

# Gramener Case Study

Group Name:

1. Isha Jotshi
2. Pravin Pawar
3. Aashish Sharma
4. Avinash Kothari

# Agenda

- **Problem Statement**
- Data Understanding
- Approach for the Analysis
- Derived Variables
- EDA (Exploratory Data Analysis) (Univariate and Bi-variate)
- Inferences
- Conclusion

# Problem Statement

The data is provided for a peer-to-peer lending company which connects lenders with borrowers in an online marketplace, providing competitive rates to both parties. The lenders who are willing to provide finance are enlisted on the website post 'opening an account' with the company. The company grants loans to borrowers post necessary due diligence and verification.

As per the data provided the key challenge currently is to identify factors impacting charged off loans. These charged off loans contribute to 14% of the total loans dispersed. The company now wants to understand the driver behind loan defaults, so it can utilize this knowledge for its portfolio and risk assessment.

# Agenda

- Problem Statement
- **Data Understanding**
- Approach for the Analysis
- Derived Variables
- EDA (Exploratory Data Analysis) (Univariate and Bi-variate)
- Inferences
- Conclusion

# Data Understanding

The data received is a customer single view data with 111 variables and 39,717 observations. The key attributes of the data is as follows:

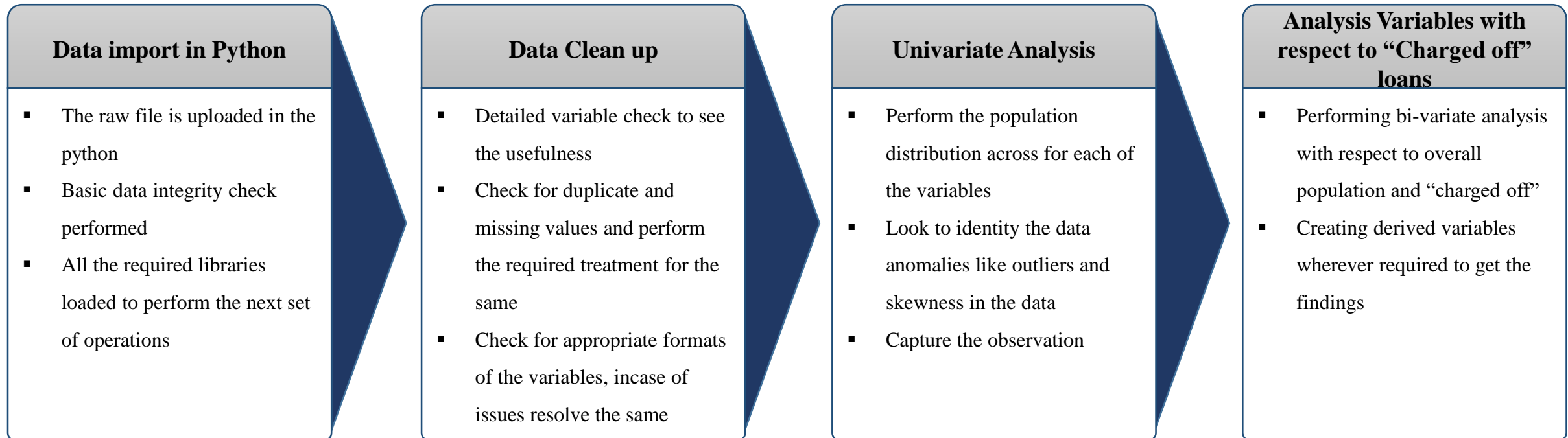
1. **Duration (Issue Date):** July 2007 to December 2011
2. **Customer data** like (id, member\_id, emp\_title, emp\_length, home\_ownership, zip\_code, addr\_state, etc.)
3. **Customer credit worthiness** related variables like (grade, subgrade, annual\_inc, home\_ownership, dti, delinq\_2yrs, earliest\_cr\_line, inq\_last\_6mths , open\_acc, revol\_bal, last\_credit\_pull\_d, pub\_rec\_bankruptcies, etc.)
4. **Loan details** variables like (loan\_amnt, funded\_amnt, funded\_amnt\_inv, term, int\_rate, installment, grade, sub\_grade, loan\_status, purpose, title, etc.)

# Agenda

- Problem Statement
- Data Understanding
- **Approach for the Analysis**
- Derived Variables
- EDA (Exploratory Data Analysis) (Univariate and Bi-variate)
- Inferences
- Conclusion

# Approach for the Analysis

## Follow of the analysis:



***Note:** During the data clean up phase 68 variables dropped as there was high share of missing values or null values in those columns.*

# Agenda

- Problem Statement
- Data Understanding
- Approach for the Analysis
- **Derived Variables**
- EDA (Exploratory Data Analysis) (Univariate and Bi-variate)
- Inferences
- Conclusion



# Derived Variables

Below is the list of variables basis which derived variables\*\* are created to have rich data insights:

- **funded\_amnt\_group**: The funded\_amt grouped from <1k to >30k with an interval of 3k.
- **funded\_amnt\_inv\_group**: The funded\_amt grouped from <1k to >30k with an interval of 3k.
- **Int\_rate\_group**: The int\_rate grouped from 4% to >24% with an interval of 2%
- **installment\_group**: instalment is grouped from <50 to >1200 with interval of 50
- **annual\_inc\_group**: annual\_inc is grouped from <30k to >100k with interval of 10k
- **open\_acc\_group**: open\_acc all the accounts >20 have been clubbed together as it contributes only to 2% of the total accounts.
- **revol\_util\_%\_group**: revol\_util% is categorised from 0% to 100% with interval of 10%
- **total\_pymnt\_group**: total\_pymnt is grouped from <3k to >39k with interval of 3k
- **total\_pymnt\_inv\_group**: total\_pymnt\_inv is grouped from <3k to >39k with interval of 3k
- **total\_rec\_prncp\_group**: total\_rec\_prncp is grouped from <2k to >26k with interval of 2k
- **total\_rec\_int\_group**: total\_rec\_int is grouped from <500 to 10k with interval of 500

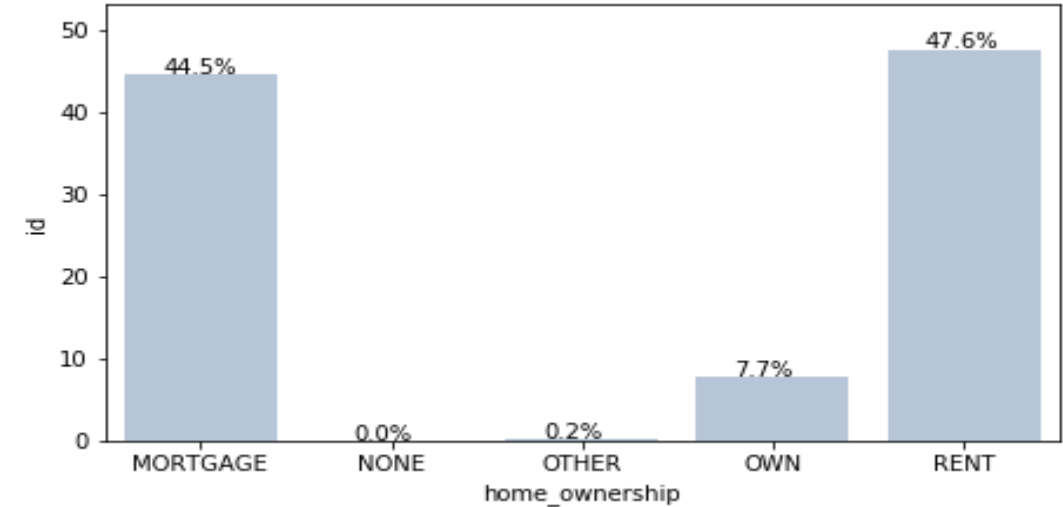
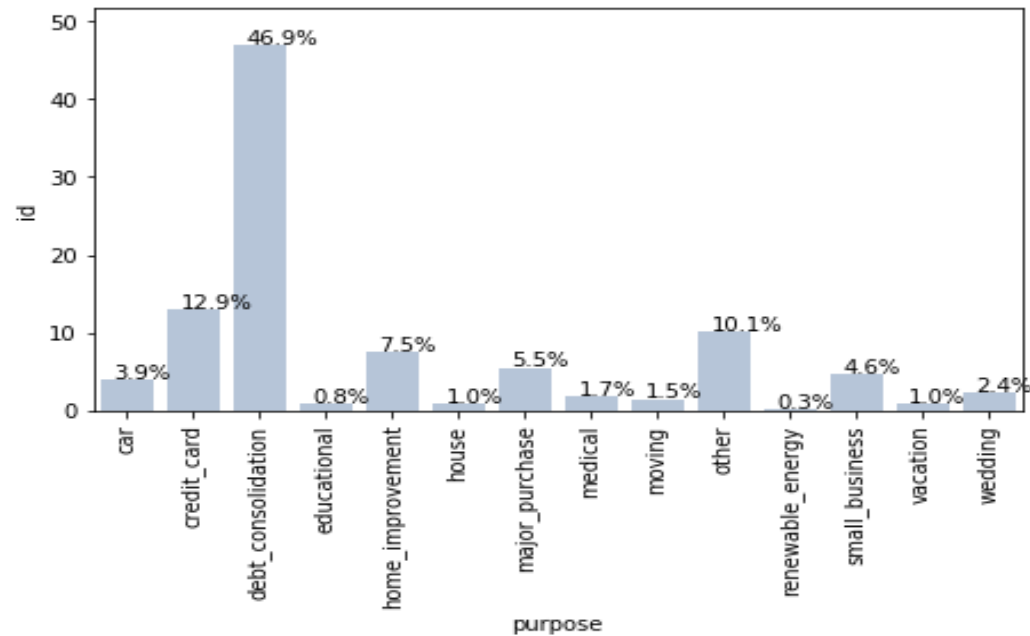
\*\*All the derived variables are created post checking measures of the central tendency for each of the variables and comparing the values for at every 10<sup>th</sup> percentile

