DATA CLEANING AND ANALYSIS DOCUMENTATION

Project: Loan Disbursement Analysis

---

1. Project Overview

The purpose of this project is to analyze loan disbursements and monitor portfolio health. Main objectives:

1. Calculate **monthly disbursements in EUR** for the last 6 months using `principal_amount_eur`.
2. Visualize **disbursements per loan officer**, showing both count of loans and total volume disbursed for the last month.
3. Calculate **Portfolio at Risk (PAR)** as a KPI:

   PAR = \frac{\text{Sum of principal_amount_outstanding_eur with overdue_days > 30}}{\text{Total principal_amount_outstanding_eur}}

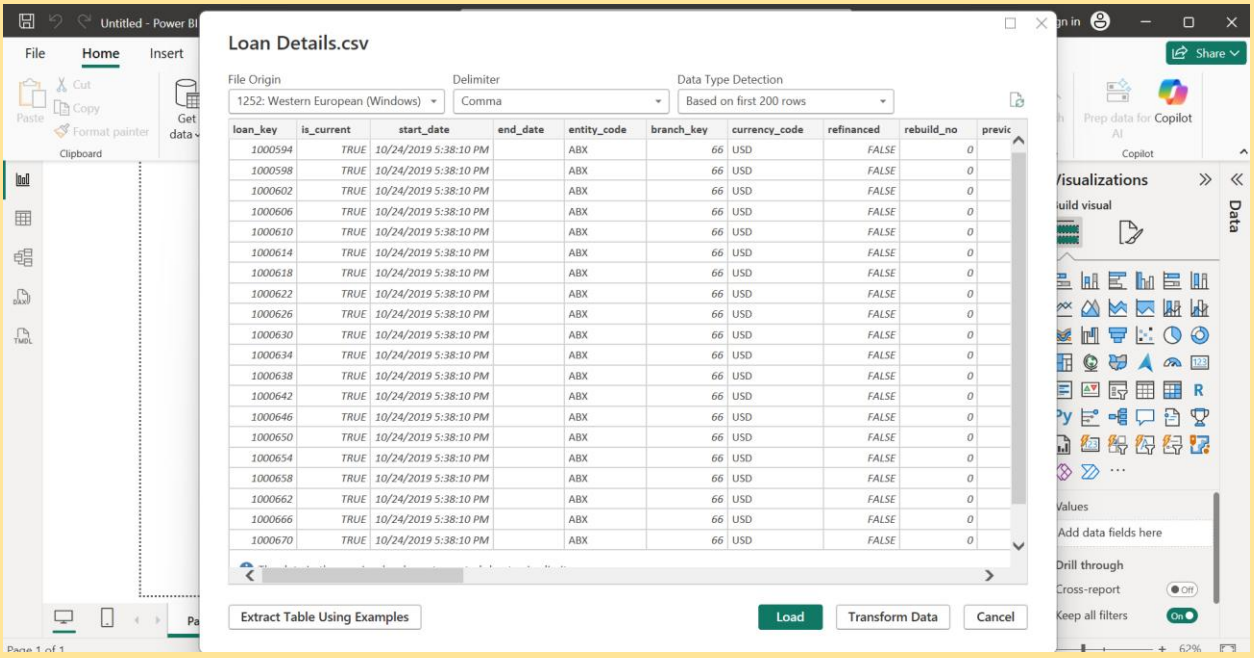
4. Add filters to breakdown by `branch_key`, `product`, and `loan_officer`.

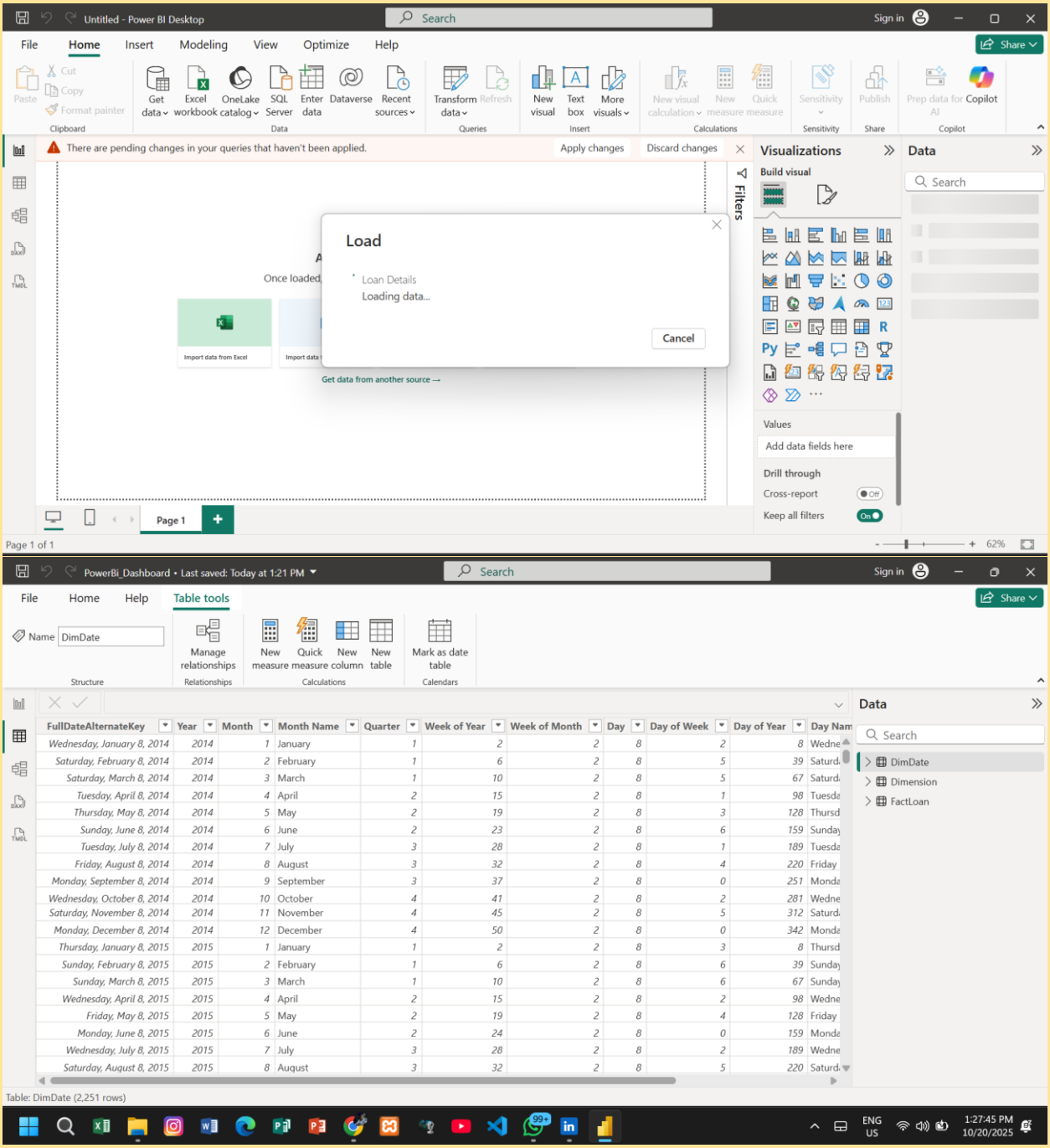
---

2. Loading Data into Power Query

2.1 Steps

1. Open **Power BI Desktop → Home → Get Data**.
2. Select data source and load.
3. Click **Transform Data** to open **Power Query Editor**.


