



# **CASE STUDY FOR CREDIT ANALYSIS USING (EDA)**

**Priyanka Narode**





# INTRODUCTION

- ▶ This assignment aims to give an idea of applying EDA(exploratory data analysis) in a real business scenario.
- ▶ The basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.




# Steps to Do

- PROBLEM STATEMENT UNDERSTANDING
- DATA UNDERSTANDING(APPLICATION DATA)
- DATA CLEANING: IDENTIFICATION OF MISSING VALUES AND OUTLIER
- DATA IMBALANCE
- SANITY CHECKS (checking negative and null values)
- UNIVARIATE ANALYSIS
- BIVARIATE ANALYSIS



# Data Understanding (Application Data)

- Checking Shape, Info, D-Types , size and Describe of the application data and pervious application data to get quick understanding of the data.
- 

# Describe (application data)

[7]:

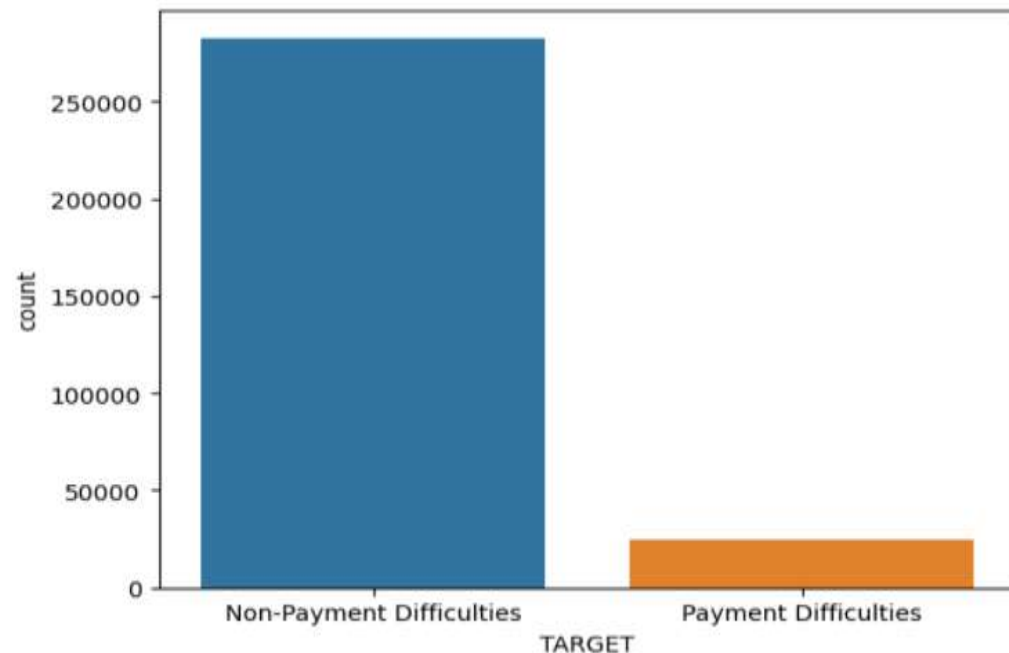
	SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	DAYS_BIRTH
count	307511.00	307511.00	307511.00	307511.00	307511.00	307499.00	307233.00	307511.00	307511.00
mean	278180.52	0.08	0.42	168797.92	599026.00	27108.57	538396.21	0.02	-16037.00
std	102790.18	0.27	0.72	237123.15	402490.78	14493.74	369446.46	0.01	4363.99
min	100002.00	0.00	0.00	25650.00	45000.00	1615.50	40500.00	0.00	-25229.00
25%	189145.50	0.00	0.00	112500.00	270000.00	16524.00	238500.00	0.01	-19682.00
50%	278202.00	0.00	0.00	147150.00	513531.00	24903.00	450000.00	0.02	-15750.00
75%	367142.50	0.00	1.00	202500.00	808650.00	34596.00	679500.00	0.03	-12413.00
max	456255.00	1.00	19.00	117000000.00	4050000.00	258025.50	4050000.00	0.07	-7489.00

