**Introduction**

In 2013, the global P2P lending industry reached at $3.5 billion USD. Analysts predict that P2P’s global value will be $1 trillion USD in 2050. P2P lending market in the US valued higher than $32.8 billion in 2017.[[1]](#footnote-1) As you can see, the global peer-to-peer lending market shows the significant growth especially in personal purposes due to advancement in technology. The technology makes users can access easily to offering. For this reason, the demand for peer-to-peer lending market is boosting.[[2]](#footnote-2)

However, the default rate is getting be a problem because there is a high possibility to rise the default rate.[[3]](#footnote-3) Even though there is a high risk of default, the demand of P2P lending is increasing. Therefore, the most important issue for lending company will decide who can be clients as borrowers with lower risk of default. Finding a way to decrease default rate is important. For this, we need to identify which factors or traits will be the potential risk. Therefore, analysis of income, credit rating, the purpose of lending money, etc. of current customers will be useful for selecting good potential customers. Because we can adapt that current customers’ traits to people who want to be borrowers. It means we can predict potential customers who would be fully paid or defaulted. Based on this, lending company can get insight to decrease risk and hold more profit.

To predict who will be the good customers, we use the data from Lending Club. Lending Club is the world’s largest peer to peer lending platform in the US. They offer lending service by connecting the borrowers with investors through an online marketplace. Lending Club data contains details information about borrowers and type of loans they request; based on borrower’s credit score, credit history, desired loan amount and the borrower’s debt-to-income ratio, Lending Club determines whether the borrower is credit worthy and assigns to its approved loans a credit grade that determines payable interest rate and fees. All loans are unsecured personal loans and can be between $1,000 - $40,000. The standard loan period is three years; a five-year period is available at a higher interest rate and additional fees.[[4]](#footnote-4)

**Business Questions**

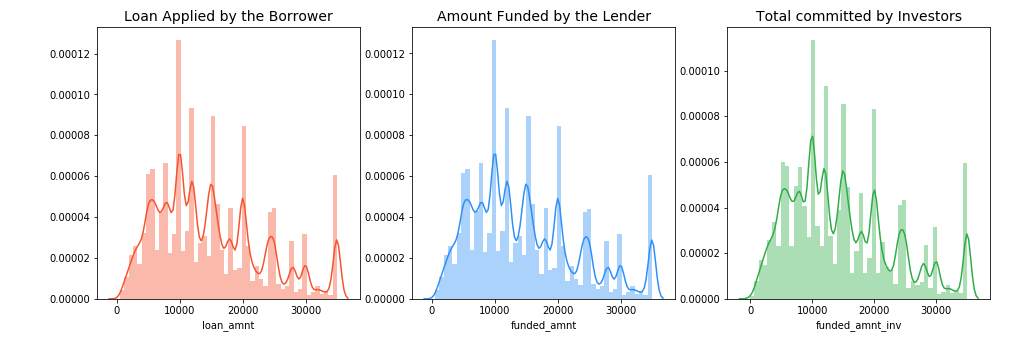
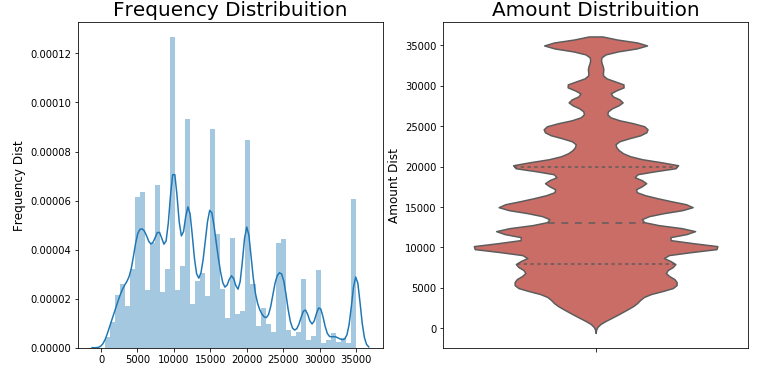
To recognize what kinds of people and group will pay back well, we need to distinct the traits of them. For this, we will make some business question based on 3 keywords; 1) credit rate 2) default 3) purpose. Firstly, for getting good customers in the future we need to figure out what factors that affect the credit rating greatly. Furthermore, loan status will be highly linked with income so check that the amount of income and the income frequency will be meaningful.

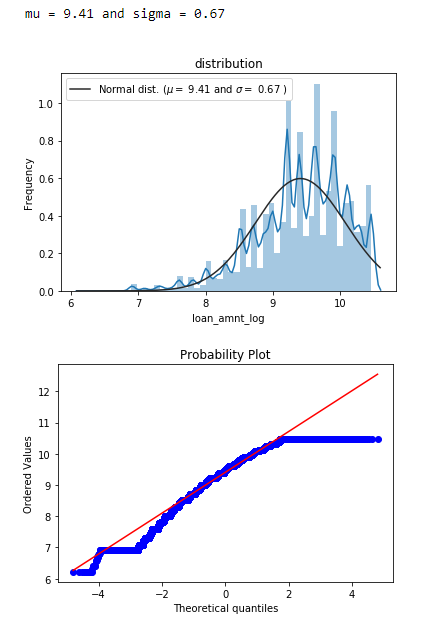
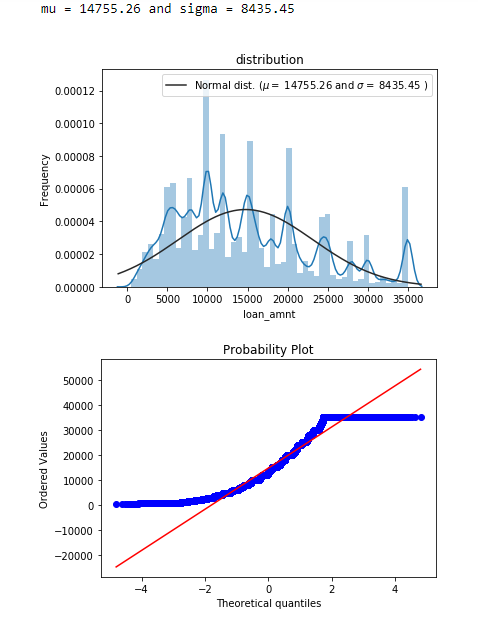
Next, analysis of gap between a group who are fully paid and the other group who are defaulted will be helpful question to avoid risk of delinquency. Besides, it will be helpful to make a standard which can judge whether new customers’ ability and conditions are qualified or not.

