

5. DATA CONVERSION:

5.1 Overview

prep_converter.py is a **CSV Data Conversion and Processing Tool** that transforms unified metrics data into a structured format separating quantitative and qualitative business metrics. This data processing utility serves as a critical post-processing step in the business intelligence pipeline, taking the complex analytical outputs from insight_prep.py and converting them into clean, standardized data formats that can be easily consumed by downstream systems and stakeholders.

5.2 Main Purpose

The script serves as a **post-processing converter** that takes the output from insight_prep.py (unified metrics CSV and legend CSV files) and converts them into a final, structured CSV file that properly separates and validates quantitative and qualitative metrics. Acting as the final transformation layer in the analytical pipeline, it ensures that raw business intelligence data is converted into a format suitable for advanced analytics, machine learning models, and executive reporting.

5.3 Key Functionalities

5.3.1 Data Input Processing

The converter begins by processing the input data sources that form the foundation of its operations. It reads the unified metrics CSV file, which contains the comprehensive analytical outputs generated by the insight preparation pipeline. Simultaneously, it processes the legend CSV file that provides crucial metadata and business context for each metric. The system includes robust validation mechanisms to check for file existence, proper formatting, and data integrity before proceeding with the conversion process.

5.3.2 Quantitative Metric Extraction

At its core, the converter performs sophisticated quantitative metric extraction from multiple data columns within the source files. It focuses on extracting numeric values from several key columns including Key_Metric, Secondary_Metric_1, Secondary_Metric_2, etc. The extraction process employs context-aware parsing that leverages the legend definitions to intelligently identify and extract various types of quantitative data. This includes currency values denoted by dollar signs, percentages, ratios such as LTV:CAC ratios, customer counts, score-based metrics on a 0-10 scale, and time durations measured in seconds, months, or days.

5.3.3 Intelligent Data Type Detection

The converter implements an advanced data type detection system that prioritizes the metric type definitions from the legend CSV over automatic pattern detection. This **legend-driven parsing** approach ensures accuracy by treating data according to its predefined business context rather than relying solely on format patterns. For instance, if the legend specifies a metric as "Percentage," the converter will extract plain numeric values and treat them as percentages regardless of their original formatting, providing consistent and reliable data interpretation.

```
# Example: If legend says "Percentage", it treats numbers as percentages
# Example: If legend says "Currency", it extracts $ values regardless of
format
```

5.3.4 Data Validation & Quality Control

A comprehensive data validation and quality control system ensures the accuracy and reliability of the converted metrics. The converter includes mathematical impossibility detection algorithms that flag unrealistic values such as success rates exceeding 100% or negative counts for inherently positive metrics. It implements unit standardization to convert various unit representations like thousands (K), millions (M), billions (B), and provides consistent formatting across all metrics. The error classification system categorizes issues by severity levels including LOW, MEDIUM, HIGH, and FATAL, allowing for appropriate handling and prioritization of data quality concerns.

5.3.5 Output Generation

The converter generates a comprehensive final CSV file with a well-structured format designed to meet diverse analytical needs. This output includes columns such as Metric_Type to distinguish between qualitative and quantitative metrics, Metric_Name for clean metric identification, Value for numeric data points, and Unit for standardized unit representation. Business intelligence scores are preserved through columns like Valence_Score, Significance_Score, and Risk_Level, while data lineage tracking is maintained through Source_Metric and Source_Column fields. A Mathematical_Validation_Flag column provides quality assurance by highlighting any data validation issues that were detected during processing.

To re-state and enumerate, it creates a **final CSV** with these columns:

- Metric_Type : 'Qualitative' or 'Quantitative'
- Metric_Name : Clean metric name
- Value : Numeric value (for quantitative metrics)

- Unit : Standardized unit (\$, %, ratio, score, etc.)
- Valence_Score , Significance_Score , Risk_Level : Business scores
- Source_Metric , Source_Column : Data lineage tracking
- Mathematical_Validation_Flag : Flags data quality issues

5.3.6 Robust Error Handling

The system incorporates multiple layers of robust error handling to ensure reliable operation even with challenging data conditions. A circuit breaker pattern prevents cascade failures by temporarily halting processing when error rates become excessive. An exponential backoff retry mechanism handles transient errors gracefully, automatically attempting recovery with increasing delays between attempts. Comprehensive logging captures detailed error information and processing statistics, while the recovery suggestion system provides actionable guidance for resolving encountered issues.

5.3.7 Processing Statistics

The converter generates detailed statistical reports that provide insights into the conversion process and data quality. These reports include success and failure rates for metric extractions, unit distribution summaries that show the prevalence of different measurement types, and error classification breakdowns that help identify patterns in data quality issues. Processing performance metrics track execution times and resource utilization, providing valuable information for optimization and scaling considerations.

5.4 Example Use Case

When `insight_prep.py` generates a complex metric entry such as a revenue analysis with multiple data points, the converter transforms this into a structured format. For example, if the original metric shows `Key_Metric: $125,430`, `Secondary_Metric_1: 15.7%`, and `Secondary_Metric_2: 2.3:1`, the converter creates both qualitative and quantitative representations. The qualitative row preserves the original business context and metadata, while the quantitative rows break down the individual components into separate, analyzable records such as revenue values in dollars, growth rates as percentages, and profit ratios as standardized ratio values.

To re-state, if `insight_prep.py` generated a metric like:

```
Metric: Revenue Analysis
Key_Metric: $125,430
Secondary_Metric_1: 15.7%
Secondary_Metric_2: 2.3:1
```

The converter would create:

1. **Qualitative row:** Original metric with business context
2. **Quantitative rows:**
 - Revenue Analysis - Revenue: 125430 (\$)
 - Revenue Analysis - Growth Rate: 15.7 (%)
 - Revenue Analysis - Profit Ratio: 2.3 (ratio)

5.5 Purpose in the Workflow

This converter acts as the final processing step that transforms raw business analysis data into a clean, structured format suitable for advanced analytical applications. The processed data becomes immediately compatible with **data visualization platforms, business intelligence dashboards, and further statistical analysis workflows**. It serves as the bridge between complex analytical processing and practical business applications, enabling **machine learning model input, executive reporting, and automated decision-making systems**. The script is designed to handle large datasets with robust error recovery, making it production-ready for processing business metrics at scale across enterprise environments. ed customer profiles that enable dynamic pricing, personalized marketing, and proactive support based on complete customer journey data.

Ultimately, the E-Commerce Data Hub transforms fragmented e-commerce data landscapes into unified, intelligent systems that actively enhance business intelligence and operational efficiency. Its plug-and-play nature ensures rapid deployment while AI-driven optimization adapts continuously to evolving business needs, creating a foundation for sustained competitive advantage in the digital commerce landscape.