



Lending Club Case Study

PURNA CHANDRARAO V
RAMAKRISHNA K

The problem

Company

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

Borrowers can easily access lower interest rate loans through a fast online interface.

Context

Lending Club wants to understand the **driving factors** behind loan default, i.e. the **driver variables** which are strong indicators of default.

The company can utilise this knowledge for its portfolio and risk assessment.

Problem statement

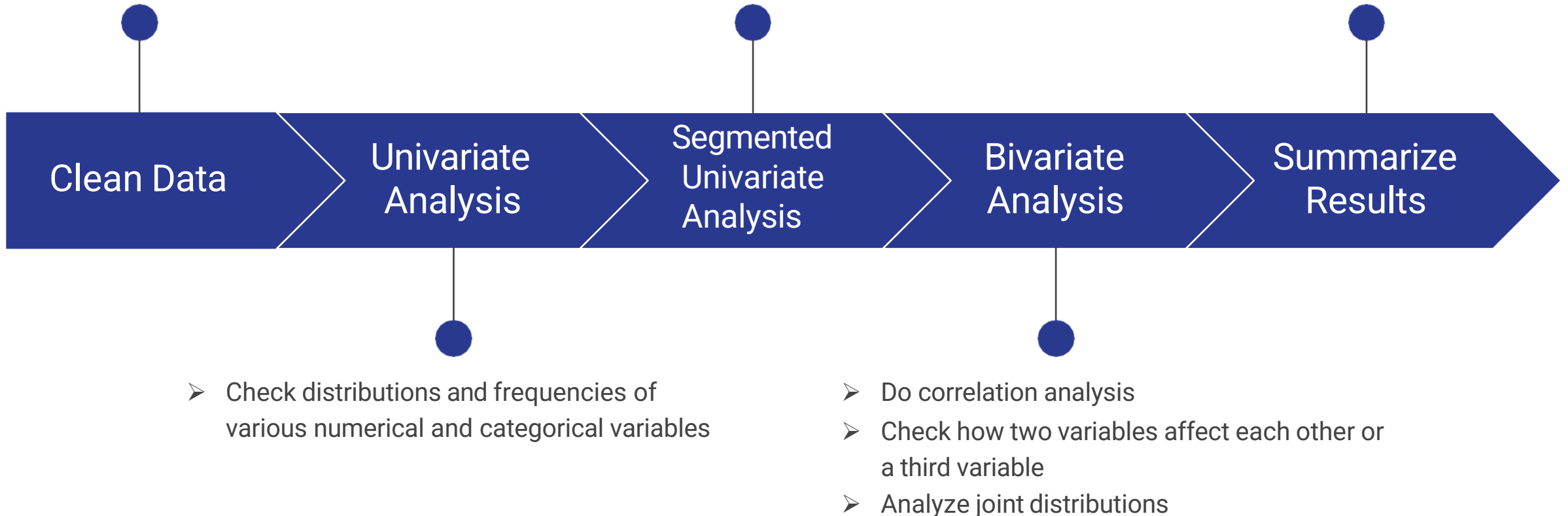
As a data scientist working for Lending Club analyze the dataset containing information about past loan applicants using EDA to understand how consumer attributes and loan attributes influence the tendency of default

Analysis Approach

- Drop columns with null values, unwanted columns, all random values or single category value
- Convert values to proper int, float, date representations

- Analyze variables against segments of other variables
- Create derived variables

- Publish insights and observations



Data Clean up

Observation	Drop Column Count
All Null/NA values	44
All Unique values for all rows	3
Only 1 value for all the rows without any missing values	6
Only 1 unique value as 0.0 for more than 99.9% of the rows	3

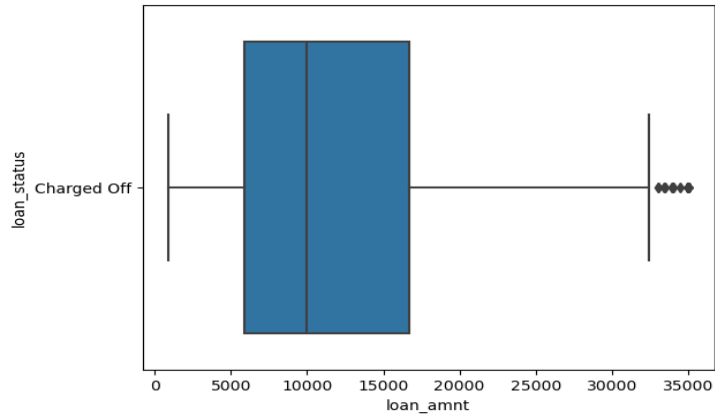
Handling Incorrect Data Types	Column Names
Columns converted to Numeric	term, int_rate, revol_util
Columns converted to Date	issue_d, earliest_cr_line, last_credit_pull_d

