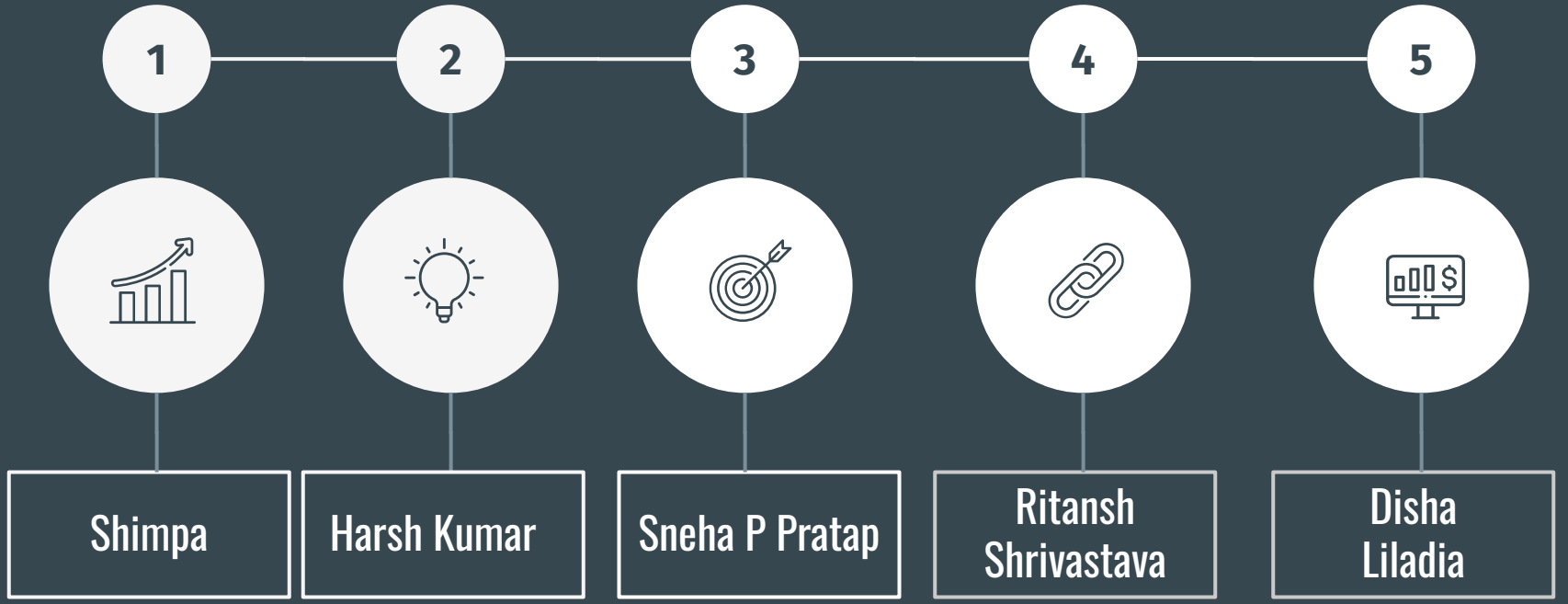
MAJOR PROJECT

Presented by : Team K

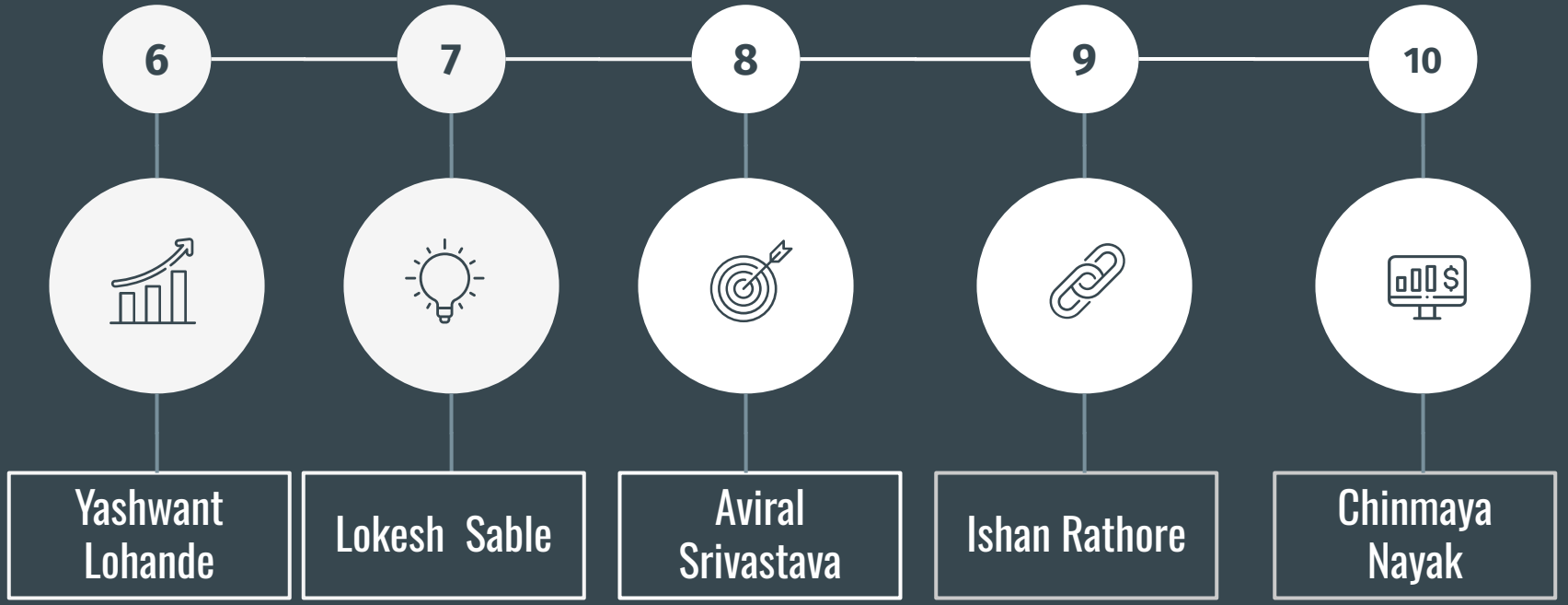
Team Leads

	**Team Lead**	Manish Kumar Sahu
	**Team Lead**	Shravya Venishetty
	**Co-Lead**	Rudra Nair
	**Co-Lead**	Amit Kumar Patra

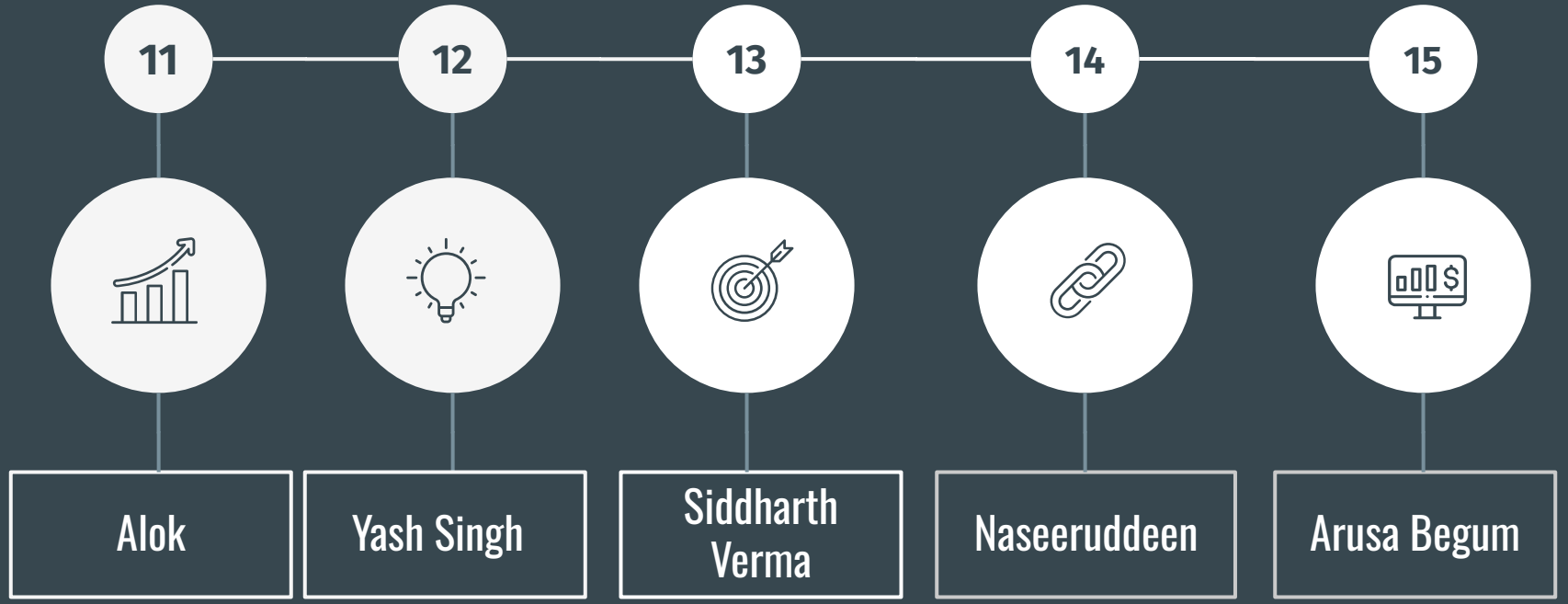
Extended Team Members



Extended Team Members



Extended Team Members



Extended Team Members

16

Anand Kumar Sahu

17

Komal Shyam Pathade

18

Prarthana S Poojary

19

Aishawarya Dadasaheb Bhusari

20

Rushikesh Narayan Amrute

Objectives

Our objective is to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

We will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of the following case study.

