

Nhóm 6

Thành viên



Nguyễn Thu Trang



Nguyễn Đặng
Trường An



Trần Văn Ninh



Nguyễn Hồng
Phương Nghi

Phần 1

Giới thiệu chung về file dữ liệu

Phần 2

Phân tích dữ liệu

Phần 3

Tổng kết

Phần 1: Giới thiệu chung về file dữ liệu

Thông tin chung về dữ liệu

Thông tin khách hàng vay tiền tại một ngân hàng ở Ấn Độ

Number	Income	Age	Experience	Married/Single	House Ownership	Car Ownership	Profession	State	Risk Flag
0	1303834	23	3	single	rented	no	Mechanical engineer	Madhya Pradesh	0
1	7574516	40	10	single	rented	no	Software developer	Maharashtra	0
2	3991815	66	4	married	rented	no	Technical writer	Kerala	0
3	6256451	41	2	single	rented	yes	Software developer	Odisha	1
4	5768871	47	11	single	rented	no	Civil servant	Tamil Nadu	1
5	6915937	64	0	single	rented	no	Civil servant	Maharashtra	0
6	3954973	58	14	married	rented	no	Librarian	Tamil Nadu	0

43,190 dòng dữ liệu

Tên cột	Ý nghĩa	Giới hạn dữ liệu	Kiểu dữ liệu	Chi tiết
Income	Thu nhập kiếm được	10,310 - 9,998,280	Định tính	Rời rạc
Age	Độ tuổi	21 - 79	Định tính	Liên tục
Experience	Số năm kinh nghiệm làm việc	0 - 20	Định tính	Liên tục
Married/Single	Tình trạng hôn nhân	Độc thân, Đã kết hôn	Định lượng	Định danh True/False
House_Ownership	Tình trạng sở hữu nhà	Sở hữu, thuê, không sở hữu/thuê	Định lượng	Định danh
Car_Ownership	Tình trạng sở hữu xe	Có hoặc không	Định lượng	Định danh True/False
Profession	Nghề nghiệp	51 nghề	Định lượng	Định danh
STATE	Khu vực đang sinh sống	Ấn Độ, 28 vùng	Định lượng	Định danh
Risk_Flag	Được đánh giá có rủi ro hay không khi cho vay	0 (Không rủi ro), 1 (có rủi ro)	Định tính	Định danh True/False

Điều kiện làm sạch dữ liệu

1 2 3 4 5 6 7 8 9

Number	Income	Age	Experience	Married/Single	House Ownership	Car Ownership	Profession	State	Risk Flag
0	1303834	23	3	single	rented	no	Mechanical engineer	Madhya Pradesh	0
1	7574516	40	10	single	rented	no	Software developer	Maharashtra	0
2	3991815	66	4	married	rented	no	Technical writer	Kerala	0
3	6256451	41	2	single	rented	yes	Software developer	Odisha	1
4	5768871	47	11	single	rented	no	Civil servant	Tamil Nadu	1
5	6915937	64	0	single	rented	no	Civil servant	Maharashtra	0
6	3954973	58	14	married	rented	no	Librarian	Tamil Nadu	0

37,658 dòng dữ liệu

1. Giá trị là not na (bao gồm not null).
2. Giá trị trong các cột Income, Age, Experience, Risk_Flag đều chuyển được ra số và có giá trị dương.
3. Thông tin trong các cột đều khác "empty".
4. Thông tin trong các cột 4,5,6,9 đều phải nằm trong giới hạn dữ liệu.

Mục tiêu phân tích

Tập trung phân tích nhóm khách hàng có *rủi ro* khi cho vay.

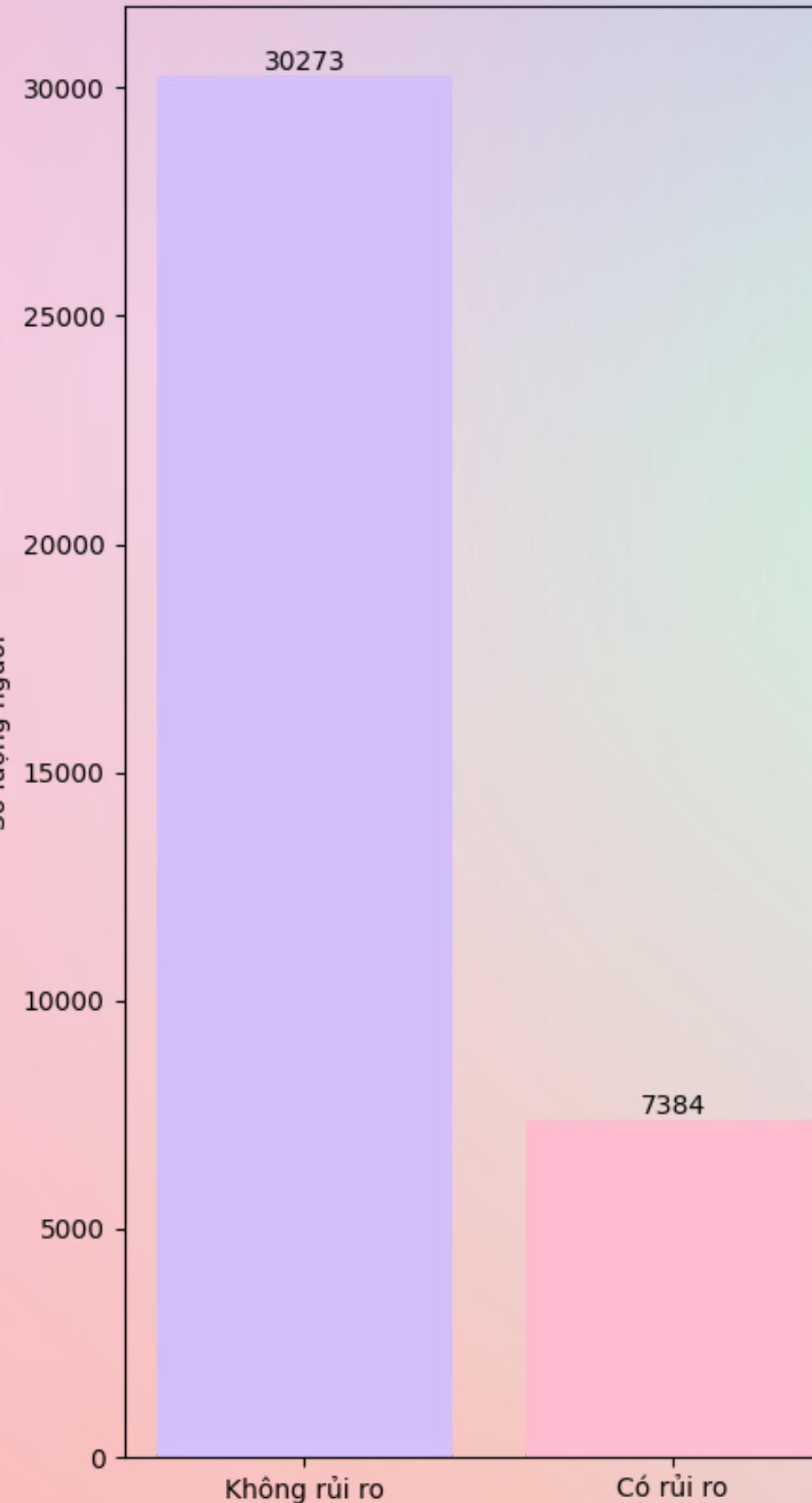
=> Tìm ra những *mối liên quan* trong trường thông tin của nhóm này.



