DATA SCIENCE AND MACHINE LEARNING

MODULE – 3 MICROSOFT POWER BI

CAPSTONE PROJECT

PROJECT TITLE: BANK LOAN PERFORMANCE ANALYSIS

MENTOR:
ARCHANA S
[DSML]

BY
M.BALA SARAVANAN
(TN DSML APR 2024)
[23-07-2024 - Tuesday]

Acknowledgements

I would like to express my heartfelt gratitude to my mentor, Archana S, for her invaluable guidance and unwavering support throughout this project. Her insightful feedback and encouragement have been instrumental in shaping my research and helping me achieve my goals. I am deeply appreciative of her dedication and commitment to my academic growth. Thank you, madam, for being an exceptional mentor and a source of inspiration.

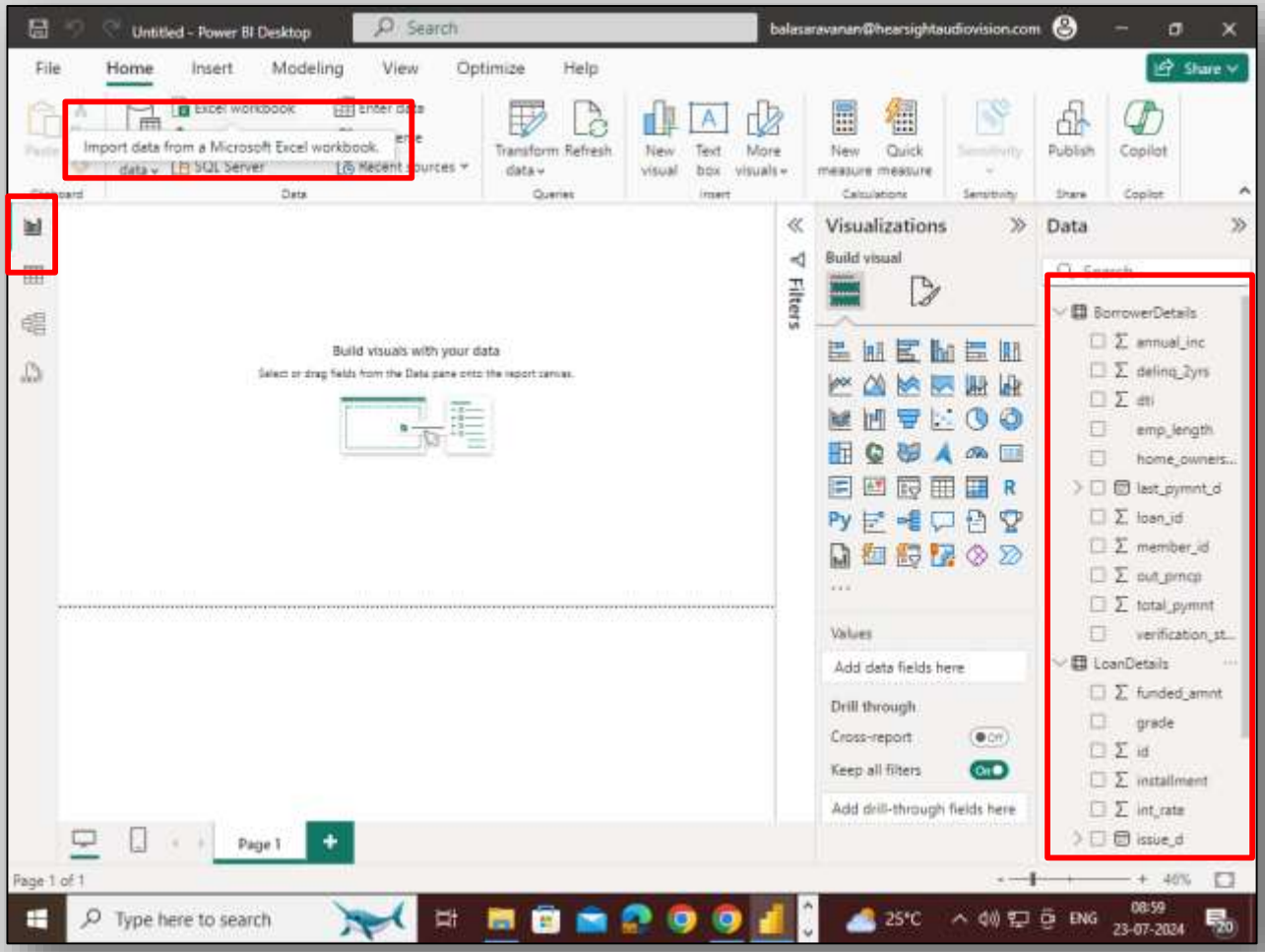
ENTRI elevate தமிழ்

Table Of Content

ENTRI

1) Importing Data

Import the "LoanDetails" and "BorrowerDetails" sheets from the "bank loan.xlsx“ file into Power BI.



2) Transformation Using Power Query

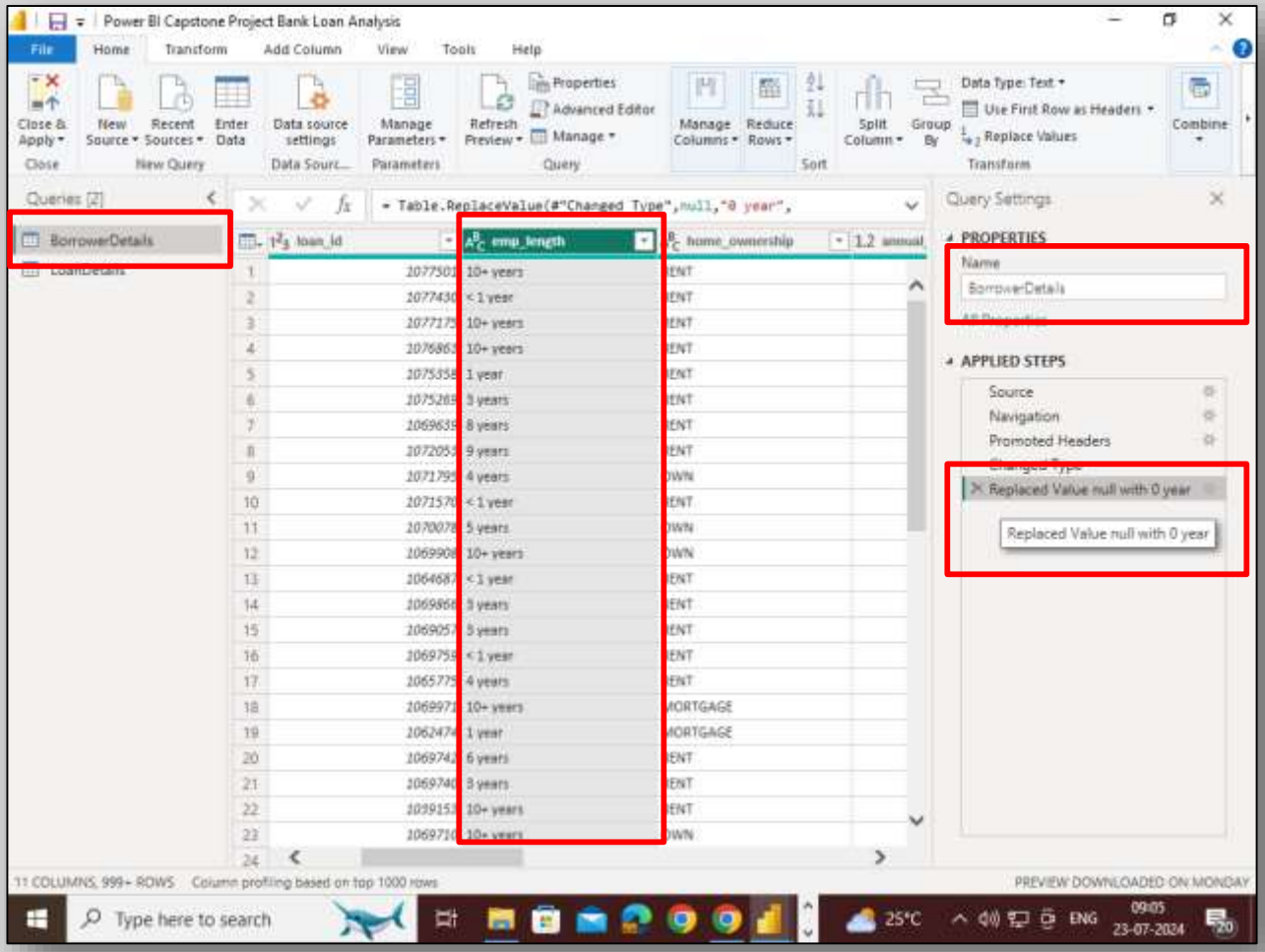
A. Data Cleaning.

