EDA CASE STUDY

LOAN DATA ANALYSIS

BY MAANIK SHARMA

BUSINESS BACKGROUND & PROBLEM

We are analyzing data from a **consumer finance company**, the **largest online loan marketplace**.
It offers:

- **Personal loans**
- **Business loans**
- **Medical financing**

Problem Statement:

Can we identify what factors affect loan repayment or default?

OBJECTIVE & ANALYSIS APPROACH

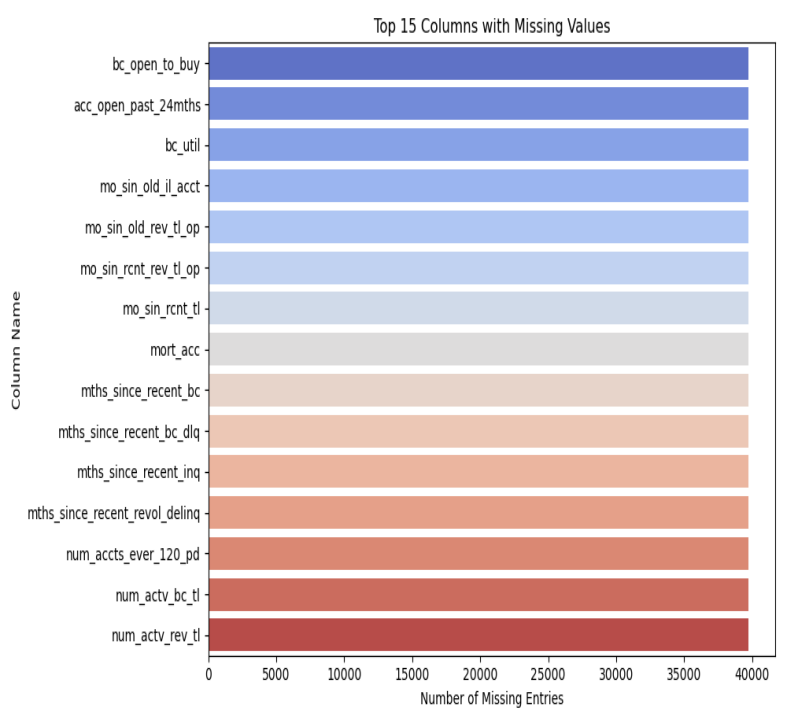
- **Goal:**
- Understand borrower behavior
- Find patterns linked to **loan default**
- Use EDA to help improve **loan approval decisions**
- **Approach:**
- Load and clean data
- Perform **univariate** & **bivariate** analysis
- Summarize **business-level findings**

DATASET OVERVIEW

	loan_amnt	term	int_rate	grade	emp_length	home_ownership	annual_inc	loan_status
0	5000	36 months	10.65	B	10+ years	RENT	24000.0	Fully Paid
1	2500	60 months	15.27	C	< 1 year	RENT	30000.0	Charged Off
2	2400	36 months	15.96	C	10+ years	RENT	12252.0	Fully Paid
3	10000	36 months	13.49	C	10+ years	RENT	49200.0	Fully Paid
4	3000	60 months	12.69	B	1 year	RENT	80000.0	Current

- Dataset has over **X rows** and **Y columns**
- Includes: Loan amount, interest rate, grade, employment, state, and loan status
- We removed irrelevant and highly missing columns

