

Cleaned & Refined Messy Rows Dataset

Data Cleaning Project

Project Type: Data Cleaning & Transformation

Project Overview

01.

Real-world messy dataset cleaning & transformation

02.

Converted raw, unstructured data → analysis-ready format

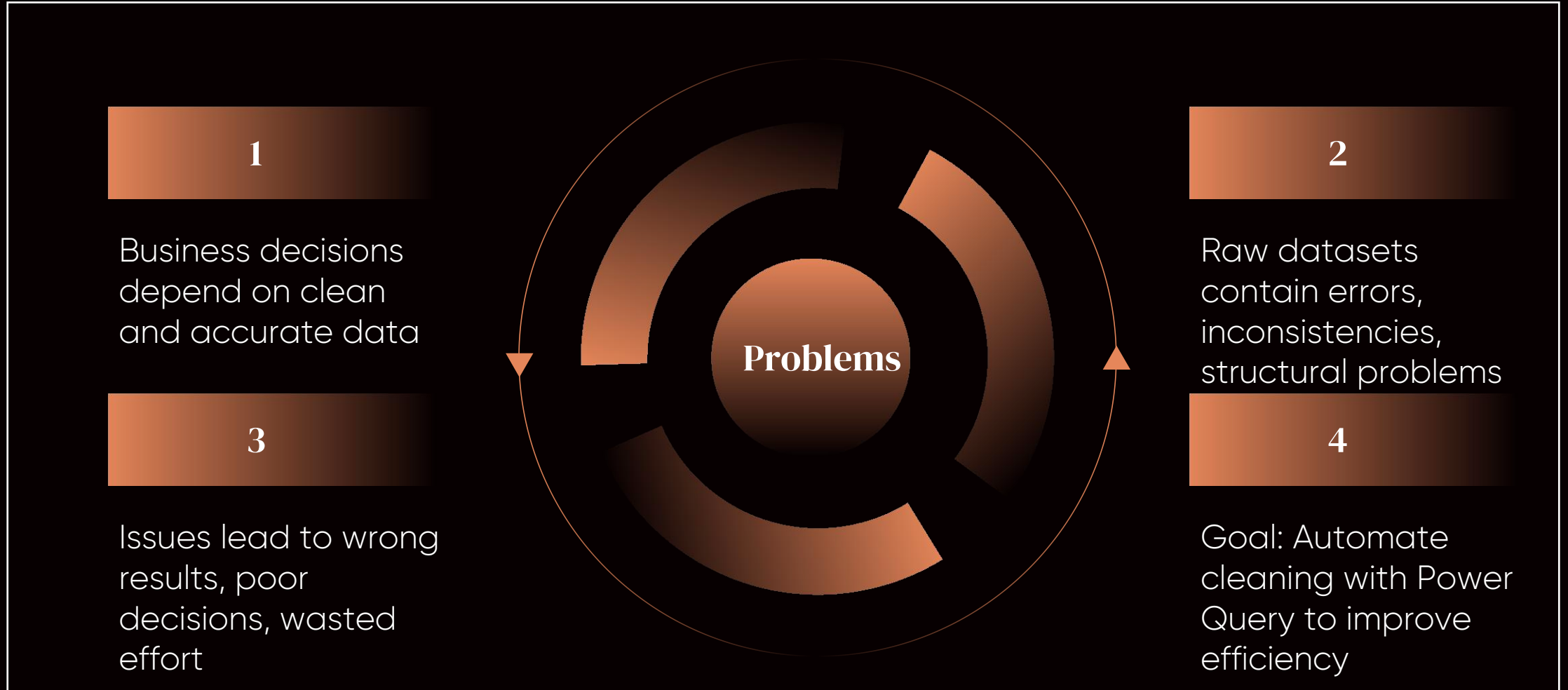
03.

Ensured accuracy, clarity, and consistency across fields

04.

Solved duplicate columns, casing, nulls, city fragments, data types

Problem Statement



Data Issues Identified

01.

