Loan Data Analysis

Name: Harsh Gupta

Email: dsharshgupta@gmail.com

Mobile No: 8890817327

Link to Jupiter Notebook: -
https://colab.research.google.com/drive/1EYMy_6LliyvkEllUYzsWRHX8L0Zw5Ovf?usp=sharing

• Introduction:

In this report, we present a comprehensive analysis of loan data obtained from DPDzero. The dataset provided consists of information on various borrowers within a loan portfolio, including their pending loan amounts, demographic details such as state and city, tenure, interest rates, bounce behaviour, disbursed amounts, and unique loan identifiers.

The problem statement itself suggests some of the analyses which include risk labelling, tenure labelling, ticket size labelling and channel recommendations. Here is a brief overview of what we have done in the analysis part.

The report is structured as follows:

Risk Analysis: We categorize borrowers into risk segments based on their bounce behaviour, providing a summary of the distribution along with visuals.

Ticket Size Segmentation: Borrowers are grouped into three cohorts based on the size of their amount pending.

Tenure Completion Analysis: Borrowers are segmented according to their tenure status, providing insights into the stage of their loan repayment journey.

Channel Spend Recommendations: We propose optimized channel spend strategies, considering factors such as borrower behaviour, communication effectiveness, and cost efficiency.

Tools: We have used tools such as Python, Pandas and NumPy for the analysis.

Here is how the data looks before and after analysis.

Before Analysis:

Amount Pending	State	Tenure	Interest Rate	City	Bounce String	Disbursed Amount	Loan Number
963	Karnataka	11	7.69	Bangalore	SSS	10197	JZ6FS
1194	Karnataka	11	6.16	Bangalore	SSB	12738	RDIOY
1807	Karnataka	14	4.24	Hassan	BBS	24640	WNW4L
2451	Karnataka	10	4.70	Bangalore	SSS	23990	6LBJS
2611	Karnataka	10	4.41	Mysore	SSB	25590	ZFZUA

After Analysis:

Amount Pending	State	Tenure	Interest Rate	City	Bounce String	Disbursed Amount	Loan Number	Risk Label	Tenure Label	Ticket Size	Metropolitan city	Score	Channel label
423	Maharashtra	11	11.84	Sangli	FEMI	4389	HEMS0	Unknown risk	Early tenure	Low ticket size	No	2	Whatsapp bot
444	Tamil Nadu	11	12.23	VIRUDHUNAGAR	FEMI	4598	1BYJD	Unknown risk	Early tenure	Low ticket size	No	2	Whatsapp bot
451	Maharashtra	7	37.92	Pune	LSSSSB	2793	7COLC	High Risk	Late tenure	Low ticket size	Yes	7	Voice bot
522	Karnataka	11	12.83	Bagalkot	FEMI	5390	587TX	Unknown risk	Early tenure	Low ticket size	No	2	Whatsapp bot
522	Maharashtra	11	12.83	Pune	S	5390	5QJN0	Low risk	Early tenure	Low ticket size	Yes	3	Whatsapp bot

- ## Risk Analysis:

Criteria for segmentation:

Unknown Risk (New Customers): Borrowers who are new to the system and lack sufficient historical data to assess their risk profile.

Low Risk: Borrowers who have exhibited consistent repayment behaviour by not bouncing in the last 6 months.

Medium Risk: Borrowers who have experienced occasional bounces, occurring less than twice in the last 6 months, with no bounce recorded in the most recent month.

High Risk: Borrowers who do not fall into the above categories and thus pose a higher risk of default**.**

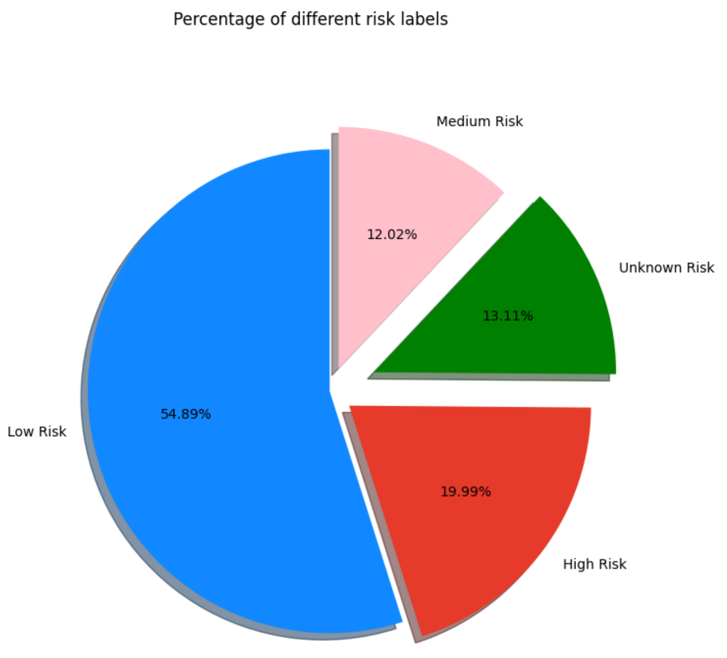
Risk Segmentation Summary:

Summary Statistics:

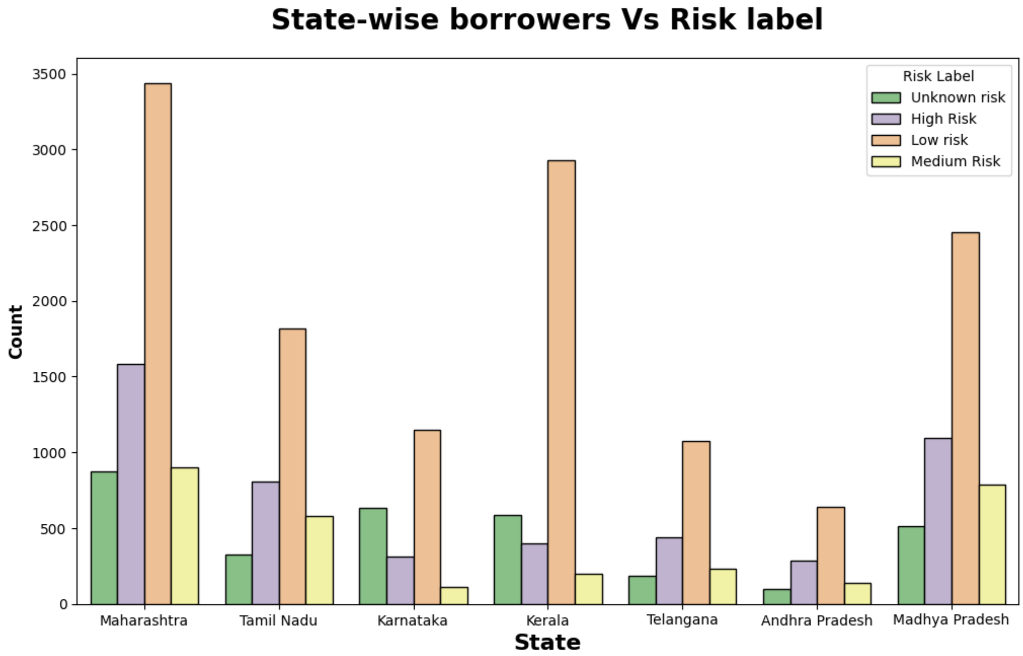
Title	Number
Total Number of Borrowers	24582
Number of Low-Risk Borrowers	13492
Number of High-Risk Borrowers	4913
Number of Medium-Risk Borrowers	2955
Number of New Customers (Unknown-Risk)	3222

Visualization:

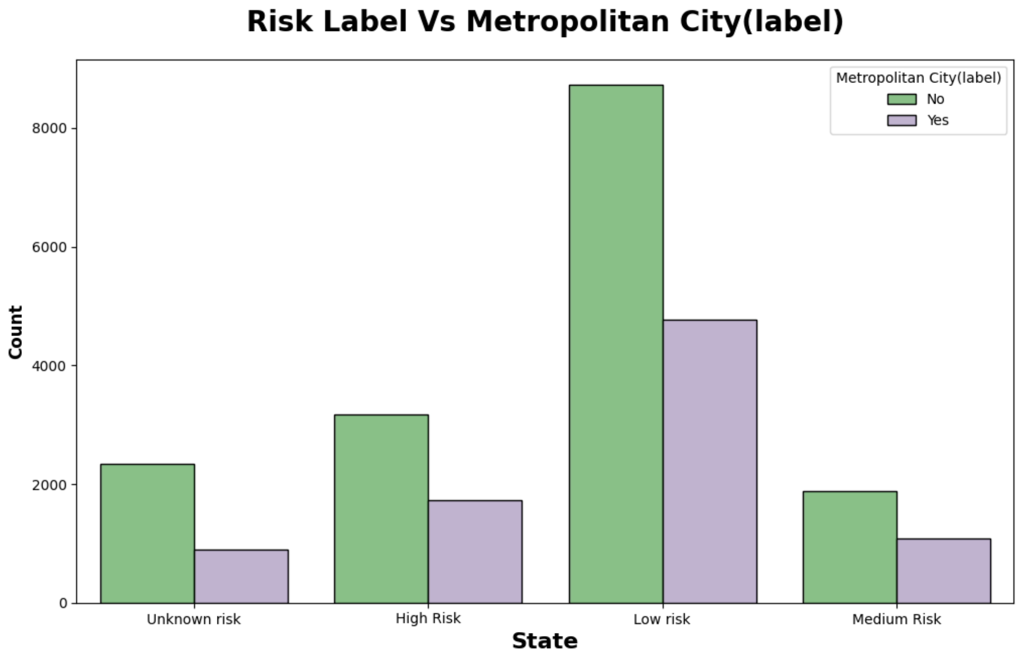
1. **Percentage of different borrowers Vs Risk label**