B. Data Transformation.

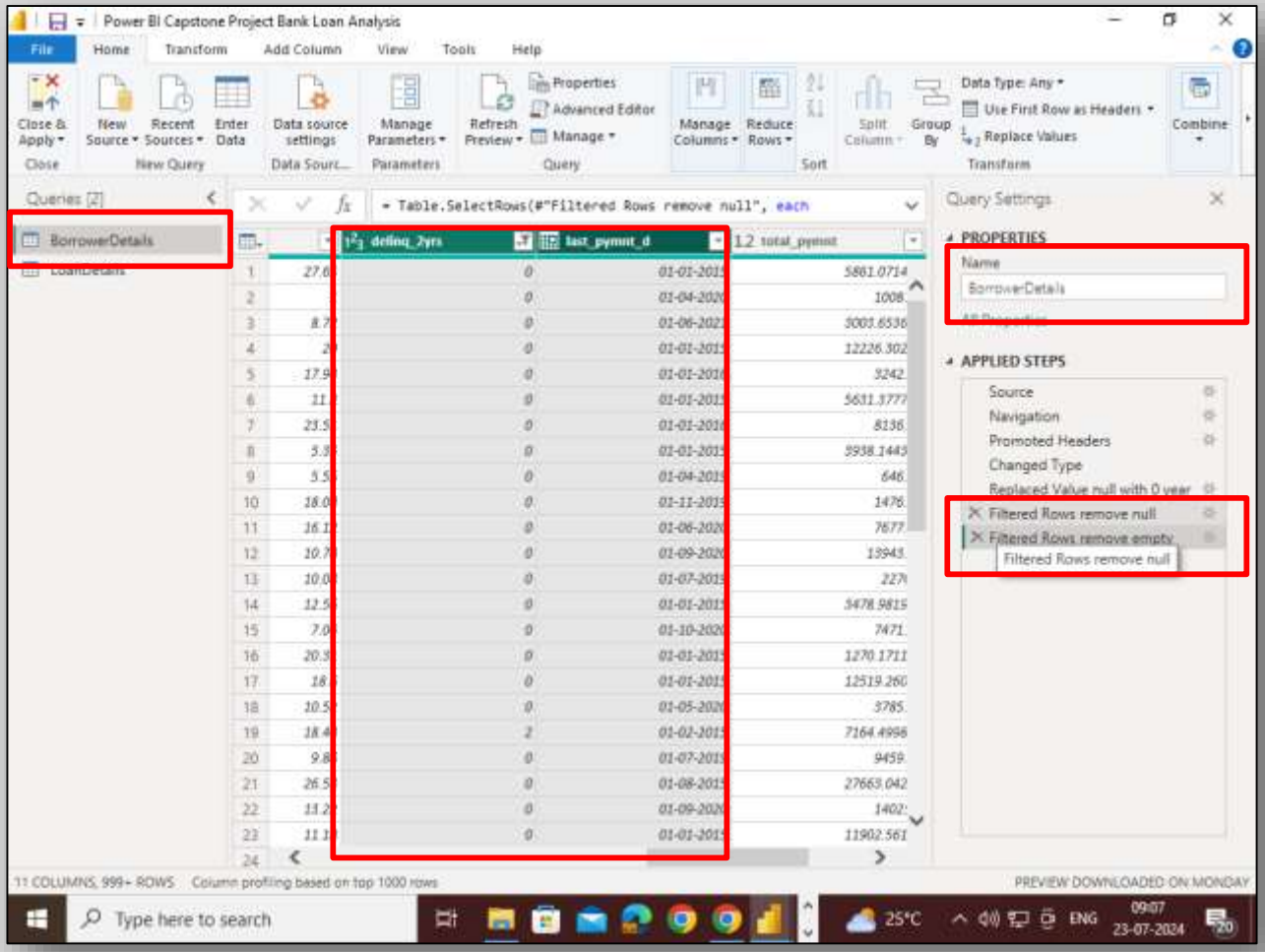
2.A) Data Cleaning

I. *Handling Missing Values and Duplicates.*
II. *Dealing with Inconsistencies.*

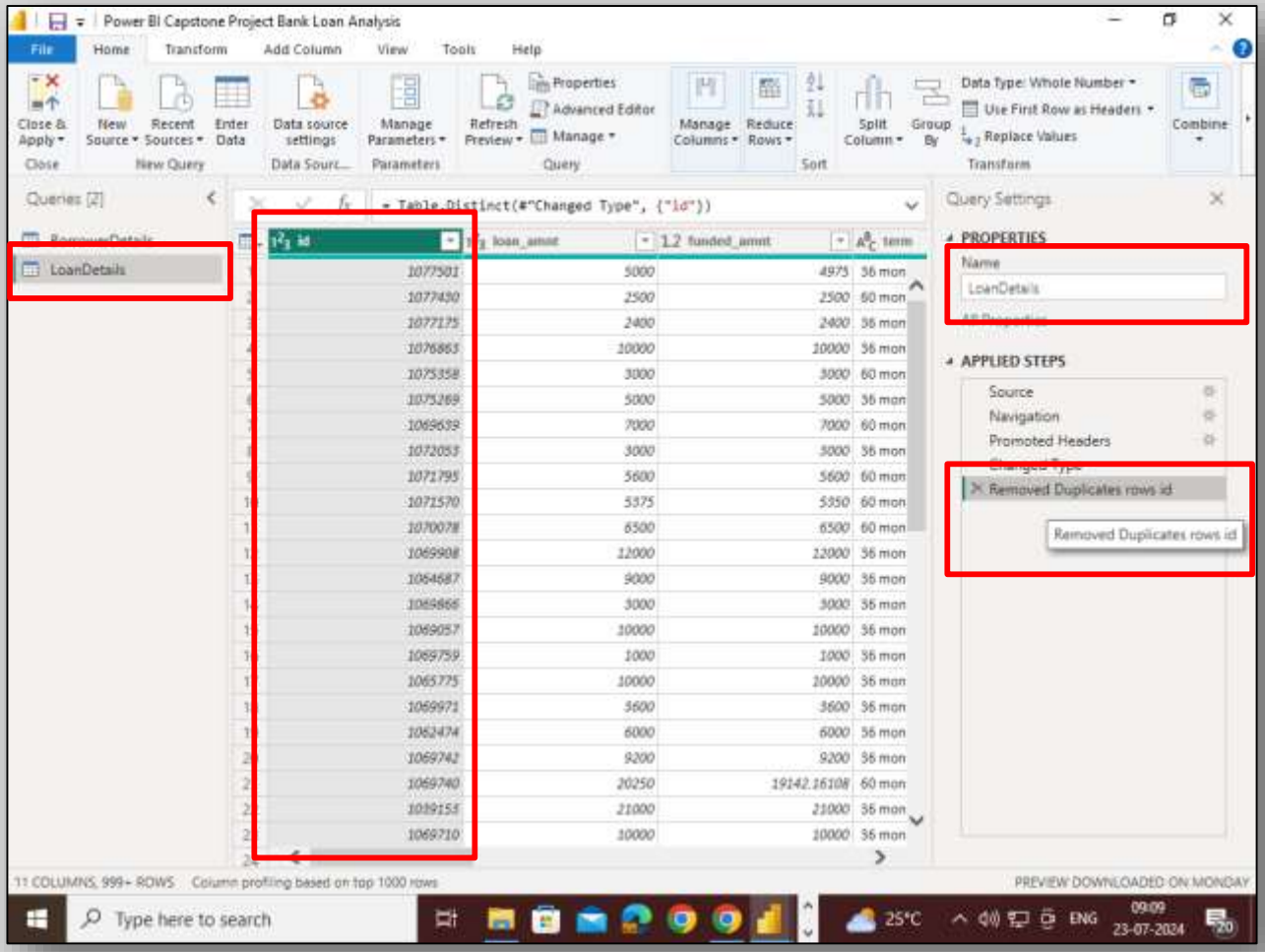
2.A I) Replace missing values (null) in the 'emp_length' column of the "BorrowerDetails" table with '0 year'.



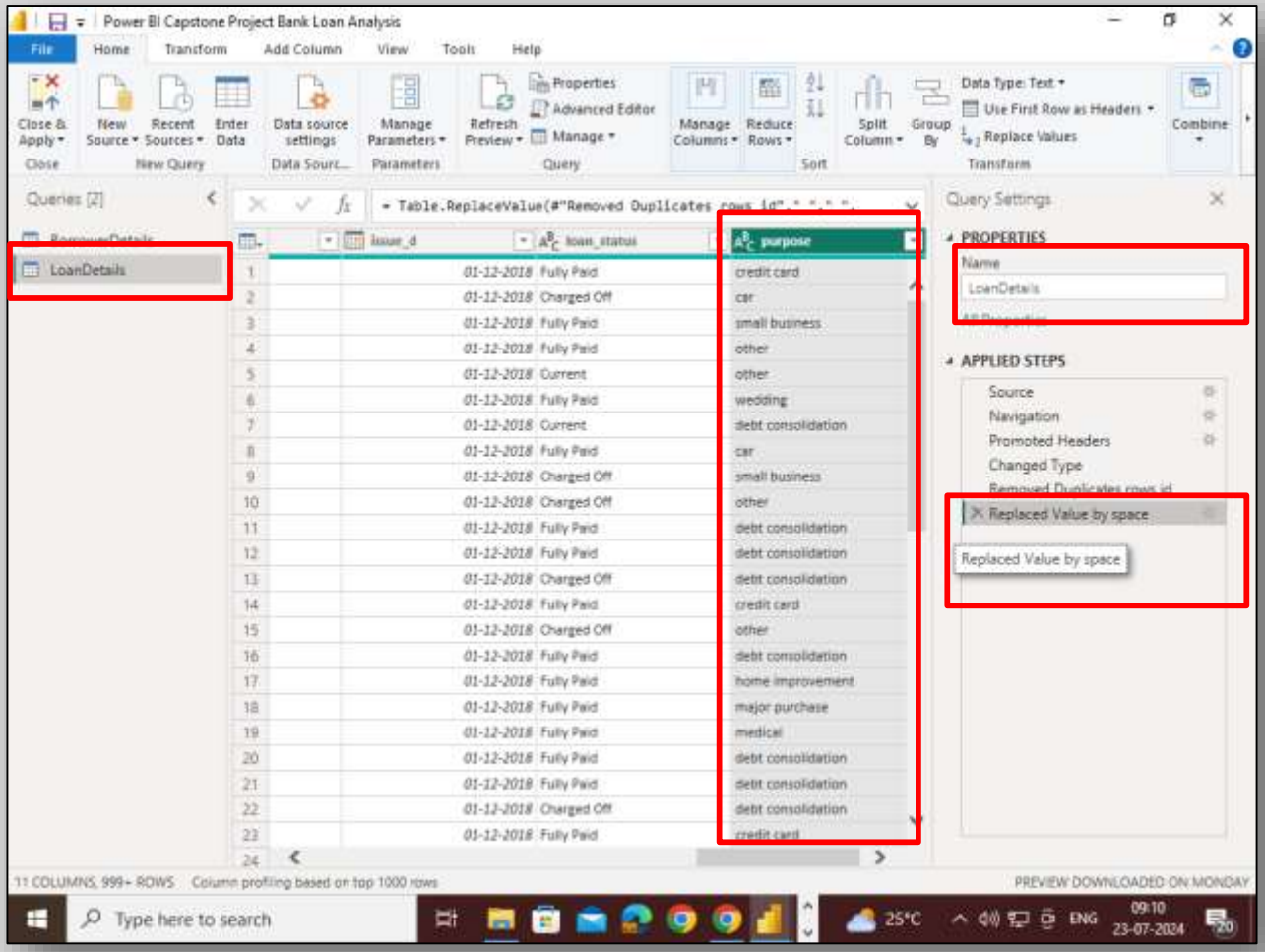
2.A I) Remove rows with missing values in the 'last_pymnt_d' and 'delinq_2yrs' columns.



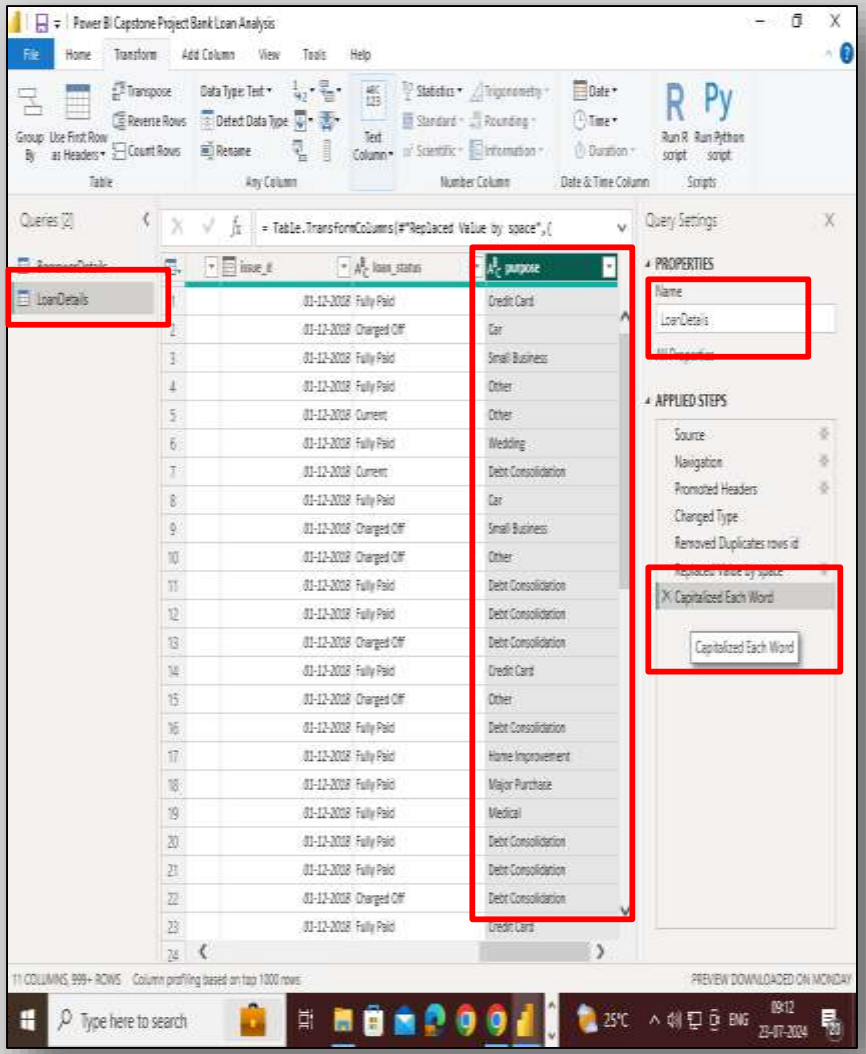
2.A I)Remove duplicate rows in the 'id' column of the "LoanDetails" table.

