Bank Looking For Clients

Explanatory Analysis of a dataset with 5000 customers
to get an idea how to sell a new product in a more efficient way.

What is the case?

The case is The Bank has a customers Data with various characteristics of the customers. The management built a new product - Personal Loan, and ran a small campaign towards selling the New Product to their clients.

After some time, 9% of customers have Personal Loan from The Bank.

What is the Goal?

- To sell more Personal Loan products to Bank customers.
- To devise campaigns to better target marketing to increase the success ratio with a minimal budget.
- To identify the potential customers who have a higher probability of purchasing the loan.

What are the Questions?

As soon as we got 9% of customers who bought the Product, we got the following questions:

- Is there some associations between personal characteristics and the fact that customer bought the Product? If so:
- What are those Main Characteristics that have an association with the Product and what is the strength of the association?
- What are the Segments of Main Characteristics, that have a higher strength of association with the Product?
- What is the sample of Data with customers from Main Segments?

Explanatory analysis

We made a simple step-by-step analysis of customers' characteristics to identify patterns to effectively choose the subset of customers who have a higher probability of buying a new product.

We performed the following steps:

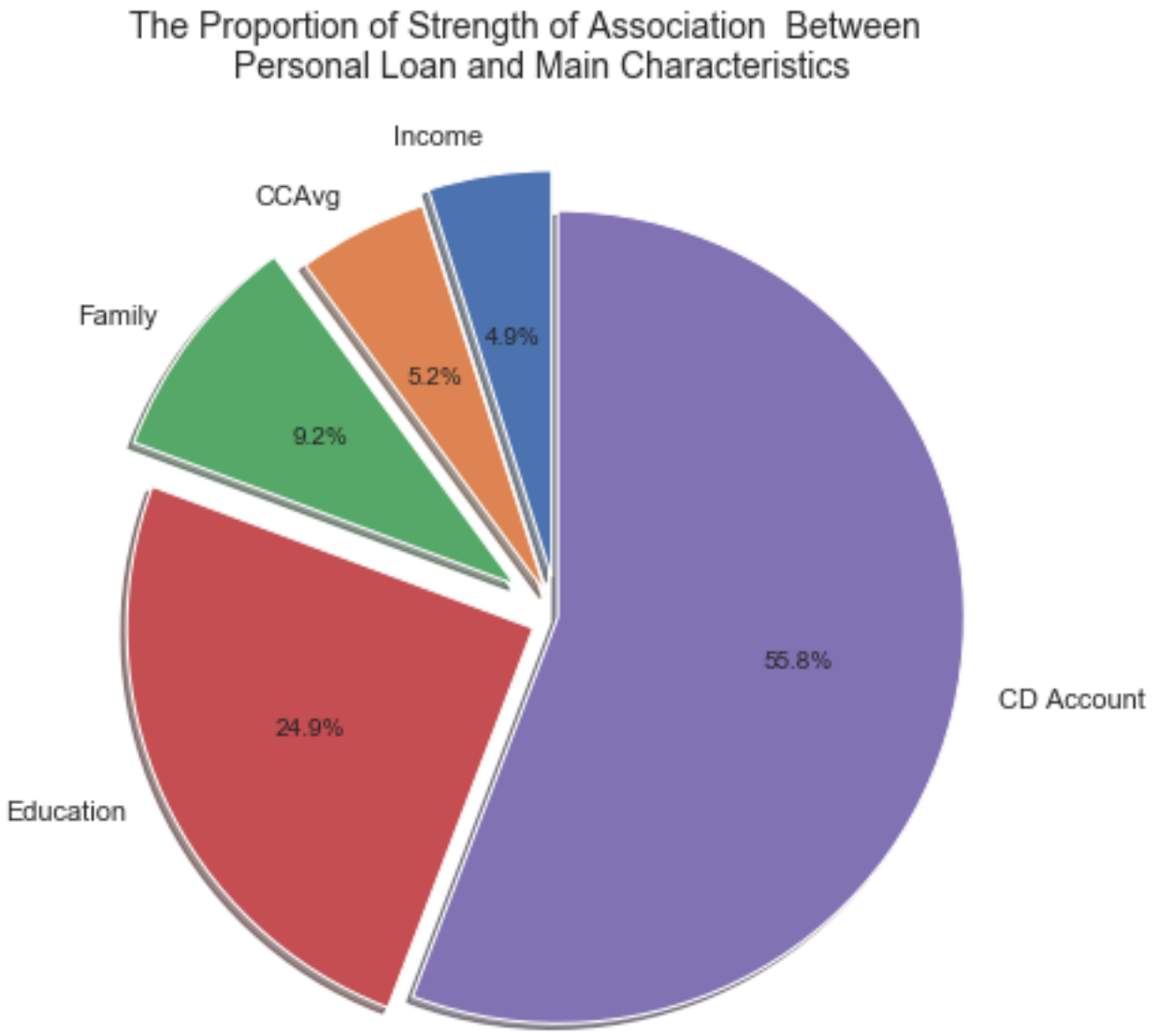
- Checked all twelve characteristics whether or not each of them has an association with the fact that the product has been sold.
- Found FIVE main characteristics that have higher than moderate strength of association with the product.
- Analyzed main characteristics and segmented each one by different strengths of association with the product.
- Tried to make a subset of customers with ideal characteristics who have the highest probability of buying the product. Unfortunately, our dataset does not contain such information. So...
- We build a simple algorithm to make a subset of data to get the customers' IDs who have a high probability to buy the product.

Strength of association

- We ran logistic regression models through the whole dataset and found FIVE Main Characteristics.
- Each of these Main Characteristics have **highest odds** to increase the chance to sell the product with an increased value of characteristic by one unit, leaving all other characteristics the same.
- We named the size of odds as strength of association.

Five main characteristics

NAME	ODDS
CD Account	12.02
Education	5.35
Family size	1.99
CCAvg	1.11
Income	1.05



What are the Segments of Main Characteristics?

