REPORT SERIES WITH DLOOKR

Exploratory Data Analysis Report

Author:
dlookr package

Version:
0.4.0

March 27, 2021

Contents

Chapter 1

Introduction

The EDA Report provides exploratory data analysis information on objects that inherit data.frame and data.frame.

1.1 Information of Dataset

The dataset that generated the EDA Report is an 'data.frame' object. It consists of 20,000 observations and 22 variables.

1.2 Information of Variables

Table 1.1: Information of Variables

variables	types	missing_count	missing_percent	unique_count	unique_rate
tot_credit_debt	numeric	0	0.00	19978	0.999
avg_card_debt	numeric	0	0.00	19607	0.980
credit_age	numeric	0	0.00	410	0.020
credit_good_age	numeric	0	0.00	243	0.012
card_age	numeric	0	0.00	383	0.019
non_mtg_acc_past_due_12_months_num	character	0	0.00	5	0.000
non_mtg_acc_past_due_6_months_num	character	0	0.00	3	0.000
mortgages_past_due_6_months_num	character	0	0.00	2	0.000
credit_past_due_amount	numeric	0	0.00	605	0.030
inq_12_month_num	numeric	0	0.00	11	0.001
card_inq_24_month_num	numeric	0	0.00	19	0.001
card_open_36_month_num	character	0	0.00	3	0.000
auto_open_.36_month_num	character	0	0.00	3	0.000
uti_card	numeric	0	0.00	20000	1.000
uti_50plus_pct	numeric	0	0.00	20000	1.000
uti_max_credit_line	numeric	0	0.00	20000	1.000
uti_card_50plus_pct	numeric	2055	10.27	17946	0.897
ind_acc_XYZ	character	0	0.00	2	0.000
rep_income	numeric	1570	7.85	118	0.006
States	factor	0	0.00	7	0.000
Default_ind	numeric	0	0.00	2	0.000
out	factor	0	0.00	2	0.000

The target variable of the data is 'out', and the data type of the variable is factor.

1.3   About EDA Report

EDA reports provide information and visualization results that support the EDA process. In particular, it provides a variety of information to understand the relationship between the target variable and the rest of the variables of interest.

Chapter 2

Univariate Analysis

2.1 Descriptive Statistics

edaData

22 Variables   20000 Observations

---

tot_credit_debt

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	19978	1	94564	26555	55824	64443	78744	94671	110329	124707	133415

```
lowest :    2367.43   3664.49   4662.60   6898.50  11363.34
highest: 175998.38 179084.56 182094.91 182858.99 188890.96
```

---

avg_card_debt

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	19607	1	14088	4913	8454	9555	11322	13244	15196	16935	18039

```
lowest :  2363.12  2521.21  2814.66  3074.70  3148.68, highest: 19945.05 19955.42 19959.03 19960.61 99999.00
```

Value	2000	3000	4000	5000	6000	7000	8000	9000	10000	11000	12000	13000
Frequency	1	6	8	51	123	258	591	909	1452	2017	2511	2803
Proportion	0.000	0.000	0.000	0.003	0.006	0.013	0.030	0.045	0.073	0.101	0.126	0.140

Value	14000	15000	16000	17000	18000	19000	20000	100000
Frequency	2638	2261	1801	1158	695	396	109	212
Proportion	0.132	0.113	0.090	0.058	0.035	0.020	0.005	0.011

For the frequency table, variable is rounded to the nearest 1000

---

credit_age

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	410	1	296.7	69.64	195	217	255	297	339	375	398

```
lowest :  54  78  79  80  82, highest: 521 527 537 539 545
```

---

credit_good_age

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	243	1	149.8	38.34	94	106	127	150	172	193	205

```
lowest :  21  26  27  28  31, highest: 279 280 281 283 296
```

---

card_age

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	383	1	268	67.04	171	191	227	268	308	344	365

```
lowest :  41  56  62  71  75, highest: 481 484 494 516 520
```

---

non_mtg_acc_past_due_12_months_num

n	missing	distinct
20000	0	5

```
lowest : 0 1 2 3 4, highest: 0 1 2 3 4
```

Value	0	1	2	3	4
Frequency	18502	918	446	119	15
Proportion	0.925	0.046	0.022	0.006	0.001

---

non_mtg_acc_past_due_6_months_num

n	missing	distinct
20000	0	3

Value	0	1	2
Frequency	19481	490	29
Proportion	0.974	0.024	0.001

---

mortgages_past_due_6_months_num

n	missing	distinct
20000	0	2

Value	0	1
Frequency	19396	604
Proportion	0.97	0.03

---

credit_past_due_amount

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	605	0.088	329.3	643.9	0	0	0	0	0	0	0

```
lowest :    0.00   316.39   434.70   602.68   695.96, highest: 27229.53 27726.89 28644.74 29392.72 32662.98
```

---

inq_12_month_num

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	11	0.948	1.763	1.893	0	0	0	1	3	4	5

```
lowest :  0  1  2  3  4, highest:  6  7  8  9 10
```

Value	0	1	2	3	4	5	6	7	8	9	10
Frequency	6696	3541	3475	2871	1824	968	423	153	37	11	1
Proportion	0.335	0.177	0.174	0.144	0.091	0.048	0.021	0.008	0.002	0.001	0.000

---

card_inq_24_month_num

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	19	0.984	3.41	3.237	0	0	1	3	5	8	9

```
lowest :  0  1  2  3  4, highest: 14 15 16 17 18
```

Value	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Frequency	3936	2452	2654	2401	2093	1809	1503	1092	824	521	341	189	93	58
Proportion	0.197	0.123	0.133	0.120	0.105	0.090	0.075	0.055	0.041	0.026	0.017	0.009	0.005	0.003

Value	14	15	16	17	18
Frequency	18	5	6	3	2
Proportion	0.001	0.000	0.000	0.000	0.000

---

card_open_36_month_num

n	missing	distinct
20000	0	3

Value	0	1	2
Frequency	16865	3009	126
Proportion	0.843	0.150	0.006

---

auto_open_.36_month_num

n	missing	distinct
20000	0	3

Value	0	1	2
Frequency	17191	2798	11
Proportion	0.860	0.140	0.001

---

uti_card

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	20000	1	0.5032	0.1233	0.3238	0.3628	0.4296	0.5028	0.5774	0.6443	0.6816

```
lowest : 0.06512047 0.06563675 0.07869497 0.10148322 0.11754010
highest: 0.89357072 0.90489927 0.92232634 0.92532315 0.96928868
```

---

uti_50plus_pct

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	20000	1	0.511	0.128	0.3254	0.3653	0.4352	0.5099	0.5884	0.6566	0.6975

```
lowest : 0.03374933 0.07398763 0.08376058 0.11596965 0.12081086
highest: 0.89448028 0.89499581 0.90084806 0.90509788 0.98896404
```

uti_max_credit_line

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
20000	0	20000	1	0.5076	0.1226	0.3290	0.3680	0.4335	0.5072	0.5814	0.6467	0.6874

```
lowest : 0.005173925 0.091742468 0.098516713 0.115342939 0.117451965
highest: 0.894630428 0.903665489 0.912962710 0.971640159 1.000000000
```

uti_card_50plus_pct

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50
17945	2055	17945	1	0.4896	0.1348	0.2923	0.3380	0.4098	0.4901

.75	.90	.95
0.5690	0.6431	0.6855

```
lowest : 0.000000000 0.005784274 0.032522037 0.065678794 0.068748893
highest: 0.918661007 0.929283466 0.931222261 0.949958864 0.970775774
```

ind_acc_XYZ

n	missing	distinct
20000	0	2

Value	0	1
Frequency	14829	5171
Proportion	0.741	0.259

rep_income

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
18430	1570	117	1	75500	18465	49000	55000	64000	75000	86000	97000	102000

```
lowest :  12000  18000  19000  20000  22000, highest: 130000 131000 132000 134000 150000
```

States

n	missing	distinct
20000	0	7

```
lowest : AL FL GA LA MS, highest: GA LA MS NC SC
```

Value	AL	FL	GA	LA	MS	NC	SC
Frequency	2893	2857	2857	2849	2827	2898	2819
Proportion	0.145	0.143	0.143	0.142	0.141	0.145	0.141

Default_ind

n	missing	distinct	Info	Sum	Mean	Gmd
20000	0	2	0.219	1586	0.0793	0.146

out

n	missing	distinct
20000	0	2

Value	0	1
Frequency	18414	1586
Proportion	0.921	0.079

2.2   Normality Test of Numerical Variables

2.2.1   Statistics and Visualization of (Sample) Data

tot_credit_debt

* normality test : Shapiro-Wilk normality test
- statistic : 0.99942, p-value : 0.117589

Table 2.1: skewness and kurtosis : tot_credit_debt

type	skewness	kurtosis
original	-0.0759	2.9708
log transformation	-1.1031	5.7589
sqrt transformation	-0.5117	3.5680
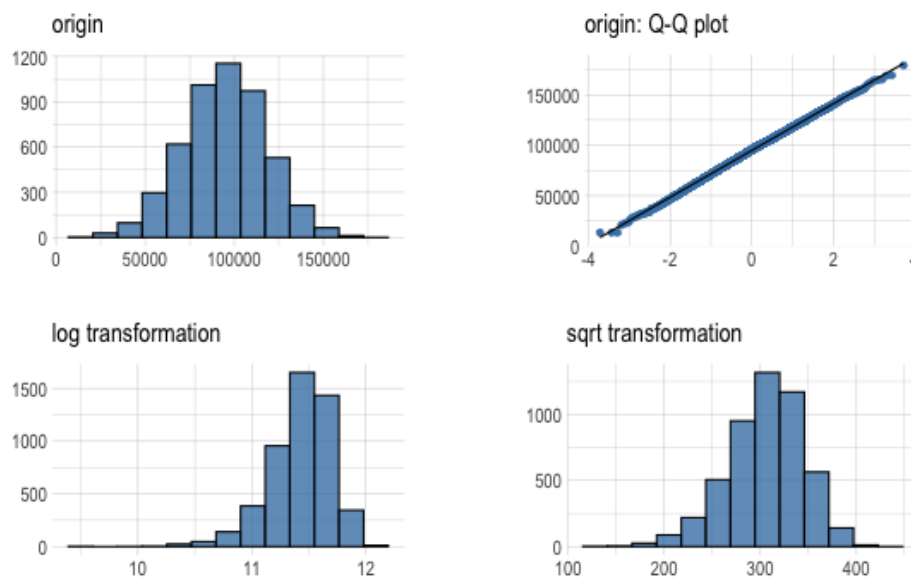
Figure 2.1: tot_credit_debt

avg_card_debt

* normality test : Shapiro-Wilk normality test
- statistic : 0.29595, p-value : 1.00664E-87

Table 2.2: skewness and kurtosis : avg_card_debt

type	skewness	kurtosis
original	8.1285	73.4388
log transformation	2.5735	21.0594
sqrt transformation	5.8720	49.0470
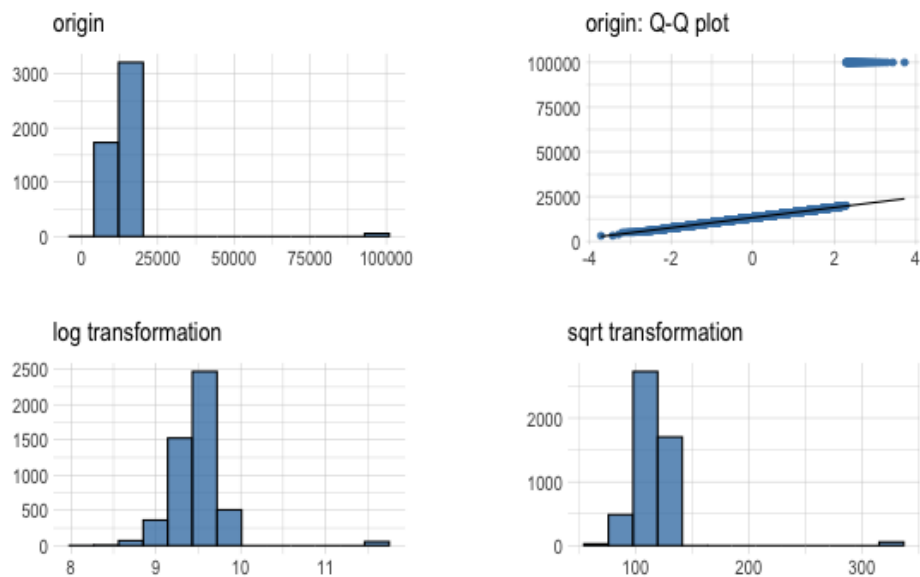
Figure 2.2: avg_card_debt

credit_age

* normality test : Shapiro-Wilk normality test
- statistic : 0.9996, p-value : 0.405466

