

Lending Club Case Study

Sakti Saurav
Avaneesh Srivastava

Index

1. Preparation of data - clean ups, manipulations, converting in suitable formats, derived columns
2. Data Analysis: findings on what influence defaults
 - a. Consumer Attributes
 - b. Loan Attributes
3. Recommendations

Preparation of data

| Aspect | Rationale | Impacted Attribute |
|---------------------------|--|--|
| Clean ups | Columns with high percentage of missing values and NA were removed | 54 columns were removed. Eg all_util, total_rev_hi_lim etc |
| | Removed columns like URL | url |
| Manipulations, formatting | Truncated dti (removed decimal part) for easy analysis | Dti |
| | Removed Outliers from Annual Income | |
| | Interest rates have % as suffix – cleaned | Int_rate |
| New derived columns | Bucketized interest rate into ordered categorical values according to quartiles (low/middle/upper middle/high) | |
| | Bucketized annual income into ordered categorical values according to quartiles | |
| | Bucketized Employee Duration | |

Data Analysis (1/6) : %age of defaults & trend by grade

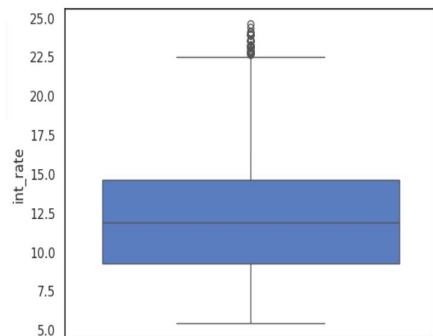
14% of loans have defaulted

| <i>loan_status</i> | Count of id | %age contribution |
|--------------------|-------------|-------------------|
| Charged Off | 5627 | 14% |
| Current | 1140 | 3% |
| Fully Paid | 32950 | 83% |

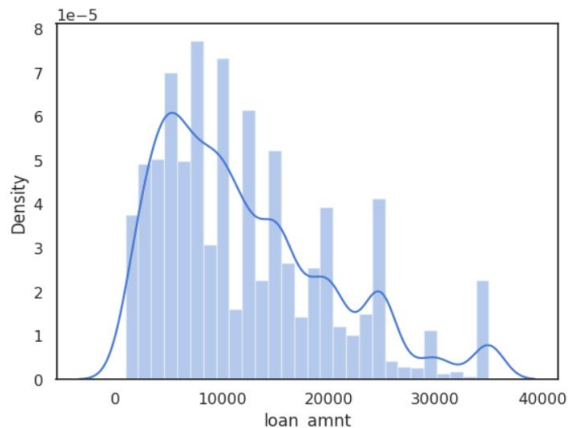
Loan grades B, C and D contribute more to defaults

| <i>Count of id</i> | <i>loan_status</i> | |
|--------------------|--------------------|-----------------------------|
| <i>grade</i> | Charged Off | Contribution in Charged Off |
| A | 602 | 11% |
| B | 1425 | 25% |
| C | 1347 | 24% |
| D | 1118 | 20% |
| E | 715 | 13% |
| F | 319 | 6% |
| G | 101 | 2% |

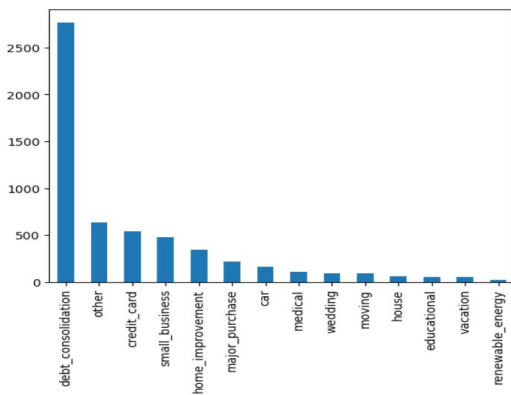
Data Analysis (2/6) : Univariate Analysis



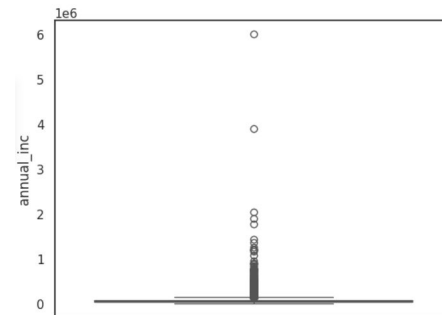
Quartile distribution of interest rates



75% of loans that defaulted had loan amount <16.5K



Debt consolidation & CC payment are top purposes for loans that defaulted.



