

Office Hours

BUILD AN AI MODEL TO FORECAST TRENDS AND
DETECT OPPORTUNITIES IN RETAIL BANKING



Exploratory Data Analysis

1. Structure of the Data

a. Number of rows and columns

(17511, 26)

b. Columns in the data

```
['ACCT_OPN_DATE', 'SCHM_TYPE', 'SANCT_LIM', 'COLLATERAL_TYPE',
'LOAN_TENURE_MONTHS', 'REP_FREQ', 'INTTRATE', 'PREPAYMENT_HIST',
'OVD_AMT', 'CUST_DOB', 'MONTHLY_AVG_DEP', 'FD_AMT',
'MONTHLY_AVG_CR_AMT', 'MONTHLY_AVG_DR_AMT', 'MONTHLY_NO_OF_TXNS',
'DISTRICT', 'PROVINCE', 'PROFESSION', 'RISK_RATING', 'DORMANT_STATUS',
'NO_OF_QR_TXNS', 'NO_OF_POS_TXNS', 'FONELOAN_OS', 'CC_OS',
'account_name', 'AS_ON'],
```

Exploratory Data Analysis

2. Data Types

Numerical: Variables that represent measurable quantities.

- a. Continuous
- b. Discrete

Categorical: Variables that represent groups

- a. Ordinal: Has an order. Example: small, medium, large
- b. Nominal. Example: Color(Red, Green, Blue).

Need to be represented in a format that allows for numerical calculations.

Datetime:

Can be treated as either numerical or categorical.

ACCT_OPN_DATE	datetime64[ns]
SCHM_TYPE	object
SANCT_LIM	float64
COLLATERAL_TYPE	object
LOAN_TENURE_MONTHS	float64
REP_FREQ	object
INTTRATE	float64
PREPAYMENT_HIST	object
OVD_AMT	float64
CUST_DOB	datetime64[ns]
MONTHLY_AVG_DEP	float64
FD_AMT	float64
MONTHLY_AVG_CR_AMT	float64
MONTHLY_AVG_DR_AMT	float64
MONTHLY_NO_OF_TXNS	float64
DISTRICT	object
PROVINCE	object
PROFESSION	object
RISK_RATING	object
DORMANT_STATUS	object
NO_OF_QR_TXNS	float64
NO_OF_POS_TXNS	float64
FONELOAN_OS	float64
CC_OS	float64
MAPPED_FORACID	object
MAPPED_CIF_ID	object
MAPPED_OP_AC	object
AS_ON	datetime64[ns]
dtype: object	

Exploratory Data Analysis

3. Remove features

4. Handle irregularities in the data

a. Missing Values:

i. Machine Learning Algorithms make numerical calculations with the data and can crash when there is no value present.

ii. Can't directly remove rows as there may be some valuable data there

b. Duplicates

ACCT_OPN_DATE	0
scheme_type	0
SANCT_LIM	0
COLLATERAL_TYPE	16236
LOAN_TENURE_MONTHS	10449
REP_FREQ	10449
INTTRATE	0
PREPAYMENT_HIST	0
OVD_AMT	10636
CUST_DOB	0
MONTHLY_AVG_DEP	0
FD_AMT	17151
MONTHLY_AVG_CR_AMT	11009
MONTHLY_AVG_DR_AMT	10257
MONTHLY_NO_OF_TXNS	10160
DISTRICT	114
PROVINCE	114
PROFESSION	765
RISK_RATING	0
DORMANT_STATUS	0
NO_OF_QR_TXNS	17004
NO_OF_POS_TXNS	17432
FONELOAN_OS	17409
CC_OS	16668
AS_ON	0

Exploratory Data Analysis

Handle Missing Data

- For categorical data:

Add a new category called unknown or fill in the most used category

- For numerical data:

a. **Remove** if majority of the data is missing unless absolutely needed.

b. **Simple Imputation**

Mean: For data with no outliers

Median: For data with outliers

c. **Advanced Imputation**

KNN Imputation: Find K similar rows to the row with missing data

d. **Use is missing feature**

For some features, logging the missing values or using them as a feature itself could be useful.

Remove:

NO_OF_QR_TXNS

NO_OF_POS_TXNS

FONELOAN_OS

CC_OS

Add unknown category / or the most frequent category:

COLLATERAL_TYPE

PROFESSION

PROVINCE

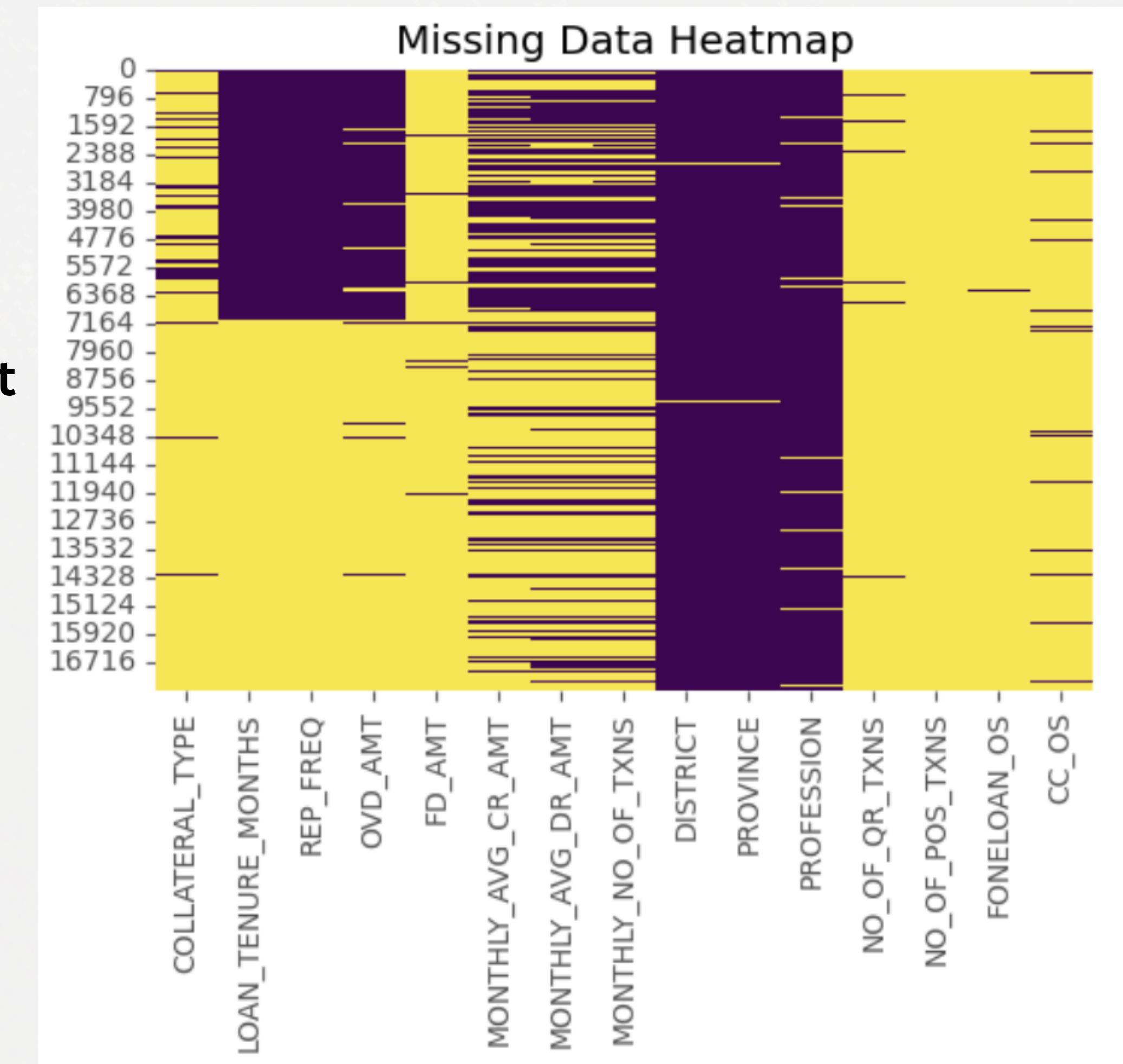
DISTRICT

REP_FREQ

Imputation:

LOAN_TENURE_IN_MONTHS

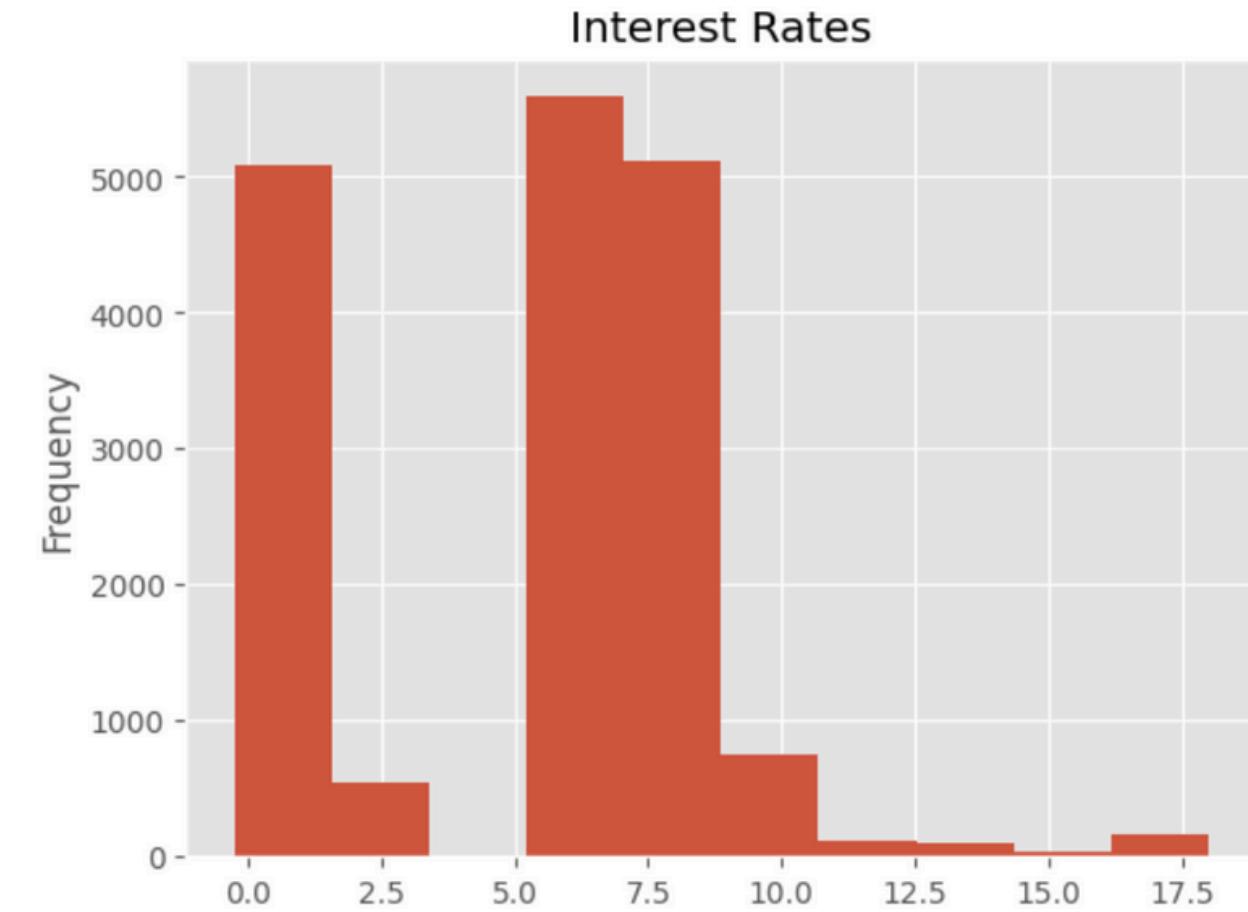
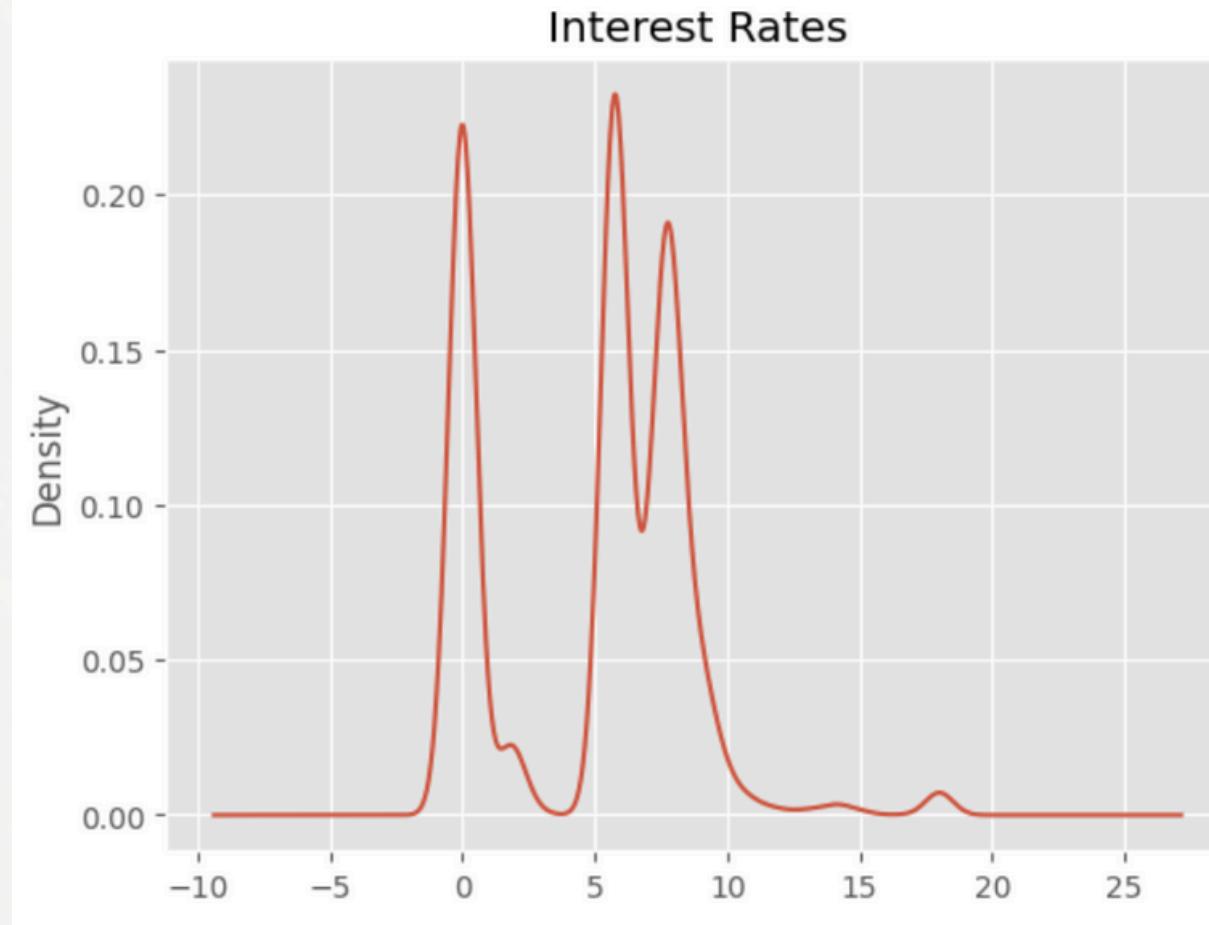