Table 2.3: skewness and kurtosis : credit_age

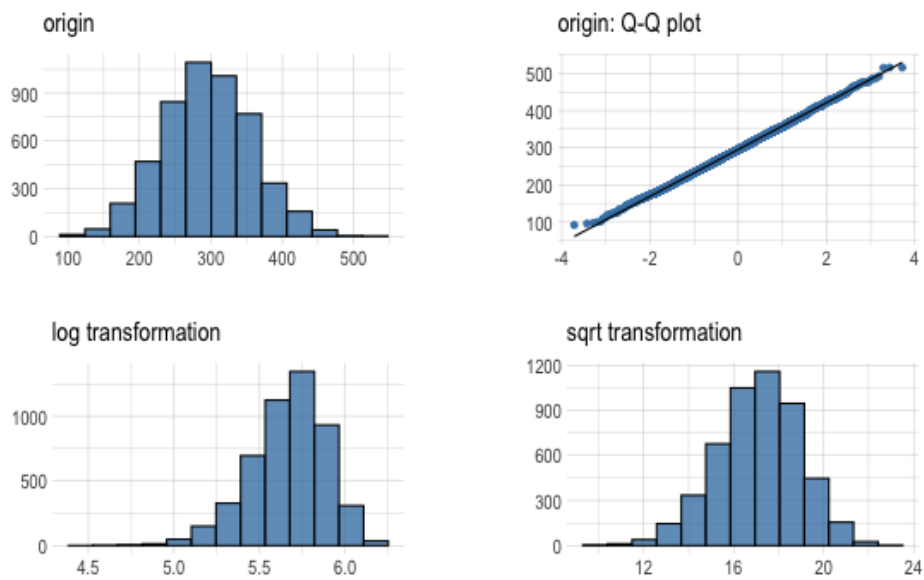| type	skewness	kurtosis
original	0.0472	2.9451
log transformation	-0.6656	3.9106
sqrt transformation	-0.2859	3.1577

Figure 2.3: credit_age

credit_good_age

* normality test : Shapiro-Wilk normality test
- statistic : 0.99923, p-value : 0.0272846

Table 2.4: skewness and kurtosis : credit_good_age

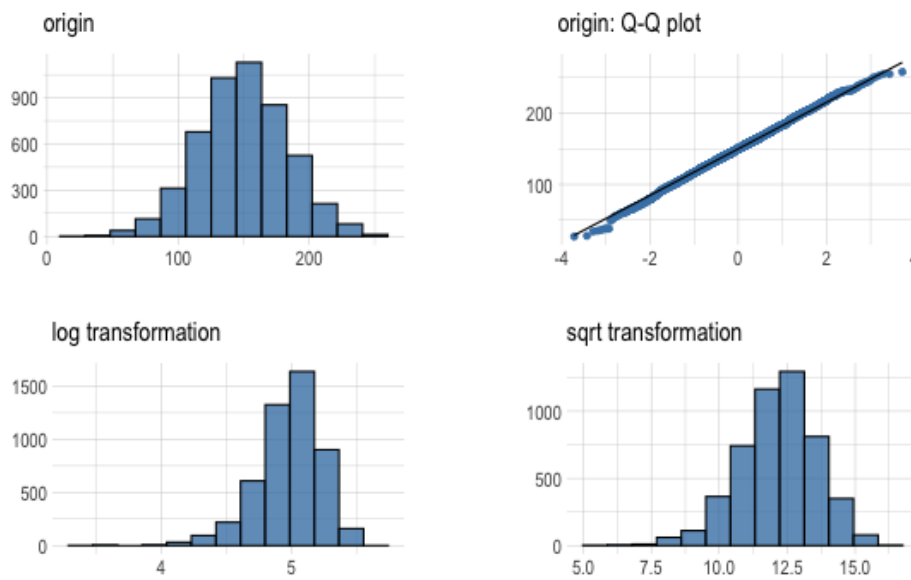| type	skewness	kurtosis
original	-0.0505	3.0563
log transformation	-1.0304	5.7109
sqrt transformation	-0.4696	3.6797

Figure 2.4: credit_good_age

card_age

* normality test : Shapiro-Wilk normality test
- statistic : 0.99944, p-value : 0.134592

Table 2.5: skewness and kurtosis : card_age

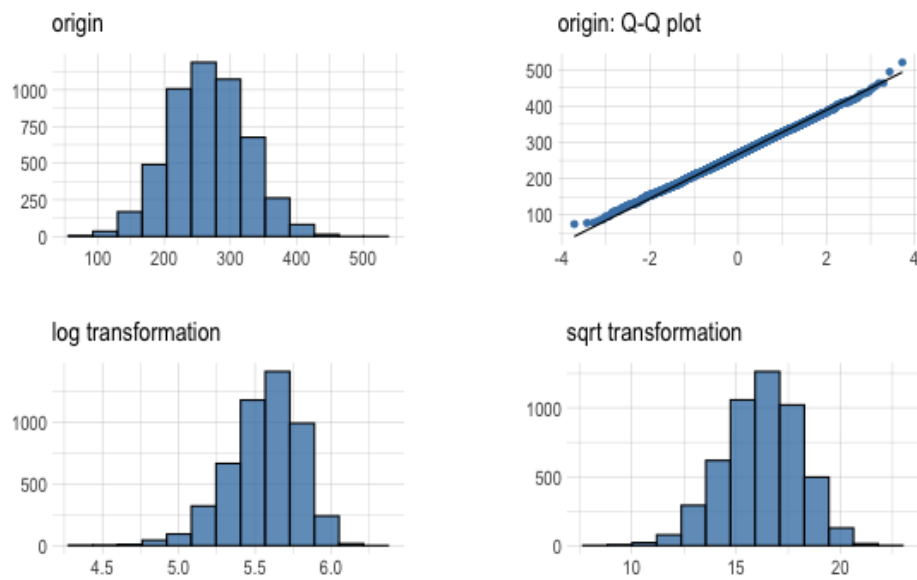| type	skewness	kurtosis
original	0.0516	2.9401
log transformation	-0.7002	4.0567
sqrt transformation	-0.2953	3.1696

Figure 2.5: card_age

credit_past_due_amount

* normality test : Shapiro-Wilk normality test
- statistic : 0.14897, p-value : 4.53231E-92

Table 2.6: skewness and kurtosis : credit_past_due_amount

type	skewness	kurtosis
original	7.3235	62.6301
log+1 transformation	5.5162	31.5245
sqrt transformation	5.9842	38.7520
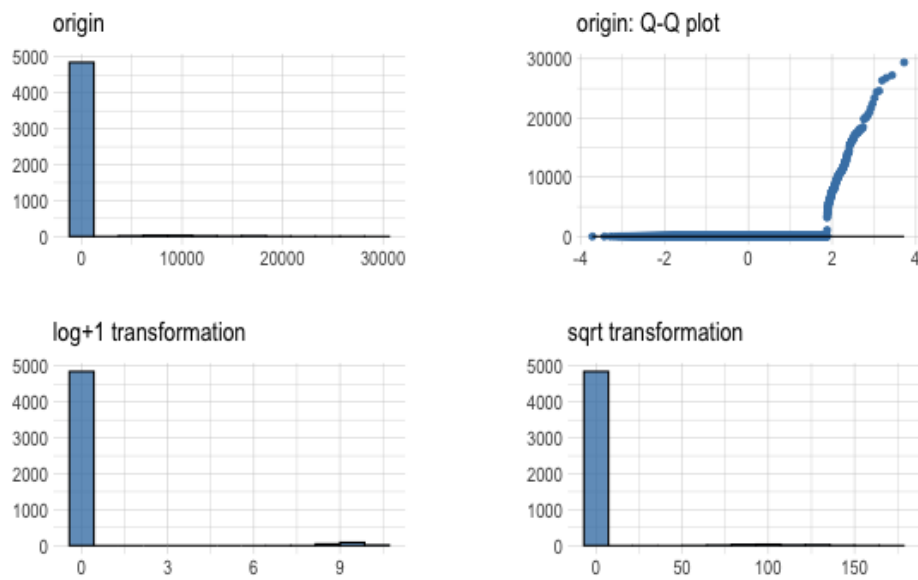
Figure 2.6: credit_past_due_amount

inq_12_month_num

* normality test : Shapiro-Wilk normality test
- statistic : 0.87073, p-value : 1.25395E-53

Table 2.7: skewness and kurtosis : inq_12_month_num

type	skewness	kurtosis
original	0.8443	3.0945
log+1 transformation	0.0210	1.6272
sqrt transformation	-0.0770	1.6445
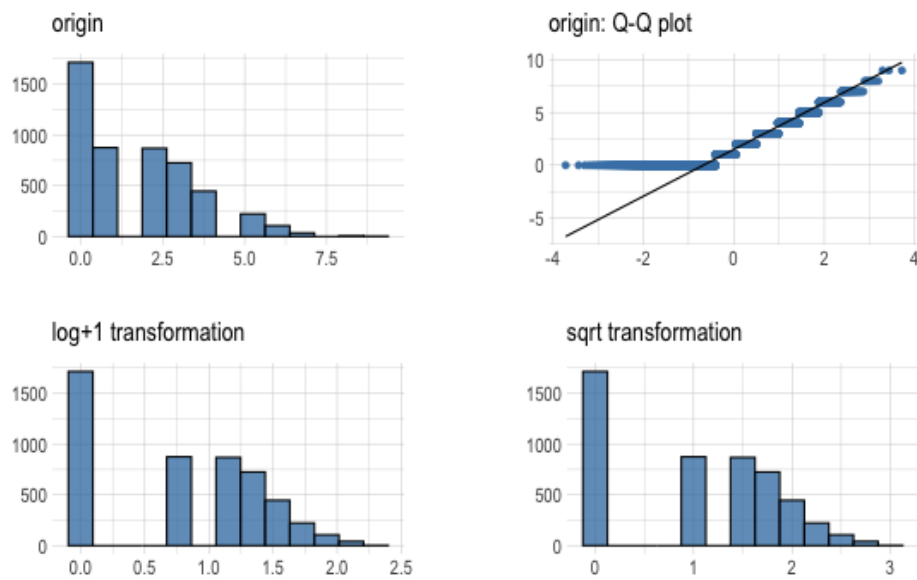
Figure 2.7: inq_12_month_num

card_inq_24_month_num

* normality test : Shapiro-Wilk normality test
- statistic : 0.91452, p-value : 1.31877E-46

Table 2.8: skewness and kurtosis : card_inq_24_month_num

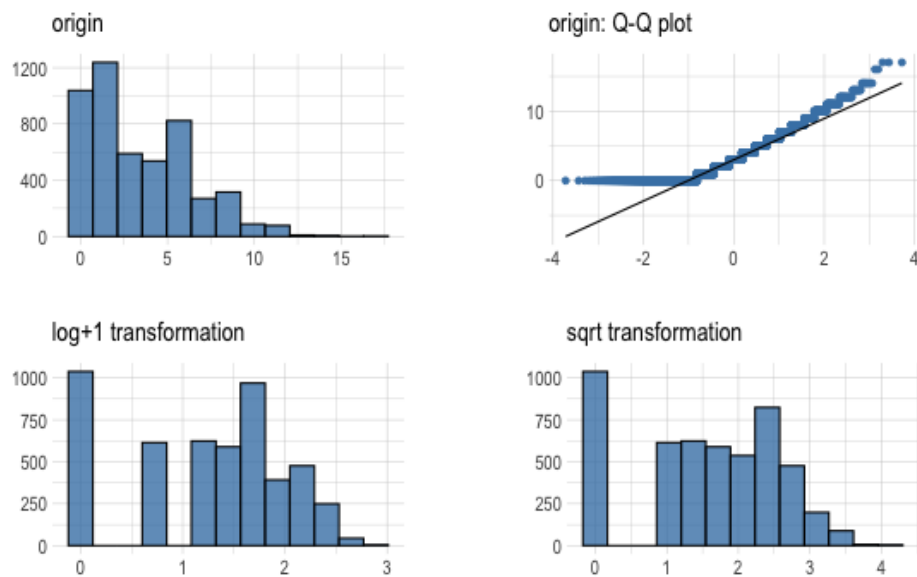| type	skewness	kurtosis
original	0.8279	3.3036
log+1 transformation	-0.3249	1.9364
sqrt transformation	-0.2828	2.1173

Figure 2.8: card_inq_24_month_num

uti_card

* normality test : Shapiro-Wilk normality test
- statistic : 0.99965, p-value : 0.57014

