



Lending Club Case Study

- Ashok Vashist &
- Bineesh Panangat

Agenda

- Business Objective
- Technical Solution
- Approach
- Analysis
- Conclusion

Business Objectives



Understand the Driving factors behind Loan Defaults



Identify Risky Loan Applicants



Reduce the Credit Loss



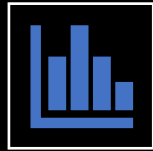
Exhaustive Data Driven Analysis for Decision Making

Technical Solution

We would be leveraging Exploratory Data Analysis (EDA) technique often used by Data Scientist to analyze and investigate datasets and summarize their main characteristics using Data Visualization Methods. Few techniques used are as follows



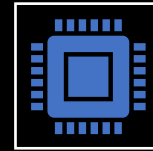
Data Cleaning –
imputation, managing
missing statistics, and
removing outliers



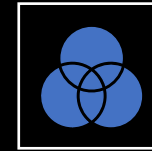
Descriptive statistics
Measures like suggest,
median, mode, preferred
deviation, range, and
percentiles are usually
used



Data Visualization
histograms, box plots,
scatter plots, line plots,
heatmaps, and bar charts



Feature Engineering
Normalization, binning,
encoding, etc



Correlation and
Relationships -



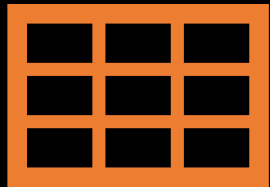
Data Segmentation -
Dividing the information
into significant segments
based totally on sure
standards or traits



Approach

- Data Quality Steps
 - Data Cleanup
 - Dropping Rows
 - Find Duplicate Rows and Drop them
 - Remove header, footer or summary rows if any
 - Drop rows with No Column Values
 - Rows with high percentage of missing values
 - Dropping Columns
 - Drop columns that do not contribute to analysis such
 - Long text (Eg ,emp_title)
 - Redundant columns such as Id and URL
 - Removing Nulls
 - With Multiple Zero Values
 - With NA Values

Approach

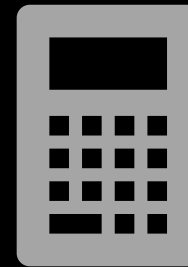


Data Manipulation

Strip % from int_rate column and convert into float

pub_rec_bankruptcies fillna = 0

Outlier removal Eg . Annual_inc



Introduce Additional columns

Extract Month ,Year & quarter from issue_d
column

Analysis on critical columns

Employment Length(emp_length) – Higher the tenure more the financial stability and greater chances of loan payment – But As per given data this is quite opposite

Annual Income (annual_inc) - Annual income of the customer, higher the income, more chances of loan pass

State (addr_state) - There could be higher delinquency or defaulters demographical

Debt to Income (dti) - The percentage of the salary which goes towards paying loan. Lower DTI, higher the chances of a loan pass

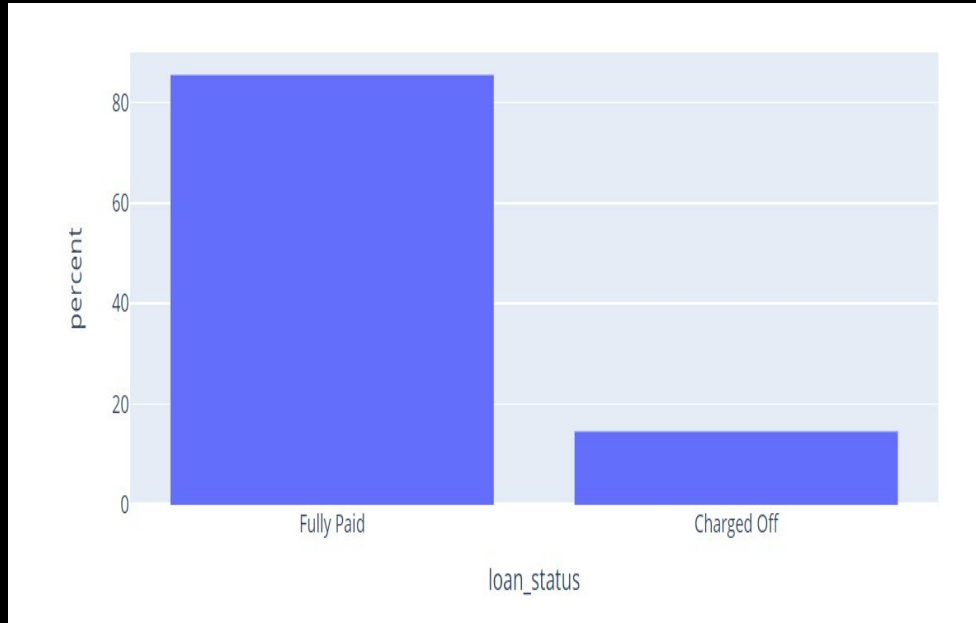
Home Ownership (home_ownership) - Owning a home adds a collateral which increases the chances of loan payment , As per Data even Mortgage category was defaulting less



