Project 5 - Bank Loan

Group 3:
Michel Brun – Paul Melin – Zijing Xue

Dataset - Individual Bank Loan

67463 rows × 35 columns

Independent Variable

- Grade & Subgrade
- Inquries
- Public Record
- Loan Title
- Debit to Income
- Employment Duration

Dependent Variable

- Loan Amount
- Interest Rate
- Open account

Challenges

- Enormous database: cleaning and decision
- Understanding: proper nouns in bank industry
- Make hypothesis and test: if the result reject the hypo => another time
- Time management
- Plot with Streamlit: combine the code in spider using container

Process

PLANNING

- ➔ Create GitHub
- ➔ Review dataset
- ➔ Data cleaning
- ➔ Clarify tasks
- ➔ Kanban planning

PREPARATION

- ➔ Understanding of variables
- ➔ Data Cleaning
- ➔ EDA
- ➔ Correlation

VISUALIZATION

- ➔ Make hypothesis
- ➔ Plot
- ➔ Use Streamlit
- ➔ Combine plots from 3 of us
- ➔ Slides

Highlights & Learnings

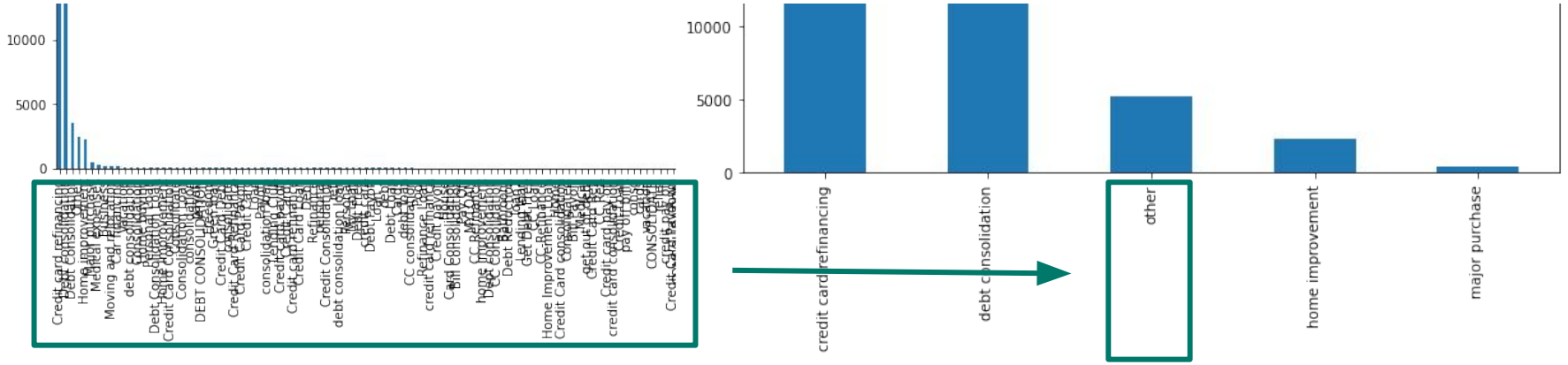
- Deal with Streamlit : Not the same logic of creation then working on Jupyter
- Online Streamlit : Requirements and local/online path
- Kanban : Split the tasks and show the progress of the project

Improvements

- Better understanding of the dataset
- Deeper dive into the relationship among variables
- Have plot with more significant correlation
- Optimize visualisation

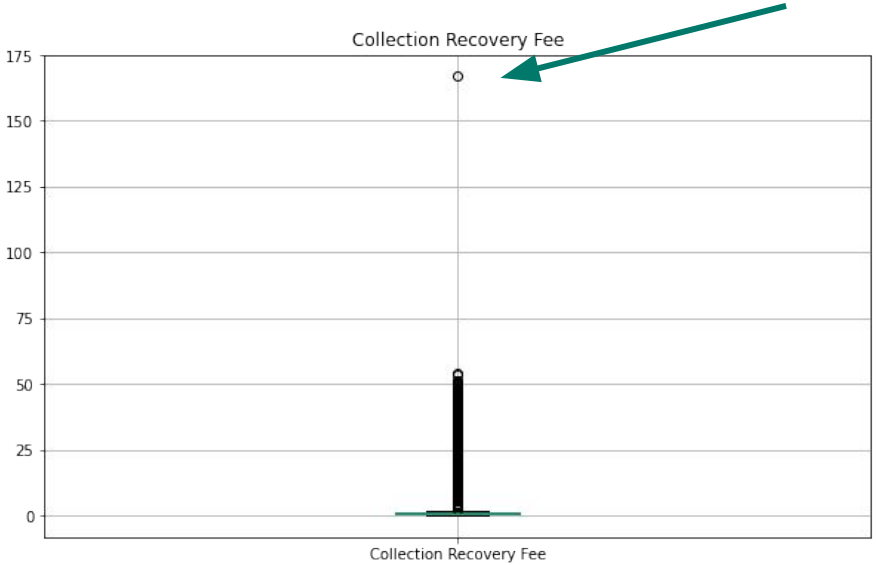
Comparison of the initial and final datasets

