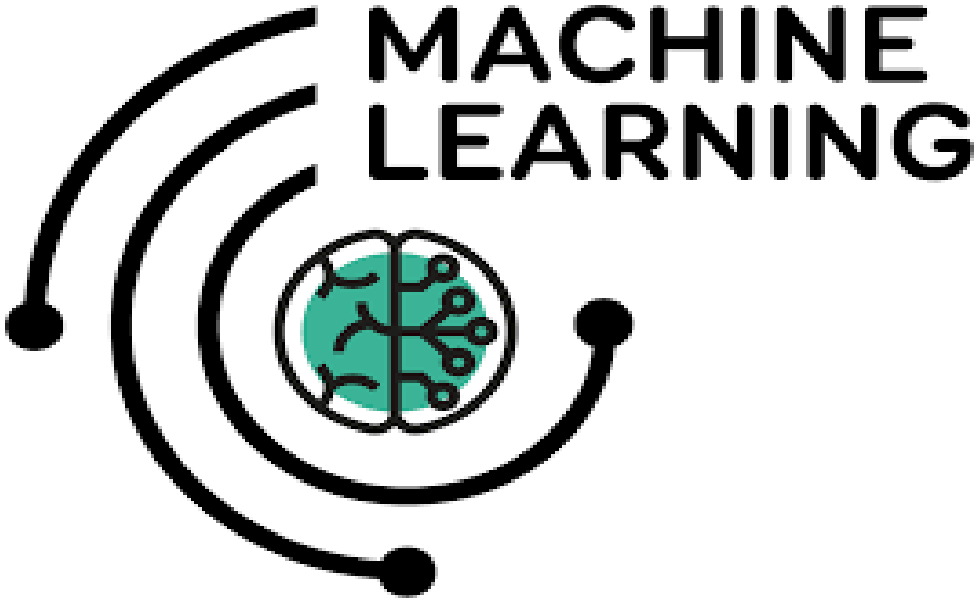


Lending Club Case Study

Dr. Manjunath M Email: manjuu.bms66@gmail.com Phone: 9535096576

Work Flow

1 -Introduction To EDA

2 – Introduction to Case Study

3 – Problem Statement

4 -Procedure to Solve Case Study

5 – Univariate Analysis

6 – Bivariate Analysis

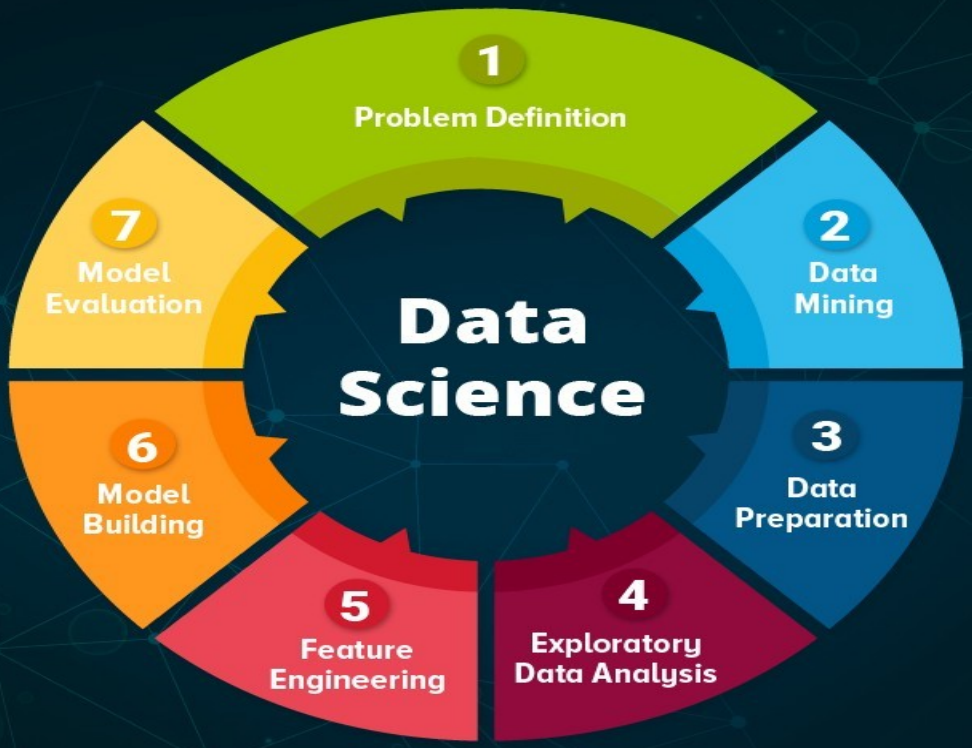
7 - Derived Metrics Analysis

8 – Conclusion

9 – Question and Answer

Dr. Manjunath M
Email: manjuu.bms66@gmail.com
Phone: 9535096576

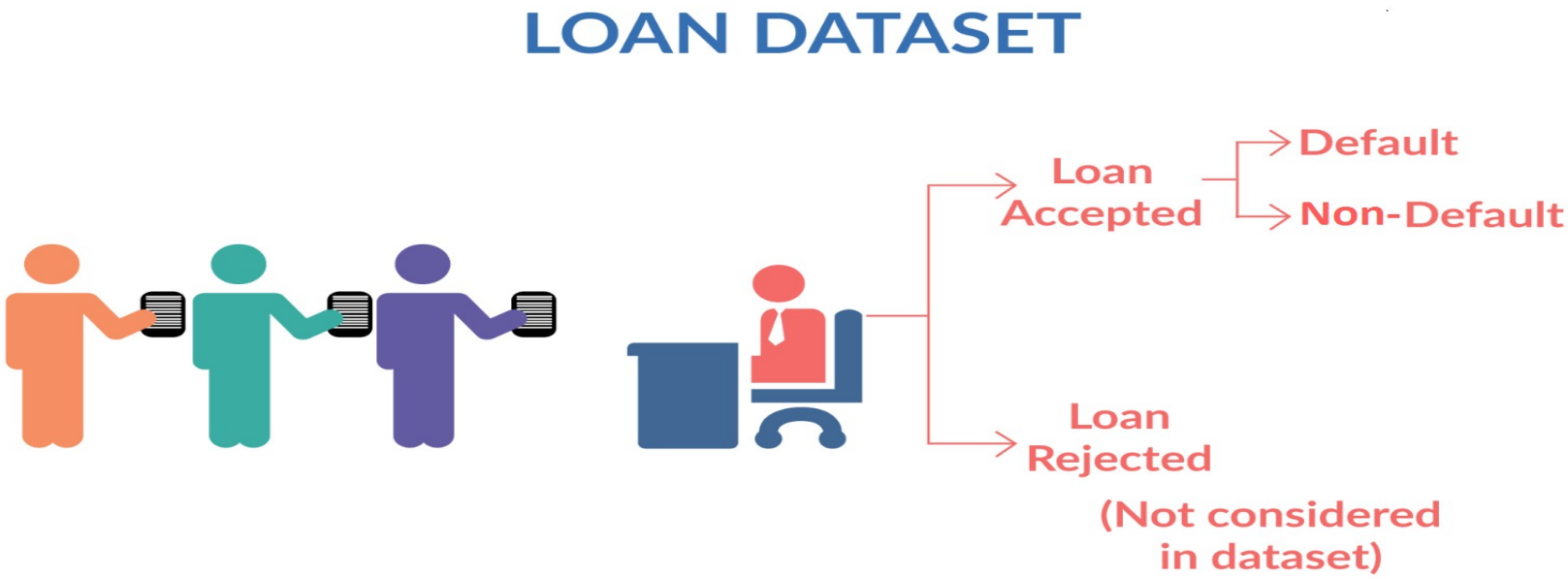
What is Exploratory Data Analysis (EDA) ?



- EDA stands for Exploratory Data Analysis
- EDA is an approach to analyze the data using visual techniques
- It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations

Dr. Manjunath M Email: manjuu.bms66@gmail.com Phone: 9535096576

Problem Statement:



- **The consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.**
- **Two types of risks are associated with the bank's decision:**
  - **If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company**
  - **If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company**

