



# Credit Data Analysis

Academi Week 2

# Background, Goals, and Metadata



## BACKGROUND

People with various background apply for loan to the bank. For various reasons, some of these people do not pay their credit. Because of this the bank suffered loss. We are trying to avoid that by analyze history of clients and determine roughly what factors affect credit payments.

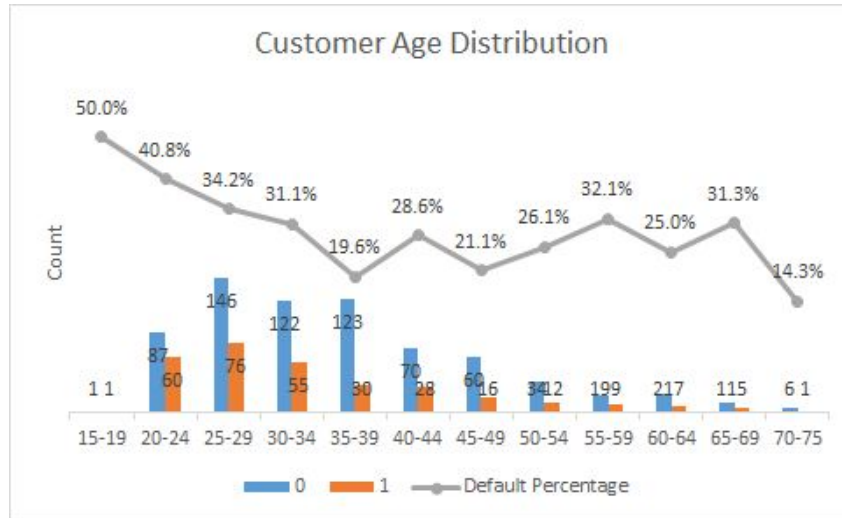
## GOALS

Analyze and point out important variables that affect loan credit default.

## METADATA

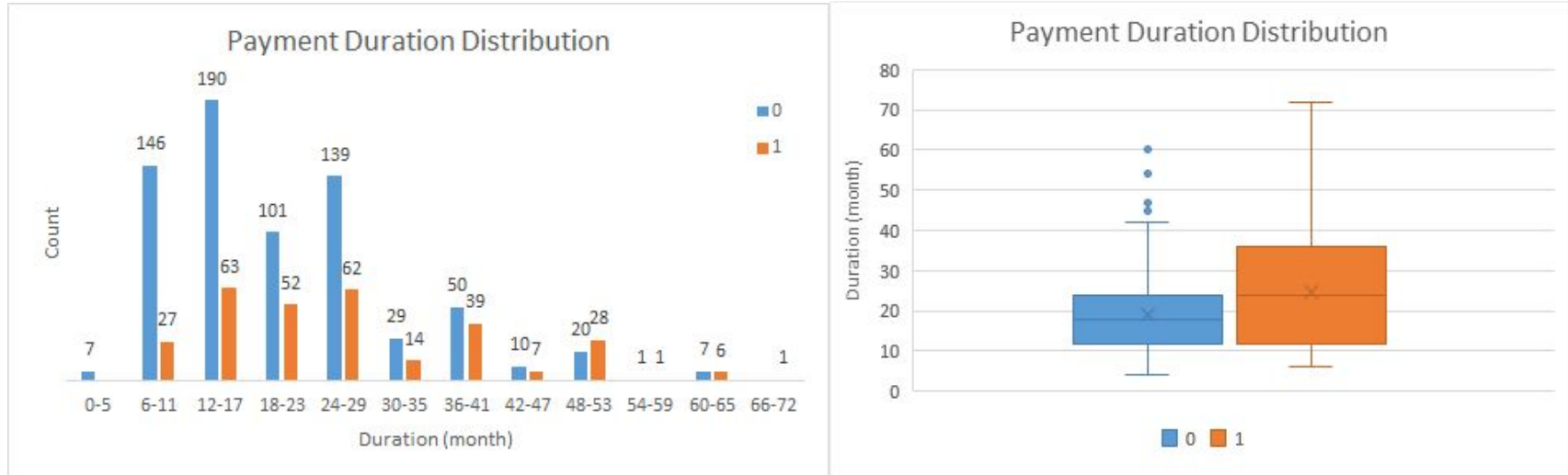
- The data contains 20 indepent variables / features, and the target variable (default): the classification whether an applicant is considered a Good or a Bad credit risk.
- The data consist of 1000 applicants data.
  - Default : **300** (1), 700 (0)
  - Female: 310 -- Male: 690
  - Single: 548 -- Not single: 452