- Duplicate and unnecessary columns
- Blank and null values in key fields

02.

- Inconsistent text formatting
- City names split into multiple columns

03.

- Merged cells with unrelated data
- Columns arranged in random, non-logical order

04.

- Incorrect data types (e.g., text stored as numbers)

Step-by-Step Cleaning – Part 1

1. Convert raw dataset into Excel table

2. Remove duplicate and blank columns

3. Import into Power Query Editor

4. Merge fragmented city columns (e.g., 'New'+ 'York')



5. Apply Capitalize Each Word for consistency

6. Split merged columns using delimiter

7. Handle missing values with N/A or placeholders

Step-by-Step Cleaning – Part 2

8. Use 'Columns from Examples' for extraction

9. Reorder fields logically (IDs → Names → Contact → Location → Metrics)

10. Detect and assign correct data types



11. Ensure uniformity in IDs and Emails

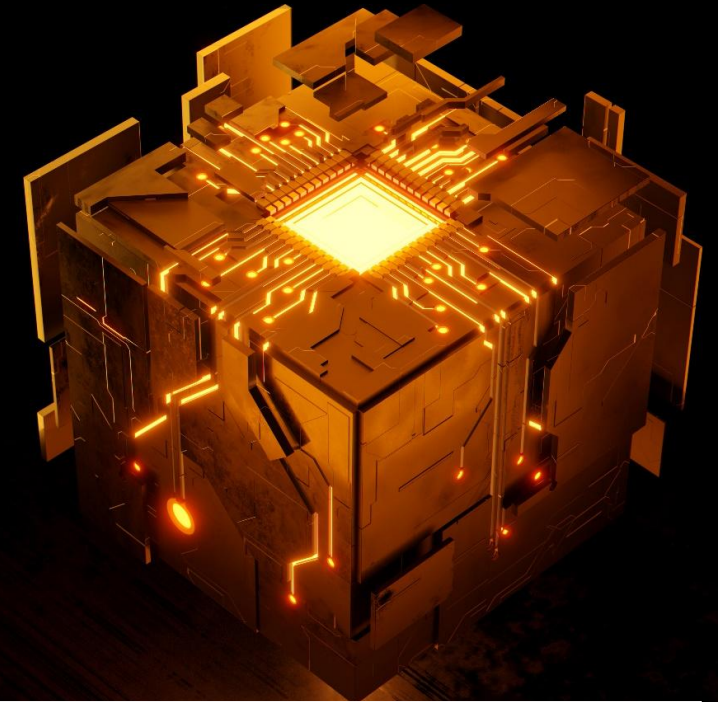
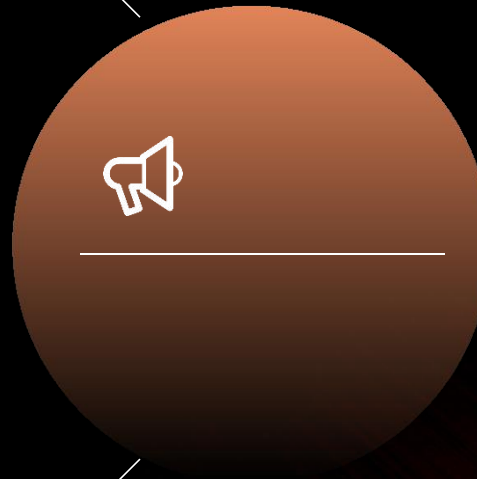
12. Export final clean dataset into Excel workbook

