Questions:

1. Explain the data

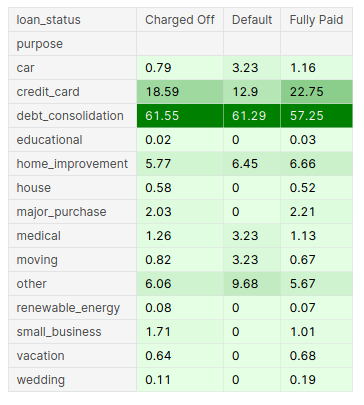
Data contain complete loan data for all loans issued through the 2007-2015, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information. The file containing loan data through the "present" contains complete loan data for all loans issued through the previous completed calendar quarter. Additional features include credit scores, number of finance inquiries, address including zip codes, and state, and collections among others

1. Clean the data (remove columns with null values, all random values or single categorical value + convert the data to proper int, float or date representations)

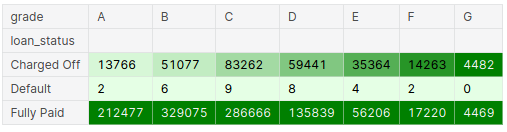
---Done--

1. Perform the following analysis (show appropriate plots) and state your hypothesis:
   1. Overall loan status (study total loans per category and total loans recovered)

People who gets loan for credit card or debt tends to charge off without paying

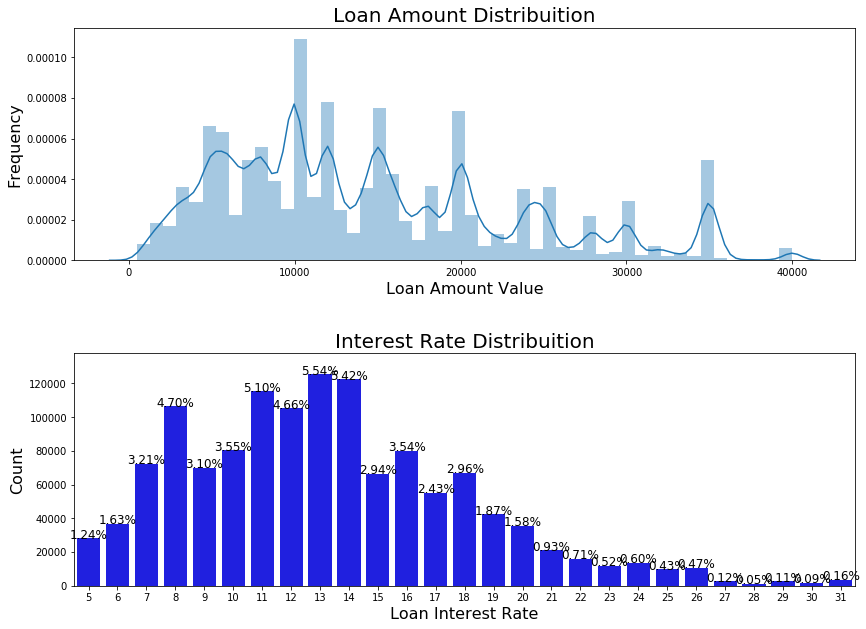
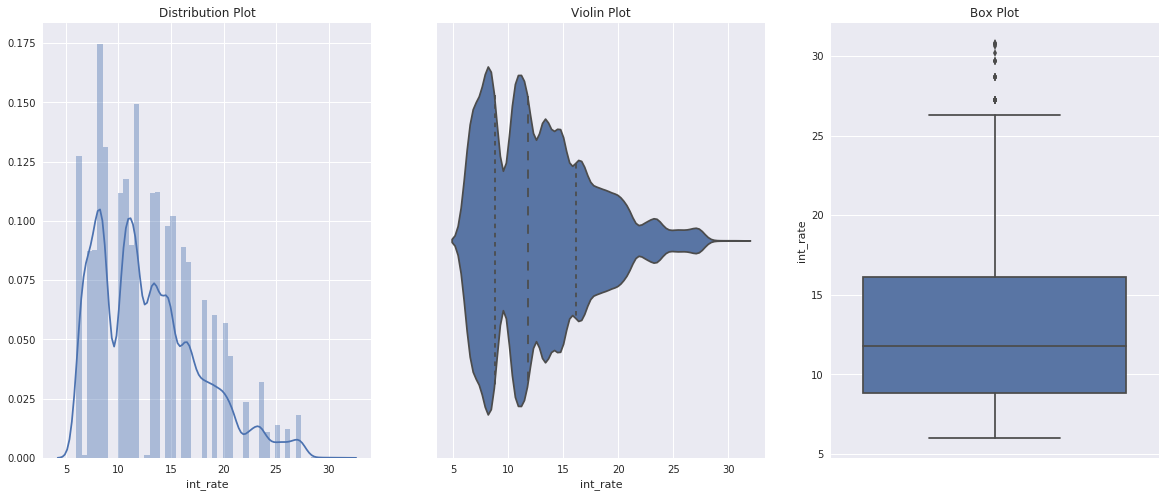


* 1. Understand loan based on grade

People of Grade A and B are more likely to pay back while Grade G around 50% of the time have the chance of charged off with loan.