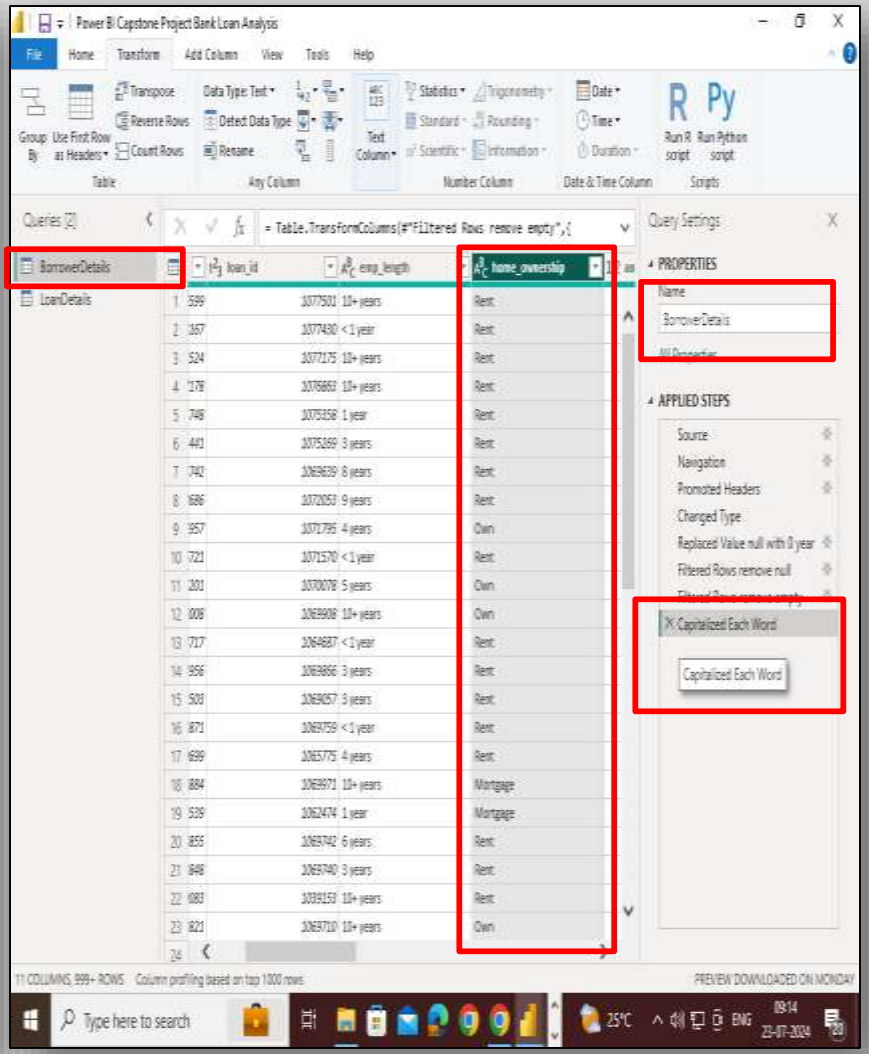


2.A II) Ensure words in the 'purpose' column are separated by spaces instead of underscores (e.g., "credit card" instead of "credit_card").



2.A II) Format the 'purpose' and 'home_ownership' columns to proper case.





2.B) Data Transformation

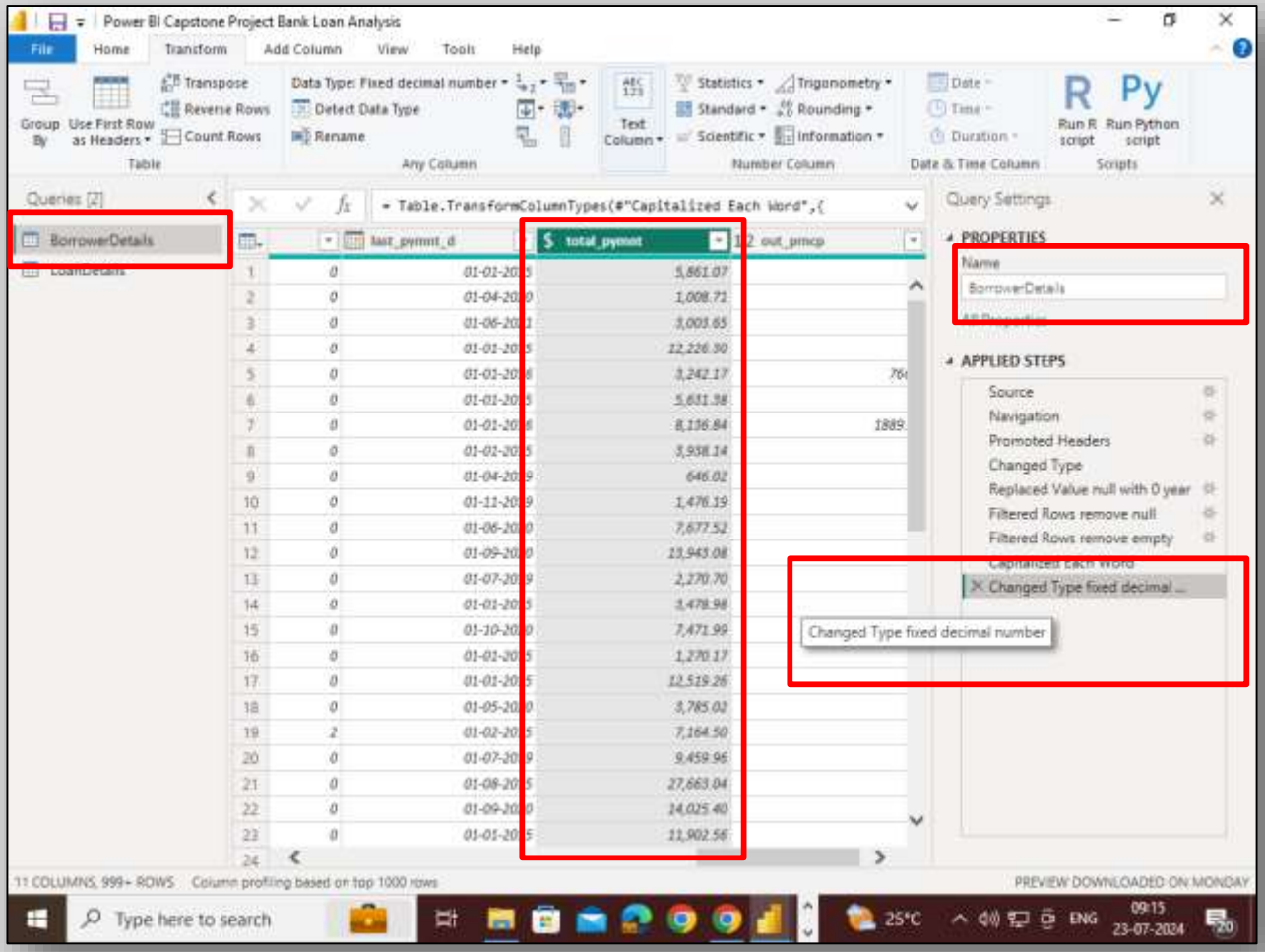
I. Column Transformation.

II. Column Renaming.

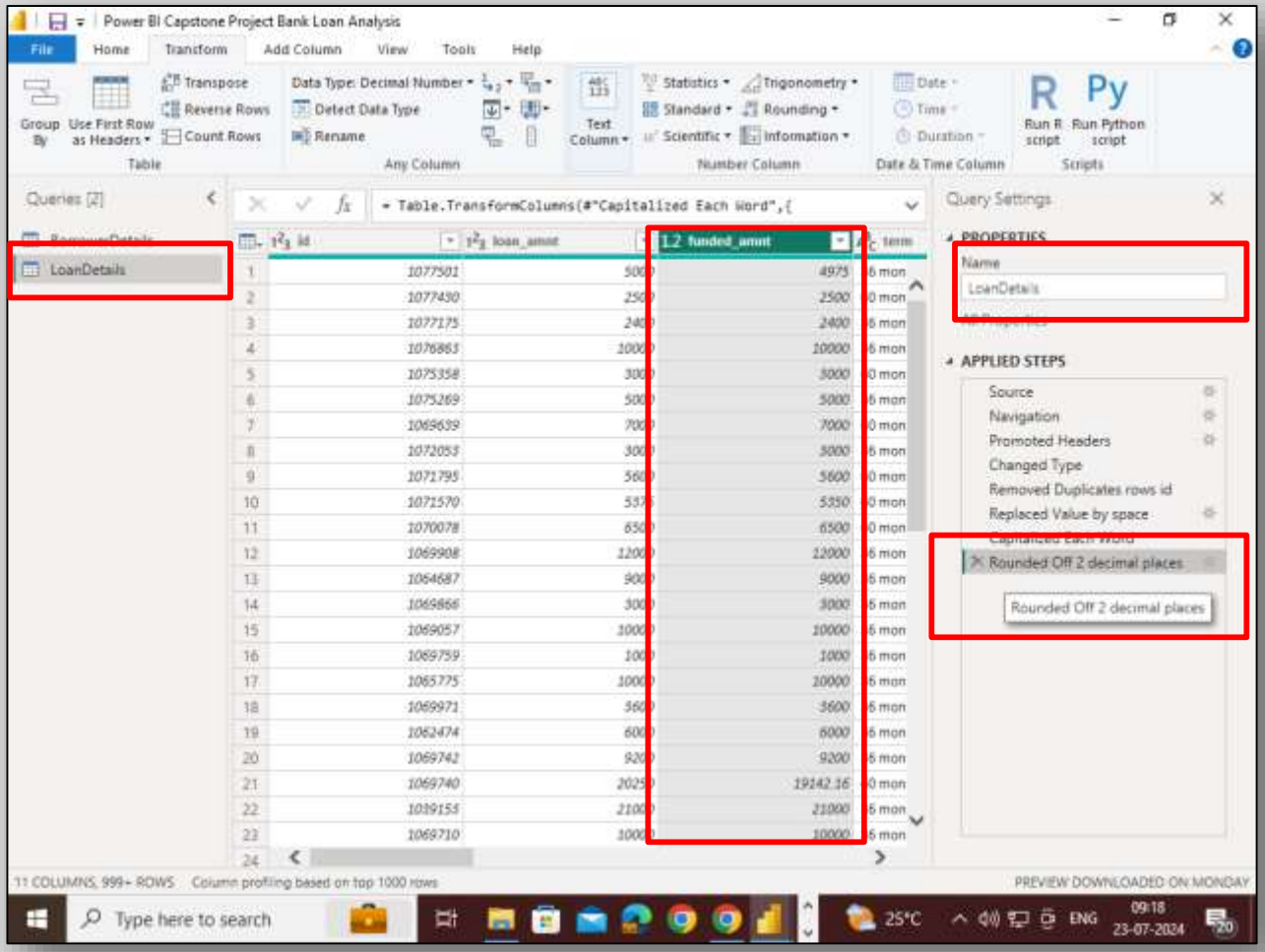
III. Creating New Columns.

