

Lending Club - Case Study

upGrad - Submission on Lending Club - Case Study

Approach Taken

- ▶ Below is the high-level approach taken to solve the case study. Each steps is defined in detail in individual slides.
- 1. Data Cleansing
- 2. Field Creation
- 3. Univariate Analysis
- 4. Bivariate Analysis
- 5. Observations

Data Cleansing

First approach is to clean the data based on below factors

- ▶ Empty Fields
- ▶ Irrelevant Fields
- ▶ Irrelevant records
- ▶ Outliers

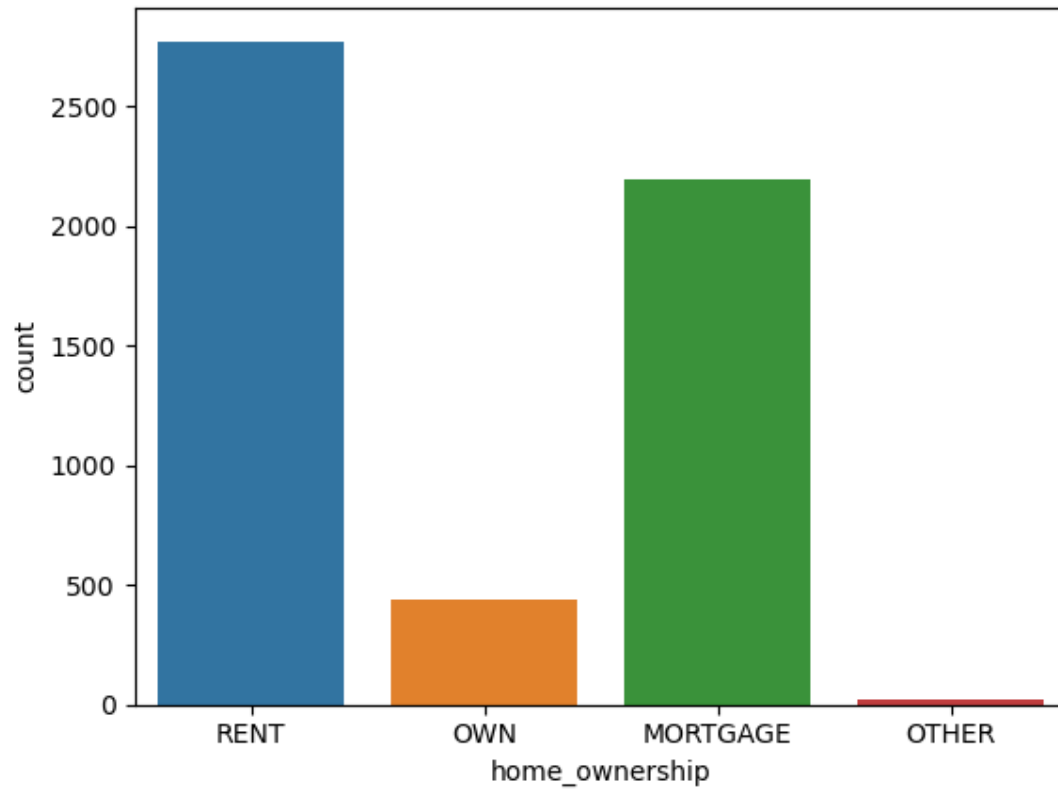
Field Creation

- ▶ Creation of additional fields from existing ones like extracting 'month' and 'year' from date field and adding it as separate columns