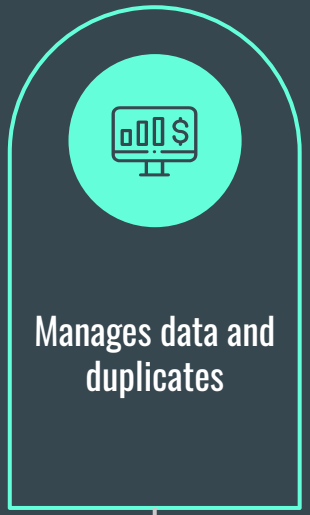
In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

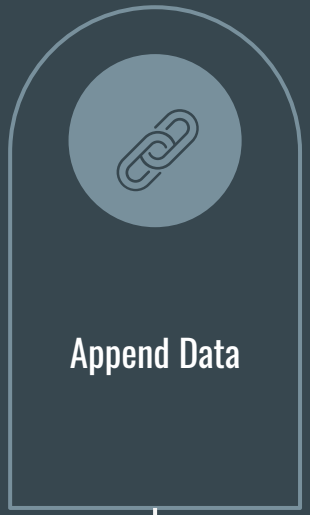
Data Cleaning

By eliminating duplicate entries, adding missing values, standardizing data formats, fixing data types, and spotting outliers, we improved our dataset through data cleaning.

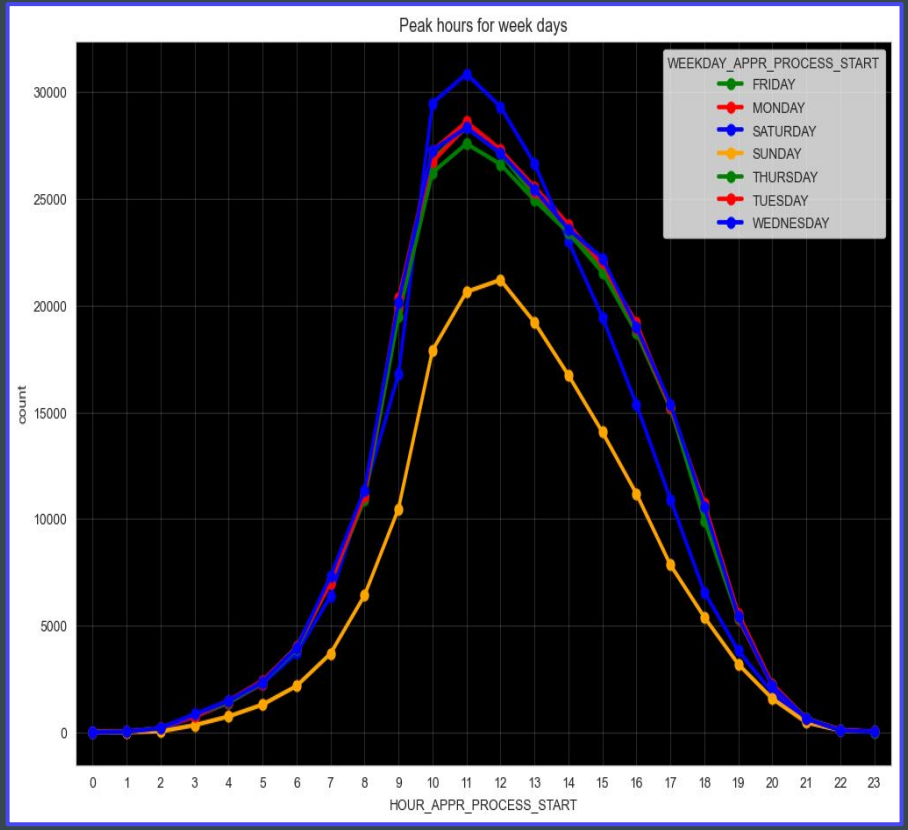
These procedures guaranteed the accuracy, consistency, and dependability of our data for analysis.