IV. Column Dropping.

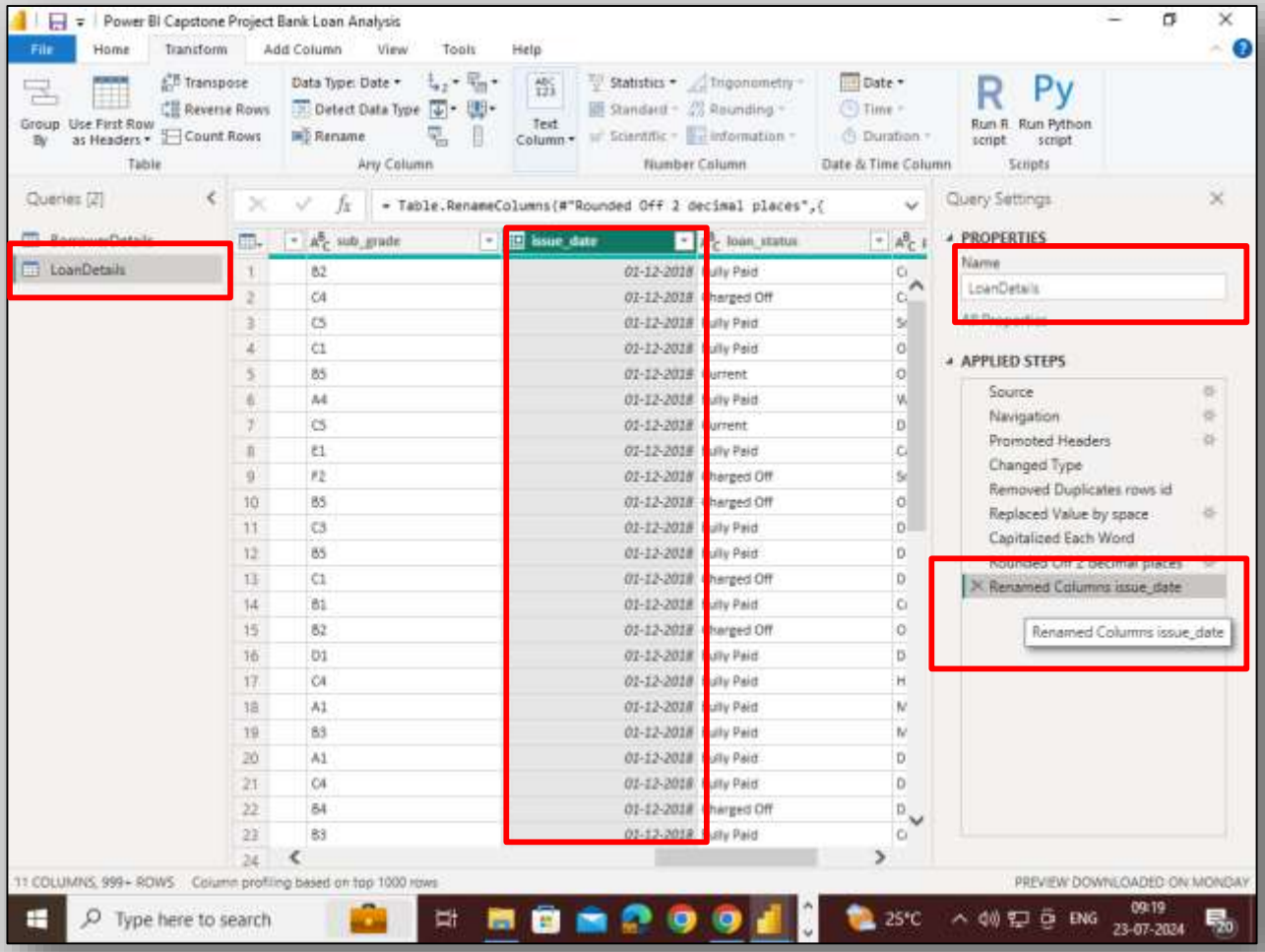
2.B I) Change the data type of the 'total_pymnt' column to 'Fixed decimal number'.



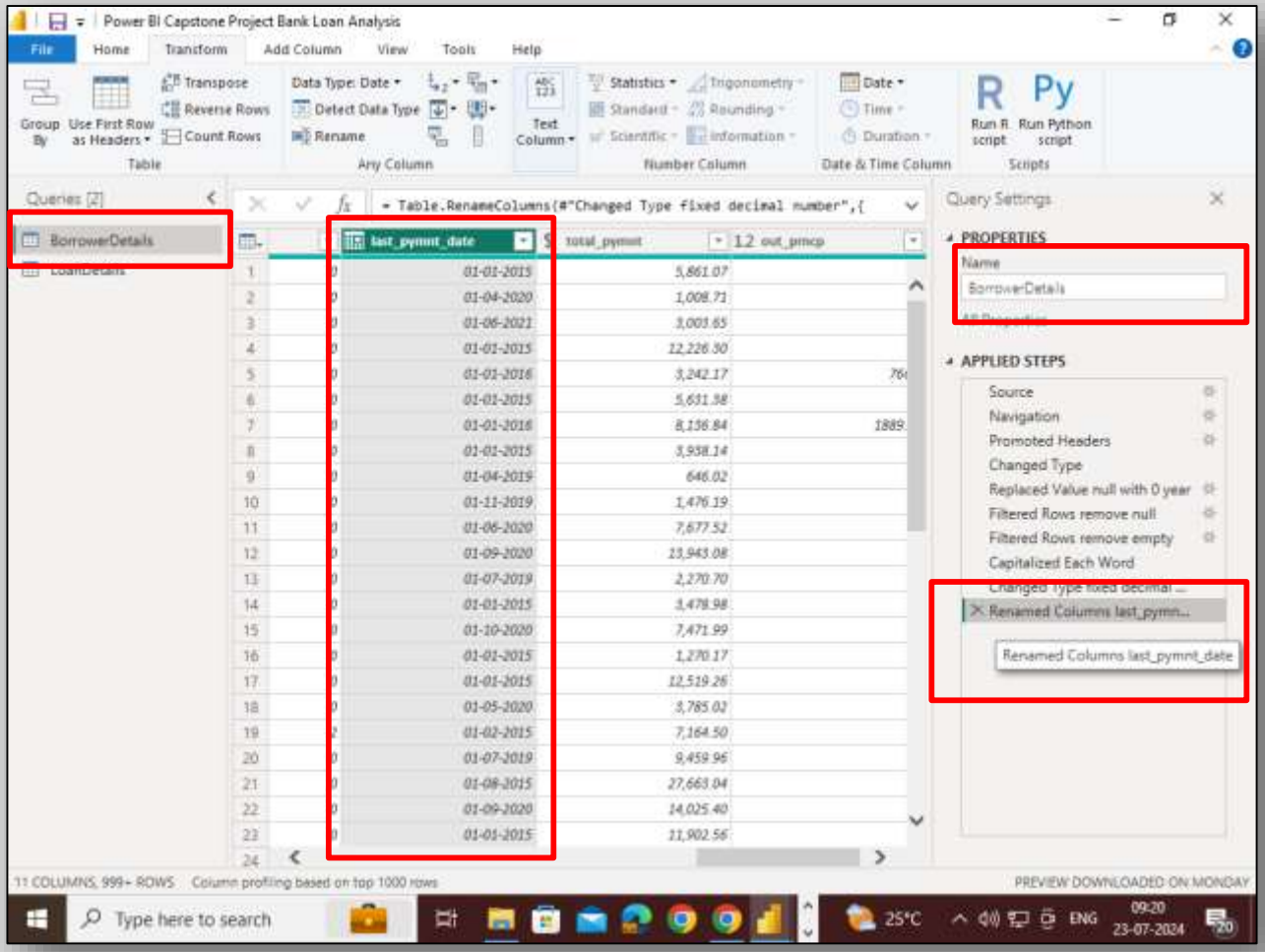
2.B I) Round off the numbers in the 'funded_amnt' column to 2 decimal places.



2.B II) Rename the column 'issue_d' to 'issue_date'.

