

PROJECT

2) Transformation Using Power Query

Data Cleaning:

Handling Missing Values and Duplicates:

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

	ber_id	loan_id	emp_length	home_ownership	annual_inc	verification_status
1	1302032	1067644	0	OWN	12,000.00	Source Verified
2	1290516	1058717	0	MORTGAGE	60,000.00	Verified
3	1298394	1064582	0	RENT	24,000.00	Source Verified
4	1298164	1064366	0	RENT	18,408.00	Not Verified
5	1296491	1063912	0	MORTGAGE	31,500.00	Verified
6	1295380	1062897	0	MORTGAGE	37,800.00	Not Verified
7	1294654	1062781	0	MORTGAGE	70,000.00	Source Verified
8	1294602	1062535	0	MORTGAGE	42,000.00	Not Verified
9	1266720	1036838	0	MORTGAGE	32,000.00	Verified
10	1292992	1061194	0	OWN	41,100.00	Not Verified
11	1292813	1060848	0	MORTGAGE	46,212.00	Verified
12	1292242	1060485	0	MORTGAGE	45,696.00	Verified
13	1290942	1059128	0	MORTGAGE	25,000.00	Not Verified
14	1287582	1056011	0	MORTGAGE	17,352.00	Verified
15	1282516	1051228	0	OWN	34,848.00	Verified
16	1246619	1018385	0	MORTGAGE	24,000.00	Verified
17	1289493	1057726	0	RENT	60,000.00	Verified
18	1288882	1057318	0	RENT	35,000.00	Not Verified
19	1146931	926180	0	OWN	40,000.00	Source Verified
20	1287654	1056081	0	OWN	57,000.00	Source Verified
21	1287282	1055716	0	RENT	27,050.00	Source Verified
22	1287125	1055566	0	RENT	84,855.00	Verified

➤ Remove rows with missing values in the 'last_pymnt_d' and 'delinq_2yrs' columns.

\$ dti	delinq_2yrs	last_pymnt_d	\$ total_pymnt
7.50	1	01-05-2019	4,185.93
16.17	1	01-02-2021	5,651.83
17.52	1	01-02-2020	23,731.05
14.94	1	01-12-2021	16,666.20
18.14	1	01-05-2020	2,888.13
12.22	1	01-09-2021	7,604.58
21.20	1	01-11-2020	4,896.44
7.53	1	01-12-2021	11,263.02
11.08	1	01-12-2019	6,967.17
17.85	1	01-01-2019	586.98
10.40	2	01-01-2021	7,657.70
24.12	2	01-11-2021	12,071.89
13.96	1	01-10-2019	5,050.50
17.31	1	01-11-2021	7,034.14
9.47	1	01-10-2019	8,570.86
2.40	1	01-05-2020	10,773.46
13.50	1	01-11-2021	22,396.83
9.14	1	01-05-2020	18,768.16
20.22	2	01-09-2020	3,335.80
16.97	2	01-10-2021	12,347.08
21.08	2	01-04-2015	55,362.54
24.29	1	01-09-2019	33,137.53

➤ Remove duplicate rows in the 'id' column of the "LoanDetails" table.

id	loan_amnt	\$ funded_amnt	term	\$ int_rate	\$ installment
1077501	5000	4,975.00	36 months	10.65	
1077430	2500	2,500.00	60 months	15.27	
1077175	2400	2,400.00	36 months	15.96	
1076863	10000	10,000.00	36 months	13.49	
1075358	3000	3,000.00	60 months	12.69	
1075269	5000	5,000.00	36 months	7.90	
1069639	7000	7,000.00	60 months	15.96	
1072053	3000	3,000.00	36 months	18.64	
1071795	5600	5,600.00	60 months	21.28	
1071570	5375	5,350.00	60 months	12.69	
1070078	6500	6,500.00	60 months	14.65	
1069908	12000	12,000.00	36 months	12.69	
1064687	9000	9,000.00	36 months	13.49	
1069866	3000	3,000.00	36 months	9.91	
1069057	10000	10,000.00	36 months	10.65	
1069759	1000	1,000.00	36 months	16.29	

PROPERTIES
Name
LoanDetails
All Properties
APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
Filtered Rows
Changed Type1
X Removed Duplicates