# Customer Age



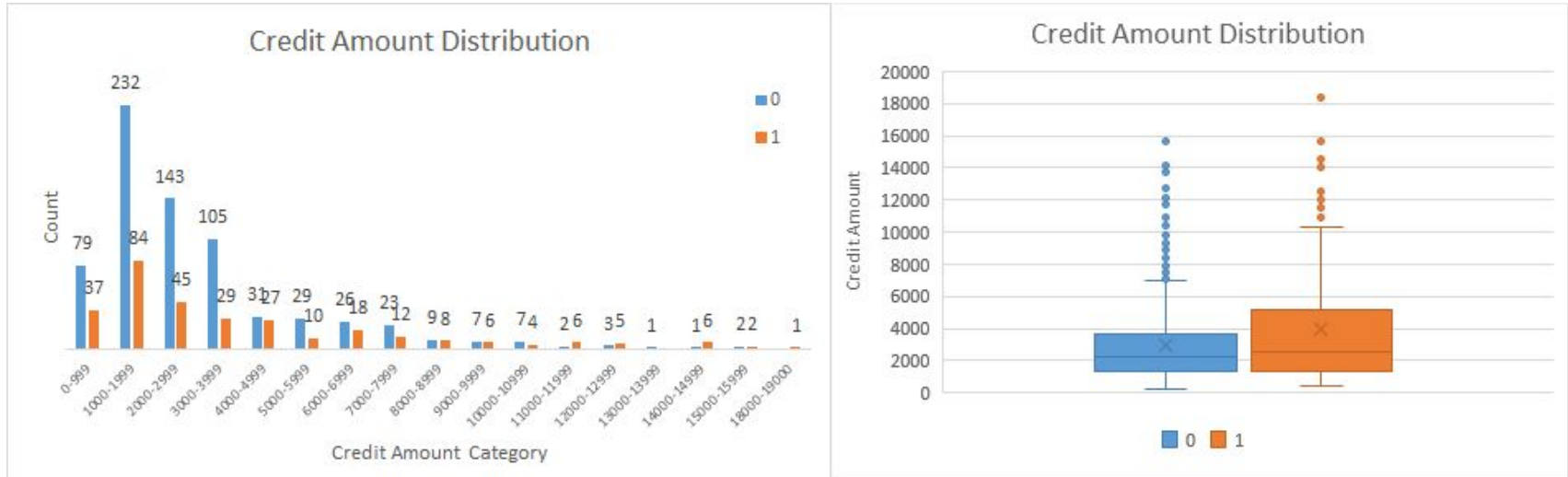
- Youth and young adults have tendency to have default credit especially on age 15 -34 (More than 30% chance)
- People with age 35 or more probably already have stable jobs