Quartile distribution of Annual income and Outliers

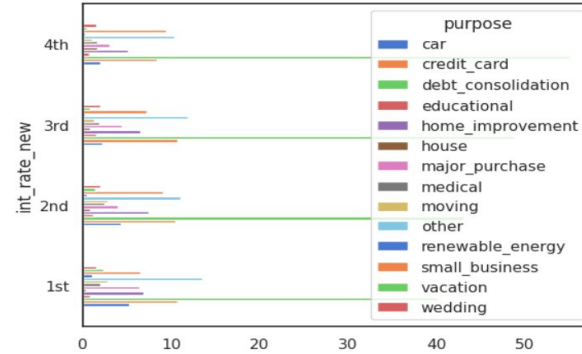
Data Analysis (3/6) : Bivariate Analysis

| loan_status | Charged Off | Current | Fully Paid |
|--------------|-------------|---------|------------|
| int_rate_new | | | |
| 1st | 600 | 42 | 9385 |
| 2nd | 1192 | 217 | 8799 |
| 3rd | 1524 | 305 | 7820 |
| 4th | 2311 | 576 | 6945 |

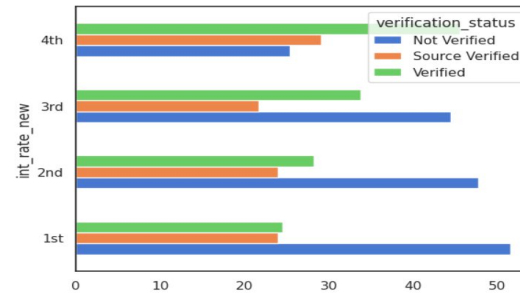
Interest rates of charged off loans were high

| annual_inc_new | low | middle | uppermiddle | rich |
|----------------|------|--------|-------------|------|
| loan_status | | | | |
| Charged Off | 1751 | 1461 | 1345 | 1063 |
| Current | 179 | 305 | 299 | 356 |
| Fully Paid | 7947 | 8247 | 8254 | 8461 |

As income increases, default decreases



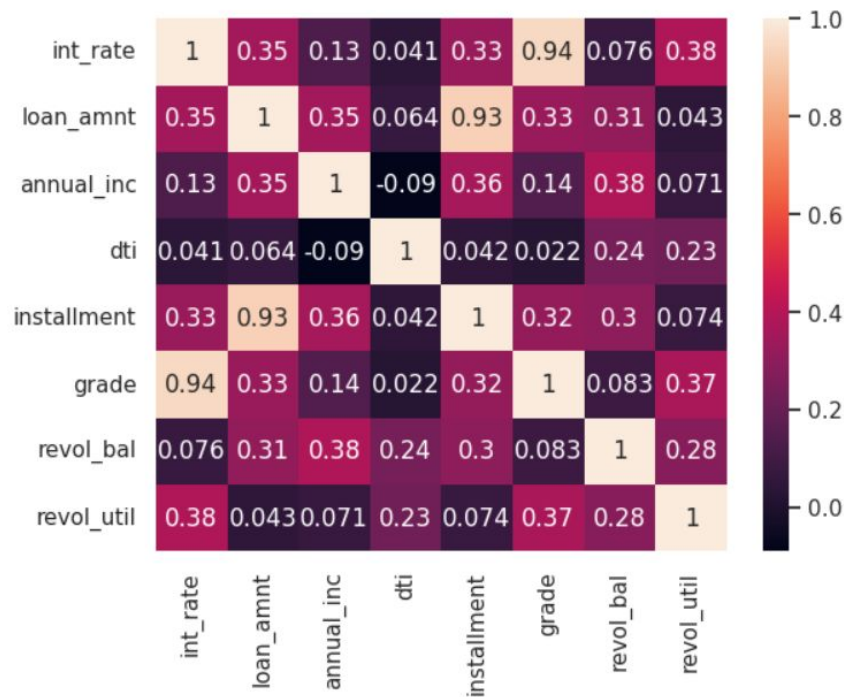
Debt consolidation & CC payment were major purposes for charged off loans



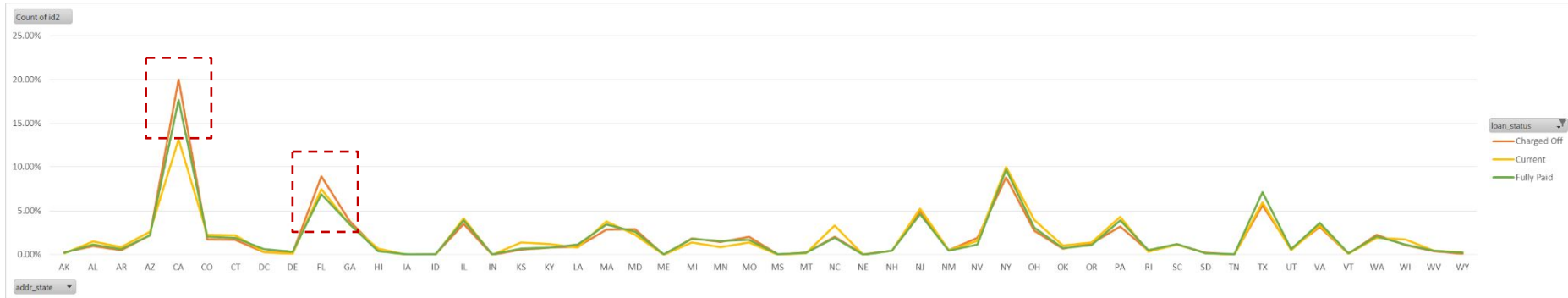
Higher interest rates are more verified compared to lower interest rates for Charged off applicants

