IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11568

ANALYTICS OF LENDING

Harsh Gupta[1], Garwit Choudhary[2], Shraddha Srivastava[3]

[1,2,3]Department of Computer Science and Engineering, Inderprastha Engineering College, Ghaziabad

Abstract: Data in our world is like a gold mine which has to be processed first to make something out of it; In our project data needs to be analysed so as produce good result. There are many companies where they pay their consumers for reviewing their product and these reviews plays a major role to analyse the factor which influences the review rating. Here, we have used EDA i.e., Exploratory Data Analysis where data interpretations can be done in row and column format. We have used python language for data analysis, it is object oriented, interpreted and interactive programming language and it is open source.
Lending club receives a loan application, and it has to decide whether to approve the loan or reject it based on the application. Based on the decision there are two types of risks that can occur which will either result in loss of business or financial loss. Our research paper tried to find out the factors that can reduce the occurring of above factors. We have used EDA to understand how consumer attributes and loan attributes influence the tendency to default.

Keywords: Exploratory Data Analysis, Python, Jupyter, Numpy and Pandas

I. INTRODUCTION

Data are growing very faster in today's world. It is not so easy to process the data manually and keep the updates with the growing data. Data analysis and visualization programs allows us to reach even deeper understanding of the data. Using the programming language Python, with its English commands and easy-to-follow syntax, offers a powerful open-source alternative to the old traditional techniques and applications.
Data analytics allows the companies to understand their efficiency and performance and ultimately helps the business make more informed decisions.
We have used data analytics to identify variables which strongly indicates that the said attributes in such people leads to default of loan. Identifying such attributes will greatly help our decision. It will greatly reduce loss of business and financial losses of the company. It will smoothen the process as we can filter the information based on the attributes and fasten the process.

If one is ready to spot these risky loan candidates, then such loans is reduced thereby lowering the quantity of credit loss. To identify such applicants' using EDA is the main aim of this case study. In different words, the corporate needs to grasp the driving factors (or driver variables) behind loan default, i.e., the variables that are sturdy indicators of default. the corporate will use the conclusions for its portfolio and risk assessment.

II. LITERATURE REVIEW

There has been a large amount of research done in the area of peer-to-peer lending. As it is a popular phenomenon, the literature is still growing. There have been both theoretical and empirical articles written on the subject, with the majority of them being empirical in nature. We have read many of these literature in which different types of methods have been used such as Excel, Rain Forest Algorithm, ML and AI (Every algorithm), Decision tree, Random Forest, Naïve Bayes and Decision Tree, Blockchain, Cryptocurrencies, Artificial Intelligence, Big Data, Principal Component Analysis, Linear Regression, SVM, MLP, Logistic Regression, Feature Engineering, Linear Discriminate Analysis etc. These methods and technologies gave us the different insight towards the project.

III. PROPOSED SYSTEM

This company is an online loan marketplace, facilitating personal loans, business loans, housing loans and financing of medical procedures. Borrowers/Customers can easily access lower interest rate loans through a fast online interface. Like most other companies, lending loans to a 'risky' applicant is the largest source of financial loss (called credit loss). The credit loss is defined as the amount of money lost by the lender when the borrower/customer refuses to pay or runs away with the money he/she owed. In other words, borrowers/customers who default cause the largest amount of loss to the lenders. In this case, the borrowers/customers labelled as 'charged-off' are the 'defaulters'.
In each step of our methodology, we employed the use of Python version 3. We choose Python because of the availability of extensive data analysis libraries. We used Pandas, NumPy, Seaborn, Jupyter-Notebook and Matplotlib in our study.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11568

Data Source: We used Lending Club's data for this analysis. The dataset we are using is for the time period from 2007 to 2011. There are over 42000 observations and over a hundred variables. It's usually terribly laborious to run analysis with all the variables and observations. We found the data on GitHub. So, we cleaned this data.

Data Understanding: After sourcing the data we went through the data dictionary and the actual data to understand the data we are dealing with and to draw rough estimates of what we can achieve from this. We roughly went through each variable and their values to understand what type of data are we dealing with. We discovered that most of the values are missing also many attributes aren't relevant to our goal.

