Lending Club Casestudy

ANUSHA CHAUDHURI

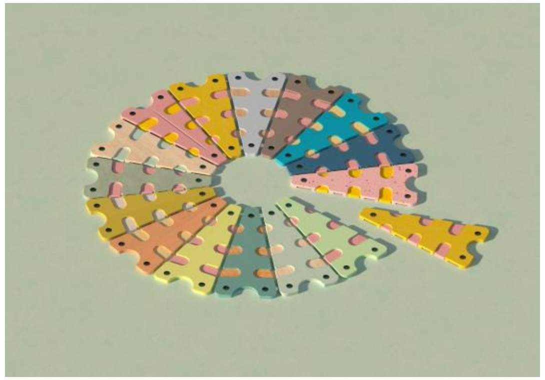
BRAMHANAYAGHE ARUMUGAM

Company Context

Largest online marketplace
Facilitates personal loans, business loans, and medical procedure financing
Offers lower interest rates through fast online interface

The Challenge

- Significant credit loss due to loan defaults
- 'Charged-off' customers are primary defaulters
- Need to identify and mitigate risky loans



Our Objectives

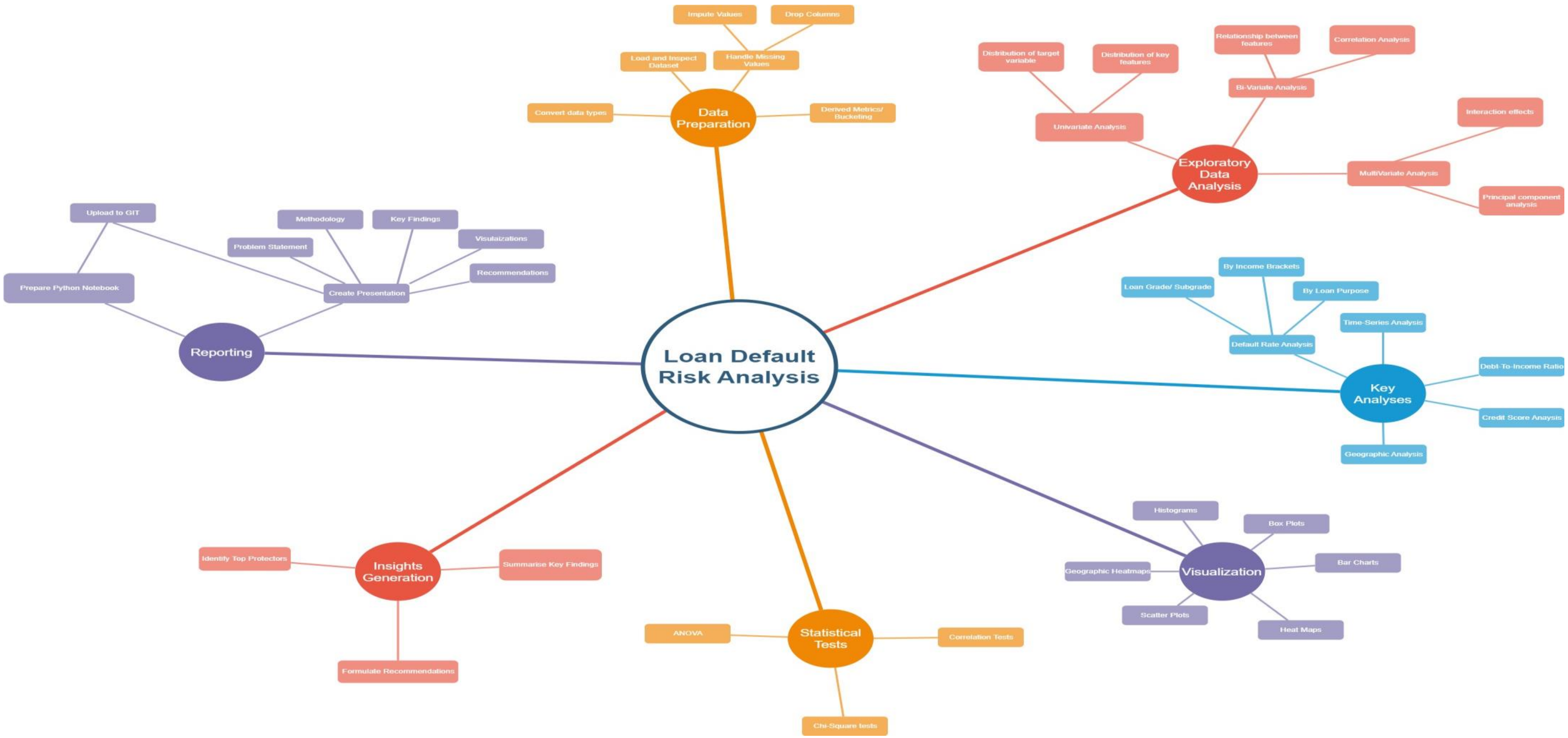
- Identify key driving factors behind loan defaults
- Understand variables that strongly indicate default risk
- Develop insights to improve portfolio and risk assessment

Expected Outcome

- Reduced credit loss through targeted risk management
- Enhanced ability to identify high-risk loan applicants

"Turning Data into Actionable Insights for Smarter Lending"

