**Introduction:**
The motive is to understand the factors that influence the repayment of loans by the clients, the degree to which these factors play a role and to make sure that the customers who are capable of paying their loans on time are not rejected by the bank.

**Details of the dataset:**
We have two seperate datasets which are classified as follows:
1. Information regarding the current application of the client. It includes variables as to where the client lives, what is the credit rating, the number of family members, the income of the client, the amount that the client has asked for as a loan from the bank as so on. This data is partly given by the client during the application process and partly sourced from third parties who possess expertise in client crediit rating.

2. Information regarding the previous loan history of the client. It includes variables such as the status of the previous application, whether it was approved by the bank or not, if yes then the downpayments made, the rate of interest decided upon, the final amount credited to the customer as a loan and so on.

**VARIABLES THAT ARE STRONG INDICATORS OF DEFAULT:**

**I. Occupation of the clients:**
Below is the table of each occupation and the percentage of defaulters i.e clients who have missed at least one payment on time.

***Top 5 Occupations by % defaulters:***

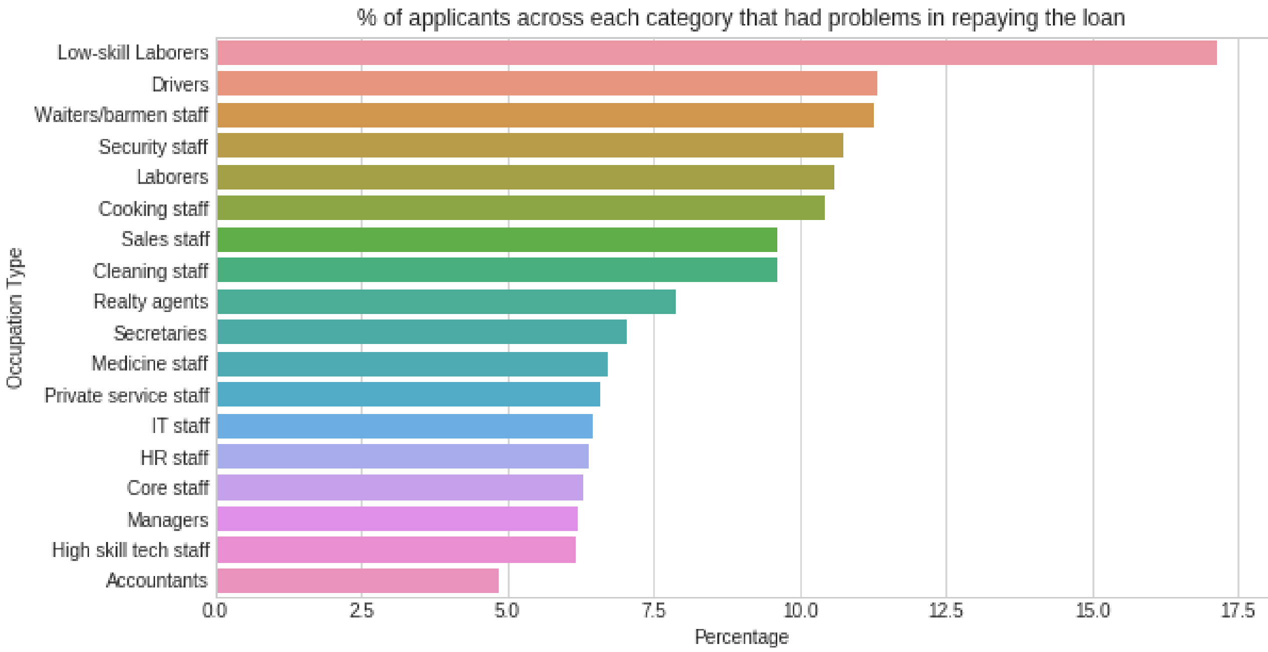
| Total number of clients | Occupation | Defaulter clients | Default percentage |
|---|---|---|---|
| 2093 | Low-skill Laborers | 359 | 17.15 |
| 18603 | Drivers | 2107 | 11.33 |
| 1348 | Waiters/barmen staff | 152 | 11.28 |
| 6721 | Security staff | 722 | 10.74 |
| 55186 | Laborers | 5838 | 10.58 |
| | | **Average %** | **12.22** |

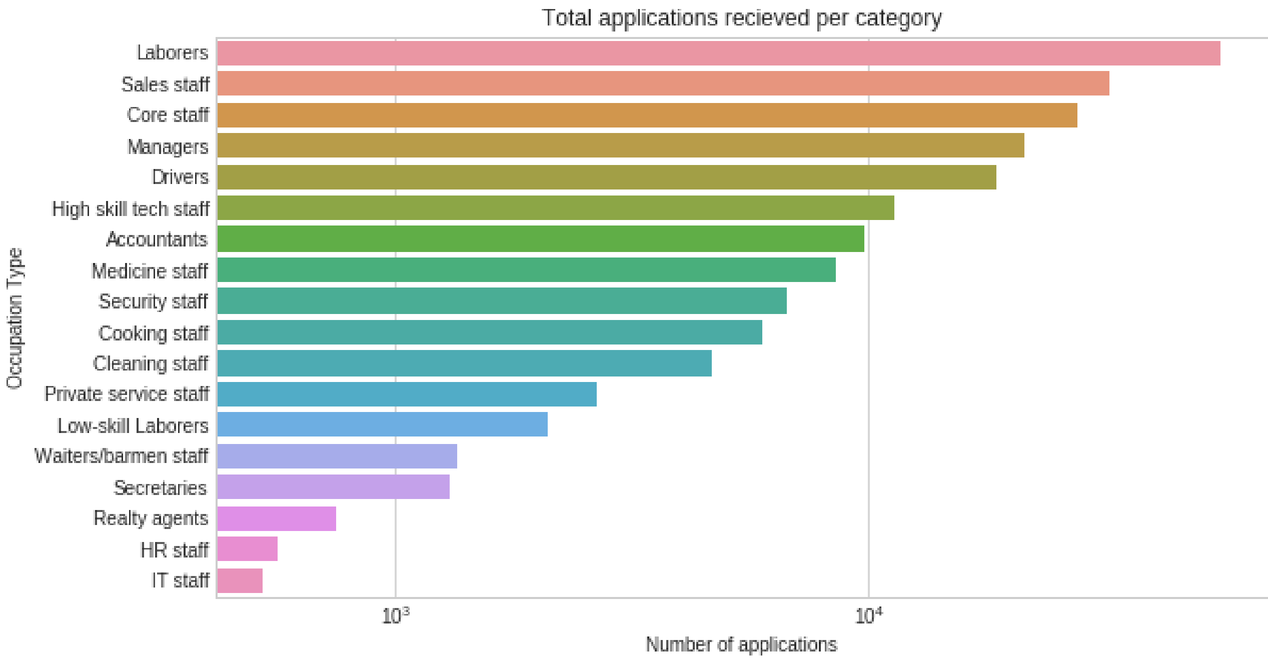
It can be seen that the blue collar workers comprise the majority of the total loan applications and also the rate of default os really high for them whichaverages around 12%.

***Bottom 5 Occupations by % defaulters:***

| Total number of clients | Occupation | Defaulter clients | Default percentage |
|---|---|---|---|
| 563 | HR staff | 36 | 6.39 |
| 27570 | Core staff | 1738 | 6.30 |
| 21371 | Managers | 1328 | 6.21 |
| 11380 | High skill tech staff | 701 | 6.16 |
| 9813 | Accountants | 474 | 4.83 |
| | | **Average %** | **5.98** |

White collar workers  who are highly skilled at their jobs on the other hand seem less likely to default with the average default rate of about 6% which is nearly half that of the blue collar workers.
Therefore it can be inferred that the loan applications of blue collar workers must be highly scrutinized before lending the loan since they are the biggest section of people who are likely to default and thus the bank incurring loss on the application.

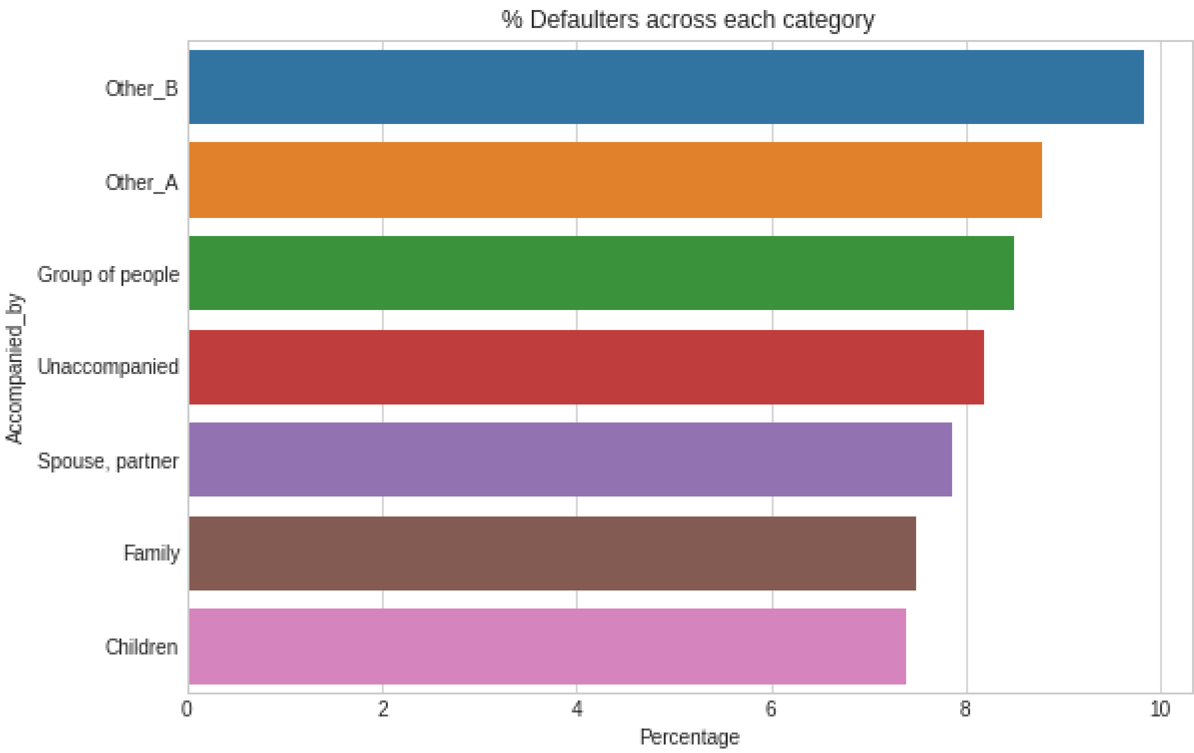



Above is the graphic visual to see the percentage of applicants falling in each category and the percentage of defaulters across the same.

