The goal of this project

To organize, clean and analyze for the following questions:

is the Loan ammount applied for affected by gender?

Does education affect credit history?

does education affect weekly applicant income?

If you are married do u have a higher weekly applicant income?

if u have a higher weekly applicant income do u get a higher loan ammount?

LOAN DATA PROJECT

The uncleaned dataset was taken from kaggle the link is: https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset

What the dataset is used for?

The loan approval dataset is a collection of financial records and associated information used to determine the eligibility of individuals or organizations for obtaining loans from a lending institution.

It includes various factors such as cibil score, income, employment status, loan term, loan amount, assets value, and loan status. This dataset is commonly used in machine

learning and data analysis to develop models and algorithms that predict the likelihood of loan approval based on the given features.

Problems Identified early by scoping the dataset.

we can see here a number of things have to be changed.

dependents column 3+ should be 3 and state in footer 3 means more than 3 dependents.

Nan values in self employed change that to Data not provided.

Change loan ammount times 1000 to show real ammount.

loan ammount term shows weekly so rename column.

change 0.0 in credit history to 0.

create a column for yearly, weekly and daily repayments.

missing values in gender dependents self employed loanammount loan ammount term and credit history

```
In []: df.head(20)
    # we can see here a number of things have to be changed.
    # dependents column 3+ should be 3 and state in footer 3 means more than 3 dependents.
    # Nan values in self employed change that to Data not provided.
    # Change loan ammount times 1000 to show real ammount.
    # loan ammount term shows weekly so rename column.
    # change 0.0 in credit history to 0.
# create a column for yearly, weekly and daily repayments.
```

Out[]:	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_

	_			•			• •	• • •		_
	0 LP0010	15 Male	Yes	0	Graduate	No	5720	0	110.0	
	1 LP00102	22 Male	Yes	1	Graduate	No	3076	1500	126.0	
	2 LP00103	31 Male	Yes	2	Graduate	No	5000	1800	208.0	
	3 LP00103	35 Male	Yes	2	Graduate	No	2340	2546	100.0	
	4 LP00105	51 Male	No	0	Not Graduate	No	3276	0	78.0	
	5 LP0010	54 Male	Yes	0	Not Graduate	Yes	2165	3422	152.0	
	6 LP0010	55 Female	No	1	Not Graduate	No	2226	0	59.0	
	7 LP0010	56 Male	Yes	2	Not Graduate	No	3881	0	147.0	
	8 LP0010	59 Male	Yes	2	Graduate	NaN	13633	0	280.0	
	9 LP00106	67 Male	No	0	Not Graduate	No	2400	2400	123.0	
1	0 LP00107	78 Male	No	0	Not Graduate	No	3091	0	90.0	
1	1 LP00108	32 Male	Yes	1	Graduate	NaN	2185	1516	162.0	
1	2 LP00108	33 Male	No	3+	Graduate	No	4166	0	40.0	
1	3 LP00109	94 Male	Yes	2	Graduate	NaN	12173	0	166.0	
1	4 LP00109	96 Female	No	0	Graduate	No	4666	0	124.0	
1	5 LP00109	99 Male	No	1	Graduate	No	5667	0	131.0	
1	6 LP00110	05 Male	Yes	2	Graduate	No	4583	2916	200.0	
1	7 LP00110	07 Male	Yes	3+	Graduate	No	3786	333	126.0	
1	8 LP00110	08 Male	Yes	0	Graduate	No	9226	7916	300.0	
1	9 LP0011	15 Male	No	0	Graduate	No	1300	3470	100.0	

In [174... df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 367 entries, 0 to 366
Data columns (total 12 columns):

Data	Cotumns (total 12	cocumins).	
#	Column	Non-Null Count	Dtype
0	Loan_ID	367 non-null	object
1	Gender	356 non-null	object
2	Married	367 non-null	object
3	Dependents	357 non-null	object
4	Education	367 non-null	object
5	Self_Employed	344 non-null	object
6	ApplicantIncome	367 non-null	int64
7	CoapplicantIncome	367 non-null	int64
8	LoanAmount	362 non-null	float64
9	Loan_Amount_Term	361 non-null	float64
10	Credit_History	338 non-null	float64
11	Property_Area	367 non-null	object
	67 (64/5)	64(6)	

dtypes: float64(3), int64(2), object(7)
memory usage: 34.5+ KB

In [175... #see stats for your data.
df.describe()

Out[175...