EDA Procedure:

- Exploratory Data Analysis is a **data analytics process** to understand the **data in depth** and learn the different data characteristics, often with visual means.
- This allows you to get a better feel of your data and find useful patterns in it.
- **Steps** Involved in Exploratory Data Analysis
  - **Data sourcing:**
    - It refers to the process of finding and loading data into our system
  - **Data cleaning:**
    - Data cleaning refers to the process of removing unwanted variables and values from the dataset
  - **Univariate analysis:**
    - In Univariate Analysis, analyzsis of data is done for one variable
  - **Bivariate analysis**
    - In Bivariate Analysis, analyzsis of data is done for two variable
  - **Derived metrics:**
    - New variables could be created based on your business understanding

Univariate Analysis:

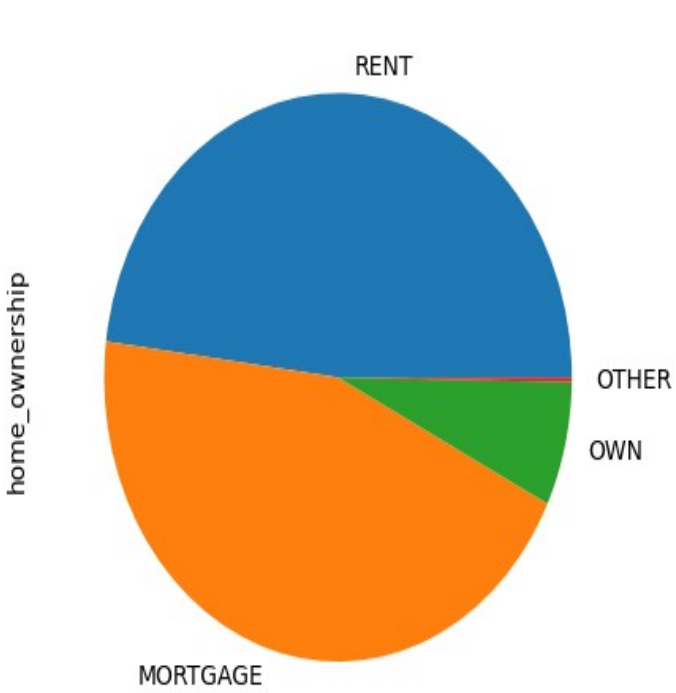
- **Univariate analysis is the technique of comparing and analyzing the dependency of a single predictor and a response variable**
- **Example:**
  - **Analysing the home_ownership column and its relationship**