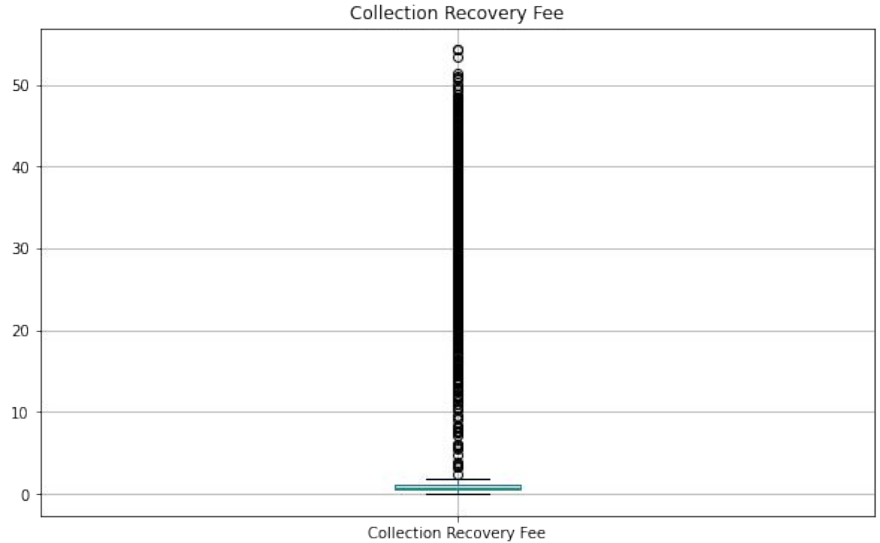
- ★ No duplicates - No missing value - Few columns dropped because no exploitable Data
- ★ Loan Title Cleaning => mixed cases, repetitive definition of same type of title, title with few frequency - group into ‘other’



Comparison of the initial and final datasets

★ Collection Recovery Fee replaced by new Max





Dashboard

Bank Loan Dashboard on Streamlit online

Thank you :)

Question?

Annex 0: Overview

Scatter Chart: Explore Relationship among Grades :

X-Axis

Loan Amount

Y-Axis

Funded Amount

Grade

A

Bar Chart: Average Data per Loan Title :