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
count	367.000000	367.000000	362.000000	361.000000	338.000000
mean	4805.599455	1569.577657	136.132597	342.537396	0.825444
std	4910.685399	2334.232099	61.366652	65.156643	0.380150
min	0.000000	0.000000	28.000000	6.000000	0.000000
25%	2864.000000	0.000000	100.250000	360.000000	1.000000
50%	3786.000000	1025.000000	125.000000	360.000000	1.000000
75%	5060.000000	2430.500000	158.000000	360.000000	1.000000
max	72529.000000	24000.000000	550.000000	480.000000	1.000000

```
Out[176... Loan ID
                     Gender
                                                                   11
                     Married
                                                                     0
                     Dependents
                                                                   10
                     Education
                                                                     0
                     Self Employed
                                                                   23
                     ApplicantIncome
                                                                     0
                     CoapplicantIncome
                                                                     0
                     LoanAmount
                                                                     5
                                                                     6
                     Loan_Amount_Term
                                                                   29
                     Credit History
                     Property_Area
                                                                     0
                     dtype: int64
In [179... # find all missing gender values and delete them
Out[179...
                                 Loan_ID Gender Married Dependents Education
                                                                                                                                Self_Employed ApplicantIncome CoapplicantIncome LoanAmount Loan
                       22 LP001128
                                                        NaN
                                                                                                               Graduate
                                                                                                                                                                                     3909
                                                                                                                                                                                                                                0
                                                                                                                                                                                                                                                   101.0
                                                                                                                        Not
                       51 LP001287
                                                                                                                                                                                     3500
                                                                                                                                                                                                                            833
                                                                                                                                                                                                                                                    120.0
                                                        NaN
                                                                           Yes
                                                                                                     3+
                                                                                                                                                       No
                                                                                                               Graduate
                     106 LP001563
                                                        NaN
                                                                                                      0
                                                                                                                                                                                     1596
                                                                                                                                                                                                                           1760
                                                                                                                                                                                                                                                   119.0
                                                                           Nο
                                                                                                               Graduate
                                                                                                                                                       Nο
                     138
                              LP001769
                                                        NaN
                                                                                                  NaN
                                                                                                               Graduate
                                                                                                                                                                                     3333
                                                                                                                                                                                                                           1250
                                                                                                                                                                                                                                                    110.0
                                                                           No
                                                                                                                                                       No
                                                                                                                        Not
                             LP002165
                                                                                                                                                                                     2038
                                                                                                                                                                                                                                                   100.0
                     209
                                                        NaN
                                                                           No
                                                                                                                                                       No
                                                                                                                                                                                                                          4027
                                                                                                               Graduate
                    231 LP002298
                                                        NaN
                                                                           No
                                                                                                      0
                                                                                                               Graduate
                                                                                                                                                      Yes
                                                                                                                                                                                     2860
                                                                                                                                                                                                                          2988
                                                                                                                                                                                                                                                    138.0
                     245 LP002355
                                                        NaN
                                                                          Yes
                                                                                                      0
                                                                                                                                                                                     3186
                                                                                                                                                                                                                          3145
                                                                                                                                                                                                                                                   150.0
                                                                                                               Graduate
                                                                                                                                                       No
                    279 LP002553
                                                        NaN
                                                                           No
                                                                                                      0
                                                                                                               Graduate
                                                                                                                                                       No
                                                                                                                                                                                   29167
                                                                                                                                                                                                                                                    185.0
                             LP002614
                                                                                                                                                                                     6478
                                                                                                                                                                                                                                0
                                                                                                                                                                                                                                                    108.0
                                                        NaN
                                                                           No
                                                                                                               Graduate
                                                                                                                                                       No
                                                                                                                        Not
                     303 LP002657
                                                        NaN
                                                                                                                                                      Yes
                                                                                                                                                                                       570
                                                                                                                                                                                                                          2125
                                                                                                                                                                                                                                                     68.0
                                                                           Yes
                                                                                                                Graduate
                                                                                                                        Not
                    318 LP002775
                                                        NaN
                                                                                                                                                       No
                                                                                                                                                                                     4768
                                                                                                                                                                                                                                0
                                                                                                                                                                                                                                                    125.0
                                                                                                               Graduate
In [180...
                   # drop rows there the name column has missing nan values.
                    df_copy = df_copy.dropna(subset=["Gender"])
In [181... df_copy[df_copy['Gender'].isnull()]
                     # there is no null values in gender.
                        Loan_ID Gender Married Dependents Education Self_Employed ApplicantIncome CoapplicantIncome LoanAmount Loan_Amount Loan_Amount
In [183... df.info()
                  <class 'pandas.core.frame.DataFrame'>
                  RangeIndex: 367 entries, 0 to 366
                  Data columns (total 12 columns):
                    #
                           Column
                                                                     Non-Null Count Dtype
                    0
                            Loan ID
                                                                      367 non-null
                                                                                                         object
                            Gender
                                                                      356 non-null
                                                                                                         object
                            Married
                                                                      367 non-null
                    2
                                                                                                         object
                            Dependents
                                                                      357 non-null
                                                                                                         object
                    4
                            Education
                                                                      367 non-null
                                                                                                         object
                    5
                            Self Employed
                                                                      344 non-null
                                                                                                         object
                            ApplicantIncome
                                                                      367 non-null
                    6
                                                                                                         int64
                            CoapplicantIncome
                                                                      367 non-null
                                                                                                         int64
                    8
                                                                      362 non-null
                                                                                                         float64
                            LoanAmount
                    9
                            Loan Amount Term
                                                                                                          float64
                                                                      361 non-null
                    10
                          Credit_History
                                                                      338 non-null
                                                                                                         float64
                    11 Property_Area
                                                                      367 non-null
                                                                                                         object
                  dtypes: float64(3), int64(2), object(7)
                  memory usage: 34.5+ KB
```

Solutions to clean:

In [176... # check for missing values in all columns

next is dependents

change nan info not provided

dependents column 3+ should be 3 and state in footer 3 means more than 3 dependents.

change dependents to 0 for Nan in dependents. we assume its 0 because no information is given.

now u have changed this change the data type for dependents to numerical so u can analyze this

now we deal with loan ammount drop all the missing entries for loan ammount because the accountant has screwed up obviously and the application needs to be redone.

remove all entries with missing values in loan ammount term as customer will have to redo application

remove all missing values for loan ammount as application will have to be redone.

now we only have missing values for credit history.

change all missing values to 0.

there are some values in credit history with 0.0

replace all instances of 0.0 in credit history with 0.

rename loan ammount term column to weekly loan ammount term

drop the .0 in Loan ammount term by switching to integer.

credit history should also be one integer

applicant income is per week as well as coapplicant income. so rename columns

need to times the loan ammount by a thousand for each loan ammount in whole dataset.

pick the loan ammount column and change to times 1000 answer.

change loan ammount to integer

CREATE 4 NEW COLUMNS WITH DATA CALCULATIONS FOR YEARLY MONTHLY WEEKLY AND DAILY REPAYMENTS FOR THE LOANS.

create a new copy of the changed dataframe, then do a IQR analysis to deal with the error.

after this do graphs to answer the questions.

```
In [184… # next is dependents
                     # change nan info not provided
                     # dependents column 3+ should be 3 and state in footer 3 means more than 3 dependents.
                     #change dependents to 0 for Nan in dependents. we assume its 0 because no information is given.
                     \# now u have changed this change the data type for dependents to numerical so u can analyze this .
In [185... # next is delf employed
                     # change all NAn in self employed to no as no data is provided meaning that the customer didnt feel the need to
In [187…] # now we deal with loan ammount drop all the missing entries for loan ammount because the accountant has screwed
                     # remove all entries with missing values in loan ammount term as customer will have to redo application
In [188… # there are missing values in loan ammount term, in the same vein the application will have to be redone we need
In [189... df copy.isnull().sum()
Out[189... Loan ID
                                                                          0
                      Gender
                                                                          0
                      Married
                                                                         0
                      Dependents
                                                                         0
                                                                         0
                      Education
                      Self Employed
                                                                         0
                      ApplicantIncome
                                                                         0
                      CoapplicantIncome
                                                                         0
                      LoanAmount
                                                                         0
                      Loan Amount Term
                                                                         0
                      Credit History
                                                                       28
                      Property Area
                                                                         0
                     dtype: int64
In [190… # now we only have missing values for credit history.
                     \# change all missing values to 0.
                     # there are some values in credit history with 0.0
                     # replace all instances of 0.0 in credit history with 0.
                   \verb|C:\Users\setminus Ser\AppData \setminus Coal\Temp\ipykernel\_19124 \setminus 2315742299.py: 3: Future Warning: Setting an item of incompatible incompatible in the property of the 
                   dtype is deprecated and will raise an error in a future version of pandas. Value '0' has dtype incompatible with
```

In [191… # dealt with all missing values and data cleaning lets now look at the data.

float64, please explicitly cast to a compatible dtype first.

df copy.loc[df copy['Credit History'].isnull(), 'Credit History'] = '0'

0 Out[191... Loan_ID Gender 0 Married 0 Dependents 0 Education Self Employed 0 ApplicantIncome 0 CoapplicantIncome LoanAmount 0 Loan Amount Term 0 Credit_History 0 Property_Area dtype: int64

In [194... # rename loan ammount term column to weekly loan ammount term

drop the .0 in Loan ammount term by switching to integer.

credit history should also be one integer

applicant income is per week as well as coapplicant income. so rename columns

need to times the loan ammount by a thousand for each loan ammount in whole dataset.

pick the loan ammount column and change to times 1000 answer.

change loan ammount to integer

In [195... df copy.head(10)

columns have now been renamed see below.