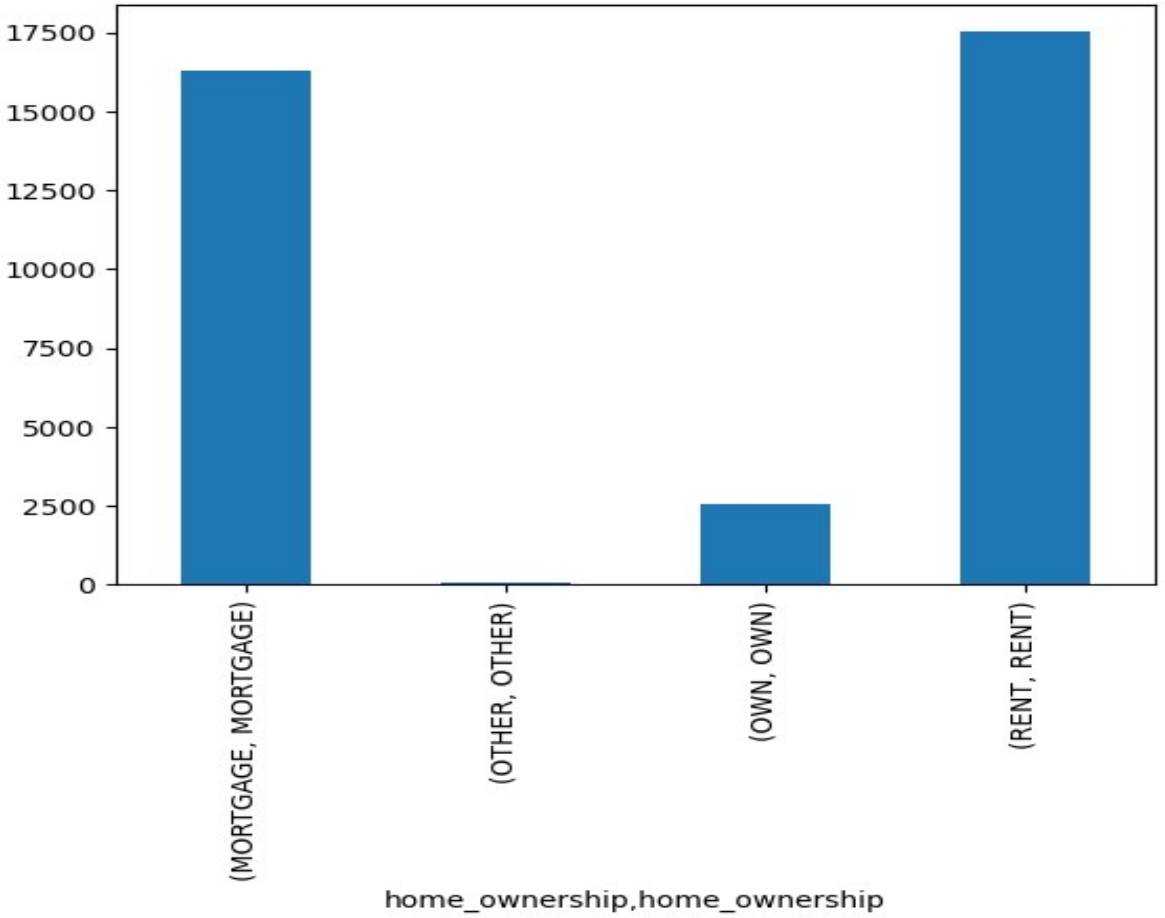
```
[40]: df["home_ownership"].value_counts()

[40]: RENT        17512
      MORTGAGE    16315
      OWN          2581
      OTHER          94
      Name: home_ownership, dtype: int64
```

Univariate Analysis:

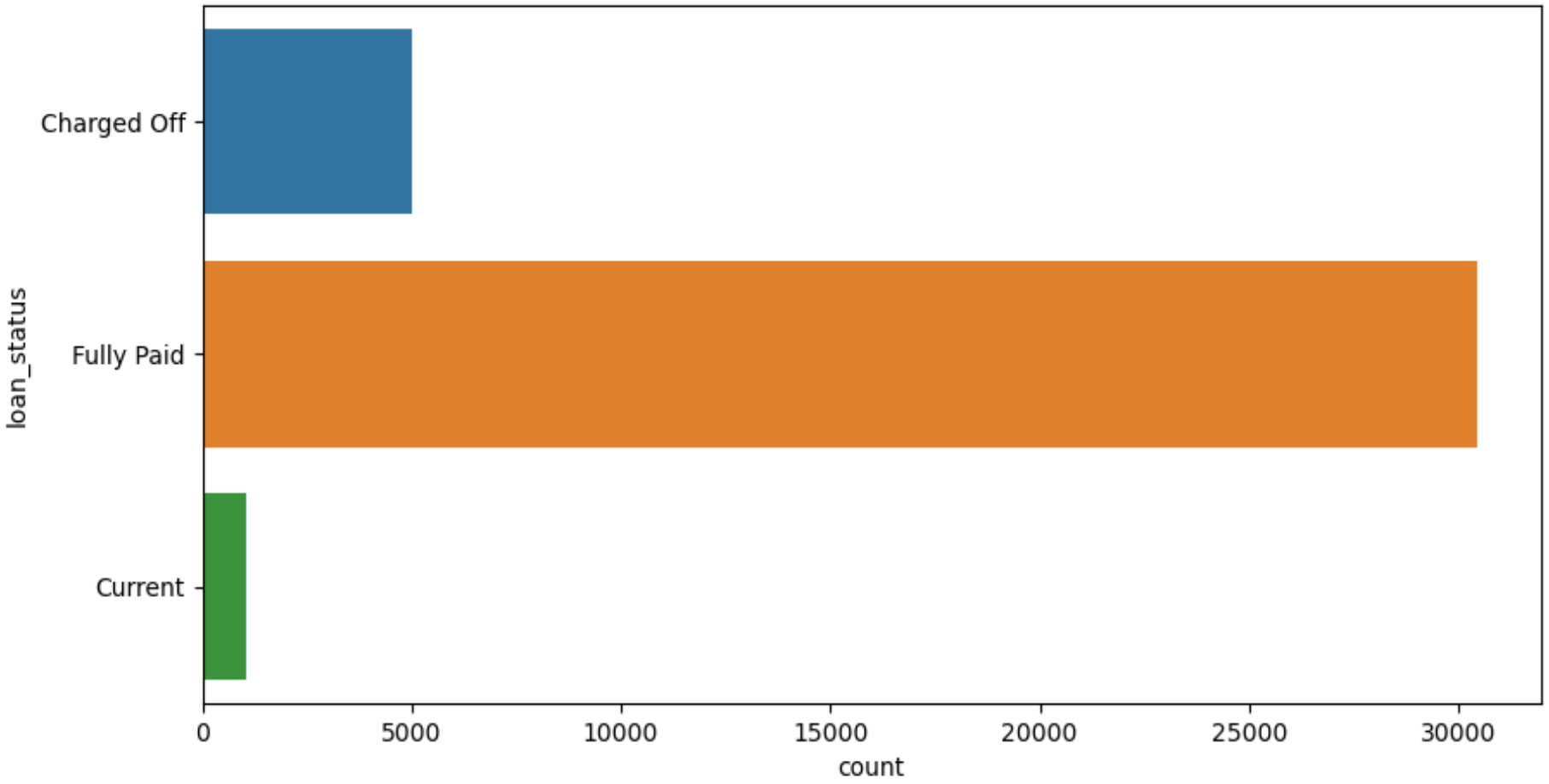
- Analysing the home_ownership column using Pie chart and Bar Chart





Univariate Analysis:

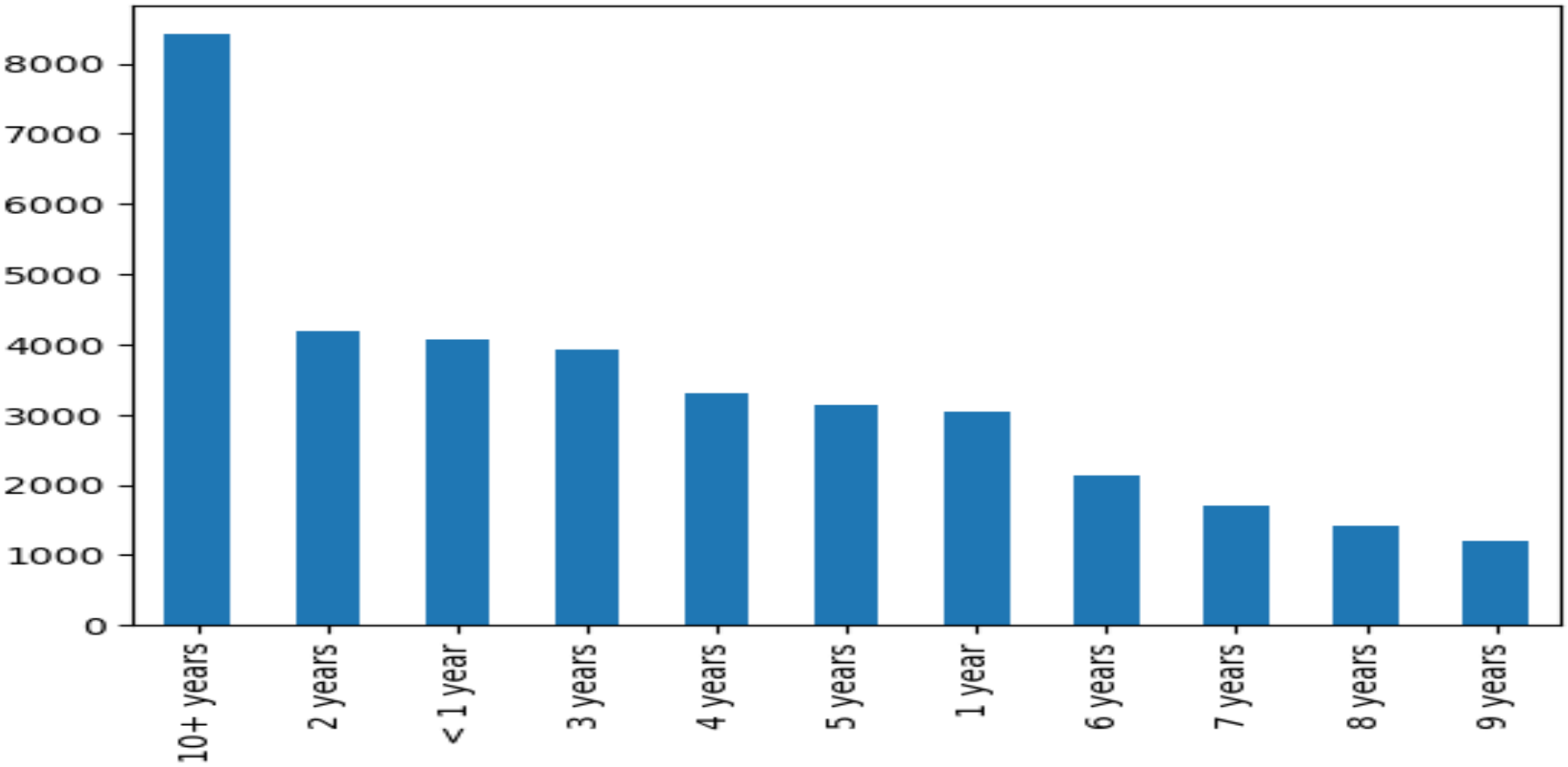
- Analysing the loan_status column using Bar Chart



Dr. Manjunath M
Email: manjuu.bms66@gmail.com
Phone: 9535096576

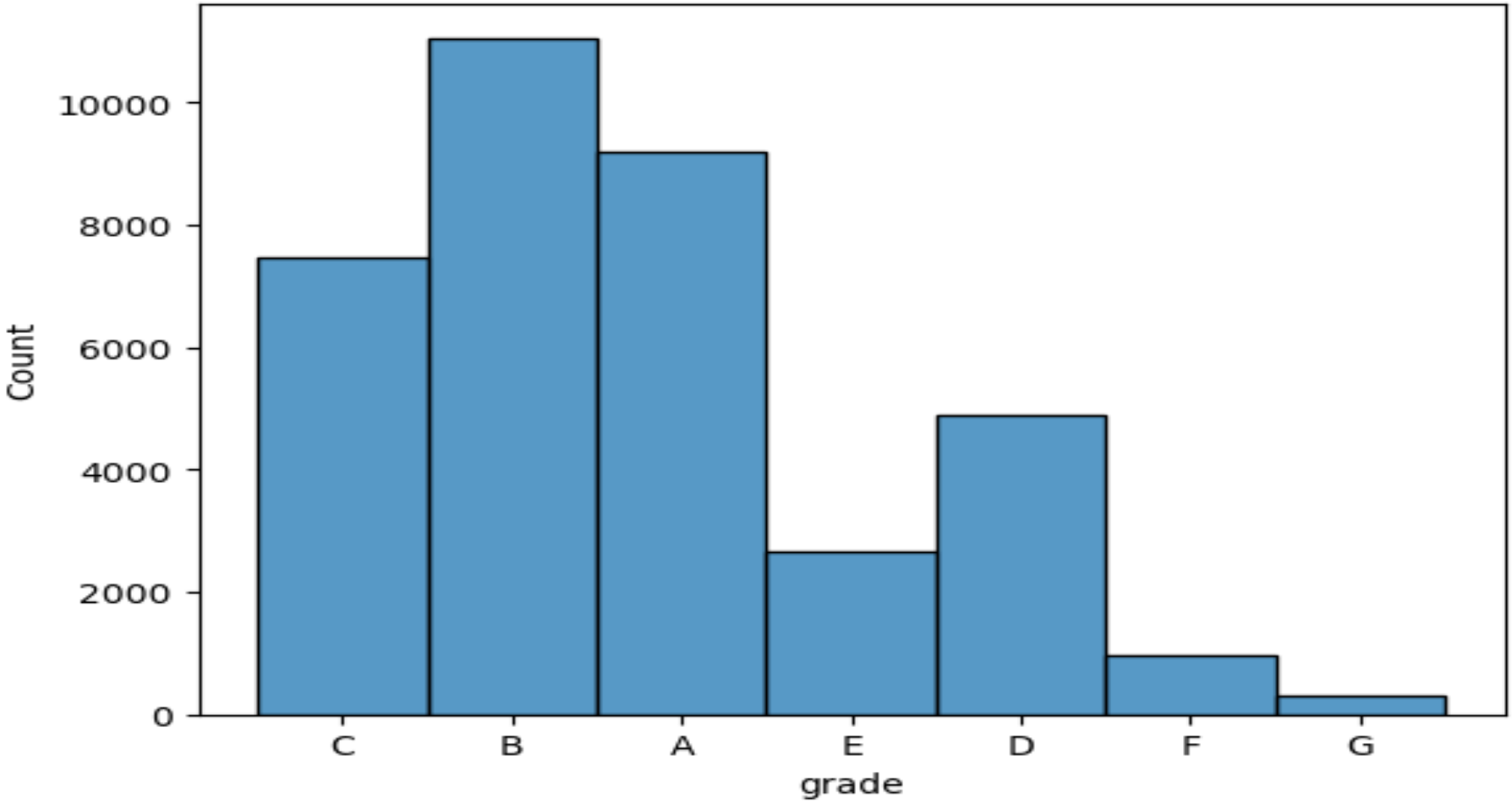
Univariate Analysis:

- Analysing the emp_length column using Bar Chart



Univariate Analysis:

- Analysing the grade column using histogram Chart



Dr. Manjunath M Email: manjuu.bms66@gmail.com Phone: 9535096576

Bivariate Analysis:

- Bivariate analysis helps in analysing the relationship between two variables.
- Example:
  - Analysing the home_ownership and loan_status columns relationship

3.5.2 Anaysisng the home_ownership column with loan_status for better understanding

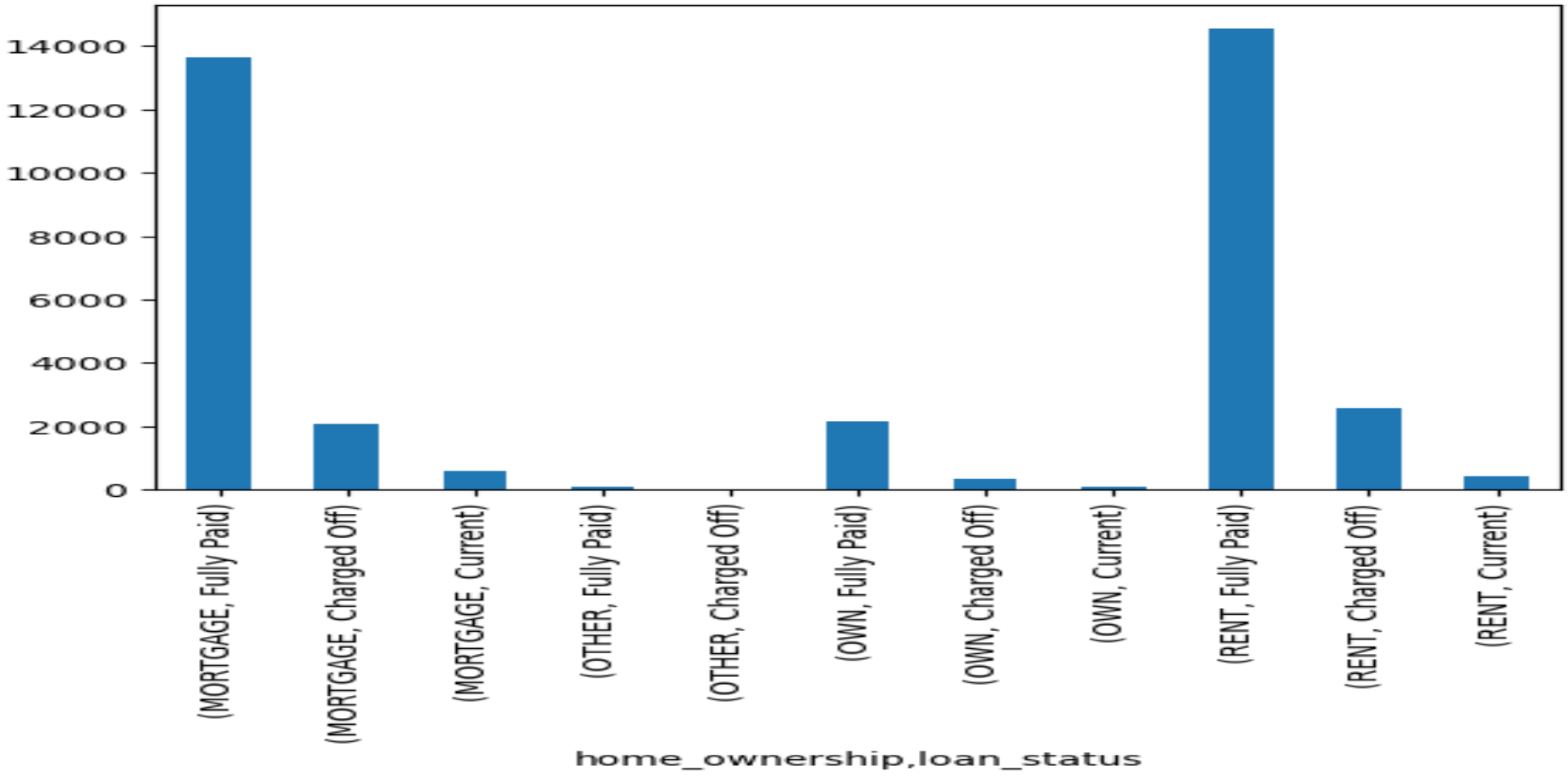
[57]:
```
# Running analysis on home ownership type
df_pivoted_home=df.
 ↪pivot_table(index=["home_ownership"],values="member_id",columns="loan_status",aggfunc="count'
df_pivoted_home.reset_index(inplace=True)
df_pivoted_home
```

