**\*\*Manual Tasks:\*\***

Review Data Quality Reports: Manually review data quality reports generated from the automated processes to identify any issues that require manual intervention.

Handle Exceptions: Manually handle exceptions or special cases that cannot be handled through automated processes.

Long-Term Improvements and Roadmap:

Continuous Monitoring: Implement a system for continuous monitoring of data quality, with alerts for anomalies or issues.In my compaany I used Azure data Factory ETL pipelines and monitors to ensure data quality, while loading the data.

Feedback Loop: Established a feedback loop where manual interventions are used to improve automated processes over time.

Data Quality Framework: Developed a data quality framework that includes standards, processes, and tools for ensuring and maintaining data quality.

**\*\*Scaling the Process:\*\***

Parallel Processing: Use parallel processing techniques to handle data validation and cleansing for hundreds or thousands of SKUs concurrently.

Batch Processing: Implement batch processing to handle large volumes of data efficiently.

Distributed Computing: Used distributed computing frameworks like Apache Spark, AWS Glue services and many more to scale the data validation and cleansing process across multiple nodes.

**\*\*Automated Approaches:\*\***

Data Conversion: I used Pandas' functions like, to\_numeric, to\_datetime function to convert  columns to the appropriate data type, handling errors as needed.

Missing Value Handling: Use dropna to remove rows with missing values.

Outlier Detection: Use statistical methods IQR to detect and handle outliers, either by removing them or replacing them with a suitable value.

Data Type Validation: Used pandas' dtypes attribute to validate that columns have the correct data types.

Data Consistency Checks: Used regular expressions or specific logic to check for consistency in column values (e.g. date formats, valid SKUs).

Data Integrity Checks: Performed checks to ensure data integrity, such as ensuring that start dates are before end dates and others.