Out[1

195		Loan_ID	Gender	Married	Dependents	Education	Self_Employed	Weekly_Applicant_Income	Weekly_Coapplicant_Income	Lo
	0	LP001015	Male	Yes	0	Graduate	No	5720	0	
	1	LP001022	Male	Yes	1	Graduate	No	3076	1500	
	2	LP001031	Male	Yes	2	Graduate	No	5000	1800	
	3	LP001035	Male	Yes	2	Graduate	No	2340	2546	
	4	LP001051	Male	No	0	Not Graduate	No	3276	0	
	5	LP001054	Male	Yes	0	Not Graduate	Yes	2165	3422	
	6	LP001055	Female	No	1	Not Graduate	No	2226	0	
	7	LP001056	Male	Yes	2	Not Graduate	No	3881	0	
	8	LP001059	Male	Yes	2	Graduate	No	13633	0	
	9	LP001067	Male	No	0	Not Graduate	No	2400	2400	

CREATE 4 NEW COLUMNS WITH DATA CALCULATIONS FOR YEARLY MONTHLY WEEKLY AND DAILY REPAYMENTS FOR THE LOANS.

In [197...

```
Loan_ID Gender
                                Married Dependents
                                                     Education Self_Employed Weekly_Applicant_Income Weekly_Coapplicant_Income Loa
          0 I P001015
                                                                                                                                0
                          Male
                                                      Graduate
                                                                          Nο
                                                                                                  5720
                                   Yes
             LP001022
                                                      Graduate
                                                                                                  3076
                                                                                                                             1500
                          Male
                                   Yes
                                                                          No
            LP001031
                                                  2
                                                      Graduate
                                                                          No
                                                                                                  5000
                                                                                                                             1800
                          Male
                                   Yes
                                                  2
                                                                                                  2340
             LP001035
                          Male
                                   Yes
                                                      Graduate
                                                                          No
                                                                                                                             2546
                                                           Not
             LP001051
                                    No
                                                  0
                                                                          No
                                                                                                  3276
                                                                                                                                0
                                                      Graduate
                                                           Not
            LP001054
                          Male
                                   Yes
                                                                          Yes
                                                                                                  2165
                                                                                                                             3422
                                                      Graduate
          6 LP001055
                        Female
                                                                                                  2226
                                                                                                                                0
                                    No
                                                  1
                                                                          Nο
                                                      Graduate
             LP001056
                                                                                                                                0
                                                  2
                                                                                                  3881
                          Male
                                   Yes
                                                                          No
                                                      Graduate
                                                                                                                                0
          8 LP001059
                          Male
                                   Yes
                                                  2
                                                      Graduate
                                                                          Nο
                                                                                                 13633
                                                           Not
             LP001067
                                    No
                                                                                                  2400
                                                                                                                             2400
                                                                           No
                                                      Graduate
In [198...
          # Modify the copy (original DataFrame remains unchanged)
          # Display the original and copied DataFrame
          # print("Original DataFrame:")
          # print(df)
          # print("\nCopied DataFrame (formatted):")
          # print(df_copy)
In [199... df_copy.head(5)
                                                     {\bf Education} \quad {\bf Self\_Employed} \quad {\bf Weekly\_Applicant\_Income} \quad {\bf Weekly\_Coapplicant\_Income}
              Loan_ID Gender
                                Married Dependents
                                                                                                                                  Loa
          0 LP001015
                                                      Graduate
                                                                                                  5720
                                                                                                                                0
                          Male
                                   Yes
                                                  0
                                                                          No
             LP001022
                          Male
                                   Yes
                                                      Graduate
                                                                          No
                                                                                                  3076
                                                                                                                             1500
          2 LP001031
                          Male
                                   Yes
                                                  2
                                                      Graduate
                                                                          No
                                                                                                  5000
                                                                                                                             1800
                                                  2
             LP001035
                                                                                                  2340
                                                                                                                             2546
                          Male
                                   Yes
                                                      Graduate
                                                                          No
                                                           Not
          4 LP001051
                          Male
                                    No
                                                  0
                                                                          No
                                                                                                  3276
                                                                                                                                0
                                                      Graduate
         # rename monthly column to Monthly repayments
In [201… # Rounded all repayment columns to the nearest 0.10
          # the data is clean now. Lets account for outliers.
In [202... df_copy.head(5)
              Loan_ID Gender Married Dependents
                                                     Education Self_Employed Weekly_Applicant_Income Weekly_Coapplicant_Income
                                                                                                                                   Loa
          0 LP001015
                                                                                                                                0
                          Male
                                   Yes
                                                      Graduate
                                                                          No
                                                                                                  5720
             LP001022
                          Male
                                   Yes
                                                  1
                                                      Graduate
                                                                           No
                                                                                                  3076
                                                                                                                             1500
          2 LP001031
                                                  2
                                                      Graduate
                                                                                                  5000
                                                                                                                             1800
                          Male
                                   Yes
                                                                          Nο
                                                      Graduate
                                                                                                  2340
          3 LP001035
                          Male
                                   Yes
                                                  2
                                                                          Nο
                                                                                                                             2546
                                                           Not
          4 LP001051
                                                  0
                                                                                                  3276
                                                                                                                                0
                          Male
                                    No
                                                                          No
                                                      Graduate
In [203… # As we are not testing hypothesis we can account for outliers now
          # first remove outliers from only numerical columns using Interquartile range
          # Step 1: Read data from a CSV file
          # Replace 'your file.csv' with the path to your CSV file
          # file path = 'data-1.csv'
          # Read the CSV file into a DataFrame
          # df = pd.read csv('data-1.csv')
```

```
# Step 2: Specify the columns to analyze
 # Replace with the list of columns you want to process
 # Ensure the specified columns exist in the DataFrame
 # Step 3: Remove outliers from each specified column
  # Create a copy of the original DataFrame to store filtered data
     # Calculate IQR for the column
     # Define the lower and upper bounds for outliers
     # Filter out the outliers for the column
 # Step 4: Save the filtered DataFrame (optional)
 # Uncomment the following line to save the filtered data to a new CSV file
 # df filtered.to csv('filtered data.csv', index=False)
 # Print the original and filtered DataFrames
 # Print the number of rows removed
Original DataFrame:
     Loan ID Gender Married Dependents
                                            Education Self_Employed \
    LP001015
              Male
                     Yes 0
                                            Graduate
1
    LP001022
               Male
                        Yes
                                      1
                                             Graduate
                                                                No
2
    LP001031
               Male
                        Yes
                                      2
                                             Graduate
                                                                No
                                     2
3
    LP001035
               Male
                        Yes
                                            Graduate
                                                                No
                                    0 Not Graduate
    LP001051
               Male
                        No
                                                                No
                . . .
                        . . .
                                    3 Not Graduate
362 LP002971
               Male
                        Yes
                                                               Yes
363 LP002975
               Male
                        Yes
                                     0
                                         Graduate
                                                                Nο
364 LP002980
               Male
                        No
                                     0
                                            Graduate
365 LP002986
                                            Graduate
              Male
                        Yes
                                      0
                                                                Nο
366 LP002989
              Male
                        No
                                      0
                                             Graduate
                                                               Yes
    Weekly_Applicant_Income Weekly_Coapplicant_Income LoanAmount \
0
                                                            110000
                       5720
                                                    0
1
                       3076
                                                  1500
                                                            126000
2
                                                  1800
                                                            208000
                       5000
                       2340
                                                  2546
                                                            100000
4
                                                            78000
                       3276
                                                    0
                       4009
                                                  1777
                                                            113000
362
                                                   709
363
                       4158
                                                            115000
                                                  1993
                                                            126000
                       3250
364
365
                       5000
                                                  2393
                                                            158000
                       9200
                                                             98000
366
    Weekly_Loan_Ammount_Term Credit_History Property_Area \
0
                                           1
1
                         360
                                           1
                                                    Urban
2
                         360
                                           1
                                                    Urban
3
                         360
                                           0
                                                     Urban
4
                         360
                                          1
                                                     Urban
                                         . . .
362
                         360
                                         1
                                                     Urban
                         360
                                                     Urban
363
                                           1
364
                         360
                                           0
                                                 Semiurban
365
                         360
                                          1
                                                    Rural
366
                         180
                                                    Rural
```

