

EXPLORATORY DATA ANALYSIS - PANDAS

AGE

	Default
age_group	
middle age borrowers	0.118453
old borrowers	0.060629
young borrowers	0.193859

AGE:

1. Lenders often perceive younger borrowers (21-30) as higher risk due to limited credit history,
2. while older borrowers (51+) may face stricter eligibility requirements and potentially higher interest rates due to nearing retirement.
3. Mid-age borrowers (31-50) generally have more established credit and income, making them more attractive to lenders.

INTEREST RATE

	min	mean	max
age_group			
middle age borrowers	2.01	15.921583	25.0
old borrowers	2.00	16.016339	25.0
young borrowers	2.00	15.817147	25.0

	min	mean	max
age_group			
middle age borrowers	2.0	13.192753	25.0
old borrowers	2.0	13.310854	25.0
young borrowers	2.0	12.921670	25.0

INTEREST RATE:

1. For defaulters we can see that the interest rate is higher than global average of interest rates which can be one of the reasons behind defaulting.
2. For non-defaulters we can see relatively lower interest rates. This can be one of the reasons that they didn't default

INCOME

	Default
income_group	
high income	0.090548
low income	0.179581
lower-mid income	0.109399
upper-mid income	0.096820

INTEREST RATE:

1. Most of the defaulters are from low income group.

CREDIT SCORE vs MARITAL STATUS vs DEFAULTERS

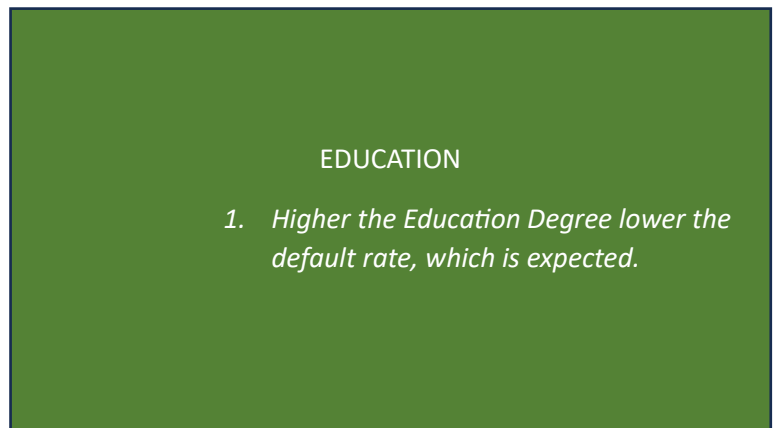
		Default
credit_score_group	MaritalStatus	
exceptional	Divorced	0.105494
	Married	0.085880
	Single	0.103258
fair	Divorced	0.122221
	Married	0.104071
	Single	0.116486
good	Divorced	0.116173
	Married	0.093843
	Single	0.108953
poor	Divorced	0.134495
	Married	0.111925
	Single	0.127897
very good	Divorced	0.114537
	Married	0.093737
	Single	0.106709

CREDIT SCORE vs MARITAL STATUS vs DEFAULTERS

1. for poor credit score people who are either divorced or single have high default rate, same in case of exceptional as well.
2. so married people tend to have settled in life and pay their loans on time and thus maintain the credit score

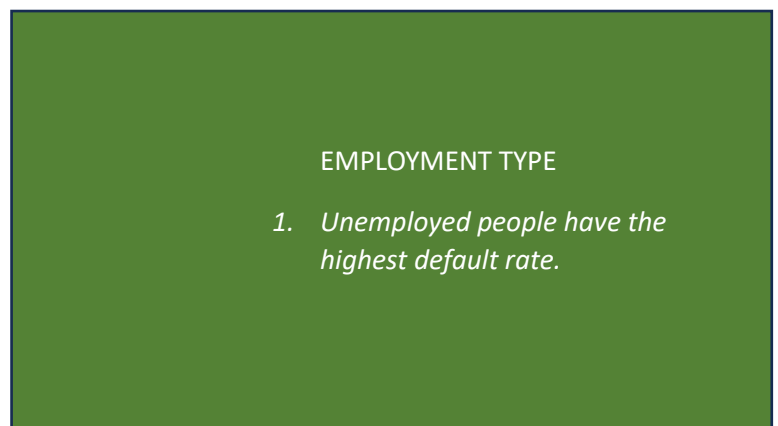
EDUCATION

	Default
Education	
Bachelor's	0.121011
High School	0.128789
Master's	0.108717
PhD	0.105860



EMPLOYMENT TYPE

	Default
EmploymentType	
Full-time	0.094634
Part-time	0.119652
Self-employed	0.114620
Unemployed	0.135529



LOAN PURPOSE

	Default
LoanPurpose	
Auto	0.118814
Business	0.123260
Education	0.118381
Home	0.102348
Other	0.117885

LOAN PURPOSE

1. *People who take loan for the purpose of BUSINESS, tend to DEFAULT more (mostly incase the business doesn't work)*

DTI RATIO

	Default
dti_ratio_group	
high risk	0.119052
low risk	0.111641
moderate risk	0.112081
very high risk	0.121455
very low risk	0.103631

