Upgrad Executive PG Program in ML and AI

ML Lending Club Case Study
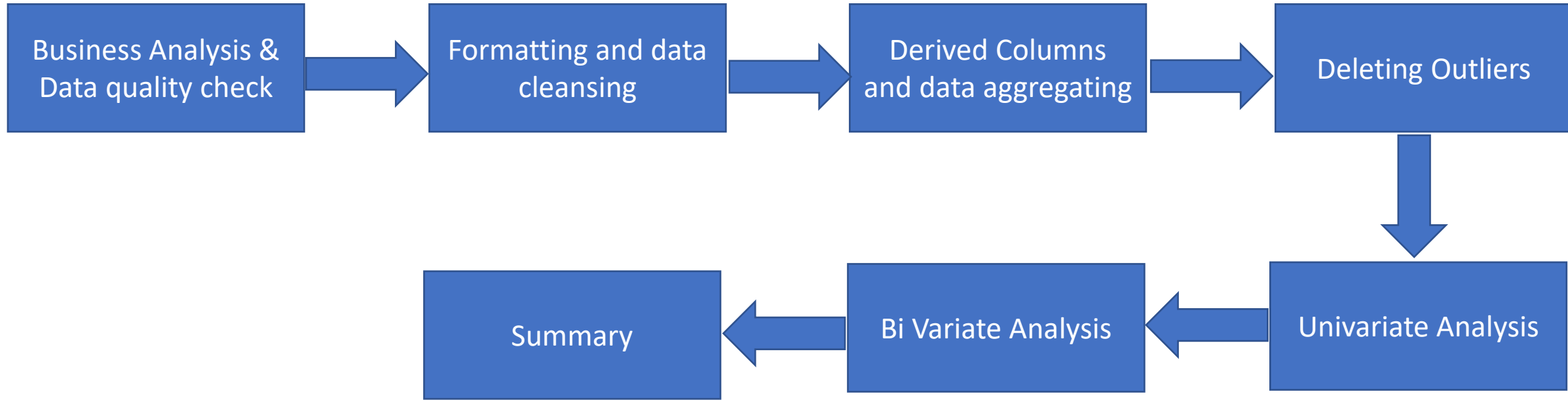
Group: Biswa P Samanta

Table of Contents:

- Objective
- Approach
- Data analysis
- Data Cleansing
- Data Analysis
- Summary
- Recommendations

Objective:

- Lending Club is a marketplace for personal loans. It matches the borrowers who are seeking a - loan with investors looking to lend money for a profitable return.

- There is chance of risk while funding loan to the borrowers which can lead to financial loss to the Lending Club.

- The main objective of this Case Study is to analyze the driving factors or variables leading to default risk and provide valuable recommendations to prevent the same.

Approach:

Business Analysis & Data quality check → Formatting and data cleansing → Derived Columns and data aggregating → Deleting Outliers → Univariate Analysis → Bi Variate Analysis → Summary

Data Analysis:

- There are three types of applicants
  - Fully Paid
  - Current
  - Charged off
- Current status is ignored as the applicants are in process of paying back the loan
- Charged off – The charged off applicants are risky as they have not paid off the

Data Cleansing:

- Identifying missing values in columns
- Analyze min and max annual incomes.
- Removal of columns with missing and redundant values
- Removing outliers
- Replacing the nulls with appropriate values
- Deriving columns for grouping data
- Data cleansing or massaging the data (removing "%" sign)