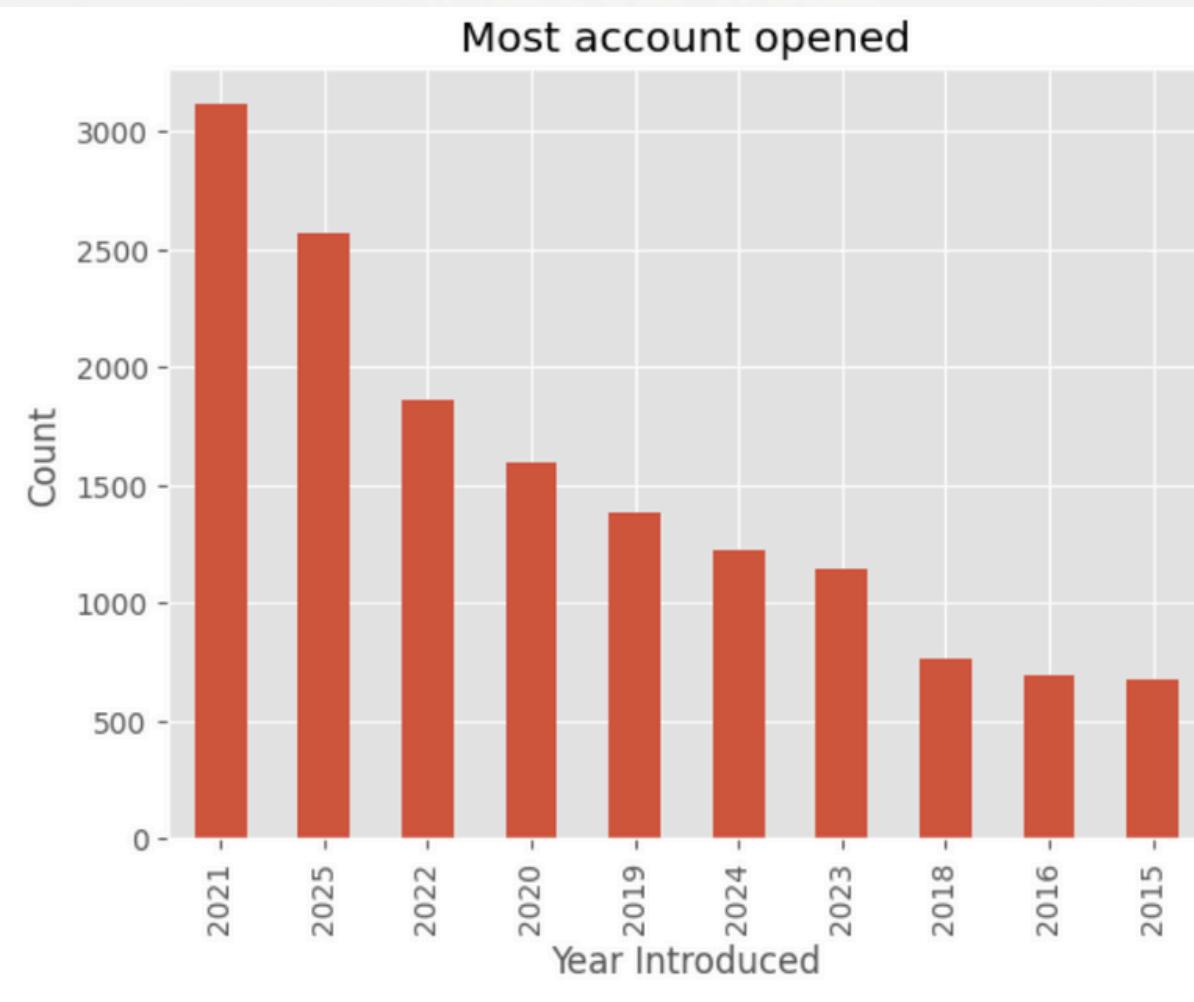
And other numerical datas



