

The issue

A critical data integrity issue surfaced—our reports were compromised by missing and inaccurate data, jeopardising the reliability of our insights and our credibility with the client.

In response, a rapid and decisive data validation process was implemented, aligning datasets between the client's system (Senpiper) and our UK server.

This swift intervention restored data accuracy and safeguarded our business operations and client trust.

The solution

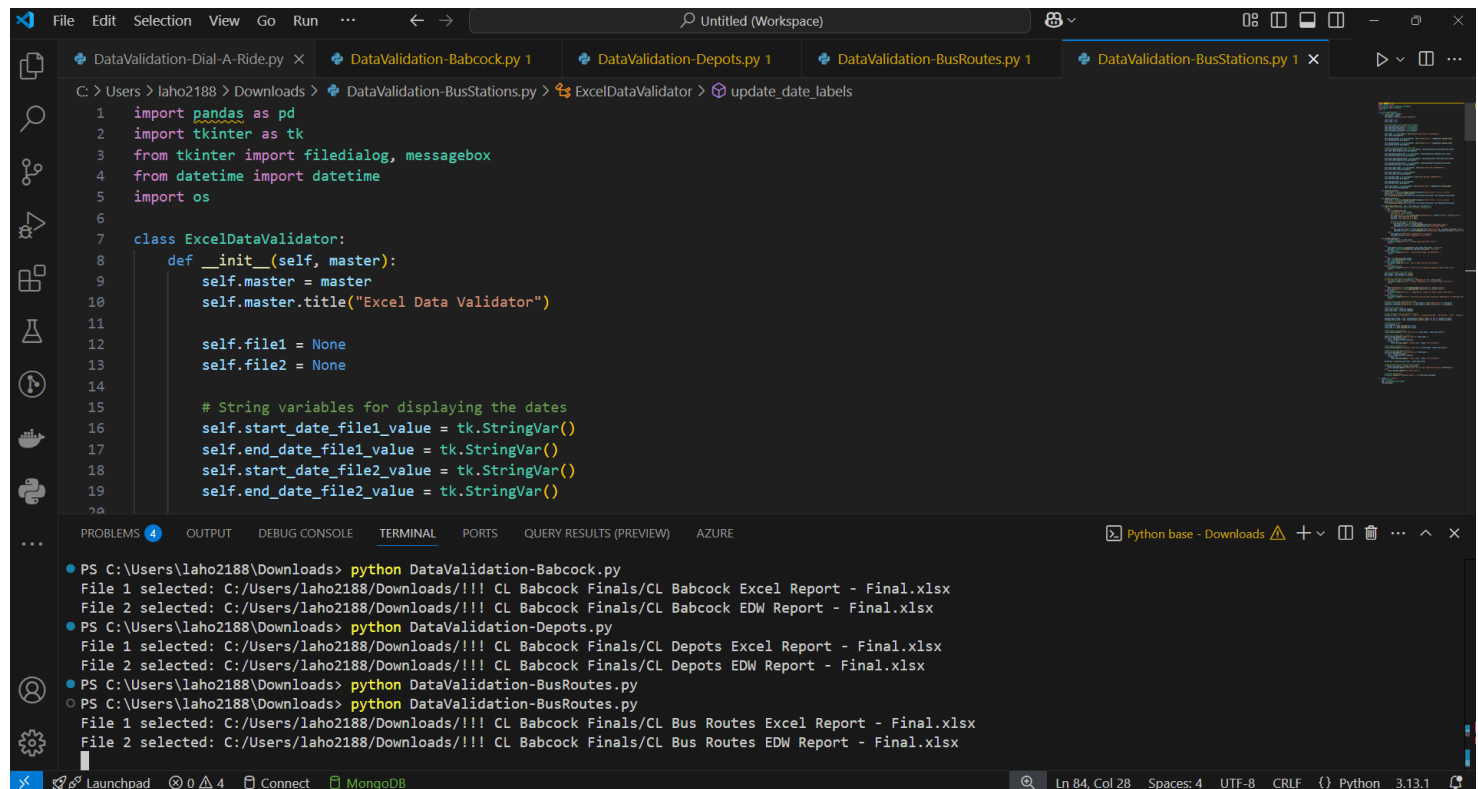
- ⚡ **Date Range Confrontation** – Cross-checked timelines to confirm that both datasets covered the same periods. Any inconsistency could distort trends and insights.
- ⚡ **Row Count Showdown** – Verified that both datasets held the same number of records. Any mismatch could signal data loss or duplication.
- ⚡ **Hunt for Missing Data** – Scanned for blank mandatory fields, identifying critical gaps that compromised our reports.
- ⚡ **Column Name Lockdown** – Ensured every column matched perfectly between sources—no room for ambiguity or misinterpretation.

