EDA lending case study

Explore the world of EDA lending and dive into a captivating case study that uncovers valuable insights and outcomes.

By ROOPESH H S,
BINDUJA P KADEMANI

Introduction

EDA lending, or Exploratory Data Analysis lending, is a crucial process in the financial industry. It involves analyzing patterns in customer attributes and loan data to make informed loan approval decisions.

- A consumer finance company faces the challenge of balancing the risk of lost business with the risk of financial losses due to loan defaults.
- EDA techniques can be used to uncover patterns and relationships between borrower attributes and loan performance, reducing the risk of non-repayment while still capturing profitable lending opportunities.
- The goal is to identify patterns in consumer attributes and loan attributes associated with loan defaults, understand the relationship between these attributes and loan defaults, and develop insights that can be used to predict whether a new loan applicant is likely to default.

Background

The lending industry is constantly evolving, facing numerous challenges such as increasing default rates and changing customer preferences. Understanding these challenges is essential for successful loan approval processes.

- Loan lending is a crucial aspect of the financial industry, enabling individuals and businesses to access credit.
- Loan defaults occur when borrowers fail to make scheduled payments, leading to financial losses for lenders.
- Borrower-related factors influencing loan defaults include credit history, debt-to-income ratio, and employment stability.
- Loan-related factors influencing loan defaults include loan amount, interest rate, and loan terms.
- Loan-related factors influencing loan defaults include loan amount, interest rate, and loan terms. EDA plays a vital role in loan lending by identifying patterns, trends, and correlations between borrower attributes, loan attributes, and loan defaults.
- Key objectives of EDA in loan lending are to identify patterns, understand relationships, and develop predictive insights for new loan applicants.

Problem Statement

A consumer finance company is looking for patterns in customer and loan attributes that are associated with loan defaults. By identifying these patterns, they can improve their loan approval decisions and mitigate risks.

Through comprehensive EDA, we aim to gain a deeper understanding of the following aspects:

- Identifying Patterns in Consumer Attributes and Loan Attributes Associated with Loan Defaults: Uncovering the specific characteristics of borrowers and loan terms that are more closely linked to defaults.
- Understanding the Relationship Between Consumer Attributes and Loan Attributes with Loan Defaults: Exploring the interplay between borrower characteristics, such as credit history, income levels, and debt-to-income ratios, and loan attributes, such as loan amounts, interest rates, and repayment terms, in influencing default rates.
- Developing Insights that Can Be Used to Predict Whether a New Loan Applicant is Likely to Default: Utilizing the insights gained from EDA to develop predictive models that can assess the creditworthiness of new loan applicants, enabling more informed loan approval decisions.

Data Analysis

Data analysis plays a crucial role in EDA lending. By leveraging data, financial institutions can gain valuable insights into customer behavior, loan performance, and market trends, enabling them to make informed decisions.The goal of this analysis is to develop insights that can be used to predict whether a new loan applicant is likely to default.

These are the following steps needed to perform Data Analysis :

1. Data Extraction.
2. Data Cleaning.
3. Exploratory Data Analysis (EDA) :

- Univariate Analysis.
- Segmented Univariate Analysis.
- Bivariate Analysis.
- Derived Metrics

Data Extraction And Cleaning

Loan lending data has been extracted and converted to csv file. It contains the complete loan data for all loans issued through the time period 2007 to 2011.We have imported the Data to Exploratory Data Analysis .

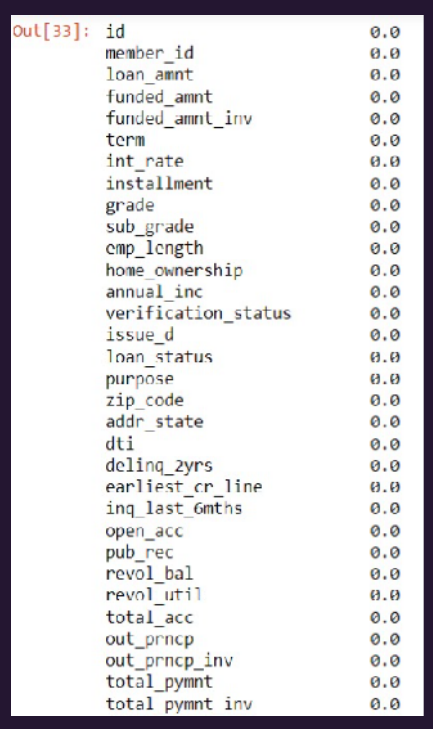
Data Extraction :

Extract the suitable Datasets required for Case-study.

Data Cleaning :

1. Identify and handle missing values in the data :

- In the dataset, we encountered missing values, unwanted columns, and null values. We have dropped unwanted columns, removed missing value rows, imputed values to make the data-frame concise .
- As we can see in the following image, the data-frame does not contain null or missing values after the cleaning is done.

