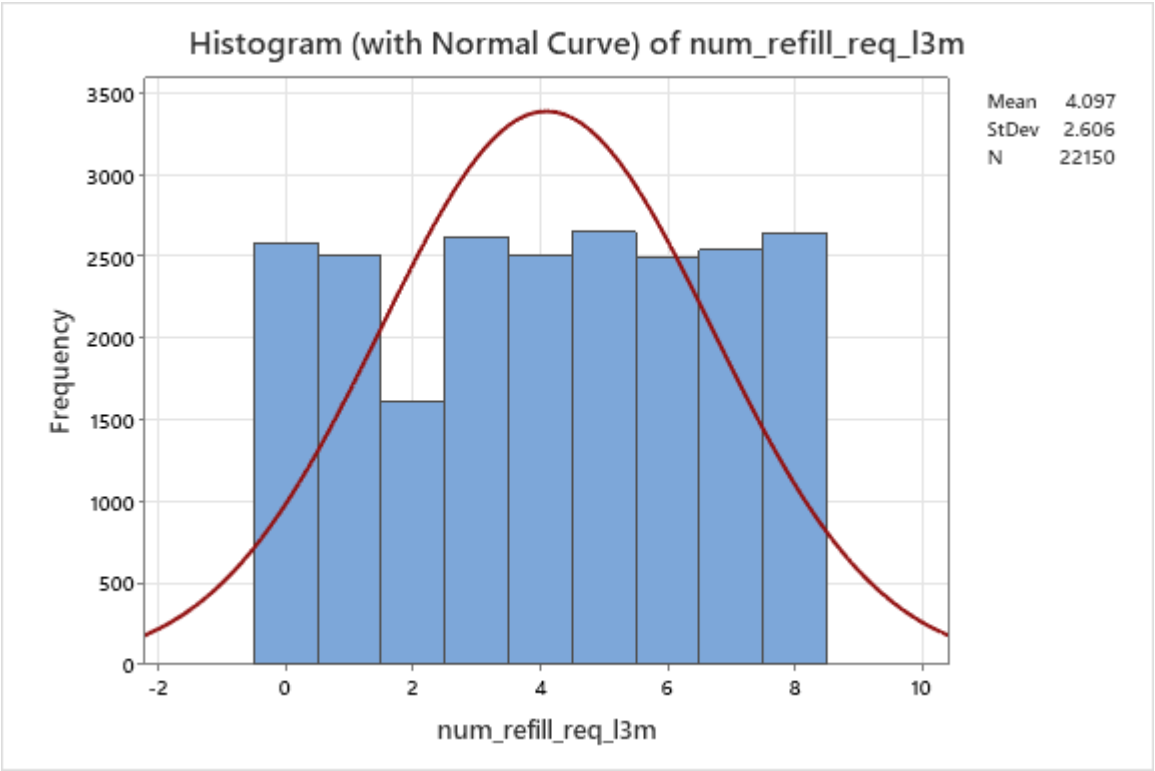
DATA ANALYTICS REPORT

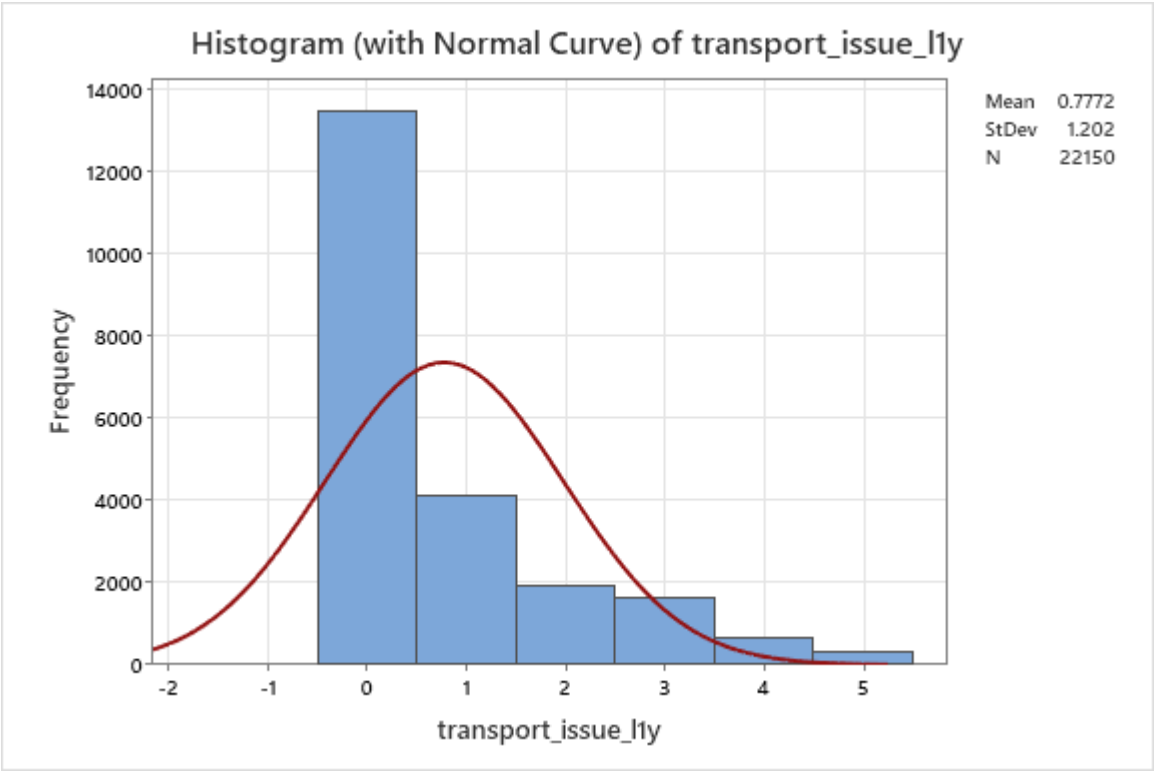
1. DESCRIPTIVE STATS:

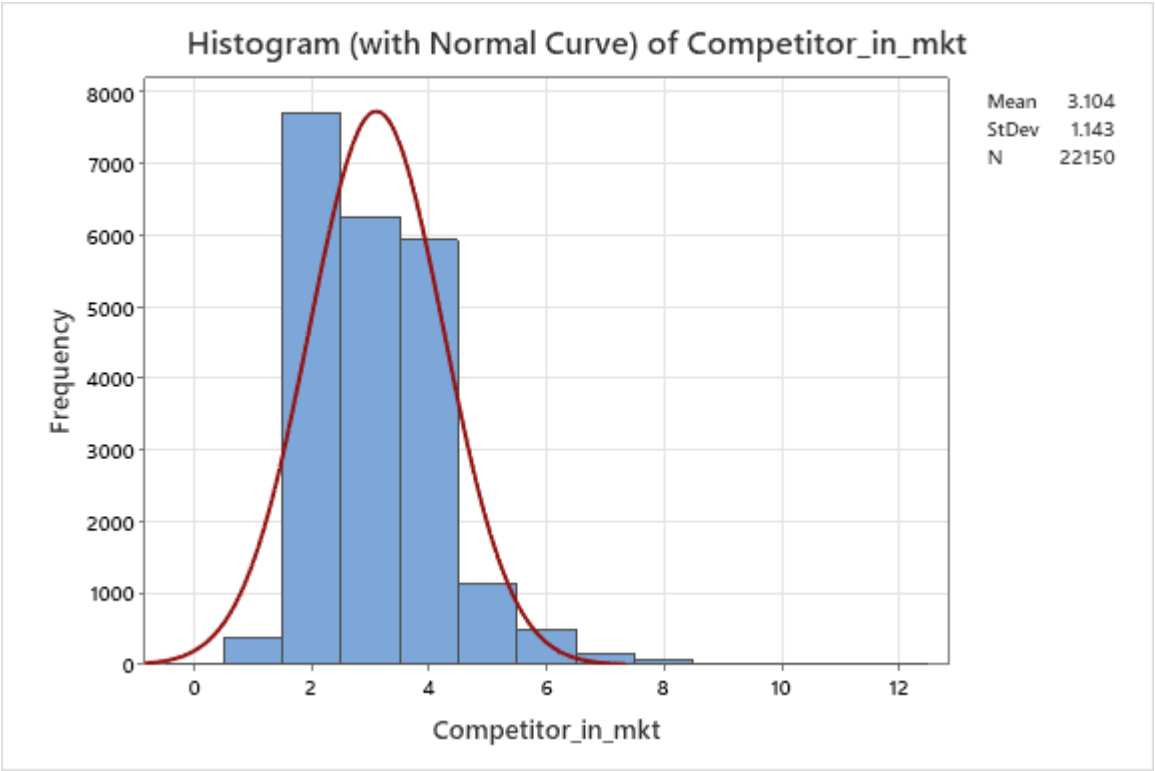
Statistics

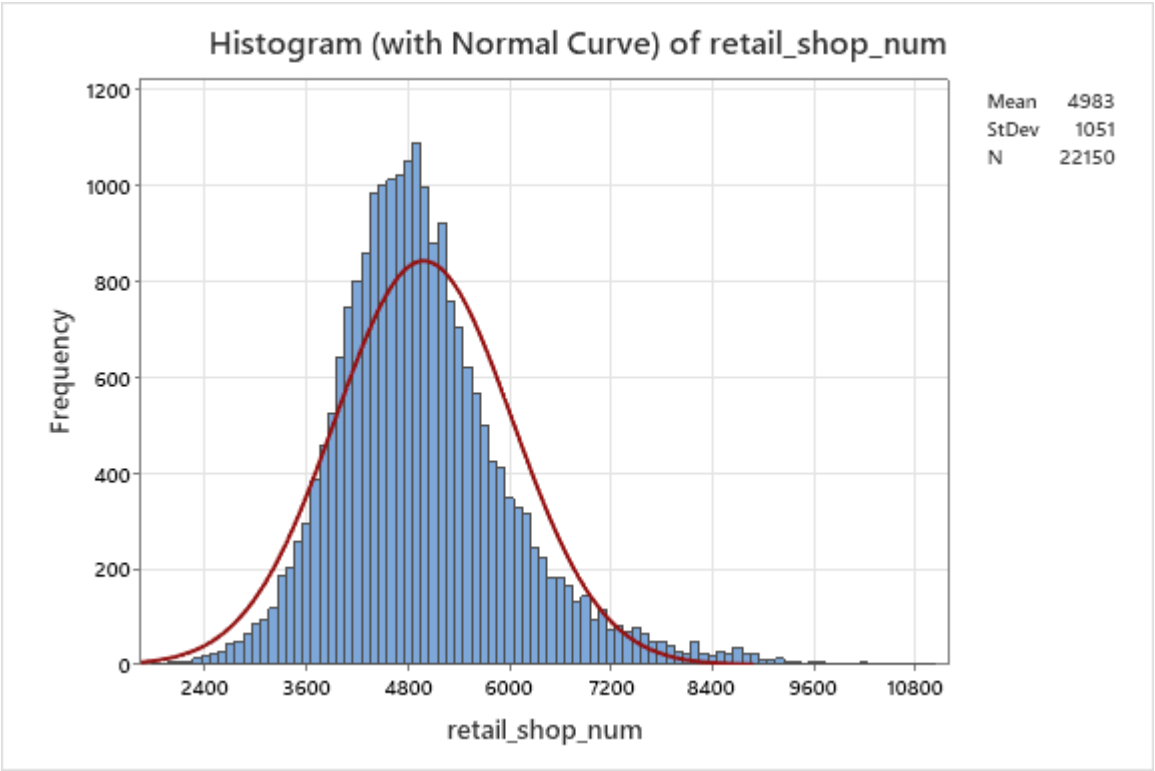
Variable	N	N*	Mean	SE Mean	StDev	Variance	Minimum
num_refill_req_l3m	22150	0	4.0970	0.0175	2.6063	6.7927	0.0000
transport_issue_l1y	22150	0	0.77720	0.00807	1.20175	1.44420	0.00000
Competitor_in_mkt	22150	0	3.1039	0.00768	1.1429	1.3062	0.00000
retail_shop_num	22150	0	4983.1	7.06	1050.6	1103832.3	1821.0
distributor_num	22150	0	42.387	0.108	16.058	257.851	15.000
flood_impacted	22150	0	0.09869	0.00200	0.29825	0.08895	0.00000
flood_proof	22150	0	0.05449	0.00153	0.22699	0.05153	0.00000
electric_supply	22150	0	0.65607	0.00319	0.47503	0.22565	0.00000
dist_from_hub	22150	0	163.61	0.421	62.66	3926.36	55.00
workers_num	21273	877	28.936	0.0538	7.843	61.519	10.000
wh_est_year	11605	10545	2009.4	0.0699	7.53	56.7	1996.0
storage_issue_reported_l3m	22150	0	17.117	0.0616	9.174	84.166	0.0000
temp_reg_mach	22150	0	0.30420	0.00309	0.46008	0.21167	0.00000
wh_breakdown_l3m	22150	0	3.4878	0.0114	1.6917	2.8617	0.0000
govt_check_l3m	22150	0	18.768	0.0581	8.645	74.732	1.000
product_wg_ton	22150	0	22087	78.1	11626	135168348	2065