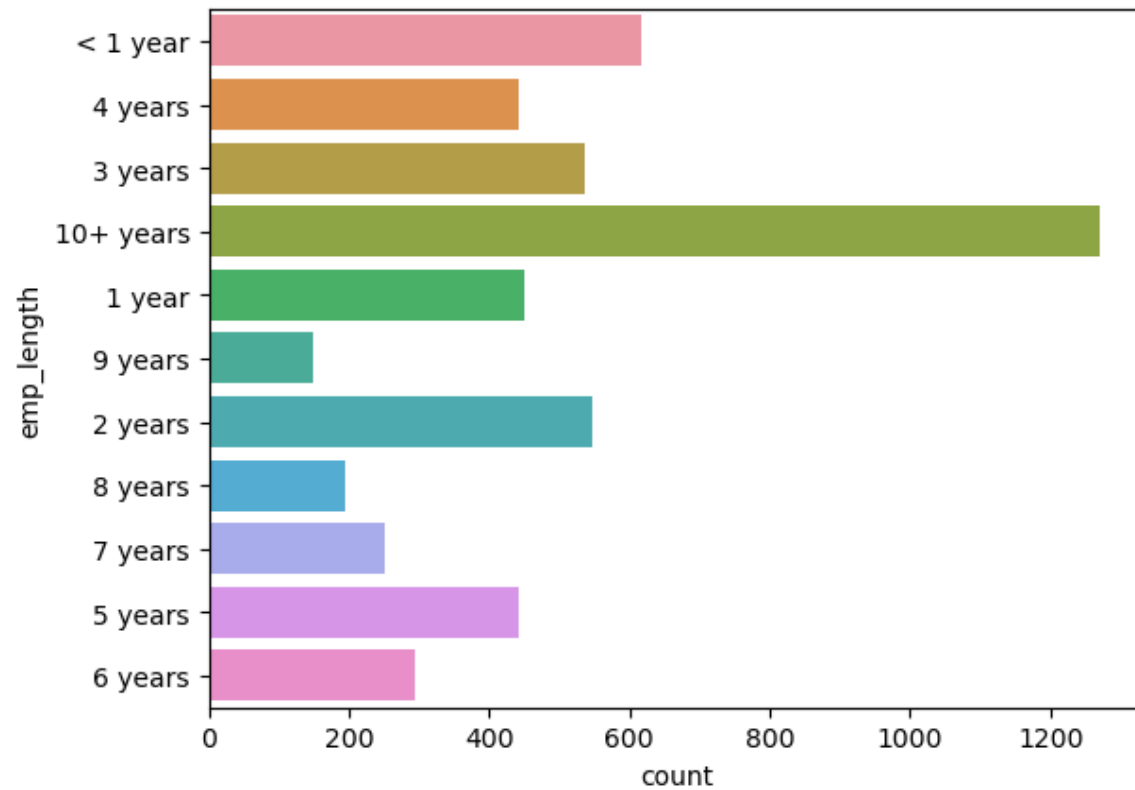
Univariate Analysis

- ▶ Univariate analysis will mostly be done on the 'Charged Off' records and by using the count of 'charged off' loans against different fields we will try to understand patterns.
- ▶ As part of this slide only relevant charts will be included

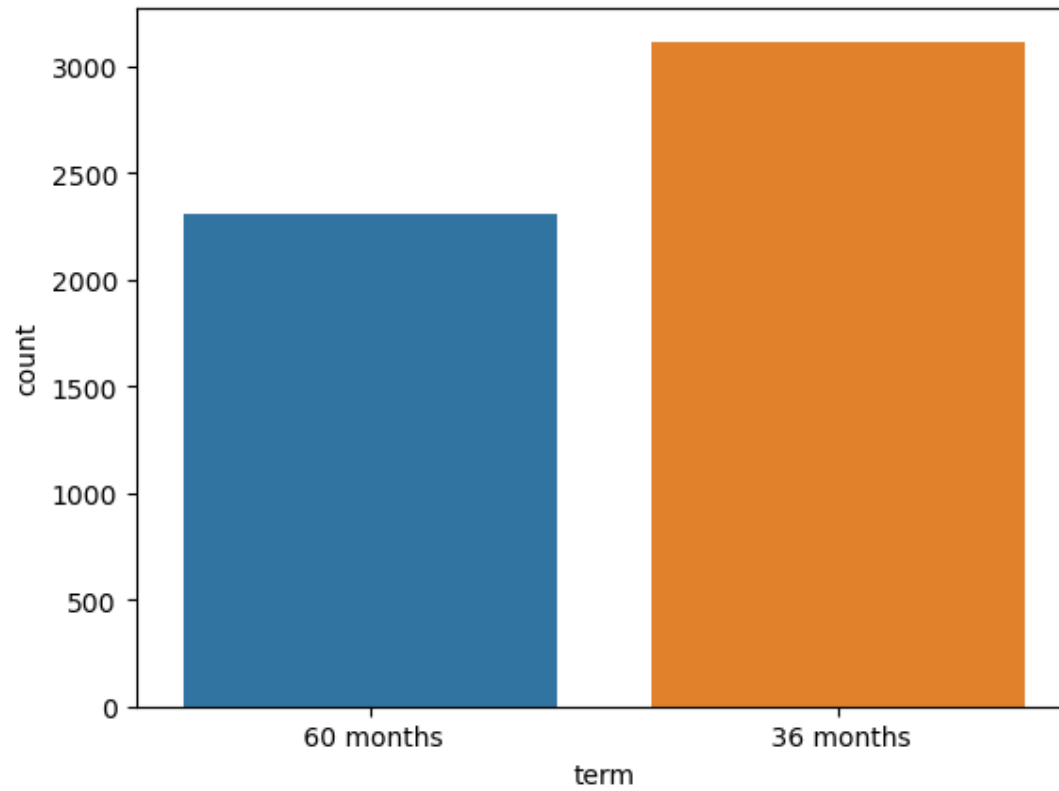
Univariate Analysis (Contd..)



