LendingClub

A case study

Apr, 2023

Abhinav Sinha
Sachin Sachdeva

PROBLEM STATMENT

- Lending Club is the largest online loan marketplace, providing loans for medical procedures, personal, and business purposes.
- Borrowers can easily access loans at lower interest rates through a fast online process.
- Identifying risky loan applicants and understanding the factors that drive default is crucial for the company to reduce losses and manage risk effectively. This is the focus of the present case study.

Lets look at the data ..

Data Deep-dive

the file provided ‘loan.csv’ had:

1. 111 Columns
2. 39,717 Rows

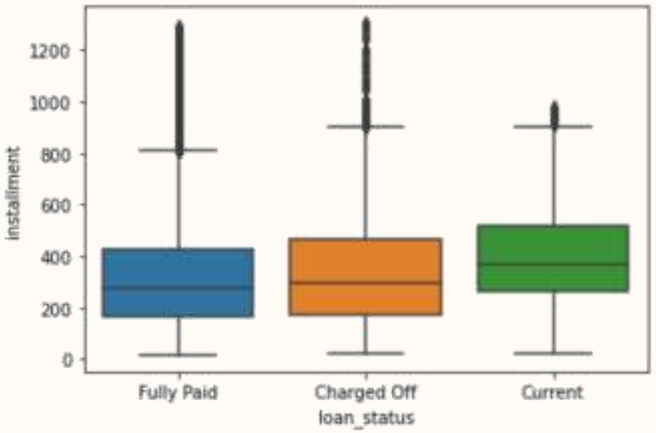
The Data was analyzed and cleaned through multiple channels including but not limited to:

1. Missing value treatment.
2. Relevancy Check.
3. Sanity Check.
4. Value conversion.
5. Appending support columns.

Missing Values
(Seaborn Heatmap)

- Missing values visualized on a heatmap to quickly infer the columns that had no values, were subsequently dropped.
- Columns with more than 70% missing values were also dropped since imputation could’ve caused an skew.

Relevancy/Impact quick check
(Boxplot)



No significant impact of installments

- A quick check on the data reflects the impact of a variable on the dependent variable i.e., loan_status.
- Columns that had no impact on the dependent variable were dropped to preserve relevant structure.

Data Deep-dive

the file provided ‘loan.csv’ had:

1. 111 Columns
2. 39,717 Rows

The Data was analyzed and cleaned through multiple channels including but not limited to:

1. Missing value treatment.
2. Relevancy Check.
3. Sanity Check.
4. Value conversion.
5. Appending support columns.

Sanity Check

```
loan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 38521 entries, 0 to 39716
Data columns (total 41 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   loan_amnt             38521 non-null  int64
 1   funded_amnt           38521 non-null  int64
 2   funded_amnt_inv       38521 non-null  float64
 3   term                  38521 non-null  object
 4   int_rate              38521 non-null  object
 5   installment           38521 non-null  float64
 6   grade                 38521 non-null  object
 7   sub_grade             38521 non-null  object
 8   emp_length            38521 non-null  object
 9   home_ownership        38521 non-null  object
 10  annual_inc            38521 non-null  float64
 11  verification_status   38521 non-null  object
 12  issue_d               38521 non-null  object
 13  loan_status           38521 non-null  object
 14  pymnt_plan            38521 non-null  object
 15  purpose               38521 non-null  object
 16  title                 38521 non-null  object
 17  zip_code              38521 non-null  object
 18  addr_state            38521 non-null  object
 19  dti                   38521 non-null  float64
```

- Quick check reveals the data types that were read wrongly during the file load.
- These data types were separately treated.

Value Conversion

```
## Converting date types from object type
loan[['issue_d','last_pymnt_d','last_credit_pull_d']] = loan[['issue_d',
loan[['issue_d','last_pymnt_d','last_credit_pull_d']] = loan[['issue_d',
```

```
[31]: ## Converting number types: integer and float
loan['term'] = loan['term'].apply(lambda x:int(x[:3]))
loan['int_rate'] = loan['int_rate'].apply((lambda x:float(x[:-1])))
loan['revol_util'] = loan['revol_util'].apply((lambda x:float(x[:-1])))
```

- The datatypes were corrected to maintain a cohesive data structure.