Table 2.9: skewness and kurtosis : uti_card

type	skewness	kurtosis
original	0.0407	3.0536
log transformation	-0.8274	4.9620
sqrt transformation	-0.3413	3.4234
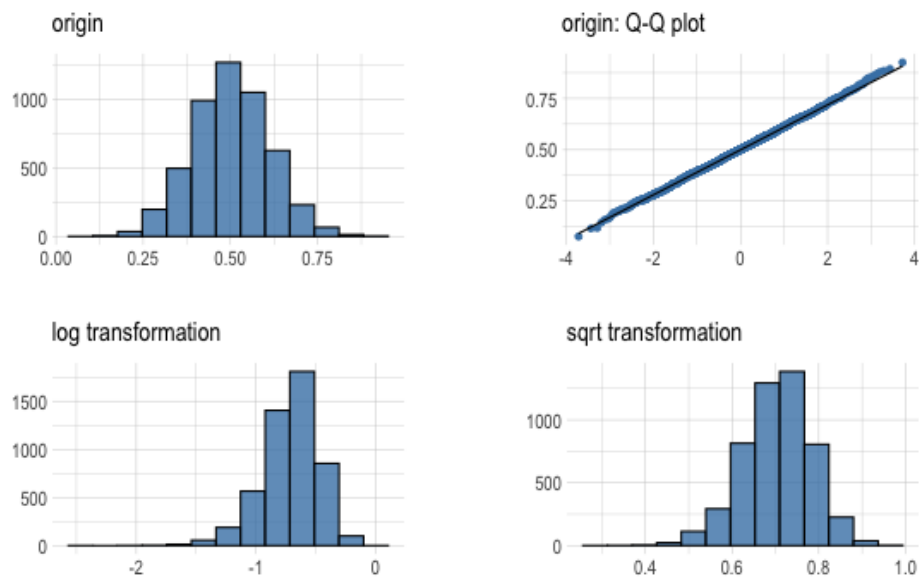
Figure 2.9: uti_card

uti_50plus_pct

* normality test : Shapiro-Wilk normality test
- statistic : 0.99928, p-value : 0.0405855

Table 2.10: skewness and kurtosis : uti_50plus_pct

type	skewness	kurtosis
original	0.0676	2.9642
log transformation	-0.7094	4.1106
sqrt transformation	-0.2912	3.2281
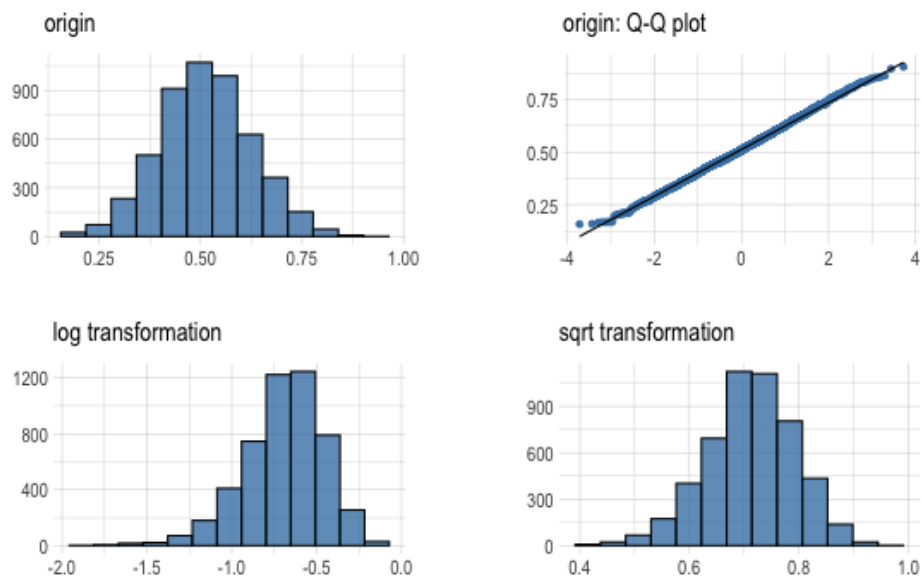
Figure 2.10: uti_50plus_pct

uti_max_credit_line

* normality test : Shapiro-Wilk normality test
- statistic : 0.99932, p-value : 0.0562949

Table 2.11: skewness and kurtosis : uti_max_credit_line

type	skewness	kurtosis
original	0.0382	2.8555
log transformation	-0.6523	3.7194
sqrt transformation	-0.2866	3.0474
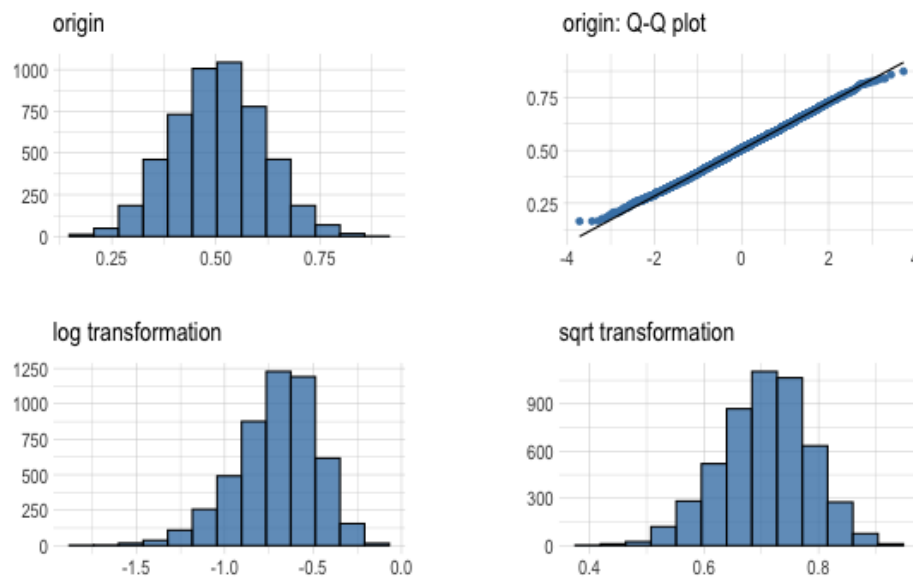
Figure 2.11: uti_max_credit_line

uti_card_50plus_pct

* normality test : Shapiro-Wilk normality test
- statistic : 0.99936, p-value : 0.072726

Table 2.12: skewness and kurtosis : uti_card_50plus_pct

type	skewness	kurtosis
original	-0.0003	3.2029
log transformation	-1.2078	6.8115
sqrt transformation	-0.4967	3.8964
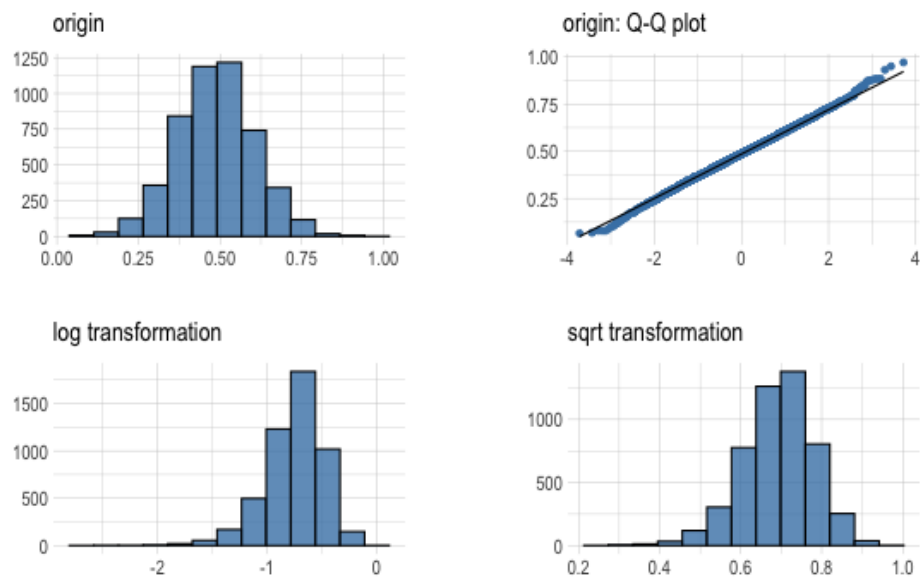
Figure 2.12: uti_card_50plus_pct

rep_income

* normality test : Shapiro-Wilk normality test
- statistic : 0.99914, p-value : 0.013105

Table 2.13: skewness and kurtosis : rep_income

type	skewness	kurtosis
original	-0.0266	3.0332
log transformation	-0.8558	4.6151
sqrt transformation	-0.4035	3.4454
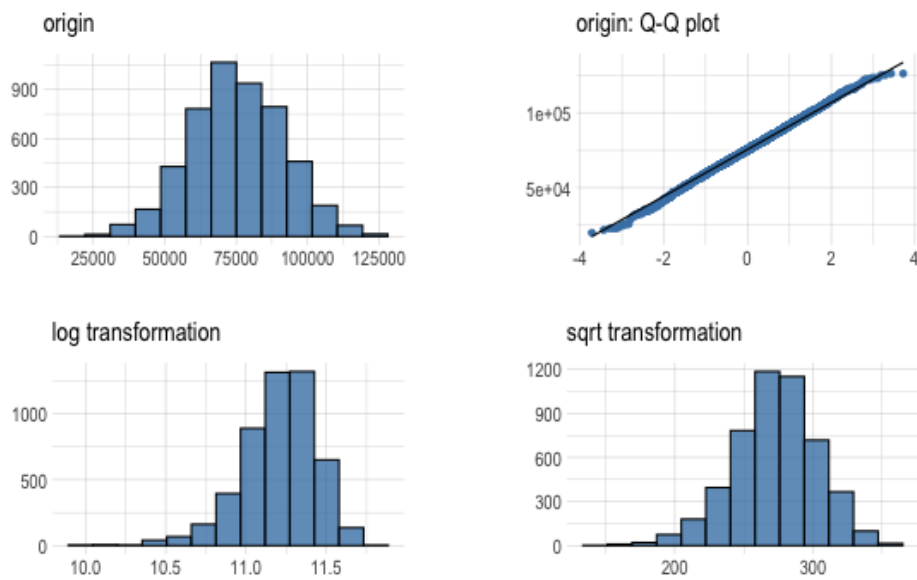
Figure 2.13: rep_income

Default_ind

* normality test : Shapiro-Wilk normality test
- statistic : 0.30025, p-value : 1.38568E-87

Table 2.14: skewness and kurtosis : Default_ind

type	skewness	kurtosis
original	3.0963	10.587
log+1 transformation	3.0963	10.587
sqrt transformation	3.0963	10.587
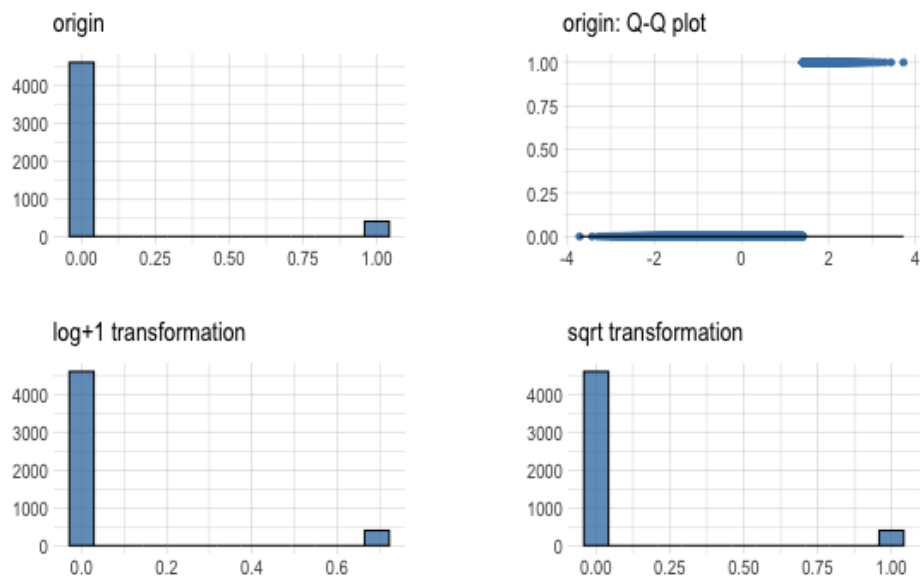
Figure 2.14: Default_ind

Chapter 3

Relationship Between Variables

3.1 Correlation Coefficient

3.1.1 Correlation Coefficient by Variable Combination

Table 3.1: The correlation coefficients (0.5 or more)

Variable1	Variable2	Correlation Coefficient
card_age	credit_age	0.937
card_inq_24_month_num	inq_12_month_num	0.859
uti_card_50plus_pct	uti_card	0.847
credit_good_age	credit_age	0.787
uti_50plus_pct	uti_card	0.748
uti_max_credit_line	uti_card	0.746
card_age	credit_good_age	0.736
uti_card_50plus_pct	uti_50plus_pct	0.635
uti_card_50plus_pct	uti_max_credit_line	0.634
uti_max_credit_line	uti_50plus_pct	0.555