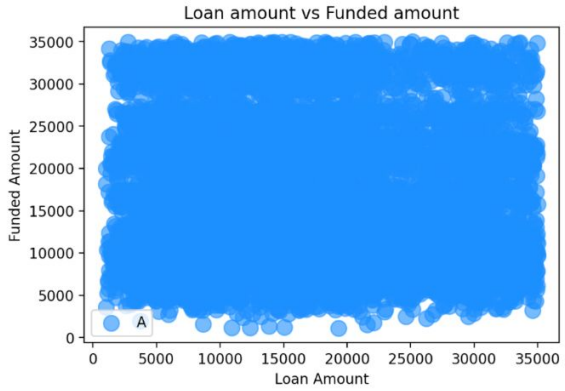
Average Data per Loan Title

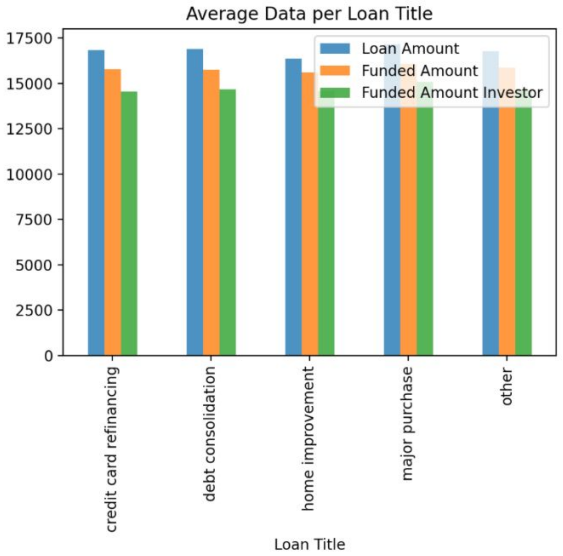
Loan Amount ×

Funded Amount ×

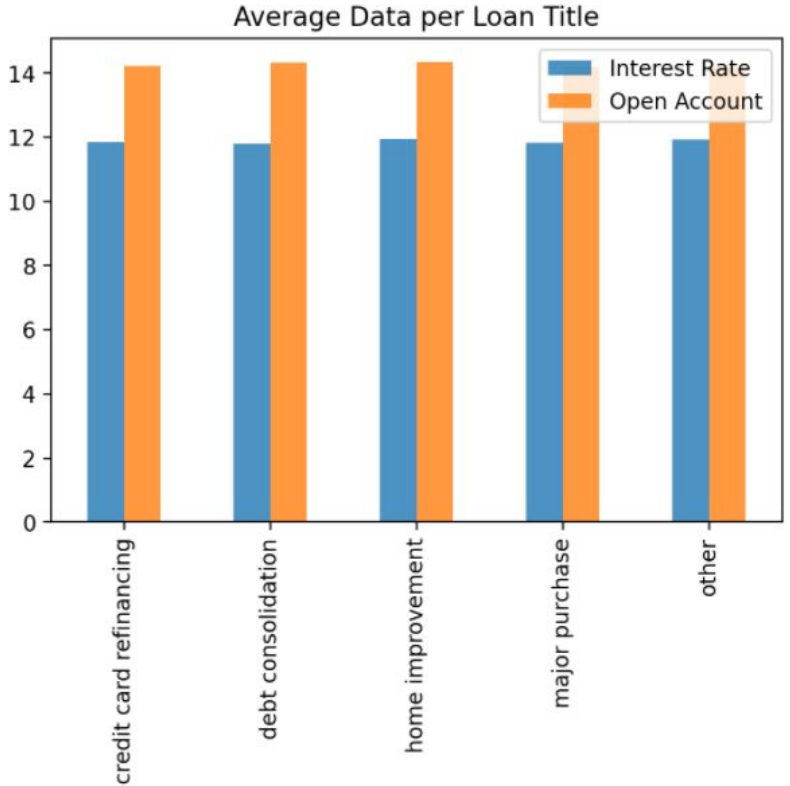
Funded Amount Inv… ×

Bank Loan Analysis