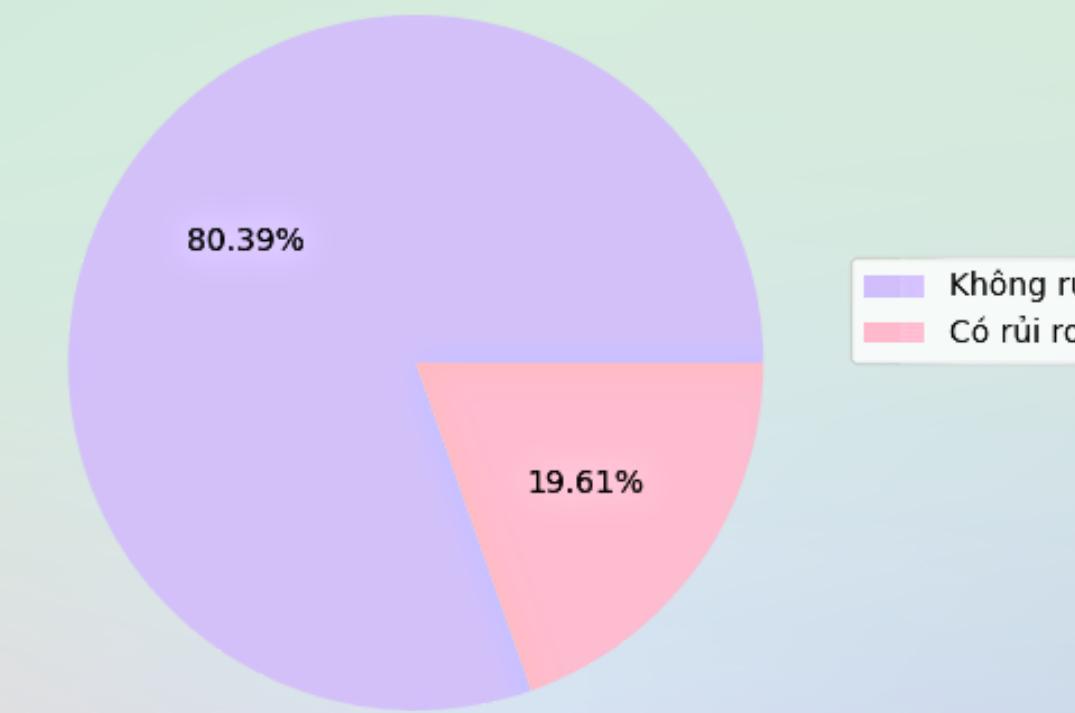
Phần 2: Phân tích dữ liệu

Tần số, tần suất người có rủi ro khi cho vay

Tần số người vay có rủi ro và không rủi ro



Tần suất người vay có rủi ro và không rủi ro



- Số người vay có rủi ro ~20%
- Tuy nhiên, cần lưu ý kỹ nhóm đối tượng này để tránh nợ xấu.