# Agenda

- Problem Statement
- Data Understanding
- Approach for the Analysis
- Derived Variables
- **EDA (Exploratory Data Analysis) (Univariate and Bi-variate)**
- Inferences
- Conclusion

# Univariate Analysis for Categorical variables

Purpose/Home Ownership

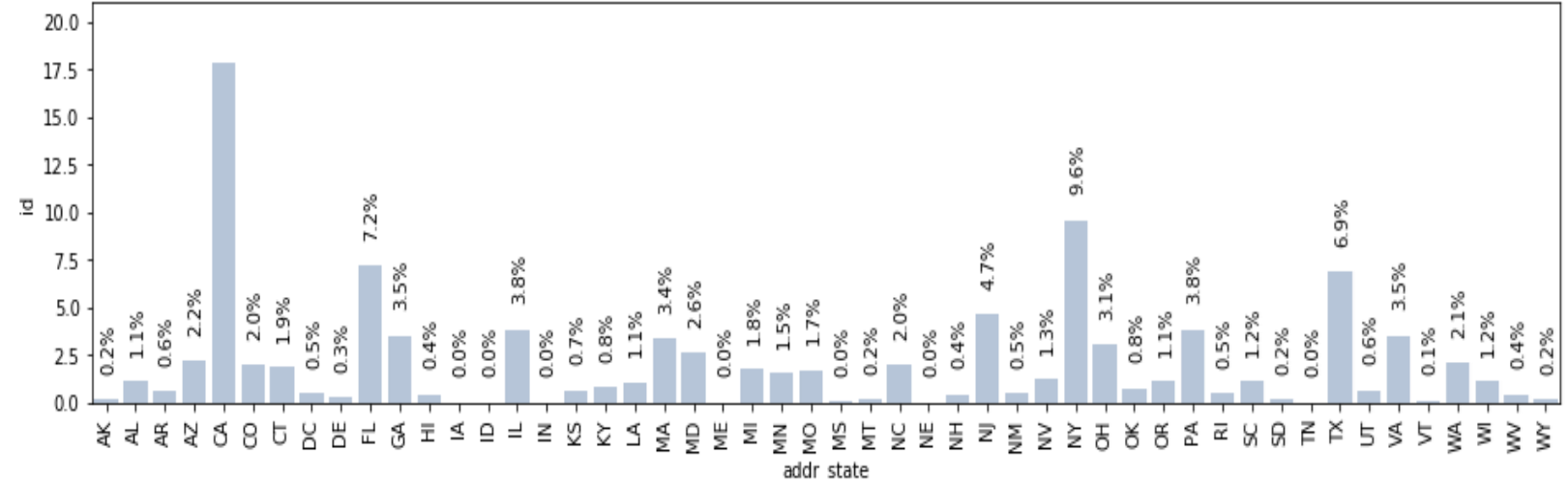
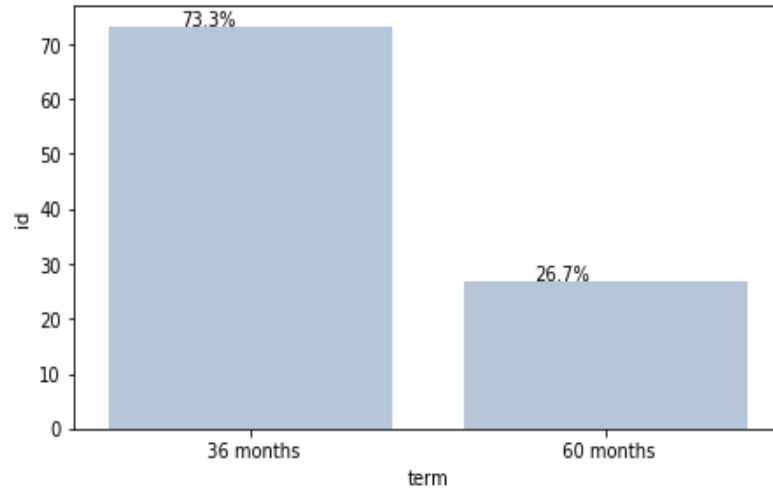


## Key Findings:

- 47% of the borrowers have requested loan for debt consolidation
- Borrowers living on rent or having a housing mortgage form 88% of the total borrower population

# Univariate Analysis for Categorical variables (contd.)

Term/State

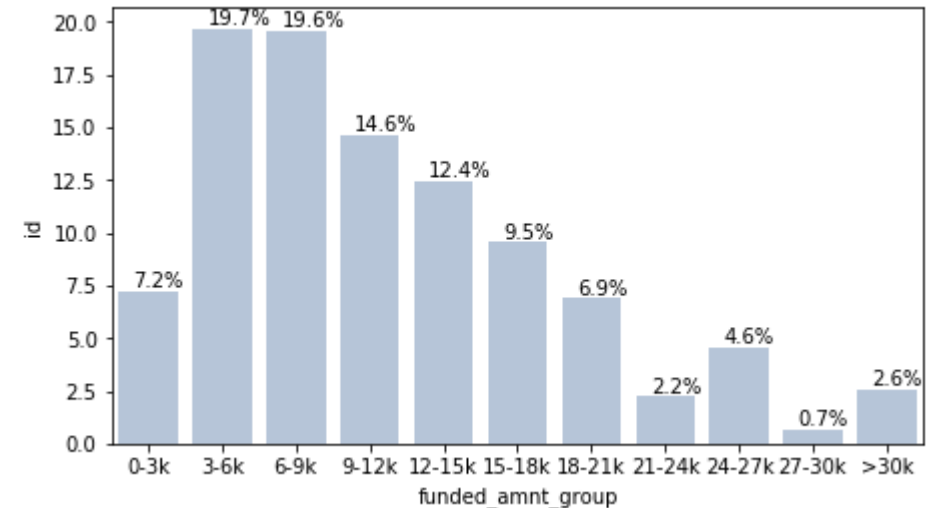
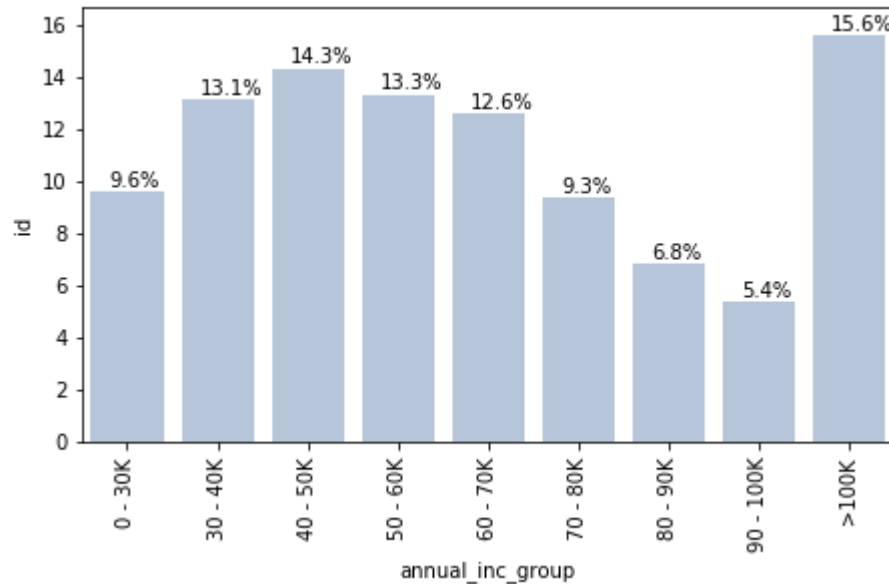


## Key Findings:

- 73% of the borrower's loan term is 36 months
- 50% of loan borrowers are from the states of CA, NY, FL, TX and NJ

