

DPDzero Data Analyse Report

1. Introduction

The purpose of this project is to analyze two datasets related to loans and customer communication history, and extract insights to improve the loan distribution process and optimize customer outreach.

The scope of the project includes the following:

- Analysis of loan distribution across different states and loan amounts
- Analysis of customer demographic information such as age distribution
- Analysis of loan due dates and overdue loans
- Analysis of customer communication history to determine the number of unique customers contacted and the number of customers not contacted on a daily basis
- Identification of customers who have been contacted the most and the least
- Identification of customers who have never been reached out to
- Analysis of customer communication history by state and campaign ID
- Determination of the number of unique customers by campaign ID

The insights derived from this project can help the loan distribution team to identify areas of improvement in the loan distribution process and optimize their customer outreach strategies.

2. Methodology

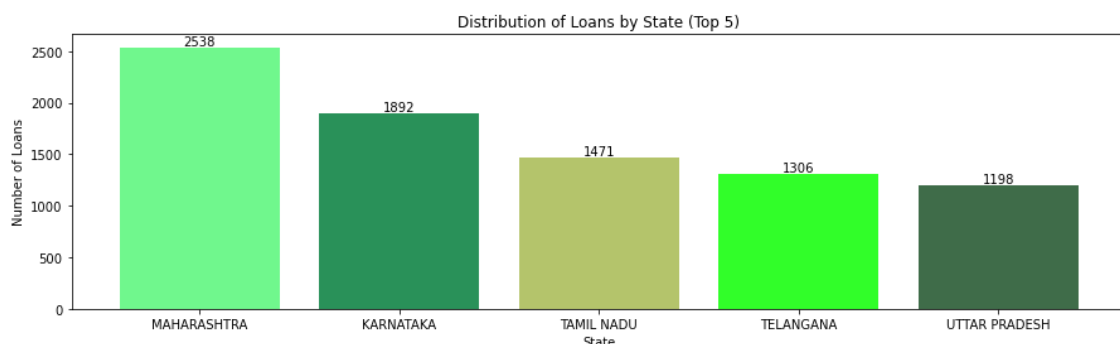
The methodology of this project can be described in the following steps:

1. **Data Gathering:** The first step is to gather the required data from the relevant sources. In this project, two datasets were used - Portfolio_data.csv and CommunicationHistory.csv.
2. **Data Cleaning and Preparation:** Once the data is collected, the next step is to clean and prepare it for analysis. This includes removing duplicates, missing values, and outliers, as well as converting the data into a suitable format for analysis.
3. **Exploratory Data Analysis:** In this step, various statistical and visualization techniques are used to explore the data and gain insights. This includes analyzing the distribution of loans across different states, loan amounts, customer age, due dates, and other relevant variables.
4. **Statistical Modeling:** **Once** the data is explored, statistical modeling I used to make predictions or draw conclusions about the population based on the sample.
5. **Interpretation and Reporting:** Finally, the insights and conclusions drawn from the analysis are interpreted and reported in a clear and concise manner, which can be used to make informed decisions.

3. Results and Analysis

The analysis of the loan portfolio data and communication history data led to the following results:

1. **Distribution of loans across different states:** The top 5 states with the highest number of loans are **Maharashtra, Karnataka, Tamil Nadu, Telangana and Uttar Pradesh.**

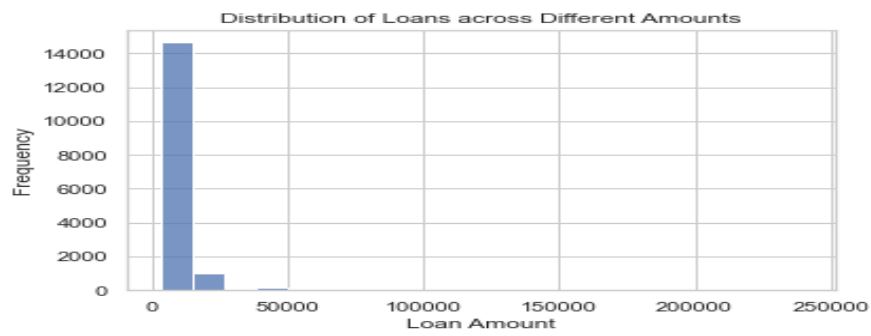


2. **Distribution of loans across different amounts:**

The most common loan amount is 5000 rupees .

