LENDING CLUB CASE STUDY

PAVAN NARRA

JHESER GUZMAN

Problem Statement

- A consumer finance company which specialize in lending various types of loans to urban customers, receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

- Two types of risks are associated with the bank's decision:

- 1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.

- 2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Business Objective

▶ This company 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.

▶ Lending loans to 'Risky' applicants is the largest source of financial loss (called credit loss). Borrowers who default cause the largest amount of loss to the lenders and such Customers are labelled as 'charged-off' and are the 'defaulters'.

▶ The Business Objective of this Study is to identify Driving Factors (or driver variables) which are strong indicators of default and potentially use the insights in Approval / Rejection decision making

▶ The company can utilize this knowledge for its Portfolio and Risk Assessment and can reduce the Risky Loan Approvals and hence Cutting Down the Amount of its Credit Loss.

Exploratory Data Analysis (EDA)

▶ We have been provided with one Excel File contains the complete loan data for all loans issued through the time period 2007 t0 2011 and the Data Dictionary of the Data Fields, which describes the meaning of each field in the Loans Data File.

▶ The data given contains the information about past loan applicants and whether they 'defaulted' or not.

▶ The aim is to identify patterns which indicate if a person is likely to default, which can be used by the Lending Club for taking actions such as listed below or more :

▪ Denying the loan

▪ Reducing the amount of loan

▪ Lending (to risky applicants) at a higher interest rate

Understanding the Data Set

- The given data set contains 3 types of Data

   1. Customer (Applicant) Demographic Data (
Applicant's   Specific/Personal Information)

   2. Loan Specific Data (Loan Information)

   3. Customer Behavioural Data (once the loan is granted)

Types Of Data

Customer Demographic Data
Employment Title
Employment Length
Annual Income
Description
Zip code
State Address
Home Ownership

Loan Information Data
Loan Amount
Funded Loan Amount
Loan Interest Rate
Issue Date
Loan Status
Grade
Term

Customer Behavioral Data
No of Delinquent Accounts
Delinquency 2yrs
Earliest Credit Line
Installment
Revolving Balance
Recoveries
Application Type

Data Cleaning

- Checked the percentage of missing values of each column

- Removed the columns which have all null values

- Identified the Outliers in the annual income column and removed them

- Casted few columns to numeric data type

- Some Columns are derived from the existing Columns for doing the Analysis

Data Analysis Features and Label Selction

- The Label we need to predict is identified as 'Loan Status'
- The Features are identified and remaining columns were dropped for the anlyses purpose
- The features are selected from the loan information data
- The features has continuous variables and categorical variables
- The categorical variables are identified as e term, emp length, grade, sub grade, home ownership, and verification status
- The continuous like loan amount , funded amount , inter_rate etc are binned and create new Categorical Variables

Univariate Analysis

▶ Loan status
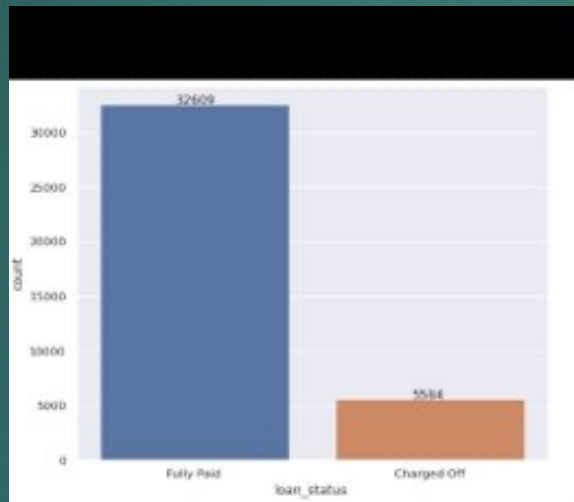
Most of the loans are Fully Paid. About 15% of loan are having status as defaulters. There is high imbalance of Classes here