# Quantitative Univariate Analysis

## Annual Income / Funded Amount

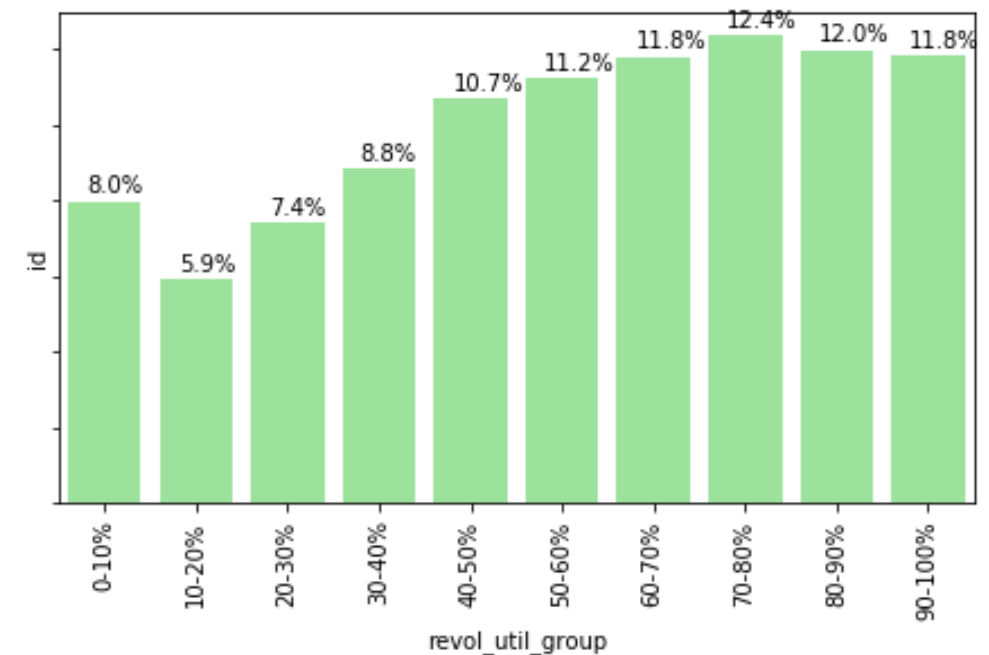
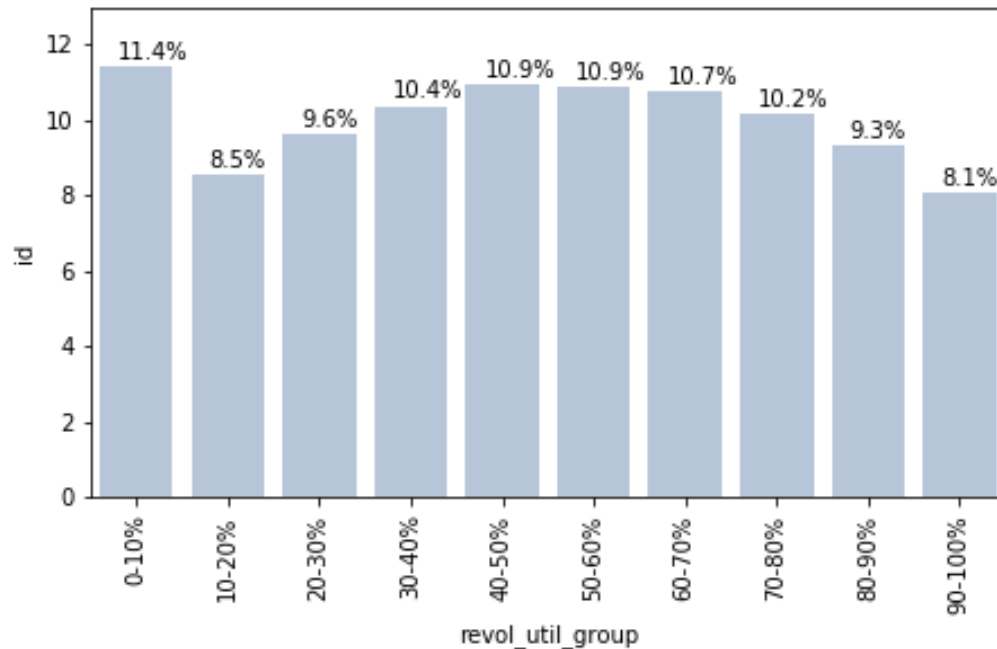


### Key Findings:

- Borrowers having income in the range of 30 -80K form 53% of the total population
- 20% of the borrowers are in the bucket of 3-6k funded amount

# Segmented Univariate Analysis for Quantitative variables

## Revolving Credit Utilization Rate for Charged off loans

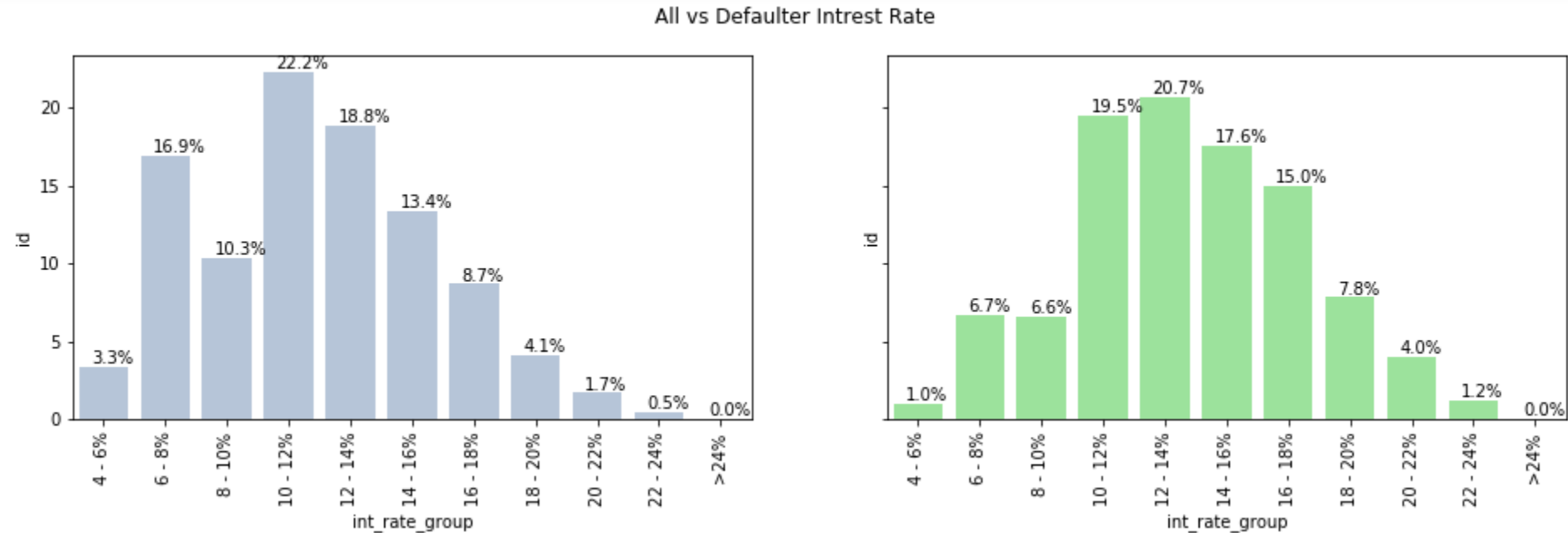


### Key Findings:

- Borrowers with revolving credit utilization rate greater than 60% show a trend of higher default rate in comparison to the overall population

# Segmented Univariate Analysis for Quantitative variables(contd.)

## Interest Rate

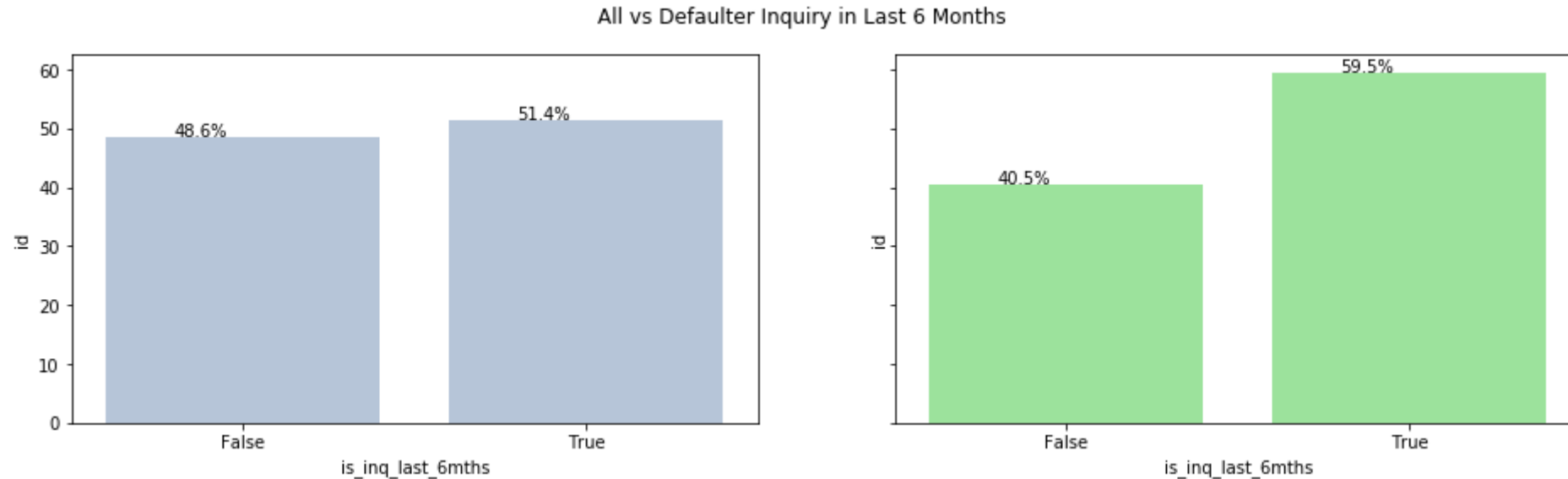


### Key Findings:

- Borrowers loans with interest rate bracket of 12 to 22 % have higher charged off % as compared to total population