Untitled - Power BI
File
Home
Insert
Share
Loan Details.csv
File Origin
1252: Western European (Windows)
Delimiter
Comma
Data Type Detection
Based on first 200 rows
loan_key | is_current | start_date | end_date | entity_code | branch_key | currency_code | refinanced | rebuild_no | previc
1000594 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000598 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000602 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000606 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000610 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000614 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000618 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000622 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000626 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000630 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000634 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000638 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000642 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000646 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000650 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000654 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000658 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000662 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000666 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
1000670 | TRUE | 10/24/2019 5:38:10 PM | | ABX | 66 | USD | FALSE | 0
Extract Table Using Examples
Load
Transform Data
Cancel
Visualizations
Data
Values
Add data fields here
Drill through
Cross-report
Keep all filters



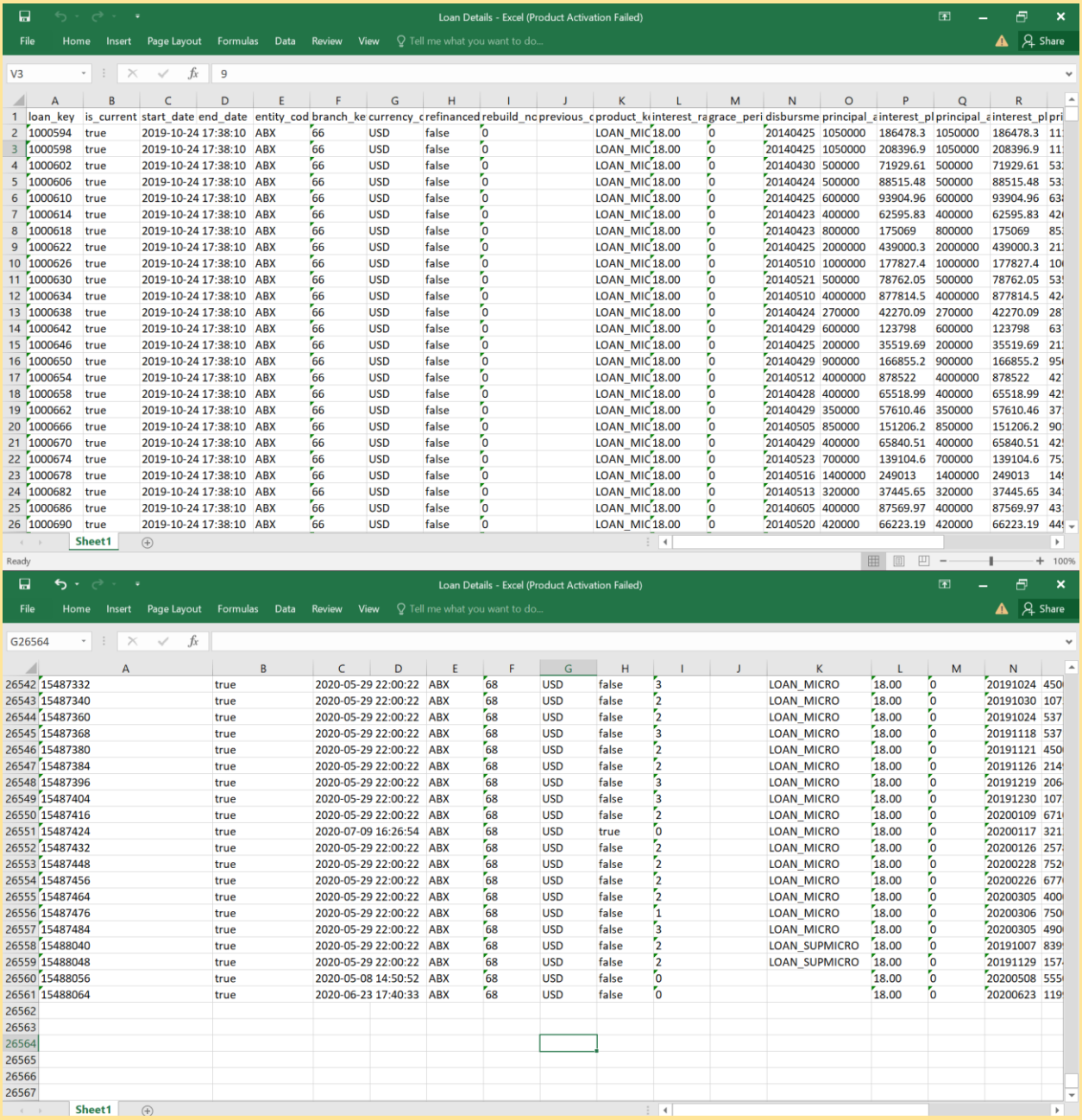
2.2 Initial Data Structure

The dataset contains the following columns:

Column	Description
loan_id	Unique identifier for each loan
principal_amount_eur	Loan principal in EUR

Column	Description
principal_amount_outstanding_eur	Outstanding amount in EUR
disbursement_date	Date of loan disbursement
loan_officer	Officer who disbursed the loan
branch_key	Branch identifier
product	Loan product type
overdue_days	Days overdue for repayment

