LENDING CLUB CASE STUDY

Submitted By:

Santokh Singh, Raghu Ram Ratan Cherukumilli

Objective

To identify key factors or markers contributing to charge off while issuing a loan by the firm in order to reduce financial loss created by a defaulted loans providing better risk assessment and profitable portfolio.

Methodology

Exploratory Data Analysis (EDA) is the methodology used in case study.

Exploratory Data Analysis (EDA)

The analysis is directed towards finding column relationships towards objective.

- Removal of empty columns and Rows
- Removal of outliers and duplicate rows
- Reformated data for further analysis
- Removals of columns not providing any information
- Deriving columns out of existing ones

Data Cleaning

- Analysis of different columns
- Checking pattern of distribution in a column
- Plotting pie charts, box plots, bar plots, dist plots etc. for data visualization
- Segmented Univariate Analysis

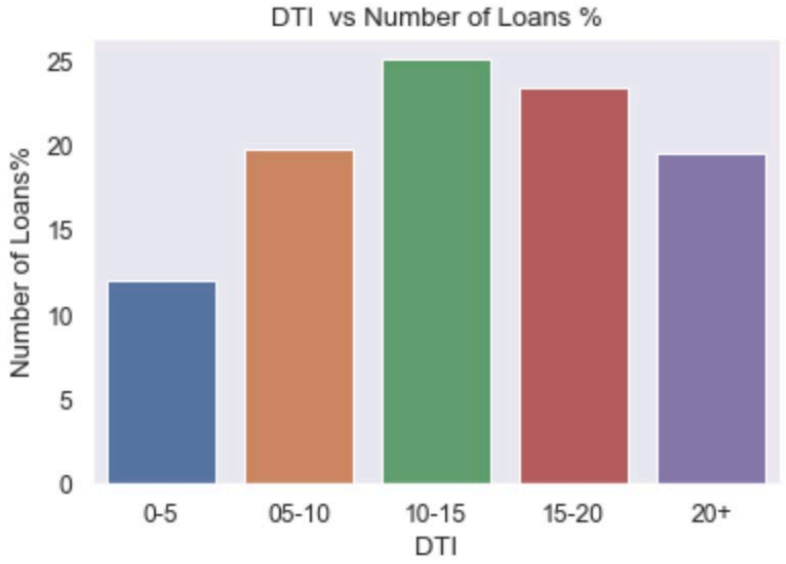
Univariate Analysis

- Identification of columns for analysis
- Pivoting over columns and aggregating data.
- Plotting heatmaps, bar plots, count plots etc. and look for pattern and insights.

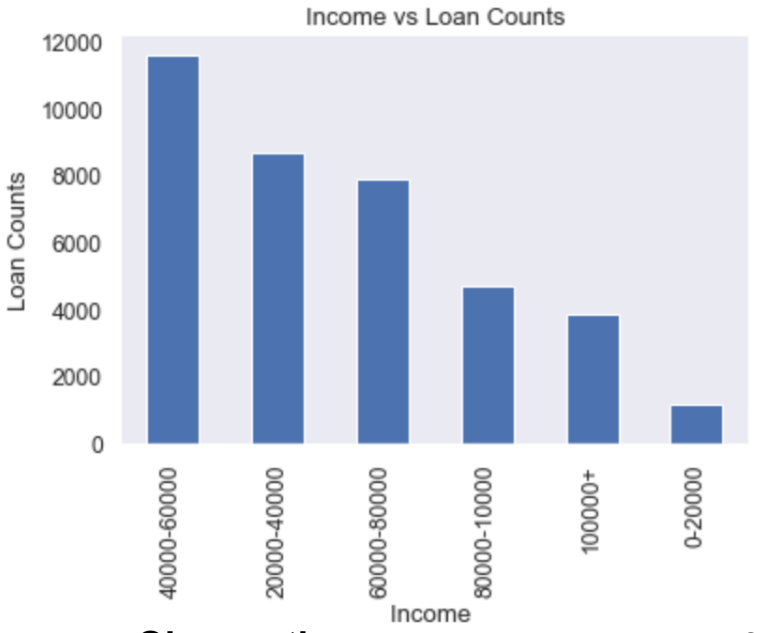
Bivariate Analysis

Observations

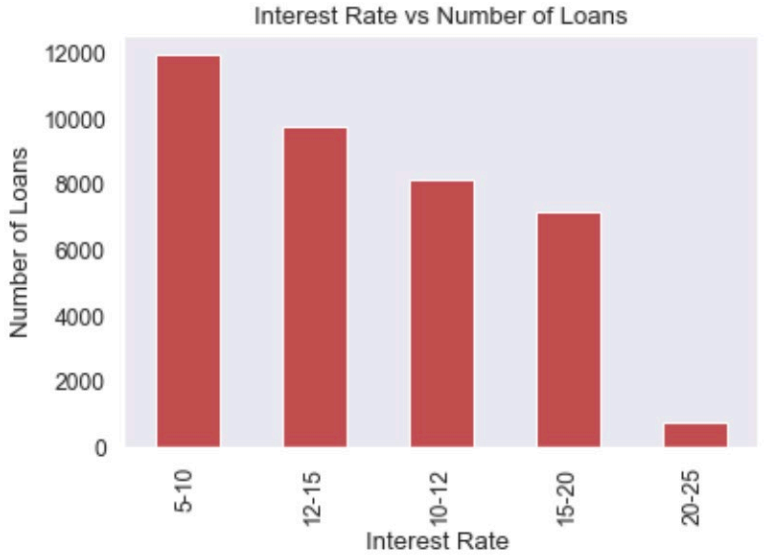
General trends observed in data

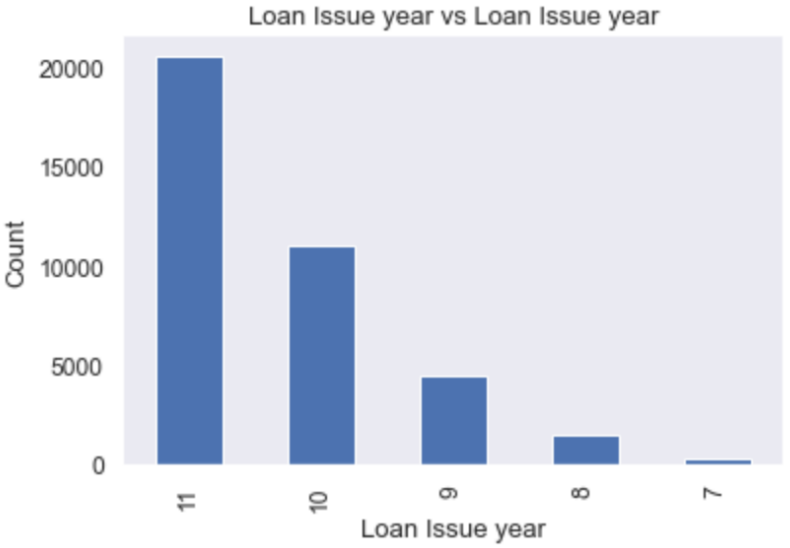


Observation: Debt to Income (DTI) ratio falls in range of 10-15 % for most of the loans issued.



Observation: Annual income range of 40k-60k take maximum number of loans among all income category.





Observation: Maximum number of loans get interest rate range between 5-10 %.