Data Cleaning: Data cleaning is very important to draw conclusive results and to reduce redundancy of data. We have to remove missing columns and set the data according to the data types we want. There are many types of missing values such as MCAR: Missing Completely At Random. It is the highest level of randomness. This also means that the variables with the missing values are not dependent on any other variables/features values. MAR: Missing At Random. This means that the missing values in any column or feature are dependent on other feature values. MNAR: Missing Not At Random. Missing not at random data is a more serious issue and, in this case, it is advisable to check the data gathering process further and understand the reason behind missing data. After determining the type of missing data, we remove or handle the values as per the situation.

Data Preparation: To prepare the date we have to first categorize each column into two types that is numerical and categorical and then we proceed to form pair of these columns for further steps. Also, there are many variables whose values are not in specified data- types like employee's employment length which should be an integer but was object so we check each attribute and their data- types and rectified the problem.

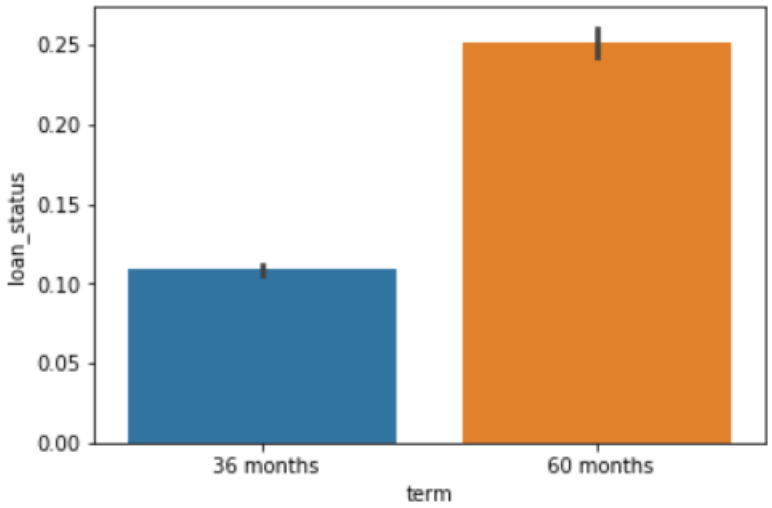
Exploratory Data Analysis: We first separated the Customer behaviour variables (those which are generated after the loan is approved such as delinquent 2 years, revolving balance, next payment date etc.). Now, the customer behaviour variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval so, we removed them. Now comes the main work for our data in which we have to determine which type of analysis can be performed in our dataset in our case we mainly used univariate analysis and few times bivariate analysis. In this section we plot various graphs and determine the conclusions depicted upon them.

IV. CONCLUSION

Overall default rate is 14% of our data. The top products and their interest rate are:

1. Credit card- 11.62%
2. Debt consolidation- 12.4%
3. Home improvement- 11.29%
4. Major purchase- 10.8%

```
plot_cat('term')
```



The loans give with the time period of 60 months defaults more than 36 months.