Lastly, the reason people want to borrow loan is related to their daily life or personal reason. By analyzing the purpose we can get information related to people’s interest and the reason why they need money. This would provide another chance of business strategies.

**Analysis**

1.대출 정보:우선 대출금액의 분포를 살펴본 뒤 언제 대출금액이 크게 증가했는지 살펴보자

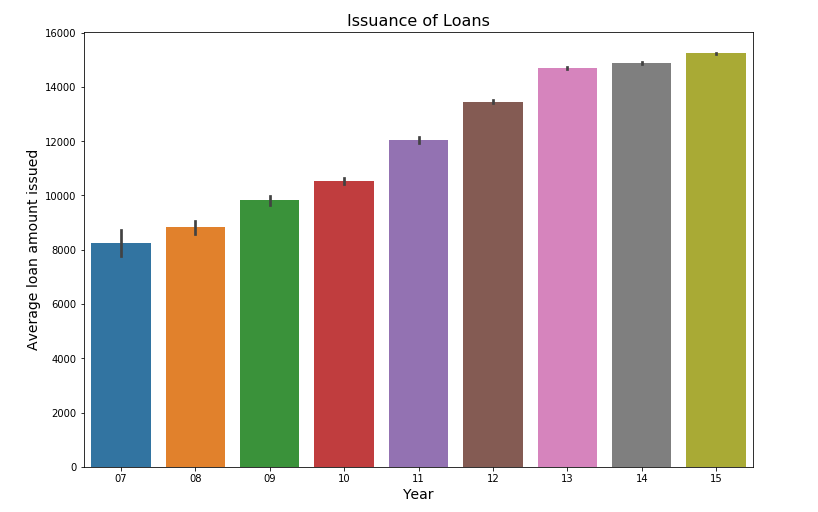
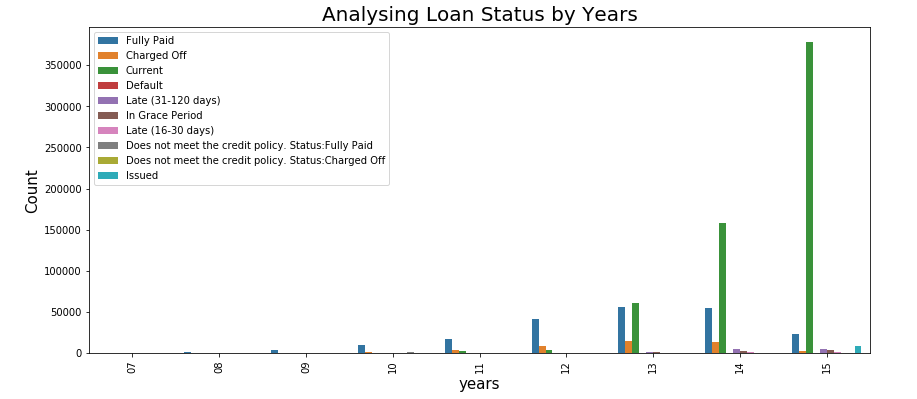




The above figure used the probability distribution, quantile-quantile plot to see if the normalization distribution was followed.

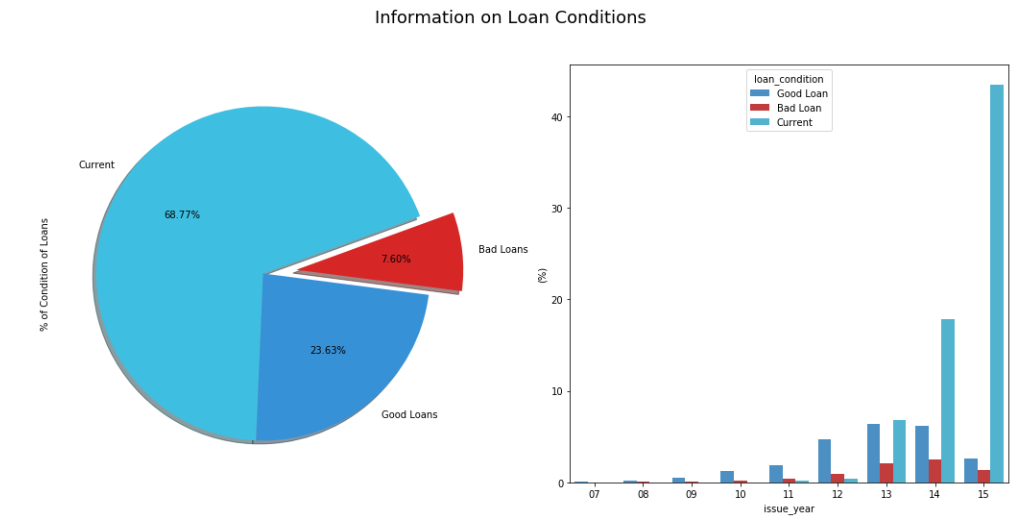
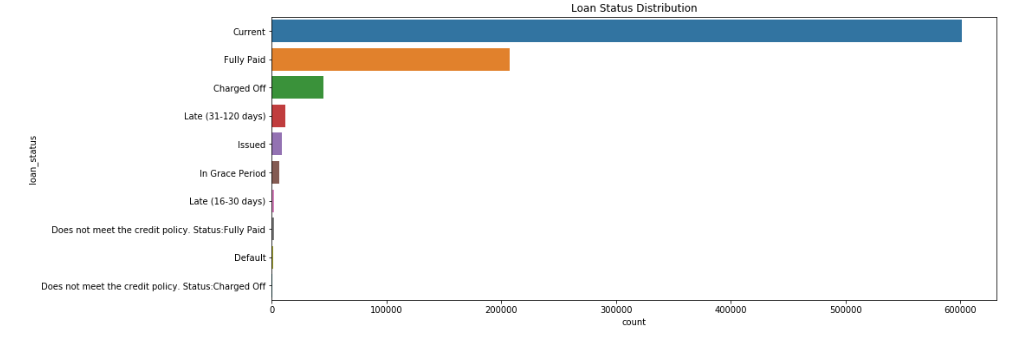
You can see that the Loan variable does not follow the normal distribution.

