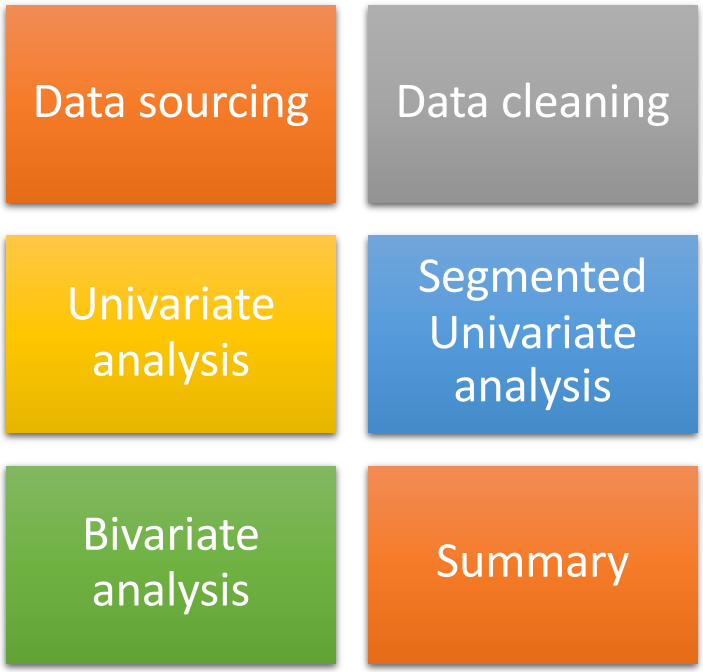
Lending Club Case Study

Objective

- A consumer finance company has to make a decision for loan approval based on the applicant's profile.
- Given the past loan applicants and whether they 'defaulted' or not, using EDA, identify the consumer / loan attributes influence the tendency of default.

Exploratory Data Analysis (EDA)



Data Sourcing

Data set downloaded from Upgrad's website.

Meaning of the variables in data set can be found here

Data Cleaning – Steps followed

- Deleted unnecessary / irrelevant columns for analysis
- Dropped columns which has significant no. of missing values (> 90%)
- Dropped rows if missing values in a column are <= 10%
- Verified no missing values in any columns of data frame
- Standardized values wherever applicable -> converting object to int / float
- Renaming columns for better understanding
- Created derived columns – type driven, and business driven (columns are highlighted in next slide)
- Filtering data that ‘s not required for analysis

Data Cleaning - Results

```
loan_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 35376 entries, 1 to 39680
Data columns (total 49 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   id                          35376 non-null  int64
 1   member_id                   35376 non-null  int64
 2   loan_amnt                   35376 non-null  int64
 3   funded_amnt                 35376 non-null  int64
 4   funded_amnt_inv             35376 non-null  float64
 5   term_in_months              35376 non-null  int64
 6   int_rate_percent            35376 non-null  float64
 7   installment                 35376 non-null  float64
 8   grade                       35376 non-null  object
 9   sub_grade                   35376 non-null  object
 10  emp_title                   35376 non-null  object
 11  emp_length                  35376 non-null  int64
 12  home_ownership              35376 non-null  object
 13  annual_inc                  35376 non-null  float64
 14  verification_status         35376 non-null  object
 15  issue_d                     35376 non-null  object
 16  loan_status                 35376 non-null  object
 17  purpose                     35376 non-null  object
 18  zip_code                    35376 non-null  object
 19  addr_state                  35376 non-null  object
 20  dti                         35376 non-null  float64
 21  delinq_2yrs                 35376 non-null  int64
 22  earliest_cr_line            35376 non-null  object
 23  inq_last_6mths              35376 non-null  int64
 24  open_acc                    35376 non-null  int64
 25  pub_rec                     35376 non-null  int64
 26  revol_bal                   35376 non-null  int64
 27  revol_util                  35376 non-null  object
 28  total_acc                   35376 non-null  int64
 29  out_prncp                   35376 non-null  float64
 30  out_prncp_inv               35376 non-null  float64
 31  total_pymnt                 35376 non-null  float64
 32  total_pymnt_inv             35376 non-null  float64
 33  total_rec_prncp             35376 non-null  float64
 34  total_rec_int               35376 non-null  float64
 35  total_rec_late_fee          35376 non-null  float64
 36  recoveries                  35376 non-null  float64
 37  collection_recovery_fee     35376 non-null  float64
 38  last_pymnt_d                35376 non-null  object
 39  last_pymnt_amnt             35376 non-null  float64
 40  last_credit_pull_d          35376 non-null  object
 41  pub_rec_bankruptcies        35376 non-null  float64
 42  issue_d_month               35376 non-null  object
 43  issue_d_year                35376 non-null  int64
 44  last_pymnt_d_month          35376 non-null  object
 45  last_pymnt_d_year           35376 non-null  int64
 46  last_credit_pull_d_month    35376 non-null  object
 47  last_credit_pull_d_year     35376 non-null  int64
 48  is_profit                   35376 non-null  bool
dtypes: bool(1), float64(16), int64(15), object(17)
memory usage: 13.3+ MB
```