3.1.2 Correlation Plot of Numerical Variables
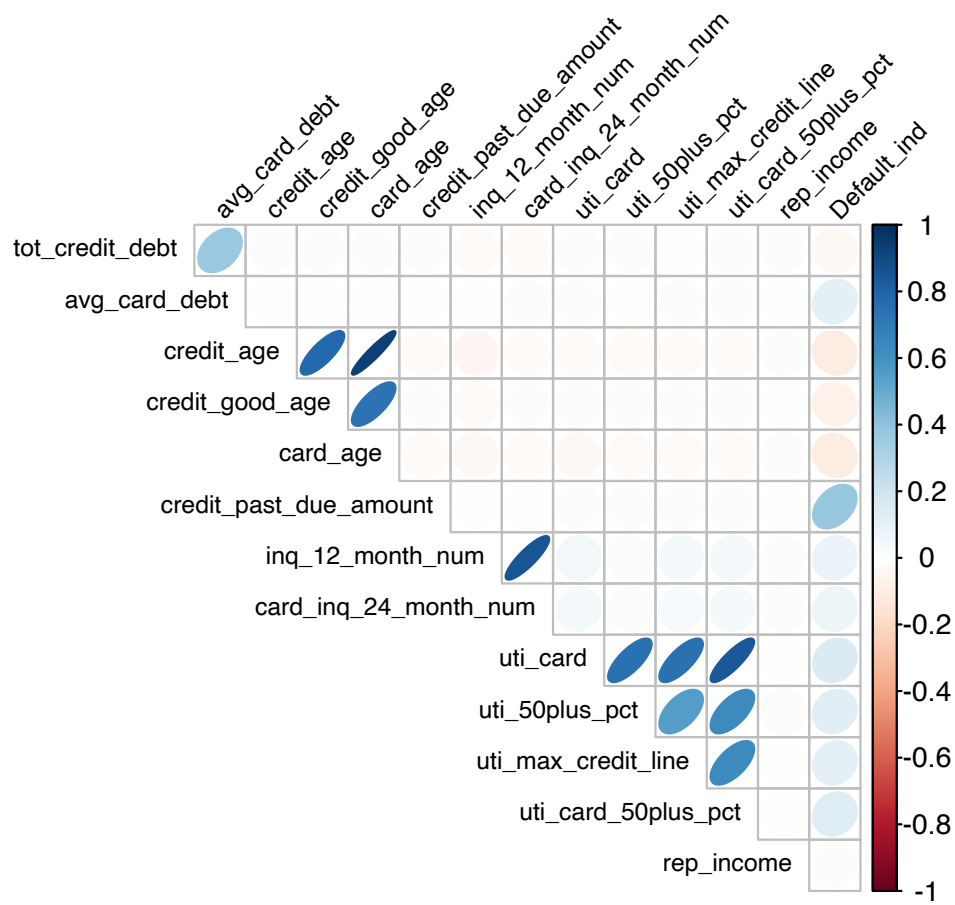
Figure 3.1: The correlation coefficient of numerical variables

Chapter 4

Target based Analysis

4.1 Grouped Descriptive Statistics

4.1.1 Grouped Numerical Variables

tot_credit_debt

Table 4.1: tot_credit_debt

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	91,915.57	94,791.79
sd	29,920.36	22,901.39
se(mean)	751.30	168.77
IQR	42,384.09	30,896.17
skewness	0.10	0.00
kurtosis	-0.36	0.00
0%	6,898.50	2,367.43
1%	28,207.73	41,242.98
5%	43,753.43	56,900.48
10%	52,634.15	65,620.24
20%	65,514.08	75,536.63
25%	70,589.52	79,292.61
30%	74,912.33	82,816.56
40%	83,236.90	88,955.41
50%	91,235.59	94,912.08
60%	98,518.52	100,622.41
70%	106,885.74	106,667.40
75%	112,973.61	110,188.78
80%	119,288.16	114,062.32
90%	131,825.55	124,160.09
95%	140,839.70	132,503.44
99%	160,888.36	148,647.02
100%	188,890.96	182,858.99
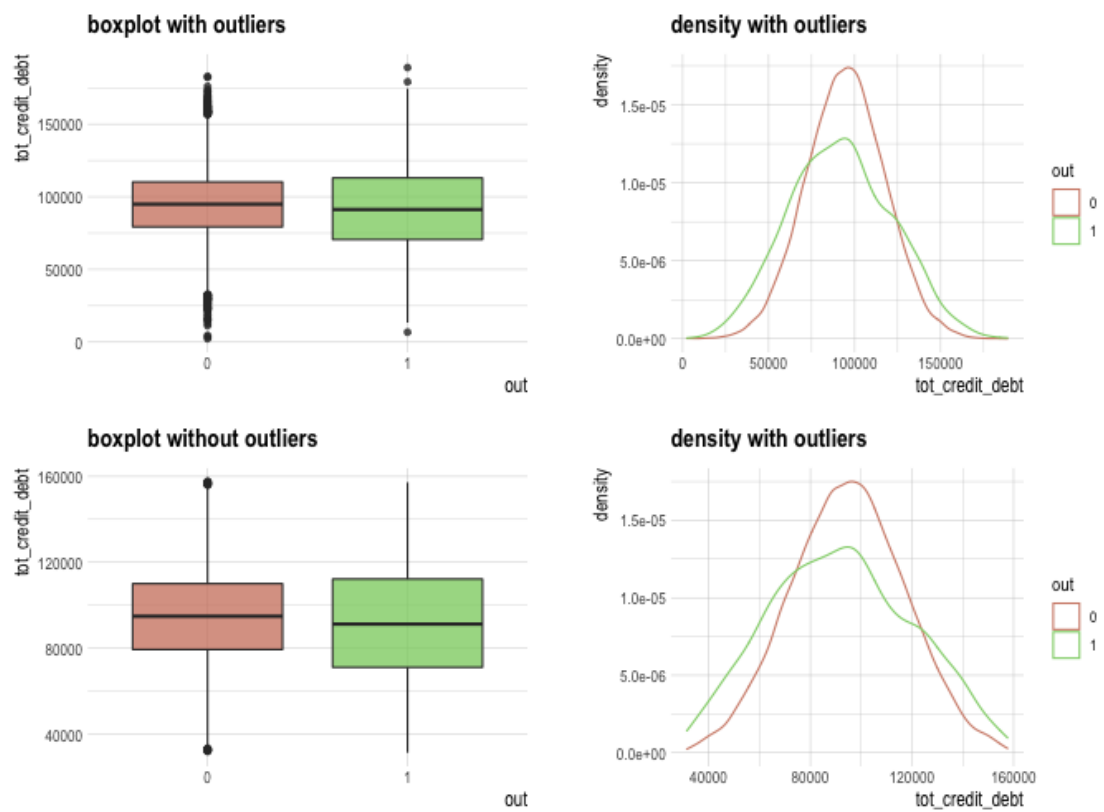
Figure 4.1: tot_credit_debt

avg_card_debt

Table 4.2: avg_card_debt

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	17,880.43	13,761.61
sd	21,195.06	7,363.26
se(mean)	532.21	54.26
IQR	6,492.14	3,736.47
skewness	3.49	10.09
kurtosis	10.71	115.62
0%	2,363.12	4,595.02
1%	4,567.20	7,097.11
5%	6,100.97	8,737.49
10%	7,451.09	9,741.53
20%	9,005.12	10,953.43
25%	9,864.23	11,401.83
30%	10,563.44	11,813.67
40%	11,805.47	12,563.48
50%	13,125.53	13,251.24
60%	14,343.11	13,948.46
70%	15,648.05	14,714.47
75%	16,356.37	15,138.30
80%	16,973.85	15,607.82
90%	18,979.66	16,782.99
95%	99,999.00	17,775.06
99%	99,999.00	19,546.38
100%	99,999.00	99,999.00
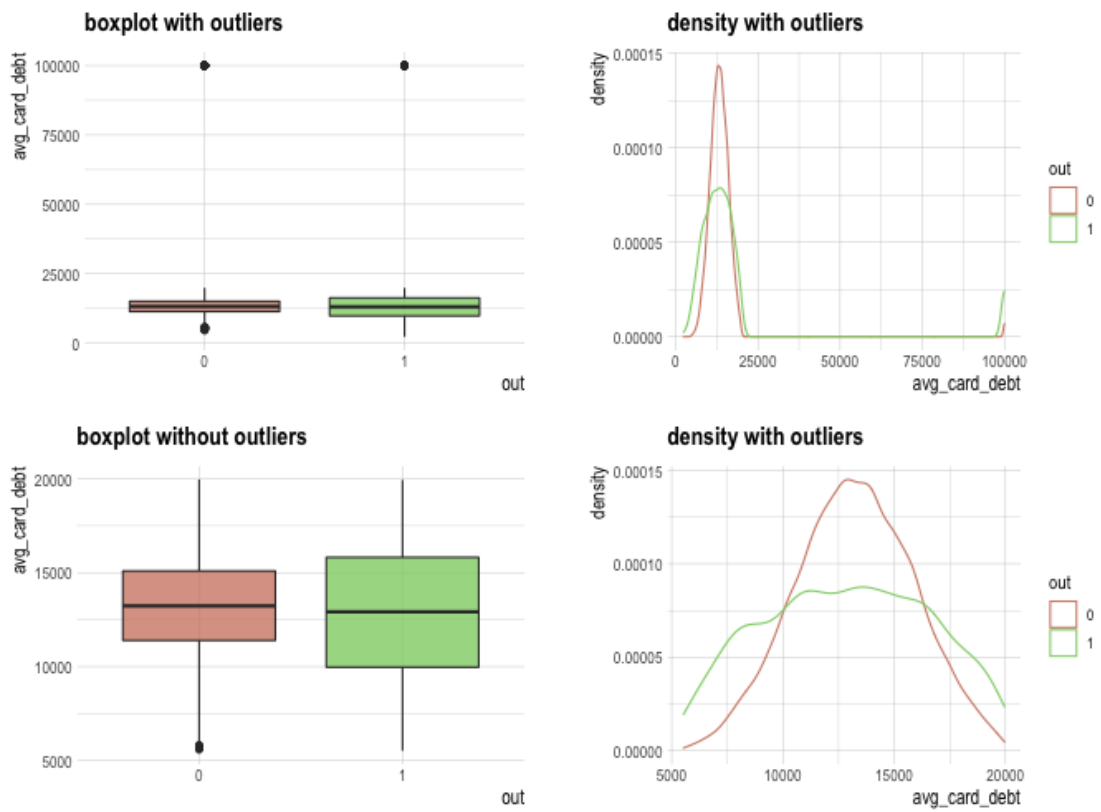
Figure 4.2: avg_card_debt

credit_age

Table 4.3: credit_age

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	275.16	298.55
sd	61.99	61.34
se(mean)	1.56	0.45
IQR	83.00	83.00
skewness	0.14	0.04
kurtosis	0.16	0.00
0%	54.00	78.00
1%	136.85	157.00
5%	177.00	198.00
10%	200.00	219.00
20%	222.00	247.00
25%	233.00	257.00
30%	241.00	266.00
40%	258.00	283.00
50%	274.00	299.00
60%	289.00	314.00
70%	305.00	331.00
75%	316.00	340.00
80%	326.00	350.00
90%	357.00	377.00
95%	381.75	400.00
99%	417.15	443.87
100%	521.00	545.00
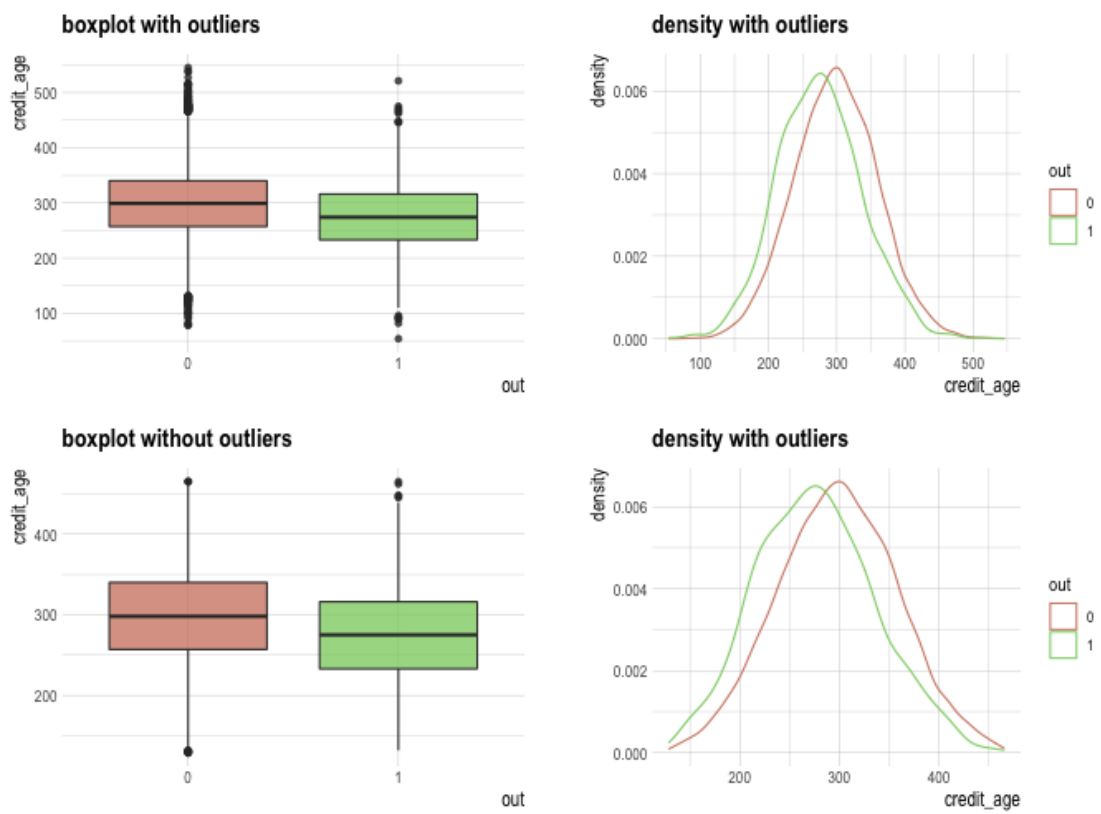
Figure 4.3: credit_age

