Exploratory Data Analysis Report

Report by dlookr package

2019-11-09

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 **3,527 observations** and **33 variables**.

1.2 Information of Variables

The variable information of the data set that generated the EDA Report is shown in the following table.:

Information of Variables

variables	types	missing_count	missing_percent	unique_count	unique_rate
Age	integer	0	0	4	0.0011341
Gender	integer	0	0	2	0.0005671

variables	types	missing_count	missing_percent	unique_count	unique_rate
Marital_status	integer	0	0	2	0.0005671
Num_comp_worked	integer	0	0	4	0.0011341
Work_Experience	integer	0	0	4	0.0011341
Last_salary	integer	0	0	5	0.0014176
Salary_hike	integer	0	0	5	0.0014176
Work_hours	integer	0	0	3	0.0008506
Traveling_required	integer	0	0	2	0.0005671
NoticePeriod_buyout	integer	0	0	2	0.0005671
Job_type	integer	0	0	2	0.0005671
Whether_comp_Possess_vehicle	integer	0	0	2	0.0005671
Work_location	integer	0	0	3	0.0008506
Traveltowork_distance_._time	integer	0	0	3	0.0008506
Work_area_living_cost	integer	0	0	3	0.0008506
Public_transport_nearby	integer	0	0	2	0.0005671
Current_employment	integer	0	0	2	0.0005671
Offer_letter_Processtime.week.	integer	0	0	5	0.0014176
Bonus	integer	0	0	2	0.0005671
Allowance	integer	0	0	2	0.0005671
Overtime_pay	integer	0	0	2	0.0005671
Emp_shares_scheme	integer	0	0	2	0.0005671
Performance_incentive	integer	0	0	2	0.0005671
Annual_leave	integer	0	0	2	0.0005671
Medical_Family	integer	0	0	2	0.0005671
Child_day_care	integer	0	0	2	0.0005671
Nursing_room	integer	0	0	2	0.0005671
Parking	integer	0	0	2	0.0005671
Comp_size	integer	0	0	5	0.0014176
Comp_rating	integer	0	0	5	0.0014176
Work_life	integer	0	0	5	0.0014176
Stresslevel	integer	0	0	4	0.0011341
Accept_offer	integer	0	0	2	0.0005671

The target variable of the data is **'NULL'**, and the data type of the variable is **NULL(You did not specify a target variable)**.

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.

2 Univariate Analysis

2.1 Descriptive Statistics

edaData

33 Variables   3527 Observations

Age

n	missing	distinct	Info	Mean	Gmd
3527	0	4	0.856	0.8035	0.8049

```
Value         0     1     2     3
Frequency  1384  1515   565    63
Proportion 0.392 0.430 0.160 0.018
```

Gender

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.745	1902	0.5393	0.4971

Marital_status

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.739	1549	0.4392	0.4927

Num_comp_worked

n	missing	distinct	Info	Mean	Gmd
3527	0	4	0.733	0.5574	0.804

```
Value         0     1     2     3
Frequency  2225   872   196   234
Proportion 0.631 0.247 0.056 0.066
```

Work_Experience

n	missing	distinct	Info	Mean	Gmd
3527	0	4	0.88	1.297	0.9593

```
Value         0     1     2     3
Frequency   632  1623   866   406
```

```
Proportion 0.179 0.460 0.246 0.115
```

Last_salary

n	missing	distinct	Info	Mean	Gmd
3527	0	5	0.821	1.383	0.9841

```
Value         0     1     2     3     4
Frequency   438  1951   683   260   195
Proportion 0.124 0.553 0.194 0.074 0.055
```

Salary_hike

n	missing	distinct	Info	Mean	Gmd
3527	0	5	0.867	1.246	0.9917

```
Value         0     1     2     3     4
Frequency   704  1718   745   254   106
Proportion 0.200 0.487 0.211 0.072 0.030
```

Work_hours

n	missing	distinct	Info	Mean	Gmd
3527	0	3	0.771	0.5415	0.7112

```
Value         0     1     2
Frequency  2077   990   460
Proportion 0.589 0.281 0.130
```

Traveling_required

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.75	1725	0.4891	0.4999

NoticePeriod_buyout

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.691	1268	0.3595	0.4607

Job_type

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.704	2198	0.6232	0.4698

Whether_comp_Possess_vehicle

n	missing	distinct	Info	Sum	Mean	Gmd

| 3527 | 0 | 2 | 0.384 | 2995 | 0.8492 | 0.2562 |

Work_location

n	missing	distinct	Info	Mean	Gmd
3527	0	3	0.824	1.053	0.7163

```
Value          0     1     2
Frequency    735  1870   922
Proportion 0.208 0.530 0.261
```

Traveltowork_distance_._time

n	missing	distinct	Info	Mean	Gmd
3527	0	3	0.856	0.7757	0.7757

```
Value          0     1     2
Frequency   1423  1472   632
Proportion 0.403 0.417 0.179
```

Work_area_living_cost

n	missing	distinct	Info	Mean	Gmd
3527	0	3	0.855	1.201	0.7683

```
Value          0     1     2
Frequency    637  1543  1347
Proportion 0.181 0.437 0.382
```

Public_transport_nearby

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.746	1642	0.4656	0.4978

Current_employment

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.421	2932	0.8313	0.2806

Offer_letter_Processtime.week.

n	missing	distinct	Info	Mean	Gmd
3527	0	5	0.784	2.195	0.9004

```
Value          0     1     2     3     4
Frequency     89   453  2097   458   430
Proportion 0.025 0.128 0.595 0.130 0.122
```

Bonus

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.748	1848	0.524	0.499

Allowance

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.479	704	0.1996	0.3196

Overtime_pay

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.579	922	0.2614	0.3863

Emp_shares_scheme

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.327	439	0.1245	0.218

Performance_incentive

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.608	997	0.2827	0.4057

Annual_leave

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.364	3029	0.8588	0.2426

Medical_Family

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.441	631	0.1789	0.2939

Child_day_care

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.255	331	0.09385	0.1701

Nursing_room

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.735	2011	0.5702	0.4903

Parking

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.749	1686	0.478	0.4992

Comp_size

n	missing	distinct	Info	Mean	Gmd
3527	0	5	0.91	2.763	1.307

```
Value        0     1     2     3     4
Frequency   350   207   517  1309  1144
Proportion 0.099 0.059 0.147 0.371 0.324
```

Comp_rating

n	missing	distinct	Info	Mean	Gmd
3527	0	5	0.915	2.385	1.122

```
Value        0     1     2     3     4
Frequency    84   572  1378   889   604
Proportion 0.024 0.162 0.391 0.252 0.171
```

Work_life

n	missing	distinct	Info	Mean	Gmd
3527	0	5	0.877	2.495	1.214

```
Value        0     1     2     3     4
Frequency   377   270   647  1696   537
Proportion 0.107 0.077 0.183 0.481 0.152
```

Stresslevel

n	missing	distinct	Info	Mean	Gmd
3527	0	4	0.918	2.845	1.091

```
Value        1     2     3     4
Frequency   387   891  1130  1119
Proportion 0.110 0.253 0.320 0.317
```

