# Insights

- We have 122 features(16 categorical, rest numerical) in the ‘application_data’ file and the data is majorly imbalanced.
- People who are unemployed or on Maternity Leave are more likely the ones who will suffer difficulty in paying credit loans.
- The lower the education level, the higher the chances of becoming a defaulter.
- People having ‘Single’ / ’Civil Marriage’ marital status have the highest chances of becoming a defaulter.
- “Rented apartments” and “With Parents” categories have the highest probability of being defaulters.
- Females take more loans, and males take less quantity of loans. However, Men have more trouble in paying their loans/credits.
- 30-40 is the age group where there is a maximum chance of being a defaulter.
- Strange Behavior: This is strange, owning a property has nothing to do with defaulting ideally if a person owns a property, then the chances of defaulting should be significantly less.
- those who don't own a car are more likely to default.
- Low-skill laborers, Security Staff, Waiters/Barmen staff, Drivers, and Cooking Staff are the occupations with the highest defaulters ratio.
- People who have 3 - 5 years of experience usually have Payment difficulties.
- ‘cash loans’ is having higher number of credits than ‘Revolving loans’ contract type.
- Clients who have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’, ‘Other’ , ‘Medicine’ and ‘Government’. Less clients are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.
- The features of Income, credit amount, and good price have a high correlation with each other. The larger the applicant's income, the larger the credit amount, similar to the value of the good price.
- Repeater is the type of customer with the highest rejection and cancellation rate, about 20%
- Applicant without Portfolio information will usually get loan canceled
- Regional / Local and Stone channel has a higher loan approval rate than other channels
- Here we can see that Female is getting more Refused more approved more canceled more unused but in case of male it is having average in every category.
- Working type people are applying more loans as compare to others.

# Conclusions:

- Orgs should focus on the **'Student,' 'Pensioner,' and 'Businessman' contract types**, especially for **housing types other than 'Co-op apartment,'** to ensure successful payments.
- A reduced emphasis on the **'Working' income type** is advisable, given its higher incidence of unsuccessful payments.

- Additionally, caution is warranted when considering **loan purposes**, with particular attention to the **'Repair' category** due to its higher frequency of unsuccessful payments.
- A strategic focus on clients residing in housing types **'With parents'** and **“Rented Apartments”** is recommended, as they exhibit the lowest occurrence of unsuccessful payments.
- Double-check the clients who are **Unemployed** and/or on **Maternity Leave**.
- According to the analysis of data, we can see that Owning another property has nothing to do with the success rate. All are equally likely defaulters. Orgs should not fall into that trap.
- Attention to **‘Single’ / ’Civil Marriage’ marital status** clients.
- **Low-skill laborers, Security Staff, Waiters/Barmen staff, Drivers, and Cooking Staff** are the **occupations** that should be given attention.
- Applicants **without Portfolio** information - Avoid

# Future Work

- **Statistics:** Some of the factors may not provide reliable insights. To ensure the validity of our hypotheses, we can employ statistical tests such as the Test of Means, Test of Variance, and ANOVA tests.
- **Feature Engineering:** Implement a stats model and find the odds ratio of each column to find out which columns are most important. So we can remove other colinear columns, to avoid multi-colinearity.
- **Handle imbalanced data**:
    - Undersampling: RandomUnderSampler, TomekLink
    - Oversampling: SMOTE, ADASYN, SMOTEENN, SMOTETomek