

# Phân tích dữ liệu cho vay của ngân hàng

## Objective:

Phân tích các khoản vay tốt và xấu cùng với các yếu tố liên quan như mục đích vay, thời gian vay, và độ dài thời gian làm việc của người vay,... Nhằm đưa ra các đề xuất giúp tối ưu hóa danh mục cho vay và quản lý rủi ro hiệu quả.

## Key Performance Indicators (KPIs) Requirements:

### Task 1: Total Loan Application

Total Loan Applications

Tổng số hợp đồng vay.

```
select count(distinct id) as Total_Loan_Applications from bank_loan
```

Total_Loan_Applications	
1	38576

### Month-To-Date

Vì dữ liệu của chúng ta không tính đến hiện tại nên chúng ta sẽ chỉ truy vấn theo từng tháng một.

Ví dụ như tháng 12 sẽ được code như sau:

```
select count(distinct id) as Total_Loan_Applications_MTD from bank_loan
where month(issue_date)=12
```

Total_Loan_Applications	
	4314

### Month To Month

Công thức tính MoM sẽ là  $\frac{MTD - PMTD}{PMTD}$ , tôi sẽ không biểu diễn công thức cụ thể bằng SQL vì chúng có thể thực hiện nhanh hơn trong Excel, ví dụ về MTD và PMTD sẽ được biểu diễn bằng đoạn code sau .

```
select count(distinct id) as PMTD_Total_Loan_Applications from bank_loan
where month(issue_date)=11
select count(distinct id) as MTD_Total_Loan_Applications from bank_loan
where month(issue_date)=12
```

### Task 2: Total Funded Amount

Tổng số tiền mà ngân hàng đã giải ngân:

```
select sum(loan_amount) as Total_Funded_Amount from bank_loan
```

Total_Funded_Amount	
1	435757075

Tổng số tiền đã giải ngân MTD, ví dụ với tháng 12.

```
select sum(loan_amount) as Total_Payment_Received_MTD from bank_loan
where month(issue_date)=12
```

	Total_Funded_Amount
1	53981425

Tổng số tiền giải ngân MoM, ví dụ với tháng 12 và tháng trước đó, công thức sẽ tương tự như ở trên và tôi xin phép không đề cập ở phần SQL.

```
select sum(loan_amount) as MTD_Total_Payment_Received from bank_loan
where month(issue_date)=12
select sum(loan_amount) as PMTD_Total_Payment_Received from bank_loan
where month(issue_date)=11
```

### Task 3: Total Amount Received

Tổng số tiền mà ngân hàng thu lại được:

```
select sum(total_payment) as Total_Payment_Received from bank_loan
```

	Total_Payment_Received
1	473070933

Tổng số tiền thu lại được MTD, ví dụ với tháng 12.

```
select sum(total_payment) as Total_Payment_Received_MTD from bank_loan
where month(issue_date)=12
```

	Total_Payment_Received_MTD
1	58074380

Tổng số tiền thu lại được MoM, ví dụ với tháng 12 và tháng trước đó, công thức sẽ tương tự như ở trên và tôi xin phép không đề cập ở phần SQL.

```
select sum(total_payment) as Total_Payment_Received_MTD from bank_loan
where month(issue_date)=12
select sum(total_payment) as Total_Payment_Received_PMTD from bank_loan
where month(issue_date)=11
```

### Task 4: Average Interest Rate

Lãi suất trung bình:

```
select avg(int_rate)*100 as Average_Interest_Rate from bank_loan
```

	Average_Interest_Rate
1	12.0488314172048

Lãi suất trung bình MTD, ví dụ với tháng 12:

```
select avg(int_rate)*100 as Average_Interest_Rate_MTD from bank_loan
where month(issue_date)=12
```

	Average_Interest_Rate
1	12.3560408676042

Lãi suất trung bình MoM, ví dụ với tháng 12 và tháng trước đó, công thức sẽ tương tự như ở trên và tôi xin phép không đề cập ở phần SQL.

```
select avg(int_rate)*100 as Average_Interest_Rate from bank_loan
where month(issue_date)=12
select avg(int_rate)*100 as PMTD_Average_Interest_Rate from bank_loan
where month(issue_date)=11
```

