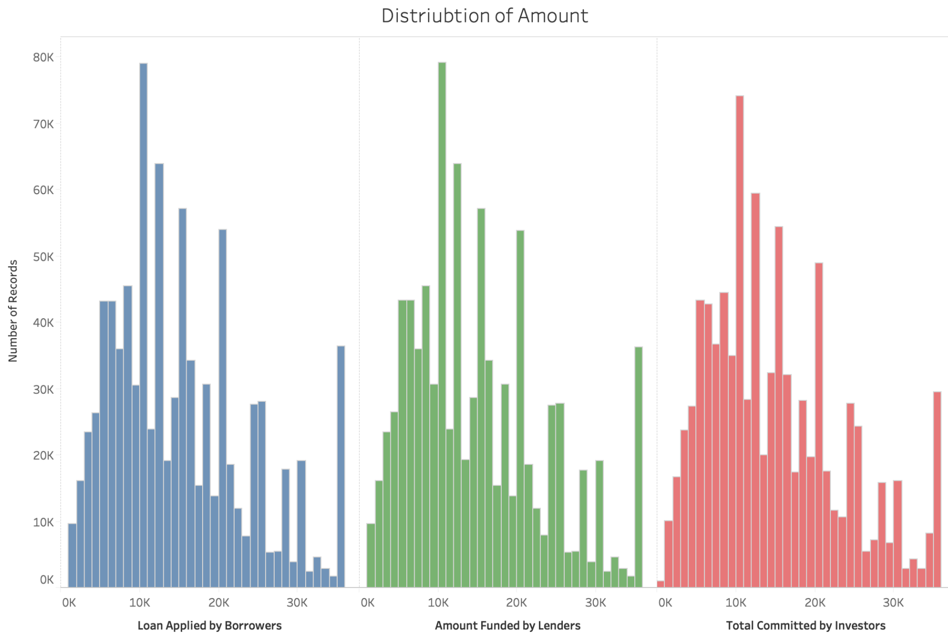
Descriptive Analysis

Introduction

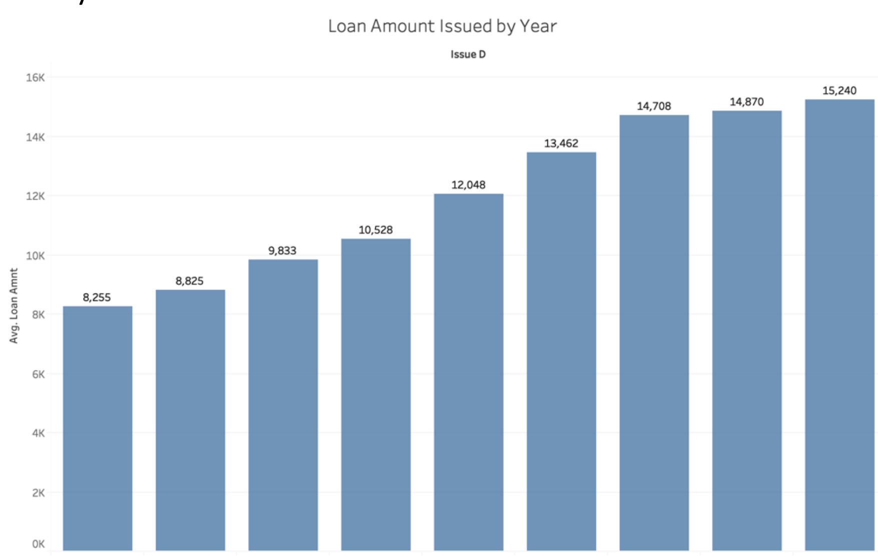
a) **General Information**
887379 rows * 74 columns

b) **Similar Distribution**

- As the distribution graph showed, most of the loans issued to borrowers were around $10,000 - $20,000. And all these three distributions (borrower/lender/investor) are similar, which means the qualified borrowers may have a great change to get the loan they applied for.



- The average loan amount issued per year is in an increasing trend, possible because the economic recovery in U.S. Year 2015 issued the most loans.



Good Loans vs. Bad Loans

a) Type of Loans

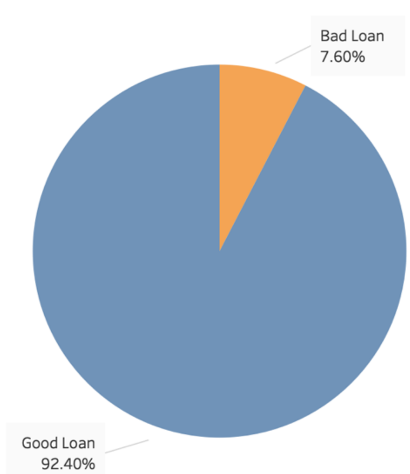
- There are 10 types of loan status, and we divided them as good loans or bad loans just as the below table showed. One point has to be clear is that, the current and issued loans for now are considered as good loans. But we do not know what they will be in future, thus the amount of bad loans may increase as the days pass by.

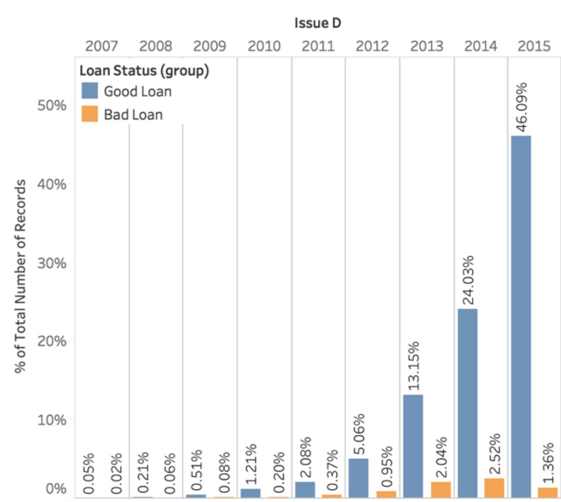
Loan Status Summary

Loan Status..	Loan Status	
Bad Loan	Charged Off	45,248
	Default	1,219
	Does not meet the credit ..	761
	In Grace Period	6,253
	Late (16-30 days)	2,357
	Late (31-120 days)	11,591
Good Loan	Current	601,779
	Does not meet the credit ..	1,988
	Fully Paid	207,723
	Issued	8,460

- Bad loans make up 7.6% of total loans for now, there may be an increase of it due to the current and issued loans which have risk falling into default. There are more and more loans (number of records) issued as the year passes.