Visualization

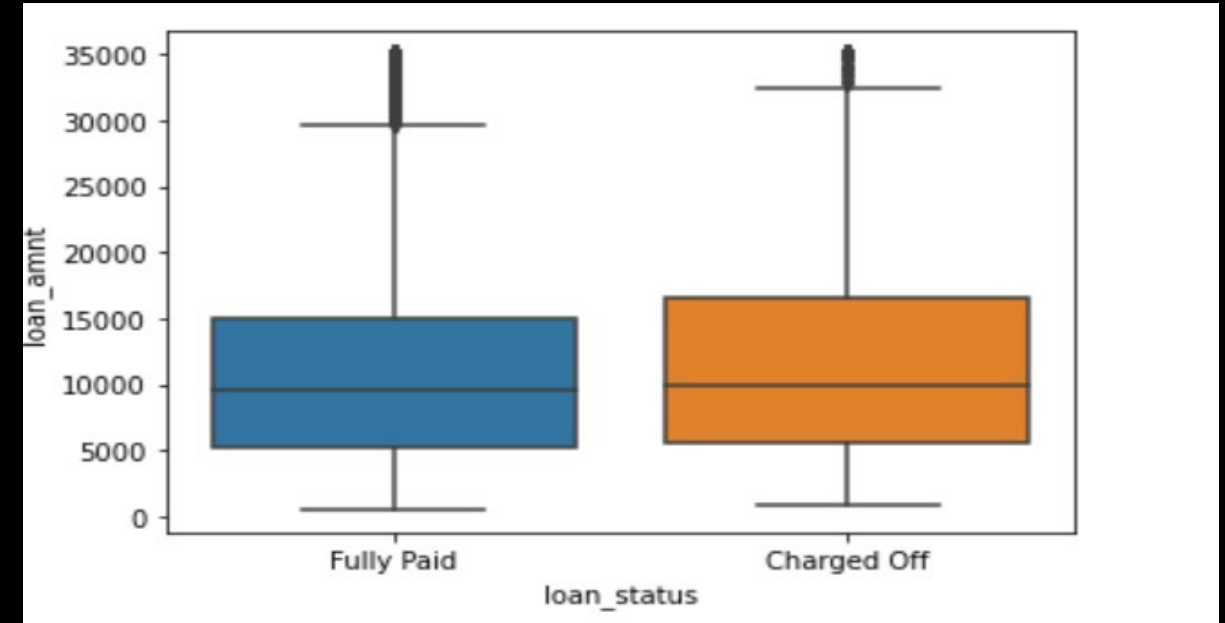


Loan Status and Amount



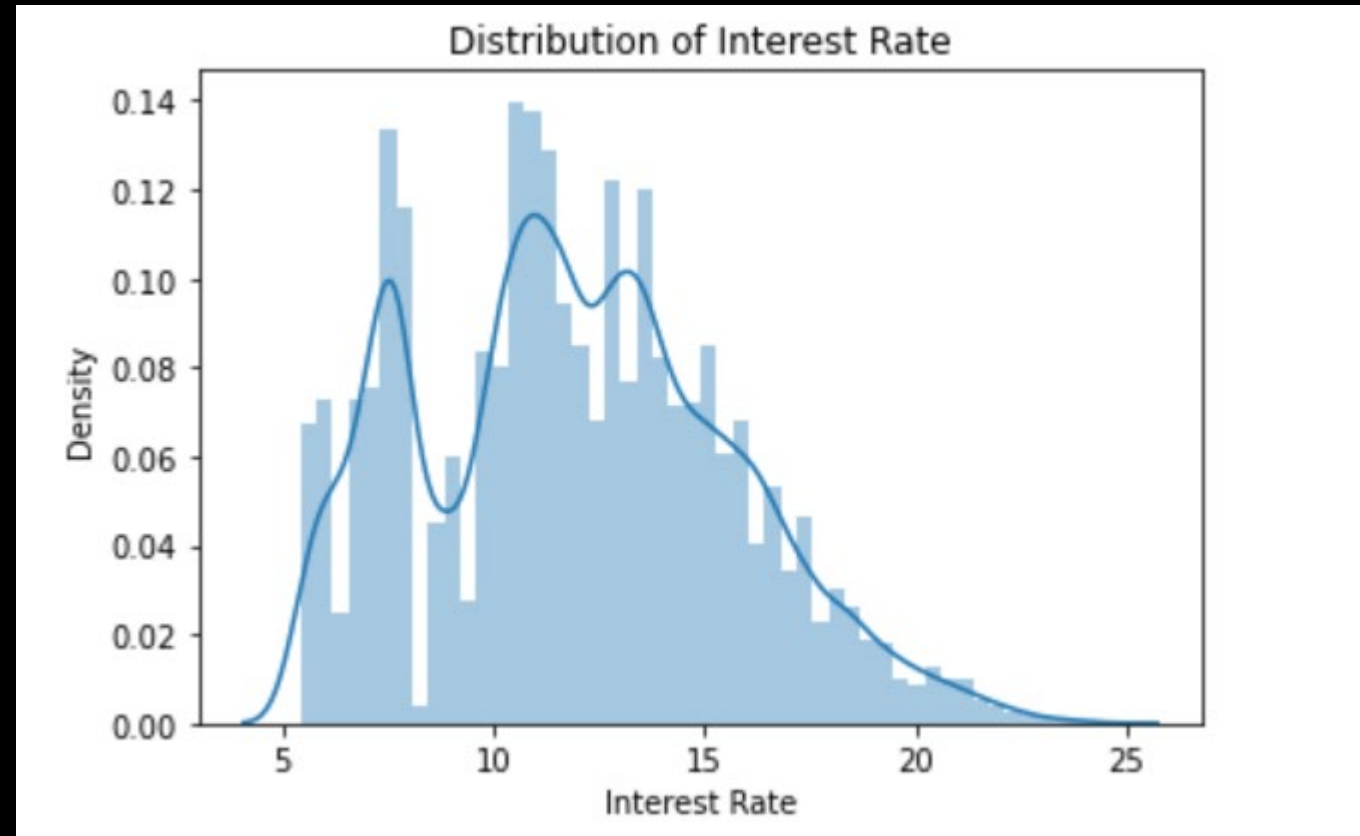
- **Loan Status:**

- * 83 % of the loans were fully paid
- * 14 % were defaulted
- * 3 % is current loan



- **Loan Amount:** It varies from 5000 to 30000 with a