2.3 Our un organized data we had



Loan Details - Excel (Product Activation Failed)

loan_key	is_current	start_date	end_date	entity_cod	branch_ke	currency_c	refinanced	rebuild_nc	previous_c	product_k	interest_ra	grace_peri	disbursme	principal_a	interest_pl	principal_a	interest_pl	pri
1000594	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140425	1050000	186478.3	1050000	186478.3	11
1000598	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140425	1050000	208396.9	1050000	208396.9	11
1000602	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140430	500000	71929.61	500000	71929.61	53
1000606	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140424	500000	88515.48	500000	88515.48	53
1000610	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140425	600000	93904.96	600000	93904.96	63
1000614	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140423	400000	62595.83	400000	62595.83	42
1000618	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140423	800000	175069	800000	175069	85
1000622	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140425	2000000	439000.3	2000000	439000.3	21
1000626	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140510	1000000	177827.4	1000000	177827.4	10
1000630	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140521	500000	78762.05	500000	78762.05	53
1000634	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140510	4000000	877814.5	4000000	877814.5	42
1000638	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140424	270000	42270.09	270000	42270.09	28
1000642	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140429	600000	123798	600000	123798	63
1000646	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140425	200000	35519.69	200000	35519.69	21
1000650	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140429	900000	166855.2	900000	166855.2	95
1000654	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140512	4000000	878522	4000000	878522	42
1000658	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140428	400000	65518.99	400000	65518.99	42
1000662	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140429	350000	57610.46	350000	57610.46	37
1000666	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140505	850000	151206.2	850000	151206.2	90
1000670	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140429	400000	65840.51	400000	65840.51	42
1000674	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140523	700000	139104.6	700000	139104.6	75
1000678	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140516	1400000	249013	1400000	249013	14
1000682	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140513	320000	37445.65	320000	37445.65	34
1000686	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140605	400000	87569.97	400000	87569.97	43
1000690	true	2019-10-24 17:38:10	ABX	66	USD	false	0		LOAN_MIC	18.00	0	20140520	420000	66223.19	420000	66223.19	44

Loan Details - Excel (Product Activation Failed)

26542	15487332	true	2020-05-29 22:00:22	ABX	68	USD	false	3		LOAN_MICRO	18.00	0	20191024	450
26543	15487340	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20191030	107
26544	15487360	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20191024	537
26545	15487368	true	2020-05-29 22:00:22	ABX	68	USD	false	3		LOAN_MICRO	18.00	0	20191118	537
26546	15487380	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20191121	450
26547	15487384	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20191126	214
26548	15487396	true	2020-05-29 22:00:22	ABX	68	USD	false	3		LOAN_MICRO	18.00	0	20191219	206
26549	15487404	true	2020-05-29 22:00:22	ABX	68	USD	false	3		LOAN_MICRO	18.00	0	20191230	107
26550	15487416	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20200109	671
26551	15487424	true	2020-07-09 16:26:54	ABX	68	USD	true	0		LOAN_MICRO	18.00	0	20200117	321
26552	15487432	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20200126	257
26553	15487448	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20200228	752
26554	15487456	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20200226	677
26555	15487464	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_MICRO	18.00	0	20200305	400
26556	15487476	true	2020-05-29 22:00:22	ABX	68	USD	false	1		LOAN_MICRO	18.00	0	20200306	750
26557	15487484	true	2020-05-29 22:00:22	ABX	68	USD	false	3		LOAN_MICRO	18.00	0	20200305	490
26558	15488040	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_SUPMICRO	18.00	0	20191007	839
26559	15488048	true	2020-05-29 22:00:22	ABX	68	USD	false	2		LOAN_SUPMICRO	18.00	0	20191129	157
26560	15488056	true	2020-05-08 14:50:52	ABX	68	USD	false	0			18.00	0	20200508	555
26561	15488064	true	2020-06-23 17:40:33	ABX	68	USD	false	0			18.00	0	20200623	119

3. NULL VALUE HANDLING

3.1 Identifying Null Values

Each column was analyzed for nulls.

Column	Null Count	Percentage
principal_amount_eur	_____	____%
principal_amount_outstanding_eur	_____	____%
disbursement_date	_____	____%
loan_officer	_____	____%
branch_key	_____	____%
product	_____	____%
overdue_days	_____	____%

3.2 Null Value Treatment Strategy

Critical Columns (Cannot be null – remove rows):

- `principal_amount_eur`
- `disbursement_date`

Non-Critical Columns (Replace nulls with defaults):

- `loan_officer` → "Unassigned"
- `branch_key` → "Unknown Branch"
- `product` → "Not Specified"
- `overdue_days` → 0

3.3 Removing Nulls in Critical Fields

Step 1: Filter out null `disbursement_date`

Power Query Formula:

```
= Table.SelectRows(#"Filtered Rows", each [disbursement_date] <> null)
```

Result: Removed _____ rows

3.4 Replacing Nulls in Non-Critical Fields

Step 2: Replace null `loan_officer` values

```
= Table.ReplaceValue(#"Filtered Rows1", null, "Unassigned",
Replacer.ReplaceValue, {"loan_officer"})
```

Result: Replaced _____ null values

Step 3: Replace null `branch_key` values

```
= Table.ReplaceValue(#"Replaced Value", null, "Unknown Branch",
Replacer.ReplaceValue, {"branch_key"})
```

Result: Replaced _____ null values

Step 4: Replace null `product` values

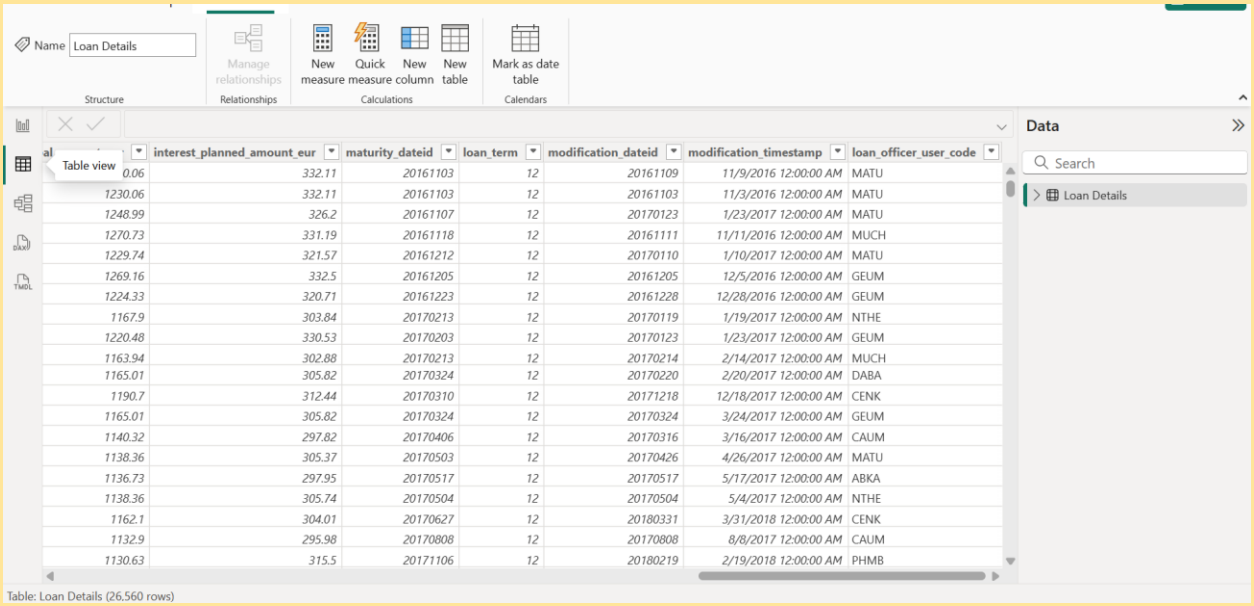
```
= Table.ReplaceValue(#"Replaced Value1", null, "Not Specified",
Replacer.ReplaceValue, {"product"})
```