Mối quan hệ giữa sự rủi ro với thu nhập, độ tuổi và số năm kinh nghiệm

- Mối quan hệ giữa rủi ro và thu nhập

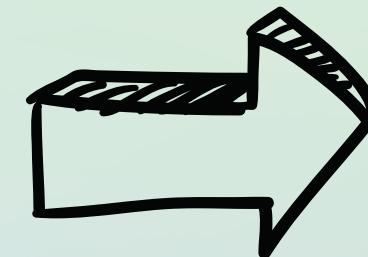
Hệ số tương quan : **0.001**

- Mối quan hệ giữa rủi ro và tuổi

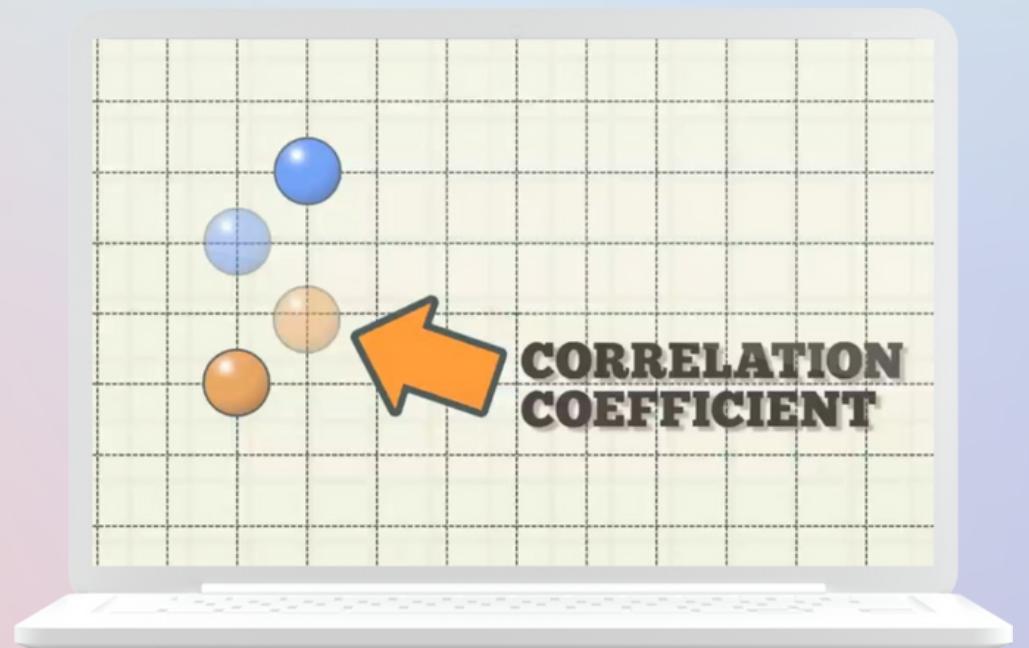
Hệ số tương quan : **0.00047**

- Mối quan hệ giữa rủi ro và số năm kinh nghiệm

Hệ số tương quan: **-0.01**

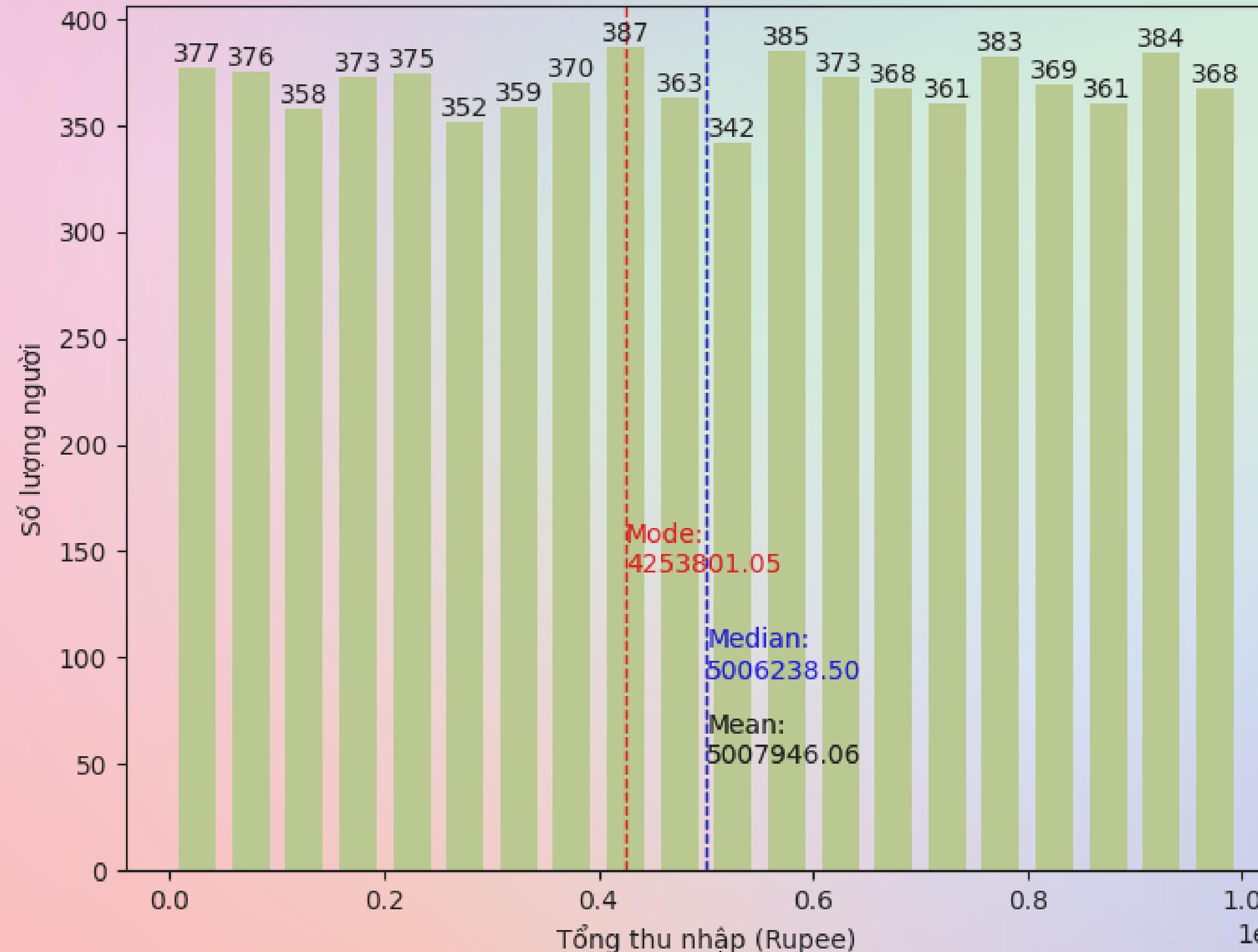


**GẦN NHƯ KHÔNG CÓ
MỐI QUAN HỆ ĐỒNG
BIẾN HAY NGHỊCH
BIẾN**



Sự phân bố về thu nhập

Sự phân bố về thu nhập

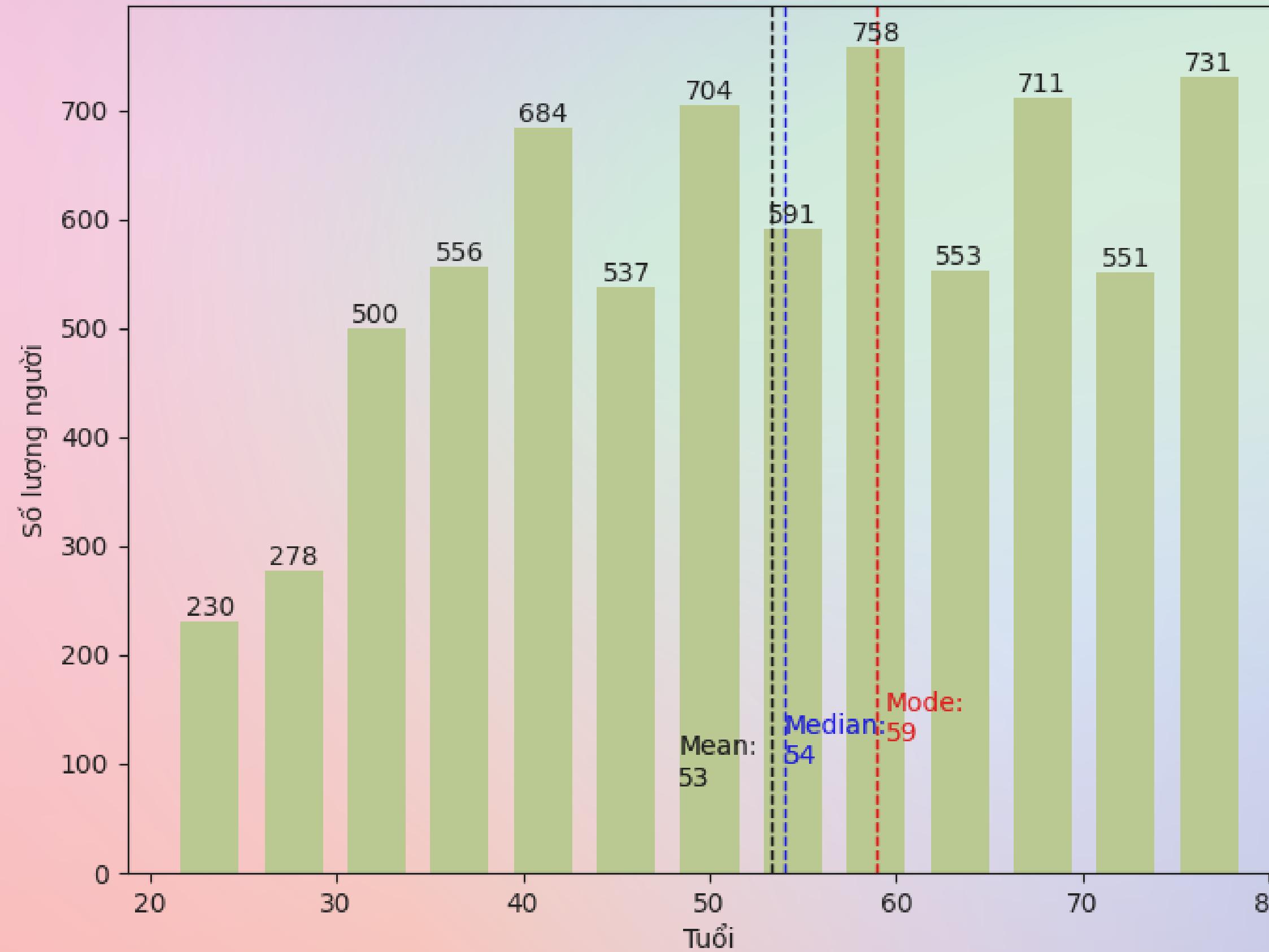


- Min: 10,675
- Max: 9,994,501
- Phân vị:
[25%, 50%, 75%] =
[2,492,112 ; 5,006,238 ; 7,529,020]

