Gramener Case Study

Solution Deck

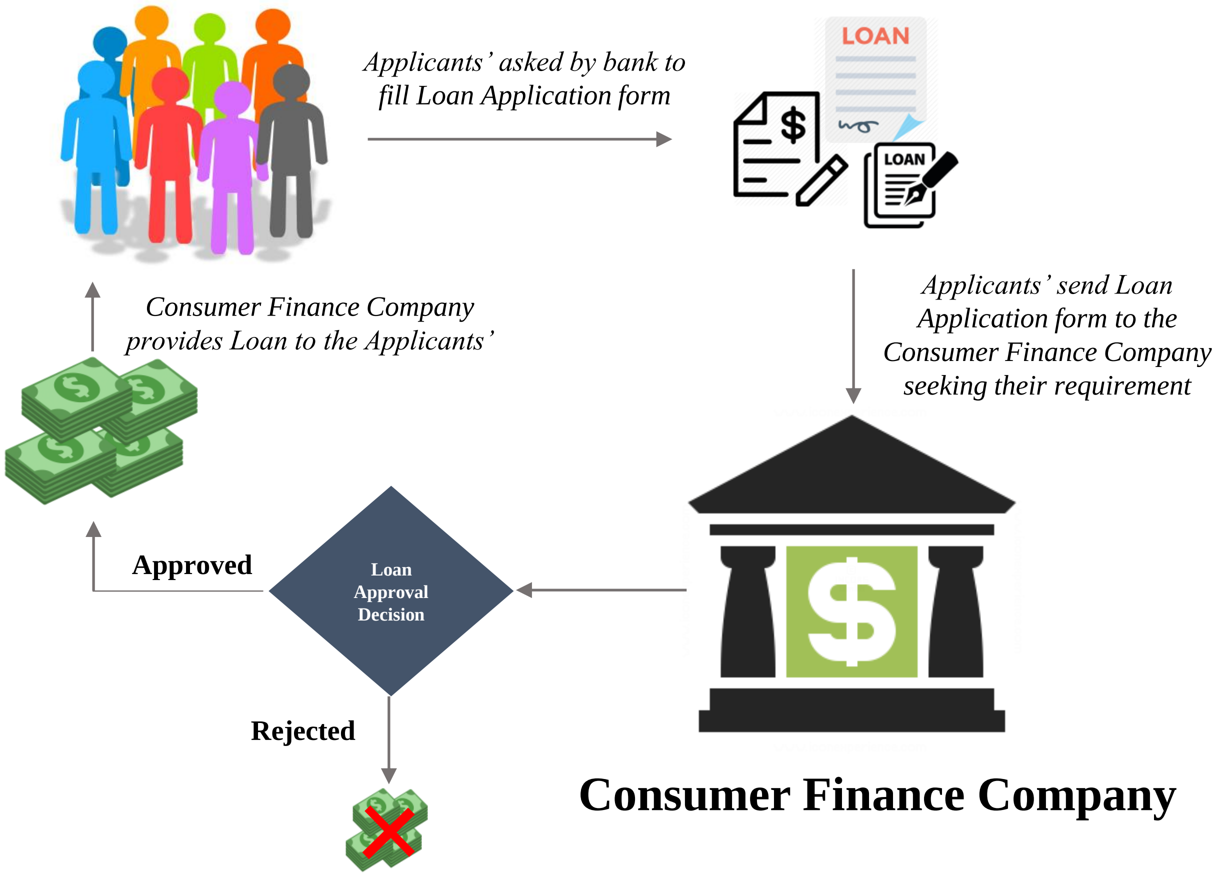
Group Members –

1. Brijesh Singh
2. Sourjya Sen
3. Debraj Ray
4. Nilanjan Dutta

Loan Approval Process

Current Scenario

Approval Process



Challenges / Risks

- If the Applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company

- If the Applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

Analysis Objective

Strategy

Driver Analysis

through available sources

- Data landscaping through **Univariate** & **Bivariate** means to identify potential drivers led by the applicants who are at default

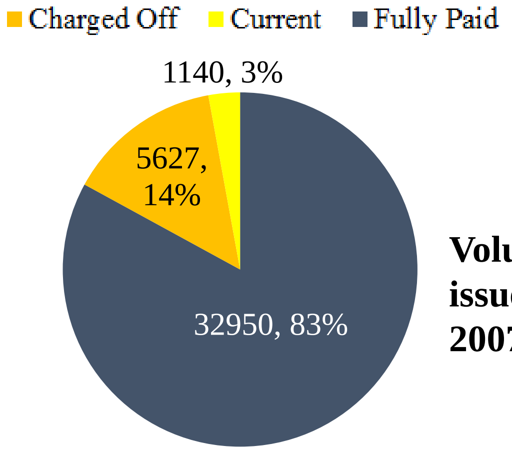
Driver Analysis

through derived measures

- Creating **derived metrics** using available resources to find patterns around applicants who are at default