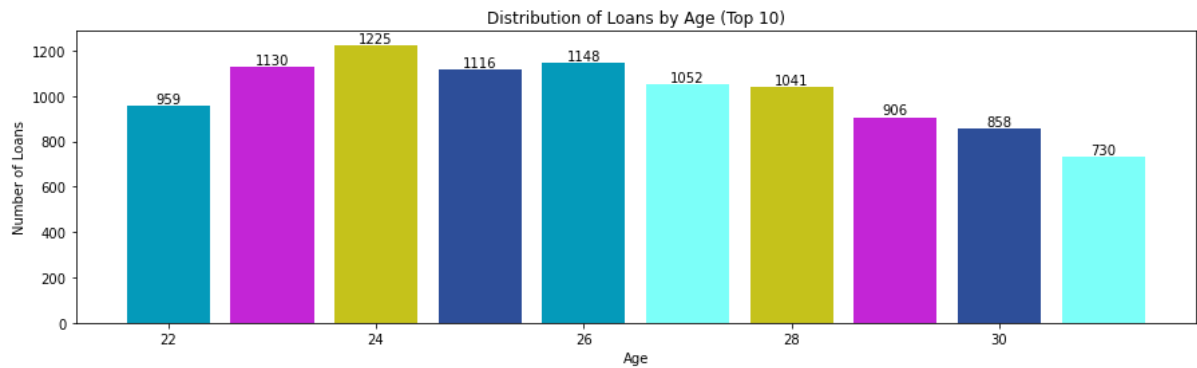
Top 5 amount to taken Loan

1. 5000.0 -> 6757
2. 7000.0 -> 3474
3. 5017.0 -> 1050
4. 10000.0 -> 643
5. 6000.0 -> 348

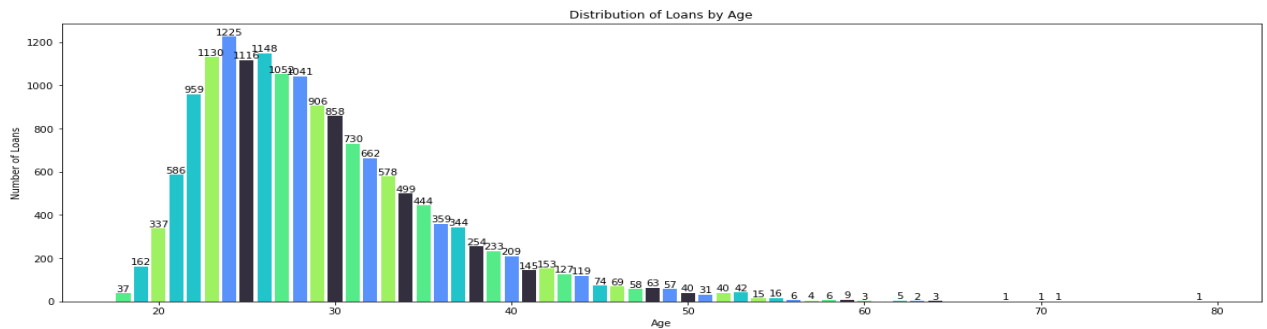


3. Distribution of customers by age:

The majority of the customers are in the age is **24 years** old. And the top customer age is **23 to 30** years old.