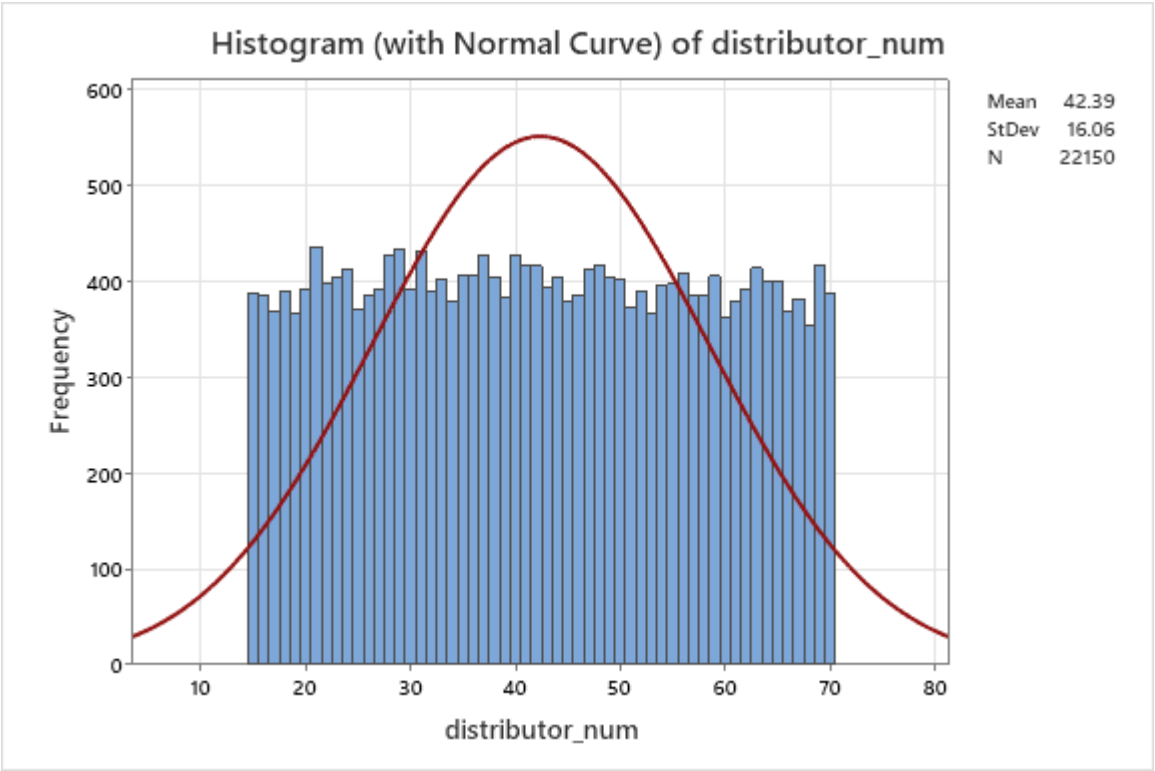
Variable	Q1	Median	Q3	Maximum	IQR	Skewness	Kurtosis
num_refill_req_l3m	2.0000	4.0000	6.0000	8.0000	4.0000	-0.08	-1.22
transport_issue_l1y	0.00000	0.00000	1.00000	5.00000	1.00000	1.61	1.82
Competitor_in_mkt	2.0000	3.0000	4.0000	12.0000	2.0000	0.99	1.80
retail_shop_num	4309.0	4859.0	5499.0	11008.0	1190.0	0.91	1.85
distributor_num	29.000	42.000	56.000	70.000	27.000	0.02	-1.19
flood_impacted	0.00000	0.00000	0.00000	1.00000	0.00000	2.69	5.24
flood_proof	0.00000	0.00000	0.00000	1.00000	0.00000	3.93	13.41
electric_supply	0.00000	1.00000	1.00000	1.00000	1.00000	-0.66	-1.57
dist_from_hub	109.00	164.00	218.00	271.00	109.00	-0.01	-1.20
workers_num	24.000	28.000	33.000	98.000	9.000	1.04	3.29
wh_est_year	2003.0	2009.0	2016.0	2023.0	13.0	0.01	-1.18
storage_issue_reported_l3m	10.000	18.000	24.000	39.000	14.000	0.12	-0.68
temp_reg_mach	0.00000	0.00000	1.00000	1.00000	1.00000	0.85	-1.28
wh_breakdown_l3m	2.0000	3.0000	5.0000	6.0000	3.0000	-0.07	-0.95
govt_check_l3m	11.000	20.000	26.000	32.000	15.000	-0.36	-1.06
product_wg_ton	12151	22099	30102	55151	17951	0.34	-0.50

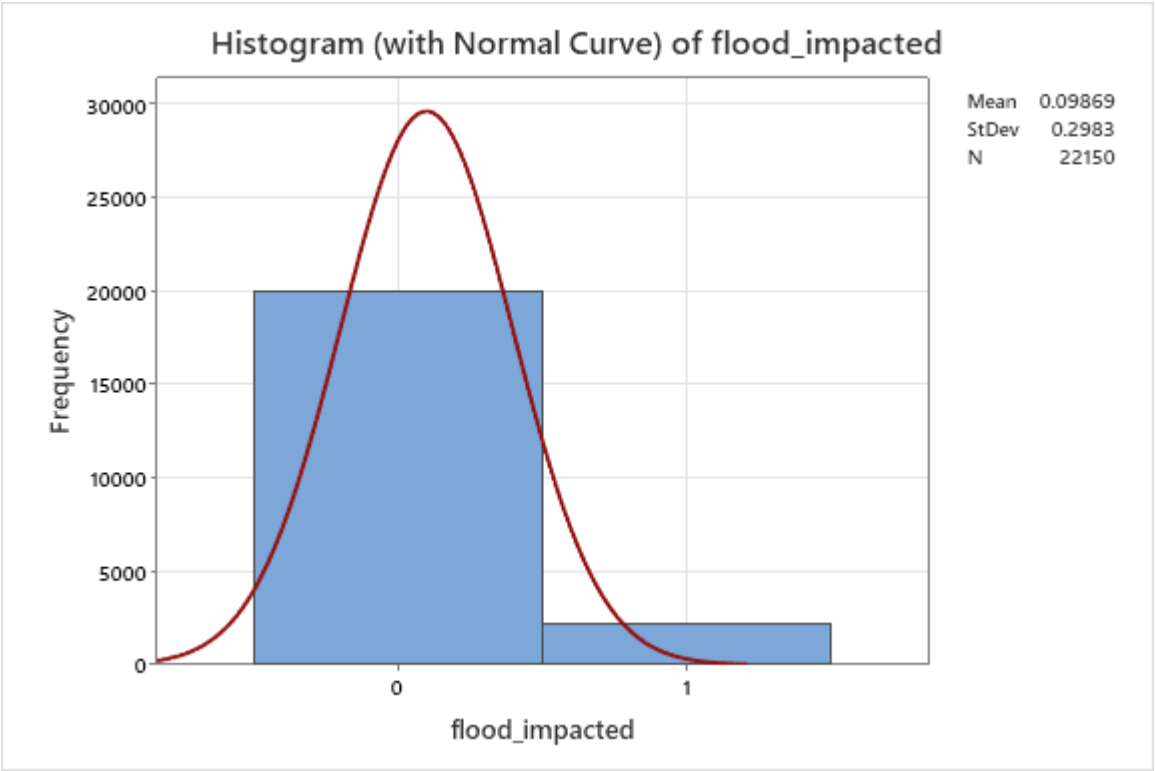

Histogram (with Normal Curve) of num_refill_req_l3m
Mean 4.097
StDev 2.606
N 22150
Frequency
num_refill_req_l3m
0
500
1000
1500
2000
2500
3000
3500
-2
0
2
4
6
8
10