Term
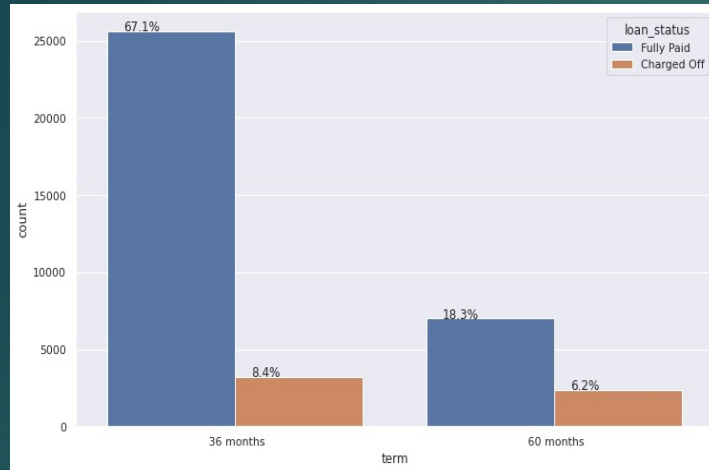
Grade

36 Month term is the most preferd loan but the number of charged off loans are more in 60 months period when compared to the total borrowers in that priod
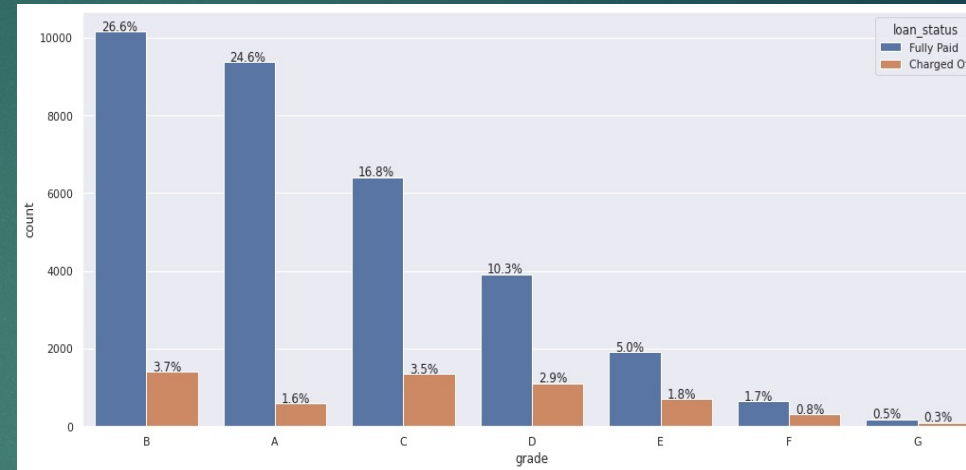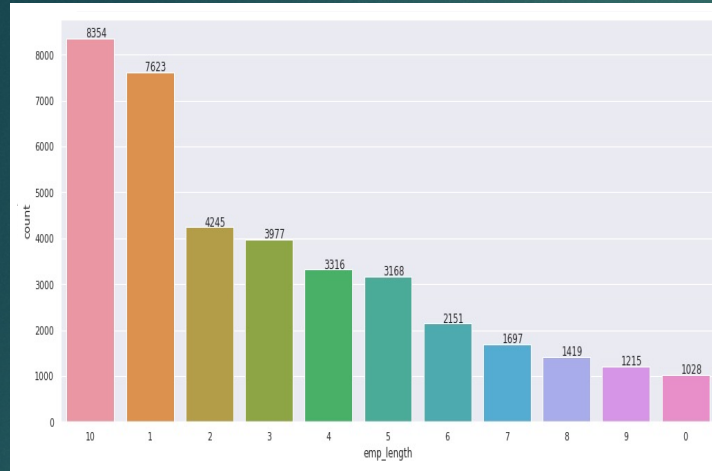
Most of the loans are categorized into groupB followed by group A ,C,D,E,F,G.G grade loans has the highest charged off loans .Where as A grade has the lowest charged off followed by B grade Loan