- Dataset now analysis-ready

Tools & Technologies Used

1. Advanced Excel – For table creation, exporting

2. Power Query – For automating cleaning & transformations



Power Query

3. Data Profiling Tools – For detecting nulls, duplicates, inconsistencies

Files in Repository

Cleaned & Refined Messy Rows
Dataset.pdf – Cleaning steps
documentation

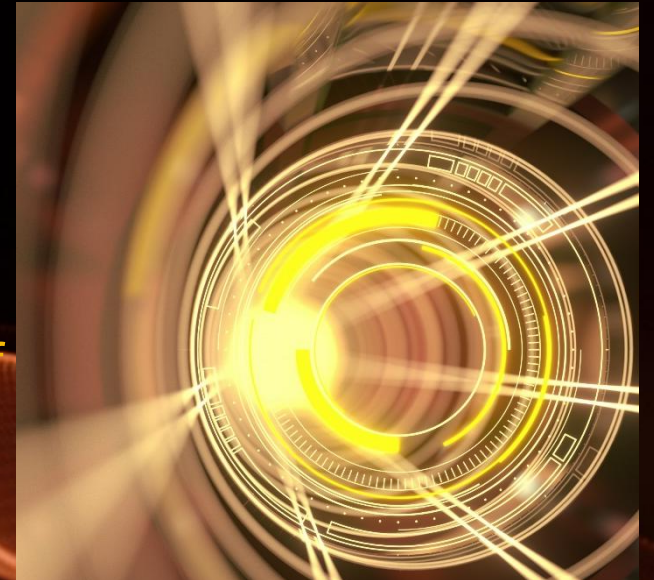


Messy Employee Dataset
2000.csv – Raw dataset (CSV
format)



Messy Employee Dataset
2000.xlsx – Raw dataset (Excel
format)

Data Cleaning.png – Snapshot of
cleaned dataset



Key Learning Outcomes

01.

- Practical Power Query application for messy datasets

02.

- Skills in text case standardization & null handling

03.

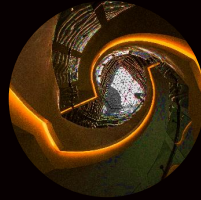
- Detecting and assigning correct data types
- Automation of repetitive cleaning tasks

04.

- Ability to convert messy, unstructured data into clean dataset
- Improved efficiency and accuracy in reporting

Future Enhancements

Automate entire process using
Power BI Dataflows



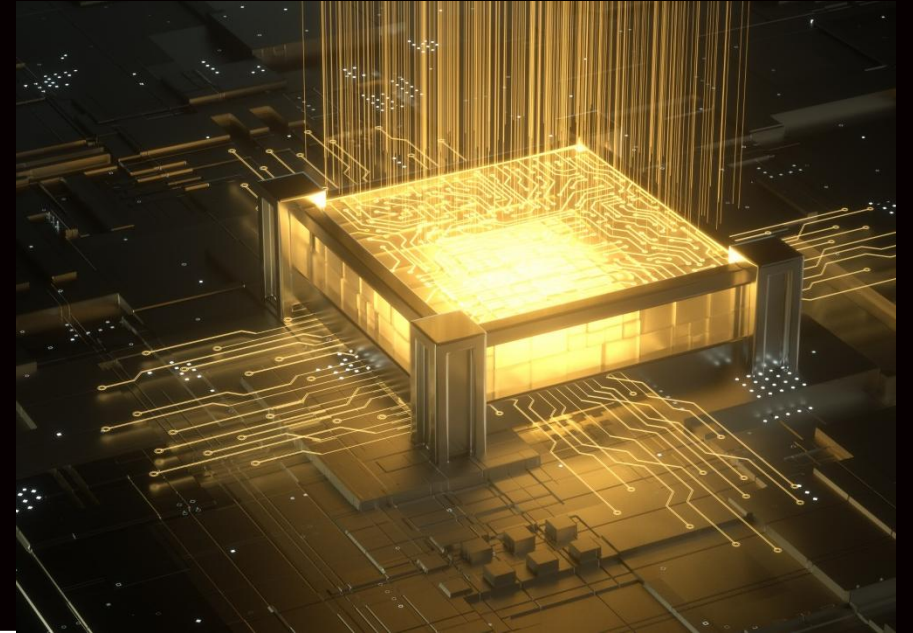
Add custom validation rules for
incoming data



Build Power BI dashboard to
visualize before-and-after results



Project Overview



Thank you!

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