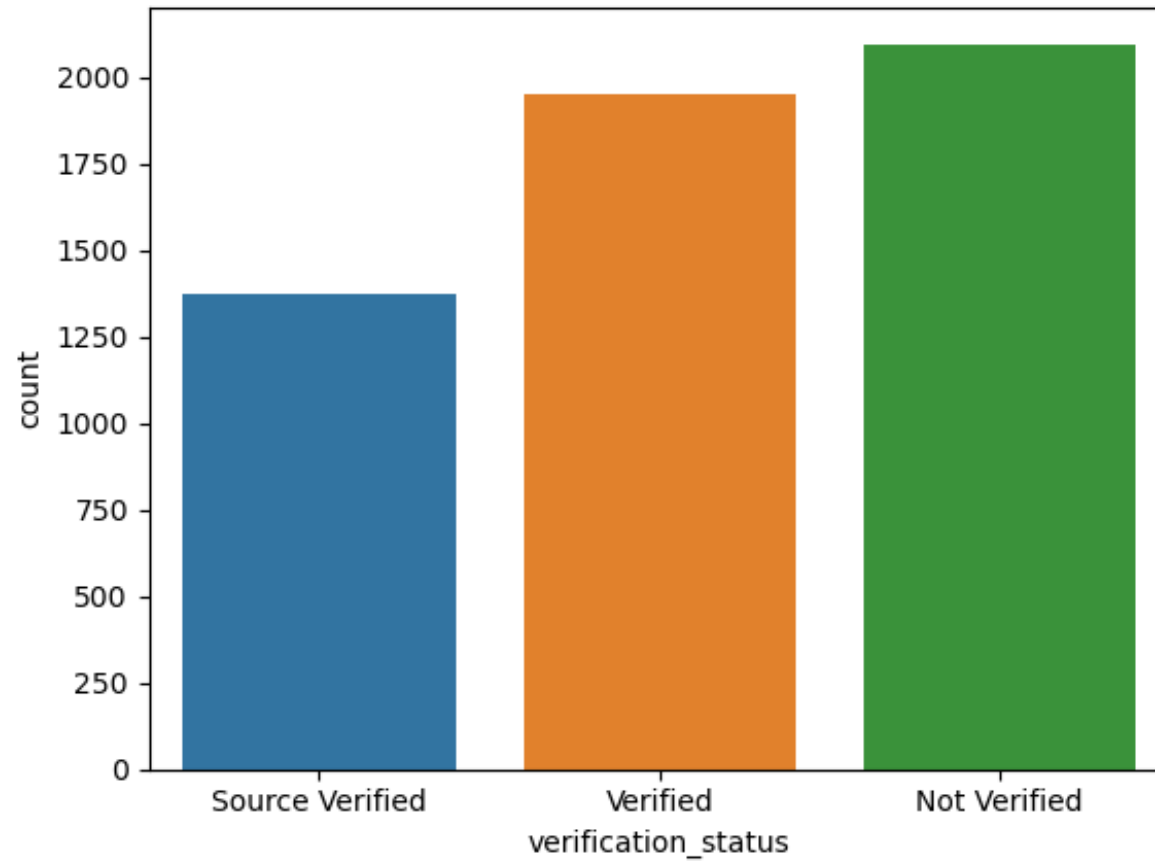
Univariate Analysis (Contd..)



