

The background features three vertical bars on the left: a wide pink bar, a medium blue bar, and a narrow light beige bar. In the top right and bottom right corners, there are decorative patterns of small pink dots arranged in a grid-like fashion, with some dots missing to create a sparse effect.

DATA CLEANING PROJECT

POWER QUERY

IDENTIFYING DATA PROBLEMS

This project is about cleaning a messy finance center data sheet that have 3 main problems:

1. Data for one item divided into two rows
2. Presence of null values and empty columns
3. Unwanted values (e.g., totals) in the middle of the table

UNIQUE	L_CCY	BALANCE-LCY	BALANCE-	M	M
ISS_DATE	EXP_DATE	APPLICANT			
IMBG201911000018		INR	732,977.00	732,977.00	IN
	12/31/22	SHERI NOWMER			
MBG6601001	3/15/22	INR	730,656.89	730,656.89	IN
	7/31/23	SHERI NOWMER			
MBG6602004		INR	318,362.31	318,362.31	IN
	5/26/22	SHERI NOWMER			
MBG6603759		INR	509,679.99	509,679.99	IN
	10/5/22	DERRICK WHELPY			
MBG6603930		INR	115,922.15	115,922.15	IN
	10/23/22	JEANNE DERRY			
MBG6604844		INR	121,044.86	121,044.86	IN
	12/25/22	DERRICK WHELPY			
MBG6604960		INR	0.01	0.01	IN
	1/3/23	SHERI NOWMER			
MBG6700337		INR	45,923.52	45,923.52	IN
	2/9/23	SHERI NOWMER			
MBG6701527		INR	500,000.00	500,000.00	IN
	5/11/23	SHERI NOWMER			
MBG6701529		INR	3,000,000.00	3,000,000.00	IN
	5/11/23	SHERI NOWMER			
				TOTAL :	BY NEGOTIATIO
				6,074,566.73	INR 0.00

SOLUTION APPROACH STEPS

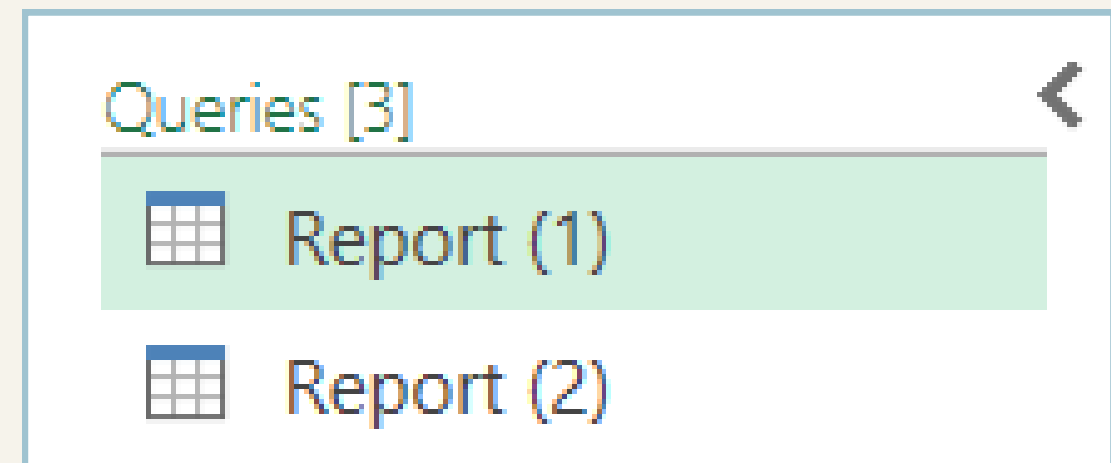
1

In Power Query, I divide the original sheet into two separate sheets:

- Report (1) containing data from row 1
- Report (2) containing data from row 2

This helps me clean each row's data separately, as they represent different parts of the same item.

Once the cleaning is done, I combine them into one sheet again.



IMBG201911000018		INR	732,977.00	732,977.00
	12/31/22	SHERI NOWMER		
MBG6601001	← Row 1	INR	730,656.89	730,656.89
3/15/22	7/31/23	SHERI NOWMER	← Row 2	
MBG6602004		INR	318,362.31	318,362.31
5/26/22	5/31/23	SHERI NOWMER		
MBG6603759		INR	509,679.99	509,679.99
10/5/22	6/30/23	DERRICK WHELPLY		
MBG6603930		INR	115,922.15	115,922.15
10/23/22	6/30/23	JEANNE DERRY		
MBG6604844		INR	121,044.86	121,044.86

SOLUTION APPROACH STEPS

2

In Report (1): 1. Remove null rows

2. Promote headers

3. Filter rows on the data on Row 1

4. Remove null columns

	ABC 123 Column1	ABC 123 Column2	ABC 123 Column3	ABC 123 Column4	ABC 123 Column5
1	null	null	null	null	null
2	null	null	null	null	null
3	null	null	null	null	null
4	null	null	null	null	null
5	UNIQUE NUMBER	null	L_CCY	BALANCE-LCY	BALANCE-INR
6	ISS_DATE	EXP_DATE	APPLICANT NAME	null	null
7	null	null	null	null	null
8	null	null	null	null	null
9	IMBG201911000018	null	INR	732977	732977
10	null	12/31/2022	SHERI NOWMER	null	null
11	MBG6601001	null	INR	730656.89	730656.89