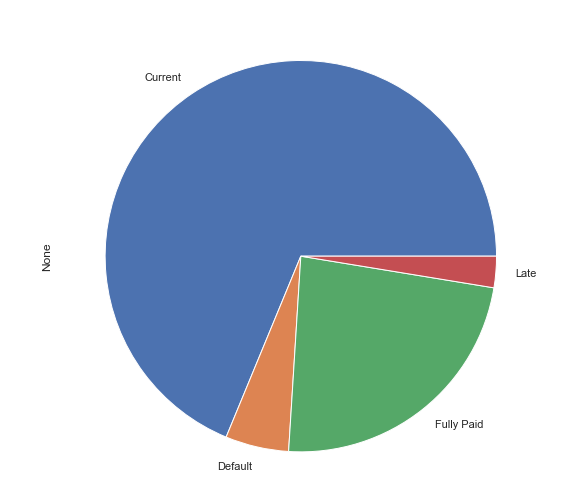
variable is right skewed. As linear models love normally distributed data , tried to transform this variable and make it more normally distributed.



발행된 대출금의 대부분은 미화 1만 달러에서 2만 달러 범위였다. 2015년은 대부분의 대출이 발행된 해였다. 대출이 계속 늘어나는 것을 볼수있다 (미국 경제의 회복으로 인한 요인을 가정할수 있고 렌딩클럽이 성장하는 추세라는 것을 예상할수 있다.) 잠재 대출자가 요구하는 대출, 렌더가 펀딩한 금액, 투자자가 출자한 금액도 비슷하게 분산되어 있어 통제된 데이터 세트만 공개되었다는 것을 의미한다. 대출자의 요구와 다른 대출가치가 있다면 회사의 대출승인요건을 예측할 수 있을 것이다.

2. Type of loan

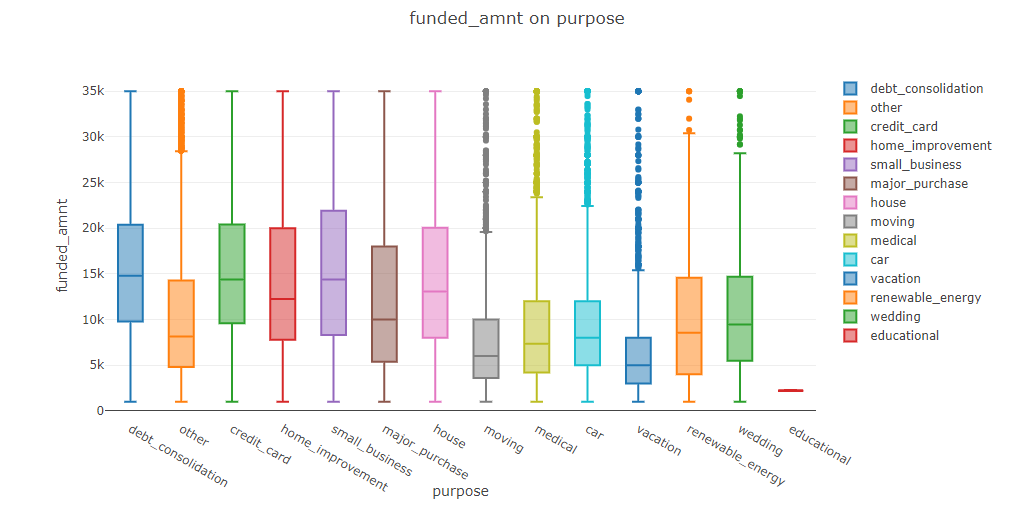
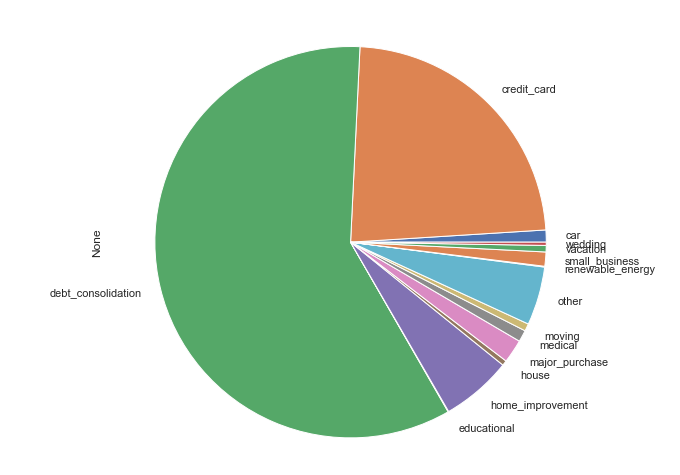
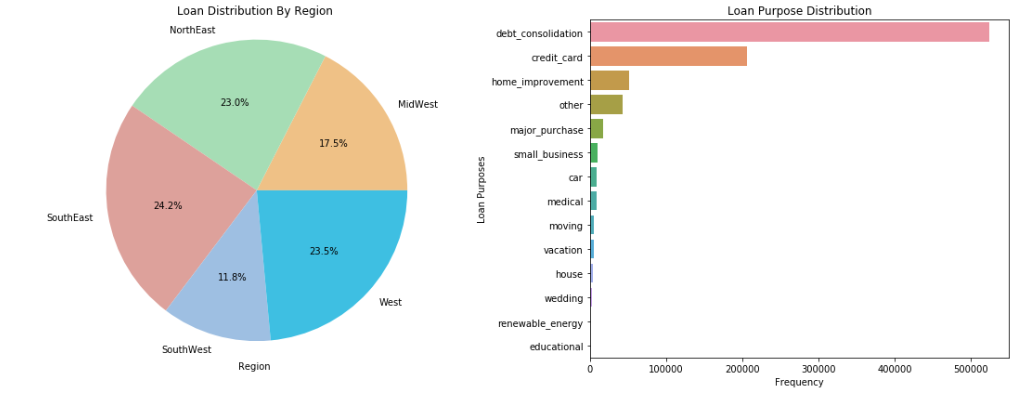