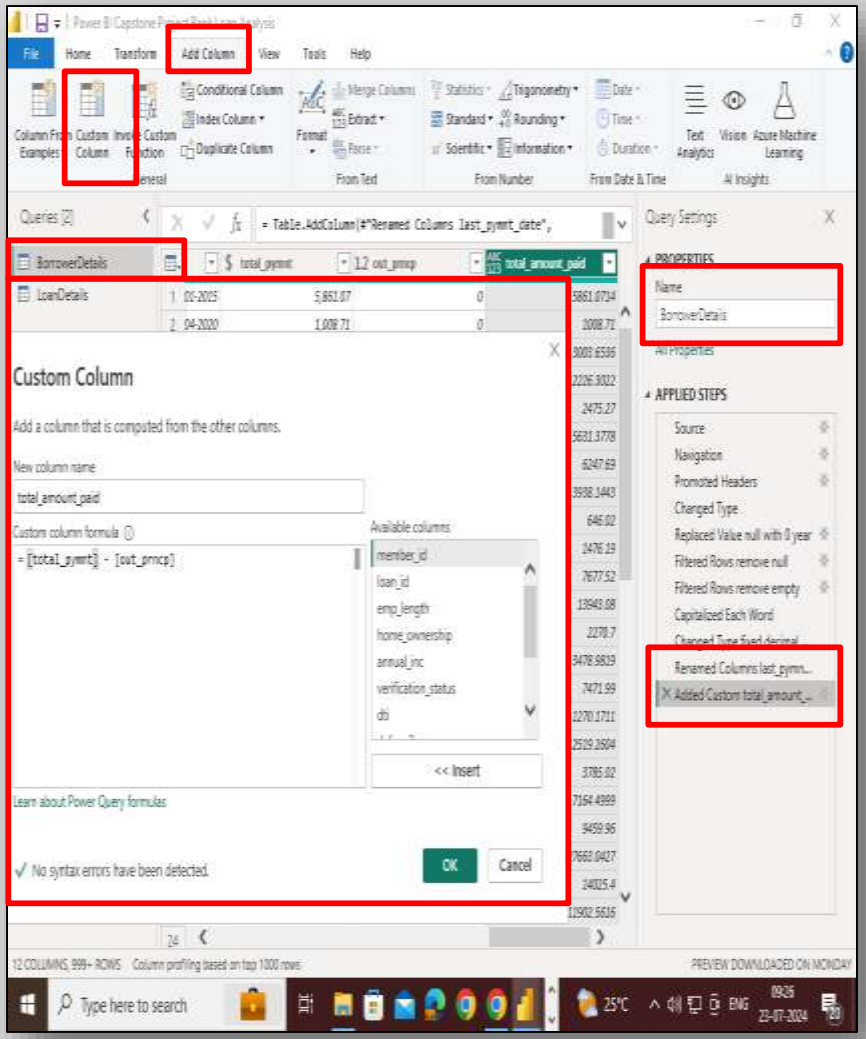


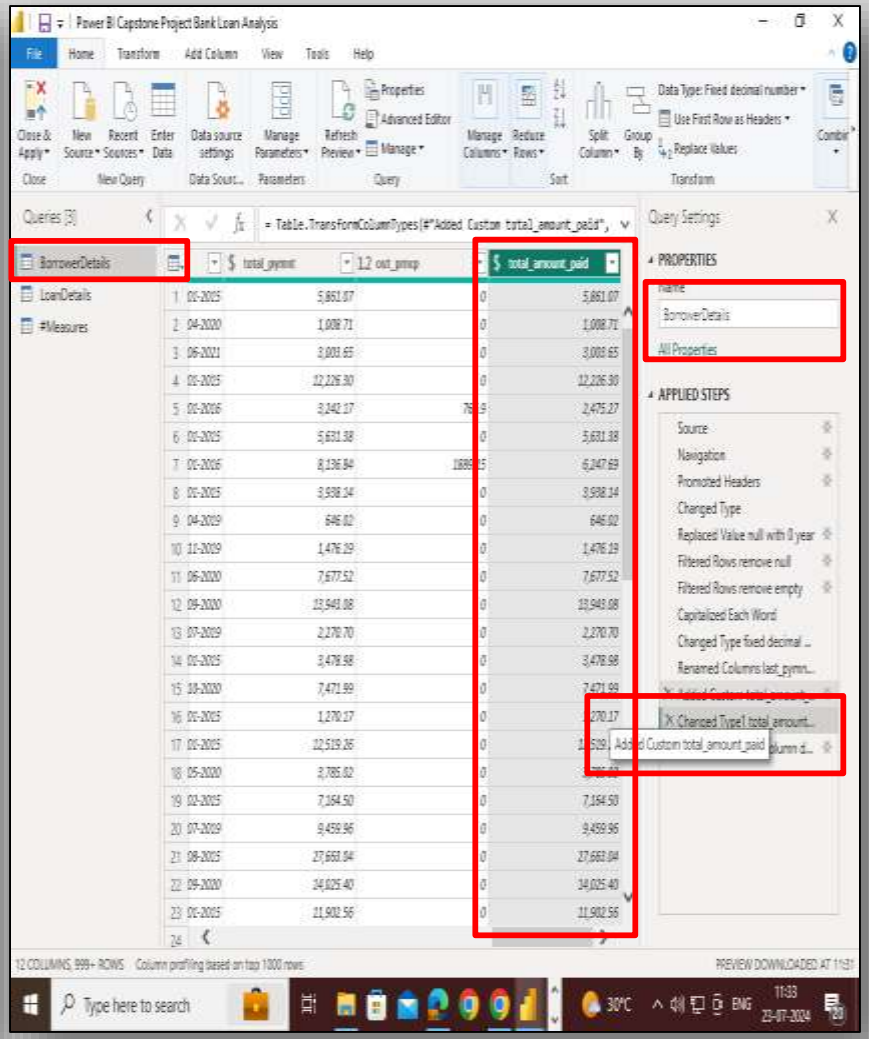
2.B II) Rename the column 'last_pymnt_d' to 'last_pymnt_date'.



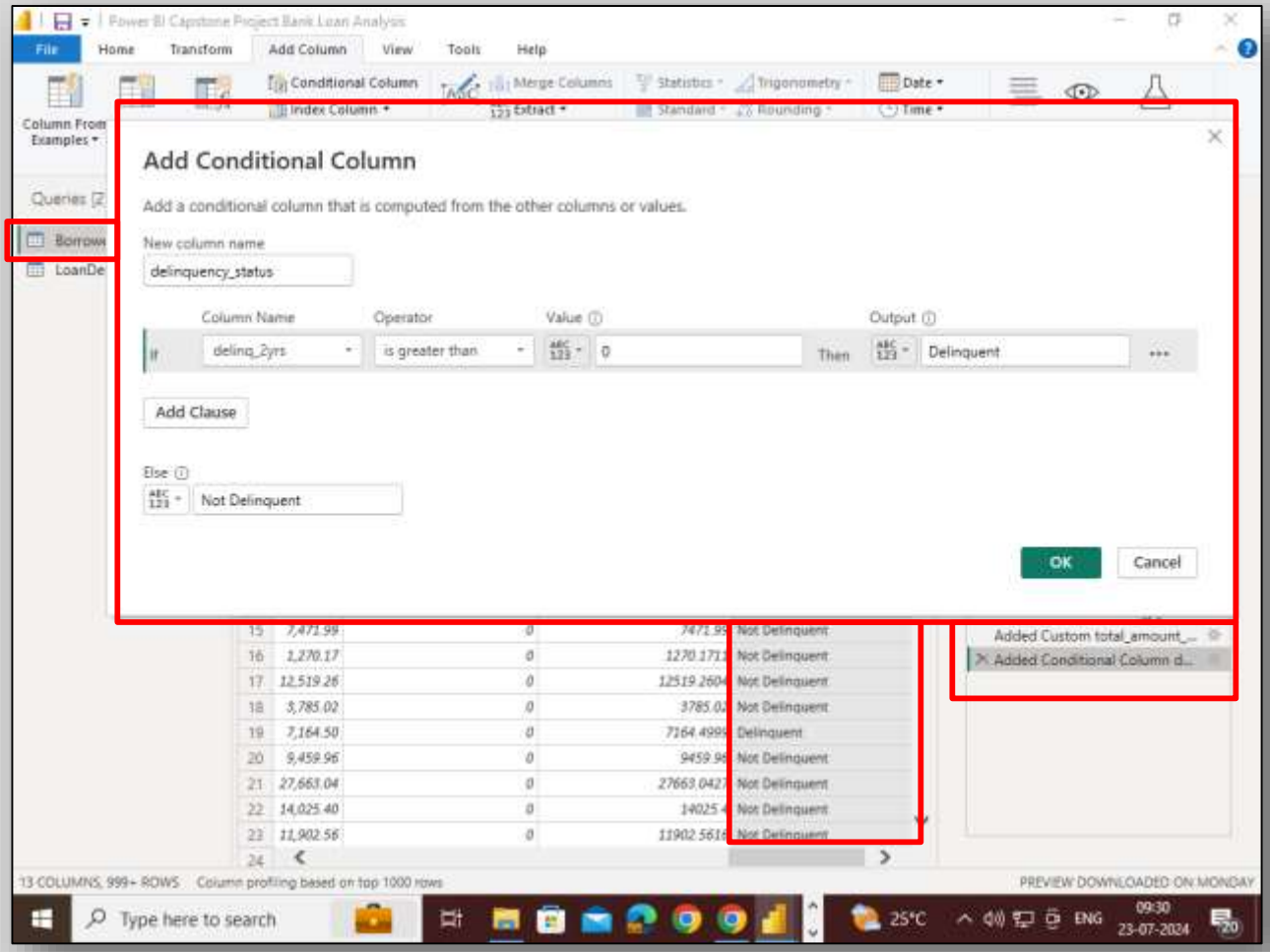
2.B III) Create a new custom column named 'total_amount_paid' to calculate the total amount paid by each borrower by subtracting 'out_prncp' from 'total_pymnt'.

total_amount_paid = [total_pymnt] - [out_prncp]

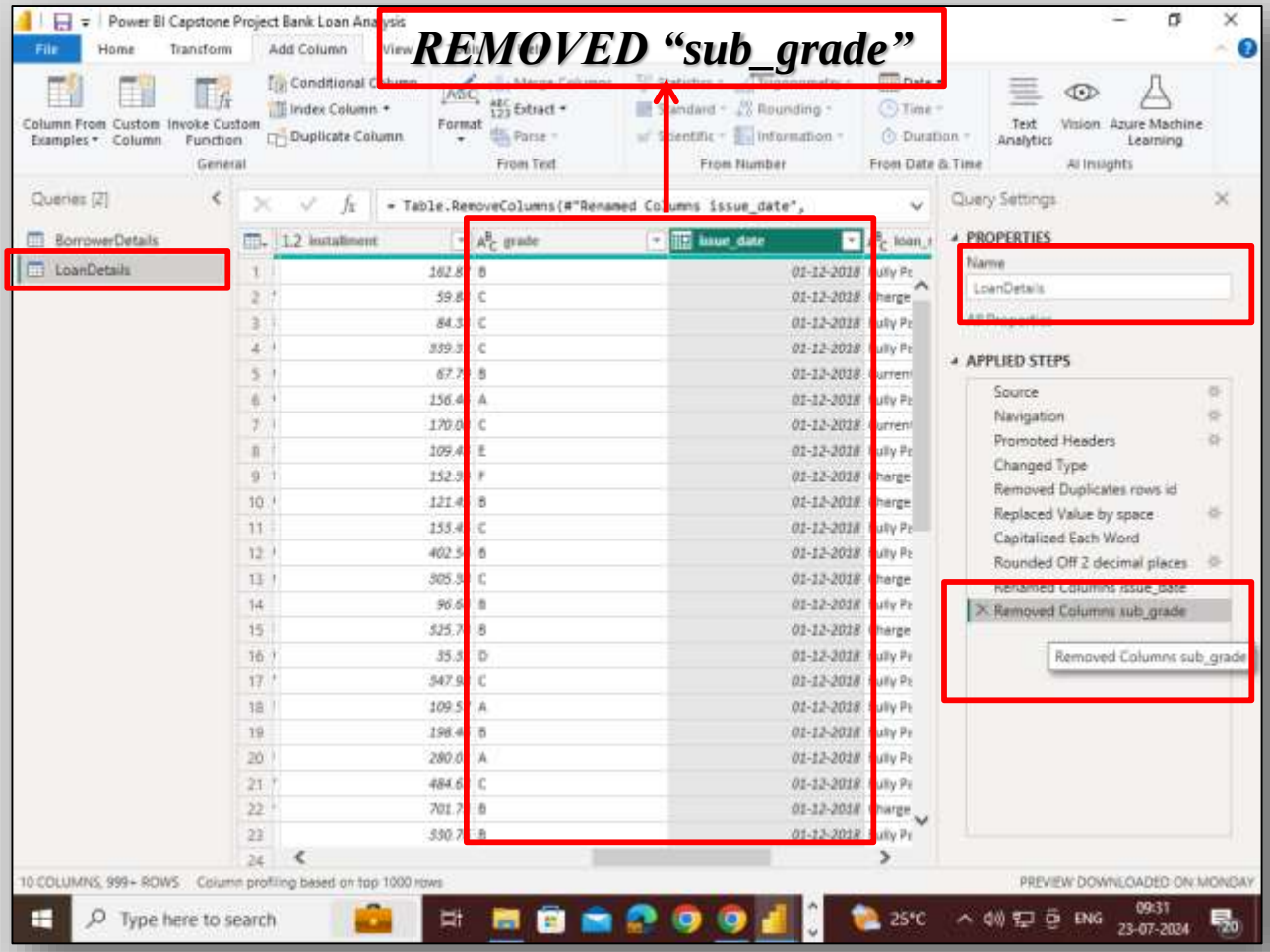




2.B III) Add a new conditional column named 'delinquency_status' to identify if the borrower has any delinquencies. If the number of delinquencies in 'delinq_2yrs' is greater than 0, the status should be "Delinquent", otherwise "Not Delinquent".