credit_good_age

Table 4.4: credit_good_age

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	140.54	150.57
sd	33.56	33.94
se(mean)	0.84	0.25
IQR	45.00	45.00
skewness	-0.04	0.02
kurtosis	0.34	0.04
0%	21.00	27.00
1%	52.85	72.00
5%	87.00	95.00
10%	99.00	107.00
20%	113.00	122.00
25%	118.00	128.00
30%	123.00	133.00
40%	132.00	142.00
50%	140.00	151.00
60%	148.00	159.00
70%	158.00	168.00
75%	163.00	173.00
80%	169.00	179.00
90%	184.00	194.00
95%	195.00	206.00
99%	220.00	230.00
100%	274.00	296.00
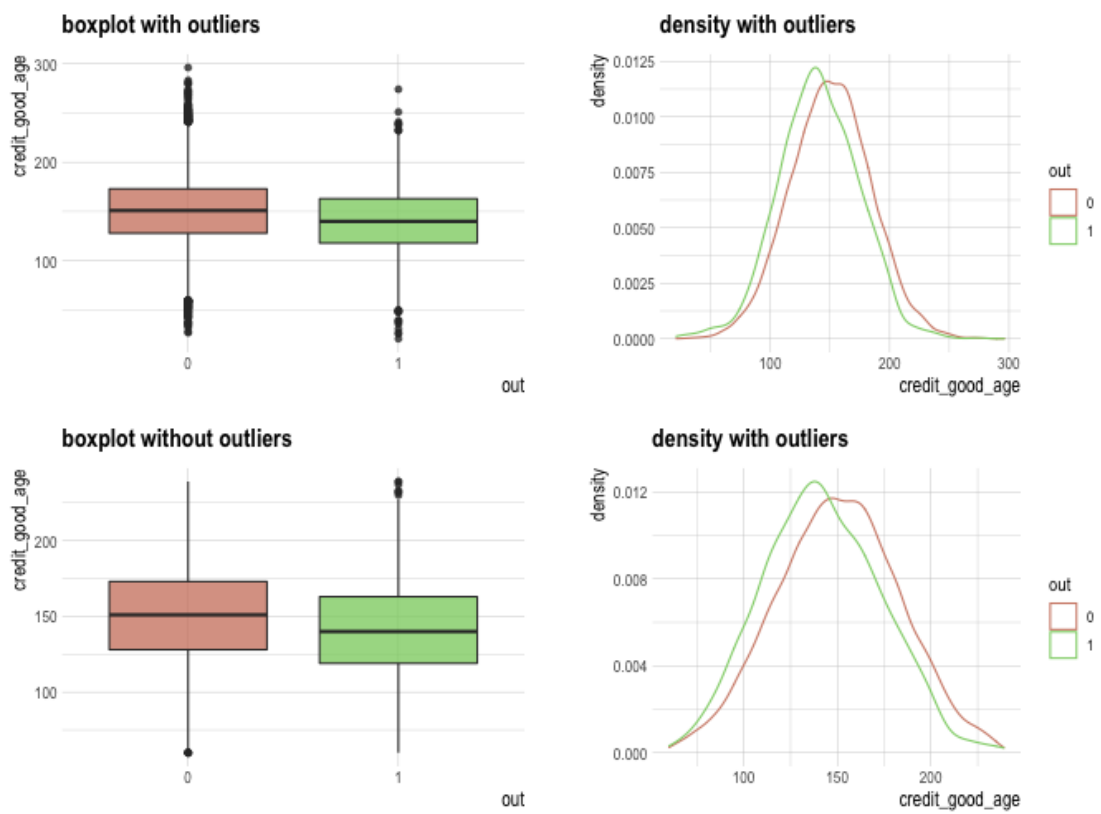
Figure 4.4: credit_good_age

card_age

Table 4.5: card_age

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	247.71	269.76
sd	58.60	59.11
se(mean)	1.47	0.44
IQR	82.00	81.00
skewness	0.10	0.03
kurtosis	0.03	-0.03
0%	41.00	56.00
1%	120.70	134.00
5%	155.00	172.00
10%	175.50	194.00
20%	198.00	220.00
25%	206.00	229.00
30%	215.00	239.00
40%	231.00	255.00
50%	247.00	270.00
60%	263.00	285.00
70%	277.00	301.00
75%	288.00	310.00
80%	298.00	320.00
90%	321.00	345.00
95%	346.00	366.00
99%	392.00	409.00
100%	463.00	520.00
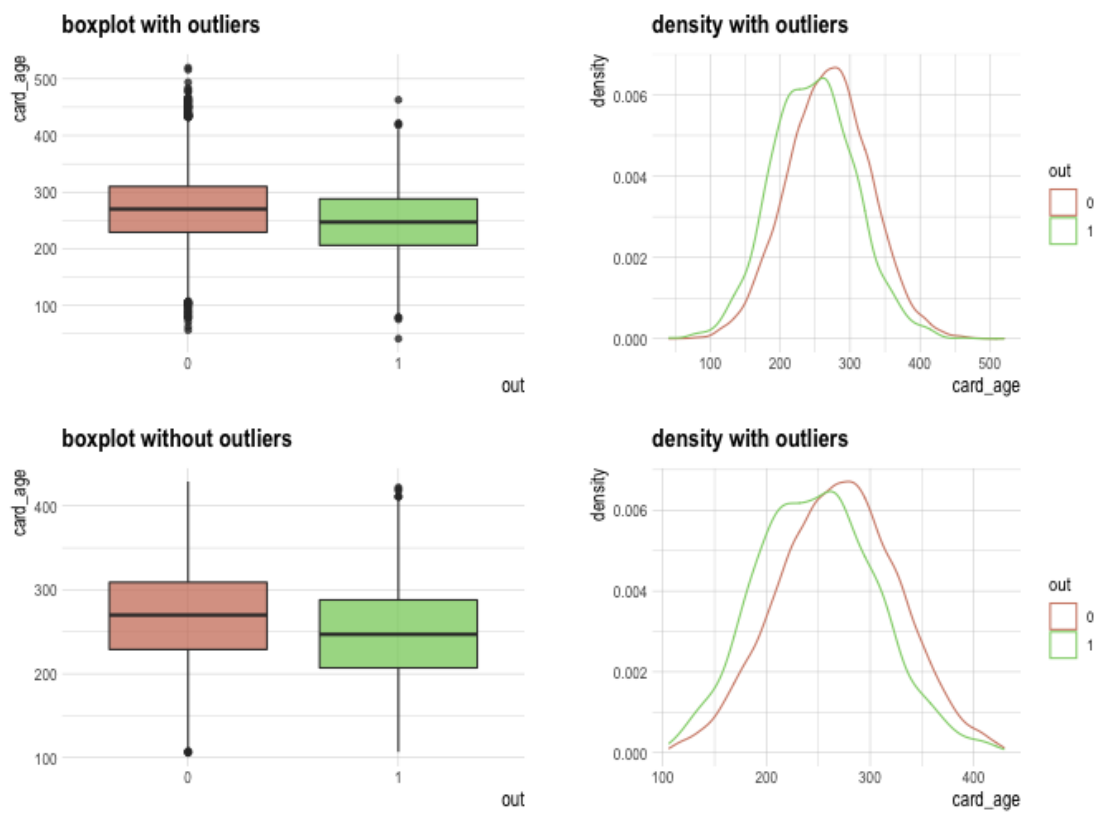
Figure 4.5: card_age

credit_past_due_amount

Table 4.6: credit_past_due_amount

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	3,060.92	94.01
sd	5,777.48	1,048.88
se(mean)	145.07	7.73
IQR	4,905.16	0.00
skewness	1.90	12.69
kurtosis	2.98	177.99
0%	0.00	0.00
1%	0.00	0.00
5%	0.00	0.00
10%	0.00	0.00
20%	0.00	0.00
25%	0.00	0.00
30%	0.00	0.00
40%	0.00	0.00
50%	0.00	0.00
60%	0.00	0.00
70%	0.00	0.00
75%	4,905.16	0.00
80%	7,559.12	0.00
90%	11,918.01	0.00
95%	16,653.25	0.00
99%	22,470.85	0.00
100%	32,662.98	24,211.55
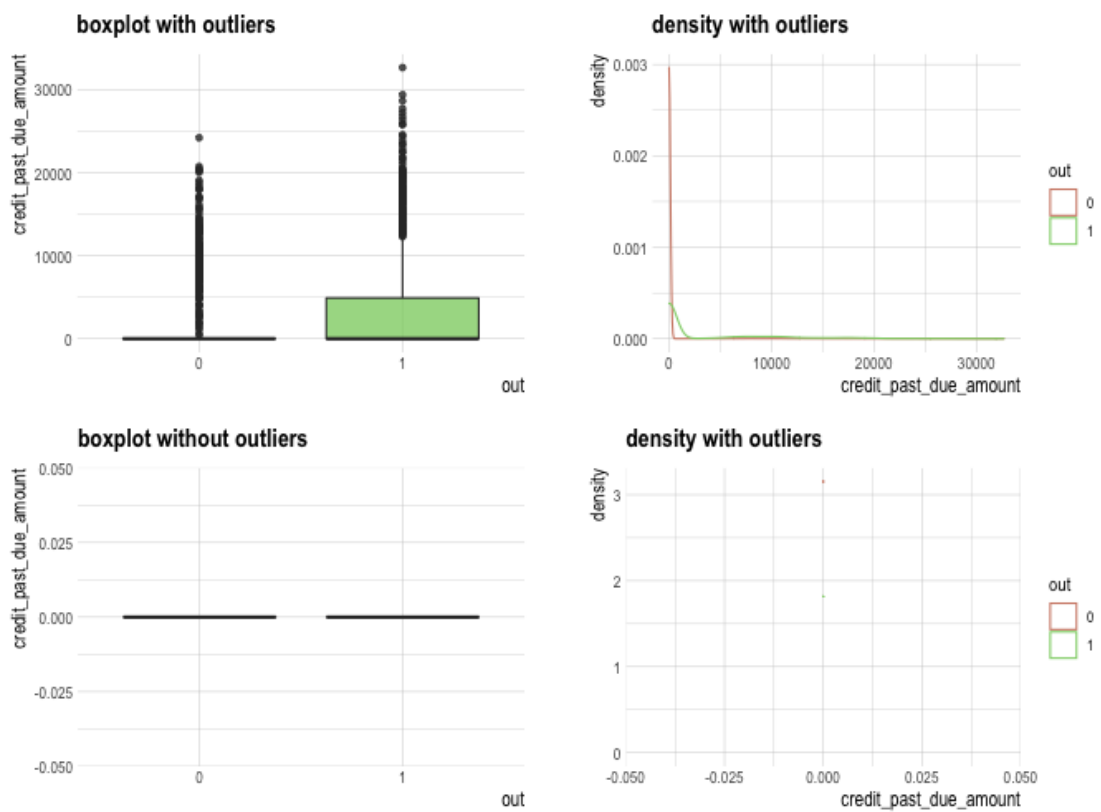
Figure 4.6: credit_past_due_amount

