Data Visualization Final Project

Shivee Singh

Introduction

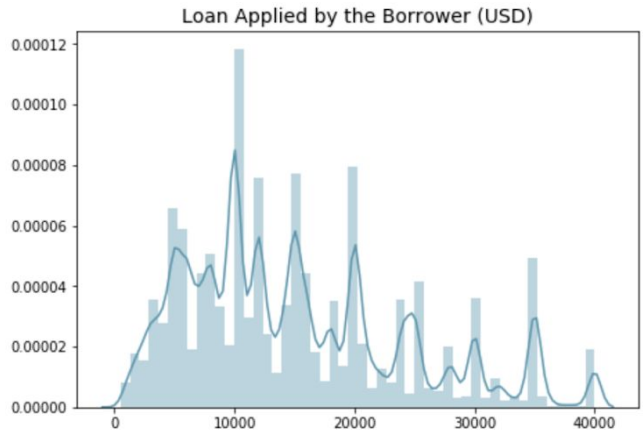
1. What is this dataset ?
   The dataset consists of loan details of all loans issued by Lending Club through 2007-2015. Lending club is a peer to peer lending company based in the United States. It is an innovative business model where investors can lend to borrowers and earn a profit depending on the risk they take (the borrowers credit score).

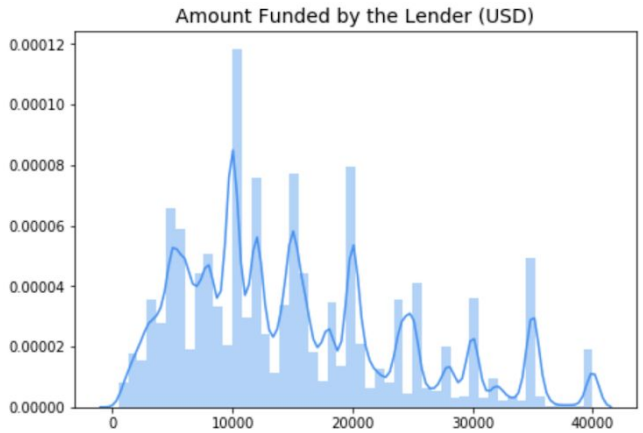
2. Where did you get it from ?
   I got this dataset from Kaggle datasets.

3. Why did you choose this particular data and what type of questions were you hoping to explore ?
   I have keen interest in understanding about personal credit and how peer to peer lending industry is supporting personal credit transactions. By analyzing this dataset I can find out various things about lending club as a company and also various aspects of peer to peer (p2p) lending such as how fast peer to peer lending is growing, who are the main customers, how the loans are performing, how risky are such loans, what are the main categories for which people borrow on such platforms etc.