median of 10000. Loan amount is majorly small and very few clients have taken large loans and larger it goes we have higher chance of defaulting.

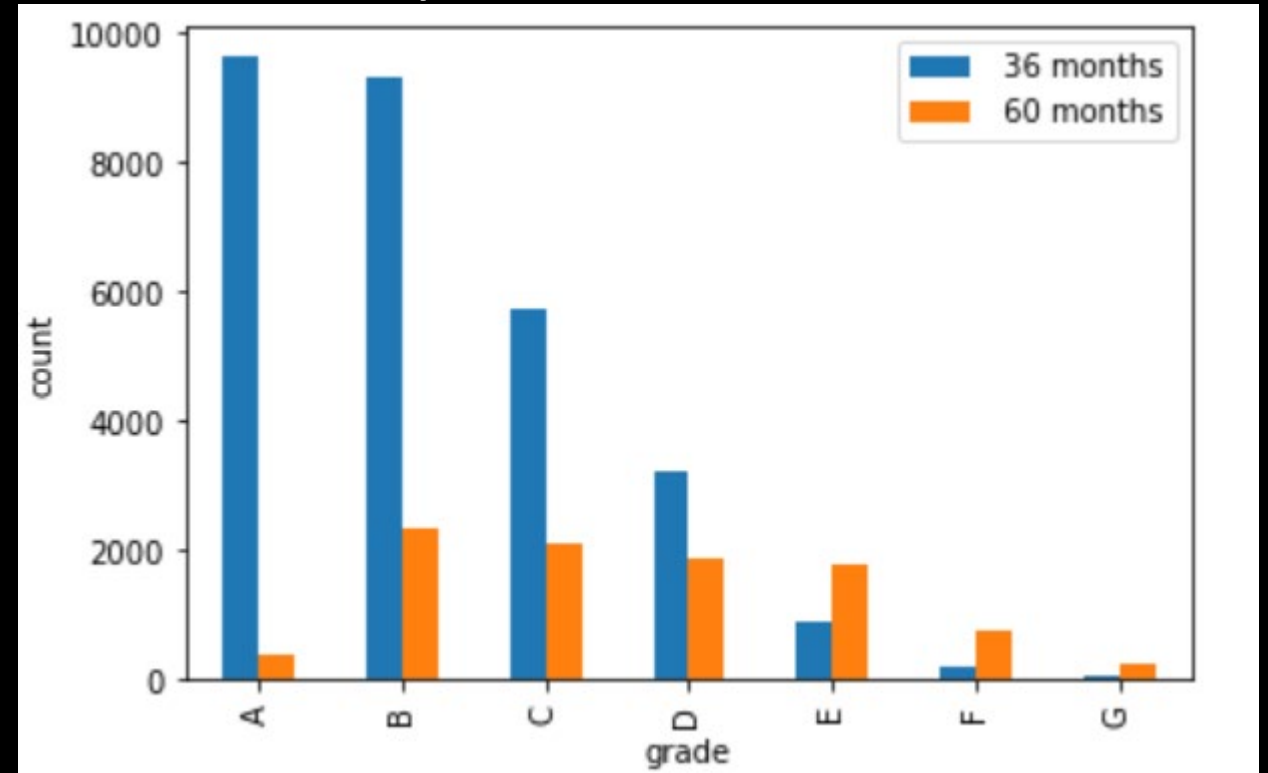
Term and Interest Rate



- **Interest Rate:** The chance of defaulting increases with interest rate.

