Excel Capstone Project

Heisenberg

financial services

Kareem Shaaban
HEISENBERG Albuquerque, New Mexico

Introduction

Heisenberg is a financial services company specializing in providing loans to individuals and small businesses. The company is looking to better understand its loan portfolio, identify key risk factors, and improve customer satisfaction. My task is to create a Loan Data Analysis Dashboard that will provide insights into the company's loan performance, customer demographics, and potential areas for improvement. This dashboard will be used by the management team to make data-driven decisions about loan approval processes, interest rates, and risk management.

So let me take you through this project phases....

Outline

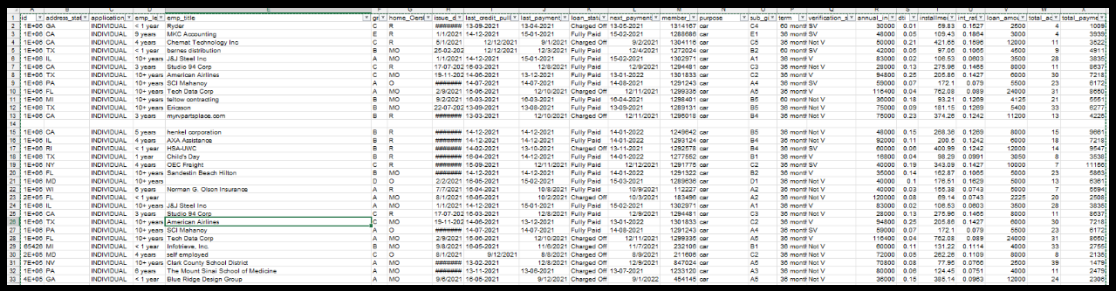
1. Exploring Phase
2. Cleaning Phase
3. Analysis Phase
4. Dashboard
5. Findings & Recommendations

Meet Our Staff

Exploring Phase

During the **exploration** phase, I carefully reviewed the task and the data dictionary to gain a thorough understanding of the entities and attributes. My focus was on identifying how to structure the fact table with relevant dimensions and determining the best approach for cleaning the data.

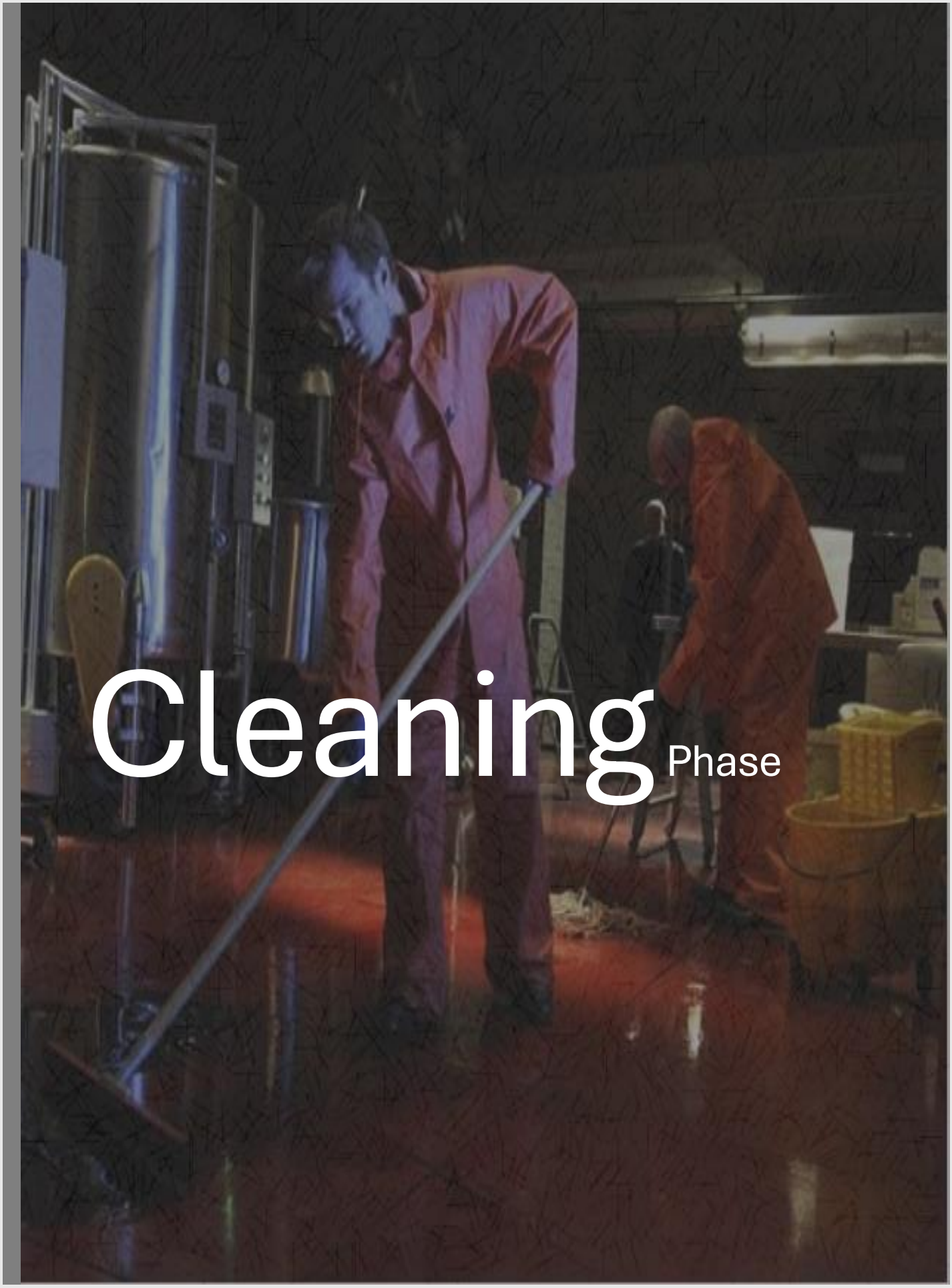
Data Sample



So, I decided to split the table into two: one for loan data and the other for borrower data. Additionally, I will create a separate table for the full state names. I'll use Power Query to clean and consolidate the data, and then establish connections between them all using Power Pivot.

Data Dictionary

	A	B	C	D	E	F	G	H	I	J
1	Column Name	Description								
2	id	Unique identifier for each loan record.								
3	address_state	Abbreviation of the state where the borrower resides.								
4	application_type	Type of loan application, e.g., INDIVIDUAL.								
5	emp_length	Length of employment of the borrower, in years.								
6	emp_title	Job title of the borrower.								
7	grade	Grade assigned to the loan, indicating the risk level (e.g., A, B, C).								
8	home_ownership	Home ownership status of the borrower (R-RENT,MO- MORTGAGE, O-OWN).								
9	issue_date	Date the loan was issued.								
10	last_credit_pull_date	Date of the last credit pull on the borrower.								