inq_12_month_num

Table 4.7: inq_12_month_num

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	2.29	1.72
sd	1.87	1.72
se(mean)	0.05	0.01
IQR	3.00	3.00
skewness	0.54	0.86
kurtosis	-0.35	0.14
0%	0.00	0.00
1%	0.00	0.00
5%	0.00	0.00
10%	0.00	0.00
20%	0.00	0.00
25%	1.00	0.00
30%	1.00	0.00
40%	2.00	1.00
50%	2.00	1.00
60%	3.00	2.00
70%	3.00	3.00
75%	4.00	3.00
80%	4.00	3.00
90%	5.00	4.00
95%	6.00	5.00
99%	7.00	6.00
100%	9.00	10.00
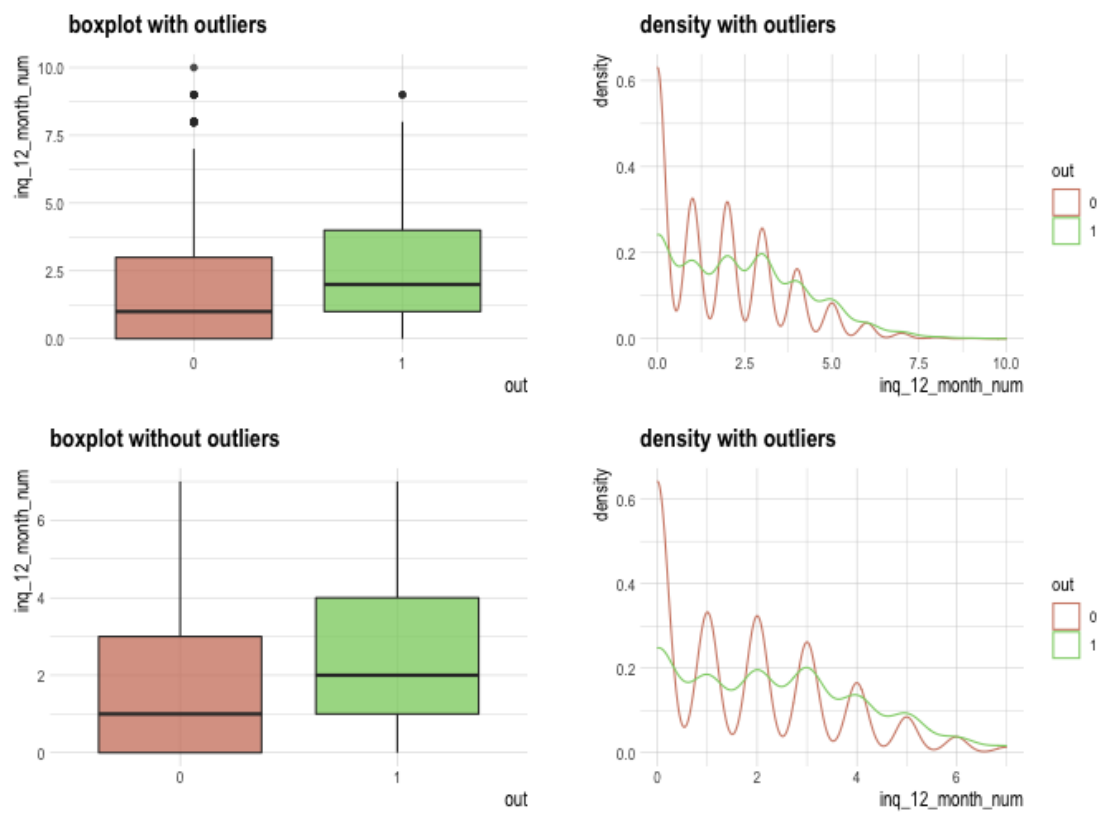
Figure 4.7: inq_12_month_num

card_inq_24_month_num

Table 4.8: card_inq_24_month_num

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	4.16	3.34
sd	3.14	2.90
se(mean)	0.08	0.02
IQR	4.00	4.00
skewness	0.51	0.83
kurtosis	-0.45	0.28
0%	0.00	0.00
1%	0.00	0.00
5%	0.00	0.00
10%	0.00	0.00
20%	1.00	0.00
25%	2.00	1.00
30%	2.00	1.00
40%	3.00	2.00
50%	4.00	3.00
60%	5.00	4.00
70%	6.00	5.00
75%	6.00	5.00
80%	7.00	6.00
90%	9.00	7.00
95%	10.00	9.00
99%	12.00	11.00
100%	15.00	18.00
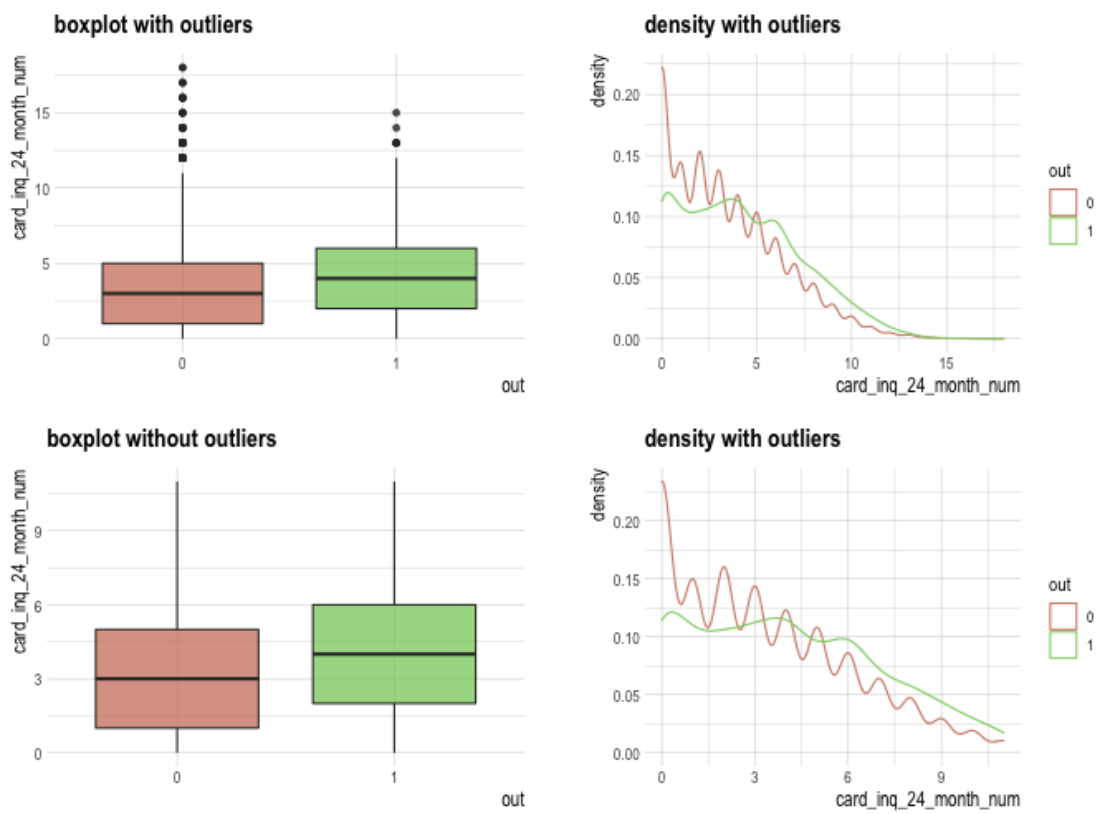
Figure 4.8: card_inq_24_month_num

uti_card

Table 4.9: uti_card

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	0.56	0.50
sd	0.11	0.11
se(mean)	0.00	0.00
IQR	0.15	0.15
skewness	-0.04	-0.02
kurtosis	0.02	0.04
0%	0.21	0.07
1%	0.30	0.25
5%	0.37	0.32
10%	0.42	0.36
20%	0.47	0.41
25%	0.49	0.43
30%	0.50	0.44
40%	0.53	0.47
50%	0.56	0.50
60%	0.59	0.53
70%	0.62	0.55
75%	0.64	0.57
80%	0.65	0.59
90%	0.70	0.64
95%	0.74	0.67
99%	0.82	0.74
100%	0.97	0.92
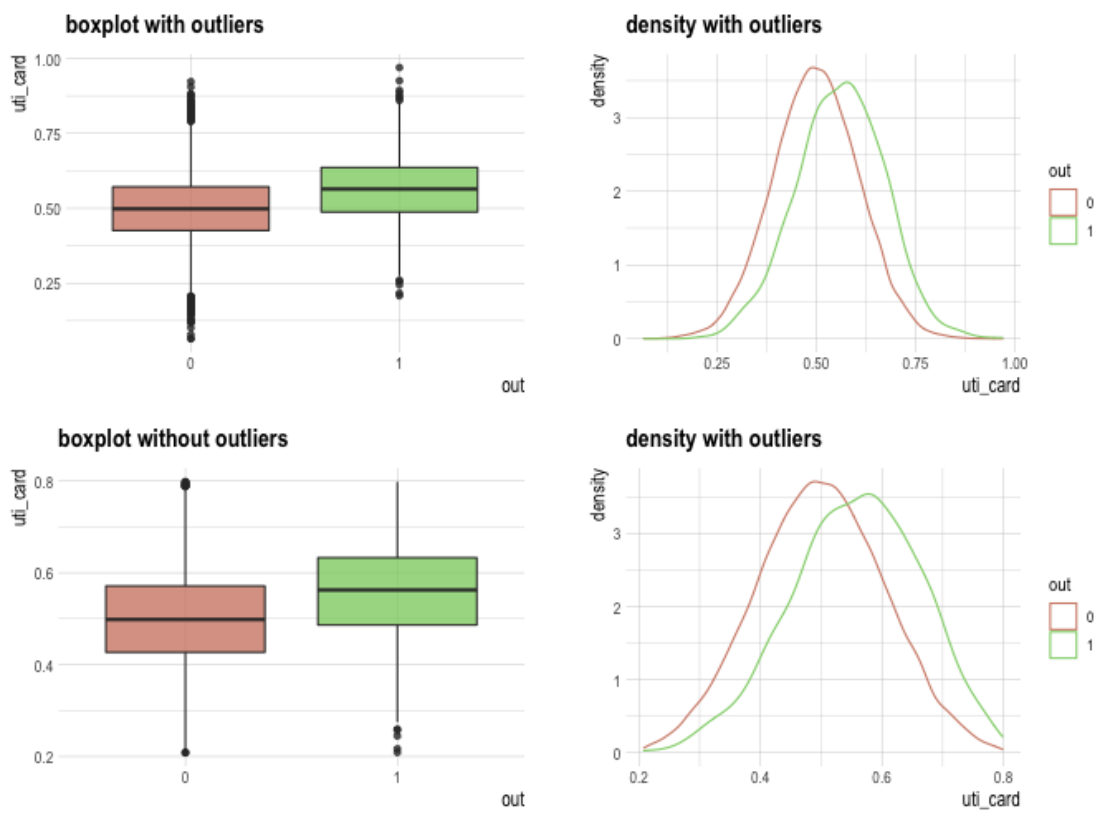
Figure 4.9: uti_card

uti_50plus_pct

Table 4.10: uti_50plus_pct

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	0.56	0.51
sd	0.11	0.11
se(mean)	0.00	0.00
IQR	0.16	0.15
skewness	0.00	0.00
kurtosis	-0.14	0.01
0%	0.16	0.03
1%	0.29	0.25
5%	0.38	0.32
10%	0.41	0.36
20%	0.46	0.41
25%	0.48	0.43
30%	0.49	0.45
40%	0.53	0.48
50%	0.56	0.51
60%	0.59	0.53
70%	0.62	0.57
75%	0.64	0.58
80%	0.66	0.60
90%	0.70	0.65
95%	0.75	0.69
99%	0.82	0.77
100%	0.89	0.99
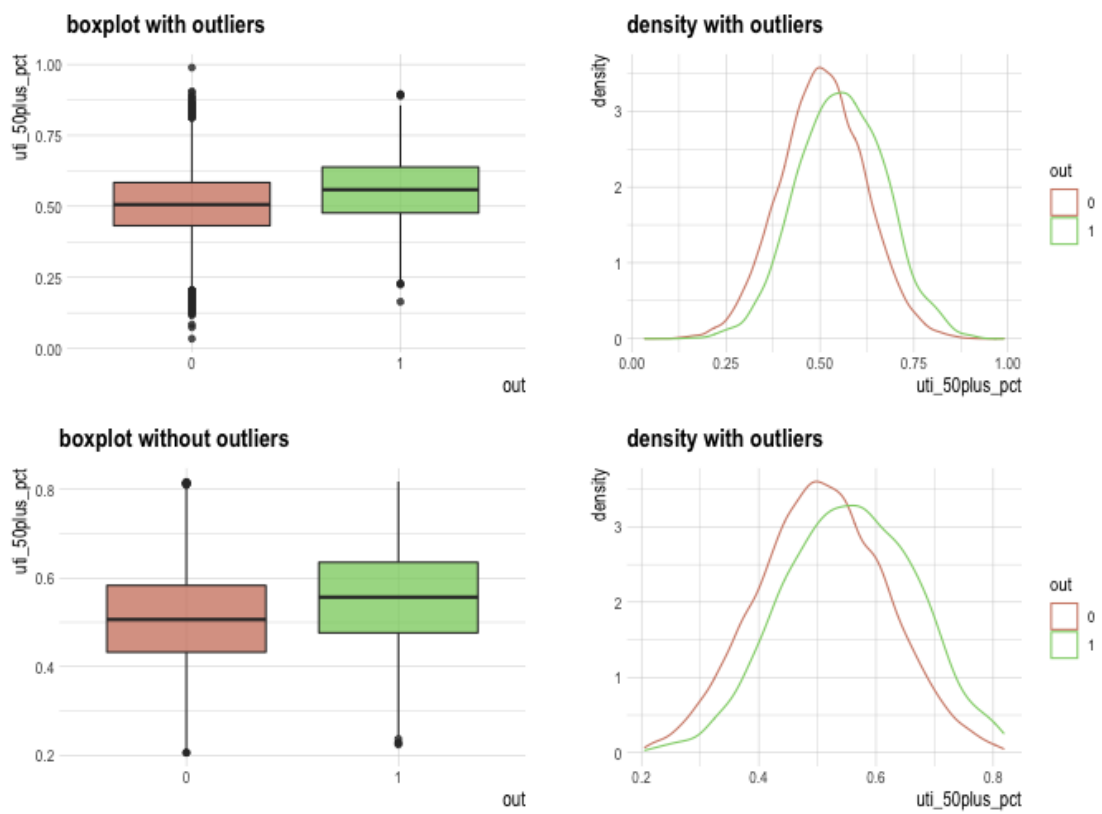
Figure 4.10: uti_50plus_pct