# Segmented Univariate Analysis for Quantitative variables(contd.)

## Inquiry in Last 6 Months



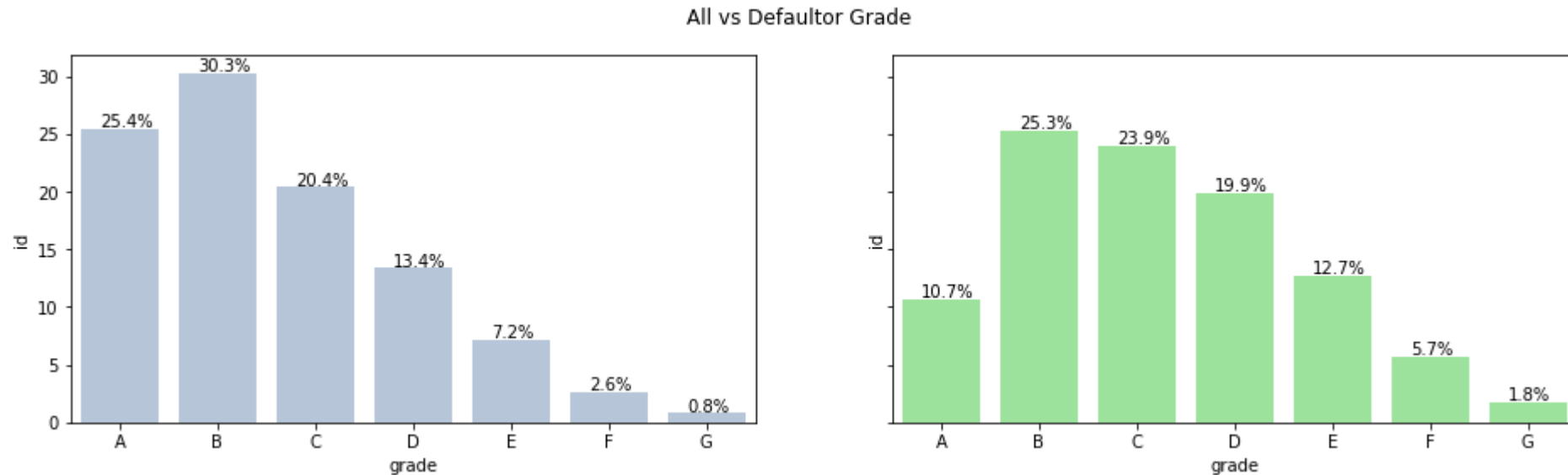
### Key Findings:

- 51% of borrowers have made an inquiry for loan/credit at least once in the last 6 months, out of which ~60% of the borrowers have defaulted



# Bivariate Analysis

## Grades



### Key Findings:

- Analyzing the assigned grades to each borrower shows that borrowers with grades like C,D and E are defaulting comparatively more than the overall population which is in contrast with lower grades F and G where the default rate is lower than the overall population

# Agenda

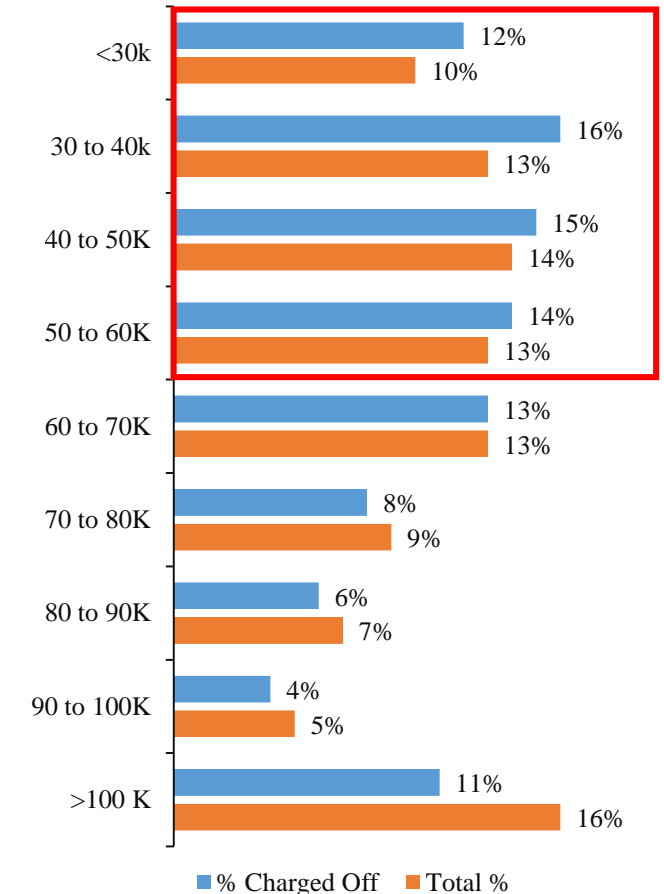
- Problem Statement
- Data Understanding
- Approach for the Analysis
- Derived Variables
- EDA (Exploratory Data Analysis) (Univariate and Bi-variate)
- **Inferences**
- Conclusion

# Annual Income Analysis

Total Loan Distribution (Count)					Total Loan Distribution (%)			% Total	
Annual_inc_bucket	Charged Off	Current	Fully Paid	Grand Total	Charged Off	Current	Fully Paid	Charged%	Total %
<30k	700	56	3043	3799	18%	1%	80%	12%	10%
30 to 40k	895	102	4205	5202	17%	2%	81%	16%	13%
40 to 50K	854	169	4663	5686	15%	3%	82%	15%	14%
50 to 60K	775	158	4360	5293	15%	3%	82%	14%	13%
60 to 70K	713	145	4135	4993	14%	3%	83%	13%	13%
70 to 80K	478	108	3125	3711	13%	3%	84%	8%	9%
80 to 90K	328	107	2270	2705	12%	4%	84%	6%	7%
90 to 100K	240	74	1814	2128	11%	3%	85%	4%	5%
>100 K	644	221	5335	6200	10%	4%	86%	11%	16%
<b>Grand Total</b>	<b>5627</b>	<b>1140</b>	<b>32950</b>	<b>39717</b>	<b>14%</b>	<b>3%</b>	<b>83%</b>	<b>100%</b>	<b>100%</b>

## Key Findings:

- Borrowers with annual income less than 60 k tend to default 7% more as compared to the income group > 60k
- The key bucket which needs attention is the income group 30 to 40k as there is a gap of 3% between the quantity of loan taken and loan charged off