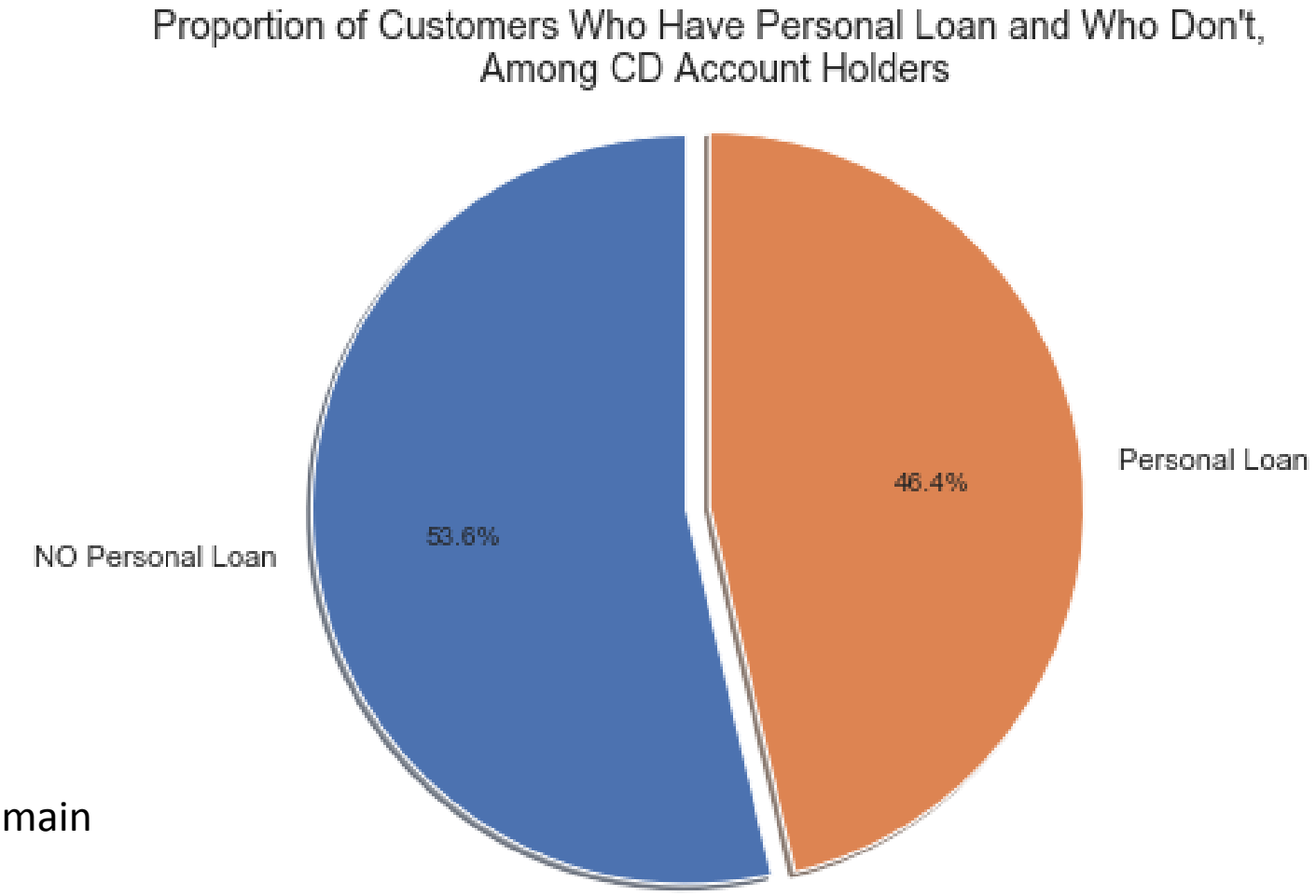
For each characteristic, we ran a step-by-step analysis to identify the segments with the highest strengths of association.

CD Account

We can say that the proportion of customers who have Personal Loan among them and have CD account is quite high.

Let's get a closer look…

CD Account

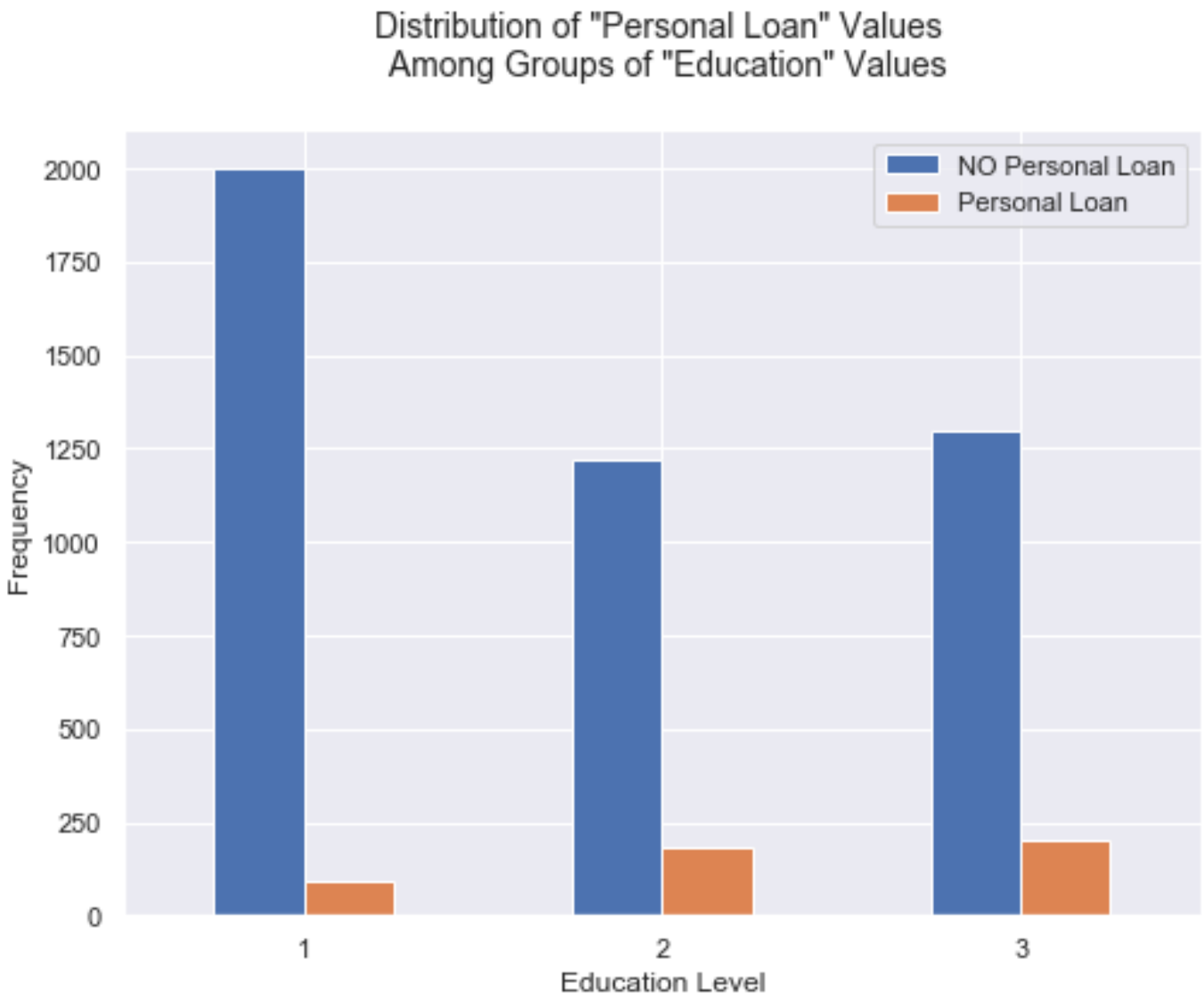


Conclusion

- 46.4% of CD Account Holders have Personal Loan.
- For 'CD Account' characteristic - the main segment to sell Personal Loan is the customers who already have a CD.
- Target value of 'CD Account' variable = 1