Distribution terms for Grades (Business Driven Matrix)

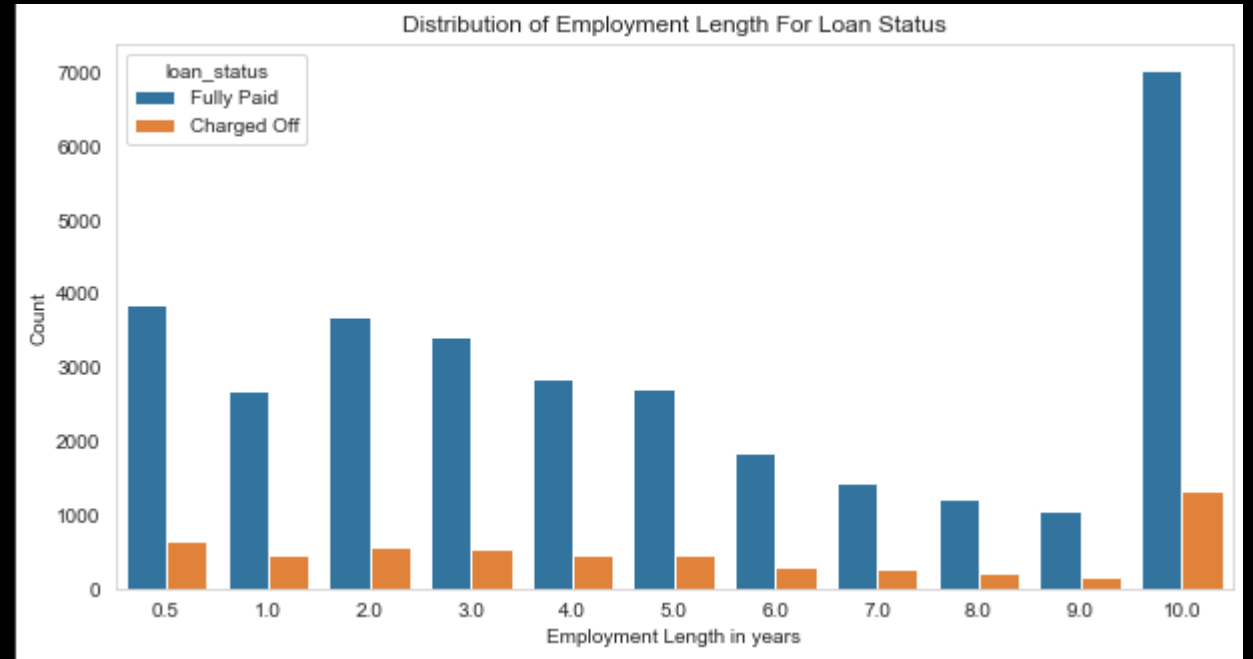
loan_status	Charged Off	Fully Paid
grade		
G	33.779264	66.220736
F	32.684426	67.315574
E	26.849418	73.150582
D	21.986234	78.013766
C	17.194281	82.805719
B	12.205567	87.794433
A	5.993031	94.006969



- **Grade:** Higher(A) the grade greater the chance of paying loan.

Employment Length

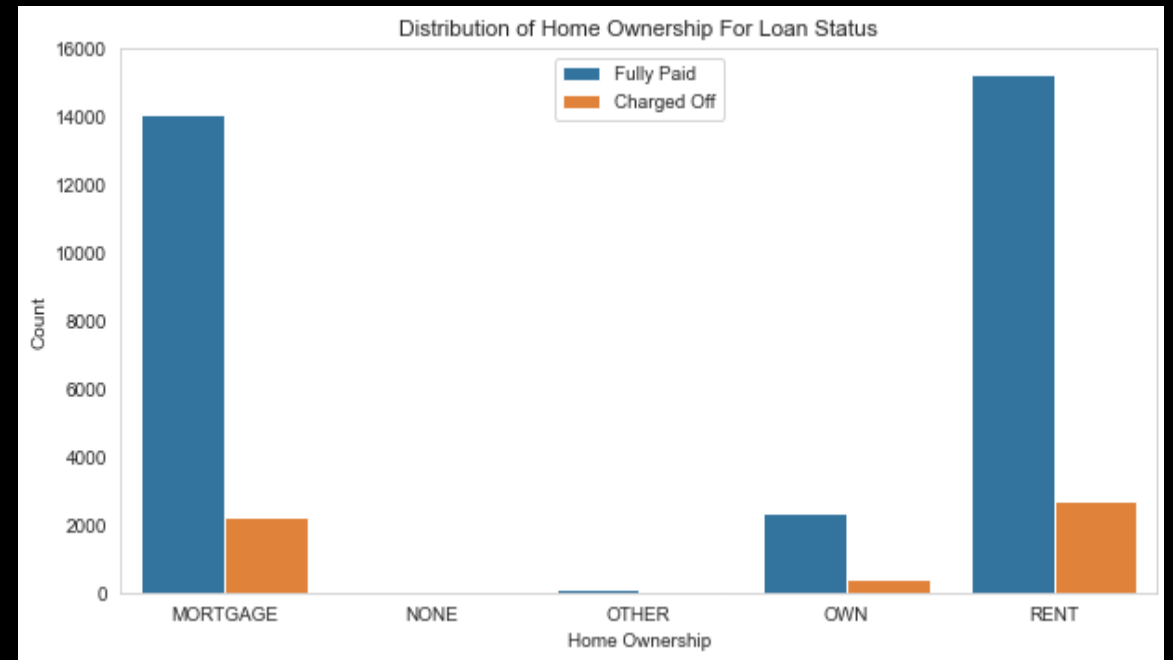
loan_status	Charged Off	Fully Paid
emp_length		
10+ years	15.680961	84.319039
7 years	15.371128	84.628872
1 year	14.389397	85.610603
5 years	14.339386	85.660614
< 1 year	14.174800	85.825200
6 years	14.160517	85.839483
8 years	14.146341	85.853659
3 years	13.833500	86.166500
4 years	13.824057	86.175943
2 years	13.213703	86.786297
9 years	12.887439	87.112561



