

Project 2 – your title here

MTH 161 – Fall 2024

Lily Mae Arenz

2024-11-10

Research Question

My research questions is as follows; Can Credit Score (predictor) effect the Loan Approval (outcome)?

Rationale for the Research Question

This research question allows one to evaluate how credit score could potentially impact loan approval. The answer to this question (if the null is likely true), could encourage society, especially younger generations to find ways to improve their credit score; making sure they can pay back their debt in a timely manner by not taking out various loans or more credit cards then they can handle. If the null is likely not true and the alternative is likely true this could encourage researchers to investigate other variables that could impact loan approval.

Hypotheses

Null: Credit score does not effect the outcome of the loan (approval or rejection). There is no correlation between the two variables. Alternative: Credit score does effect the loan approval. The variables share a positive correlation where increased credit score can lead to increased probability of loan approval.

Background Information on the DATA

The data utilized to awnser this question is from OpenIntro. OpenIntro provides publicly sourced data for educational purposes. The data that I found was from a Lending Club. The lending club evaluates a clients “credit worthiness” assigning them a loan grade and interest rate. Out of the loans, those that are approved are funded by investors who earn

interest on repayments.

The data set evaluates the following characteristics; data characteristics, loan details, credit information, loan outcomes and other financial metrics. Each of the categories are explained below.

Borrower characteristics provide information about who applies for the loans - their income, employment status, and housing situation. The variable `annual_inc` (numerical) measures the borrower's annual income in dollars. The variable, `emp_length` (numerical) shows their time of employment in years. The variable `home_ownership` (categorical), shows housing status (whether the borrower rents, owns, or has a mortgage). Additionally, the variable `purpose` (categorical) describes the purpose of the loan - reduction of debt, home improvement or credit card payments.

The loan details focuses on the specifics of the loan issued. The variable `loan_amnt` (numerical) represents the amount of money the borrower is requesting. The variable `int_rate` (numerical) is the annual interest rate of the loan, as a percentage. The repayment variable (categorical), is the time in which it takes to repay the loan which is marked by 36 or 60 months. Loan creditworthiness is measured through grade (categorical), ranging from A (highest creditworthiness) to G (lowest creditworthiness). In each of these grades there is also a sub grade A1, A2, or C5.

The credit information highlights the borrower's credit history and financial behavior. The variables `fico_range_high` and `fico_range_low` (both numerical) define the borrower's creditworthiness by specifying the range of their FICO score range. The variable `inq_last_6mths` (numerical) indicates the number of recent credit inquiries in the borrower's financial history. Additionally, `revol_util` (numerical) represents the credit utilization rate, as a percentage.

The loan outcome is essentially the status of loans. The variable `loan_status` (categorical) describes the current state of the loan, such as Fully Paid, Charged Off, Default, Late, or Current. Another variable, is the `total_pymnt` (numerical), which measures the total amount repaid by the borrower including interest.

Finally, the financial metrics includes characteristics of borrowers' financial health. The variable `dti` (numerical), shows the percentage of the borrower's income dedicated to debt payments. `installment` (numerical) shows the fixed monthly payment amount the borrower needs to make for their loan. Additionally, `delinq_2yrs` (numerical) records the number of delinquent credit accounts the borrower has had in the past two years.

Sources

Sources; Loan dataset: https://www.openintro.org/data/index.php?data=loans_full_schema

```
loans_data <- read.csv("Datasets/loans_full_schema.csv")  
  
glimpse(loans_data)
```