Data Deep-dive

the file provided 'loan.csv' had:

1. 111 Columns
2. 39,717 Rows

The Data was analyzed and cleaned through multiple channels including but not limited to:

1. Missing value treatment.
2. Relevancy Check.
3. Sanity Check.
4. Value conversion.
5. Appending support columns.

Support columns issue through function

```
loan['issue_yr'] = pd.DatetimeIndex(loan['issue_d']).year
```

```
loan['issue_yr'] = loan['issue_yr'].apply(str)
```

```
def emp_length_facet(i):
        if i == '10+ years':
            return 'Managerial'
        elif i in ['9 years','8 years','7 years'] :
            return 'Senior'
        elif i in ['4 years' , '3 years' ,'2 years' , '1 years']:
            return 'Junior'
        else:
            return 'Entry'

loan['emp_level'] = loan['emp_length'].apply(emp_length_facet)
```

- Support columns were issued to simplify the analysis.
- This helps us identify patterns and relationships between different variables that might not have been immediately obvious otherwise

ANALYSIS & INSIGHTS

Univariate

statistical analysis of a single variable, in order to understand its distribution and properties

Bivariate

statistical analysis of two variables simultaneously, in order to determine the empirical relationship between them

Univariate: Purpose & Exp Level

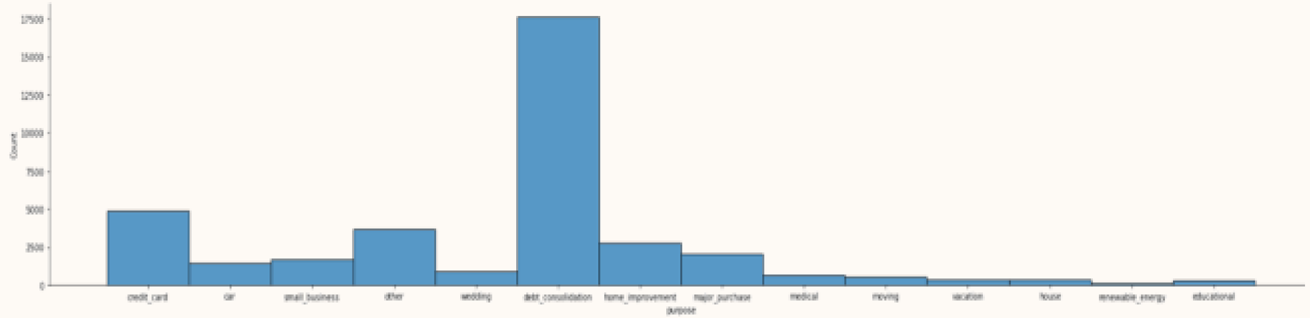
(statistical analysis of a single variable, in order to understand its distribution and properties)

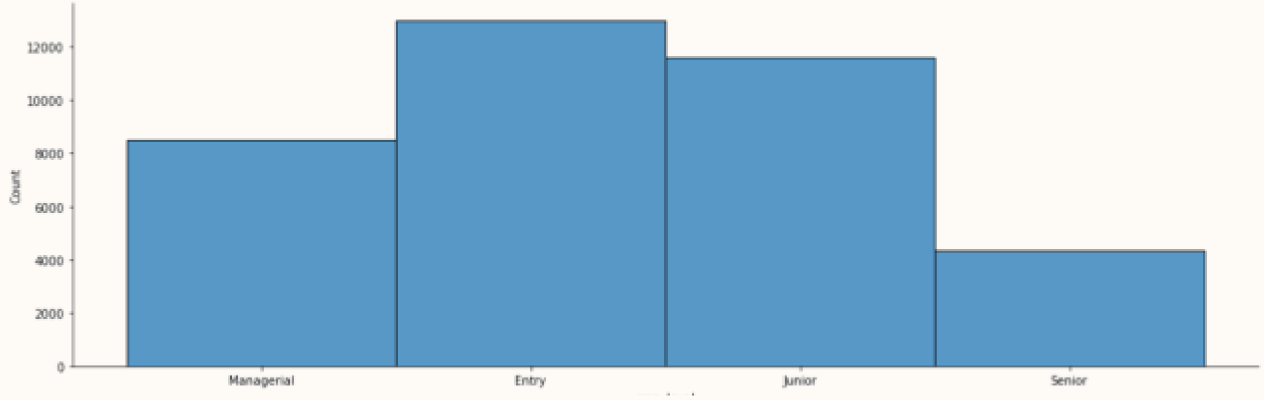
ANALYSIS

Most loans were sanctioned towards:

<u>Purpose</u>: debt consolidation

<u>Experience Level</u>: Entry (0 Yrs exp)





Univariate: Verification & Loan Issue

ANALYSIS

verification_status	funded_amnt_inv
Not Verified	0.319453
Source Verified	0.234964
Verified	0.445583

32%

Of all loans have been sanctioned to unverified sources, which is alarming

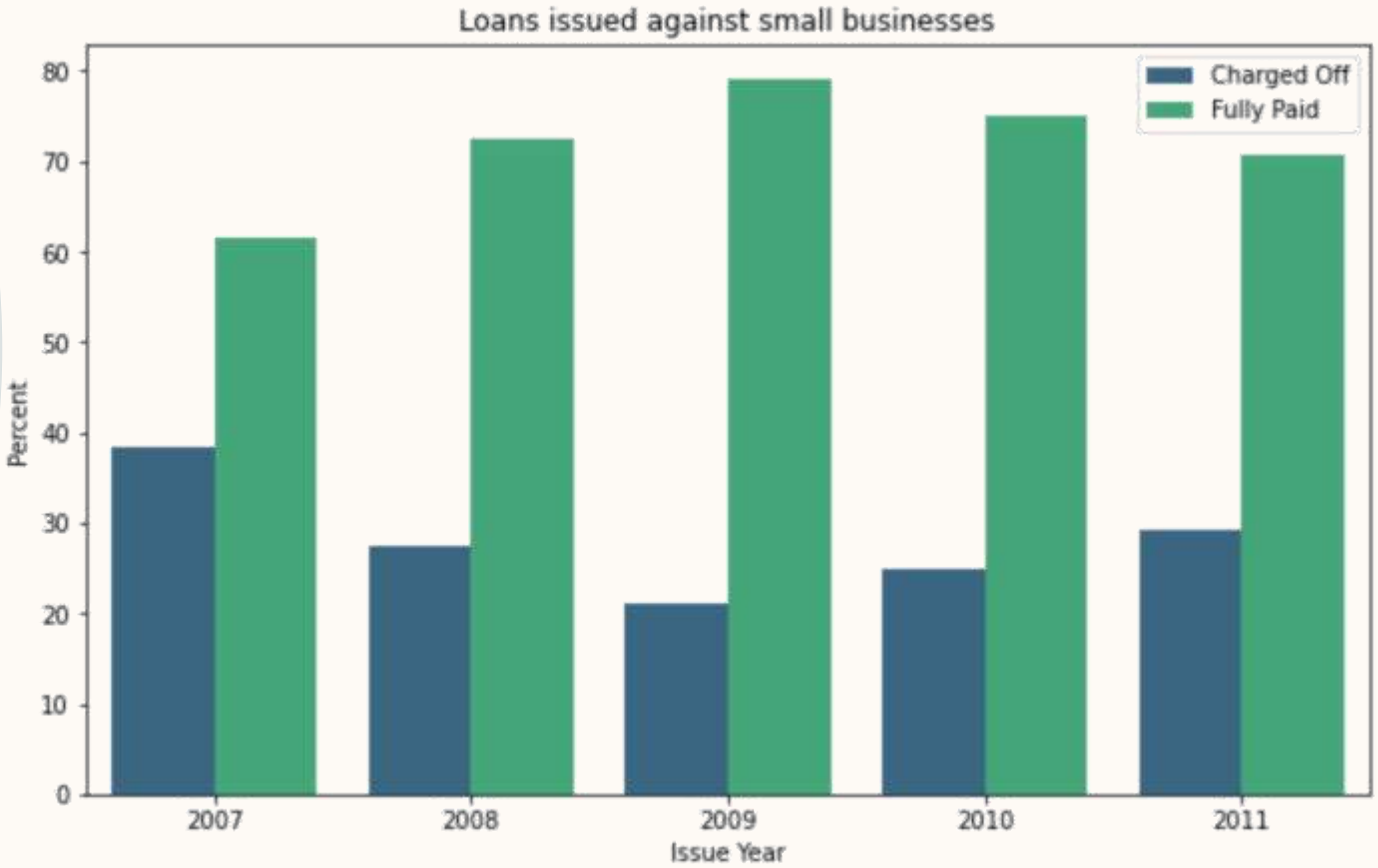
issue_yr	funded_amnt_inv
2007	0.000795
2008	0.013034
2009	0.106195
2010	0.288505
2011	0.591471

59%

Of all loans have been sanctioned just in 2011. Indication of a rise in loan demand.