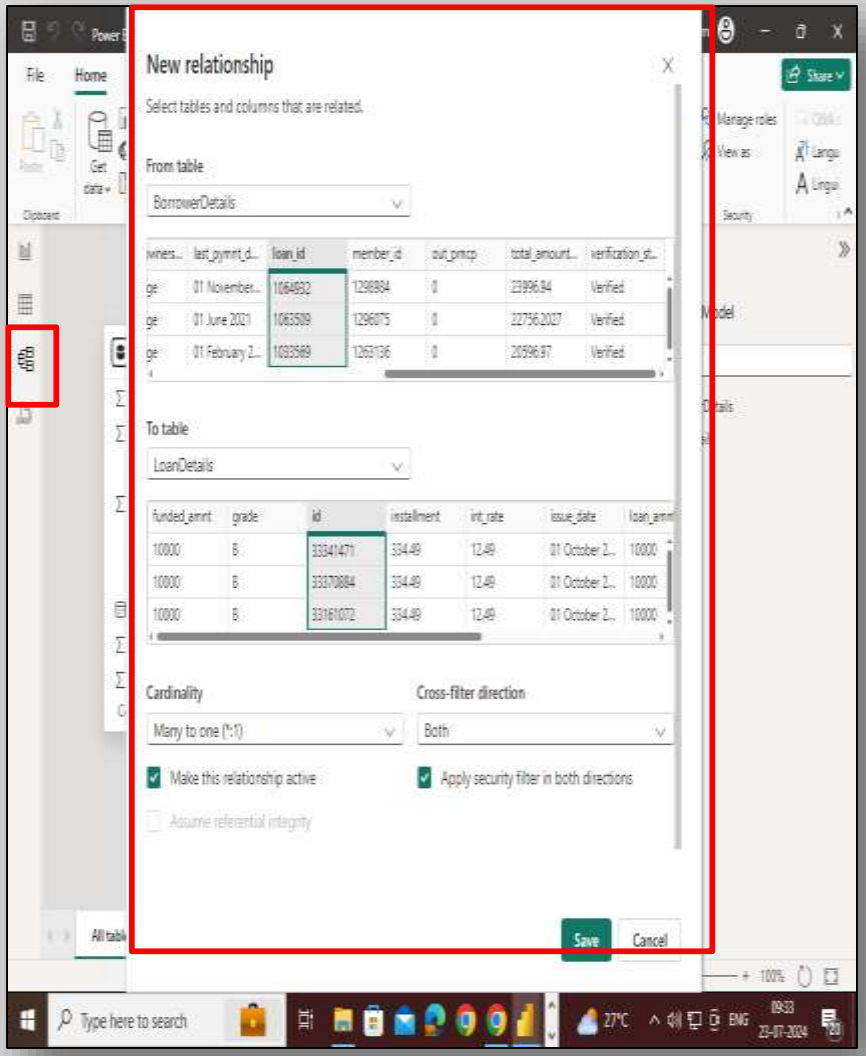


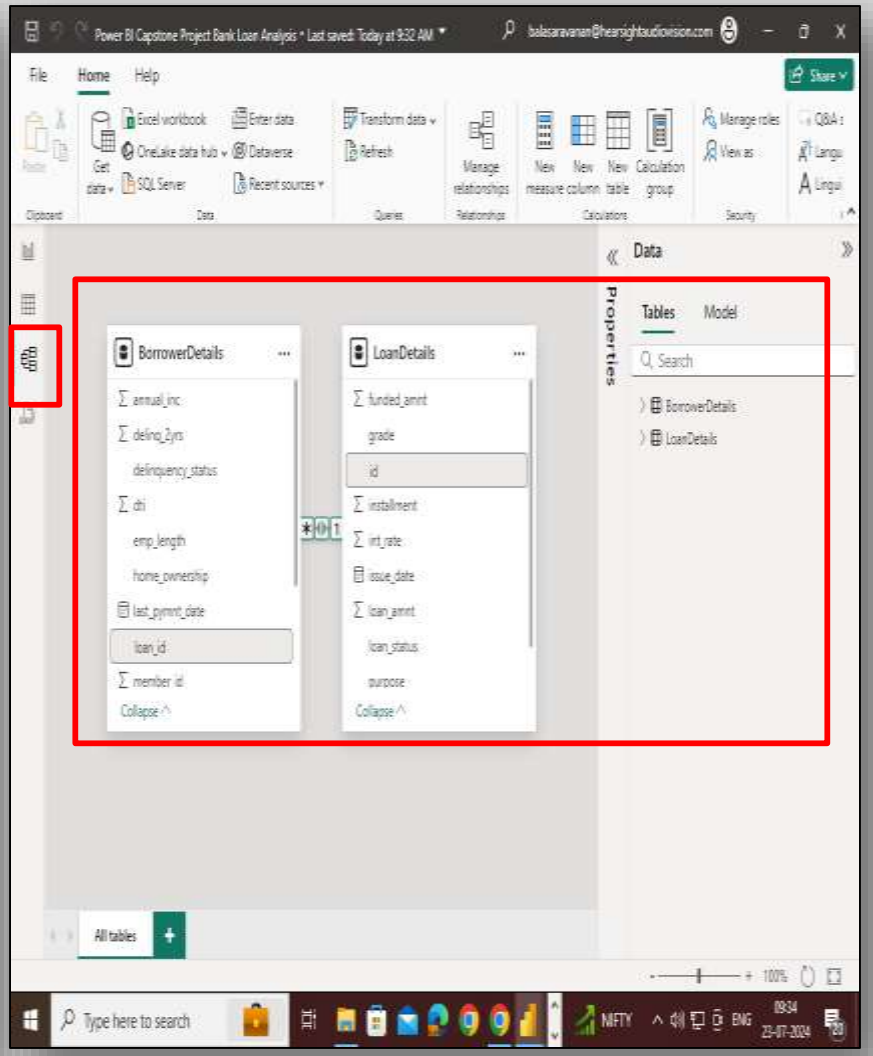
2.B IV) Remove the 'sub_grade' column as that does not significantly contribute to the analysis.



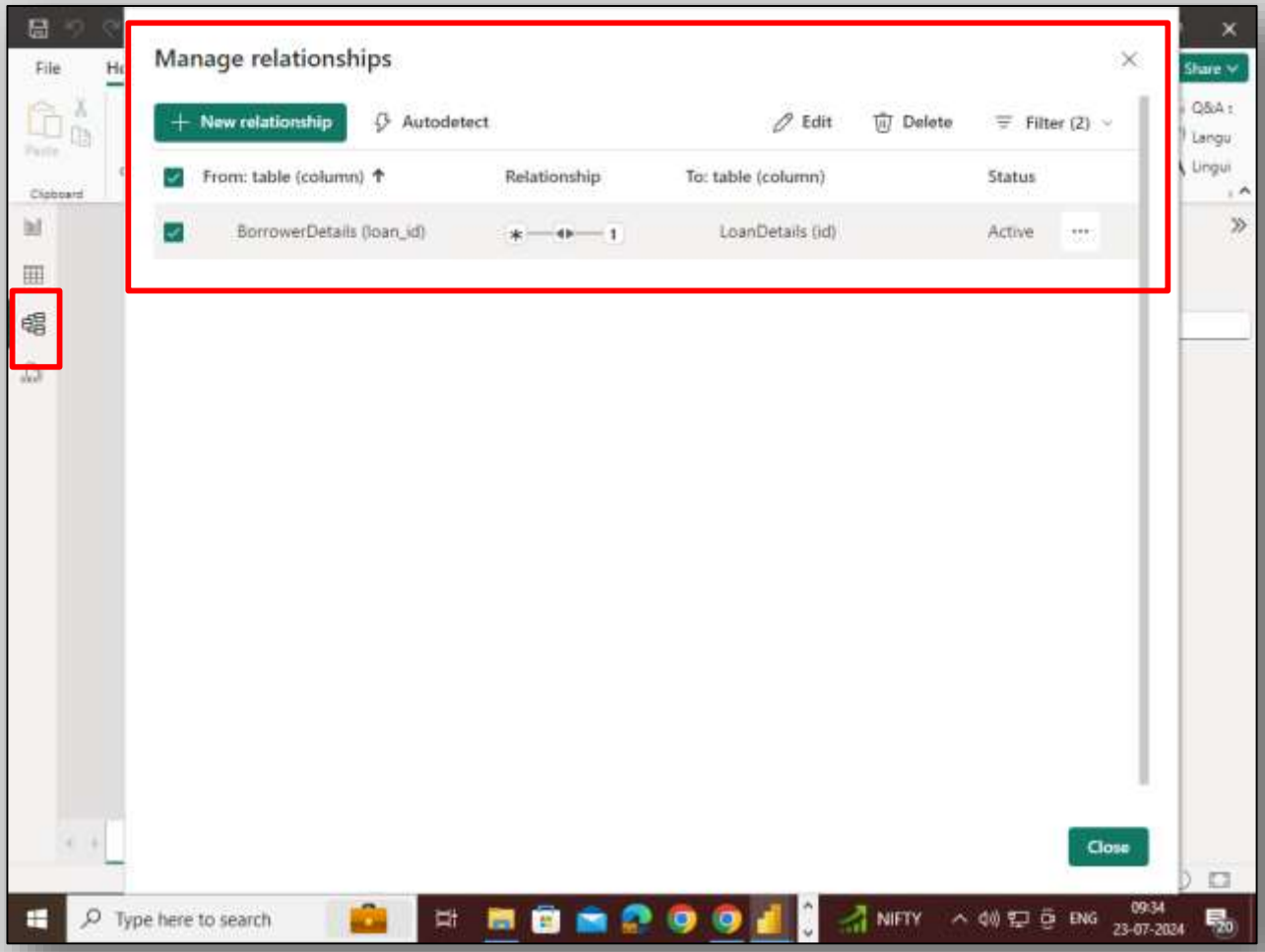
3) Data Modeling

Identify the common column between both the tables and establish relationships between the two tables. Ensure the cross-filter direction is set to "Both". This step is crucial for enabling cross-table analysis and ensuring data integrity within the dataset.