### Task 5: Average Debt to Income (DTI)

DTI trung bình:

```
select avg(dti) as Average_DTI from bank_loan
```

	Average_DTI
1	0.133274331211432

DTI trung bình MTD, ví dụ với tháng 12:

```
select avg(dti) as Average_DTI_MTD from bank_loan  
where month(issue_date)=12
```

	Average_DTI
1	0.136655377880425

DTI trung bình bình MoM, ví dụ với tháng 12 và tháng trước đó, công thức sẽ tương tự như ở trên và tôi xin phép không đề cập ở phần SQL.

```
select avg(dti) as Average_DTI_MTD from bank_loan  
where month(issue_date)=12  
select avg(dti) as Average_DTI_PMTD from bank_loan  
where month(issue_date)=11
```

## Good Loans vs Bad Loans KPI

### Good Loans

#### Good Loan Applications Percentage:

Tỉ lệ nợ tiêu chuẩn:

```
select (count(case when loan_status='Fully Paid' or loan_status='Current' then 1 end)*100)/count(distinct id) as  
Good_Loan_Application_Percentage  
from bank_loan
```

	Good_Loan_Application_Percentage
1	86

#### Good Loan Applications

Số lượng nợ tiêu chuẩn:

```
select COUNT(distinct id) as Good_Loan_Applications from bank_loan  
where loan_status='Fully Paid' or loan_status='Current'
```

	Good_Loan_Applications
1	33243

#### Good Loan Funded Amount

Tổng lượng tiền đã giải ngân cho nợ tiêu chuẩn:

```
select sum(loan_amount) as Good_Loan_Funded_Amount  
from bank_loan  
where loan_status='Fully Paid' or loan_status='Current'
```

	Good_Loan_Funded_Amount
	370224850

#### Good Loan Total Received Amount

Tổng tiền đã thu hồi được đối với nợ tiêu chuẩn:

```
select sum(total_payment) as Good_Loan_Funded_Amount  
from bank_loan  
where loan_status='Fully Paid' or loan_status='Current'
```

	Good_Loan_Funded_Amount
1	435786170

### Bad Loans

### Bad Loan Applications Percentage:

Tỉ lệ nợ xấu:

```
select (count(case when loan_status='Charged Off' then 1 end)*100)/count(distinct id) as  
Bad_Loan_Application_Percentage  
from bank_loan
```

	Bad_Loan_Application_Percentage
1	13

### Bad Loan Applications:

Số lượng nợ xấu:

```
select COUNT(distinct id) as Bad_Loan_Applications from bank_loan  
where loan_status='Charged Off'
```

	Bad_Loan_Applications
1	5333

### Bad Loan Funded Amount:

Tổng số tiền đã giải ngân cho nợ xấu:

```
select sum(loan_amount) as Bad_Loan_Funded_Amount  
from bank_loan  
where loan_status='Charged Off'
```

	Bad_Loan_Funded_Amount
1	65532225

### Bad Loan Total Received Amount:

Tổng số tiền đã thu hồi được từ nợ xấu:

```
select sum(total_payment) as Bad_Loan_Funded_Amount  
from bank_loan  
where loan_status='Charged Off'
```

	Bad_Loan_Funded_Amount
1	37284763

### Bank Loan Overview:

#### MONTH:

```
SELECT  
    MONTH(issue_date) AS Month_Munber,  
    DATENAME(MONTH, issue_date) AS Month_name,  
    COUNT(id) AS Total_Loan_Applications,  
    SUM(loan_amount) AS Total_Funded_Amount,  
    SUM(total_payment) AS Total_Amount_Received  
FROM bank_loan  
GROUP BY MONTH(issue_date), DATENAME(MONTH, issue_date)  
ORDER BY MONTH(issue_date)
```