Univariate Analysis

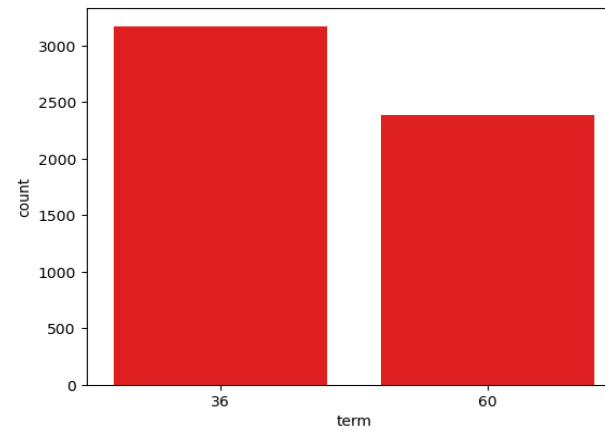
Observation	Column
14% of the loan borrower are charged off	Loan_status
75% of Charged off customers have taken loan < 20k	Loan_amnt
Charged off customers count by loan term <ul style="list-style-type: none">• 36 Months – 3173• 60 Months – 2382	Term
Most of the Chargedoff customers interest_rate divided into buckets of 5% is between 10%-20%	int_rate
Most of the ChargedOff customers are 10+years exp	emp_length
debt_consolidation is Major reason impacting Charge Off	purpose
40-60k range income has has ChargedOffs followed by 20-40k	annual_inc