Accept_offer

n	missing	distinct	Info	Sum	Mean	Gmd
3527	0	2	0.748	1677	0.4755	0.4989

2.2 Normality Test of Numerical Variables

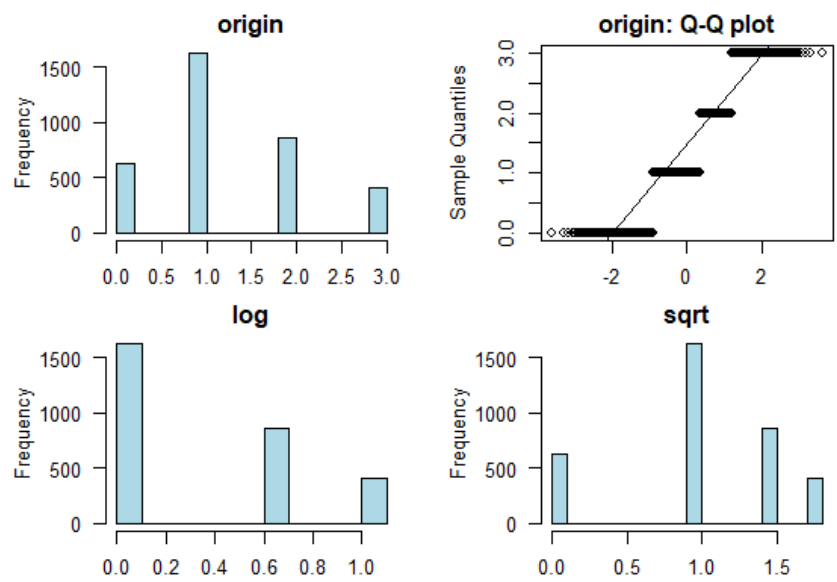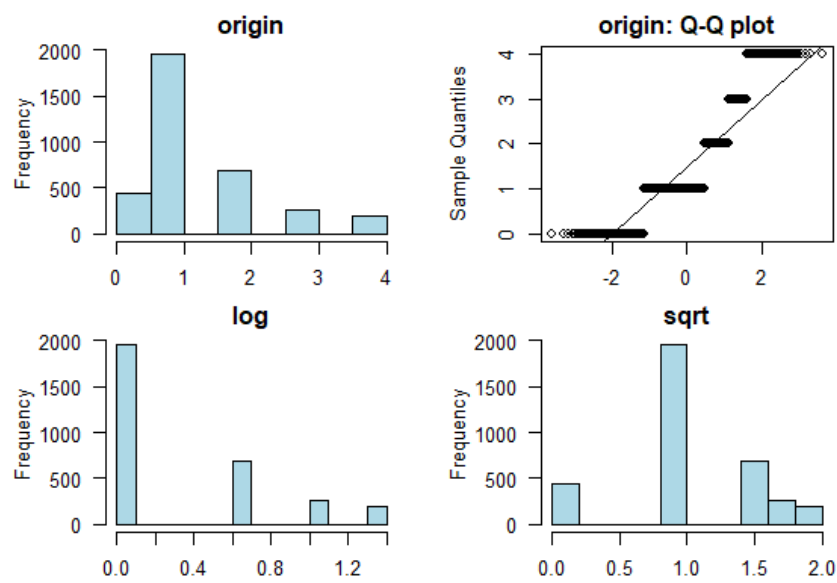## 2.2.1 Statistics and Visualization of (Sample) Data

[Age]

normality test : Shapiro-Wilk normality test

statistic : 0.81069, p-value : 1.11659E-53

skewness and kurtosis

type	skewness	kurtosis
original	0.5879396	2.650072
log transformation	NaN	NaN
sqrt transformation	-0.1681584	1.434451
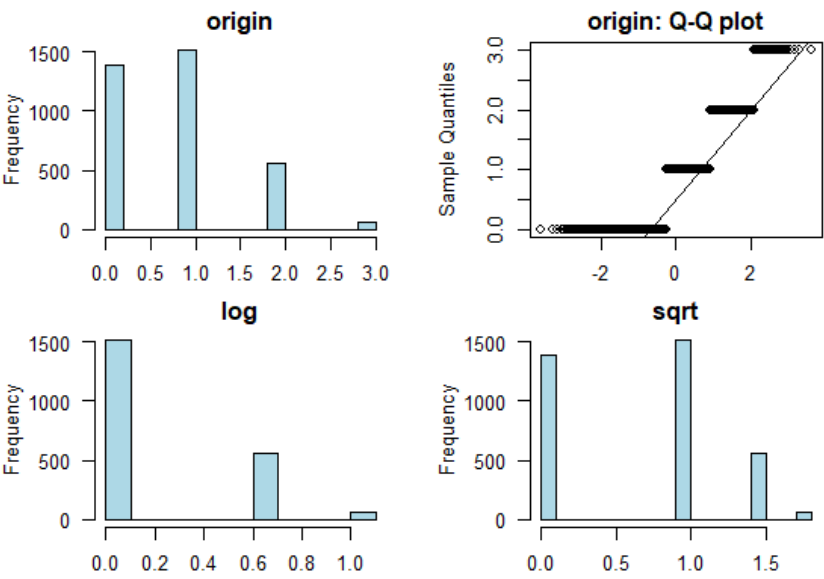
[Gender]

normality test : Shapiro-Wilk normality test

statistic : 0.63438, p-value : 9.15558E-66

skewness and kurtosis

type	skewness	kurtosis
original	-0.1575607	1.024825
log transformation	NaN	NaN
sqrt transformation	-0.1575607	1.024825
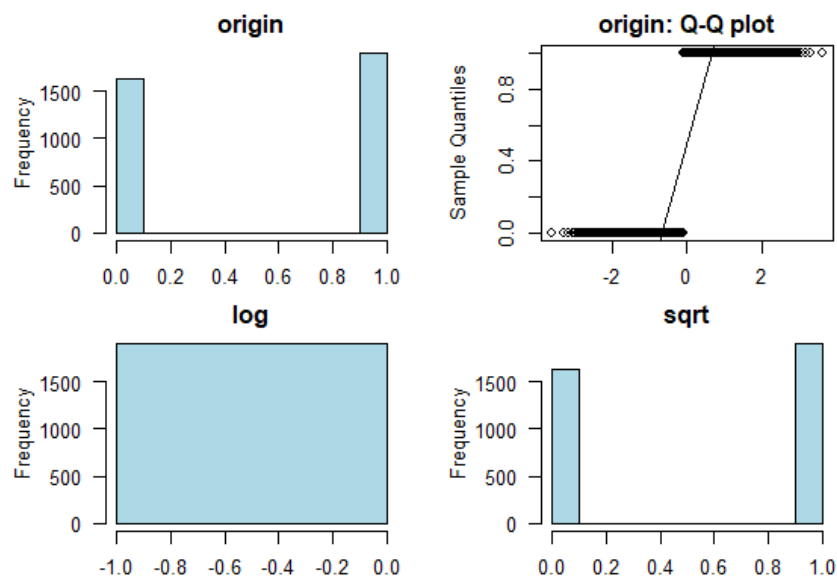
[Marital_status]

normality test : Shapiro-Wilk normality test

```
statistic : 0.63122,  p-value : 6.24644E-66
```

skewness and kurtosis

type	skewness	kurtosis
original	0.245086	1.060067
log transformation	NaN	NaN
sqrt transformation	0.245086	1.060067
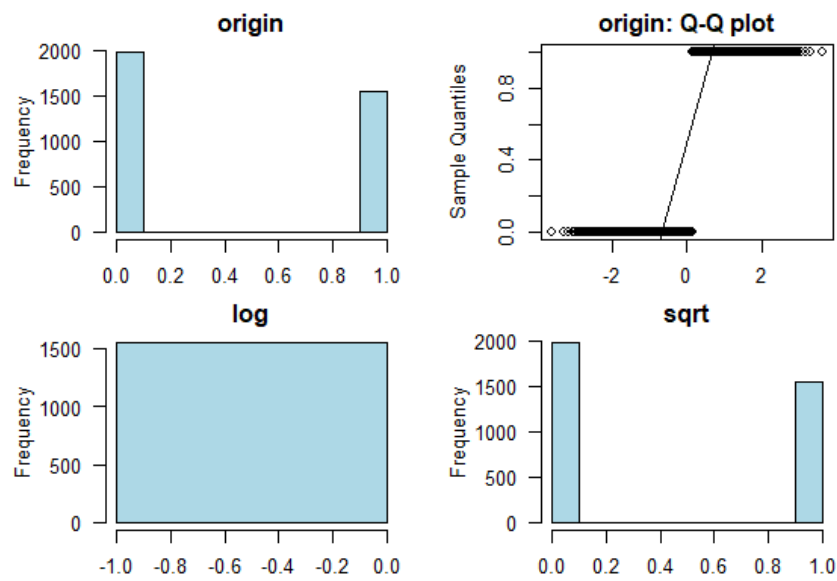
[Num_comp_worked]

normality test : Shapiro-Wilk normality test

```
statistic : 0.66381,  p-value : 3.74087E-64
```

skewness and kurtosis

type	skewness	kurtosis
original	1.591329	4.677313
log transformation	NaN	NaN
sqrt transformation	0.837825	2.140986
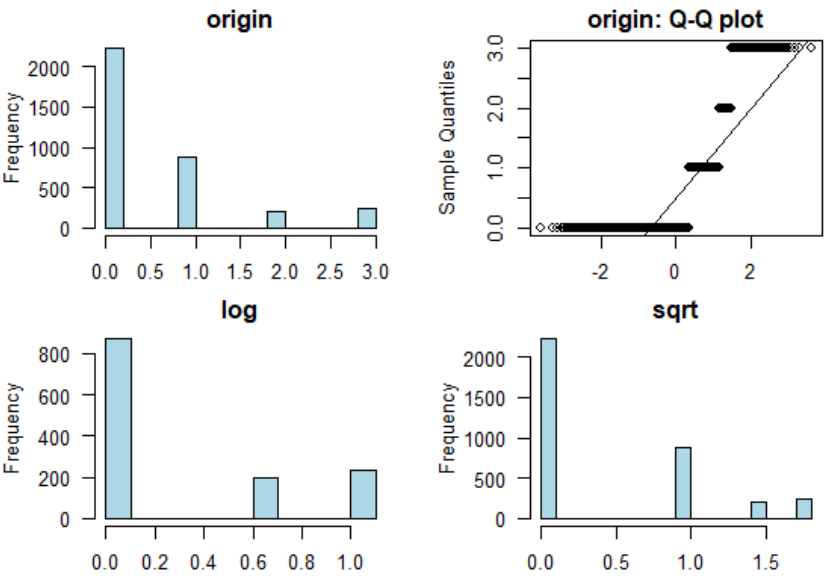
[Work_Experience]

normality test : Shapiro-Wilk normality test

```
statistic : 0.86459,  p-value : 5.19507E-48
```

skewness and kurtosis

type	skewness	kurtosis
original	0.3538369	2.421953
log transformation	NaN	NaN
sqrt transformation	-0.8127897	2.781679