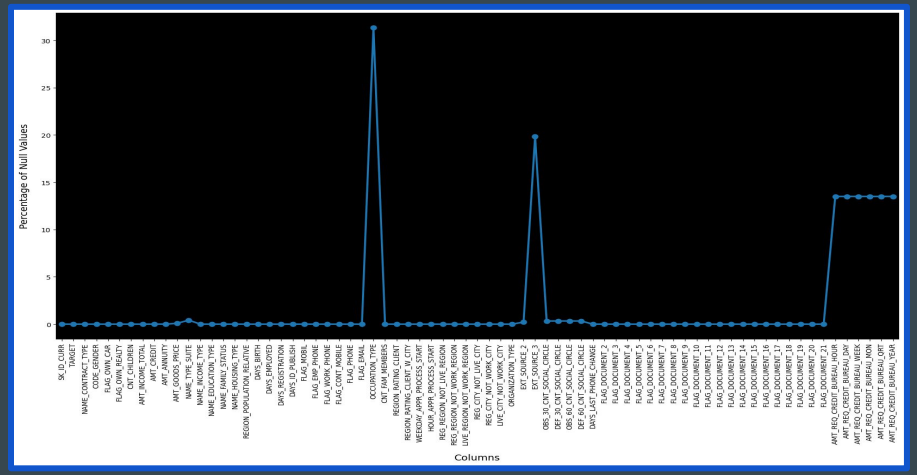
5 Steps to Cleaner Data

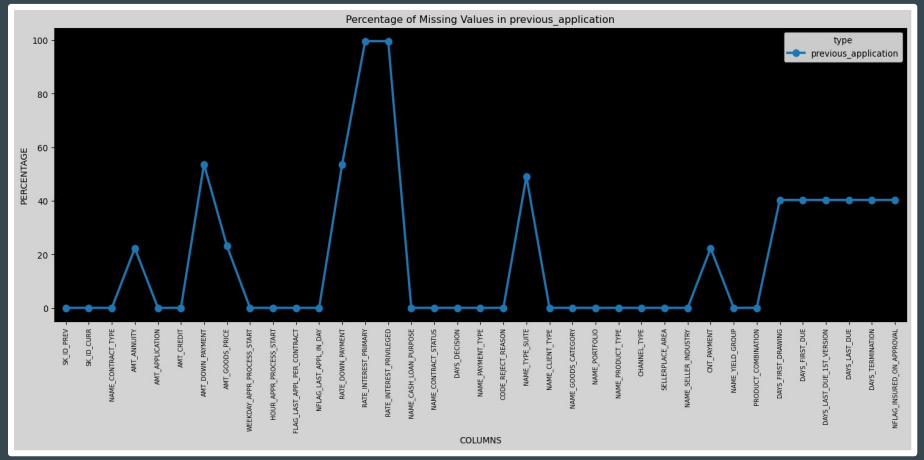


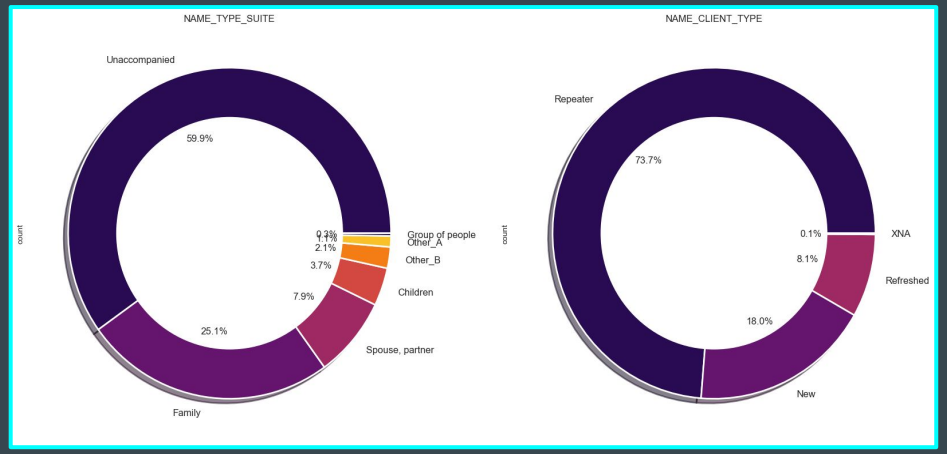


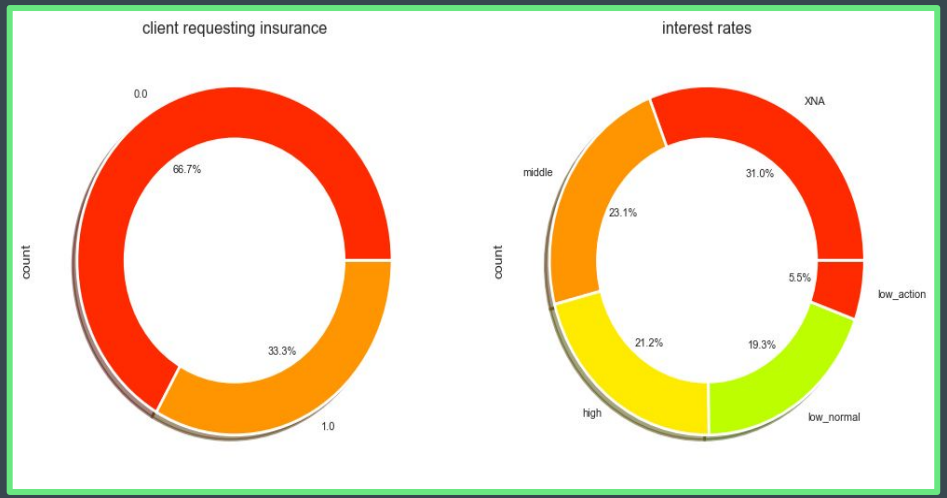