Univariate: Most frequent default Purpose

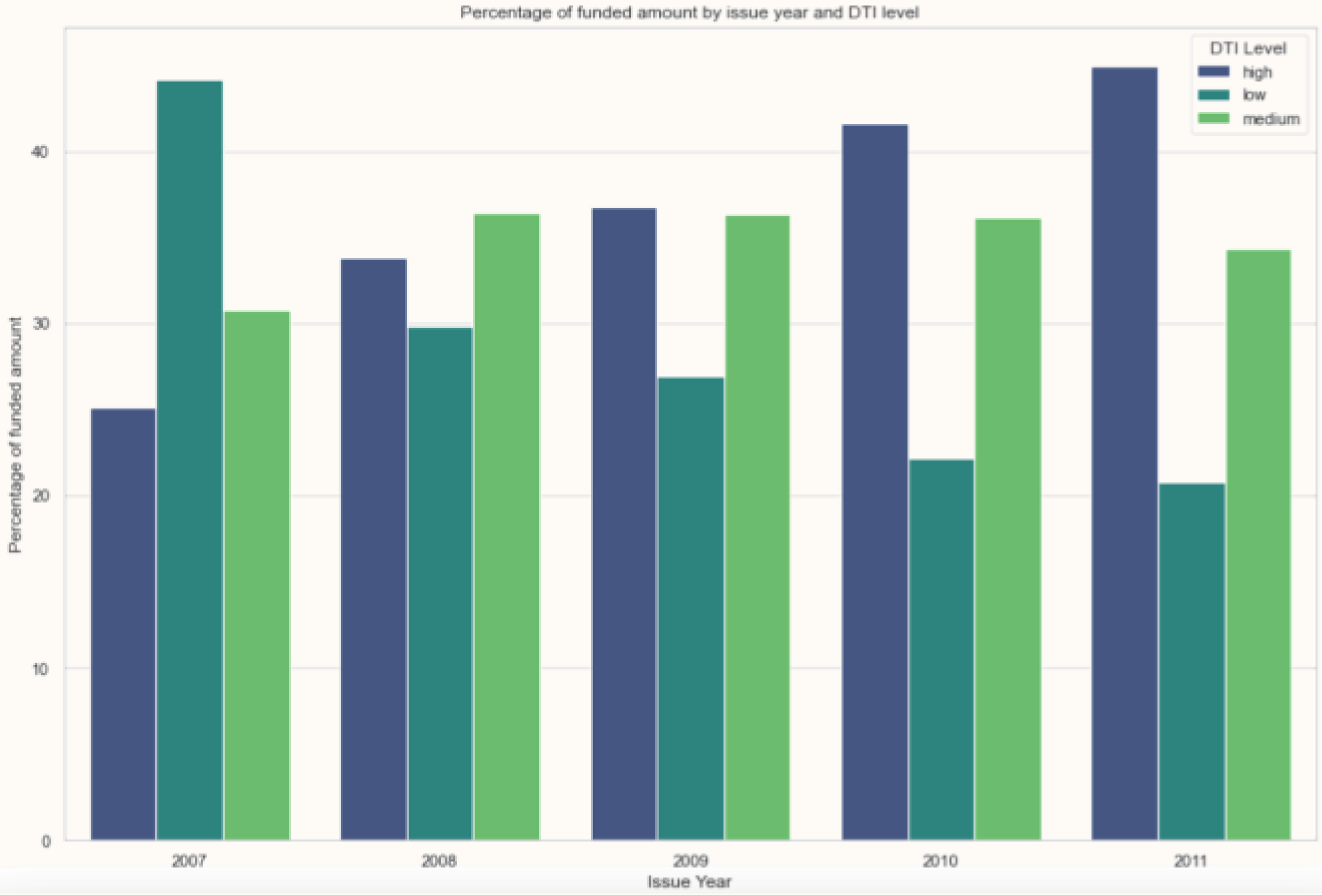
ANALYSIS



Small Business have been defaulting more frequently

Univariate: DTI levels

ANALYSIS



High DTI ratio personnel have been funded loans more frequently through the years, which is concerning.