# Data Cleaning

- Identification of Missing Values and Treatment:
- There were several Columns with missing value percentage greater than 40% so we dropped them.
- Remaining columns with missing values we imputed them with mean/median/mode as required.
- Outliers: There were several columns which have outliers present in them. Depicted through boxplot in columns named as :
  - AMT\_ANNUITY
  - AMT\_GOODS\_PRICE
  - NAME\_TYPE\_SUITE
  - OCCUPATION\_TYPE

# DATA IMBALANCE

- We observed the data set it was highly imbalanced almost 91.9% clients were Non-Payment Difficulties(0) and about 8.1% for Payment Difficulties(1).





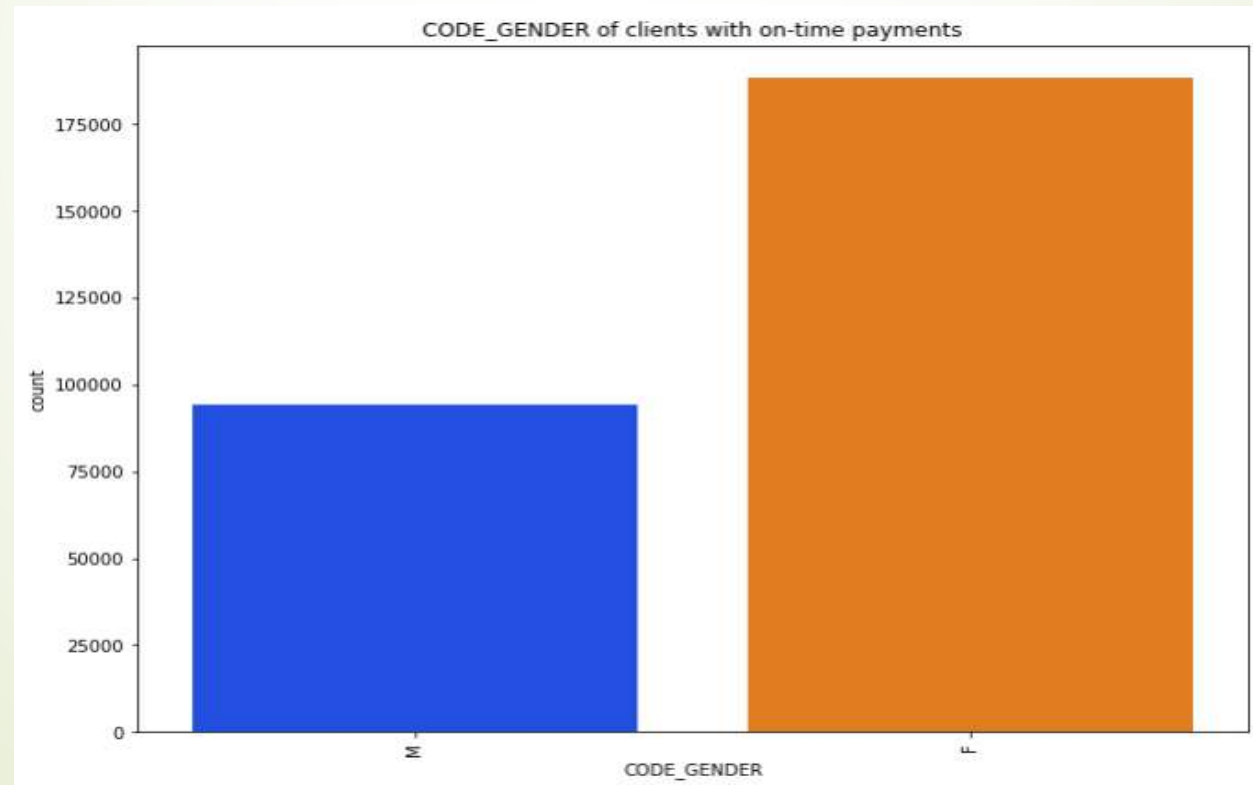
# Missing Values

- Doing analysis through the data observed some columns have negative values, Missing Values, Validate Data Type. So to correct them we did sanity checks in these columns named as;
- "DAYS\_BIRTH",
- "DAYS\_EMPLOYED",
- "DAYS\_REGISTRATION",
- "DAYS\_ID\_PUBLISH",
- "DAYS\_LAST\_PHONE\_CHANGE"



