



Group 11

POWERPOINT

Loan Prediction Based on Customer Behavior

Group Member: Jiazheng Qian, Huili Han



Catalogue

CONTENTS

01

Data set description

Group eleven

02

Problems and Analysis

Group eleven

03

Further study

Group eleven

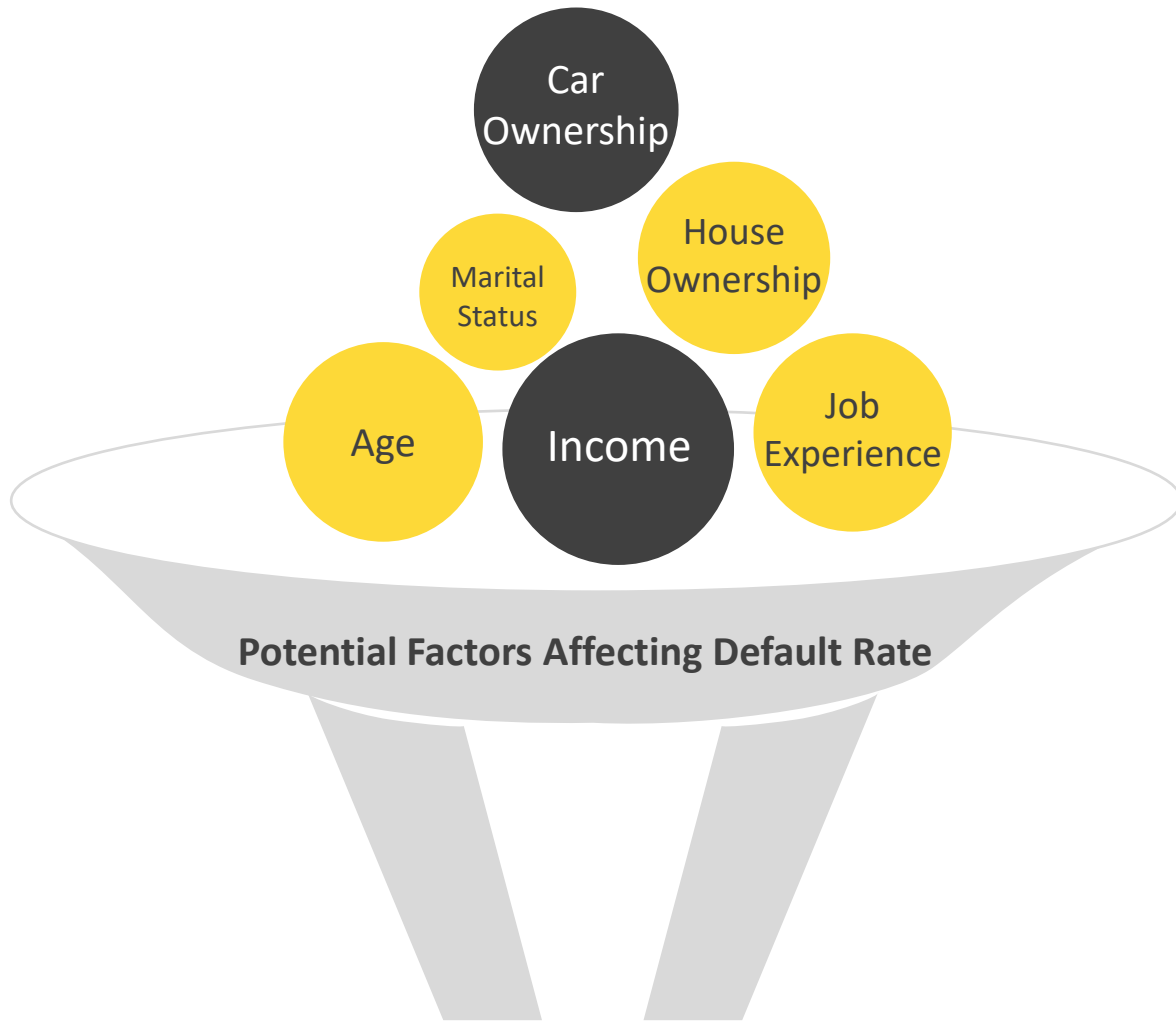
04

data app demo

Group eleven

DATASITE DESCRIPTION

The purpose and commercial value of research.



01

Research Background

An organization wants to predict who possible defaulters are for the consumer loans product. They have data about historic customer behavior based on what they have observed. Hence when they acquire new customers they want to predict who is riskier and who is not.

