



Multi-Extension Folder-Based Data Cleaning Project

Power Query





Project Overview

01

Managed cleaning & consolidation
of multiple file formats
(.csv, .txt, .xlsx, .pdf, .accdb)

02

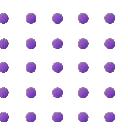
Master folder with disorganized
multi-year & multi-shift data

03

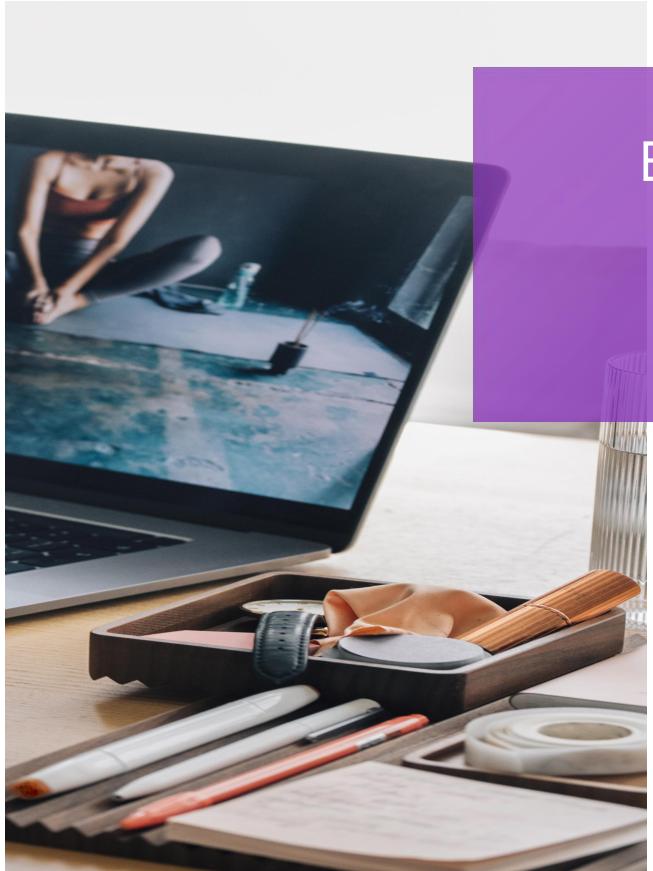
Challenges: inconsistent names,
varied structures, format-specific
issues

04

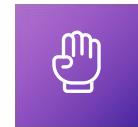
Goal: Unified Excel workbook with
automated data pipeline



Project Objectives



Build scalable ETL (Extract, Transform, Load) process



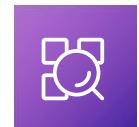
Automate import & consolidation of mixed-format files