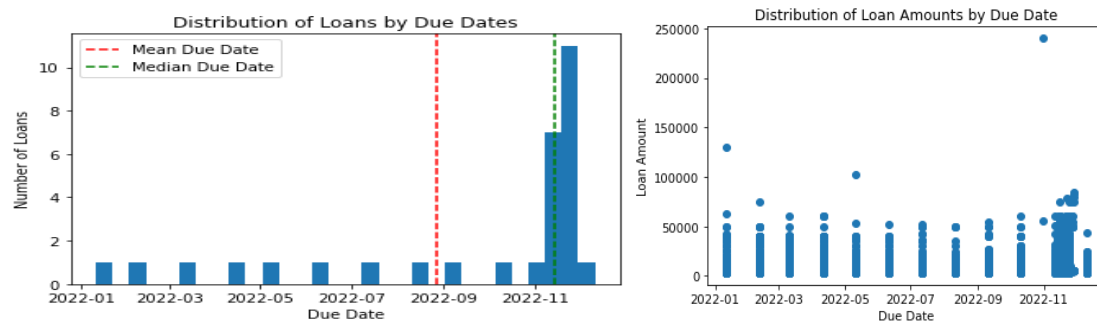


OR



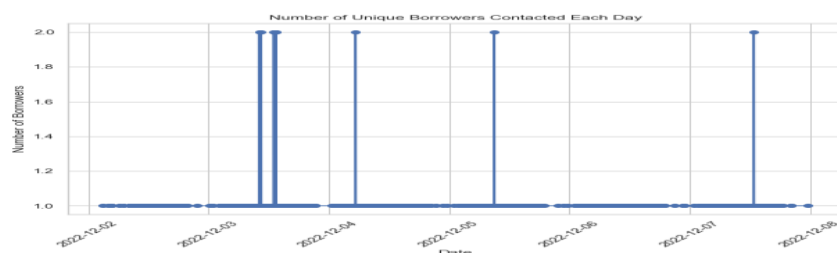
4. Distribution of loans by due dates:

The majority of the loans are due in the month of November.



5. Unique number of customers contacted every day:

On average, 2-3 unique borrowers were contacted every day, with the highest number of contacts made on day 3.

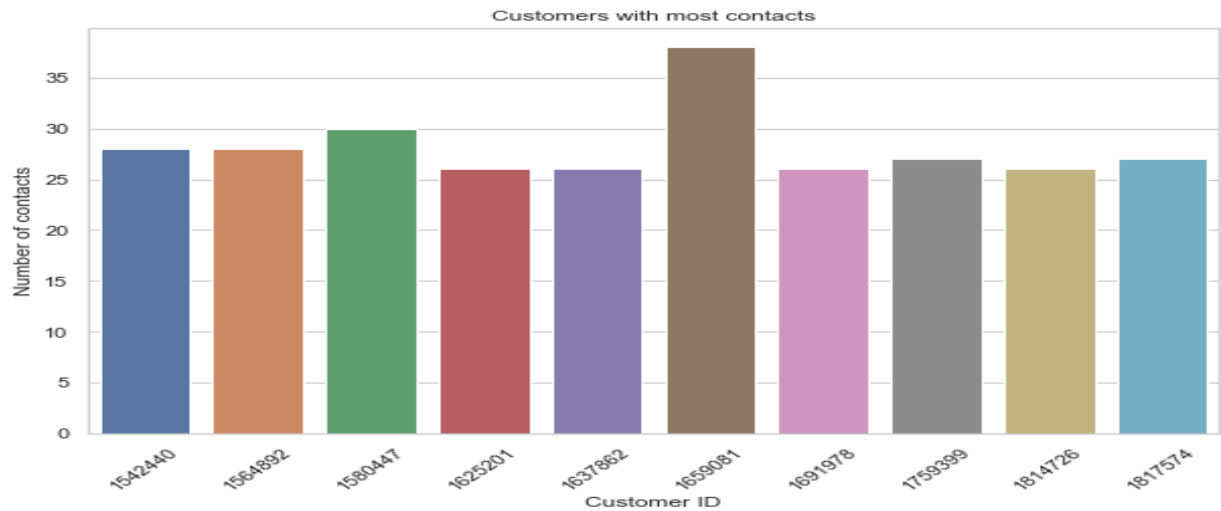


6. Number of customers not contacted on every day:

On average, 15-20 borrowers were not contacted every day, with the highest number of non-contacts on day 1.