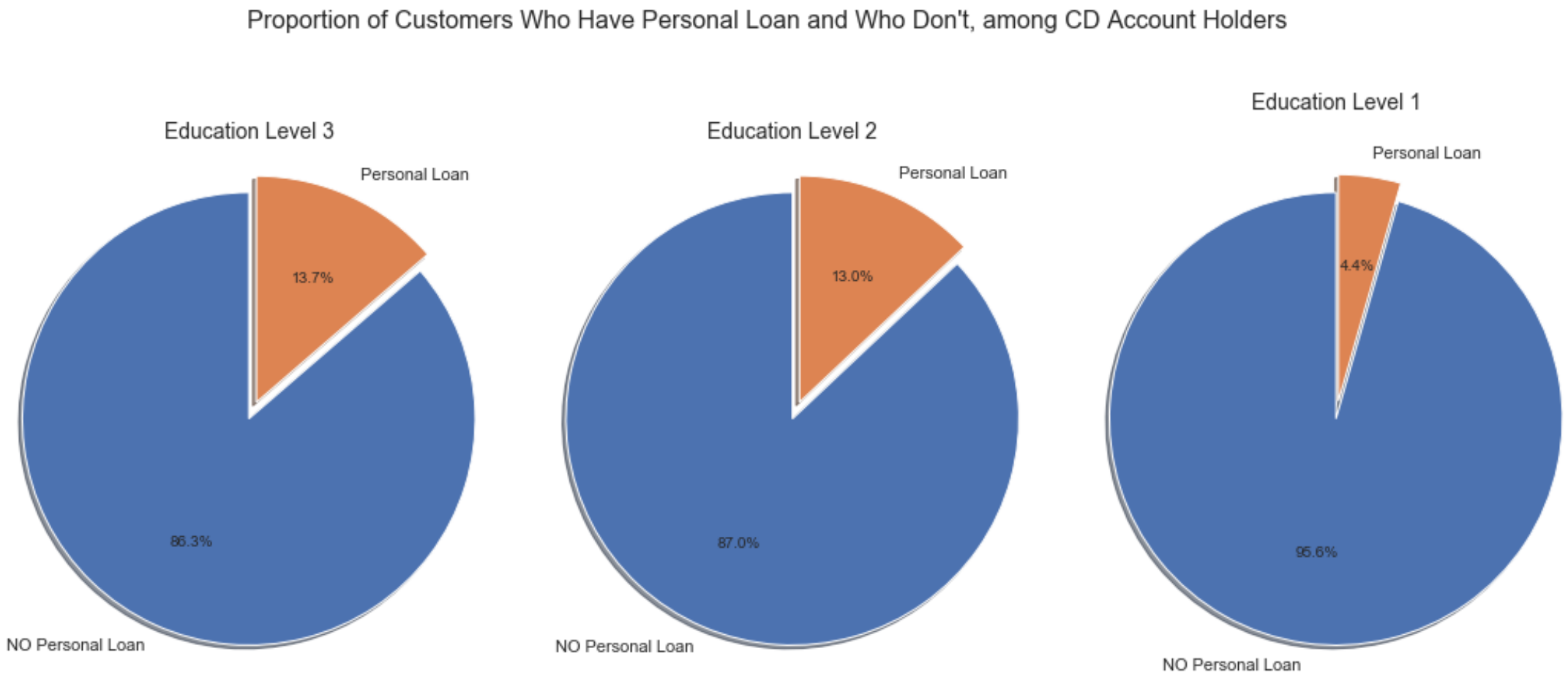
Education

We can say that the proportion of customers who have Personal Loan to those who do not among customers with 2 and 3 Levels of Education **is higher** than the proportion among customers who have 1 level of Education.