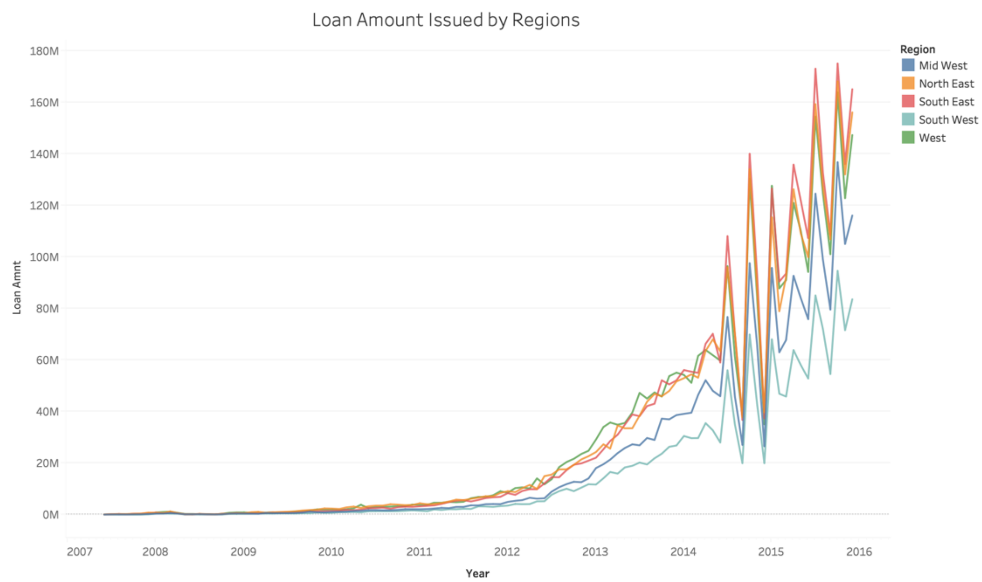
Comparison of Loan Status



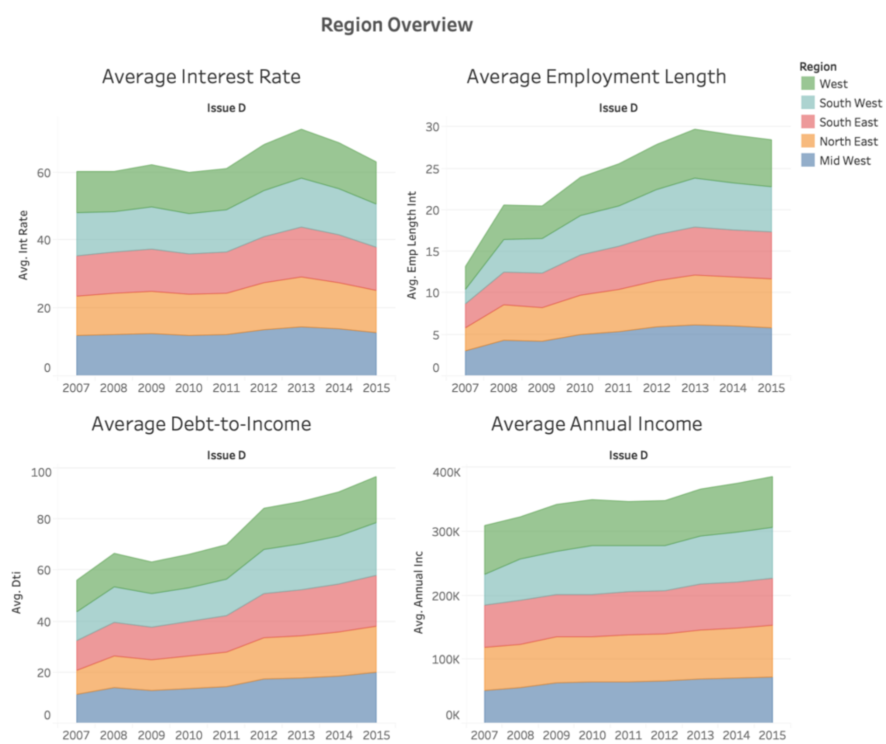


b) Loans Issued by Region

- Southeast seems to be the most attractive region for the funding loans to borrowers, since it basically has the highest loan amount through the years. And the Northeast and West also ranks top of the loan amount issued than others.



- Average interest rates for all regions have decreased since 2013, and this may affect the increase in the volume of loans. And as the area chart showed, West and Southwest had a rapid decrease in interest rates.
- West and Southwest have higher Employment length than the other regions.
- West and Southwest had a rapid increase in debt-to-income since 2012, this may be caused by the rapid decrease in interest rate in these regions. And compared to these, Midwest and Northeast have a gentler situation of debt-to-income, no drastic change in it
- And for average annual income, Southwest and West have a slight increase in it through the years.



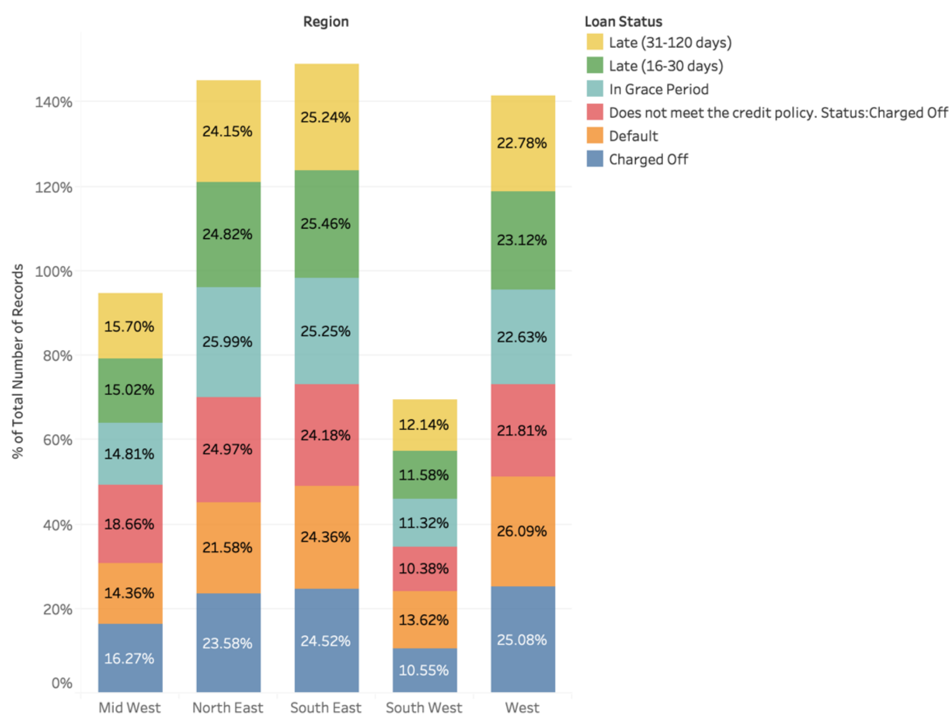
c) A Deeper Look into Bad Loans by Region

- Southeast, West and Northeast had more bad loans than others and had a higher percentage in most of the loan statuses.
- Northeast had higher percentage in Grace Period and Does not meet the credit policy: Charged Off these two statuses.
- Southeast and West have a slightly higher percentage in specific Default status compared to Northeast.
- Basically, we could conclude that Southeast, West and Northeast have higher risk level of falling into bad loans than other regions.