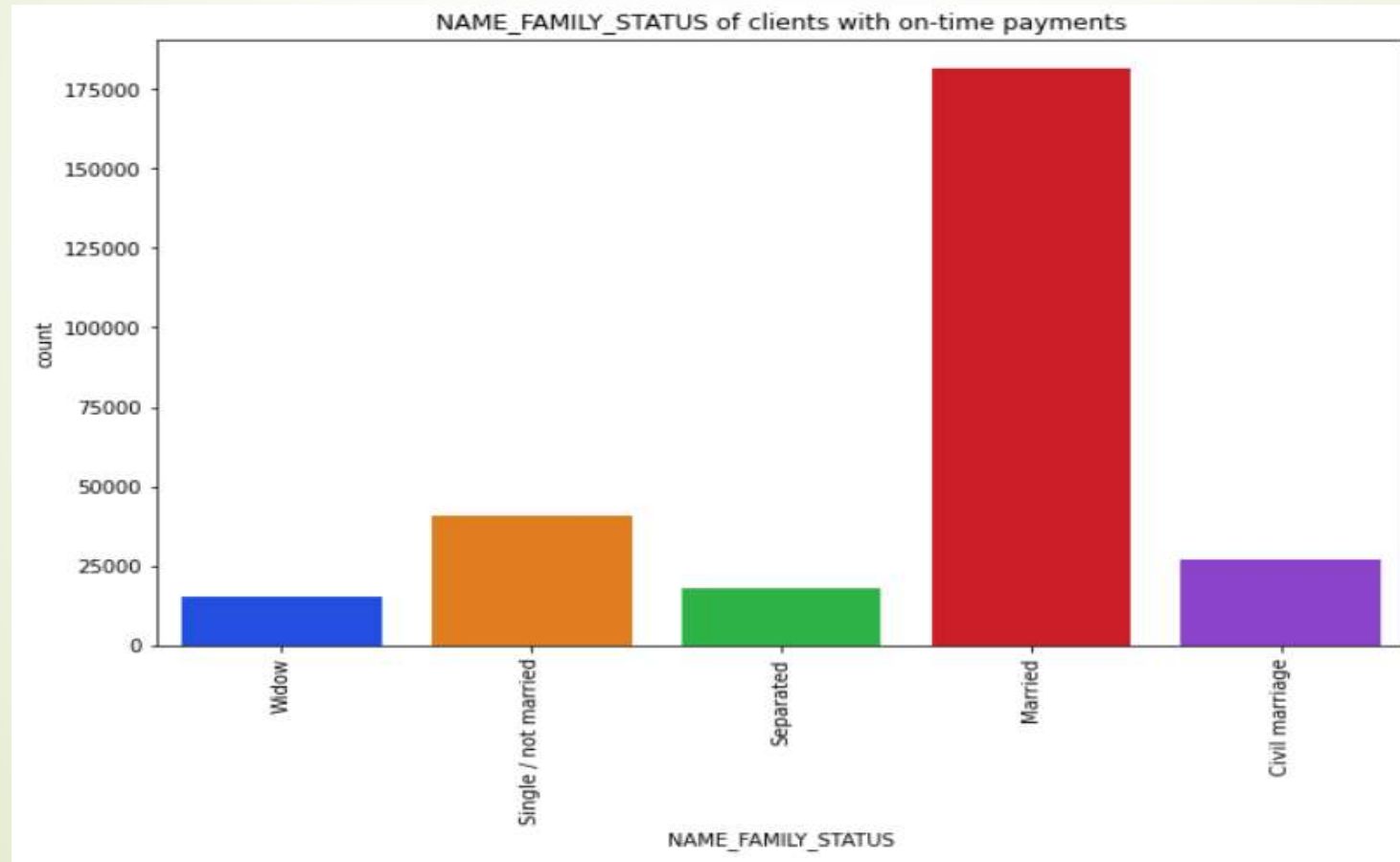
# Bivariate Analysis

- We can say that most client are females, working, have secondary education and are married.
- We can say taking loan from the bank someone accompanied with them as well