=> Độ phân bố tương đối đều

Sự phân bố về độ tuổi

Sự phân bố về tuổi



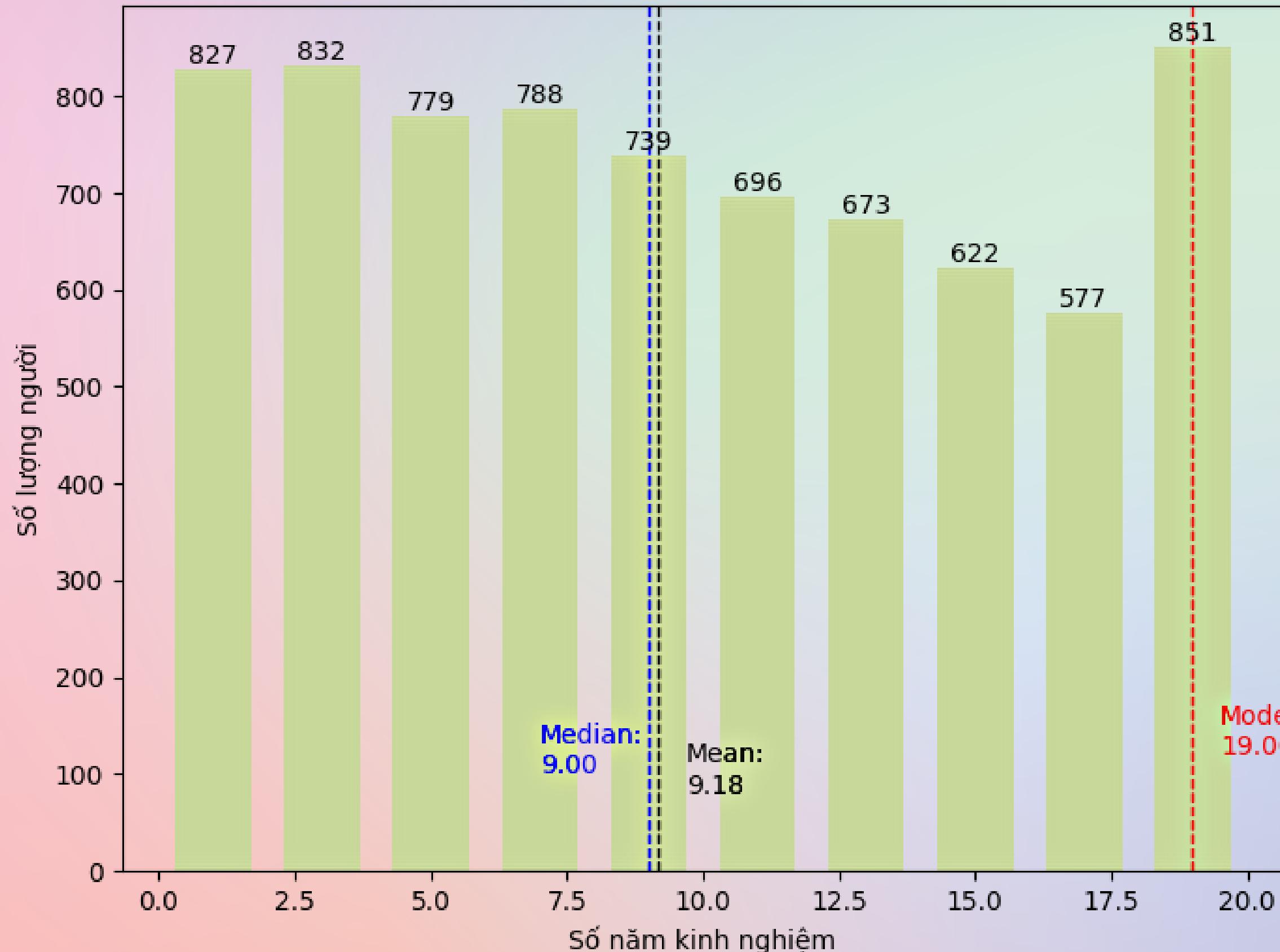
- Min: 21
- Max: 79

- Phân vị:
 $[25\%, 50\%, 75\%] = [40 ; 54 ; 66]$

=> *Những người ở độ tuổi trung niên trở lên có mức độ rủi ro cao hơn*

Sự phân bố về số năm kinh nghiệm

Sự phân bố về năm kinh nghiệm

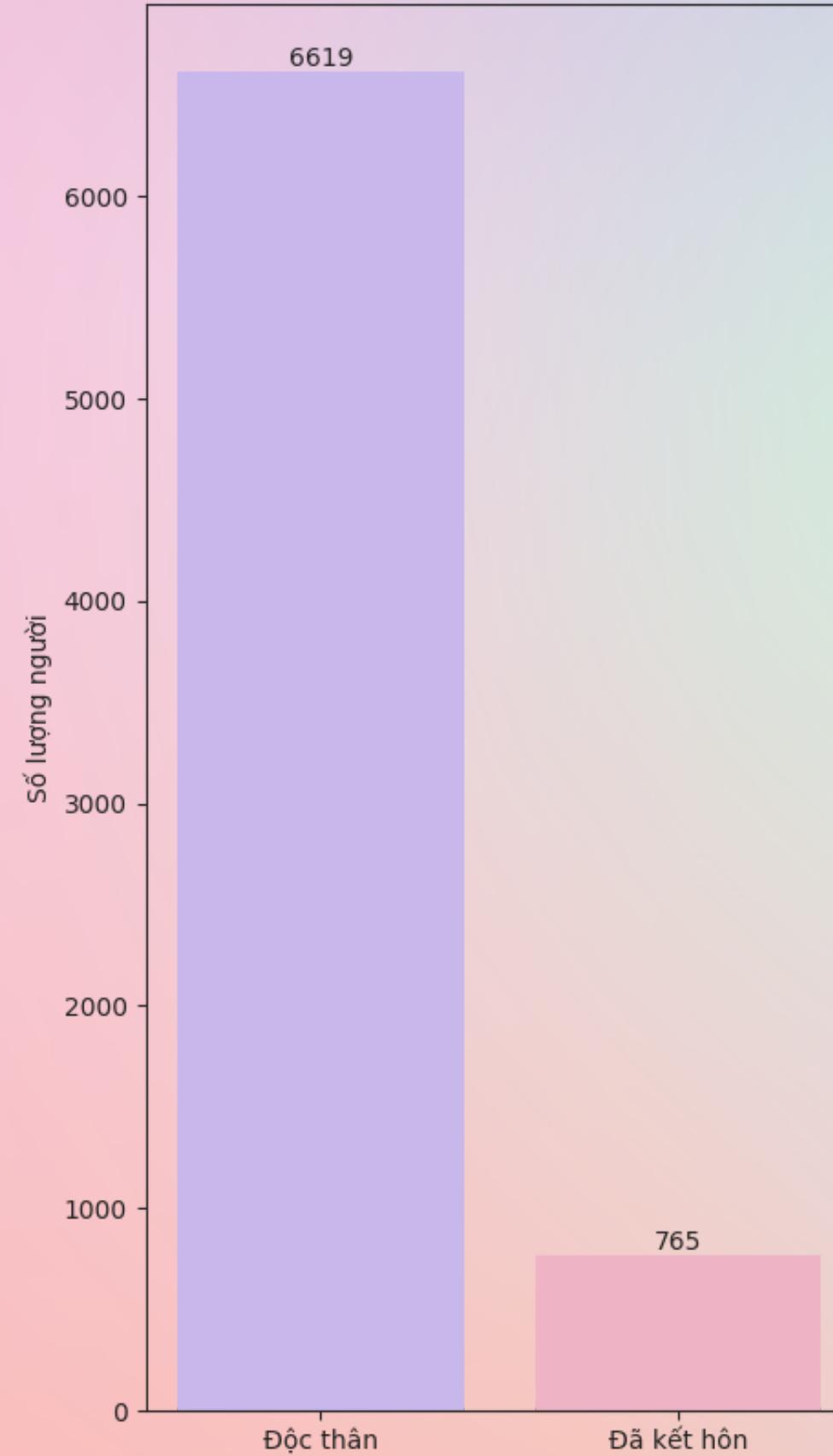


- Min: 0
- Max: 20