Summary of Data

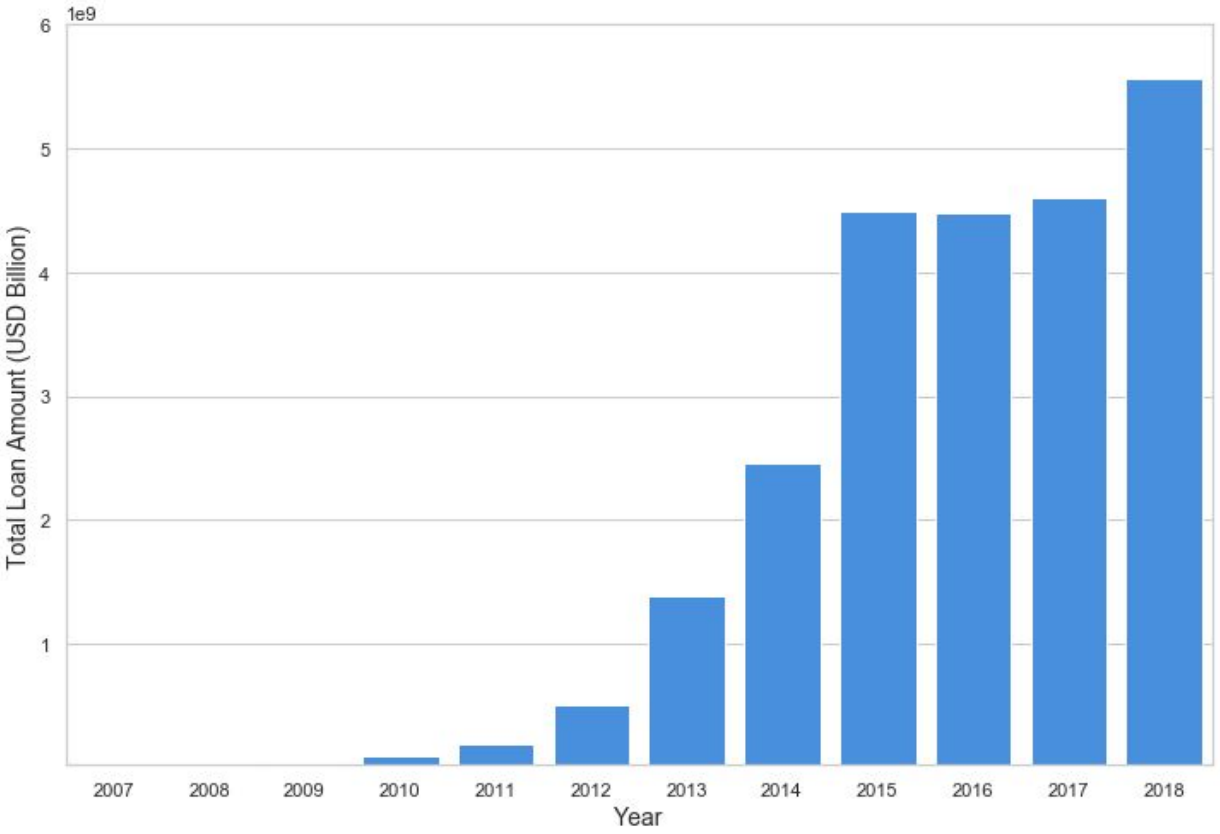
Histogram



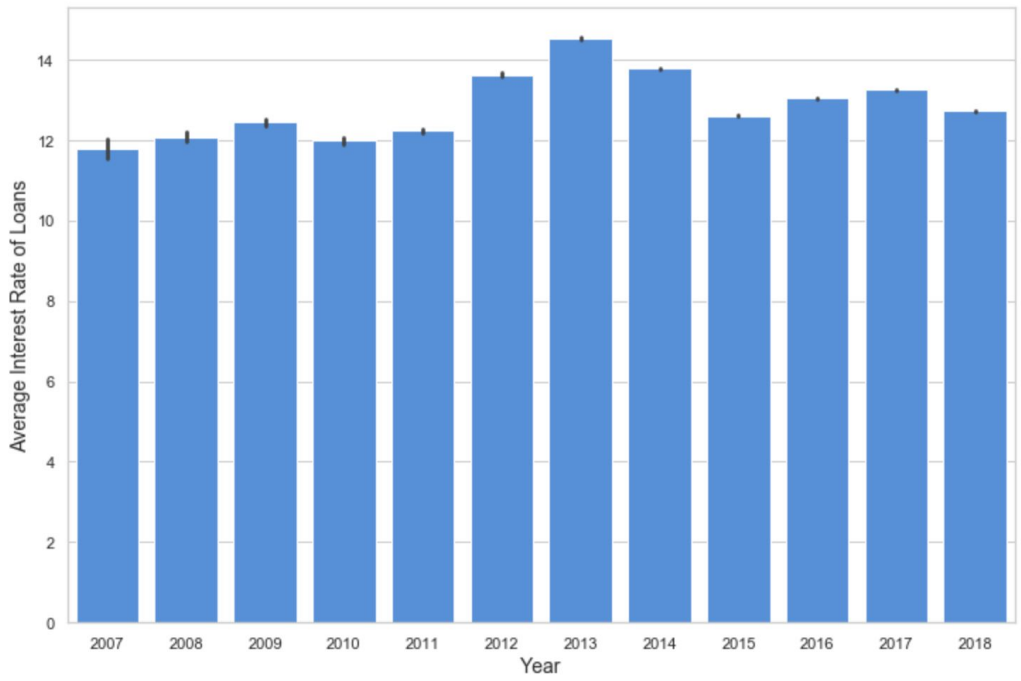


The chart above shows distribution of the loan amount requested by borrower and amount funded by
the investors. The distributions are almost identical indicating that the amount requested was fulfilled
the investor funds. We can see that the most common loan amount is 10k USD. The most common range of loans is 10k-20k.

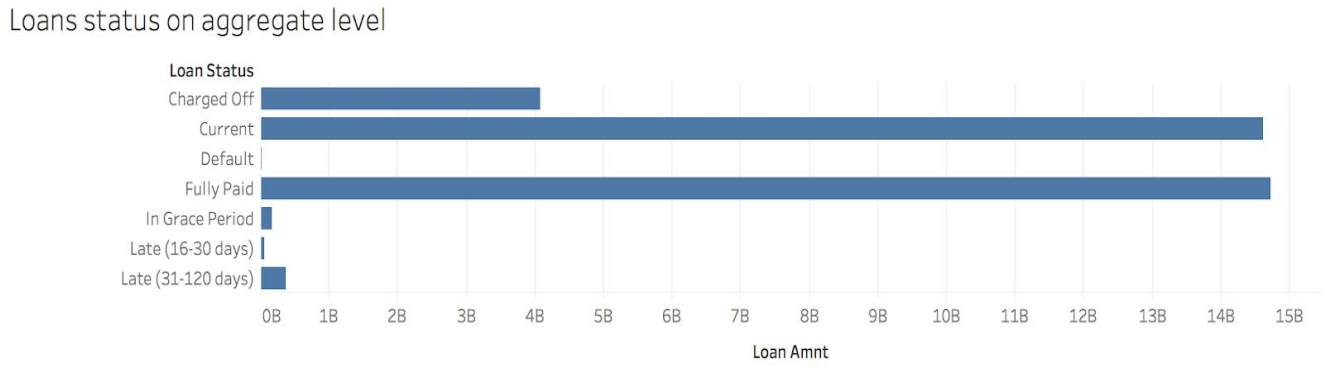
Barplot



The chart above shows total amount of loans extended through Lending Club over the years. There has been phenomenal growth from 2010 to 2015 (almost exponential growth). Post 2015, the growth slowed down from the historic fast pace. It is possible that too many people defaulted on their loans in 2015 which made people averse to investing in a p2p lending or it is possible that other players came into the market.