**II. Client's company while applying for the loan:**
Another metric that can be used to filter the potential defaulter clients is the people that they are accompanied by while aplying for the loan.
These are divided into 6 categories:
1. Family
2. Spouse / Partner
3. Children
4. Other_A
5. Other_B
6. Group of people

Below is the percentage of defaulters across each category:




As is evident from the above figure, there are two points that can be inferred from it:
a.) The clients who arrive with their families are least likely to default on their loan payments. Even amongst families, the clients who are accompanied by their kids only default 7% of the time.

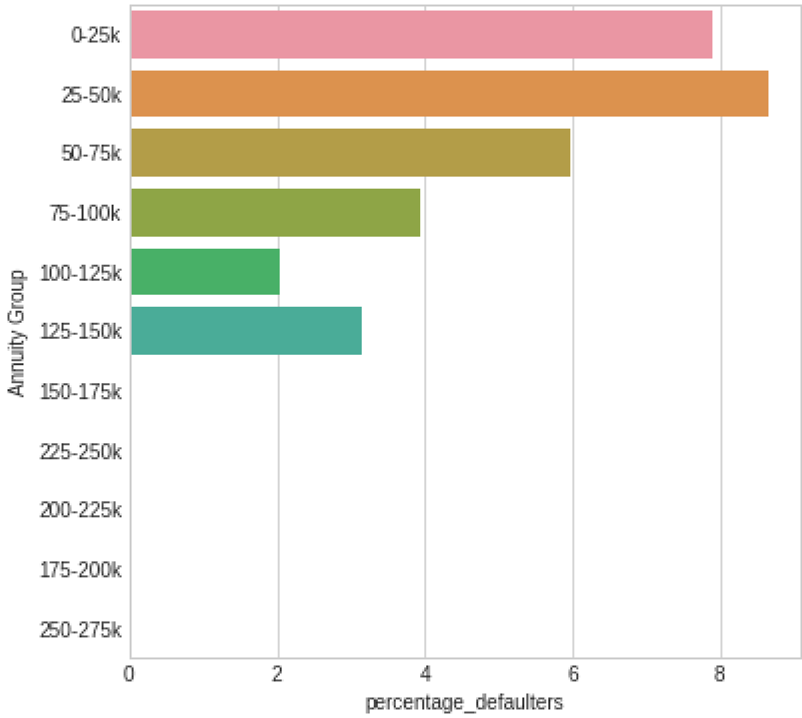
b.) Clients who are accompanied by someone who is not related to them are more likely to default on the payments with the average rate of default being around 9% which is 2 percent higher than their counterparts who are accompanied by their family.

**III. Annuity Amount:**
The total amount of money that the client needs to pay back to the bank every year also has a significant impact on the client defaulting the payments.

For the sake of simplicity, the annuity amount has been divided into chunks of Rs. 25,000. Below is the distribution of applicants across each category:

| Annuity Group | Count_total |
|---|---|
| 0-25k | 154867 |
| 25-50k | 131330 |
| 50-75k | 19326 |
| 75-100k | 1471 |
| 100-125k | 342 |
| 125-150k | 95 |
| 150-175k | 24 |
| 225-250k | 24 |
| 200-225k | 11 |
| 175-200k | 8 |
| 250-275k | 1 |

We can see that the majority of the applicants fall under the Rs 75,000 per annum. Below is the graphic that show the distribution of defaulters across each category.




The clients who default the most have their annuity between Rs 50,000 – 75,000, with their default rate being above 8%. This is closely followed by the applicants whose annuity is less than Rs. 25,000 per annum.
It is interesting to note that the group which defaults the least has annuity between Rs. 1 lakh to Rs 1,25,000. Even at annuities higher than 1.5 lakhs there are literally no defaults.

This phenomena can be attributed to the fact that there are a great deal of blue-collar workers amongst the applicants and that the default rate for them averages around 12%. In addition to that the income of blue collar workers is low and therefore it makes sense that the lower annuity groups tend to default the most.

**IV. Family Size:**

Below is the report showing the top 5 highest percentage default across each category of applicant.

| Family Size | Number of Defaulters | Number of applicants | % Defaulters |
|---|---|---|---|
| 11.0 | 1 | 1 | 100.00 |
| 13.0 | 1 | 1 | 100.00 |
| 10.0 | 1 | 3 | 33.33 |
| 8.0 | 6 | 20 | 30.00 |
| 6.0 | 55 | 408 | 13.48 |
| 5.0 | 327 | 3478 | 9.40 |

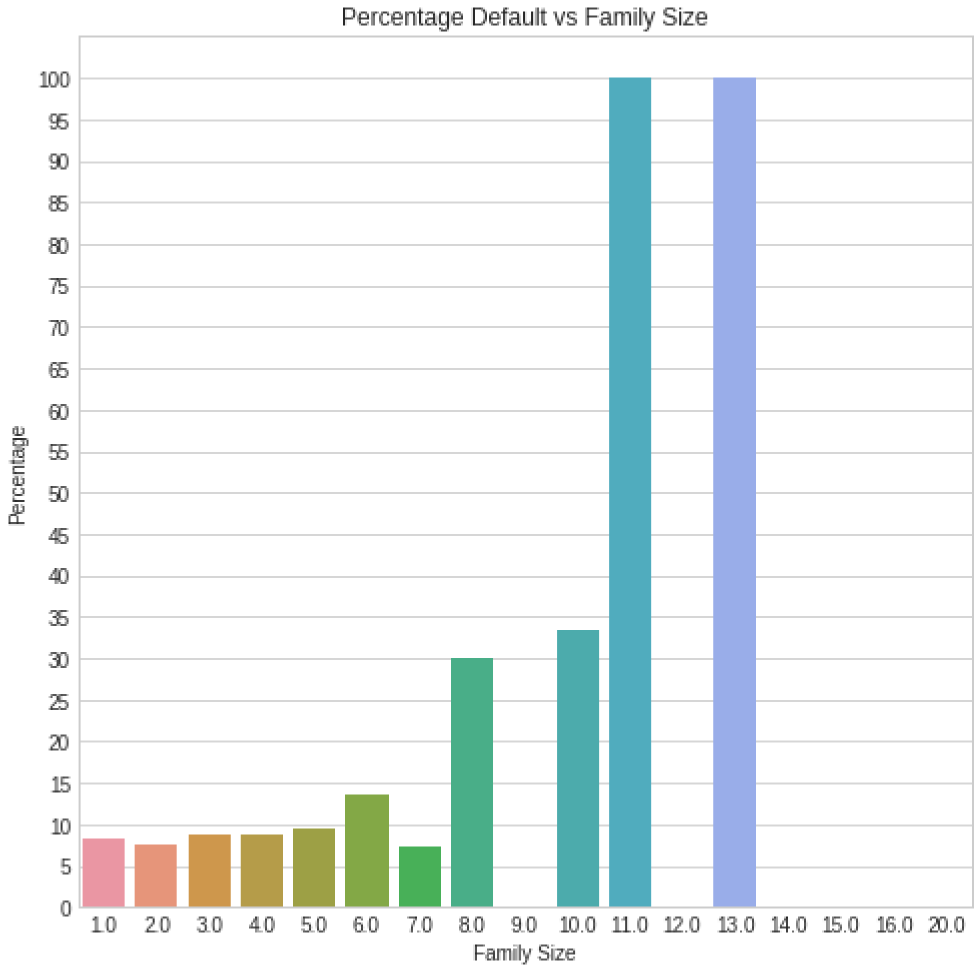
The applicants whose family size is greater than 5 members tend to default the most with one-third of the clients defaulting when the family size is 10.
It is to be noted that we cannot really make any inferences about the clients whose family size is greater than 10 because we simply do not have enough data, however looking at the trend, it is safe to assume that had more data been present, the results would have been in sync.

*Applicants who are least likely to default by family size:*

| No. of members | Number of Defaulters | Number of applicants | % Defaulters |
|---|---|---|---|
| 3.0 | 4608 | 52601 | 8.76 |
| 4.0 | 2136 | 24697 | 8.65 |
| 1.0 | 5675 | 67847 | 8.36 |
| 2.0 | 12009 | 158357 | 7.58 |
| | | **% Average** | **8.34** |

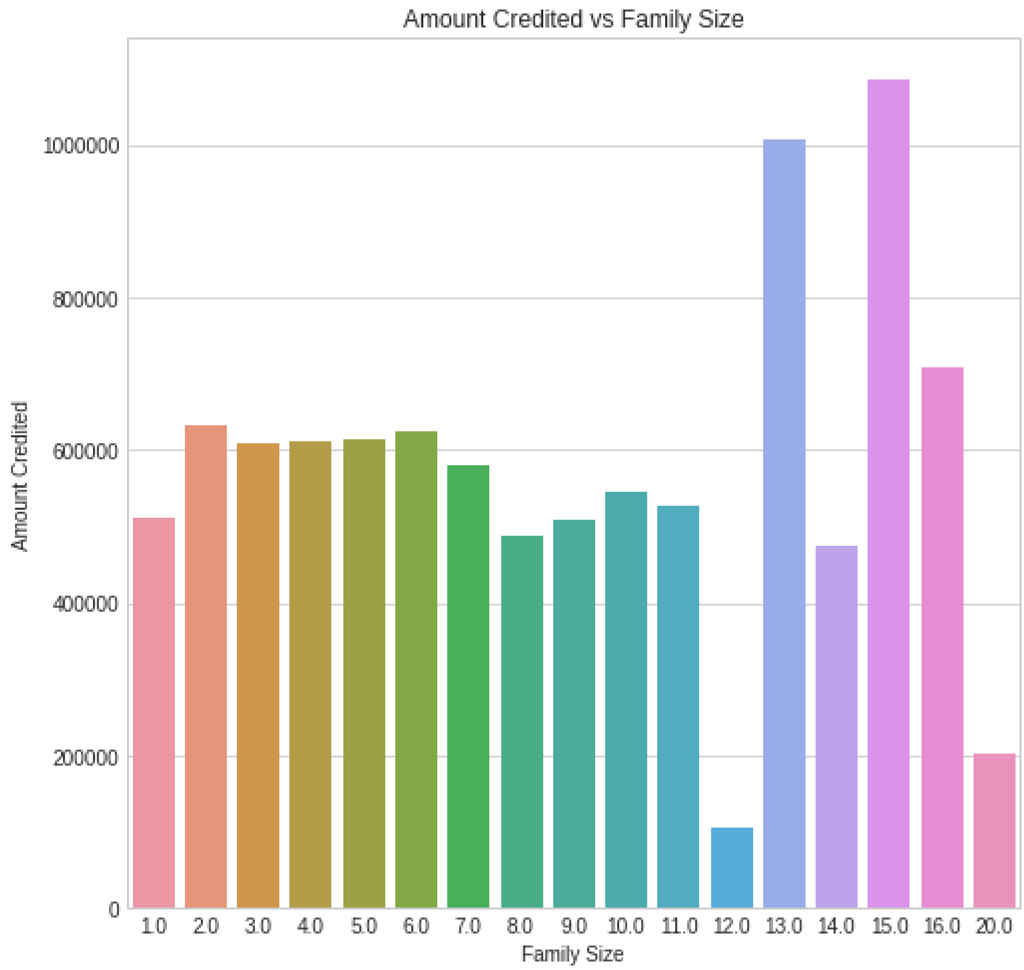
The clients with family size less than 4 are the least likely to default with the average rate of default being 8.34%

Below is the plot for the same,




At the first glance the graph might seem odd, but as the family size increases, the number of applicants falling in that category drastically falls to single digit number. Therefore it would not be correct to take the values shown by this graph to be correct for family size greater than 6.

