CREDIT EDA ANALYSIS BY AISHWARYA GOSAVI

PROBLEM STATEMENT:

- ▶ The loan providing companies find it hard to give loans to people due to their insufficient or non-existent credit history.
- ▶ When the company receives a loan application, the company has to decide for loan approval based on the application profile.
- ▶ There are 2 types of risks-
- 1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- 2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

BUSINESS OBJECTIVE:

- ▶ This case study aims to identify patterns which indicates 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. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study
- ▶ 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 utilize this knowledge for its portfolio and risk assessment.

STEPS INVOLVED:

- 1. DATA SOURCING (Provided already in assignment)-
 - 1. Application data CSV file
 - 2. Previous data CSV file
 - 3. Application Description

2. DATA UNDERSTANDING -

- 1. Importing necessary libraries.
- 2. Checking each column, Index, header etc.
- 3. Identifying data Quality issues.

3. DATA MANIPULATION -

- 1. Checking the structure of the data.
- 2. Missing value imputation analysis.
- 3. Checking datatypes to date, time etc for ease case of analysis.

4. OUTLIERS CHECK &DATA IMBALANCE CHECK -

- Checking the data for outliers that would cause the analysis to be biased.
- Checking for imbalances, ratio, percentage of imbalance.

5.DATA ANALYSIS -

- Business requirement oriented analysis.
- Correlation between columns
- Univariate Analysis
- Bivariant Analysis
- HEATMAP

6.CONCLUSION-

- How EDA is helpful for taking future course of action for business.

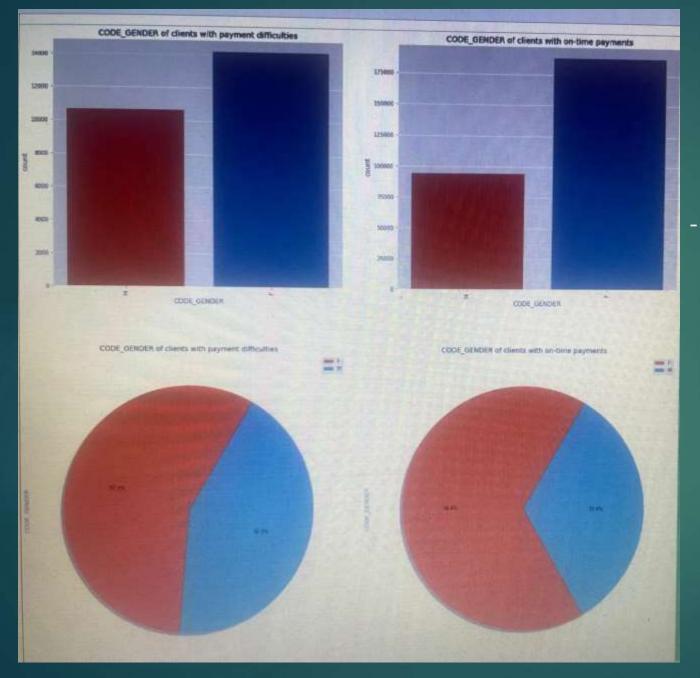
Checking imbalance ratio of TARGET variable 250000 -200000 -150000 -100000 -50000 -0 -0 - On-time Payment 1 - Payment Difficulty

CHECKING IMBALANCE FOR TARGET COLUMN 'Target'-

- ▶ The ratio for people who have difficulty paying to on-time payers is 11.38.
- ▶ 1 in every 11 applicant has payment difficulty.
- print(df2['TARGET'].value_counts(norm alize=True)) shows