[Last_salary]

normality test : Shapiro-Wilk normality test

    statistic : 0.8142, p-value : 2.36104E-53

skewness and kurtosis

type	skewness	kurtosis
original	1.0454673	3.867493
log transformation	NaN	NaN
sqrt transformation	-0.6078941	3.571923

[Salary_hike]

normality test : Shapiro-Wilk normality test

statistic : 0.85253, p-value : 2.01249E-49

skewness and kurtosis

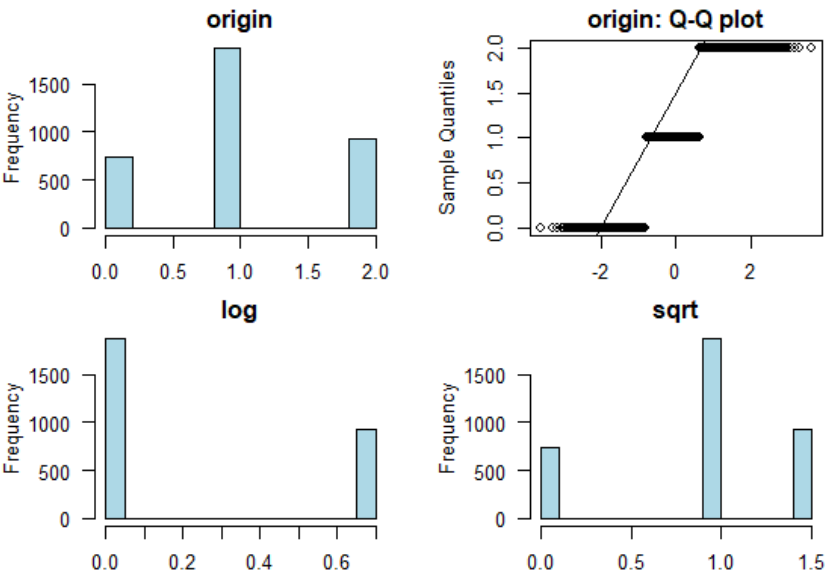| type	skewness	kurtosis
original	0.8239704	3.585294
log transformation	NaN	NaN
sqrt transformation	-0.5949362	2.639159
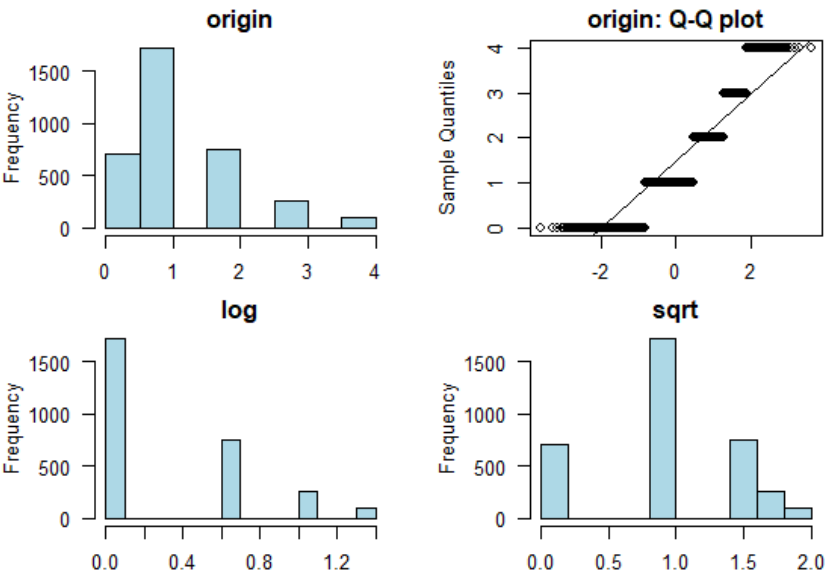
[Work_hours]

normality test : Shapiro-Wilk normality test

statistic : 0.70948, p-value : 2.12392E-61

skewness and kurtosis

type	skewness	kurtosis
original	0.9308180	2.519874
log transformation	NaN	NaN
sqrt transformation	0.5132925	1.478600
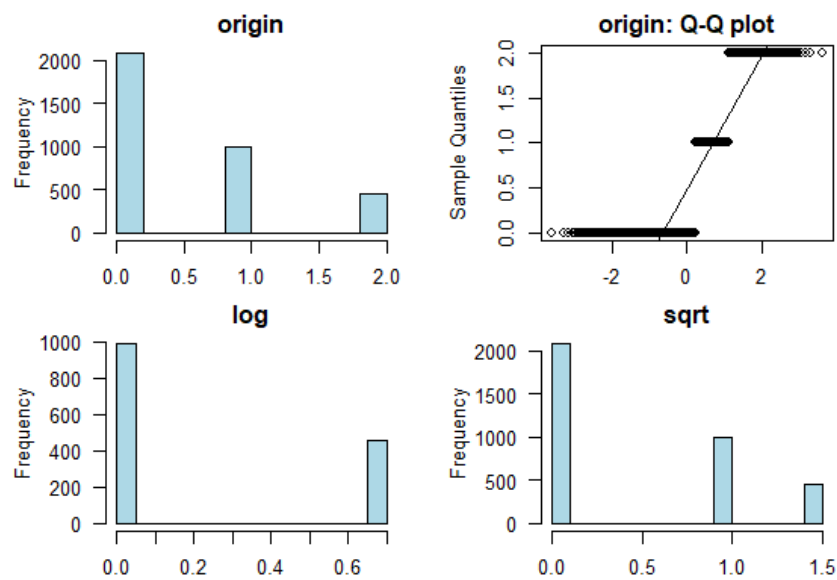
[Traveling_required]

normality test : Shapiro-Wilk normality test

```
statistic : 0.63645,  p-value : 1.17887E-65
```

skewness and kurtosis

type	skewness	kurtosis
original	0.0436736	1.001907
log transformation	NaN	NaN
sqrt transformation	0.0436736	1.001907
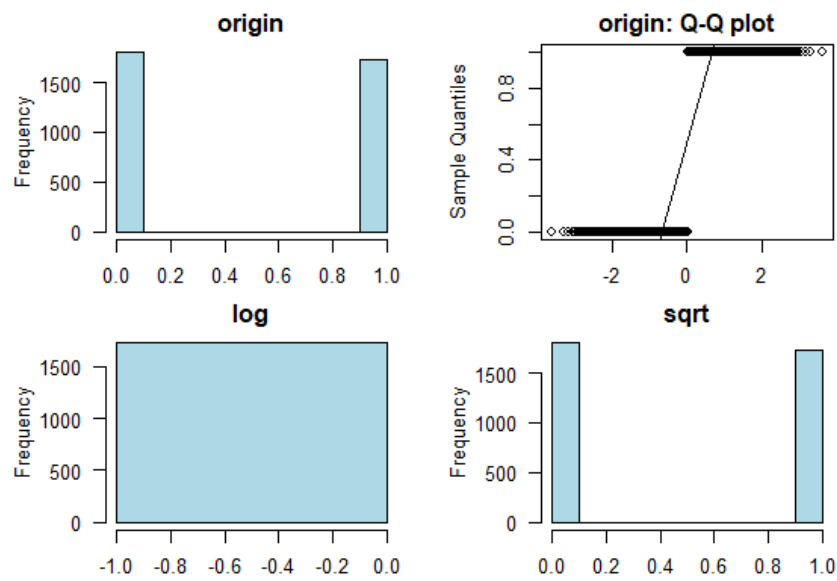
[NoticePeriod_buyout]

normality test : Shapiro-Wilk normality test

```
statistic : 0.60727,  p-value : 3.7369E-67
```