Employement Years

Employement years vs loan status

10 years and above employemnt years are the highest loan takers then followed by the one year experience
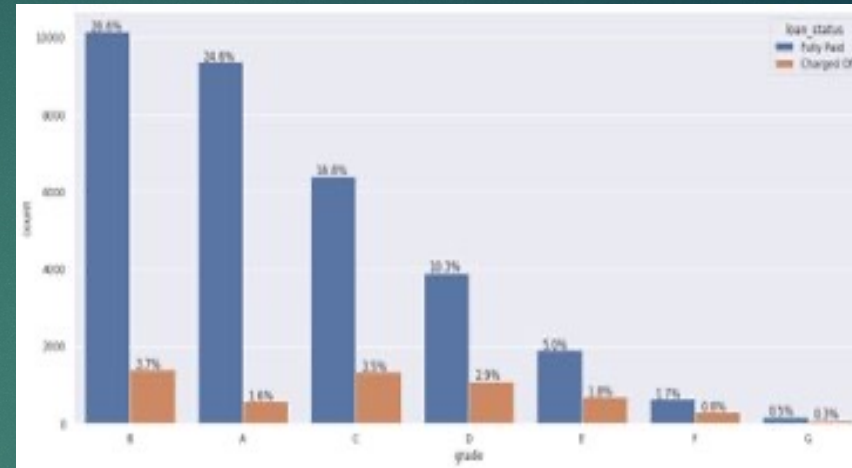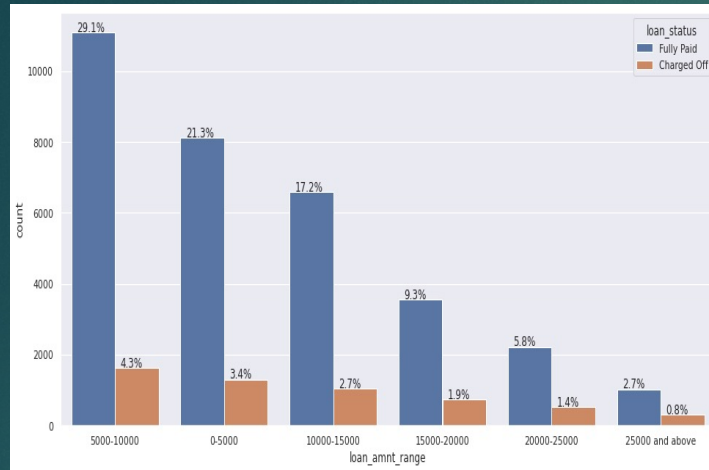
0-1 year has highest charged off loans percentage comapred to the total borrowers in that group followed by 10 years

Loan amount range vs Loan status
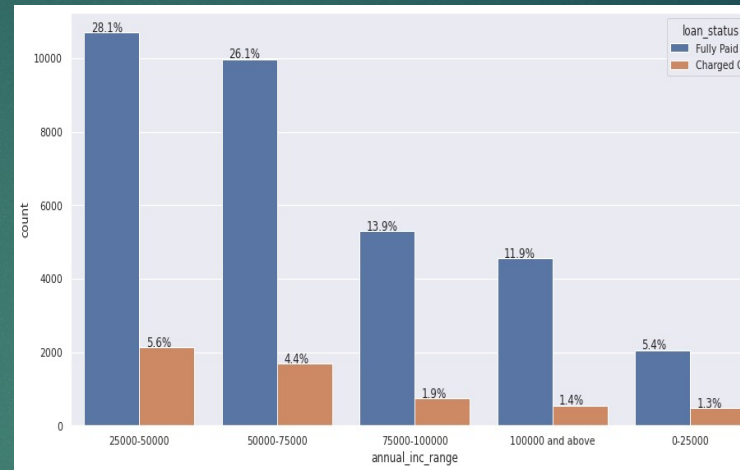
Annual income vs loan status

Between 5000-10000 is the most prefered loan amount than followed by 0-5000,10000-15000

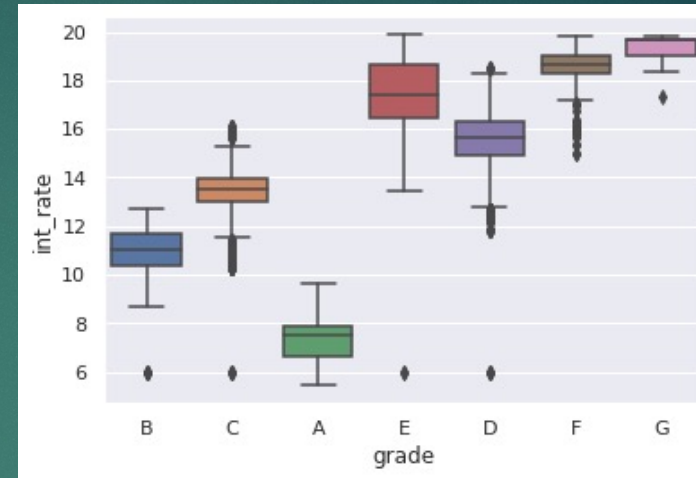The 25000-5000 is the highest annual income followed by the 50000-75000