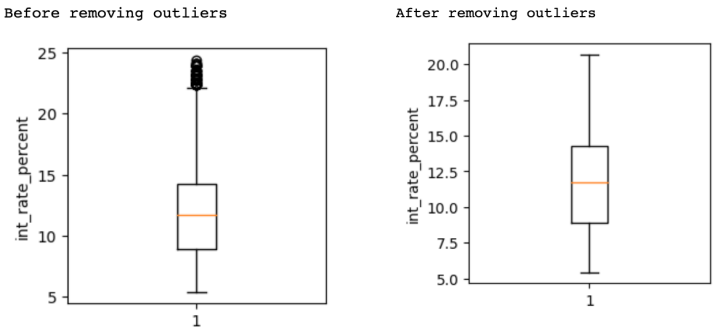
```
loan_data.shape
```

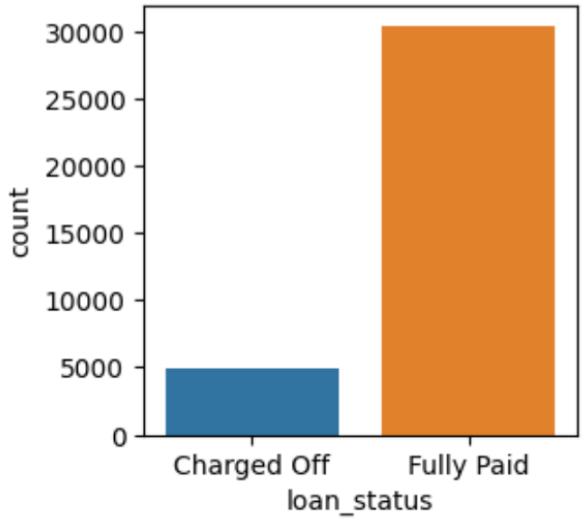
```
(35376, 49)
```

Univariate Analysis – Loan Status & Interest rate

- Appx. 14% of loan records are defaulters and others are fully paid.
- This tells us that there's a skew in dataset

- Outliers were there in interest rate variable – before & after (removed outliers) are displayed using box plot

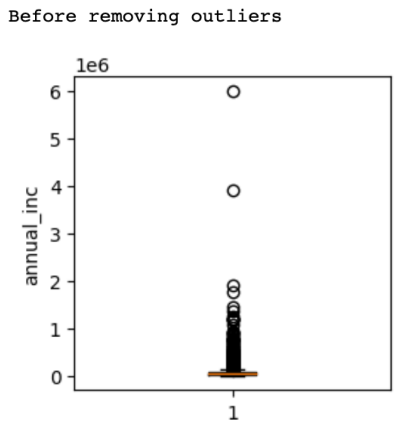


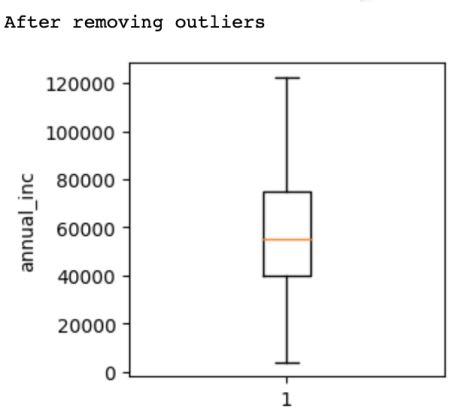


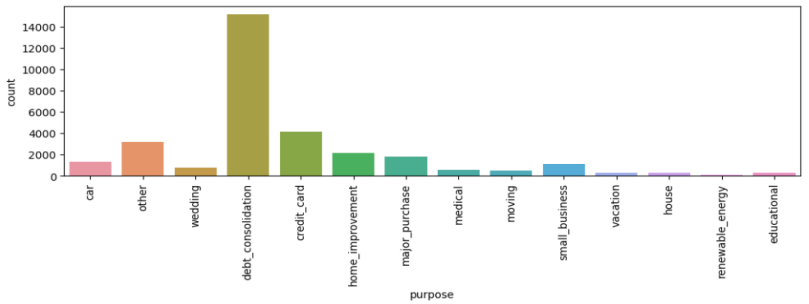
Univariate Analysis – Annual Income & Loan Purpose