- Phân vị:
 $[25\%, 50\%, 75\%] = [4 ; 9 ; 14]$

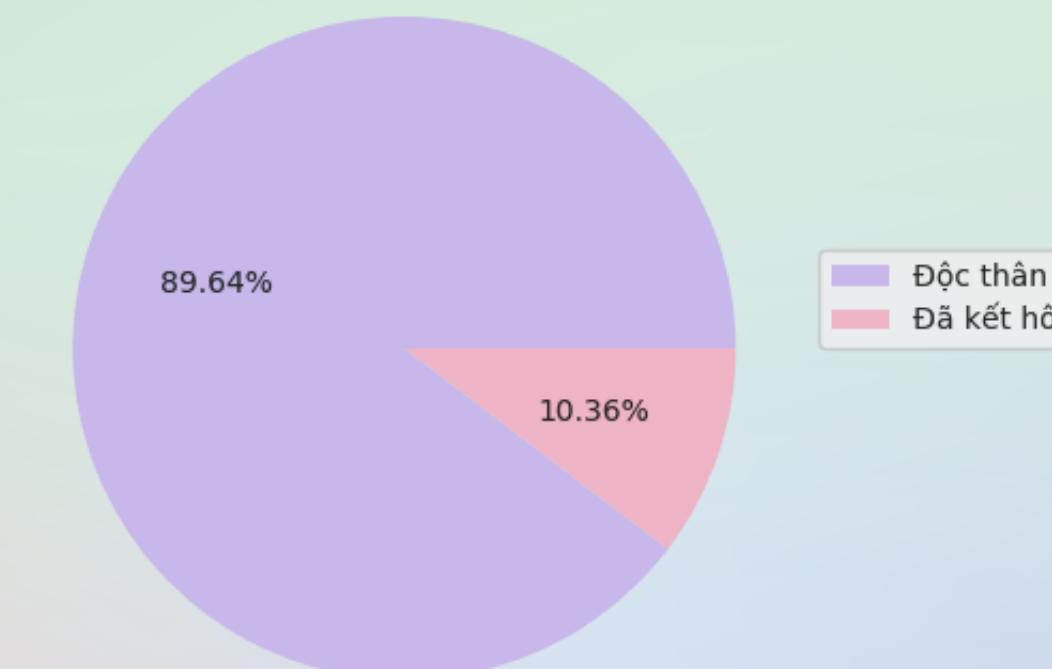
=> *Những người có rủi ro tập
trung vào nhóm có kinh nghiệm
làm việc lâu năm*

Tình trạng hôn nhân của nhóm có rủi ro

Tần số về tình trạng hôn nhân
của người có rủi ro khi cho vay



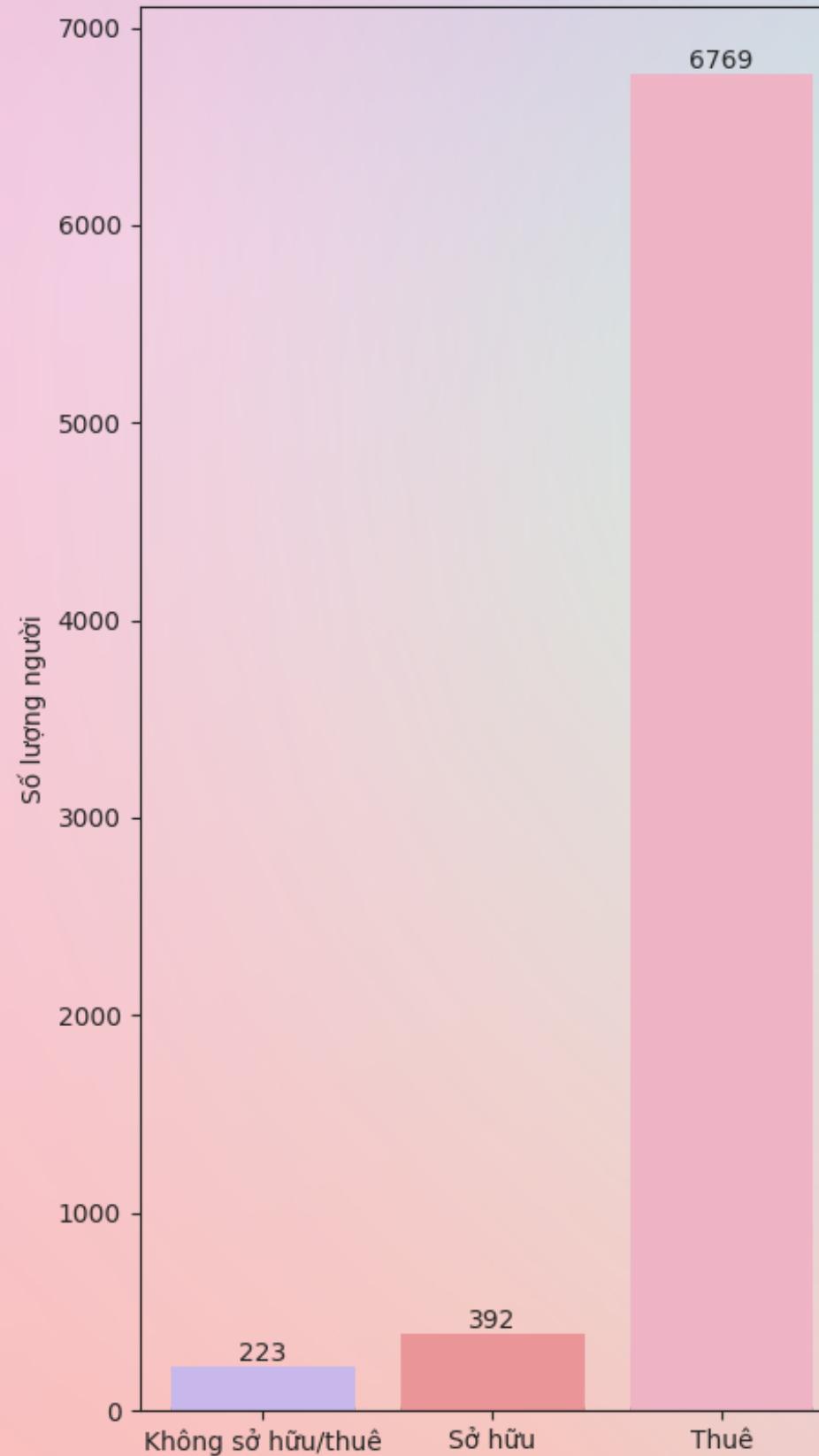
Tần suất về tình trạng hôn nhân
của người có rủi ro khi cho vay