skewness and kurtosis

type	skewness	kurtosis
original	0.5855391	1.342856
log transformation	NaN	NaN
sqrt transformation	0.5855391	1.342856
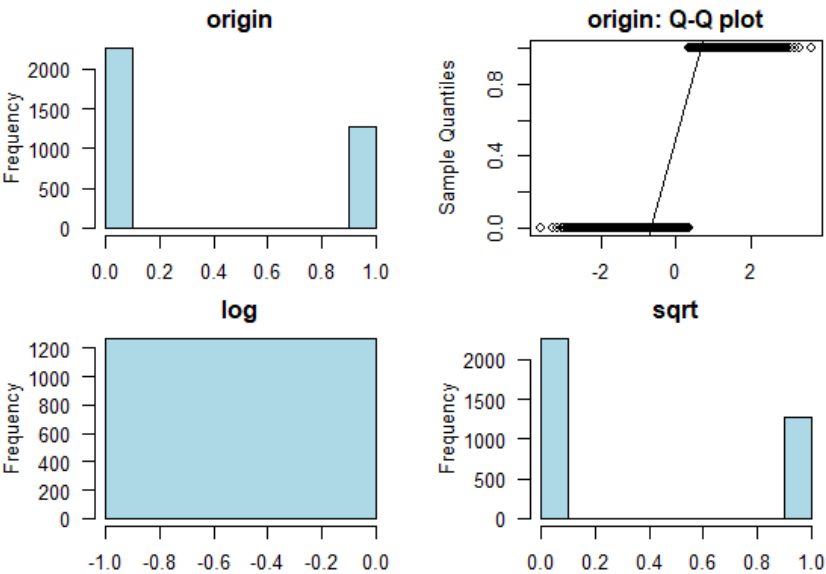
[Job_type]

normality test : Shapiro-Wilk normality test

```
statistic : 0.61417,  p-value : 8.28552E-67
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.5084444	1.258516
log transformation	NaN	NaN
sqrt transformation	-0.5084444	1.258516
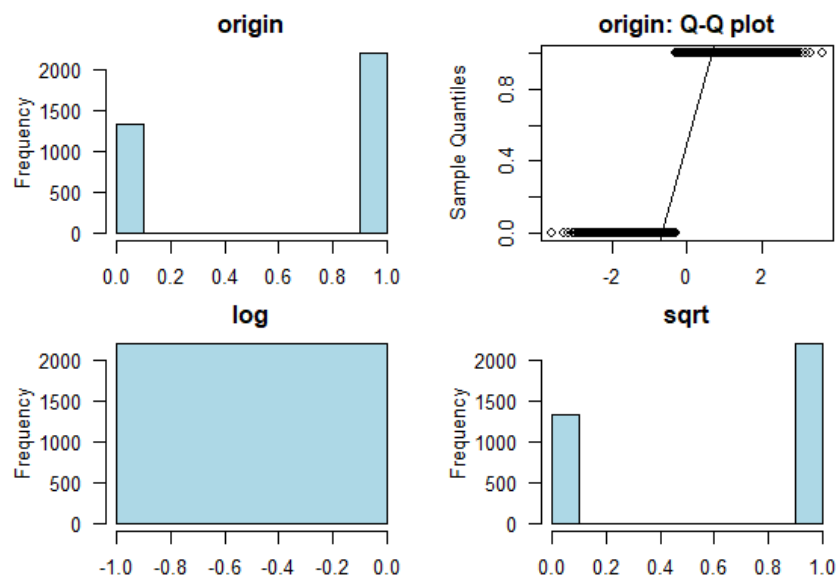
[Whether_comp_Possess_vehicle]

normality test : Shapiro-Wilk normality test

statistic : 0.42755, p-value : 1.01625E-74

skewness and kurtosis

type	skewness	kurtosis
original	-1.951238	4.807329
log transformation	NaN	NaN
sqrt transformation	-1.951238	4.807329
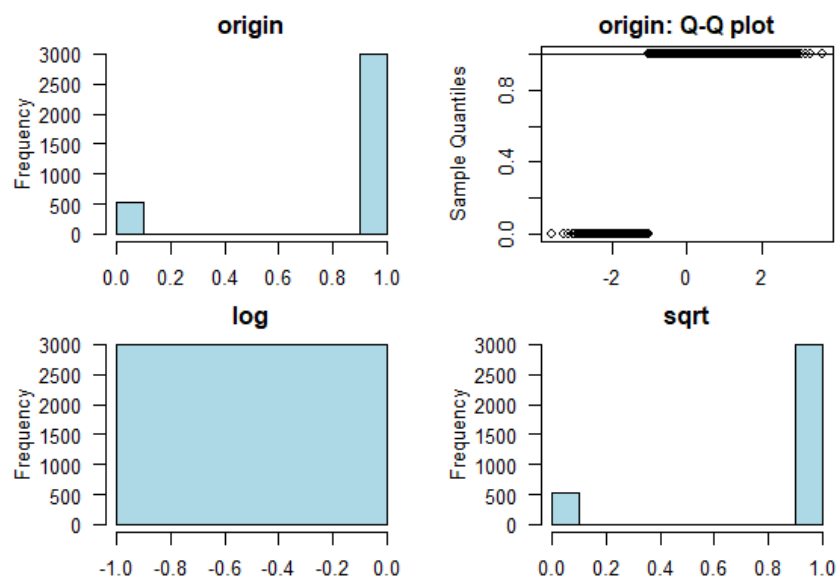
[Work_location]

normality test : Shapiro-Wilk normality test

> statistic : 0.80245, p-value : 2.0116E-54

skewness and kurtosis

type	skewness	kurtosis
original	-0.0670852	2.138914
log transformation	NaN	NaN
sqrt transformation	-0.9650894	2.620237

[Traveltowork_distance_._time]

normality test : Shapiro-Wilk normality test

> statistic : 0.79085, p-value : 1.98781E-55

skewness and kurtosis

type	skewness	kurtosis
original	0.3737775	1.939689
log transformation	NaN	NaN
sqrt transformation	-0.1787093	1.331311
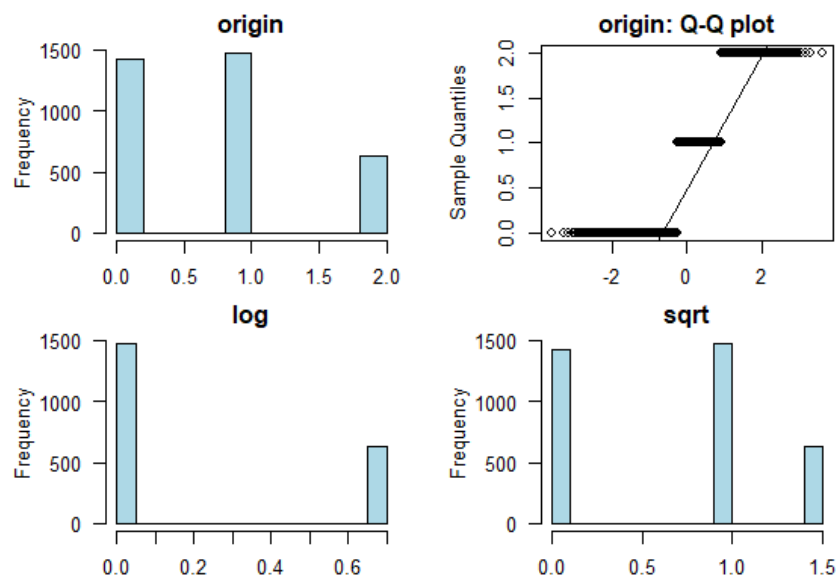
[Work_area_living_cost]

normality test : Shapiro-Wilk normality test

statistic : 0.79521, p-value : 4.689E-55

skewness and kurtosis

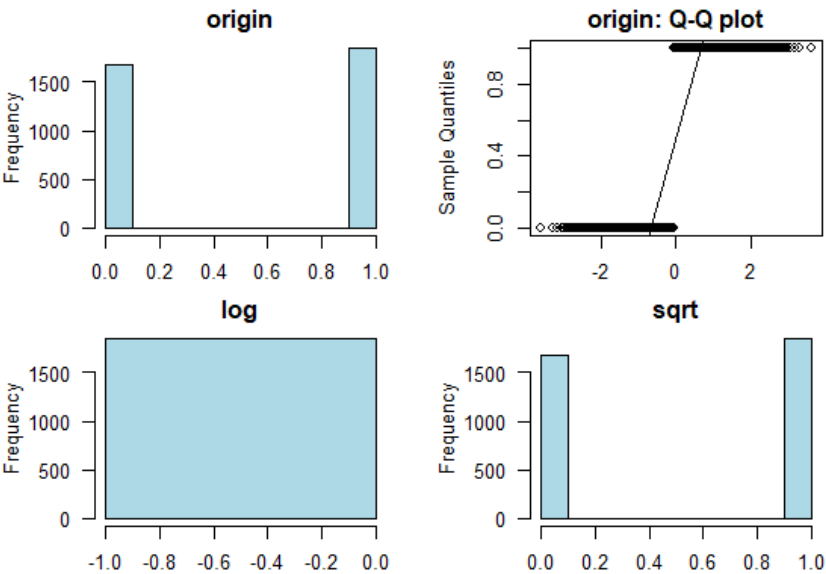| type	skewness	kurtosis
original	-0.3237349	1.953436
log transformation	NaN	NaN
sqrt transformation	-1.1248630	2.964018
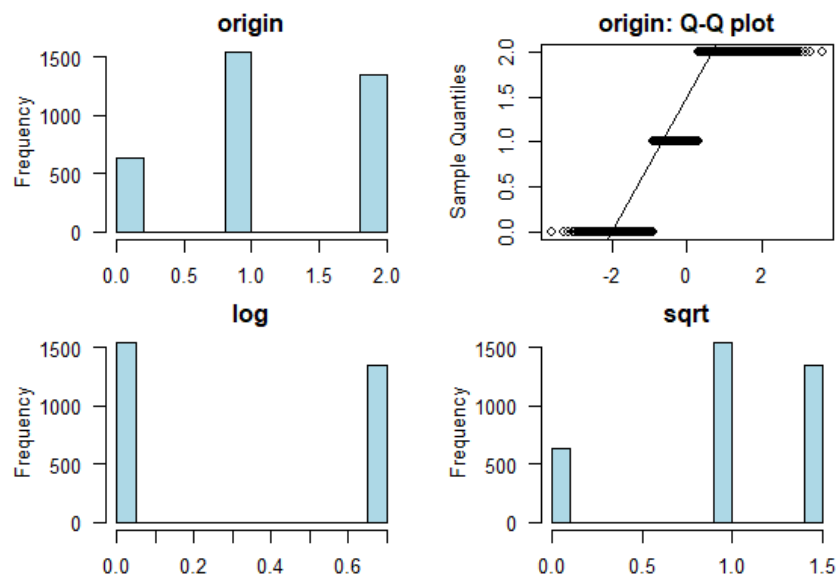
[Public_transport_nearby]

normality test : Shapiro-Wilk normality test

> statistic : 0.63489, p-value : 9.75175E-66

skewness and kurtosis

type	skewness	kurtosis
original	0.1381224	1.019078
log transformation	NaN	NaN
sqrt transformation	0.1381224	1.019078
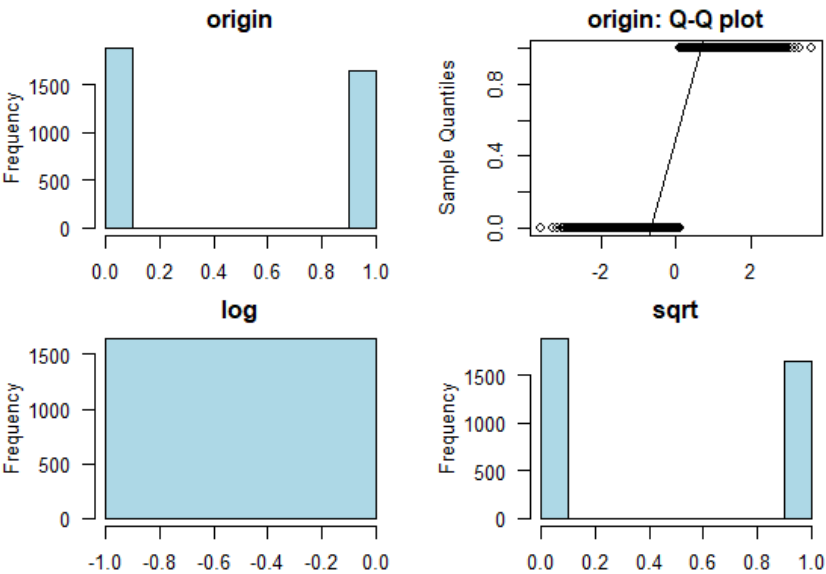
[Current_employment]

normality test : Shapiro-Wilk normality test

> statistic : 0.4521, p-value : 8.10082E-74

skewness and kurtosis

type	skewness	kurtosis
original	-1.769368	4.130664
log transformation	NaN	NaN
sqrt transformation	-1.769368	4.130664
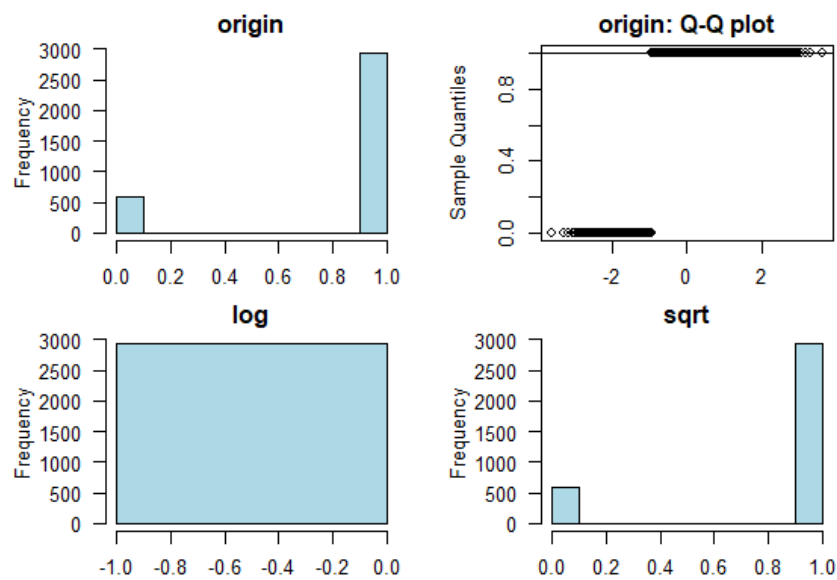
[Offer_letter_Processtime.week.]

normality test : Shapiro-Wilk normality test