Rate vs Loan Status
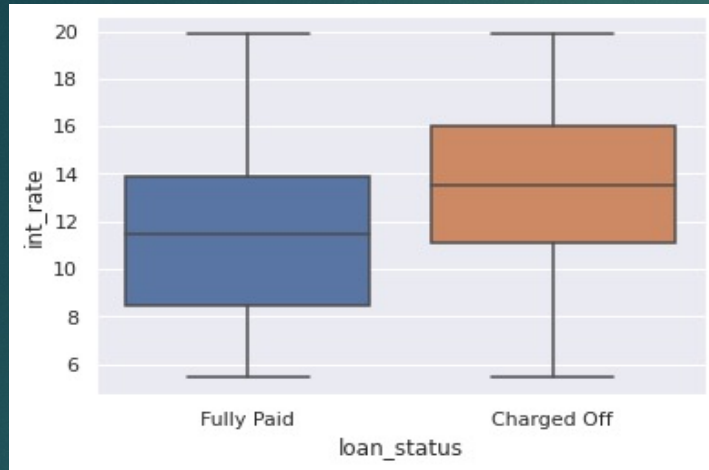
Grade vs int rate

The median value of interest rates for charged of loans is higher than the paid off loans

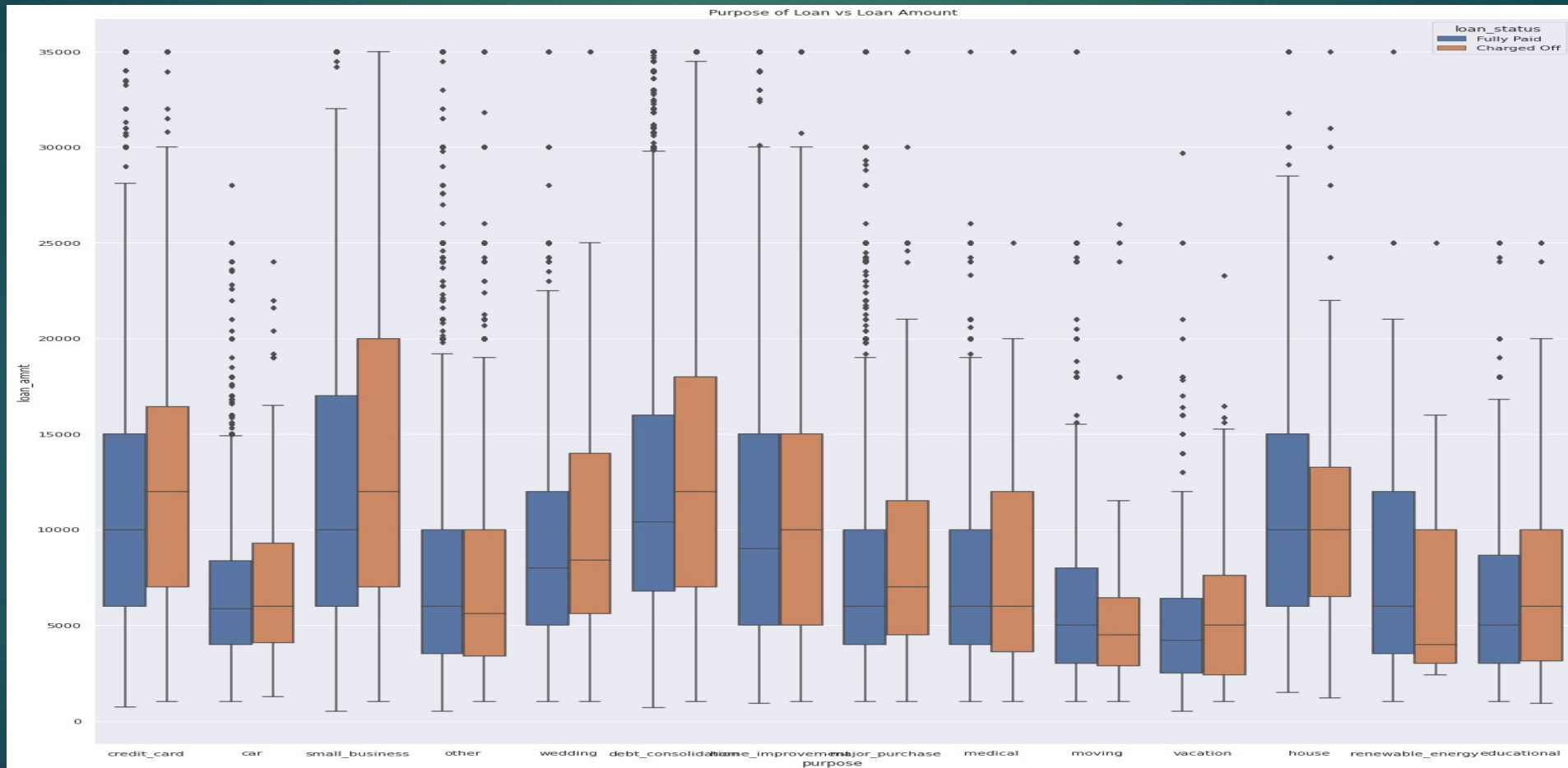Grade A has the lowest interest rate followed by the Grade B