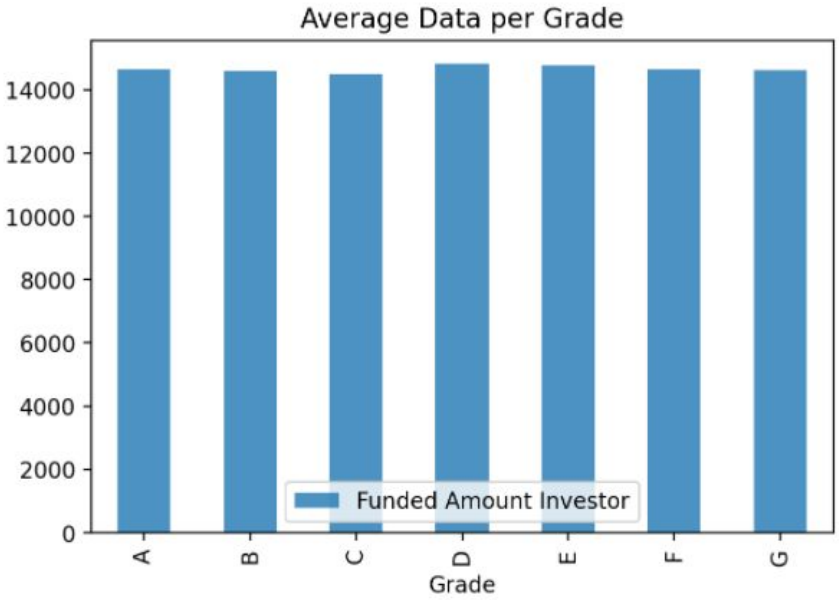


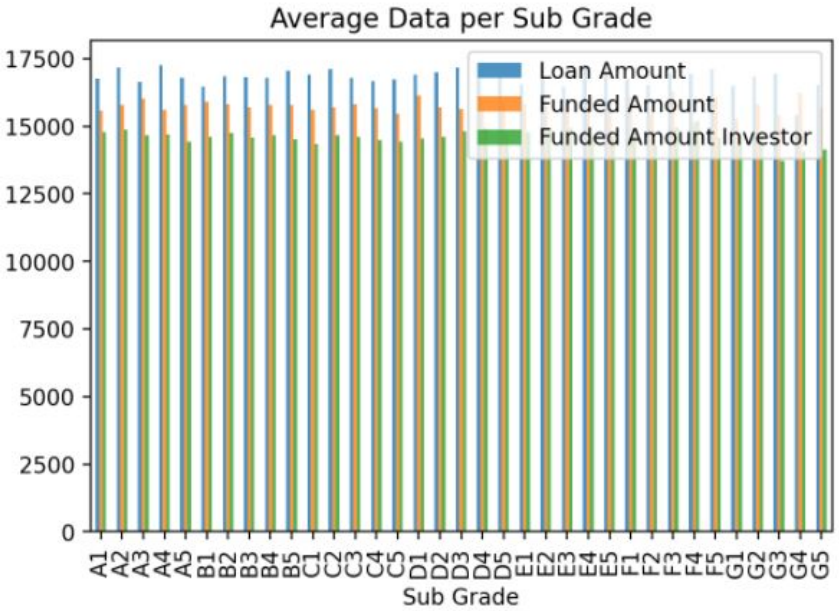


Annex 1: Average Data with Loan Title

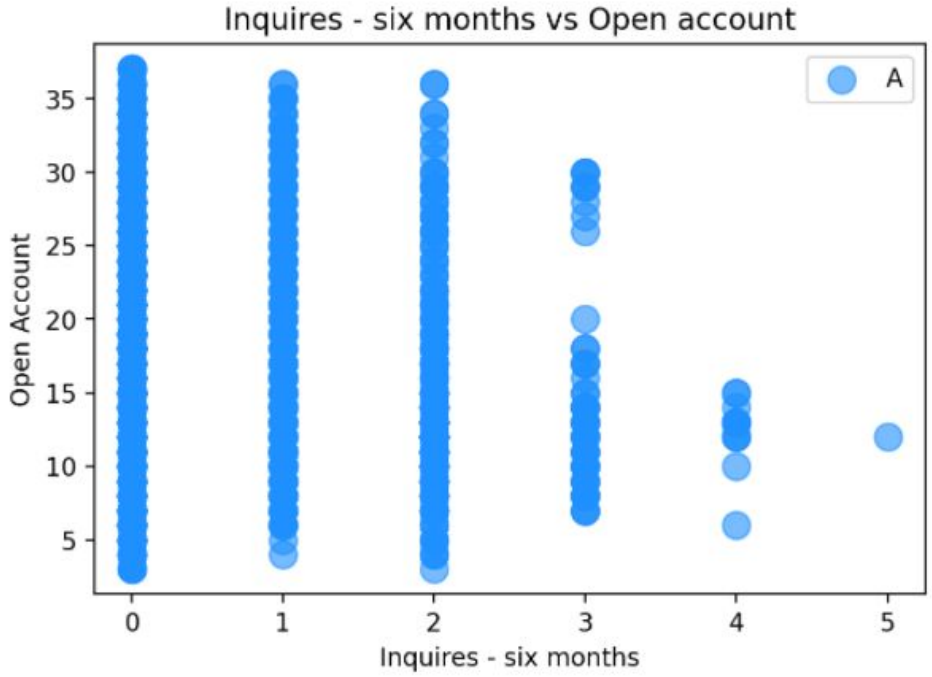


Annex 2: Average Data with Grade and Subgrade