렌딩클럽이 지금까지 선언한 부실채권 액수가 얼마인지 볼수있다. 물론 앞으로도 채무불이행 위험이 있는 대출이 여전히 존재한다는 것을 이해해야 한다. 우리가 알아야 할 것은 부실대출의 액수가 날이 갈수록 증가할 수 있다는 것이다. 왜냐하면 최신으로 갈수록 더 많은 대출이 발생하고 있기 때문이다.

이것은 물론 대부분이 수익으로 연결될테지만, 현재는 대출상태만 고려하는 중이다. 대출이 늘어나는 비율과 채무불이행 되는 비율을 고려해보면 어느정도 예측이 가능할것이다.  
  
요약:.  
현재 부실대출은 전체 대출의 7.60%를 차지하고 있지만 부실대출의 위험이 있는 현재 대출은 여전히 남아 있다는 점을 보자 (그러므로 이 비율은 가능한 변경의 대상이 된다.) 2012년 이후 평균금리는 하락했지만 이는 대출금액의 증가를 설명할 수 있을 것이다.

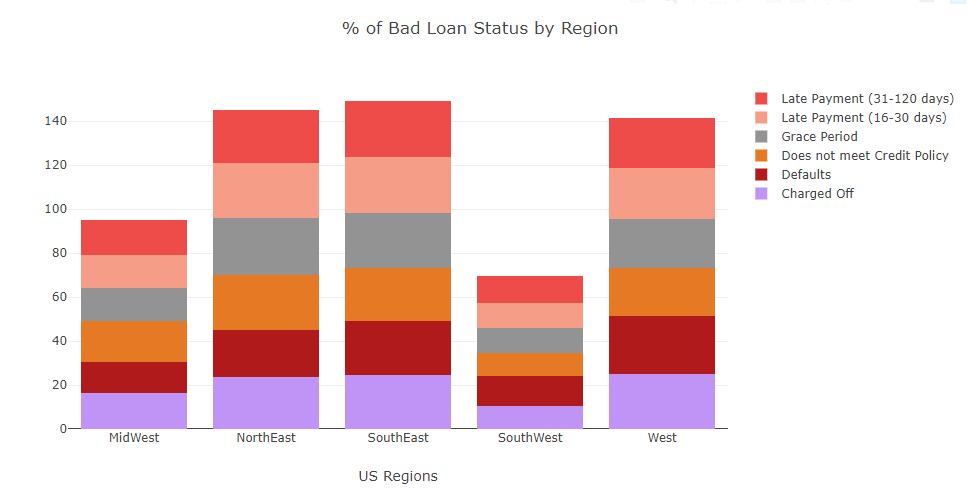
3. Loans Issued by Region and purpose



Summary:

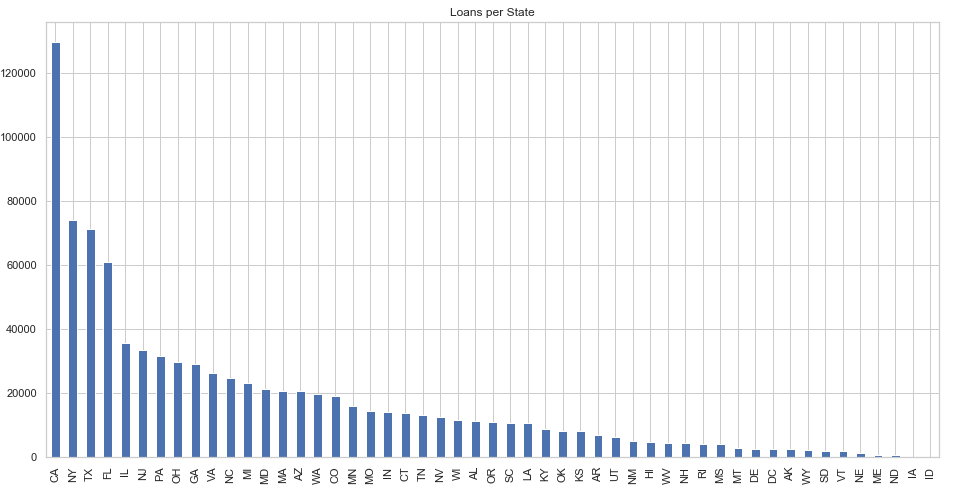
SouthEast , West and NorthEast지역은 대출금이 많은 반면 SouthWest과 MidEast 지역은 대출금이 낮은 편이다.  
목적에 따른 대출금액 정도 : 빛 상환, 신용카드 지불, 비즈니스를 위한 대출이 높고 교육,휴가 를 위한 대출은 적다.