Looking at the amount that the applicant asked for depending upon the family size.




Family size and the amount credited are directly proportional with some exceptions. It can be seen that clients with bigger families on average take the highest amount of loans.

From the previous plot, we know that the clients with bigger families are more likely to default on the payments and therefore steps should be taken to limit the amount credited to their account, since the client might default and the bank will incur heavy losses.

**V. Type of Loan:**
Loans are broadly classified into 2:
1. Cash loans : Taken for a fized tenure (home loan, card loan, etc)
2. Revolving loans : Line of credit that can be availed any time (Credit Card)

Below is the breakup of how likely it is that applicants across the two categories will default on their payments.

| Type of Loan | Total clients | Defaulter clients | Percentage default |
|---|---|---|---|
| Cash loans | 278232 | 23221 | 8.35 |
| Revolving loans | 29279 | 1604 | 5.48 |

It can be seen that the majority of the applicants have opted for “Cash loans” and the same are more likely to default on their payments. On the other hand, clients opting for “Revolving loans” are less likely to default.

This can be explained by the following argument that since it is possible for the clients to pay a minimum amount on the credit card and keep enjoying the benefits, therefore not many people have defaulted in this case. However, in case of Cash loans, the clients are required to pay a fixed amount every month, failing to meet which they will be categorized as defaulters.

This tells us that “Revolving loans” are less risky for the bank to give out as compared to the “Cash loans”. However, the profit earned in the form of interests will obviously be higher in the case of “Cash loans”. The bank must strike a balance between the two to optimize the profit.

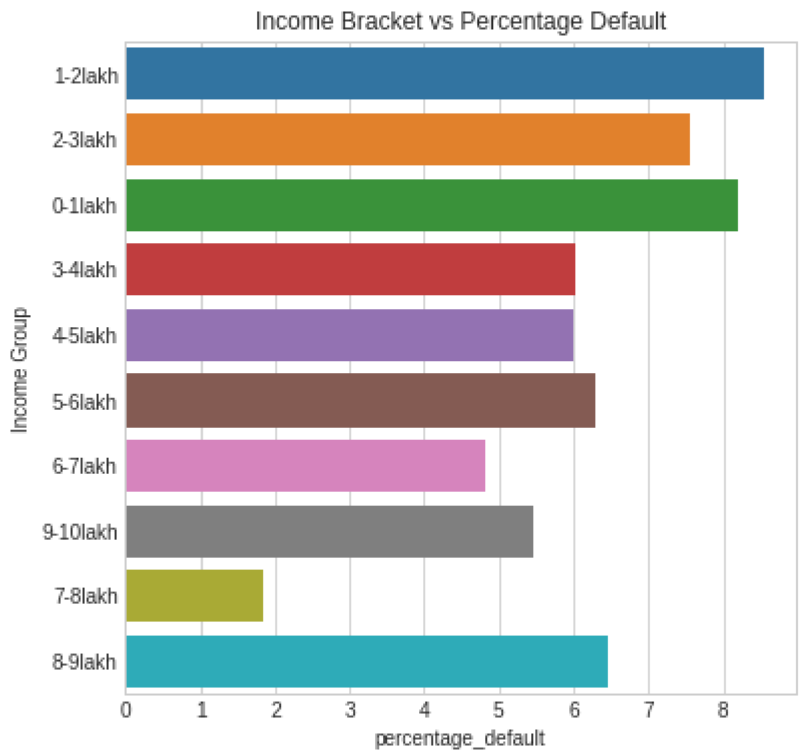
**VI. Total Income of the client:**

Below is a table representing the the various income groups, the frequency of each group and the rate of default.

| Income Group | Total applicants | Number of Defaulters | Percentage Defaulters |
|---|---|---|---|
| 1-2lakh | 155898 | 13326 | 8.55 |
| 2-3lakh | 65176 | 4921 | 7.55 |
| 0-1lakh | 63698 | 5225 | 8.20 |
| 3-4lakh | 14676 | 885 | 6.03 |
| 4-5lakh | 5361 | 322 | 6.01 |
| 5-6lakh | 1095 | 69 | 6.30 |
| 6-7lakh | 869 | 42 | 4.83 |
| 9-10lakh | 202 | 11 | 5.45 |
| 7-8lakh | 162 | 3 | 1.85 |
| 8-9lakh | 124 | 8 | 6.45 |

It can be seen that the number of applicants is the highest in the income bracket Rs. 1 to 2 lakh, and the majority of the clientele falls under the income bracket of Rs. 3 lakh.

Below is a barplot representing each category.



It can be seen that the rate of default is highest amongst the lower income groups (below 3 lakh). It can be attributed to the fact that the majority of these belong to the blue-collar work force and given their low income they are more likely to default.

It is also interesting to to see that the rate of default decreases till 8 lakh and then suddenly increases with the income. This shows that Rs. 8 lakh marks the boundary between the middle class housholds and the upper-middle class households. A typical middle class salaryman is therefore the least likely to default on loans.

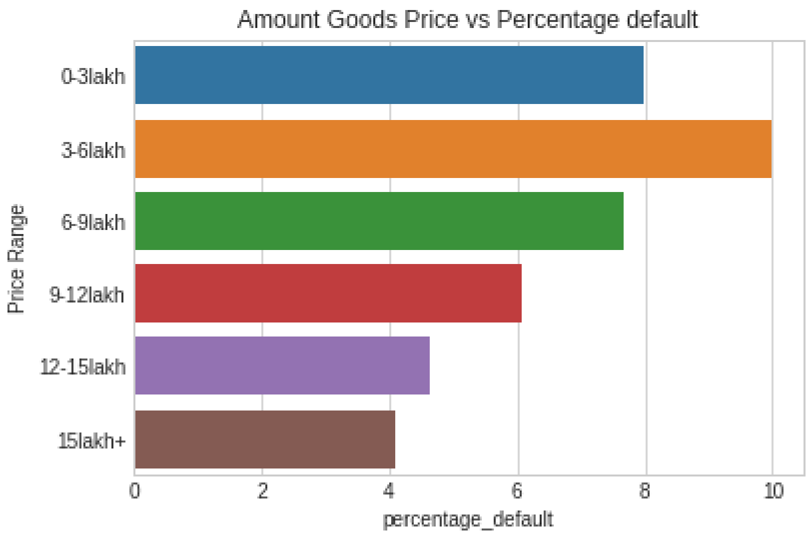
**VII. Collateral offered:**

Below table shows the percentage default rate corresponding to the price of the collateral.

| Price Range | Count_total | Count_default | percentage_default |
|---|---|---|---|
| 0-3lakh | 104426 | 8343 | 7.99 |
| 3-6lakh | 91309 | 9110 | 9.98 |
| 6-9lakh | 54202 | 4162 | 7.68 |
| 9-12lakh | 39694 | 2412 | 6.08 |
| 12-15lakh | 10688 | 494 | 4.62 |
| 15lakh+ | 6914 | 283 | 4.09 |

The majority of the collateral offered is within the mark of Rs. 6 lakhs.

Below is a plot representing each category of collateral.




The rate of default is the highest for the price range Rs. 3-6 lakh and then it gradually decreases as the price of the collateral increases.

This behaviour can be attributed to the fact that the well-to-do clientele owns properties which are highly valued and since they possess the means to pay back the loan therefore the rate of default decreases.

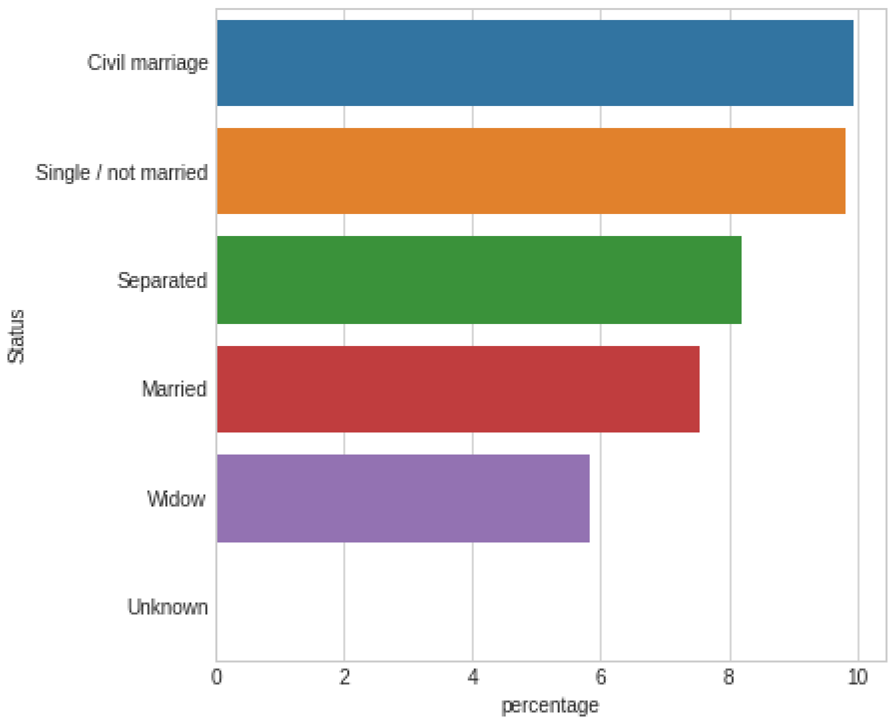
Whereas, in the lower price range of the collateral, since it is not highly valued, the applicant might be willing to risk it if it gets the loan off from his/her back.

The bank must look out for clientele whose collateral exceeds that of Rs. 10 lakh mark since they are least likely to default.