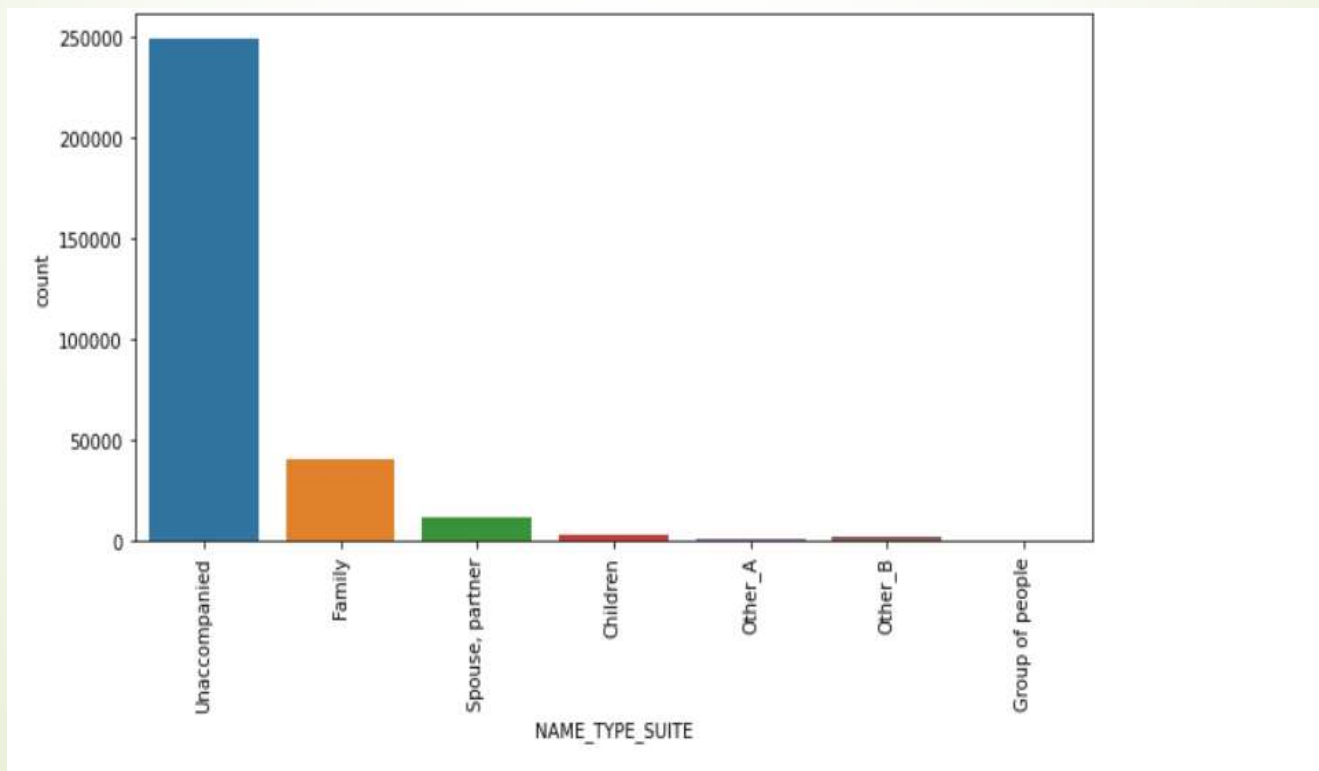
# NAME\_FAMILY\_STATUS

- Count of married people applying for a loan is higher than the rest.