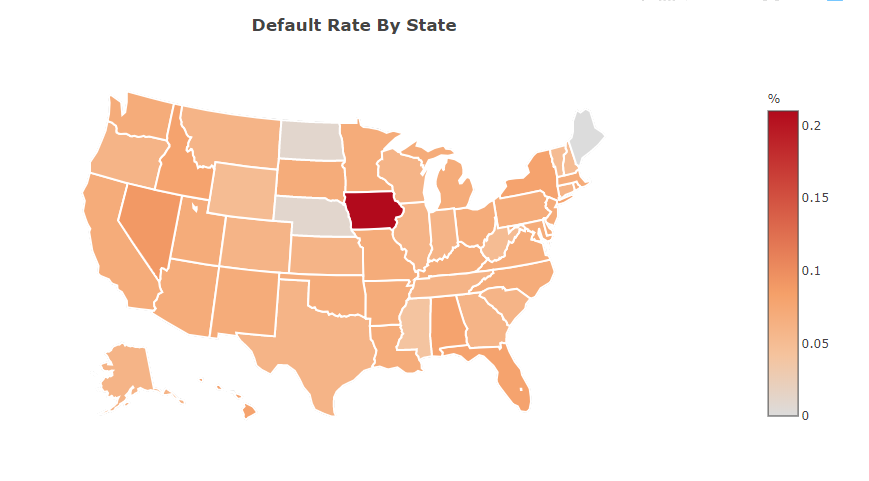
지역에 따른 불량대출 상태비율



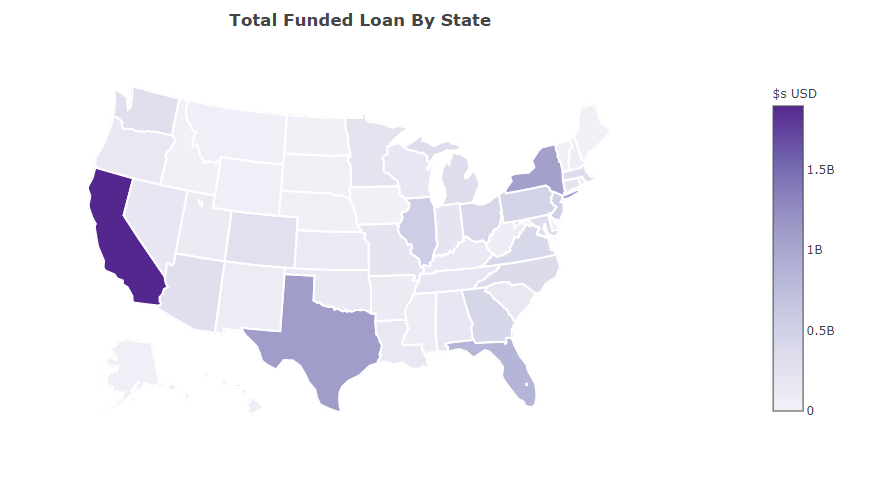
west와 SouthEast 지역이 default 나 charged off가 높아 가장 바람직하지 않은 대출 상태를 가지고 있다고 결론지을 수 있지만, 다른 지역에 비해 약간의 차이일 뿐이다.

대출건수를 보면 CA,NY,TX,FL 등이 높다. 특히 캘리포니아는 12만건을 넘어 평균적인 대출건수의 3배이상 높다.





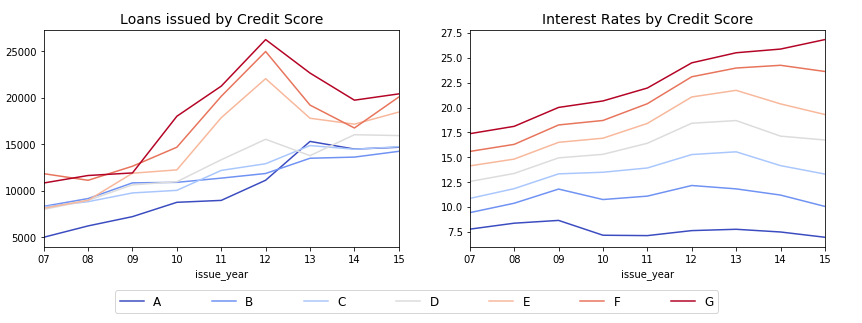
IOWA는 가장 높은 수준의 채무불이행 비율을 가지고 있지만, 대출건수가 작고 채무불이행 건수가 작기에 비율만 가지고 위험도가 높은 주라고 볼순 없다.



