



# DATA MINING PROJECT REPORT

Maddalena Amendola

Daniele Gadler

Riccardo Manetti

Gemma Martini

October 14, 2018

# Contents

<b>1</b>	<b>Data Understanding</b>	<b>2</b>
1.1	BoxPlots by Gemma . . . . .	2
1.1.1	Age . . . . .	2
1.1.2	Limit . . . . .	2
1.1.3	Billing amount . . . . .	3
1.1.4	Amount of previous payments . . . . .	3
1.1.5	Payment status . . . . .	4

# 1 Data Understanding

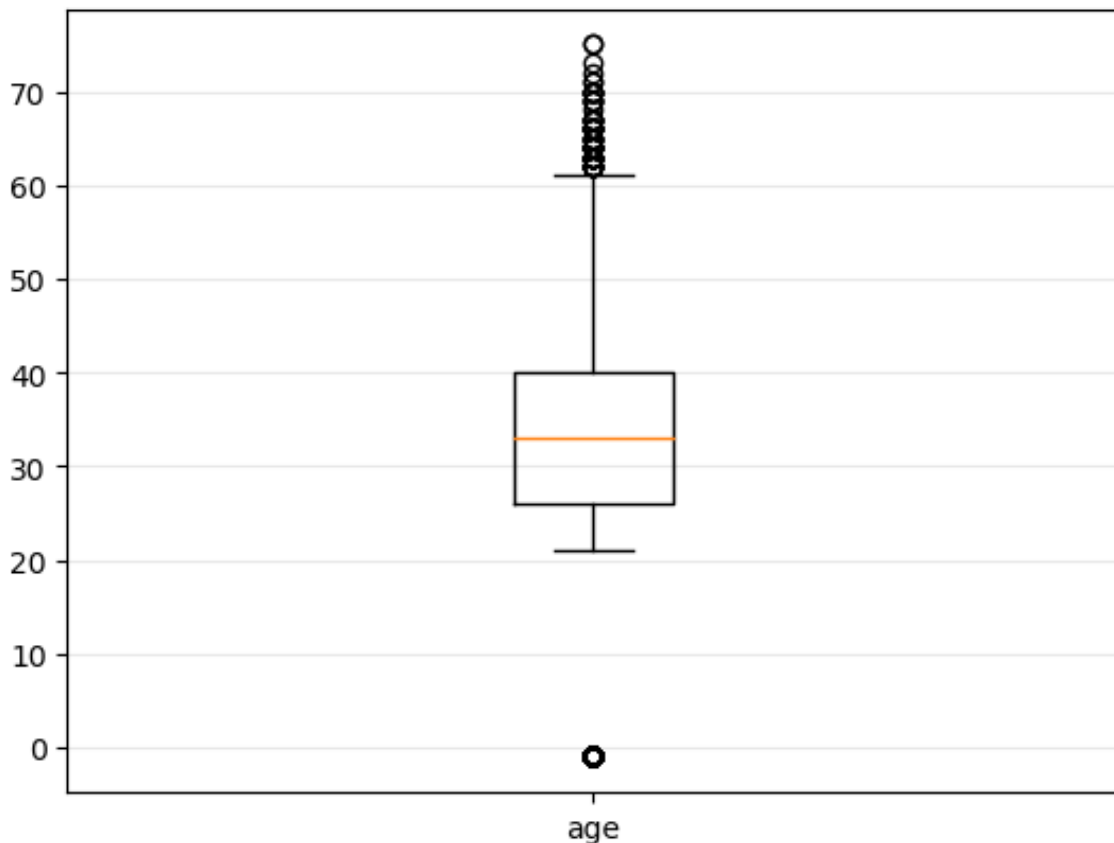
This part contains a first analysis of the data contained in the Taiwan credit card dataset. The collected data take into account users movements from April 2005 to September 2005.

## 1.1 BoxPlots by Gemma

These boxplots are created with Python and give us the chance to understand the shape of the given dataset, in order to find out possible outlayers, which may give ah hint on the many characteristics of the data.