- **Employment Length:** Majority of clients have 10+ years of experience and has highest proportion of defaulted loan.

Home Ownership

loan_status	Charged Off	Fully Paid
home_ownership		
OTHER	18.367347	81.632653
RENT	15.362554	84.637446
OWN	14.890756	85.109244
MORTGAGE	13.671347	86.328653
NONE	NaN	100.000000



Home ownership : people who has home ownership status as “Mortgage” or “Owner” has high chances of paying loan

Annual Income & Purpose

	loan_status	Charged Off	Fully Paid
annual_inc_qranges			
	VL	18.312597	81.687403
	L	15.692585	84.307415
	M	14.962886	85.037114
	H	12.965065	87.034935
	VH	10.928524	89.071476

■ Annual Income :

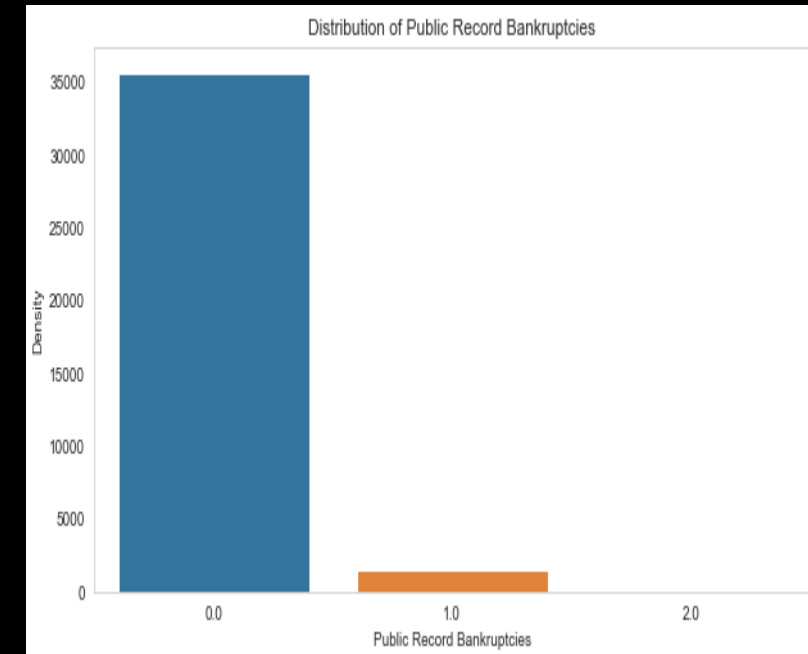
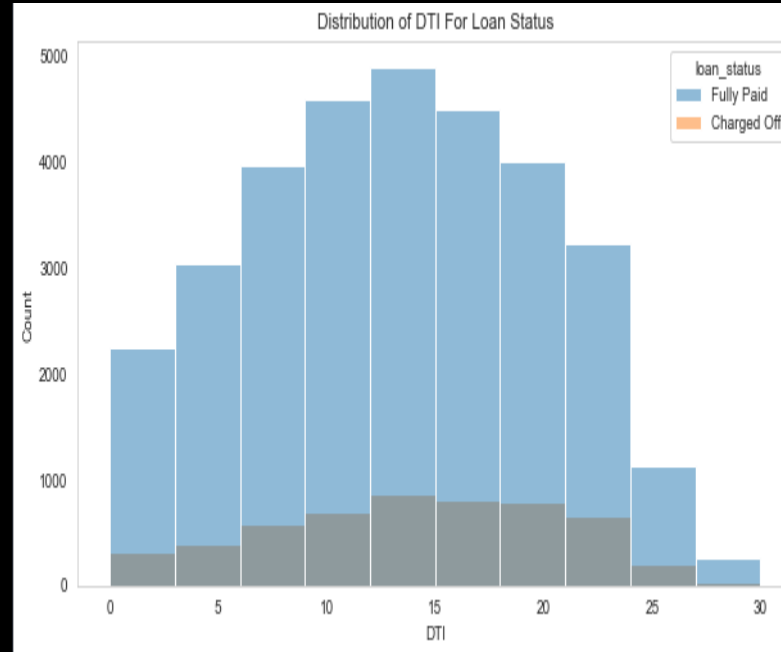
Very low (VL) Income group has maximum default