02

Research Purpose

Based on historical data, our group explore the factors that influence default rate, such as job seniority, asset ownership and age range.

03

Commercial Value

The research is helpful for enterprises to assess the risks of product delivery objects according to their attributes.

DATASITE DESCRIPTION

The scope and objects of dataset.

Research Scope & Object

- The data set covers 26 states around the world, including groups with an annual income between 10 thousands and 10 millions.

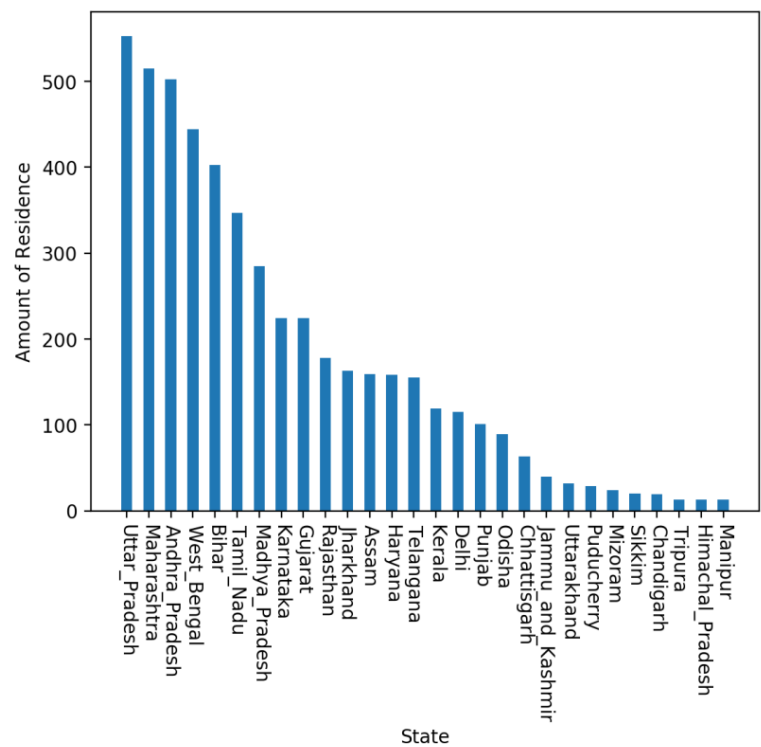


Fig 1-1 The Research Scope of Dataset

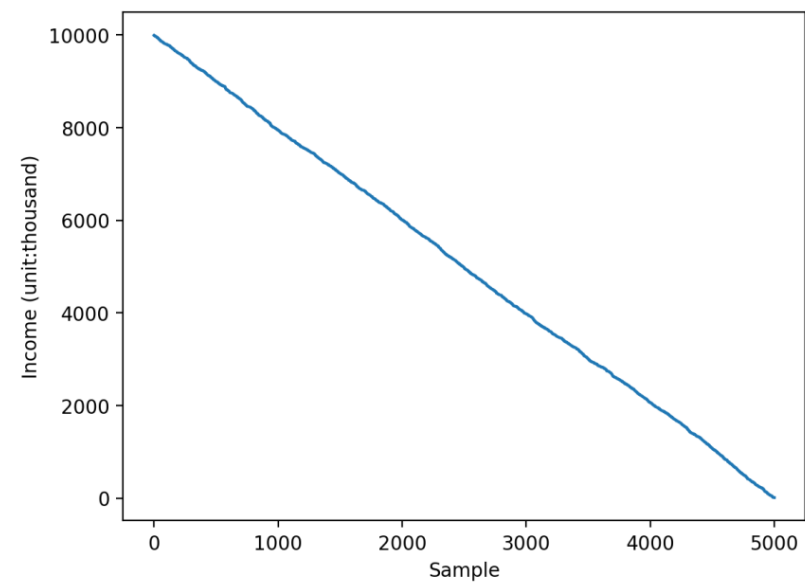


Fig 1-2 The Research Objects of Dataset

DATASITE DESCRIPTION

The attributes of dataset.