```
statistic : 0.82523,  p-value : 2.68829E-52
```

skewness and kurtosis

type	skewness	kurtosis
original	0.405142	3.357904
log transformation	NaN	NaN
sqrt transformation	-1.347298	7.810792
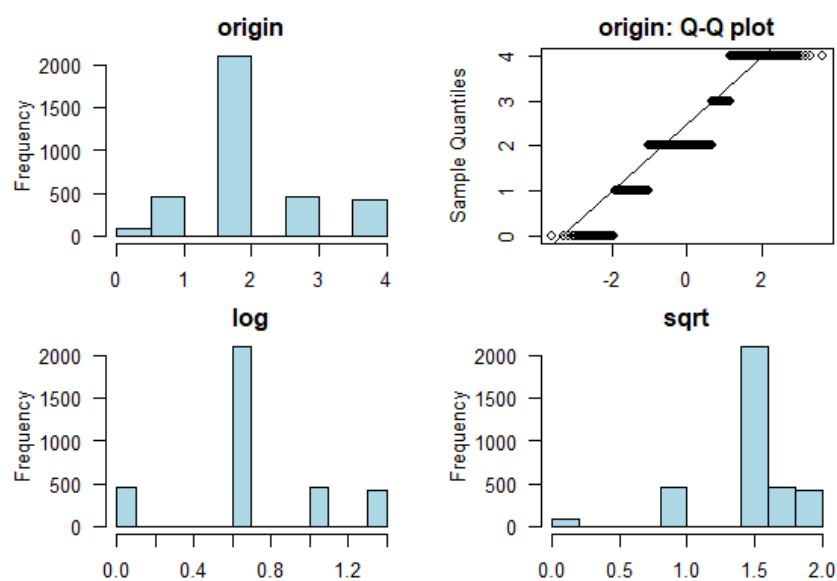
[Bonus]

normality test : Shapiro-Wilk normality test

> statistic : 0.63579, p-value : 1.08731E-65

skewness and kurtosis

type	skewness	kurtosis
original	-0.0959424	1.009205
log transformation	NaN	NaN
sqrt transformation	-0.0959424	1.009205

[Allowance]

normality test : Shapiro-Wilk normality test

> statistic : 0.4894, p-value : 2.22041E-72

skewness and kurtosis

type	skewness	kurtosis
original	1.503105	3.259323
log transformation	NaN	NaN
sqrt transformation	1.503105	3.259323
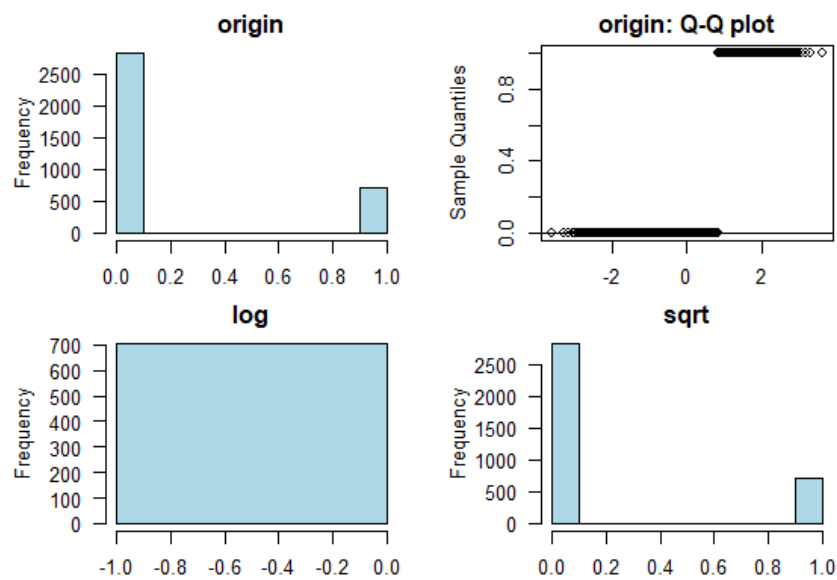
[Overtime_pay]

normality test : Shapiro-Wilk normality test