Monthly

15888.888889 1324.074074

18200.000000 1516.666667 30044.444444 2503.703704

14444.444444 1203.703704 11266.666667 938.888889

16322.22222 1360.185185 16611.111111 1384.259259

18200.000000 1516.666667

22822.22222 1901.851852

28311.111111 2359.259259

Yearly Loan Ammount Term Yearly repayments

6.923077

6.923077

6.923077

6.923077

6.923077

6.923077

6.923077

6.923077

6.923077

3.461538

0

1

2

3

4

362

363

364

365

366

```
0
            305.55556
                                 43.650794
1
            350.000000
                                 50.000000
2
            577.77778
                                 82.539683
3
            277.777778
                                 39.682540
4
            216.666667
                                 30.952381
                                 44.841270
            313.888889
362
363
            319.444444
                                 45.634921
            350.000000
                                 50.000000
364
             438.888889
                                 62.698413
365
366
            544.44444
                                 77.77778
[345 rows x 17 columns]
Filtered DataFrame (without outliers):
      Loan ID Gender Married Dependents
                                                Education Self Employed \
0
     LP001015
                Male
                          Yes
                                         0
                                                 Graduate
                                                                      Nο
1
     LP001022
                Male
                          Yes
                                         1
                                                 Graduate
                                                                      No
                                                 Graduate
2
                Male
     LP001031
                          Yes
                                         2
                                                                      No
     LP001035
                Male
                          Yes
                                         2
                                                 Graduate
4
     LP001051
                                            Not Graduate
                Male
                           No
                                         0
                                                                      No
                          . . .
                                                                      . . .
     LP002969
                          Yes
361
                Male
                                                 Graduate
                                         1
                                                                      No
     LP002971
                Male
                                            Not Graduate
362
                          Yes
                                         3
                                                                     Yes
     LP002975
                Male
                                                 Graduate
363
                          Yes
                                         0
                                                                      No
364
     LP002980
                Male
                           No
                                         0
                                                 Graduate
                                                                      No
    LP002986
365
                Male
                          Yes
                                         0
                                                 Graduate
                                                                      No
     Weekly\_Applicant\_Income \ \ Weekly\_Coapplicant\_Income \ \ LoanAmount \ \ \backslash \\
0
1
                         3076
                                                      1500
                                                                 126000
2
                         5000
                                                      1800
                                                                 208000
3
                         2340
                                                      2546
                                                                 100000
4
                         3276
                                                         0
                                                                  78000
361
                         2269
                                                      2167
                                                                  99000
                                                                 113000
362
                         4009
                                                      1777
363
                         4158
                                                       709
                                                                 115000
364
                         3250
                                                      1993
                                                                 126000
365
                         5000
                                                      2393
                                                                 158000
     Weekly Loan Ammount Term
                                 Credit History Property Area
0
                           360
                                                         Urban
                                              1
1
                           360
                                               1
                                                         Urban
2
                           360
                                              1
                                                         Urban
3
                           360
                                               0
                                                         Urban
4
                           360
                                                         Urban
                                               1
                            . . .
361
                           360
                                              1
                                                     Semiurban
362
                           360
                                               1
                                                         Urban
363
                           360
                                                         Urban
                                               1
364
                           360
                                               0
                                                     Semiurban
365
                           360
                                                         Rural
     Yearly_Loan_Ammount_Term
                                 Yearly repayments
                                                         Monthly
0
                                      15888.888889 1324.074074
                      6.923077
1
                      6.923077
                                      18200.000000
                                                     1516.666667
2
                      6.923077
                                      30044.44444
                                                     2503.703704
3
                      6.923077
                                      14444.44444
                                                     1203.703704
4
                      6.923077
                                      11266.666667
                                                      938.888889
                      6.923077
                                      14300.000000 1191.666667
361
                                      16322.22222 1360.185185
362
                      6.923077
                                      16611.111111 1384.259259
363
                      6.923077
                                      18200.000000
364
                      6.923077
                                                     1516.666667
365
                      6.923077
                                      22822.22222 1901.851852
     Weekly_repayments Daily_repayments
0
            305.55556
                                 43.650794
            350.000000
                                 50.000000
1
2
            577.77778
                                 82.539683
                                 39.682540
3
            277.77778
4
                                 30.952381
            216.666667
361
            275.000000
                                 39.285714
            313.888889
                                 44.841270
362
363
            319.444444
                                 45.634921
364
            350.000000
                                 50.000000
365
            438.888889
                                 62.698413
[311 rows x 17 columns]
```