DTI RATIO

1. *here we can see that DTI ratio with ratio greater than 0.5 or 50% which indicates very high risk, have mostly defaulters*

FEATURE ENGINEERED COLUMNS:

LOAN to INCOME RATIO

	Default
LoanToIncomeRatio_group	
high	0.083782
low	0.077209
moderate	0.080392
very high	0.128272
very low	0.075314

LOAN to INCOME RATIO

1. Very high bin indicates that loan amount taken is more than the income, and thus most defaulters are present there.

EMPLOYMENT LENGTH

	Default
EmploymentLengthGroup	
<1yr	0.169076
1-5yr	0.134950
5-10yr	0.089161
10-20yr	NaN
20+yr	NaN

EMPLOYMENT LENGTH

1. People with <1year of employment tend to DEFAULT more.

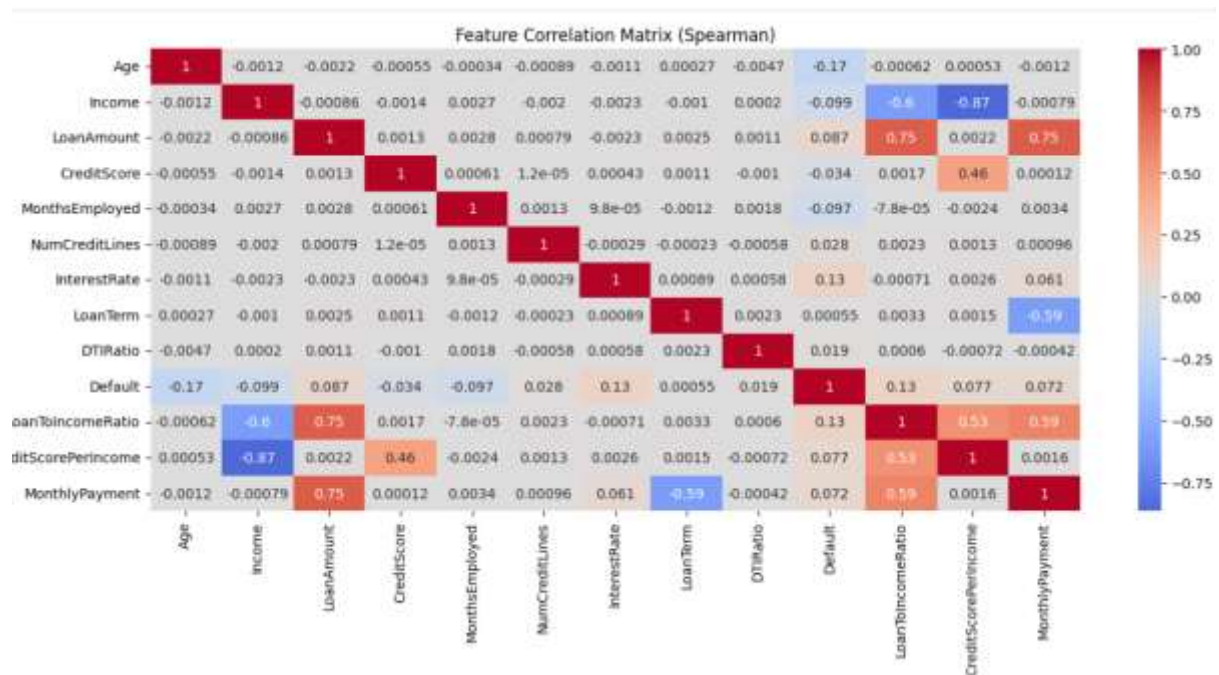
MONTHLY PAYMENT

	Default
MonthlyPayment_group	
high	0.093547
low	0.080092
moderate	0.085563
very high	0.129392
very low	0.071941

MONTHLY PAYMENT

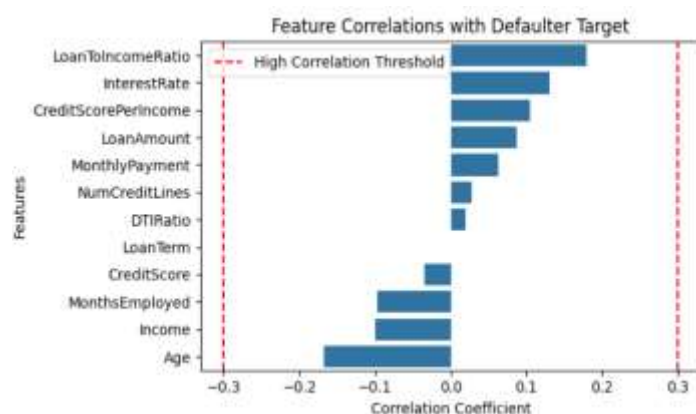
1. Higher monthly payment burden leading to more DEFAULTERS.

CORRELATION



CORRELATION

- Age, income, loan amount and credit score are correlated.



FEATURE CORRELATION with DEFAULTER

- Loan to income ratio, interest rate, age and income play a very significant role.