[57]:

	loan_status	home_ownership	Charged Off	Current	Fully Paid
0		MORTGAGE	2075	597	13643
1		OTHER	18	0	76
2		OWN	349	70	2162
3		RENT	2562	399	14551

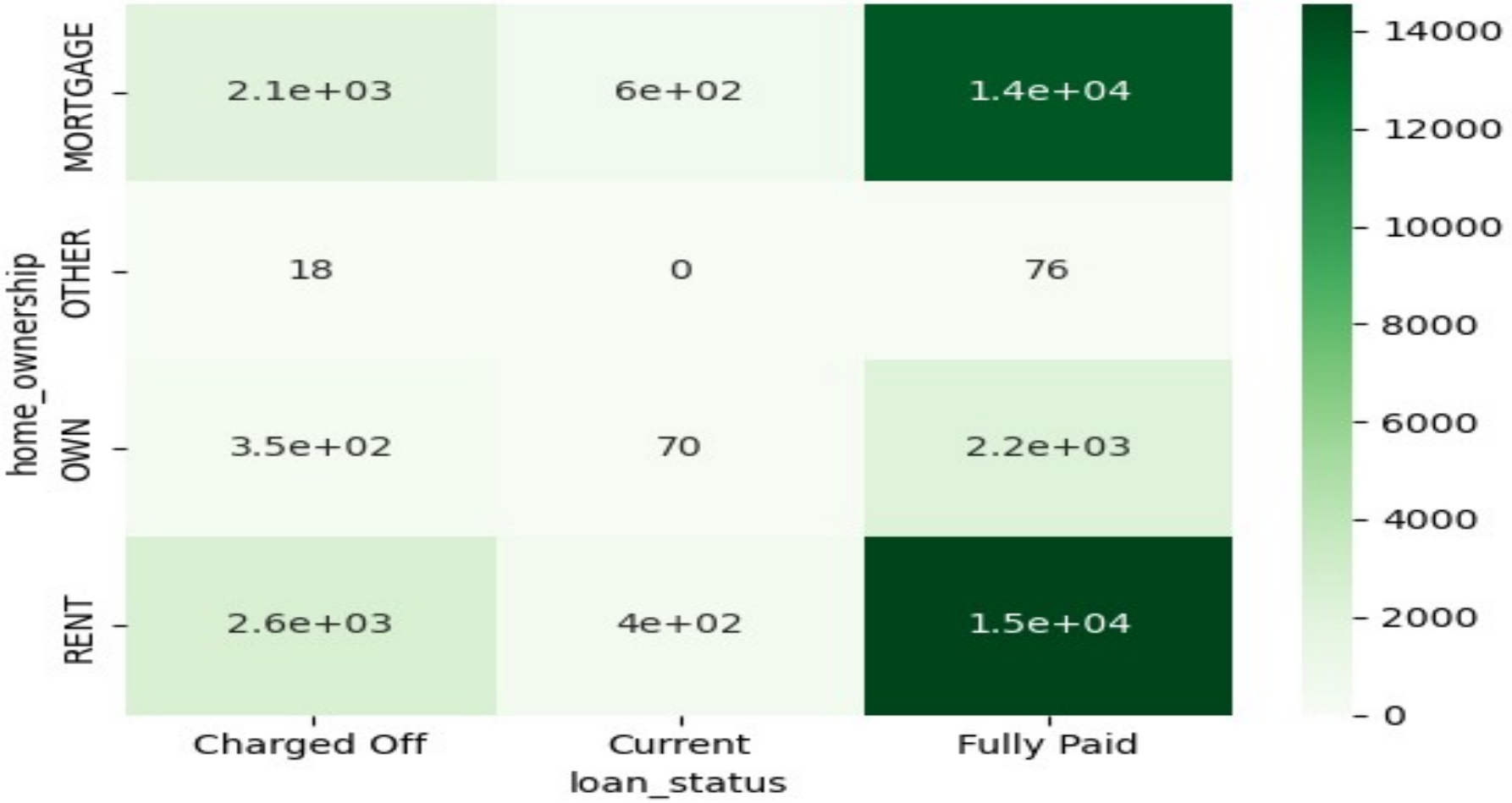
Bivariate Analysis:

- Analysing the home_ownership and loan_status column using Bar Chart



Bivariate Analysis:

- Analysing the **home_ownership** and **loan_status** column using **Heat Map**



Dr. Manjunath M Email: manjuu.bms66@gmail.com Phone: 9535096576

Derived Metrics:

- A derived metric is a calculation based on the data included in the report definition
- Example:
  - Deriving the new column called "Risk"

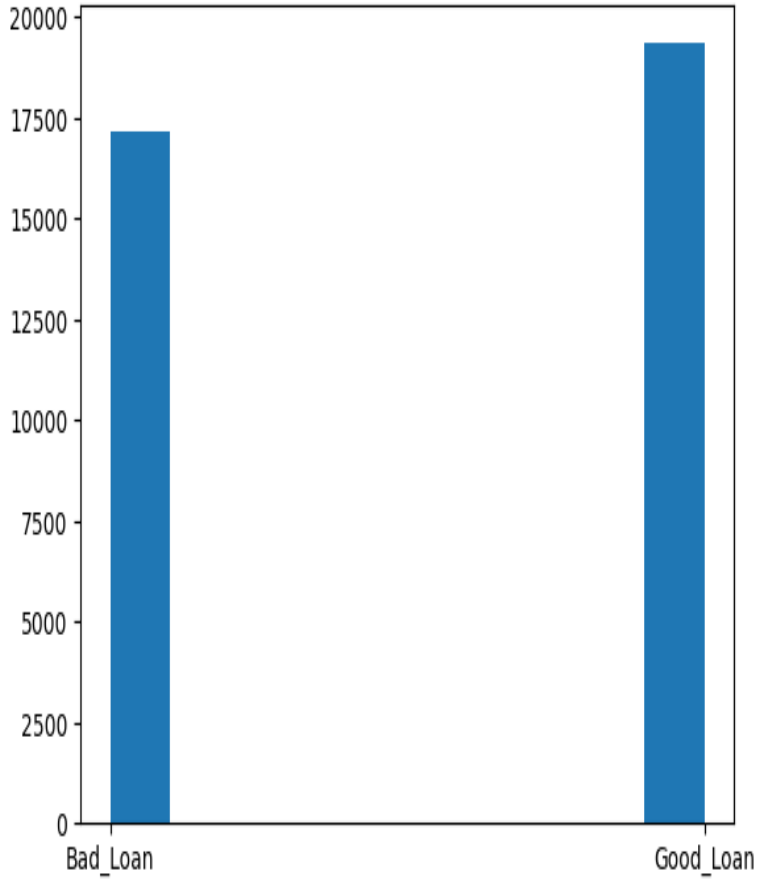
```
Grade=["A", "B", "C"]
Sub_Grade = ["A1", "A2", "A3", "A4", "A5", "B1", "B2", "B3", "B4", "B5", "C1",␣
 ↪"C2", "C3", "C4", "C5", "D1", "D2", "D3", "D4", "D5"]
Home = ["RENT", "OWN", "MORTGAGE"]
loan_status = ["Current", "Fully Paid"]
verification = ["Verified", "Source Verified"]
df["Risk"] = (df["home_ownership"].isin(Home) ) & (df["grade"].isin(Grade)) &␣
 ↪(df["loan_status"].isin(loan_status)) & (df["annual_inc"]>= 40000)
```