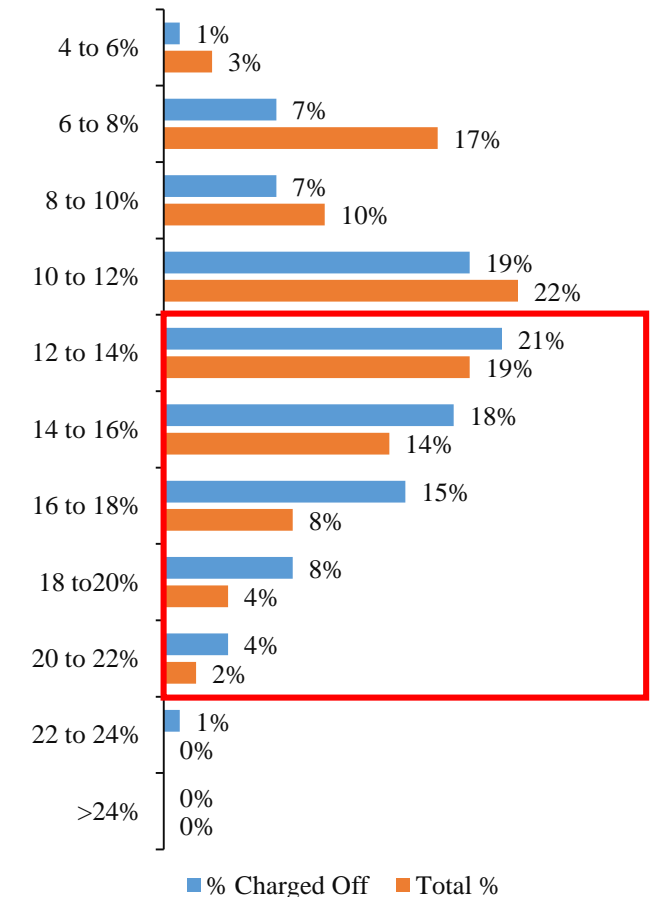


# Loan Interest Analysis

int_rate_bucket	Total Loan Distribution (Count)				Total Loan Distribution (%)			% Total	
	Charged Off	Current	Fully Paid	Grand Total	Charged Off	Current	Fully Paid	Charged%	Total %
4 to 6%	55	2	1291	1348	4%	0%	96%	1%	3%
6 to 8%	385	13	6487	6885	6%	0%	94%	7%	17%
8 to 10%	390	62	3708	4160	9%	1%	89%	7%	10%
10 to 12%	1066	231	7279	8576	12%	3%	85%	19%	22%
12 to 14%	1164	210	6095	7469	16%	3%	82%	21%	19%
14 to 16%	1005	161	4244	5410	19%	3%	78%	18%	14%
16 to 18%	826	191	2340	3357	25%	6%	70%	15%	8%
18 to 20%	440	168	1032	1640	27%	10%	63%	8%	4%
20 to 22%	225	72	388	685	33%	11%	57%	4%	2%
22 to 24%	69	29	84	182	38%	16%	46%	1%	0%
>24%	2	1	2	5	40%	20%	40%	0%	0%
<b>Grand Total</b>	<b>5627</b>	<b>1140</b>	<b>32950</b>	<b>39717</b>	<b>14%</b>	<b>3%</b>	<b>83%</b>	<b>100%</b>	<b>100%</b>

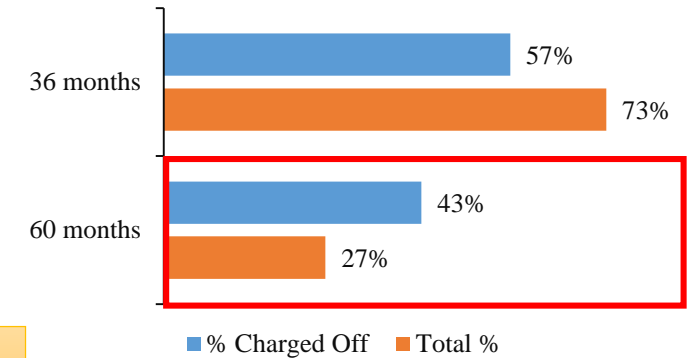
## Key Findings:

- Borrowers getting loan with interest rate between 12 to 22% have high share of charged off loans.
- Assuming that these borrowers are given loans at higher rate to avoid default. But this policy doesn't seem to be working as there are 65% of charged off loans with interest rate between 12 and 22%



# Term Analysis

Total Loan Distribution (Count)					Total Loan Distribution (%)			% Total	
Term	Charged Off	Current	Fully Paid	Grand Total	Charged Off	Current	Fully Paid	Charged%	Total %
36 months	3227		25869	29096	11%	0%	89%	57%	73%
60 months	2400	1140	7081	10621	23%	11%	67%	43%	27%
Grand Total	5627	1140	32950	39717	14%	3%	83%	100%	100%



As observed loans with tenure with 60 month have 16% more share of charged off loans as compared to the volume of loans

## Drill Down of loans with 60 months tenure

term	Int_rate_Bucket	Count		% of Total	
		Charged Off	Grand Total	Charged Off	Grand Total
60 months	16 to 18%	557	2022	10%	5%
	12 to 14%	410	1981	7%	5%
	14 to 16%	375	1673	7%	4%
	18 to 20%	366	1311	7%	3%
	10 to 12%	328	1982	6%	5%
	20 to 22%	201	619	4%	2%
	22 to 24%	67	175	1%	0%
	8 to 10%	67	491	1%	1%
	6 to 8%	26	346	0%	1%
	>24%	2	5	0%	0%
	4 to 6%	1	16	0%	0%
60 months Total		2400	10621	43%	27%

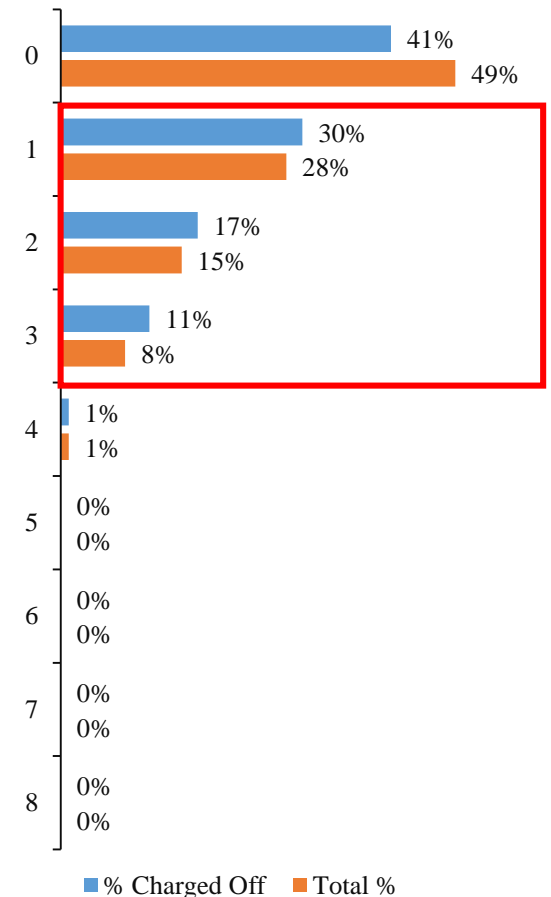
31% of the charged off loans with 60 month tenure are with interest rate between 12 to 20%

# Inquiry in past 6 Months Analysis

Total Loan Distribution (Count)					Total Loan Distribution (%)			% Total	
Inq_6_months	Charged Off	Current	Fully Paid	Grand Total	Charged Off	Current	Fully Paid	Charged%	Total %
0	2280	591	16429	19300	12%	3%	85%	41%	49%
1	1677	311	8983	10971	15%	3%	82%	30%	28%
2	943	158	4711	5812	16%	3%	81%	17%	15%
3	619	65	2364	3048	20%	2%	78%	11%	8%
4	51	10	265	326	16%	3%	81%	1%	1%
5	28	2	116	146	19%	1%	79%	0%	0%
6	16	1	47	64	25%	2%	73%	0%	0%
7	10	1	24	35	29%	3%	69%	0%	0%
8	3	1	11	15	20%	7%	73%	0%	0%
Grand Total	5627	1140	32950	39717	14%	3%	83%	100%	100%

## Key Findings:

- Customers with any credit inquiry during the past 6 month tend to default more as compared to the customer with no credit enquiry
- Customers with 1 to 3 credit inquiry contributes to 58% of charged off loans and where contribute only 50% of the total loans