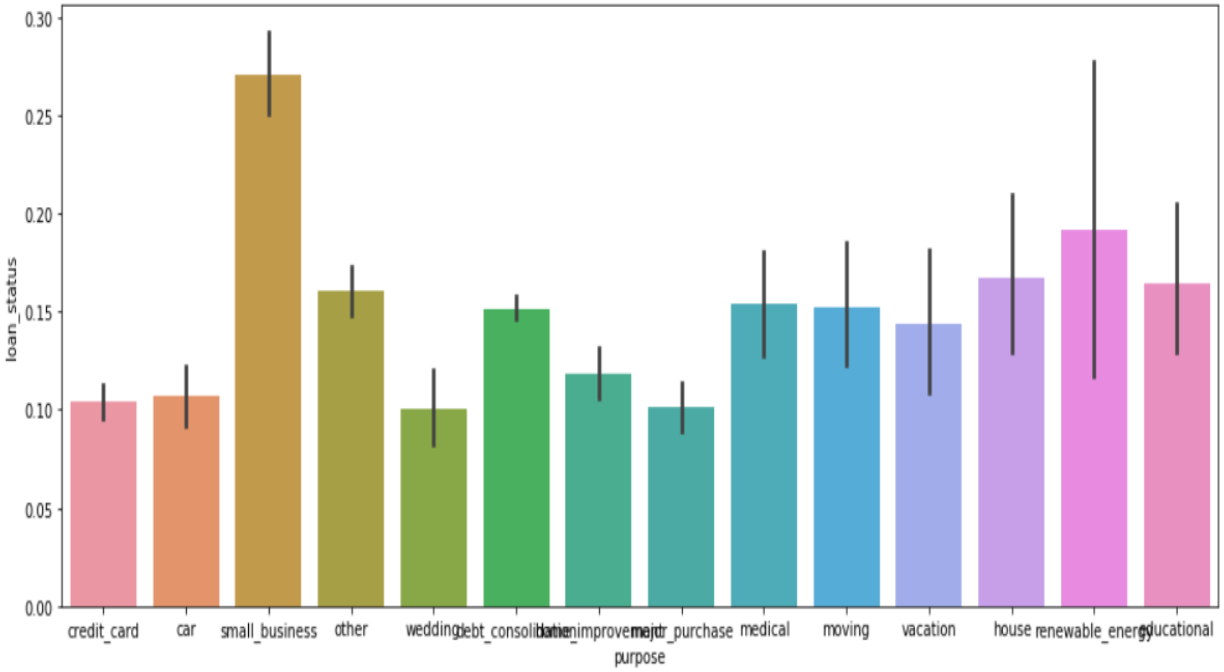
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

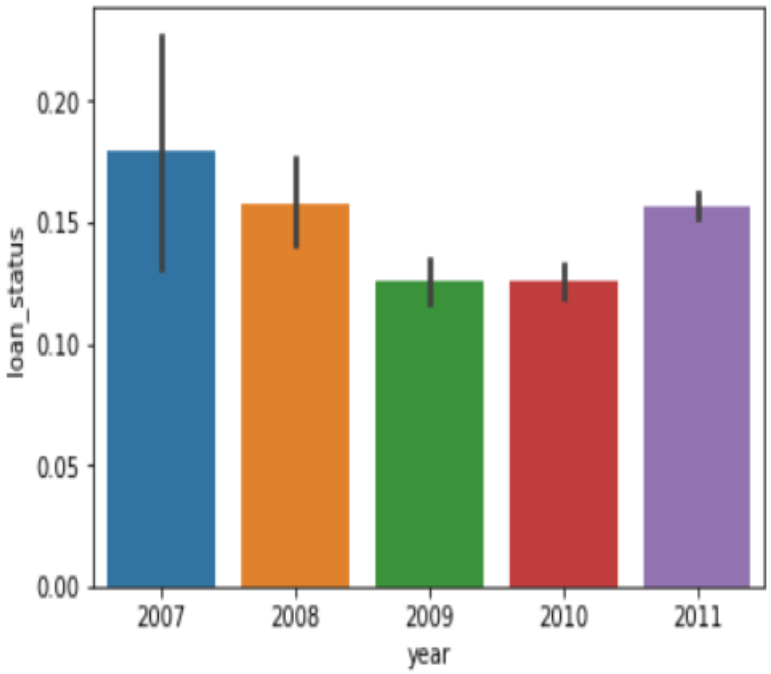
DOI: 10.17148/IJARCCE.2022.11568

```
plt.figure(figsize=(16, 6))
plot_cat('purpose')
```



Small business loans default the most followed by Renewable energy and Education.

```
plot_cat('year')
```



