Business context

Company XYZ sells products online and collects transactional data from various sources. This data goes into a data warehouse and is used by the analytics department to create reports and predictive models. Recently, a number of inaccuracies have been noticed in the reports due to data errors such as missing values, duplicate records, incorrect date formats, etc.

The task is to design a process that automatically detects and reports on data quality issues.

There are provided three sample CSV files containing transaction and product data:

transactions.csv - transaction data (transaction number, transaction date, product ID, transaction amount).

products.csv - list of products (product ID, name, category, list price).

customers.csv - customer data (customer ID, name, surname, email address).

The files contain some quality errors, which should be identified and a report saved.

Required steps:

1. **Data analysis**

There should be perform analyses of the data for recognizing quality issues.

2. **Data quality tests**

Data quality tests should be scheduled to detect recognized problems.

3. **Implementation**

Please create a script (e.g. in Python) that automatically detects the above problems and generates a report with the results.

4. **Solution proposals**

Once errors have been detected, please propose remedial solutions,

5. **Documentation**

There should be prepared documentation that includes a description of the task.