Cleaning Phase

cleaning phase is crucial in any analytics project, as clean data is essential for producing accurate results. As a result, it is the most time-consuming phase of this project.

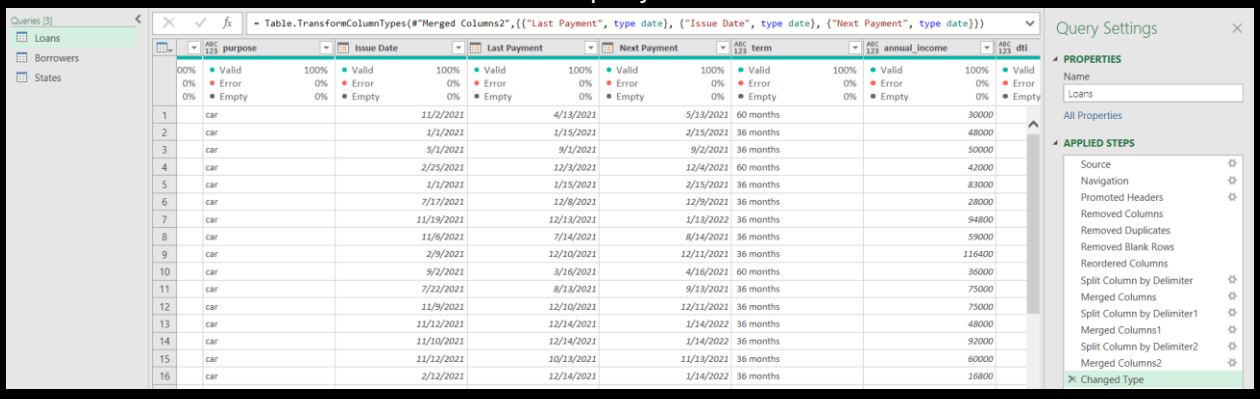
So, I opened a new Excel file and began by importing the data from the source. After selecting the dataset, I clicked on 'Transform Data' to proceed.

I duplicated the query to create two new tables: one for loans and one for borrowers. I also added a table for states and their codes. The loans table will serve as the fact table, while the others will act as dimensional tables.

Next, the cleaning process began. I started with the fact table by removing unnecessary columns. After that, I removed duplicates and blank rows, followed by reordering the columns.

I then noticed issues with all the date columns, so I split each one by delimiter into day, month, and year columns. Afterward, I merged them back correctly in the "MM/DD/YYYY" format.

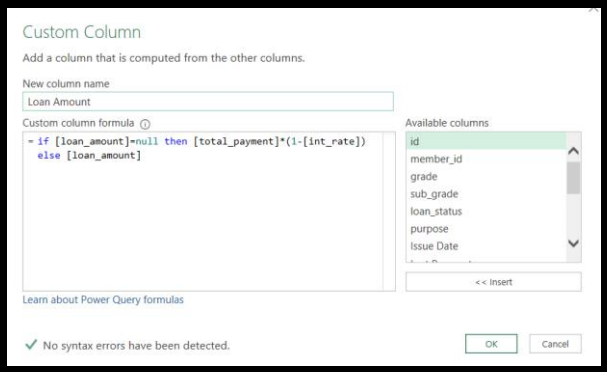
Power query view



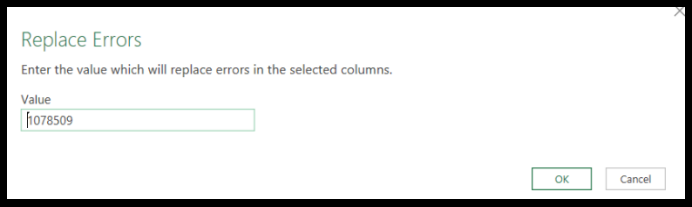
After completing the previous steps, I found missing values in the loan amount, annual income, and total payment columns. To address this, I plan to calculate the missing values using related columns such as installment, interest rate, and debt-to-income ratio (DTI)

- To get the loan amount = Total payment*(1- interest rate)
- To get total payment:
  - If the loan status “fully paid” then it equals
    - Installments * Duration terms
  - If the loan status “charge off” or “currently” it equals
    - Installments * ((last payment date-Issued date)/30)

Here’s how I replaced the missing values.



After that, I discovered an error in the loan ID column. To resolve it, I generated a new ID using ChatGPT and replaced the erroneous values with it.

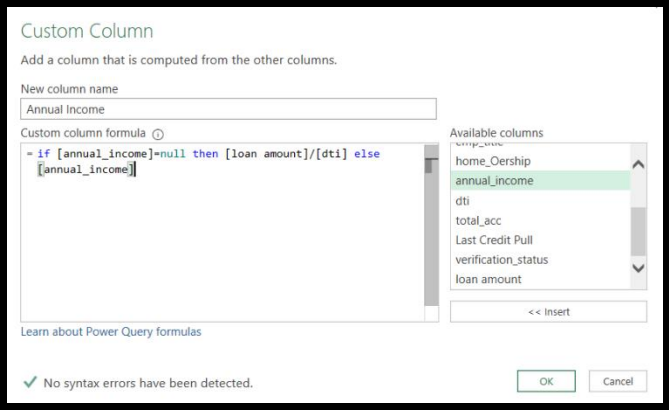


✓ Finally, the cleaning and adjustments for the loans table were completed.

Next, the cleaning process continued with the Borrowers table. I began by removing unnecessary columns, followed by eliminating duplicates and blank rows to ensure data consistency.

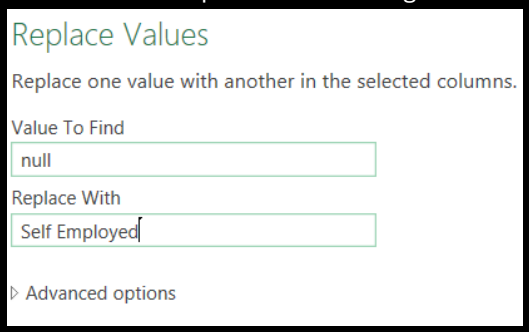
I found missing values in the annual income column. To address this, I will calculate them using a specific formula based on the available data.

- Annual income = loan amount / (TDI) ratio

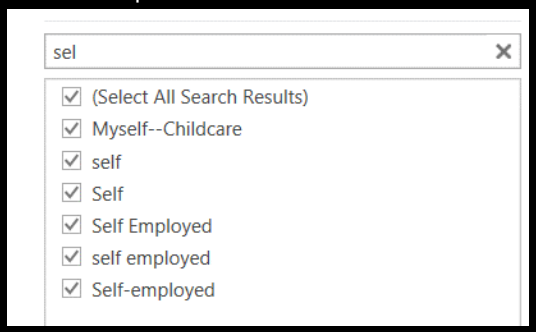


Afterward, I addressed the null values and corrected certain entries in the employer column to ensure the data was clean and accurate.

