**Requirement Specification Document**

**1. Introduction**

This document outlines the requirements for a Java-based system to process sensor data and generate reports. The system will read data from CSV files, process the data to calculate monthly averages, maximum and minimum values, and identify outliers based on predefined thresholds.

**2. System Overview**

The system will consist of the following components:

* **Data Ingestion Module**: Reads sensor data and threshold data from CSV files.
* **Processing Module**: Calculates monthly averages, maximum, and minimum values for each sensor.
* **Outlier Detection Module**: Identifies outliers based on threshold values.
* **Reporting Module**: Generates two CSV reports: one for monthly statistics and one for outliers.
* **Error Handling Module**: Manages errors and provides appropriate error codes and messages.

**3. Functional Requirements**

3.1 Data Ingestion

* **Input Files**:
  + sensor\_data.csv: Contains sensor readings.
    - Fields: date, sensor\_type, value, unit, location\_id
  + thresholds.csv: Contains threshold values for sensors.
    - Fields: sensor\_type, min\_threshold, max\_threshold
* **Functionality**:
  + Read and parse the CSV files.
  + Validate the data format and content.

3.2 Processing Module

* **Monthly Statistics Calculation**:
  + Calculate the monthly average, maximum, and minimum values for each sensor.
  + Generate a CSV file (`monthly\_stats.csv`) with the following fields:
    - sensor\_type, month, avg\_value, max\_value, min\_value

3.3 Outlier Detection

* **Outlier Identification**:
  + Compare sensor values against the thresholds.
  + Identify values that fall outside the min and max thresholds.
  + Generate a CSV file (`outliers.csv`) with the following fields:
    - date, sensor\_type, value, unit, location\_id, threshold\_exceeded [Min/Max]

3.4 Reporting Module

* **Output Files**:
  + monthly\_stats.csv: Contains monthly average, maximum, and minimum values.
  + outliers.csv: Contains sensor readings that are outliers.

**4. Non-Functional Requirements**

* **Performance**: The system should process data efficiently and handle large CSV files.
* **Scalability**: The system should be able to scale to accommodate additional sensors and data volume.
* **Reliability**: The system should handle errors gracefully and provide meaningful error messages.
* **Resiliency**: System should log errors and continue processing even if
  + **No thresholds defined for given sensor type**
  + **Incorrect data format for one of the row in the input csv**

**5. Error Handling**

* **Error Codes and Messages**:
  + ERR001: "File not found" - The specified CSV file could not be located.
  + ERR002: "Invalid data format" - The data in the CSV file does not match the expected format.
  + ERR003: "Processing error" - An error occurred during data processing.
  + ERR004: "Thresholds not defined" - Threshold values for a sensor type are missing.

**6. Assumptions and Constraints**

* The CSV files are updated regularly and contain accurate data.
* The system will run on a server with sufficient resources to handle the data processing tasks.

**7. Glossary**

* **CSV**: Comma-Separated Values, a file format used to store tabular data.
* **Sensor**: A device that detects and measures physical properties.
* **Outlier**: A data point that differs significantly from other observations.