Lending Case Study using EDA

By

Renuka Mokka

VeeraNagi Reddy Mekala

Problem Statement

- We need to help a **consumer finance company** which specialises in lending various types of loans to urban customers in their **risk analytics**. This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures .
- The input data contains information about past loan applicants and whether they ‘defaulted’ or not. The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

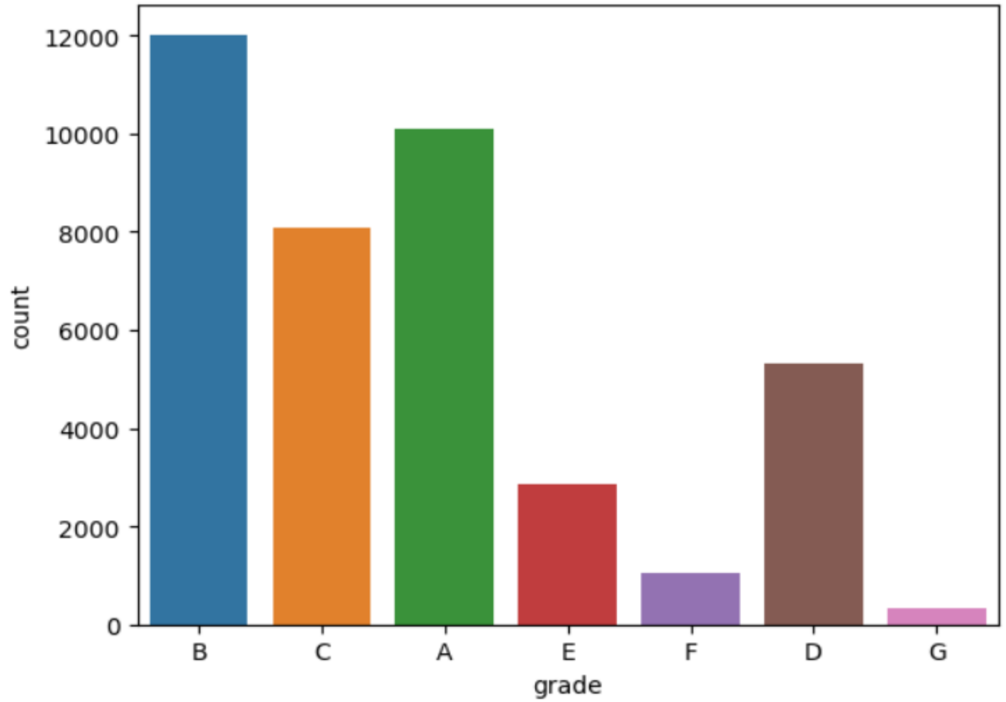
Data Cleaning

- Checked the missing value percentage of input columns and dropped the columns having 40% and above null value percentage.
- Imputed the columns having less percentage of missing values using median for continuous variables and “Missing” for categorical variables.
- Type casted few columns from int to string object (emp id, id) , object to date-time for date related objects, object to int for percentage values(interest rate)
- Created **Derived column** for ‘Charged Off’ column having 1 and 0 values based on loan status.