Bad Loans Summary

Region	Charged Off	Default	Does not meet the credit policy. Status:..	In Grace Period	Late (16-30 days)	Late (31-120 days)	Grand Total
Mid West	7,361	175	142	926	354	1,820	**10,778**
North East	10,671	263	190	1,625	585	2,799	**16,133**
South East	11,094	297	184	1,579	600	2,925	**16,679**
South West	4,774	166	79	708	273	1,407	**7,407**
West	11,348	318	166	1,415	545	2,640	**16,432**

Percentage of Bad Loan Status by Region



The Business Perspective

a) Understanding the Operative Side of Business by State

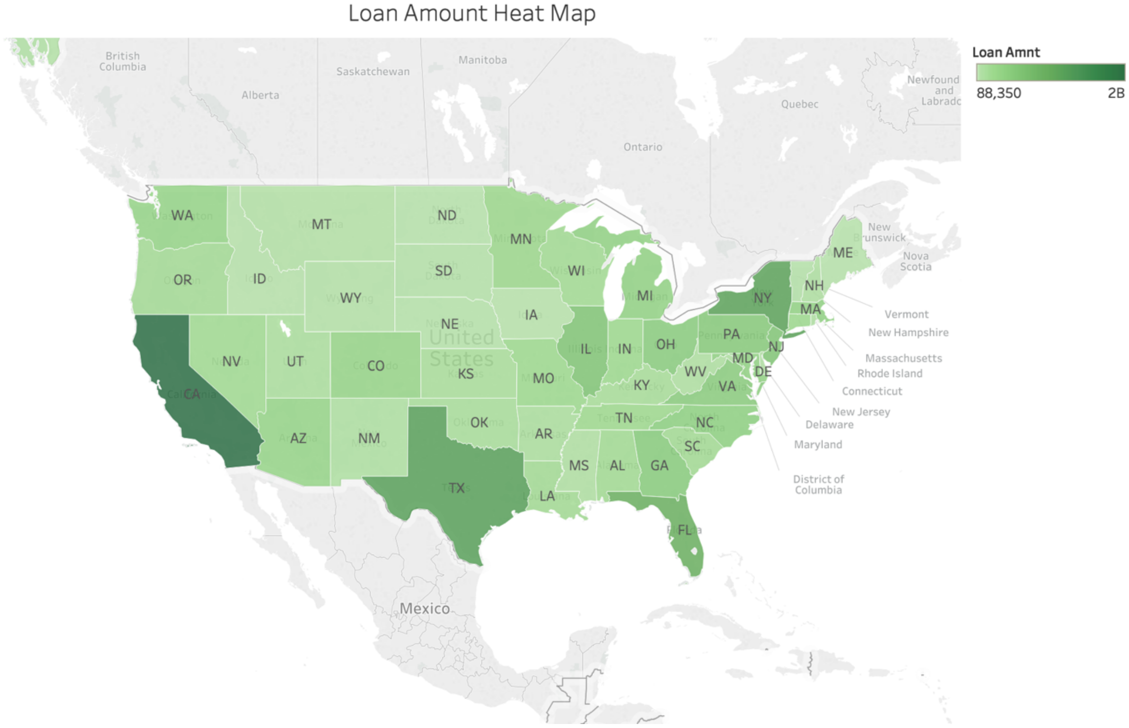
- Focus on three major aspects, total loans amount issued, average interest rate and average annual income of borrowers by State. We'd like to find the states which give high returns with low risk level. The state summary table shows the top 10 loan amount issued states.

- Based on the heat map, California, Texas, New York and Florida have higher loan amount issued through the years.
- All these 4 states have the similar average interest rate level as the entire U.S, which is around 13%.
- The average annual income of the top 3 loans amount issued states (CA, TX, NY) are all above the average level, which may be the reason why most loans are issued in these states.

State Summary

Addr State	Loan Amnt	Avg. Int Rate	Avg. Annual Inc
Grand Total	**256,735,528**	**13.20**	**71,067**
CA	1,898,145,250	13.16	79,294
TX	1,109,129,575	13.16	80,471
NY	1,076,941,975	13.33	78,295
FL	848,989,075	13.27	70,293
IL	539,068,450	13.10	76,898
NJ	515,253,000	13.15	85,160
PA	455,642,875	13.33	71,102
GA	434,709,600	13.27	75,096
OH	422,675,375	13.33	66,931
VA	411,222,050	13.29	80,938

Loan Amount Heat Map



b) Analysis by Income Category

- Based on the previous analysis, average annual income may be a reason indicate the loan amount issued. Thus, we'd like to give a deeper view on the patterns they may exist. First cut the annual income amount into low, medium and high levels. (Low <= $100K, $100K < Medium <=$200K, High >= $200K)

- Borrowers with high income usually has more loan amounts issued than the low income and medium income groups. Of course, the loan amounts can be issued depend on the annual income of each borrower.
- Borrowers which were in the low income category had higher chance to make loans issued fall into bad loans.
- Low income group usually had shorter employment length and higher interest rate than the medium and high income group.

Assessing Risks

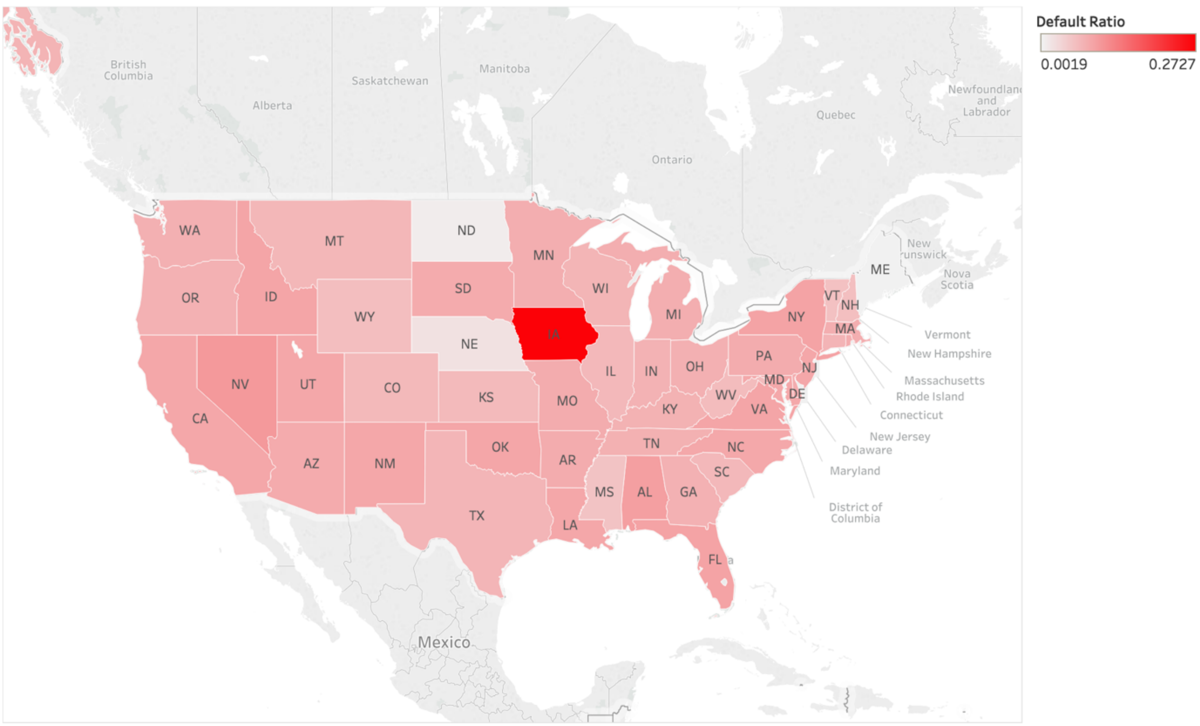
a) Understanding the Risky Side of Business