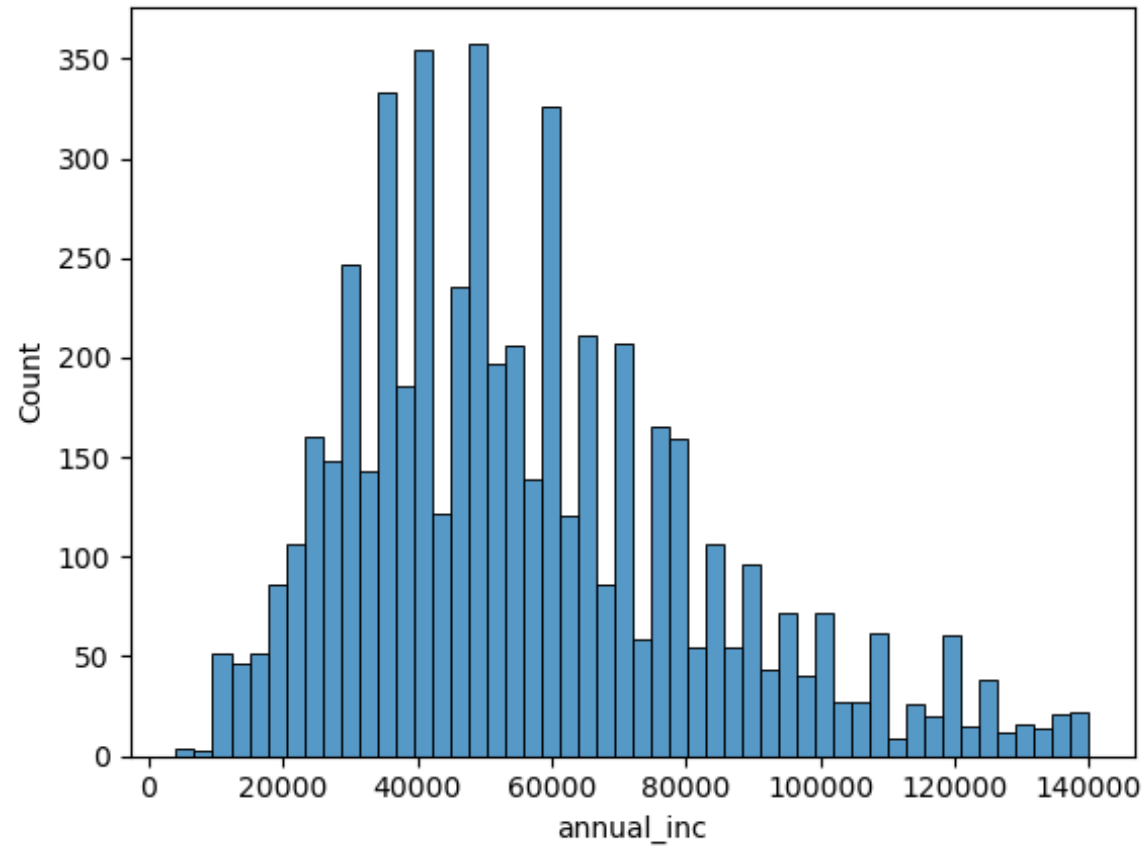
Univariate Analysis (Contd..)



