Example ETL Normalization Prompt

Prompt for ETL Normalization of Categorical and Numeric Datasets

Objective: Create a comprehensive Extract, Transform, Load (ETL) process to normalize two datasets with potentially different categorical and numeric representations.

Task Instructions:

Dataset Analysis Phase: Carefully compare the two input datasets Identify all categorical and numeric fields in both datasets Create a comprehensive mapping strategy for: a) Categorical field transformations b) Numeric field standardization c) Field name alignments

Normalization Strategy Development: Develop a systematic approach to map categorical values Create rules for handling: Different categorical value representations Variations in field naming conventions Potential data inconsistencies

Transformation Rules: Establish clear, consistent rules for: Value mapping Data type conversions Handling of missing or undefined values Ensure no data loss during transformation Preserve the original meaning and context of data

Implementation Considerations: Use a robust programming language (recommended: Python with Pandas) Implement comprehensive error handling Create a transparent, reversible transformation process Include detailed logging of all transformation steps

Validation Requirements: Develop validation checks to ensure: All data points are successfully transformed No critical information is lost Transformed data maintains statistical integrity Transformed dataset matches target dataset structure

Documentation: Provide clear documentation of: Transformation methodology Mapping rules Potential limitations Edge case handling

Deliverables:

Comprehensive ETL transformation script

Detailed transformation mapping document

Validation report

Transformation log

Evaluation Criteria:

Completeness of transformation

Data integrity preservation

Clarity of transformation logic

Robustness of error handling

Efficiency of transformation process