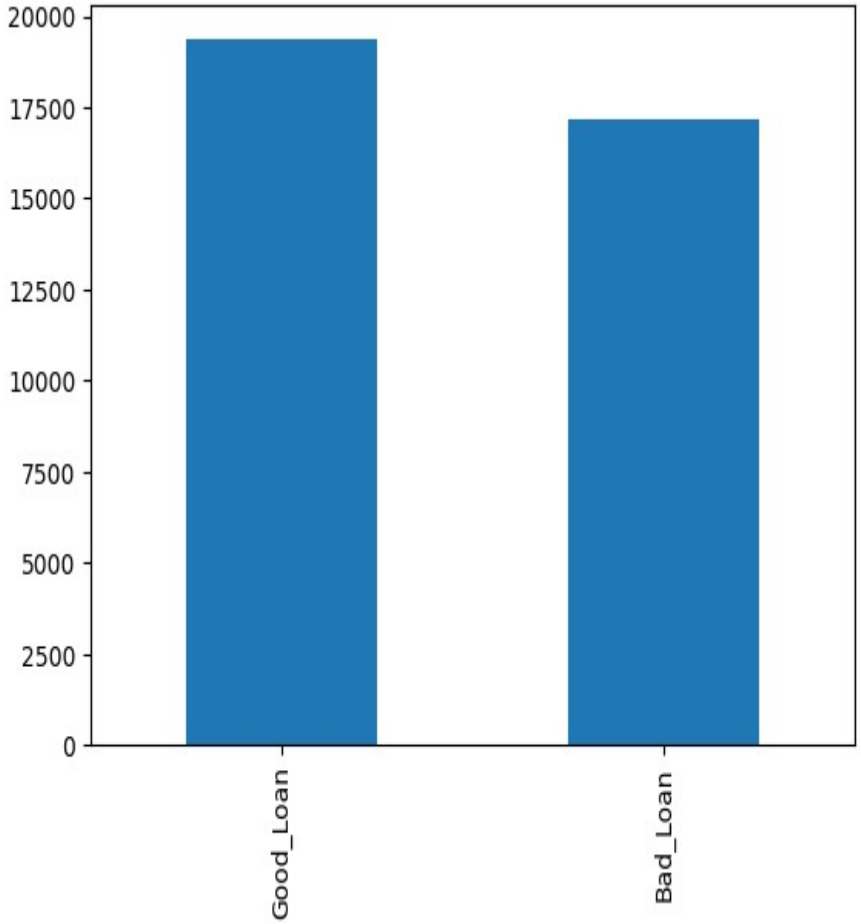
```
df["Risk"].value_counts()
```

```
True     19356
False    17146
Name: Risk, dtype: int64
```

Derived Metrics:

- Analysing the Risk column using histogram and Bar Chart

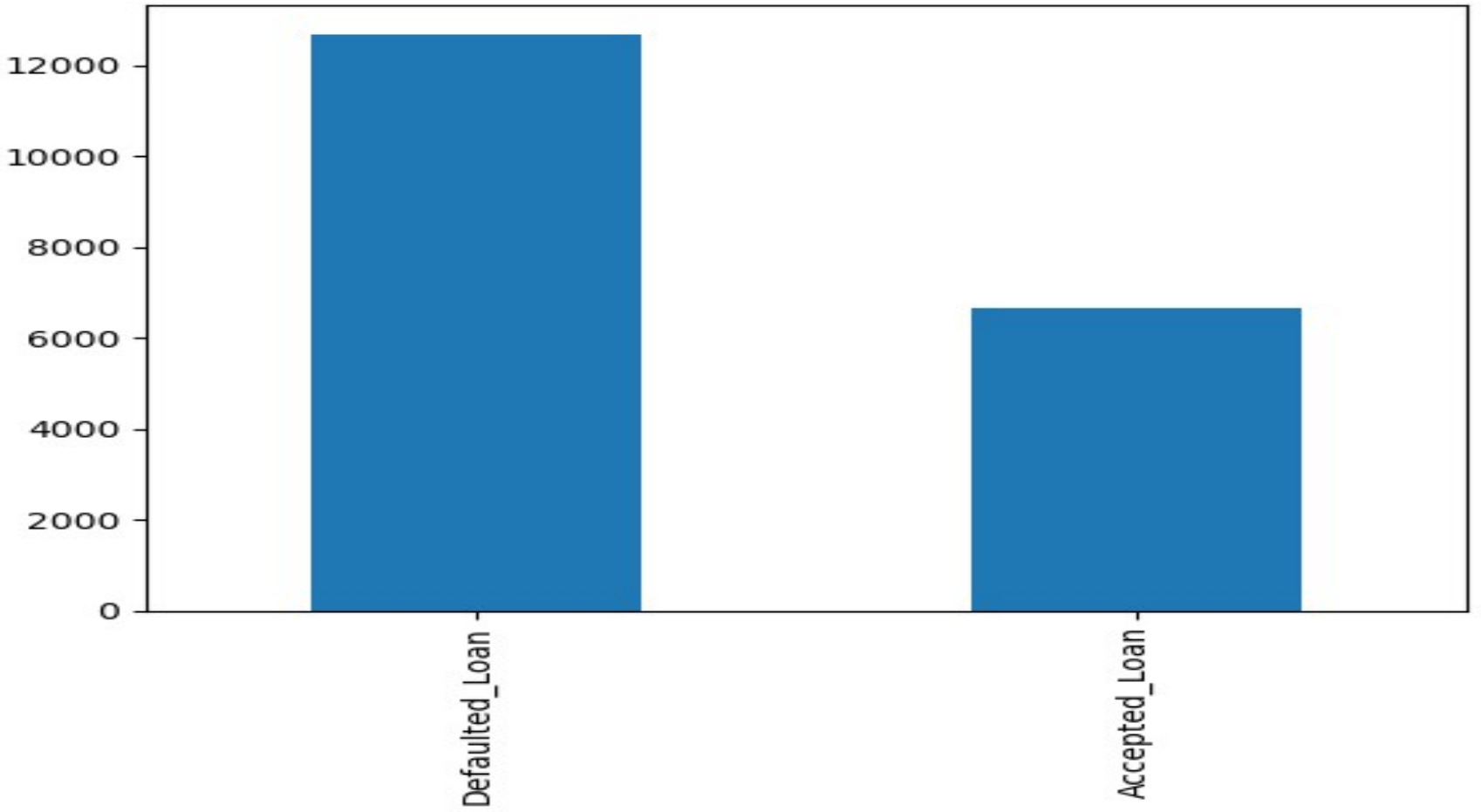




Dr. Manjunath M Email: manjuu.bms66@gmail.com Phone: 9535096576

Derived Metrics:

- Analysing the **Accepted** and **Defaulted** Loan using **Bar Chart**



Derived Metrics:

- Analysing the **Accepted** and **Defaulted Loan** using **Heat Map**

Dr. Manjunath M Email: manjuu.bms66@gmail.com Phone: 9535096576

Statistics of Accepted Loan

In [109]:
```
df_Acce = df_Res["Accepted"]=="Accepted_Loan"
df_Res.loc[df_Acce, :]
```

Out[109]:

id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership
1075358	1311748	3000	3000	3000.000000	60 months	12.69%	67.79	B	B5	University Medical Group	1.0	RENT
1069908	1305008	12000	12000	12000.000000	36 months	12.69%	402.54	B	B5	UCLA	11.0	OWN
1062474	1294539	6000	6000	6000.000000	36 months	11.71%	198.46	B	B3	Connection Inspection	1.0	MORTGAGE
1069710	1304821	10000	10000	10000.000000	36 months	11.71%	330.76	B	B3	Value Air	11.0	OWN
1069697	1273773	15000	15000	15000.000000	36 months	9.91%	483.38	B	B1	Winfield Pathology Consultants	2.0	MORTGAGE
...	...	...	...	...	...	...	...	...	...	...	...	...
355680	358791	1000	1000	92.173793	36 months	7.37%	31.05	A	A1	Retired	11.0	OWN
355467	360172	7500	5550	0.000000	36 months	8.32%	174.74	A	A4	L-3 Communications Holdings	1.0	MORTGAGE
351964	354815	8000	8000	2141.029177	36 months	10.96%	261.76	B	B5	Kapstone	5.0	MORTGAGE
323288	323280	7500	7500	1758.843849	36 months	10.39%	243.38	B	B4	H&S	0.0	MORTGAGE
308498	308484	25000	18175	14903.250000	36 months	10.08%	587.14	B	B3	Emergency Medical Associate	2.0	MORTGAGE

6666 rows × 43 columns

Statistics of Defaulted Loan

In [110]:
```
df_Def= df_Res["Accepted"]=="Defaulted_Loan"
df_Res.loc[df_Def, :]
```

Out[110]:

id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	emp_title	emp_length	home_ownership
1076863	1277178	10000	10000	10000.00	36 months	13.49%	339.31	C	C1	AIR RESOURCES BOARD	11.0	RENT
1069639	1304742	7000	7000	7000.00	60 months	15.96%	170.08	C	C5	Southern Star Photography	8.0	RENT
1070078	1305201	6500	6500	6500.00	60 months	14.65%	153.45	C	C3	Southwest Rural metro	5.0	OWN
1065775	1299699	10000	10000	10000.00	36 months	15.27%	347.98	C	C4	Chin's Restaurant	4.0	RENT
1069971	1304884	3600	3600	3600.00	36 months	6.03%	109.57	A	A1	Duracell	11.0	MORTGAGE
...	...	...	...	...	...	...	...	...	...	...	...	...
223308	223192	7500	7500	1000.00	36 months	10.78%	244.76	C	C1	Tradelink	1.0	MORTGAGE
222829	222675	14400	14400	1510.69	36 months	9.51%	461.35	B	B2	County of San Diego	11.0	RENT
200600	200597	7500	7500	1599.78	36 months	9.83%	241.41	B	B3	UCLA Medical Center	7.0	RENT
186572	186568	12000	12000	725.00	36 months	9.01%	381.66	B	B2	Bank of America Corp.	6.0	MORTGAGE
158706	158450	12375	12375	1000.00	36 months	10.91%	404.62	C	C3	Fullmoon Software	2.0	RENT

12690 rows × 43 columns

Conclusion

- The analysis presented above is related to Loan dataset.
- The aim of this analysis to reduce the loss percentage of Consumer Finance Company by providing the loan to loan applicants
- This detailed analysis of loan dataset using EDA addresses the two key challeges
  - If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
  - If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company
- The analysis presented above, identifies the number of Good_loan and Bad_loans using various EDA process and, also *identify the staticstics to approve Full loan and Defaulted loan based on the loan applications.*
- Hence, by identifing staticstics to Full approved loan and dfaulted loan, *reduces the loss percentage for the Finance Company*
-

Dr. Manjunath M *Email: manjuu.bms66@gmail.com* *Phone: 9535096576*


Thank
you