Dealing with Inconsistencies:

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

issue_d	loan_status	purpose	PROPERTIES
01-12-2018	Fully Paid	credit card	Name LoanDetails
01-12-2018	Charged Off	car	All Properties
01-12-2018	Fully Paid	small business	APPLIED STEPS
01-12-2018	Fully Paid	other	Source
01-12-2018	Current	other	Navigation
01-12-2018	Fully Paid	wedding	Promoted Headers
01-12-2018	Current	debt consolidation	Changed Type
01-12-2018	Fully Paid	car	Filtered Rows
01-12-2018	Charged Off	small business	Changed Type1
01-12-2018	Charged Off	other	Removed Duplicates
01-12-2018	Fully Paid	debt consolidation	X Replaced Value
01-12-2018	Fully Paid	debt consolidation	
01-12-2018	Charged Off	debt consolidation	
01-12-2018	Fully Paid	credit card	
01-12-2018	Charged Off	other	
01-12-2018	Fully Paid	debt consolidation	
01-12-2018	Fully Paid	home improvement	
01-12-2018	Fully Paid	major purchase	


➤ Format the 'purpose' and 'home_ownership' columns to proper case.

home_ownership	purpose
Rent	Credit Card
Mortgage	Car
Mortgage	Small Business
Rent	Other
Mortgage	Other
Rent	Wedding
Rent	Debt Consolidation
Rent	Car
Own	Small Business
Own	Other
Rent	Debt Consolidation
Rent	Debt Consolidation
Mortgage	Debt Consolidation
Mortgage	Credit Card
Rent	Other
Mortgage	Debt Consolidation
Rent	Home Improvement
Mortgage	Major Purchase
Mortgage	Medical
Mortgage	Debt Consolidation

Data Transformation:

Column Transformation:

- Change the data type of the 'total_pymnt' column to 'Fixed decimal number'.

 last_pymnt_d	\$ total_pymnt
01-05-2019	4,185.93
01-02-2021	5,651.83
01-02-2020	23,731.05
01-12-2021	16,666.20
01-05-2020	2,888.13
01-09-2021	7,604.58
01-11-2020	4,896.44
01-12-2021	11,263.02
01-12-2019	6,967.17
01-01-2019	586.98
01-01-2021	7,657.70
01-11-2021	12,071.89
01-10-2019	5,050.50
01-11-2021	7,034.14
01-10-2019	8,570.86
01-05-2020	10,773.46
01-11-2021	22,396.83
01-05-2020	18,768.16
01-09-2020	3,335.80
01-10-2021	12,347.08
01-04-2015	55,362.54
01-09-2019	33,137.53

- Round off the numbers in the 'funded_amnt' column to 2 decimal places.

23 loan_amnt	\$ funded_amnt
5000	4,975.00
2500	2,500.00
2400	2,400.00
10000	10,000.00
3000	3,000.00
5000	5,000.00
7000	7,000.00
3000	3,000.00
5600	5,600.00
5375	5,350.00
6500	6,500.00
12000	12,000.00
9000	9,000.00
3000	3,000.00
10000	10,000.00
1000	1,000.00
10000	10,000.00
3600	3,600.00
6000	6,000.00
9200	9,200.00
20250	19,142.16
21000	21,000.00

Column Renaming:

- Rename the column 'issue_d' to 'issue_date'.

^A _C sub_grade	^B _C issue_date
B2	01-12-2018
C4	01-12-2018
C5	01-12-2018
C1	01-12-2018
B5	01-12-2018
A4	01-12-2018
C5	01-12-2018
E1	01-12-2018
F2	01-12-2018
B5	01-12-2018
C3	01-12-2018
B5	01-12-2018
C1	01-12-2018
B1	01-12-2018

➤ Rename the column 'last_payment_d' to 'last_payment_date'.