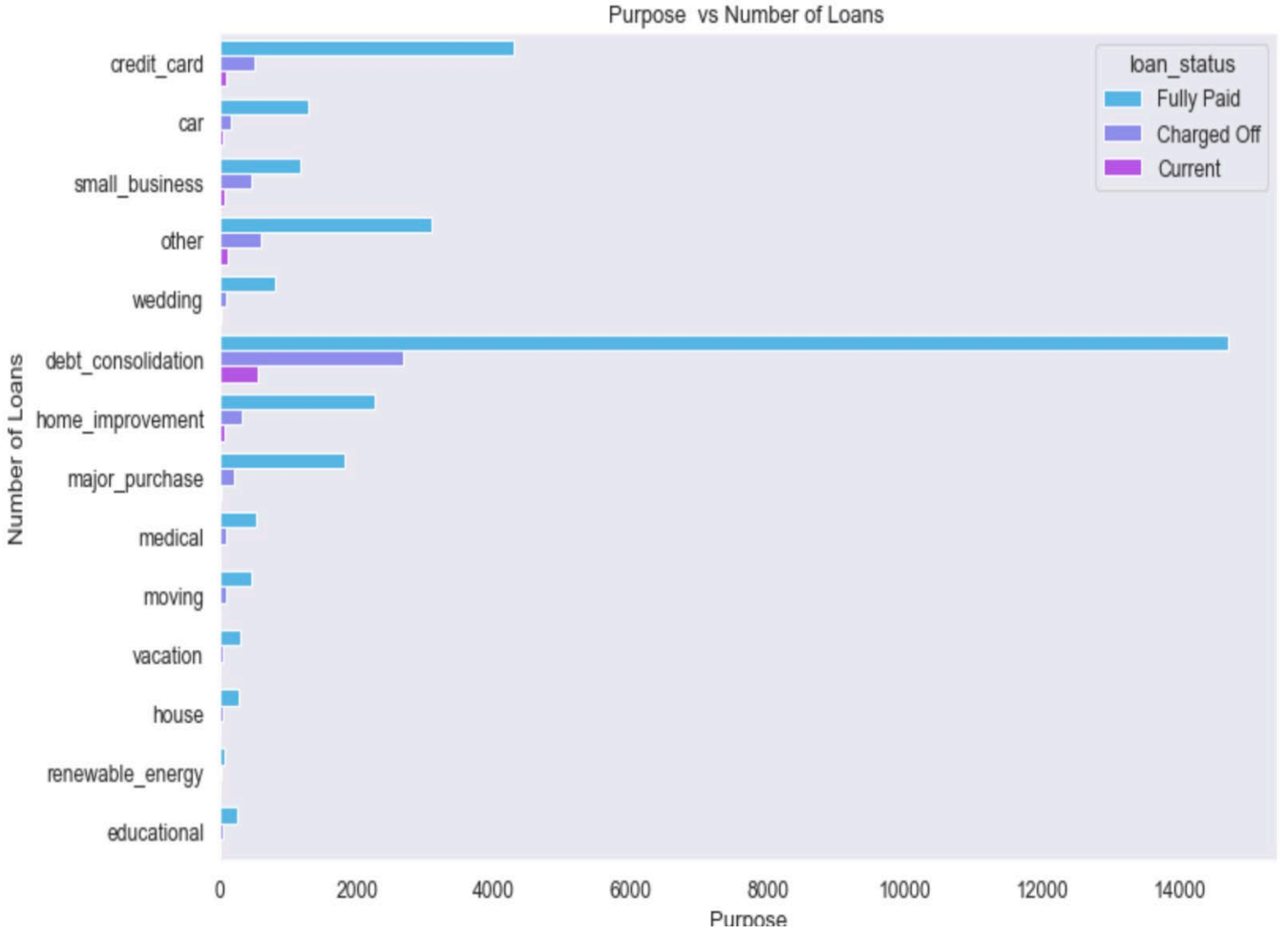
Observation: Its observable that the number of loans issued are almost getting doubled every subsequent year from 2007 to 2011

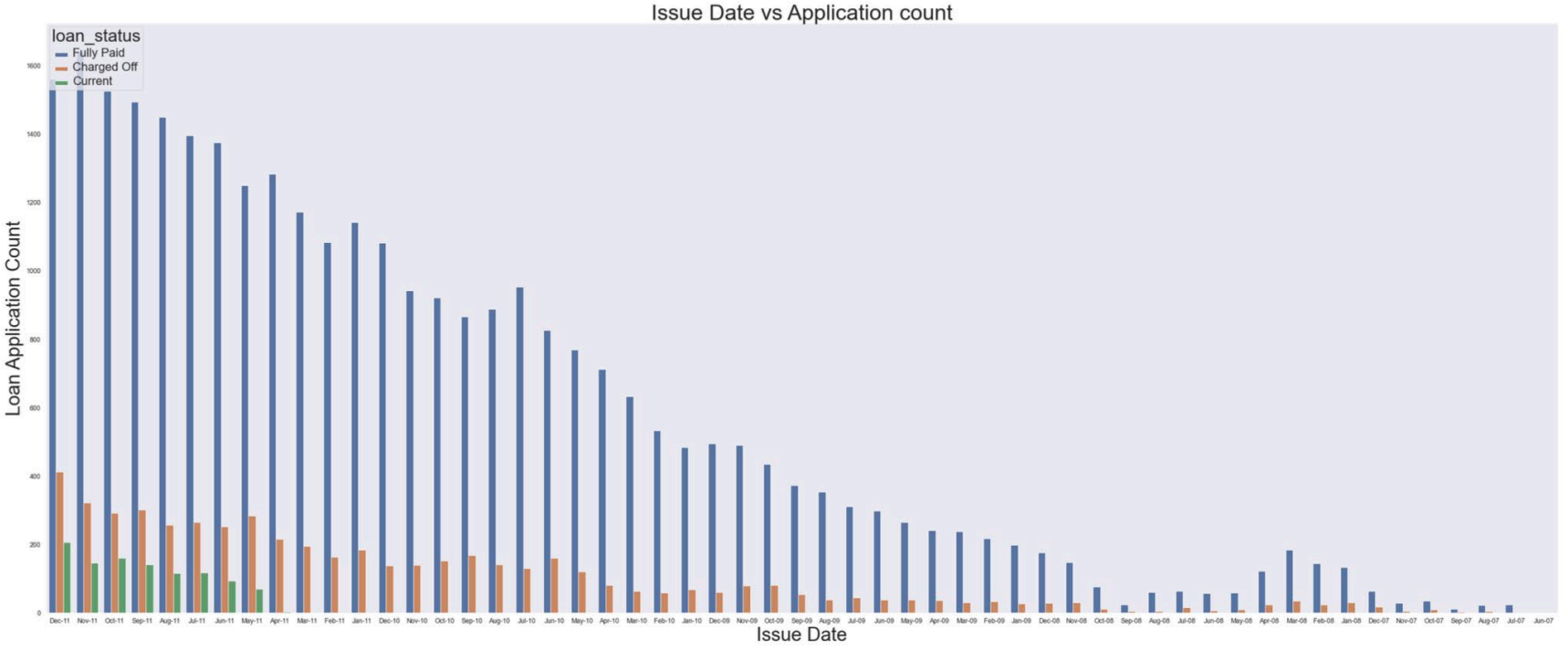
Observation:

1. Debt Consolidation and credit card are the major loan purposes.

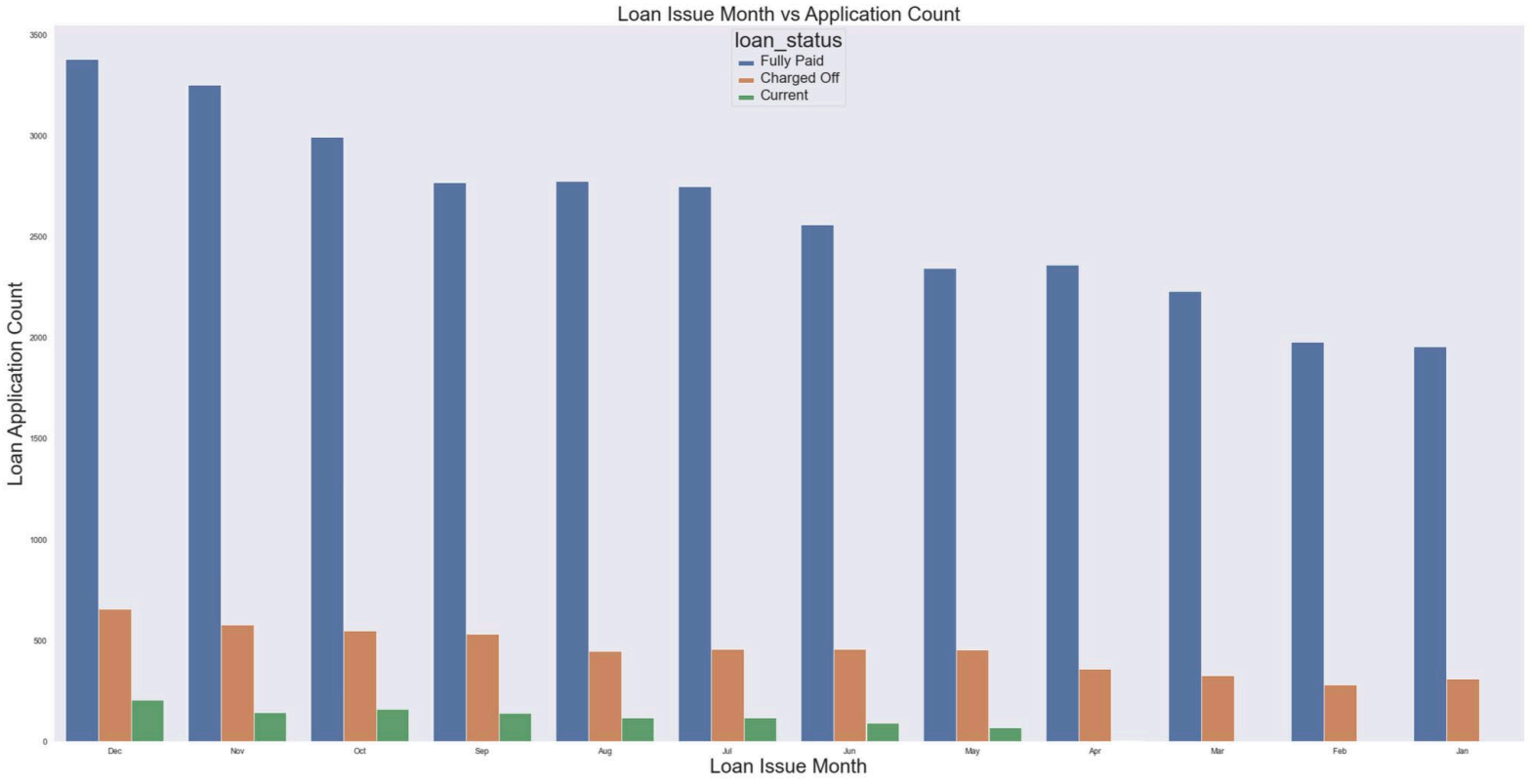
2. After that home improvement, other and major purchases purposes have high number of loans.

3. Debt consolidation has highest contribution to charge off by number because maximum number of loans are also taken in this category. Credit Card also has a similar case.





Observation: There is strong growth in number of loans from 2007 to 2011; although there is some slow down from April 2008 to Oct 2008



Observation: Maximum number of loans are taken in the 4th quarter, it almost remain same in 3rd quarter and keep rising in first 2 quarters.

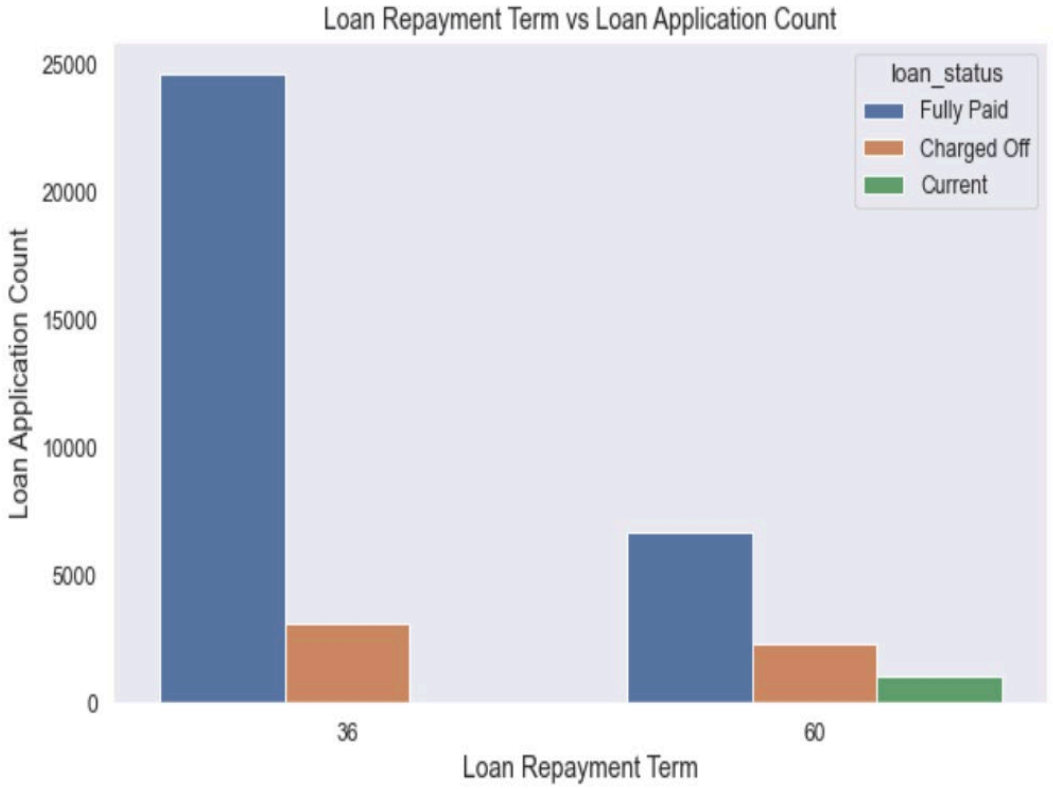
Inferences

Key factors contributing to charged off

Inference

Loan Term of 60 months is near to 12% more likely to charge off compared to 36 months.

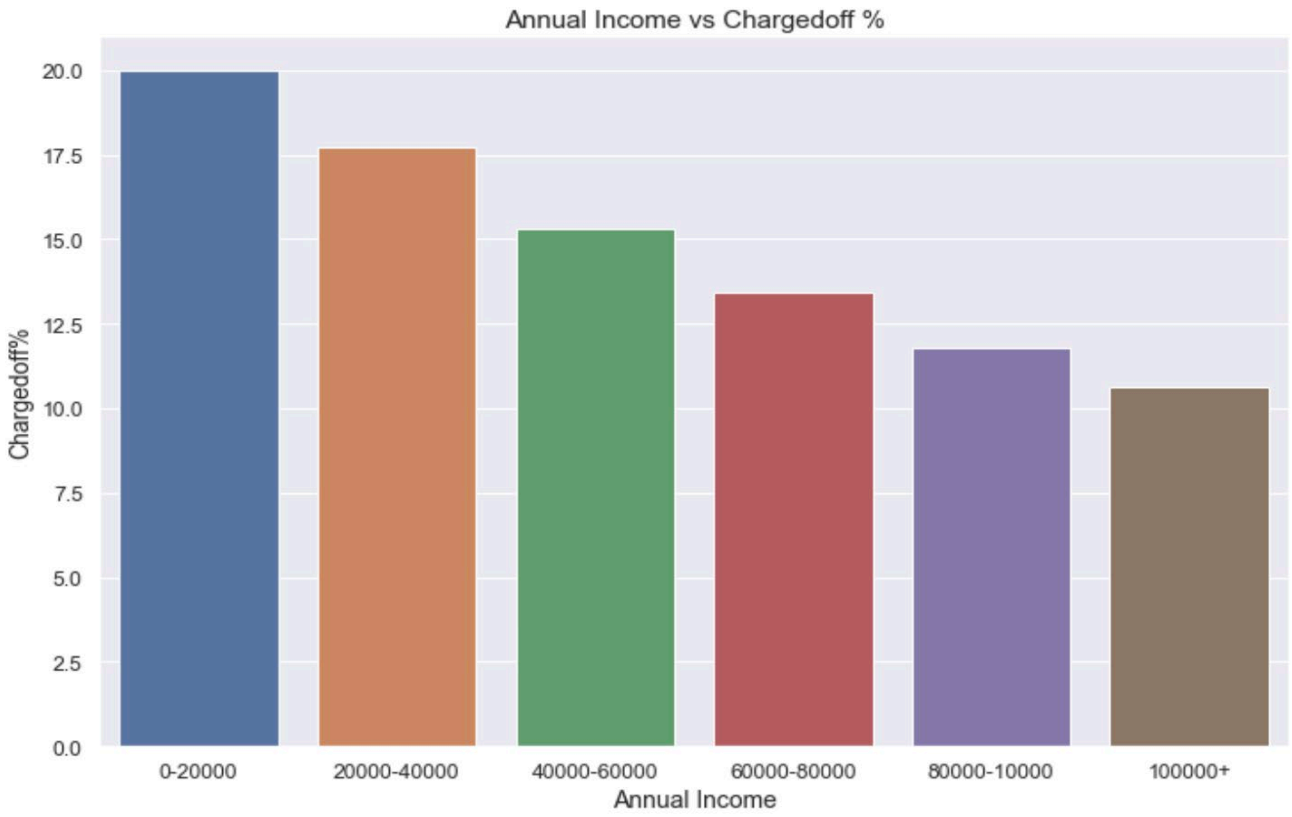
Reducing the term length will reduce the risk of charge off.