Univariate: in Conclusion

INSIGHTS

Loan sanctions have gotten aggressive through the year with an unprecedented rise in demand

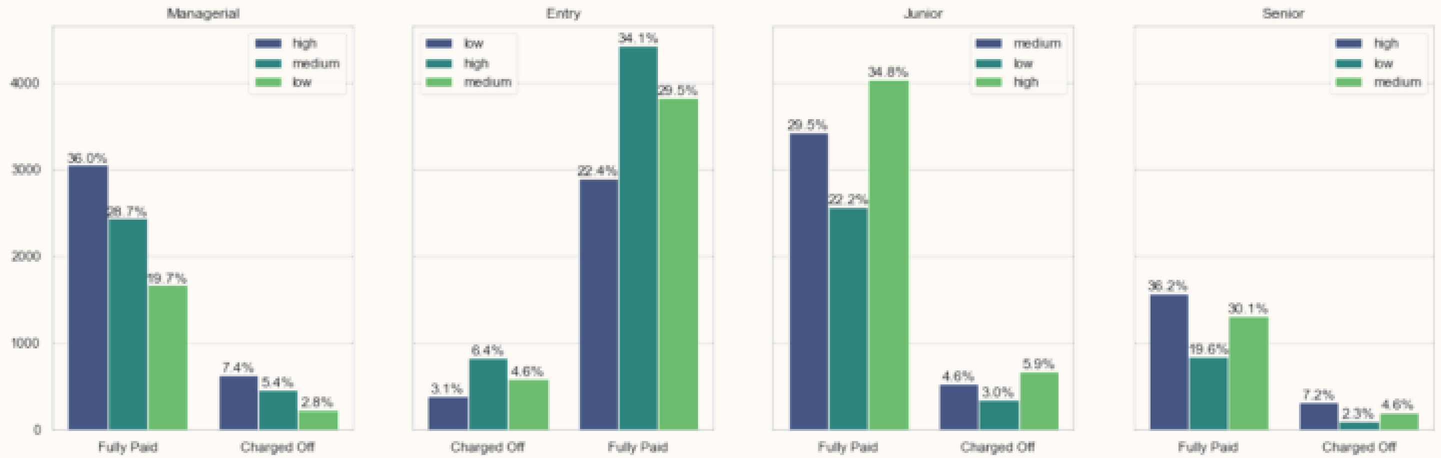
To cater to a high demand, basic audit measures such as source verification and DTI levels are often overlooked.

Small Businesses have been the highest loan defaulter when it comes to closing. There needs to be a specific measure in place to monitor small businesses closely.

ANALYSIS

Bivariate: Exp Level V/s DTI* against loan status

(statistical analysis of two variables simultaneously, in order to determine the empirical relationship between them)