- Default ratio = number of bad loans /number of good loans, which will be a metric to evaluate the risk level of each state.
- Debt-to-income is also an important metric which show the level of debt of each individual consumer with respect to its total income.
- The employment length also helps to assess the level of risk due to it connected with the labor market.
- IA has the highest default ratio, and it's much higher than the other states, but the volume of loans issued by IA is too low. There are only 3 bad loans occurred in IA.

State Summary Part 2

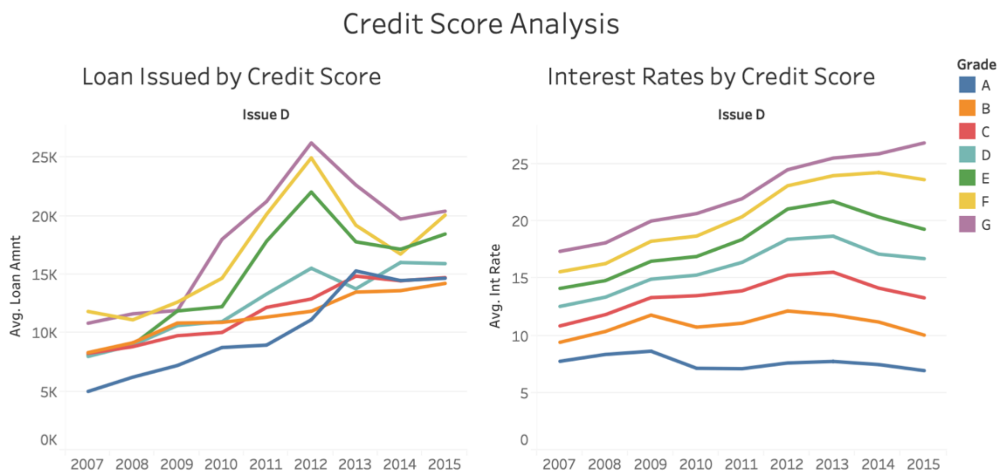
Addr State	Default Ratio	Count Bad Loan	Percentage of Bad Loans	Avg. Dti	Avg. Emp Length Int
IA	0.273	3	0.004%	13.656	3.357
NV	0.104	1,173	1.740%	18.296	5.501
HI	0.103	427	0.633%	19.237	6.276
AL	0.097	993	1.473%	20.042	6.153
NY	0.093	6,289	9.327%	16.542	5.700
FL	0.092	5,148	7.635%	18.410	5.413
ID	0.091	1	0.001%	15.388	2.917
VA	0.090	2,167	3.214%	18.282	5.753
OK	0.090	665	0.986%	19.493	5.698
NM	0.089	402	0.596%	19.761	5.715

Default Rates Heat Map

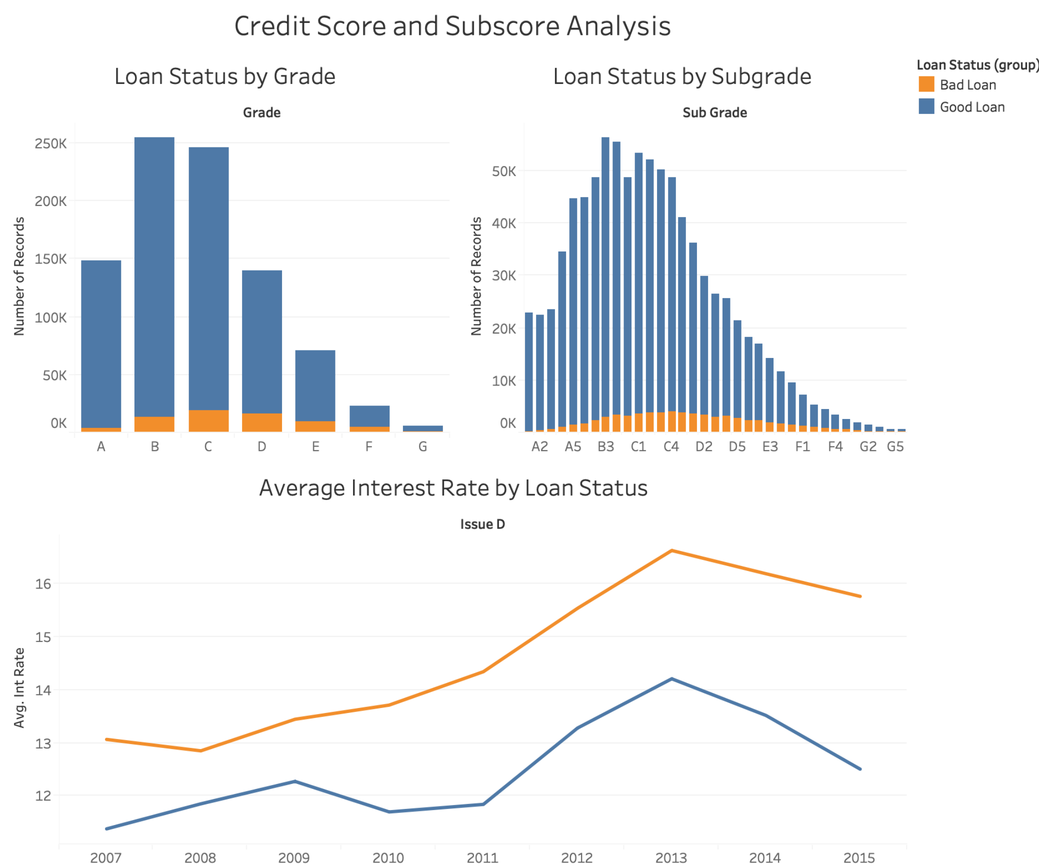


b) The Importance of Credit Scores

- Credit score will be an important metric to evaluate the risk level of each customer. The lower the grade of the credit score, the higher the risk for investors.
- The group with lower credit score had been issued more loan amounts, this may because there is a bigger customer base for the lower credit score group. This might had contributed to a higher level of risk.
- Lower credit score will generate higher interest rate for borrowers.