Inference

The **low annual income** range of less than 20K has the highest charge off near to 20% and the charge off decreases with higher income.

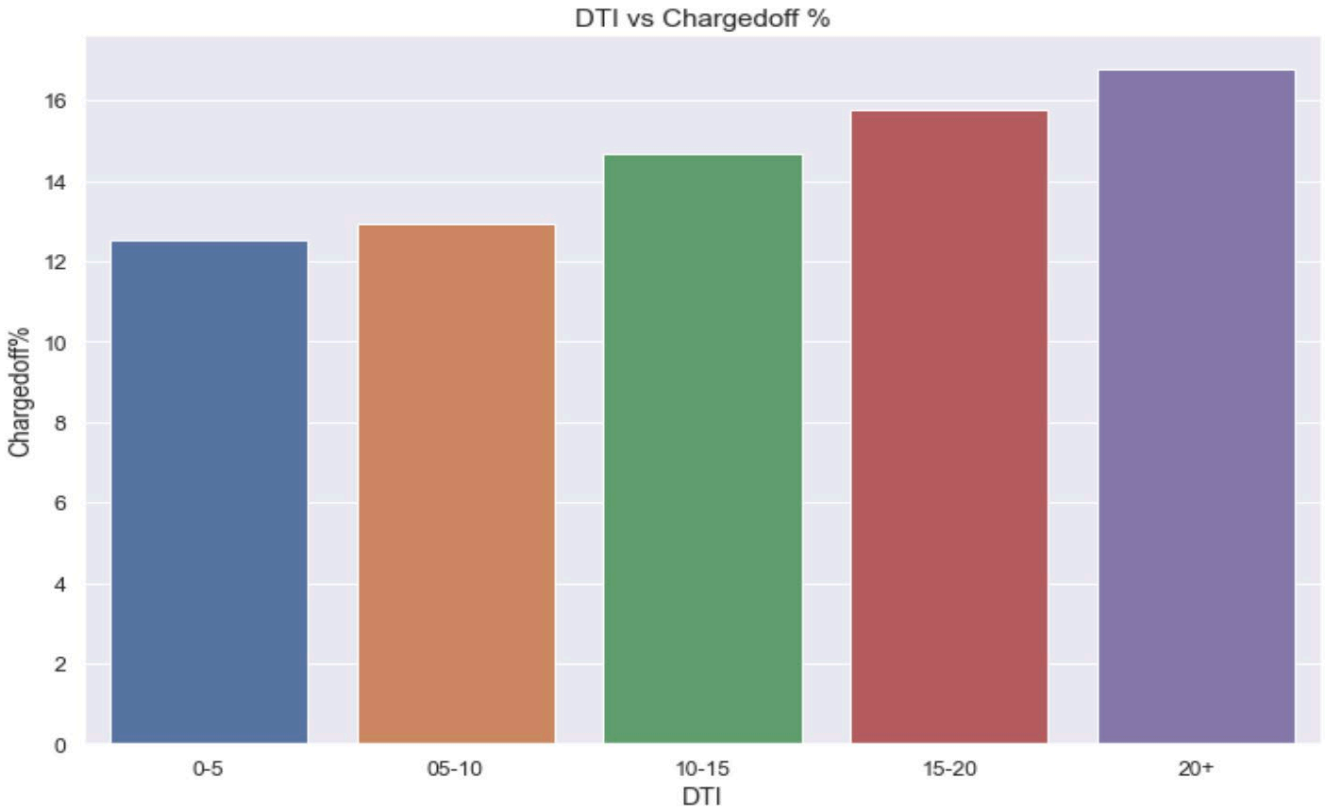
Capping loan amount based on income reduces risk.



Inference

Charged Off consistently increases with **DTI** ie. existing debt to income percentage.

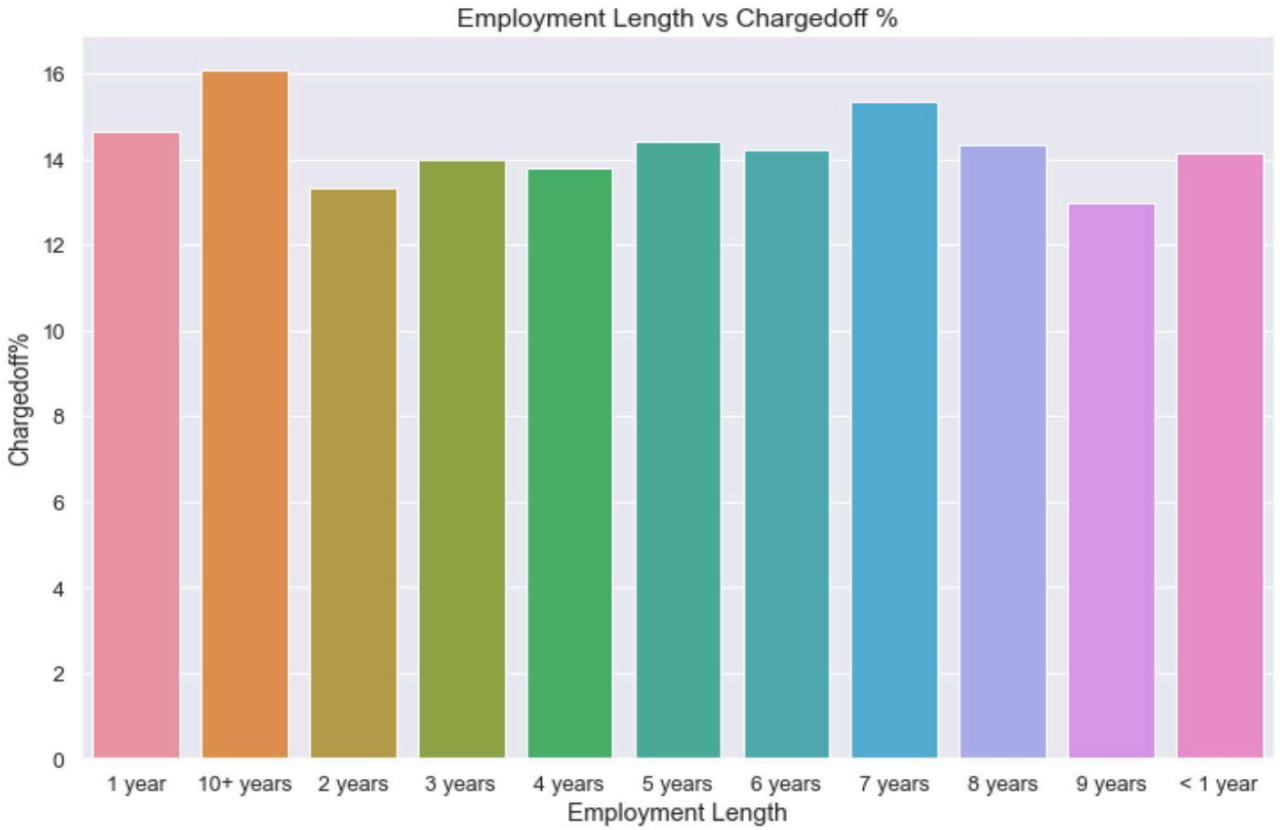
Higher the DTI higher the chances of Charged Off



Inference

People with work experience of 10+ year and 7+ year have high charged off %, also people with these work experience take maximum number of loans.

