5 | 07-01-2025 |
| 8 | 11 | 22 | 4 | 5 | 26-08-2025 |
| 9 | 45 | 12 | 5 | 1 | 21-03-2025 |
| 10 | 8 | 13 | 4 | 4 | 29-07-2025 |
| 11 | 43 | 2 | 4 | 1 | 02-09-2024 |
| 12 | 10 | 32 | 2 | 5 | 24-05-2025 |
| 13 | 25 | 20 | 4 | 2 | 18-01-2025 |
| 14 | 33 | 9 | 2 | 3 | 28-02-2025 |
| 15 | 36 | 32 | 3 | 3 | 18-05-2025 |
| 16 | 46 | 17 | 4 | 3 | 27-07-2025 |
| 17 | 27 | 19 | 4 | 1 | 20-08-2025 |
| 18 | 18 | 31 | 1 | 1 | 09-03-2025 |
| 19 | 43 | 15 | 1 | 1 | 08-01-2025 |
| 20 | 21 | 14 | 4 | 3 | 02-09-2024 |
| 21 | 15 | 15 | 5 | 5 | 16-04-2025 |