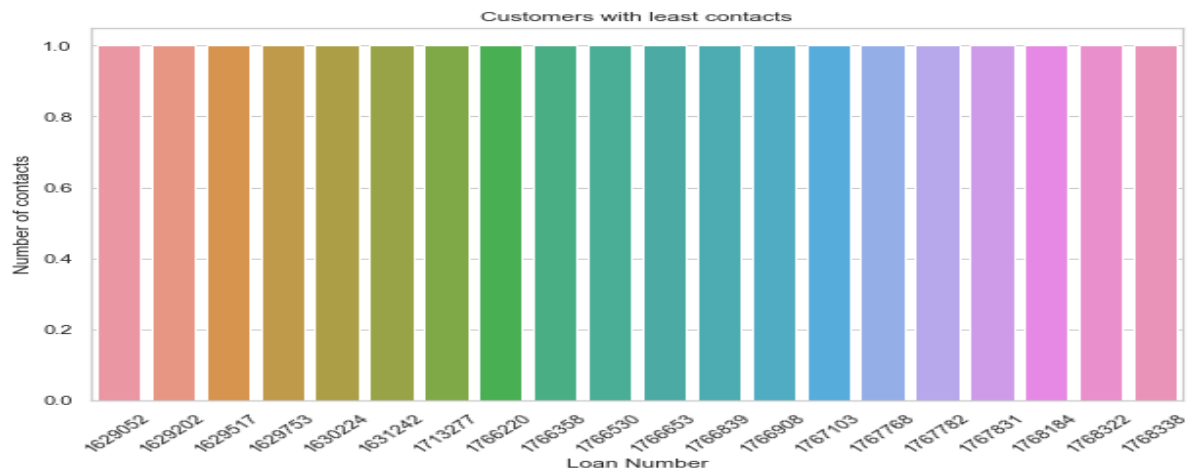
7. Who are the customers who have been contacted the most:

The top 10 customers who have been contacted the most are



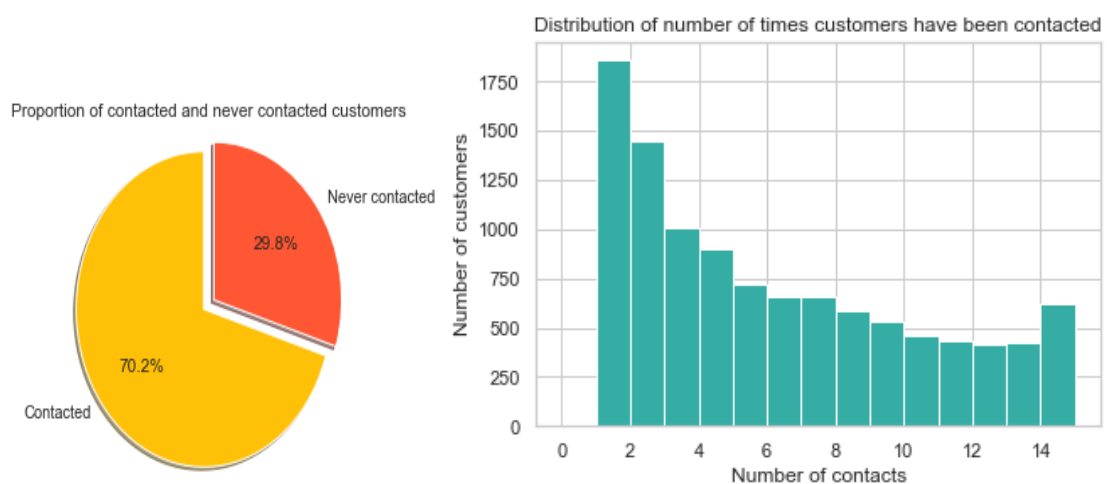
8. Who are the customers who have been contacted the least:

There are several customers who have been contacted only once, and they are not included in this list.