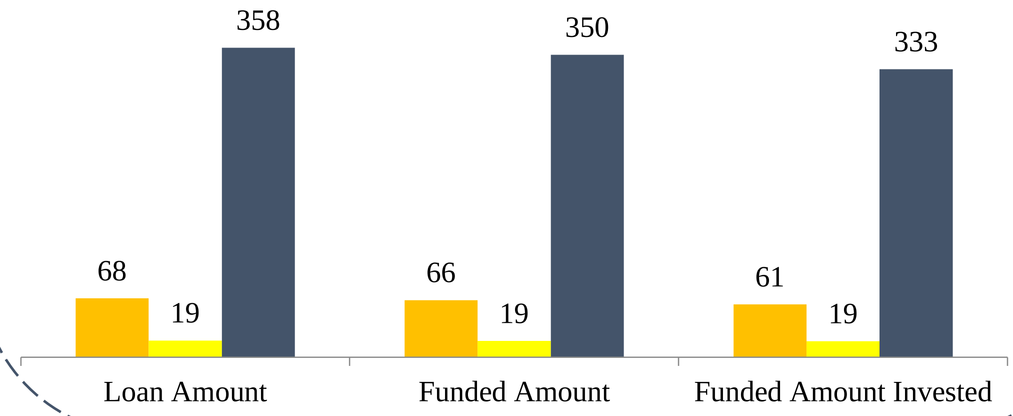
Understanding the Data**

Landscaping – Overall



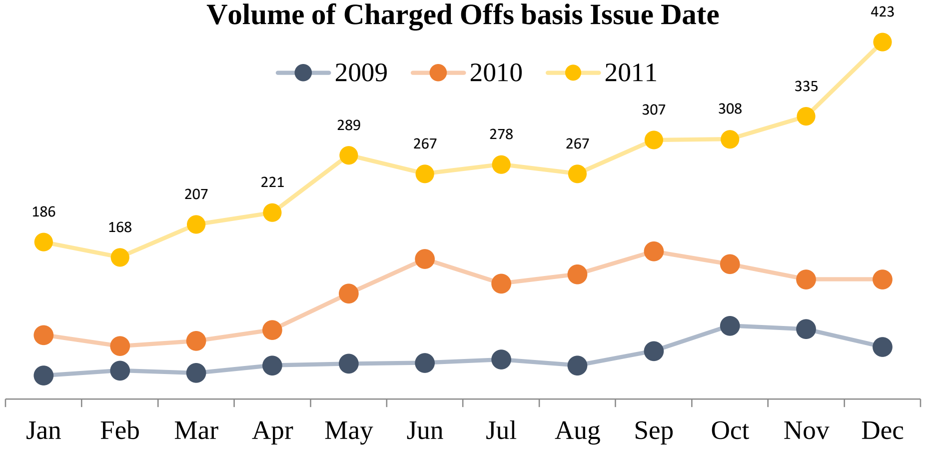
Volume of Loans issued across 2007 – 2011

Transaction Amount of Loans across different metrics (in MM)



- Across all the Loans issued by the Consumer Finance Company, **14%** Loans were categorized as Default whose Loan Amount equivalent to **~68 MM**

Volume of Charged Offs basis Issue Date



- There has been an increasing trend in the **Charged offs** observed in the recent years

[YoY %increase counts to **119%** for the most recent year]

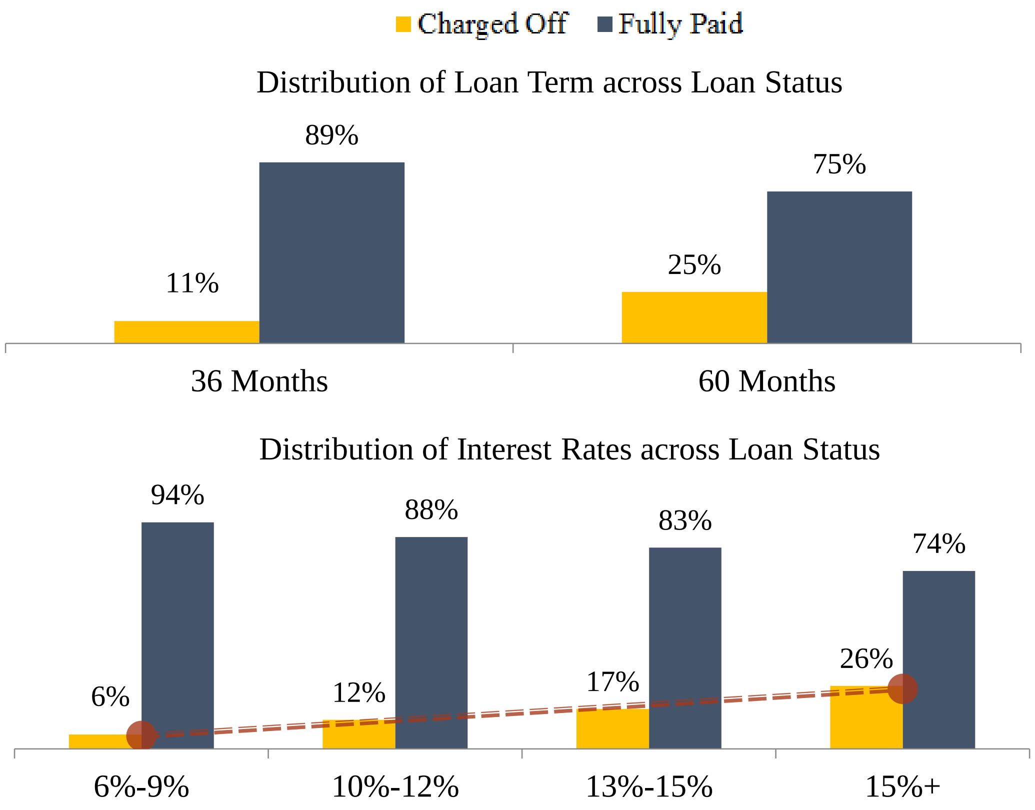
Loan Data provided for the **Consumer Finance Company and the Loans issued from **2007-2011****

Driver Analysis *by Categories*
[through available sources]

Objective around doing Driver Analysis is to create multiple hypotheses around the available data;
to see what factors having a significant impact on the Loan default

Hypothesis – I

Is there any pattern around the Loan Term?