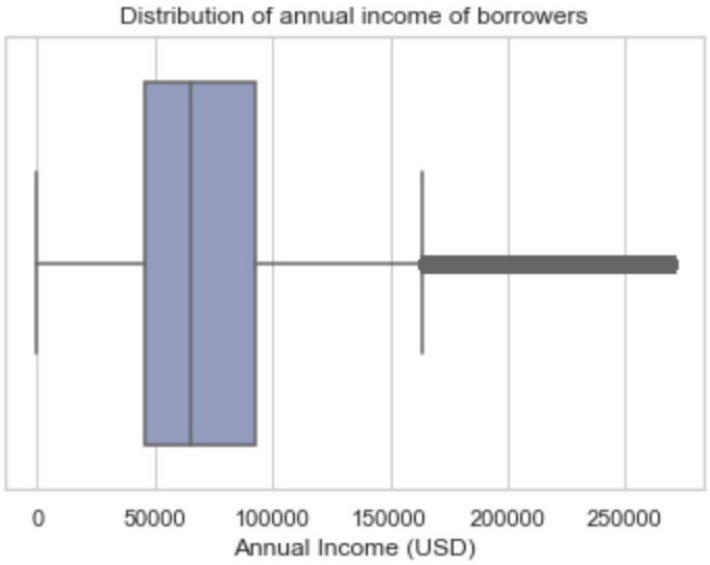


The chart above shows how average interest rates have varied over the years. On an average the rates increased during 2012-2013 but have softened a bit since then. No major trend can be seen here.

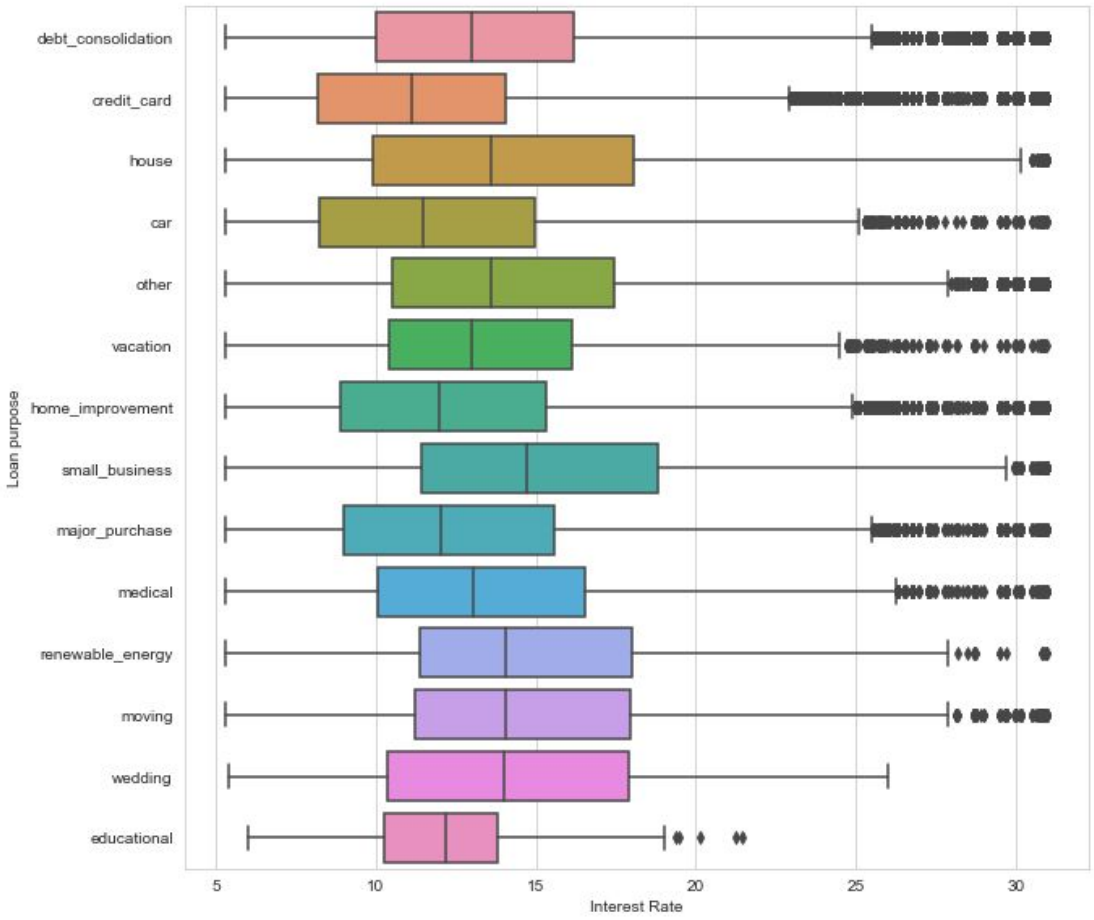


The chart above shows how lending club loans have performed till date. The loan amount is on x axis and loan status on y axis. We see that majority of the loans are good because they are either fully paid or current. But the charged off loans is ~4B USD which is roughly 11% of the total loans funded. It is not a small number that can be ignored.