# Credit Duration



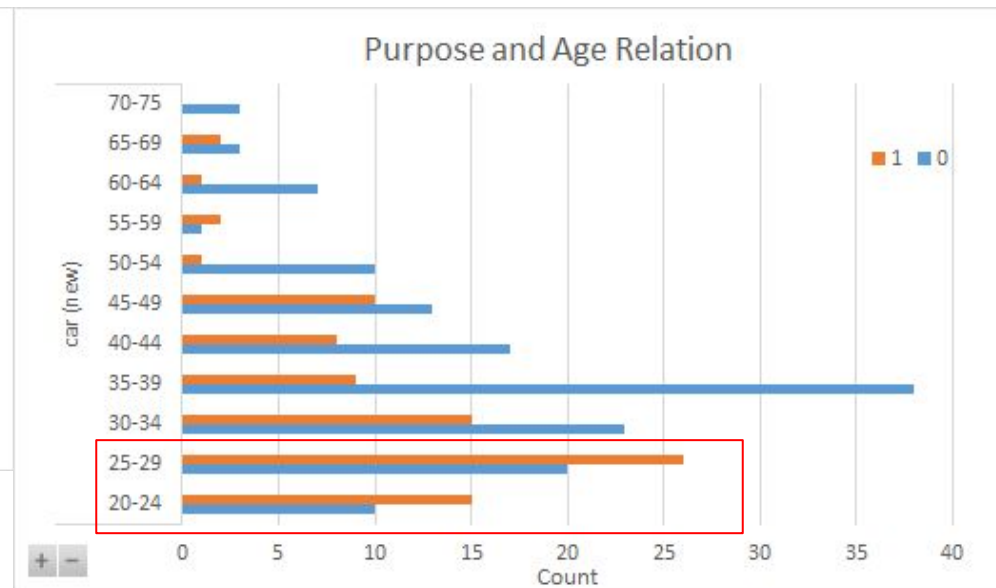
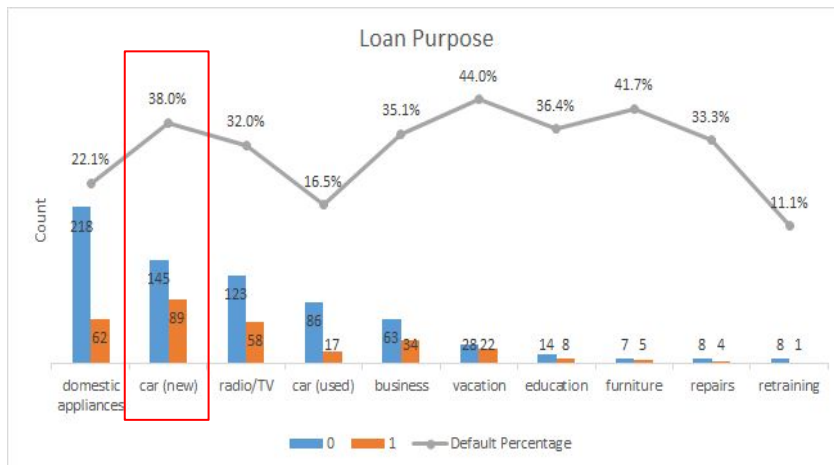
- Most of the non default applicant have below 30 month payment duration
- Most of the default applicant have below 42 month payment duration
- People with credit default tends to have higher duration payment