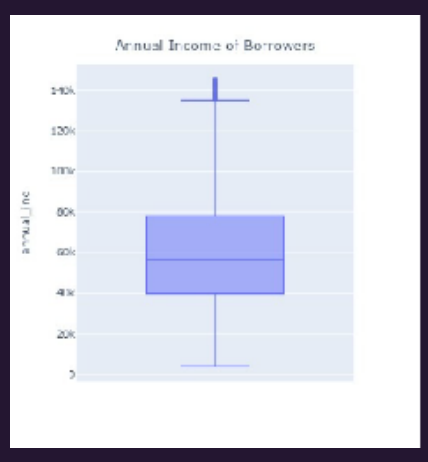


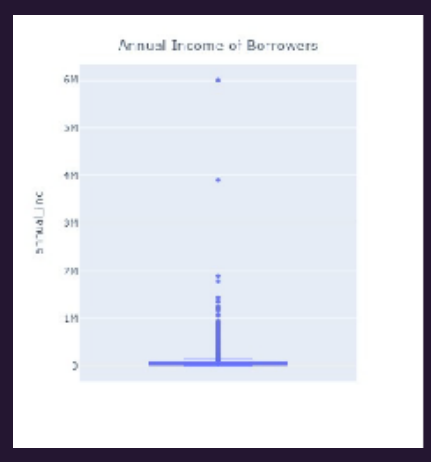
2. Standardize and normalize numerical variables if necessary :

- In the data-frame, for column employee length(work experience) we observed unstandardized value. Since work experience value should be in integer, we removed unstandardized values and converted the column to integer.

3. Check for outliers and inconsistencies in the data :

- In the Data-frame, we observed outliers in the annual income column. To identify these outliers, we plotted a bar plot and removed them from the dataset.
- We also removed inconsistent column Description (desc) since they were added by the borrower and were not relevant to the analysis.
- In following set of images, we can see how the annual income column looks before(Left) and after(right) removing outliers.





Exploratory Data Analysis (EDA)

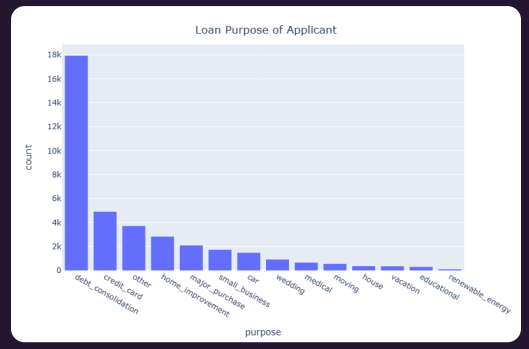
Univariate Analysis

It is a statistical technique that involves the examination of a single variable at a time. It is a useful tool for understanding the basic characteristics of a dataset and for identifying patterns and trends.

To apply univariate analysis for the given dataset, we have identified the following variables,

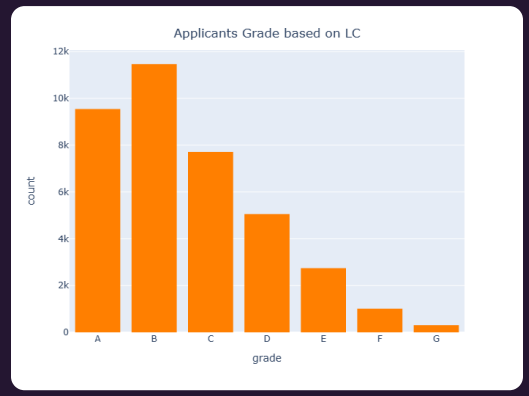
- Term, Grade, Sub-Grade, Home-Ownernship , Verification Status , Loan-Status , Address-State etc.
- We have used Data Visualization to plot graph.
- The following are the some of the graph that we have plotted using univariate analysis.

Univariate Analysis Graphs



1. Loan Purpose Histogram

- The graph shows the Loan given to Applicants.
- As we can see , highest number of loan is give to the debt_consolidation and various fields with different ratio of loans provided.



2. Applicant Grade Histogram

- The graph shows the grade given to Applicants based on loan merit.
- As we can see, grade B has highest number of Applicants and decreases trend of number of Applicants as the grade increases .

3. Loan Amount Histogram

- The graph shows the Loan amount given to Applicants.
- As we can see, range of loan amount with different values like median , quartile range etc...