Boxplot

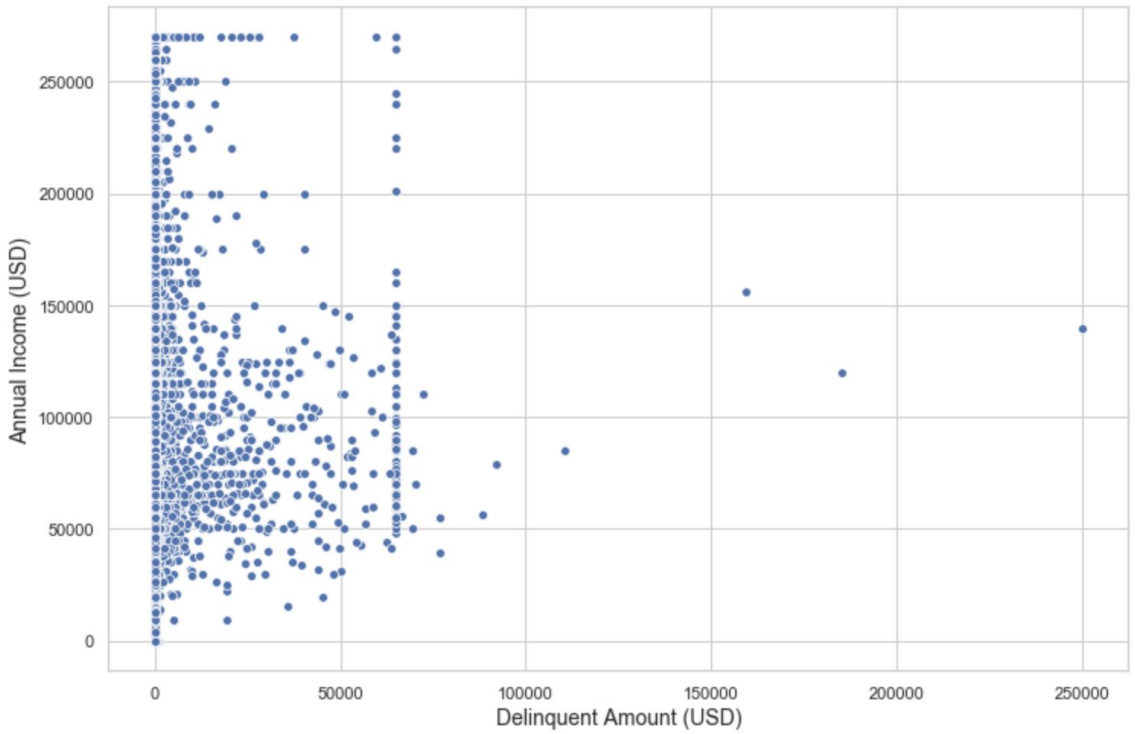


The chart above shows the distribution of annual income of the borrowers. The median annual income of borrowers on lending club platform is $ 60K USD. Majority of the borrowers have annual incomes between $ 40K USD - 90K USD. It is also seen that some data is missing or incorrect since there are records with zero income. I believe it is not possible for anyone to get loan with zero income.



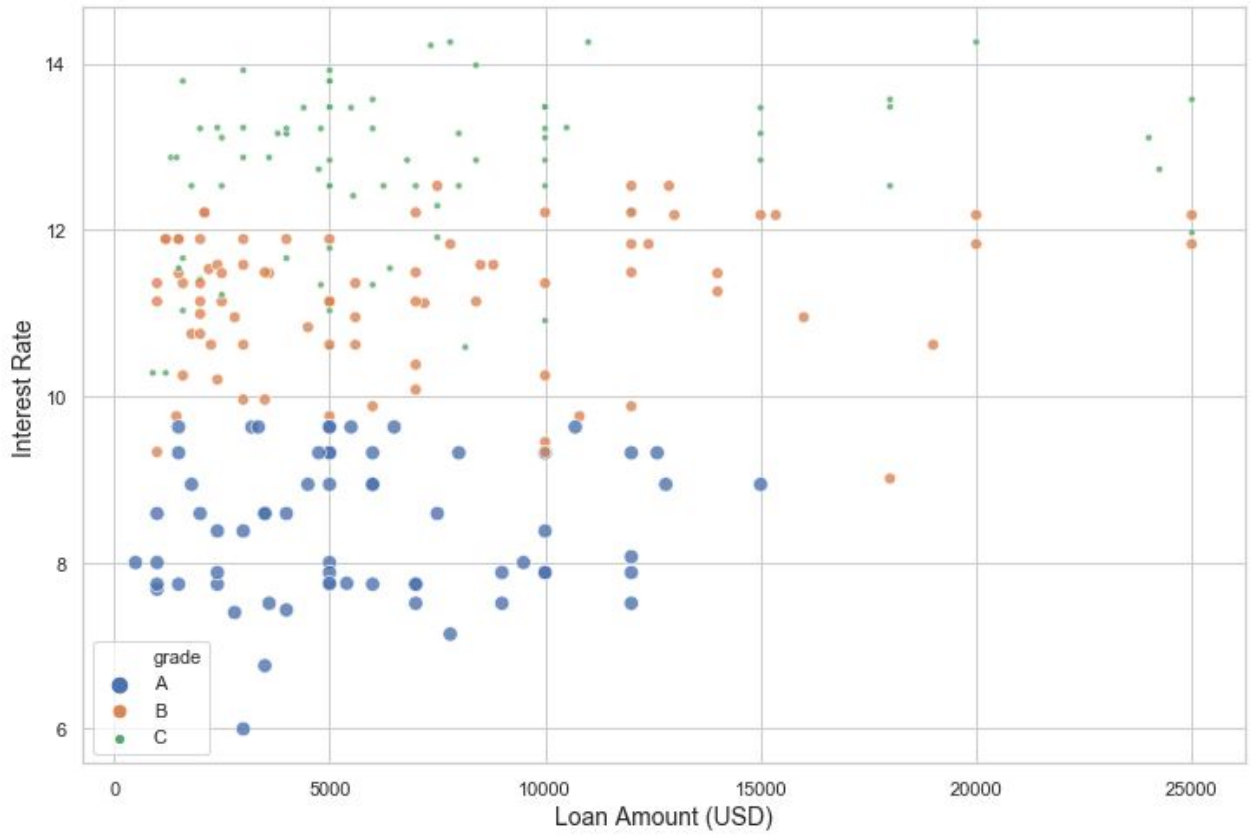
The chart above shows the distribution of interest rates across different loan purposes. Educational loans have a low variability as compared to house or small business loans.