Bivariate Analysis
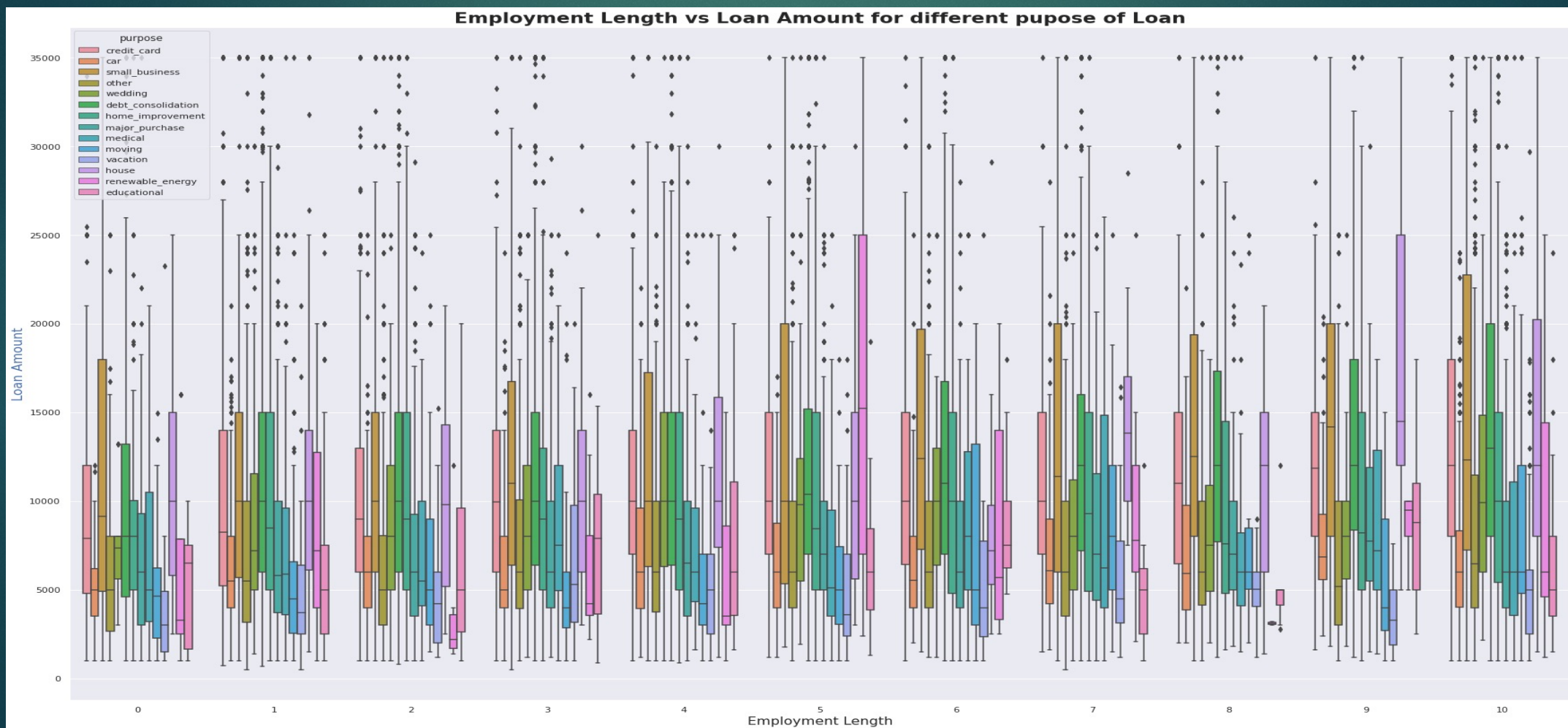# Purpose of Loan vs Loan Amount