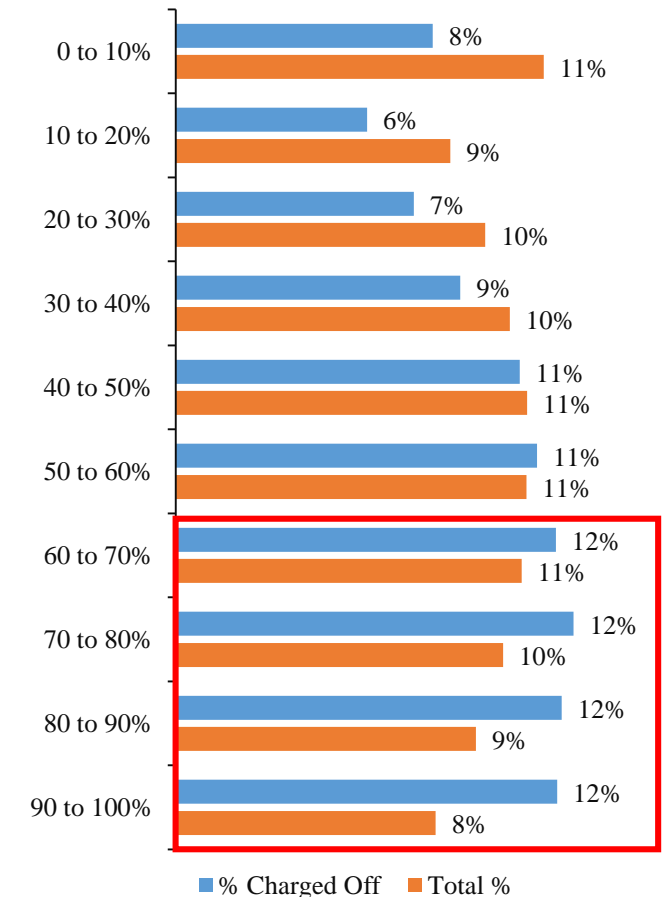


# Revolving Credit Percentage Analysis

revol_utl_ %	Total Loan Distribution (Count)				Total Loan Distribution (%)			% Total	
	Charged Off	Current	Fully Paid	Grand Total	Charged Off	Current	Fully Paid	Charged%	Total %
0 to 10%	449	85	4002	4536	10%	2%	88%	8%	11%
10 to 20%	334	71	2980	3385	10%	2%	88%	6%	9%
20 to 30%	416	101	3299	3816	11%	3%	86%	7%	10%
30 to 40%	497	119	3503	4119	14%	3%	83%	9%	10%
40 to 50%	601	118	3614	4333	15%	3%	82%	11%	11%
50 to 60%	631	136	3556	4323	16%	4%	81%	11%	11%
60 to 70%	664	154	3447	4265	17%	3%	79%	12%	11%
70 to 80%	695	134	3206	4035	18%	3%	78%	12%	10%
80 to 90%	674	129	2898	3701	21%	3%	76%	12%	9%
90 to 100%	666	93	2445	3204	12%	3%	85%	12%	8%
<b>Grand Total</b>	<b>5627</b>	<b>1140</b>	<b>32950</b>	<b>39717</b>	14%	3%	83%	100%	100%

## Key Findings:

- Borrower's with revolving credit % greater than 60% have higher charged off loan % at 48% whereas they contribute to only 38% of the loans



# State

**% Total**

Charged%	Total %
20%	18%
9%	10%
9%	7%
6%	7%
5%	5%
4%	4%
3%	4%
3%	4%
4%	4%
3%	3%
3%	3%
3%	3%
2%	2%
2%	2%
2%	2%
2%	2%
2%	2%
2%	2%
1%	2%
14%	15%
100%	100%

- 
- | State  | % Charged Off | Total % |
|--------|---------------|---------|
| CA     | 20%           | 18%     |
| NY     | 9%            | 10%     |
| FL     | 9%            | 7%      |
| TX     | 6%            | 7%      |
| NJ     | 5%            | 5%      |
| IL     | 4%            | 4%      |
| PA     | 3%            | 4%      |
| VA     | 3%            | 4%      |
| GA     | 4%            | 4%      |
| MA     | 3%            | 3%      |
| OH     | 3%            | 3%      |
| MD     | 3%            | 3%      |
| AZ     | 2%            | 2%      |
| WA     | 2%            | 2%      |
| CO     | 2%            | 2%      |
| NC     | 2%            | 2%      |
| CT     | 2%            | 2%      |
| MI     | 2%            | 2%      |
| MO     | 2%            | 2%      |
| MN     | 1%            | 2%      |
| Others | 14%           | 15%     |

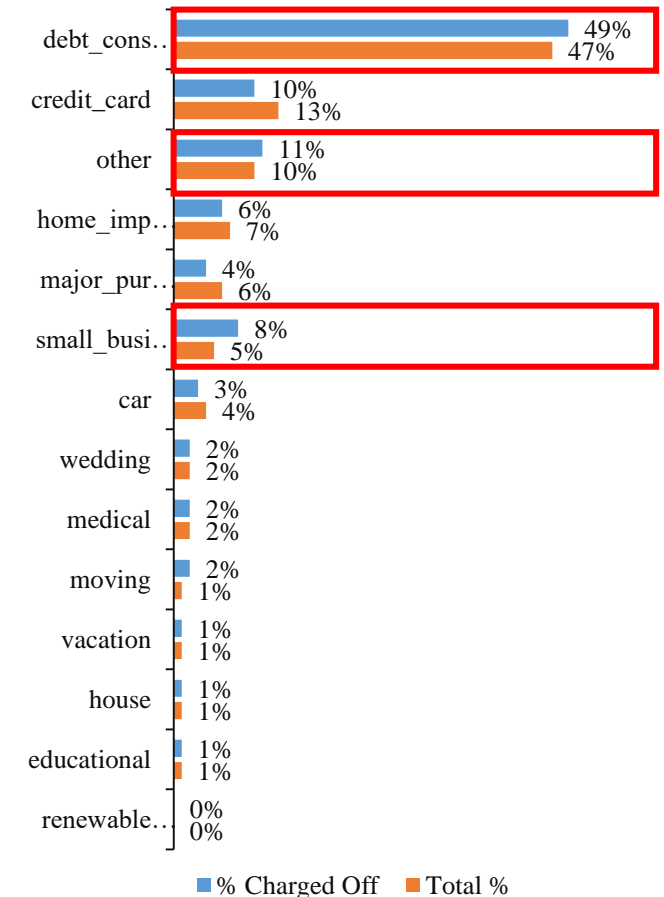


# Purpose

Total Loan Distribution (Count)					Total Loan Distribution (%)			% Total	
purpose	Charged Off	Current	Fully Paid	Grand Total	Charged Off	Current	Fully Paid	Charged%	Total %
debt_consolidation	2767	586	15288	18641	15%	3%	82%	49%	47%
credit_card	542	103	4485	5130	11%	2%	87%	10%	13%
other	633	128	3232	3993	16%	3%	81%	11%	10%
home_improvement	347	101	2528	2976	12%	3%	85%	6%	7%
major_purchase	222	37	1928	2187	10%	2%	88%	4%	6%
small_business	475	74	1279	1828	26%	4%	70%	8%	5%
car	160	50	1339	1549	10%	3%	86%	3%	4%
wedding	96	21	830	947	10%	2%	88%	2%	2%
medical	106	12	575	693	15%	2%	83%	2%	2%
moving	92	7	484	583	16%	1%	83%	2%	1%
vacation	53	6	322	381	14%	2%	85%	1%	1%
house	59	14	308	381	15%	4%	81%	1%	1%
educational	56		269	325	17%	0%	83%	1%	1%
renewable_energy	19	1	83	103	18%	1%	81%	0%	0%
<b>Grand Total</b>	<b>5627</b>	<b>1140</b>	<b>32950</b>	<b>39717</b>	14%	3%	83%	100%	100%

## Key Findings:

- 49% of the charged off loans are for debt consolidation
- Borrowers with purpose of small business have 8% charged off loans



# Total Payment

Total Loan Distribution (Count)

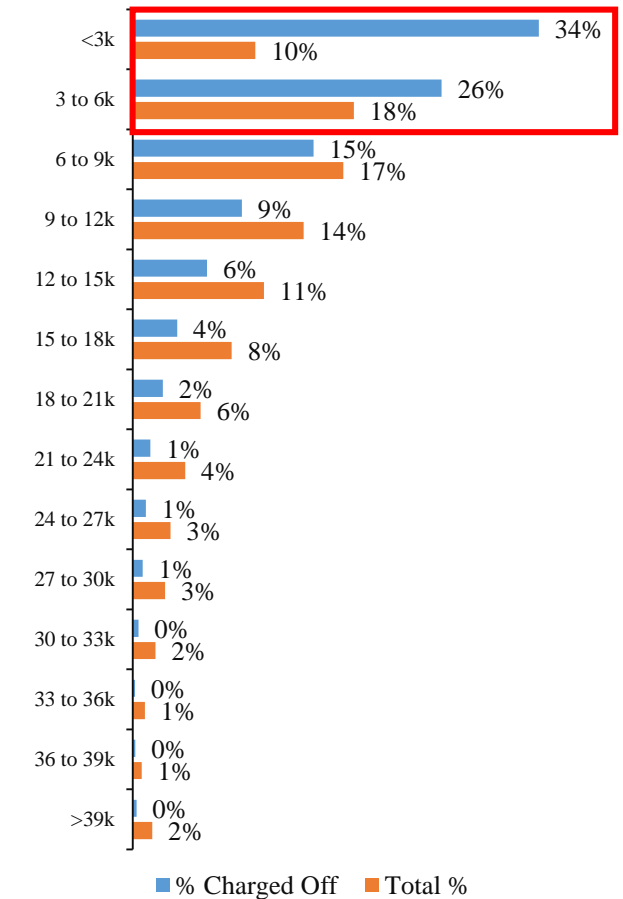
total_pymnt_group	Charged Off	Current	Fully Paid	Grand Total
<3k	1891	13	2126	4030
3 to 6k	1439	57	5770	7266
6 to 9k	842	88	5989	6919
9 to 12k	508	65	5041	5614
12 to 15k	346	104	3861	4311
15 to 18k	207	139	2903	3249
18 to 21k	140	131	1963	2234
21 to 24k	82	95	1545	1722
24 to 27k	61	99	1083	1243
27 to 30k	46	94	919	1059
30 to 33k	26	70	646	742
33 to 36k	10	57	332	399
36 to 39k	12	38	242	292
>39k	17	90	530	637
<b>Grand Total</b>	<b>5627</b>	<b>1140</b>	<b>32950</b>	<b>39717</b>