```
statistic : 0.54796,  p-value : 6.26542E-70
```

skewness and kurtosis

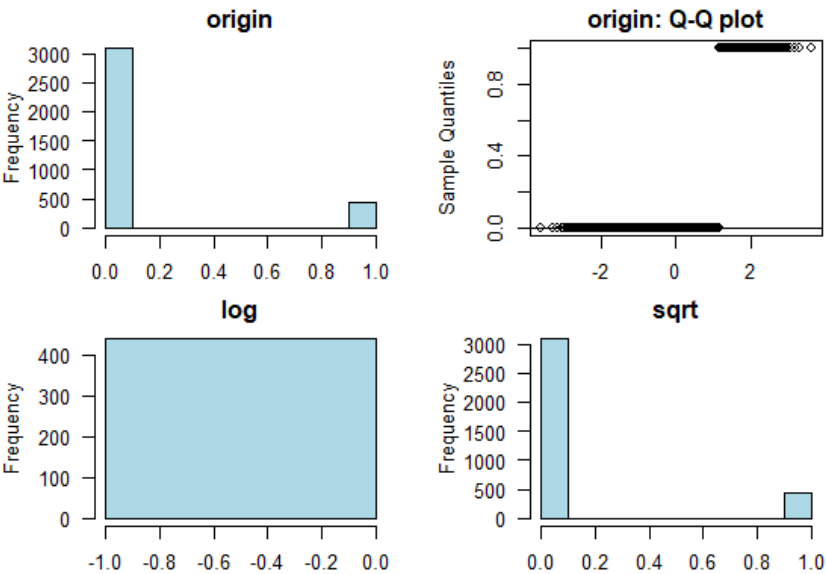| type	skewness	kurtosis
original	1.085962	2.179314
log transformation	NaN	NaN
sqrt transformation	1.085962	2.179314
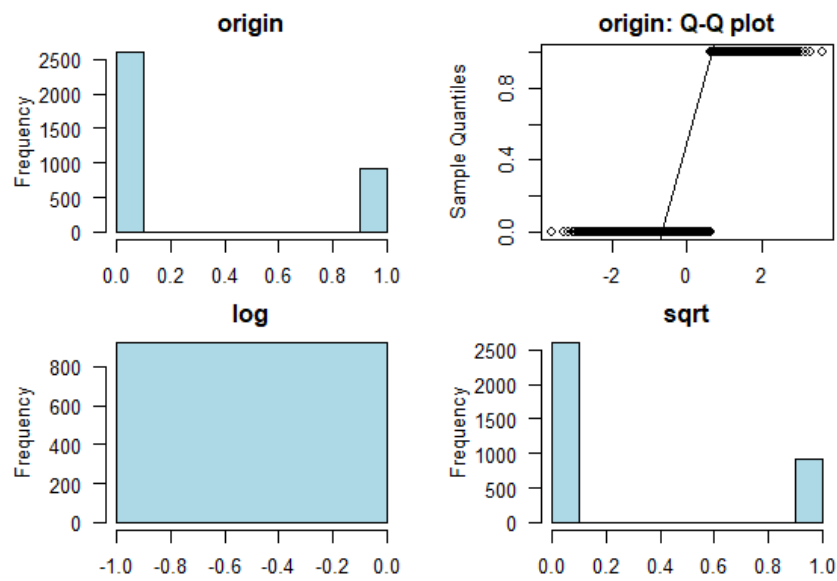
[Emp_shares_scheme]

normality test : Shapiro-Wilk normality test

statistic : 0.3865, p-value : 3.72852E-76

skewness and kurtosis

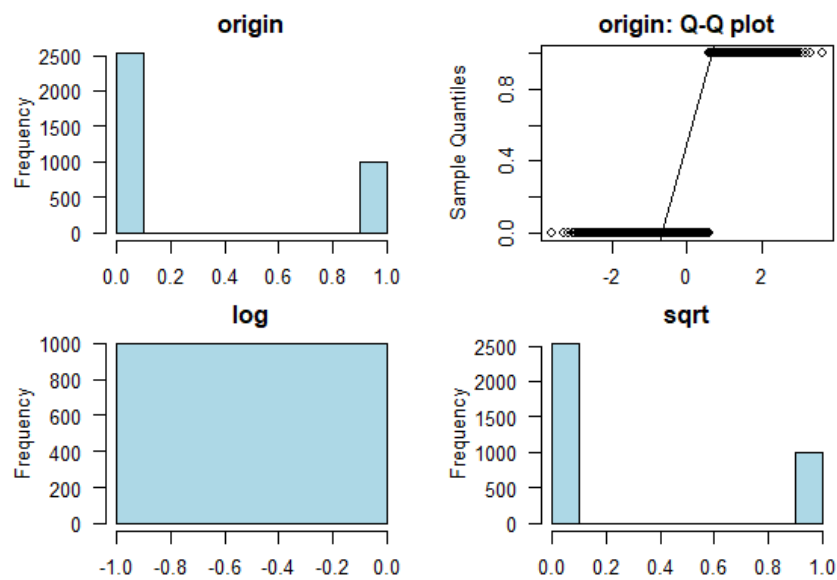| type	skewness	kurtosis
original	2.275155	6.176332
log transformation	NaN	NaN
sqrt transformation	2.275155	6.176332

[Performance_incentive]

normality test : Shapiro-Wilk normality test

statistic : 0.56398, p-value : 3.26897E-69

skewness and kurtosis

type	skewness	kurtosis
original	0.9652378	1.931684
log transformation	NaN	NaN
sqrt transformation	0.9652378	1.931684

origin

origin: Q-Q plot
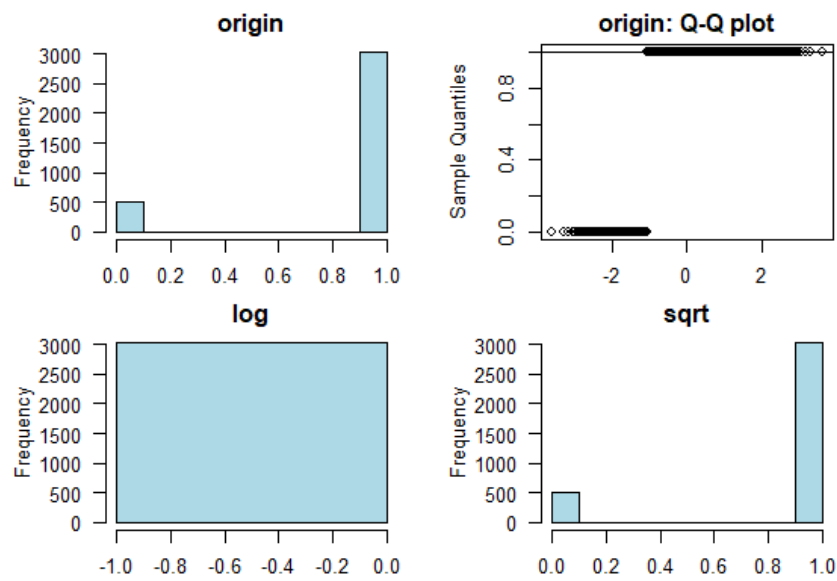

log

sqrt

[Annual_leave]

normality test : Shapiro-Wilk normality test