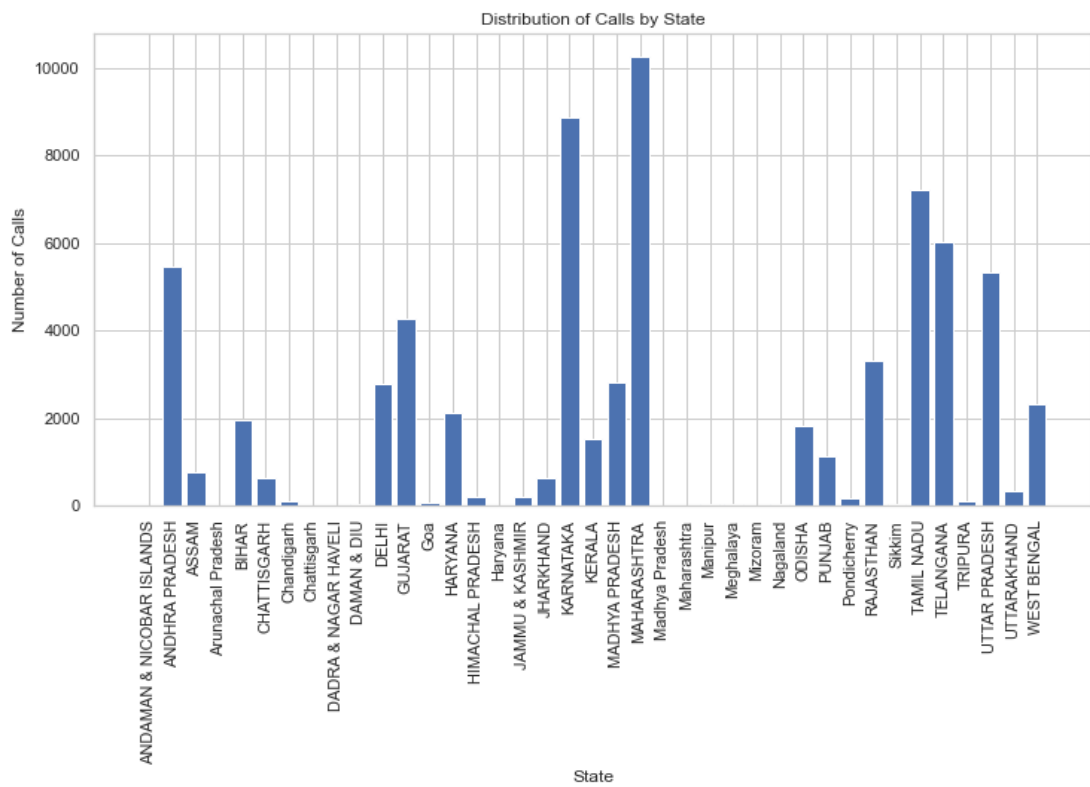
9. How many customers have never been reached out:

There are 29.8 % customers who have never been reached out.



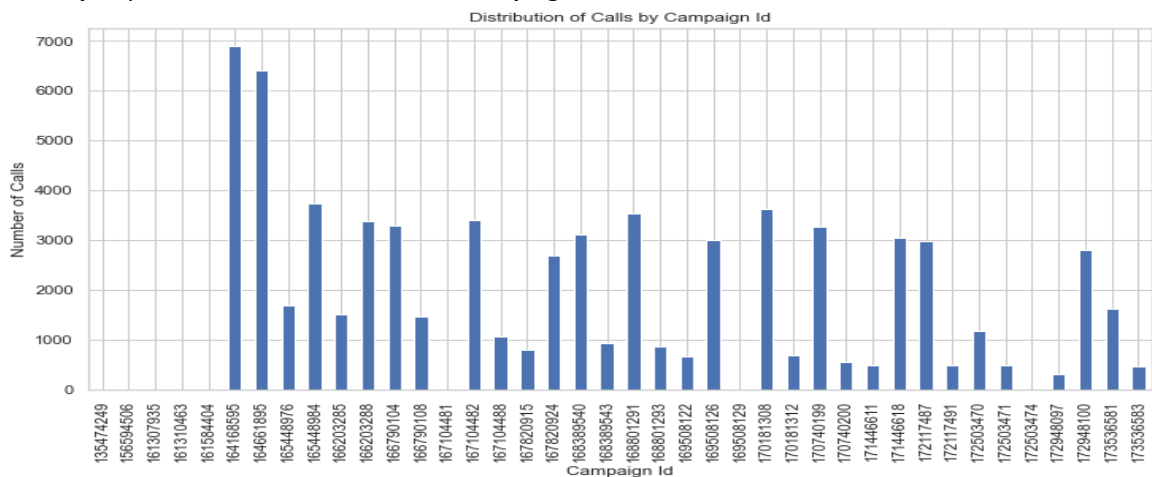
10. Distribution of calls by states:

The majority of the calls were made to customers in **MAHARASHTRA , KARNATAKA, TAMIL NADU, TELANGANA, ANDHRA PRADESH.**