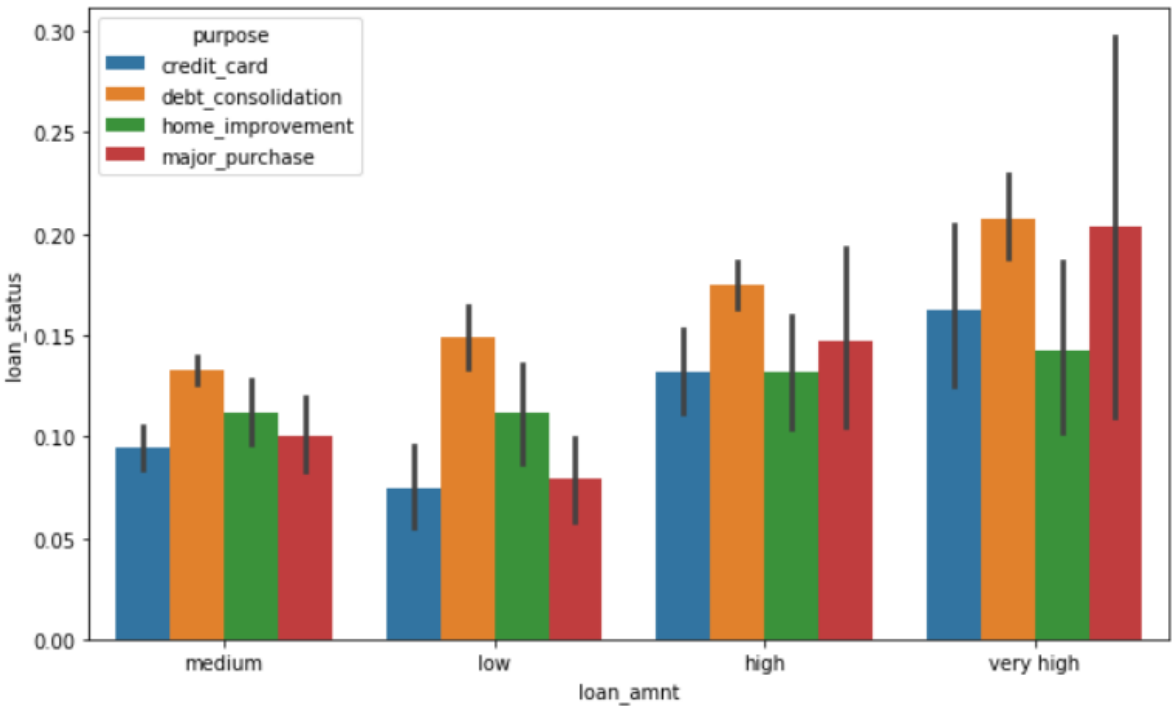
The default rate was steadily going down from 2007 to 2010 but suddenly there was a hike in number of defaulters in 2011; more data is needed to know the specific reason. The only conclusion we reached that the most defaulted loan in 2011 is debt consolidation.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

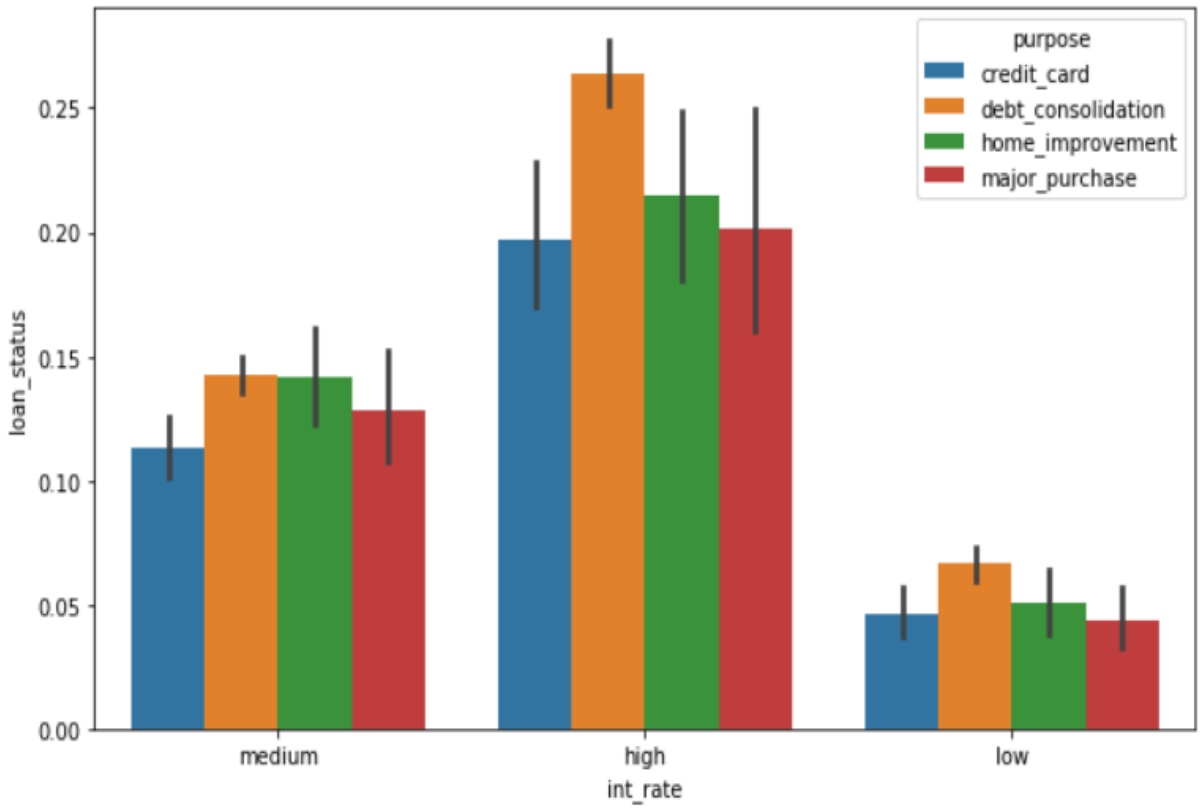
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11568