Employment Length vs Loan Amount for different pupose of Loan

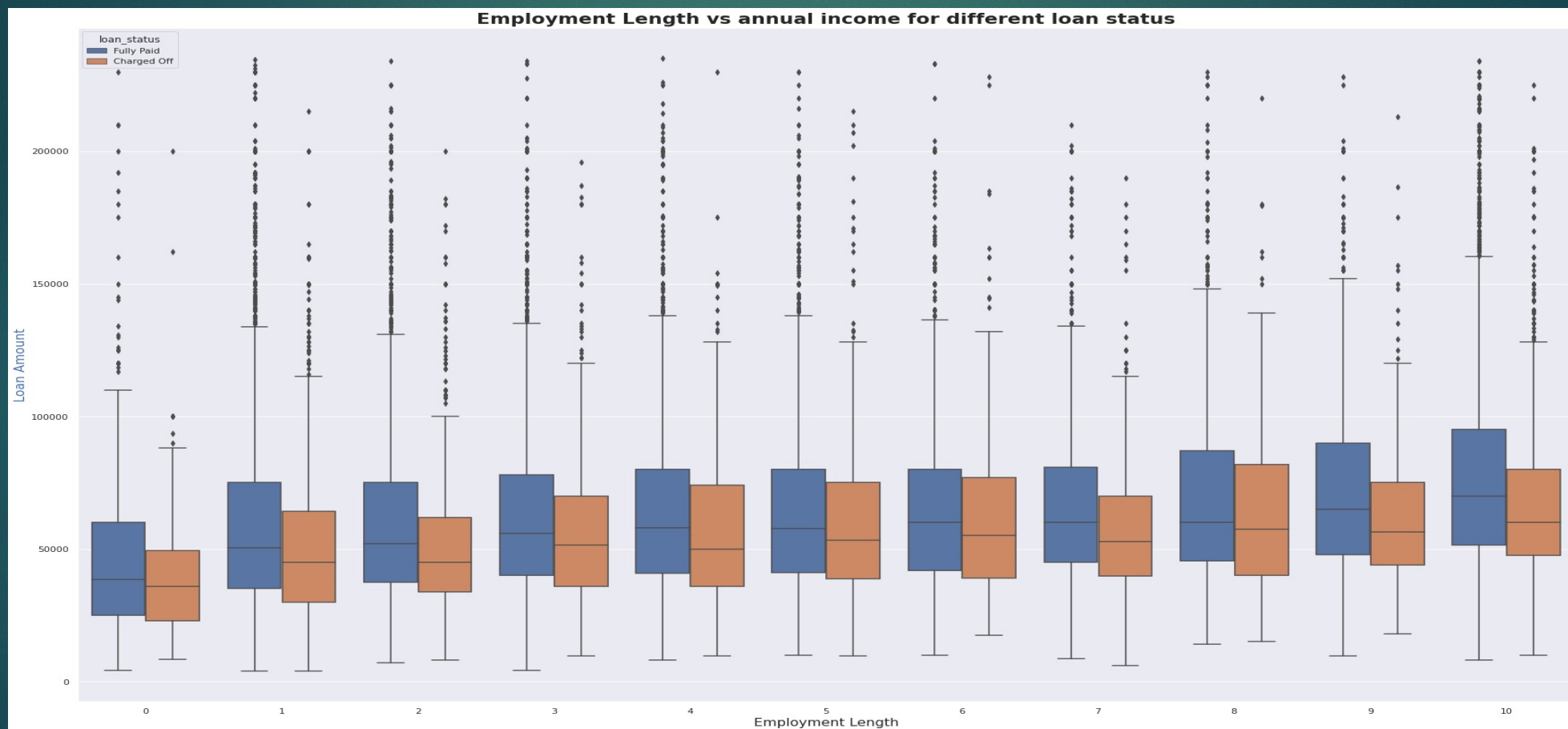# Employment Length vs annual income for different loan status

Employment Length vs annual income for different loan status

Observations

- Grade A loan has the lowest charged off loans because of lowest interest rates .As the interest rates increases the charged off loans are increasing

- The employees with 0-1 year and 10 –10+years,followed by 7,8 years experience have more charged off loans and the common pattern bewteen them is higher mean value of loan on smaller business

- The Median of the annual income is increasing as the experince increase ,the mean annual income of employes who defaults loan is less when compared to the employess who fully paid the loan

- The small bussiness,credit card,debit consolidation has high mean of loan amount which indicates investors should be extra couscious when giving the loan for those purposes