Exploratory Data Analysis

1. Univariate Analysis (One variable at a time)

- a. Look at one variable at a time.
- b. Helpful in spotting outliers
- c. Understand the distribution of the data

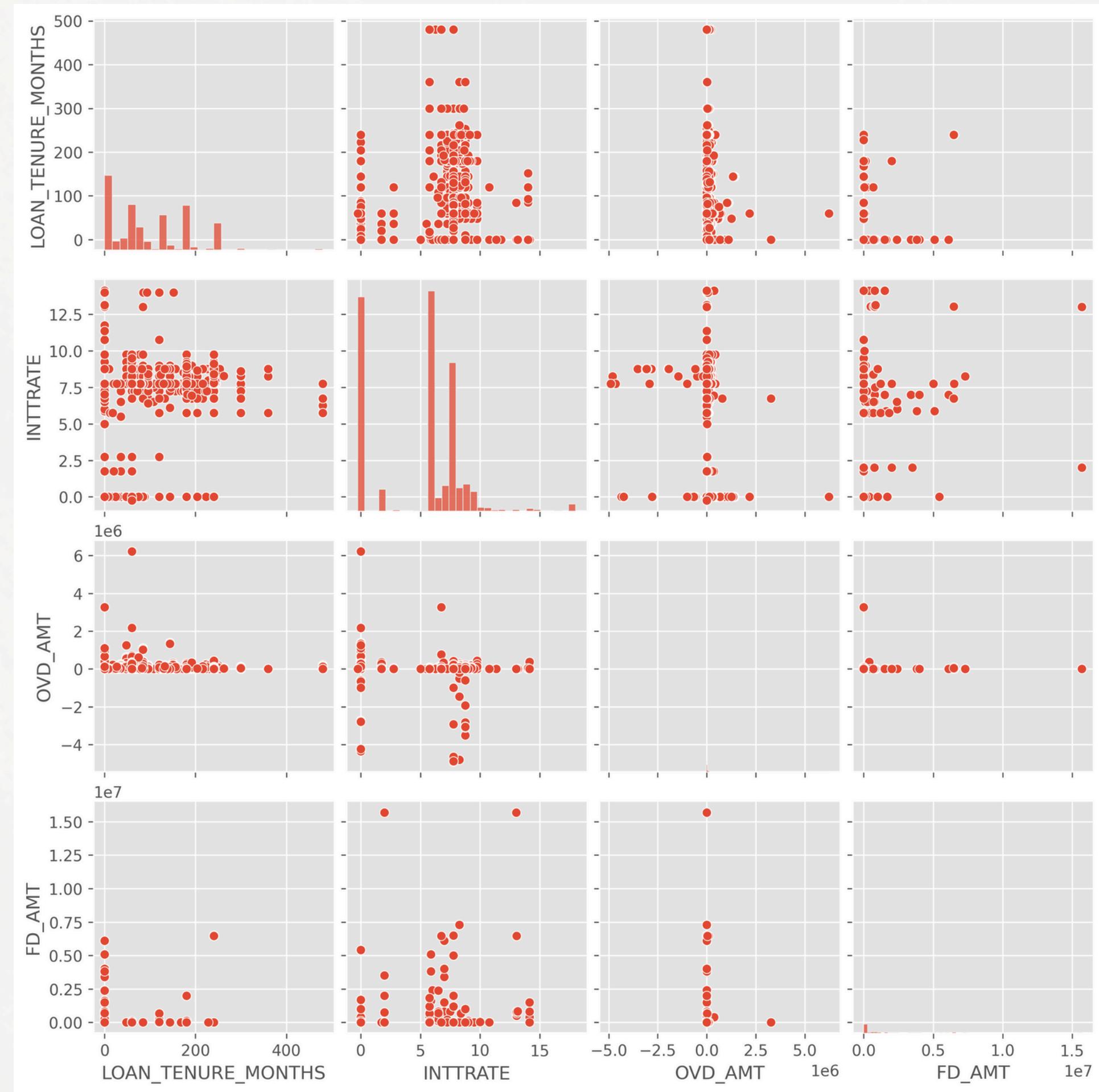


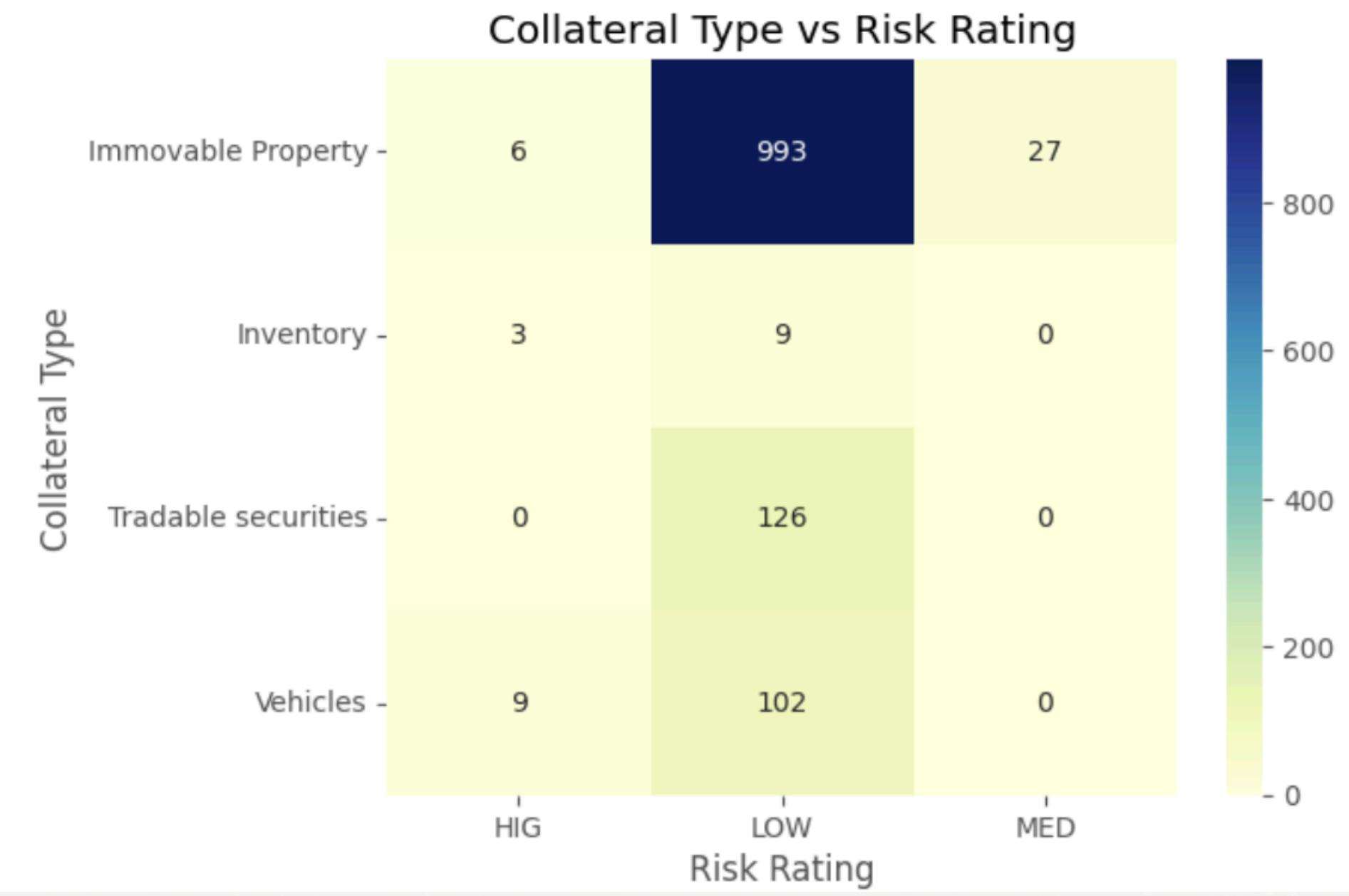
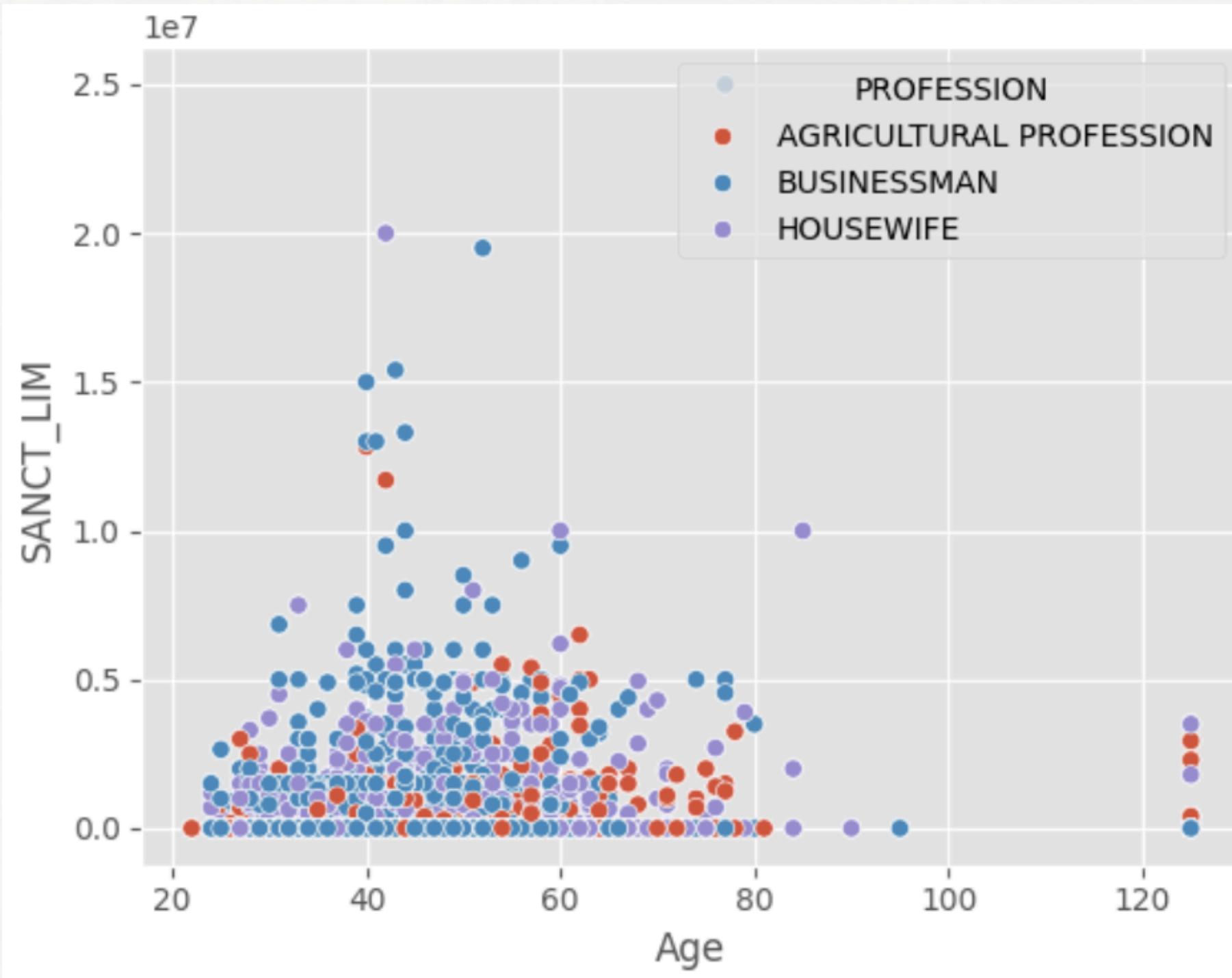
—