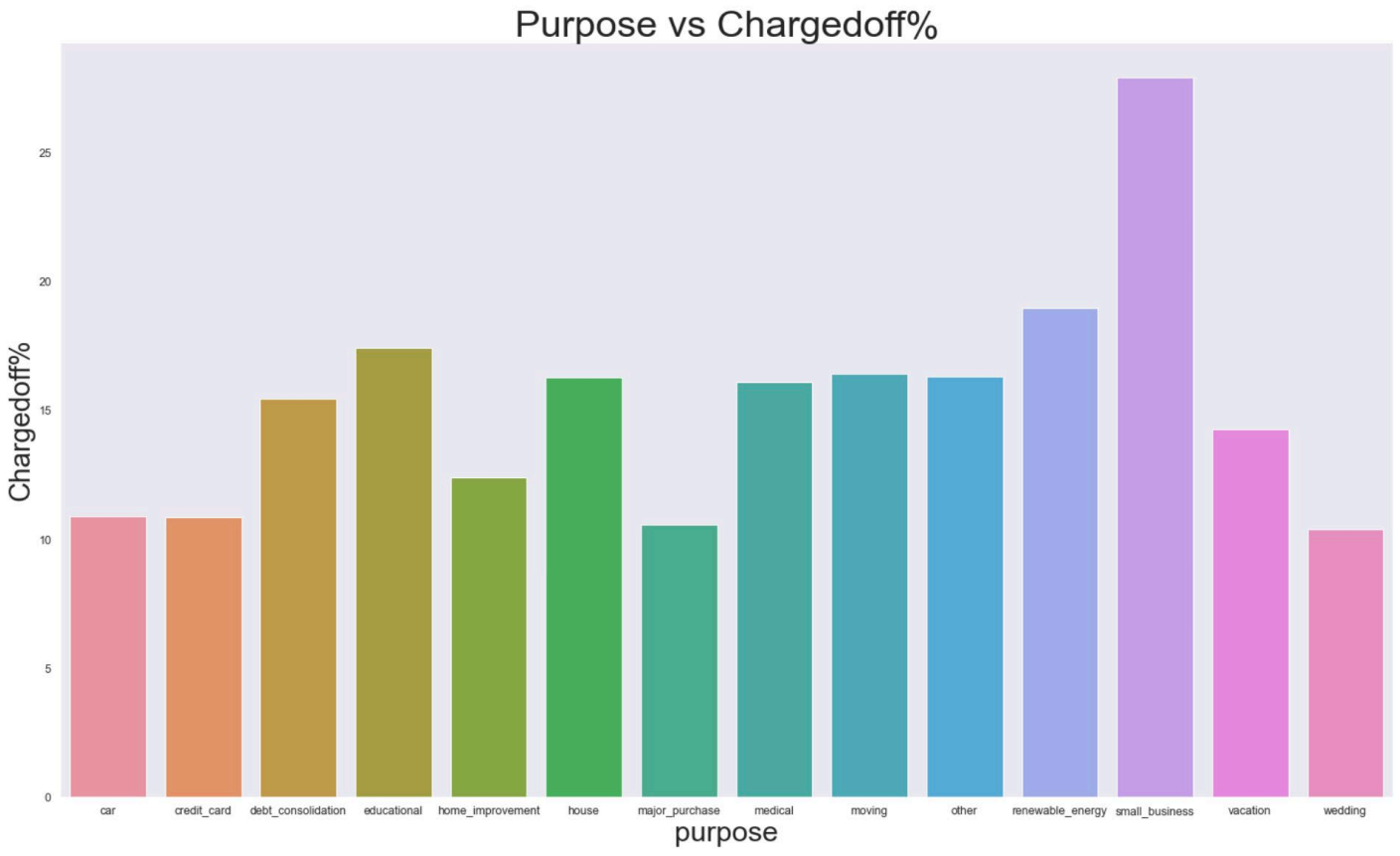
Other than that 1 year or below combined forms the third major charged off category.



Inference

The **loan purpose** with major charged off is taken by **small business** category in proportion.

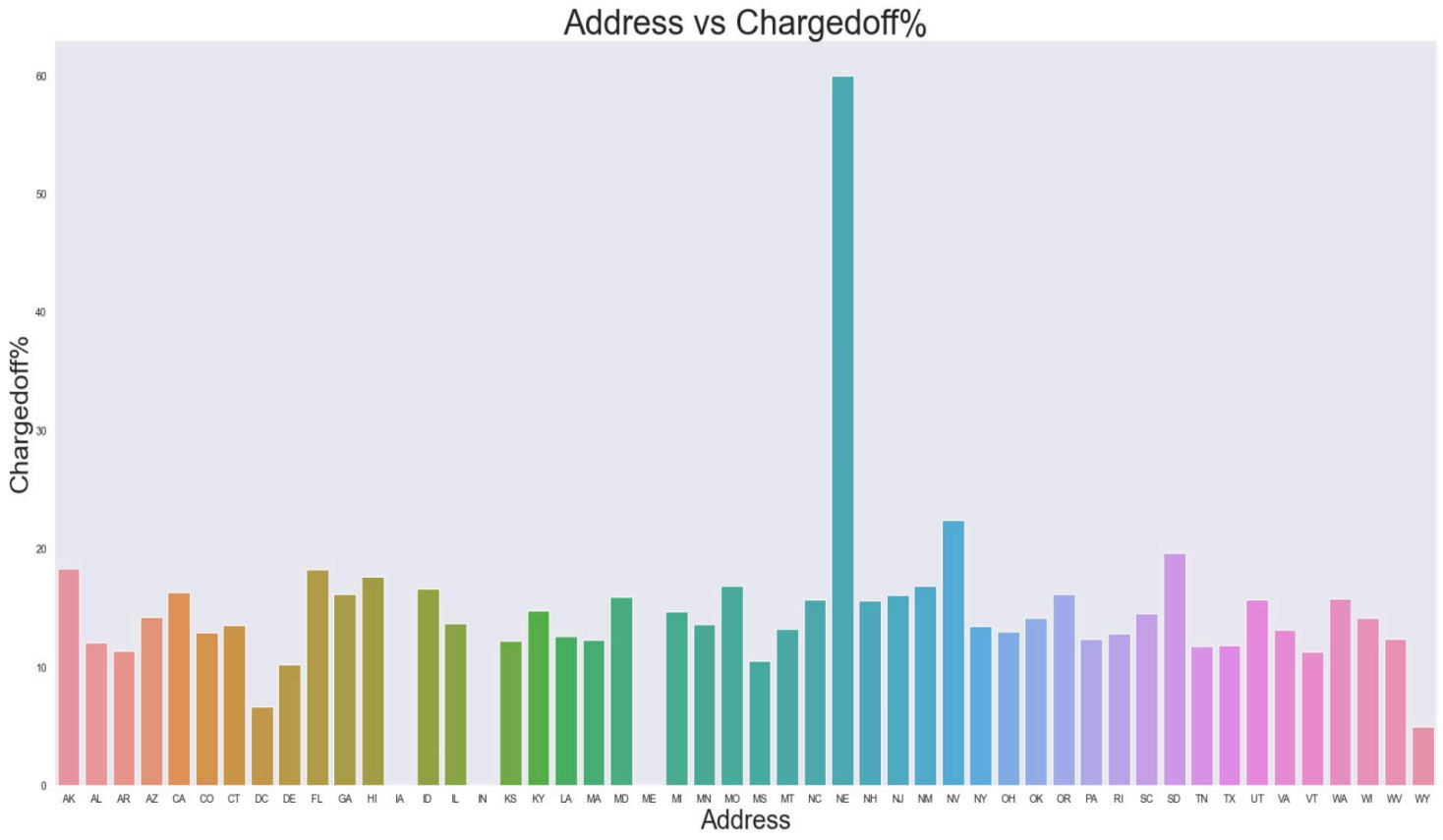
Loans to Small Business have higher chances of charged off.