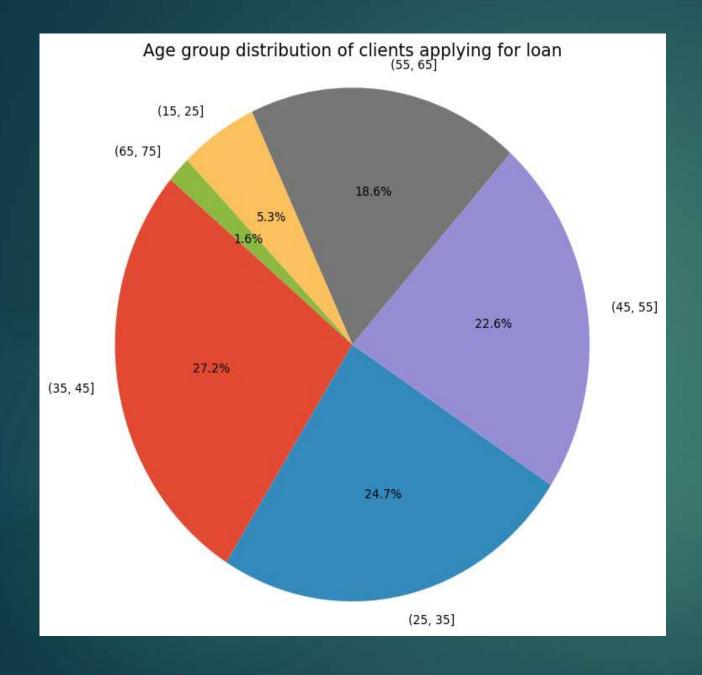
TARGET

- $0 \quad 0.919271$
- $1 \quad 0.080729$



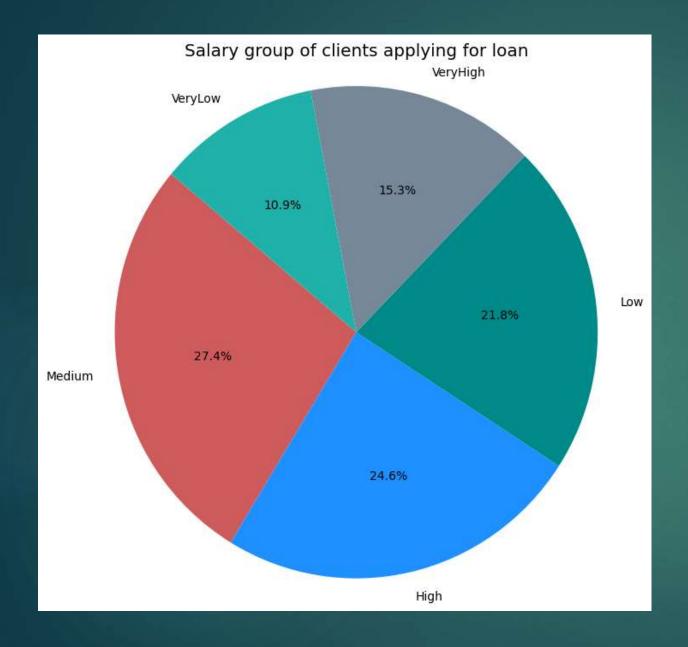
GENDER FACTOR

- Male applicants are more defaulting than Female factor.
 - Female have processed more loans comparison to males.
 - Female clients who are working are more likely to make on time payment.



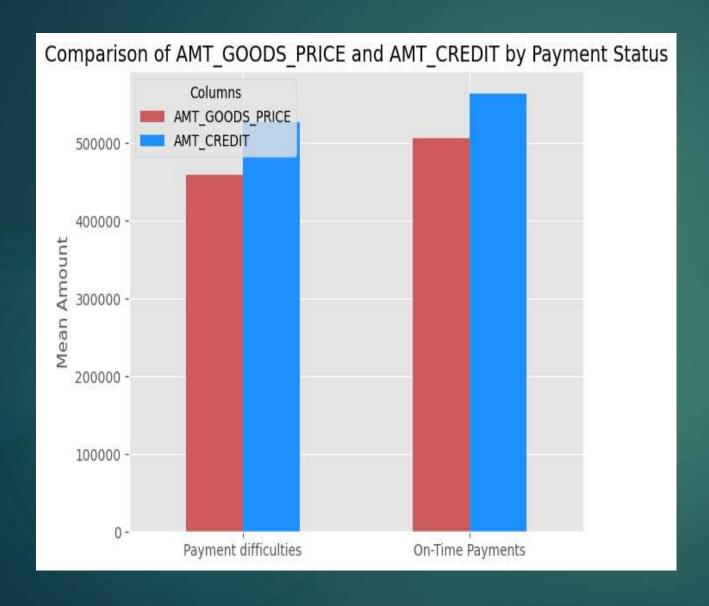
Age is factor when determining whether client can pay on time

- For YEARs_BIRTH between 20 and 40,there are more clients with payment difficulties.
- 35-40 age group is the largest group of age applying loans.
- -Clients in the age range 30-40 and 40-50 is ideal target for loans.