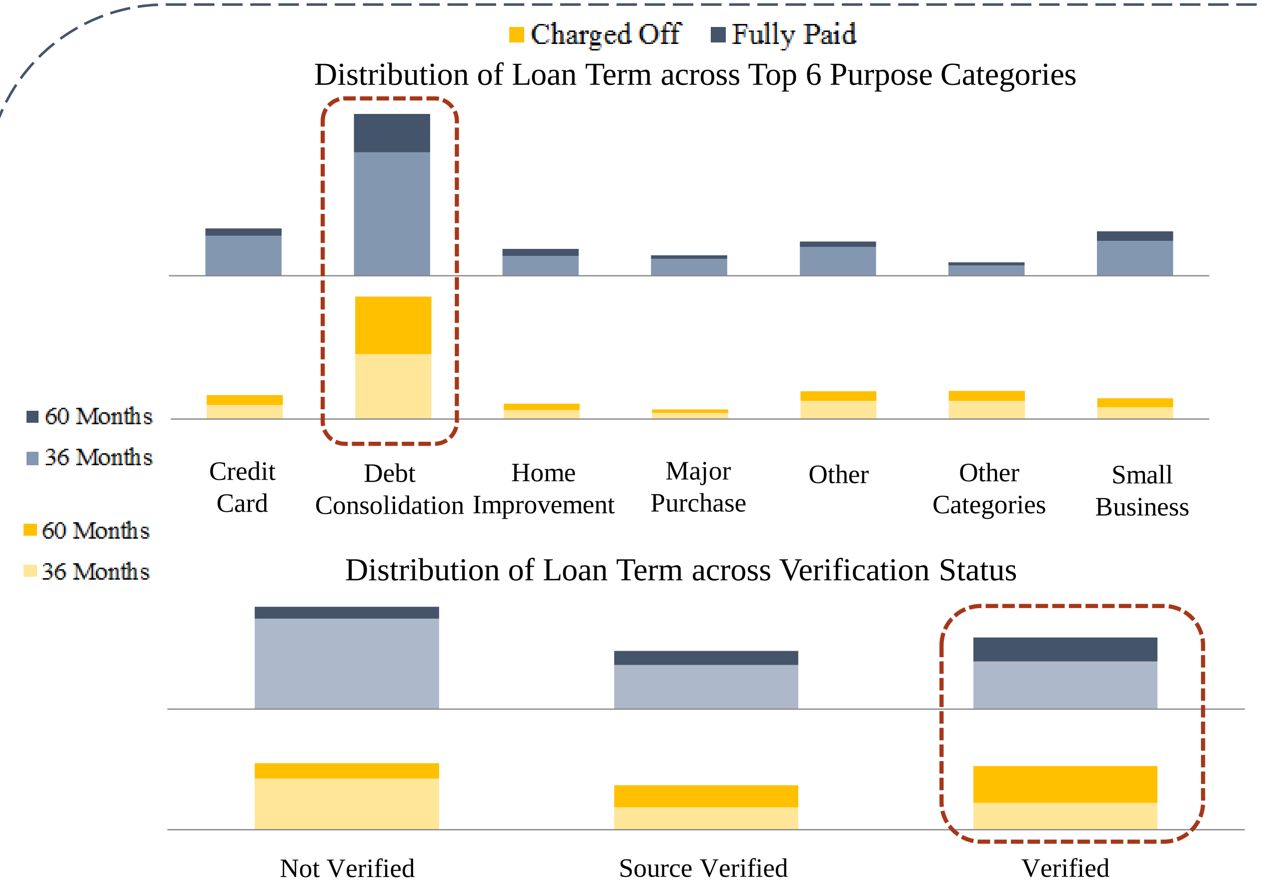


- It was observed that Applicants with a **longer** Loan Term tends to **default more** than shorter ones
- To address the above cause, we looked into the Interest Rates for the Applicants and observed that with the **increase** in **Interest Rates**, the chances of **getting default** also **increases**.

"With the Increase in Interest rate, the chances of default increases as a Loan with Higher Interest rate generally goes for a longer term. ***Consumer Finance Company*** *can take a look at optimizing the Interest rates across types to reduce the proportion of Charge Offs as the Applicants' Income stability might vary across years."*

Hypothesis – II

Does Loan Term also affect other factors?

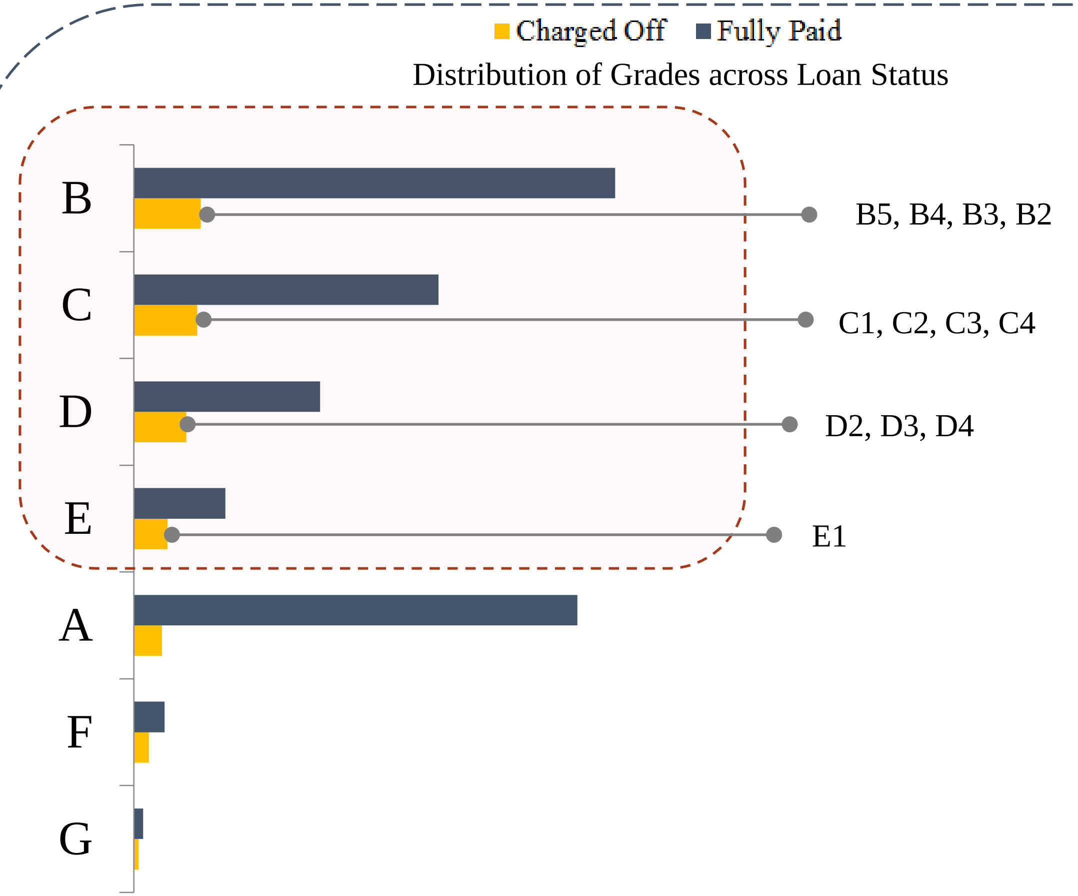


- It was observed that Applicants having **Debt Consolidation** as the purpose of Loan tends to default with a higher chance over a Loan Term of **60 Months**
- Applicants whose Verification Status showed **Verified** were captured more in the Charged off list for the Loan Term of **60 Months**

*"Debt Consolidation captured to be a very common purpose for Loan Application. However, applicants opting for a Loan Term of **60 Months** have higher chances of getting at default. Also, when looked at the Verification Status, majority of the Applicant's income sources were found Verified for the same term. **Consumer Finance Company** can scrutinize the verification process to understand specific trait of Borrower's at default."*

Hypothesis – III

Does Grades impact the Loan Status?