Data Analysis

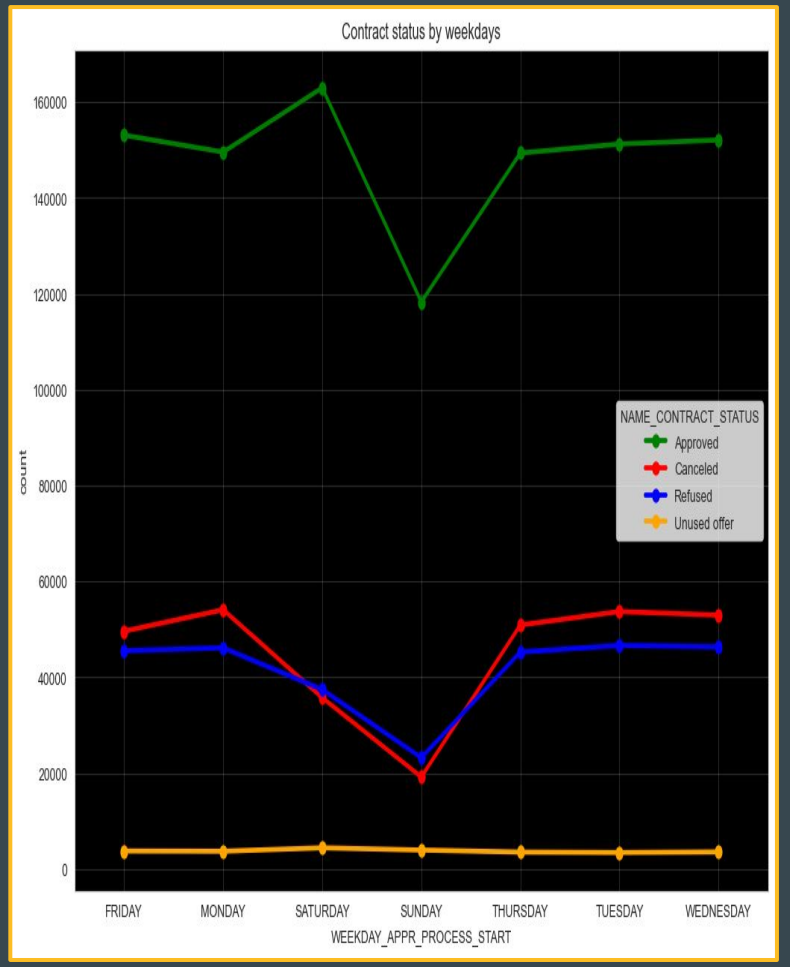


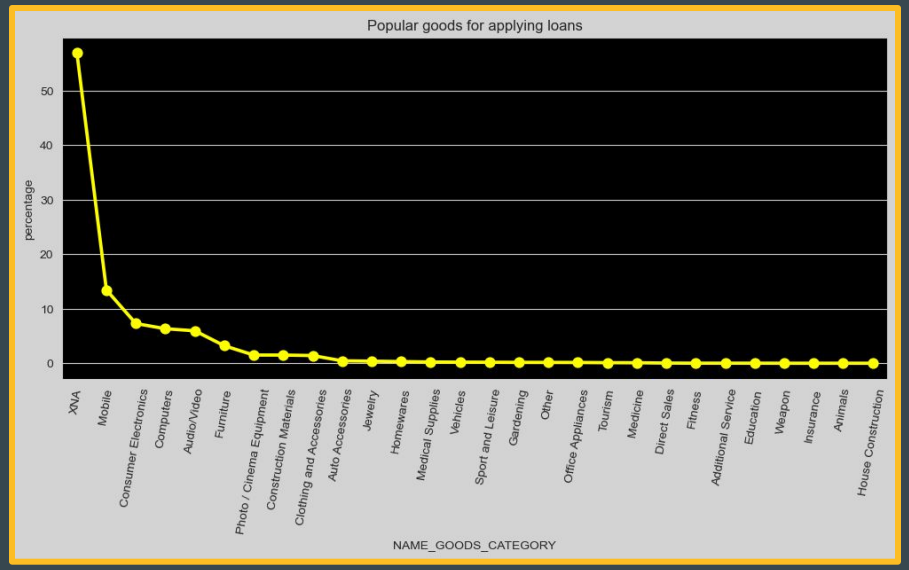


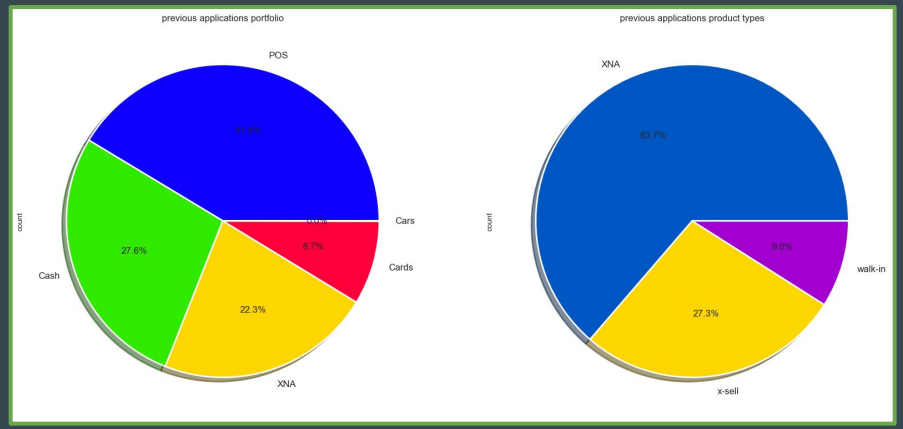


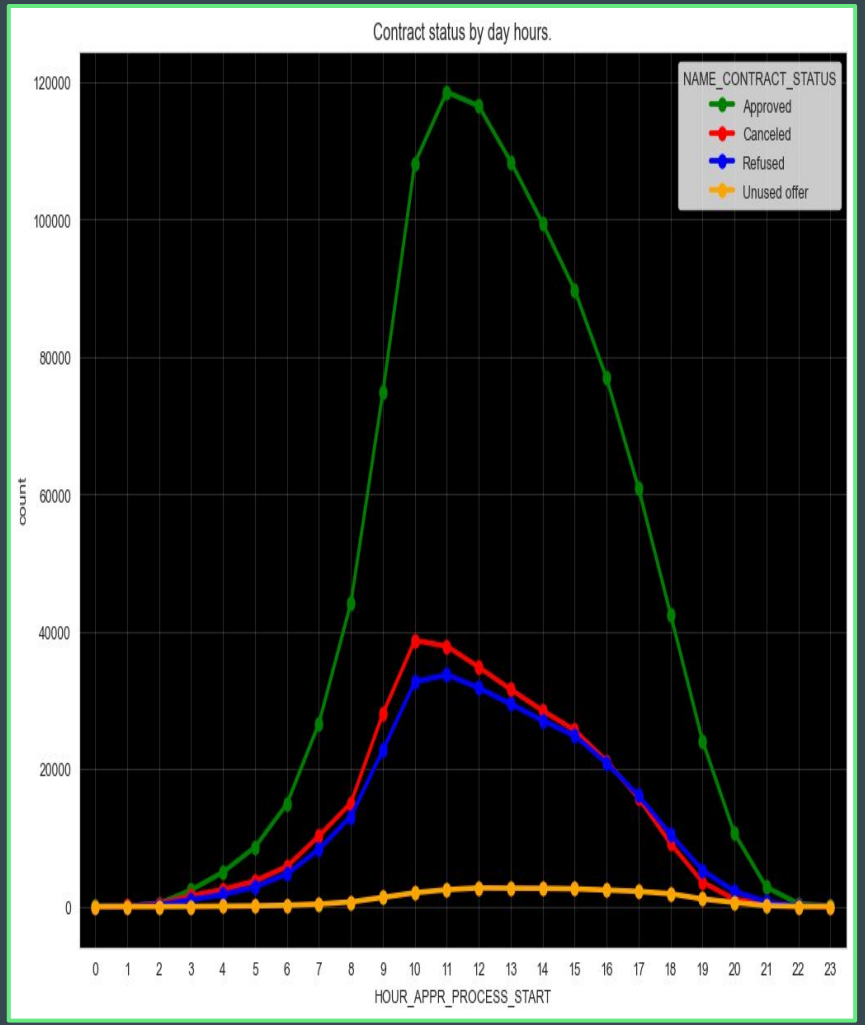

NAME_TYPE_SUITE
Unaccompanied
59.9%
Group of people
Other_A
Other_B
Children
Spouse, partner
Family
0.3%
1.7%
2.1%
3.7%
7.9%
25.1%
count
NAME_CLIENT_TYPE
Repeater
73.7%
XNA
Refreshed
New
0.1%
8.1%
18.0%
count