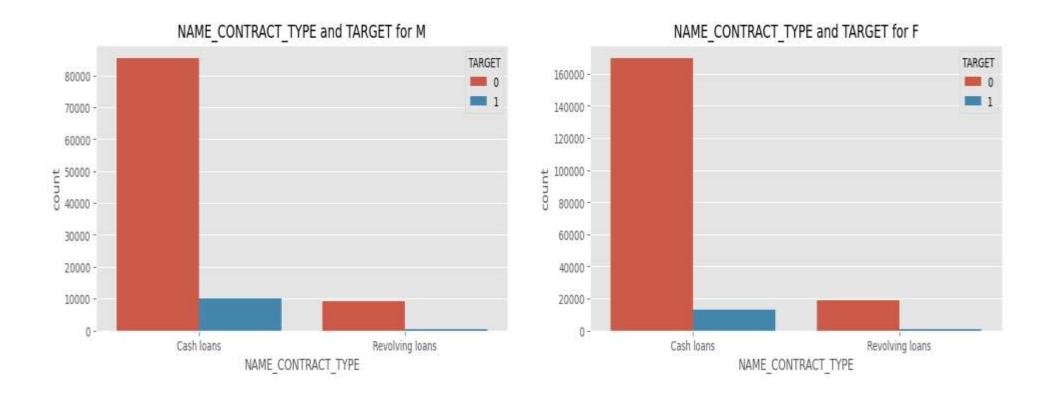


Salary group of clients applying for loan-

'Medium' income group is the largest group applying for loans. Followed by "High" income group. "Very Low" income group is the smallest group applying for loan.



- On average, individuals with ontime payments have both higher AMT_GOODS_PRICE and AMT_CREDIT amounts compared to those with payment difficulties.
- The gap between AMT_GOODS_PRICE and AMT_CREDIT is more pronounced for those with payment difficulties, indicating a potential risk factor where higher credit amounts relative to goods prices might correlate with payment issues.



- Cash Loans Dominance: Both males and females predominantly have cash loans, with a significantly higher count compared to revolving loans.
- **Default Rates**: For both genders, the proportion of defaults (TARGET=1, shown in blue) is higher in cash loans compared to revolving loans.
- **Gender Comparison**: The overall count of loans (both cash and revolving) is higher for females than males, indicating a higher loan acquisition rate among females.