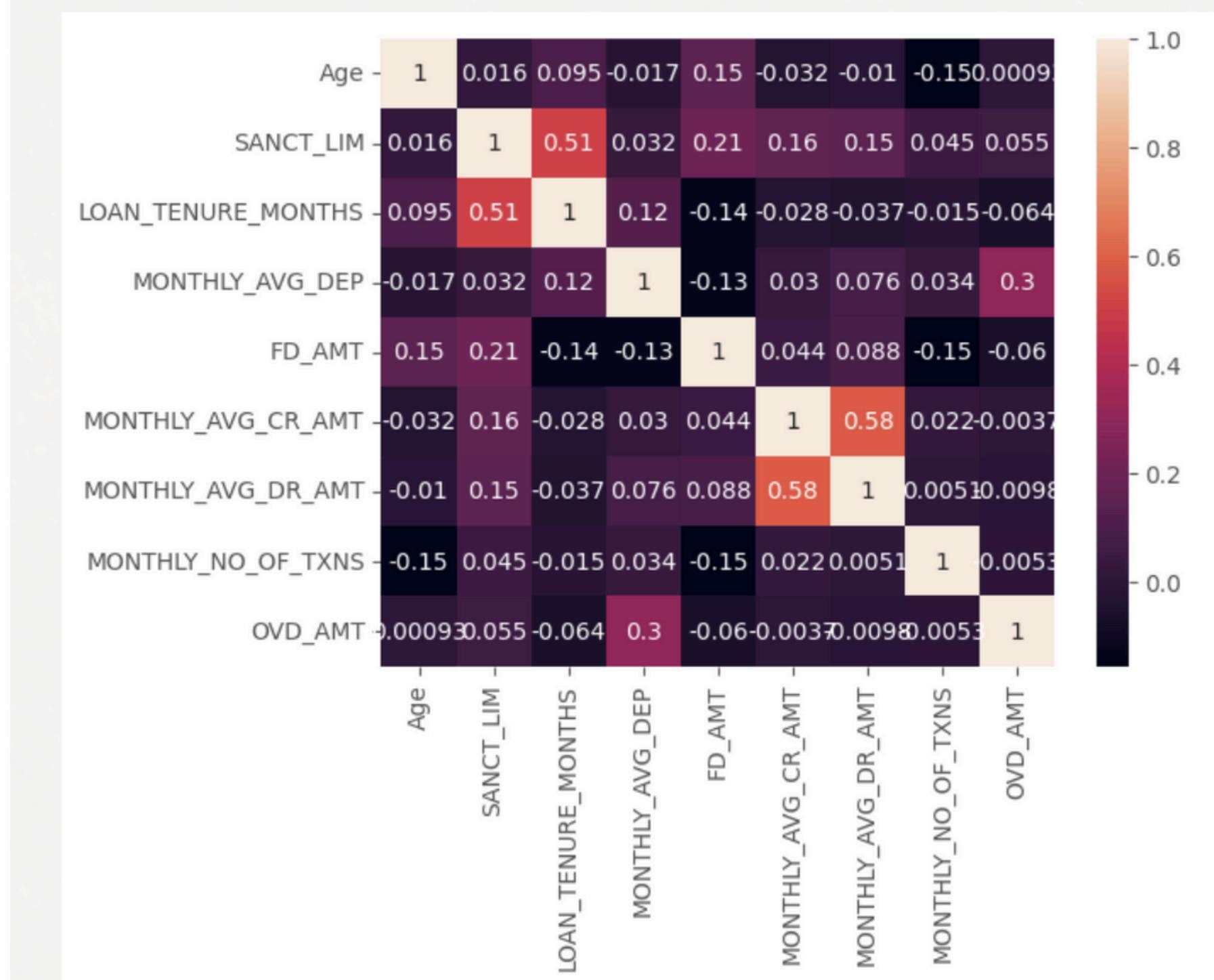
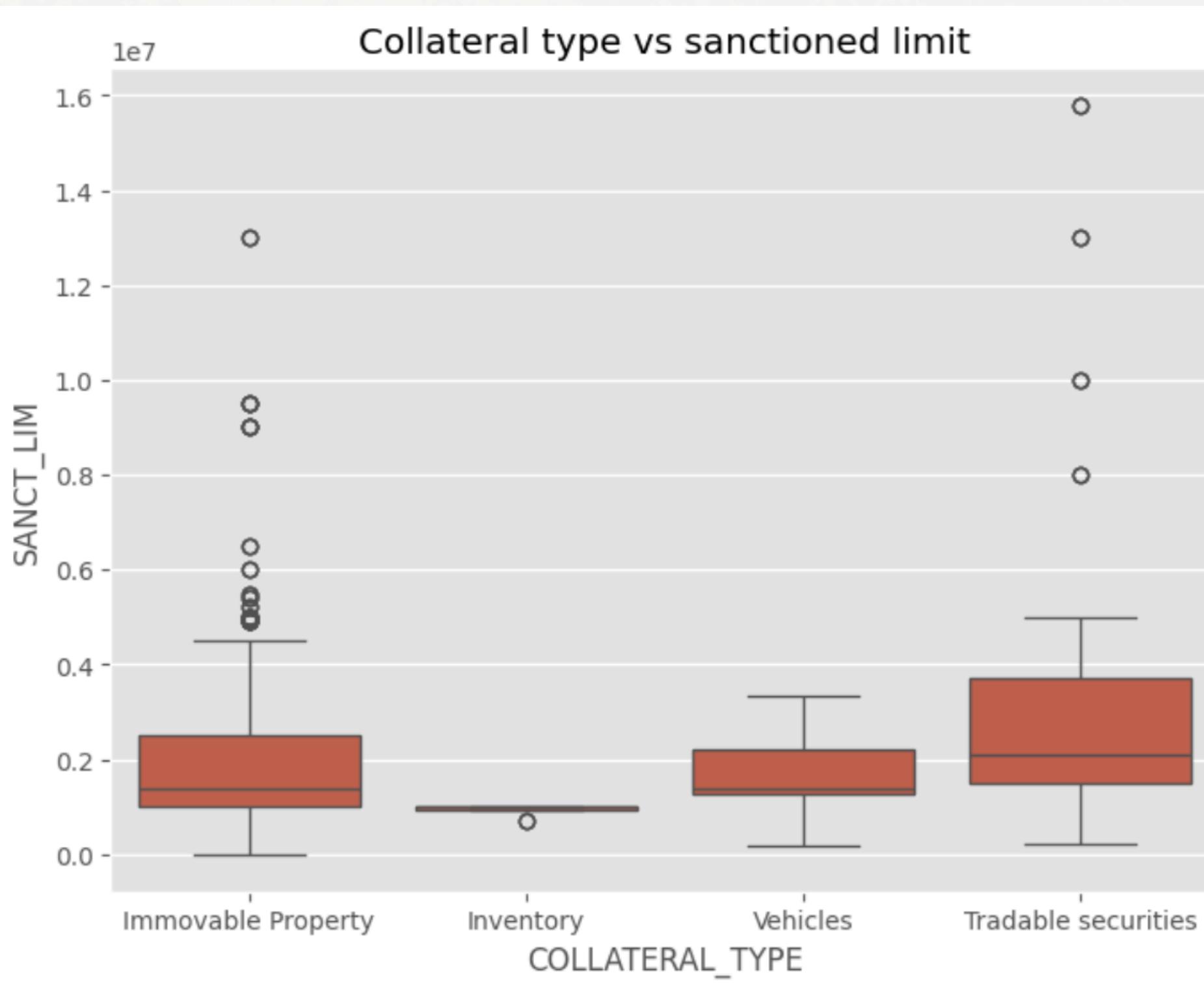
Exploratory Data Analysis