Total Loan Distribution (%)

Charged Off	Current	Fully Paid
47%	0%	53%
20%	1%	79%
12%	1%	87%
9%	1%	90%
8%	2%	90%
6%	4%	89%
6%	6%	88%
5%	6%	90%
5%	8%	87%
4%	9%	87%
4%	9%	87%
3%	14%	83%
4%	13%	83%
3%	14%	83%
14%	3%	83%

% Total

Charged%	Total %
34%	10%
26%	18%
15%	17%
9%	14%
6%	11%
4%	8%
2%	6%
1%	4%
1%	3%
1%	3%
0%	2%
0%	1%
0%	1%
0%	2%
0%	1%
0%	1%
0%	2%
100%	100%



## Key Findings:

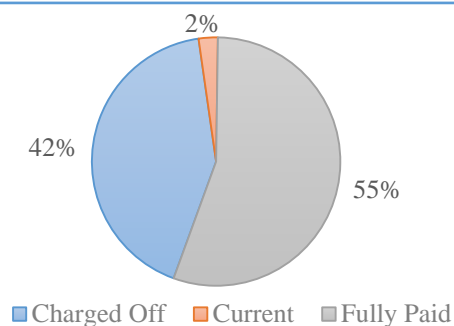
- Small payments which are <6k are hurting the most
- 60% of the charged off loans are below 6K

# Late Fee, Recovery and Collection Fee

## Late Fees

There are 2046 loans which are levied late fees.

loan_status	Loan Count	% of Loan Count
Charged Off	863	42%
Current	51	2%
Fully Paid	1132	55%
<b>Grand Total</b>	<b>2046</b>	<b>100%</b>



- 42% loans are charged off.
- Borrowers who are failing to pay the installment on time have high chances to default (charged off).

## Recoveries

There are 4218 loans which are recoveries charges.

home_ownership	Loan Count	% of Loan Count
RENT	2208	52%
MORTGAGE	1658	39%
OWN	340	8%
OTHER	12	0%
<b>Grand Total</b>	<b>4218</b>	<b>100%</b>

- All the 4218 loans have the loan status as Charged off
- 52% borrowers with recoveries have rented home ownership
- 39% borrowers with recoveries have mortgaged home

## Collection Recovery Fees

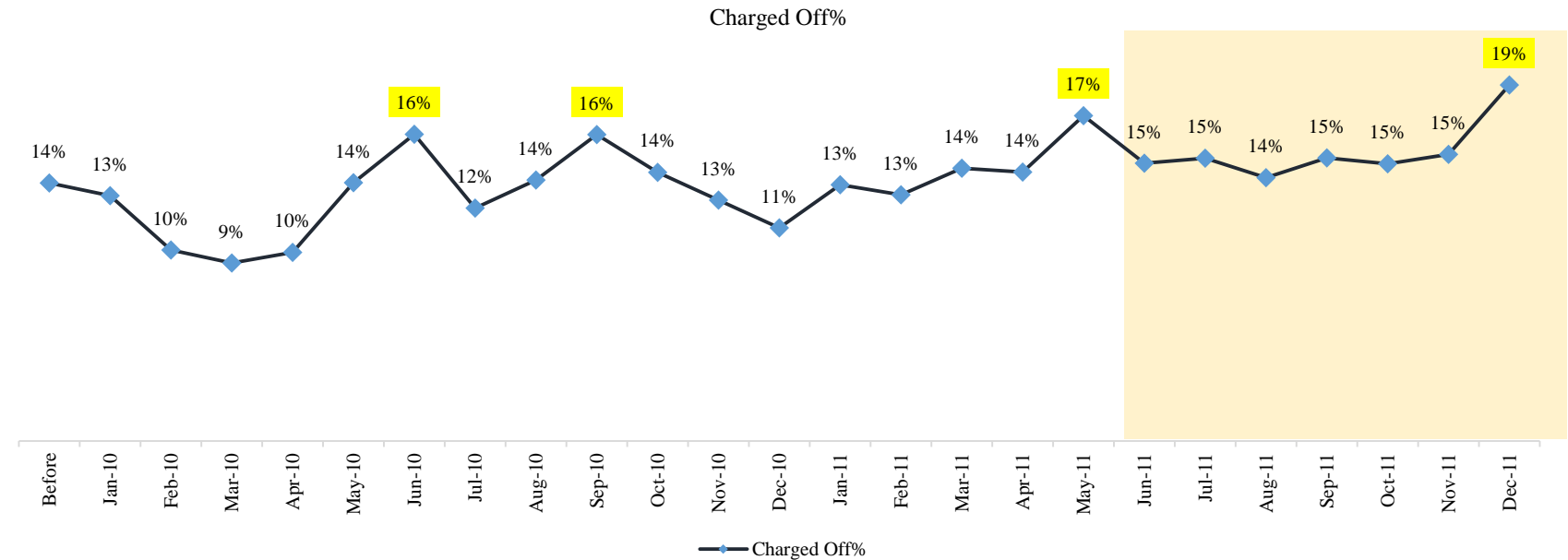
There are 3782 loans which are collection charges.

home_ownership	Loan Count	% of Loan Count
RENT	2039	54%
MORTGAGE	1421	38%
OWN	310	8%
OTHER	12	0%
<b>Grand Total</b>	<b>3782</b>	<b>100%</b>

- All the 3782 loans have the loan status as Charged off
- 54% borrowers with recoveries have rented home ownership
- 38% borrowers with recoveries have mortgaged home

# Issue Date

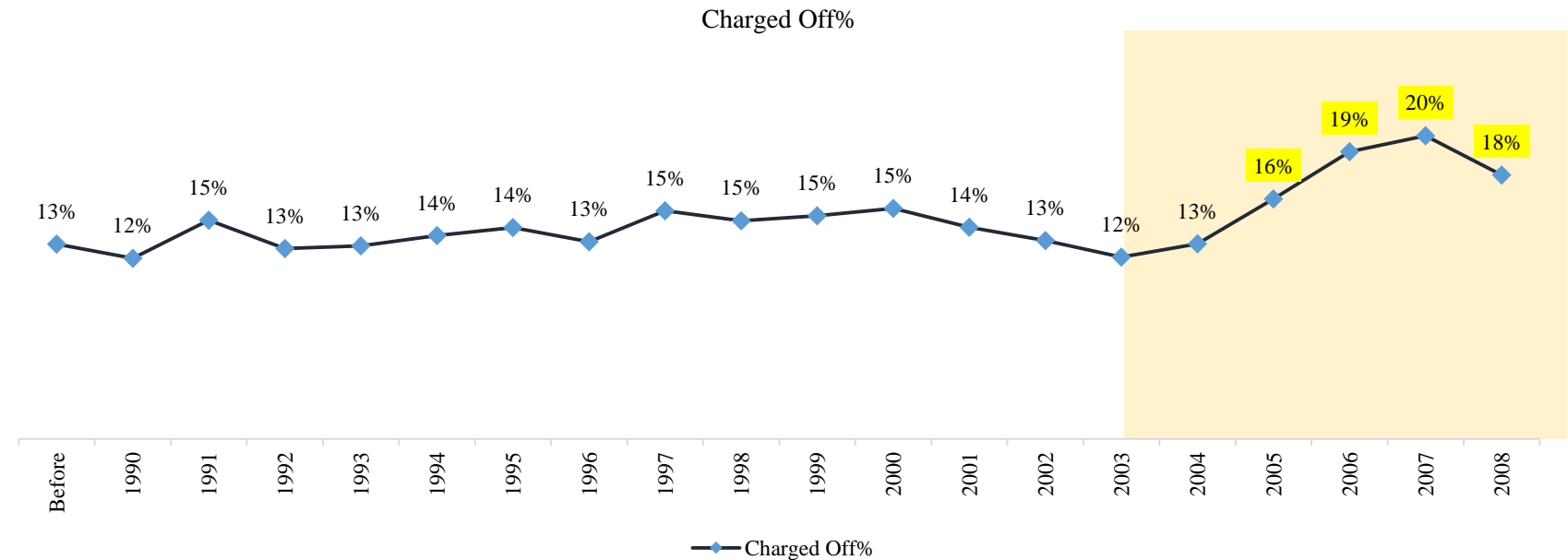
Issue Date	Total (#)	Charged off(#)	Charged Off%
Before	6529	886	14%
Jan-10	589	76	13%
Feb-10	627	63	10%
Mar-10	737	69	9%
Apr-10	827	82	10%
May-10	920	125	14%
Jun-10	1029	166	16%
Jul-10	1119	137	12%
Aug-10	1078	148	14%
Sep-10	1086	175	16%
Oct-10	1132	160	14%
Nov-10	1121	142	13%
Dec-10	1267	142	11%
Jan-11	1380	186	13%
Feb-11	1297	168	13%
Mar-11	1443	207	14%
Apr-11	1562	221	14%
May-11	1689	289	17%
Jun-11	1827	267	15%
Jul-11	1870	278	15%
Aug-11	1928	267	14%
Sep-11	2063	307	15%
Oct-11	2114	308	15%
Nov-11	2223	335	15%
Dec-11	2260	423	19%
<b>Grand Total</b>	<b>39717</b>	<b>5627</b>	<b>14%</b>