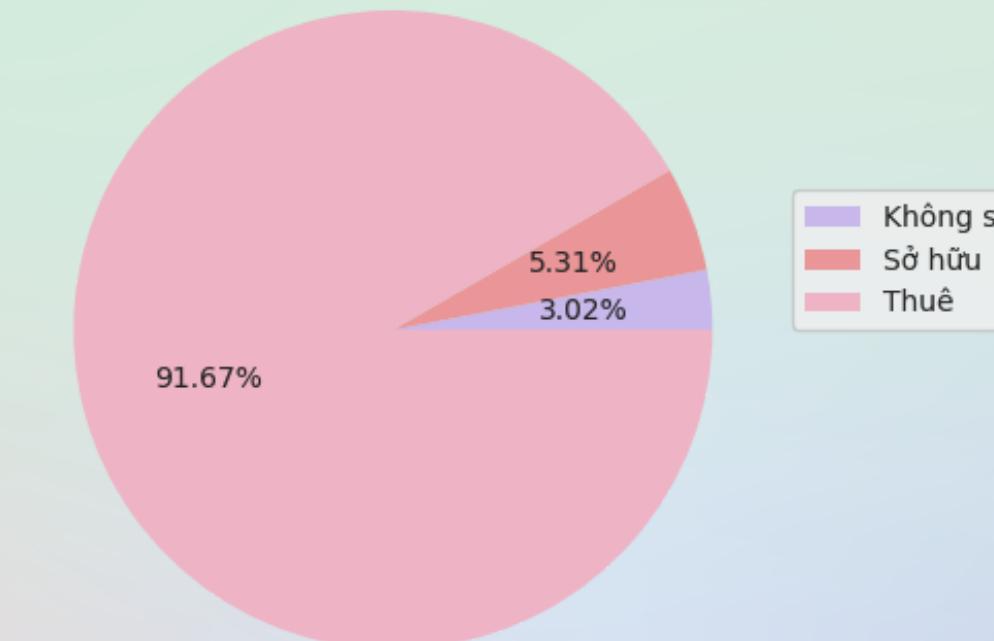
=> **Đa số là những người độc thân (~ 90%)**

Tình trạng sở hữu nhà của nhóm có rủi ro

Tần số về tình trạng sở hữu nhà
của người có rủi ro khi cho vay



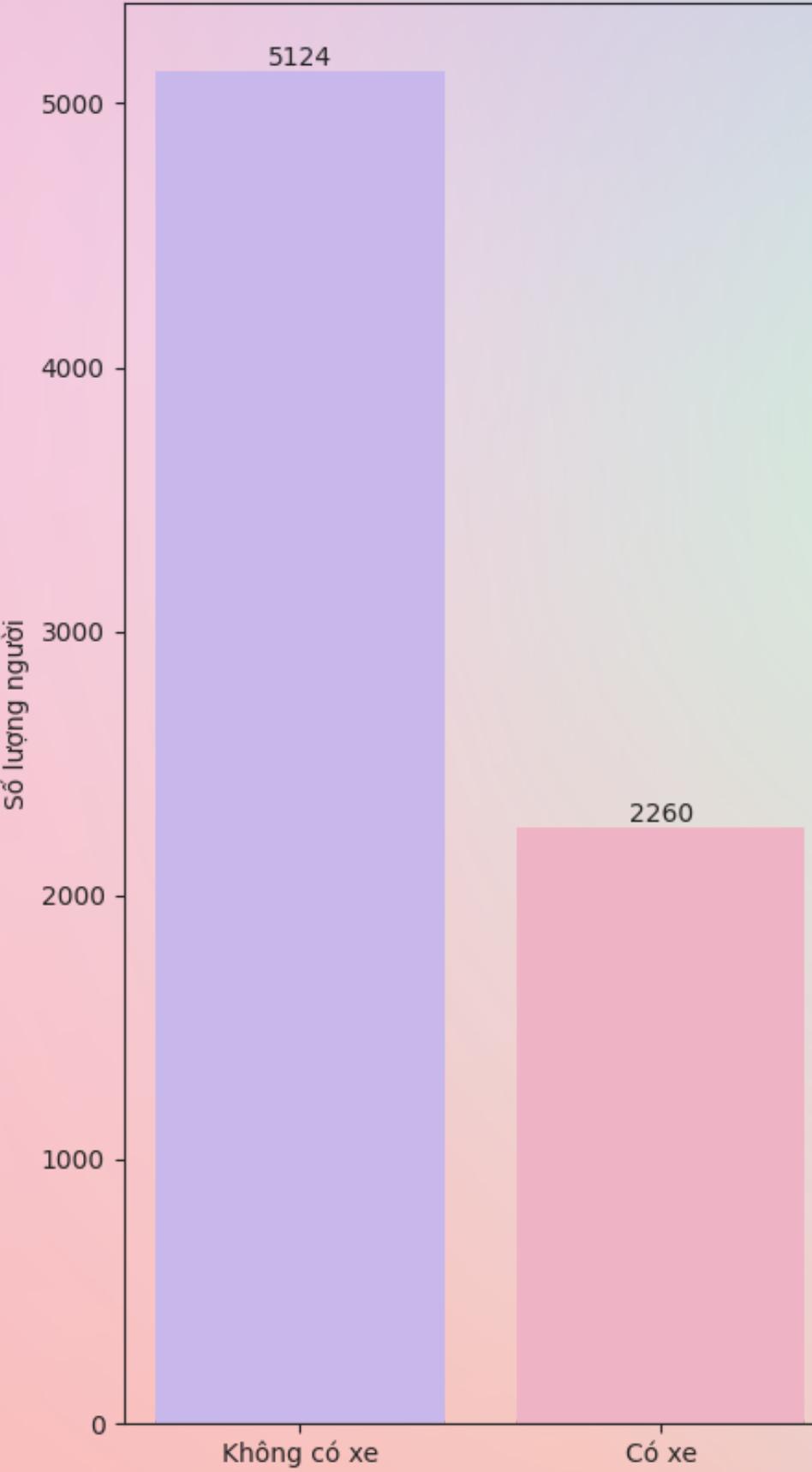
Tần suất về tình trạng sở hữu nhà
của người có rủi ro khi cho vay