Bigger (above 15,000) the amount of loan higher the chances of default. In which debt consolidation loan and major purchase loan were the ones that default the most.



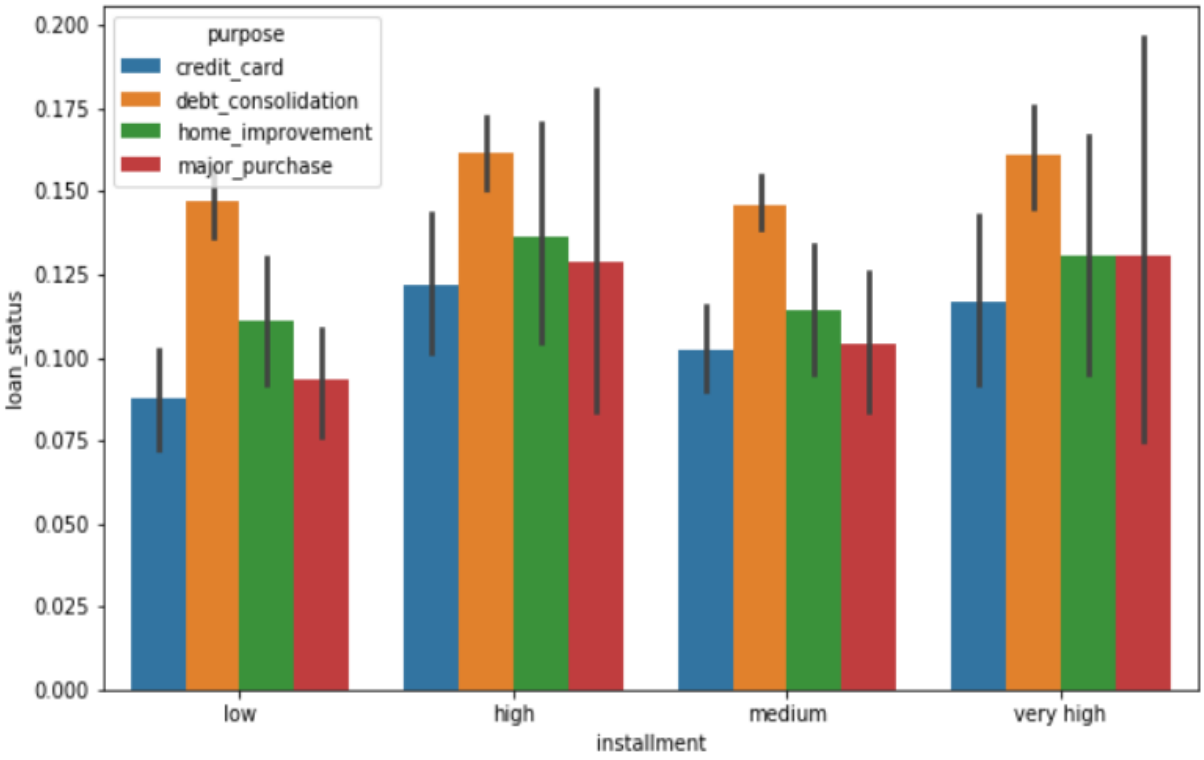
High interest rates (above 15%) have more defaults. In which debt consolidation loan and home improvement loan were the ones that default the most.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