- Outliers were there in annual income variable – before & after change are displayed using box plot

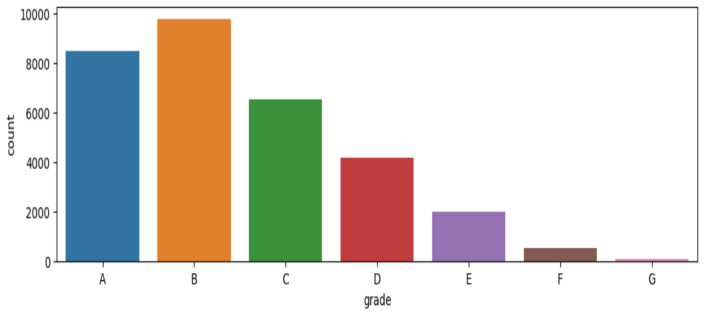
- Most of the loans are given for debt consolidation purpose



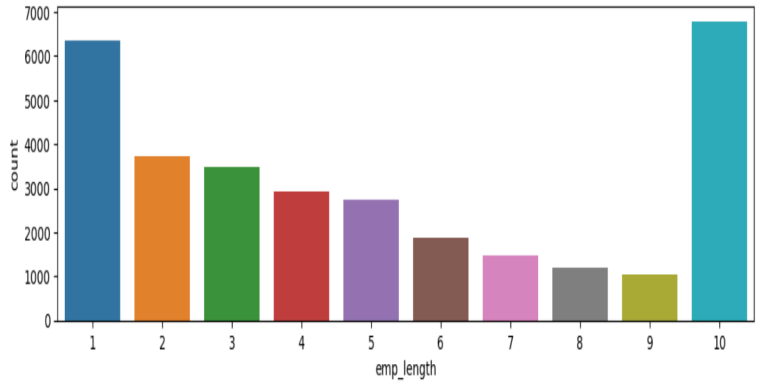




Univariate Analysis – Grade & Employee Length



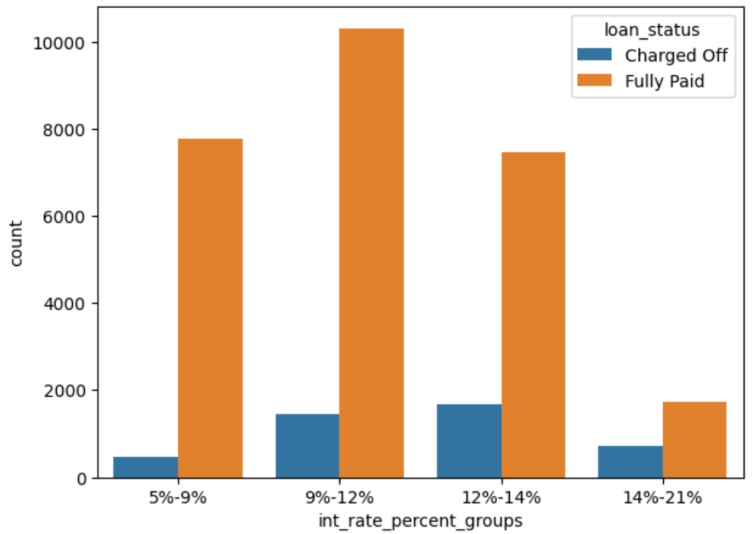
- Most of the loans have grade B followed by grade A and C



- Most of the loans are borrowed by applicants having 10+ years of emp experience followed by people having less than one-year emp experience

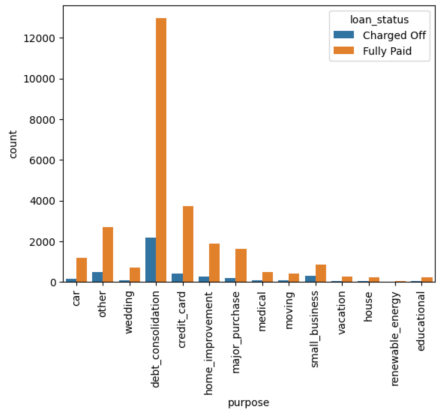
Segmented univariate analysis – Interest Rate