The dataset contains **11** different attributes:

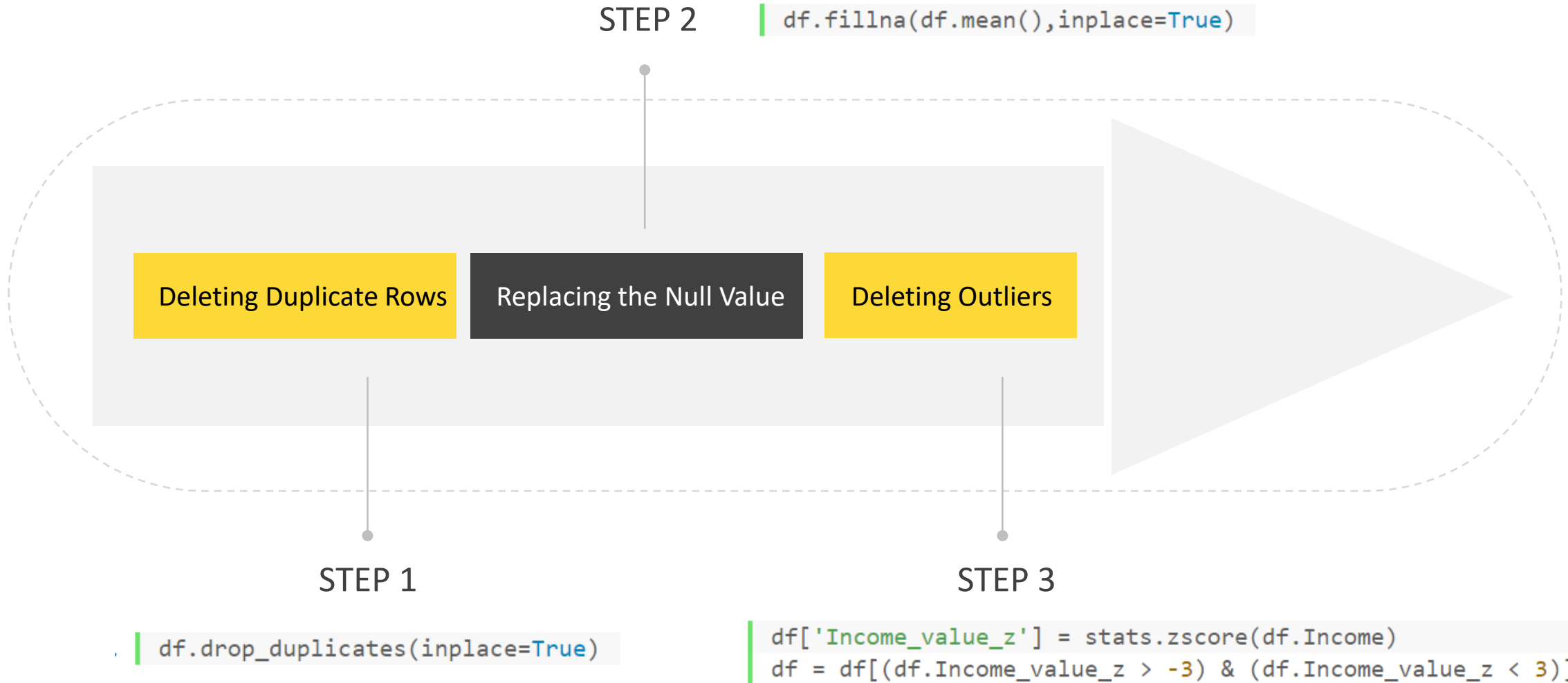
Column	Description	Type
income	Income of the user	int
age	Age of the user	int
experience	Professional experience of the user in years	int
profession	Profession	string
married	Whether married or single	string
house_ownership	Owned or rented or neither	string
car_ownership	Does the person own a car	string
risk_flag	Defaulted on a loan	string
Current_job_years	Years of experience in the current job	int
Current_house_years	Number of years in the current residence	int
city	City of residence	string
state	State of residence	string

Fig. 1-3 The Attributes of Dataset

Our research mainly focus on the impact of **income, age, job qualifications, marital status** and **ownership** factors on the risk of default.

DATASITE PREPROCESSING

Preprocessing the data to draw conclusions.



Problem 1.1

Explore the potential factors affecting the default rate.