	loan_status	Charged Off	Fully Paid
purpose			
	small_business	27.080958	72.919042
	renewable_energy	18.627451	81.372549
	educational	17.230769	82.769231
	other	16.377749	83.622251
	house	16.076294	83.923706
	moving	15.972222	84.027778
	medical	15.565345	84.434655
	debt_consolidation	15.325395	84.674605
	vacation	14.133333	85.866667
	home_improvement	12.069565	87.930435
	credit_card	10.781778	89.218222
	car	10.673783	89.326217
	wedding	10.367171	89.632829
	major_purchase	10.325581	89.674419

- Purpose: Bank should be vigilant in providing loan to small business and renewable energy as they have maximum charge-off

DTI ratio & Bankruptcy

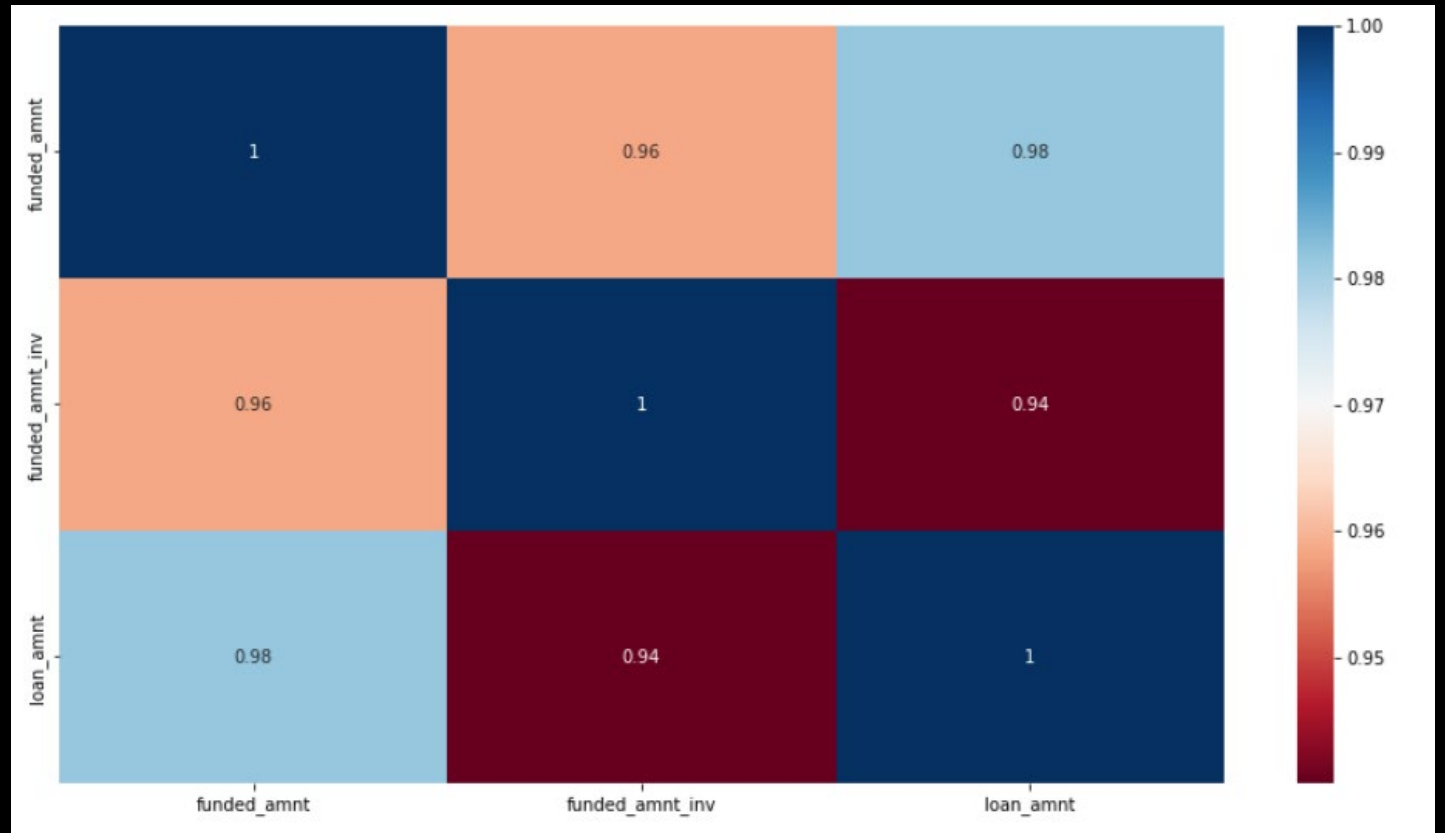
loan_status	Charged Off	Fully Paid
dti_ranges		
20-25	16.996047	83.003953
15-20	15.764385	84.235615
10-15	14.558671	85.441329
25-30	13.964687	86.035313
05-10	12.765147	87.234853
00-05	12.223587	87.776413



- **DTI:** The large percentage of Clients have a large Debt to Income ratio which shows that lending to such clients can be very risky.