Rows: 10,000

Columns: 55

\$ emp_title	<chr> "global config engineer ", "warehouse~
\$ emp_length	<int> 3, 10, 3, 1, 10, NA, 10, 10, 10, 3, 1~
\$ state	<chr> "NJ", "HI", "WI", "PA", "CA", "KY", "~
\$ homeownership	<chr> "MORTGAGE", "RENT", "RENT", "RENT", "~
\$ annual_income	<dbl> 90000, 40000, 40000, 30000, 35000, 34~
\$ verified_income	<chr> "Verified", "Not Verified", "Source V~
\$ debt_to_income	<dbl> 18.01, 5.04, 21.15, 10.16, 57.96, 6.4~
\$ annual_income_joint	<dbl> NA, NA, NA, NA, 57000, NA, 155000, NA~
\$ verification_income_joint	<chr> "", "", "", "", "Verified", "", "Not ~
\$ debt_to_income_joint	<dbl> NA, NA, NA, NA, 37.66, NA, 13.12, NA,~
\$ delinq_2y	<int> 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0~
\$ months_since_last_delinq	<int> 38, NA, 28, NA, NA, 3, NA, 19, 18, NA~
\$ earliest_credit_line	<int> 2001, 1996, 2006, 2007, 2008, 1990, 2~
\$ inquiries_last_12m	<int> 6, 1, 4, 0, 7, 6, 1, 1, 3, 0, 4, 4, 8~
\$ total_credit_lines	<int> 28, 30, 31, 4, 22, 32, 12, 30, 35, 9,~
\$ open_credit_lines	<int> 10, 14, 10, 4, 16, 12, 10, 15, 21, 6,~
\$ total_credit_limit	<int> 70795, 28800, 24193, 25400, 69839, 42~
\$ total_credit_utilized	<int> 38767, 4321, 16000, 4997, 52722, 3898~
\$ num_collections_last_12m	<int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
\$ num_historical_failed_to_pay	<int> 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0~
\$ months_since_90d_late	<int> 38, NA, 28, NA, NA, 60, NA, 71, 18, N~
\$ current_accounts_delinq	<int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
\$ total_collection_amount_ever	<int> 1250, 0, 432, 0, 0, 0, 0, 0, 0, 0, 0,~
\$ current_installment_accounts	<int> 2, 0, 1, 1, 1, 0, 2, 2, 6, 1, 2, 1, 2~
\$ accounts_opened_24m	<int> 5, 11, 13, 1, 6, 2, 1, 4, 10, 5, 6, 7~
\$ months_since_last_credit_inquiry	<int> 5, 8, 7, 15, 4, 5, 9, 7, 4, 17, 3, 4,~
\$ num_satisfactory_accounts	<int> 10, 14, 10, 4, 16, 12, 10, 15, 21, 6,~
\$ num_accounts_120d_past_due	<int> 0, 0, 0, 0, 0, 0, 0, 0, NA, 0, 0, 0, 0, ~
\$ num_accounts_30d_past_due	<int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
\$ num_active_debit_accounts	<int> 2, 3, 3, 2, 10, 1, 3, 5, 11, 3, 2, 2,~
\$ total_debit_limit	<int> 11100, 16500, 4300, 19400, 32700, 272~
\$ num_total_cc_accounts	<int> 14, 24, 14, 3, 20, 27, 8, 16, 19, 7, ~
\$ num_open_cc_accounts	<int> 8, 14, 8, 3, 15, 12, 7, 12, 14, 5, 8,~
\$ num_cc_carrying_balance	<int> 6, 4, 6, 2, 13, 5, 6, 10, 14, 3, 5, 3~
\$ num_mort_accounts	<int> 1, 0, 0, 0, 0, 3, 2, 7, 2, 0, 2, 3, 3~
\$ account_never_delinq_percent	<dbl> 92.9, 100.0, 93.5, 100.0, 100.0, 78.1~
\$ tax_liens	<int> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0~
\$ public_record_bankrupt	<int> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0~
\$ loan_purpose	<chr> "moving", "debt_consolidation", "othe~
\$ application_type	<chr> "individual", "individual", "individu~
\$ loan_amount	<int> 28000, 5000, 2000, 21600, 23000, 5000~

```

$ term                <int> 60, 36, 36, 36, 36, 36, 60, 60, 36, 3~
$ interest_rate        <dbl> 14.07, 12.61, 17.09, 6.72, 14.07, 6.7~
$ installment          <dbl> 652.53, 167.54, 71.40, 664.19, 786.87~
$ grade                <chr> "C", "C", "D", "A", "C", "A", "C", "B~
$ sub_grade            <chr> "C3", "C1", "D1", "A3", "C3", "A3", "~
$ issue_month          <chr> "Mar-2018", "Feb-2018", "Feb-2018", "~
$ loan_status          <chr> "Current", "Current", "Current", "Cur~
$ initial_listing_status <chr> "whole", "whole", "fractional", "whol~
$ disbursement_method  <chr> "Cash", "Cash", "Cash", "Cash", "Cash~
$ balance              <dbl> 27015.86, 4651.37, 1824.63, 18853.26,~
$ paid_total           <dbl> 1999.330, 499.120, 281.800, 3312.890,~
$ paid_principal       <dbl> 984.14, 348.63, 175.37, 2746.74, 1569~
$ paid_interest        <dbl> 1015.19, 150.49, 106.43, 566.15, 754.~
$ paid_late_fees       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~