# Credit Amount



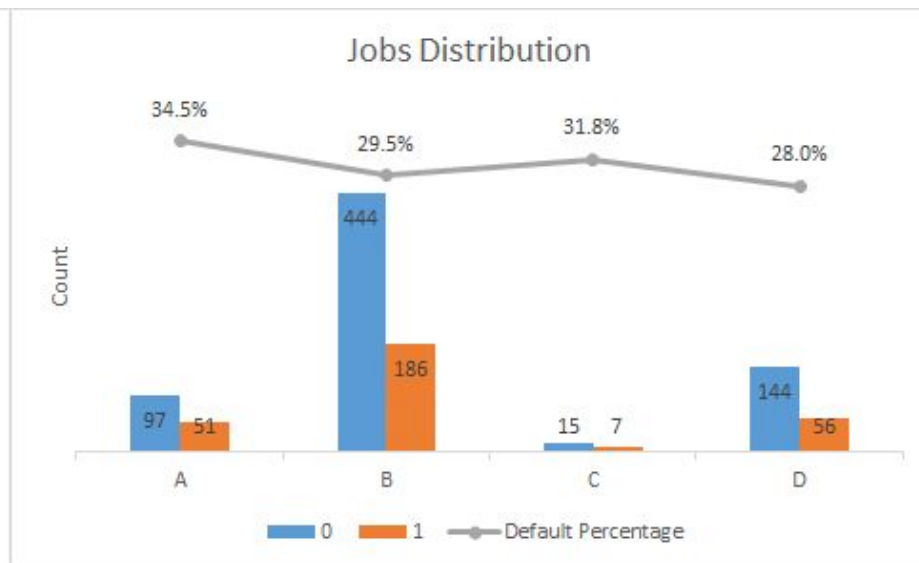
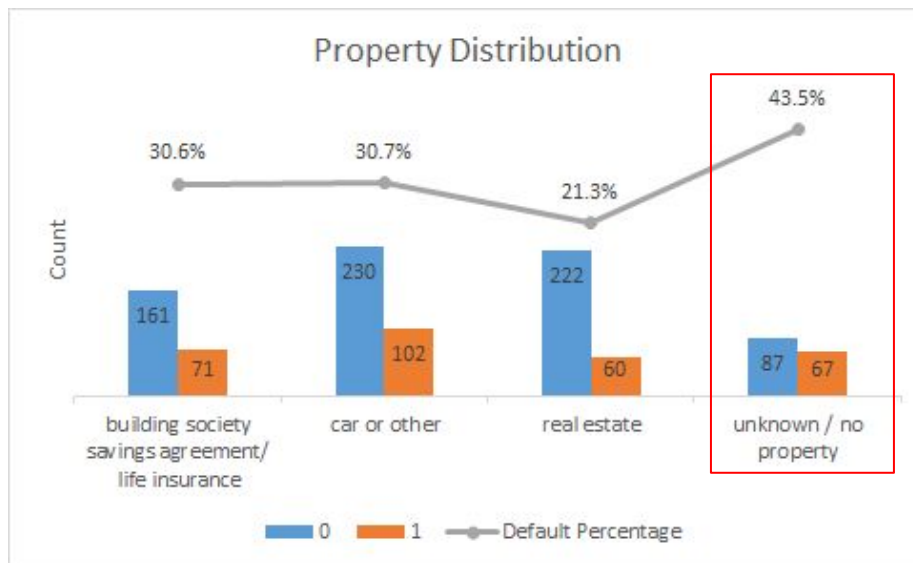
- Most of the non default customer have below 4k credit amount
- Most of the default customer have below 6k credit amount
- People with credit default tends to have higher credit amount