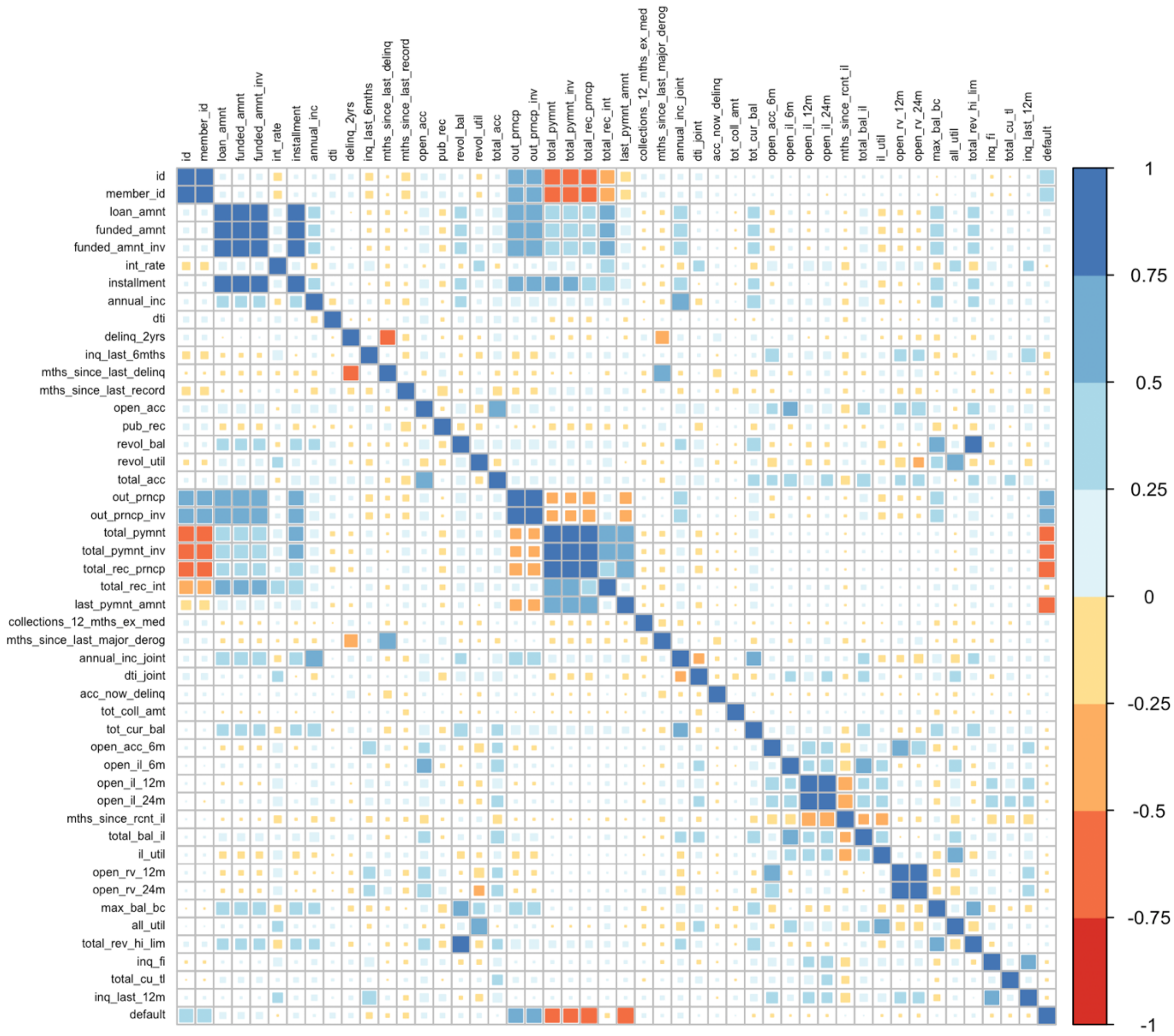
- The borrowers with credit score of "C" has the highest possibility to fall into default, especially for C4 group.
- Logically, the bad loans have higher interest rate than the good ones.



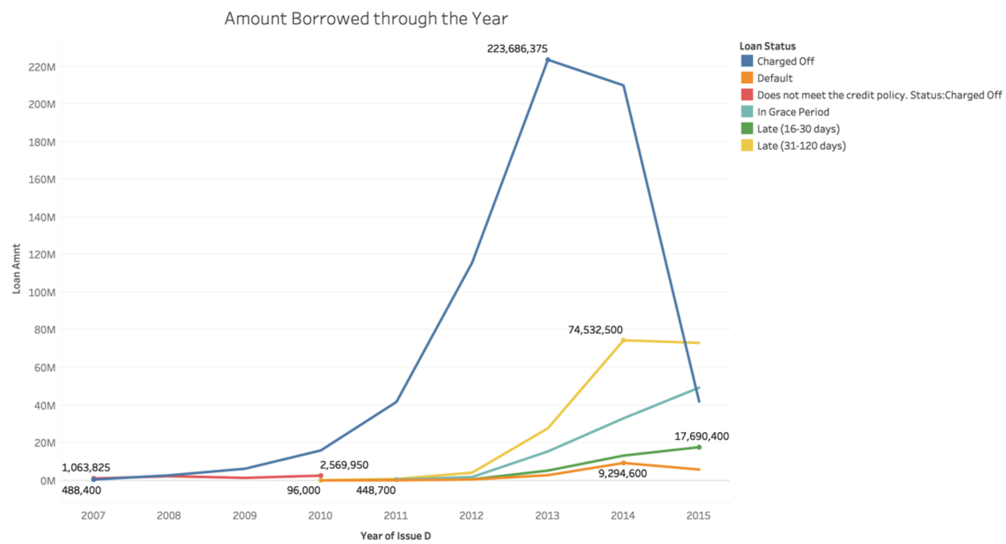
c) What Determines a Bad Loan?

- Logically, assume possible factors that will increase the risk level: low annual income, high interest rate, low debt-to-income and low credit score.