	Month_Munber	Month_name	Total_Loan_Applications	Total_Funded_Amount	Total_Amount_Received
1	1	January	2332	25031650	27578836
2	2	February	2279	24647825	27717745
3	3	March	2627	28875700	32264400
4	4	April	2755	29800800	32495533
5	5	May	2911	31738350	33750523
6	6	June	3184	34161475	36164533
7	7	July	3366	35813900	38827220
8	8	August	3441	38149600	42682218
9	9	September	3536	40907725	43983948
10	10	October	3796	44893800	49399567
11	11	November	4035	47754825	50132030
12	12	December	4314	53981425	58074380

**STATE:**

**SELECT**

address\_state **AS** State,

**COUNT**(id) **AS** Total\_Loan\_Applications,

**SUM**(loan\_amount) **AS** Total\_Funded\_Amount,

**SUM**(total\_payment) **AS** Total\_Amount\_Received

**FROM** bank\_loan

**GROUP BY** address\_state

**ORDER BY** address\_state

	State	Total_Loan_Applications	Total_Funded_Amount	Total_Amount_Received
12	HI	170	1850525	2080184
13	IA	5	56450	64482
14	ID	6	59750	65329
15	IL	1486	17124225	18875941
16	IN	9	86225	85521
17	KS	260	2872325	3247394
18	KY	320	3504100	3792530
19	LA	426	4498900	5001160
20	MA	1310	15051000	16676279
21	MD	1027	11911400	12985170
22	ME	3	9200	10808
23	MI	685	7829900	8543660
24	MN	592	6302600	6750746
25	MO	660	7151175	7692732
26	MS	19	139125	149342
27	MT	79	829525	892047
28	NC	759	8787575	9534813
29	NE	5	31700	24542
30	NH	161	1917900	2101386
31	NJ	1822	21657475	23425159
32	NM	183	1916775	2084485
33	NV	482	5307375	5451443
34	NY	3701	42077050	46108181
35	OH	1188	12991375	14330148
36	OK	293	3365725	3712649
37	OR	436	4720150	4966903
38	PA	1482	15826525	17462908
39	RI	196	1883025	2001774
40	SC	464	5080475	5462458
41	SD	63	606150	656514
42	TN	17	162175	141522
43	TX	2664	31236650	34392715
44	UT	252	2849225	2952412
45	VA	1375	15982650	17711443
46	VT	54	504100	534973
47	WA	805	8855525	9531739
48	WI	446	5070450	5485161
49	WV	167	1830525	1991936
50	WY	79	890750	1046050

**Term:**

**SELECT**

term AS Term,  
COUNT(id) AS Total\_Loan\_Applications,  
SUM(loan\_amount) AS Total\_Funded\_Amount,

SUM(total\_payment) AS Total\_Amount\_Received  
 FROM bank\_loan  
 GROUP BY term  
 ORDER BY term

	Term	Total_Loan_Applications	Total_Funded_Amount	Total_Amount_Received
1	36 months	28237	273041225	294709458
2	60 months	10339	162715850	178361475

## Employment Length:

SELECT  
 emp\_length AS Employment\_Length,  
 COUNT(id) AS Total\_Loan\_Applications,  
 SUM(loan\_amount) AS Total\_Funded\_Amount,  
 SUM(total\_payment) AS Total\_Amount\_Received  
 FROM bank\_loan  
 GROUP BY emp\_length  
 ORDER BY emp\_length

	Employee_Length	Total_Loan_Applications	Total_Funded_Amount	Total_Amount_Received
1	< 1 year	4575	44210625	47545011
2	1 year	3229	32883125	35498348
3	10+ years	8870	116115950	125871616
4	2 years	4382	44967975	49206961
5	3 years	4088	43937850	47551832
6	4 years	3428	37600375	40964850
7	5 years	3273	36973625	40397571
8	6 years	2228	25612650	27908658
9	7 years	1772	20811725	22584136
10	8 years	1476	17558950	19025777
11	9 years	1255	15084225	16516173