- As observed from the above trend month in July-10, Sep-10, May-10 and Dec-11 have seen a spike in terms of charged of loan%
- December 2011 has the highest Charged off % of loans at 19%

# Earliest Credit Line

Issue Date	Total (#)	Charged off(#)	Charged Off%
Before	5548	721	13%
1990	1038	125	12%
1991	980	143	15%
1992	1063	135	13%
1993	1514	195	13%
1994	2027	275	14%
1995	2150	303	14%
1996	2274	299	13%
1997	2456	374	15%
1998	2851	415	15%
1999	2968	442	15%
2000	3250	500	15%
2001	2521	356	14%
2002	1927	255	13%
2003	1856	225	12%
2004	1761	229	13%
2005	1493	239	16%
2006	1169	224	19%
2007	712	144	20%
2008	159	28	18%
<b>Grand Total</b>	<b>39717</b>	<b>5627</b>	<b>14%</b>



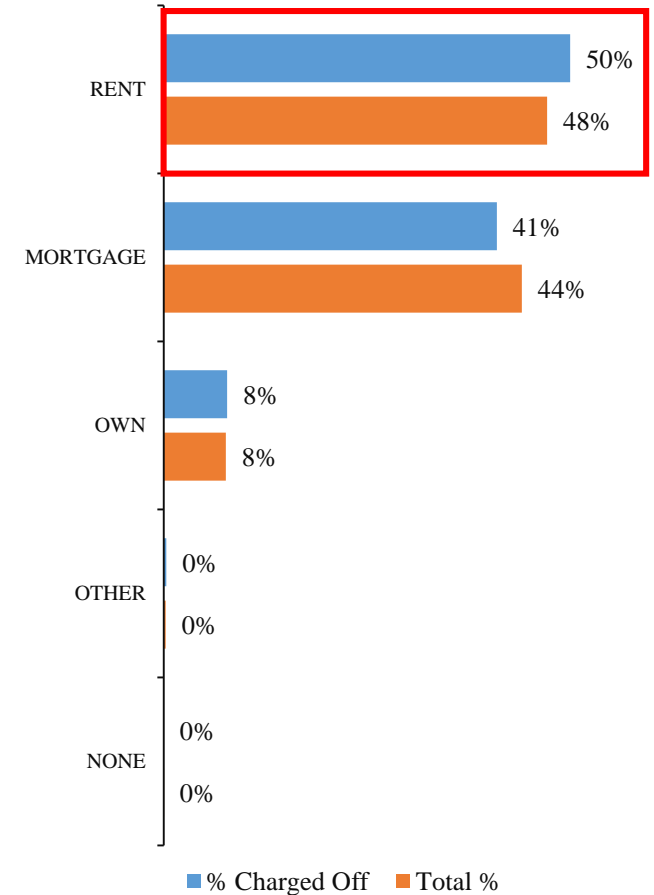
- As observed from the above trend month post 2003 the charged off % have been on an increasing trend and continued till 2007
- Even though there is a drop beyond 2007 but still there is difference of 4% as compared to overall average of (14%) charged off loans

# Home Ownership

Total Loan Distribution (Count)					Total Loan Distribution (%)			% Total	
home_ownership	Charged Off	Current	Fully Paid	Grand Total	Charged Off	Current	Fully Paid	Charged%	Total %
RENT	2839	419	15641	18899	15%	2%	83%	50%	48%
MORTGAGE	2327	638	14694	17659	13%	4%	83%	41%	44%
OWN	443	83	2532	3058	14%	3%	83%	8%	8%
OTHER	18		80	98	18%	0%	82%	0%	0%
NONE			3	3	0%	0%	100%	0%	0%
<b>Grand Total</b>	<b>5627</b>	<b>1140</b>	<b>32950</b>	<b>39717</b>	<b>14%</b>	<b>3%</b>	<b>83%</b>	<b>100%</b>	<b>100%</b>

## Key Findings:

- Borrowers with Rented home ownership tend to take more loans and also have significantly higher % of charged off loans
- Borrowers with Mortgaged ownership take 44% loans but only contribute to 41% of charged off loans.

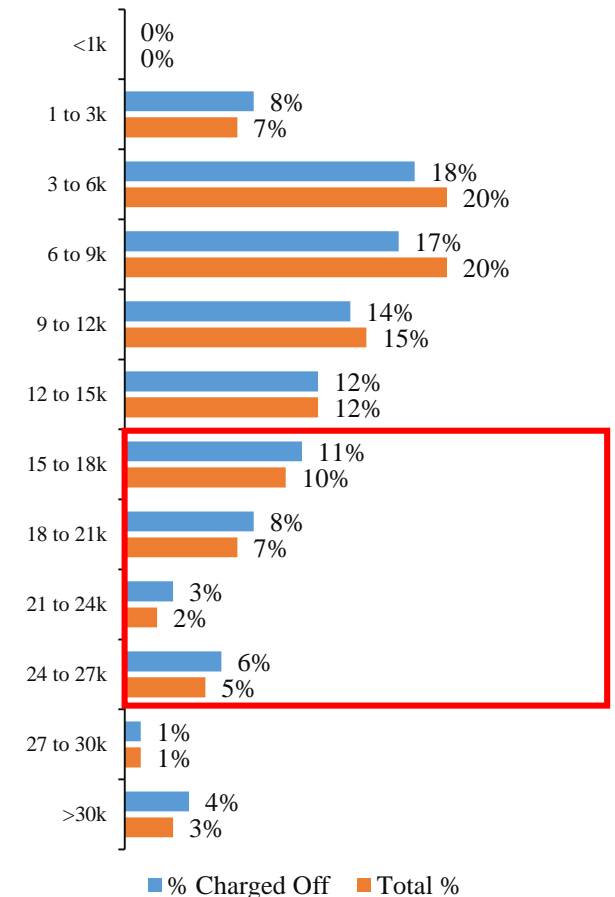


# Funded Amount

Total Loan Distribution (Count)					Total Loan Distribution (%)			% Total	
funded_amnt_group	Charged Off	Current	Fully Paid	Grand Total	Charged Off	Current	Fully Paid	Charged%	Total %
<1k	1		11	12	8%	0%	92%	0%	0%
1 to 3k	426	20	2414	2860	15%	1%	84%	8%	7%
3 to 6k	1031	92	6683	7806	13%	1%	86%	18%	20%
6 to 9k	970	103	6699	7772	12%	1%	86%	17%	20%
9 to 12k	761	77	4978	5816	13%	1%	86%	14%	15%
12 to 15k	656	224	4060	4940	13%	5%	82%	12%	12%
15 to 18k	595	159	3032	3786	16%	4%	80%	11%	10%
18 to 21k	456	154	2127	2737	17%	6%	78%	8%	7%
21 to 24k	164	84	634	882	19%	10%	72%	3%	2%
24 to 27k	326	98	1393	1817	18%	5%	77%	6%	5%
27 to 30k	36	26	209	271	13%	10%	77%	1%	1%
>30k	205	103	710	1018	20%	10%	70%	4%	3%
<b>Grand Total</b>	<b>5627</b>	<b>1140</b>	<b>32950</b>	<b>39717</b>	<b>14%</b>	<b>3%</b>	<b>83%</b>	<b>100%</b>	<b>100%</b>

## Key Findings:

- There is no major variation amount the amount funded and the % charged off loans
- There is a very marginal increase in charged off loan % for loans with funding amount between 15k to 27k.



# Agenda

- Problem Statement
- Data Understanding
- Approach for the Analysis
- Derived Variables
- EDA (Exploratory Data Analysis) (Univariate and Bi-variate)
- Inferences
- **Conclusion**



# Summary

## *Factors impacting the Charged off Loans*

- Customer with total payment less than 6k is impacting the charged off loan % **by 31%**
- Loans with the interest rate between 12 to 22 % is driving the charged off loans **by 18%**
- Loan tenure of 60 months is impacting the charged off loans by **16%**
- Borrowers with revolving credit percentage greater than 60% has an impact of **10%**
- Borrowers with 1 to 3 inquires in past 6 months is impacting charged off loans **by 8%**
- Borrowers with annual income less than 60k is driving the charged off loan % **by 7%**
- Others and small business loans are impacting the charged off loans **by 5%**
- CA and FL jointly impact the charged off loan % **by 4%**

Thank You