Scatterplot



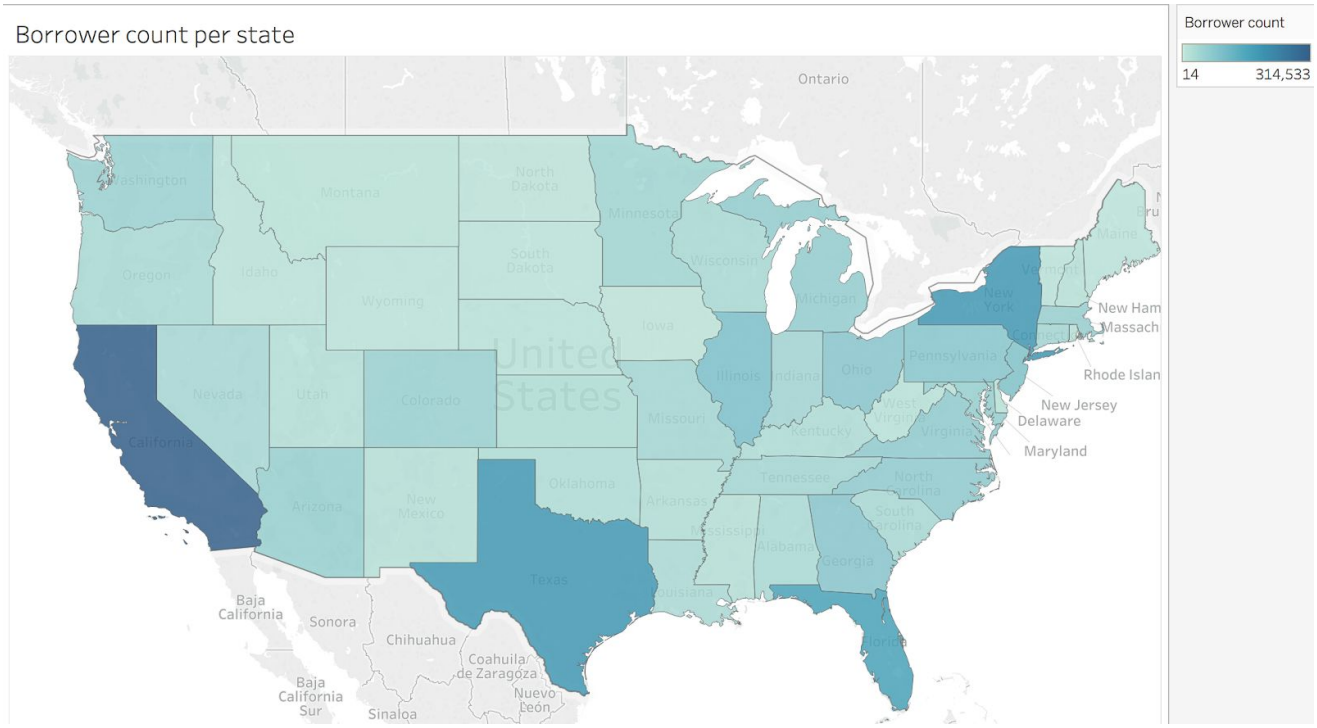
The chart above shows the relationship between delinquent amount (over*due*) and the annual income of borrower. Although there is no clear pattern but we can still see that the higher overdue amounts are from people whose incomes are lower. Which is expected.

Bubble Map



The chart above shows the distribution of loans in the ‘educational’ category. The size and color of the bubbles indicate the ‘risk’ category of the loan. The grade ‘A’ is considered to be safest , ‘B’ is riskier and so on. As expected the interest rate for risky loans is higher. We also observe that with an increase in loan amount, the riskiness increases. ‘A’ grade educational loans range maximum upto 15K USD. The range for ‘B’ and ‘C’ grade is higher ~25K USD.