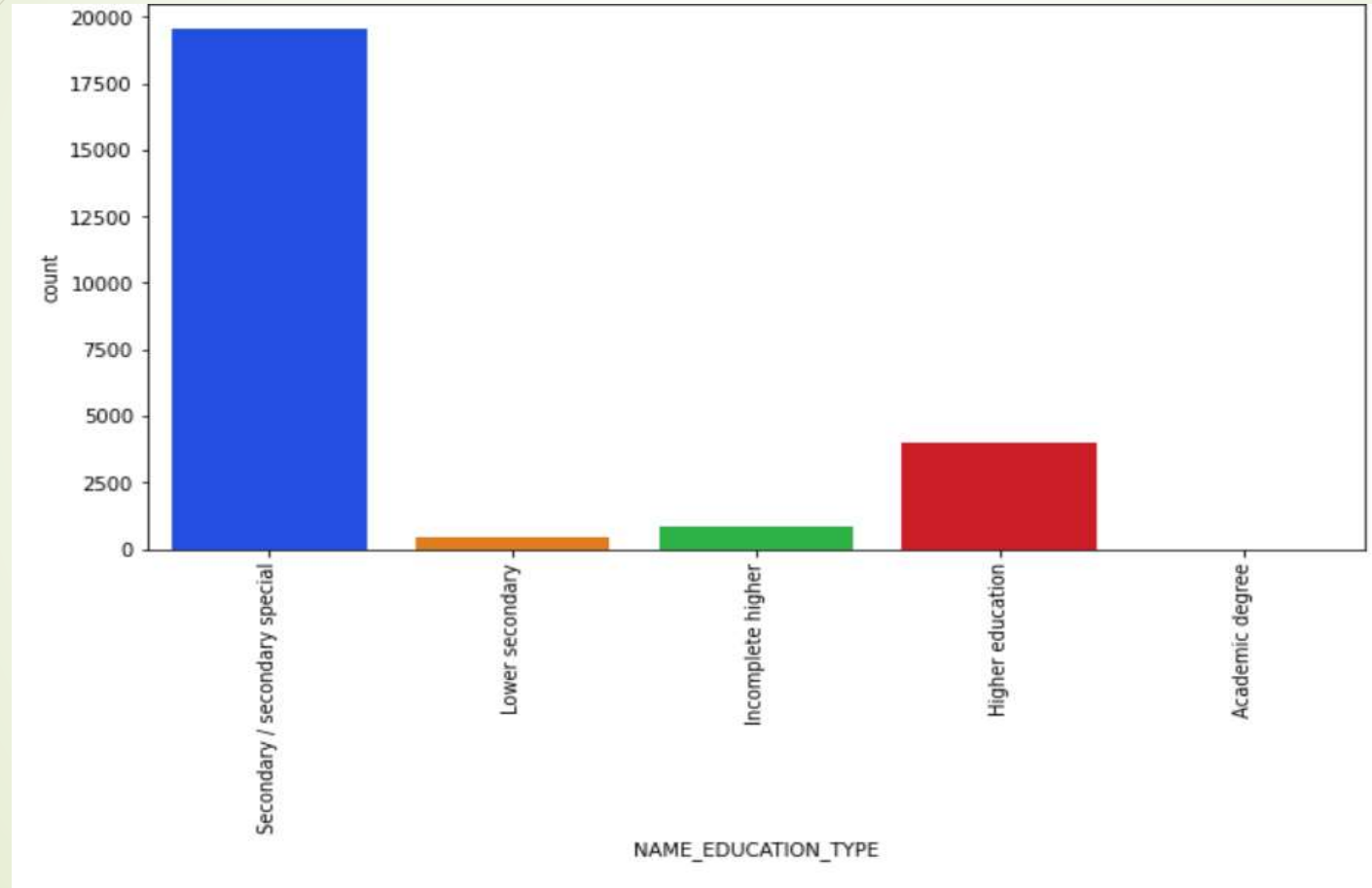


# NAME\_TYPE\_SUITE

- The majority of customers are unaccompanied, with family being the next largest group
- This shows that most customers tend to be alone rather than in groups.