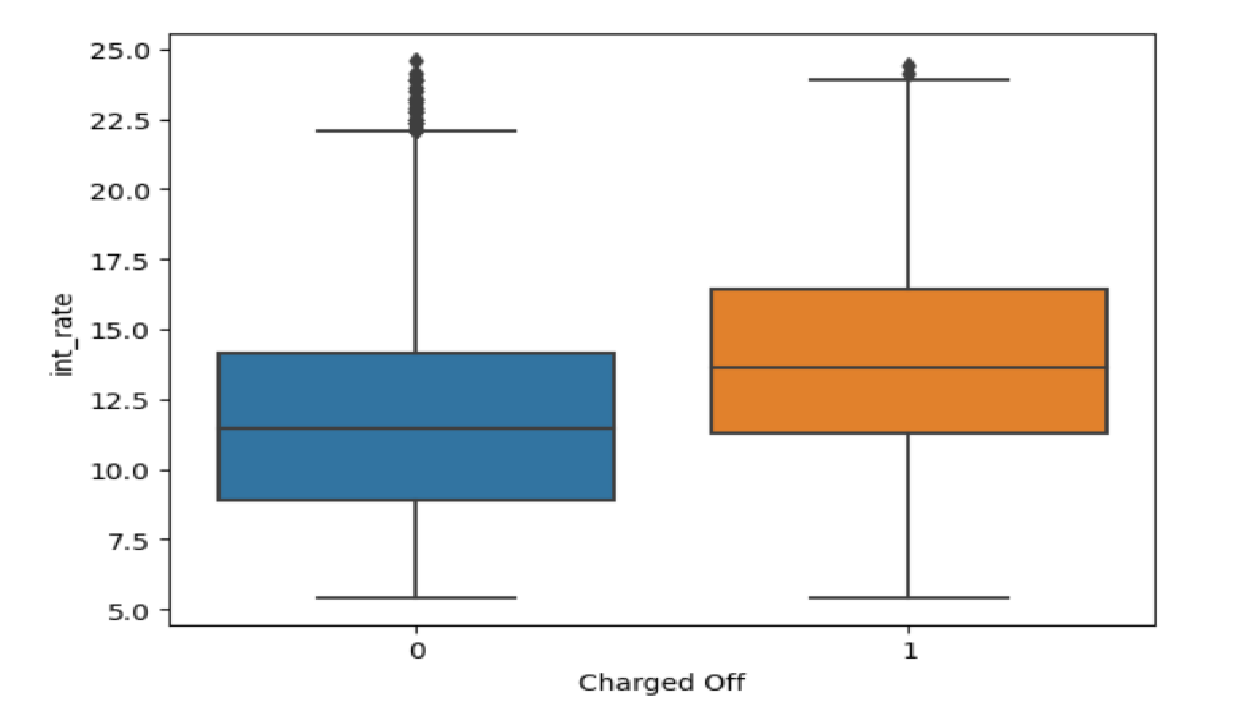
Univariate Analysis

Categorical

```
sns.countplot(x=loan["grade"])
plt.show()
```

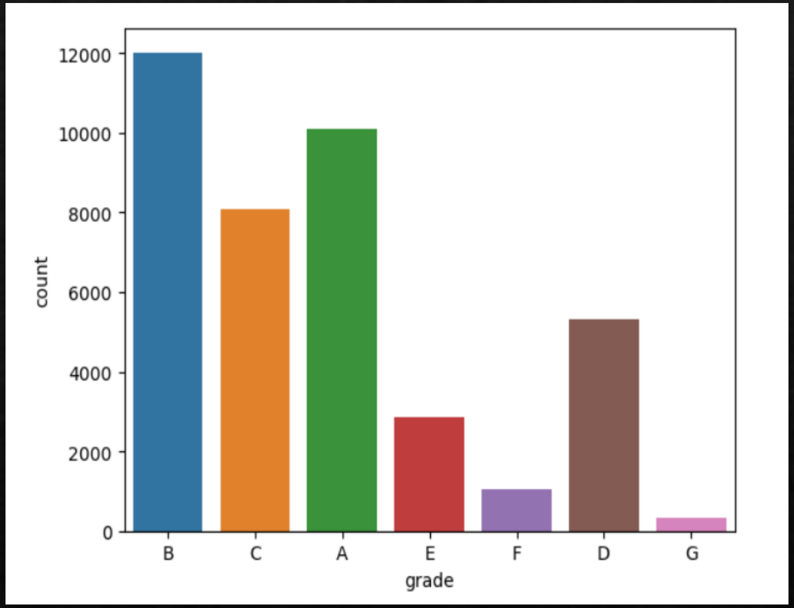


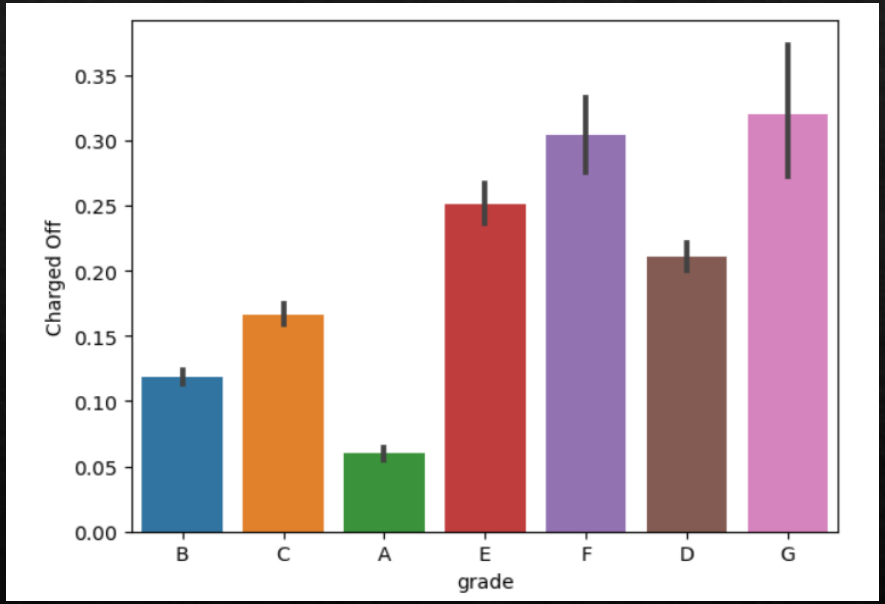
Bivariate Analysis
Charged off vs Interest Rate (int_rate)



Conclusion : Higher chances of Charged Off when higher the interest rate.
i.e. people who are willing to go for high interest rates, have more probability of being defaulters.