Approach Overview



Data Preparation

Load Data

```
df=pd.read_csv('loan.csv')

df.head(10)
```

	id	member_id	loan_amnt	funded_amnt	funded_amnt_inv	term	int_rate	installment	grade	sub_grade	...	num_tl_90g_dpd_24m	num_tl_op_p
0	1077501	1296599	5000	5000	4975.0	36 months	10.65%	162.87	B	B2	...	NaN	
1	1077430	1314167	2500	2500	2500.0	60 months	15.27%	59.83	C	C4	...	NaN	
2	1077175	1313524	2400	2400	2400.0	36 months	15.96%	84.33	C	C5	...	NaN	
3	1076863	1277178	10000	10000	10000.0	36 months	13.49%	339.31	C	C1	...	NaN	
4	1075358	1311748	3000	3000	3000.0	60 months	12.69%	67.79	B	B5	...	NaN	

Check Null Percentages

```
print(df.isnull().sum()/len(df)*100)
```

```
id                               0.000000
member_id                        0.000000
loan_amnt                        0.000000
funded_amnt                      0.000000
funded_amnt_inv                  0.000000
                                   ...
tax_liens                        0.098195
tot_hi_cred_lim                100.000000
total_bal_ex_mort              100.000000
total_bc_limit                 100.000000
total_il_high_credit_limit     100.000000
Length: 111, dtype: float64
```

Data Preparation

Find High Null Columns

```
Columns with null proportion > 40% :
 ['mths_since_last_delinq', 'mths_since_last_record', 'next_pymnt_d', 'mths_since_last_major_derog', 'annual_inc_joint', 'dti_joint', 'verificatio
Columns with null proportion <= 40% :
 ['id', 'member_id', 'loan_amnt', 'funded_amnt', 'funded_amnt_inv', 'term', 'int_rate', 'installment', 'grade', 'sub_grade', 'emp_title', 'emp_len
```

```
null_pctg=df.isnull().sum()/len(df)*100

high_null_cols=null_pctg[null_pctg>40].index.tolist()

low_null_cols=null_pctg[null_pctg<=40].index.tolist()
```

Removing High Null Columns

```
df_cleaned = df.drop(columns=high_null_cols)

print(f"Dropped {len(high_null_cols)} columns. New shape: {df_cleaned.shape}")

print("Dropped columns:", high_null_cols)

df = df_cleaned
```

Data Preparation

Changing Datatypes

```
# converting int_rate to float

df['int_rate']=df['int_rate'].str.rstrip('%').astype('float')

print(df['int_rate'].dtype)

print(df['int_rate'].head(10))

# converting revol_util to float

df['revol_util']=df['revol_util'].str.rstrip('%').astype(float)

print(df['revol_util'].dtype)

print(df['revol_util'].head(10))
```

Data Preparation

Cleanup redundancy

```
# dropping rows with loan_status 'Current'

df=df[df['loan_status']!='Current']

df.shape
```

Remove Bias

```
df['emp_title']=df['emp_title'].fillna('Unknown')
```

Imputing

```
# imputing last_pymnt_d with 'Not paid'

df['last_pymnt_d']=df['last_pymnt_d'].fillna('Not paid')
```

Data Preparation

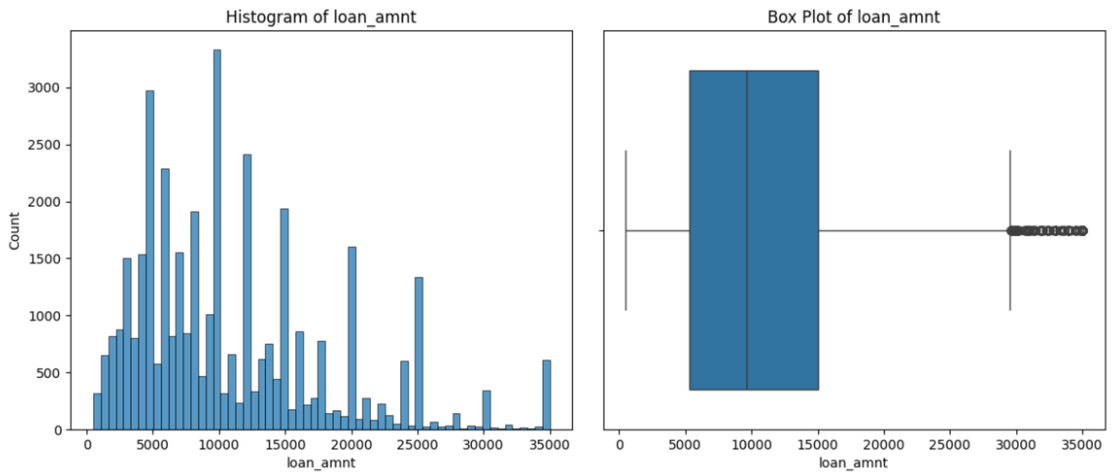
Derived / Feature Extraction- Determining credit history length of applicants

```
# Calculate the credit score using the normalized columns

df['credit_score'] = (

    (1 - df['delinq_2yrs_norm']) * 0.35 +    # Invert since lower is better

    (1 - df['pub_rec_bankruptcies_norm']) * 0.35 +  # Invert

    (1 - df['total_rec_late_fee_norm']) * 0.35 +  # Invert

    (1 - df['collection_recovery_fee_norm']) * 0.35 +  # Invert

    (1 - df['revol_util_norm']) * 0.30 +      # Invert since lower is better

    (df['credit_history_length_norm'] * 0.15) +  # Keep as is

    (df['total_acc_norm'] * 0.10)              # Keep as is

)
```