# Purpose to Use Loan



- Three biggest purpose is new car, domestic appliances, and radio/television
- From the three purpose, new car is the one with biggest risk to default (about 38%)
- From other purpose, the one with biggest risk are vacation (44%)
- People below 30 years old have a really high tendency to default if the purpose of their loan is new car

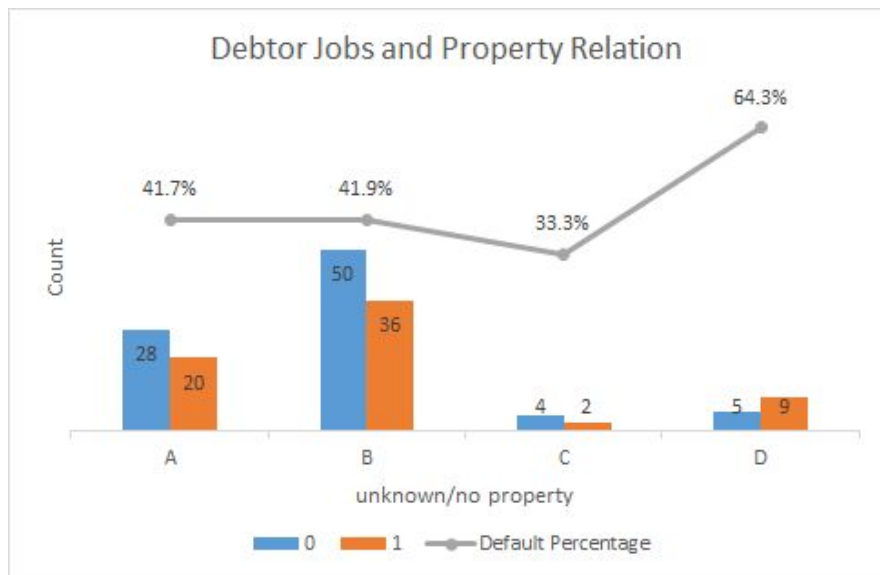
# Property and Jobs



- People with unknown property have higher chance to default
- Most of customer jobs are skilled employee/official

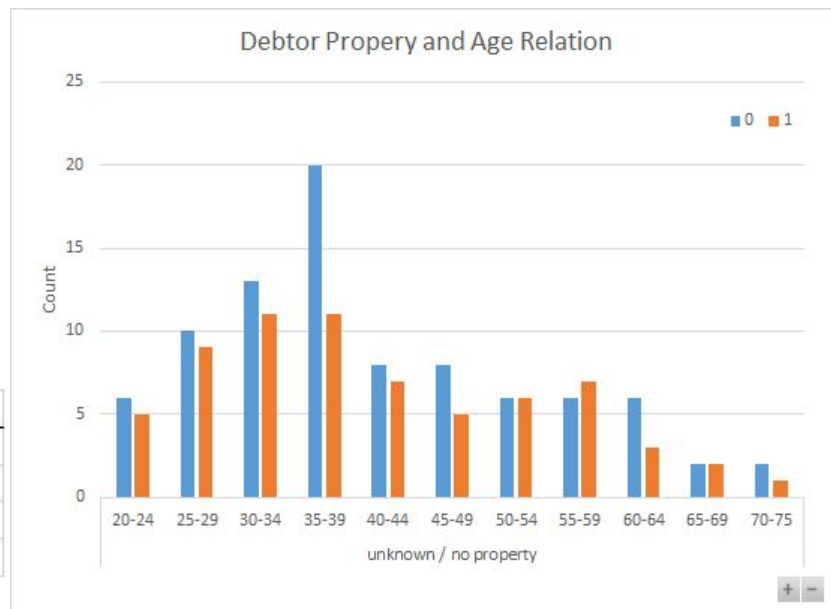
Category	Jobs
A	management/self-employed/highly qualified employee/officer
B	skilled employee/official
C	unemployed/unskilled - non-resident
D	unskilled - resident