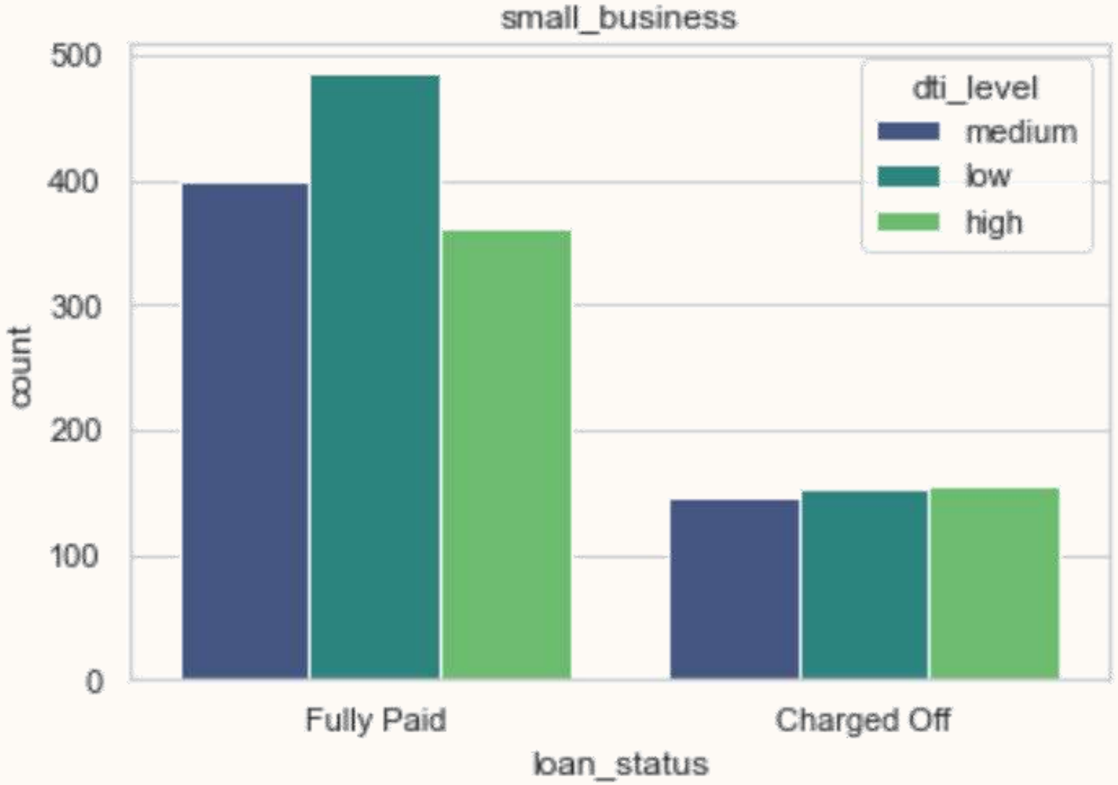


Managerial levels: at High DTI are more likely to default.
Senior levels: at High DTI are more likely to default.
Junior levels: at Medium & High DTI are more likely to default.
Entry levels: at High DTI are more likely to default.

DTI stands for Debt-to-Income ratio, which is a measure of a borrower's monthly debt payment divided by their gross monthly income, used to assess their ability to manage their current and future debt payments. Lower DTI the better

ANALYSIS

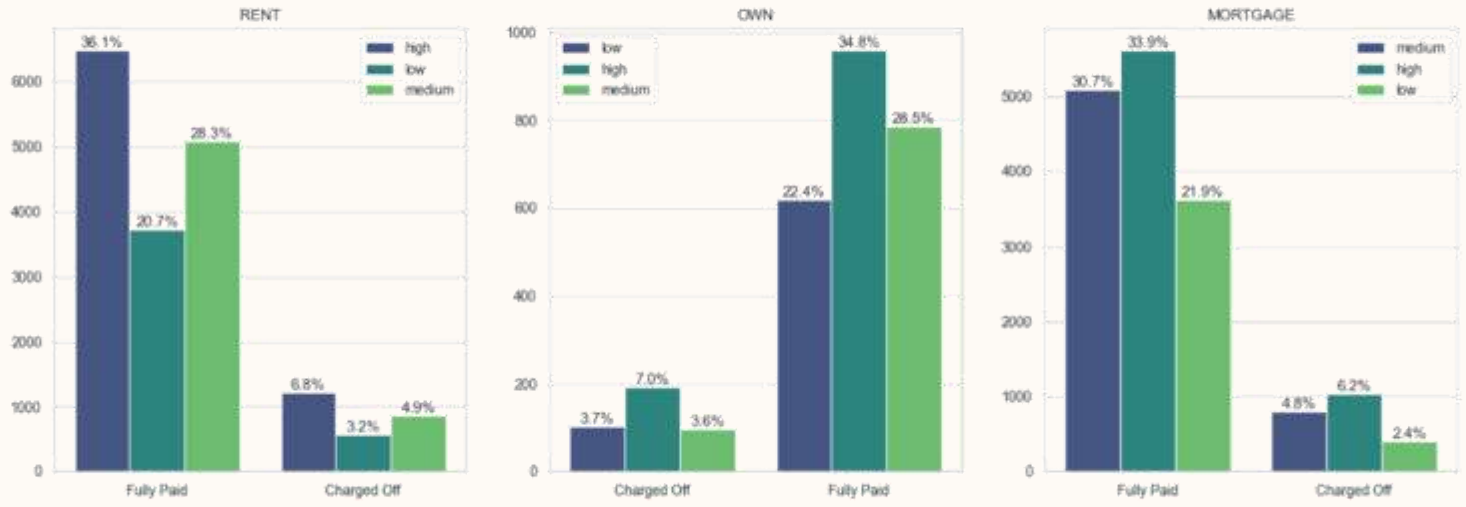
Bivariate: Loan Purpose with DTI against Loan Status