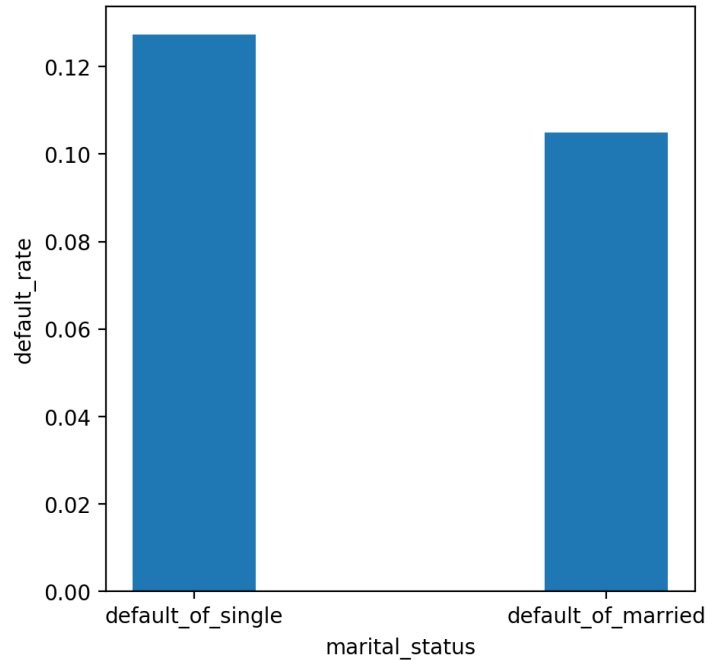


Fig. 1-4 Default Rate of Groups with Different Marital Status

- **Fig. 1-4** shows that the default rate of single people is higher than that of married people (the gap is about 0.022%).
- **Fig. 1-5** shows that the default rate of the group that does not own the car (0.713%) is about 2.5 times higher than that of the group who own the car (0.287%).

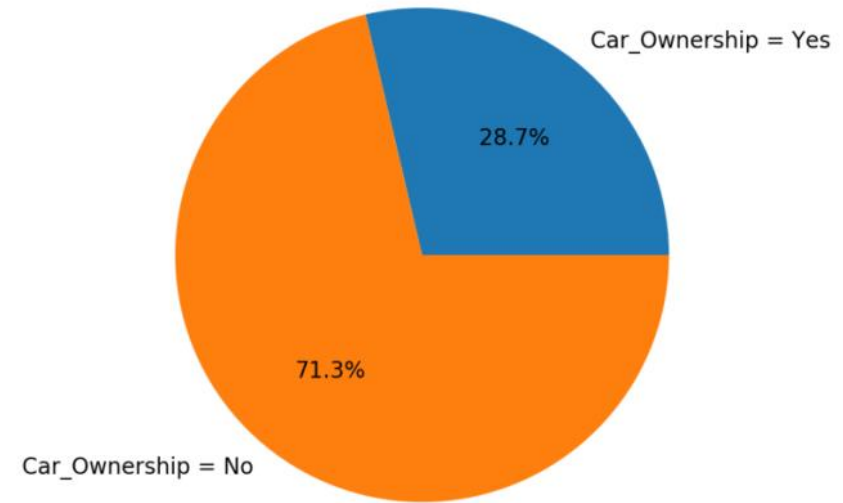


Fig. 1-5 Default Rate of Groups with Car Ownership & Without

Problem 1.2

Explore the potential factors affecting the default rate.

