PYTHON AUTOMATION PROJECT – DOCUMENTATION PACK

FUNCTIONAL REQUIREMENTS DOCUMENT (FRD)

Project: Python Data Automation Script
Author: Punit Dave

1. Functional Scope

This FRD defines functional requirements for the Python script that processes raw operational data, cleans it, analyzes it, and exports standardized output.

2. Input Data Requirements

Input Type	Description
Raw CSV file	Contains raw transaction or operational data
Excel file (.xlsx)	Optional input format
Column list	Expected columns: date, amount, status, customer_id, etc.

3. Functional Requirements

3.1 Data Cleaning Requirements

Requirement ID	Description
FR-01	Remove duplicate rows
FR-02	Remove rows with null mandatory fields
FR-03	Convert date fields to proper datetime format
FR-04	Standardize column names (snake_case)
FR-05	Validate numeric fields

3.2 Transformation Requirements

Requirement ID	Description
FR-06	Create new calculated fields (e.g., total_amount)
FR-07	Extract Year, Month, Day columns
FR-08	Categorize values (e.g., High/Medium/Low)
FR-09	Map codes to labels (optional dictionary mapping)

3.3 Summary Report Requirements

Metric	Definition
Total Records	Count of all rows
Unique Customers	DISTINCT COUNT of customer_id
Total Amount	SUM(amount)
Success Rate	(Success Transactions / Total Transactions)

3.4 Script Output Requirements

Output File	Description
Cleaned CSV file	Cleaned and transformed dataset
Summary report CSV	KPI summary metrics
Log file (optional)	Shows errors or warnings

4. Python Script Expected Workflow

1. Load raw CSV file
2. Remove duplicates
3. Replace or drop null values
4. Standardize columns
5. Apply transformations
6. Generate summary metrics
7. Export outputs
8. Log completion status

5. Non-Functional Requirements

Category	Requirement
Performance	Script should complete processing in < 5 seconds for 50k rows
Usability	Script should work with drag-and-drop file or file path input
Scalability	Should support large datasets (up to 1M rows)
Reliability	Should run without manual intervention
Maintainability	Code should be modular and readable