3) Data Modeling – Manage Relationships



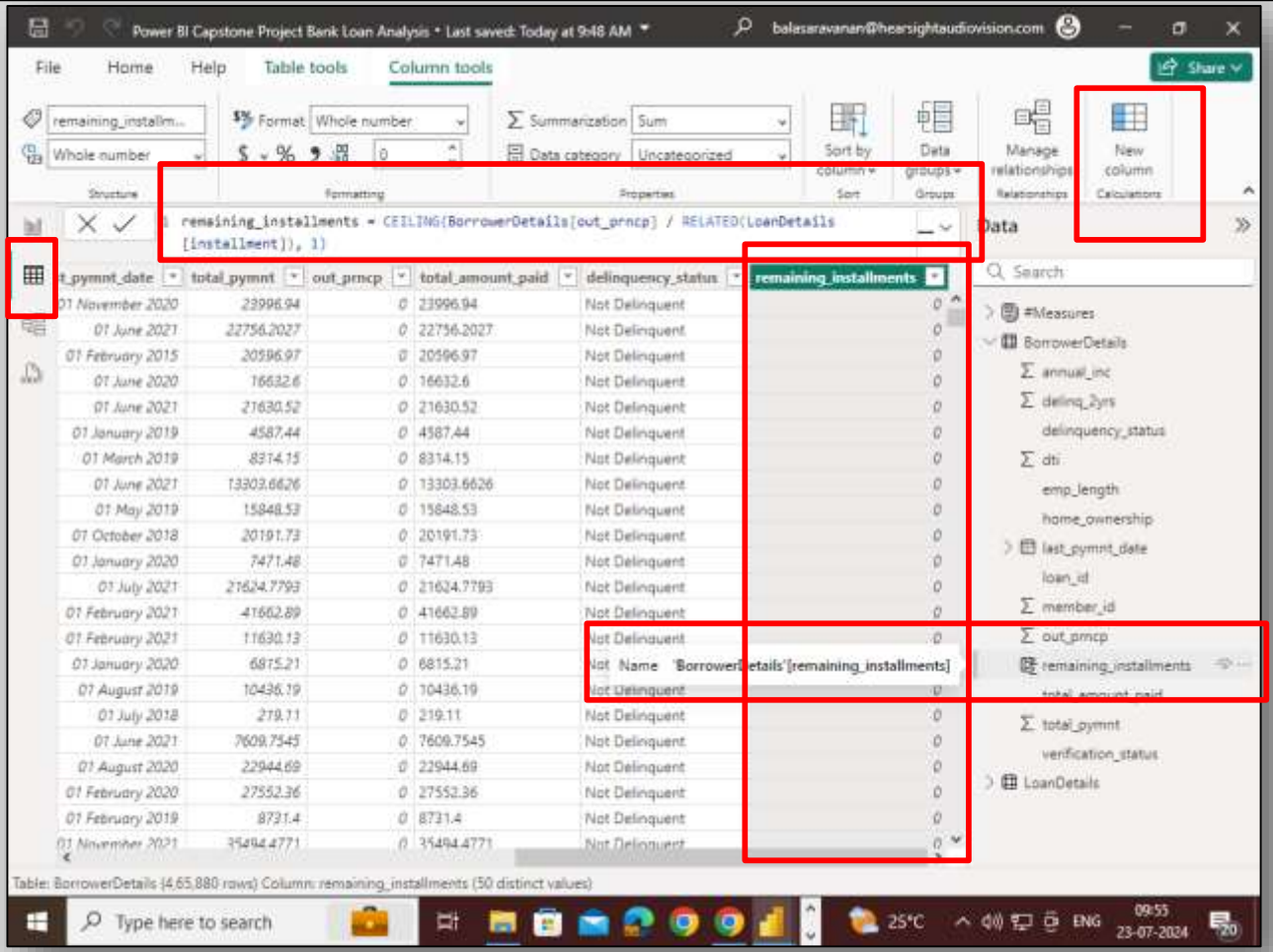
4) Creating Measures and Calculated Columns using DAX

A. *Calculated Columns.*

B. *Measures.*

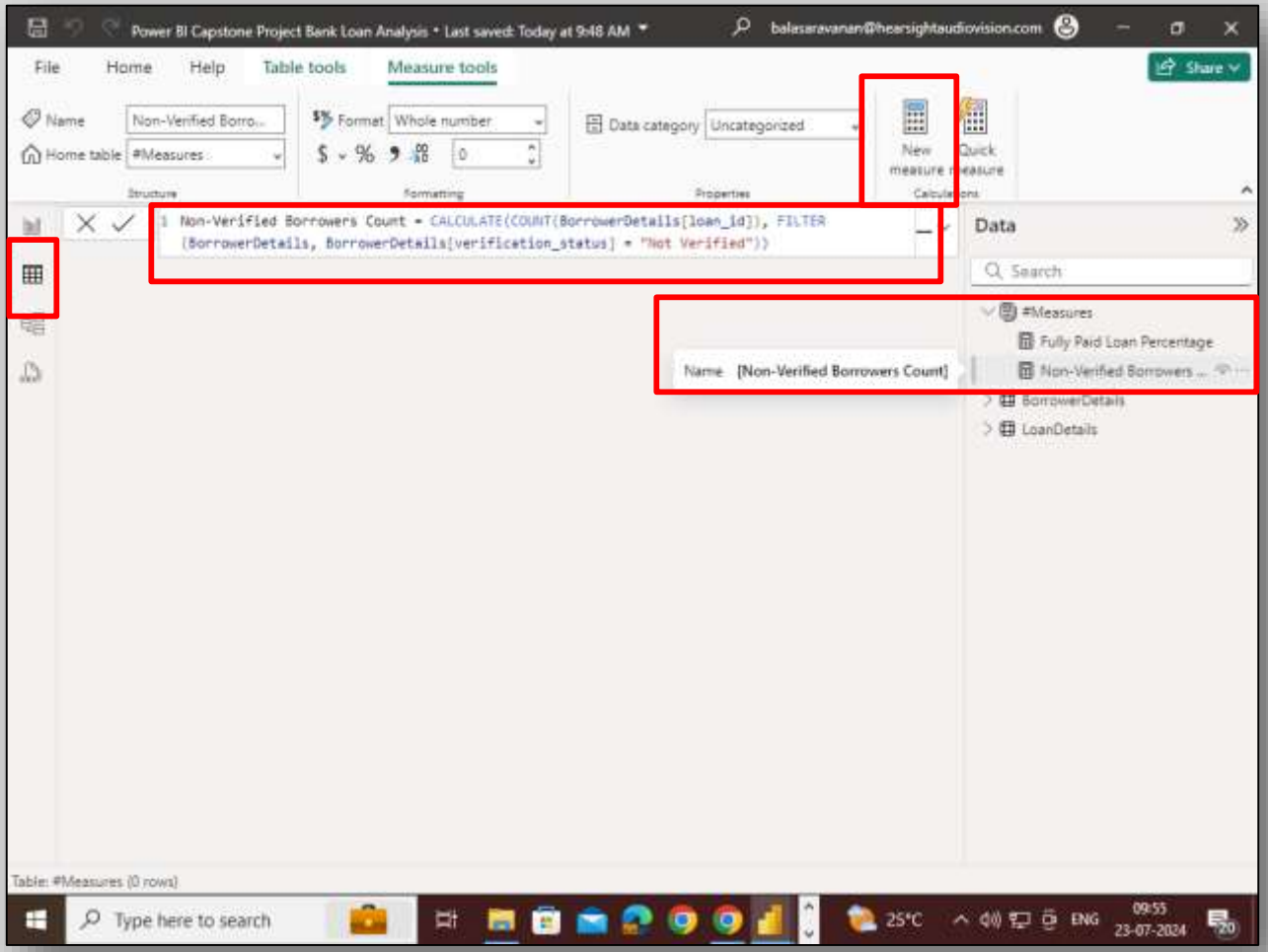
4.A) Create a new calculated column named 'remaining_installments' using DAX in the "BorrowerDetails" table to calculate the number of remaining installments by dividing the remaining principal amount ('out_prncp') by the monthly installment amount ('installment') and round up the result using the CEILING() function to account for any partial payments.

remaining_installments = CEILING(BorrowerDetails[out_prncp] / RELATED(LoanDetails[installment]), 1) - Calculated Column



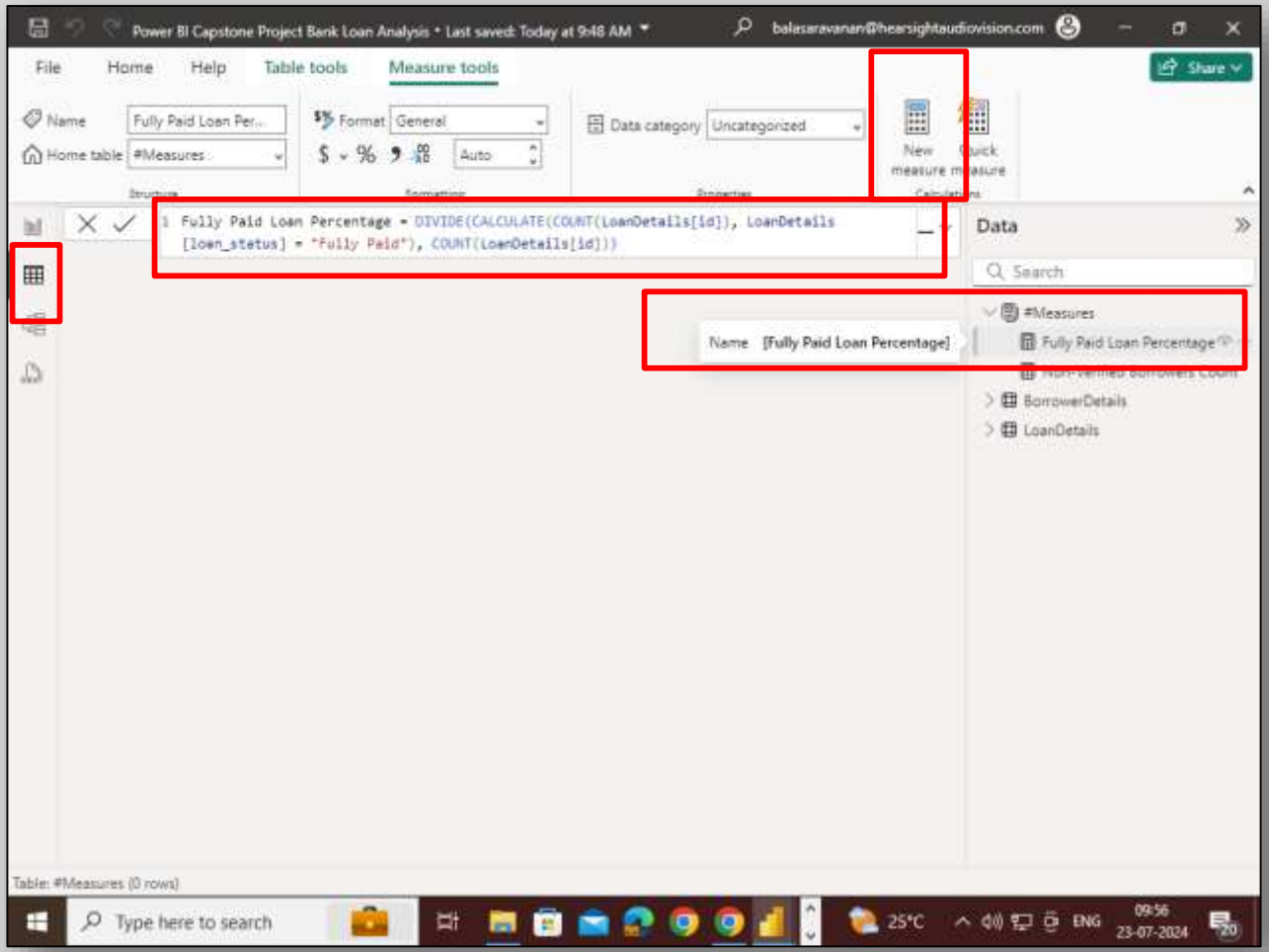
4.B I) Create a measure named 'Non-Verified Borrowers Count' using DAX to count the number of loans that have been 'Not Verified'.