## Purpose:

SELECT  
 purpose AS PURPOSE,  
 COUNT(id) AS Total\_Loan\_Applications,  
 SUM(loan\_amount) AS Total\_Funded\_Amount,  
 SUM(total\_payment) AS Total\_Amount\_Received

FROM bank\_loan  
GROUP BY purpose  
ORDER BY purpose

	PURPOSE	Total_Loan_Applications	Total_Funded_Amount	Total_Amount_Received
1	car	1497	10223575	11324914
2	credit card	4998	58885175	65214084
3	Debt consolidation	18214	232459675	253801871
4	educational	315	2161650	2248380
5	home improvement	2876	33350775	36380930
6	house	366	4824925	5185538
7	major purchase	2110	17251600	18676927
8	medical	667	5533225	5851372
9	moving	559	3748125	3999899
10	other	3824	31155750	33289676
11	renewable_energy	94	845750	898931
12	small business	1776	24123100	23814817
13	vacation	352	1967950	2116738
14	wedding	928	9225800	10266856

## HOME OWNERSHIP:

SELECT

home\_ownership AS Home\_Ownership,  
COUNT(id) AS Total\_Loan\_Applications,  
SUM(loan\_amount) AS Total\_Funded\_Amount,  
SUM(total\_payment) AS Total\_Amount\_Received

FROM bank\_loan

GROUP BY home\_ownership

ORDER BY home\_ownership

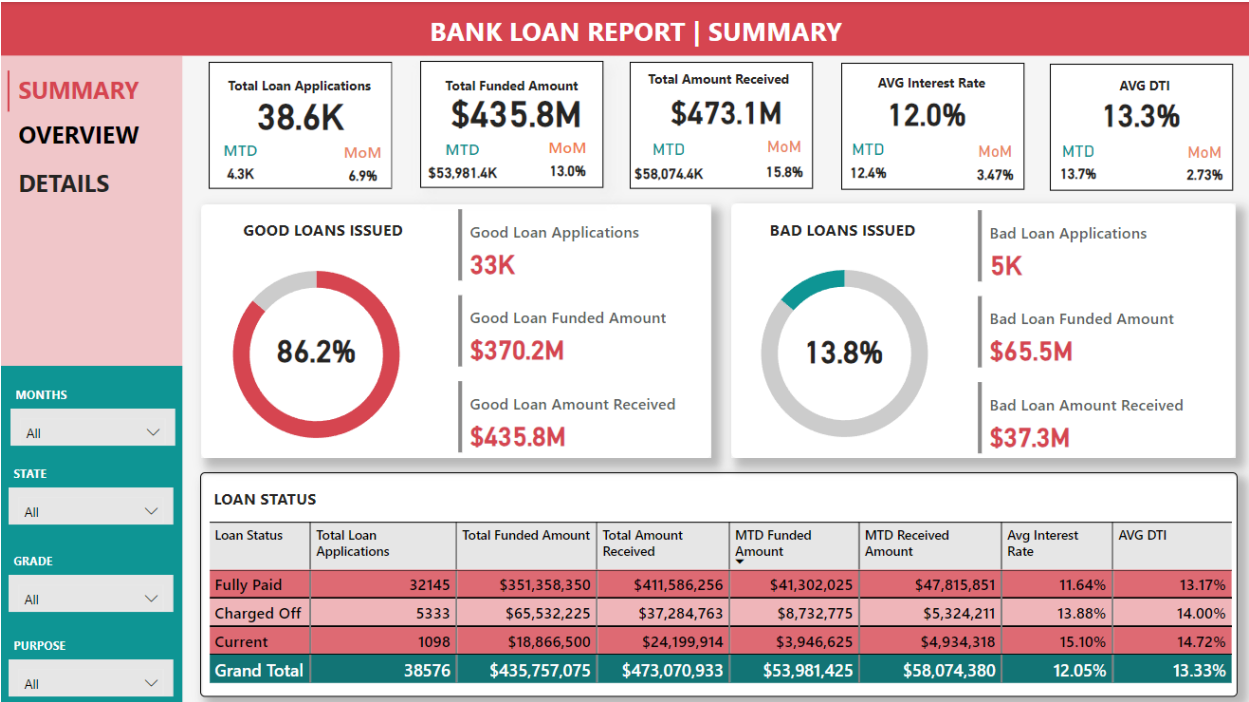
	Home_Ownership	Total_Loan_Applications	Total_Funded_Amount	Total_Amount_Received
1	MORTGAGE	17198	219329150	238474438
2	NONE	3	16800	19053
3	OTHER	98	1044975	1025257
4	OWN	2838	29597675	31729129
5	RENT	18439	185768475	201823056