Here's how I replaced the missing values.



I replaced those values too.



A sample of the data I was working with.

Apple, Inc.	Apple Inc
Zurich Of North America	Zurich North America
24 Hour Fitness, Inc.	24 Hour Fitness
24 Hr. Fitness	24 Hour Fitness
Abercrombie And Fitch	Abercrombie & Fitch
3M Co.	3M
ABC Supply Co. Inc	ABC Supply
ABM	ABM Industries

Power Query Advanced Editor View

```
#"Self Employed" = Table.ReplaceValue(#"Renamed Columns",each [Employer],
      each if [Employer] = null then "Self Employed"
      else if [Employer] = "self" then "Self Employed"
      else if [Employer] = "Self" then "Self Employed"
      else if [Employer] = "Self-employed" then "Self Employed"
      else if [Employer] = "self employed" then "Self Employed"
      else if [Employer] = "Myself--Childcare" then "Self Employed"
      else [Employer]

,Replacer.ReplaceValue,{"Employer"}),
```

```
#"Handling Employer Column" = Table.ReplaceValue(#"Self Employed",each [Employer],
        each if [Employer] = "Apple, Inc." then "Apple Inc"
        else if [Employer] = "24 Hour Fitness, Inc." then "24 Hour Fitness"
        else if [Employer] = "24 hour fitness" then "24 Hour Fitness"
        else if [Employer] = "24 Hr. Fitness" then "24 Hour Fitness"
        else if [Employer] = "Zurich of North America" then "Zurich North America"
        else if [Employer] = "Zurich Financial Services - Global Corporate" then "Zurich North America"
        else if [Employer] = "Zurich NA" then "Zurich North America"
        else if [Employer] = "Zurich Insurance" then "Zurich North America"
        else if [Employer] = "Abercrombie and Fitch" then "Abercrombie & Fitch"
        else if [Employer] = "3m" then "3M Company"
        else if [Employer] = "3M" then "3M Company"
        else if [Employer] = "ABM" then "ABM Industries"
        else if [Employer] = "Abc Supply Co. Inc" then "ABC Supply"
        else if [Employer] = "abc supply" then "ABC Supply"
        else if [Employer] = "Abc Supply" then "ABC Supply"
        else if [Employer] = "" then ""

        else [Employer]

    ,Replacer.ReplaceValue,{"Employer"}),

    // I tried as much as I could
```

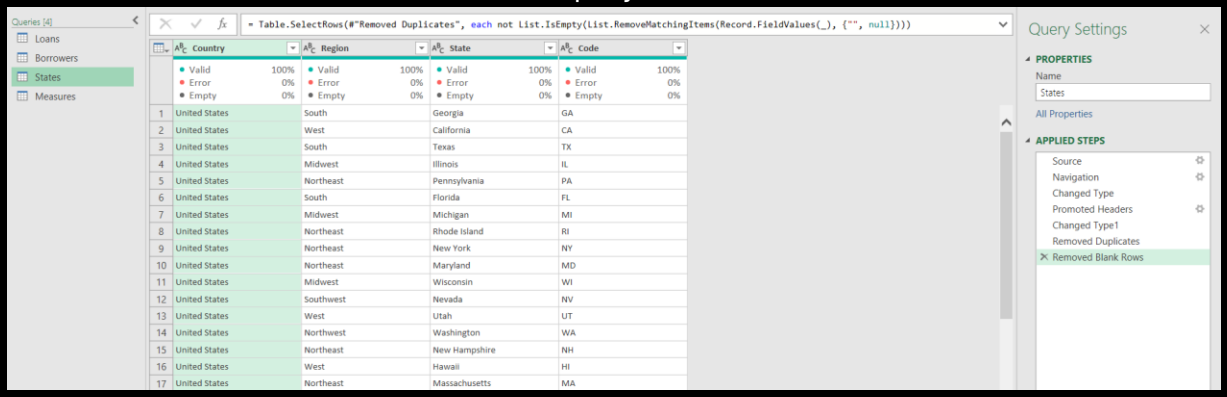
Power Query Advanced Editor View

```
#"Employment Length" = Table.ReplaceValue(#"Renamed Columns1",each [Employment Length],

    each if [Employment Length] = "< 1 year" then "-1 'Less Than One Year'"
    else if [Employment Length] = "10+ years" then "More Than 10 Years"

    else [Employment Length]

    ,Replacer.ReplaceValue,{"Employment Length"}),

#"Home Ownership Status" = Table.ReplaceValue(#"Employment Length",each [Home Ownership Status],

    each if [Home Ownership Status] = "O" then "OWN"
    else if [Home Ownership Status] = "R" then "RENT"
    else if [Home Ownership Status] = "MO" then "MORTGAGE"

    else [Home Ownership Status]

    ,Replacer.ReplaceValue,{"Home Ownership Status"}),

#"Verification Status" = Table.ReplaceValue(#"Home Ownership Status",each [Verification Status],

    each if [Verification Status] = "SV" then "Source Verified"
    else if [Verification Status] = "Not V" then "Not Verified"
    else if [Verification Status] = "V" then "Verified"

    else [Verification Status]

    ,Replacer.ReplaceValue,{"Verification Status"}),
```

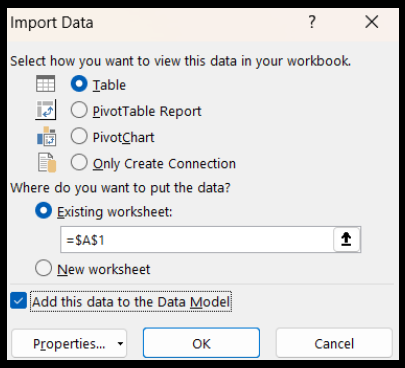
- ✓ Finally, the cleaning and adjustments for the Borrowers table were completed.

Finally, the cleaning process continued with the last table “States”. I began by removing duplicates and blank rows to ensure data consistency and promoting first row as a header.