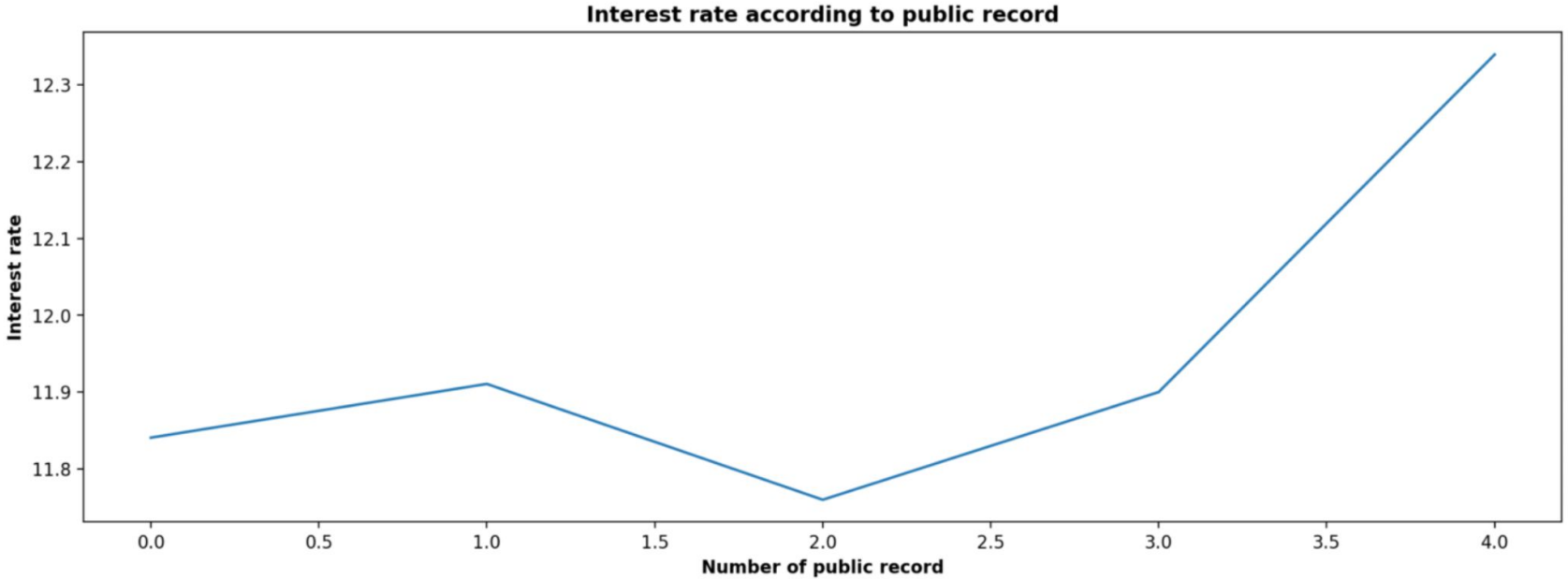




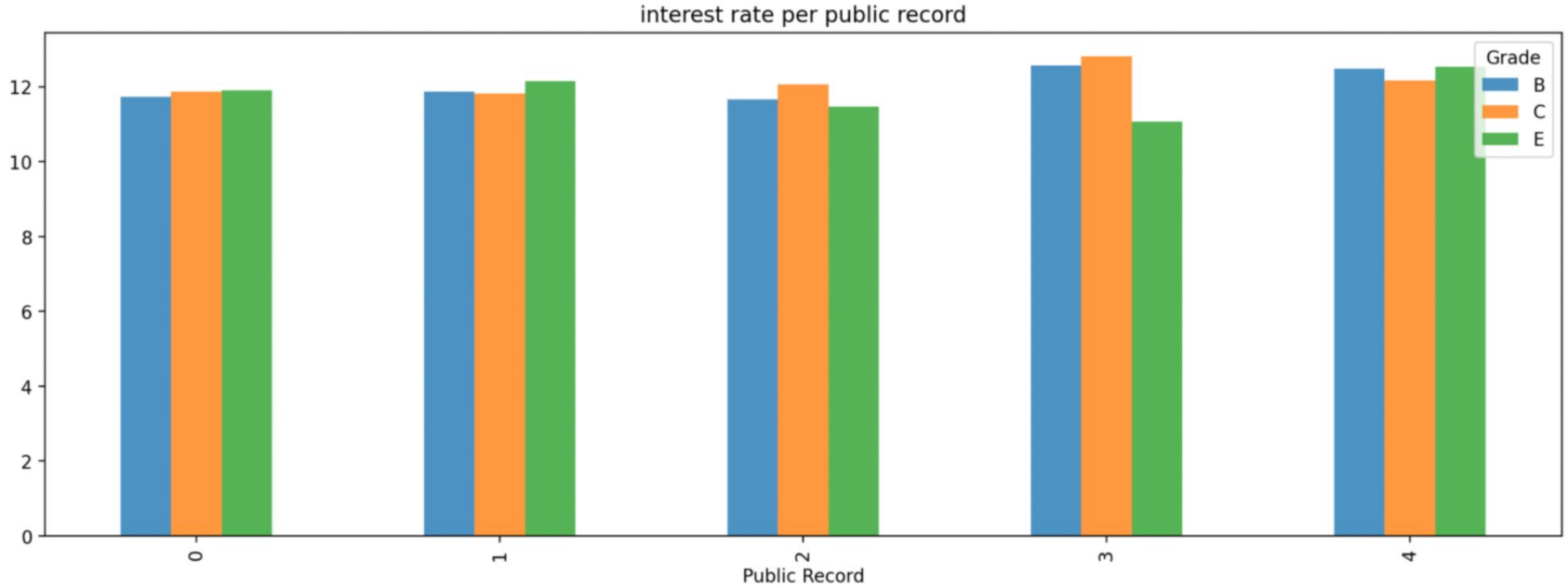
Annex 3: Average Data with Inquires



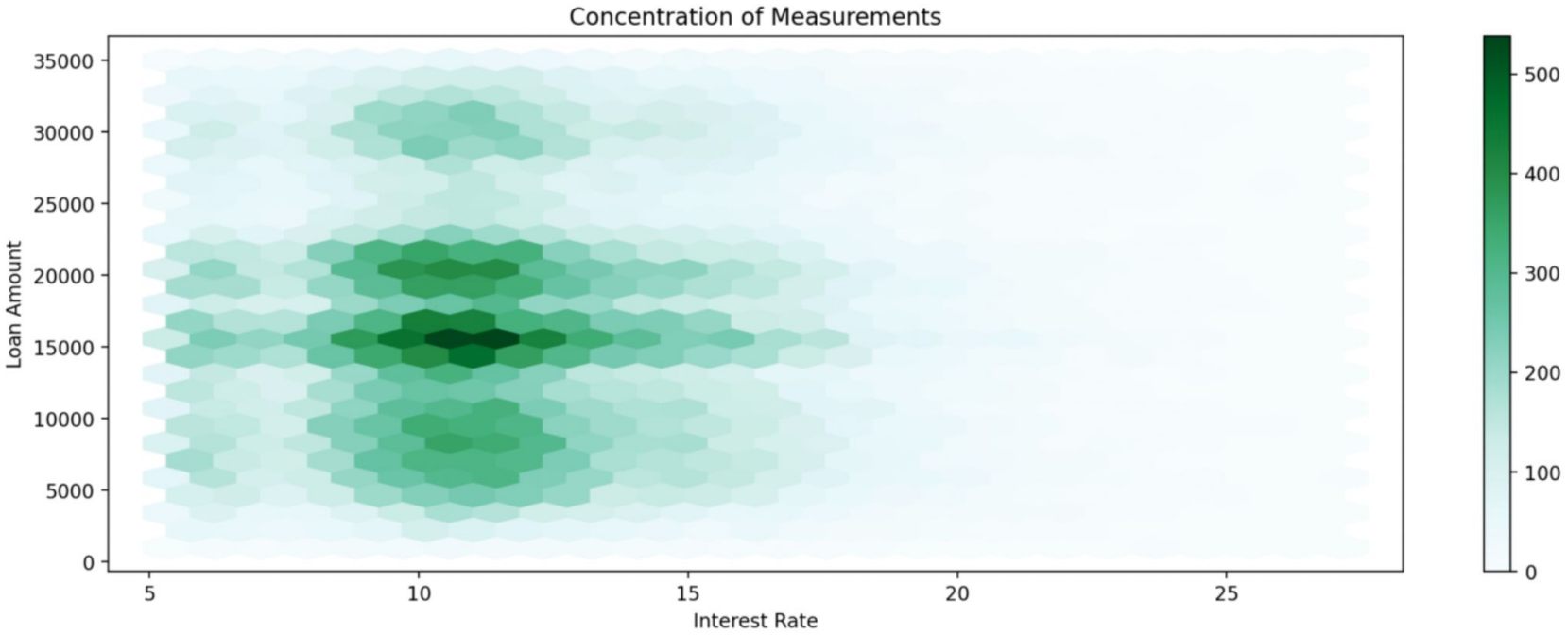
Annex 5:


Interest rate according to public record
Interest rate
12.3
12.2
12.1