Let's get a closer look…

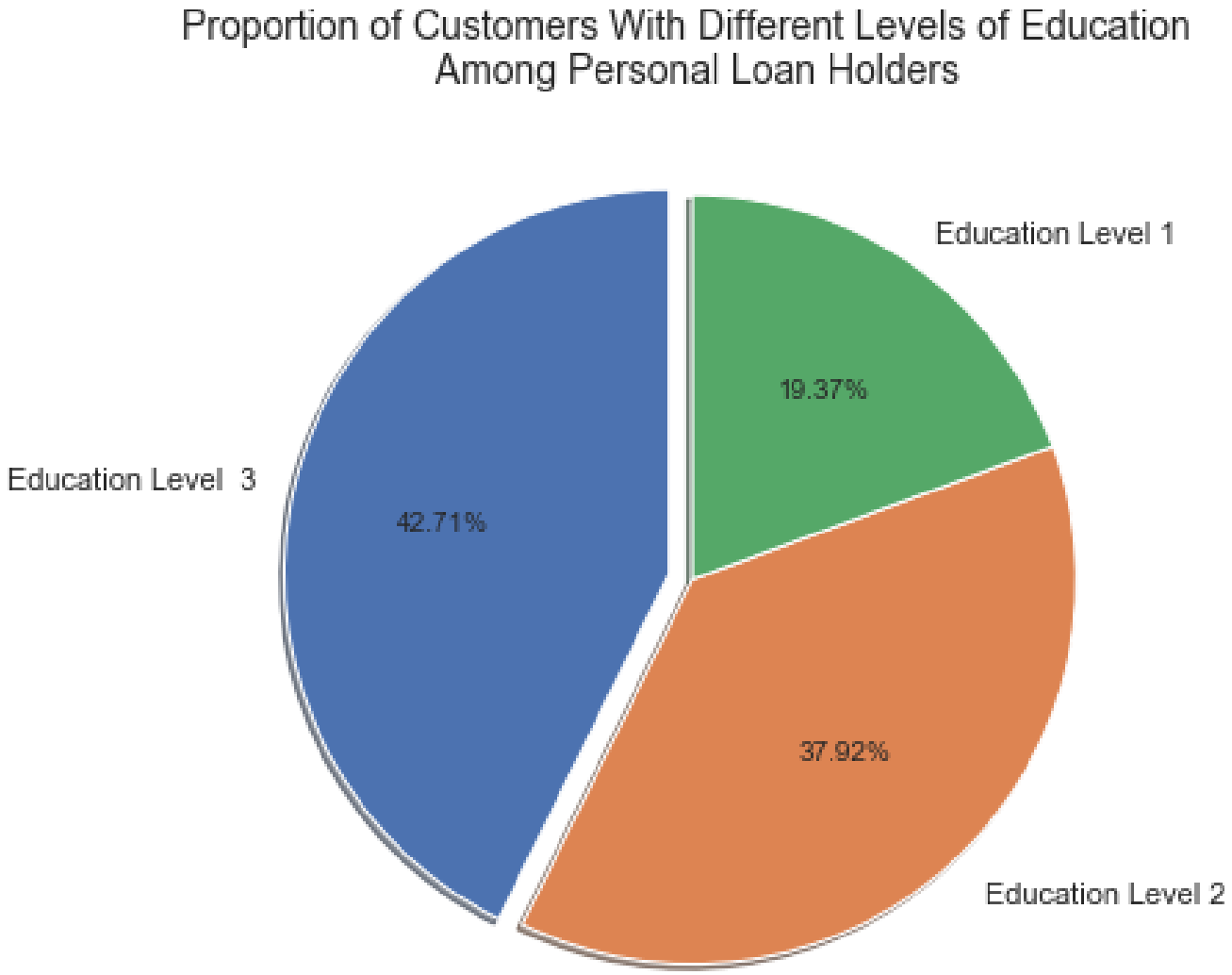
Education


Proportion of Customers Who Have Personal Loan and Who Don't, among CD Account Holders
Education Level 3
Personal Loan
13.7%
86.3%
NO Personal Loan
Education Level 2
Personal Loan
13.0%
87.0%
NO Personal Loan
Education Level 1
Personal Loan
4.4%
95.6%
NO Personal Loan

Education

Conclusion

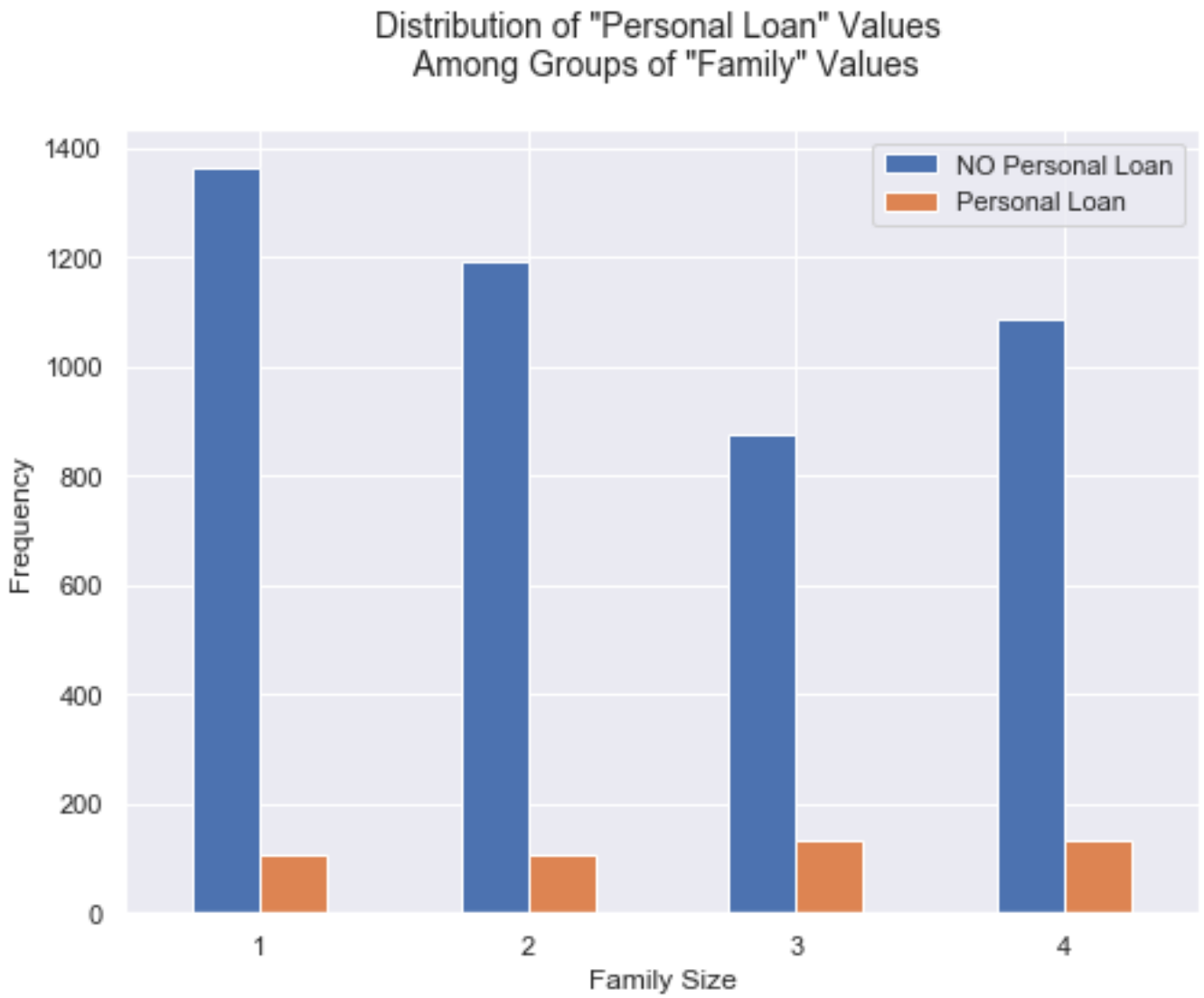
- 42.7% and 37.9% of customers who have Personal Loan, have Education Level 3 and Level 2 respectively.
- For 'Education' characteristic - the main segments to sell Personal Loan are the customers who have 2nd and 3rd levels of education.
- Target values of 'Education' variable are 3 and 2 in descending order of priority.