NO. RECEIVED. THE RESERVE OF THE PARTY OF THE D Q R. 1000, 1100, 1120, 1100 COOK 1 40 1918 HOES NO CNT CHILDREN 4 0.001.001.000.000.000.000.100.100.0 AMT_INCOME_TOTAL .000.03F.000.03F.00 0.000.030.00 ARIT CREDIT 21.000 750.00 ON ON DE CONTROL DE CO 030.030.030.000.03 ANT ANNUTTY -TO THE PARTY OF THE REAL PROPERTY OF THE PARTY OF THE PAR CHARLES CONTRACTOR AWY CODDS PRICE 0 MOTERN OF THE OWNER OF THE OWNER I DOMESTIC ON OR DO SEVEN SEVEN SE VALUE ON DESCRIPTION OF THE PROPERTY OF THE OWN DESCRIPTION OF THE ON OWNERS WHERE A COLUMN TO SERVICE DE LA CONTRACTION DEL CONTRACTION DE LA CONTRA 0.75 REGION POPULATION RELATIVE # 050.030.00 016.030.440.45 1 0 020.020.030.030.030.01 170.010.030.030.030.030.030.03 000.030.0E0000000 BUTTO DE LA CORTA DE COMPANSA DE CONTRACTO DE nio set nin net m DAYER BUREFUL - CO. -0.5 dt 3 dt 030 030 040 0 dt 030 040 040 040 040 441 2 dt 2 5 5 6 0 030 040 040 040 040 030 030 BO 65 5 DHD 140 02 ON OR OR OR OR OR OR OR OR O'CO. DAYS EMPLOYED 4003 1. OR 2 D DE 020 D B BCD DON ORD ORD ORD ORD 251 2 IS BBO B D DON O D DEB DOG ORD 25 250 000.030.030.030.010.200.201.000.21 .00 1 0.000.040.03 030,030,030,030,000 030 030,000,000 200,100 00 DAYS REGISTRATION . TO 200, 0 30, 1 40, 1 30, 1 100, 1 10, 1 2 . HE .- O. C. OR . OR .- OR .-0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.000 190 0 0 0 0 0 0 DAYS RD PUBLISH FLAG_MOBIL B 8130 00 -0.030 000 546 048 157 V n20.62 ng - 1-0.0 0.00 CO 0.00 D 0 FLAG EMP PHIDNE -0.50 or formemaned as "eewer - comen except on one CONTRACTOR OF THE PROPERTY OF SECURE OF SECURE OF THE PROPERTY PLACE WORK PHONE 4 THE OWNER ON OWNER 140 2 MILES OF 000 18 010 030 06 500 00:000 040 000 000 000 00:00 00:00 030,000,000,000,00 PLAD CONT MODILE 4 0.000 020.000 030.000 040.000.000 040.000.030.030.000.0 0.033.030.00.00 FLAGS PHIDINE OR OR OR OR OTHER WINDS 111-0 OF OUR DAD ON DED DED DES UND DOS DOS DED DE PLAG EMAIL 010.010.01 0.010.030.00.040.020.03 030-030-040-030-03 REGION NATING CLIENT -0 040 040 040:070 070:440:000 000:100:00 0.00.030.000.010.02 RECICIN MATING CLIENT W CITY -DHO CHELO EL DINI DIO 4 NO DEL 1100. S NEL 11 THE COLUMN TWO TO SEE A SECTION ASSESSMENT AND ASSESSMENT OF THE COLUMN trastina itaio oro na HAD SHELDED DESCORE THE RESERVE OF REPORT OF THE P. 020.00,010,030.01 HOUR APPR PROCESS START & REG_REGION_NOT_LIVE_REGION 4 D20.040.000.020.0 0.25 THE REPORT OF THE PARTY OF THE ORGANIST BUREUM COMPAN. DATE OF THE OWNER. PERS PRESIDEN NOT WORK PROJON 030 090 DHLU31 (I) LIVE REISIGN NOT WORK REGION -OTHER DESIGNATION OF THE OWNER, OF THE OWNER, OWNER MARKET OF OUR OUR WILLIAM SHOULD UNKNOWN OVER A CO. COMPOSITE DE COMPOSITION DE COMPOSIT REG_CITY_BOT_LIVE_CITY -6 1 to ann out take own own own and min of 1 COOR CHESTORY OF THE PARTY 10 to 0 000 010 W F 0 (0) REG. CITY, NOT, WORK, CITY 4 OR SEE CLE ONLINE DOO DOO BED 100 DEC OU LIVE CITY NOT WORK CITY 4 050 000 020 010 020 020 120 200 000 0 030 1 8 010 030 03 DESCRIPTION OF THE PARTY OF THE PROPERTY OF NOTICE AND ADDRESS OF SECTION AND ADDRESS OF A PROPERTY OF COLORO DE COLORO DE PERSONA DE LA COLORO DE CO EKT_SOURCE_2 -000 OD 030 090.03 000 000 0 m n to 000 000 000 000 000 000 000 000 13.1343 OBS 30 CNT SOCIAL CIRCLE -0.000 OCD 0.000 OCD 0.000 OCD 0.000 OCD DEF_10_CNT_SOCIAL_CHICLE 4 OND C 2010 OND D 0 10 OND C 2010 OND C COND C 2010 OND C 2010 O C 2010 OND C 2010 OND C C 2010 OND 000,000,030,030,0 ON'S NO CAT SOCIAL CHICLE -0 OF O DE COMO DEC CESO DE COMO DE COMO DE COMO D20.00.000.00.00 and our ore one one one one one one one one DEP 60 CHT SOCIAL CIRCLE 4003 ON ON OH OH OH OR CO. ON OH OH ON ON OH OH OH O' 030 040 030 030 040 040 040 050 030 030 030 030 0 DAYS LAST PHONE CHANGE -0 0.000.001 101 100 10.000 10.000 10.00 0 No.000,000,000,00 FLAG DOCUMENT 2 -0 U O PLOCO OF BOOK OWN O PLUTO ON U SO U OWN ONE DED AND U FLAG DOCUMENT 3 010.000.001110:040:040.130.230.000 0.230(440.030.530.3) 0.00 O (0.00 PLAG DOCUMENT 4 -0.25 PLACE DIDCUMENT S 4 THE OWN THE PART OF THE OWN THE PART OF THE а одината т т са вар ови и дължи при при и по оправи пред нар изроди однови ови у LOCALD-ED DED. COM. 113 many closed many to be those or we many their those or we many their t 130 030 030 030 030 030 030 031 0.23 011.000.000.070.03 PLAD DOCUMENT & 4 -0.620 1 Trom 010 0 Trom 020 0 Crom 020 0 Crom 0 Trom 1 Trom 0 Crom 0 Cr HIRITINA 100 P. DITE FLAG DOCUMENT 7 -PLAD DOCUMENT # 4 CONTRACTOR - CHORD CONTRACTOR OFFICE EDWINDLING OR CONTRACTOR OF COURT OF CHARGE SECTION OF THE PARTY. FLAB_DOCUMENT_S -000.000.030.030.020.030.030.030.030 0.000.000.000.0 PLACE DICHELIMENT AC PLACE DOCUMENT 11 0 ON OWNER OWNERS ON OWNERS ON O -0.50FLAG DOCUMENT 12 FLAG DOCLIMENT, 19 -0-III ODE THE THE OWN DED THE REAL PROPERTY OF THE PERSON CHARLEST CONTRACTOR OF DESCRIPTION OF THE OWN THE PROPERTY OF THE OWNER. PLAG DOCUMENT 14 0.00 A DISCOUNT OF THE CONTRACT OF 000 030 000 030 Or PLAG_DOCUMENT_15 -ON ON ON ON UND HOLD ON HER STORE HER ON ON ON ON ON ON ON ON ON 20 (ET. 20 (ET PLAG DOCUMENT IN -0 DOT COLOR DE COLORDO D .030.030.0m.030.01 om nation and national and national nation of the contract of ONE OWN DOD ONLOG FLAG DOCUMENT-17 -0.01 1.000 000 000 000 000 000 000 000 000 0 FEAG DOCUMENT 18 -0.03 DOO ON HO GROWN HO ON DW. ON H COMPOSITION DED. ONC. UK PLACE DISCUSSENT 19 -0.00 DAY OND OWN USED ON USE USED WITH USED ONE OWN OWN USED USED USED USED OWN OWN OWN OWN. OWN GOOD OWN U FLAG DOCUMENT 20 4 000.000.000.000.00 1000.000.000 - 140.000 H | 1 PLACE DISCUMENT 21 48 TO SEE AND 15-E WITH MAD ONLY IN MICH. WHEN IN ANY ONLY ONE DAY OF THE WAY ON THE CASE OF THE CASE OF DOMESTIC AND ADDRESS OF THE PART OF THE PROPERTY OF THE n (1 5 to 10m (10m (10m (1) YEARS_BIRTH 4 281-003-1-01020-1-0102-000-500a serie um aso per les asoles asoles asoles de zerie. Il a qualitate de l'adja t YEARS EMPLOYED -0.03 0.130 CD 000.00.030.020.020.000.1 0.000 0.000 100.02 030,030,030,030,030,030,030,030,030,000 VEGRA, RESISTRATION -0.03 HAME SHE WAS THE OWN THE PROPERTY OF 1 19,000,000 to 0 nm 100,100 nm nm nm nm nm 0 00 00 00 00 nm nm nm nm nm nm nm nm .m/--- n.non.nsp.nr CONTRACTOR OF THE OWN DESCRIPTION OF THE OWN YEARS LAST PHONE CHANGE 4 8 0 0 0 5 0 8 5 8 LISE DOCUMENT B RUS DOCUMENT 3