Inference

State/city **NE** has the highest proportion of charged off, near to 60%.

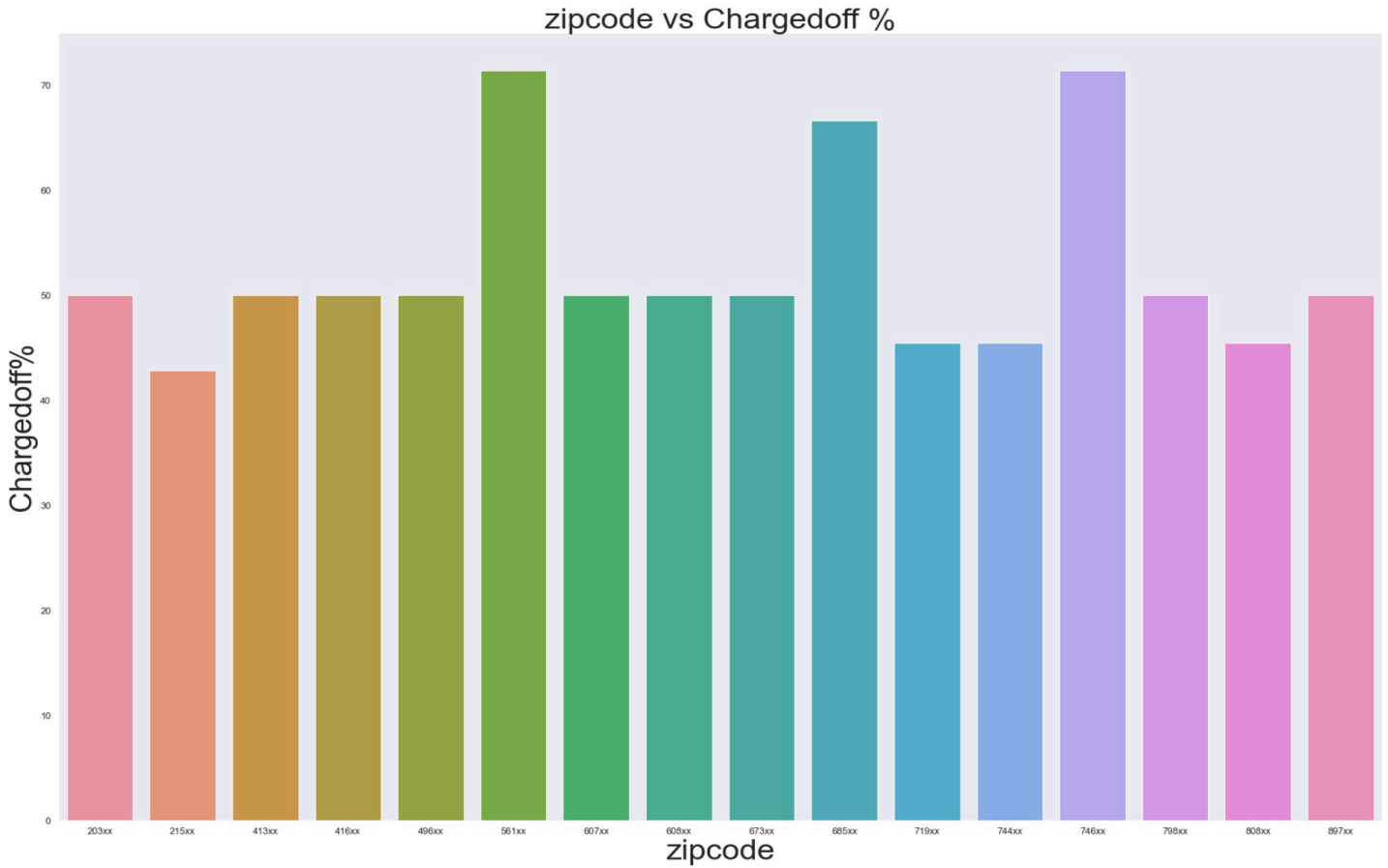
Loan issued in NE has very high chance of charge off.



Inference

Apart from the state/city, areas with zip codes 561xx, 685xx and 746xx have more than 60% occurrences of charge offs.

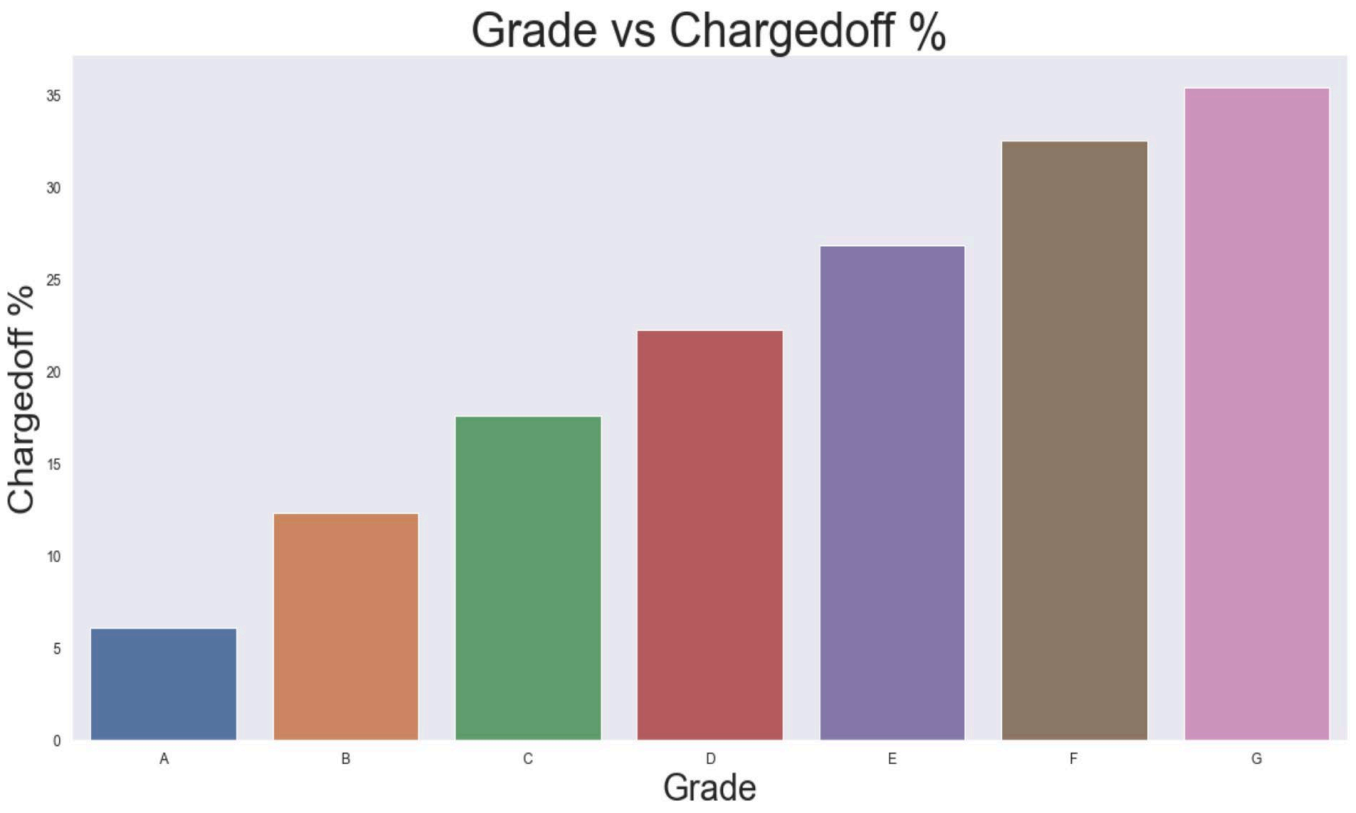
Its very high risk to issue loan in above zip code areas.