Weekly_repayments Daily_repayments

Tn [284	df filtered	

Loan_ID Gender Married Dependent Education Self_Employed Weekly_Applicant_Income Weekly_Coapplicant_Income		_filtered							
1 LP001022 Male Yes 1 Graduate No 3076 1500 2 LP001031 Male Yes 2 Graduate No 5000 1800 3 LP001035 Male Yes 2 Graduate No 2340 2546 4 LP001051 Male No 0 Graduate No 3276 0 4 LP001051 Male Yes 1 Graduate No 3276 0 361 LP00296 Male Yes 1 Graduate No 2289 2167 362 LP002971 Male Yes 3 Graduate No 2289 2167 362 LP002971 Male Yes 0 Graduate No 3250 1993 363 LP00296 Male No 0 Graduate No 3250 1993 364 LP002980 Male No 0 Graduate No 3250 1993 365 LP002980 Male No 0 Graduate No 3250 1993 367 LP002980 Male No 0 Graduate No 3250 1993 368 LP002980 Male No 0 Graduate No 5000 2393 311 rows × 17 columns # Rounded atl repayment columns to the nearest 8.18 df_filtered.head(5) Loan_ID Gender Married Dependents Education Self_Employed Weekly_Applicant_Income Weekly_Coapplicant_Income Lo Columns to Male Yes 2 Graduate No 3076 1500 2 LP001035 Male Yes 1 Graduate No 3076 1500 2 LP001035 Male Yes 2 Graduate No 3076 1500 2 LP001031 Male Yes 2 Graduate No 3076 1500 3 LP001035 Male Yes 2 Graduate No 3076 0 1500 4 LP001035 Male Yes 2 Graduate No 3076 0 1500 4 LP001031 Male Yes 2 Graduate No 3076 0 1500 5 LP001031 Male Yes 2 Graduate No 3076 0 1500 5 LP001031 Male Yes 2 Graduate No 3076 0 1500 4 LP001051 Male No 0 No 0 No 0 No 0 0 No No 0 No 0 No 0 No No 0 No 0 No No 0 No 0 No No No No No 0 No No No No No 0 No		Loan_ID	Gender	Married	Dependents	Education	Self_Employed	Weekly_Applicant_Income	Weekly_Coapplicant_Income
2 LP001031 Male Yes 2 Graduate No 5000 1800 3 LP001035 Male Yes 2 Graduate No 2340 2546 4 LP001051 Male No 0 Rotal No 3276 0 4 LP001051 Male No 0 Rotal No 3276 0 581 LP002969 Male Yes 1 Graduate No 2269 2167 582 LP002971 Male Yes 3 Rotal Yes 4009 1777 583 LP002975 Male Yes 0 Graduate No 3250 1993 585 LP002980 Male No 0 Graduate No 3250 1993 585 LP002980 Male Yes 0 Graduate No 3250 1993 585 LP002980 Male Yes 0 Graduate No 5000 2333 511 rows x 17 columns 51 Adjust for changes that didnt reflect in df filtered. # rename monthly column to Nonthly_repayments # Rounded all repayment columns to the nearest 0.10 6 LF001015 Male Yes 0 Graduate No 5720 0 6 LP001016 Male Yes 0 Graduate No 5720 0 6 LP001016 Male Yes 0 Graduate No 5720 0 6 LP001016 Male Yes 0 Graduate No 5720 0 7 LP001017 Male Yes 0 Graduate No 5720 0 8 LP001031 Male Yes 0 Graduate No 5720 0 8 LP001031 Male Yes 0 Graduate No 5720 0 9 LP001031 Male Yes 0 Graduate No 5720 0 1 LP001031 Male Yes 0 Graduate No 5720 0 1 LP001031 Male Yes 0 Graduate No 5720 0 1 LP001031 Male Yes 2 Graduate No 5720 0 1 LP001031 Male Yes 2 Graduate No 5720 0 1 LP001031 Male Yes 2 Graduate No 5720 0 2 LP001031 Male Yes 2 Graduate No 5720 0 3 LP001035 Male Yes 2 Graduate No 5720 0 4 LP001031 Male Yes 2 Graduate No 5720 0 5 LP001031 Male Yes 2 Graduate No 5720 0 6 LP001031 Male Yes 2 Graduate No 5720 0 8 LP001031 Male Yes 2 Graduate No 5720 0 8 LP001031 Male Yes 2 Graduate No 5720 0 8 LP001031 Male Yes 2 Graduate No 5720 0 8 LP001031 Male Yes 2 Graduate No 5720 0 8 LP001031 Male Yes 2 Graduate No 5720 0 8 LP001031 Male Yes 2 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 5720 0 8 LP001031 Male Yes 3 Graduate No 57		0 LP001015	Male	Yes	0	Graduate	No	5720	0
3 LP001035 Male Yes 2 Graduate No 2340 2546 4 LP001051 Male No 0 Graduate No 3276 0 364 LP002990 Male Yes 1 Graduate No 2299 2167 365 LP002997 Male Yes 0 Graduate No 4158 709 365 LP002998 Male No 0 Graduate No 4158 709 365 LP002998 Male No 0 Graduate No 3250 1993 365 LP002998 Male No 0 Graduate No 3250 1993 367 LP002998 Male No 0 Graduate No 5000 2393 311 rows x 17 columns # adjust for changes that didnt reflect in df filtered. # rename monthly column to Monthly repsyments # Rounded all repsyment columns to the nearest θ.10 df_filtered.head(S) Loan_ID Gender Married Dependents Education Self_Employed Weekly_Applicant_Income Meekly_Ceapplicant_Income Loa 1 Graduate No 3078 1500 LP001015 Male Yes 0 Graduate No 3078 1500 LP001035 Male Yes 2 Graduate No 3078 1500 3 LP001035 Male Yes 2 Graduate No 3078 1500 3 LP001035 Male Yes 2 Graduate No 3078 0 1600 3 LP001035 Male Yes 2 Graduate No 3078 0 1600 4 LP001051 Male No 0 Round No 3240 2546 4 LP001051 Male No 0 Round No 3256 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0		1 LP001022	Male	Yes	1	Graduate	No	3076	1500
## LP001051 Male No 0 0 Graduate No 3276 0 0 0 0 0 0 0 0 0		2 LP001031	Male	Yes	2	Graduate	No	5000	1800
### LP002969 Male Yes 1 Graduate No 2289 2167 362 LP002971 Male Yes 3 Graduate No 2289 2167 363 LP002975 Male Yes 0 Graduate No 4159 709 364 LP002980 Male No 0 Graduate No 3250 1993 365 LP002980 Male No 0 Graduate No 5000 2383 365 LP002988 Male Yes 0 Graduate No 5000 2383 311 rows × 17 columns ###################################		3 LP001035	Male	Yes	2	Graduate	No	2340	2546
381 LP002969 Male Yes		4 LP001051	Male	No	0		No	3276	0
382 LP002971 Maile Yes 3 Signaturate No 4158 709									
363 LP002975 Male Yes 0 Graduate No 4158 709	36	1 LP002969	Male	Yes	1	Graduate	No	2269	2167