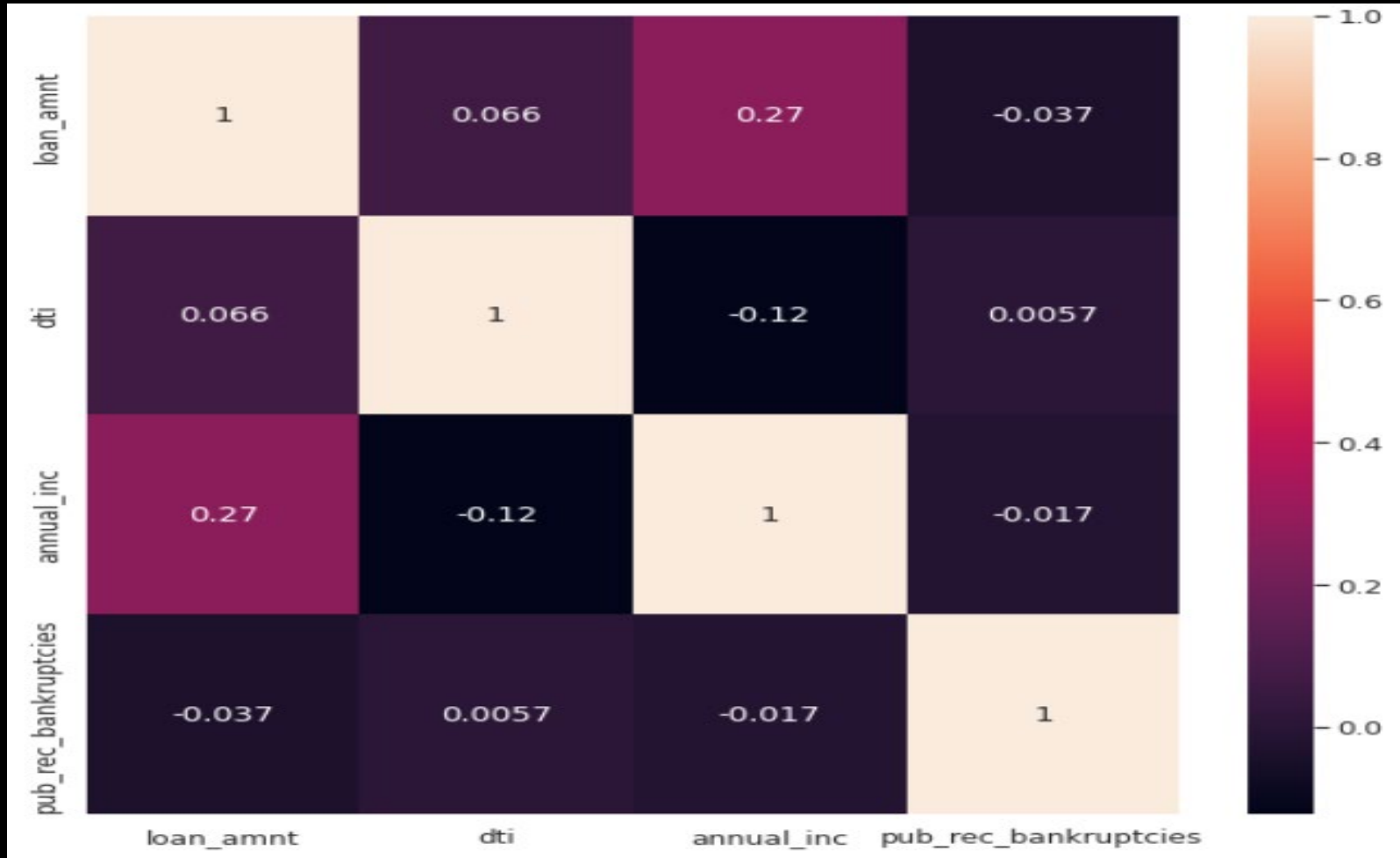
- **Public Recorded Bankruptcy:** Majority of clients have no record of declaring bankruptcy.