2. **State-wise borrowers Vs Risk label**



3. **Risk label Vs Metropolitan City(label)**



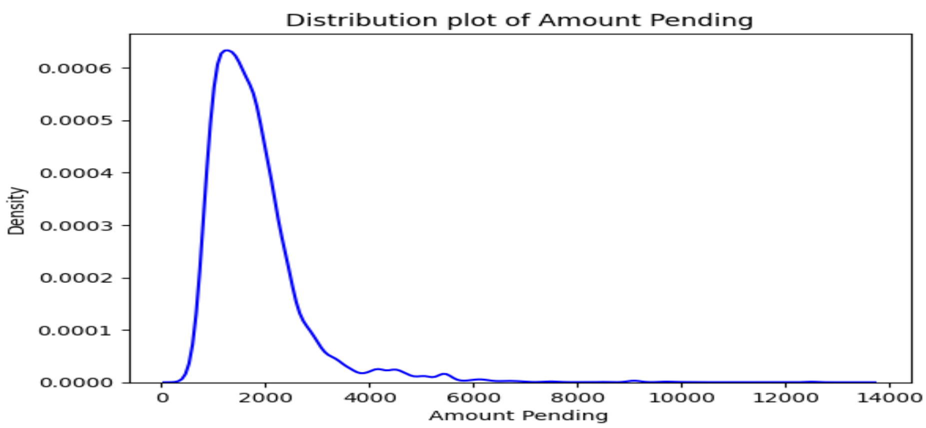
Insights:

1. The pie chart suggests that Low-risk borrowers make up about 54.9% of total borrowers, High-risk borrowers are slightly higher than medium and unknown-risk borrowers with 19.99% and medium and unknown-risk borrowers have 12.02% and 13.11% respectively.
2. The second-hued bar chart suggests that most of the loans are distributed in Maharashtra, Kerela and Madhya Pradesh, and every state data follows the same pattern where we have more low-risk borrowers. In Karnataka, we see a trend of new customers as there are more unknown-risk borrowers.
3. The third-hued bar chart suggests that most of the loans are distributed in non-metropolitan cities and all risk labels share the same distribution.

- ## Ticket Size Segmentation:

Criteria for segmentation:

For segmentation of ticket size, we checked the distribution curve of the amount pending and the mean of the amount pending (1791.17), by seeing the distribution and mean we can say that the distribution is slightly right skewed and based on this we decided on the following criteria.



Low ticket size: If the borrower has an amount pending less than 0.9*(mean of amount pending column) then it will be considered a low-ticket size borrower.

High ticket size: If the borrower has an amount pending greater than 1.28*(mean of amount pending column) then it will be considered a high-ticket size borrower.

Medium ticket size: If the borrower does not fall in the above two categories, then it will be considered a medium-ticket size borrower.

Ticket-size Segmentation Summary:

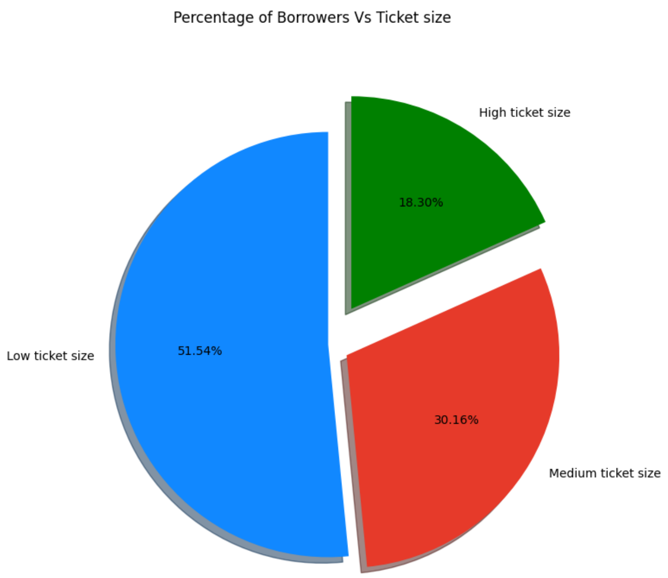
Summary Statistics:

Title	Number
Sum of the pending amount	44030607
Sum of low-ticket size pending amount	15272848
Sum of medium-ticket size pending amount	14211011
Sum of high-ticket size pending amount	14546748

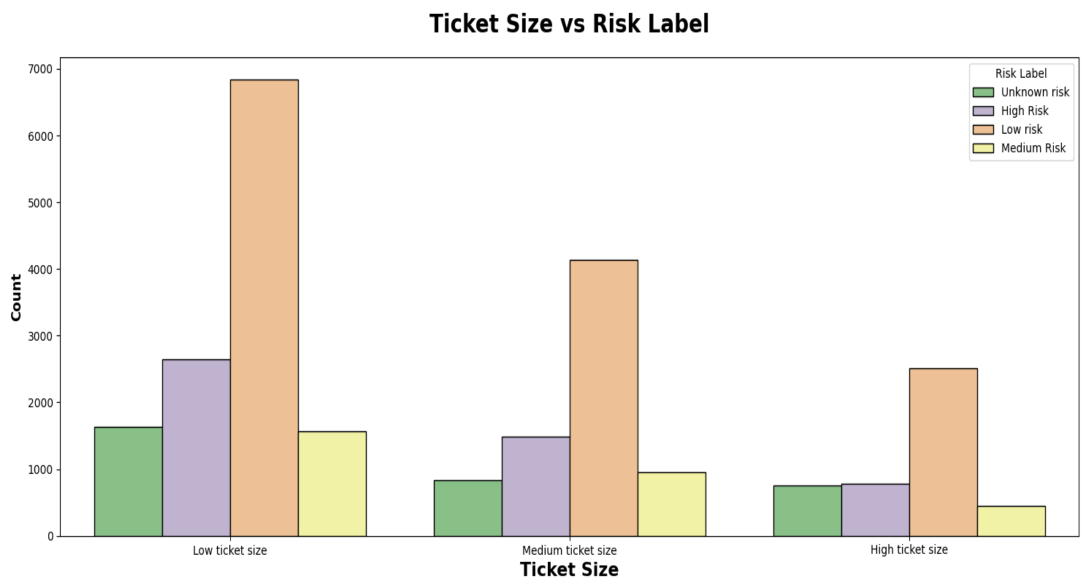
All ticket sizes have approximately equal sum as defined in the problem statement.

Visualization:

1. **Percentage of borrowers Vs Ticket size**



2. **Ticket size Vs Risk label**



Insights:

1. The pie chart suggests that there are more low-ticket borrowers which consist of 51.54% and medium and high-ticket size has 30.16% and 18.30% of borrowers respectively.

2. The second-hued chart suggests that there are more low-ticket and low-risk borrowers this is kind of a good sign and we need to focus on borrowers who have high-ticket and are at high risk and all ticket sizes follow the same kind of distribution.

- **Tenure segmentation:**

Criteria for segmentation:

We categorize borrowers into three tenure segments based on the duration of their loan tenure:

Early Tenure: Borrowers who are in the initial phase of their loan tenure, specifically those who have been in the book for 3 months.
Late Tenure: Borrowers who are nearing the end of their loan tenure, precisely those who are 3 months away from closing the loan.
Mid Tenure: All other borrowers who do not fall into the above categories, indicating they are in the intermediate phase of their loan tenure.