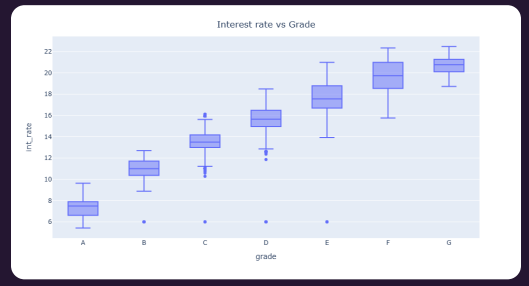
Segmented-Univariate Analysis

Segmented univariate analysis is an extension of univariate analysis that involves dividing a dataset into segments based on a categorical variable and then performing univariate analysis on each segment. This allows you to compare and contrast the characteristics of a variable across different groups or categories.

To apply segmented-univariate Analysis for the given data-set, we have identified the following variables.

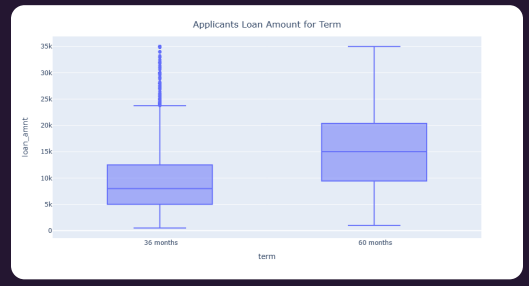
- Interest rate vs Grade, Applicants verification_status for loan_amnt, Applicants Loan Amount for Term etc.
- We have used Data Visualization to plot graph.
- The following are the some of the graph that we have plotted.

Segmented-Univariate Analysis Graphs



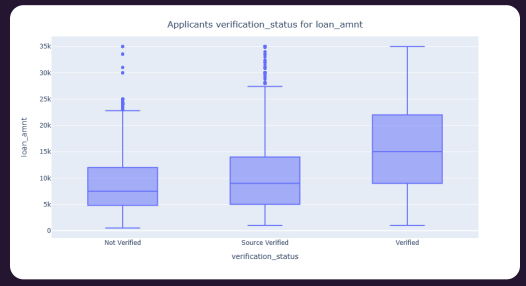
Interest rate vs Grade

- The plot shows Applicants interest rate range for the grade.
- As we can observe in graph, the interest rate increases rapidly as the grade for which Applicants have taken loan.



Applicants Loan Amount for Term

- The plot shows Applicants loan amount range for the selected term.
- As we can see in the graph, the loan amount has more variations for term 36 months than 60 months.



Applicants verification-status for loan amount

From the plot, we can see that the Applicants who are verified have taken larger loan while source verified and not verified borrowers have taken lower amounts.

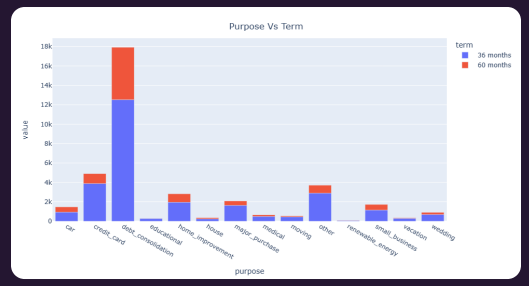
Bivariate Analysis

Bivariate analysis explores the relationship between two variables. It helps us understand how changes in one variable affect the other, revealing patterns, correlations, and potential associations.

To apply Bivariate Analysis for the given dataset, we have identified the following variables :

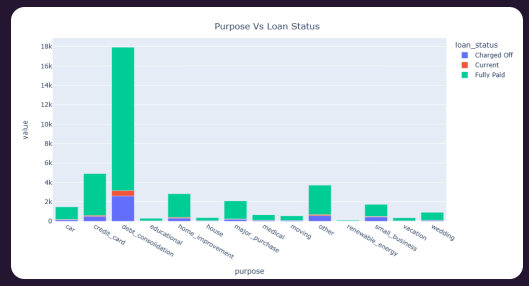
- Purpose Vs Term, Purpose, Purpose Vs Loan Status , correlation of suitable datasets etc.
- We have used Data Visualization to plot graph for Two different variables.
- The following are the some of the graph that we have plotted.

Bivariate Analysis Graphs



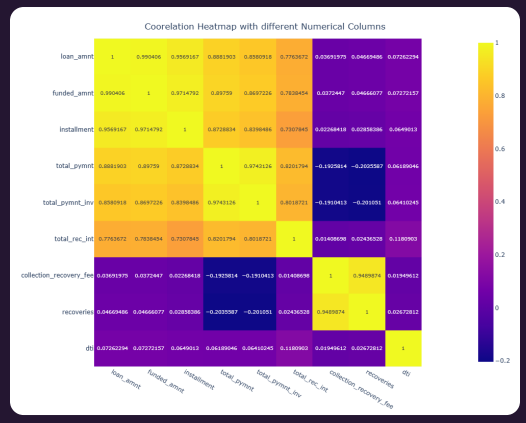
Purpose Vs Term

- In the bar graph, we can see most number of loans taken by the Applicants are for debt consolidation and most of them are for 36 months terms.
- While we saw debt consolidation and small business Applicants had similar loan amount spread in segmented univariate analysis graph we can see here that number of Applicants are less for small business compared debt consolidation.