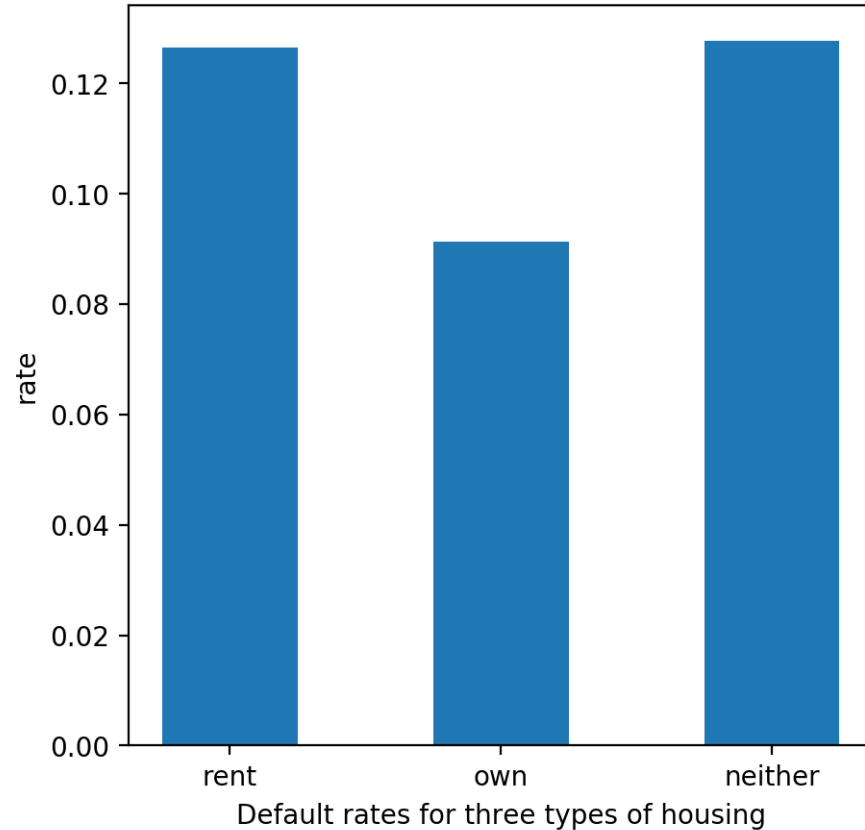


Fig. 1-6 Default Rate of Groups with House Ownership & Without

· **Fig. 1-6** shows that the default rate of rental groups (12.655%) is higher than that of homeowners (9.127%).

Conclusion

Marital status and asset ownership (house and car) are correlated with default rate, and groups with lack of asset ownership or single marital status are more likely to be default.

Problem 2.1

Relationship Between Profession Experience and Default Possibility

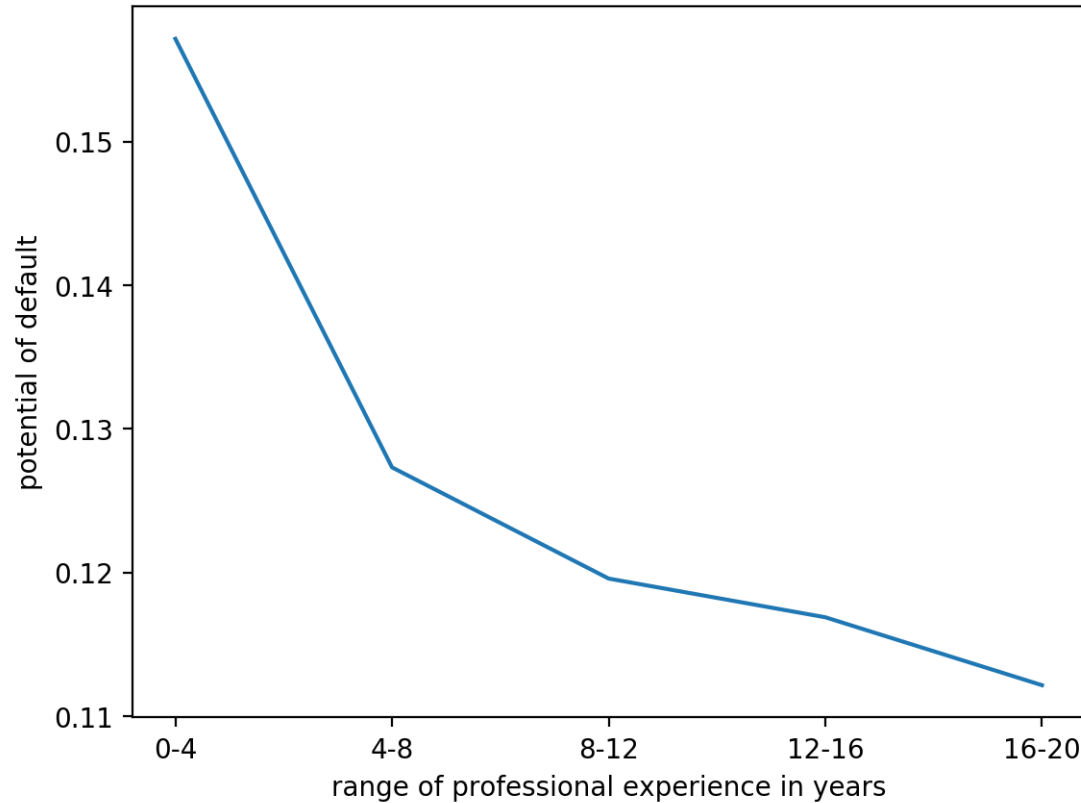
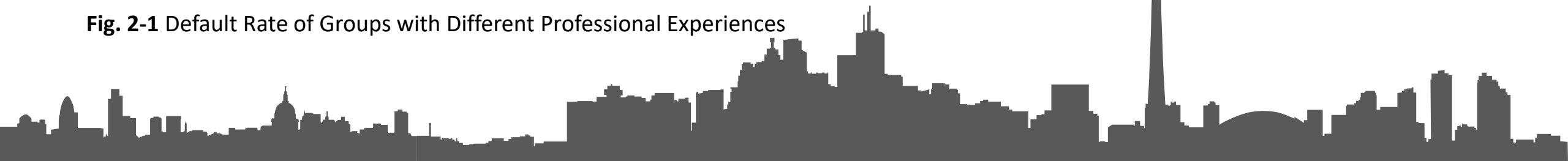


Figure Analysis

- ☆ The figure (**Fig. 2-1**) shows that there is a significant negative correlation between years of work experience and default rate.
- ☆ The group with 0-4 years of work experience had the highest default rate (15.716%), and the group with 16-20 years of work experience had the lowest default rate (11.214%).