**VIII. Marital Status:**
Below is the distribution the marital status of the applicants.

| Status | Count_total | Count_default | percentage |
|---|---|---|---|
| Married | 196432 | 14850 | 7.56 |
| Single / not married | 45444 | 4457 | 9.81 |
| Civil marriage | 29775 | 2961 | 9.94 |
| Separated | 19770 | 1620 | 8.19 |
| Widow | 16088 | 937 | 5.82 |
| Unknown | 2 | 0 | 0.00 |

Plot:



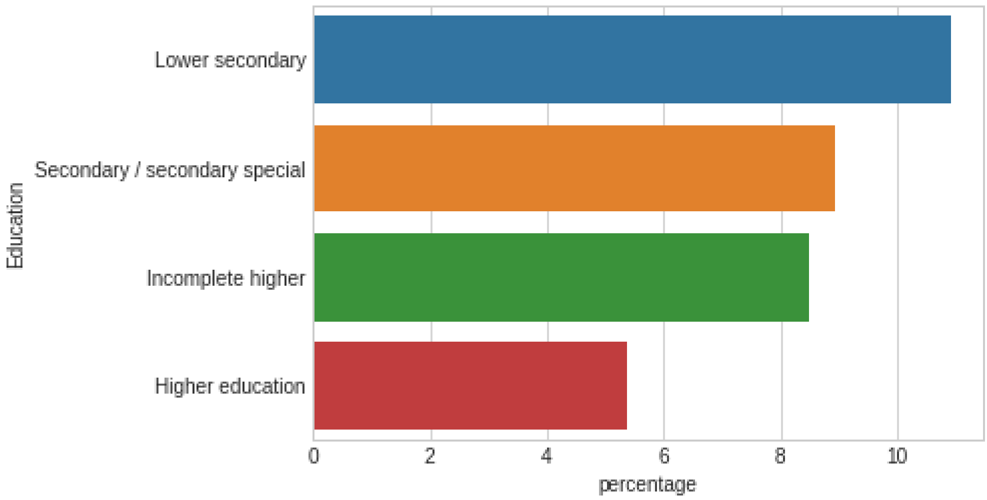

Applicants with “Civil marriages” tend to default the highest. One reason for this could be that the applicants forged their marriage certificate just in order to take loans and willingly default afterwards. Therefore, the applicants with “Civil Marriage” in their application should be thoroughly vetted before approving their loan.

The second highest category is for “Single people” with an average default rate of around 10%. The reason for this could be many, one of which can be that the majority of unmarried people are in their 20’s. It might be the case that these applicants had applied for education loans and due to high rates of unemployment they are unable to secure a job and hence the chances of default are high for them.

For the remaining categories, the rate of default decreases with the lowest being observed for the category “Widower”. Since most of the applicants falling into this category are of advanced age, therefore it is highly likely that they have a good enough corpus maintained so that their loan payments are done on time.

**IX. Level of Education:**
Below is the distribution of education with respect to the rate of default.

| Education | Total Count | Defaulter Count | Percentage_defaulters |
|---|---|---|---|
| Secondary / secondary special | 218391 | 19524 | 8.94 |
| Higher education | 74863 | 4009 | 5.36 |
| Incomplete higher | 10277 | 872 | 8.48 |
| Lower secondary | 3816 | 417 | 10.93 |



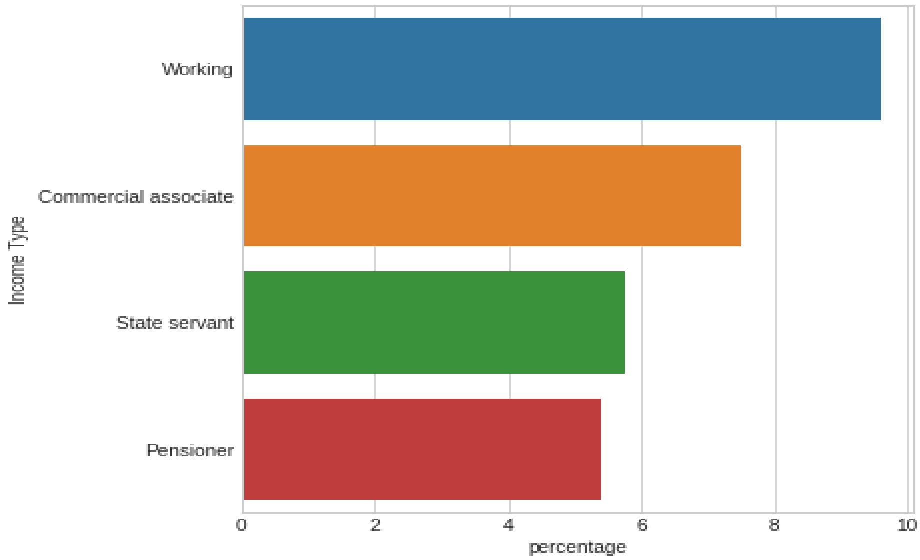

The level of education is inversely proportional to the default rate.
The sheer difference between the default rates is a testament to the role of education, with the default rate being about 11% for “Lower Secondary” and just 5% for “Higher Education” which is less than half of it.

The reason behind this phenomena is obvious, the applicants with lower levels of education end up in the not so well paying jobs and along with it comes the risk of employment safety with changing times. On the other hand, a good education provides job security and a well enough pay.

**X. Income Type of the client:**
Below is the distribution of the various types of income that the clients have.

| Income Type | Count_total | Count_default | Percentage |
|---|---|---|---|
| Working | 158774 | 15224 | 9.59 |
| Commercial associate | 71617 | 5360 | 7.48 |
| Pensioner | 55362 | 2982 | 5.39 |
| State servant | 21703 | 1249 | 5.75 |
| Unemployed | 22 | 8 | 36.36 |
| Student | 18 | 0 | 0.00 |
| Businessman | 10 | 0 | 0.00 |
| Maternity leave | 5 | 2 | 40.00 |

Plotting the distribution.



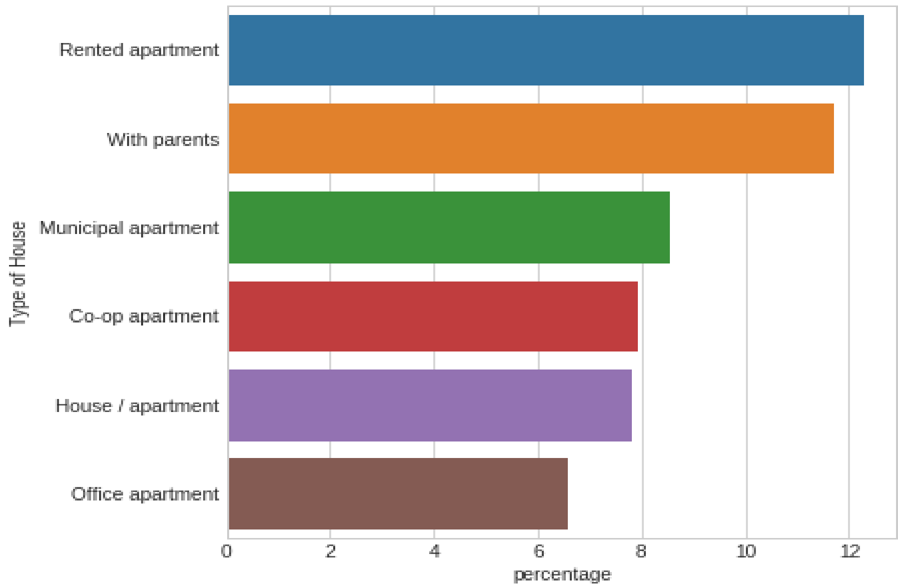

It can be seen that the working class is the most likely to default with the default rate being close to 10%.

Another interesting insight is that the clients working for private firms are more likely to default on their payments as compared to clients in Public service.
The reason for it being that the State government ensures the job security of its employees whereas that is not the case with private firms who value quality of work more than anything.
Another reason for it can be the difference in the pay scales of private and government employees. Government jobs are known for their higher entry wages whereas private sector is notorious for its low entry pay grade which has remained stagnant for years.

The best performing category amongst all is that of “Pensioners” with the average rate of default being as low as 5%. Regardless of the low default rate, the risk associated with the demise of the pension dependent clients might pose a problem for the bank.

**XI. Type of Residence:**
Below is the breakup of the different kinds of residence.

| Type of House | Count_total | Count_default | percentage |
|---|---|---|---|
| Rented apartment | 4881 | 601 | 12.31 |
| With parents | 14840 | 1736 | 11.70 |
| Municipal apartment | 11183 | 955 | 8.54 |
| Co-op apartment | 1122 | 89 | 7.93 |
| House / apartment | 272868 | 21272 | 7.80 |
| Office apartment | 2617 | 172 | 6.57 |




The general trend is that the rate of default increases with the clients who do not own their residence. The highest number of defaults come from people who have rented their house with the rate of default being above 12%.
Followed by them are the applicants who live with their parents. By general intuition, the majority of all the clients in this category belong to the marital status group "Single" and as seen before the default rate for them was high.

A curious thing to notice is that the rate of default is the least for the applicants who reside in the office provided apartments with the rate of default being less than 6.5%.

**XII. Employment history:**

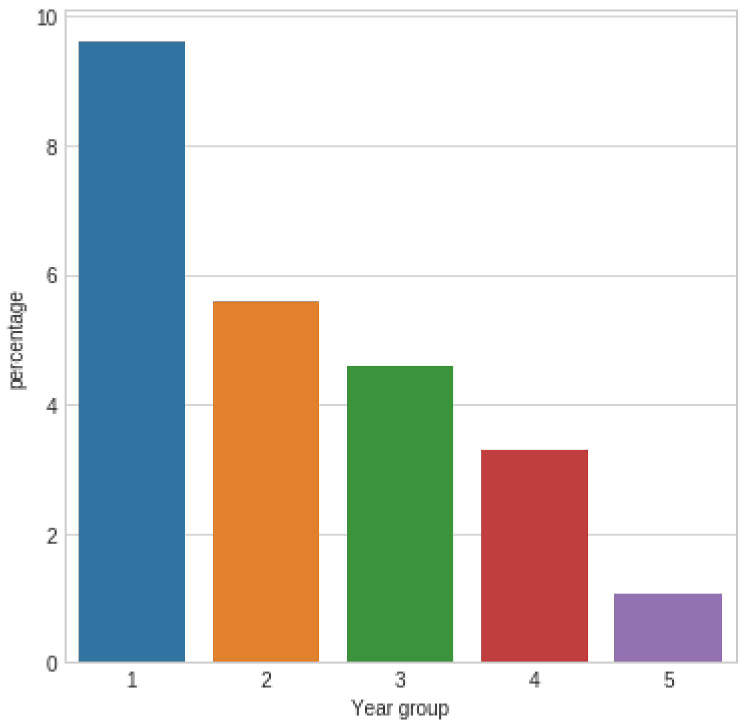
The clients have been divided into groups depending on the number of decades that they have worked. The groups are as follows:

Number of years 0 – 10 : 1
0 – 20 : 2
20 –30 : 3
30 –40 : 4
40 –50 : 5

Below is the distribution of the clients based on employment:

| Year group | Count_total | Count_default | percentage |
|---|---|---|---|
| 1 | 196721 | 18922 | 9.62 |
| 2 | 41937 | 2338 | 5.58 |
| 3 | 10720 | 491 | 4.58 |
| 4 | 2465 | 81 | 3.29 |
| 5 | 284 | 3 | 1.06 |

As is reasonable, the rate of default decreases drastically as the employment years increase.