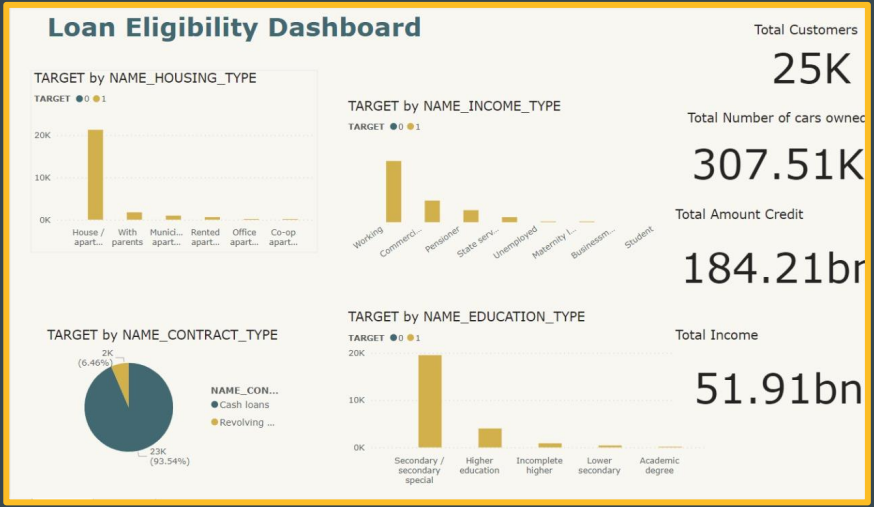

client requesting insurance
0.0
66.7%
33.3%
1.0
count
interest rates
XNA
31.0%
low_action
5.5%
low_normal
19.3%
high
21.2%
middle
23.1%
count

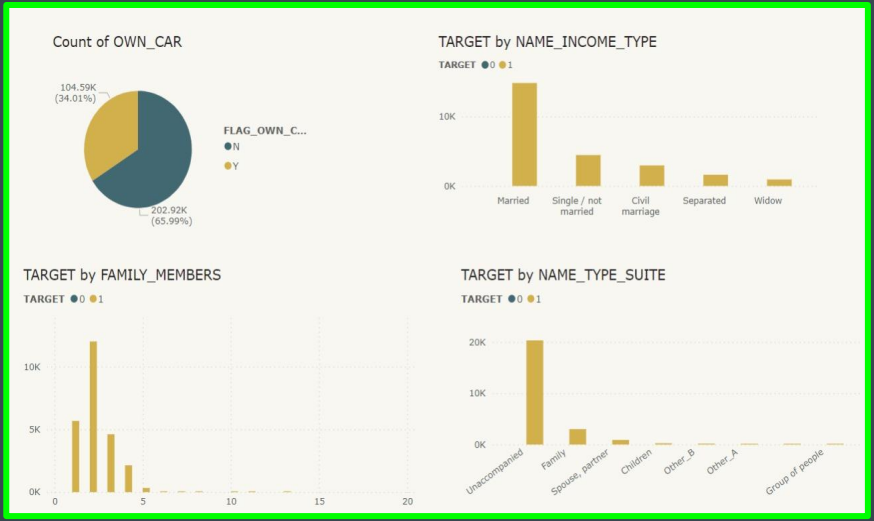

Contract status by weekdays
160000
140000
120000
100000
80000
60000
40000
20000
0
count
NAME_CONTRACT_STATUS
Approved
Canceled
Refused
Unused offer
FRIDAY
MONDAY
SATURDAY
SUNDAY
THURSDAY
TUESDAY
WEDNESDAY
WEEKDAY_APPR_PROCESS_START