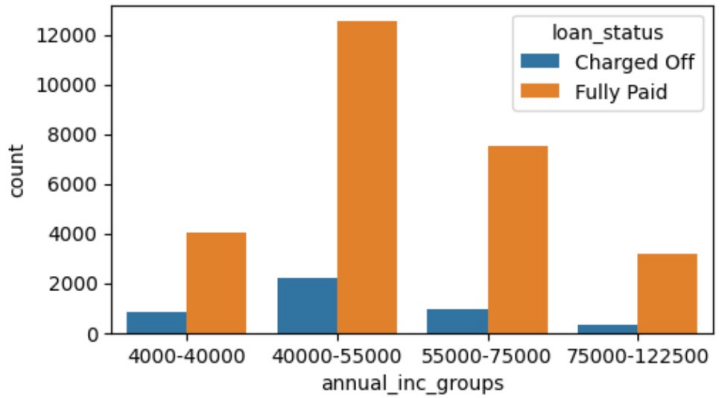
- Defaulters rate is high for higher interest rates. More specifically, interest rates between 12% and 14% have a significant defaulter's rate



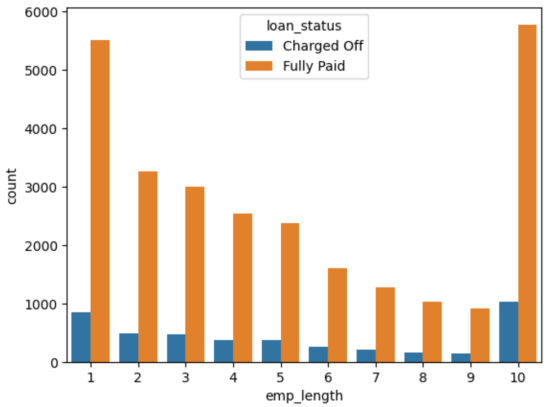
Segmented univariate analysis – Loan Purpose and Annual Income

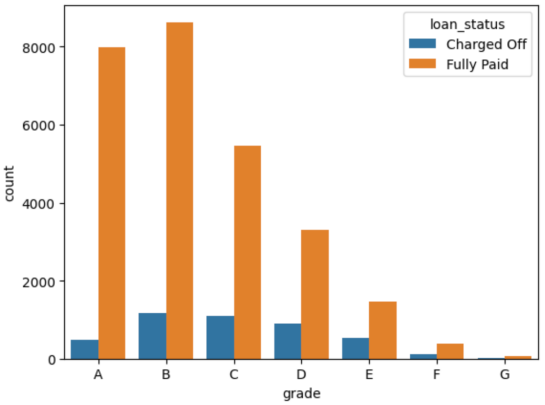
- Most no. of defaulters got loan for the purpose of debt consolidation.
- Loan applicants with less annual income have higher number of defaulters. More specifically, annual income between 40K-55K are more likely to default.





Segmented Univariate Analysis – Grade & Employee Length

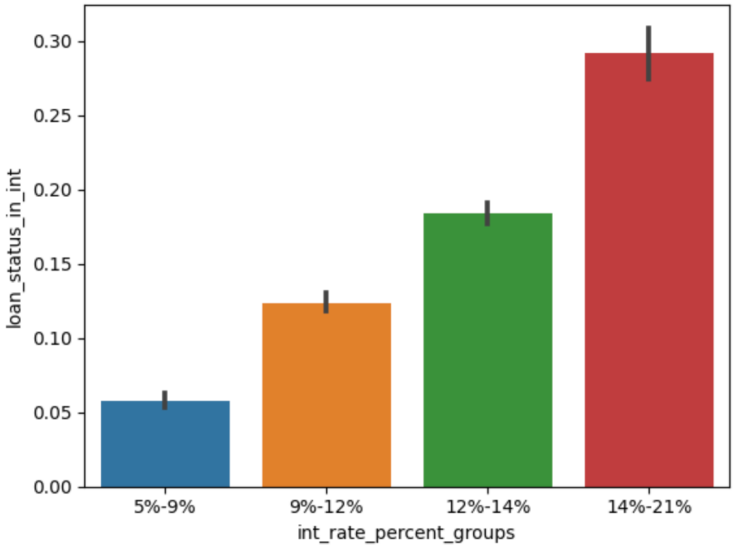




- Loans with grade B and C has more defaulters than other grades
- Most no. of defaulters have 10+years of experience followed by less than one year of experience