Clients who are unemployed or have worked for less than 10 years are more likely to default with the rate of default being almost 10%.

However, as the number of years increase, the default rate decreases with the lowest being a menial 1%.

Therefore it is profitable for the bank to extend loans to people with a good number of years.

**XIII. Age of the client:**
The age of the client has been broken up in buckets of 10 years.

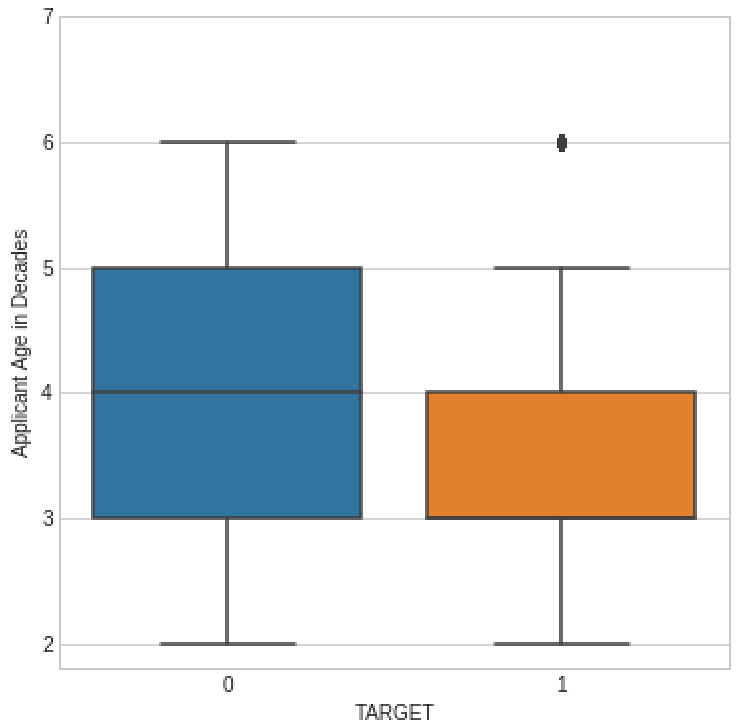
Below is the distribution:

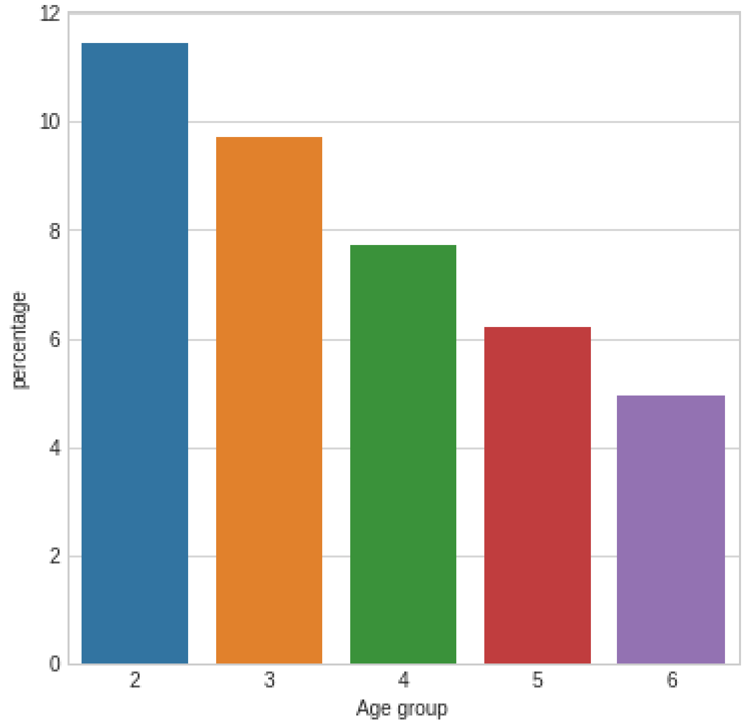
| Age group | Count_defaulters | Count_total | percentage |
|---|---|---|---|
| 2 | 4708 | 41093 | 11.46 |
| 3 | 7972 | 81944 | 9.73 |
| 4 | 5983 | 77415 | 7.73 |
| 5 | 4229 | 68039 | 6.22 |
| 6 | 1933 | 39020 | 4.95 |

There is an inverse relationship with the percentage of default and the age of the client. It can be seen that the older clients tend to default much less as compared to their younger counterparts.

The highest rate of default is 11.5% for people between 20 to 30 years of age, the causes for which have been discussed in previous pages.
The rate of default decreases significantly to 5% for the clients above 60 years of age.

The mature working population i.e clients between 40 to 60 years of age have an average default rate of around 7%.
The distribution can be visualized as follows:



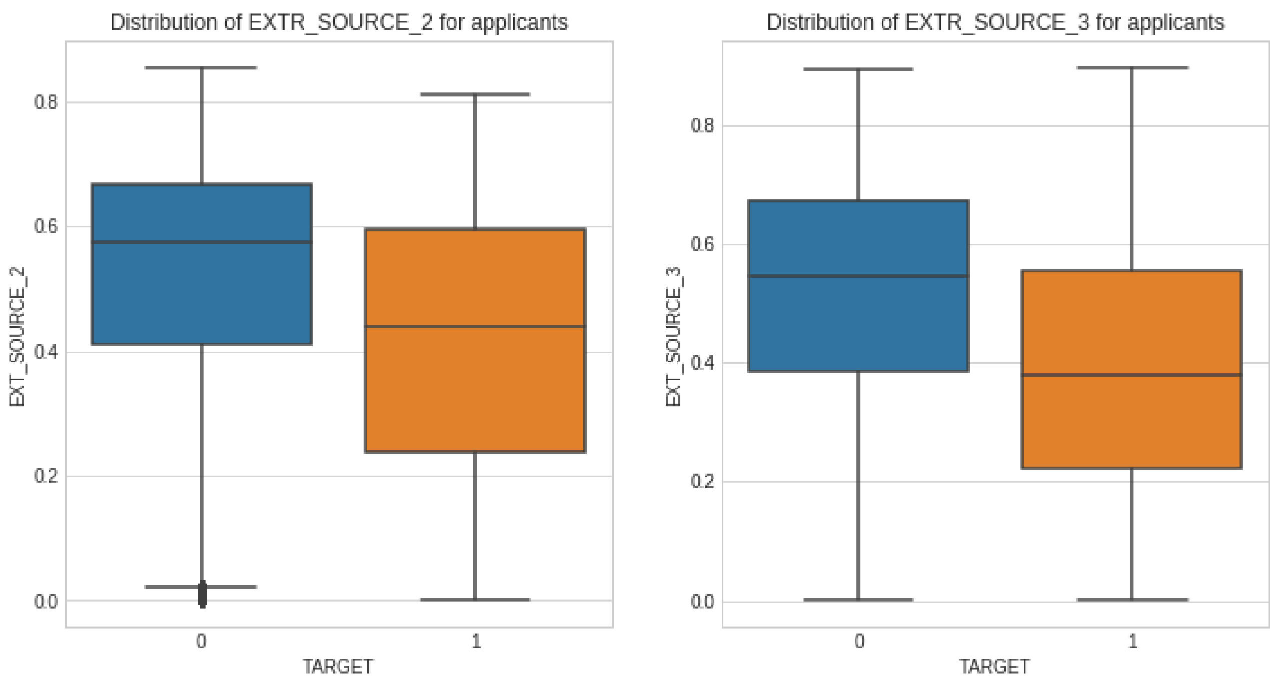



The banks should prefer extending loans to the people within the age group of 30 – 60 years of age. A thourough vetting must be done for clients below 30 years of age since they are the responsible for the majority of the loss that the bank incurs.

**XIV. Credit Rating of the client:**

Credit rating is a score that is given to an individual that gives an idea about the financial health of that person. A higher credit score would mean good financial health.

Given below is the distribution of the credit ratiing of the clientsfrom two different sources.



It can be inferred that the people who've defaulted on their payments (TARGET=1) have a lower credit rating with a median value of around 0.4 which is significantly less than that of the clients who have made their payments on time (above 0.5).

Banks should not extend loans to people with  a low credit score since they are most likely to default on their payments. However, if exceptions are to be made for some clients considering their income and other aspects, then a bargain could be struck so that the bank charges a higher rate of interest for such clients.

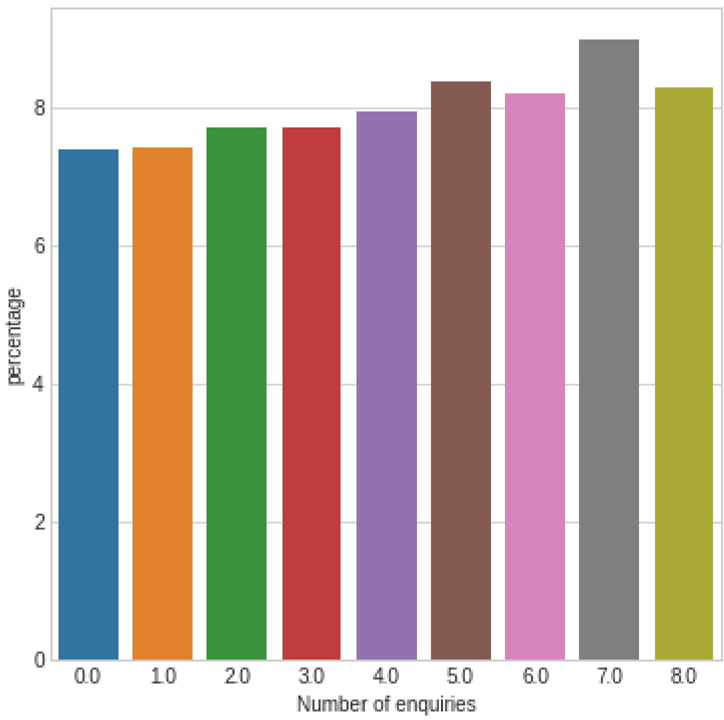
From a strictly "loss-avoiding" point of view, loans should not be extended to the clients with a low credit score.