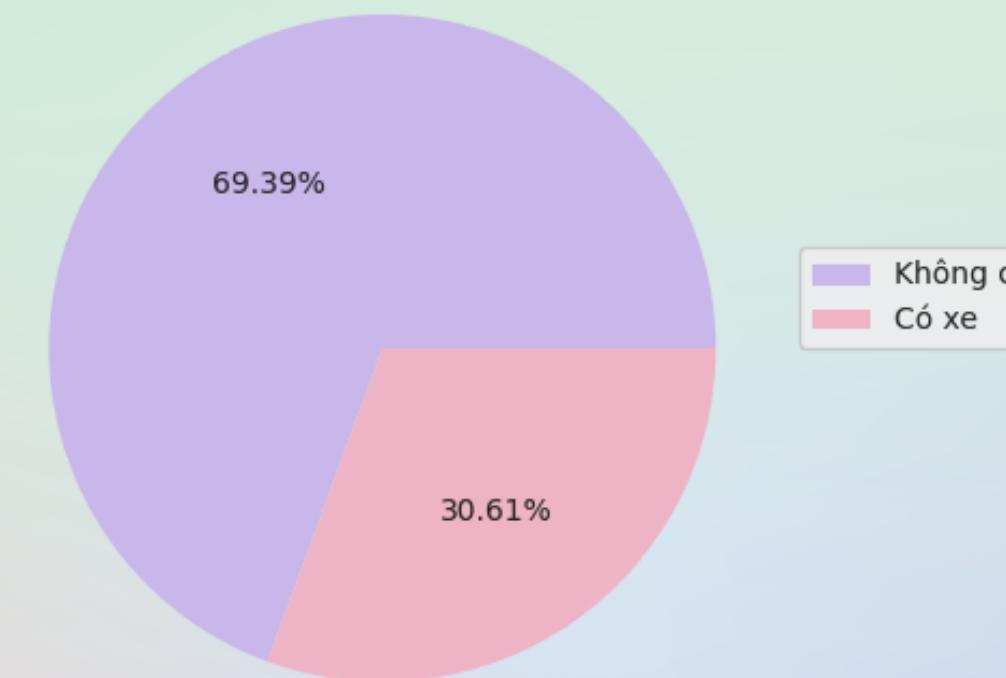
=> Đa số là những người thuê
nhà (> 90%)

Tình trạng sở hữu xe của nhóm có rủi ro

Tần số về tình trạng sở hữu xe
của người có rủi ro khi cho vay



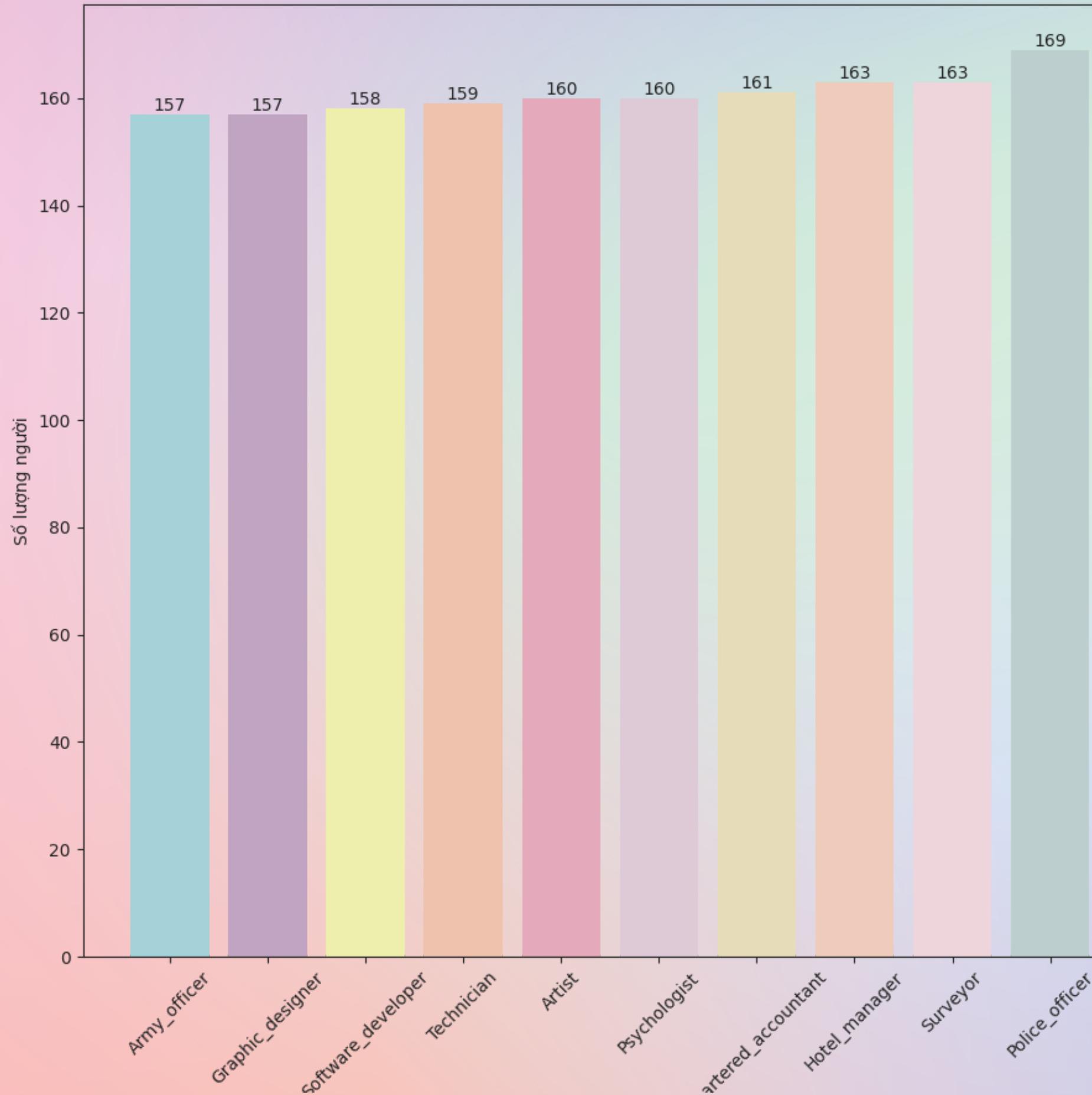
Tần suất về tình trạng sở hữu xe
của người có rủi ro khi cho vay



=> Đa số là những người không
có xe ô tô (~70%)