Data Analysis (4/6): Correlation heatmap for charged off loans

- a) The interest Rate is proportional to grade
- b) The loan amount is proportional to the installment
- c) There is a slight positive relation between annual income and loan amount/installment

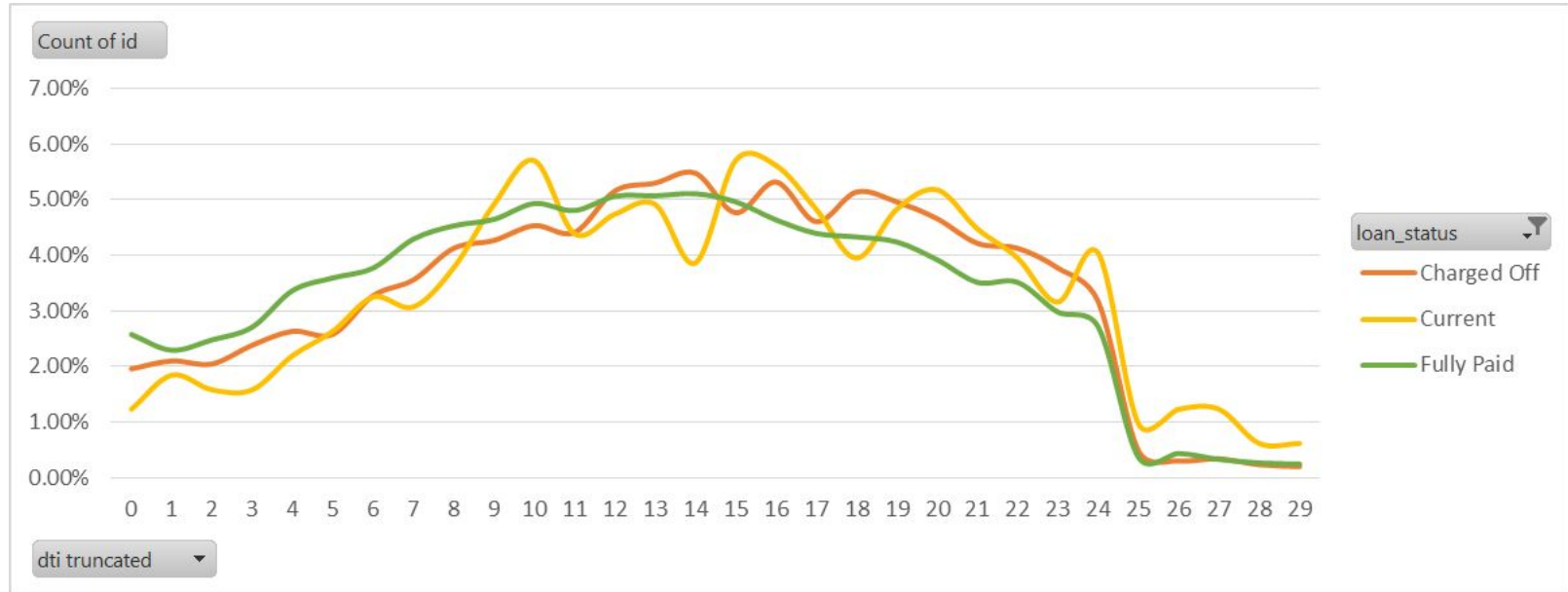


Data Analysis (5/6) : Impact of states on Loans



States : CA, FL seem to have higher defaults than paid off loans. **Alert!**

Data Analysis (6/6): Impact of DTI on Loans



High defaults for DTI >12. Curve for Charged Off loan cross curve for Fully Paid at ~12. **Alert!**

Recommendations

| | Indicators | Recommendation |
|---|--|--|
| <i>Attributes showing high default rates</i> | <ul style="list-style-type: none">• States : CA, FL• DTI : Greater than 12• Purpose: Debt Consolidation / Credit card bill repayment | <ol style="list-style-type: none">1. Exercise caution for applications from CA, FL2. Exercise caution for DTI > 123. Exercise caution for purpose – debt consolidation and CC repayment |
| <i>Attributes showing low default rates</i> | <ul style="list-style-type: none">• Salary : Greater than 82K USD | <ol style="list-style-type: none">1. Maximize on customers with high salary |