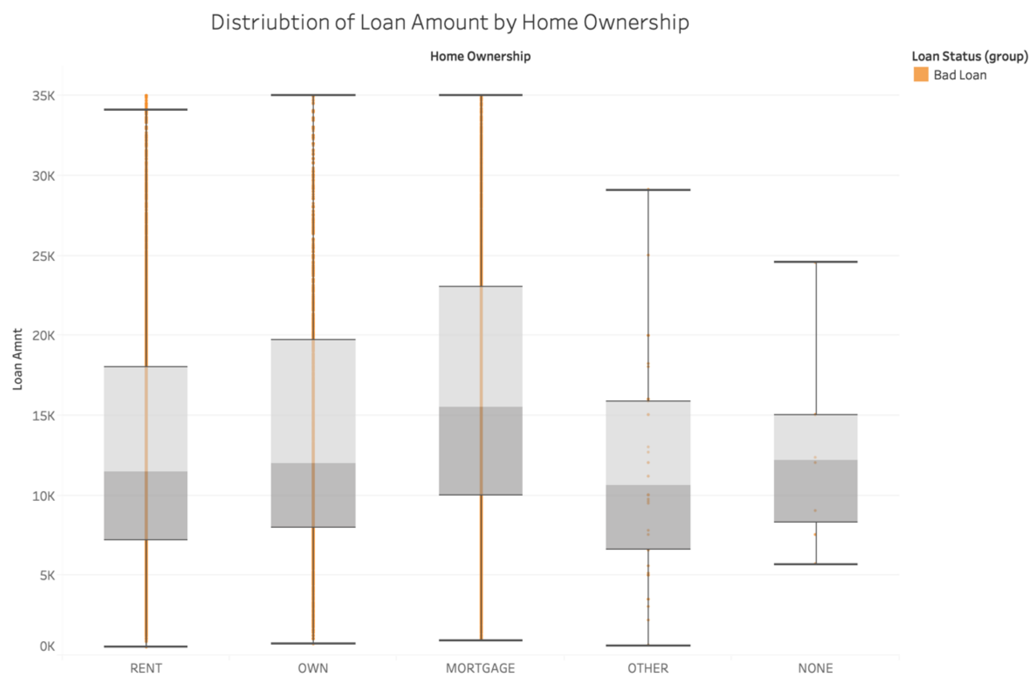
- As the correlation plot showed, loan status (default) has strong positive correlation with out_prncp, out_prncp_inv, strong negative correlation with total_pymnt, total_pymnt_inv, total_rec_prncp and last_pymnt_amnt. This result is a little messy here, try to focus on the variables more worth it that we may can find a trend with the loan status.



- As the loan amount borrowed by different bad loan types showed, in Grace Period and Late (16-30 days) had an increasing trend in the last year, all the other types had a tendency to decrease.

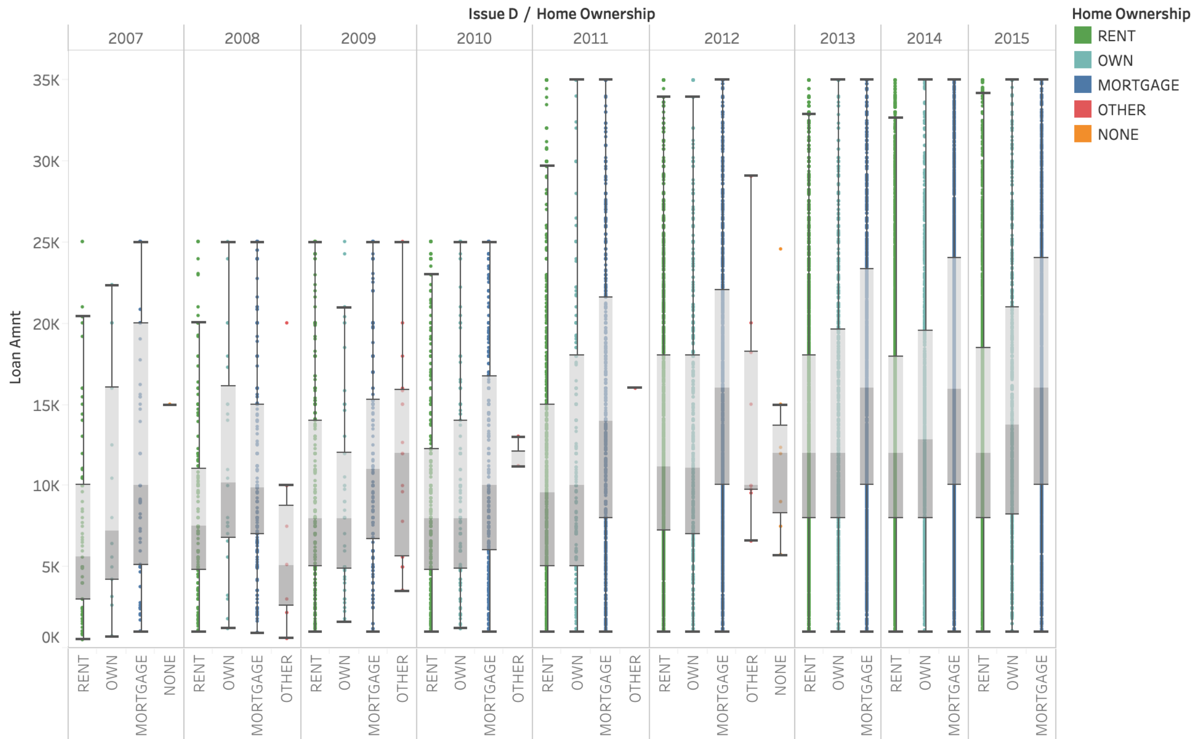


- In the home ownership category, mortgage has the highest bad loan amount issued than the other groups.



- With the trend of the bad loan amount issued by home ownership, we can see that Mortgage has a slight increase through the years.
- Thus, people who have a mortgage may have higher possibility to make the loan fall in default. (depend on all the other factors)

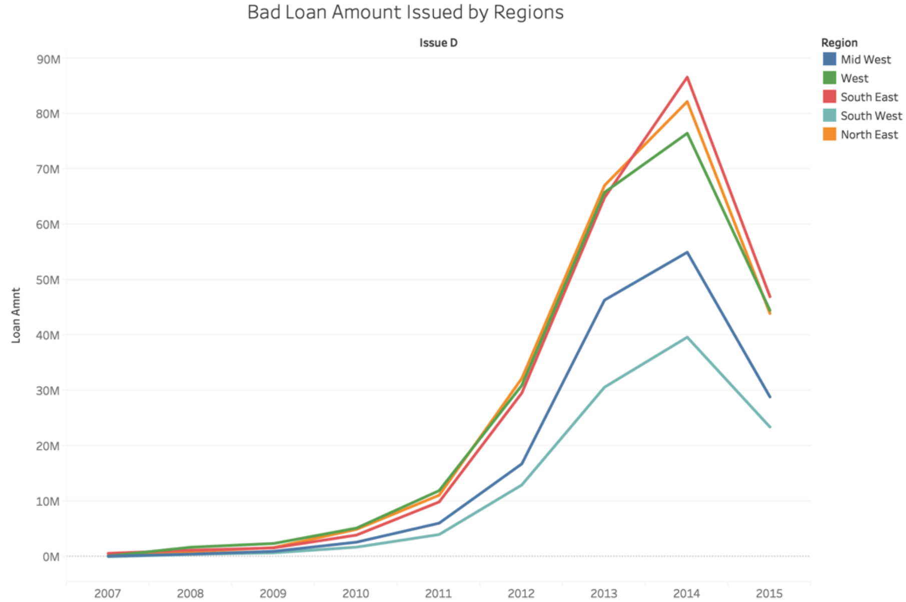
Distrituion of Amount by Home through the Years