Chloropleth Map

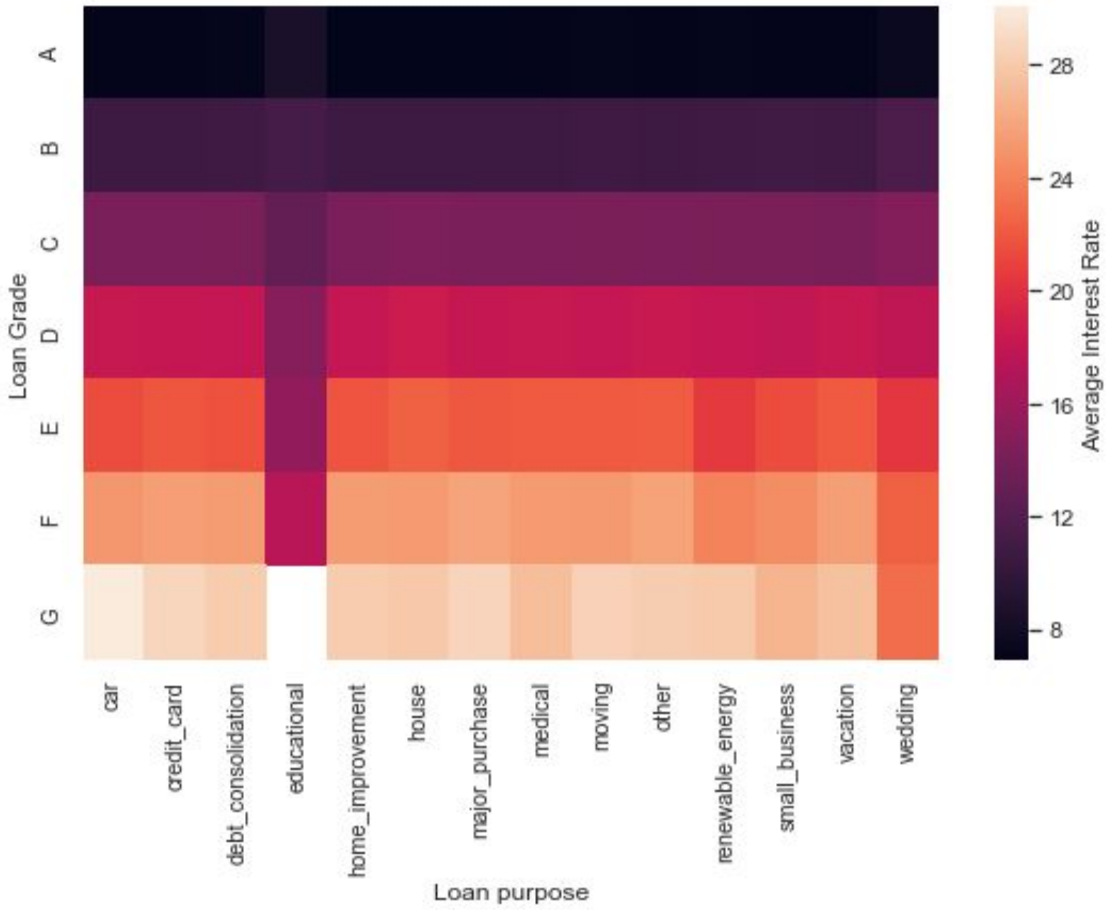


The chart above shows geographical distribution of borrowers across different states in USA. These borrowers have been served by lending club. Largest number of borrowers are from California, Texas , Florida and New York also have a significant number of customers of this platform.

Connection Map

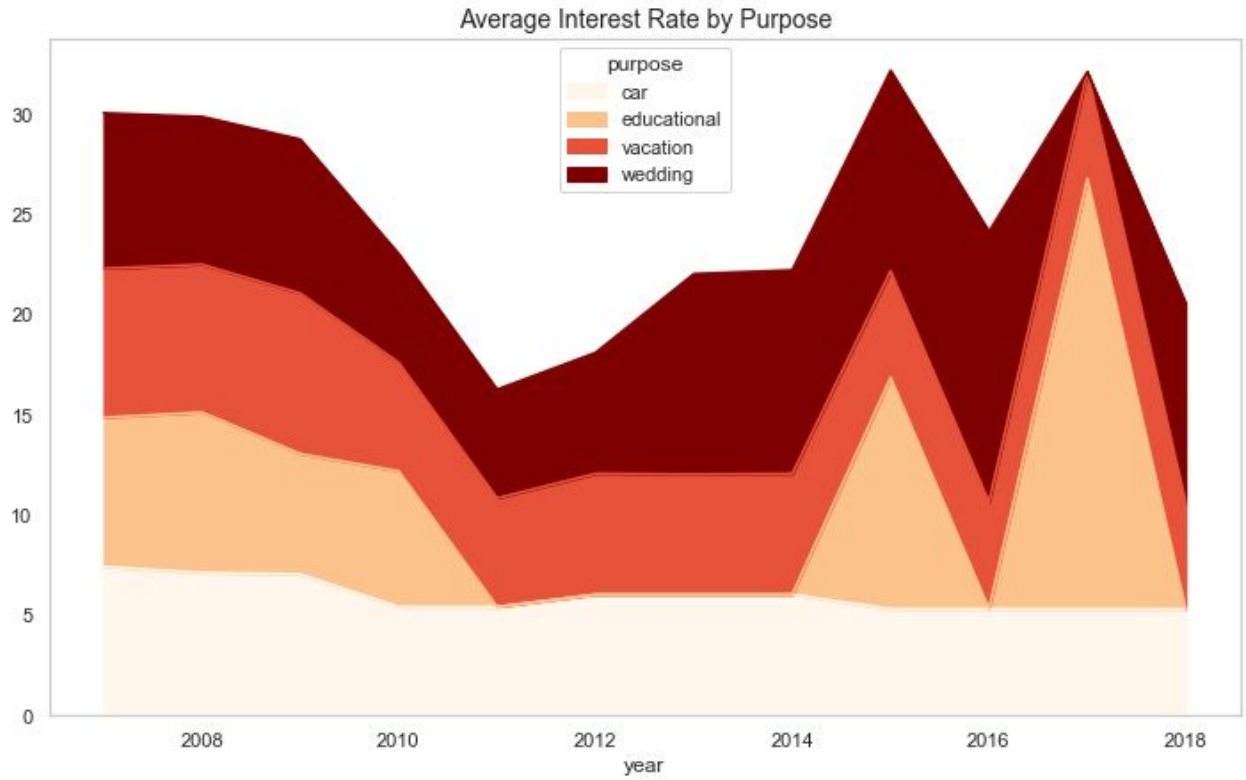
I could not find a use case to have a connection map. It is difficult to have a connection map with this dataset.