Standardize & normalize datasets



Apply format-specific cleaning techniques



Organize cleaned outputs into logical sections

Approach & Execution



Folder-Based Import Automation – Power Query Folder Connector

Auto-refresh when new files are added

File Extension-Specific Processing – duplicate queries for each type

Dynamic Transformation Logic – Invoke Custom Function

Complex Formats – Extracted tables from PDFs & imported Access tables

Data Cleaning – Standardized columns, data types, removed errors & invalids

Organized Final Output



Loaded cleaned data
into separate Excel
worksheets



Organized by file
type for clarity



Single unified
workbook ready
for analysis

Key Skills Demonstrated



Folder-Based Automation – Batch import & refresh

Extension-wise Query Duplication & Filtering

Dynamic Parsing using Invoke Custom Function

Advanced Data Extraction – Handling PDFs & Access DB

Data Cleaning & Standardization – trimming, error handling

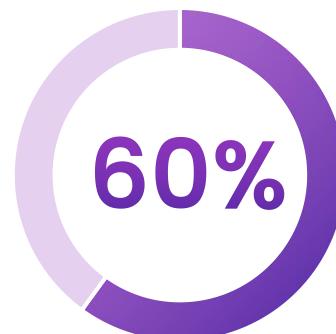
Structured Loading into Excel

Reusable, Scalable ETL Workflow

Tools Used



Microsoft Power
Query –
Automation &
transformation



Advanced Excel –
Final structured
output organization

Project Outcome



01 Robust & reusable ETL solution for multiple formats

02 Automated refresh – instant updates with new files

03 Improved efficiency, accuracy, and consistency

04 Hands-on expertise in handling multi-format data

Repository Contents

80%

/Raw Data Files –
Original messy files



Multi-Extension
Power Query
Workbook.xlsx –
Main Excel
transformations



Project
Documentation.
pdf – Detailed
steps & logic



Presentation.pptx
– Workflow
presentation



Your Title

Data
Cleaning
Report.png
– Snapshot
of results

How to Use

“

Download or Clone repository



“

Place raw files in Raw Data Files folder



“

Open the Power Query Workbook & refresh



“

Review cleaned results in Excel worksheets



Future Enhancements



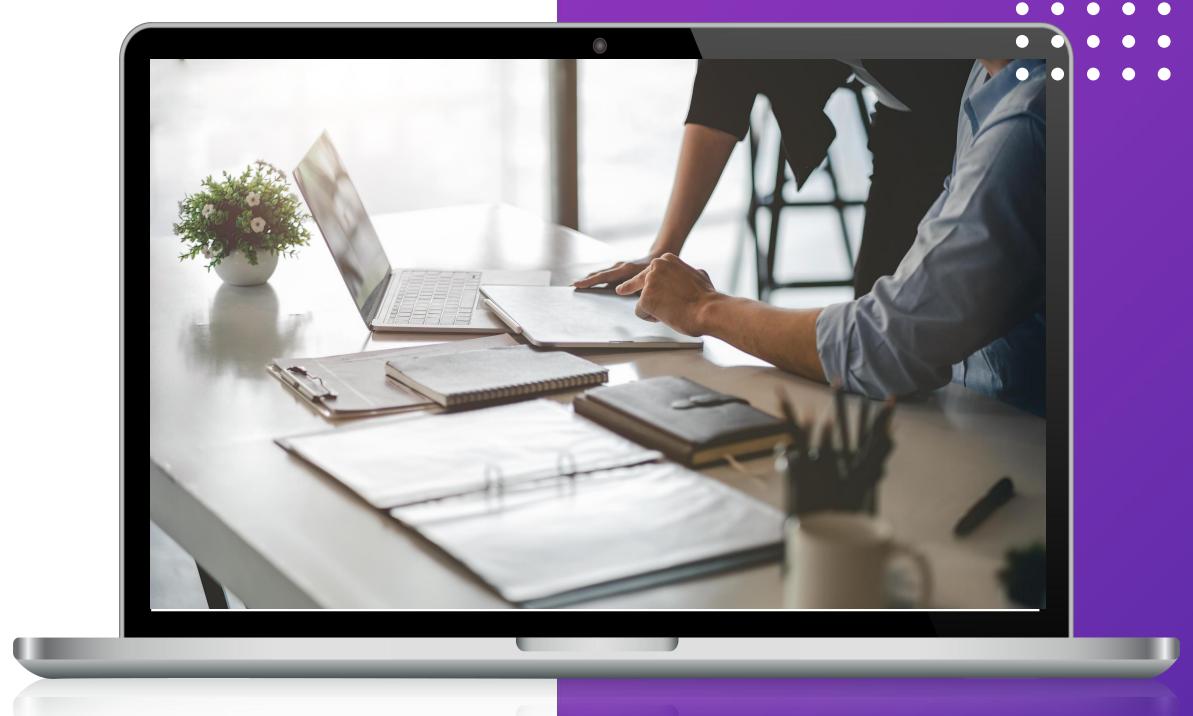
Integrate with Power BI Dataflows
for cloud-based automation



Add validation rules for
inconsistent data detection



Develop dashboard to visualize
consolidated dataset



The complete set of source files encompassing all relevant file extensions.