Result: Replaced _____ null values

Step 5: Replace null `overdue_days` values

```
= Table.ReplaceValue(#"Replaced Value2", null, 0, Replacer.ReplaceValue,
{"overdue_days"})
```

Result: Replaced _____ null values



	interest_planned_amount_eur	maturity_dateid	loan_term	modification_dateid	modification_timestamp	loan_officer_user_code
0.06	332.11	20161103	12	20161109	11/9/2016 12:00:00 AM	MATU
1230.06	332.11	20161103	12	20161103	11/3/2016 12:00:00 AM	MATU
1248.99	326.2	20161107	12	20170123	1/23/2017 12:00:00 AM	MATU
1270.73	331.19	20161118	12	20161111	11/11/2016 12:00:00 AM	MUCH
1229.74	321.57	20161212	12	20170110	1/10/2017 12:00:00 AM	MATU
1269.16	332.5	20161205	12	20161205	12/5/2016 12:00:00 AM	GEUM
1224.33	320.71	20161223	12	20161228	12/28/2016 12:00:00 AM	GEUM
1167.9	303.84	20170213	12	20170119	1/19/2017 12:00:00 AM	NTHE
1220.48	330.53	20170203	12	20170123	1/23/2017 12:00:00 AM	GEUM
1163.94	302.88	20170213	12	20170214	2/14/2017 12:00:00 AM	MUCH
1165.01	305.82	20170324	12	20170220	2/20/2017 12:00:00 AM	DABA
1190.7	312.44	20170310	12	20171218	12/18/2017 12:00:00 AM	CENK
1165.01	305.82	20170324	12	20170324	3/24/2017 12:00:00 AM	GEUM
1140.32	297.82	20170406	12	20170316	3/16/2017 12:00:00 AM	CAUM
1138.36	305.37	20170503	12	20170426	4/26/2017 12:00:00 AM	MATU
1136.73	297.95	20170517	12	20170517	5/17/2017 12:00:00 AM	ABKA
1138.36	305.74	20170504	12	20170504	5/4/2017 12:00:00 AM	NTHE
1162.1	304.01	20170627	12	20180331	3/31/2018 12:00:00 AM	CENK
1132.9	295.98	20170808	12	20170808	8/8/2017 12:00:00 AM	CAUM
1130.63	315.5	20171106	12	20180219	2/19/2018 12:00:00 AM	PHMB

---

4. DATA TYPE VALIDATION AND CORRECTION

4.1 Identifying Incorrect Data Types

Column	Detected Type	Required Type	Issue
principal_amount_eur	Text	Decimal Number	Cannot calculate

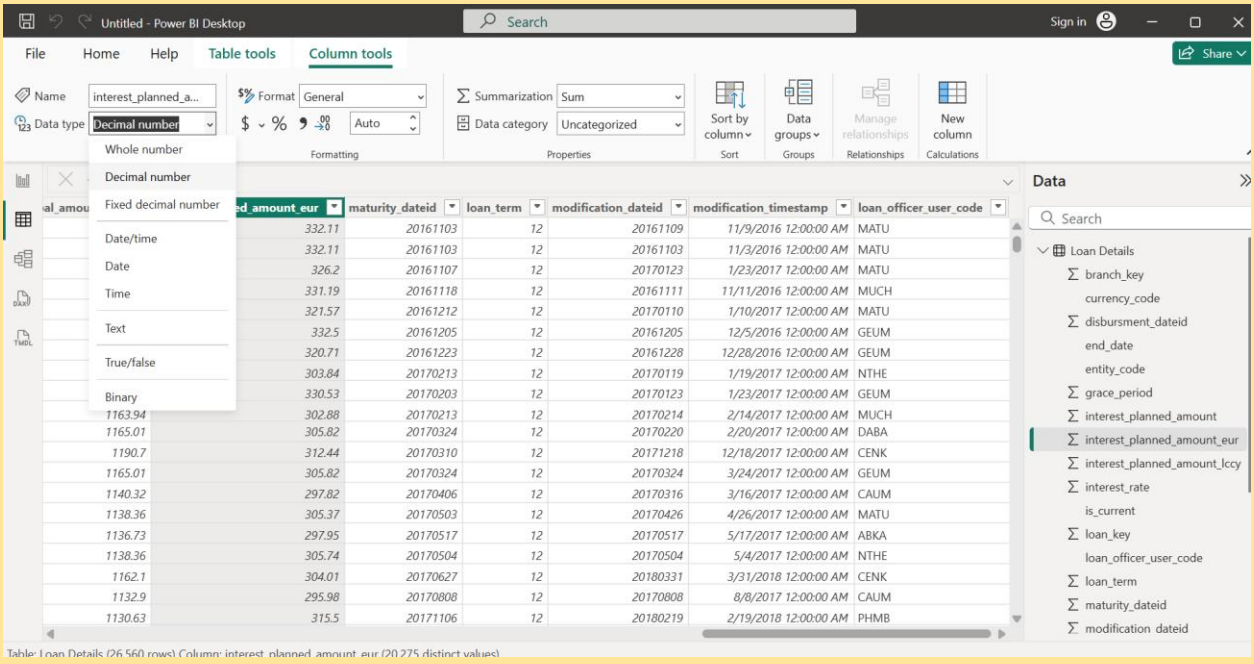
Column	Detected Type	Required Type	Issue
principal_amount_outstanding_eur	Text	Decimal Number	Cannot calculate
disbursement_date	Text	Date	Cannot filter by date
overdue_days	Text	Whole Number	Cannot calculate

4.2 Correcting Data Types

- Change `principal_amount_eur` and `principal_amount_outstanding_eur` to **Decimal**
- Change `disbursement_date` to **Date**
- Change `overdue_days` to **Whole Number**

Power Query Formulas:

```
= Table.TransformColumnTypes(#"Replaced Value3", {{"principal_amount_eur",
type number}})
= Table.TransformColumnTypes(#"Changed Type1",
{{"principal_amount_outstanding_eur", type number}})
= Table.TransformColumnTypes(#"Changed Type2", {{"disbursement_date", type
date}})
= Table.TransformColumnTypes(#"Changed Type3", {{"overdue_days",
Int64.Type}})
```



5. DUPLICATE REMOVAL

5.1 Checking for Duplicates

Duplicates were checked using `loan_id`.