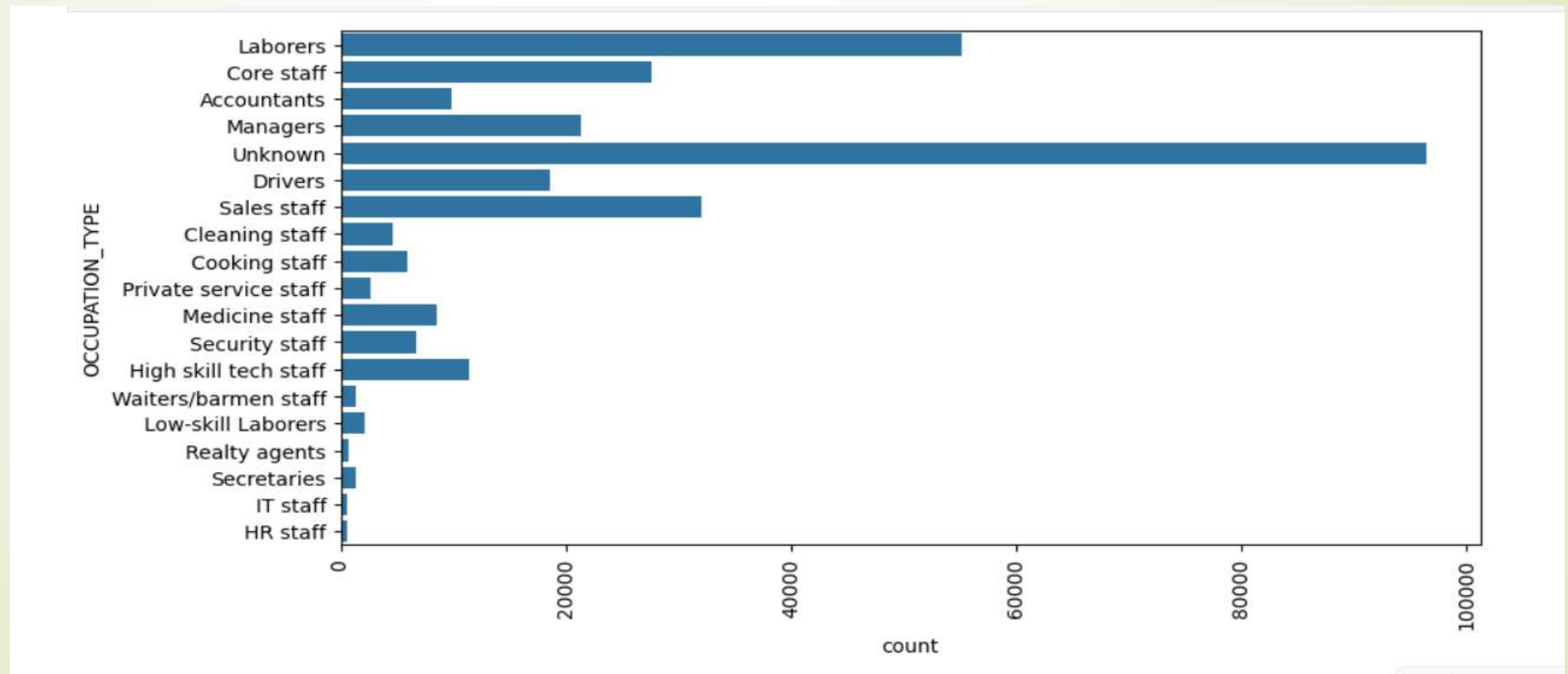


- Client with Secondary Education are more likely to apply for the loan.