Small business loans even at any DTI are defaulted at same frequency

Bivariate: House Ownership with DTI against Loan Status

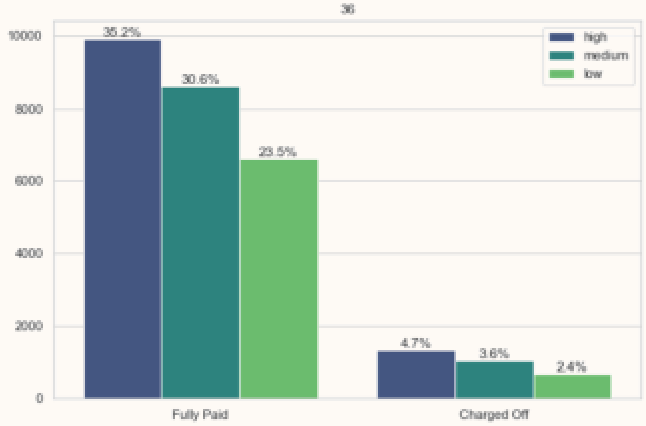
ANALYSIS

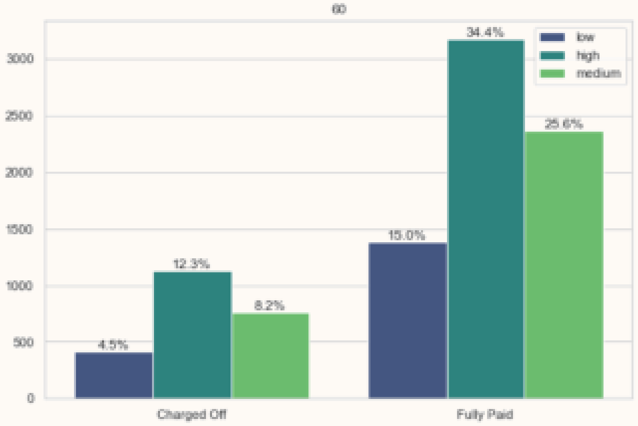


Rent ownership, even at low DTI possesses a default risk

Bivariate: Loan terms with DTI against Loan Status

ANALYSIS





Longer term loans (60 months) are sanctioned even at high DTI resulting in higher defaults

Bivariate: in Conclusion

INSIGHTS

Junior level experience have had a higher risk compared to other profiles considering the DTI levels.

Small businesses have been at a much higher risk when compared through their DTI ratio levels.

Personnel who have rent ownership are more at risk when compared to those who have own house either on Mortgage or Full.

Long term loans are at high risk of default because there's simply more time for things to go wrong over the life of the loan, including economic downturns, changes in borrower financial circumstances, and other unforeseeable events

SUMMARY

- Loan sanctions have increased in recent years to meet growing demand.
- In order to meet this demand, some basic audit measures have been overlooked in some cases.
- Small businesses have been among the highest defaulters, and there is a need to monitor them more closely.
- Junior level employees have a higher risk of default than other profiles, due to higher DTI levels.
- Small businesses with high DTI ratios have a particularly high default risk.
- People who rent their homes are at greater risk of default than people who own homes, with mortgage or freehold.
- Long-term loans carry greater default risk than short-term loans, due to the greater number of risks that can arise over the life of the loan.

Finally, Tighter audit measures and closer scrutiny of high-risk borrowers, especially small businesses and renters, may be necessary to mitigate the risks associated with recent trends in aggressive loan sanctions and rising default rates over long-term loan periods.

THANK YOU

Abhinav Sinha
Sachin Sachdeva