**XV. Number of requests made to the Credit Bureau:**
Credit bureau is a government agency which has information about the clients' pervious transactions and loan applications. Sometimes the bank might need to ask the Credit Bureau some information about a potential client. The number of such requests made are stored in this variable.

Generallym the requests are only made for clients that show arouse suspection. It could be due to some incorrect information given by the client or any other reason which leads the bank to believe that due diligence above and beyond what is done for a normal client needs to be done in this case.

Below is the distribution of the requests made:

| Number of enquiries | Count_total | Count_default | Percentage |
|---|---|---|---|
| 1.0 | 53914 | 3997 | 7.41 |
| 2.0 | 51559 | 3973 | 7.71 |
| 0.0 | 50911 | 3755 | 7.38 |
| 3.0 | 39380 | 3036 | 7.71 |
| 4.0 | 27241 | 2160 | 7.93 |
| 5.0 | 17248 | 1445 | 8.38 |
| 6.0 | 10775 | 883 | 8.19 |
| 7.0 | 6428 | 578 | 8.99 |
| 8.0 | 3834 | 318 | 8.29 |



It rate of default increases gradually as more and more enquiries are made with the rate of default peaking at 7 enquiries.
The average rate of default for enquiries exceeding 5 is around 8.5% as compared to 7% for enquiries below 5.

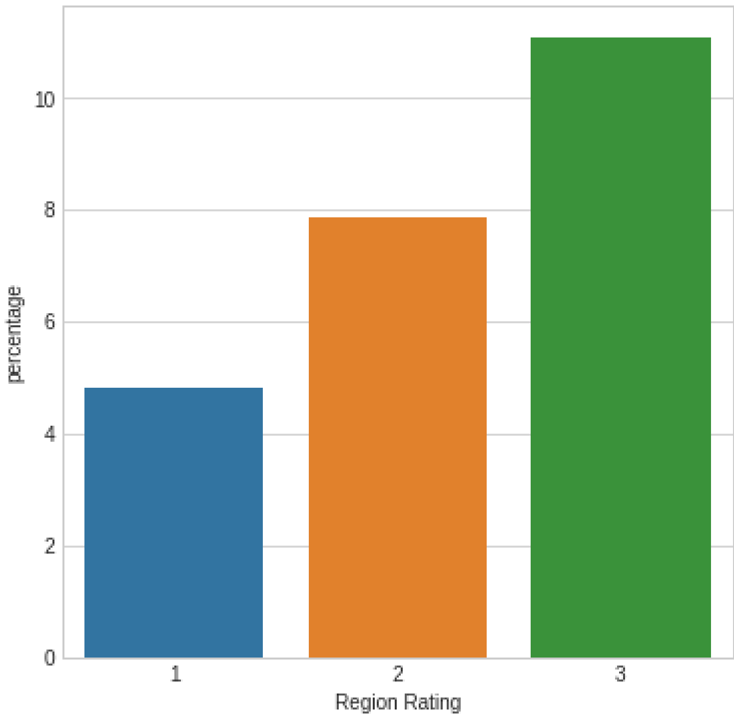
The bank should mostly not extend loans to people whose applications seem suspicious since they have higher chances of default, also this measure is very much practical since the number of such clients is very low as compared to normal applicants.

**XVI. Region Rating of the client's residency area:**
Based on the zip number that the client provides on his/her application form, the bank gives the client a rating based on the research done by the bank or some other external agency. This rating is given based on the type of region, the emplyment rate in that region, the income group of people residing in it, criminal history for the region, etc.

The rating is grouped into three categories i.e 1, 2 and 3. A higher rating would mean that the region is not suitable for extending loan and that the chances of the client defaulting from that region is more.

Below is a breakup of the clients falling in each category:

| Region Rating | Count_total | Count_default | percentage |
|---|---|---|---|
| 2 | 226984 | 17907 | 7.89 |
| 3 | 48330 | 5366 | 11.10 |
| 1 | 32197 | 1552 | 4.82 |



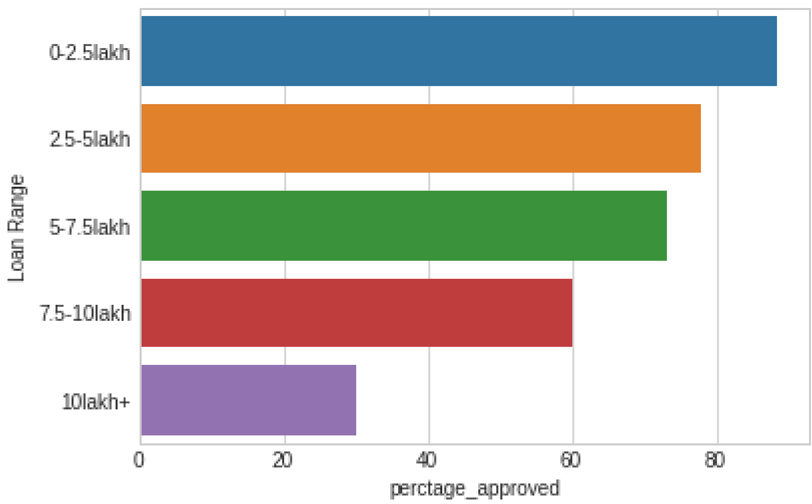

The clients belonging to region with a rating of 3 are 11% likely to default as compared to just 5% for region 1 which is less than half.

Therefore it can be said that the clients belonging to the low rated regions be strictly vetted before their loans are approved since the bank incurs loss on 11% of the applications which is quite high.

**XVII. Total amount asked by the client:**
The amount asked on the application has been categorised into buckets of Rupees 2.5 lakhs each. Below is the distribution of the amount asked by the client and the percentage of applications that were approved:

| Loan Range | Count_total | Count_approved | percentage_approved |
|---|---|---|---|
| 0-2.5lakh | 472120 | 418235 | 88.59 |
| 2.5-5lakh | 22296 | 17348 | 77.81 |
| 5-7.5lakh | 2596 | 1898 | 73.11 |
| 7.5-10lakh | 726 | 436 | 60.06 |
| 10lakh+ | 269 | 81 | 30.11 |

Visualizing the distribution:




We can see that the approval ratings drop drastically as the amount asked by the client goes up. The approval rating for below Rs. 2.5 lakh being around 90% and that for Rs.10 lakh+ being just 30%.

The banks are reluctant to loan high amounts to the people. This can be understood by the occupation distribution of the clientele as well as their income distribution. Since most of the clients are blue-collar workers therefore the approval ratings behave in such a manner.

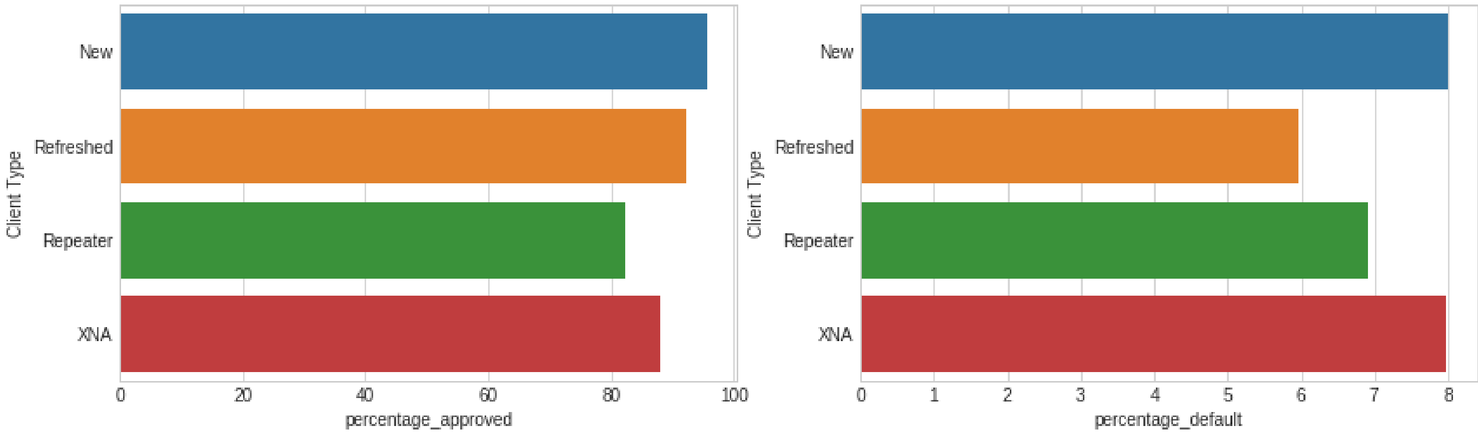
**XVIII. Client type:**
The clients have been categorized into 3 categories which are:
1. New
2. Refreshed
3. Repeater

Below is the frequency table for the same

| Client Type | Count_total | Count_approved | percentage_approved |
|---|---|---|---|
| New | 156865 | 150104.0 | 95.69 |
| Refreshed | 32958 | 30415.0 | 92.28 |
| Repeater | 236731 | 194680.0 | 82.24 |
| XNA | 251 | 221.0 | 88.05 |

The banks favour “New” Customers mainly to increase their customer base. The approval ratings for new clients are as high as 96% which means only 4 out of 100 people are rejected if they are new to the bank.
The “Refreshed” clients follow in at a close second with 92% approval ratings and the last category of “Repeaters” are less favoured in terms of approval.

Below is a plot representing the approval rates and the corresponding default rates for the different types of clients.




We can see that alothough new clients are preferred by the bank when approving the loans, it is those “New” clients which default 8% of the time.

However, if we look at the 'Refreshed' category, these applicants are less likely to default with their default rate being less than 6%. Hence 'Refreshed' clients must be given priority while approving the loans.

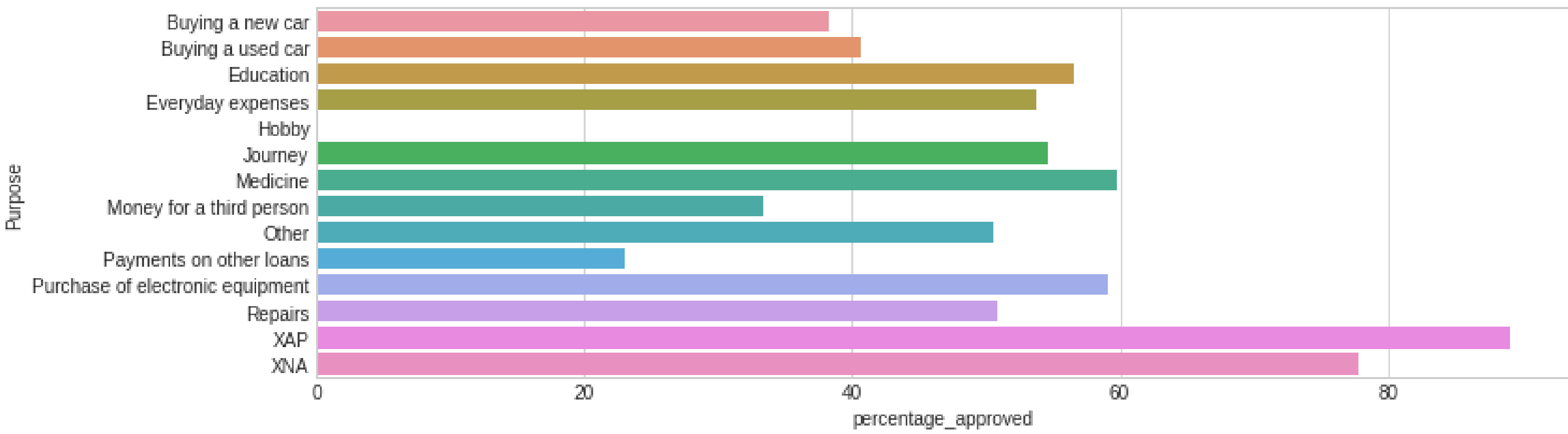
Bank's decision turns out to be right in not approving loans of 'Repeater clients' since their default rate is 8%.