```

```
head(loans_data)
```

```

      emp_title emp_length state homeownership annual_income
1 global config engineer      3    NJ      MORTGAGE      90000
2 warehouse office clerk     10    HI      RENT      40000
3           assembly         3    WI      RENT      40000
4      customer service       1    PA      RENT      30000
5 security supervisor      10    CA      RENT      35000
6                                NA    KY      OWN      34000
  verified_income debt_to_income annual_income_joint verification_income_joint
1      Verified      18.01                NA
2    Not Verified      5.04                NA
3 Source Verified     21.15                NA
4    Not Verified     10.16                NA
5      Verified     57.96             57000      Verified
6    Not Verified      6.46                NA
  debt_to_income_joint delinq_2y months_since_last_delinq earliest_credit_line
1                NA         0                38                2001
2                NA         0                NA                1996
3                NA         0                28                2006
4                NA         0                NA                2007
5             37.66         0                NA                2008
6                NA         1                 3                1990
  inquiries_last_12m total_credit_lines open_credit_lines total_credit_limit
1                 6                28                10             70795
2                 1                30                14             28800
3                 4                31                10             24193

```

4	0	4	4	25400
5	7	22	16	69839
6	6	32	12	42100
	total_credit_utilized	num_collections_last_12m	num_historical_failed_to_pay	
1	38767	0	0	
2	4321	0	1	
3	16000	0	0	
4	4997	0	1	
5	52722	0	0	
6	3898	0	0	
	months_since_90d_late	current_accounts_delinq	total_collection_amount_ever	
1	38	0	1250	
2	NA	0	0	
3	28	0	432	
4	NA	0	0	
5	NA	0	0	
6	60	0	0	
	current_installment_accounts	accounts_opened_24m		
1	2	5		
2	0	11		
3	1	13		
4	1	1		
5	1	6		
6	0	2		
	months_since_last_credit_inquiry	num_satisfactory_accounts		
1	5	10		
2	8	14		
3	7	10		
4	15	4		
5	4	16		
6	5	12		
	num_accounts_120d_past_due	num_accounts_30d_past_due		
1	0	0		
2	0	0		
3	0	0		
4	0	0		
5	0	0		
6	0	0		
	num_active_debit_accounts	total_debit_limit	num_total_cc_accounts	
1	2	11100	14	
2	3	16500	24	
3	3	4300	14	
4	2	19400	3	

5		10		32700		20
6		1		27200		27
	num_open_cc_accounts	num_cc_carrying_balance	num_mort_accounts			
1	8		6			1
2	14		4			0
3	8		6			0
4	3		2			0
5	15		13			0
6	12		5			3
	account_never_delinq_percent	tax_liens	public_record_bankrupt			
1	92.9	0				0
2	100.0	0				1
3	93.5	0				0
4	100.0	1				0
5	100.0	0				0
6	78.1	0				0
	loan_purpose	application_type	loan_amount	term	interest_rate	
1	moving	individual	28000	60	14.07	
2	debt_consolidation	individual	5000	36	12.61	
3	other	individual	2000	36	17.09	
4	debt_consolidation	individual	21600	36	6.72	
5	credit_card	joint	23000	36	14.07	
6	other	individual	5000	36	6.72	
	installment	grade	sub_grade	issue_month	loan_status	initial_listing_status
1	652.53	C	C3	Mar-2018	Current	whole
2	167.54	C	C1	Feb-2018	Current	whole
3	71.40	D	D1	Feb-2018	Current	fractional
4	664.19	A	A3	Jan-2018	Current	whole
5	786.87	C	C3	Mar-2018	Current	whole
6	153.75	A	A3	Jan-2018	Current	whole
	disbursement_method	balance	paid_total	paid_principal	paid_interest	
1	Cash	27015.86	1999.33	984.14	1015.19	
2	Cash	4651.37	499.12	348.63	150.49	
3	Cash	1824.63	281.80	175.37	106.43	
4	Cash	18853.26	3312.89	2746.74	566.15	
5	Cash	21430.15	2324.65	1569.85	754.80	
6	Cash	4256.71	873.13	743.29	129.84	
	paid_late_fees					
1	0					
2	0					
3	0					
4	0					
5	0					

```
colnames(loans_data)
```

```
[1] "emp_title"           "emp_length"
[3] "state"              "homeownership"
[5] "annual_income"      "verified_income"
[7] "debt_to_income"     "annual_income_joint"
[9] "verification_income_joint" "debt_to_income_joint"
[11] "delinq_2y"          "months_since_last_delinq"
[13] "earliest_credit_line" "inquiries_last_12m"
[15] "total_credit_lines" "open_credit_lines"
[17] "total_credit_limit" "total_credit_utilized"
[19] "num_collections_last_12m" "num_historical_failed_to_pay"
[21] "months_since_90d_late" "current_accounts_delinq"
[23] "total_collection_amount_ever" "current_installment_accounts"
[25] "accounts_opened_24m" "months_since_last_credit_inquiry"
[27] "num_satisfactory_accounts" "num_accounts_120d_past_due"
[29] "num_accounts_30d_past_due" "num_active_debit_accounts"
[31] "total_debit_limit" "num_total_cc_accounts"
[33] "num_open_cc_accounts" "num_cc_carrying_balance"
[35] "num_mort_accounts" "account_never_delinq_percent"
[37] "tax_liens" "public_record_bankrupt"
[39] "loan_purpose" "application_type"
[41] "loan_amount" "term"
[43] "interest_rate" "installment"
[45] "grade" "sub_grade"
[47] "issue_month" "loan_status"
[49] "initial_listing_status" "disbursement_method"
[51] "balance" "paid_total"
[53] "paid_principal" "paid_interest"
[55] "paid_late_fees"
```

```
summary(loans_data)
```