uti_max_credit_line

Table 4.11: uti_max_credit_line

	1	0
n	1,586.00	18,414.00
NA	0.00	0.00
mean	0.55	0.50
sd	0.11	0.11
se(mean)	0.00	0.00
IQR	0.15	0.15
skewness	0.07	-0.01
kurtosis	-0.04	-0.02
0%	0.19	0.01
1%	0.30	0.25
5%	0.38	0.33
10%	0.41	0.37
20%	0.46	0.41
25%	0.48	0.43
30%	0.49	0.45
40%	0.52	0.48
50%	0.55	0.50
60%	0.58	0.53
70%	0.61	0.56
75%	0.62	0.58
80%	0.64	0.59
90%	0.70	0.64
95%	0.73	0.68
99%	0.80	0.75
100%	1.00	0.97
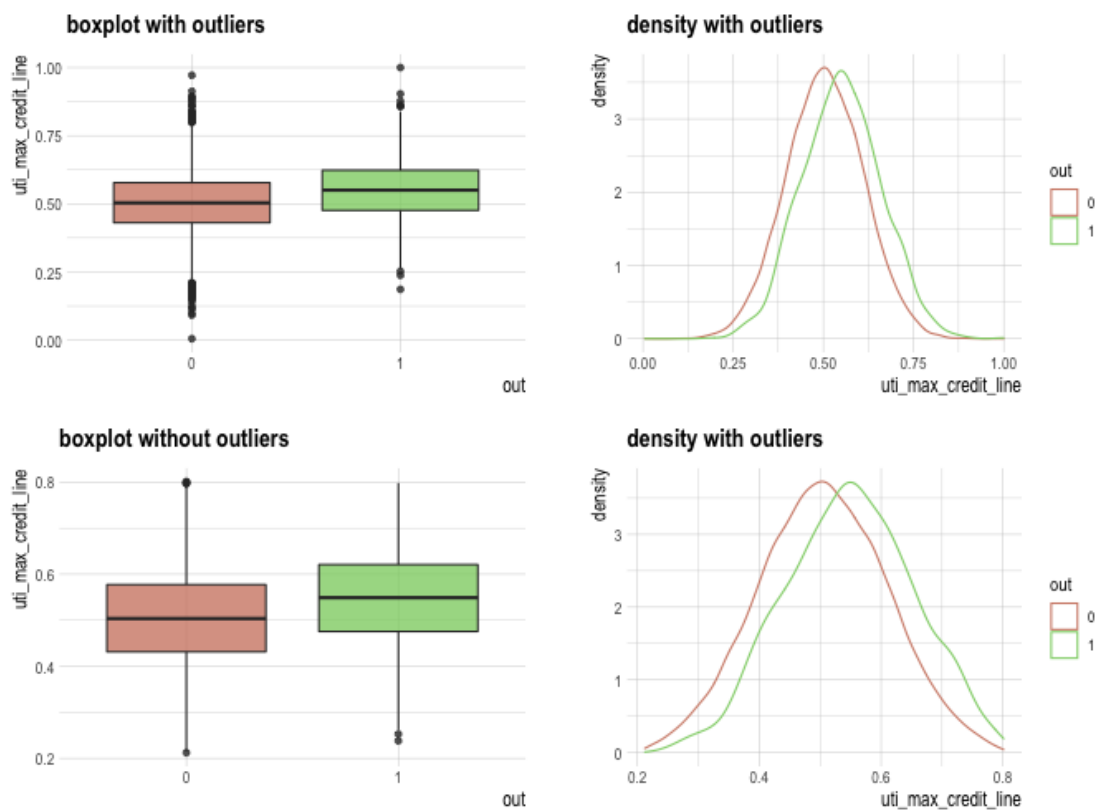
Figure 4.11: uti_max_credit_line

uti_card_50plus_pct

Table 4.12: uti_card_50plus_pct

	1	0
n	1,412.00	16,533.00
NA	174.00	1,881.00
mean	0.54	0.48
sd	0.12	0.12
se(mean)	0.00	0.00
IQR	0.17	0.16
skewness	-0.11	-0.03
kurtosis	0.12	0.09
0%	0.12	0.00
1%	0.25	0.20
5%	0.34	0.29
10%	0.39	0.34
20%	0.44	0.39
25%	0.46	0.41
30%	0.48	0.42
40%	0.52	0.46
50%	0.55	0.49
60%	0.58	0.52
70%	0.61	0.55
75%	0.63	0.56
80%	0.65	0.58
90%	0.70	0.64
95%	0.74	0.68
99%	0.84	0.76
100%	0.97	0.95
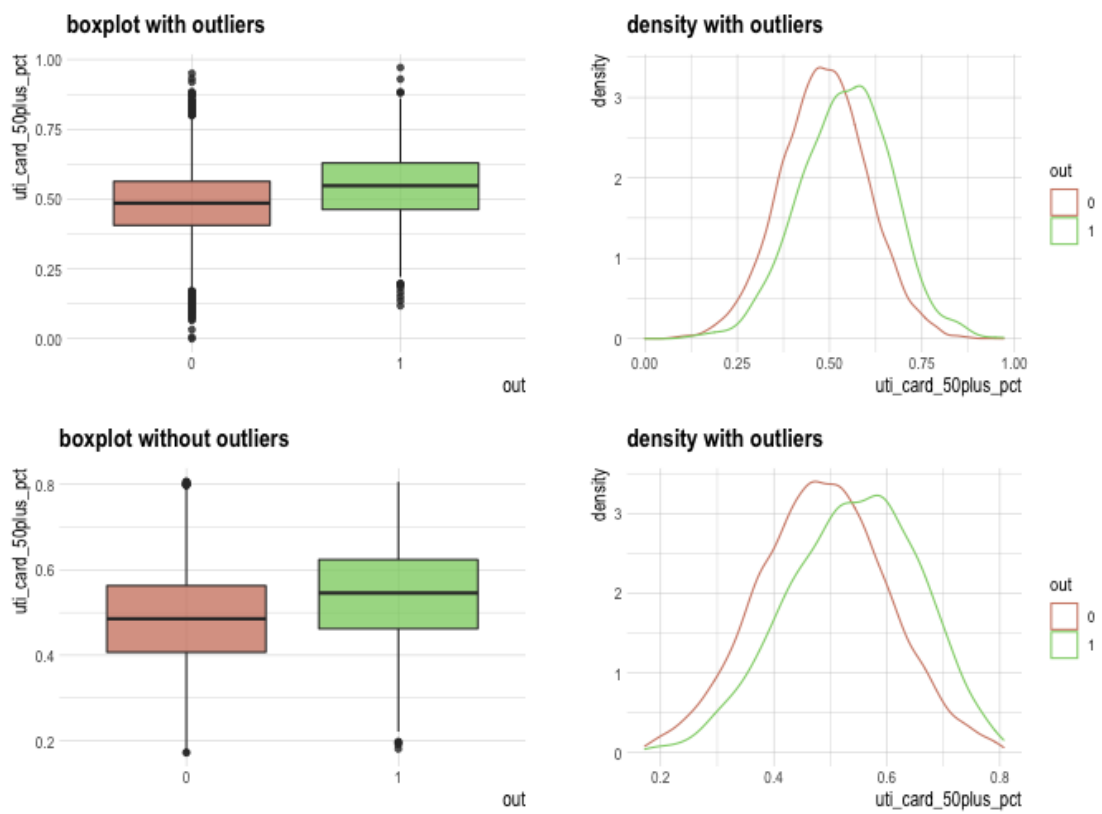
Figure 4.12: uti_card_50plus_pct

rep_income

Table 4.13: rep_income

	1	0
n	1,457.00	16,973.00
NA	129.00	1,441.00
mean	74,522.99	75,583.34
sd	16,775.13	16,323.78
se(mean)	439.48	125.30
IQR	22,000.00	22,000.00
skewness	0.11	0.00
kurtosis	-0.09	-0.01
0%	26,000.00	12,000.00
1%	37,000.00	37,720.00
5%	48,000.00	49,000.00
10%	53,000.00	55,000.00
20%	60,000.00	62,000.00
25%	63,000.00	65,000.00
30%	65,000.00	67,000.00
40%	70,000.00	71,000.00
50%	74,000.00	76,000.00
60%	78,000.00	80,000.00
70%	83,000.00	84,000.00
75%	85,000.00	87,000.00
80%	88,000.00	89,000.00
90%	96,000.00	97,000.00
95%	103,000.00	102,000.00
99%	116,000.00	114,000.00
100%	123,000.00	150,000.00
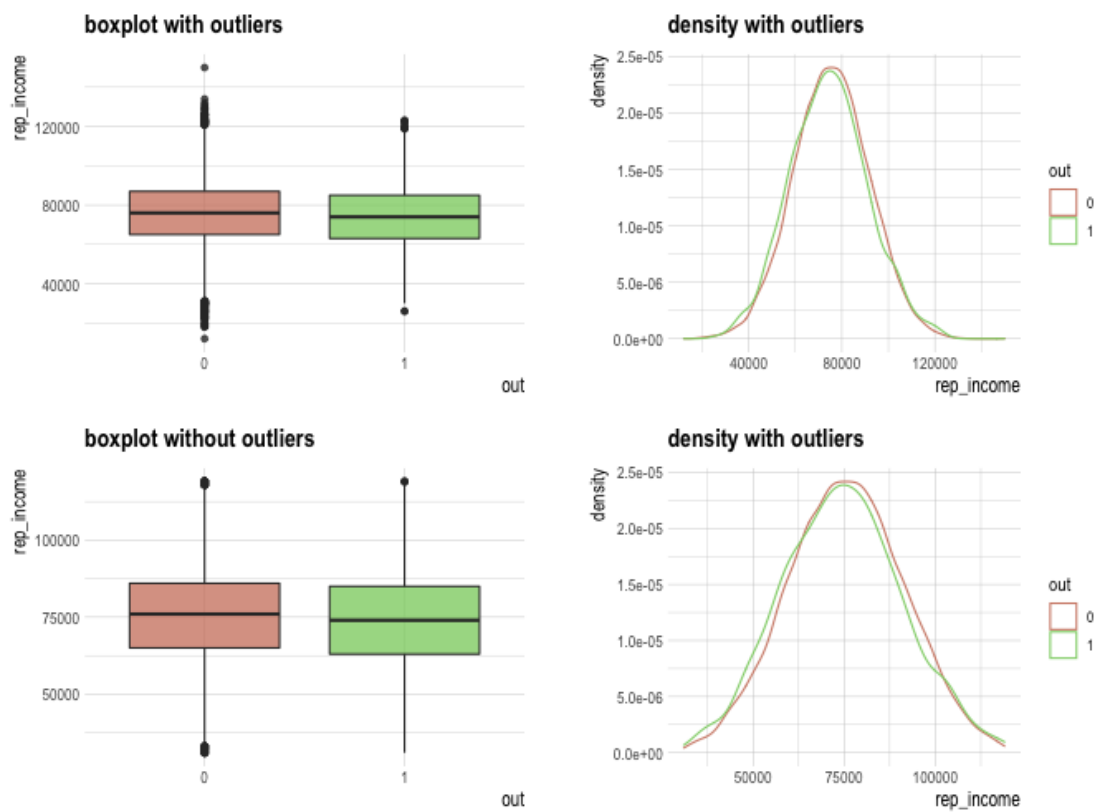
Figure 4.13: rep_income