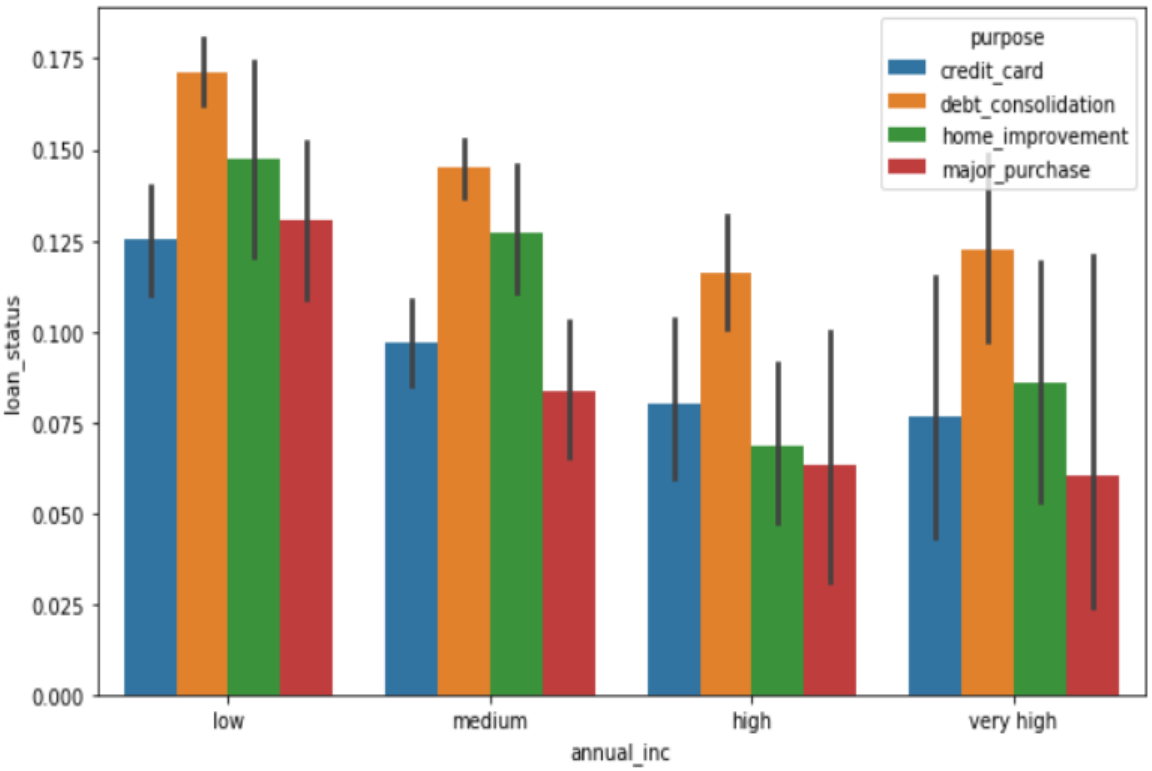
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11568



Higher the instalment amount (above 400) higher the chances of default.