Univariate Analysis Graphs

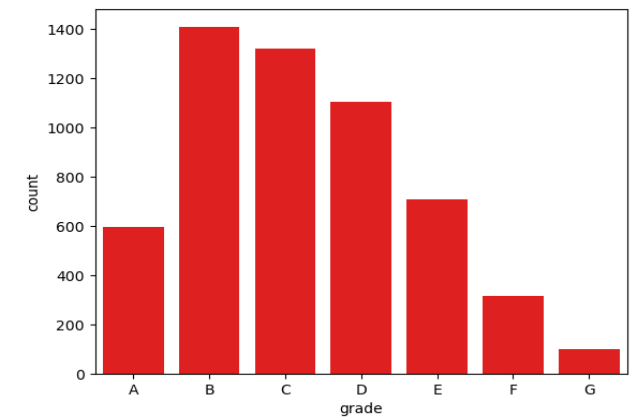
Charged Off Customers By loan_amnt



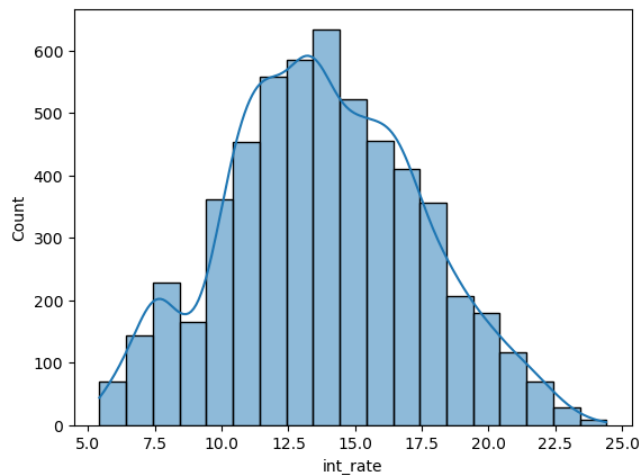
Charged Off Customers By term



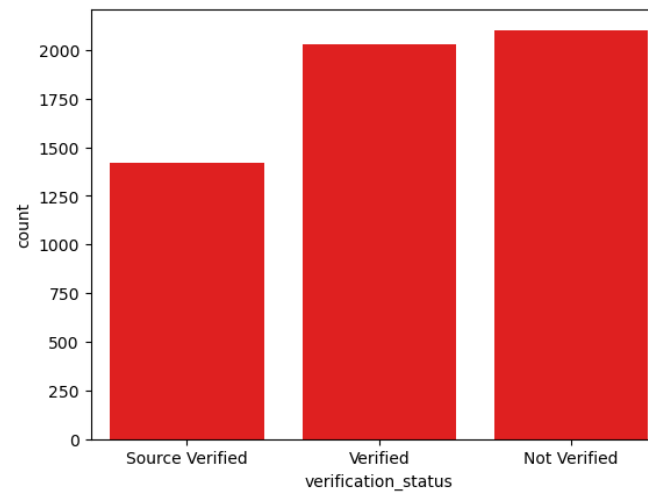
Charged Off Customers By grade



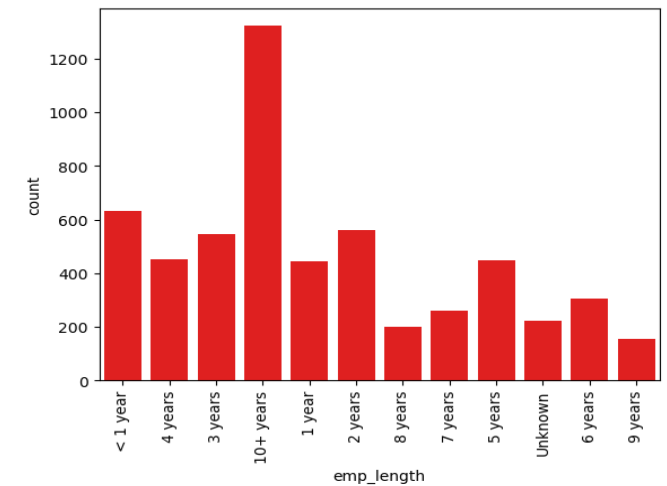
Charged Off Customers By int_rate