² ₃ delinq_2yrs	¹ ₂ last payment date
1	01-05-2019
1	01-02-2021
1	01-02-2020
1	01-12-2021
1	01-05-2020
1	01-09-2021
1	01-11-2020
1	01-12-2021
1	01-12-2019
1	01-01-2019
2	01-01-2021
2	01-11-2021
1	01-10-2019

Creating New Columns:

➤ 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	
4,185.93	
5,651.83	
23,731.05	
16,666.20	
2,888.13	
7,604.58	
4,896.44	
11,263.02	
6,967.17	
586.98	
7,657.70	
12,071.89	
5,050.50	
7,034.14	
8,570.86	
10,773.46	
22,396.83	
18,768.16	

PROPERTIES
Name
BorrowerDetails
All Properties

APPLIED STEPS

Source	✱
Navigation	✱
Promoted Headers	✱
Changed Type	
Filtered Rows	
Replaced Value	✱
Filtered Rows1	
Capitalized Each Word	
Renamed Columns	
Added Custom	✱
✕ Changed Type1	

➤ 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".

\$ Total amount paid	ABC 123 delinquency_status
5,861.07	Not Delinquent
1,008.71	Not Delinquent
3,003.65	Not Delinquent
12,226.30	Not Delinquent
2,475.27	Not Delinquent
5,631.38	Not Delinquent
6,247.69	Not Delinquent
3,938.14	Not Delinquent
646.02	Not Delinquent
1,476.19	Not Delinquent
7,677.52	Not Delinquent
13,943.08	Not Delinquent
2,270.70	Not Delinquent
3,478.98	Not Delinquent
7,471.99	Not Delinquent
1,270.17	Not Delinquent
12,519.26	Not Delinquent
3,785.02	Not Delinquent
7,164.50	Delinquent
9,459.96	Not Delinquent
27,663.04	Not Delinquent
14,025.40	Not Delinquent

Column Dropping:

➤ Remove the 'sub_grade' column as that does not significantly contribute to the analysis.

\$ installment	A ^B _C grade	issue_date
162.87	B	01-12-2018
59.83	C	01-12-2018
84.33	C	01-12-2018
339.31	C	01-12-2018
67.79	B	01-12-2018
156.46	A	01-12-2018
170.08	C	01-12-2018
109.43	E	01-12-2018
152.39	F	01-12-2018
121.45	B	01-12-2018
153.45	C	01-12-2018
402.54	B	01-12-2018
305.38	C	01-12-2018
96.68	B	01-12-2018
325.74	B	01-12-2018
35.31	D	01-12-2018
347.98	C	01-12-2018

REPORT 1

Loan Performance Analysis

The Loan Performance Analysis report aims to provide insights into the performance of loans based on various factors such as loan amount, loan status, term, interest rate, and purpose.

- **Total Funded Amount:** Create a card visual to display the total funded amount.
- **Fully Paid Loan Percentage:** Create a gauge chart to display the 'Fully Paid Loan Percentage' measure.
- **Average Interest Rate by Term:** Create a multi-row card to show the average interest rate for each term.
- **Loan Status Distribution:** Create a pie chart to visualize the sum of total payments by loan status.
- **Loan Amount by Purpose:** Create a treemap to show the average loan amount by purpose.
- **Installment Over Time:** Create a line chart to visualize the sum of installments by Year and Quarter of the issue date.
- **Maximum Total Amount Paid by Loan Status:** Create a column chart to display the maximum total amount paid by loan status.
- **Minimum Annual Income by Grade:** Create a funnel chart to show the minimum annual income by grade.
- **Issue Date Slicer:** Add a slicer for the Month of the issue date to enable dynamic data exploration.

the average debt-to-income ratio by delinquency status.

➤ **Sum of Loan Amount by Home Ownership:** Create a table to show the total loan amount by home ownership.

➤ **Average Remaining Principal by Verification Status:** Create a donut chart to display the average remaining outstanding principal by verification status.

➤ **Sum of Delinquencies by Home Ownership:** Create a bar chart to show the total number of delinquencies in the past 2 years by home ownership and filter the visual to display only Mortgage, Rent, and Own.

➤ **Max Remaining Installments by Employment Length:** Create a treemap to show the maximum remaining installments by employment length.

➤ **Total Amount Paid and Funded Amount Over Time:** Create a line chart to display the sum of total amount paid and the sum of funded amount by the year of last payment date.

➤ **Purpose Slicer:** Add a slicer for loan purpose to enable dynamic data exploration.