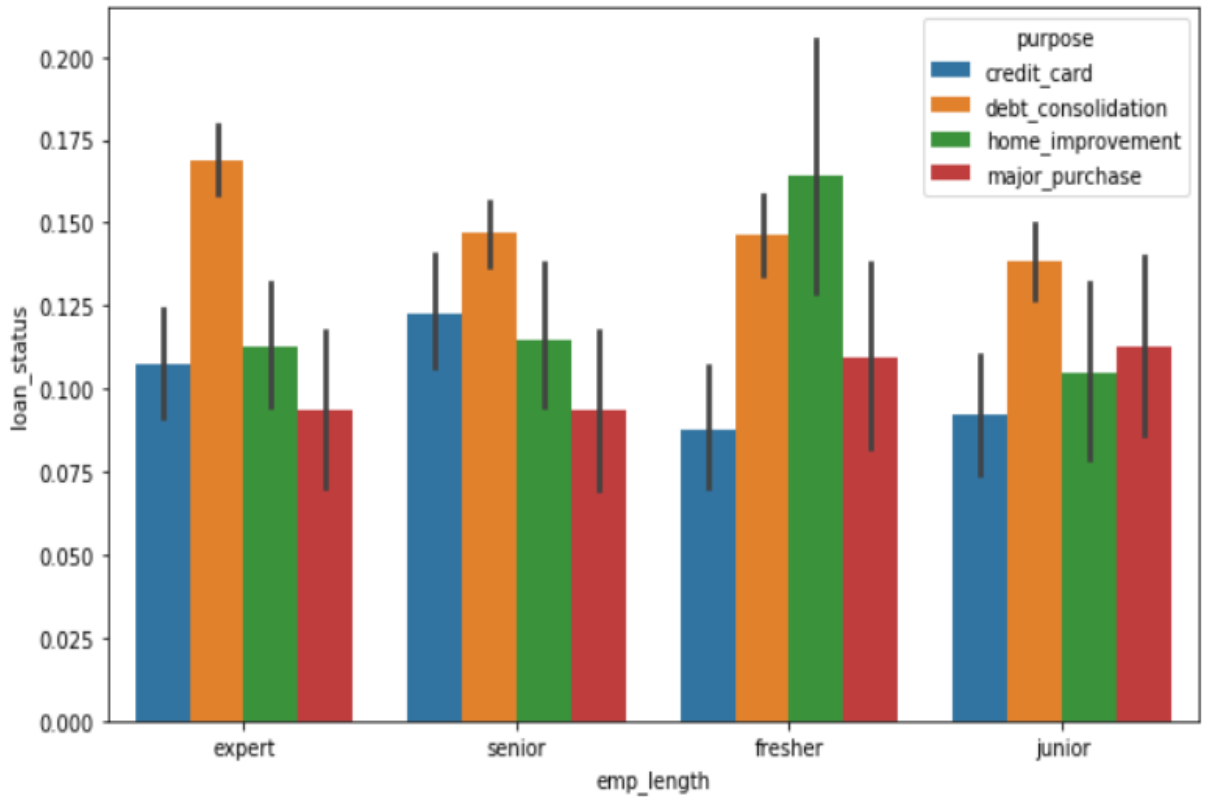
Lower the salary (below 50,000) higher the default rate.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

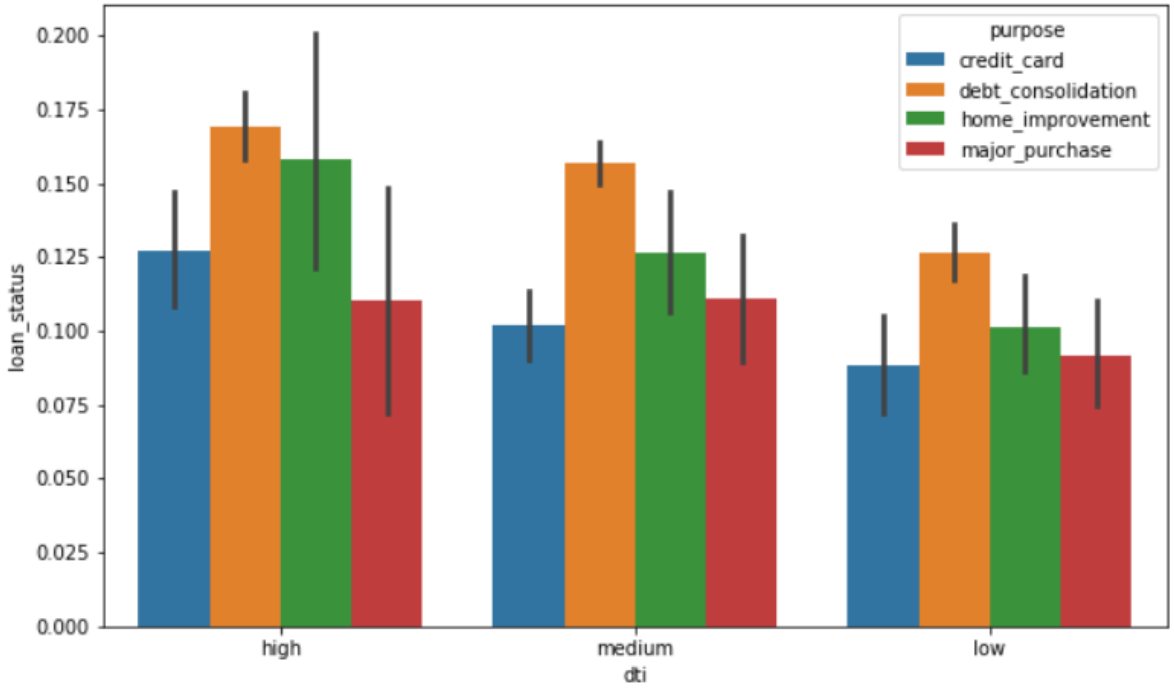
International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11568