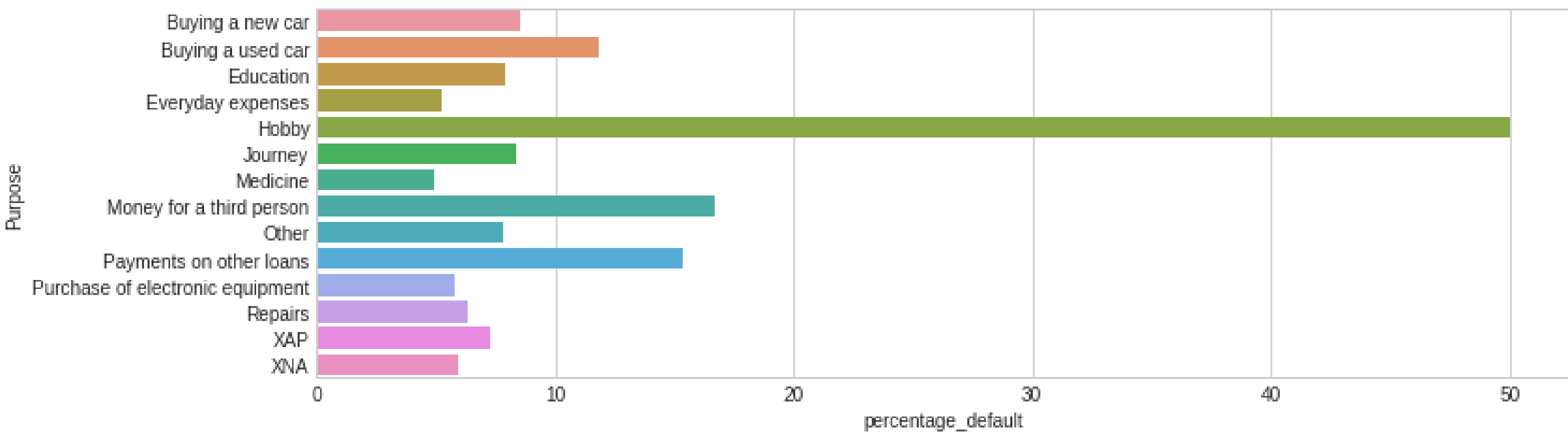
**XIX: Purpose of the loan:**

The frequency distribution of the various purposes as mentioned by the client are as follows:

| Purpose | Count_total | Count_approved | Percentage_approved |
|---|---|---|---|
| Buying a new car | 47 | 18.0 | 38.30 |
| Buying a used car | 219 | 89.0 | 40.64 |
| Education | 454 | 257.0 | 56.61 |
| Everyday expenses | 781 | 420.0 | 53.78 |
| Hobby | 2 | 0.0 | 0.00 |
| Journey | 227 | 124.0 | 54.63 |
| Medicine | 406 | 243.0 | 59.85 |
| Money for a third person | 6 | 2.0 | 33.33 |
| Other | 2152 | 1087.0 | 50.51 |
| Payments on other loans | 39 | 9.0 | 23.08 |
| Purchase of electronic equipment | 242 | 143.0 | 59.09 |
| Repairs | 1592 | 809.0 | 50.82 |
| XAP | 398400 | 354930.0 | 89.09 |
| XNA | 22238 | 17289.0 | 77.75 |

Below is a plot representing the approval rates and the default rates of the client:



The highest defaults are found in the following categories:

1.Buying a used car (12%): The aprroval rating is around 40% which needs to be reduced further so as to avoid losses.

2.Hobby(50%) : Since there are only 2 records pertaining to this category, it is rather difficult to arrive at any conclusion about it. However the approval rating is 0% for it.

3.Money for a third person(17%) : The approval rating for it is 33% that means that the bank is rejecting two-thirds of the applicants which is a healthy number.

4.Payments on other loans(15%) : The approval rating of it being 15% therefore the bank's money is right on the mark. Also it is rather ludicrous to get a new loan in order to pay off another.

**XX. Product Type opted by the client:**

The loan applications are classified into two categories which are:

1. Walk-in : Case when client comes to the bank office to apply for the loan.

2. X-sell : Case when the client applies for the loan online.

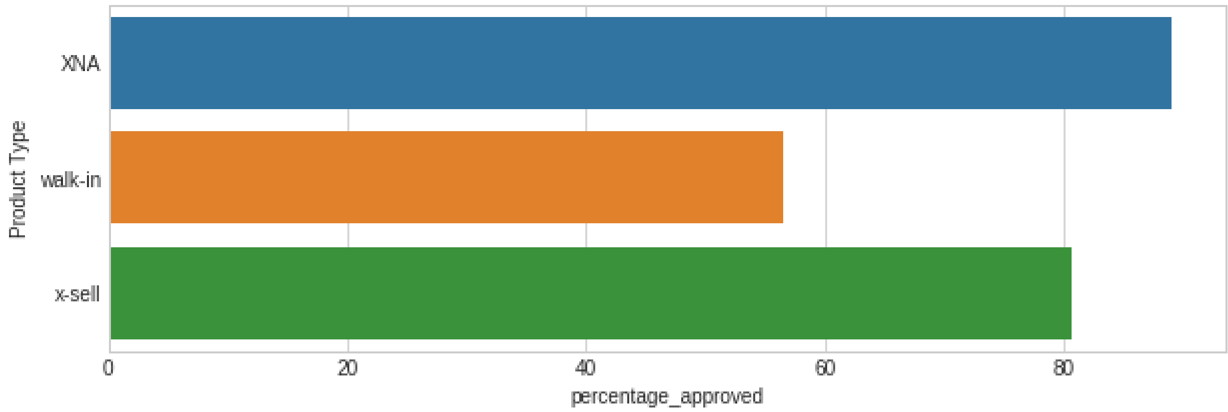
Below is the frequency distribution of the product type wrt to the default percentage:

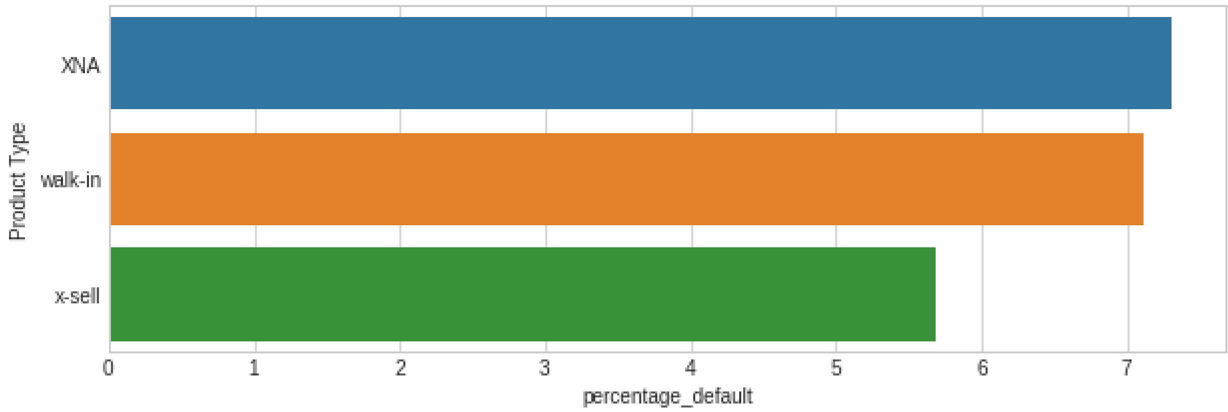
| Product Type | Count_total | Count_default | percentage_default |
|---|---|---|---|
| XNA | 398393 | 29132 | 7.31 |
| walk-in | 10026 | 714 | 7.12 |
| x-sell | 18386 | 1044 | 5.68 |

Frequency table wrt to the approval percentage:

| Product Type | Count_total | Count_approved | percentage_approved |
|---|---|---|---|
| XNA | 398393 | 354923.0 | 89.09 |
| walk-in | 10026 | 5662.0 | 56.47 |
| x-sell | 18386 | 14835.0 | 80.69 |

Visualizing



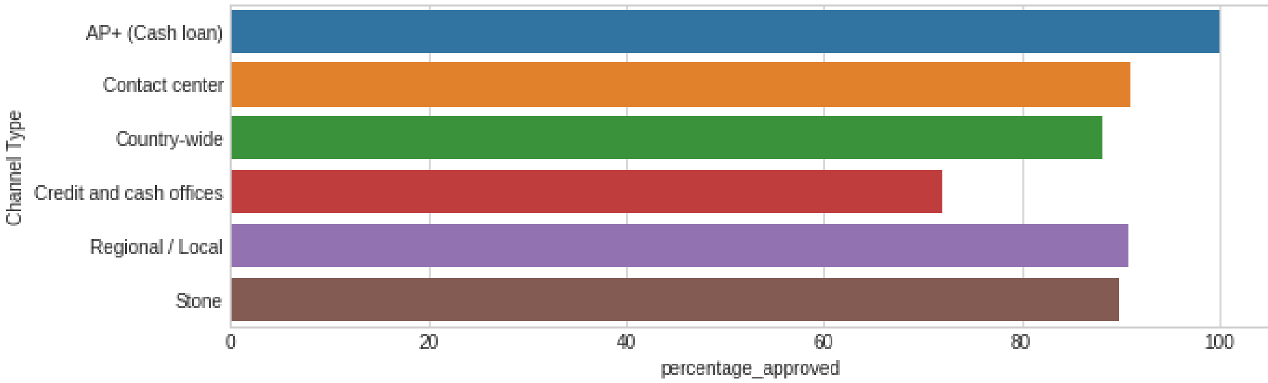

The applicants whose loans are categorized as 'walk-in' products are more likely to default with their default rating being 7%. The approval rating for the same ratings is around 56% which needs to be lowered further.
On the other hand, applicants opting for 'X-sell' kind of loans have much less likelihood of defaulting (5.5%) and their approval rating is also above 80%. One reason for this might be that 'X-sell' product is only preferred by thr literate section of applicants since it is an online product and therefore they are less likely to default.