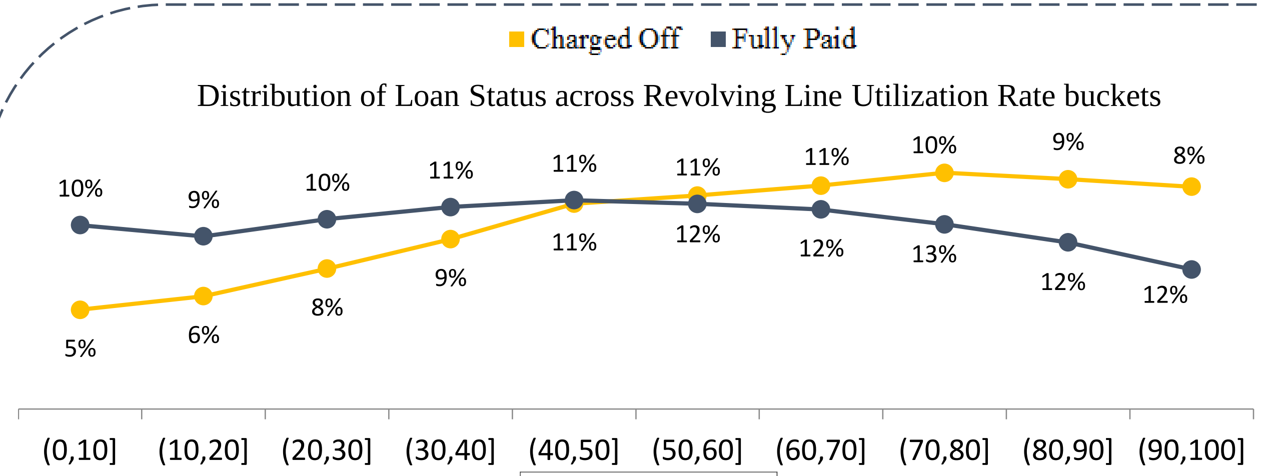


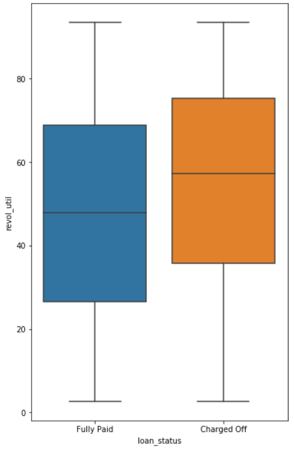
- Applicants who were rated by categories viz. **B**, **C**, **D** & **E** grades contributes to **~82%** of the total Charge off cases.
- **Majority** of the Loan Applications which were graded by these categories also had a higher volume of **Fully Paid** cases.
- There were certain **sub grades** which were significant in contributions.

*"It was observed that a very high volume of the Loan Applicants across all Loan Status' were Grade by categories such as **B**, **C**, **D** & **E**. However, when looked at **Charge Offs**, it was also found to have followed a similar pattern except for **Grade A** which had the most count of **Fully Paid** cases. It is highly recommended to the **Consumer Finance Company** to look at the Grading System before the Loan Approval"*

Hypothesis – IV

How does Revolving line utilization rate impacts Loan Status?





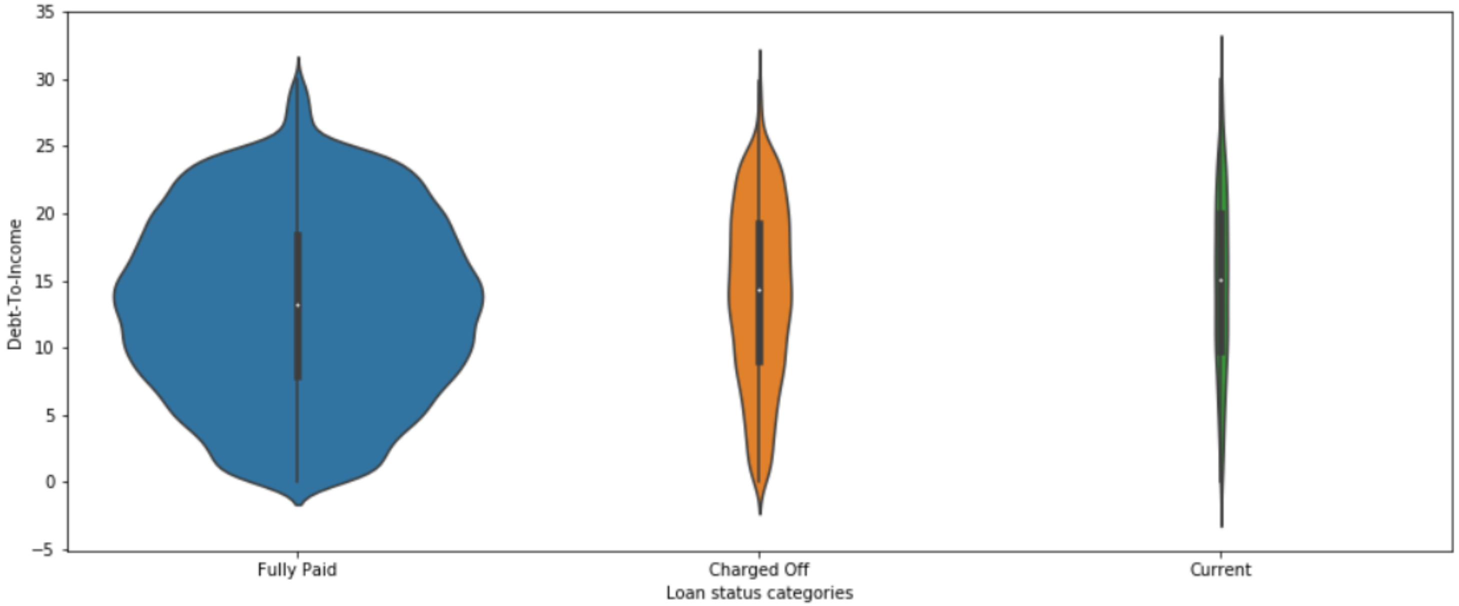
- Created **10** bins of Revolving Line Utilization rates to observe the capture of **Charged Off** & **Fully Paid** cases across the buckets.