emp_title	emp_length	state	homeownership
Length:10000	Min. : 0.00	Length:10000	Length:10000
Class :character	1st Qu.: 2.00	Class :character	Class :character
Mode :character	Median : 6.00	Mode :character	Mode :character
	Mean : 5.93		
	3rd Qu.:10.00		

	Max.	:10.00		
	NA's	:817		
annual_income	verified_income	debt_to_income	annual_income_joint	
Min. : 0	Length:10000	Min. : 0.00	Min. : 19200	
1st Qu.: 45000	Class :character	1st Qu.: 11.06	1st Qu.: 86834	
Median : 65000	Mode :character	Median : 17.57	Median : 113000	
Mean : 79222		Mean : 19.31	Mean : 127915	
3rd Qu.: 95000		3rd Qu.: 25.00	3rd Qu.: 151546	
Max. :2300000		Max. :469.09	Max. :1100000	
		NA's :24	NA's :8505	
verification_income_joint	debt_to_income_joint	delinq_2y		
Length:10000	Min. : 0.32	Min. : 0.000		
Class :character	1st Qu.:14.16	1st Qu.: 0.000		
Mode :character	Median :19.72	Median : 0.000		
	Mean :19.98	Mean : 0.216		
	3rd Qu.:25.50	3rd Qu.: 0.000		
	Max. :39.98	Max. :13.000		
	NA's :8505			
months_since_last_delinq	earliest_credit_line	inquiries_last_12m		
Min. : 1.00	Min. :1963	Min. : 0.000		
1st Qu.: 19.00	1st Qu.:1997	1st Qu.: 0.000		
Median : 34.00	Median :2003	Median : 1.000		
Mean : 36.76	Mean :2001	Mean : 1.958		
3rd Qu.: 53.00	3rd Qu.:2006	3rd Qu.: 3.000		
Max. :118.00	Max. :2015	Max. :29.000		
NA's :5658				
total_credit_lines	open_credit_lines	total_credit_limit	total_credit_utilized	
Min. : 2.00	Min. : 0.0	Min. : 0	Min. : 0	
1st Qu.:14.00	1st Qu.: 7.0	1st Qu.: 51594	1st Qu.: 19186	
Median :21.00	Median :10.0	Median : 114667	Median : 36927	
Mean :22.68	Mean :11.4	Mean : 183606	Mean : 51049	
3rd Qu.:29.00	3rd Qu.:14.0	3rd Qu.: 267550	3rd Qu.: 65421	
Max. :87.00	Max. :51.0	Max. :3386034	Max. :942456	
num_collections_last_12m	num_historical_failed_to_pay	months_since_90d_late		
Min. :0.0000	Min. : 0.0000	Min. : 2.00		
1st Qu.:0.0000	1st Qu.: 0.0000	1st Qu.: 29.00		
Median :0.0000	Median : 0.0000	Median : 47.00		
Mean :0.0138	Mean : 0.1671	Mean : 46.11		
3rd Qu.:0.0000	3rd Qu.: 0.0000	3rd Qu.: 63.00		
Max. :3.0000	Max. :52.0000	Max. :128.00		
		NA's :7715		
current_accounts_delinq	total_collection_amount_ever			