Top 10 nghề nghiệp có rủi ro nhiều nhất

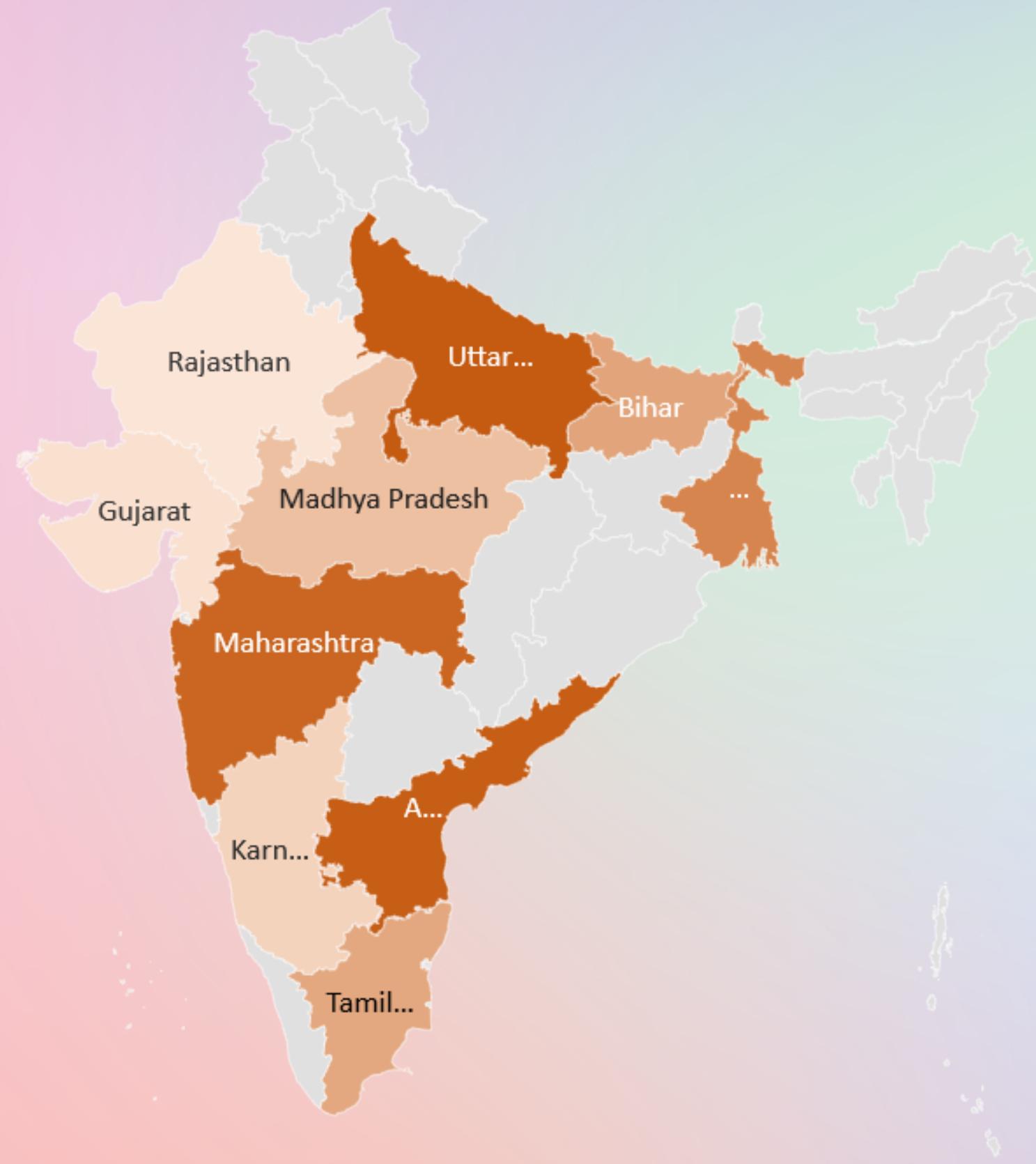
Top 10 nghề nghiệp có người có rủi ro khi cho vay



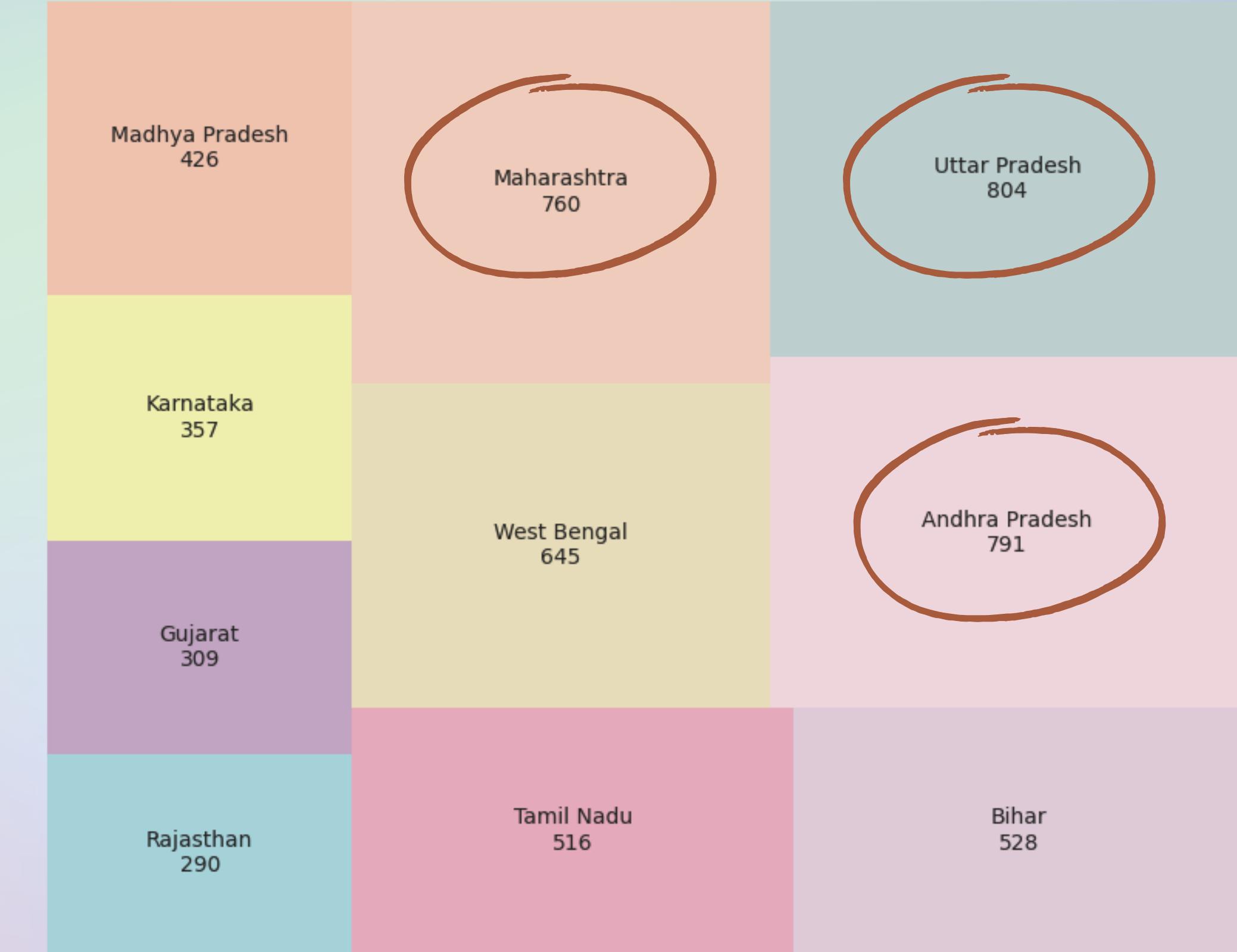
Top 3:

- Công an
- Người làm khảo sát
- Quản lý khách sạn

Top 10 khu vực có rủi ro nhiều nhất



Nơi cư trú của những người có rủi ro khi cho vay



Phần 3: Tổng kết

Tổng kết

- Nhóm người có rủi ro chiếm ~20% , là nhóm cần được chú ý để giảm nguy cơ cho ngân hàng khi cho vay.
- Không có mối tương quan đồng biến hay nghịch biến giữa độ rủi ro với thu nhập, tuổi và số năm kinh nghiệm.
- Tuy nhiên, nhóm người có rủi ro thường nằm ở độ tuổi trung niên trở lên và có nhiều năm kinh nghiệm làm việc.
- Những người có rủi ro chủ yếu là những người độc thân, thuê nhà, và không có xe ô tô.
- Công an, Người làm khảo sát, Quản lý khách sạn là những nghề tập trung nhiều người có rủi ro nhất.
- Uttar Pradesh, Andhra Pradesh, Maharashtra là những khu vực tập trung nhiều người có rủi ro nhất.



Thank you