Exploratory Data Analysis

Nature of Dataset

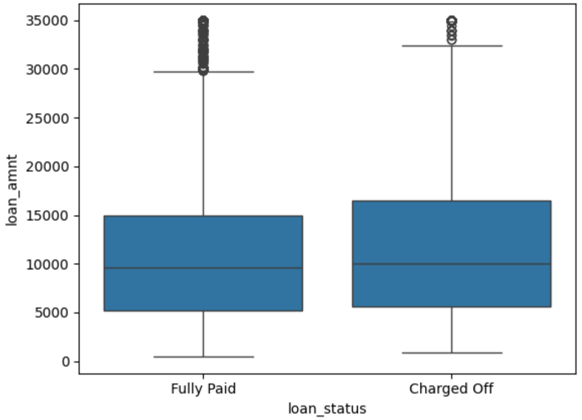
```
# Histogram
sns.histplot(data=df, x=col, ax=axes[i, 0])
axes[i, 0].set_title(f'Histogram of {col}')

# Box plot
sns.boxplot(data=df, x=col, ax=axes[i, 1])
axes[i, 1].set_title(f'Box Plot of {col}')
```



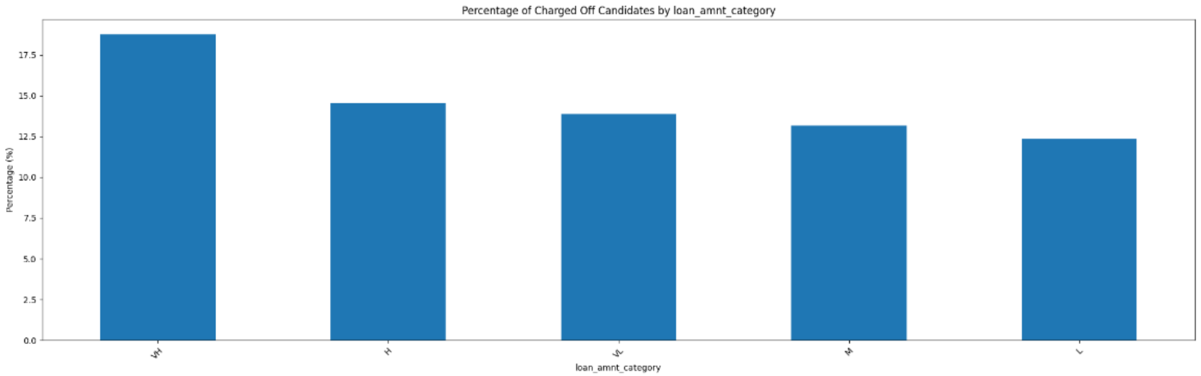
Exploratory Data Analysis

Percentage of Charged Off Loans

```python
# Calculate the percentage of charged off loans for each colu
charged_off_percentage = (
    df.groupby(col)['loan_status']
    .value_counts(normalize=True)
    .unstack()
    .fillna(0) * 100
)
```



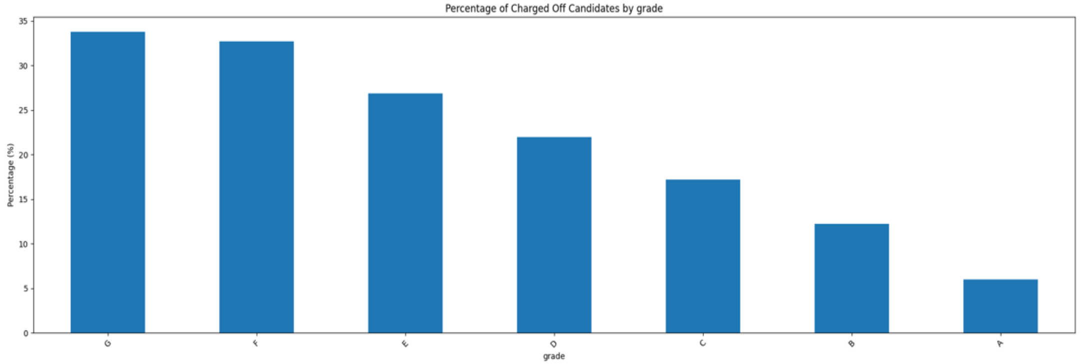
“Higher charged off cases are observed for higher loan amounts”



Exploratory Data Analysis

Grade vs ChargeOff

```
fig, axes = plt.subplots(num_cols, 1, figsize=(10, 5 * num_cols))

for i, col in enumerate(filtered_columns):

    sns.countplot(data=df, x=col,
ax=axes[i],order=df[col].value_counts().sort_values(ascending=False).index)

    axes[i].set_title(f'Count Plot of {col}')

    axes[i].tick_params(axis='x', rotation=90)

plt.tight_layout()

plt.show()
```