364 LP002980 Male No 0 Graduate No 3250 1993 365 LP002986 Male Yes 0 Graduate No 5000 2393 311 rows x 17 columns 311 rows x 17 columns 311 rows x 17 columns 312 rows x 17 columns 313 rows x 17 columns 314 rows x 17 columns 315 LP001015 Male Yes 0 Graduate No 6000 Weekly_Applicant_Income Meckly_repayments 4 Rounded all repayment columns to the nearest 0.10 31 LP001015 Male Yes 0 Graduate No 5720 0 32 LP001031 Male Yes 1 Graduate No 5720 0 33 LP001035 Male Yes 1 Graduate No 5720 0 4 LP001051 Male Yes 2 Graduate No 5000 1800 4 LP001051 Male No 0 Graduate No 5000 0 5 Craduate No 5000 0 6 Craduate No	36	2 LP002971	Male	Yes	3		Yes	4009	1777
365 LP002986 Male Yes 0 Graduate No 5000 2393 311 rows × 17 columns # adjust for changes that didnt reflect in df filtered. # rename monthly column to Monthly_repayments # Rounded all repayment columns to the nearest 0.10 df_filtered.head(5) Loan_ID Gender Married Dependents Education Self_Employed Weekly_Applicant_Income Weekly_Coapplicant_Income Loa	36	3 LP002975	Male	Yes	0	Graduate	No	4158	709
# adjust for changes that didnt reflect in df filtered. # rename monthly column to Monthly repayments # Rounded all repayment columns to the nearest 0.10 df_filtered.head(5) Loan_ID Gender Married Dependents Education Self_Employed Weekly_Applicant_Income Veekly_Coapplicant_Income Lox 0 LP001015 Male Yes 0 Graduate No 3076 1500 1 LP001022 Male Yes 1 Graduate No 3076 1500 2 LP001031 Male Yes 2 Graduate No 5000 1800 3 LP001035 Male Yes 2 Graduate No 3240 2546 4 LP001051 Male No 0 Not Graduate No 3276 0 df_filtered.isnutl().sum() Loan ID 0 Gender 0 Married 0 Dependents 0 Monthly repayments 0 Deployements 0	36	4 LP002980	Male	No	0	Graduate	No	3250	1993
Rounded all repayment columns to the nearest 0.10 Rounded all repayment columns to the nearest 0.10 Coan_ID Gender Married Dependents Education Self_Employed Weekly_Applicant_Income Weekly_Coapplicant_Income Loa Dependents Depe	36	5 LP002986	Male	Yes	0	Graduate	No	5000	2393
# adjust for changes that didnt reflect in df filtered. # rename monthly column to Monthly_repayments # Rounded all repayment columns to the nearest 0.10 df_filtered.head(5) Loan_ID Gender Married Dependents Education Self_Employed Weekly_Applicant_Income Weekly_Coapplicant_Income Loa 0 LP001015 Male Yes 0 Graduate No 5720 0 1 LP001022 Male Yes 1 Graduate No 5000 1800 2 LP001031 Male Yes 2 Graduate No 5000 1800 3 LP001035 Male Yes 2 Graduate No 3376 0 4 LP001051 Male No 0 Roraduate No 3376 0 6 Graduate No 3376 0 Complete No 3476 0	311	I rows × 17 cc	lumns						
# Rounded all repayment columns to the nearest 0.10 Coan_ID Gender Married Dependents Education Self_Employed Weekly_Applicant_Income Loa	4								
D LP001015 Male Yes 0 Graduate No 5720 0 1 LP001022 Male Yes 1 Graduate No 3076 1500 2 LP001031 Male Yes 2 Graduate No 5000 1800 3 LP001035 Male Yes 2 Graduate No 2340 2546 4 LP001051 Male No 0 Not Graduate No 3276 0 4 LP001051 Male No 0 Not Graduate No 3276 0 5 Gender 0 Gend	df			larried D	ependents E	Education S	Self Employed W	Veekly Applicant Income V	Veekly Coapplicant Income L
1 LP001022 Male Yes 1 Graduate No 3076 1500 2 LP001031 Male Yes 2 Graduate No 5000 1800 3 LP001035 Male Yes 2 Graduate No 2340 2546 4 LP001051 Male No 0 Not Graduate No 3276 0 df_filtered.isnull().sum() Loan_ID									
3 LP001035 Male Yes 2 Graduate No 2340 2546 4 LP001051 Male No 0 Not Graduate No 3276 0 df_filtered.isnull().sum() Loan_ID 0 Gender 0 Married 0 Dependents 0 Education 0 Self_Employed 0 Weekly_Applicant_Income 0 Weekly_Applicant_Income 0 Weekly_Coapplicant_Income 0 Weekly_Loan_Ammount Term 0 Credit_History 0 Property_Area 0 Yearly_repayments 0 Monthly_repayments 0 Meekly_repayments 0 Meekly_repayments 0 Daily_repayments 0 dtype: int64	1	LP001022	Male	Yes	1	Graduate	No	3076	1500
4 LP001051 Male No 0 Not Graduate No 3276 0 df_filtered.isnull().sum() Loan_ID 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	2	LP001031	Male	Yes	2	Graduate	No	5000	1800
df_filtered.isnull().sum() Loan_ID	3	LP001035	Male	Yes	2	Graduate	No	2340	2546
df_filtered.isnull().sum() Loan_ID	4	LP001051	Male	No	0		No	3276	0
Loan_ID 0 Gender 0 Married 0 Dependents 0 Education 0 Self_Employed 0 Weekly_Applicant_Income 0 Weekly_Coapplicant_Income 0 LoanAmount 0 Weekly_Loan_Ammount_Term 0 Credit_History 0 Property_Area 0 Yearly_Loan_Ammount_Term 0 Yearly_repayments 0 Monthly_repayments 0 Weekly_repayments 0 Moekly_repayments 0 Motype: int64	4								
Gender 0 Married 0 Dependents 0 Education 0 Self_Employed 0 Weekly_Applicant_Income 0 Weekly_Coapplicant_Income 0 LoanAmount 0 Weekly_Loan_Ammount_Term 0 Credit_History 0 Property_Area 0 Yearly_Loan_Ammount_Term 0 Yearly_repayments 0 Monthly_repayments 0 Weekly_repayments 0 Daily_repayments 0 dtype: int64	df	_filtered.i	.snull()	sum()					
df_filtered.shape	Ge				0				
	Dec Se We We Cr Pr Ye Mc We Da	arried ependents ducation elf_Employed eekly_CoapplonAmount eekly_Loan_/ redit_Histor early_Loan_/ early_repayr enthly_repayr aily_repaymently	cant_Inc licant_I Ammount_ ry a Ammount_ ments ments ments	ncome Term	0 0 0 0 0 0 0 0 0 0				