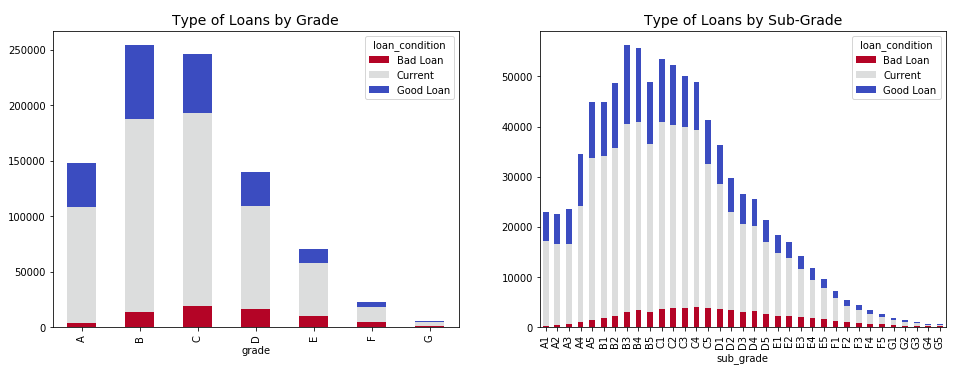
CA,NY,TX,FL 주들이 대출금액이 가장 많다. 대출건수 역시 많았고 대출금액도 많다.

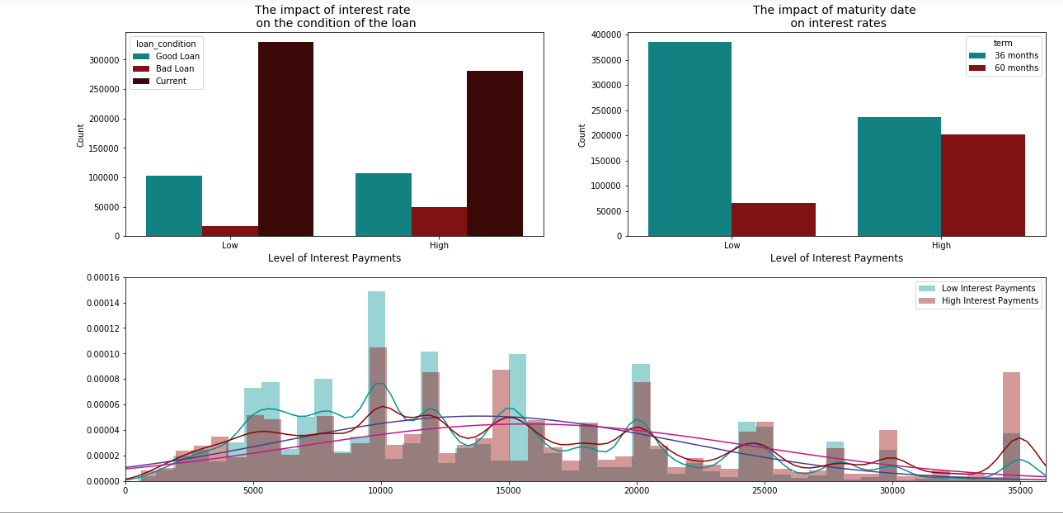
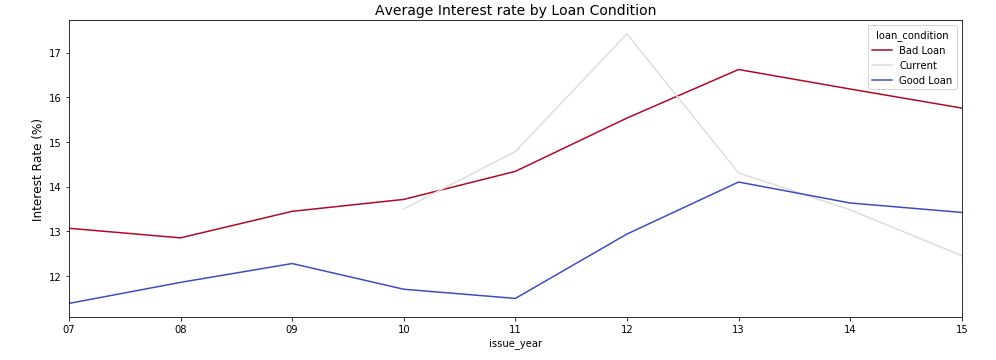
신용과 대출금액 관계: 신용이 좋을수록 대출이 적다. 반대로 신용이 나쁠수록 대출이 많다. 논리적으로, 등급이 낮을수록 이자가 더 높고 고객이 투자자들에게 갚아야 하는 이자는 더 높아진다. 그래서 대출상환이 더 어려워진다. 2012년이전까진 급상승하다 낮은 신용은 꺽이기 시작하고 고신용등급은 꾸준히 대출이 상승한다.

신용과 금리관계: 당연히 신용이 좋을수록 금리가 낮다는걸 보여준다. 특이점은 처음엔 비슷한 상승추세였지만 그후 시간이 지날수록 고신용등급일수록 순차적으로 하락하고 가장 낮은 G등급은 더 올라간다.