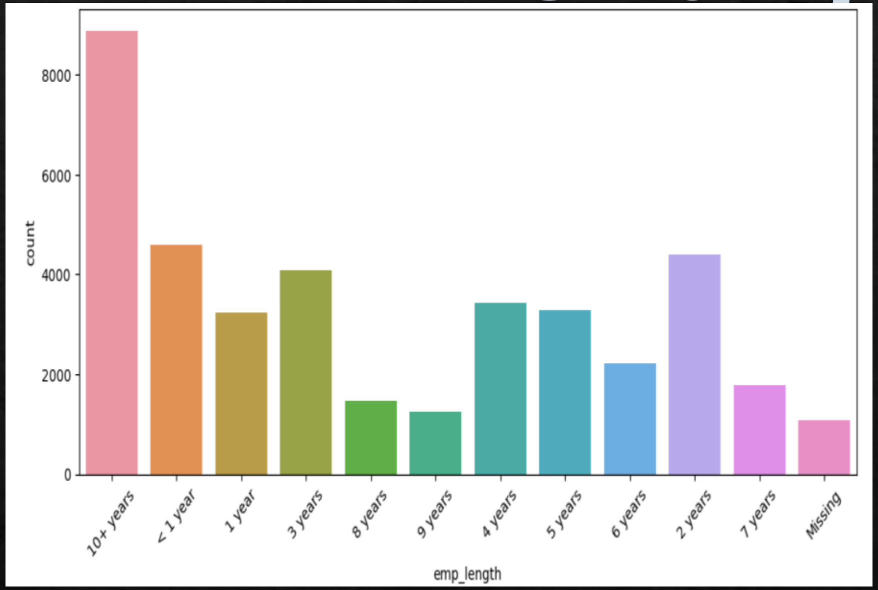
Grade vs Charged-off

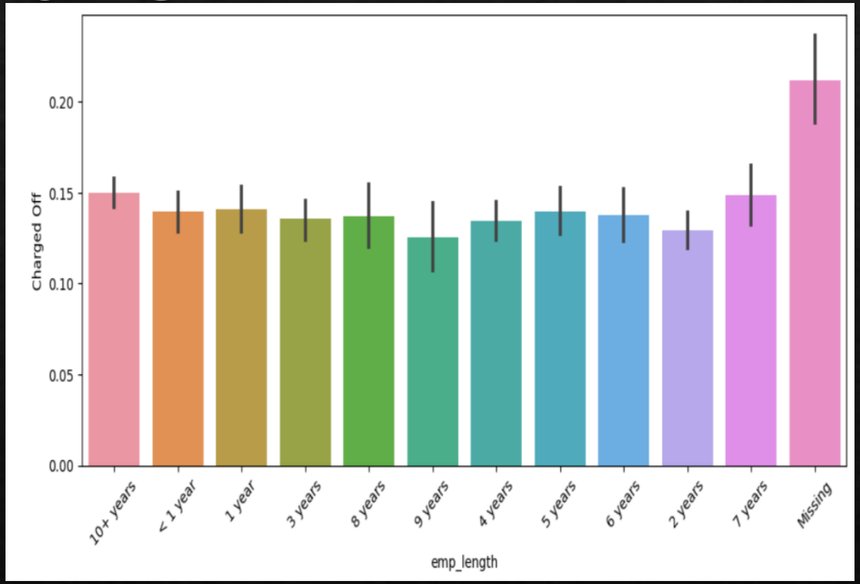




Conclusion: Evident that the "grade" is a driving factor of loan defaulter. Higher the chances of getting charged off with the lower the grade. Example G grade is most probable of defaulter and A grade is less probable to default the loan.