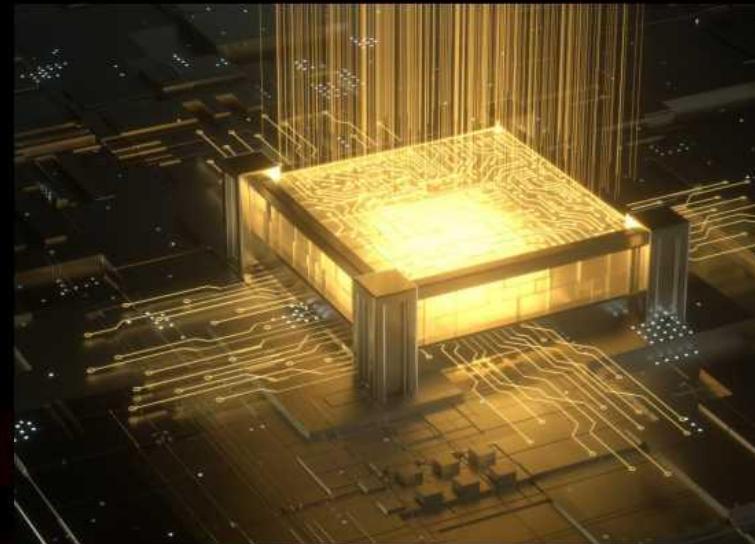
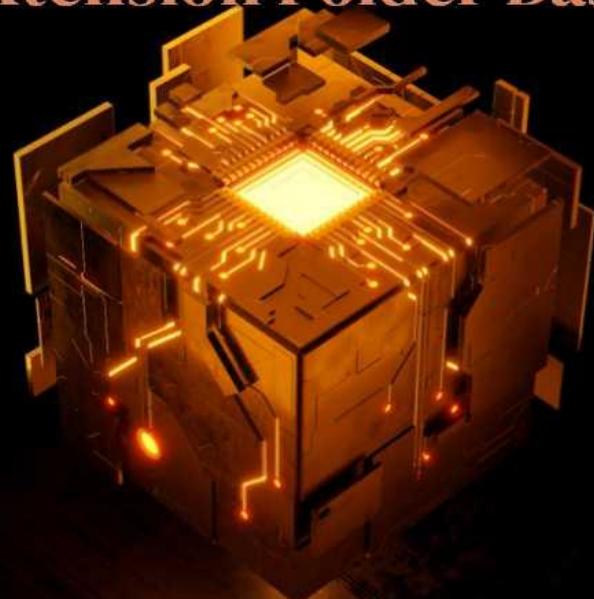
The screenshot shows a Windows File Explorer interface. The navigation pane on the left lists various locations: Quick access, Web Design, Excel Power Query, Microsoft PowerPoint, Screenshots, SQL New File, OneDrive - Personal, This PC, 3D Objects, Desktop, Documents (which is selected), Downloads, Music, Pictures, Videos, Local Disk (C:), and Local Disk (D:). The main pane displays a list of files and folders under 'This PC > Documents > csv-txt-xlsx-pdf-accdb all files'. The table has columns for Name, Date modified, Type, and Size. Notable entries include several 'Update' folders (2010, 2015, 2024, Night) and various data files in CSV, XLSX, PDF, and ACCDB formats. Many files have yellow icons above them, indicating they are backed up. The ribbon at the top includes tabs for File, Home, Share, View, Backup Tools, and a context tab for the selected folder ('Not backed up' for the folder itself, 'csv-txt-xlsx-pdf-accdb all files' for its contents).