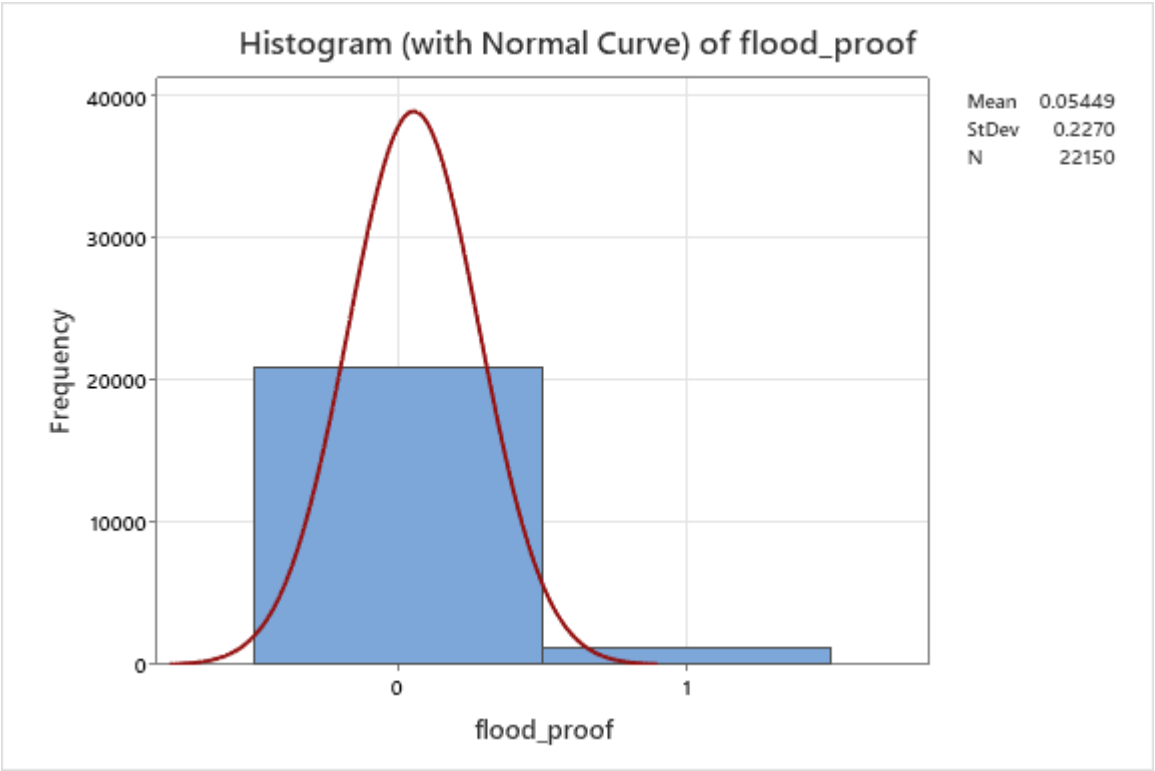

Histogram (with Normal Curve) of transport_issue_l1y
Mean 0.7772
StDev 1.202
N 22150
Frequency
transport_issue_l1y
0
2000
4000
6000
8000
10000
12000
14000
-2
-1
0
1
2
3
4
5

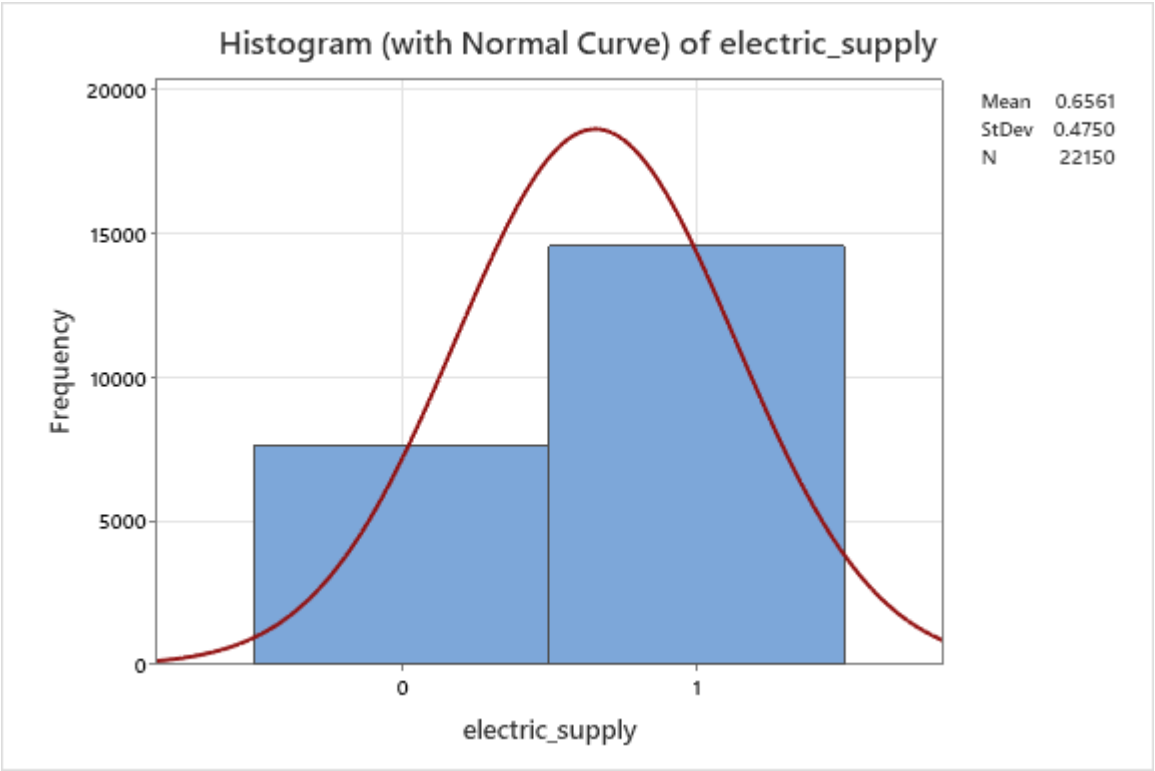

Histogram (with Normal Curve) of Competitor_in_mkt
Mean 3.104
StDev 1.143
N 22150
Frequency
Competitor_in_mkt
0
1000
2000
3000
4000
5000
6000
7000
8000
0
2
4
6
8
10
12