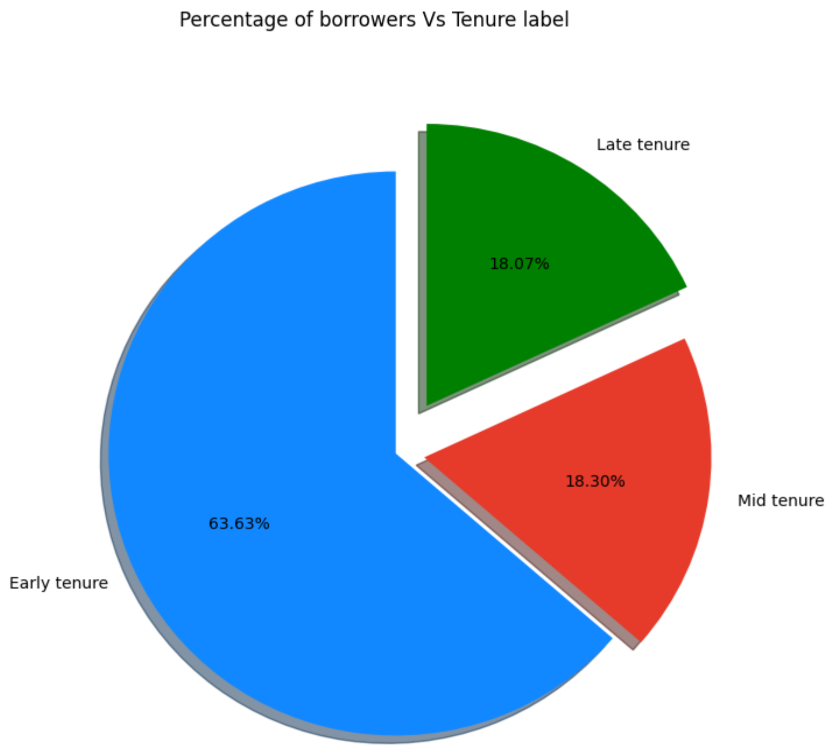
Tenure Segmentation Summary:

Summary Statistics:

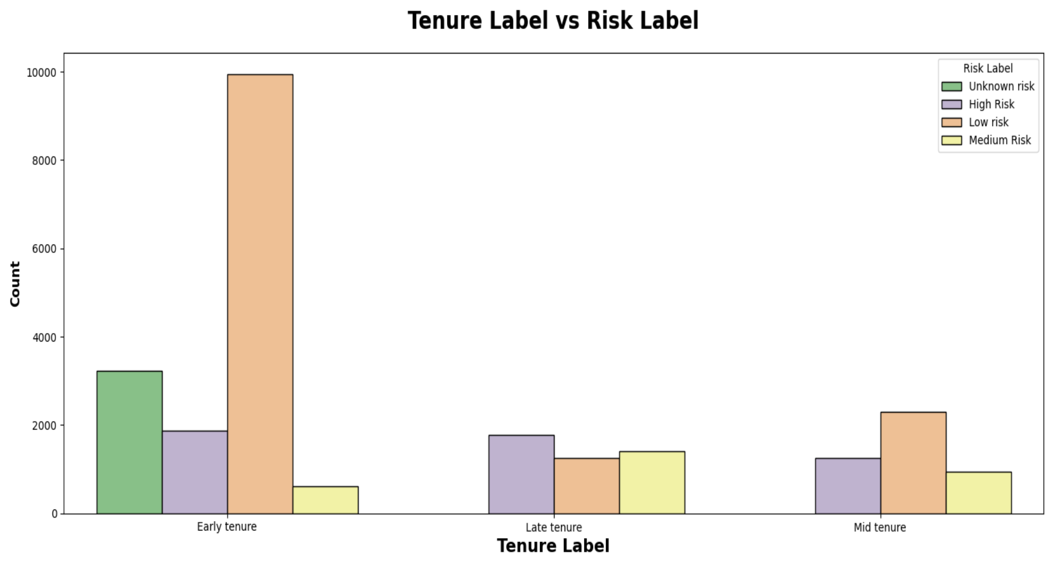
Title	Number
Total number of borrowers	24582
Number of borrowers in Early Tenure	15641
Number of borrowers in Mid Tenure	4498
Number of borrowers in Late Tenure	4443

Visualization:

1. **Percentage of borrowers Vs Tenure label**



2. Tenure Label Vs Risk Label



Insights:

1. The pie chart suggests that there are more Early-tenure borrowers which consist of 63.63% and Mid and Late-tenure

have approximately equal percentages of 18.3% and 18.07% respectively.

2. The second-hued chart shows that there are more early-tenure low-risk loans and as time passes there is more chance that borrowers may turn into high-risk borrowers as we can see there are more high-risk borrowers in late tenure.