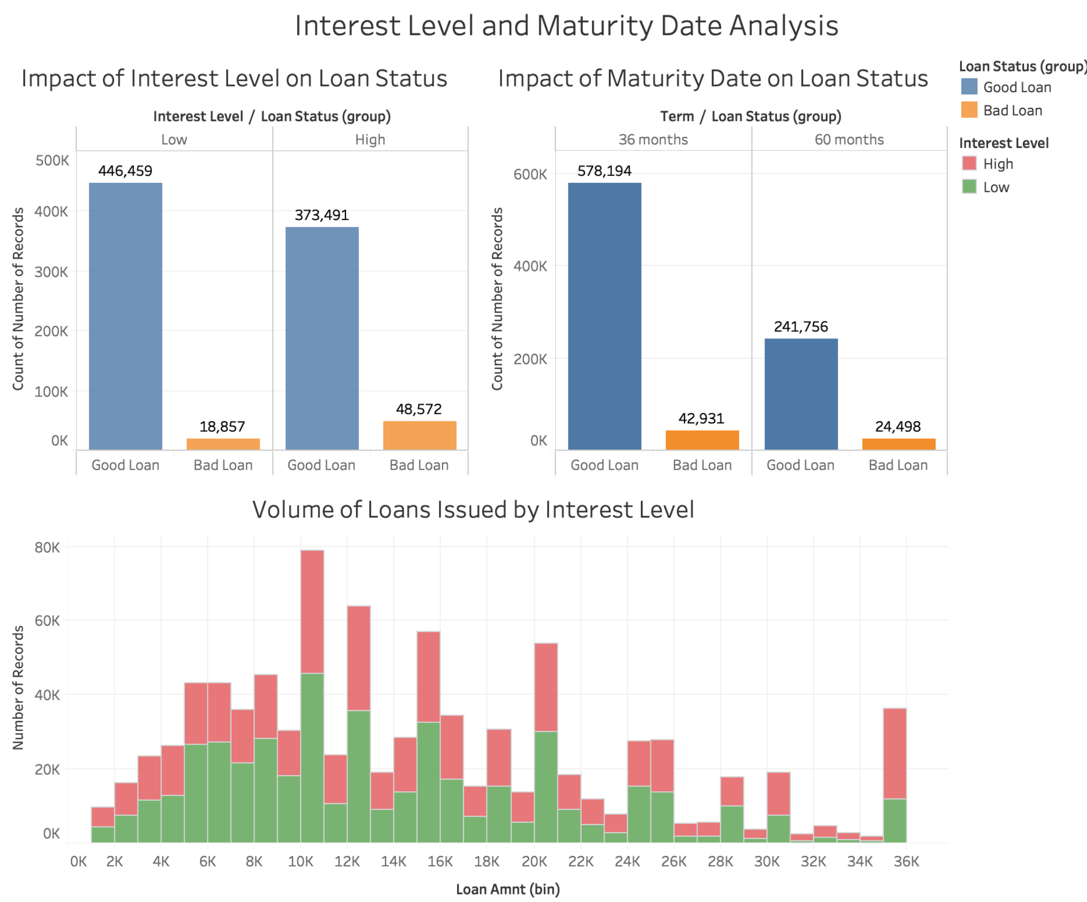
d) Defaulted Loans and Level of Risk

- The bad loans issued by all regions decreased rapidly since 2014, and in the last year recorded Southeast and West had the highest bad loan amounts.

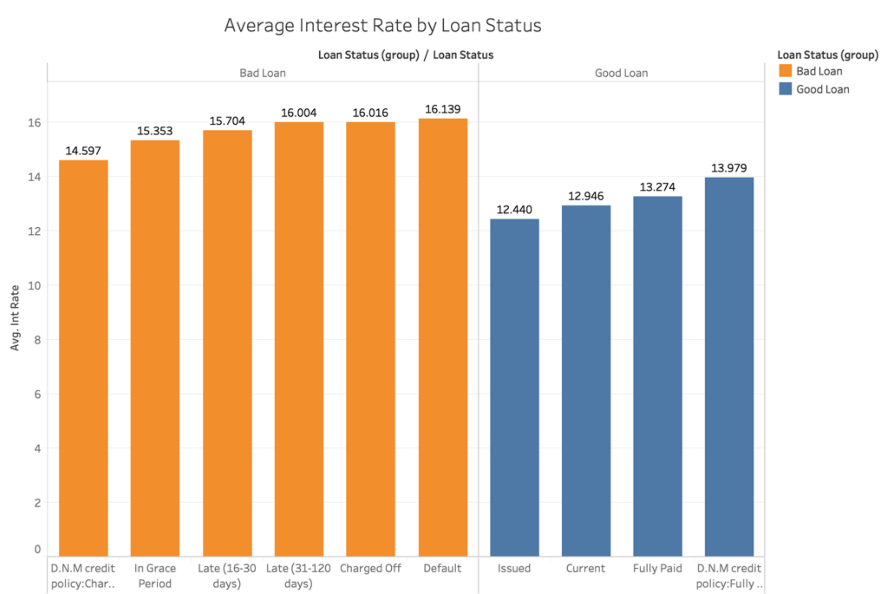
Bad Loan Amount Issued by Regions



- The loans with high interest rate(r>13.23%) have higher possibility to become a bad loan.
- The loans with shorter maturity date (36 months) are more likely to become a bad loan.