Power Query Formula:

```
= Table.Distinct(#"Changed Type4", {"loan_id"})
```

Result: Removed _____ duplicate rows

6. ANALYSIS TASKS

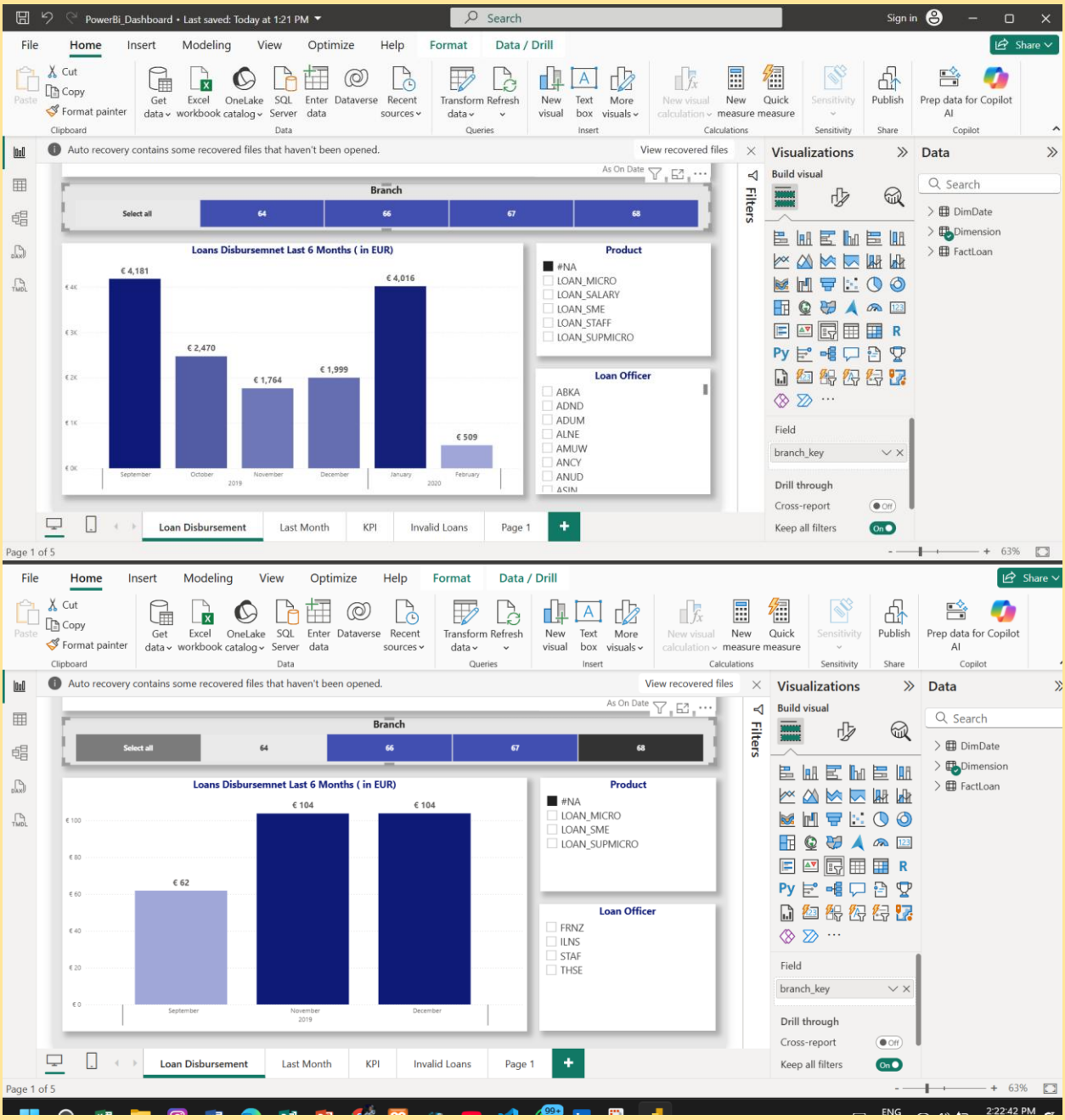
6.1 Monthly Disbursements (Last 6 Months)

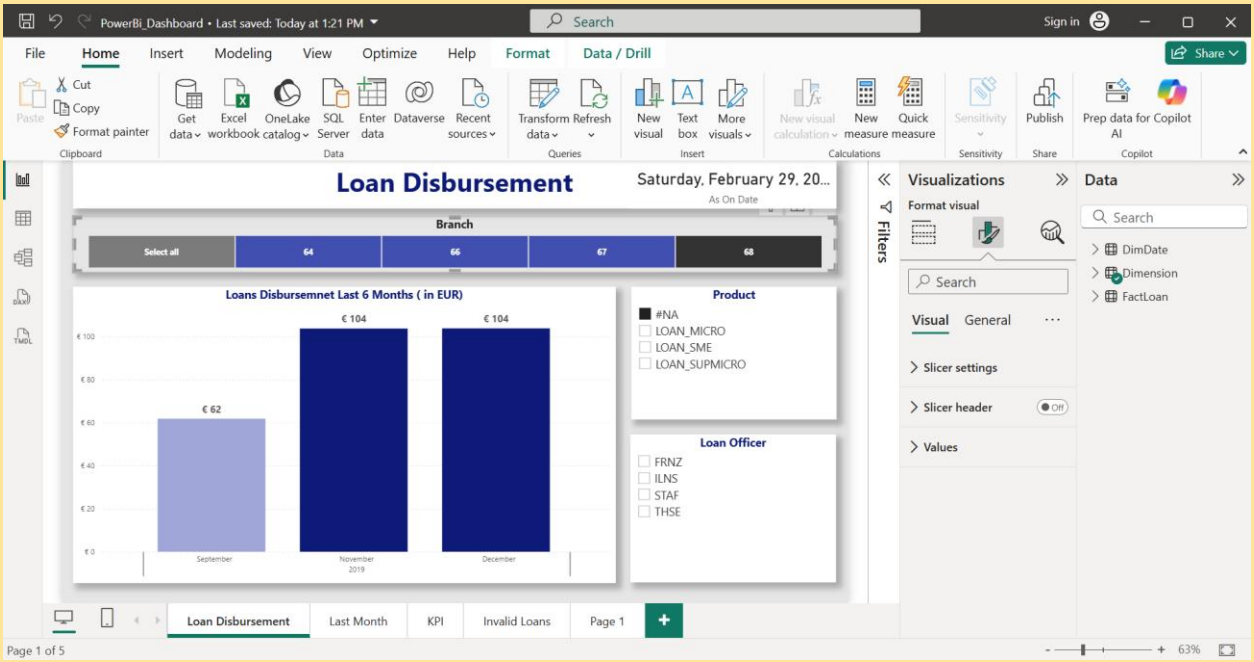
Measure in DAX:

```
Monthly_Disbursement_EUR =
CALCULATE(
    SUM(Loans[principal_amount_eur]),
    DATESINPERIOD(
        DateTable[Date],
        MAX(DateTable[Date]),
        -6,
        MONTH
    )
)
```