Out[208... (311, 17)

In [209... df_filtered.info()

```
311 non-null object
        1
            Gender
        2
            Married
                                       311 non-null
                                                      object
                                       311 non-null int32
           Dependents
        4 Education
                                       311 non-null object
                                       311 non-null
            Self Employed
                                                      object
        6
                                       311 non-null
            Weekly_Applicant_Income
                                                       int64
            Weekly_Coapplicant_Income 311 non-null
        7
                                                     int64
                                       311 non-null int32
        8
            LoanAmount
        9 Weekly_Loan_Ammount_Term10 Credit_History
                                       311 non-null
                                                      int32
                                                     int32
                                       311 non-null
                                       311 non-null object
        11 Property Area
        12 Yearly_Loan_Ammount_Term 311 non-null
                                                      float64
        13 Yearly_repayments
                                       311 non-null
                                                       float64
        14 Monthly_repayments
                                       311 non-null
                                                       float64
                                      311 non-null
        15 Weekly_repayments
                                                       float64
                                      311 non-null
        16 Daily_repayments
                                                       float64
        dtypes: float64(5), int32(4), int64(2), object(6)
        memory usage: 38.9+ KB
In [210… # 302 is bigger than 30 therefore we pay attention to z test.
         # if the sample size was smaller than 30 we pay attention to the t test.
             # df = pd.DataFrame(df_filtered)
In [211...
             Perform a z-test for two independent samples.
             # Pooled standard error
             # Z-statistic
             # P-value (two-tailed)
         def perform tests(df filtered, LoanAmount, Weekly Applicant Income, alpha=0.05):
             Perform t-test and z-test on two columns from a CSV file.
             # Load data from CSV
             # df = pd.DataFrame(df_filtered)
             # Extract the two columns
             # Perform t-test
             # Perform z-test
             # Print results
       T-Test Results:
        T-statistic: 49.76178401070807
        P-value: 1.0884762280372716e-218
        Reject the null hypothesis (significant difference between groups).
        Z-Test Results:
        Z-statistic: 49.76178401070804
        P-value: 0.0
        Reject the null hypothesis (significant difference between groups).
In [212... df_filtered
```

<class 'pandas.core.frame.DataFrame'>

Non-Null Count Dtype

311 non-null object

Index: 311 entries, 0 to 365
Data columns (total 17 columns):

#

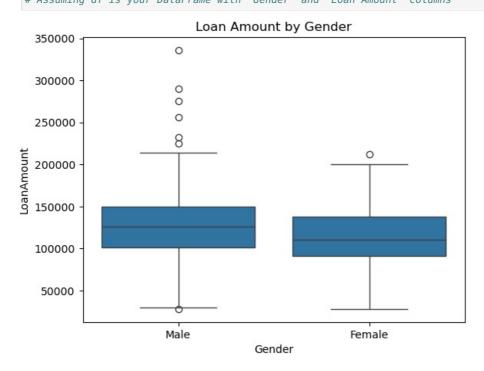
Column

0 Loan ID

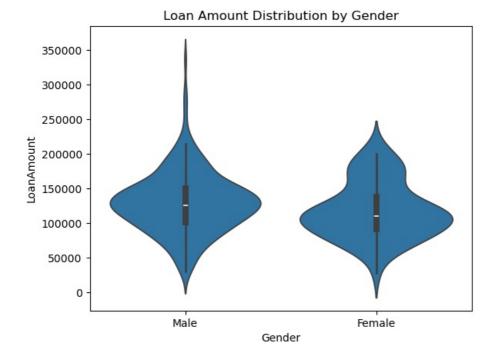
0.00 [212		Loan ID	Gondor	Married	Donandanta	Education	Salf Employed	Wookly Applicant Income	Wookly Coannigant Income	
Out[212		LOAII_ID	Gender	Warrieu	Dependents	Education	Sell_Elliployed	weekiy_Applicant_income	Weekly_Coapplicant_Income	_
	0	LP001015	Male	Yes	0	Graduate	No	5720	0	
	1	LP001022	Male	Yes	1	Graduate	No	3076	1500	
	2	LP001031	Male	Yes	2	Graduate	No	5000	1800	
	3	LP001035	Male	Yes	2	Graduate	No	2340	2546	
	4	LP001051	Male	No	0	Not Graduate	No	3276	0	
	361	LP002969	Male	Yes	1	Graduate	No	2269	2167	
	362	LP002971	Male	Yes	3	Not Graduate	Yes	4009	1777	
	363	LP002975	Male	Yes	0	Graduate	No	4158	709	
	364	LP002980	Male	No	0	Graduate	No	3250	1993	
	365	LP002986	Male	Yes	0	Graduate	No	5000	2393	

311 rows × 17 columns

In [213... # now for charts.
Assuming df is your DataFrame with 'Gender' and 'Loan Amount' columns

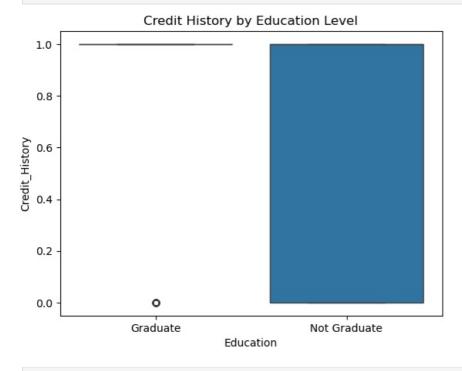


In [214...



In [215... # change credit history from object to numerical

In [216...

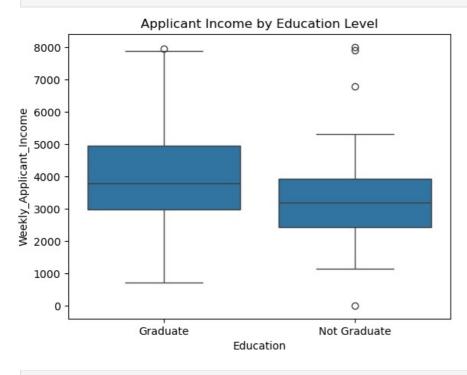


In [217...

1.25 - 1.00 - 0.75 - 0.50 - 0.00 - 0.25 - 0.00 - 0.25 - Graduate Not Graduate

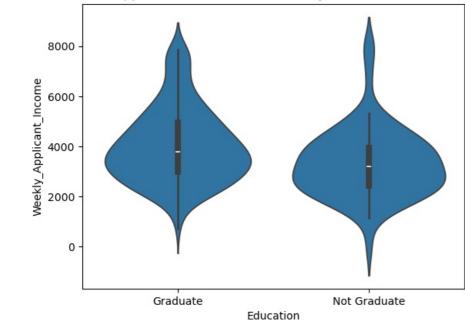
Education

In [218...