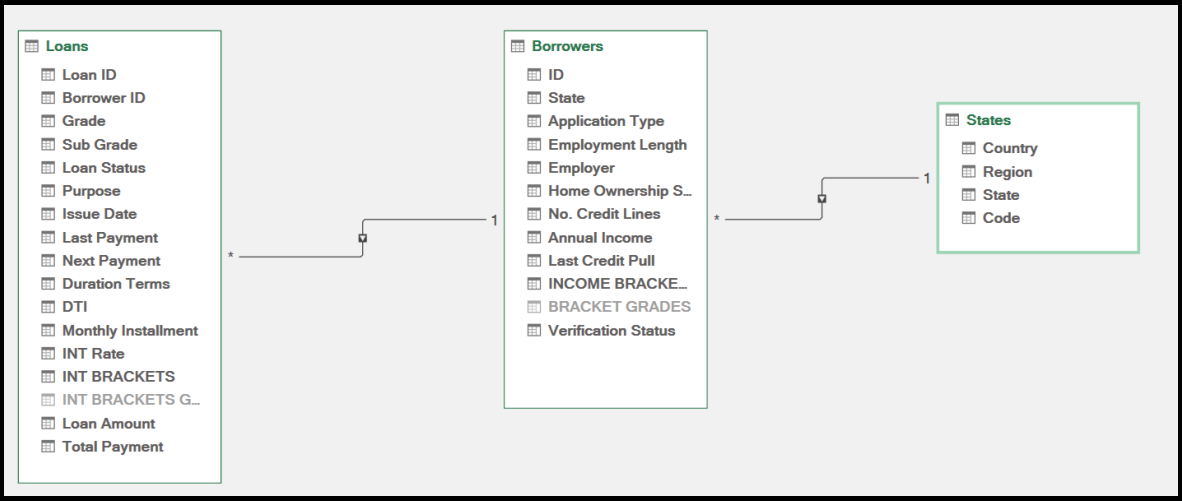
Power query view



After completing all the cleaning steps, I loaded the three queries into the workbook and added them to the data model, enabling the creation of connections between the tables.

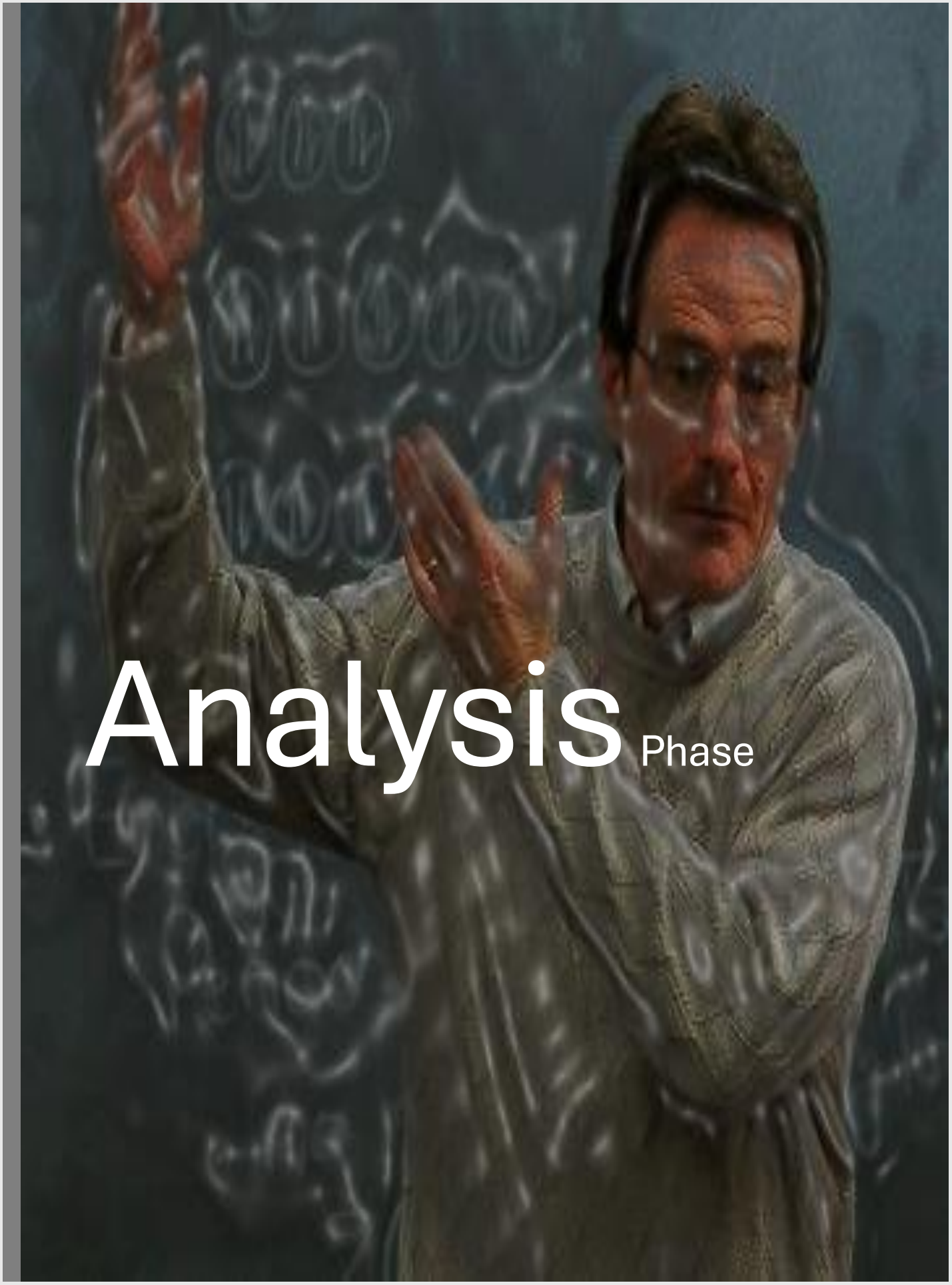


Power Pivot Diagram View



Finally, the cleaning process is complete, the longest and most critical phase of this project, as it is in any project. Here are the major challenges I faced during this phase:

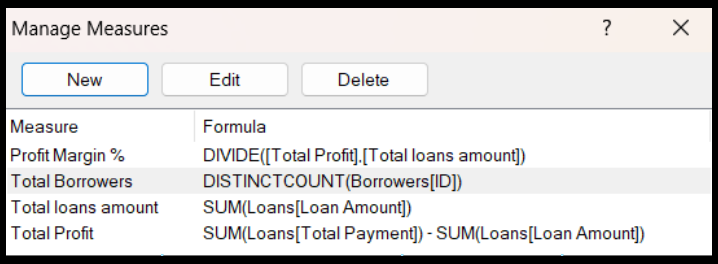
- Addressing date issues, particularly when some issue dates were found to be after the last payment date.
- Managing missing values across key columns.
- Handling inconsistencies in the employer column.


Analysis Phase

Throughout the last two phases, I worked to gain a deep understanding of the dataset, and I now have several questions that need answers. Let's begin with those questions.

- What is the total loan amount?
- How many loans have been issued?
- What is our total profit and profit margin?
- How many customers do we have?

To answer those questions, I created some measures like:



Total Borrowers	Total loans amount	Total Profit	Profit Margin %
38,574	$435.7M	$37.3M	8.56%

Key questions related to **borrowers** that need answers:

1. What is the status of information verification for borrowers?
2. How many borrowers are there in each state?
3. What is the distribution of our customers' annual income?
4. What is their homeownership status?
5. Who are the top 3 employers of our customers?
6. What is the employment length for our customers?