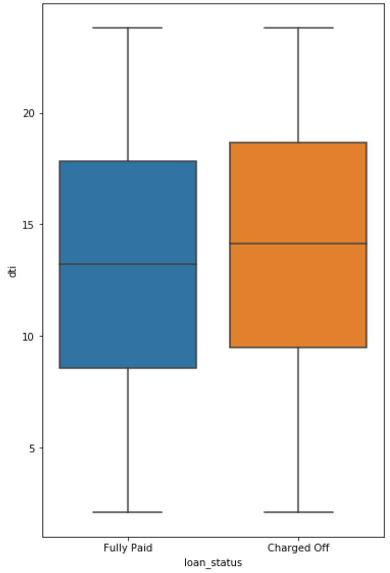
- From the plots, it can be said that higher the value of this Revolving balance utility rate , greater is the chance of loan default as we can observe the median quite high for defaulted loans.

*"With the Increase in revolving line of utilization rate, the chances of default increases. As a recommendation, the **Consumer Finance Company** while screening the Borrower's application should look at the Revolve line utilization rate at the lower side"*

Hypothesis – V

Is there any pattern around the debt to income ratio?





- It shows that there is an increase in DTI (median and mean) from **Fully Paid – Charged Off – Current**

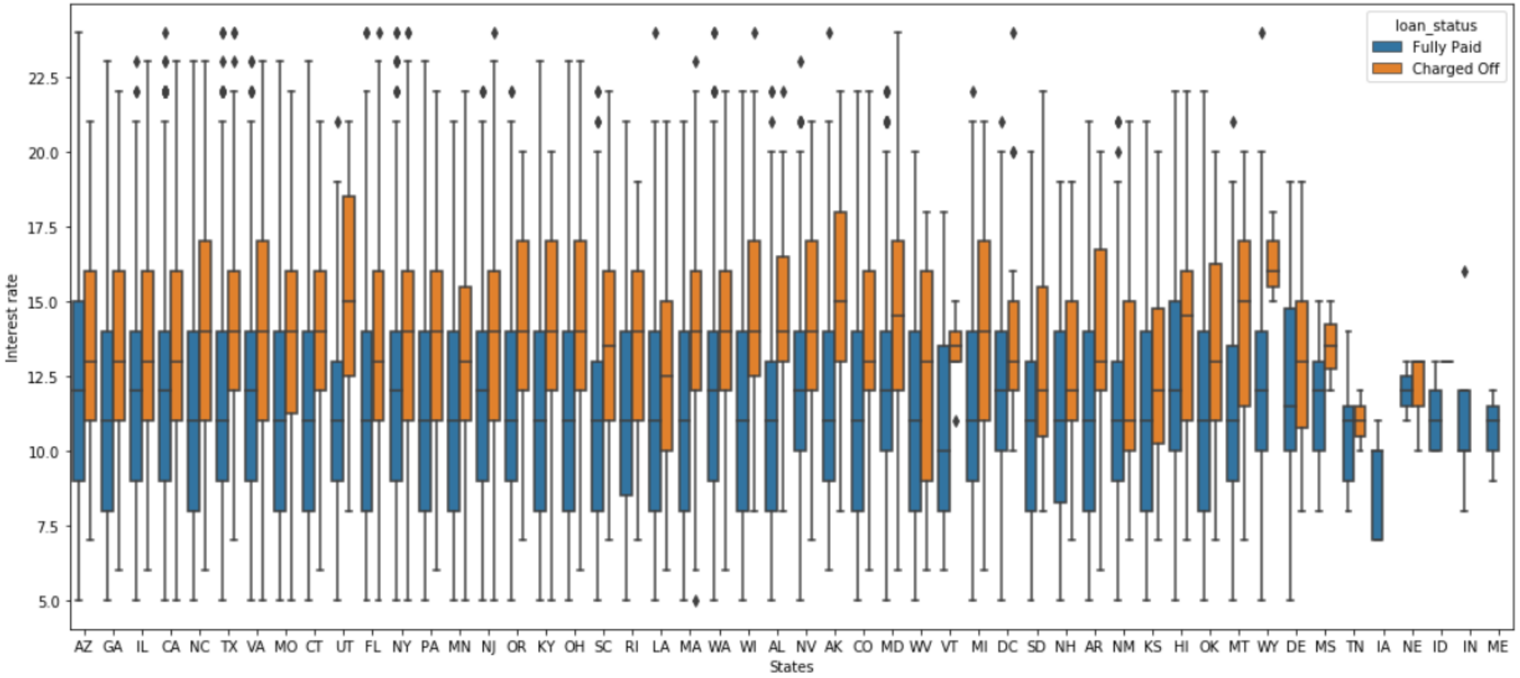
[***This is concerning for Consumer Finance Company because this indicates that despite having a few loans in the Current loan status, the DTI average is more than the Fully Paid which is nearly 30 times of Current. In other words, if we assume that the DTI rate stays the same as Consumer Finance Company accepts more loans then in the future the number of defaulters could increase significantly***]

- Higher the value of this debt to income, greater is the chance of loan default.
- For Charged off cases, there is a higher spread than the Fully paid loan implying higher the value there, greater is the risk of loan default

"With the Increase in debt to income ratio the chances of default increases is higher. As a recommendation, the ***Consumer Finance Company*** *while screening the Borrower's application should look at the debt to income ratio at the lower side"*

Hypothesis – VI

Are there Borrowers from specific States more likely to Charge offs?



- It was observed for certain states viz. **Utah**, **Ohio**, **Nevada**, **Alaska** & **Wisconsin** that with the increase in interest rates (*beyond a threshold*), the propensity of Borrowers encountering from Charge offs also increases.

*“The **Consumer Finance Company** while screening the Borrower's application from specific states (as highlighted) should implement a more robust model to capture their traits in order to minimize the charge offs”*

Driver Analysis *by Categories*

[through derived metrics]

Objective around doing Driver Analysis is to create multiple hypotheses around the derived metrics created using available data; to see what factors having a significant impact on the Loan default