"Charge off rate increases as loan grade increases from A to G"

Percentage of Charged Off Candidates by sub_grade

Key Analyses

Bivariate Analysis - Comparisons

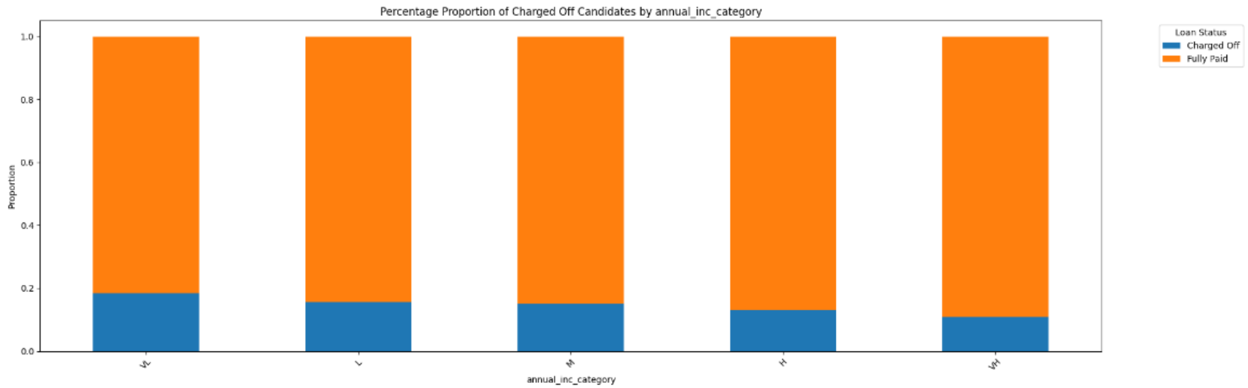
```
# Loop through each input column
for i, col in enumerate(input_columns):
    counts = df.groupby([col, 'loan_status']).size().unstack(fill_value=0)
    # Calculate the proportion of each category
    proportions = counts.div(counts.sum(axis=1), axis=0)
```

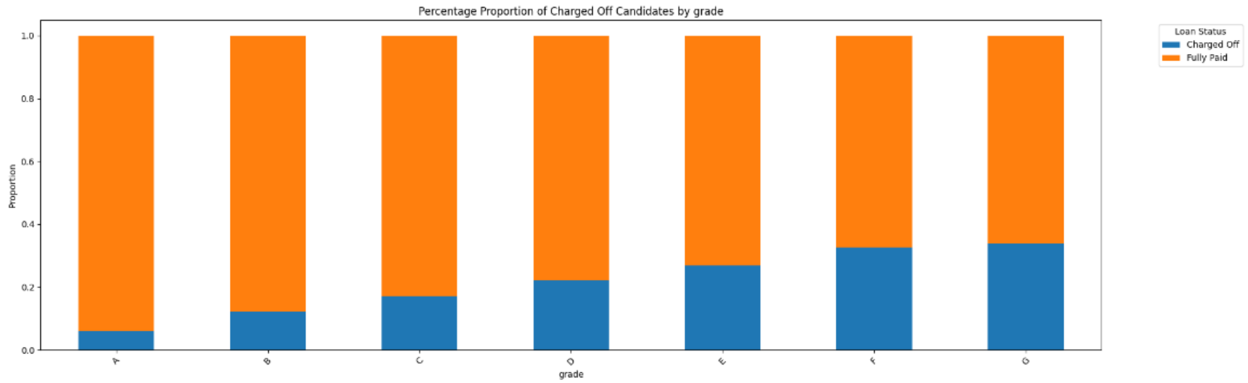
“Lower the income higher is the charge off rate”

“To reduce charge off, lower grade loans preferably

grade A need to be prioritized” (ref illustrations on next slide)