1

Row Labels	Total Borrowers
Not Verified	16,462
Verified	12,335
Source Verified	9,777

2

Row Labels	Total Borrowers
California	6,893
NewYork	3,700
Florida	2,773
Texas	2,664
New Jersey	1,822
Illinois	1,486
Pennsylvania	1,482
Virginia	1,375
Georgia	1,355
Massachusetts	1,310
Ohio	1,188
Maryland	1,027
Washington	1,019
Arizona	833
Colorado	770
North Carolina	759
Connecticut	730
Michigan	685
Missouri	660
Minnesota	592
Nevada	482
South Carolina	464
Wisconsin	446
Oregon	436
Alabama	432
Louisiana	426
Kentucky	320
Oklahoma	293
Kansas	260
Utah	252
Arkansas	236
Rhode Island	196
New Mexico	183
Hawaii	170
West Virginia	167
New Hampshire	161
Delaware	110
Wyoming	79
Montana	79
Alaska	78
South Dakota	63
Vermont	54
Mississippi	19
Tennessee	17
Indiana	9
Idaho	6
Nebraska	5
Iowa	5
Maine	3

3

Row Labels	Total Borrowers
Less than 12K$	102
12K$ TO 24K$	1,460
24K$ TO 36K$	4,803
36K$ TO 48K$	6,482
48K$ TO 60K$	6,263
60K$ TO 72K$	5,847
72K$ TO 84K$	4,008
84K$ TO 96K$	2,842
96K$ TO 108K$	1,898
108K$ TO 120K$	1,127
More than 120K$	3,742

4

Row Labels	Total Borrowers
Rent	18,437
Mortgage	17,198
Own	2,838
Other	101

5

Row Labels	Total Borrowers
Self Employed	1,569
US Army	211
Bank of America	138

6

Row Labels	Total Borrowers
-1 'Less Than One Year'	4,575
1 Year	3,229
2 Years	4,382
3 Years	4,088
4 Years	3,427
5 Years	3,273
6 Years	2,227
7 Years	1,772
8 Years	1,476
9 Years	1,255
More Than 10 Years	8,870

Key questions related to **Loans** that need answers:

1. How do loan amounts and profitability vary by loan grades?
2. What is the impact of loan duration terms on the total paid amount and profitability?
3. How does the interest rate affect the total paid amount and profitability?
4. How do loan amounts and profitability vary by loan status?
5. What are the most common purposes for these loans?

1

Row Labels	Total loans amount	Total Profit
A	**$84.3M**	**$3.8M**
B	**$130.7M**	**$10.1M**
C	**$87.5M**	**$8.5M**
D	**$63.9M**	**$6.9M**
E	**$44.2M**	**$5.0M**
F	**$18.9M**	**$2.1M**
G	**$6.3M**	**$918.0K**

2

Row Labels	Total paid amount	Total Profit
36	$294.7M	$21.7M
60	$178.4M	$15.6M

3

Row Labels	Total paid amount	Total Profit
0 TO 10%	$109.3M	$5.0M
10% TO 15%	$226.6M	$18.1M
15% TO 20%	$117.5M	$12.3M
MORE THAN 20%	$19.8M	$1.9M

4

Row Labels	Total paid amount	Total Profit
Charged Off	$37.3M	-$28.3K
Current	$24.2M	$5.3M
Fully Paid	$411.6M	$60.2M