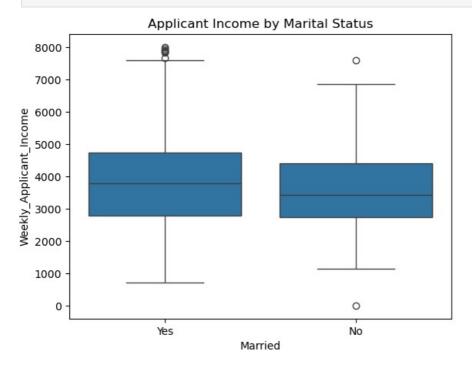


In [219...

Applicant Income Distribution by Education Level

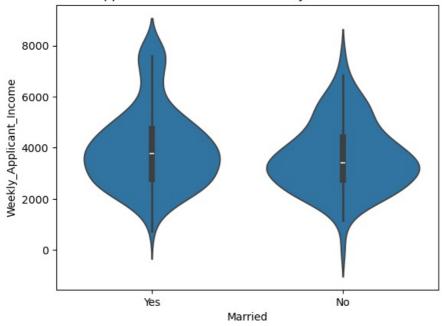


In [220...

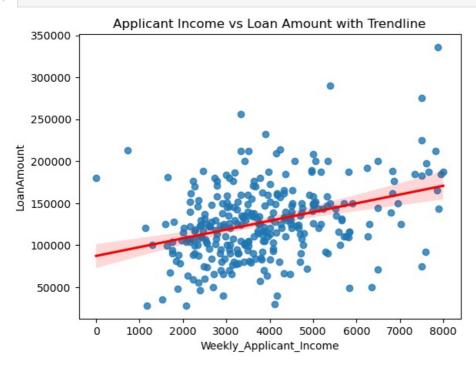


In [222...

Applicant Income Distribution by Marital Status



In []:



In [224... # can only do regression and pairplot with numerical variables.

you can see a number of trends here.

Applicant Income

applicant Income has a slightly positive relationship with Loan ammount given (the more income u have the bet # customers with a 1 credit history have a higher applicant income.

LOAN AMMOUNT GIVEN

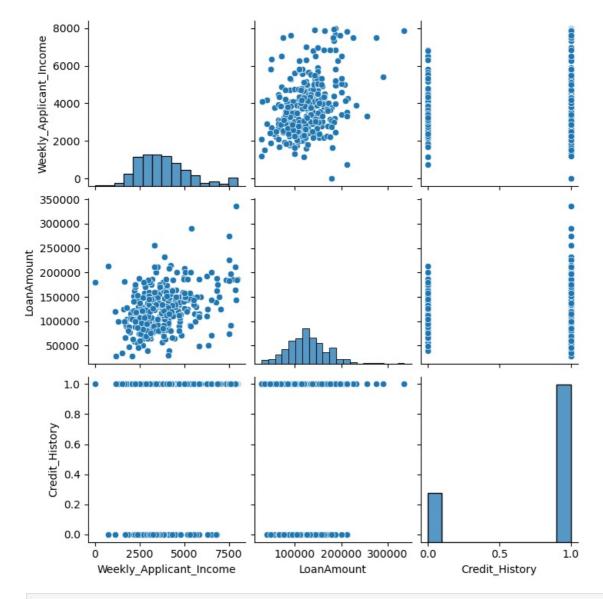
Loan ammount given doesnt have a relationship with applicant income. As it depends on applicant income.

credit history affects your loan ammount from the bank

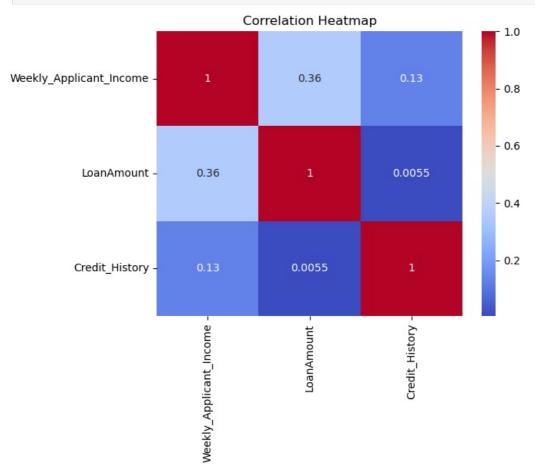
CREDIT HISTORY.

the better your credit history the higher the applicant income.

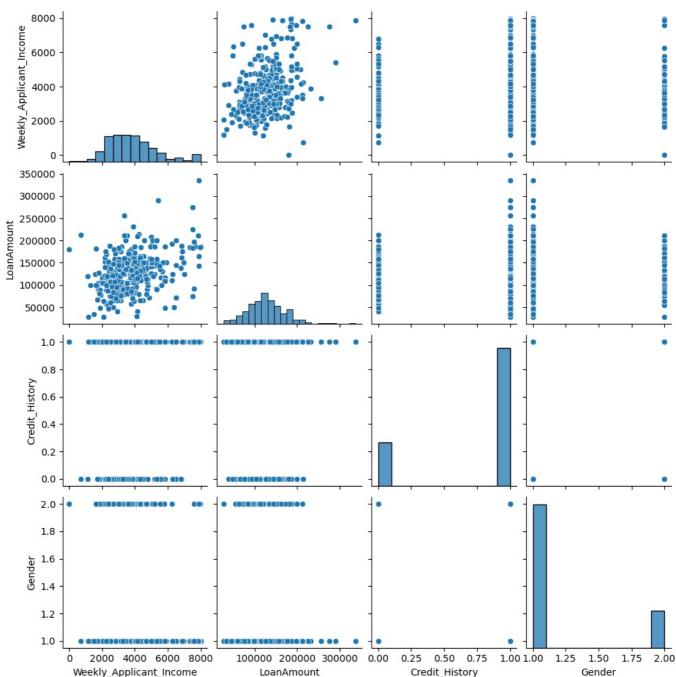
the better the credit history the higher the loan ammount given.



In [225...



Out[227... <function matplotlib.pyplot.show(close=None, block=None)>



ANSWER THE Original Questions of this project

is the Loan ammount applied for affected by gender?

In the graph above 1 is male and 2 is female. Males clearly burrow more notice the extra data pieces in the 300 000 range.

Does education affect credit history?

Yes refer to the violin plot the data is concentrated at a credit history of 1. Meaning that there is a positive correlation between eduction vs credit history.

does education affect weekly applicant income?

Yes refer to box plot Applicant income vs eduction level, the average is higher.

If you are married do u have a higher weekly applicant income?

Refer to Applicant income by Marital status viloin plot. The average and distribution at 8000 is higher. Yes if u are married u have a higher weekly applicant income.

if u have a higher weekly applicant income do u get a higher loan ammount?

Refer to Applicant Income vs Loan Amount with Trendline can apply for a higher loan ammount.	. There is a positive correlation	on if u have a higher weekly	applicant income u