```
statistic : 0.41325,  p-value : 3.14312E-75
```

skewness and kurtosis

type	skewness	kurtosis
original	-2.060762	5.24674
log transformation	NaN	NaN
sqrt transformation	-2.060762	5.24674

origin

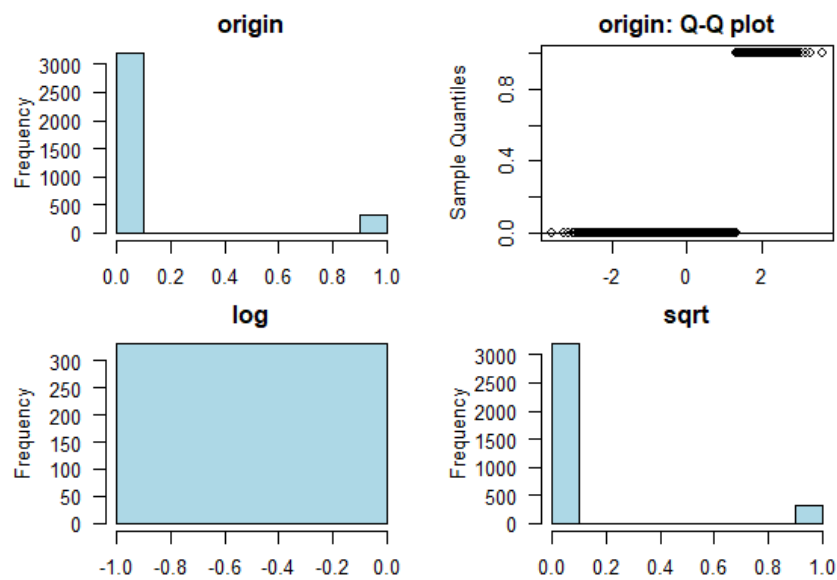## origin: Q-Q plot

log

sqrt

[Medical_Family]

normality test : Shapiro-Wilk normality test

> statistic : 0.46511, p-value : 2.51186E-73

skewness and kurtosis

type	skewness	kurtosis
original	1.675538	3.807427
log transformation	NaN	NaN
sqrt transformation	1.675538	3.807427
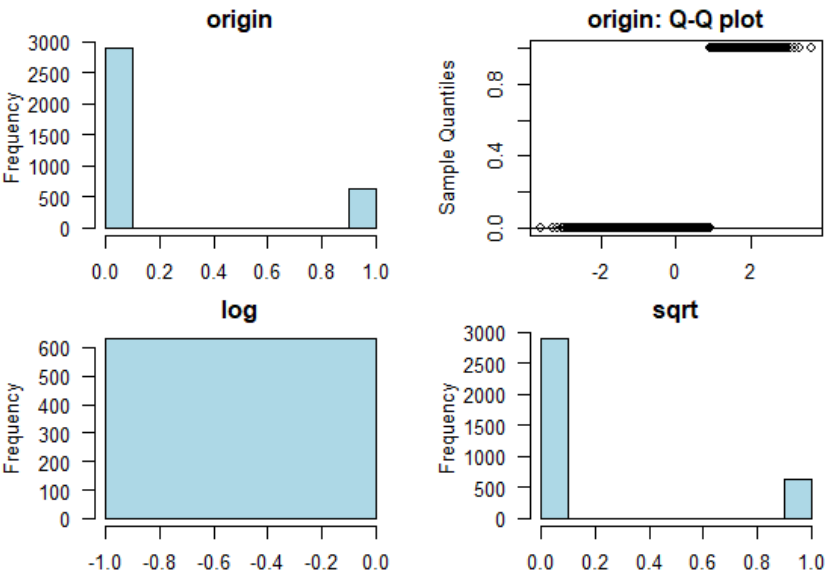
[Child_day_care]
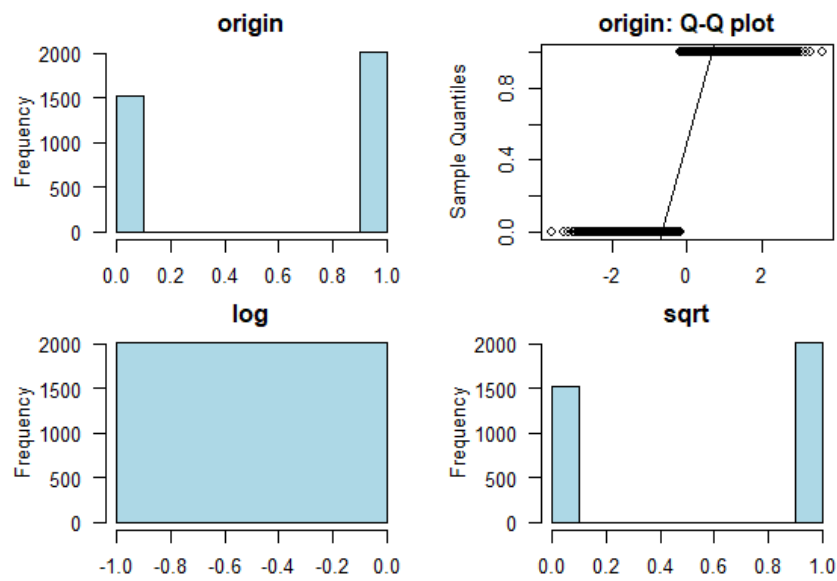
normality test : Shapiro-Wilk normality test

> statistic : 0.32988, p-value : 5.26717E-78

skewness and kurtosis

type	skewness	kurtosis
original	2.785526	8.759156
log transformation	NaN	NaN
sqrt transformation	2.785526	8.759156

origin

origin: Q-Q plot

log

sqrt

[Nursing_room]

normality test : Shapiro-Wilk normality test

```
statistic : 0.62942,  p-value : 5.02857E-66
```

skewness and kurtosis

type	skewness	kurtosis
original	-0.2834977	1.080371
log transformation	NaN	NaN
sqrt transformation	-0.2834977	1.080371

origin

origin: Q-Q plot
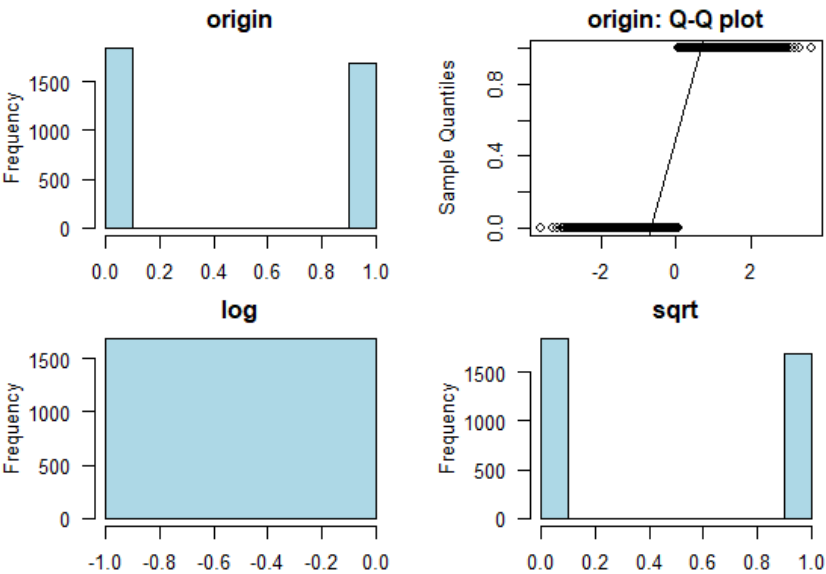

log

sqrt

[Parking]

normality test : Shapiro-Wilk normality test

```
statistic : 0.63592,  p-value : 1.10507E-65
```

skewness and kurtosis

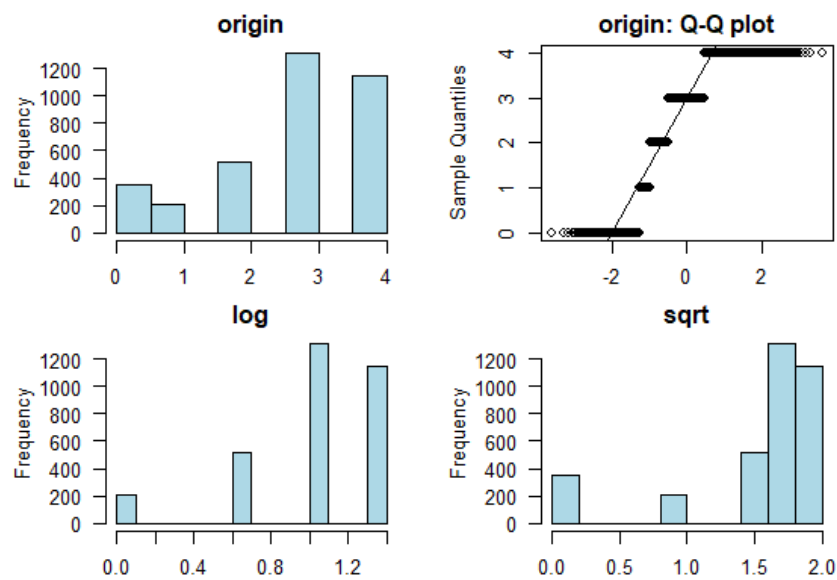| type	skewness	kurtosis
original	0.0879784	1.00774
log transformation	NaN	NaN
sqrt transformation	0.0879784	1.00774

[Comp_size]

normality test : Shapiro-Wilk normality test