Visualization: Column or line chart

- **Axis:** Month-Year
- **Values:** `Monthly_Disbursement_EUR`


PowerBi_Dashboard • Last saved: Today at 1:21 PM
Search
Sign in
File
Home
Insert
Modeling
View
Optimize
Help
Format
Data / Drill
Share
Cut
Copy
Format painter
Clipboard
Get data
Excel workbook
OneLake catalog
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New visual calculation
New measure
Quick measure
Calculations
Sensitivity
Sensitivity
Publish
Share
Prep data for Copilot AI
Copilot
Auto recovery contains some recovered files that haven't been opened.
View recovered files
As On Date
Branch
Select all
64
66
67
68
Loans Disbursemnet Last 6 Months (in EUR)
€ 4,181
€ 2,470
€ 1,764
€ 1,999
€ 4,016
€ 509
September
October
November
December
January
February
2019
2020
Product
#NA
LOAN_MICRO
LOAN_SALARY
LOAN_SME
LOAN_STAFF
LOAN_SUPMICRO
Loan Officer
ABKA
ADND
ADUM
ALNE
AMUW
ANCY
ANUD
ASIN
Filters
Visualizations
Build visual
Field
branch_key
Drill through
Cross-report
Off
Keep all filters
On
Data
Search
DimDate
Dimension
FactLoan
Loan Disbursement
Last Month
KPI
Invalid Loans
Page 1
Page 1 of 5
63%
Branch
Select all
64
66
67
68
Loans Disbursemnet Last 6 Months (in EUR)
€ 62
€ 104
€ 104
September
November
2019
December
Product
#NA
LOAN_MICRO
LOAN_SME
LOAN_SUPMICRO
Loan Officer
FRNZ
ILNS
STAF
THSE
Loan Disbursement
Last Month
KPI
Invalid Loans
Page 1
Page 1 of 5
63%
ENG
2:22:42 PM



6.2 Disbursements per Loan Officer (Last Month)

Measures in DAX:

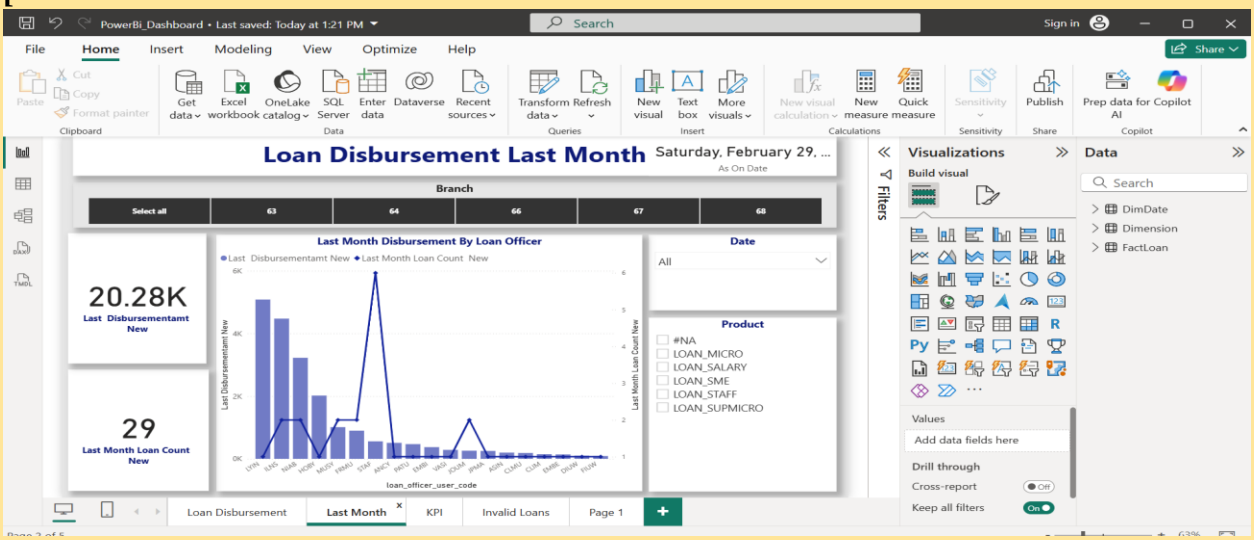
```
LastMonthLoans =
CALCULATE(
    COUNTROWS(Loans),
    FILTER(
        Loans,
        MONTH(Loans[disbursement_date]) = MONTH(TODAY())-1 &&
        YEAR(Loans[disbursement_date]) = YEAR(TODAY())
    )
)

LastMonthDisbursedVolume =
CALCULATE(
    SUM(Loans[principal_amount_eur]),
    FILTER(
        Loans,
        MONTH(Loans[disbursement_date]) = MONTH(TODAY())-1 &&
        YEAR(Loans[disbursement_date]) = YEAR(TODAY())
    )
)
```