Experience (Emp_length) vs Charged off

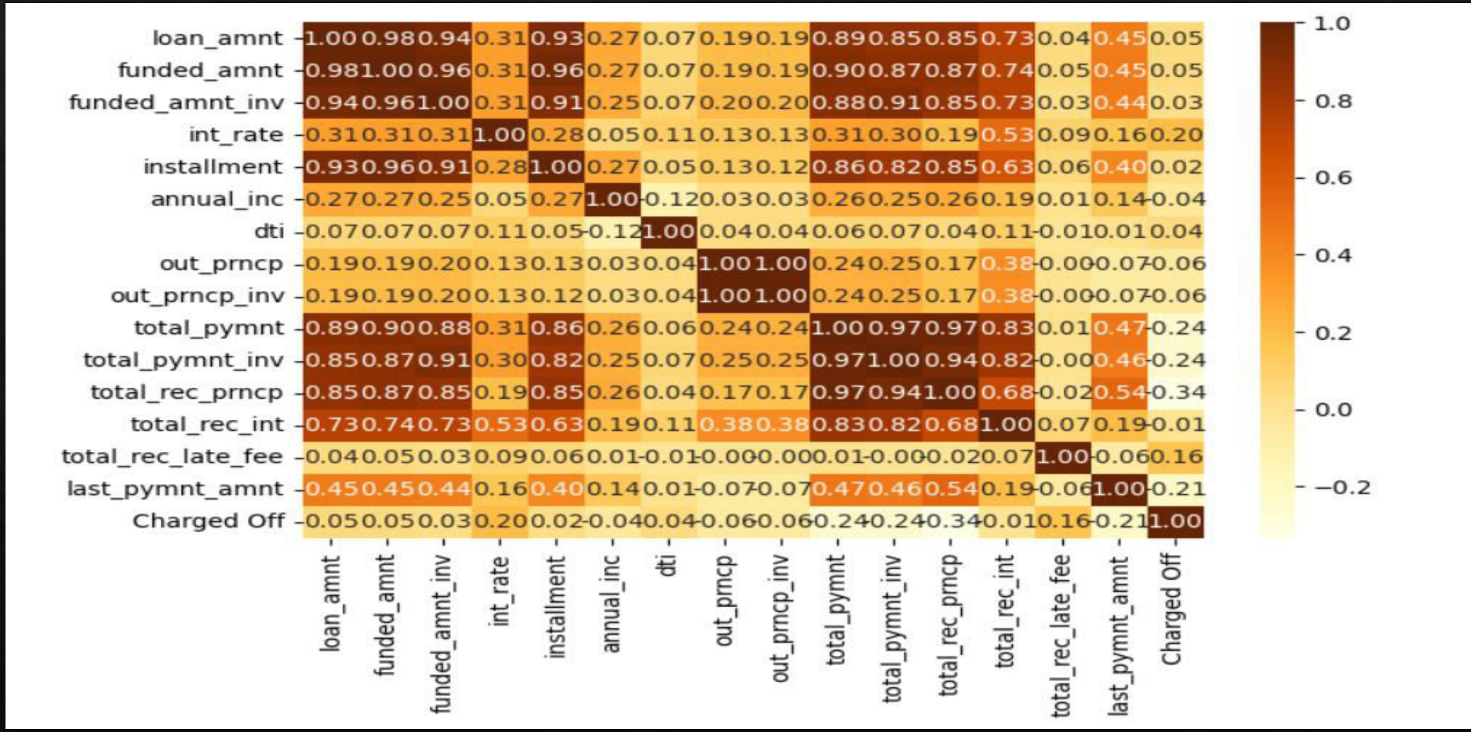




Conclusion: From the above charts experience vs charged off and experience count charts, Employee with lower experience are more probable to default the loan. (Here assumed that missing emp_length means, he is not working in a stable job anywhere, hence it's data not captured in input data). 10+ Experienced employee has huge count in the data, but their charged off seems similar with other data values, hence it shows that 10+ employees, have less probable to default the loan.

Hence, it's always advisable to offer loan to employees having stable service period, it's advised to go for employment history and then approve the loan.

Multivariate Analysis



Conclusion: From Charged off vs Interest rate and heatmap, it is understandable that Charged Off is being driven by Interest,Total principle received and last_payment_amnt. We can also observe that more the interest rate, more chances of being charged off. Also, higher the principal amount, higher chances of defaulter.

Conclusions

The Driving factors of Charged_off (loan defaulters) are:

- Grade – The lower the grade -> the higher chances of Charged off
- Interest Rate (Int_rate) – Observed that people who opt for high interest_rate, has more chances of charged_off
- Experience (Emp_length) -> Better to check for emp_length, as persons whose emp_length blank has more chances of charged_off.
- Loan Amount (Total_Principal_received and Last_Payment_Amount) – they are slightly correlated