Default_ind

Table 4.14: Default_ind

	1	0
n	1,586	18,414
NA	0	0
mean	1	0
sd	0	0
se(mean)	0	0
IQR	0	0
skewness	NaN	NaN
kurtosis	NaN	NaN
0%	1	0
1%	1	0
5%	1	0
10%	1	0
20%	1	0
25%	1	0
30%	1	0
40%	1	0
50%	1	0
60%	1	0
70%	1	0
75%	1	0
80%	1	0
90%	1	0
95%	1	0
99%	1	0
100%	1	0
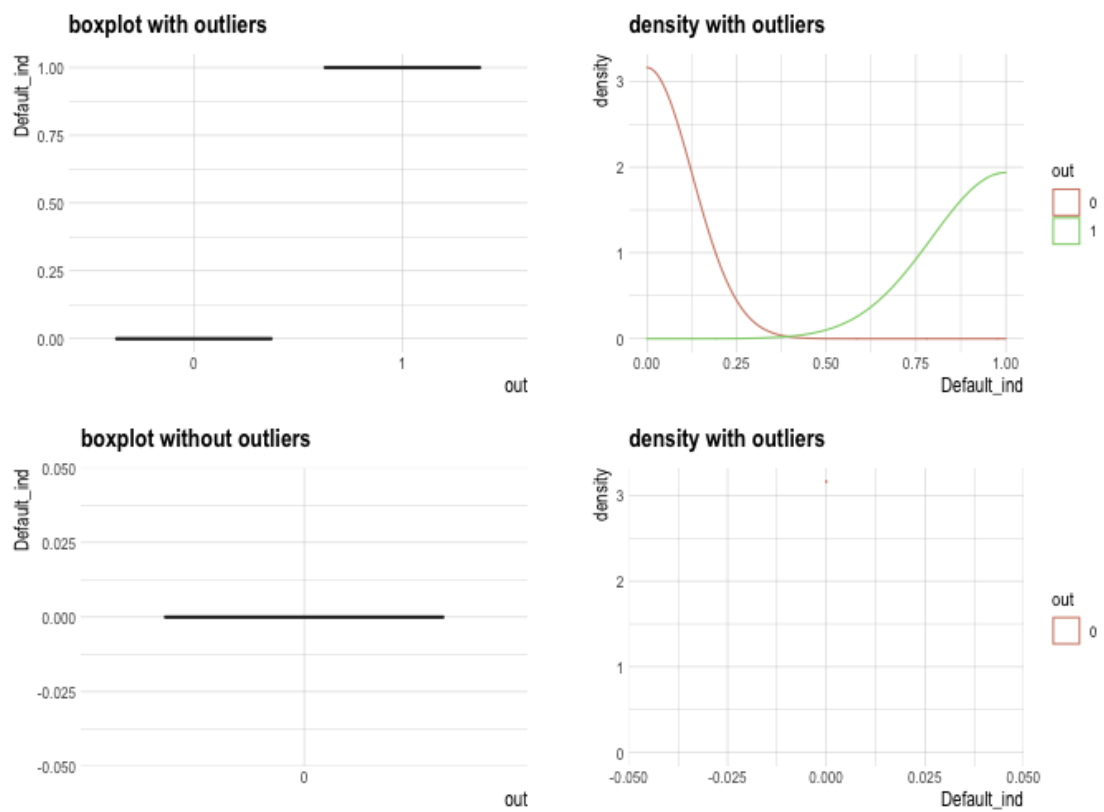
Figure 4.14: Default_ind

	0	1	Sum
AL	2,664	229	2,893
FL	2,638	219	2,857
GA	2,648	209	2,857
LA	2,621	228	2,849
MS	2,586	241	2,827
NC	2,674	224	2,898
SC	2,583	236	2,819
Sum	**18,414**	**1,586**	**20,000**

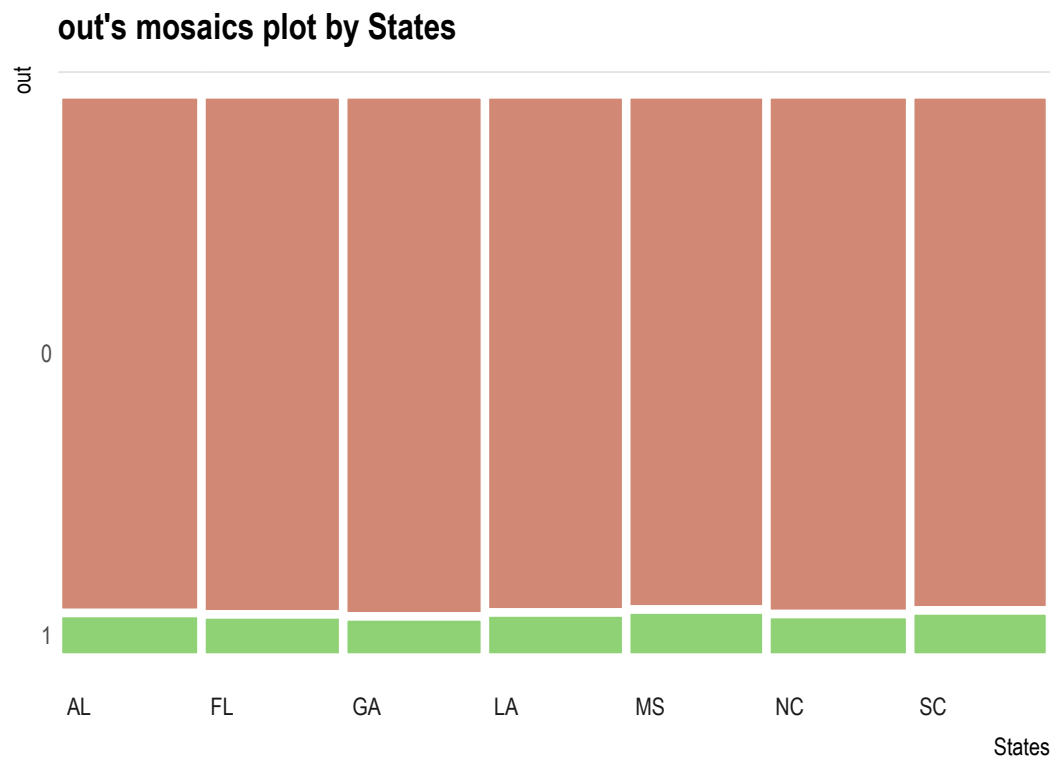|	0	1	Sum
AL	14.47	14.44	14.46
FL	14.33	13.81	14.29
GA	14.38	13.18	14.29
LA	14.23	14.38	14.24
MS	14.04	15.20	14.13
NC	14.52	14.12	14.49
SC	14.03	14.88	14.09
Sum	**100.00**	**100.00**	**100.00**

4.1.2 Grouped Categorical Variables

States

Figure 4.15: States

4.2 Grouped Relationship Between Variables

4.2.1 Grouped Correlation Coefficient

Table 4.15: The correlation coefficients (0.5 or more)

out	Variable1	Variable2	Correlation Coefficient
0	card_age	credit_age	0.936
0	card_inq_24_month_num	inq_12_month_num	0.858
0	uti_card_50plus_pct	uti_card	0.844
0	credit_good_age	credit_age	0.785
0	uti_50plus_pct	uti_card	0.743
0	uti_max_credit_line	uti_card	0.740
0	card_age	credit_good_age	0.734
0	uti_card_50plus_pct	uti_50plus_pct	0.629
0	uti_card_50plus_pct	uti_max_credit_line	0.627
0	uti_max_credit_line	uti_50plus_pct	0.545
1	card_age	credit_age	0.938
1	card_inq_24_month_num	inq_12_month_num	0.866
1	uti_card_50plus_pct	uti_card	0.848
1	credit_good_age	credit_age	0.793
1	uti_max_credit_line	uti_card	0.765
1	uti_50plus_pct	uti_card	0.751
1	card_age	credit_good_age	0.739
1	uti_card_50plus_pct	uti_max_credit_line	0.644
1	uti_card_50plus_pct	uti_50plus_pct	0.630
1	uti_max_credit_line	uti_50plus_pct	0.589
1	avg_card_debt	tot_credit_debt	0.515

4.2.2 Grouped Correlation Plot of Numerical Variables

- Grouped Correlation Case of (out == 0)
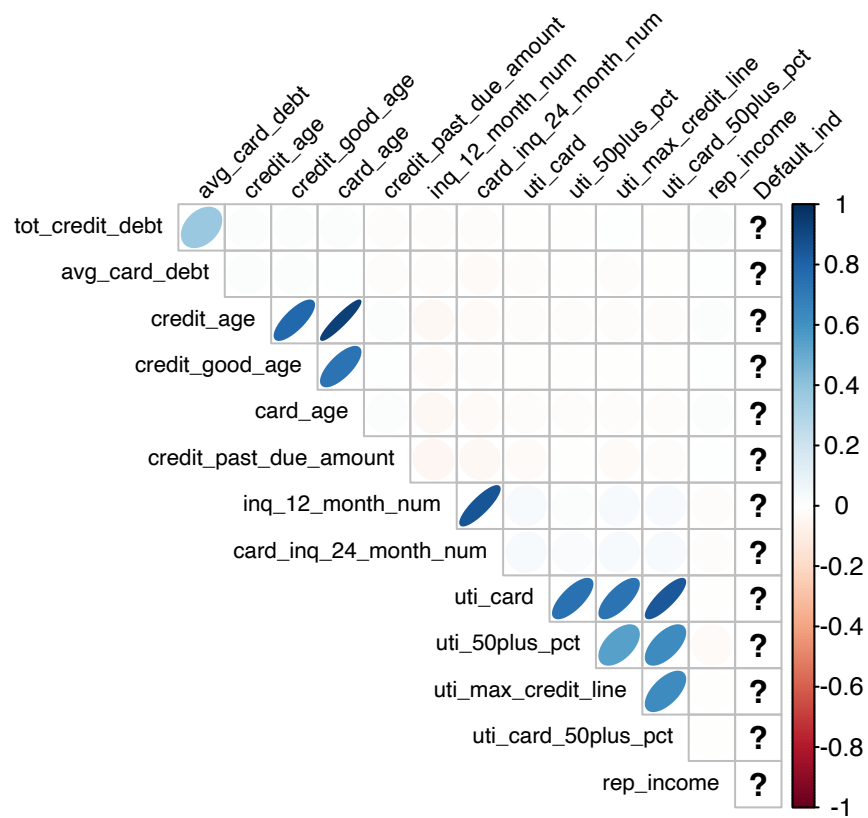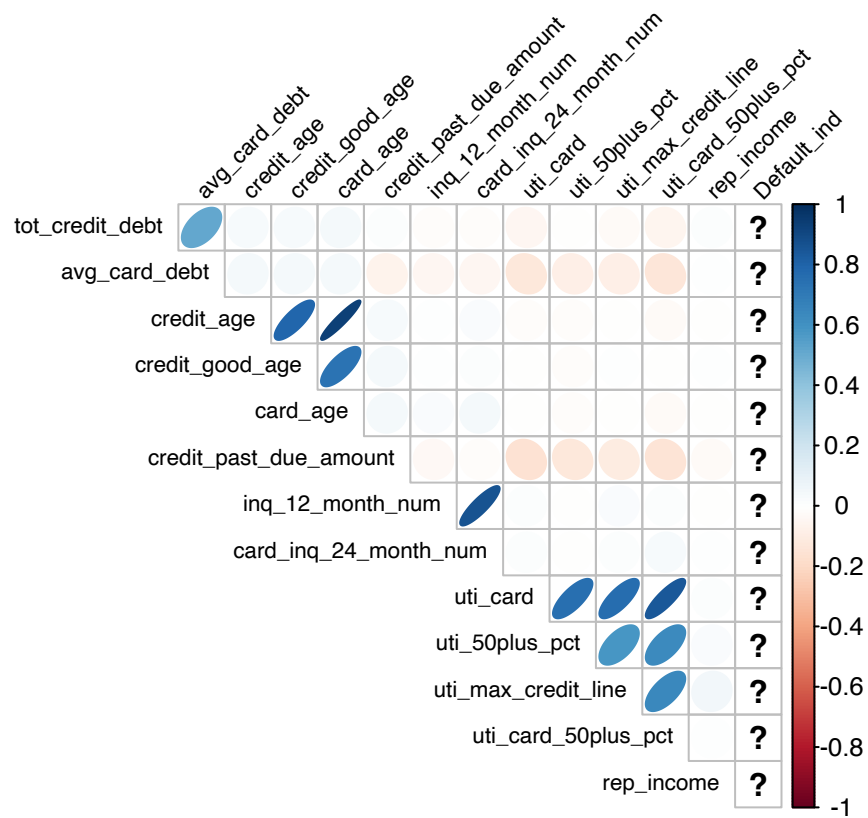- Grouped Correlation Case of (out == 1)

Figure 4.16: Correlation Matrix Plot (out == 0)

Figure 4.17: Correlation Matrix Plot (out == 1)