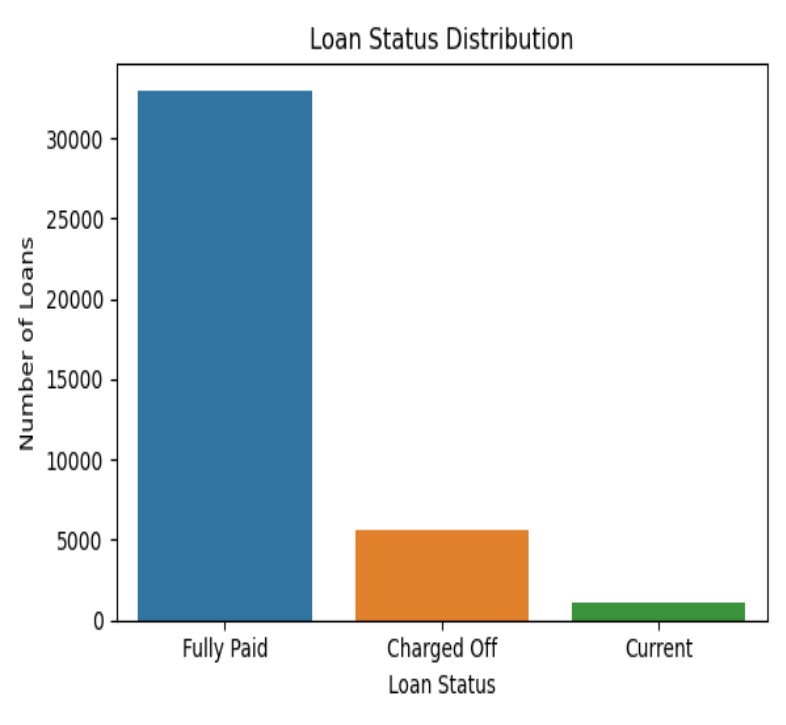
MISSING DATA



- Some columns like **employment title**, **delinquency history**, and **joint income** have missing data
- These missing values were either:
- ✅ **Dropped** if not important
- 🔧 **Filled** using techniques like "unknown" or average values
- Helps keep the dataset clean and accurate for analysis

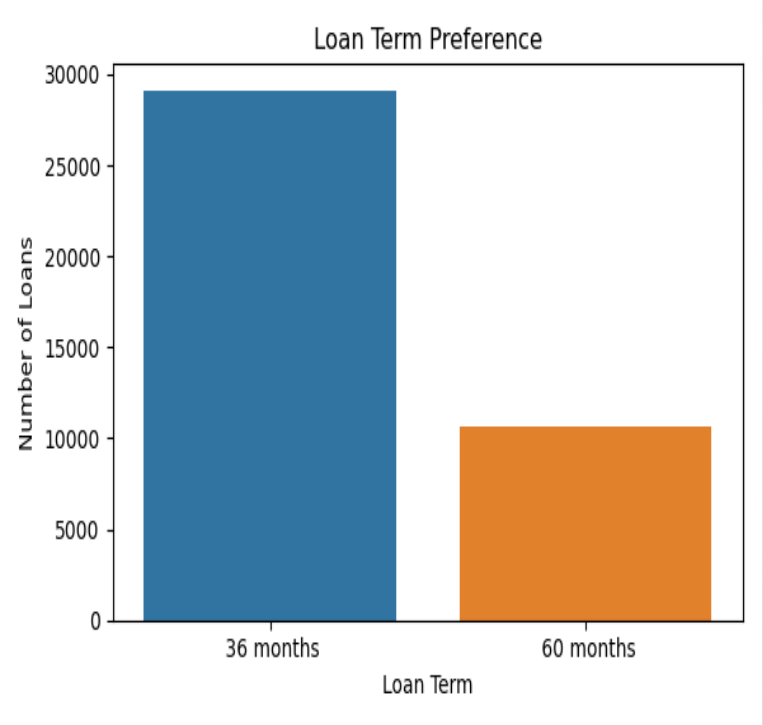
UNIVARIATE ANALYSIS

LOAN STATUS DISTRIBUTION



Majority of loans are **fully paid**, but a significant portion is **charged off**. (i.e. defaulted)