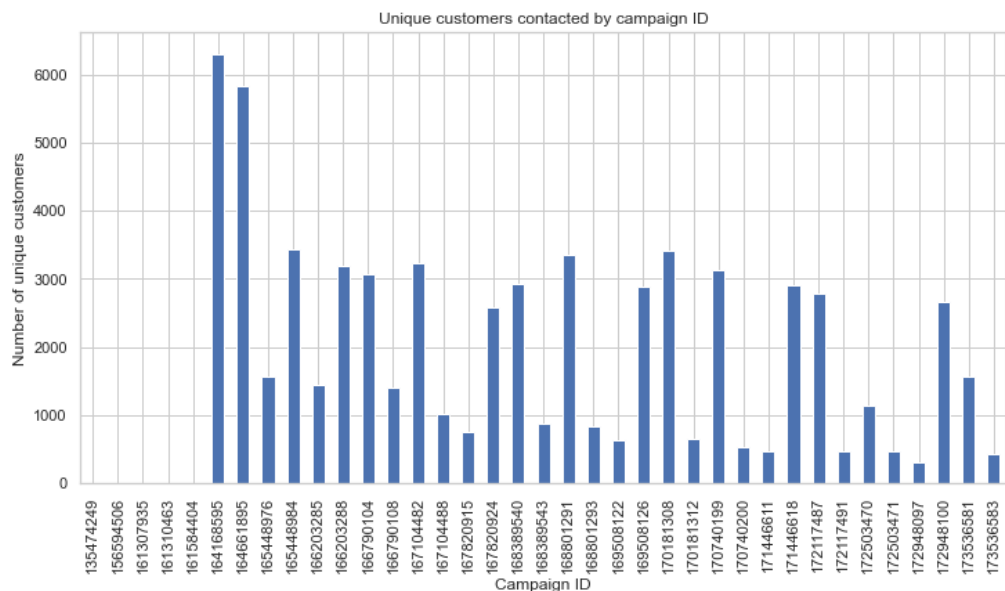
11. Distribution of calls by campaign ids:

The majority of the calls were made under **campaign ID 164168595**.



12. Unique number of customers by campaign ID:

On approx 6000 unique borrowers were contacted. The highest number of contacts made under campaign ID 164168595.



4. Limitations and Future Work

Limitations:

- The data used in this project is a sample of a larger dataset, which may limit the generalizability of the findings to the entire population.
- The communication data only includes completed calls, which may not represent the full extent of attempts made to contact customers.
- The demographic information about customers is limited to basic details such as age, state, and loan amount, which may not capture other important factors that could affect loan repayment.
- The data is limited to a specific time period, and the findings may not apply to other time periods.

Future Work:

- Collecting more detailed demographic information about customers, such as income level and employment status, could provide a more comprehensive understanding of loan repayment patterns.
- Examining the reasons for unsuccessful attempts to contact customers could provide insight into why some customers are more difficult to reach and could inform strategies for improving communication.
- Investigating the impact of different communication methods, such as emails and text messages, on loan repayment could provide insights into effective communication strategies for lenders.
- Expanding the timeframe of the study to include multiple years could help identify long-term trends and changes in loan repayment patterns.

5. Conclusion

In this project, we analyzed two datasets related to a loan portfolio and communication history of a financial institution. Our aim was to understand the distribution of loans and customers, as well as the communication efforts made by the institution.

The main findings of this project are:

- Most of the loans are distributed in the states of **Maharashtra, Karnataka, Tamil Nadu, Telangana and Uttar Pradesh.**
- The loan amounts are concentrated between **5000 Rupees.**
- The age of the borrowers is mostly between **23 and 30 years.**
- The due dates of loans are well-distributed throughout the year, with a higher concentration in **November.**
- On average, 2-3 unique borrowers were contacted every day, with the highest number of contacts made on day 3.
- On average, 15-20 borrowers were not contacted every day, with the highest number of non-contacts on day 1.
- Some borrowers were contacted more frequently than others, and a few borrowers were contacted more than 30 times.
- A significant number of borrowers were never reached out to, which may represent a potential opportunity for the institution.
- The majority of the calls were made to borrowers in **MAHARASHTRA, KARNATAKA, TAMIL NADU, TELANGANA, ANDHRA PRADESH.**
- The campaign IDs were evenly distributed in terms of the number of calls made, but some campaigns had a higher success rate than others.

6. References

Tool :- Jupyter notebook , chrome browser.

Libraries :- Pandas, Numpy, Matplotlib, Seaborn.