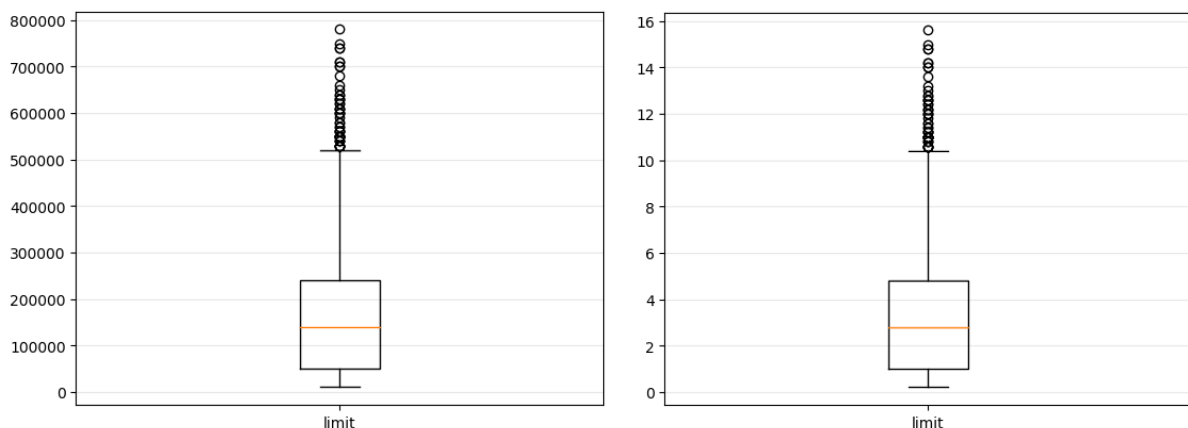
### 1.1.1 Age

From this plot we can observe that age has some values below 0. We can assume that the values allowed for this field are strictly positive.



### 1.1.2 Limit

Since the dataset takes into account Taiwanese dollars ( $\sim 0,028$  €) plotting the boxplot in terms of multiples of the average Taiwanese income (NTD 49989, about USD 1700) is more informative.