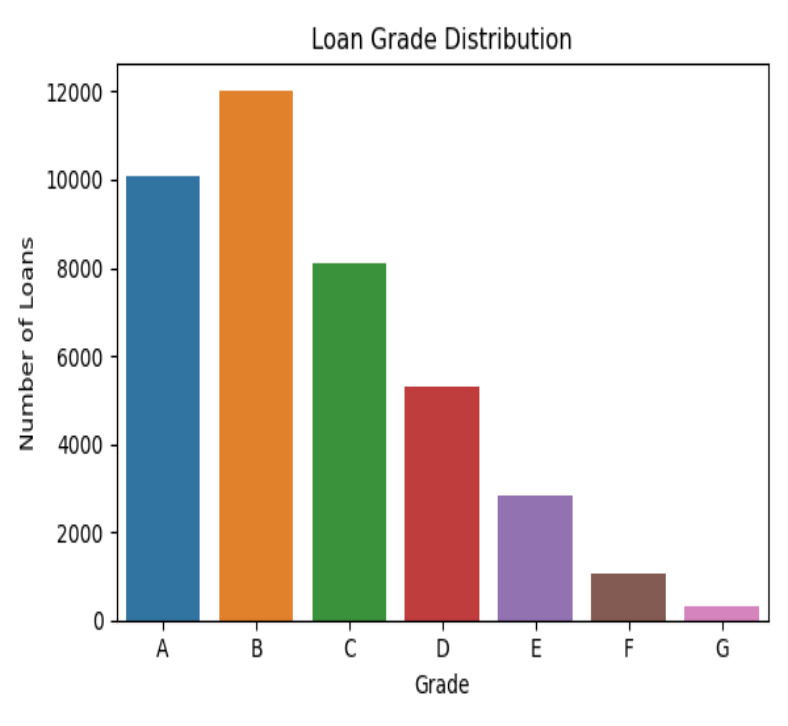
LOAN TERM PREFERENCE



Most borrowers choose **36-month loans**

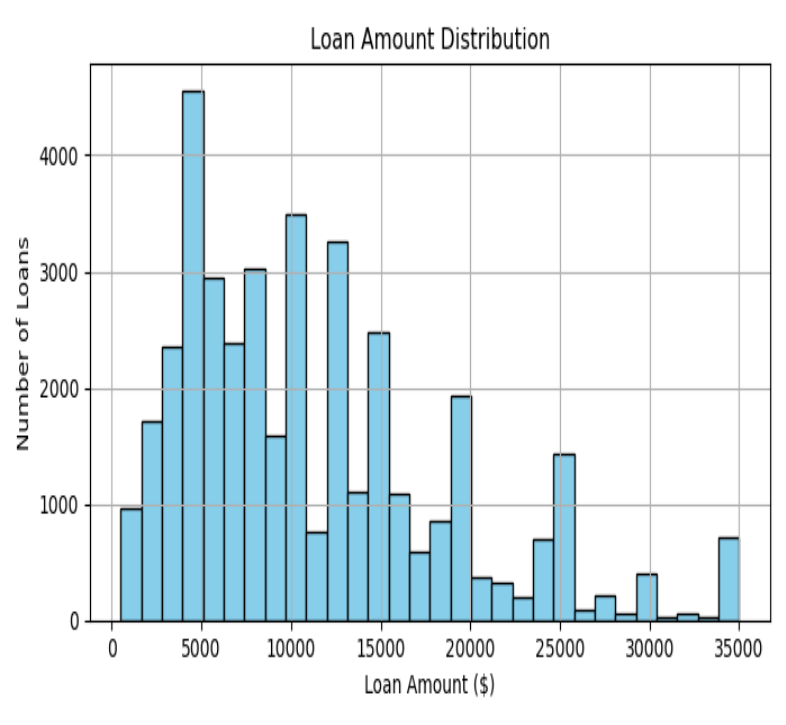
Possibly due to lower EMIs and quicker clearance

LOAN GRADE DISTRIBUTION



- Loans are categorized from **Grade A to G**
- Most loans are in **Grade B** and **Grade C**