## Phân tích từ các biểu đồ của Power Bi

Ở phần trước tôi đã dùng SQL để truy vấn các chỉ số quan trọng trong bộ dữ liệu, ở phần này tôi sẽ dùng các số liệu đó để vẽ biểu đồ và phân tích trong Power Bi. Các bạn có thể tải file Power Bi từ link github của tôi để trực tiếp tương tác với Dashboard

Dưới đây là Dashboard tổng hợp lại những chỉ số quan trọng đối với các khoản vay



Về tổng quan ta có thể thấy ngân hàng quản lý khá tốt nguồn vốn của mình khi cho vay được 435.8M và thu về được 473.1M.

Đối với **Good Loans**, có tổng số 33K hồ sơ, chiếm 86.2% tổng số hồ sơ. Tổng số tiền giải ngân của nhóm này là 370M\$, cao hơn nhiều so với nhóm Bad Loans. Số tiền thu về đạt 435.8M, cho thấy ngân hàng không chỉ giải ngân được 1 khoản lớn mà còn thu hồi được phần lớn số tiền

Đối với **Bad Loans**, chỉ có 5K hồ sơ, chiếm 13.8% tổng số hồ sơ, một dấu hiệu tốt cho thấy rủi ro tín dụng đang được quản lý khá tốt. Số tiền giải ngân cho các khoản vay này là \$65.5M, nhỏ hơn nhiều so với nhóm "Good Loans". Tuy nhiên, số tiền nhận về chỉ là \$37.3M, tạo ra một khoảng cách lớn so với số tiền đã cho vay, chứng tỏ mức độ rủi ro cao với các khoản vay xấu.

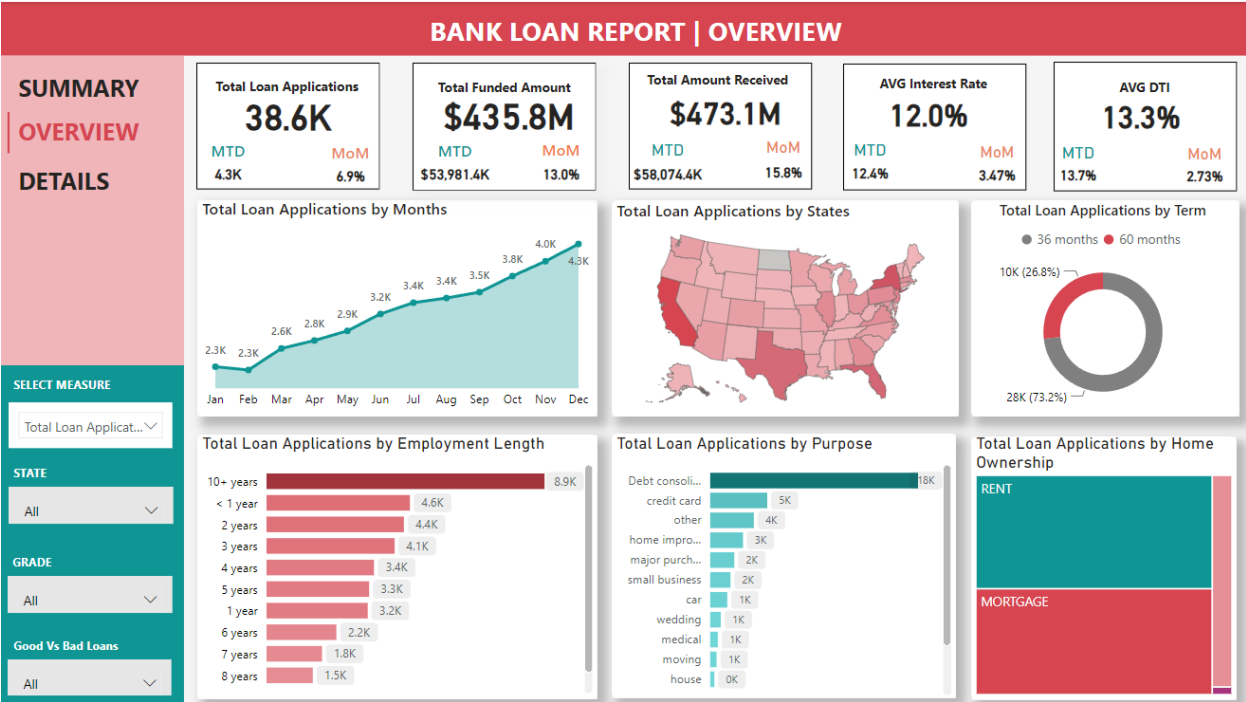
Đối với thành phần phụ của các khoản vay:

- Fully Paid:** Chiếm đa số với 32,145 hồ sơ và tổng số tiền giải ngân là \$351.3M. Điều này là một dấu hiệu tích cực, cho thấy đa phần các khoản vay "Good Loans" đã được thanh toán đầy đủ. Số tiền nhận về cũng cao nhất, đạt \$411.5M.
- Charged Off:** Có 5,333 hồ sơ đã bị **Charged Off** (nghĩa là ngân hàng coi như khoản nợ này không thể thu hồi). Đây là nhóm rủi ro cao, với tổng số tiền giải ngân \$65.5M nhưng chỉ thu được \$37.2M. Điều này đáng chú ý vì tỉ lệ thất thoát vốn ở nhóm này là hơn 43%.

- **Current:** Có 1,098 khoản vay hiện vẫn đang thanh toán, với tổng số tiền giải ngân là \$18.8M và số tiền đã thu được là \$24.1M, cho thấy khả năng thu hồi vốn vẫn còn tồn tại, nhưng cũng là điểm rủi ro cần quản lý kỹ.

### Dashboard Overview

Dưới đây là phân tích chi tiết về Dashboard Overview, tập trung vào các biểu đồ và số liệu chính, đồng thời giải thích cách chúng giúp quản lý và đánh giá hiệu quả cho vay.



### Loan Applications by Month

Biểu đồ này cho thấy xu hướng tăng trưởng ổn định trong suốt cả năm, với đỉnh điểm đạt vào tháng 12 với 4.3K hồ sơ. Sự gia tăng đều qua các tháng cũng có thể phản ánh sự phát triển ổn định của ngân hàng trong việc thu hút khách hàng và duy trì nguồn cầu liên tục.

### Loan Applications by State

- Các bang với sắc độ đậm hơn có số lượng hồ sơ vay cao hơn. Điều này cung cấp góc nhìn địa lý về hoạt động vay vốn của ngân hàng. Các bang lớn và đông dân như California, Texas có lẽ sẽ là những nơi cần được phân tích chi tiết về rủi ro và cơ hội.
- Ngân hàng có thể tập trung phát triển thêm các sản phẩm tài chính cho các bang có nhu cầu cao, đồng thời thiết lập các biện pháp quản lý rủi ro hiệu quả hơn tại các bang có hồ sơ vay cao nhưng chất lượng tín dụng thấp.

### Loan Applications by emp\_length

- Người có kinh nghiệm làm việc trên 10 năm chiếm tỷ lệ cao nhất trong số các hồ sơ vay, với 8.9K hồ sơ. Điều này có thể phản ánh rằng những người có công việc ổn định lâu dài thường có khả năng vay tín dụng tốt hơn.