Fig. 2-1 Default Rate of Groups with Different Professional Experiences



Problem 2.2

Default Ratio of Different Age Groups

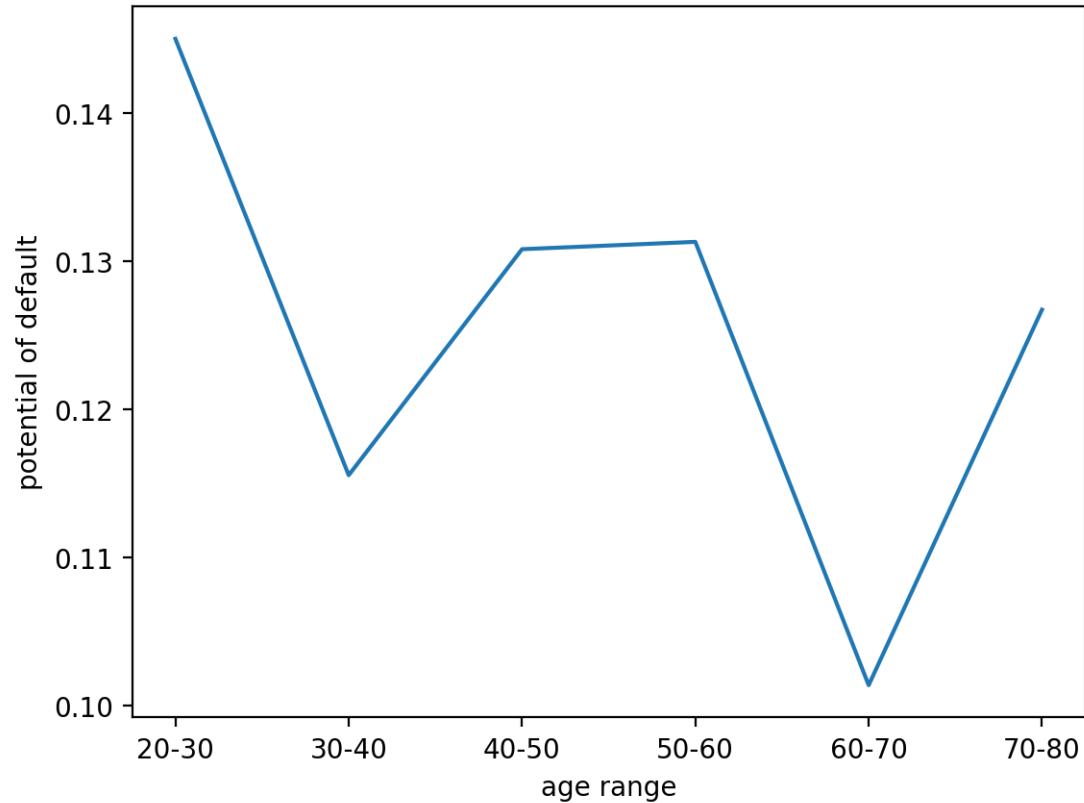
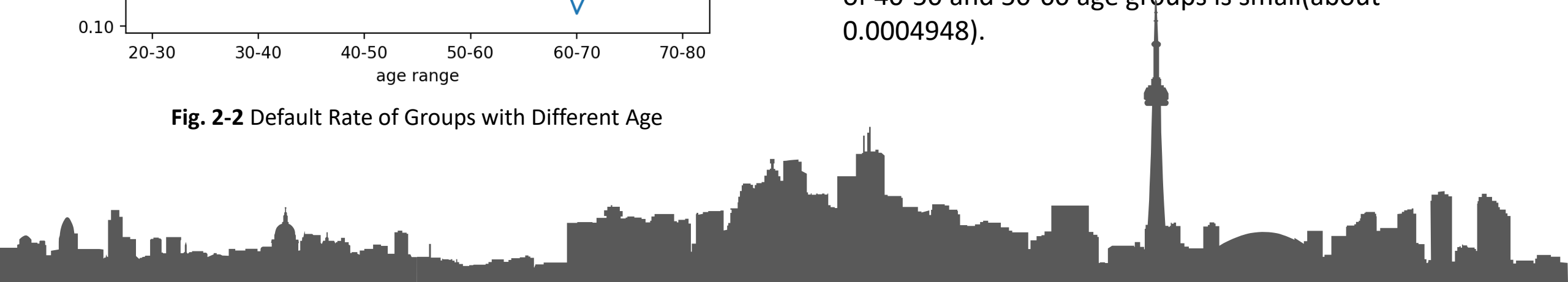