Charged Off Customers By verification_status



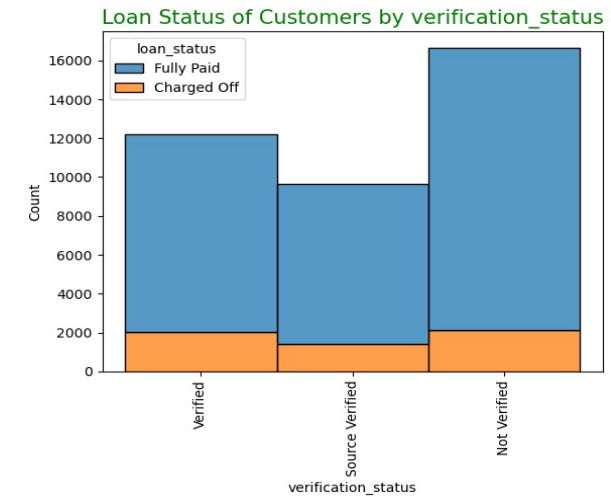
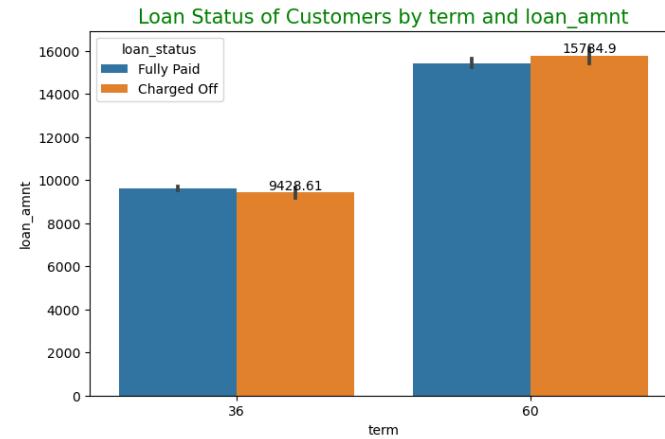
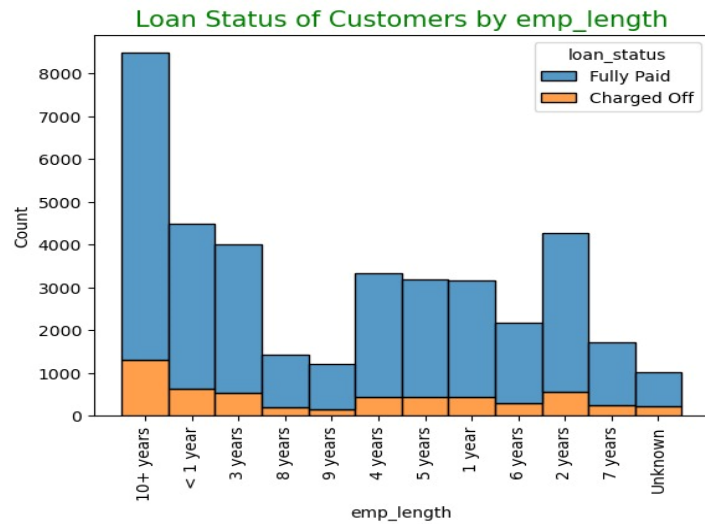
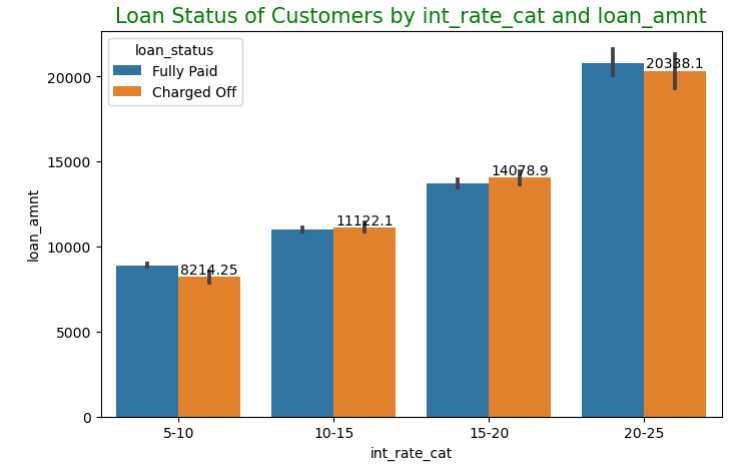
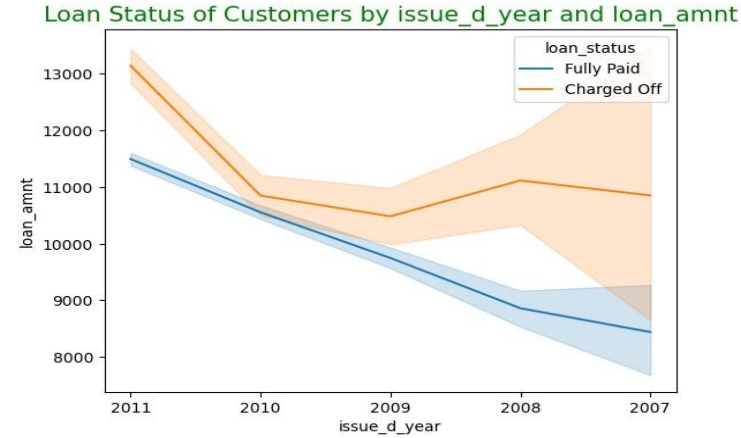
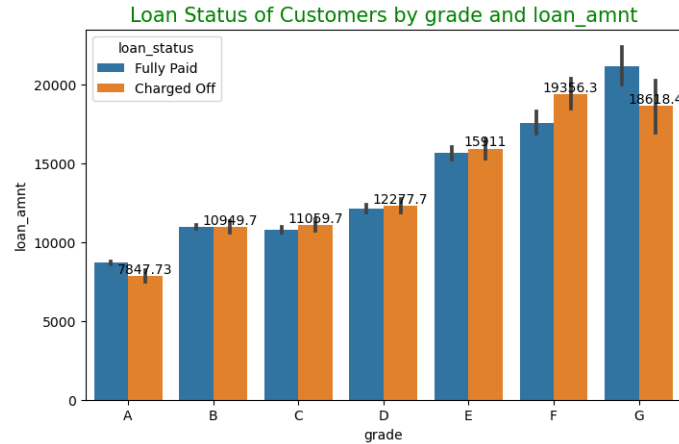
Charged Off Customers By emp_length