Key Analyses





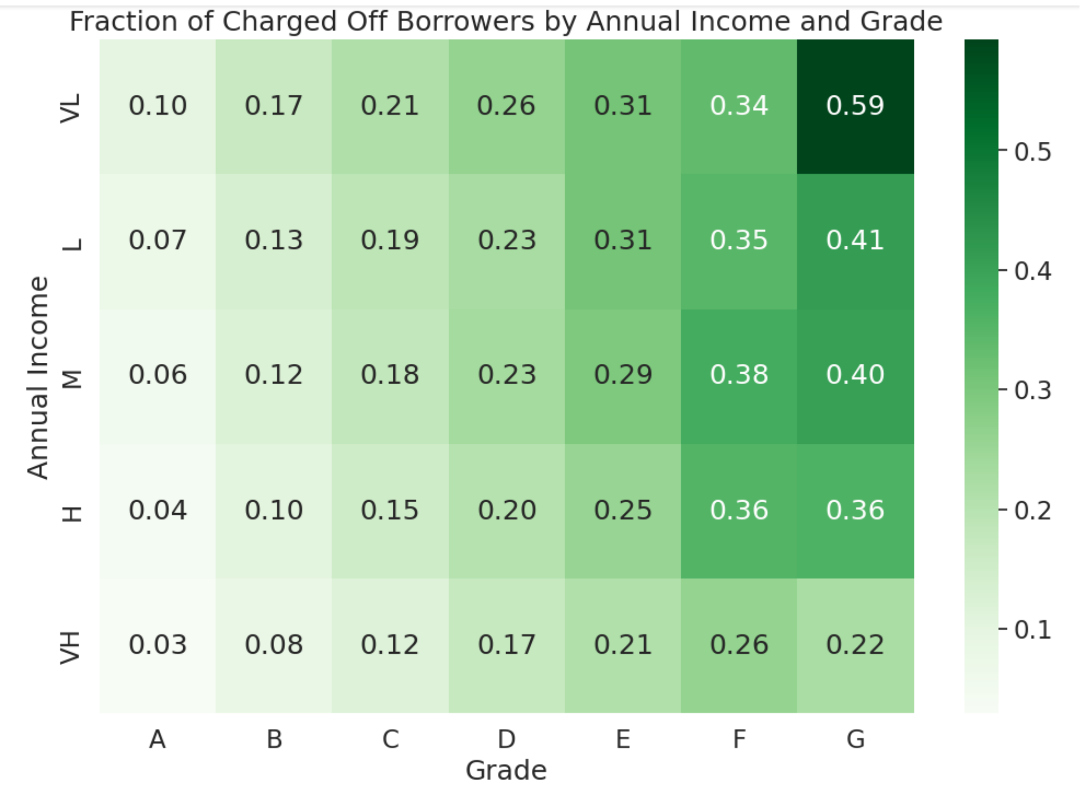
Key Analyses

Multivariate Analysis – Across Income groups and Grades

```
pivot_table = df.pivot_table(
    index='annual_inc_category',
    columns='grade',
    values='loan_status',
    aggfunc=lambda x: (x == 'Charged Off').sum() / len(x)  # Calculate the
fraction of charged off
)
```

“Lower income groups should be provided lower grade Loans (A,B), vis-à-vis Lower interest rates” (ref illustrations on next slide)

Key Analyses



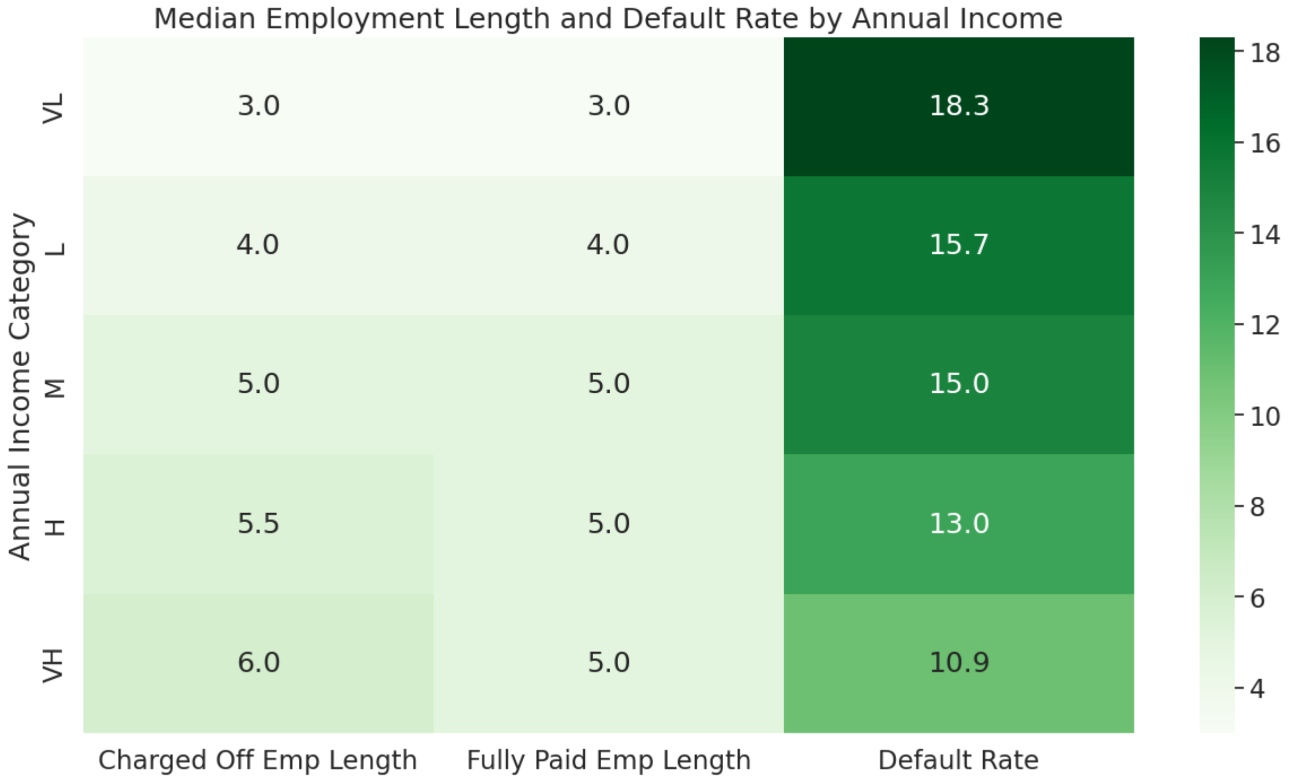
Key Analyses

Median employment length for different combinations of loan status and annual income `# Calculate default rate`

```
default_rates = (charged_off_counts / loan_counts) * 100

emp_length_pivot = df.pivot_table(

    index='annual_inc_category',

    columns='loan_status',

    values='emp_length',

    aggfunc='median'

)
```