Fig. 2-2 Default Rate of Groups with Different Age

Figure Analysis

- ☆ The picture does not show a clear relationship between default rates and age
- ☆ The default rate is the lowest in 60-70 age group(10.138%) and the highest in 20-30 age group(14.504%).The gap between the default rate of 40-50 and 50-60 age groups is small(about 0.0004948).



Further study

Random forest algorithm

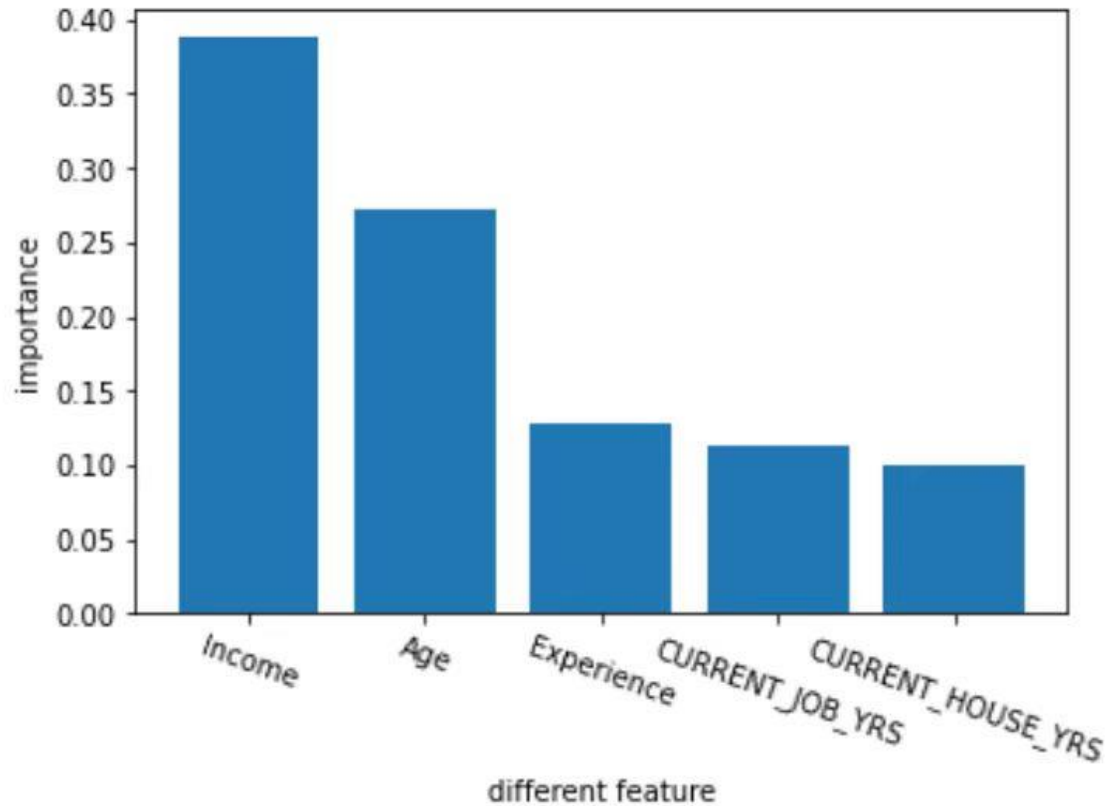


Fig. 2-3 The Importance of Different Factors Affecting Default Rate

01

Algorithm

Using the random forest algorithm to rank the importance of some features

Training set: Test set = 7:3

02

The ranking result

The importance of these features on the impact of loan default results is ranked as follows:

1.Income 2.Age 3.Experience 4.Current_job_years
5.Current_house_years

03

Predicted results

Score for the training data set :
98.29%

Score for the testing data set :
87.87%

Data app demo show

Two filters

House Ownership Selector

rented

owned

norent_noown

Profession Type (enter ALL to reset)

ALL

Apply

House Ownership Selector

rented

owned

norent_noown


Profession Type (enter ALL to reset)

Consultant

Apply

Loan Prediction Based on Customer Behavior

Predict who possible Defaulters are for the Consumer Loans Product



Introduction of Research

An organization wants to predict who possible defaulters are for the consumer loans product. They have data about historic customer behavior based on what they have observed. Hence when they acquire new customers they want to predict who is riskier and who is not.