If we see according to employment length; debt consolidation loan is the most defaulted loan across every year except the freshers (1-3 years employment length) who mostly defaults in home improvement loans.



According to debt-to-income ratio (dti) then higher the dti (above 20) greater the chances of default while giving the loan such as debt consolidation and home improvement.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 5, May 2022

DOI: 10.17148/IJARCCE.2022.11568

V. REFERENCES

[1] Jean-Francois Darre, (2015). Analysis of Lending Club's data.
[2] J.F. darre, September 30, (2015). Project 1: Lending Club's data.
[3] Credit Revolving Balance: https://www.creditcards.com/credit-card
[4] Lending Club: https://en.wikipedia.org/wiki/Lending_Club
[5] Dr. Rajeshwari M. Shettar. (2018). AN OVERVIEW STUDY ON P2P LENDING.
[6] Anahita Namvar, Mohammad Siami , Fethi Rabhi , Mohsen Naderpour. (2017). Credit risk prediction in an imbalanced social lending environment.
[7] Boris Vallee and Yao Zeng. (2018). Marketplace Lending: A New Banking Paradigm?.
[8] Qionglin Shan & Mikael Nilsson. (2018). Credit risk analysis with machine learning techniques in peer-to-peer lending market.
[9] Jonathan Ricardo Guzman. (2015). A Principal Decision: The Case of Lending Club.
[10] Carlos Serrano-Cinca, Begoña Gutiérrez-Nieto, Luz López-Palacios. (2015). Determinants of Default in P2P Lending.
[11] Bc. Michal Polena. (2016-17). Performance Analysis of Credit Scoring Models on Lending Club Data.
[12] Mohammad Mubasil Bokhari. (2019). Credit Risk Analysis in Peer to Peer Lending Data set: Lending Club.
[13] Matthew Courchene. (2014). A Theoretical Analysis of Peer-to-Peer Lending.
[14] Mohammad Rafiqul Islam, Tabitha Kemboi. (2019). Project: Lending Club Data Analysis.
[15] Alexander Bachmann, Alexander Becker, Michel Hilker. (2011). Online Peer-to-Peer Lending.
[16] Pierre-Yves FESTOC. (2013-2014). Lending Club – P2P Lending Impact Of Loan Description On Loan Performance.
[17] Lin Zhua , Dafeng Qiua , Daji Ergua , Cai Yinga , Kuiyi Liub. (2019). A study on predicting loan default based on the random forest algorithm.
[18] Guangyou Zhou, Yijia Zhang and Sumei Luo. (2018). P2P Network Lending, Loss Given Default and Credit Risks.
[19] JIAYU YAO. (2018). VALUATION OF A FINTECH COMPANY: LENDING CLUB.
[20] Haitian Lu, Bingzhong Wang, Qing Wu, and Jing Ye. (2020). Fintech and the Future of Financial Service: A Literature Review and Research Agenda.

- Data.world
- Kaggle.com
- Data.gov.in
- Github.com
- Google.com