Min. :0e+00	Min. : 0.0
1st Qu.:0e+00	1st Qu.: 0.0
Median :0e+00	Median : 0.0
Mean :1e-04	Mean : 184.3
3rd Qu.:0e+00	3rd Qu.: 0.0
Max. :1e+00	Max. :199308.0

current_installment_accounts accounts_opened_24m

Min. : 0.000	Min. : 0.000
1st Qu.: 1.000	1st Qu.: 2.000
Median : 2.000	Median : 4.000
Mean : 2.664	Mean : 4.376
3rd Qu.: 3.000	3rd Qu.: 6.000
Max. :35.000	Max. :29.000

months_since_last_credit_inquiry num_satisfactory_accounts

Min. : 0.000	Min. : 0.00
1st Qu.: 2.000	1st Qu.: 7.00
Median : 6.000	Median :10.00
Mean : 7.341	Mean :11.38
3rd Qu.:11.000	3rd Qu.:14.00
Max. :24.000	Max. :51.00
NA's :1271	

num_accounts_120d_past_due num_accounts_30d_past_due num_active_debit_accounts

Min. :0	Min. :0e+00	Min. : 0.000
1st Qu.:0	1st Qu.:0e+00	1st Qu.: 2.000
Median :0	Median :0e+00	Median : 3.000
Mean :0	Mean :1e-04	Mean : 3.595
3rd Qu.:0	3rd Qu.:0e+00	3rd Qu.: 5.000
Max. :0	Max. :1e+00	Max. :32.000
NA's :318		

total_debit_limit num_total_cc_accounts num_open_cc_accounts

Min. : 0	Min. : 2.00	Min. : 0.000
1st Qu.: 10000	1st Qu.: 7.00	1st Qu.: 5.000
Median : 19500	Median :11.00	Median : 7.000
Mean : 27357	Mean :13.03	Mean : 8.095
3rd Qu.: 36100	3rd Qu.:17.00	3rd Qu.:10.000
Max. :386700	Max. :66.00	Max. :46.000

num_cc_carrying_balance num_mort_accounts account_never_delinq_percent

Min. : 0.000	Min. : 0.000	Min. : 14.30
1st Qu.: 3.000	1st Qu.: 0.000	1st Qu.: 92.60
Median : 5.000	Median : 1.000	Median :100.00

Mean : 5.231	Mean : 1.383	Mean : 94.65
3rd Qu.: 7.000	3rd Qu.: 2.000	3rd Qu.:100.00
Max. :43.000	Max. :14.000	Max. :100.00

tax_liens	public_record_bankrupt	loan_purpose	application_type
Min. : 0.0000	Min. :0.0000	Length:10000	Length:10000
1st Qu.: 0.0000	1st Qu.:0.0000	Class :character	Class :character
Median : 0.0000	Median :0.0000	Mode :character	Mode :character
Mean : 0.0433	Mean :0.1238		
3rd Qu.: 0.0000	3rd Qu.:0.0000		
Max. :52.0000	Max. :3.0000		

loan_amount	term	interest_rate	installment
Min. : 1000	Min. :36.00	Min. : 5.31	Min. : 30.75
1st Qu.: 8000	1st Qu.:36.00	1st Qu.: 9.43	1st Qu.: 256.04
Median :14500	Median :36.00	Median :11.98	Median : 398.42
Mean :16362	Mean :43.27	Mean :12.43	Mean : 476.21
3rd Qu.:24000	3rd Qu.:60.00	3rd Qu.:15.05	3rd Qu.: 644.69
Max. :40000	Max. :60.00	Max. :30.94	Max. :1566.59

grade	sub_grade	issue_month	loan_status
Length:10000	Length:10000	Length:10000	Length:10000
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

initial_listing_status	disbursement_method	balance	paid_total
Length:10000	Length:10000	Min. : 0	Min. : 0.0
Class :character	Class :character	1st Qu.: 6679	1st Qu.: 928.7
Mode :character	Mode :character	Median :12380	Median : 1563.3
		Mean :14459	Mean : 2494.2
		3rd Qu.:20690	3rd Qu.: 2616.0
		Max. :40000	Max. :41630.4

paid_principal	paid_interest	paid_late_fees
Min. : 0.0	Min. : 0.0	Min. : 0.0000
1st Qu.: 587.1	1st Qu.: 221.8	1st Qu.: 0.0000
Median : 985.0	Median : 446.1	Median : 0.0000
Mean : 1894.5	Mean : 599.7	Mean : 0.1195
3rd Qu.: 1694.6	3rd Qu.: 825.4	3rd Qu.: 0.0000
Max. :40000.0	Max. :4216.4	Max. :52.9800

```
install.packages("dplyr")
```

Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
(as 'lib' is unspecified)

```
library(dplyr)
```