```
statistic : 0.82105,  p-value : 1.05444E-52
```

skewness and kurtosis

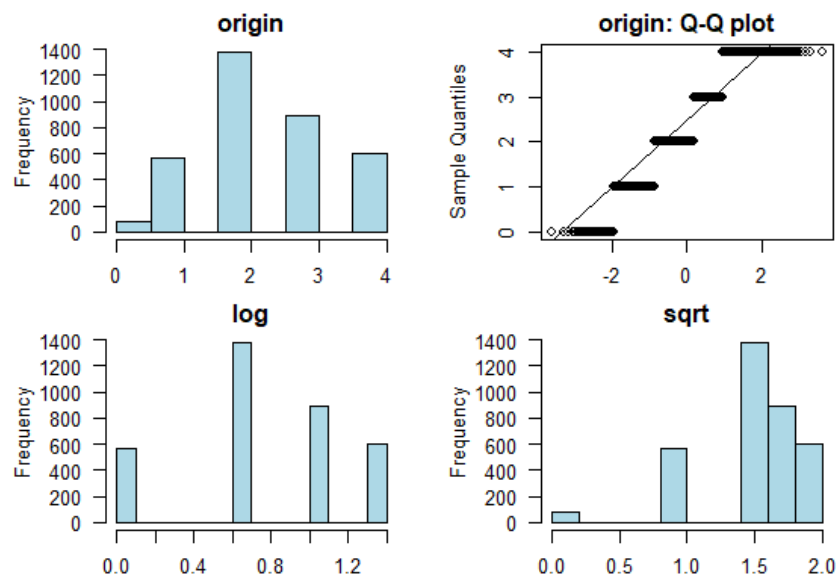| type	skewness	kurtosis
original	-0.9706788	3.008641
log transformation	NaN	NaN
sqrt transformation	-1.8187368	5.311680

[Comp_rating]

normality test : Shapiro-Wilk normality test

```
statistic : 0.89856,  p-value : 2.18127E-43
```

skewness and kurtosis

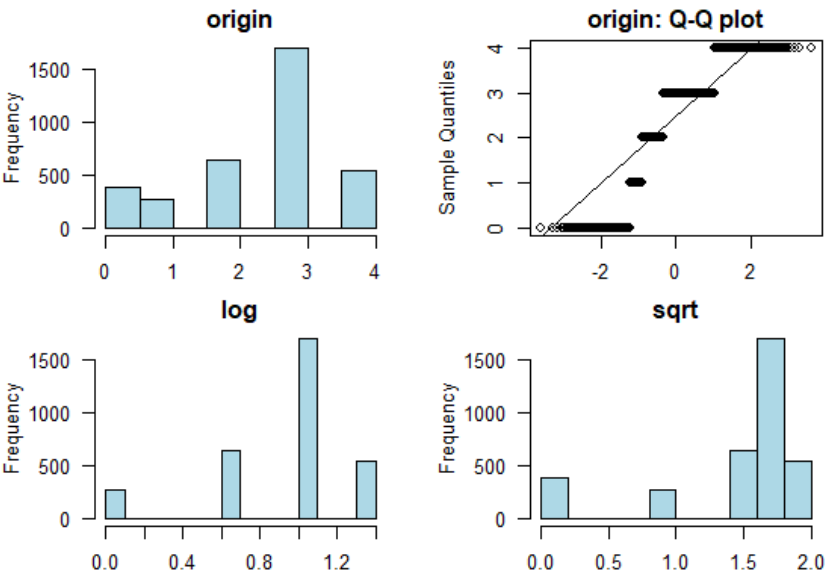| type	skewness	kurtosis
original	0.004223	2.353191
log transformation	NaN	NaN
sqrt transformation	-1.223634	5.965823

[Work_life]

normality test : Shapiro-Wilk normality test

statistic : 0.83563, p-value : 2.99917E-51

skewness and kurtosis

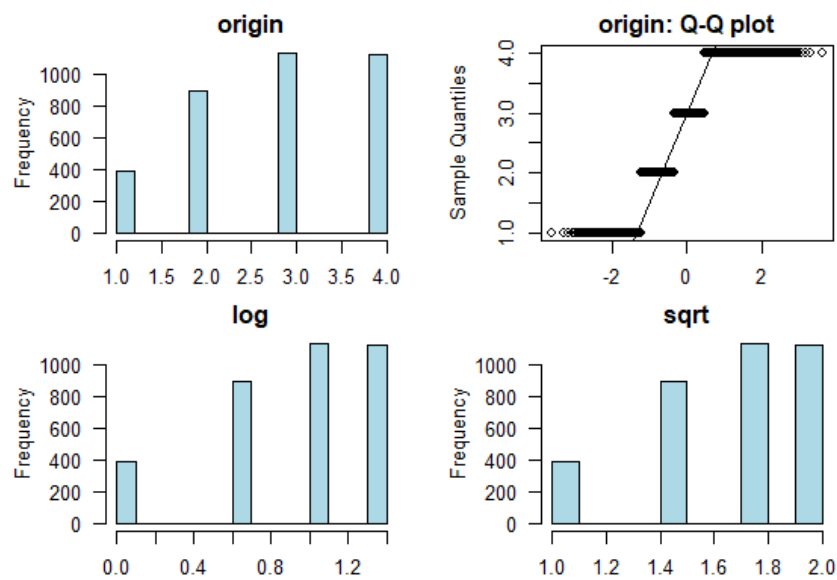| type	skewness	kurtosis
original	-0.8664152	2.939154
log transformation	NaN	NaN
sqrt transformation	-1.7303527	4.959731

[Stresslevel]

normality test : Shapiro-Wilk normality test

statistic : 0.85749, p-value : 7.46124E-49

skewness and kurtosis

type	skewness	kurtosis
original	-0.3602453	2.026329
log transformation	-1.0259904	3.206498
sqrt transformation	-0.6699032	2.497895

origin

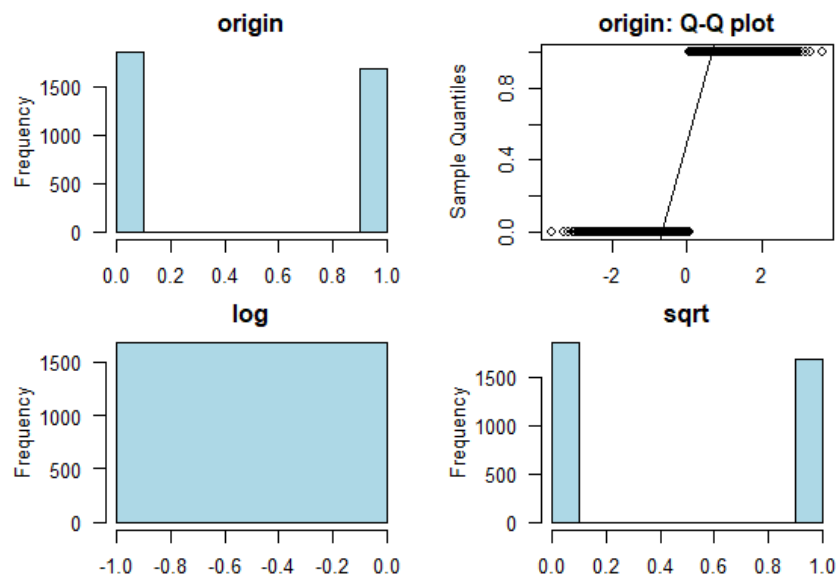### origin: Q-Q plot

log

sqrt

[Accept_offer]

normality test : Shapiro-Wilk normality test

```
statistic : 0.63575,  p-value : 1.08201E-65
```

skewness and kurtosis

type	skewness	kurtosis
original	0.0982186	1.009647
log transformation	NaN	NaN
sqrt transformation	0.0982186	1.009647

origin

origin: Q-Q plot

log

sqrt

3 Relationship Between Variables

3.1 Correlation Coefficient

3.1.1 Correlation Coefficient by Variable Combination

Table of correlation coefficients (0.5 or more)

Variable1	Variable2	Correlation Coefficient
Work_Experience	Age	0.6235723
Comp_size	Nursing_room	0.5839116
Last_salary	Age	0.5692967
Last_salary	Work_Experience	0.5513739
Medical_Family	Age	0.5123972
NoticePeriod_buyout	Salary_hike	0.5046585

3.1.2 Correlation Plot of Numerical Variables

4 Target based Analysis

4.1 Grouped Descriptive Statistics

4.1.1 Grouped Numerical Variables

There is no target variable.

4.1.2 Grouped Categorical Variables

There is no target variable.

4.2 Grouped Relationship Between Variables

4.2.1 Grouped Correlation Coefficient

```
There is no target variable.
```

4.2.2 Grouped Correlation Plot of Numerical Variables

```
There is no target variable.
```