Bivariate / Multivariate Analysis (Finding Relations)

- 1. Find correlation among variables.**
- 2. Find clusters / groups.**
- 3. Check for redundant variables**





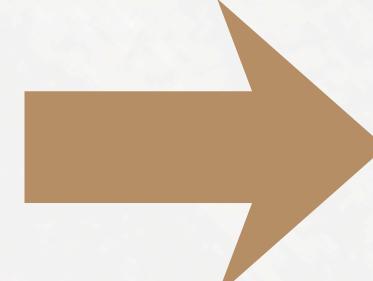


Converting Categorical Data

One-Hot Encoding

Create binary column for each unique category
Can't be used for high cardinality data i.e. with a lot of categories.

A	B	C	D	E	F
Profession		Doctor	Farmer	House-Wife	Engineer
Doctor	→	1	0	0	0
Farmer	→	0	1	0	0
House-Wife	→	0	0	1	0
Engineer	→	1	0	0	1



COLLATERAL_TYPE	COLLATERAL_TYPE_Immovable Property	COLLATERAL_TYPE_Inventory	COLLATERAL_TYPE_Tradable securities	COLLATERAL_TYPE_Vehicles
Immovable Property	True	False	False	False
NaN	False	False	False	False
Immovable Property	True	False	False	False
Immovable Property	True	False	False	False
NaN	False	False	False	False

Office Hours

BUILD AN AI MODEL TO FORECAST TRENDS AND
DETECT OPPORTUNITIES IN RETAIL BANKING