Visualization: Clustered column chart or combo chart

- **Axis:** `loan_officer`
- **Values:** Count of loans and Total Volume

[

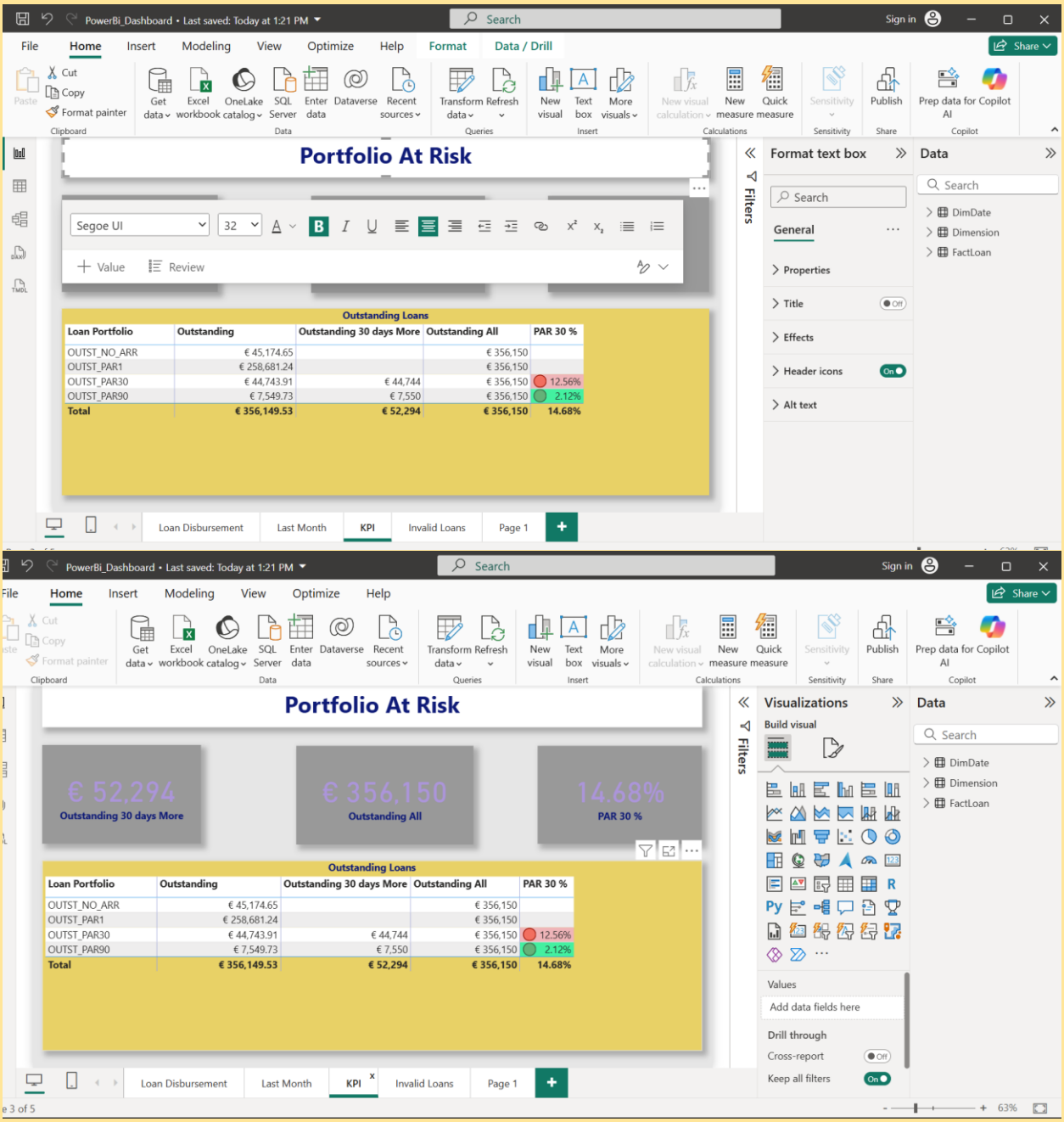


6.3 Portfolio at Risk (PAR) KPI

Measure in DAX:

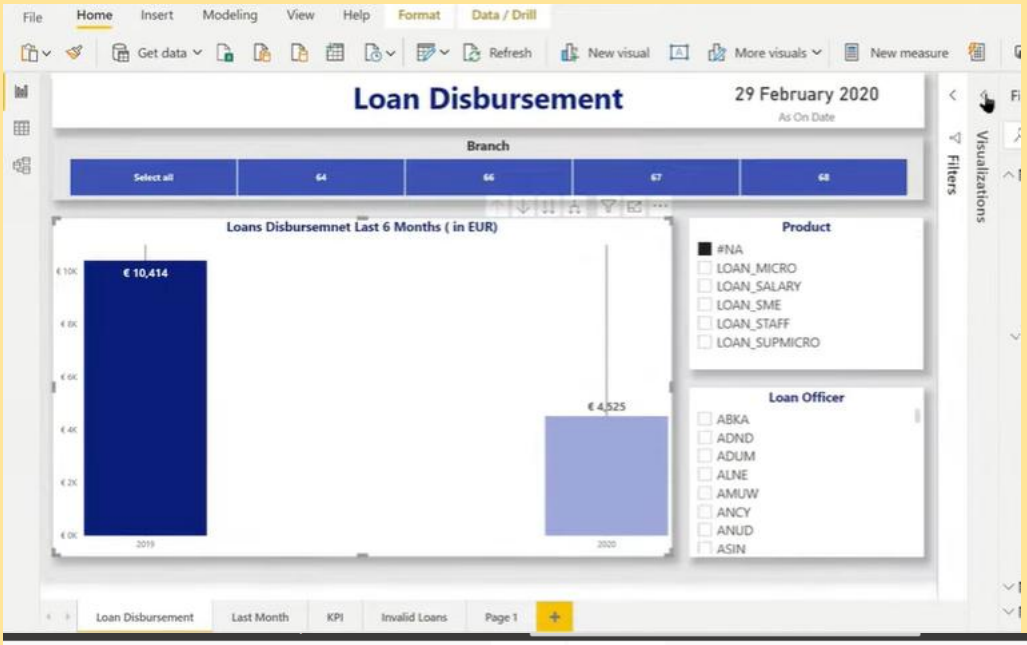
```
PAR =
DIVIDE(
    CALCULATE(
        SUM(Loans[principal_amount_outstanding_eur]),
        Loans[overdue_days] > 30
    ),
    SUM(Loans[principal_amount_outstanding_eur])
)
```