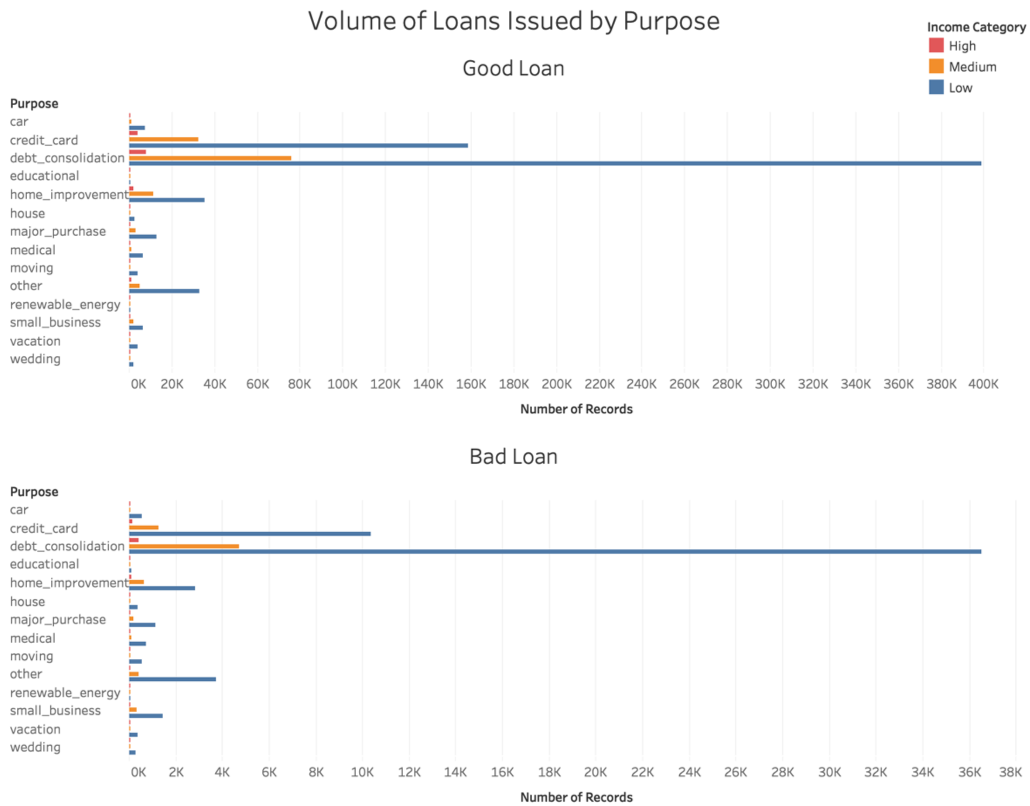


- The average interest rates of bad loans are in the range of 15%-16%, and the average interest rate of good loans are around 12%-13%.
- Not too much difference of the interest rate between bad loans and good loans, there has to be a better assessment of risk.

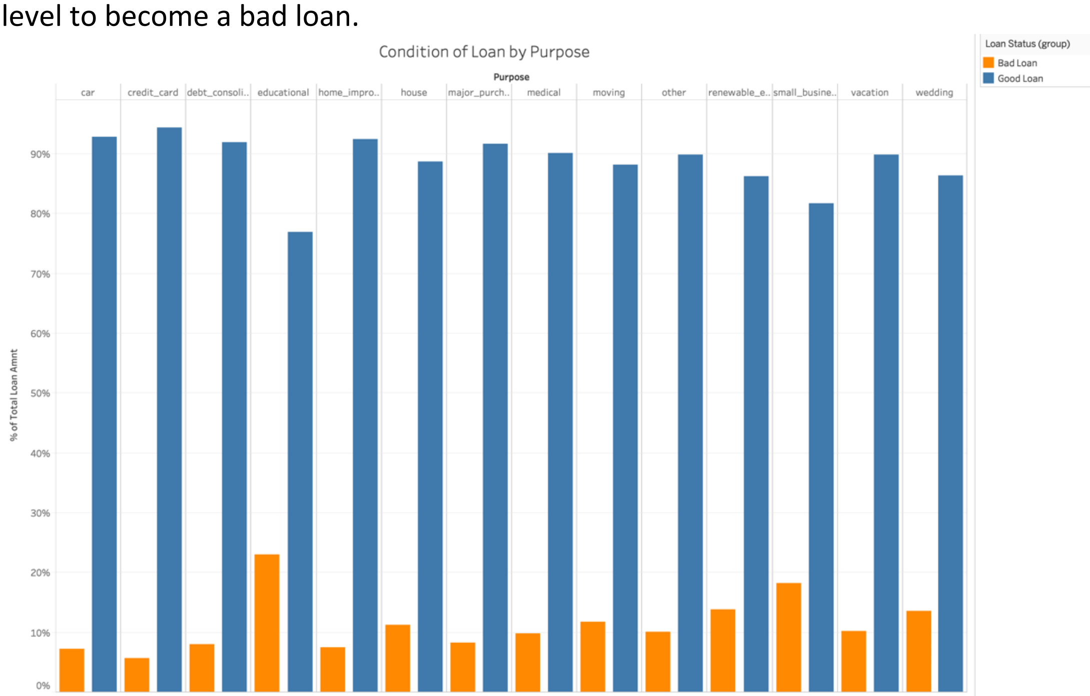


e) Risks by Purpose

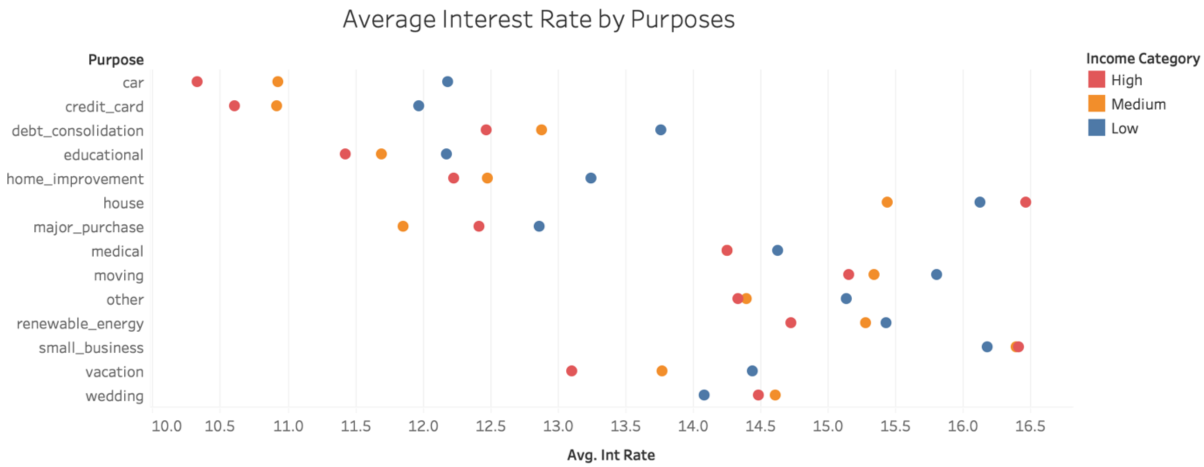
- The most frequent applied purpose is consolidate debt, and the least frequent purpose is educational.



- Borrowers who apply for educational and small business purposes loan will have a higher risk level to become a bad loan.



- For most purposes except house, small business and wedding, the low income group has a higher interest rate.



- For most all the purposes, the high income group had a lower bad/good ratio, which means high income group has lower risk level of falling default. There is one spike in the graph which is for educational purpose, since it only had two records and one is bad loan which caused the ratio up to 50%.