HeatMap



There is a high correlation between `loan_amnt` and `funded_amnt`, so one of them can be skipped

Correlation & Relationships



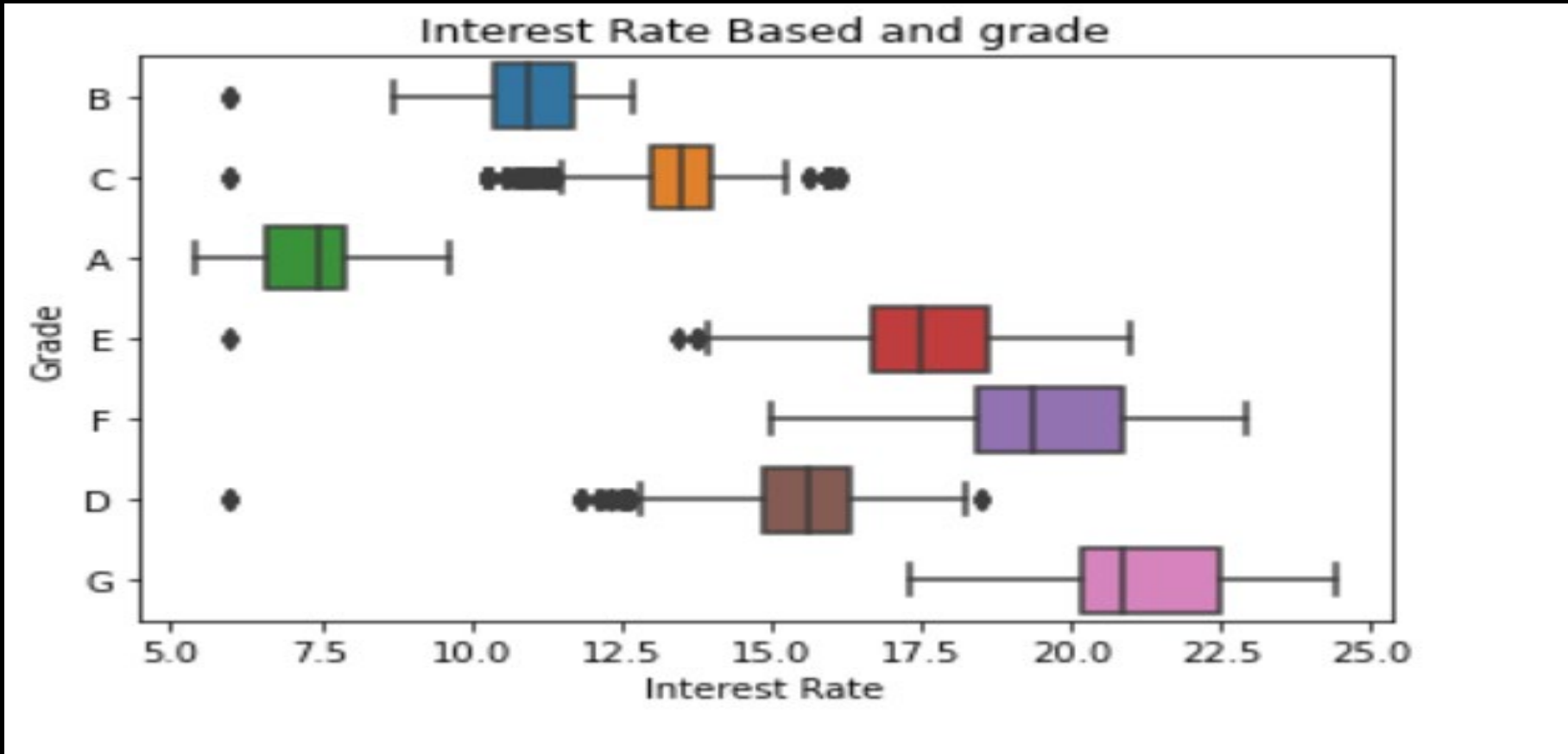
Negative Correlation

- loan_amnt has negative correlation with pub_rec_bankruptcies
- annual income has a negative correlation with dti

Strong Correlation

- annual income has a strong correlation with loan_amount

Bivariate Analysis



The Grade represent risk factor thus we can say interest rate increases with the risk.

Insights

- Major Driving factor which can be used to predict the chance of defaulting and avoiding Credit Loss:
 - 1. DTI
 - 2. Grades/Sub-Grades
 - 3. Verification Status
 - 4. Annual income
 - 5. Pub_rec_bankruptcies

Other considerations for 'defaults' :

- 1. Borrowers not from large urban cities e.g. NewYork , California etc
- 2. Borrowers having annual income in the range 50000-100000.
- 3. Borrowers having Public Recorded Bankruptcy.
- 4. Borrowers with least grades E.g. E,F,G indicates high risk.
- 5. Borrowers with very high Debt to Income value.
- 6. Borrowers with working experience 10+ years.

Conclusion

The default is highest for Other category and lowest for Mortgage category

Mortgage, Own, Rent are the best in repaying the loan and in that order

10+ years employees are defaulting about a percent more than the rest of the categories.

7 year employees also shows the similar trend

Verified category, source verified category is defaulting more when compared to not verified

Small business defaulters are way higher than the rest

DTI is able to find the defaulters in the right order. Higher the dti, higher the chance of default

The observation is same when the split/binning was done based on the ranges ['00-05', '05-10', '10-15', '15-20', '20-25', '25-30'] and based on the quantiles [0, 0.2, 0.4, 0.6, 0.8, 1]

Grade is also a strong indicator. Its shows similar trends as that of subgrades, subgrade F5 is 47% default