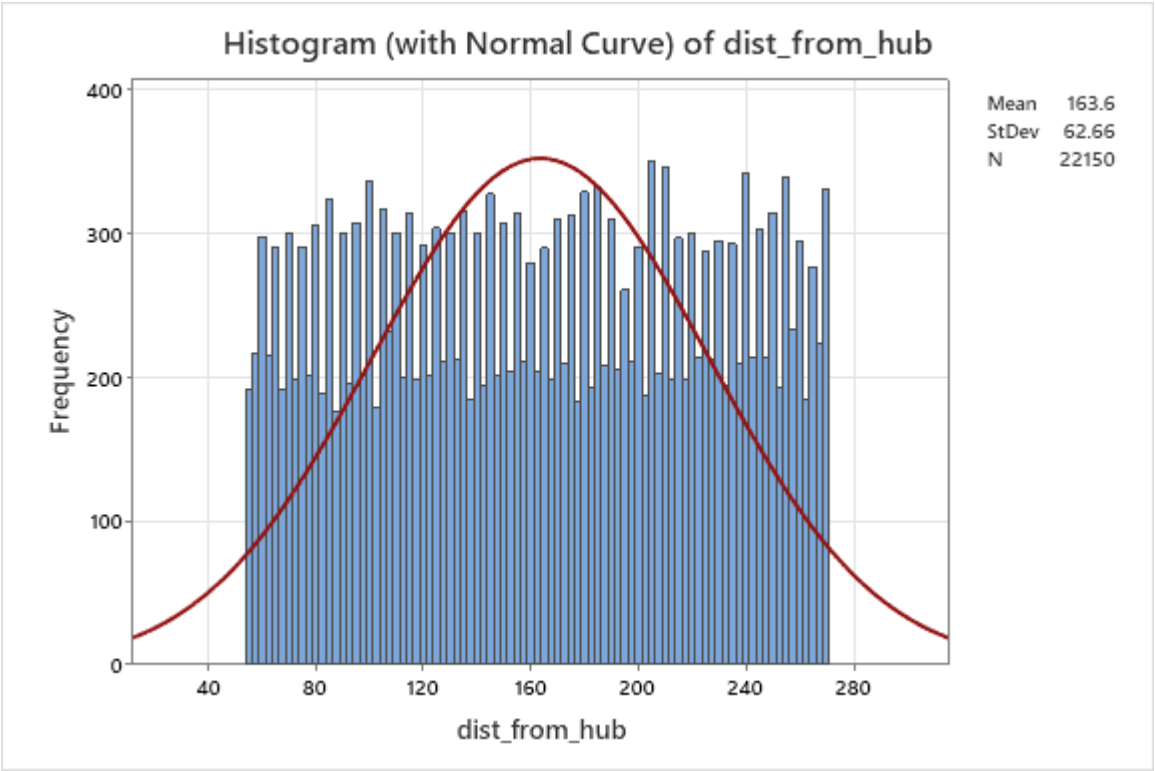

Histogram (with Normal Curve) of retail_shop_num
Mean 4983
StDev 1051
N 22150
Frequency
retail_shop_num
0
200
400
600
800
1000
1200
2400
3600
4800
6000
7200
8400
9600
10800

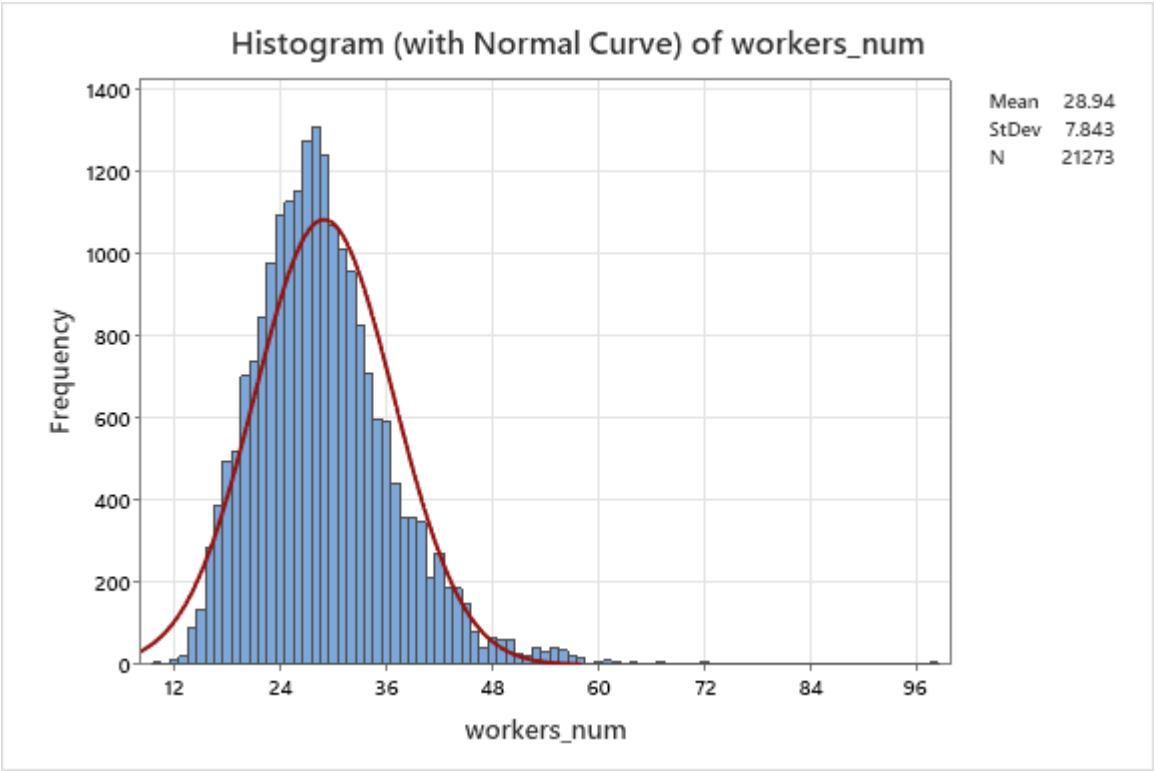

Histogram (with Normal Curve) of distributor_num
Mean 42.39
StDev 16.06
N 22150
Frequency
0
100
200
300
400
500
600
10
20
30
40
50
60
70
80
distributor_num

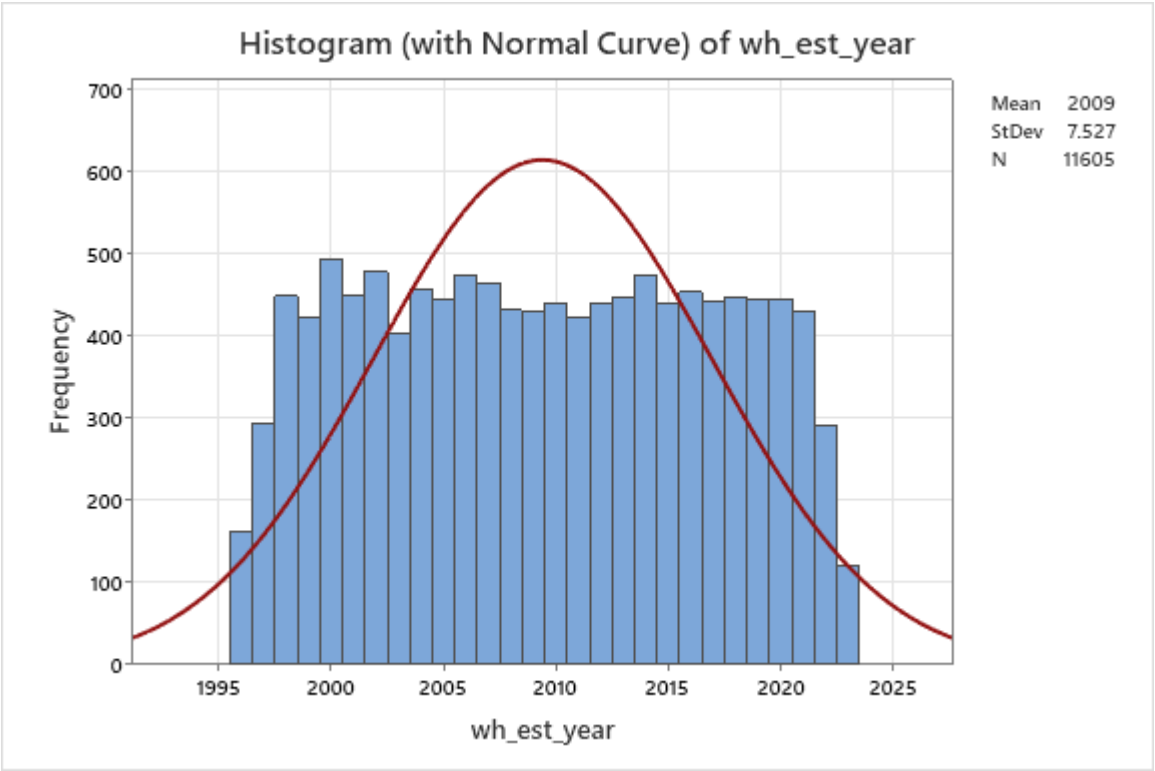

Histogram (with Normal Curve) of flood_impacted
Mean 0.09869
StDev 0.2983
N 22150
Frequency
0
5000
10000
15000
20000
25000
30000
0
1
flood_impacted