Inference

Charged off % consistently increase from grade A to G.

The risk of charged off is much higher in later grades. Grades depend on Loan Amount, Interest rate, revolving credit line etc.

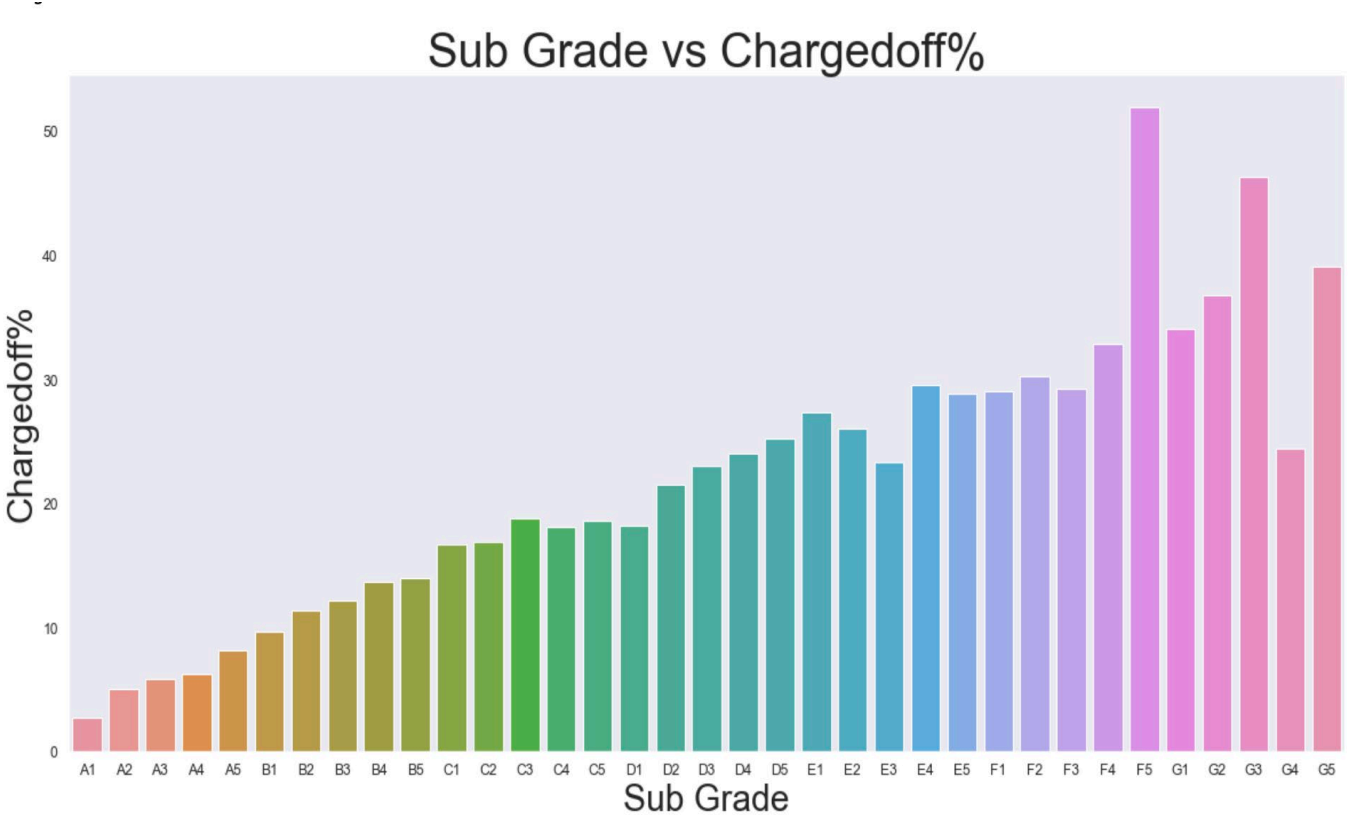


Inference

1. There is almost consistent increase in charge off from sub grade A1 to F4.

2. The categories of F5 and G3 have much higher chances of charged off.

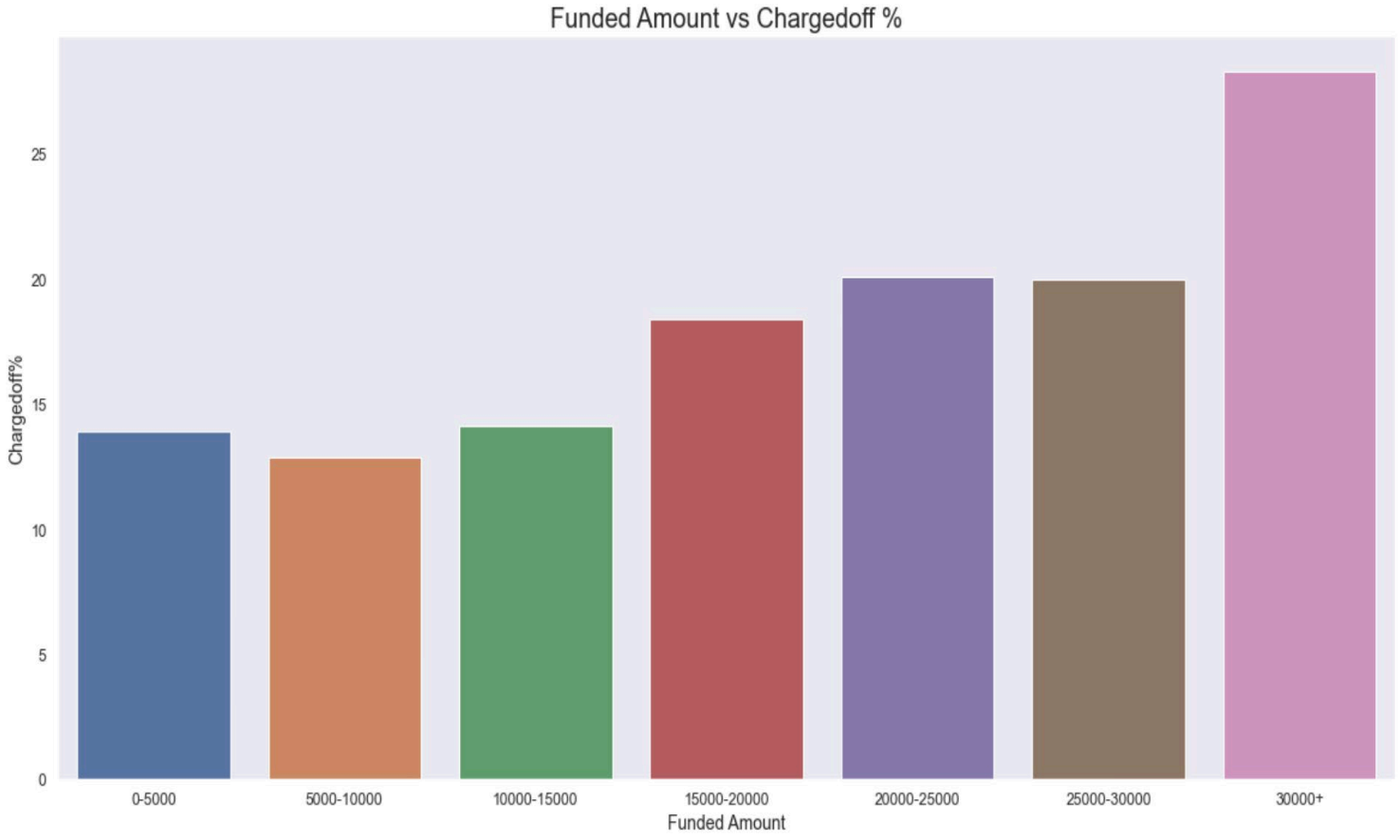
3. Irrespective of usual trend G4 has lower chances of charged off compared to its peers.