Dataset Description

The resource of dataset: <https://www.kaggle.com/datasets/subhamjain/loan-prediction-based-on-customer-behavior>

Click the check box to view the data:

Ownership and

renter

0.48% and 5.556%

renter than that of homeowners

Ownership = Yes



Data app demo show

Two filters

House Ownership Selector

rented x owned x

norent_noown x

Profession Type (enter ALL to reset)

ALL

Apply

House Ownership Selector

rented x owned x

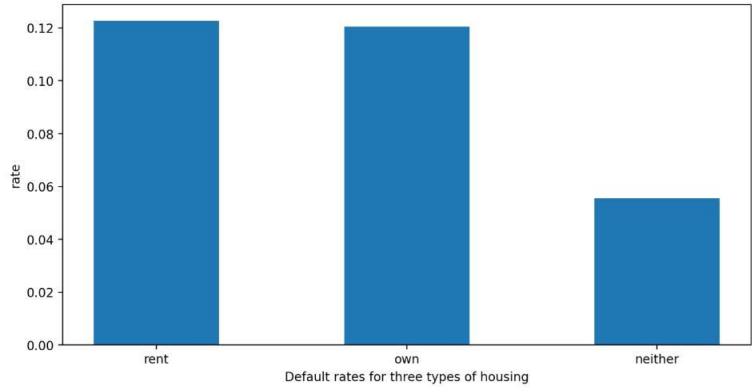
norent_noown x

Profession Type (enter ALL to reset)

Consultant

Apply

Fig3: Relationship Between Property Ownership and Default Possibility



The default rates of the three groups ("rent","own","neither") are 12.264%,12.048% and 5.556% respectively

Data Description

The figure (Fig3) shows that the default rate of rental groups (12.264%) is higher than that of homeowners (12.048%).

Fig4: Default Ratio of Car_Ownership



Fig4: Default Ratio of Car_Ownership

Data app demo show

Two filters

House Ownership Selector

rented x

owned x

norent_noown x

⌵

Profession Type (enter ALL to reset)

ALL

Apply

House Ownership Selector

rented x

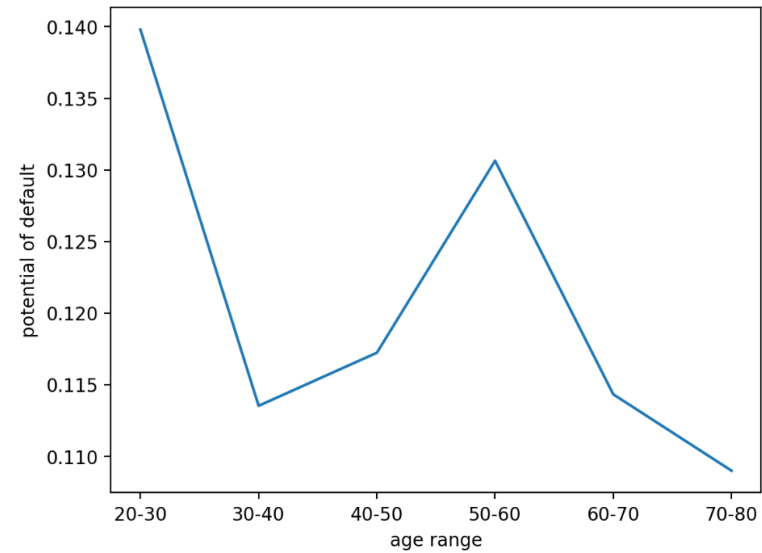
owned x

norent_noown x

Profession Type (enter ALL to reset)

ALL

Apply



The default rate of group with age range of 20-30 is: 13.980%

The default rate of group with age range of 30-40 is: 11.352%

The default rate of group with age range of 40-50 is: 11.722%

The default rate of group with age range of 50-60 is: 13.064%

The default rate of group with age range of 60-70 is: 11.432%

The default rate of group with age range of 70-80 is: 10.899%

Data Description

The figure (Fig7) shows that the default rate is the lowest in 60-70 age group (11.432%) and the highest in 50-60 age group (13.064%).





Group 11

POWERPOINT

THANK YOU