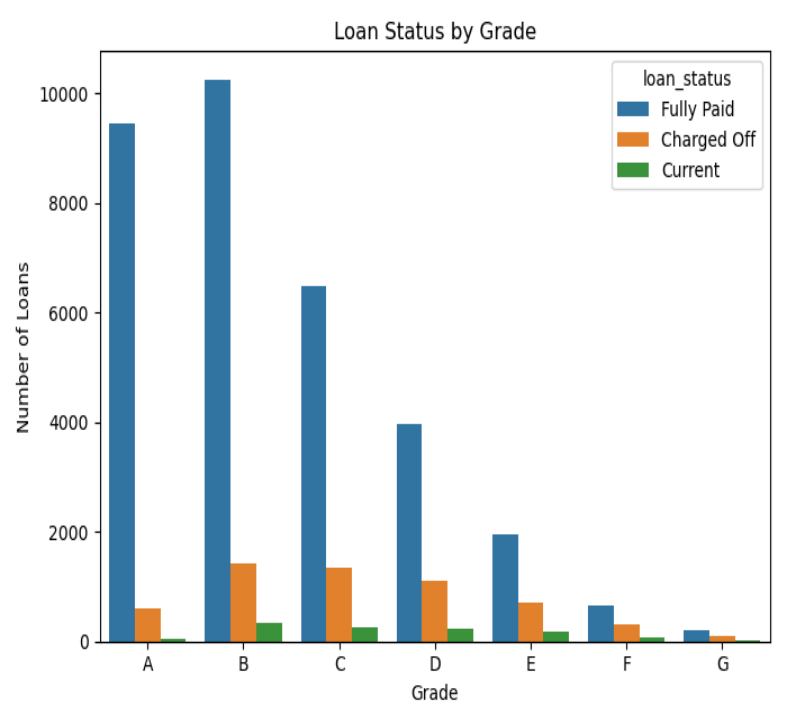
LOAN AMOUNT DISTRIBUTION



- Majority of loans are under **$15,000**
- Business focus seems on **small personal financing**