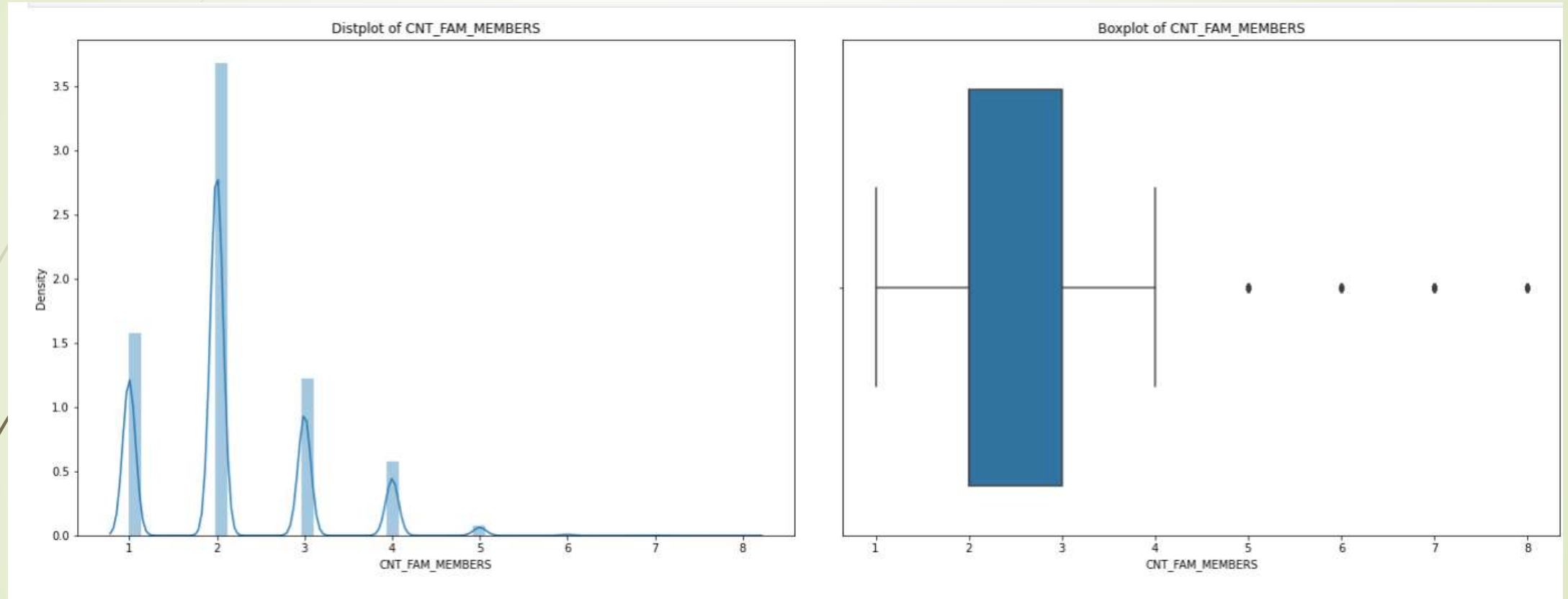


# OCCUPATION\_TYPE

- CONCLUSION FROM ANALYSIS:
- We observed that there is sharp increase in labours category in payment
- The graph is showing clearly that labourer's are the most applied applicants for loans.



# Analysis on CNT\_FAM\_MEMBERS



Applicants with 5 or more family members are clearly outliers



# CONCLUSION

- Data is Highly Imbalanced 91.9% is Client with Non-Payment Difficulties and 8.1% is client with Payment Difficulties.
- Females have higher chance of applying for a loan
- Majority of clients who applied for loan are from Working class.
- Client with Secondary Education are more likely to apply for the loan.
- Count of married people applying for a loan is higher than the rest.
- Majority of the clients have House/Apartment.
- Laborers are more frequent customer of bank.
- Most of the client who applied for loan are Unaccompanied.
- Cnt\_Fam\_Members Applicants with 5 or more family members are clearly outliers



THANKYOU