Intelligent Features

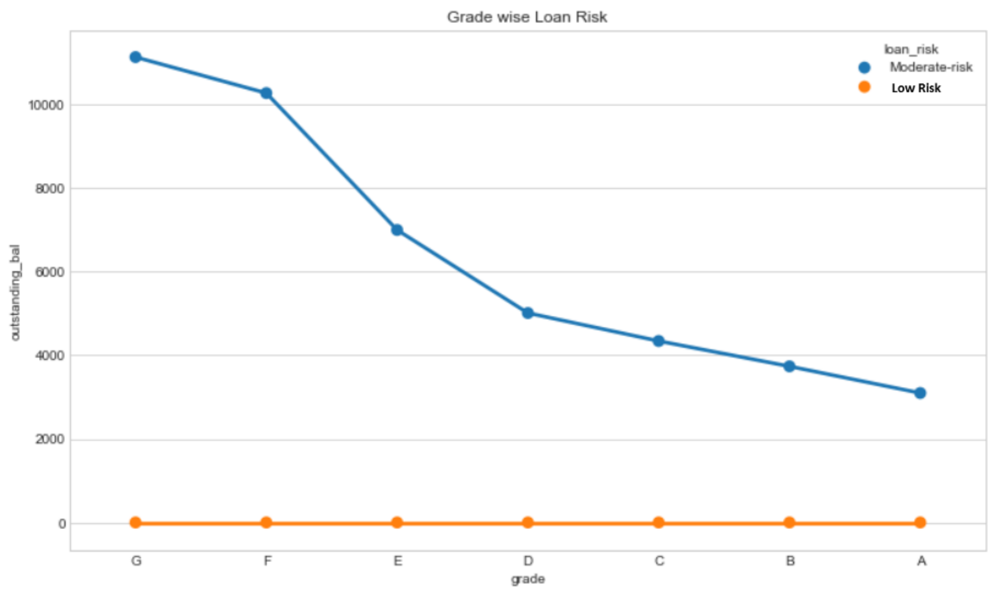
Understanding the Definition / Purpose

Feature Name	Definition/Purpose
Loan Risk Variable	Categorized the Loan Status into two parts **Low-Risk (Fully Paid)** & **Moderate Risk (Charged Offs + Current)** *It was observed that the Median distribution of Current Status for **DTI (Debt to Income)** was more than **15%**. In terms of business, more the **DTI** more goes the risk on Loan. Hence, to study a stronger impact across the variables, these two Loan Status' were combined into one*
Outstanding Balance	**[Outstanding Balance = Funded Amount – Total Principal Received Amount]** *To study the impact of Outstanding Balance on different variables*
US Regions	Divided **51** US states into **5** Regions viz. **Western, South West, South East, Mid West** & **North East** *To study the impact of US Regions across different variables*
Income Category	Categorized Borrower's Annual Incomes into 4 segments – **High (>90K)**, **High – Medium (Between 60K-90K)**, **Low-Medium (30K-60K)** & **Low(<30K)** **[***All Income in USD***]** *To study the impact of Categories across **DTI***

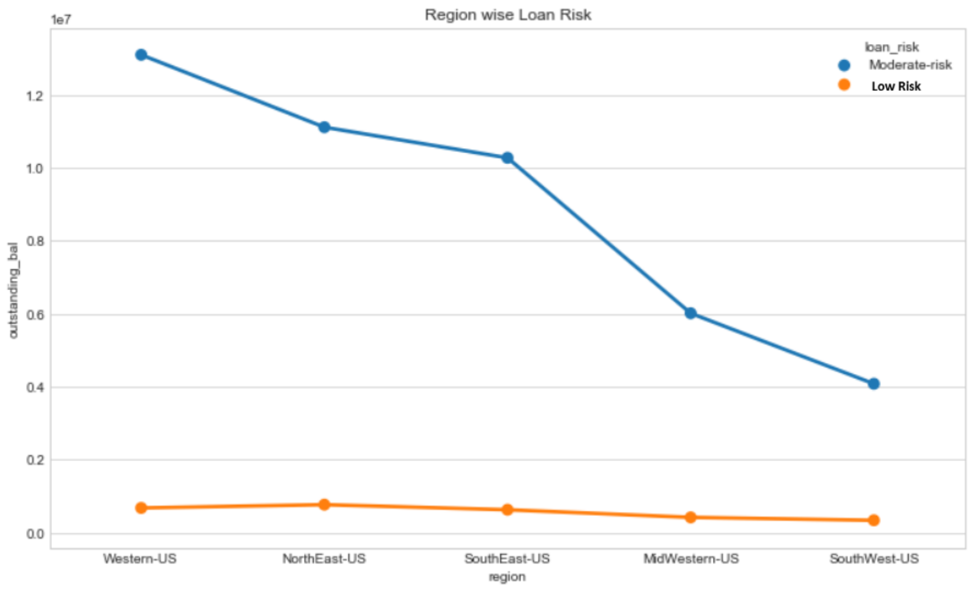
Risk Analysis for Derived Metrics

Insights from Intelligent Features – I

Distribution of Grade across Loan Risk



Distribution of US Regions across Loan Risk



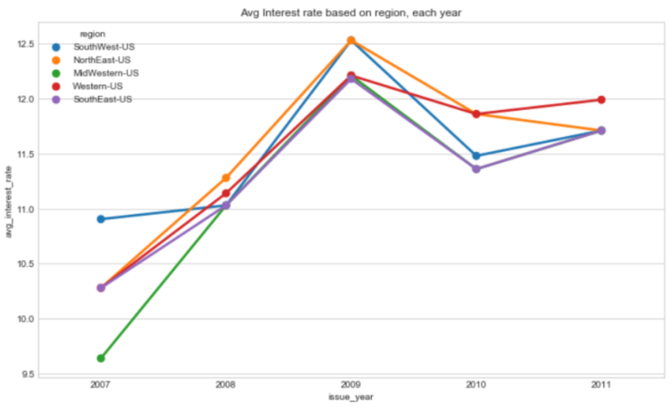
- As we can see **G,F & E** category grade are in moderate risk (higher grade category higher moderate risk)
- **Western-US** region is higher on Moderate Risk as total outstanding balance is higher in this region

Driver Analysis for Derived Metrics

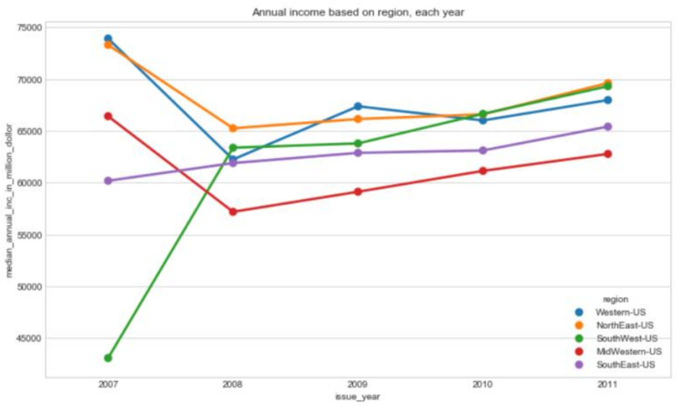
Insights from Intelligent Features – **II**