Family

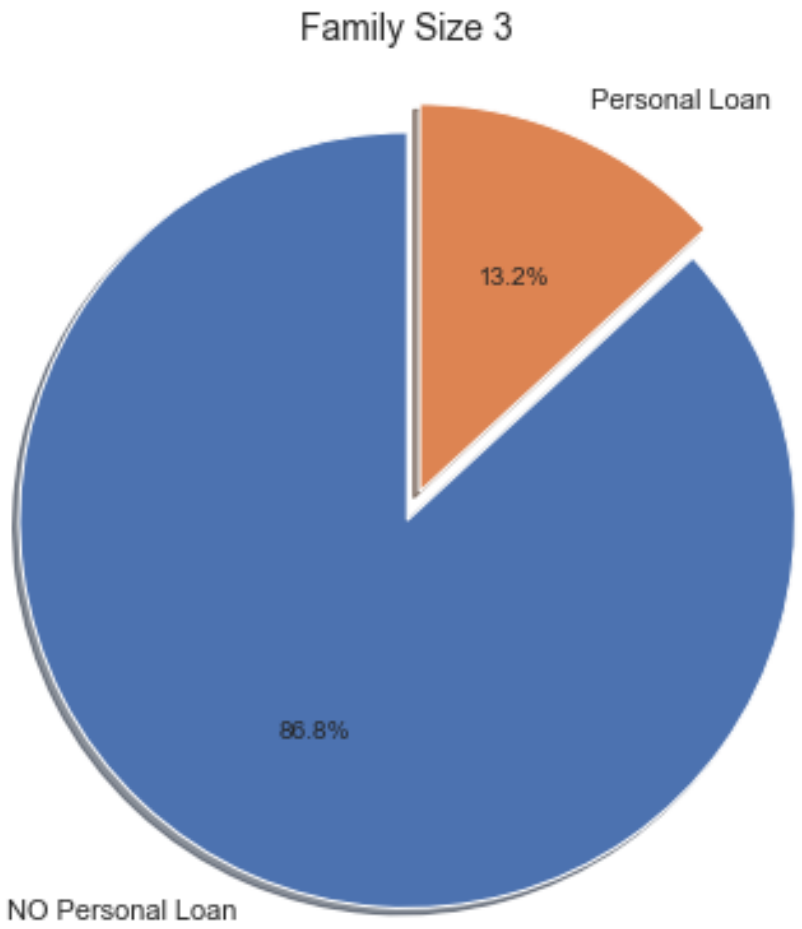
We can say that the proportion of customers who have Personal Loan among them who have Family size 2 and 3 is the highest proportion.

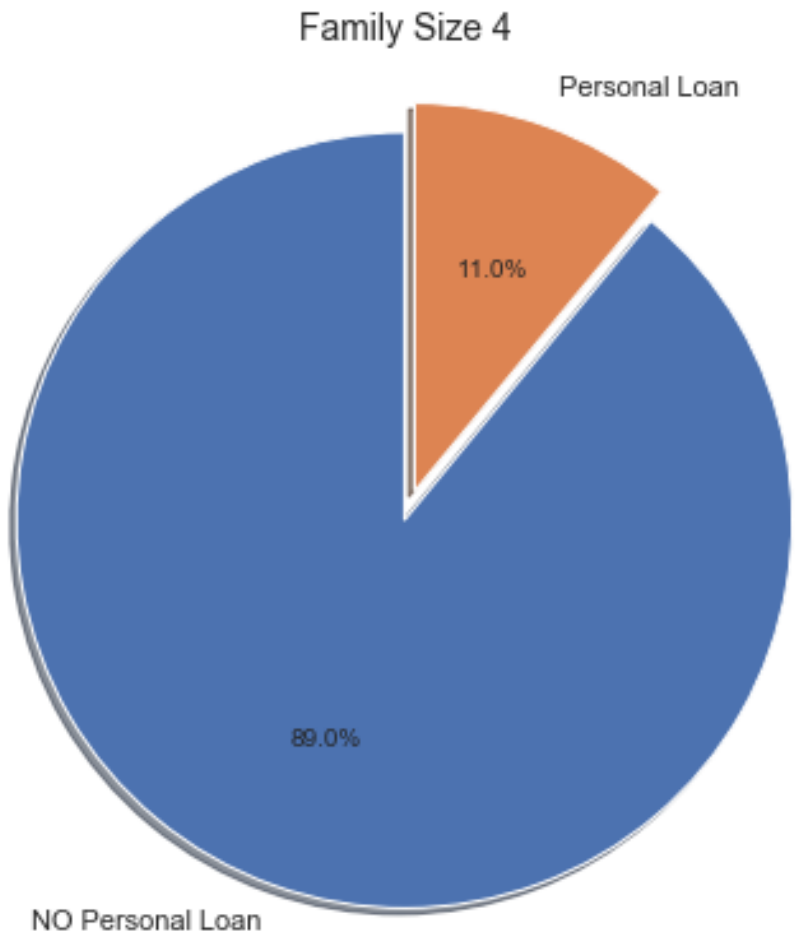
Let's get a closer look…



Family

Proportion of Customers Who Have Personal Loan and Who Don't, among Different Family Sizes

