



# 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	3
1.1.3	ba	3
1.1.4	pa	4
1.1.5	ps	4

# 1 Data Understanding

This part contains a first analysis of the data contained in the credit card dataset.

## 1.1 BoxPlots by Gemma

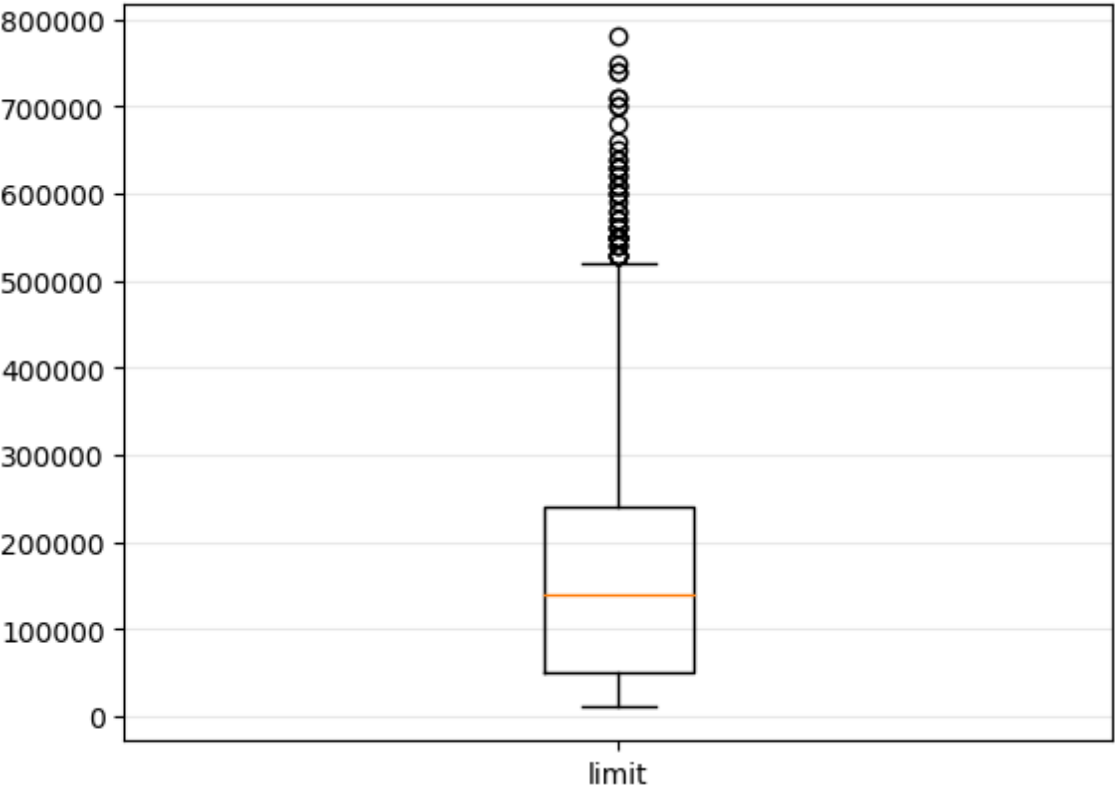
These boxplots are created with Python and give us the chance to understand the shape of the given dataset, in order to find out possible outliers, which may give ah hint on the many characteristics of the data.