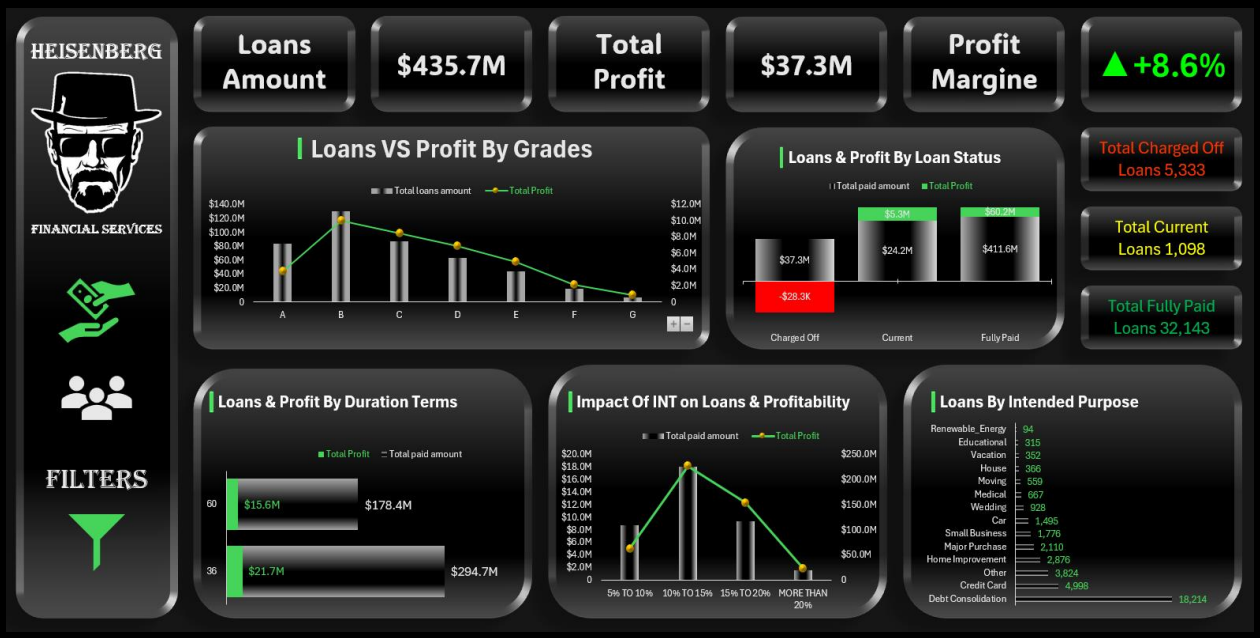
5

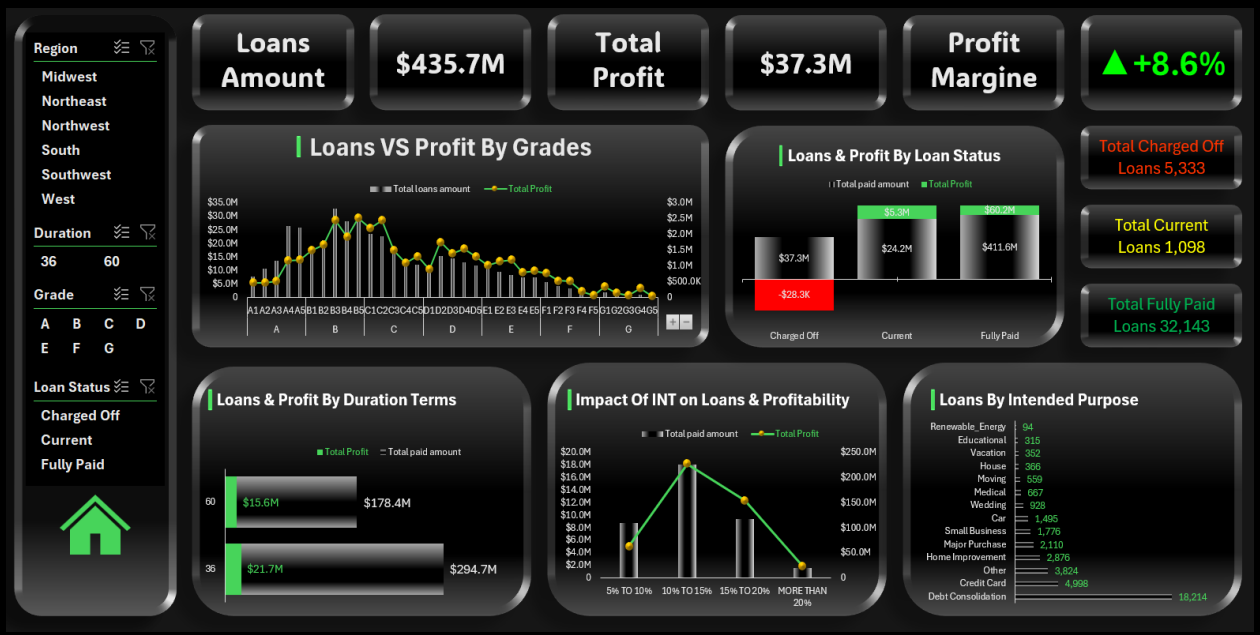
Row Labels	Count of Purpose
Debt Consolidation	18,214
Credit Card	4,998
Other	3,824
Home Improvement	2,876
Major Purchase	2,110
Small Business	1,776
Car	1,495
Wedding	928
Medical	667
Moving	559
House	366
Vacation	352
Educational	315
Renewable_Energy	94

After answering those question, I need just to visualize it by using a suitable chart to make it easier to understand after that I will go to the second phase to build my dashboard “the last phase”