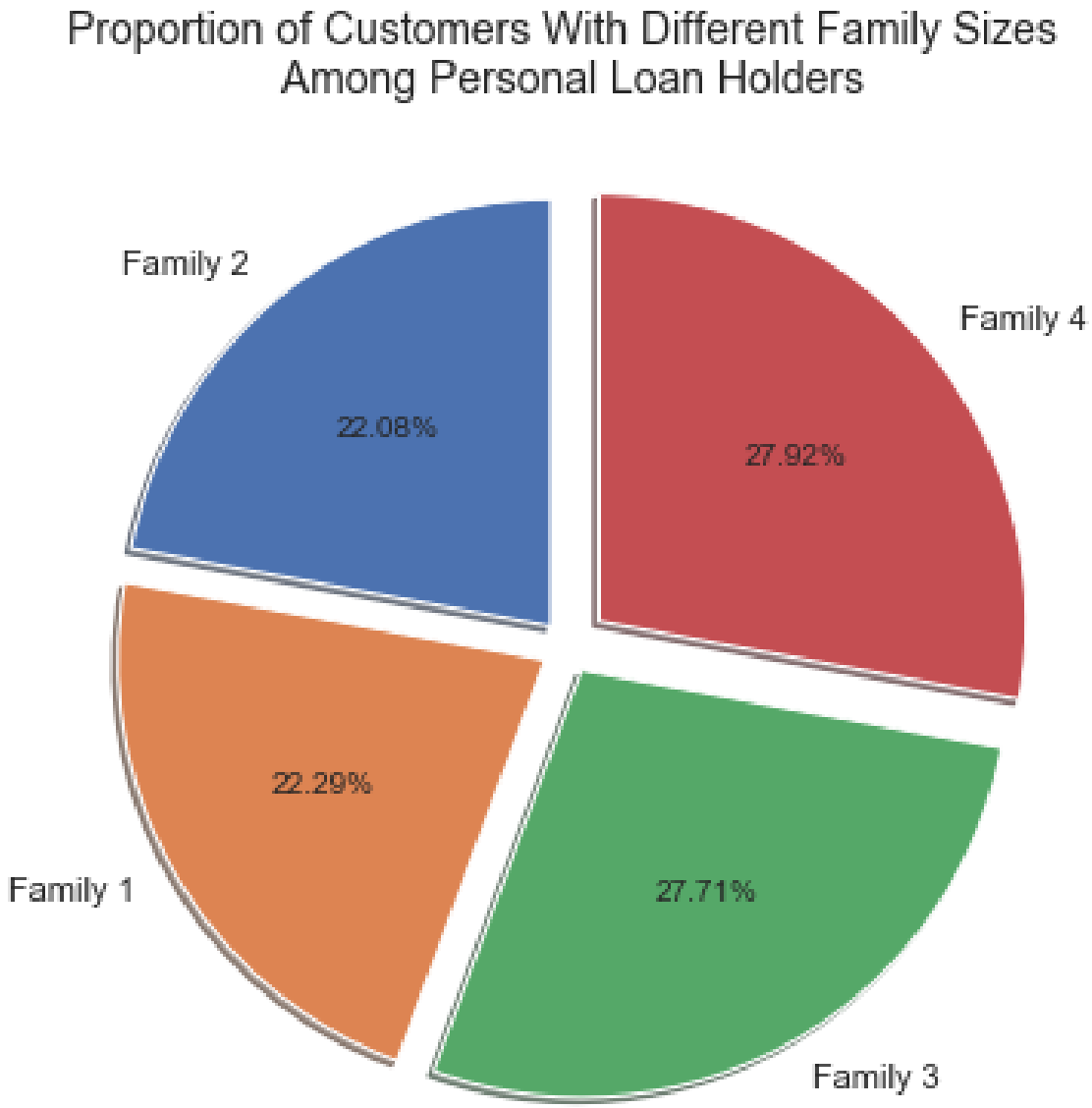




Family

Conclusion

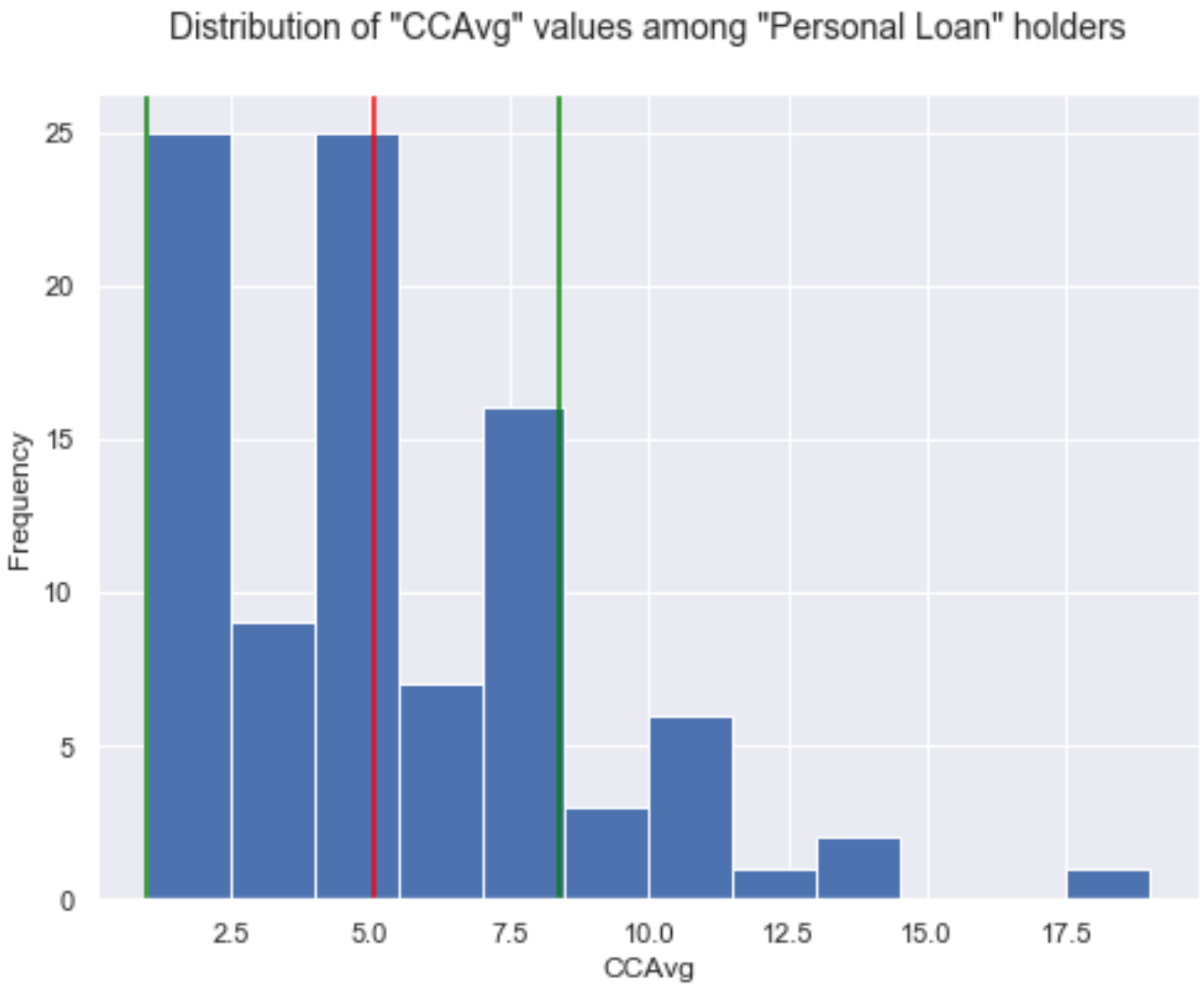
- 27.9% and 27.7% of customers who have Personal Loan, have Family size Level 4 and Level 3 respectively.
- For 'Family' characteristic, the main segments to sell Personal Loan are the customers who have Family Size 3 and 4.
- Target values of 'Family' variable are 3 and 4 in descending order of priority, since the proportion of customers who have Personal Loan is the highest with Family Size 3 (13,2%).