12.0
11.9
11.8
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Number of public record

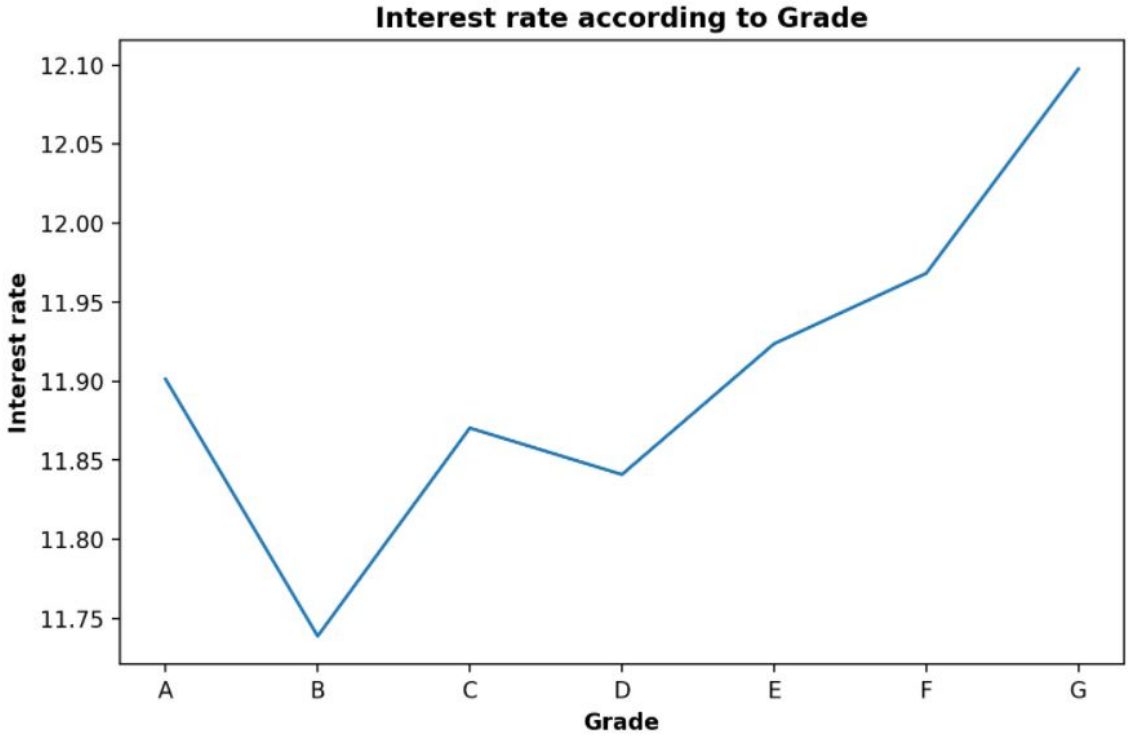
Annex 6:


interest rate per public record
Grade
B
C
E
12
10
8
6
4
2
0
0
1
2
3
4
Public Record