- Ngân hàng có thể tiếp tục phát triển các chương trình vay tín dụng dành riêng cho đối tượng này, hoặc xem xét giảm lãi suất để khuyến khích thêm nhóm đối tượng có kinh nghiệm làm việc lâu năm.

#### **Loan Applications by purpose**

- Phần lớn các khoản vay là dành cho **Debt Consolidation** (tái cơ cấu nợ) với 18K hồ sơ, sau đó là **Credit Card** với 5K hồ sơ. Điều này cho thấy người dùng thường sử dụng khoản vay để quản lý nợ cá nhân, đặc biệt là các khoản nợ tín dụng.
- Ngân hàng có thể phát triển thêm các sản phẩm hỗ trợ quản lý nợ cá nhân như các chương trình cho vay tín dụng với lãi suất thấp để giảm áp lực nợ tín dụng cao từ các khoản vay tiêu dùng.

#### **Loan Applications by term**

- Phần lớn các khoản vay có thời hạn 36 tháng (73.2%). Khoản vay ngắn hạn thường phù hợp với các nhu cầu tài chính nhỏ hơn hoặc mục tiêu ngắn hạn như tiêu dùng cá nhân hoặc sửa chữa nhà cửa.
- Ngân hàng có thể xem xét mở rộng các gói vay dài hạn với lãi suất cạnh tranh để thu hút thêm khách hàng có nhu cầu vay vốn lớn, chẳng hạn như đầu tư bất động sản hoặc kinh doanh.

#### **Loan Applications by Home Ownership**

- **Tỷ lệ người thuê (Rent):** Với 18.000 đơn vay, chiếm tỷ lệ lớn nhất. Điều này có thể chỉ ra rằng họ có nhu cầu vay tiền cao hơn, có thể do họ cần tài chính để chi trả cho các chi phí khác hoặc để cải thiện điều kiện sống. Việc không sở hữu tài sản có thể khiến họ phải tìm đến các nguồn tài chính ngoài để đáp ứng nhu cầu.
- **Tỷ lệ nhà thế chấp (Mortgage):** Với 17.000 đơn vay, nhóm này cũng có số lượng khá cao. Điều này cho thấy rằng nhiều người đã sở hữu nhà nhưng vẫn cần vay thêm tiền, có thể để cải tạo nhà, tái tài trợ hoặc đầu tư vào các dự án khác.
- **Tỷ lệ người sở hữu (Own):** Chỉ có 3.000 đơn vay từ nhóm này, cho thấy rằng những người sở hữu nhà có xu hướng ít cần vay hơn. Họ có thể đã có nguồn tài chính ổn định hơn hoặc ít phải phụ thuộc vào vay mượn.

Các chỉ số khác như Amount Received hay Funded Amount cũng thể hiện chỉ số tương tự như Loan Applications, các bạn có thể tải file Power BI về để kiểm tra

### Dashboard Details

Dưới đây là dashboard tổng hợp các chi tiết từ bộ dữ liệu và các phân tích quan trọng mà tôi đã trình bày. Nếu bạn muốn xem chi tiết hơn, bạn có thể tải file Power BI để khám phá thêm.



### Kết luận

Ngân hàng hiện tại đang có một tỷ lệ khá tốt giữa các khoản vay tốt và xấu. Tuy nhiên, nhóm "Bad Loans" vẫn có một tỷ lệ rủi ro cao, đặc biệt là nhóm **Charged Off**, nơi có khoảng cách lớn giữa số tiền giải ngân và số tiền thu hồi.

Phân tích chi tiết dữ liệu theo mục đích, thời gian vay, và tỷ lệ trả góp cũng cho thấy rằng ngân hàng có thể tối ưu hóa lợi nhuận và quản lý rủi ro tốt hơn nếu áp dụng các chương trình vay tín dụng cá nhân hóa và định hướng theo nhu cầu của khách hàng.