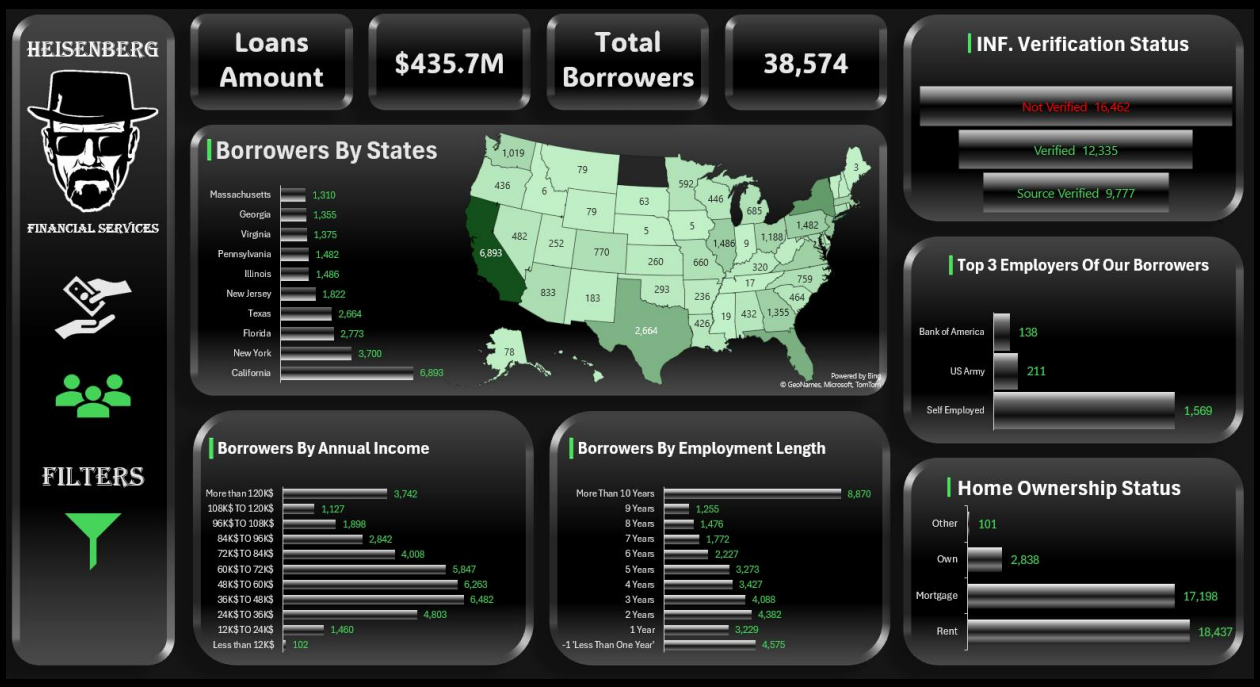

The Last Phase

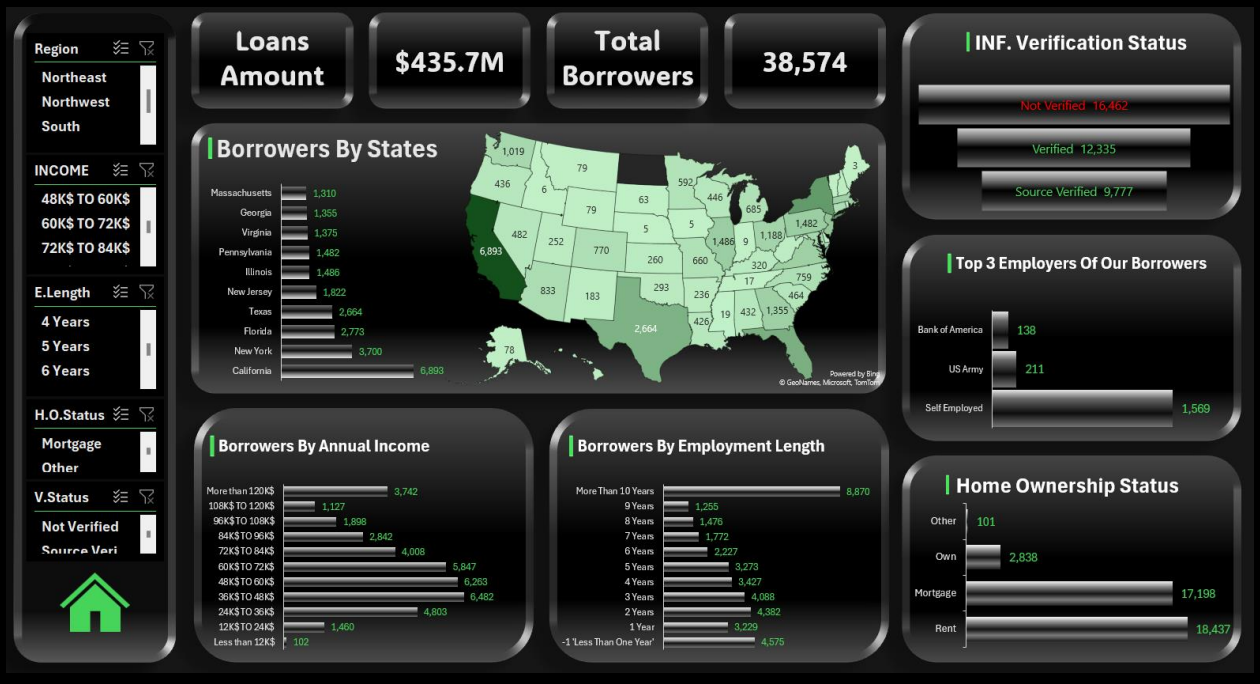
Loans View


HEISENBERG
FINANCIAL SERVICES
FILTERS
Loans Amount
$435.7M
Total Profit
$37.3M
Profit Margine
▲+8.6%
Loans VS Profit By Grades
Total loans amount
Total Profit
Loans & Profit By Loan Status
Total paid amount
Total Profit
$37.3M
-$28.3K
$5.3M
$24.2M
$60.2M
$411.6M
Charged Off
Current
Fully Paid
Total Charged Off Loans 5,333
Total Current Loans 1,098
Total Fully Paid Loans 32,143
Loans & Profit By Duration Terms
Total Profit
Total paid amount
$15.6M
$178.4M
$21.7M
$294.7M
Impact Of INT on Loans & Profitability
5% TO 10%
10% TO 15%
15% TO 20%
MORE THAN 20%
Loans By Intended Purpose
Renewable_Energy 94
Educational 315
Vacation 352
House 366
Moving 559
Medical 667
Wedding 928
Car 1,495
Small Business 1,776
Major Purchase 2,110
Home Improvement 2,876
Other 3,824
Credit Card 4,998
Debt Consolidation 18,214


Region
Midwest
Northeast
Northwest
South
Southwest
West
Duration
36
60
Grade
A B C D
E F G
Loan Status
Charged Off
Current
Fully Paid
Loans Amount
$435.7M
Total Profit
$37.3M
Profit Margine
▲+8.6%
Loans VS Profit By Grades
Total loans amount
Total Profit
Loans & Profit By Loan Status
Total paid amount
Total Profit
$37.3M
-$28.3K
$5.3M
$24.2M
$60.2M
$411.6M
Charged Off
Current
Fully Paid
Total Charged Off Loans 5,333
Total Current Loans 1,098
Total Fully Paid Loans 32,143
Loans & Profit By Duration Terms
Total Profit
Total paid amount
$15.6M
$178.4M
$21.7M
$294.7M
Impact Of INT on Loans & Profitability
5% TO 10%
10% TO 15%
15% TO 20%
MORE THAN 20%
Loans By Intended Purpose
Renewable_Energy 94
Educational 315
Vacation 352
House 366
Moving 559
Medical 667
Wedding 928
Car 1,495
Small Business 1,776
Major Purchase 2,110
Home Improvement 2,876
Other 3,824
Credit Card 4,998
Debt Consolidation 18,214

Borrowers View





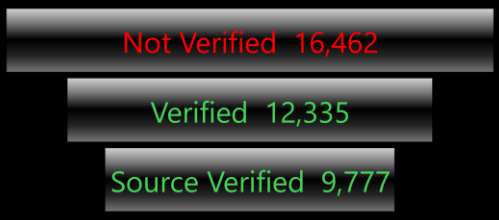
In this phase, I added columns such as "Income Brackets" and "Interest Rate Brackets" to summarize the data and make it more readable. I also used macros to hide and display the filter tab in each view, making it more dynamic. With that, the final phase is complete.


It wasn’t
The Last Phase

Findings & Recommendations

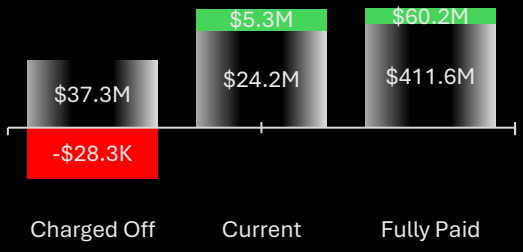
In this phase, I'll present my observations on the data through dashboards and charts, providing a clear visual representation of the insights. Afterward, I'll share my recommendations for each situation based on the data analysis and findings.

INF. Verification Status



Observation:

We have a significant amount of unverified information, which poses a high risk. This missing or unverified data is crucial for accurately evaluating loan grades. As a result, this has likely contributed to the high number of charged-off loans.



Recommendation:

Implement stricter verification processes to ensure all borrower information is accurate and complete before approving loans. This could involve enhancing automated verification tools or requiring additional documentation. Strengthening the verification process will reduce the risk associated with incomplete data and help lower the number of charged-off loans by improving loan evaluations.

Borrowers' INF.

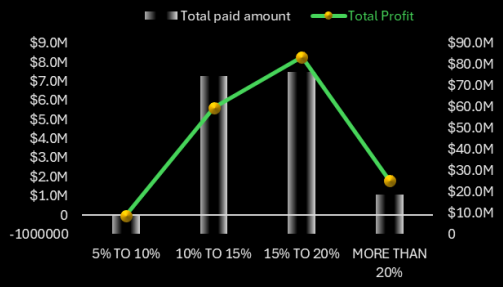
Observation:
Most of our customers have been employed for ten years, with annual incomes ranging between $24K and $84K. Many of them either rent or have a mortgage, and a significant number are self-employed. Regionally, customers are distributed across the U.S. as follows:

- South: 11,300
- West: 8,085
- Southwest: 1,498
- Northeast: 10,809
- Midwest: 5,563
- Northwest: 1,319

Recommendation:

1. Tailored Loan Products: Given the income range and employment length, consider offering loan products that cater specifically to self-employed individuals and those with moderate incomes, particularly targeting regions with higher customer volumes (South and Northeast).
2. Regional Marketing Focus: Concentrate marketing efforts in regions with high customer concentrations, like the South and Northeast, while also exploring opportunities to grow in underserved regions such as the Southwest and Northwest.
3. Flexible Homeownership Loan Options: Given that many customers either rent or have a mortgage, offering flexible loan products that cater to these homeownership statuses (e.g., loans for home improvements or refinancing options) could better serve their needs.