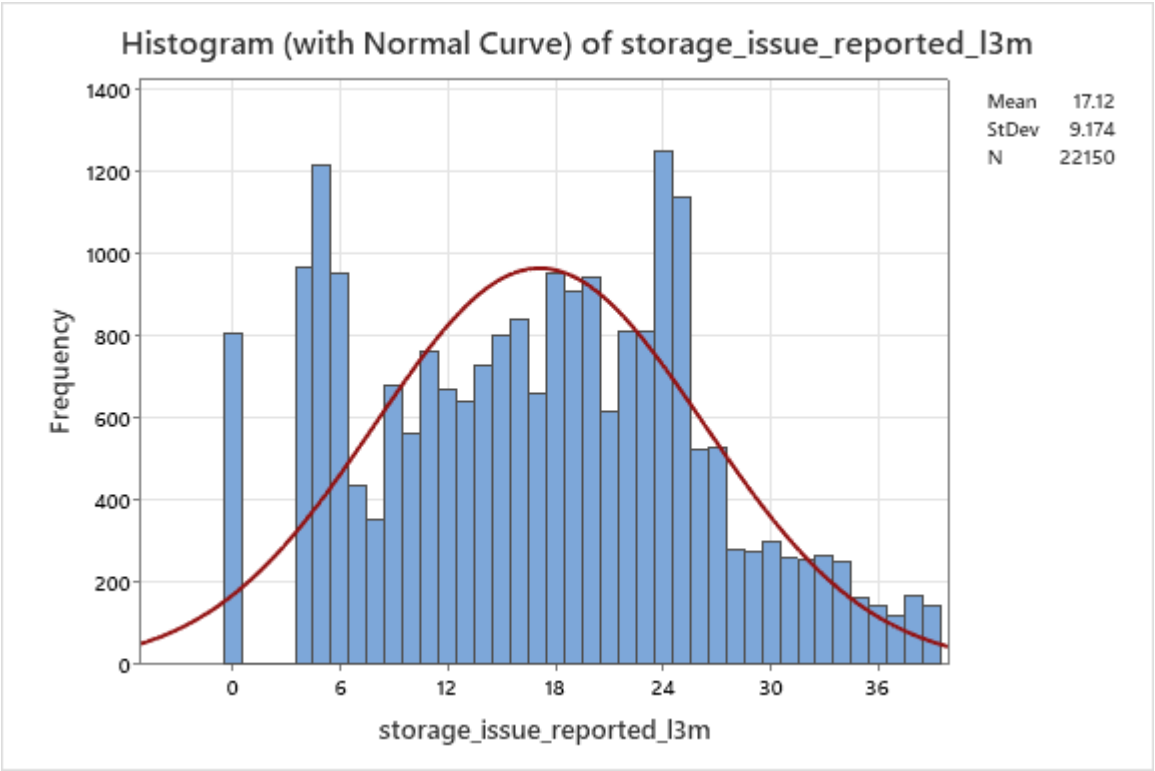

Histogram (with Normal Curve) of flood_proof
Mean 0.05449
StDev 0.2270
N 22150
Frequency
40000
30000
20000
10000
0
0
1
flood_proof

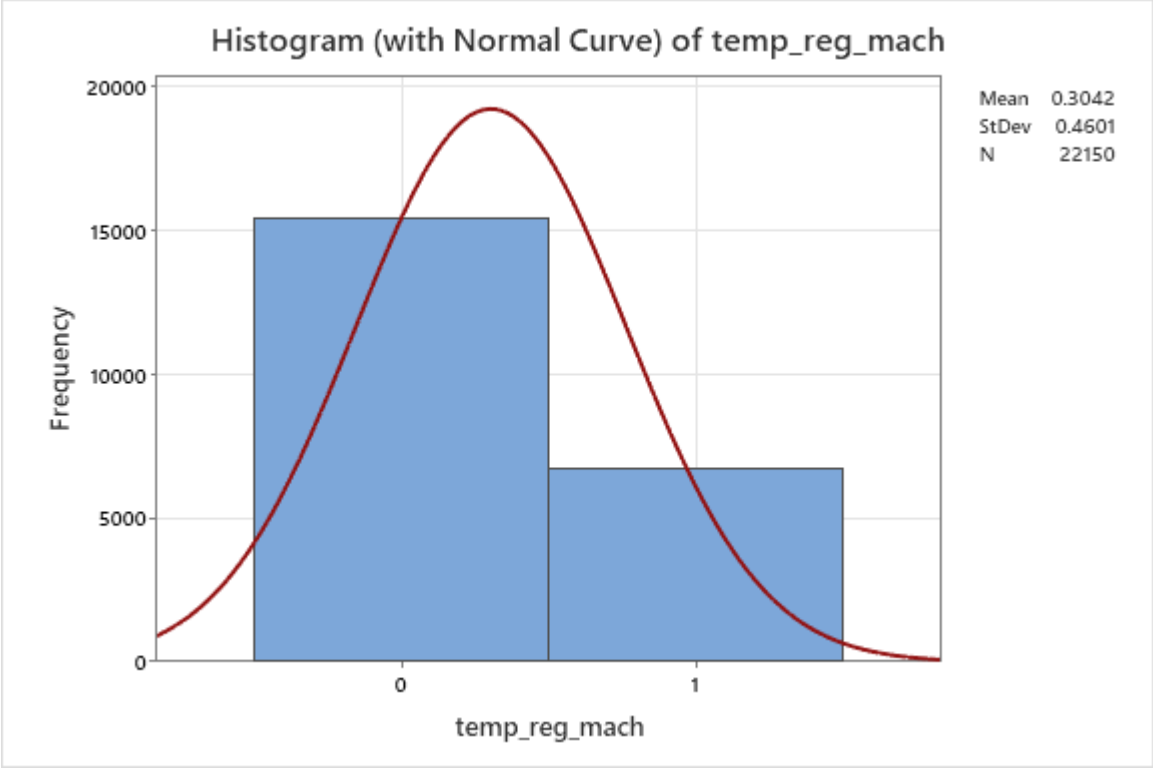

Histogram (with Normal Curve) of electric_supply
Mean 0.6561
StDev 0.4750
N 22150
Frequency
20000
15000
10000
5000
0
0
1
electric_supply