# Property, Jobs, and Age



Category	Jobs
A	management/self-employed/highly qualified employee/officer
B	skilled employee/official
C	unemployed/unskilled - non-resident
D	unskilled - resident

- People jobs and age with unknown property distribution is even
- All of people with unknown property have high risk of default



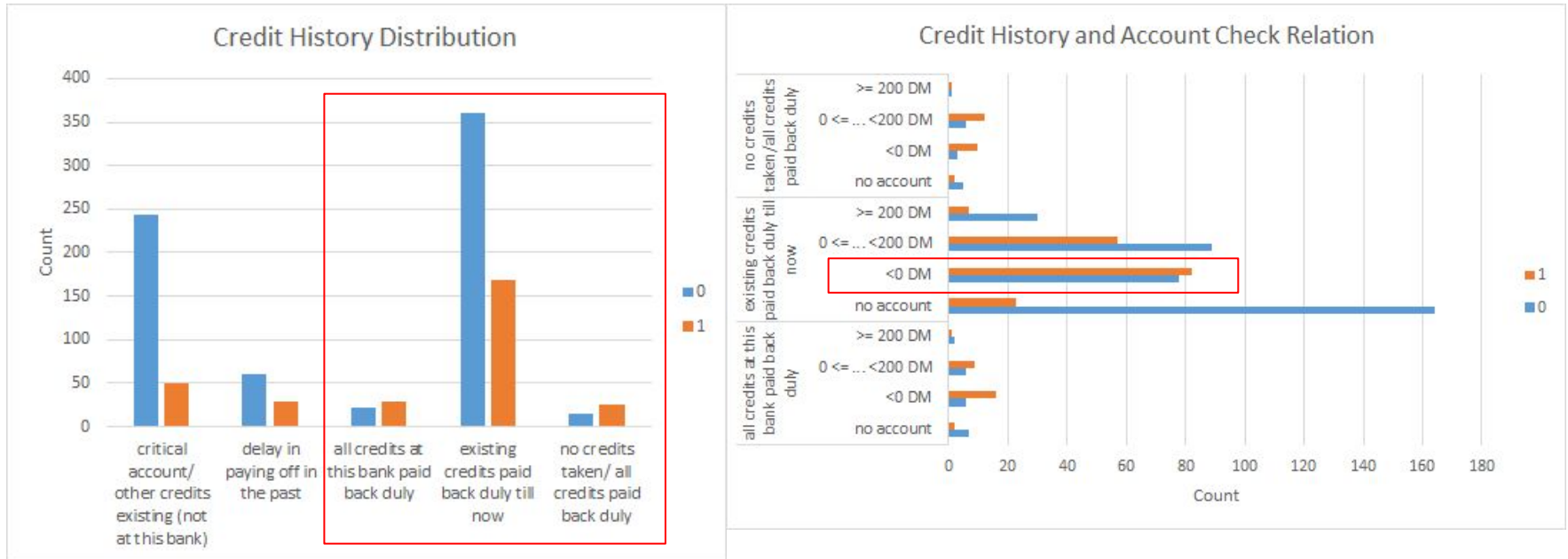


# Account Check Status



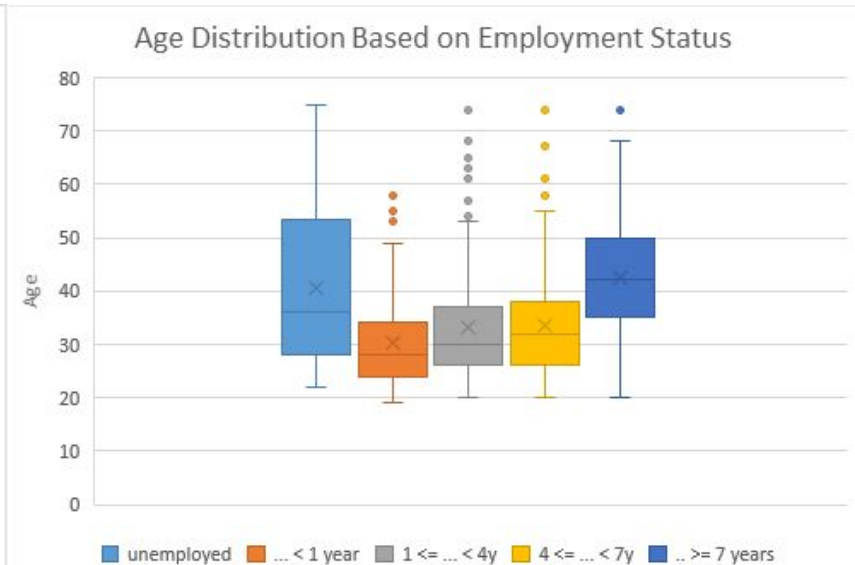
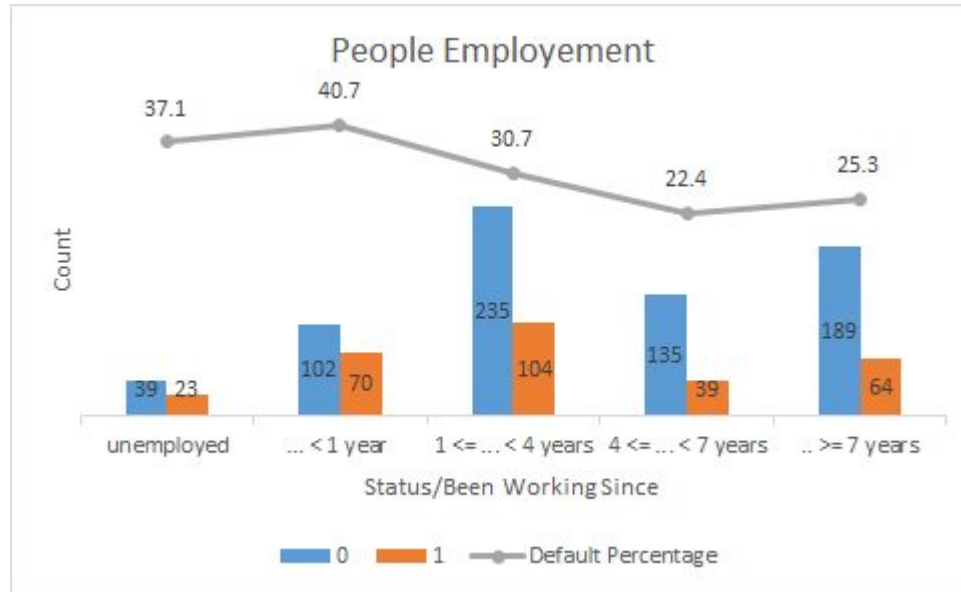
- People that have bank account and have <0 DM have higher risk to default with 135/274 people is default (49.3%) followed by 0 <= ... < 200 DM with 105/269 (39%)
- Especially people with <0 DM and work as skilled employee/official have high tendency to default (96/172)