Distribution of US Regions

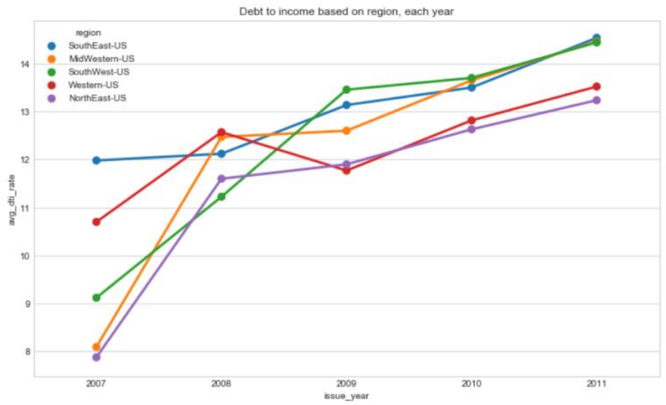
by Average Interest Rate



by Annual Income



by DTI



- MidWestern-US, SouthEast-US & SouthWest-US had a rapid increase in debt-to-income starting in 2011
- MidWestern-US, SouthEast-US and SouthWest-US had a rapid increase in interest rates (This might explain the increase in debt to income)
- DTI is inversely proportional to Annual income because dti=(monthly expenses including loan / monthly income)

Analysis of Loan Variables

Findings through Cross-Tab Analysis

	no_risk_count	moderate_risk_count	total_loans_issued	Moderate-Risk/No-Risk ratio (%)	purpose	income_category
0	228	24	252	9.52	car	High
1	366	29	395	7.34	major_purchase	High
2	970	114	1084	10.52	credit_card	High
3	2815	502	3317	15.13	debt_consolidation	High
4	39	7	46	15.22	educational	High
5	70	19	89	21.35	house	High
6	542	122	664	18.37	other	High
7	861	114	975	11.69	home_improvement	High
8	89	8	97	8.25	moving	High
9	24	3	27	11.11	renewable_energy	High
10	350	105	455	23.08	small_business	High
11	52	5	57	8.77	vacation	High
12	145	18	163	11.04	wedding	High
13	116	18	134	13.43	medical	High
14	705	128	833	15.37	home_improvement	High_med
15	318	48	366	13.11	car	High_med
16	1261	176	1437	12.25	credit_card	High_med
17	4267	883	5150	17.15	debt_consolidation	High_med
18	45	9	54	16.67	educational	High_med
19	91	17	108	15.74	house	High_med
20	461	39	500	7.8	major_purchase	High_med
21	127	28	155	18.06	medical	High_med
22	748	138	886	15.58	other	High_med
23	19	6	25	24	renewable_energy	High_med
24	351	149	500	29.8	small_business	High_med
25	76	12	88	13.64	vacation	High_med
26	94	10	104	9.62	moving	High_med
27	239	32	271	11.81	wedding	High_med
28	287	63	350	18	major_purchase	Low
29	207	37	244	15.16	car	Low
30	426	66	492	13.41	credit_card	Low
31	87	22	109	20.18	educational	Low
32	158	38	196	19.39	home_improvement	Low
33	27	8	35	22.86	house	Low
34	1538	388	1926	20.15	debt_consolidation	Low
35	71	17	88	19.32	medical	Low
36	90	29	119	24.37	moving	Low
37	550	160	710	22.54	other	Low
38	133	57	190	30	small_business	Low
39	51	11	62	17.74	vacation	Low
40	74	13	87	14.94	wedding	Low
41	13	3	16	18.75	renewable_energy	Low
42	143	31	174	17.82	vacation	Low_med
43	586	101	687	14.7	car	Low_med
44	1828	289	2117	13.65	credit_card	Low_med
45	6668	1580	8248	19.16	debt_consolidation	Low_med
46	804	168	972	17.28	home_improvement	Low_med
47	98	18	116	15.52	educational	Low_med
48	211	52	263	19.77	moving	Low_med
49	261	55	316	17.41	medical	Low_med
50	1392	341	1733	19.68	other	Low_med
51	27	8	35	22.86	renewable_energy	Low_med
52	445	238	683	34.85	small_business	Low_med
53	120	29	149	19.46	house	Low_med
54	814	128	942	13.59	major_purchase	Low_med
55	372	54	426	12.68	wedding	Low_med

Mostly done to identify distribution with easy. A couple of examples listed below –

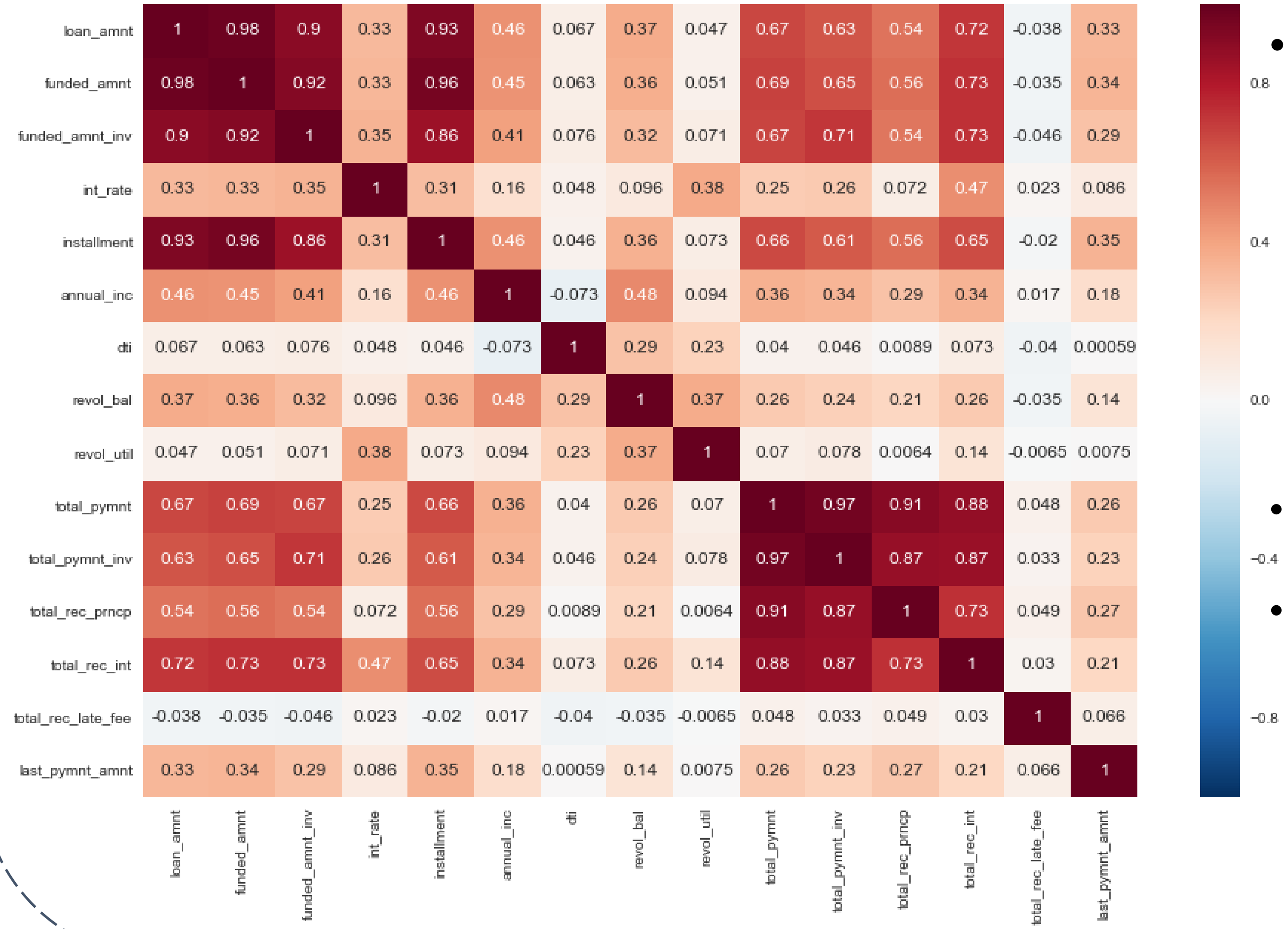
- **Capturing the High Risk Loans**– The category of applicants seeking Loans as a purpose of small business tends to have a higher risk of being a default loan. However, the severity would be less as it has a relatively l**ow frequency.**