“The difference in default rates between the lowest and highest income categories is significant (18.3% vs 10.9%), suggesting that income-based risk assessment could be valuable” (ref illustrations on next slide)

Key Analyses



Statistical Test

- **Interest Rate (int_rate)**F-statistic: **75,229.76**
- p-value: **0.000**

- **Debt-to-Income Ratio (dti)**F-statistic: **85.88**
- p-value: **2.31e-107**

ANOVA Summary

- **Revolving Utilization (revol_util)**F-statistic: **1,788.70**
- p-value: **0.000**

- **Recoveries**F-statistic: **118.86**
- p-value: **2.18e-149**

- **Total Rec. Late Fee (total_rec_late_fee)**F-statistic: **70.33**
- p-value: **1.60e-87**

- **Collection Recovery Fee**F-statistic: **38.86**
- p-value: **2.25e-47**

Insights and Conclusions

Loan Purpose
Small businesses charge –off more, requiring continuous line off credit, prioritize small loans at low interest rates

Loan Amounts
Limit high loan amounts to $17K

Interest Rates
Low interest rates are better.
Optimal rates: 5.4% – 11%..

Debt-to-Income (DTI) & Revolving Utilization
Monitor DTI > 15% and utilization > 45%.

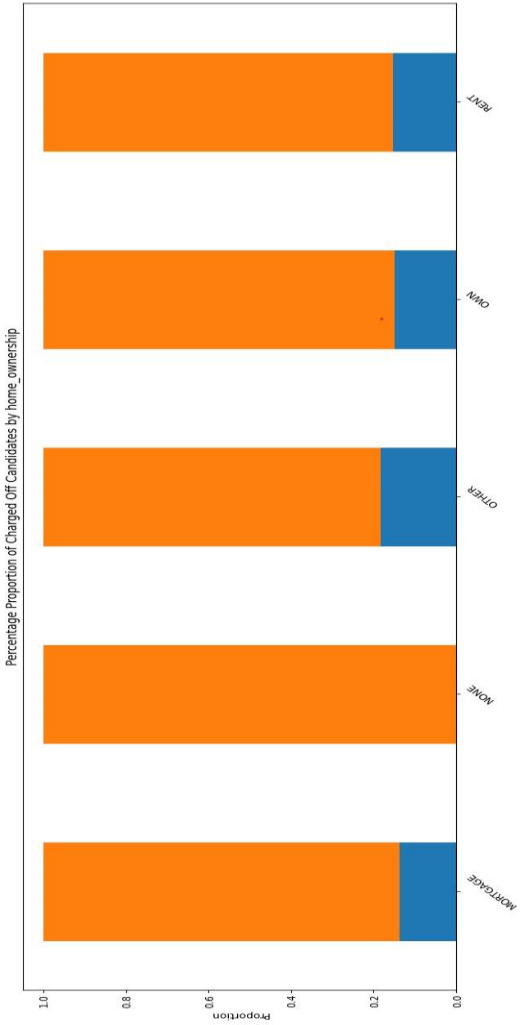
Loan Grades
Prioritize lower-grade loans

Income & Loan Term
Short-term (36-month) loans preferred.