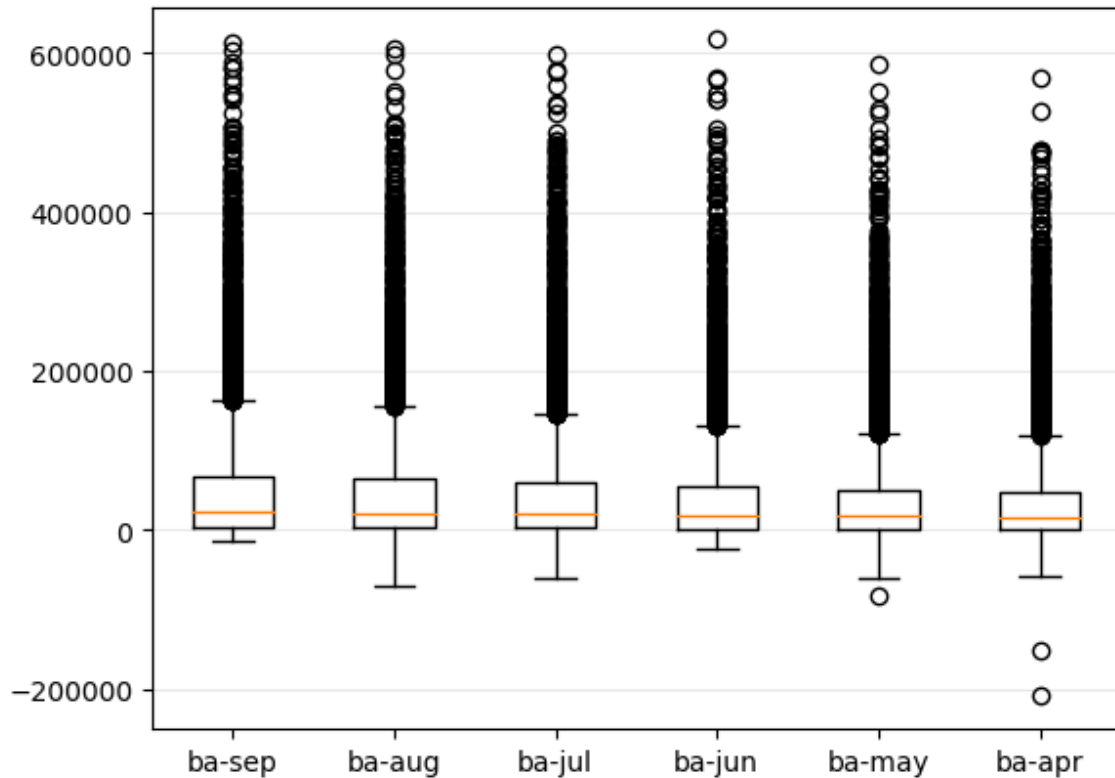


From this plot we may observe that on average the limit of the credit card plafond is 2.5 times the average income.

### 1.1.3 Billing amount

This values represents the balance of the bank account during the previous billing cycle. I can't understand what all these outliers mean!!!

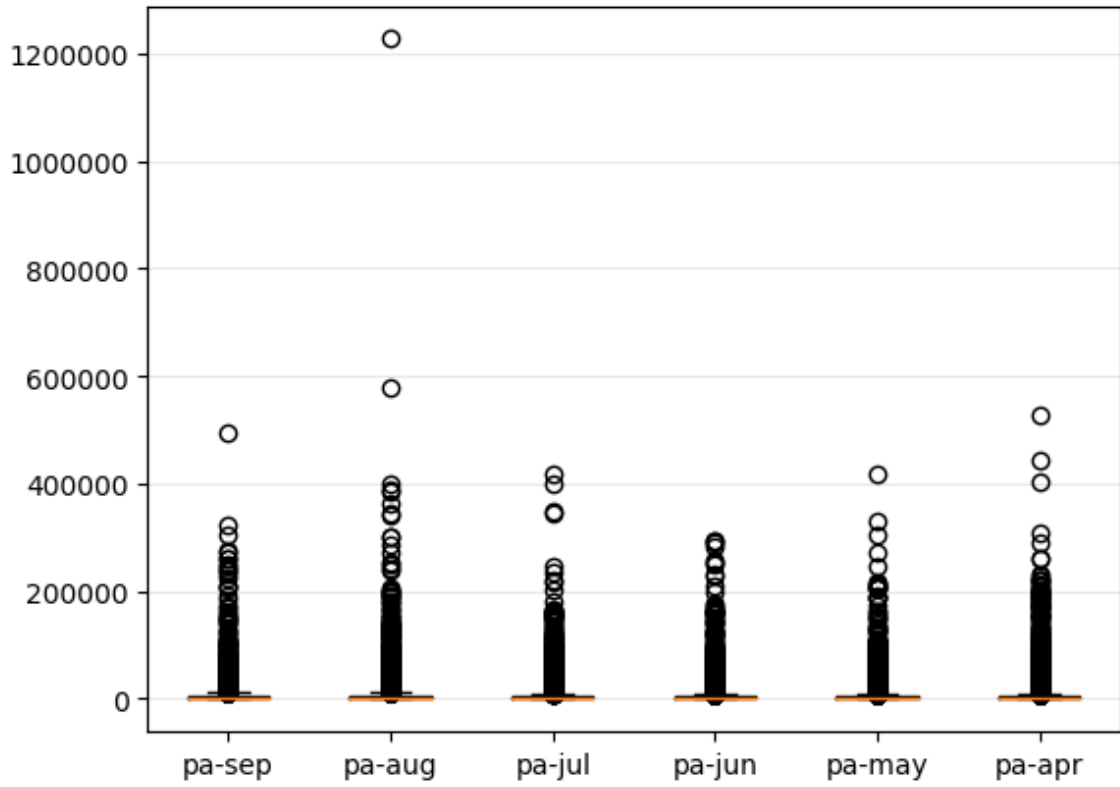
I can't understand



### 1.1.4 Amount of previous payments

This values tell us how much the user spent each month in the past 6 months, on a monthly basis.