Impact of INT Rate



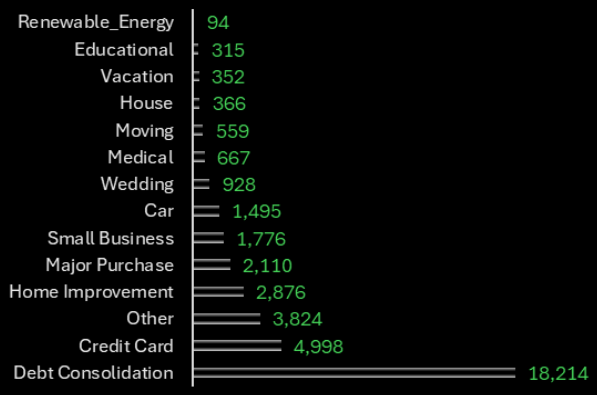
Observation:

The chart on the left shows the impact of interest rates on short-term loans (36 months), while the one on the right represents long-term loans (60 months). In the short term, interest rates between 10% and 15% are the most attractive to borrowers, with little to no demand for loans with rates above 20%. On the other hand, in the long term, interest rates between 10% and 20% are more attractive, with loans in the 15% to 20% range being more profitable than those in the 10% to 15% range, despite both ranges having a similar loan amount.

Recommendation:

1. Optimize Short-Term Loan Offerings: Focus on promoting loans with interest rates between 10% and 15% for short-term borrowers, as this is the most attractive range.
2. Maximize Long-Term Profitability: For long-term loans, target interest rates between 15% and 20% to balance profitability and borrower interest. Consider offering tailored incentives to encourage more borrowers to choose loans in this profitable range.
3. Avoid High-Interest Short-Term Loans: Since there is minimal demand for short-term loans with interest rates above 20%, consider revising or discontinuing those offers to focus resources on more attractive options for borrowers.

Intended Purpose

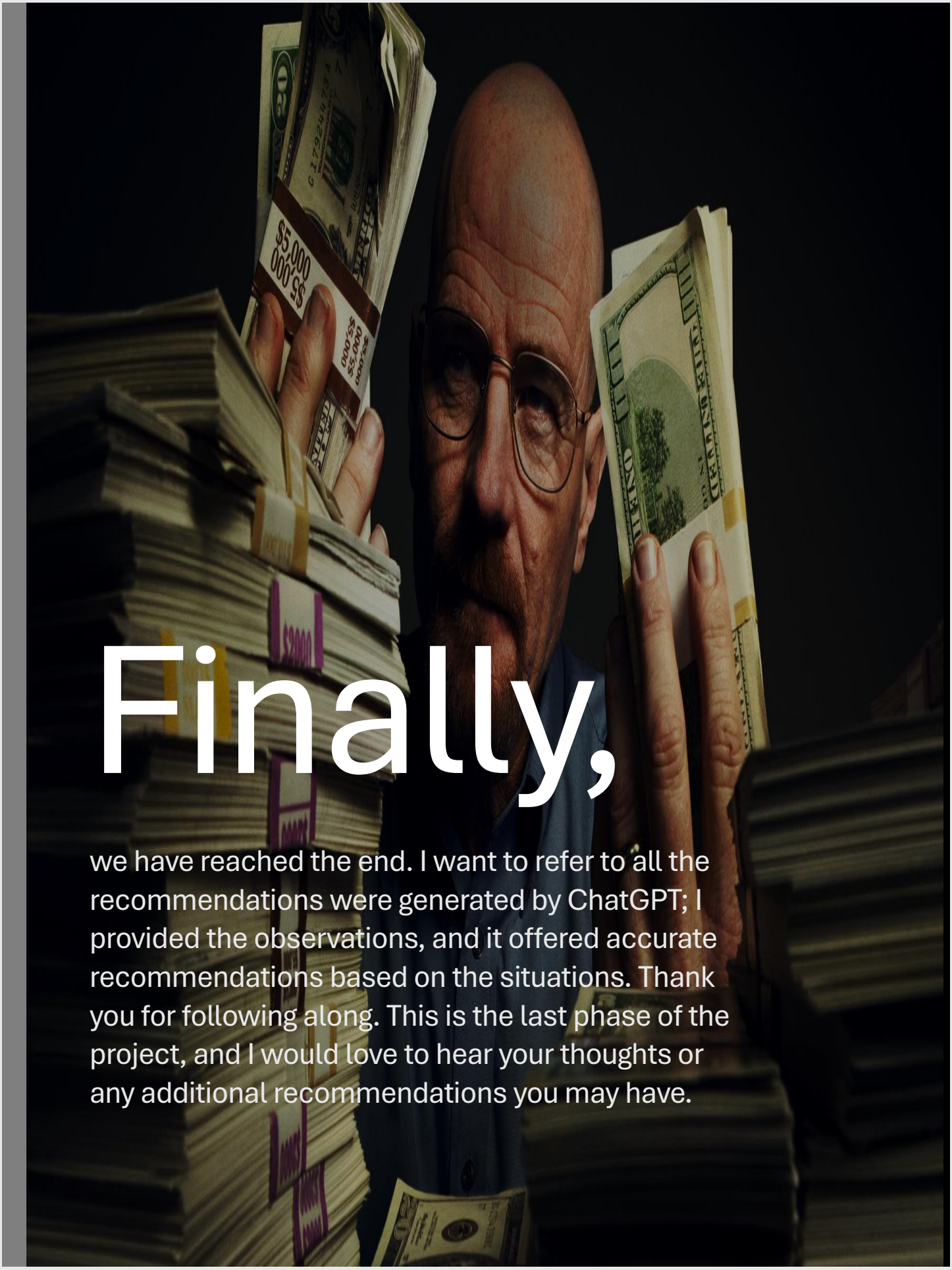


Observation:

The most common loan purpose is debt consolidation, with 18,214 loans, followed by credit card repayment with 4,998 loans. Other purposes collectively account for 15,362 loans.

Recommendation:

1. Focus on Debt Consolidation Products: Since debt consolidation is the most popular loan purpose, consider expanding and promoting tailored products in this category to capture more of this market.
2. Enhance Credit Card Repayment Offers: Given the significant number of loans for credit card repayment, offer competitive rates and flexible terms to attract more borrowers looking to consolidate credit card debt.
3. Diversify Offerings for Other Purposes: While debt consolidation and credit card repayment dominate, ensure that products for other loan purposes remain flexible and appealing to meet the diverse needs of remaining borrowers.



Finally,

we have reached the end. I want to refer to all the recommendations were generated by ChatGPT; I provided the observations, and it offered accurate recommendations based on the situations. Thank you for following along. This is the last phase of the project, and I would love to hear your thoughts or any additional recommendations you may have.