* 1. Defaults by interest rate

Most of the loans interest rates are distributed between 7% to 14%. We can note that a big part of the loans are values until USD 10,000

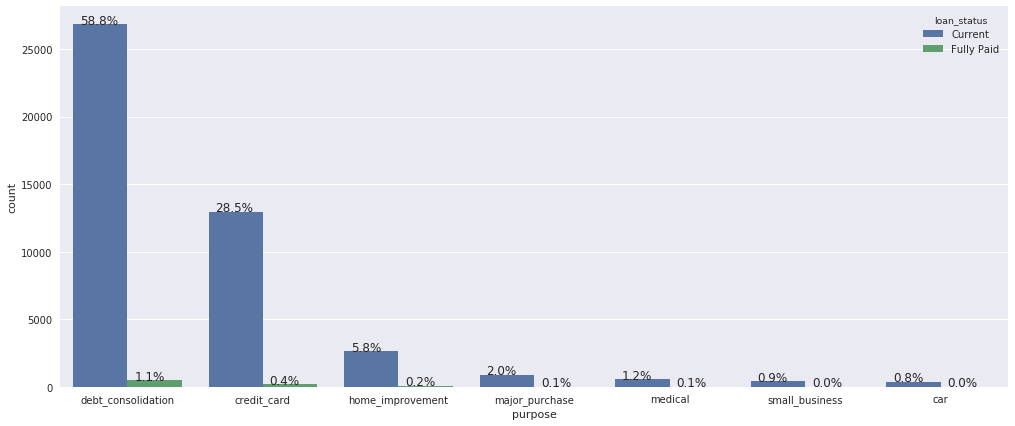


* 1. Defaults by Loan purpose

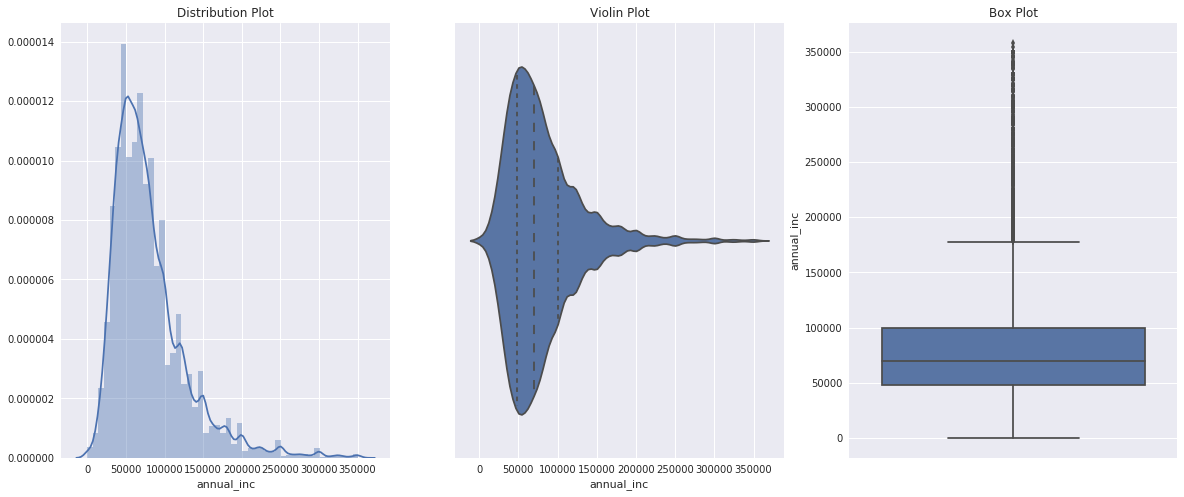
Approx 60% of the applicants applied loan for paying their Debt Consolidation

The top 3 purposes are:

* 58.8% of the Loans are to Debt Consolidation
* 28.5% are to pay Credit Card
* 5.8% to Home Improvement



* 1. Defaults by borrower’s income



Most of the applicants earns between 40000 to 90000 USD annually

* 1. Default by Revolving Line Util rate

Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| purpose | car | credit\_card | debt\_consolidation | educational | home\_improvement | house | major\_purchase | medical | moving | other | renewable\_energy | small\_business | vacation | wedding |
| loan\_status |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Charged Off | 46.57 | 59.03 | 56.07 | 46.12 | 46.74 | 42.66 | 41.65 | 49.84 | 49.88 | 49.54 | 52.55 | 47.33 | 47.29 | 51.3 |
| Default | 25.1 | 75.85 | 57.68 | 0 | 22.6 | 0 | 0 | 98.8 | 65.3 | 59.5 | 0 | 0 | 0 | 0 |
| Fully Paid | 39.2 | 55.23 | 52.42 | 37.22 | 41.85 | 35.86 | 38.11 | 45.62 | 47.2 | 46.06 | 46.21 | 42.73 | 45.39 | 46.31 |

* 1. Default by Debt to income Ratio

A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower’s self-reported monthly income.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| purpose | car | credit\_card | debt\_consolidation | educational | home\_improvement | house | major\_purchase | medical | moving | other | renewable\_energy | small\_business | vacation | wedding |
| loan\_status |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Charged Off | 17.49 | 20.51 | 20.81 | 11.05 | 17.76 | 16.72 | 16.38 | 19.56 | 18.52 | 18.6 | 17.19 | 15.72 | 19.12 | 15.17 |
| Default | 38.2 | 20.19 | 23.5 | 0 | 12.21 | 0 | 0 | 13.73 | 32.41 | 13.68 | 0 | 0 | 0 | 0 |
| Fully Paid | 14.45 | 18.05 | 18.41 | 11.27 | 15.61 | 14.01 | 14.41 | 17.02 | 15.94 | 16.5 | 15.86 | 14.05 | 17.27 | 13.83 |

1. Finally summarize your recommendations on better quality borrowers

### Top-5 variables to consider for loan prediction:

1. **Purpose of Loan**
2. **Employment Length**
3. **Grade**
4. **Interest Rate**
5. **Term**