Bivariate Analysis

Observation	
Every year the Charged off customers count as well as the number of loans are increasing	
<u>Term 60 Months</u> with Loan Amount mean 15k+ has more ChargedOffs than term 36 months	Term vs Loan_amnt vs Loan_status
<u>Purpose Home Improvement</u> having high ChargedOffs with Annual Income range 60-80k	Purpose vs Annual Income vs Loan_status
High <u>interest rate 20-25%</u> has high Charged Offs with Mean Loan Amount 20k+	Loan amnt vs int_rate vs Loan status
<u>Customer Grades F & G</u> are high likely to be Charged Offs with mean Loan Amounts 18k+	Grade vs Loan amnt vs Loan status
Annaul Income - Avg 75k & Home Ownership - Mortgage have high Charged Off	Annual Income vs Home Ownership vs Loan staus

Bivariate Analysis Graphs

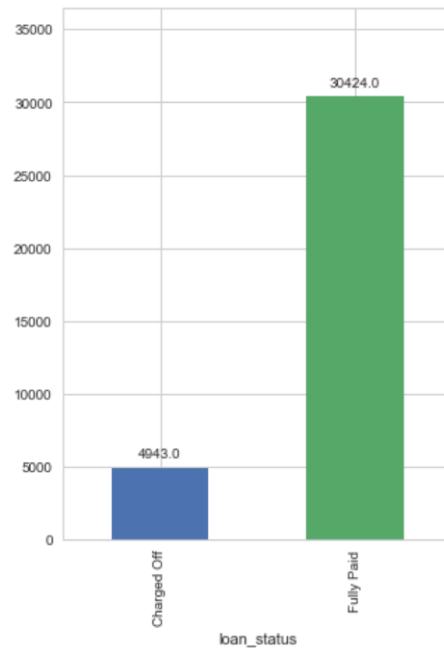


Overall Loan Status Analysis

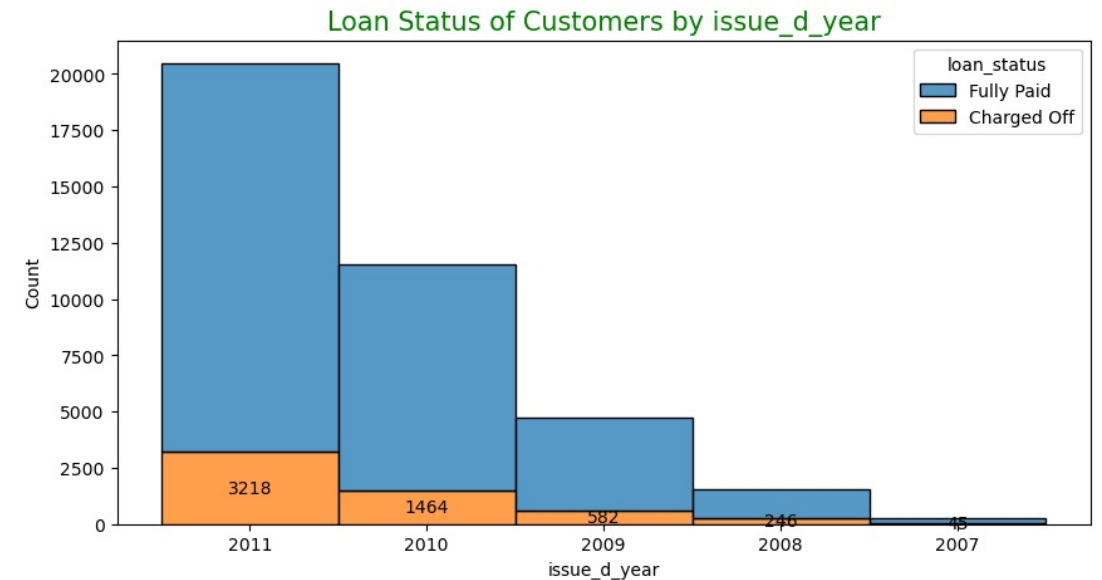
Approximately 14% of loans are charged off

Any variable that increases percentage of default to higher than 16.5% should be considered a business risk.

Total Loans



- There is huge increase in number of loans granted in year 2011
- Every year the no. of loans granted and charged off are increasing



Summary

- **Approximately 14% of loans are charged off**
- Every year the Charged off customers count as well as the number of loans are increasing
- Factors affecting charge off
 - Customer Grade (F & G) with Loan Amount 18k+
 - High interest rate 20-25%
 - Employee length 10+ year
 - Purpose – Home/House and small business
- There have been some high value loans extended to borrowers with revolving line utilization rate of higher than 75%.