Inference

Charged off proportion is highest in funded amount of 30k+ , which exposes firm to higher risk of loss.

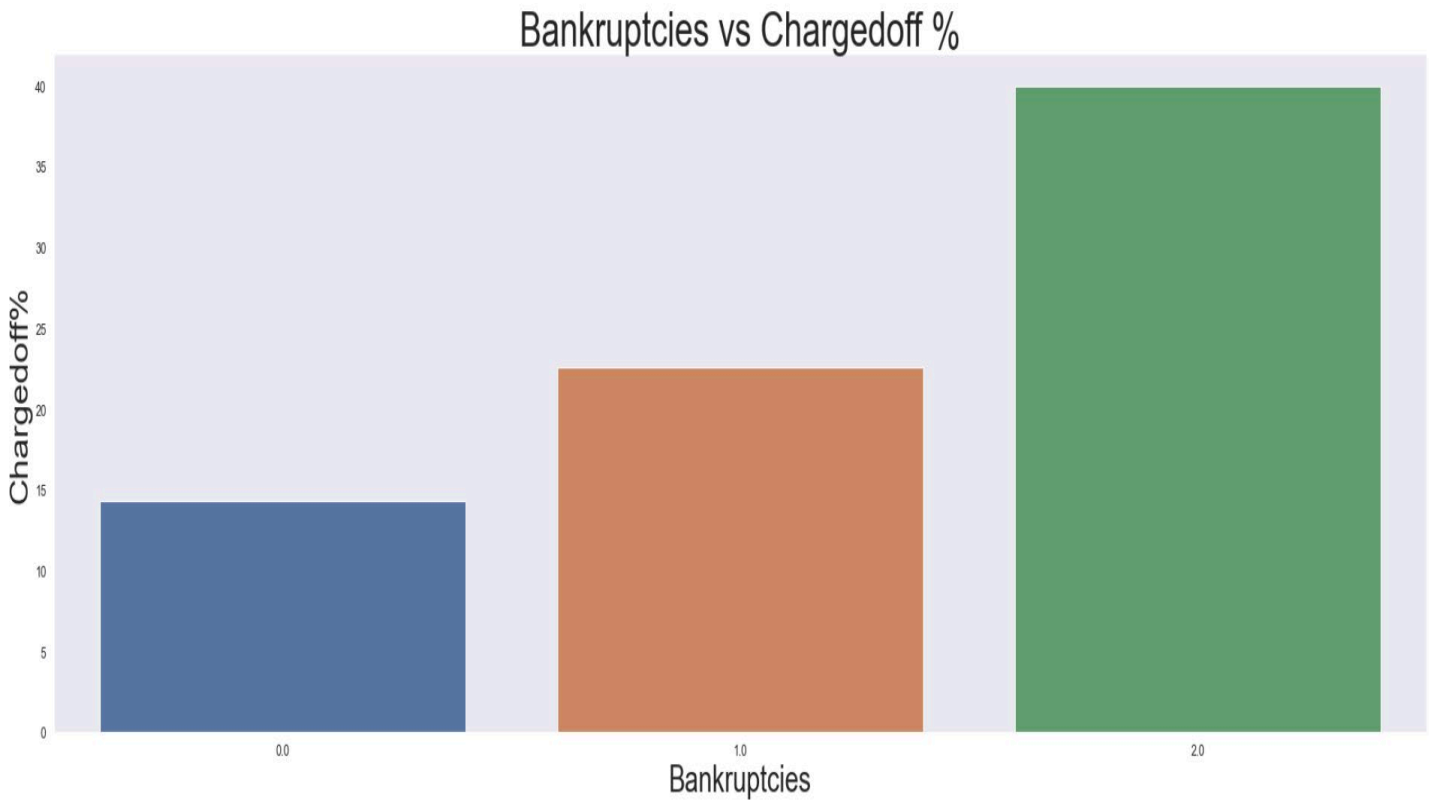
Firm needs to ensure the factors of DTI, income, bankruptcies in account before issuing loan in this category.



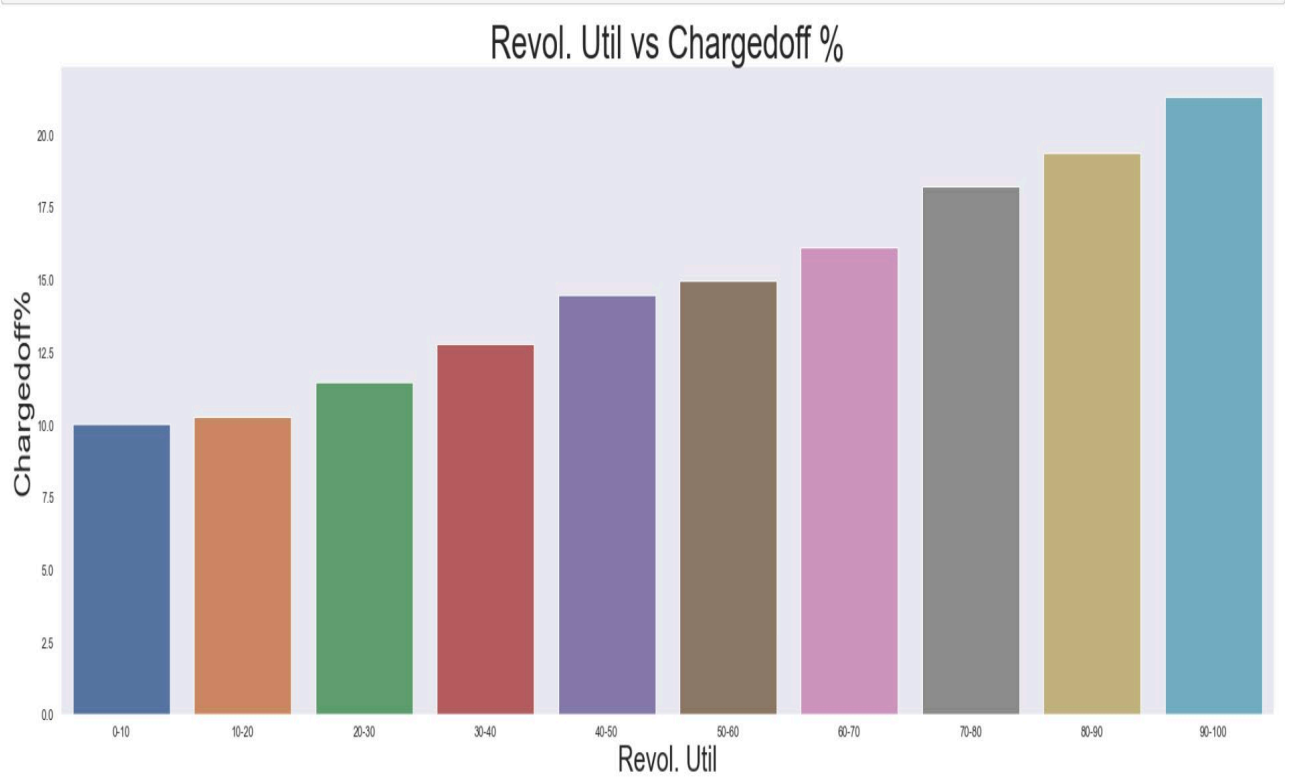
Inference

More the bankruptcies higher the chances of charged off.

Individuals who had 2 bankruptcies have high chances of defaulting the loan.



Inference

Charge Off consistently increase for increasing revolving line utilization rate.

Higher the revolving line utilization rate higher the chance charged off