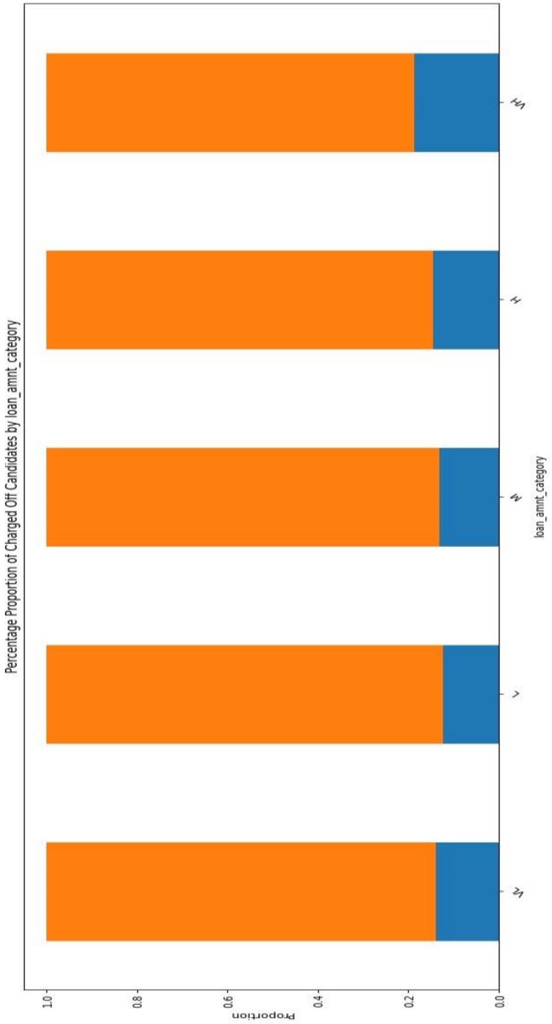
Credit Scores
Flag credit scores < 0.8

Geography & Employment
Nebraska & Walmart employees' high risk

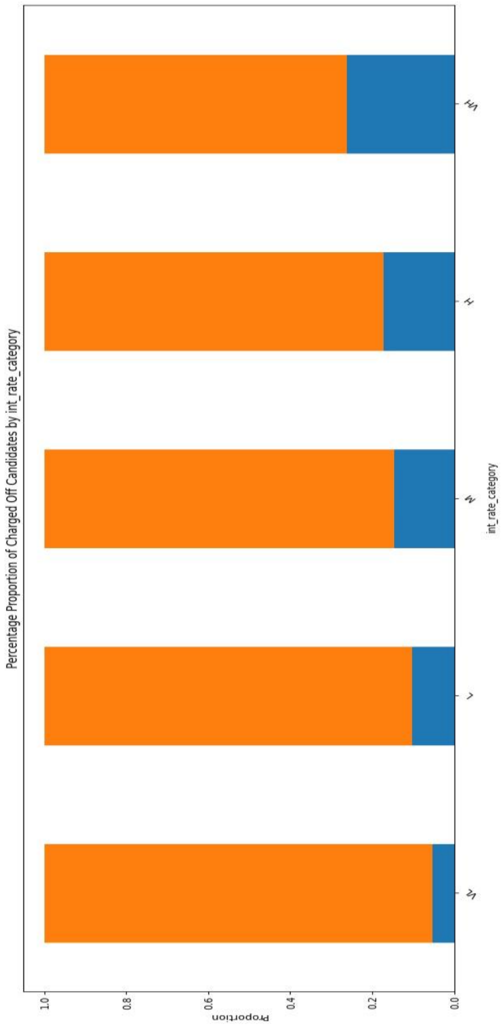
Insights and Conclusions



Loan Purpose



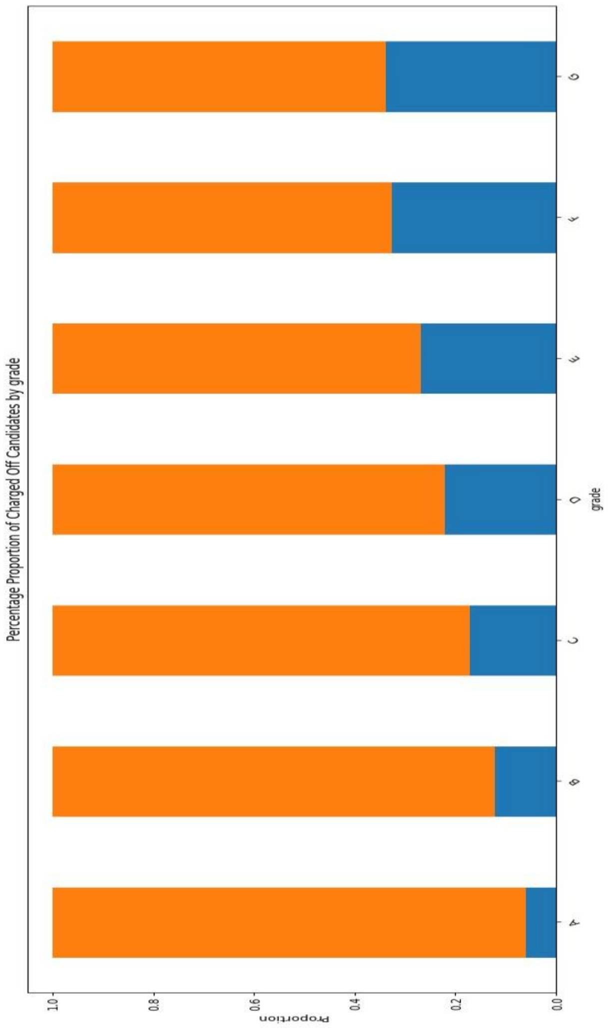