- **Capturing Frequent Purpose Categories** – It was observed that the purpose contributed by debt consolidated had the highest count.

Analysis of Loan Variables

Correlation Analysis

Correlation Matrix



	loan_amnt	funded_amnt	funded_amnt_inv	int_rate	installment	annual_inc	dti	revol_bal	revol_util	total_pymnt	total_pymnt_inv	total_rec_prncp	total_rec_int	total_rec_late_fee	last_pymnt_amnt
loan_amnt	1	0.98	0.9	0.33	0.93	0.46	0.067	0.37	0.047	0.67	0.63	0.54	0.72	-0.038	0.33
funded_amnt	0.98	1	0.92	0.33	0.96	0.45	0.063	0.36	0.051	0.69	0.65	0.56	0.73	-0.035	0.34
funded_amnt_inv	0.9	0.92	1	0.35	0.86	0.41	0.076	0.32	0.071	0.67	0.71	0.54	0.73	-0.046	0.29
int_rate	0.33	0.33	0.35	1	0.31	0.16	0.048	0.096	0.38	0.25	0.26	0.072	0.47	0.023	0.086
installment	0.93	0.96	0.86	0.31	1	0.46	0.046	0.36	0.073	0.66	0.61	0.56	0.65	-0.02	0.35
annual_inc	0.46	0.45	0.41	0.16	0.46	1	-0.073	0.48	0.094	0.36	0.34	0.29	0.34	0.017	0.18
dti	0.067	0.063	0.076	0.048	0.046	-0.073	1	0.29	0.23	0.04	0.046	0.0089	0.073	-0.04	0.00059
revol_bal	0.37	0.36	0.32	0.096	0.36	0.48	0.29	1	0.37	0.26	0.24	0.21	0.26	-0.035	0.14
revol_util	0.047	0.051	0.071	0.38	0.073	0.094	0.23	0.37	1	0.07	0.078	0.0064	0.14	-0.0065	0.0075
total_pymnt	0.67	0.69	0.67	0.25	0.66	0.36	0.04	0.26	0.07	1	0.97	0.91	0.88	0.048	0.26
total_pymnt_inv	0.63	0.65	0.71	0.26	0.61	0.34	0.046	0.24	0.078	0.97	1	0.87	0.87	0.033	0.23
total_rec_prncp	0.54	0.56	0.54	0.072	0.56	0.29	0.0089	0.21	0.0064	0.91	0.87	1	0.73	0.049	0.27
total_rec_int	0.72	0.73	0.73	0.47	0.65	0.34	0.073	0.26	0.14	0.88	0.87	0.73	1	0.03	0.21
total_rec_late_fee	-0.038	-0.035	-0.046	0.023	-0.02	0.017	-0.04	-0.035	-0.0065	0.048	0.033	0.049	0.03	1	0.066
last_pymnt_amnt	0.33	0.34	0.29	0.086	0.35	0.18	0.00059	0.14	0.0075	0.26	0.23	0.27	0.21	0.066	1

- **Two clusters** which are **highly correlated** among each other –
  - **Loan Amount**, **Funded Amount** & **Funded Amount Investment** are very highly correlated amongst each other
  - **Total Payment**, **Total Payment Investment**, **Total Received Principal** & **Total Received Interest** are highly correlated amongst each other
- The cluster variables are **positively** dependent on each other.
- Other variables which were found to be moderately Correlated –
  - A. **Installment & Total Payment**
  - B. **Annual Income & Revolving Balance.**

Summarizing the findings

Recommendations for the Consumer Finance Company

- Summarizing the analysis done on 6 Hypotheses , below are the key highlights –
  - Loan Term with longer duration tends to capture the majority of Charged Off cases. Hence, to address the cause for **higher interest rates** the company can look for an optimization.
  - **Loan Applicants' Verification System** doesn't have a good impact on the Loan Status. It needs to be **revamped**.
  - Borrowers with **higher DTI tends to default more**. It needs to be screened in a **more robust way** while processing the loan approval.
  - While screening the Borrower's application, the **Revolve line utilization rate** should be considered at the lower side to minimize the impact of default.
- Few Insights from the Derived metrics can be utilized to address the maximum impact and minimize the proportion of default –
  - Grades such as **G,F & E** are observed under moderate risk category. During the applicant's screening process, it needs to be carefully monitored.
  - **Western-US** region is higher on Moderate Risk as total outstanding balance is higher in this region. Applicants' belonging to these regions should be carefully monitored.