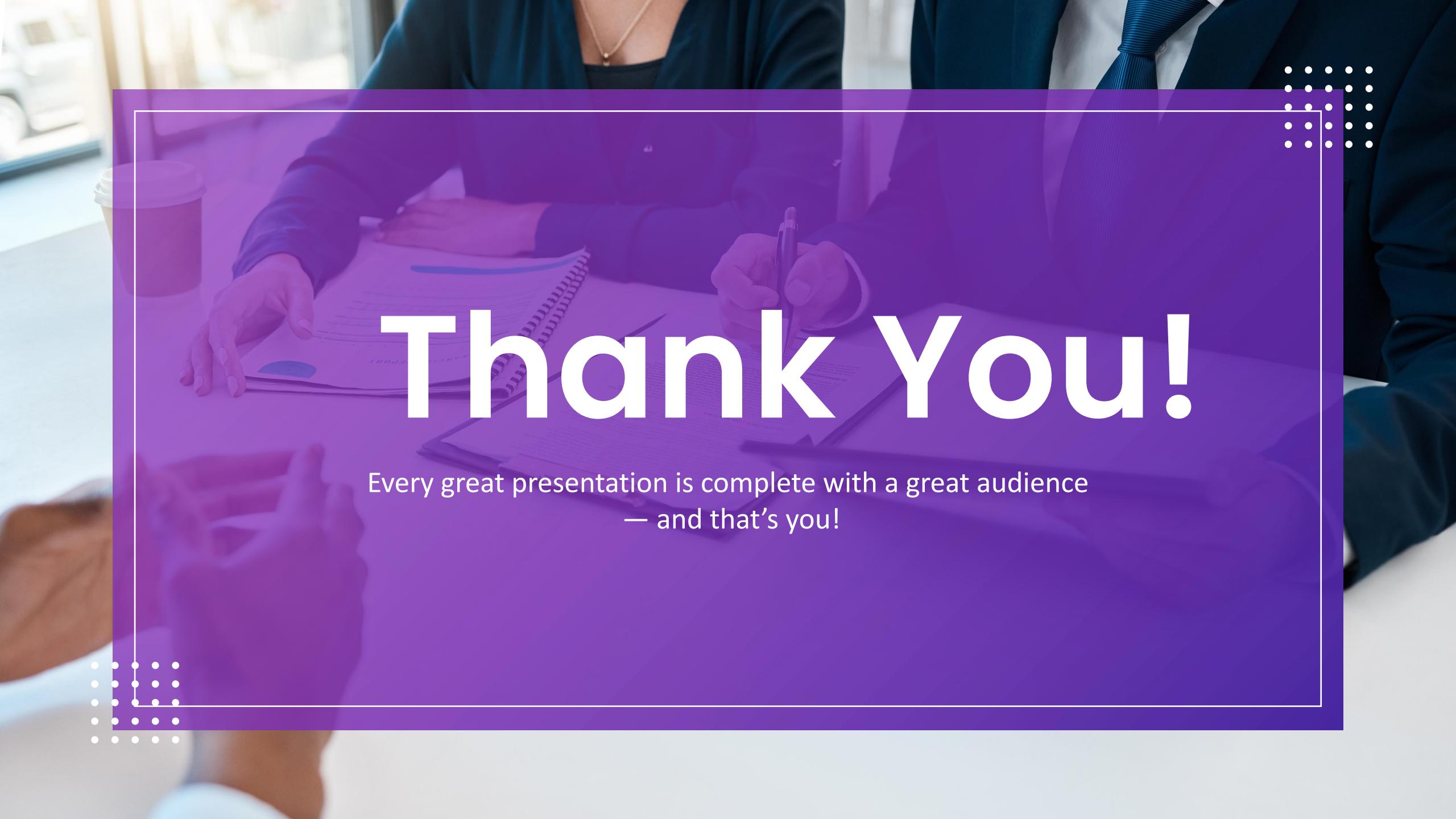
Name	Date modified	Type	Size
Update 2010 & txt Data	08-06-2025 10:07 PM	File folder	
Update 2015 & csv Data	08-06-2025 10:07 PM	File folder	
Update 2015 & pdf Data	08-06-2025 10:07 PM	File folder	
Update 2024 & xlsx Data	08-06-2025 10:07 PM	File folder	
Update Night & accdb Data	08-06-2025 10:39 PM	File folder	
2010 Order_data.txt	19-04-2025 04:19 PM	Text Document	2 KB
2015 Order_Data.csv	19-04-2025 04:20 PM	Microsoft Excel Co...	2 KB
2015 Order_data.txt	19-04-2025 04:20 PM	Text Document	2 KB
2015_Order_Data.pdf	19-04-2025 04:24 PM	PDF File	111 KB
2018 Order_Data.xlsx	19-04-2025 04:20 PM	Microsoft Excel W...	11 KB
2024 Order_Data.csv	19-04-2025 04:21 PM	Microsoft Excel Co...	2 KB
2024 Order_Data.pdf	19-04-2025 04:22 PM	PDF File	111 KB
2024 Order_Data.xlsx	19-04-2025 04:21 PM	Microsoft Excel W...	11 KB
Afternoon_Shift.accdb	13-05-2025 10:47 PM	Microsoft Access ...	736 KB
data clearing.xlsx	08-06-2025 10:44 PM	Microsoft Excel W...	72 KB
Morning_Shift.accdb	13-05-2025 10:45 PM	Microsoft Access ...	1,184 KB

Project Overview

Power Query

Unified Multi Extension Folder Based Data Cleaning





Thank You!

Every great presentation is complete with a great audience
— and that's you!