Benefits & Impact

- ✓ **Crisis Averted** – Quick action prevented a major data disaster and potential business losses.
- ✓ **Accurate Data** – Ensured clean, consistent, and reliable data for better reporting.
- ✓ **Cross-Department Use** – The app can be used in other departments to validate data from various sources.
- ✓ **Faster Decisions** – Reliable data enables quicker and smarter decision-making.
- ✓ **Time & Cost Savings** – Early detection avoided costly errors and saved troubleshooting time.
- ✓ **Client Trust** – Reinforced client confidence through proactive problem-solving.

Tools & Environment

- Programming Language:
Python
- Development Environment:
Visual Studio Code
- Libraries Used:
pandas – for data analysis
tkinter – for app interface
datetime – for dates
os – for file manipulations'



```
File Edit Selection View Go Run ... Untitled (Workspace)
DataValidation-Dial-A-Ride.py x DataValidation-Babcock.py 1 DataValidation-Depots.py 1 DataValidation-BusRoutes.py 1 DataValidation-BusStations.py 1 x
C:\Users\laho2188\Downloads> DataValidation-BusStations.py ExcelDataValidator > update_date_labels
1 import pandas as pd
2 import tkinter as tk
3 from tkinter import filedialog, messagebox
4 from datetime import datetime
5 import os
6
7 class ExcelDataValidator:
8     def __init__(self, master):
9         self.master = master
10        self.master.title("Excel Data Validator")
11
12        self.file1 = None
13        self.file2 = None
14
15        # String variables for displaying the dates
16        self.start_date_file1_value = tk.StringVar()
17        self.end_date_file1_value = tk.StringVar()
18        self.start_date_file2_value = tk.StringVar()
19        self.end_date_file2_value = tk.StringVar()
20
...
PROBLEMS 4 OUTPUT DEBUG CONSOLE TERMINAL PORTS QUERY RESULTS (PREVIEW) AZURE
Python base - Downloads
PS C:\Users\laho2188\Downloads> python DataValidation-Babcock.py
File 1 selected: C:/Users/laho2188/Downloads/!!! CL Babcock Finals/CL Babcock Excel Report - Final.xlsx
File 2 selected: C:/Users/laho2188/Downloads/!!! CL Babcock Finals/CL Babcock EDW Report - Final.xlsx
PS C:\Users\laho2188\Downloads> python DataValidation-Depots.py
File 1 selected: C:/Users/laho2188/Downloads/!!! CL Babcock Finals/CL Depots Excel Report - Final.xlsx
File 2 selected: C:/Users/laho2188/Downloads/!!! CL Babcock Finals/CL Depots EDW Report - Final.xlsx
PS C:\Users\laho2188\Downloads> python DataValidation-BusRoutes.py
File 1 selected: C:/Users/laho2188/Downloads/!!! CL Babcock Finals/CL Bus Routes Excel Report - Final.xlsx
File 2 selected: C:/Users/laho2188/Downloads/!!! CL Babcock Finals/CL Bus Routes EDW Report - Final.xlsx
Ln 84, Col 28 Spaces: 4 UTF-8 CRLF {} Python 3.13.1
```

The Data Validation app – Key Features

1. File Upload and Comparison
2. Date Range Validation
3. User Interface for Validation & Reporting
4. Row Count Check
5. Mandatory Fields/Columns Check
6. Column Name Consistency Check
7. The Difference Between Sources