Heat Map



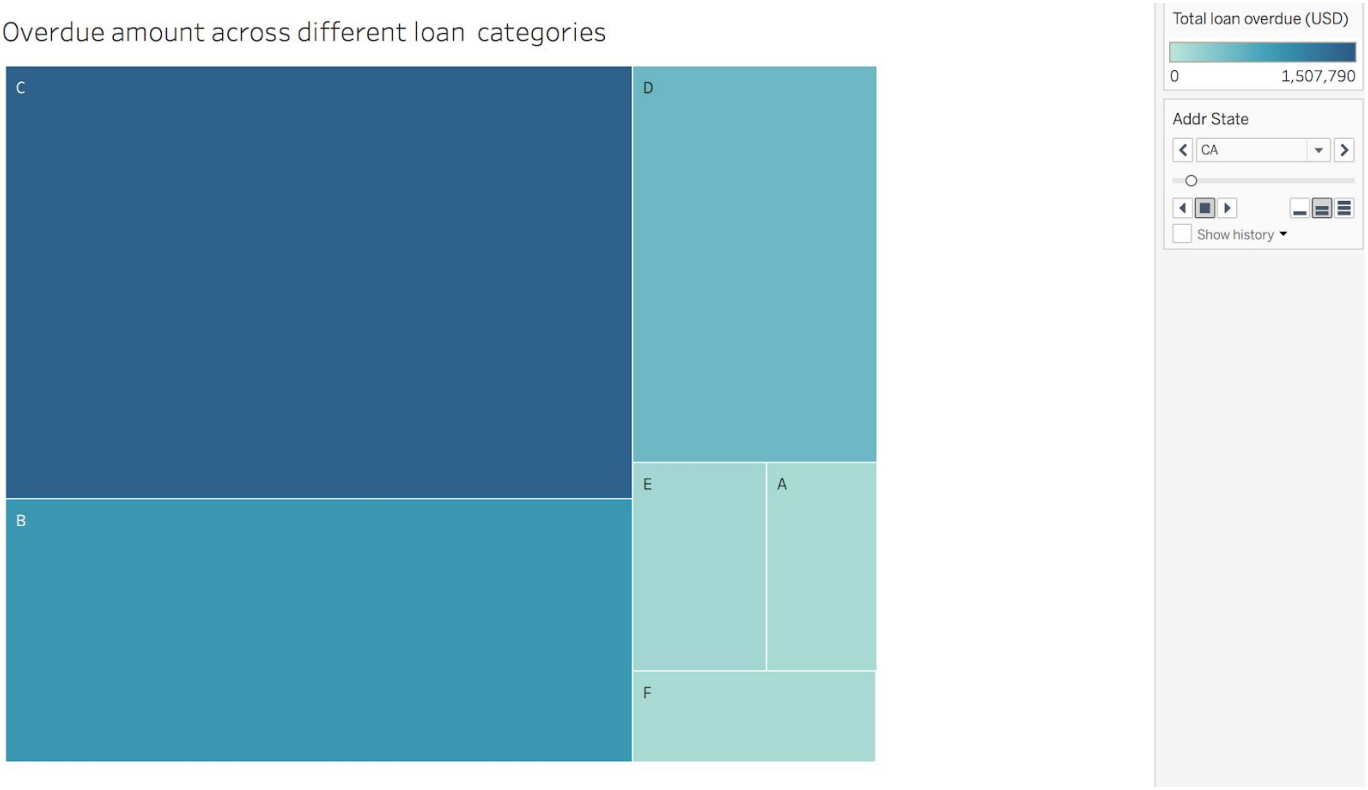
The chart above explores the effect of loan purpose and loan grade on interest rate.
It is interesting to note that safe loans (‘A’ , ‘B’, ‘C’) grade loans have low interest rates across all purposes. But the same is not true for riskier category loans such as ‘E’, F’ and ‘G’. The interest rate increase as we move towards more riskier loans. Moreover the interest rates also become more sensitive to loan purposes as we proceed along the risk grade.

Stacked Area



The chart above shows interest rates across different loan purposes. It has been plotted over the years to see trends if any. I have considered a subset of the loan purposes instead of considering all. We observe that automobile loans are the cheapest followed by educational loans, followed by vacation and wedding. Over the years car loans continued to remain cheap but after 2014, all the other categories became more expensive.