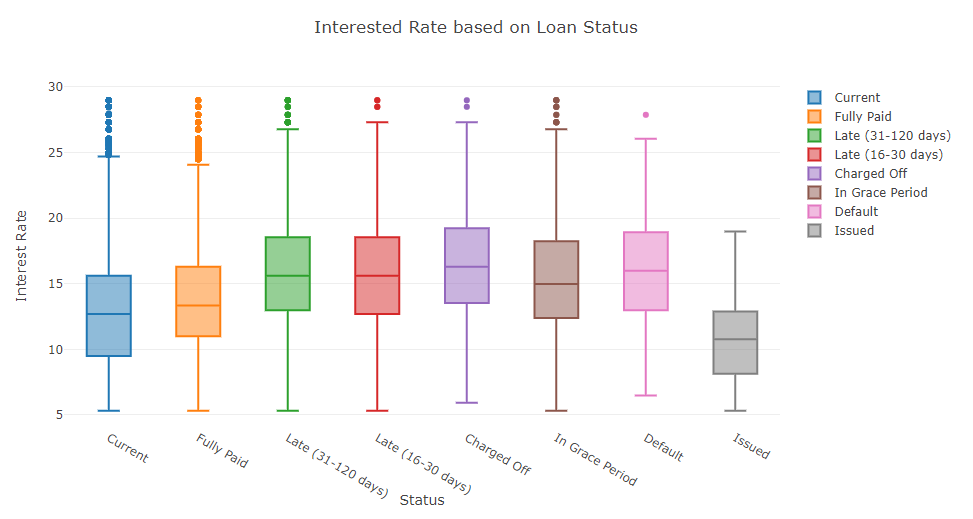
신용등급에 따른 론타입: B등급 이후부턴 잘 상환된 대출비율은 낮아지고 악성채무 비율은 높아진다. C등급의 고객들은 대출 불이행 가능성이 가장 높았다.

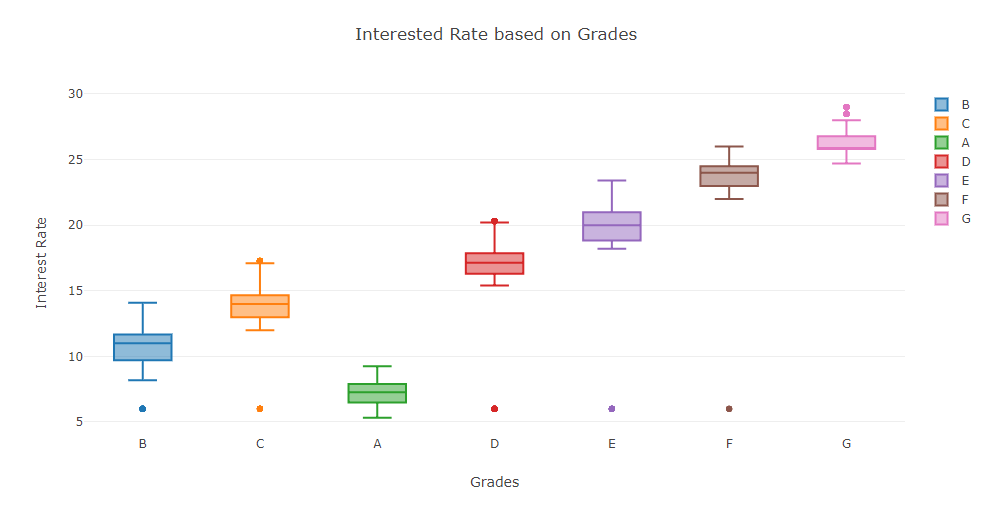




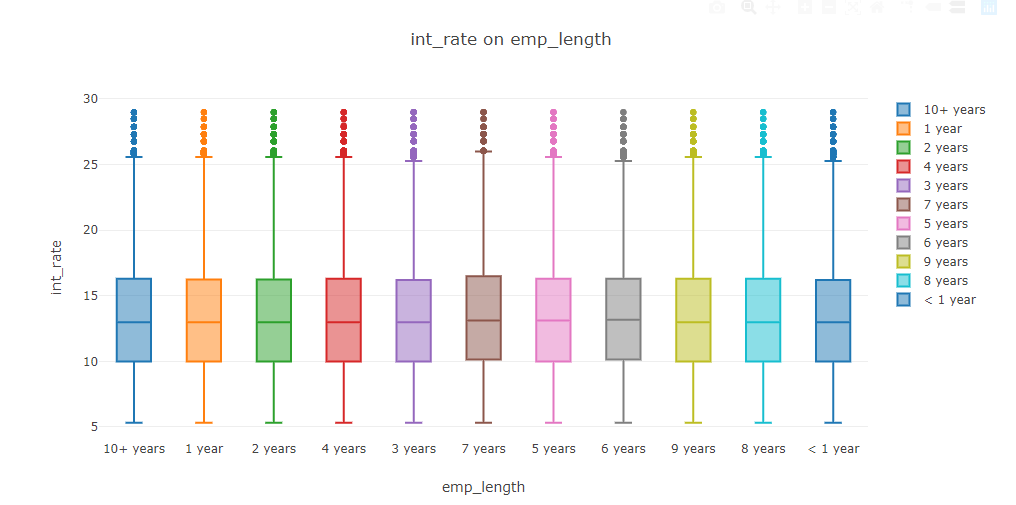
론 상태 별 시간흐름에 따른 평균금리 변화: 11년 이후 상승하다 13년 이후부터 정상,불량 두 종류 모두 대출금리가 낮아지며 일정한 금리차를 유지하고 있다.현재 대출상태를 보면 대출 건수는 늘어나지만 금리는 낮아지고 있다. 경기회복으로 인한 위험 선호심리가 살아나고 있다.

밑의 그림을보면 금리가 낮은 고신용일수록 현재 대출을 많이 받는것을 알수있고 채무불이행이 적다. 그리고 대출기간이 짧다. 반대로 저신용자 일수록 고이자에 대출기간도 길다. 그만큼 이자가 클것이다. 고신용자들은 5000~20000불 사이에 많이 있고 30000불 이상에선 오히려 저신용자들의 대출이 많다.