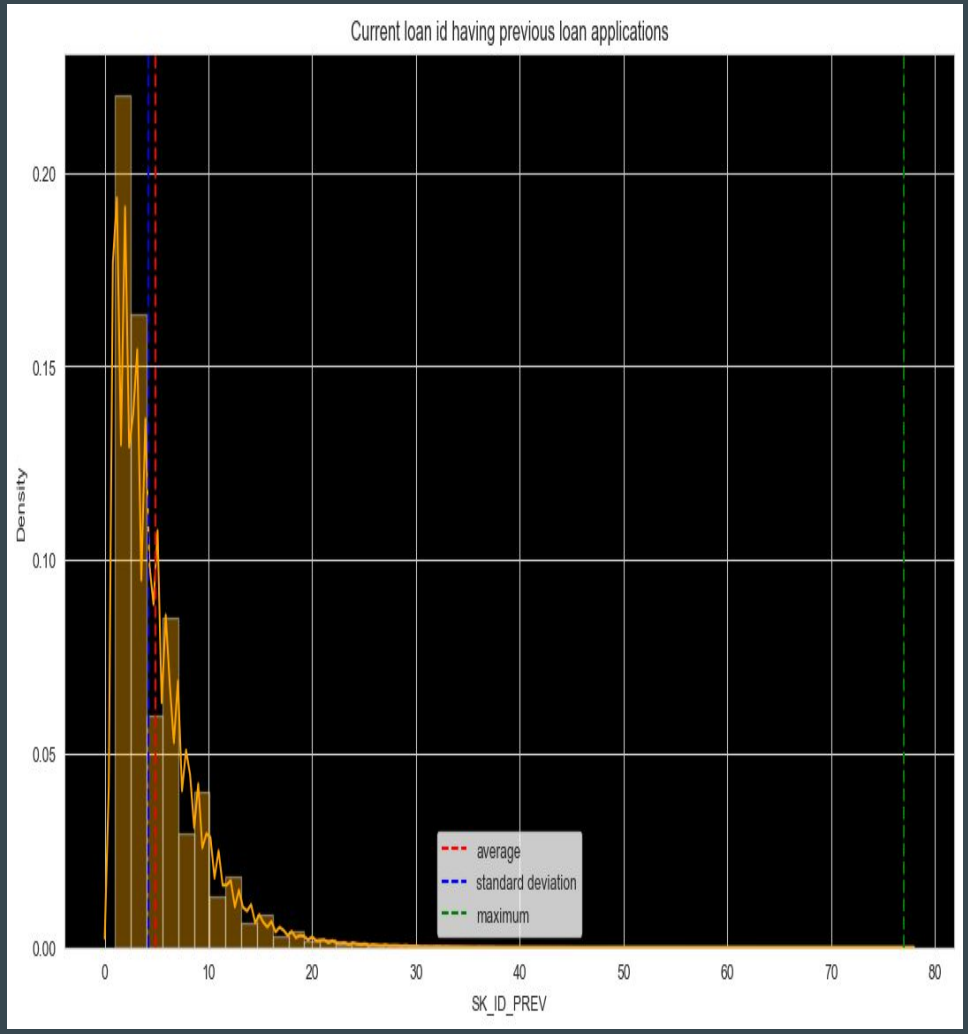

Popular goods for applying loans
percentage
0
10
20
30
40
50
XNA
Mobile
Consumer Electronics
Computers
Audio/Video
Furniture
Photo / Cinema Equipment
Construction Materials
Clothing and Accessories
Auto Accessories
Jewelry
Homewares
Medical Supplies
Vehicles
Sport and Leisure
Gardening
Other
Office Appliances
Tourism
Medicine
Direct Sales
Fitness
Additional Service
Education
Weapon
Insurance
Animals
House Construction
NAME_GOODS_CATEGORY


previous applications portfolio
POS
Cars
Cards
Cash
XNA
27.6%
8.7%
22.3%
count
previous applications product types
XNA
63.7%
9.0%
27.3%
walk-in
x-sell
count


Contract status by day hours.
NAME_CONTRACT_STATUS
Approved
Canceled
Refused
Unused offer
count
0
20000
40000
60000
80000
100000
120000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
HOUR_APPR_PROCESS_START


Loan Eligibility Dashboard
Total Customers
25K
TARGET by NAME_HOUSING_TYPE
TARGET 0 1
20K
10K
0K
House / apart...
With parents
Munici... apart...
Rented apart...
Office apart...
Co-op apart...
TARGET by NAME_INCOME_TYPE
TARGET 0 1
Working
Commerci...
Pensioner
State serv...
Unemployed
Maternity l...
Businessm...
Student
Total Number of cars owned
307.51K
Total Amount Credit
184.21br
TARGET by NAME_CONTRACT_TYPE
2K (6.46%)
23K (93.54%)
NAME_CON...
Cash loans
Revolving ...
TARGET by NAME_EDUCATION_TYPE
TARGET 0 1
20K
10K
0K
Secondary / secondary special
Higher education
Incomplete higher
Lower secondary
Academic degree
Total Income
51.91bn


Count of OWN_CAR
104.59K (34.01%)
202.92K (65.99%)
FLAG_OWN_C...
N
Y
TARGET by NAME_INCOME_TYPE
TARGET 0 1
10K
0K
Married
Single / not married
Civil marriage
Separated
Widow
TARGET by FAMILY_MEMBERS
TARGET 0 1
10K
5K
0K
0
5
10
15
20
TARGET by NAME_TYPE_SUITE
TARGET 0 1
20K
10K
0K
Unaccompanied
Family
Spouse, partner
Children
Other_B
Other_A
Group of people


Current loan id having previous loan applications
Density
0.20
0.15
0.10
0.05
0.00
0
10
20
30
40
50
60
70
80
SK_ID_PREV
average
standard deviation
maximum

Challenges

❑ What we have Done in EDA

- Understanding the dataset
- Multiple processing to identify valuable columns
- Handling non-values columns

❑ Why we are Doing EDA

- To gaining data insights

Thank You

We are now open to any questions or feedback that you may have.