Bivariate Analysis – Interest Rate

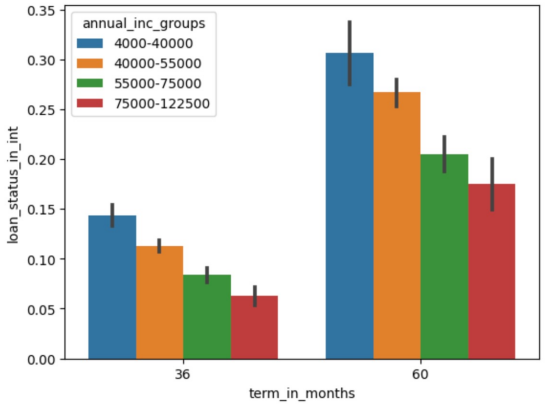
- Loan with higher interest rates (14-21%) tend to have more defaulters

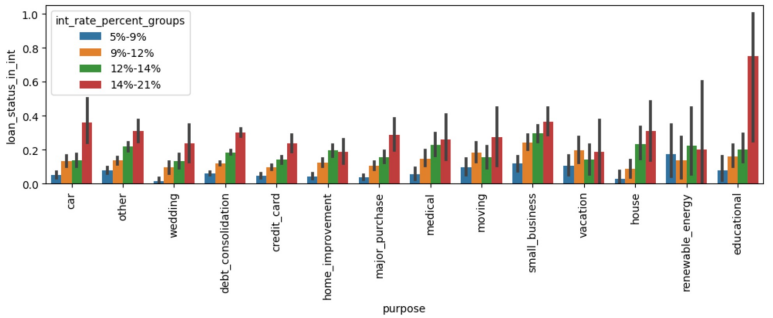


Bivariate Analysis – Loan Purpose & Annual Income

- Loan applied for educational purpose with higher interest rate (14 - 21%) tend to have more defaulters

- Loan with 60 months term tend to have more defaulters particularly people having annual income in the range of 4k to 40K

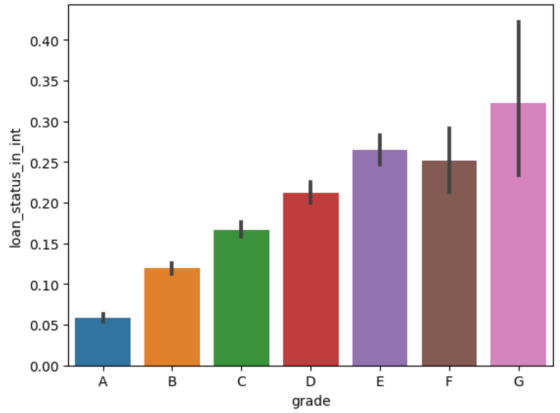


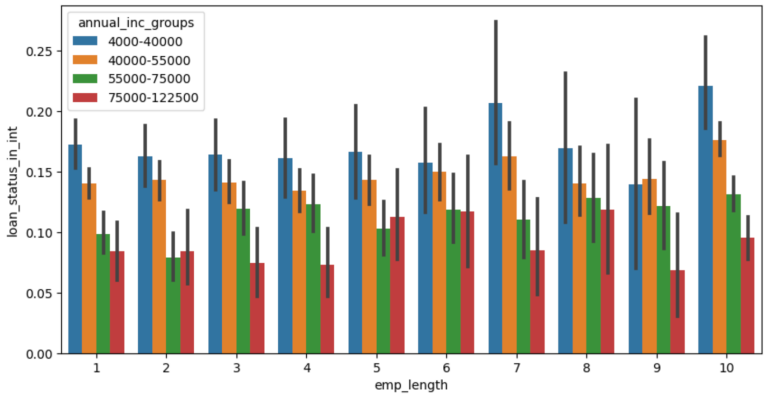


Bivariate Analysis – Grade & Employee Length

- Loan applicants with low grades have higher chances of becoming defaulters

- Loan applied by people with 10+ years of experience and having annual income in the range of 4k-40k have high chances of becoming defaulters





Summary

Following are some of the driving factors / variables behind loan default:

- grade - Loan applicants with low grades have higher chances of becoming defaulters
- int_rate - Loan applied for educational purpose with higher interest rate (14 - 21%) tend to have more defaulters
- purpose - Loan applied for 'moving' purpose tend to have more defaulters
- annual_inc - Loan with 60 months term tend to have more defaulters particularly people having annual income in the range of 4000 to 40K
- emp_length - Loan applied by people with 10+ years of experience and having annual income in the range of 4k-40k have high chances of becoming defaulters