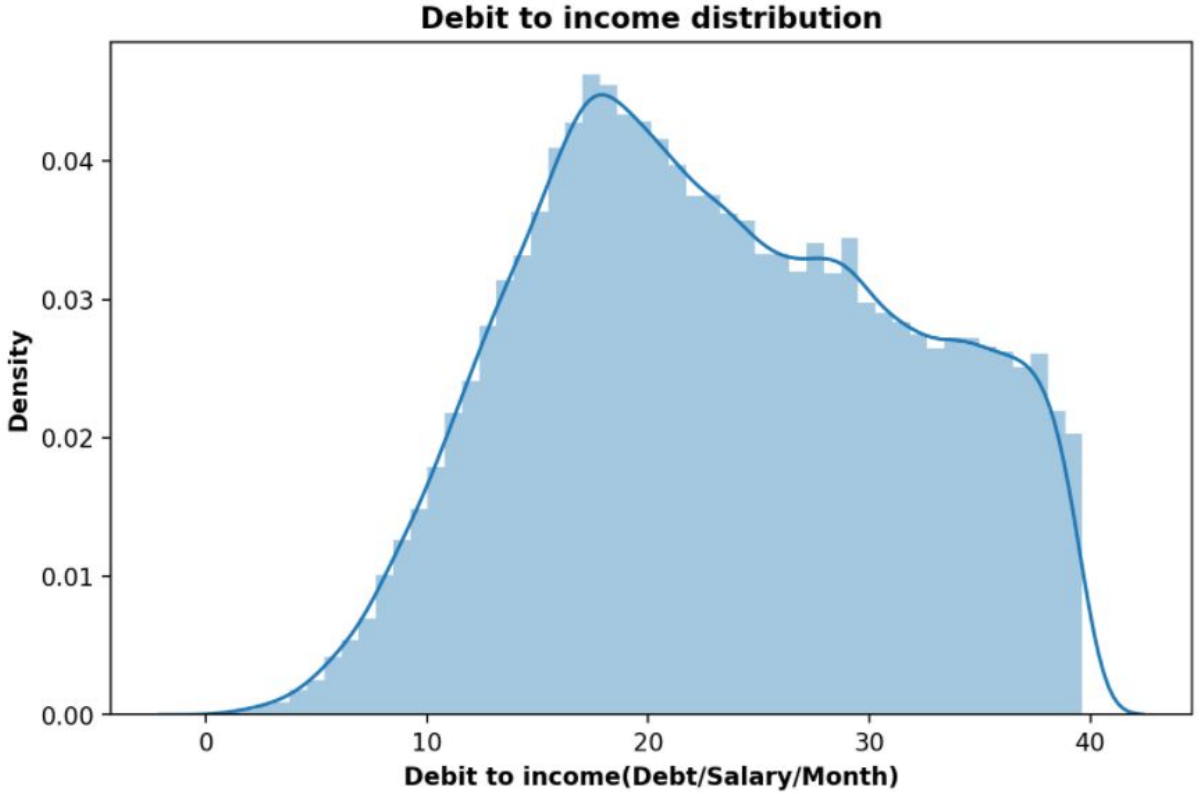
Annex 7:


Concentration of Measurements
Loan Amount
35000
30000
25000
20000
15000
10000
5000
0
5
10
15
20
25
Interest Rate
500
400
300
200
100
0

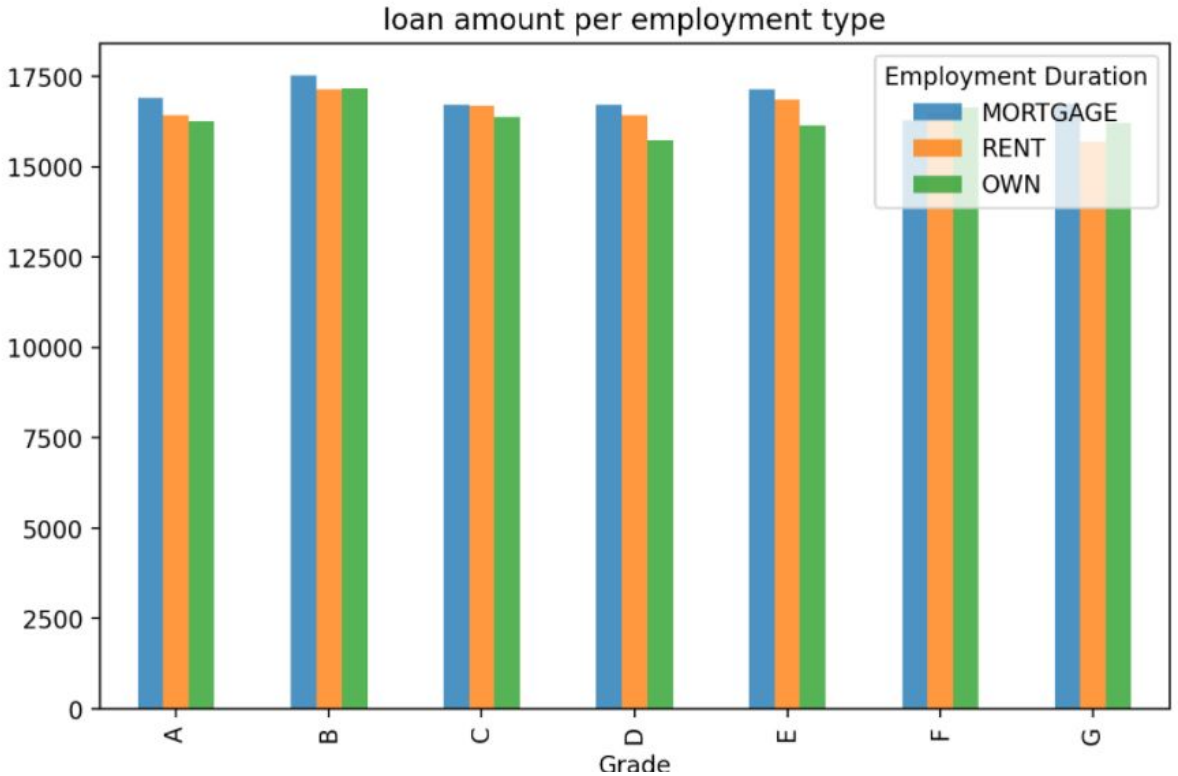
Annex 8:


Interest rate according to Grade
12.10
12.05
12.00
11.95
11.90
11.85
11.80
11.75
Interest rate
A
B
C
D
E
F
G
Grade

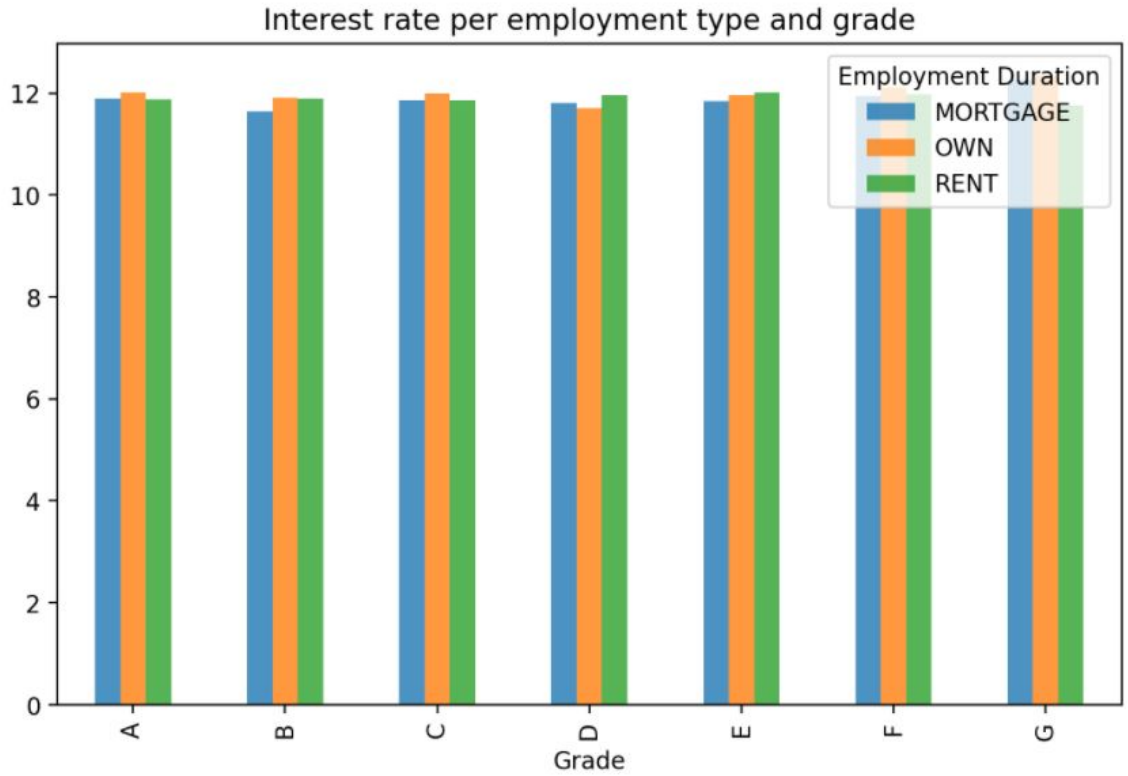
Annex 9:


Debit to income distribution
Density
0.00
0.01
0.02
0.03
0.04
0
10
20
30
40
Debit to income(Debt/Salary/Month)

Annex 10:


loan amount per employment type
Employment Duration
MORTGAGE
RENT
OWN
17500
15000
12500
10000
7500
5000
2500
0
A
B
C
D
E
F
G
Grade

Annex 11:


Interest rate per employment type and grade
Employment Duration
MORTGAGE
OWN
RENT
0
2
4
6
8
10
12
A
B
C
D
E
F
G
Grade