Loan Amount



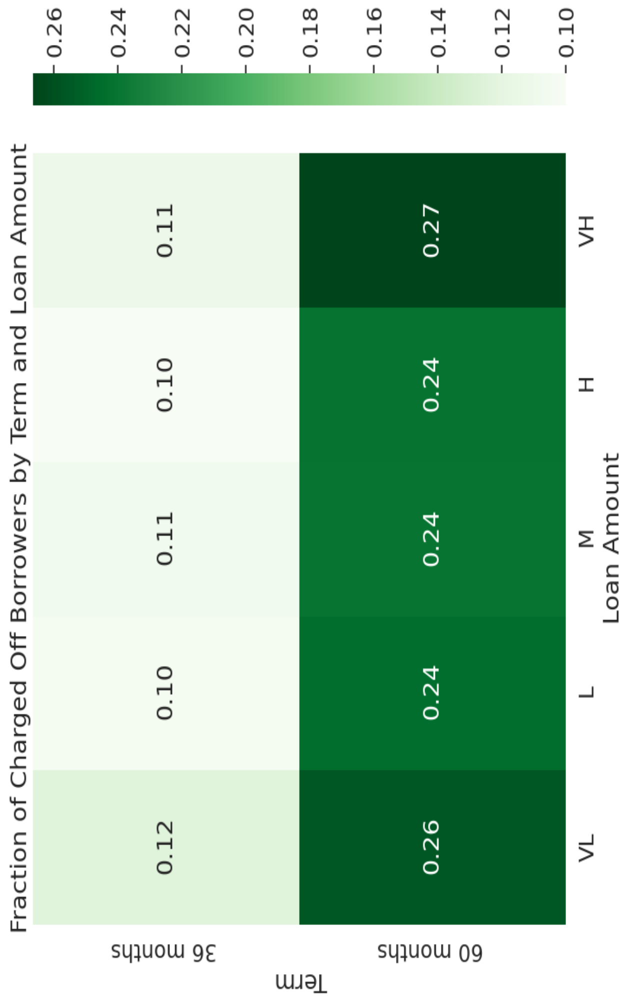
Interest Rate

DTI/ Revolving Utilization

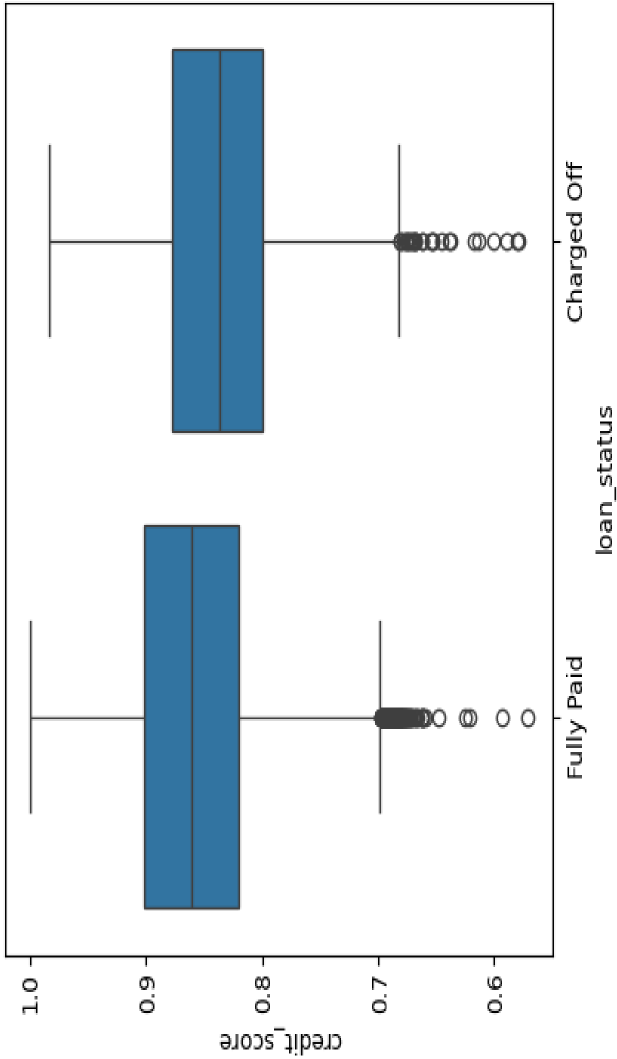
Insights and Conclusions



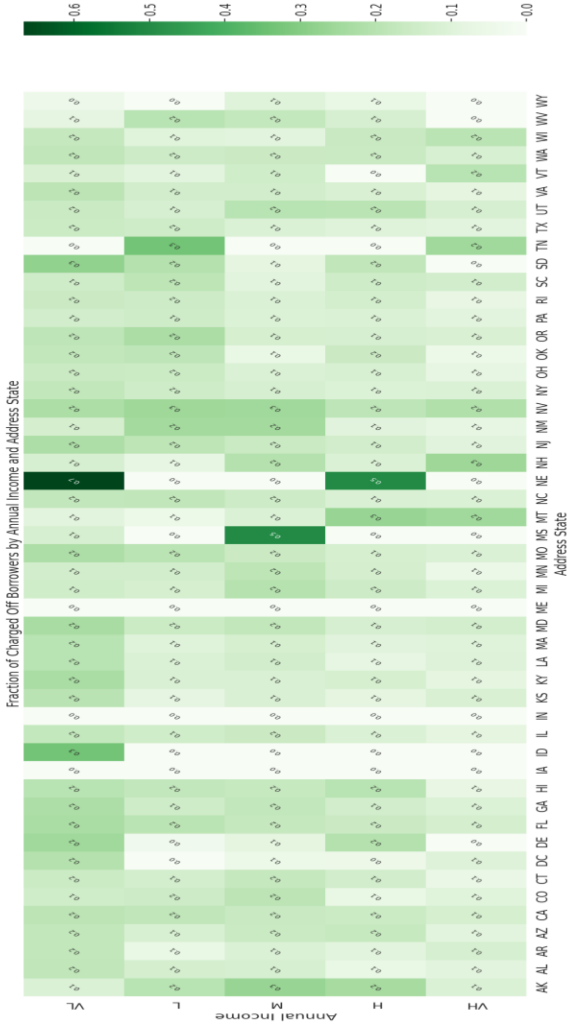
Loan Grade



Income and Loan Term



Credit Scores



Geography

End