Purpose Vs Loan Status

- In the bar graph, we can see loan status of the Applicants for which they took loan for.
- As the graph indicates , we can clearly see fully paid applicants are more when compare to other status.



Correlation of data's

The following Heat map shows correlation between the different numerical columns in the dataset .

This shows how the columns value changes while another column value changes.

- [0→1] value denotes the columns changes positively i.e. if one column value increases the other also increases.
- [0→-1]value denotes the columns changes negatively i.e if one column value increases the other also increases.
- [0] value denotes the columns are not correlated.

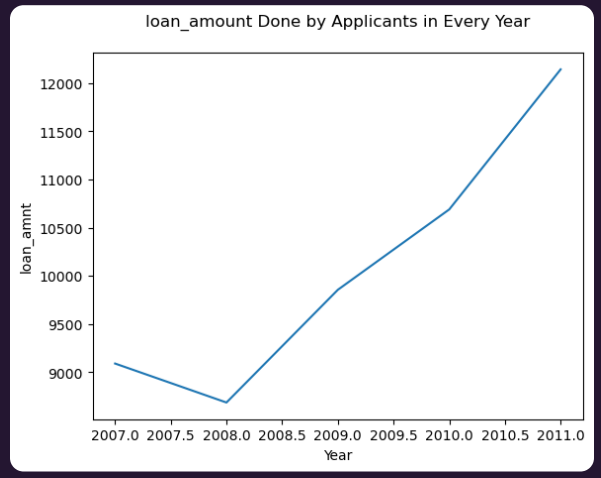
Derived Metrics

A derived metric is a new measurement created by combining or transforming existing variables in a dataset. Derived metrics are often used to gain a deeper understanding of the data and uncover hidden patterns or relationships that might not be apparent from the original variables.

To apply Derived metrics for the given dataset, we have identified the following variables :

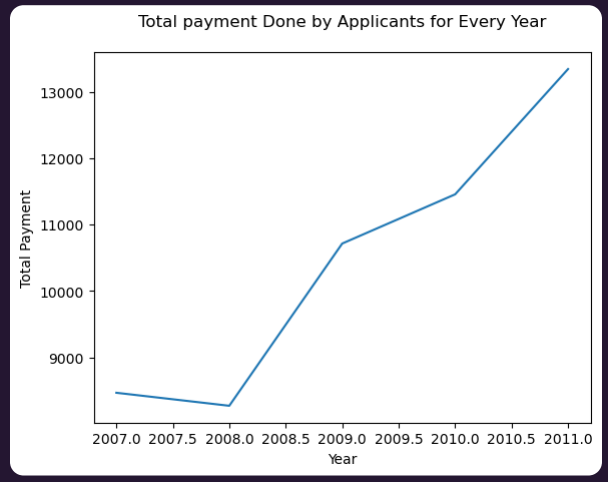
- We have extracted year and month from issue_d column and used the same for plot creation.
- We have used Data Visualization to plot graph for Year and Month grouped for different variables .
- The Following are the some of the graph that we have plotted.

Derived Metrics graphs



Total payment Done by Applicants for Every Year

In the above line graph, we can see the total payment cleared by applicants decreases till 2008. and increases as the year passes.



loan amount Done by Applicants in Every Year

In the above line graph, we can see the total payment done decreases slowly from 2007-2008 and then increases exponentially.

Results

While EDA lending offers significant benefits, there are also challenges and limitations to consider. It requires thorough data collection, processing, and interpretation. Careful analysis is essential to avoid biased decision-making.

- Exploratory Data Analysis (EDA) plays a crucial role in identifying patterns and relationships between borrower attributes, loan attributes, and loan defaults.
- The goal of EDA in loan lending is to develop insights that can be used to predict whether a new loan applicant is likely to default.
- Key outcomes of EDA for loan lending include identifying patterns in borrower attributes, understanding the relationship between borrower and loan attributes, and refining loan approval criteria.
- EDA can also help in identifying high-risk borrowers, benchmarking against industry standards, and guiding marketing campaigns.
- EDA can inform decision-making, reduce risk, and enhance the profitability of lending operations.

Conclusion

EDA can inform decision-making, reduce risk, and enhance the profitability of lending operations.

Prepared By,

- ROOPESH H S [Group Facilitator]
- BINDUJA P KADEMANI [Group member]