### 1.1.1 Age

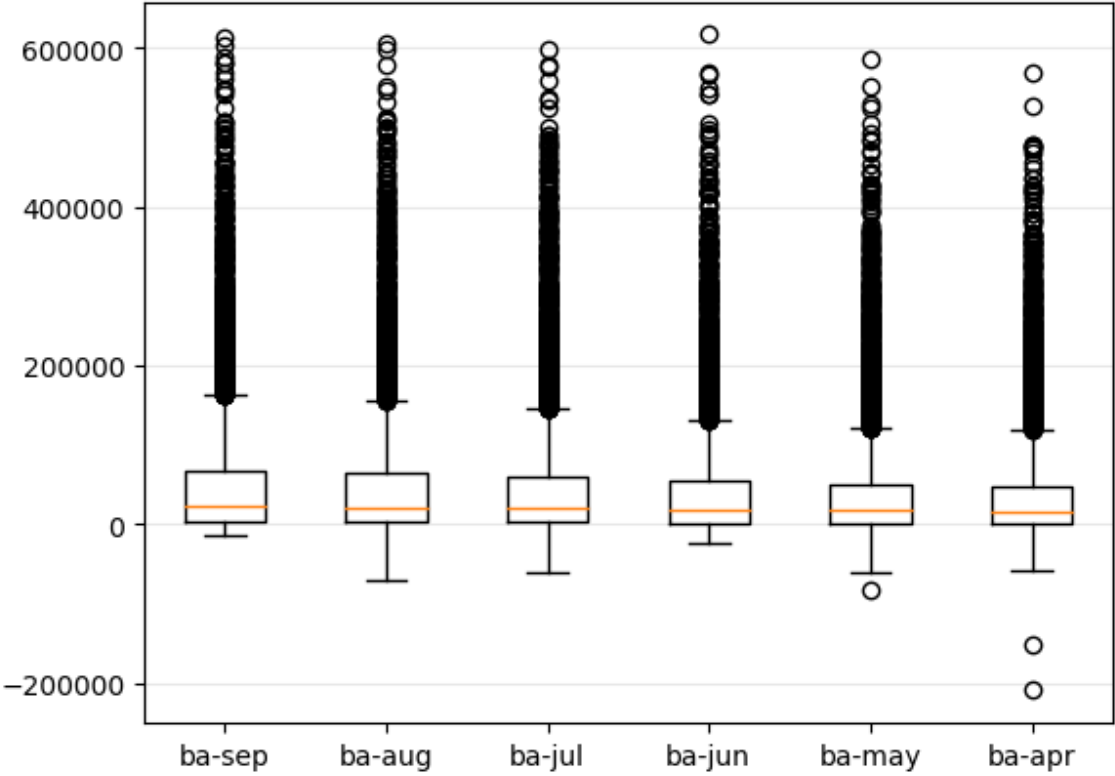
From this plot we can observe that age has some values below 0. We can assume that the values allowed for this field are strictly positive. We also notice that very low values (below ten years old) are pretty common and this is something that it's probably due to the high concentration of  $-1$  values.



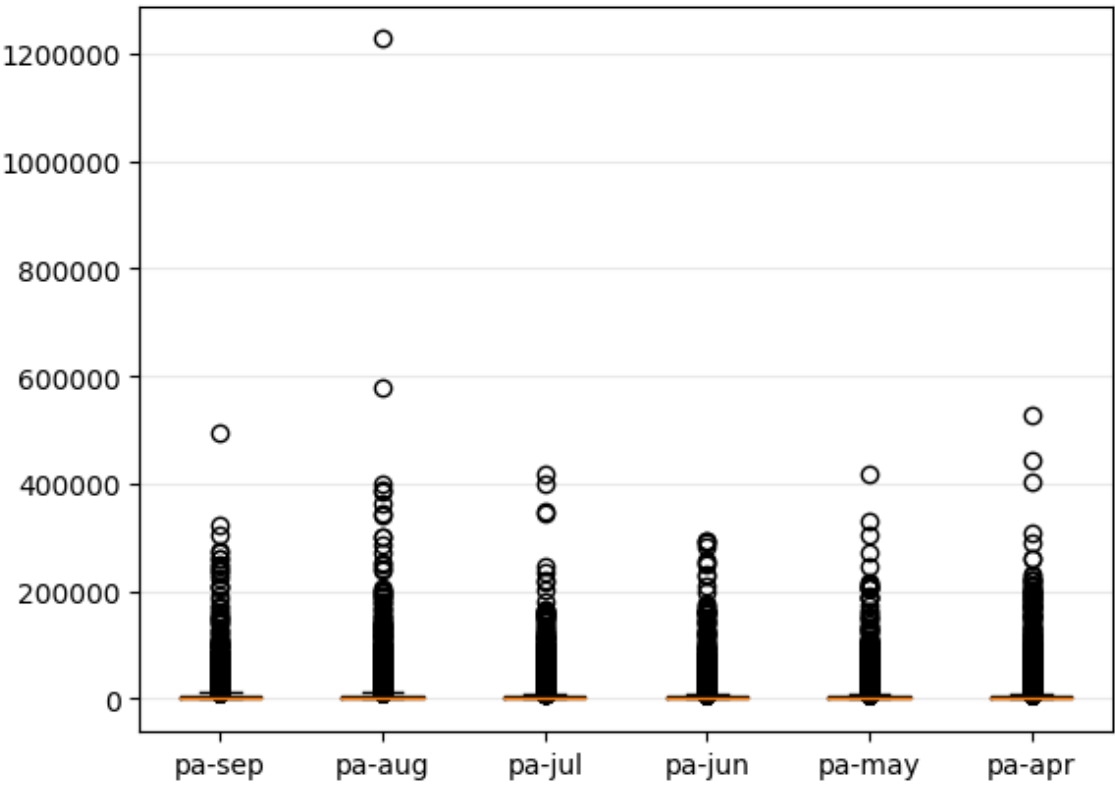
1.1.2 Limit



1.1.3 ba



1.1.4 pa



1.1.5 ps