← Row 1

← Row 2

SOLUTION APPROACH STEPS

3 In Report (2): Apply same previous steps but on Row 2

With cleaning complete, the two reports are now ready for

Report merging

AB _C UNIQUE NUMBER	AB _C L_CCY	1.2 BALANCE-LCY	ABC 123 BALANCE-INR
IMBG201911000018	INR	732977	732977
MBG6601001	INR	730656.89	730656.89
MBG6602004	INR	318362.31	318362.31
MBG6603759	INR	509679.99	509679.99
MBG6603930	INR	115922.15	115922.15
MBG6604844	INR	121044.86	121044.86
MBG6604960	INR	0.01	0.01
MBG6700337	INR	45923.52	45923.52
MBG6701527	INR	500000	500000
MBG6701529	INR	3000000	3000000



Report (2)

ISS_DATE	EXP_DATE	AB _C APPLICANT NAME
null	12/31/2022	SHERI NOWMER
3/15/2022	7/31/2023	SHERI NOWMER
5/26/2022	5/31/2023	SHERI NOWMER
10/5/2022	6/30/2023	DERRICK WHELPLY
10/23/2022	6/30/2023	JEANNE DERRY
12/25/2022	6/30/2023	DERRICK WHELPLY
1/3/2023	6/30/2023	SHERI NOWMER
2/9/2023	7/31/2023	SHERI NOWMER
5/11/2023	8/31/2023	SHERI NOWMER
5/11/2023	8/31/2023	SHERI NOWMER

MAKE IT HAPPEN

SOLUTION APPROACH STEPS

4

- To merge them into one sheet, an "Index" column was created in each report. These Index columns contained matching row numbers across the two reports.
- The merge was then performed by selecting and matching the Index columns from the 2 reports using the "Merge Queries" function in Power Query.

Merge

Select tables and matching columns to create a merged table.

Report (1)

ARGIN AMT-ADP	MARGIN AMT-SIGHT	MARGIN AMT-USNCE	TTL MARGIN-LCY	TTL MARGIN-INR	Index
0	0	0	0	0	1
0	0	135342.13	135342.13	135342.13	2
0	0	299.11	299.11	299.11	3
0	0	0	0	0	4

Report (2)

ISS_DATE	EXP_DATE	APPLICANT NAME	Index
null	12/31/2022	SHERI NOWMER	1
3/15/2022	7/31/2023	SHERI NOWMER	2
5/26/2022	5/31/2023	SHERI NOWMER	3
10/5/2022	6/30/2023	DERRICK WHELPLY	4
10/23/2022	6/30/2023	JEANNE DERRY	5

Join Kind

Left Outer (all from first, matching from second)

☐ Use fuzzy matching to perform the merge

► Fuzzy matching options

✓ The selection matches 986 of 986 rows from the first table.

OK Cancel

FINAL CONSOLIDATED DATASET

After merging, we now have a single, consolidated dataset.

The initial issues identified have been resolved:

- Data split across rows has been consolidated
- Null/empty values have been removed
- Unwanted total/subtotal values have been eliminated
- Renamed columns and change types for better clarity

	AB _C APPLICANT NAME	ISS_DATE	EXP_DATE	AB _C UNIQUE NUMBER	AB _C L_CCY	\$ BALANCE-LCY
1	SHERI NOWMER	null	12/31/2022	IMBG201911000018	INR	732,977.00
2	SHERI NOWMER	3/15/2022	7/31/2023	MBG6601001	INR	730,656.89
3	SHERI NOWMER	5/26/2022	5/31/2023	MBG6602004	INR	318,362.31
4	DERRICK WHELPLY	10/5/2022	6/30/2023	MBG6603759	INR	509,679.99
5	JEANNE DERRY	10/23/2022	6/30/2023	MBG6603930	INR	115,922.15
6	DERRICK WHELPLY	12/25/2022	6/30/2023	MBG6604844	INR	121,044.86
7	SHERI NOWMER	1/3/2023	6/30/2023	MBG6604960	INR	0.01
8	SHERI NOWMER	2/9/2023	7/31/2023	MBG6700337	INR	45,923.52
9	SHERI NOWMER	5/11/2023	8/31/2023	MBG6701527	INR	500,000.00
10	SHERI NOWMER	5/11/2023	8/31/2023	MBG6701529	INR	3,000,000.00

BENEFITS AND NEXT STEPS

- **Benefits of Cleaning and Consolidating the Data**
 - Our data's now neatly organized in rows for easy understanding.
 - We've fixed missing info and odd totals, ensuring reliable analysis.
 - It's our go-to source for all financial insights.
- **How This Helps with Better Analysis/Reporting**
 - Analysts can trust the accuracy and completeness of our data.
 - Calculations and trend-spotting are now a breeze.
 - Our reports and visuals are top-notch thanks to high-quality data.
- **Next Steps or Additional Processing**
 - Set up automated data cleaning.
 - Establish data integrity policies.

The background features three vertical stripes on the left: a wide pink stripe, a medium blue stripe, and a narrow beige stripe. The rest of the background is a light beige color, decorated with two rectangular areas of small, light pink dots in the top right and bottom right corners.

THANK YOU!