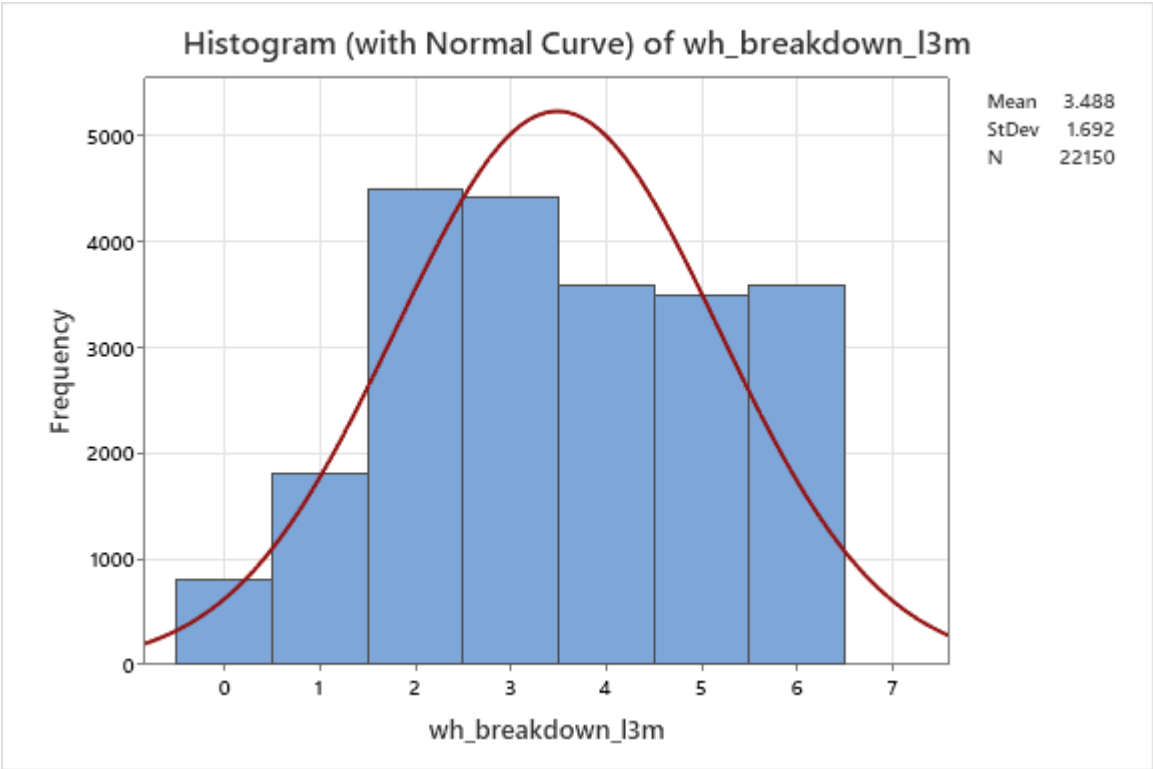

Histogram (with Normal Curve) of dist_from_hub
Mean 163.6
StDev 62.66
N 22150
Frequency
dist_from_hub
0
100
200
300
400
40
80
120
160
200
240
280

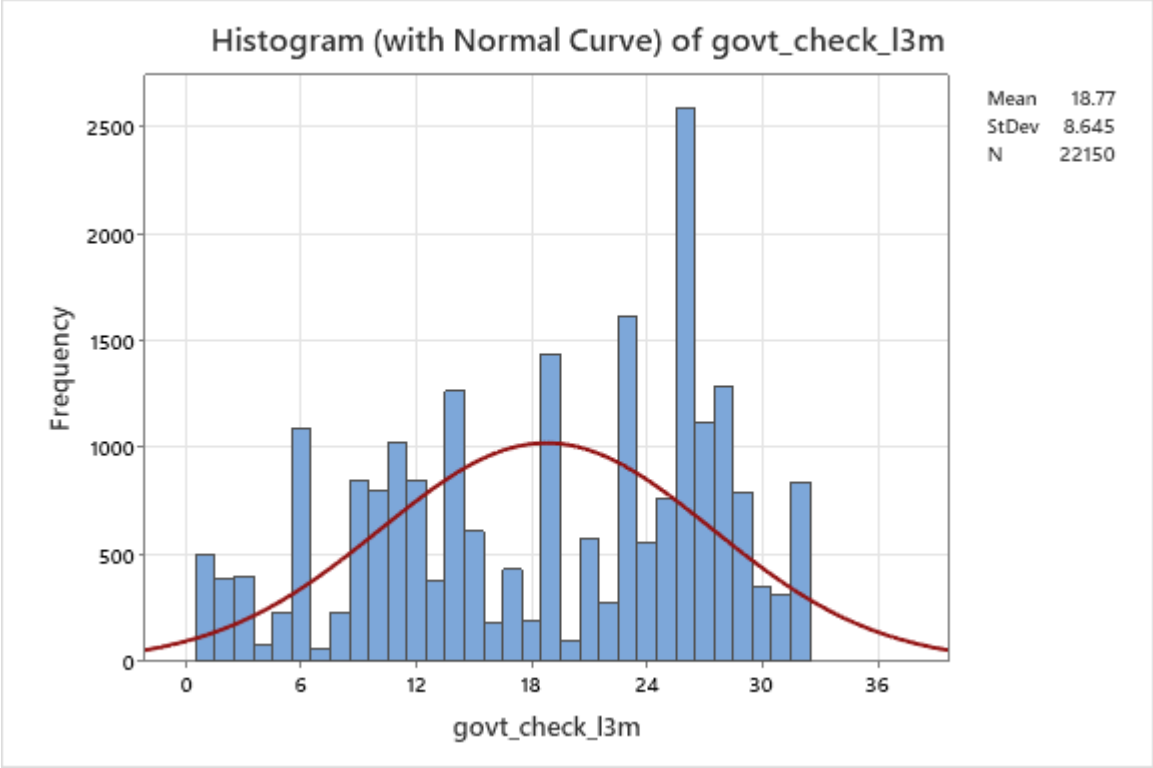

Histogram (with Normal Curve) of workers_num
Mean 28.94
StDev 7.843
N 21273
Frequency
workers_num
0
200
400
600
800
1000
1200
1400
12
24
36
48
60
72
84
96

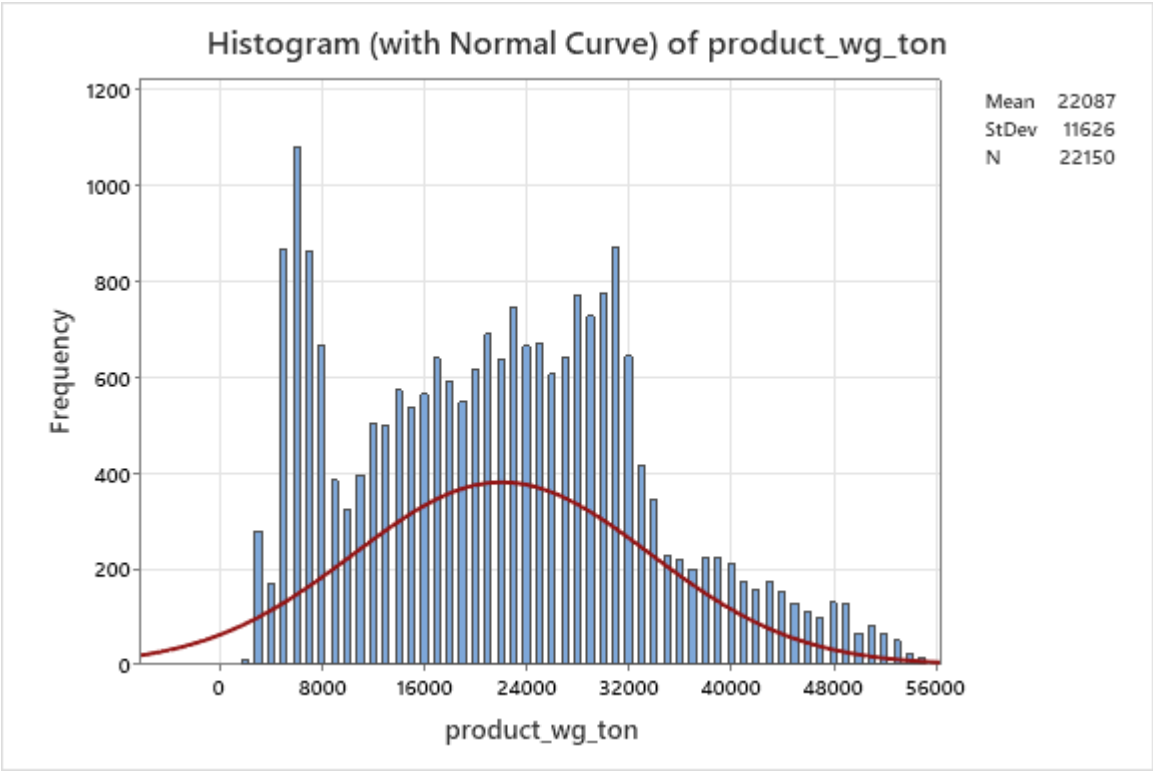

Histogram (with Normal Curve) of wh_est_year
Mean 2009
StDev 7.527
N 11605
Frequency
wh_est_year
0
100
200
300
400
500
600
700
1995
2000
2005
2010
2015
2020
2025