Excel... — □ ×

Upload Excel Files for Validation

Upload File 1

Upload File 2

CL Depots Excel Report - Final.xlsx
Earliest date: 29-09-2022

CL Depots Excel Report - Final.xlsx
Latest date: 10-02-2025

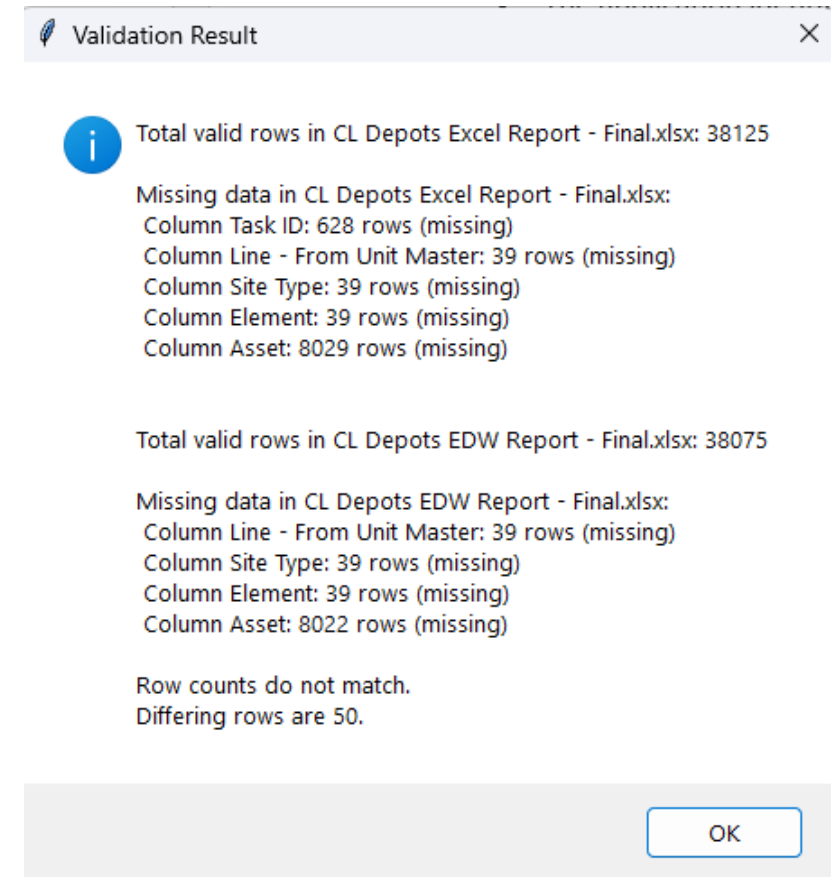
CL Depots EDW Report - Final.xlsx
Earliest date: 29-09-2022

CL Depots EDW Report - Final.xlsx
Latest date: 10-02-2025

Enter Start Date (DD-MM-YYYY):
29-09-2022

Enter End Date (DD-MM-YYYY):
10-02-2025

Validate Data



Validation Result ×

i Total valid rows in CL Depots Excel Report - Final.xlsx: 38125

Missing data in CL Depots Excel Report - Final.xlsx:
Column Task ID: 628 rows (missing)
Column Line - From Unit Master: 39 rows (missing)
Column Site Type: 39 rows (missing)
Column Element: 39 rows (missing)
Column Asset: 8029 rows (missing)

Total valid rows in CL Depots EDW Report - Final.xlsx: 38075

Missing data in CL Depots EDW Report - Final.xlsx:
Column Line - From Unit Master: 39 rows (missing)
Column Site Type: 39 rows (missing)
Column Element: 39 rows (missing)
Column Asset: 8022 rows (missing)

Row counts do not match.
Differing rows are 50.

OK

Future developments

- Possibility to add data from different sources (server/SQL and API) using user credentials
- Generate detailed reports with the missing data from submissions and mandatory columns
- Create a scheduled check and send a notification once a new issue is found