Univariate Analysis:

```
# graph
plt.figure(figsize=[30, 10])
loan_purpose = ['debt_consolidation','credit_card', 'other', 'home_improvement', 'major_purchase', 'small_business', 'car',
plt.bar(loan_purpose, percentage_loan_purpose, width= 0.9, align='center')
plt.title("Purpose of Loan")
plt.xlabel('Purpose of loan')
plt.ylabel('Percentage')
plt.show()
```
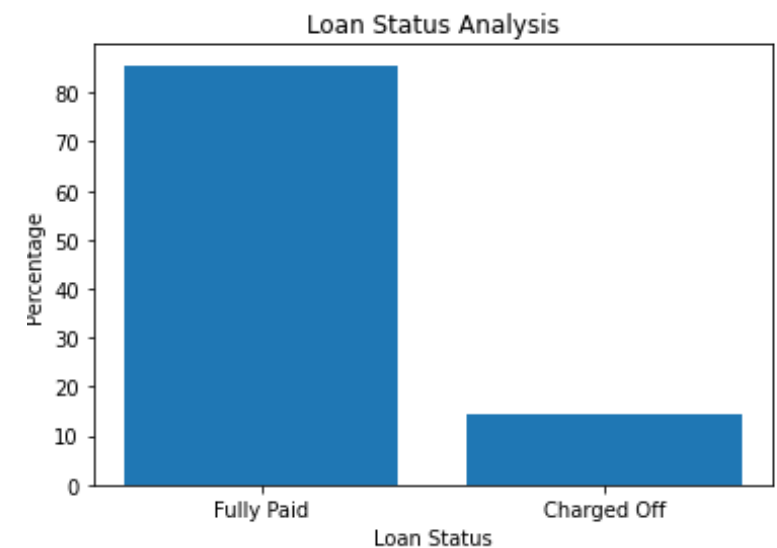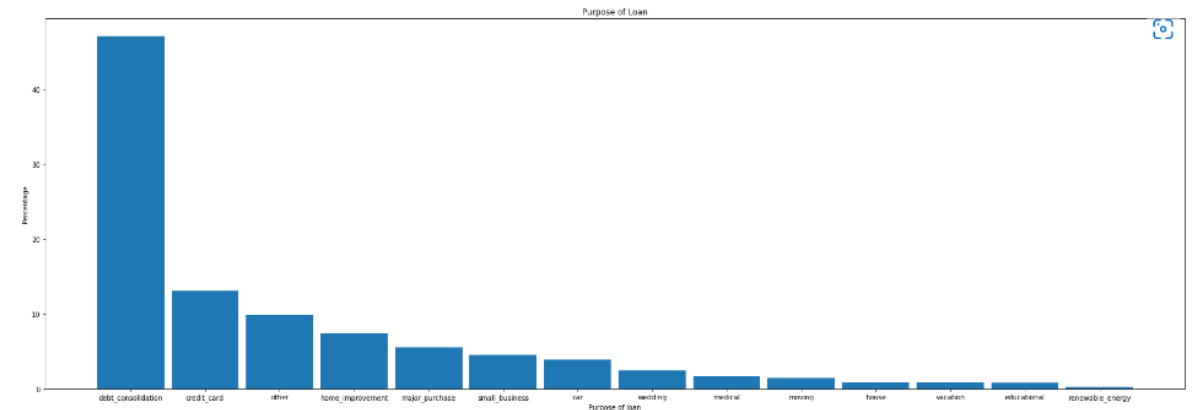
Bivariate Analysis:

loan_status	annual_inc_category	Charged Off	Fully Paid	Total	Chargedoff_Proportion
0	0-20000	191	818	1009	0.189296
1	20000-40000	1421	6709	8130	0.174785
2	40000-60000	1675	9326	11001	0.152259
3	60000-80000	1002	6503	7505	0.133511
4	80000 +	1068	8452	9520	0.112185
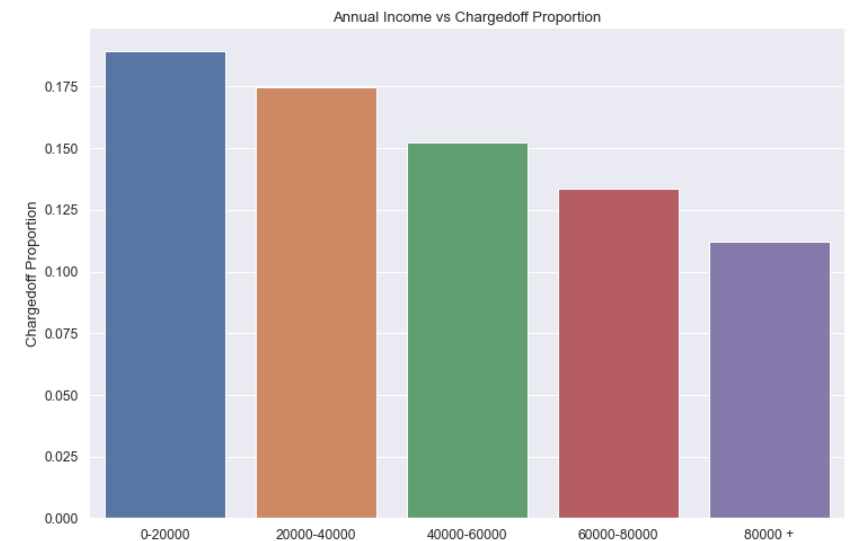
Summary:

- Hightest anual income is 6000000 and lowest annual income is 4000
- Highest loan amount is 35000 and lowest loan amount is 500
- Length of load data set is 38577 85.41% of Fully paid and 14.58% of Charged Off customers.
- Customers have lend their loans mostly for debt consolidation(nearly 47%) and for credit card(nearly   13%).
- Most of the borrowers are living in rented home or mortgazed their home.
- Borrowers who are not working or have less than 1 year of work experience have high chances of getting charged off.
- Small Business borrowers have high chances of getting charged off.
- Borrowers most likely to opt for 36 months term and have less chances to be defaulters/charged off compared to 60 months.
- Percentage of Defaulters is found to be highest in 60 months term (~ 25%) and for 30 months (~ 11%).

Recomendation:

- Applicants under 1 year employment or unemployed charged off.
- It is recommended to not provide the abovementioned applicants with higher loan amount.
- Applicants with a background bankruptcy should have more thorough check of current financial situation.
- High annual income individuals are in a better poisition to pay back the loans, hence should be considered first.