Non-Verified Borrowers Count = CALCULATE(COUNT(BorrowerDetails[loan_id]), FILTER(BorrowerDetails, BorrowerDetails[verification_status] = "Not Verified"))



4.B II) Create a measure named 'Fully Paid Loan Percentage' to calculate the percentage of fully paid loans. Divide the number of loans with a "Fully Paid" loan status by the total number of loans and then format this measure as Percentage.

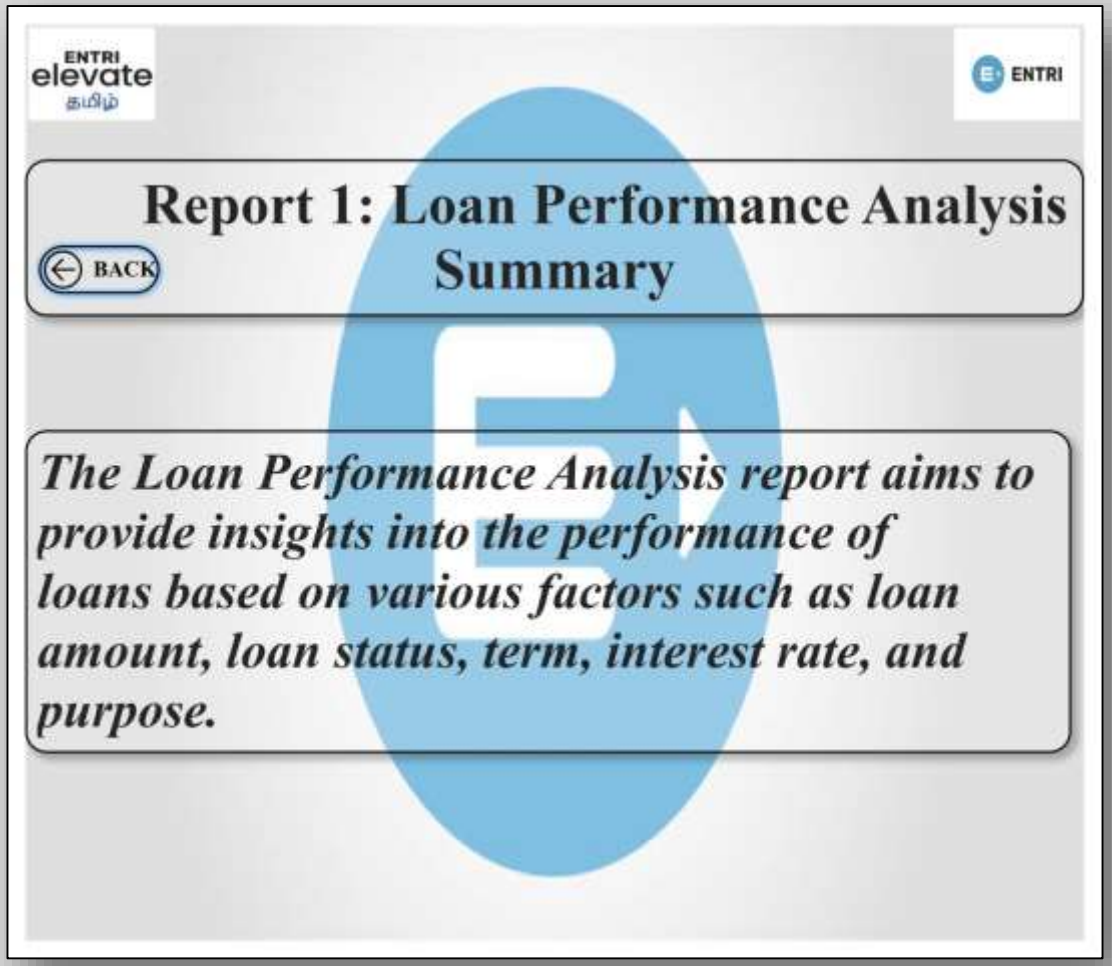
Fully Paid Loan Percentage = DIVIDE(CALCULATE(COUNT('LoanDetails'[id]), 'LoanDetails'[loan_status] = "Fully Paid"), COUNT('LoanDetails'[id]))



5) Creating Comprehensive Reports

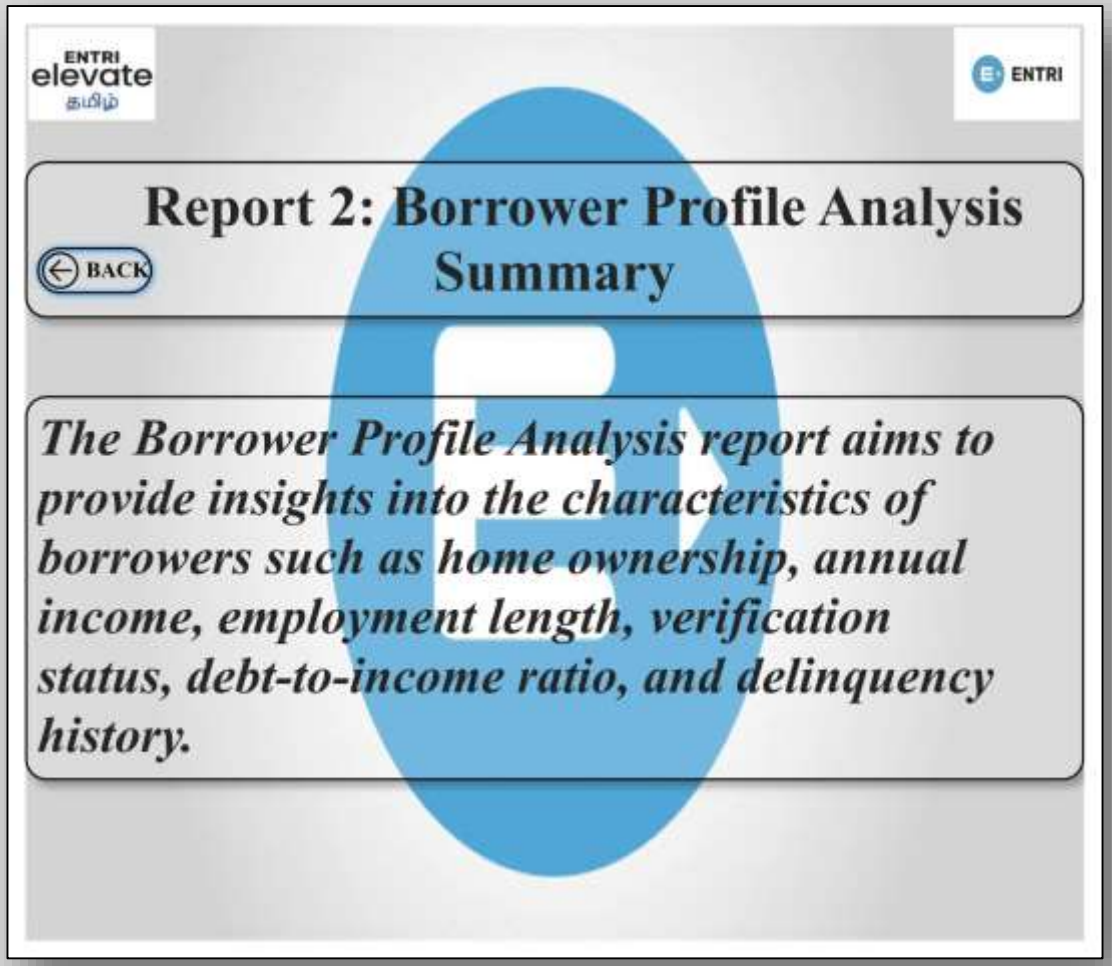
A. **Report 1: *Loan Performance Analysis.***
B. **Report 2: *Borrower Profile Analysis.***

5.A) Report 1: Loan Performance Analysis

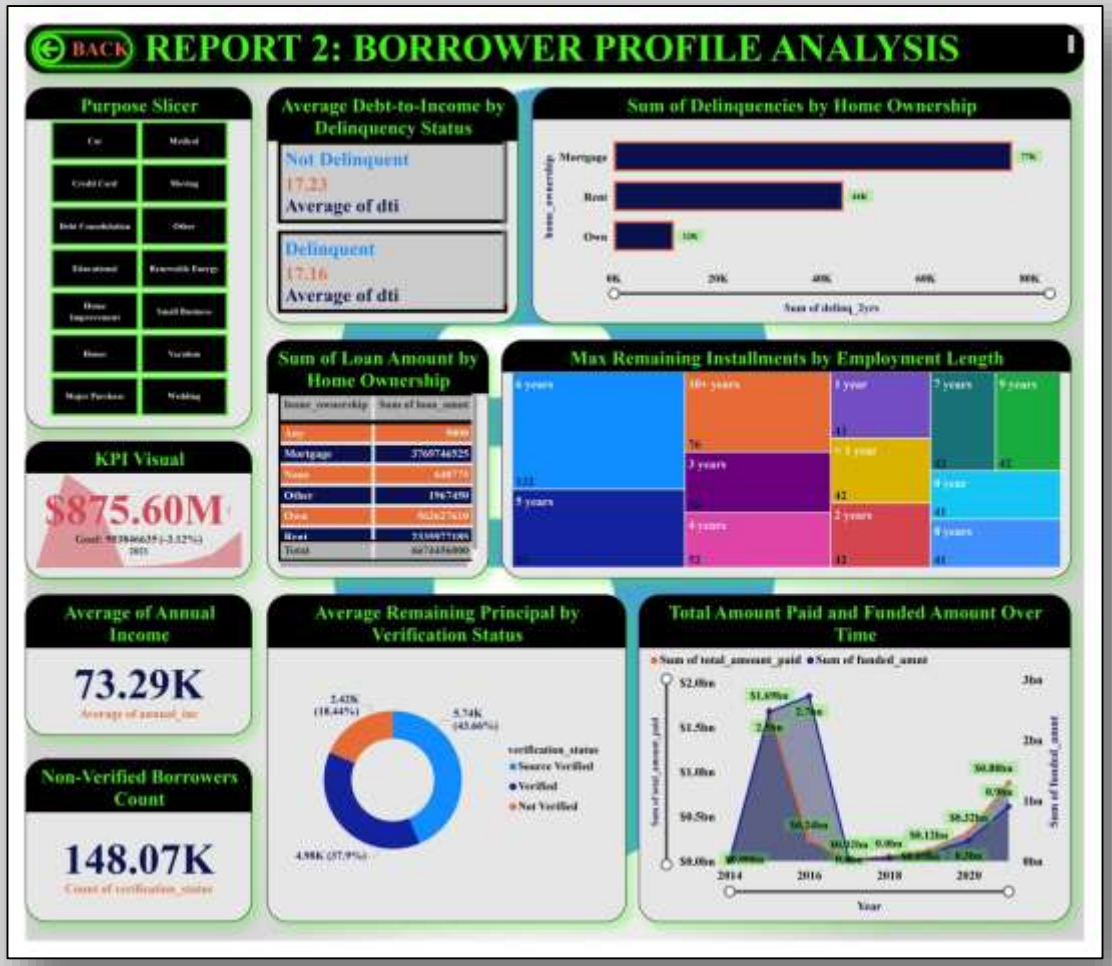


5.A) Report 1: Loan Performance Analysis

Report 2: Borrower Profile Analysis



Report 2: Borrower Profile Analysis



ENTRI elevate தமிழ்

ENTRI

THANK YOU