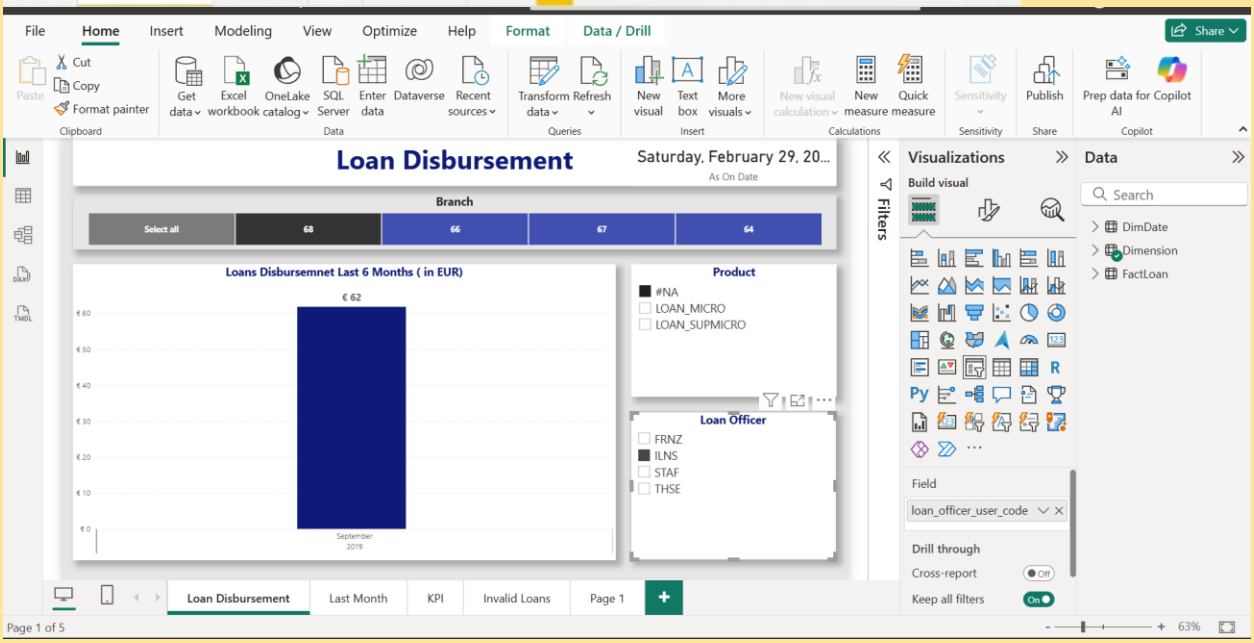
Visualization: KPI Card formatted as **Percentage**

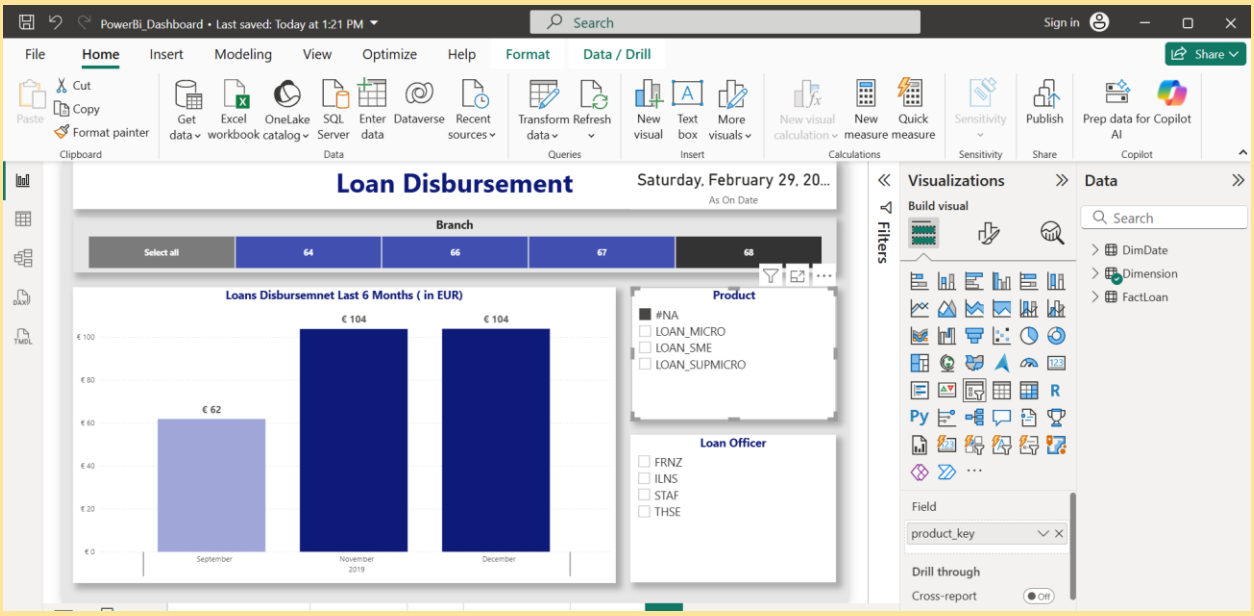


6.4 Adding Filters

- Add slicers for `branch_key`, `product`, and `loan_officer`
- Visuals update dynamically when filters are applied


File
Home
Insert
Modeling
View
Help
Format
Data / Drill
Get data
Refresh
New visual
More visuals
New measure
Loan Disbursement
29 February 2020
As On Date
Branch
Select all
64
66
67
68
Loans Disbursemnet Last 6 Months (in EUR)
€ 10,414
€ 4,525
2019
2020
Product
#NA
LOAN_MICRO
LOAN_SALARY
LOAN_SME
LOAN_STAFF
LOAN_SUPMICRO
Loan Officer
ABKA
ADND
ADUM
ALNE
AMUW
ANCY
ANUD
ASIN
Filters
Visualizations
Loan Disbursement
Last Month
KPI
Invalid Loans
Page 1


File
Home
Insert
Modeling
View
Optimize
Help
Format
Data / Drill
Share
Cut
Copy
Format painter
Paste
Clipboard
Get data
Excel workbook
OneLake catalog
SQL Server
Enter data
Dataverse
Recent sources
Data
Transform data
Refresh
Queries
New visual
Text box
More visuals
Insert
New visual calculation
New measure
Quick measure
Calculations
Sensitivity
Publish
Share
Prep data for Copilot AI
Copilot
Loan Disbursement
Saturday, February 29, 20...
As On Date
Branch
Select all
68
66
67
64
Loans Disbursemnet Last 6 Months (in EUR)
€ 62
September 2019
Product
#NA
LOAN_MICRO
LOAN_SUPMICRO
Loan Officer
FRNZ
ILNS
STAF
THSE
Filters
Visualizations
Build visual
Field
loan_officer_user_code
Drill through
Cross-report
Keep all filters
Data
Search
DimDate
Dimension
FactLoan
Loan Disbursement
Last Month
KPI
Invalid Loans
Page 1
Page 1 of 5
63%



7. Conclusion

- Dataset is cleaned, nulls handled, duplicates removed, and data types corrected.
- KPIs and charts are ready to provide insights into disbursements and portfolio risk.
- Filters allow dynamic analysis by branch, product, and loan officer.