Last slide of Heatmap insight

- ▶ The heatmap displays a correlation matrix, with dark green indicating strong positive correlations and white indicating no correlation.
- ► Notable strong correlations include OBS_30_CNT_SOCIAL_CIRCLE with DEF_30_CNT_SOCIAL_CIRCLE, and AMT_GOODS_PRICE with AMT_CREDIT.
- ► Weak correlations are observed between FLAG_OWN_CAR and CNT_CHILDREN, as well as the TARGET variable with most features.

Conclusion

- ► Chances of client having payment difficulty All the below variables were established in analysis of Application data frame as leading to default.
- ▶ Checked these against the Approved loans which have defaults, and it proves to be correct Medium income 25-35 years old s, followed by 35-45 years age group Male Unemployed Laborers, Salesman, Drivers Own House No Other IMPORTANT Factors to be considered No of Bureau Hits in last week. Month etc − zero hits is good Amount income not correspondingly equivalent to Good Bought − Income 'Low' and 'High' is a concern Previous applications with Refused, Cancelled, Unused loans also have default which is a matter of concern.
- ▶ This indicates that the financial company had Refused/Cancelled previous application but has approved the current and is facing default on these. Credible Applications refused Female applicants should be given extra weightage as defaults are lesser.
- ▶ Students and Business mean have no problem in repayment of the loan Previous applications with Refused, Cancelled, Unused loans also have cases where payments are coming on time in current application. This indicates that possibly wrong decisions were done in those cases.