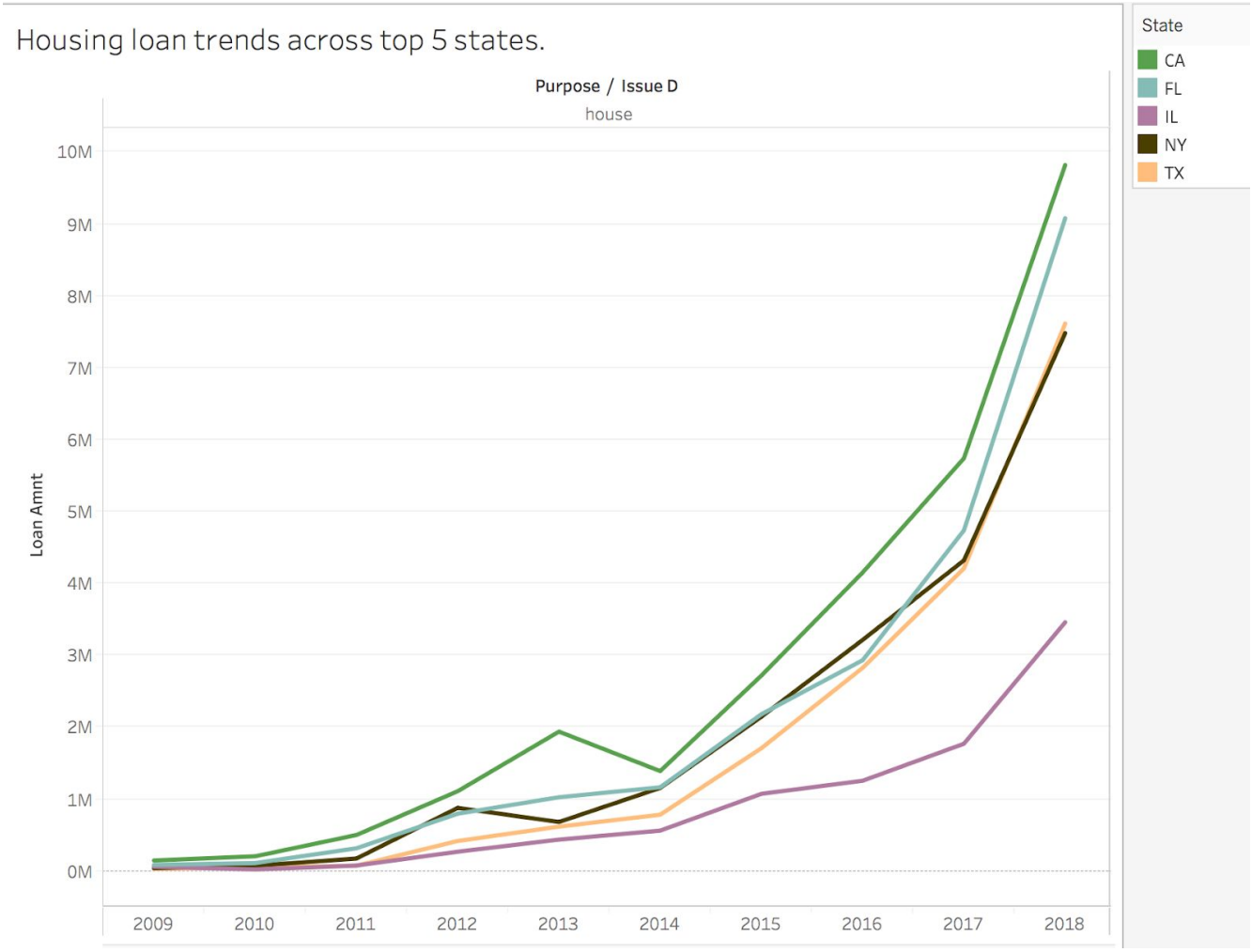
Treemapping (Interactive as well)



The chart above shows overdue amounts across different loan risk categories. The loans belonging to categories ‘C’ followed by ‘B’ and ‘D’ have the highest overdue accounts.

Story

It is a known fact that successful technology companies in USA drive play a significant role in driving economic growth. For example when a tech giant decides to expand into a new geographical area then it creates a huge demand for housing in that area.
I analyzed the housing loans across different geographical regions over the years and found very interesting insights and patterns. This can be very helpful for real estate investors.



The above chart shows year on x axis, the total amount borrowed under housing category on y axis and the different colored lines indicate different states. I chose top 5 states by total loan amount.

Insights

We observe that for a couple of years after the recession there was not enough demand/supply for housing loans, but slowly it started picking up around 2012. It is very interesting to see California was leading the housing demand consistently from 2011 upto 2017. But Florida seems to be catching up very fast. Housing demand in Florida surpassed NewYork in mid 2016 as can be seen in the graph above. This could be because a lot of tech companies have been expanding in Florida. The same is true for Texas. Housing demand in Texas has caught up with New York as of 2018. Looking at these trends I expect Florida and Texas housing market to heat up fiercely in the coming few years. It can be a great opportunity to invest.

Conclusion

- The peer to peer lending is a fast growing industry segment. Lending Club entered the industry in 2007 and was the first to enter hence it captured a large market share. The platform has seen exponential loan growth from 2010-2015 and after stagnating for a while it is again picking up since 2018.
- The most common loan amount is 10k USD. The most common range of loans is 10k-20k. The median income of borrowers is ~60K USD. Which means average debt to income ratio is 1:4 which is decent.
- The number of loans that defaulted and were charged off is quite high around 11% of the total loans. Hence investors should be vary of this fact and conduct their own due diligence is possible. Investors should note that a large number of loans in the 'B','C', 'D' category have become overdue which means even though they are ranked higher they might not be as safe as you expect them to be.
- Interest rates vary across different loan purposes. Car loans and credit card payment loans are priced lower as compared to business or home improvement loans. Hence it might be a good idea to get loan from Lending Club if you are a student or you want to pay your credit card bills. A good thing for borrowers is that interest rates have remained fairly stable over the years.
- California, Texas, New York and Florida are the states in which the highest amount of loans were issued. In the recent years housing loan demand in Florida and Texas has matched up to the level of New York. These places might be good to invest for real estate purposes.

Link to github: https://github.com/ShiveeUSF/Data-Visualization.git

Code: Uploaded jupyter notebook and tableau workbooks.

Citations:

https://www.kaggle.com/wendykan/lending-club-loan-data/
https://en.wikipedia.org/wiki/LendingClub
https://seaborn.pydata.org/