# Credit History



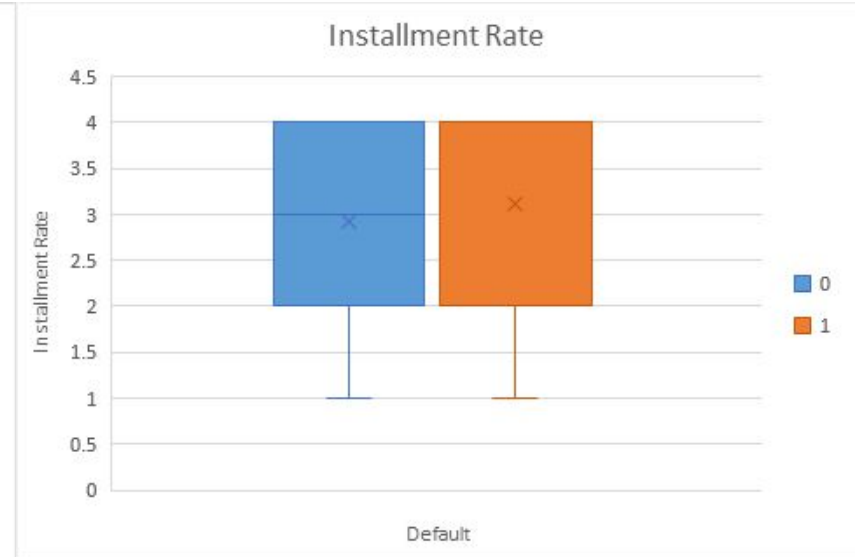
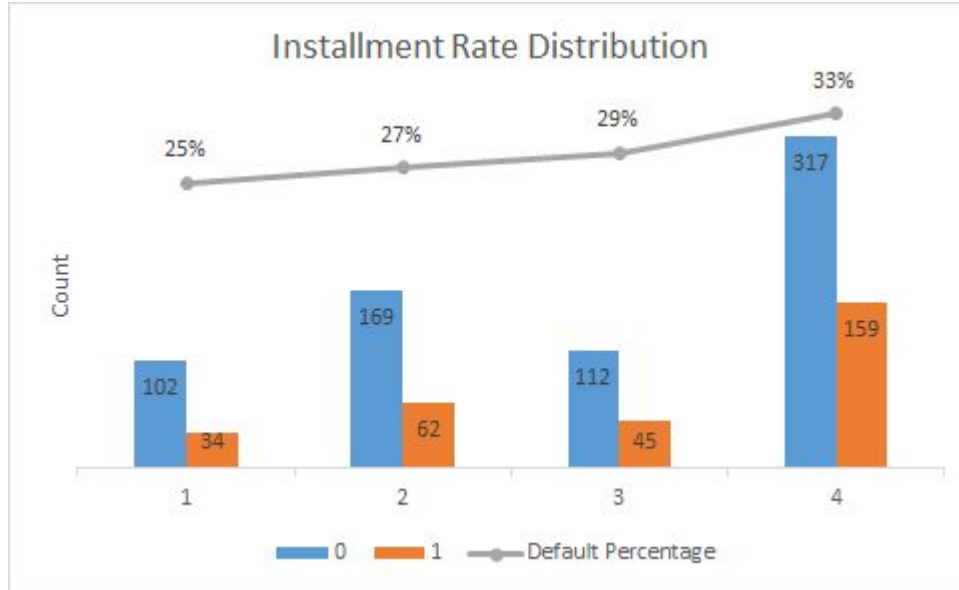
- If we compare it with their account check status, the risk of default is higher on people who paid their credit duly
- This shows that the credit history doesn't really determine people going to default or not

# Employment



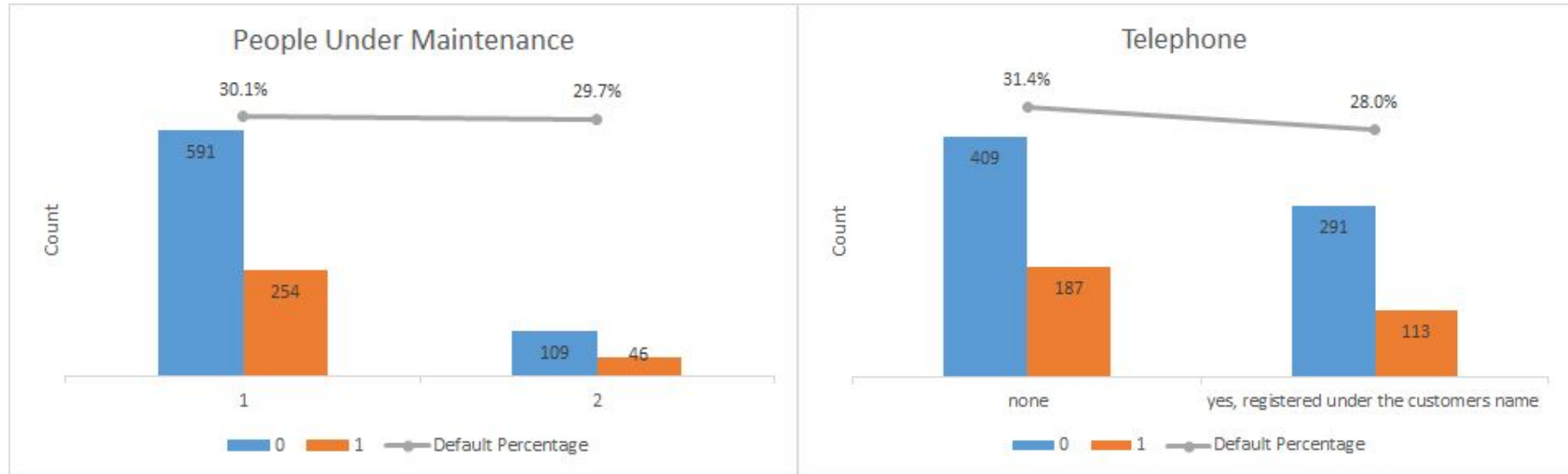
- People that unemployed and have employment age of 4 years and below has more than 30% tendency to default
- The longer a person works, the more stable their economic condition and the smaller the risk of default loans
- Unemployed people have a big range of age while other range increase following their age range