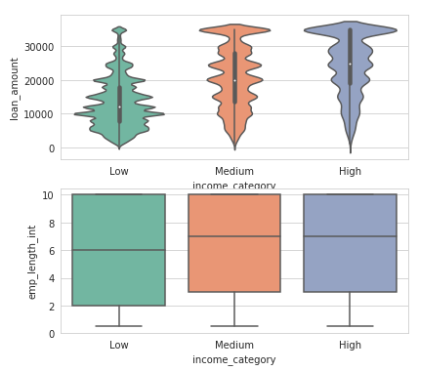


Current 나 Issued 가 이자가 가장 낮은건 흥미롭다. 렌딩클럽이 최근 대출승인을 보면 기존 정상적으로 대출상환을 잘한 그룹보다도 낮은 이자의 대출자들(고신용) 에게 대출을 허용하는것을 볼수있다.신용등급에 따른 이자율: 논리적으로 고신용자->저이율.

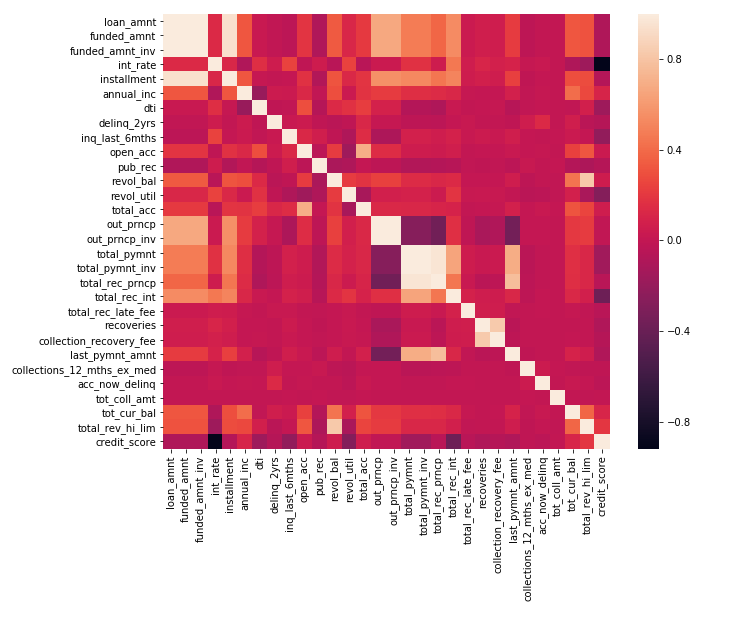
근무연수 별 금리: 근무연수와 연소득과의 어느정도 상관관계가 있기에 금리에 영향이 있을것으로 예측하기 쉽지만 실상 큰 차이가 없다.



소득 분석



저소득 범주: 연간 소득이 10만 usd 이하인 대출자.  
중소득분류:연간소득이 10만 usd 이상이지만 20만 usd 이하인 대출자.  
고소득항목: 연소득이 20만usd보다 높은 대출자.  
  
  
고소득 범주의 일부인 대출자들은 중저소득 범주의 사람들보다 더 높은 대출 금액을 받았다. 물론 연소득이 높은 사람은 더 높은 금액의 대출을 받을 가능성이 높다  
연소득이 높은 대출자는 소득이 낮은 대출자보다 근무연수가 더 길었다. 근무연수가 길수록 소득이 더 높아진다는 것을 확인할수 있다.

변수간 상관관계 

**Discussion**

The data from Lending Club show us the trends of lending business and what will be the potential risk in the future. For decreasing default, we should be careful for handling current loan which include high potential to be default. Because most of people try to lend money to payback another debt based on analysis of purpose. To track this one well, the industry can relieve default rate and lending company can keep doing their business with lower risk.

To be better analysis, firstly we need more data to predict recent trend. This data sent shows until 2015 and it is little bit out of date. That’s why we need more recent data for get accuracy. Secondly, based on analysis, we can find out the money amount which is requested by the borrower, which is invested by the investors and which is approved by the company is too similar and matched correctly. For this reason, we can tell this data is already controlled by Lending Club. More validate analysis, we need a data which is uncontrolled. It will be helpful to find default and other risks. Lastly, the debt consolidation was the main purpose of the loan (21.79%), but we cannot know which kinds of debt they want to payback so we need to research profoundly will be necessary. Also, it would be useful to analyse the population by region as it may influence the loan amounts.

By analyzing this data, we can verify the trends of P2P lending business. It has been keeping increasing of demand so the business looks great on surface. To be honest, this business has a good opportunity of growth but the result of analysis tell us we should be cautious. Most people lend money for payback of other debt so it means there is a possibility P2P will affect economy badly. Therefore, we need to keep analyzing the loan market and find a way to decrease the default rate for company and borrowers for stable economy.

1. Dvorah Rut, (2017, Dec. 28). 2017 Peer-to-Peer Lending: The Year in Pictures. *Perform*. Retrieved from http://blog.peerform.com/2017-peer-to-peer-lending/ [↑](#footnote-ref-1)
2. Transparency Market Research (TMR), (2019). https://www.marketwatch.com/press-release/peer-to-peer-lending-market-to-be-worth-us89785-billion-by-2024---tmr-2018-09-28. [online] Available at: https://www.marketwatch.com/press-release/peer-to-peer-lending-market-to-be-worth-us89785-billion-by-2024---tmr-2018-09-28 [Accessed 14 Apr. 2019]. [↑](#footnote-ref-2)
3. Marc Shoffman, (2018, Feb. 22). P2P default forecasts are rising, but should investors worry?. *Peer2Peer Finance News.* Retrieved from http://www.p2pfinancenews.co.uk/2018/02/22/p2p-default-forecasts-rising/ [↑](#footnote-ref-3)
4. Lending Club, (2019). https://www.lendingclub.com/company/about-us [online] Available at: https://www.lendingclub.com/company/about-us [Accessed 14 Apr. 2019]. [↑](#footnote-ref-4)