- ## Channel-Recommendation:

Criteria for segmentation:

We used the scoring method to categorize the channel recommendation so the repayment is higher and the cost is minimised. The score is assigned to each borrower based on risk label, ticket size, tenure status, metropolitan city or not and new customers.

```
scores = {
    'Low risk' : 1,
    'Unknown risk' : 0,
    'Medium Risk' : 2,
    'High Risk' : 3,
    'Low ticket size' : 1,
    'Medium ticket size' : 2,
    'High ticket size' : 3,
    'Early tenure' : 1,
    'Mid tenure' : 2,
    'Late tenure' : 3
}
```

This is the Python dictionary we used for the score assignment. For example, if a borrower is low-risk, medium-ticket-sized and in early tenure then the score is 1+2+1 = 4, then the borrower will be targeted using Voice bot.

WhatsApp Bot: If the score is less than or equal to 3 then the borrower will be targeted using WhatsApp Bot.

Voice Bot: If the score is less than or equal to 7 or belongs to the metropolitan city then the borrower will be targeted using Voice Bot.

Human calling: If the borrower does not fall into the above two categories, then it will be targeted as Human calling.

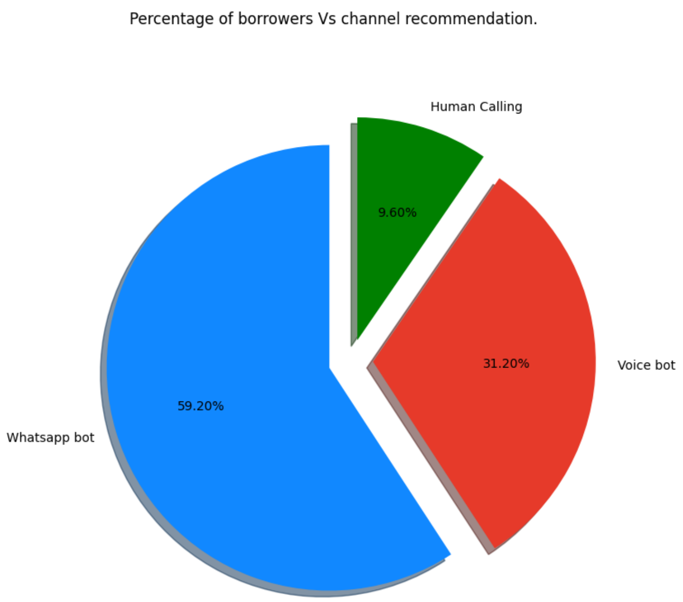
Recommendation Summary:

Summary Statistics:

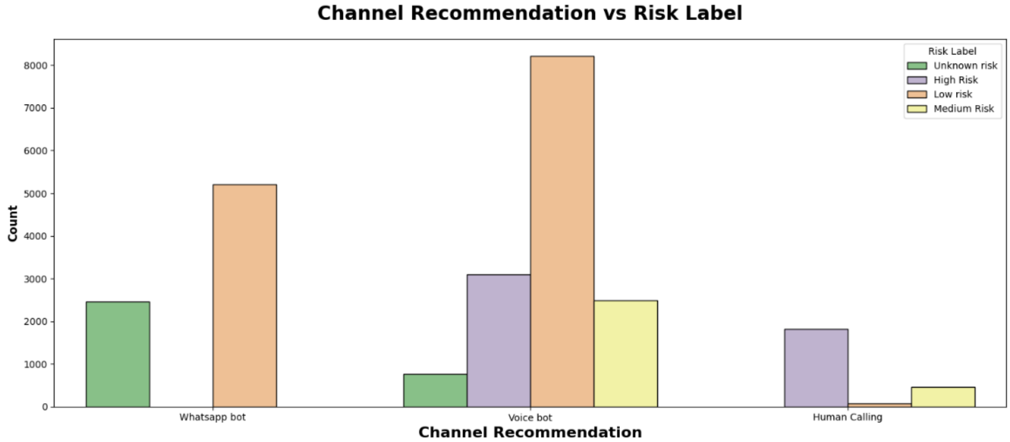
Title	Number
No. of borrowers segmented for WhatsApp Bot	14553
No. of borrowers segmented for Voice Bot	7670
No. of borrowers segmented for Human calling	2359

Visualization:

1. **Percentage of borrowers Vs channel recommendation.**



2. Channel Recommendation Vs Risk Label



Insights:

1. The pie chart above suggests that the categorization is correct up to our expectations as 59.2% of borrowers targeted using WhatsApp Bot, 31.20% using Voice Bot and 9.6% using Human Calling. This method minimised our cost to 301830 INR which is 0.06% of my total distributed amount.
2. The second-hued chart shows that no high and medium-risk borrowers were targeted by the WhatsApp bot and most of the high-risk borrowers were targeted using Human calling.