# Installment Rate



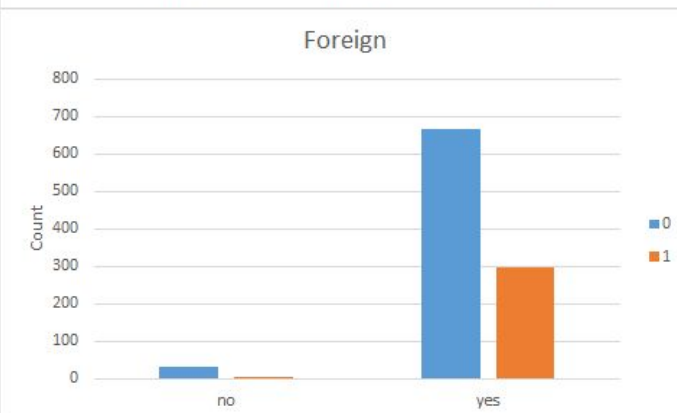
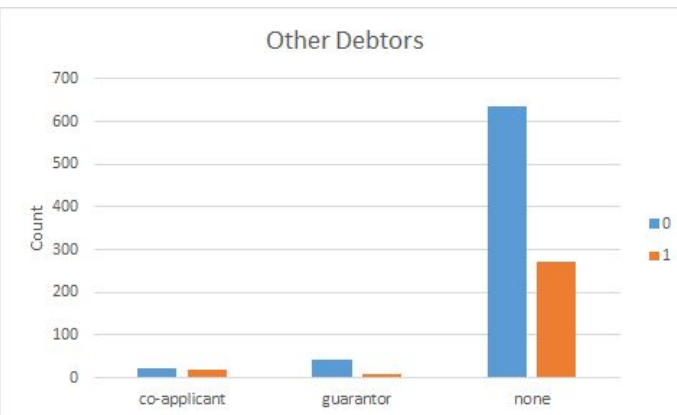
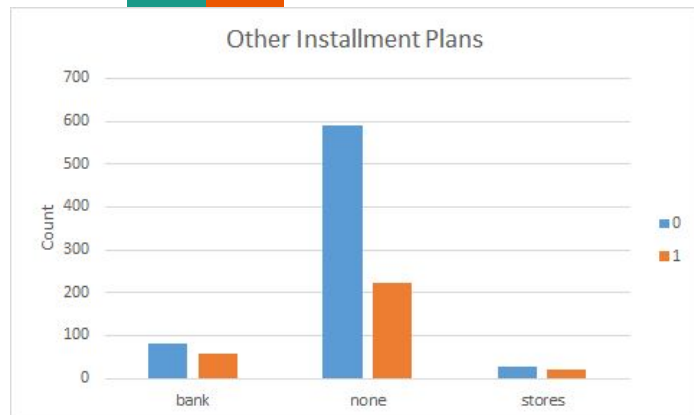
- People with more installment rate increase risk to default
- Make consideration to give a loan the higher a person's installment rate rate

# Small Difference on Default Percentage




- There is no significant difference between people on both category with about 30% of both are default so these 2 category is not really important to make decision

# Variable with Dominant Category



- These 4 variables have dominant category and make the analysis bias toward each dominant category so we can exclude analysis results from these 4

# Conclusion



Variable that affecting the loan credit default significantly :

- property (unknown/no property)
- age (people below 30)
- purpose (new car)
- account\_check\_status (< 0 DM)
- job (skilled employee/official)