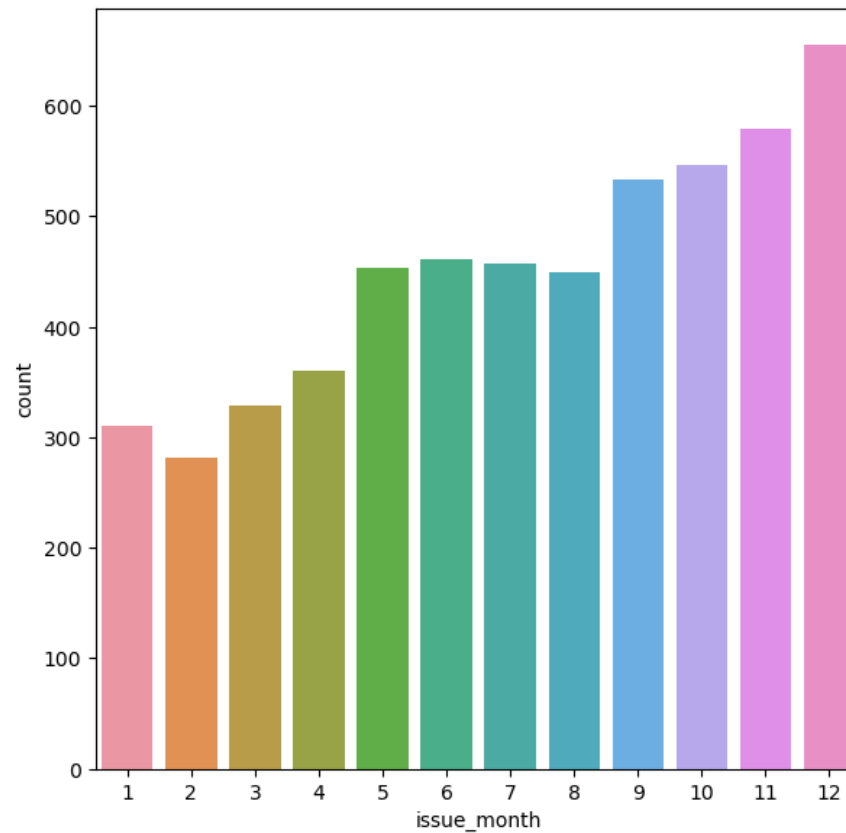
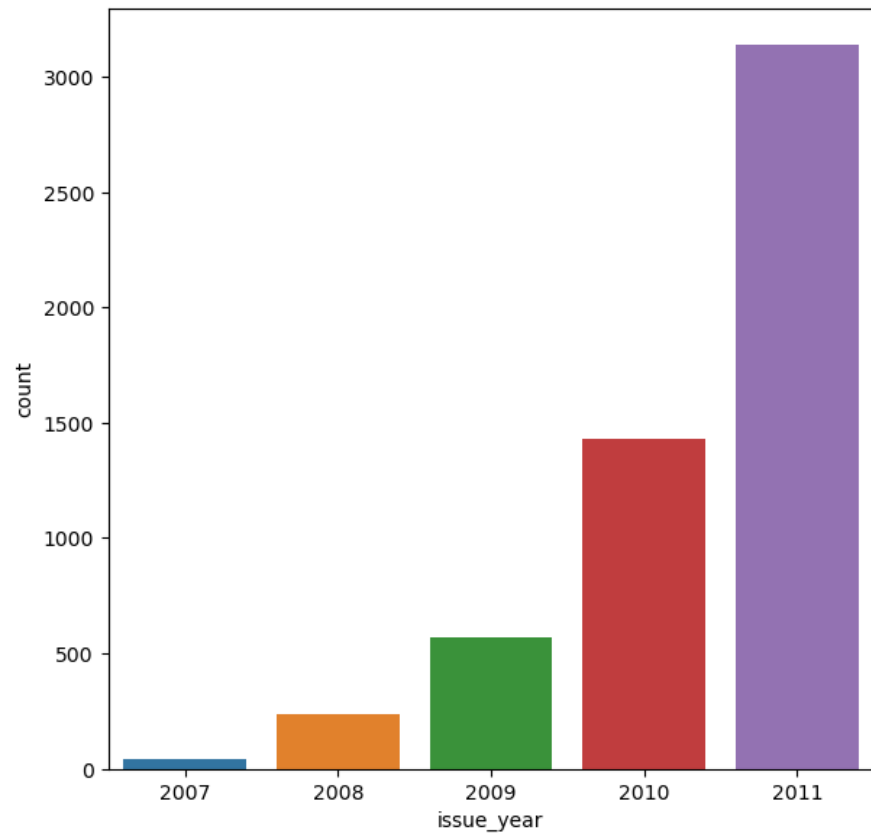
Univariate Analysis (Contd..)



