**Conclusion**

After analysing the datasets, there are few attributes of a client with which the bank would be able to identify if they will repay the loan or not. The analysis is consisted as below with the contributing factors and categorization:

**Decisive Factor whether an applicant will be Repayer:**

1. NAME\_EDUCATION\_TYPE: Academic degree has less defaults.
2. NAME-INCOME\_TYPE: Student and Businessmen have no defaults.
3. REGION\_RATING \_CLIENT: Rating is 1 is Safer.
4. ORGANISATION\_TYPE: Clients with trade Type 4 and Type 5 and Industry type 8 have defaulted less than 3%.
5. DAYS\_BIRTH: People above age 50 have low probability of defaulting
6. DAYS\_EMPLOYED: Clients with 40+ year experience having les than 1% default rate
7. AMT\_INCOME\_TOTAL: Applicant with income more than 700,000 are less likely to default.
8. NAME\_CASH\_LOAN\_PURPOSE: Loans bought for Hobby, Buying garage are being repay mostly.
9. CNT\_CHILDREN: People with zero to two children tend to repay the loans.

**Decisive Factor whether an applicant will be Defaulter:**

1. CODE\_GENDER: Men are at relatively higher default rate
2. NAME\_FAMILY\_STATUS : People who have civil marriage or who are single default a lot.
3. NAME\_EDUCATION\_TYPE: People with Lower Secondary & Secondary education
4. NAME\_INCOME\_TYPE: Clients who are either at Maternity leave OR Unemployed default a lot.
5. REGION\_RATING\_CLIENT: People who live in Rating 3 has highest defaults.
6. OCCUPATION\_TYPE: Avoid Low-skill Laborers, Drivers and Waiters/barmen staff, Security staff, Laborers and Cooking staff as the default rate is huge.
7. ORGANIZATION\_TYPE: Organizations with highest percent of loans not repaid are Transport: type 3 (16%), Industry: type 13 (13.5%), Industry: type 8 (12.5%) and Restaurant (less than 12%). Self-employed people have relative high defaulting rate, and thus should be avoided to be approved for loan or provide loan with higher interest rate to mitigate the risk of defaulting.
8. DAYS\_BIRTH: Avoid young people who are in age group of 20-40 as they have higher probability of defaulting
9. CNT\_CHILDREN & CNT\_FAM\_MEMBERS: Client who have children equal to or more than 9 default 100% and hence their applications are to be rejected.
10. AMT\_GOODS\_PRICE: When the credit amount goes beyond 3M, there is an increase in defaulters.
11. DAYS\_EMPLOYED: People who have less than 5 years of employment have high default rate.

The following attributes indicate that people from these category tend to default but then due to the number of people and the amount of loan, the bank could provide loan with higher interest to mitigate any default risk thus preventing business loss:

1. **NAME\_HOUSING\_TYPE**: High number of loan applications are from the category of people who live in Rented apartments & living with parents and hence offering the loan would mitigate the loss if any of those default.
2. **AMT\_CREDIT:** People who get loan for 300-600k tend to default more than others and hence having higher interest specifically for this credit range would be ideal.
3. **AMT\_INCOME**: Since 90% of the applications have Income total less than 300,000 and they have high probability of defaulting, they could be offered loan with higher interest compared to other income category.
4. **CNT\_CHILDREN & CNT\_FAM\_MEMBERS**: Clients who have 4 to 8 children has a very high default rate and hence higher interest should be imposed on their loans.
5. **NAME\_CASH\_LOAN\_PURPOSE**: Loan taken for the purpose of Repairs seems to have highest default rate. A very high number applications have been rejected by bank or refused by client in previous applications as well which has purpose as repair or other. This shows that purpose repair is taken as high risk by bank and either they are rejected, or bank offers very high loan interest rate which is not feasible by the clients, thus they refuse the loan. The same approach could be followed in future as well.

**Other suggestions:**

* 90% of the previously cancelled client have actually repayed the loan. Record the reason for cancellation which might help the bank to determine and negotiate terms with these repaying customers in future for increase business opportunity.
* 88% of the clients who were refused by bank for loan earlier have now turned into a repaying client. Hence documenting the reason for rejection could mitigate the business loss and these clients could be contacted for further loans

**Model Building Conclusion**

**Problem Statement 1**

Loan Preferences: Segment customers based on their loan preferences, considering factors like loan amount, term, type, and interest rates. Repayment History: Further segment these groups based on their repayment history, such as timely payments, late payments, or defaults.

we used two models

1.Random forest Classifier: accuracy 0.91

2.Gradient Bosting Classifier: accuracy 0.90

So we can conclude Random forest is best model fit for problem statement.

**Problem Statement 2**

Risk Assessment: Evaluate and quantify the risk associated with each loan applicant using the model. Approval Strategy Development: Develop strategies to balance risk minimization with approving a high number of safe loans.

Here we use two models

1.Random forest Classifier accuracy 0.92

2.Logistic Reggression accuracy 0.91

So we can conclude random forest is best model fit for problem statement