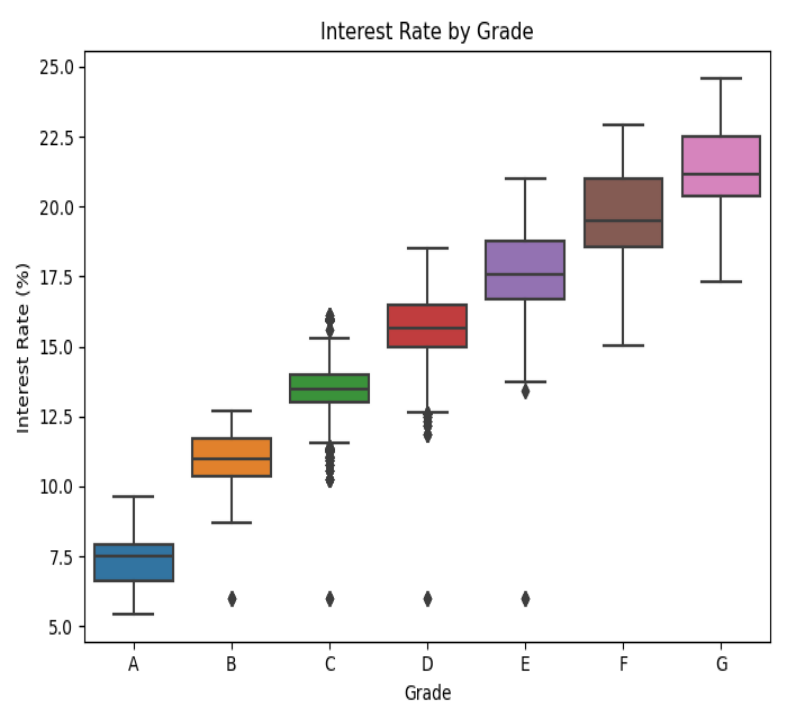
BIVARIATE ANALYSIS

GRADE VS LOAN STATUS



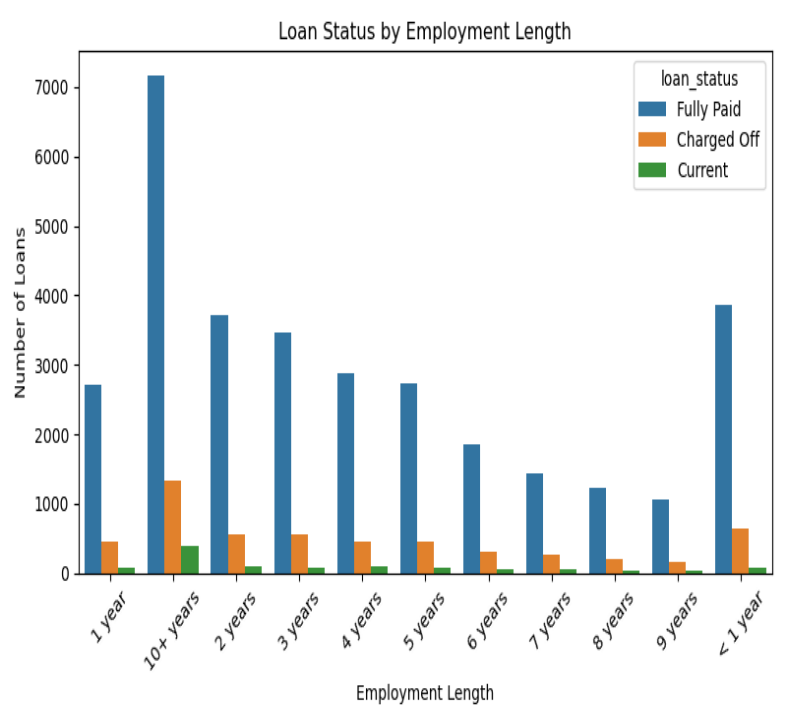
- Lower grade loans have a **higher default rate**
- Business should be cautious with D–G grades

INTEREST RATE BY GRADE



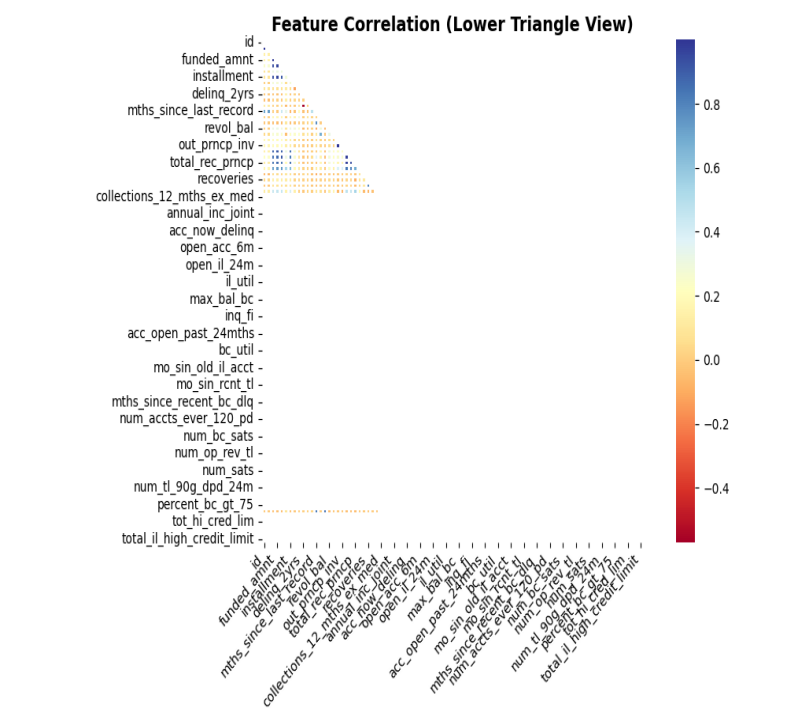
- Interest rates increase with **lower grades**
- Company prices loans higher for **risky borrowers**

EMPLOYMENT LENGTH VS DEFAULT



- Borrowers with **longer job history** default less
- Indicates **job stability** = better creditworthiness

REGIONAL TRENDS & FEATURE RELATIONSHIPS



- Some states like **CA, TX, NY** show much higher loan volumes
- Helps identify **key markets** or **regions with risk concentration**
- The correlation heatmap shows:
- **Loan amount** is strongly linked to **installment size**
- **Interest rate** is slightly higher for **larger loans and risky grades**

KEY BUSINESS INSIGHTS

- **Lower grades** → higher default risk
- **Interest rates** reflect that risk
- **Stable employment** improves repayment
- Most loans are **small-ticket, short-term**

CONCLUSION

- This analysis helped us understand which customer features lead to **loan repayment** or **default**.
- Key drivers: **grade**, **interest rate**, **employment length**, and **loan amount**
- The company can use these insights to **improve loan risk profiling** and **target the right customers**.

SUMMARY

- Data-driven decisions help improve loan targeting & reduce defaults
- Insights like these can shape lending policies and risk models