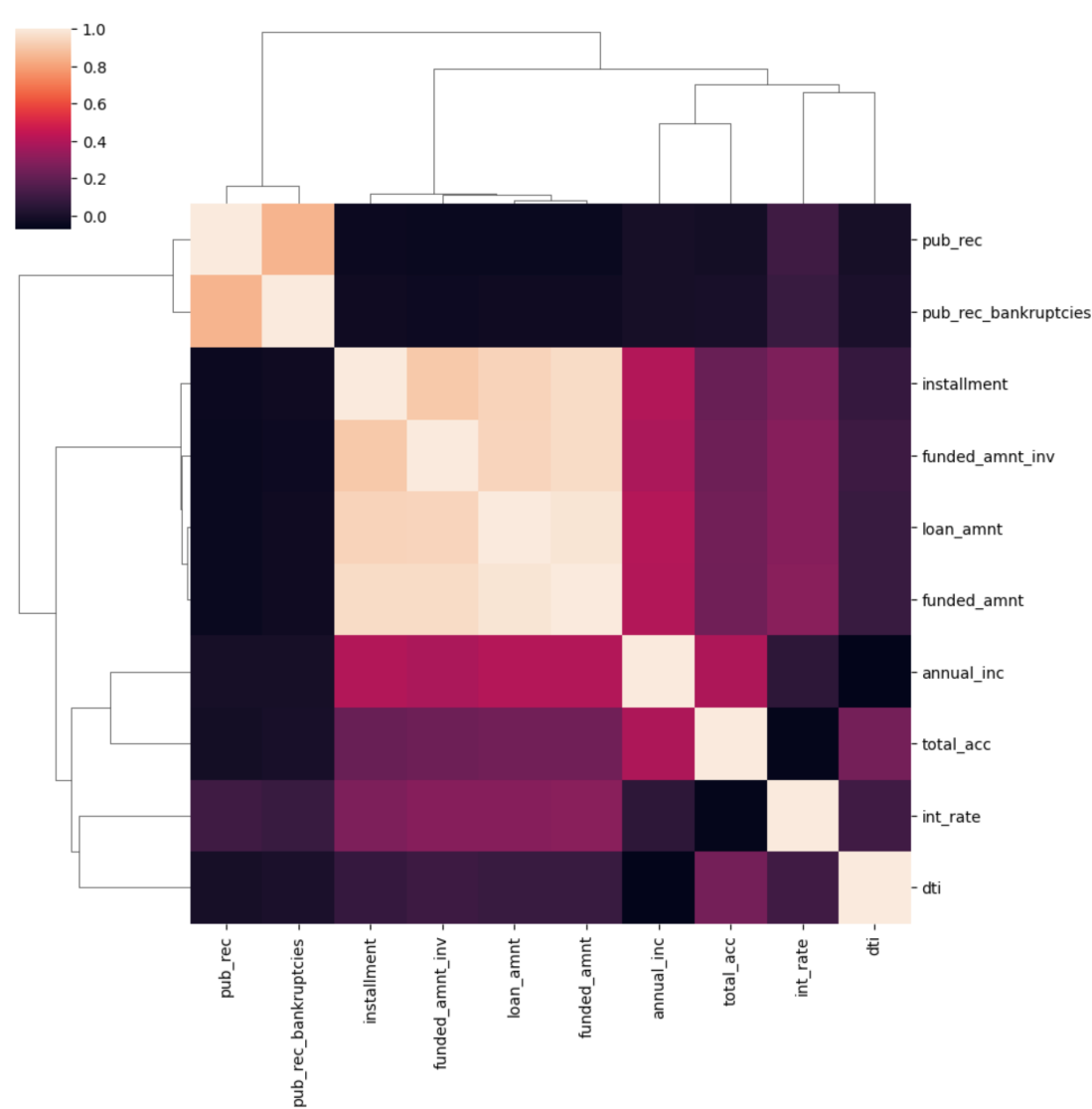
Univariate Analysis (Contd..)



Univariate Analysis (Contd..)



Bivariate Analysis



Observations

► Based on correlation graph below are the observations

- There is high degree of correlation between 'pub_rec' and 'pub_rec_bankruptcies'
- There is high degree of correlation between 'installment', 'funded_amnt_inv', 'loan_amnt' and 'funded_amnt'

► Based on above countplots the probability of customers defaulting is when

- Customers having house_ownership as 'RENT'
- When the purpose is 'debt_consolidation'
- Customers with employment length of 10
- Customers availaing term of 36 months
- When the loan status is 'Not verified'
- When the number of derogatory public records is 0
- When the 'annual_inc' is in the range of 35K - 60K appx