Non capisco :-C



#### 1.1.5 Payment status

This columns of the database contain the history of payments, more precisely:

PAYED IN TIME → value -1

PAYED ONE MONTH LATER → value 1

PAYED TWO MONTHS LATER → value 2

PAYED THREE MONTHS LATER → value 3

PAYED FOUR MONTHS LATER → value 4

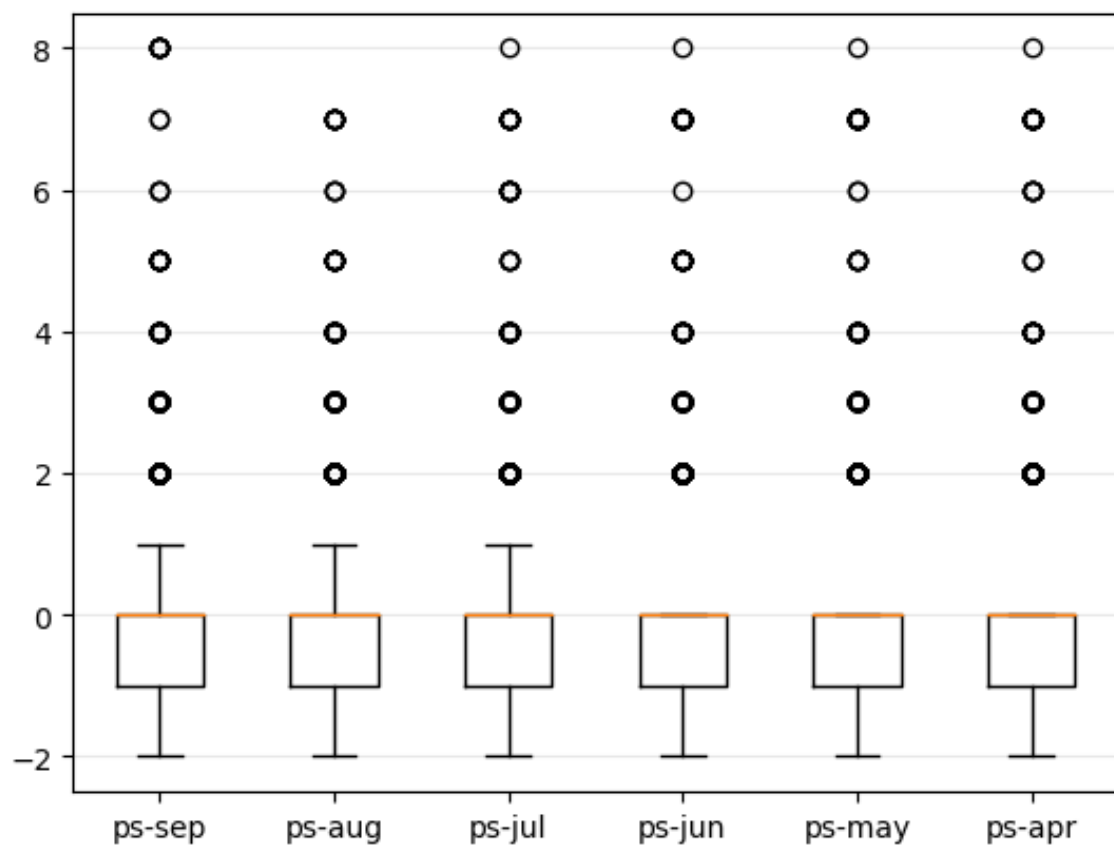
PAYED FIVE MONTHS LATER → value 5

PAYED SIX MONTHS LATER → value 6

PAYED SEVEN MONTHS LATER → value 7

PAYED EIGHT MONTHS LATER → value 8

PAYED NINE MONTHS LATER OR MORE → value 9



Direi che  
questo lo  
togliamo  
...???