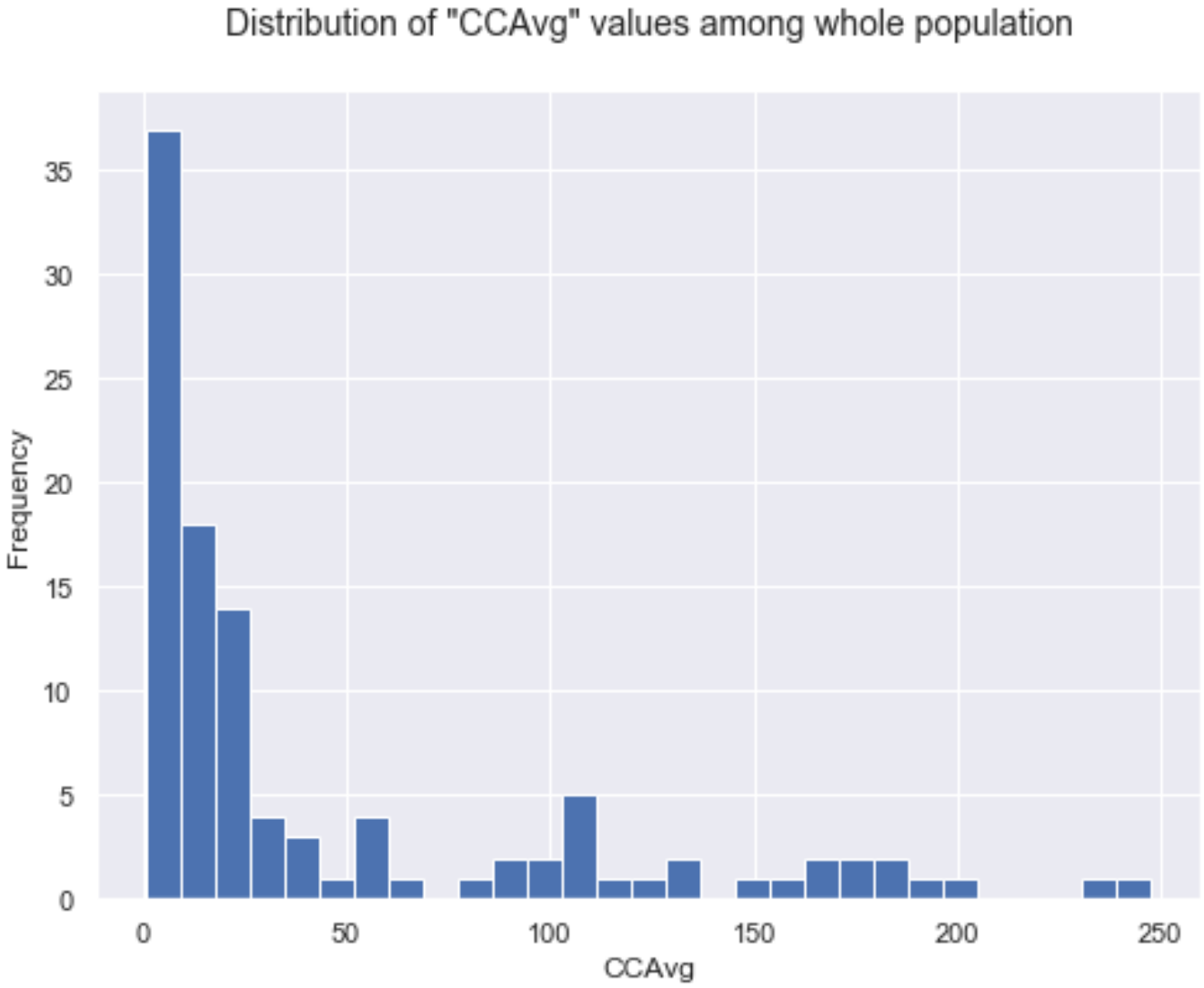
CCAvg

We can say that CCAvg characteristic values can be divided in three groups in descending order of priority consider its frequency among Personal Loan holders:

- Group I: 1 < CCAvg < 2.5
- Group II: 4 < CCAvg < 5.5
- Group III: 7 < CCAvg < 8.5



CCAvg



We can say that all of our groups of 'CCAvg' defined as priority groups to sell Personal Loan, lie inside segment with highest frequency among whole population.

CCAvg Conclusion

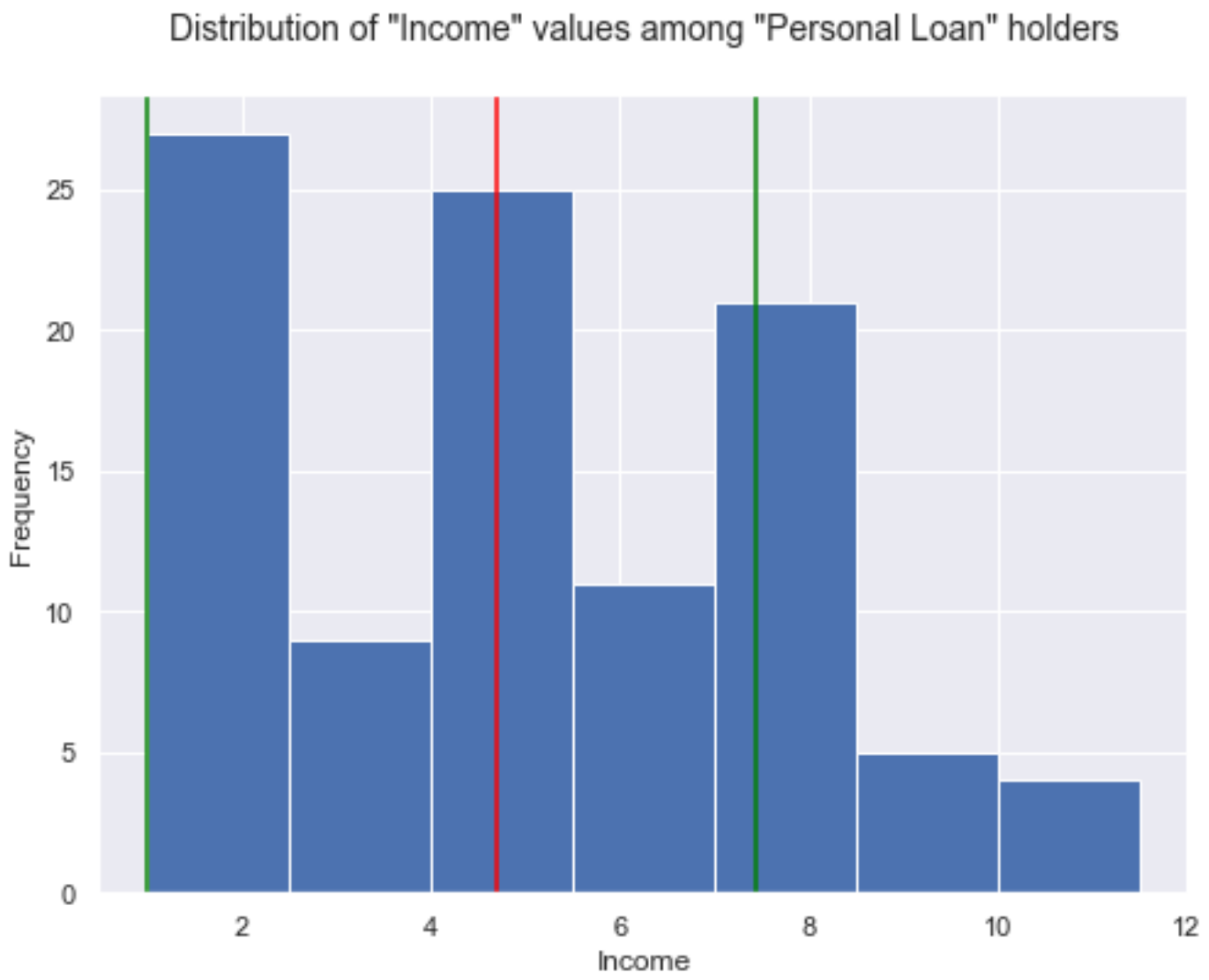
Target groups of 'CCAvg' characteristic in descending order of priority:

- Group I: 1 < CCAvg < 2.5
- Group II: 4 < CCAvg < 5.5
- Group III: 7 < CCAvg < 8.5

Income

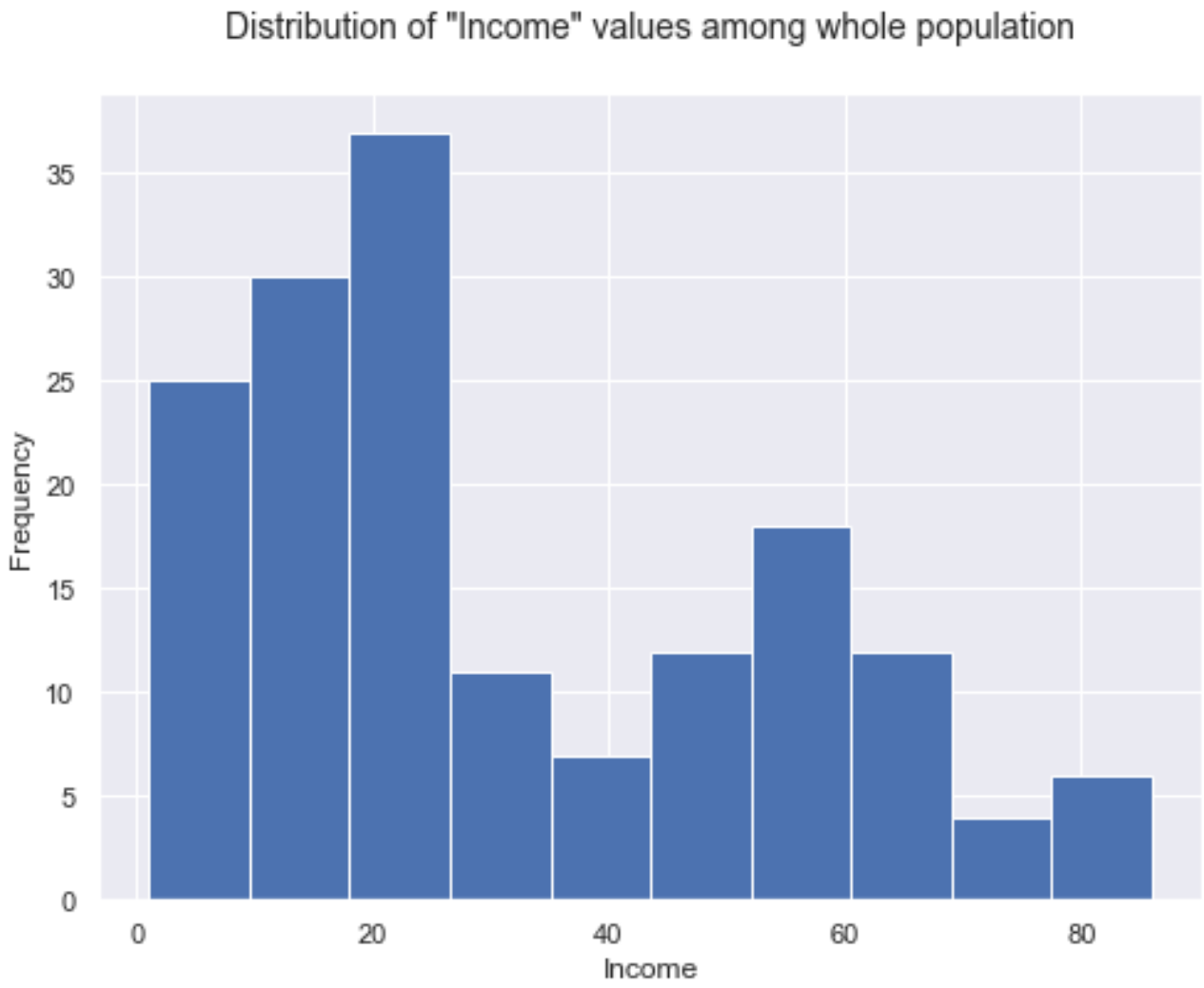
We can say that 'Income' characteristic values can be divided in three groups in descending order of priority considering its frequency among Personal Loan holders:

- Group I: 1 < Income < 2.5
- Group II: 4 < Income < 5.5
- Group III: 7 < Income < 8.5



Income

We can say, that all of our groups of 'Income' defined as priority groups to sell Personal Loan, lie inside a segment with pretty high frequency among whole population.



Income Conclusion

Target groups of 'Income' characteristic in descending order of priority:

- Group I: 1 < Income < 2.5
- Group II: 4 < Income < 5.5
- Group III: 7 < Income < 8.5

What is the sample of data with customers from Main Segments?

We found associations of the fact that a customer bought the Product and some of his/her characteristics.

We segmented each Main Characteristic with a descent order of strength of association.

With this data we are able to get subset with any combinations to get customers' IDs who have the highest probability to buy the product.

…use file "00_main.ipynb"

CD Account

- 1

Education

- 3
- 2

Family

- 3
- 4

CCAvg

- Group I: 1 < CCAvg < 2.5
- Group II: 4 < CCAvg < 5.5
- Group III: 7 < CCAvg < 8.5

Income

- Group I: 1 < CCAvg < 2.5
- Group II: 4 < CCAvg < 5.5
- Group III: 7 < CCAvg < 8.5