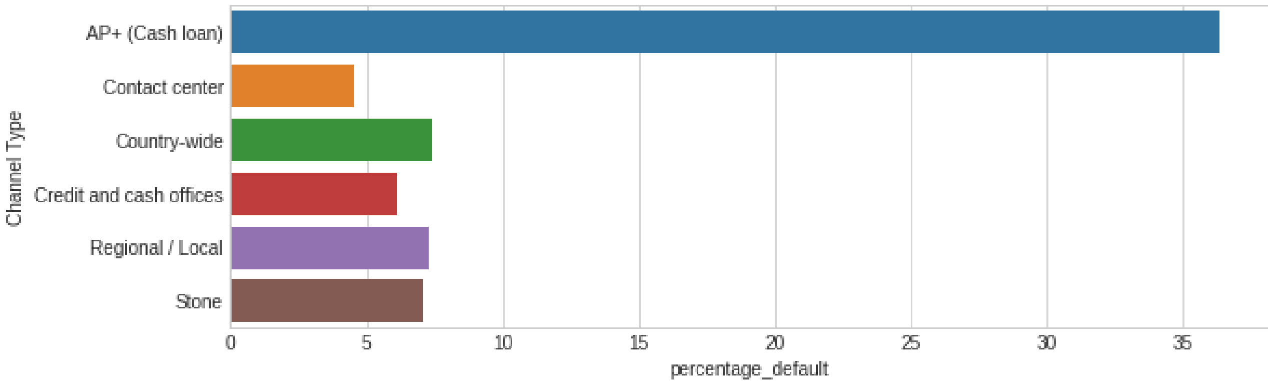
**XXI. Channel Type:**
This variable records the channel via which the customer was acquired by the bank.
Below is the frequency distribution of the product type wrt to the approval rating:

| Channel Type | Count_total | Count_approved | percentage_approved |
|---|---|---|---|
| AP+ (Cash loan) | 11 | 11.0 | 100.00 |
| Contact center | 22 | 20.0 | 90.91 |
| Country-wide | 218138 | 192297.0 | 88.15 |
| Credit and cash offices | 27978 | 20142.0 | 71.99 |
| Regional / Local | 61034 | 55395.0 | 90.76 |
| Stone | 119622 | 107555.0 | 89.91 |

Below is the frequency distribution of the product type wrt to the default rate:

| Channel Type | Count_total | Count_default | percentage_default |
|---|---|---|---|
| AP+ (Cash loan) | 11 | 4 | 36.36 |
| Contact center | 22 | 1 | 4.55 |
| Country-wide | 218138 | 16215 | 7.43 |
| Credit and cash offices | 27978 | 1716 | 6.13 |
| Regional / Local | 61034 | 4469 | 7.32 |
| Stone | 119622 | 8485 | 7.09 |




At the first glance it may seem that the bank is losing a great deal of money on 'AP+ (Cash loan)', but on closer inspection we can see that there are only 11 applicants who opted for it. Out of a total of 4 lakh+ applicants, this number is not enought to make any observations about this category. The same goes for 'Contact center' as well.

The channel 'Country-wide' has been performing well, since the default rating is 4.5% therefore an approval rating of about 91% is justified.

For the channel 'Credit and cash offices', the apprival rating is only 72%. However it can be seen that the possibility of client defaulting is only 6%. Therefore, the bank can increase the approvals for these channels which in turn will increase the bank's revenue.

**XXII. Rate of Interest:**
The rates of interest have been classified into various categories.

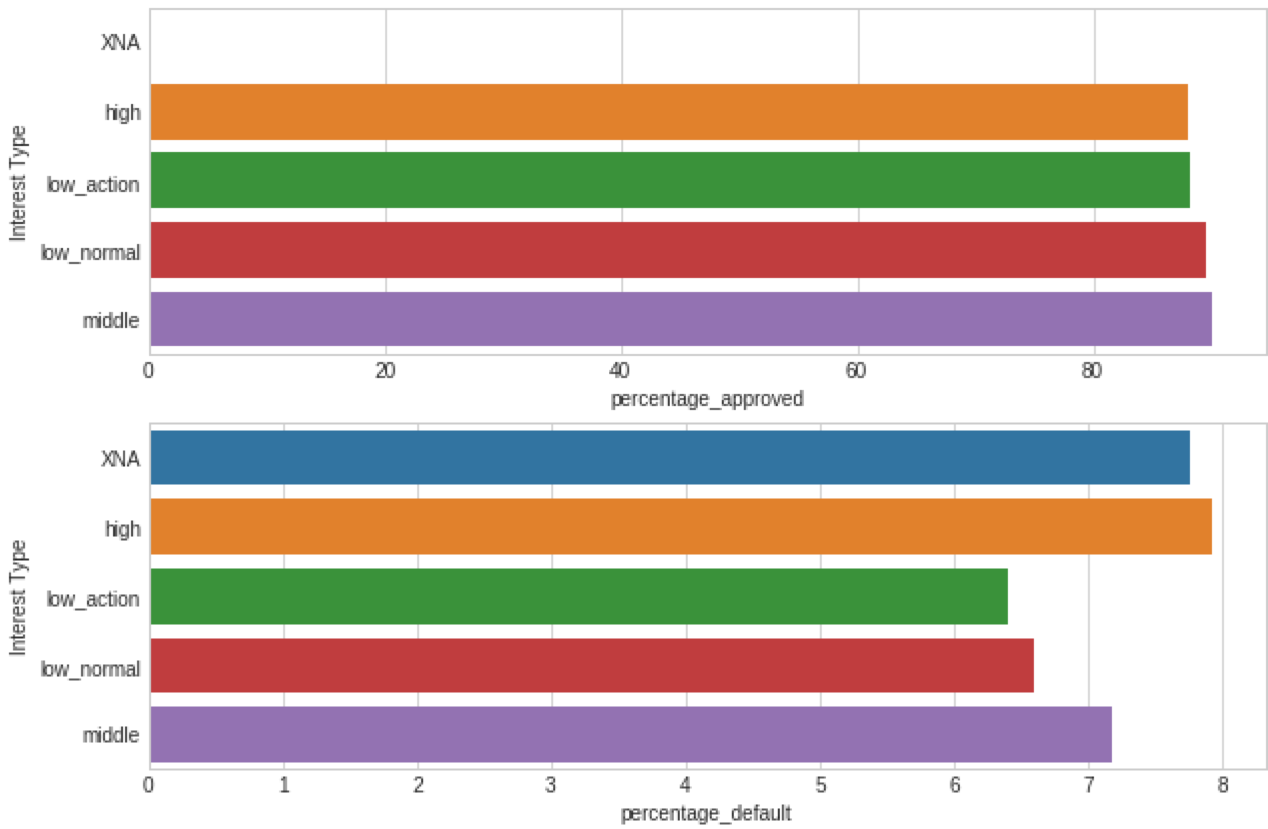
Below is the frequency distribution of the product type wrt to the default rate:

| Interest Type | Count_total | Count_default | percentage_default |
|---|---|---|---|
| XNA | 5182 | 402 | 7.76 |
| high | 156235 | 12379 | 7.92 |
| low_action | 42125 | 2698 | 6.40 |
| low_normal | 103020 | 6784 | 6.59 |
| middle | 120243 | 8627 | 7.17 |

Below is the frequency distribution of the product type wrt to the approval rate:

| Interest Type | Count_total | Count_approved | percentage_approved |
|---|---|---|---|
| XNA | 5182 | 7.0 | 0.14 |
| high | 156235 | 137681.0 | 88.12 |
| low_action | 42125 | 37183.0 | 88.27 |
| low_normal | 103020 | 92239.0 | 89.54 |
| middle | 120243 | 108310.0 | 90.08 |

Visualizing:



The interest group 'XNA' can be ignored here since it is not possible to state the interest rate of an application for which data is not available.
It can be seen that the default rate decreases from 'high' interest rates to 'low' interest rates whereas the approval rating increases slightly for the same.

To be noted that the bank risks losing money on loans with high interest as the default rate is almost 8% and the approval rating is close to 90%.
The same goes for loans with 'middle' interest rates since the approval rating is the highest at 90% and the rate of default is above 7%

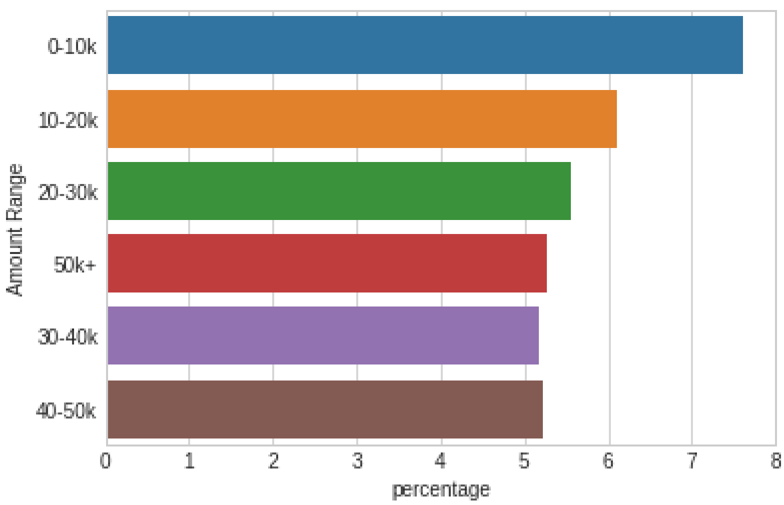
The bank's strategy on loans with 'low' interest is on the right track since the default ratings are lower and the approval ratings are high.

**XXIII. Down Payment made by the client:**
It is the amount paid by the client after the loan application is approved. The down payment is categorised into buckets for simplicity.

Frequency table wrt to the rate of default:

| Amount Range | Count_total | Count_default | percentage |
|---|---|---|---|
| 0-10k | 336924 | 25674 | 7.62 |
| 10-20k | 51872 | 3168 | 6.11 |
| 20-30k | 19069 | 1060 | 5.56 |
| 50k+ | 6616 | 348 | 5.26 |
| 30-40k | 6296 | 326 | 5.18 |
| 40-50k | 6028 | 314 | 5.21 |

Visualizing:




People with lesser down payments tend to default more with the default rate nearing 7.5%.
As the down payment goes up the default rate comes down significantly to around 5%.