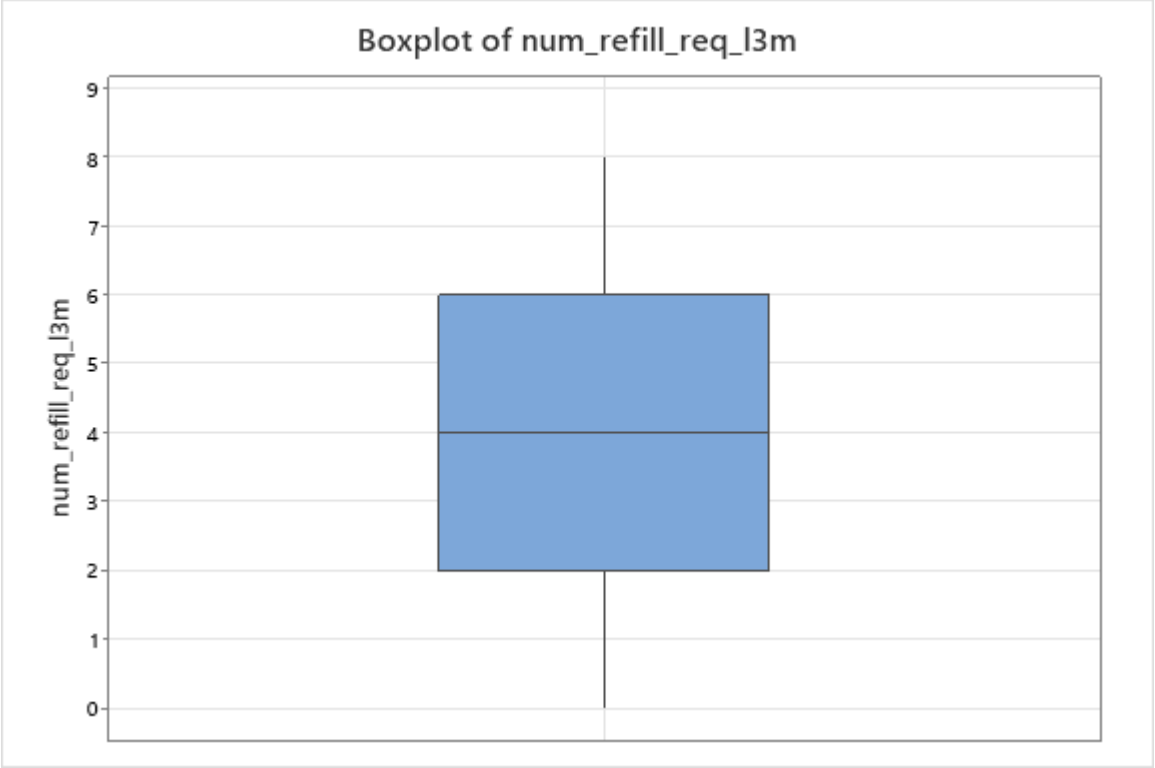

Histogram (with Normal Curve) of storage_issue_reported_l3m
Mean 17.12
StDev 9.174
N 22150
Frequency
storage_issue_reported_l3m
0
200
400
600
800
1000
1200
1400
0
6
12
18
24
30
36

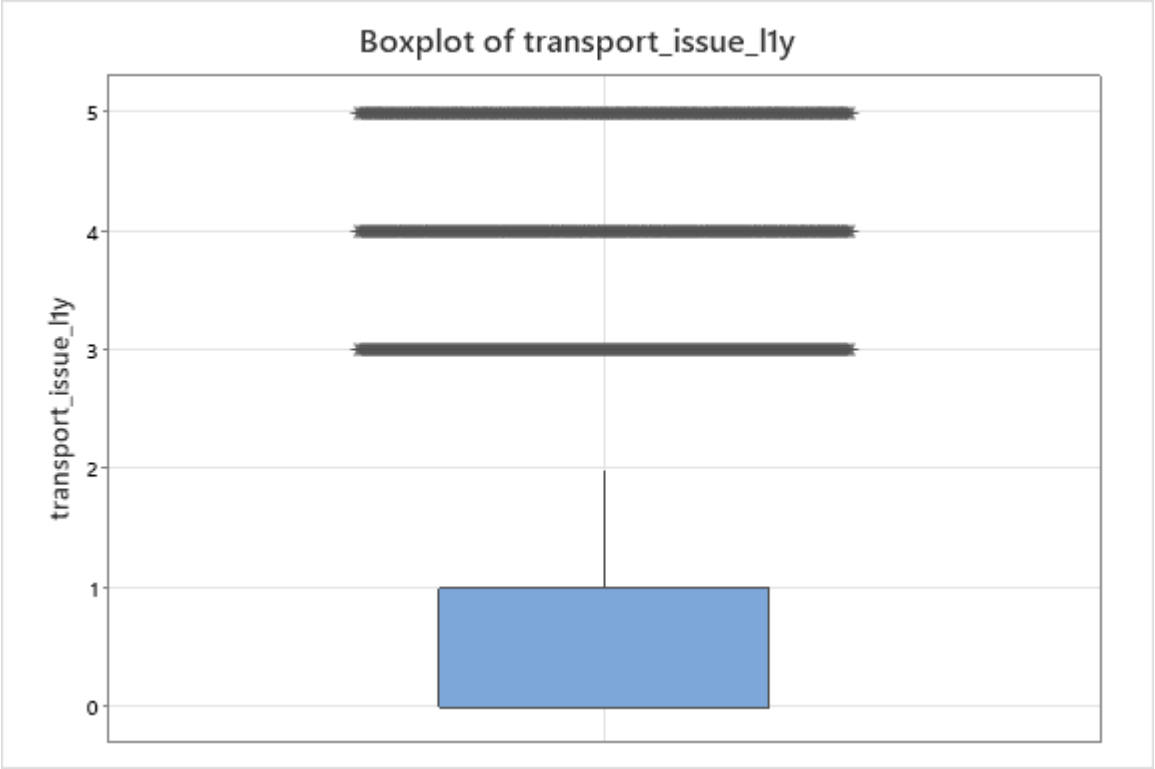

Histogram (with Normal Curve) of temp_reg_mach
Mean 0.3042
StDev 0.4601
N 22150
Frequency
20000
15000
10000
5000
0
0
1
temp_reg_mach

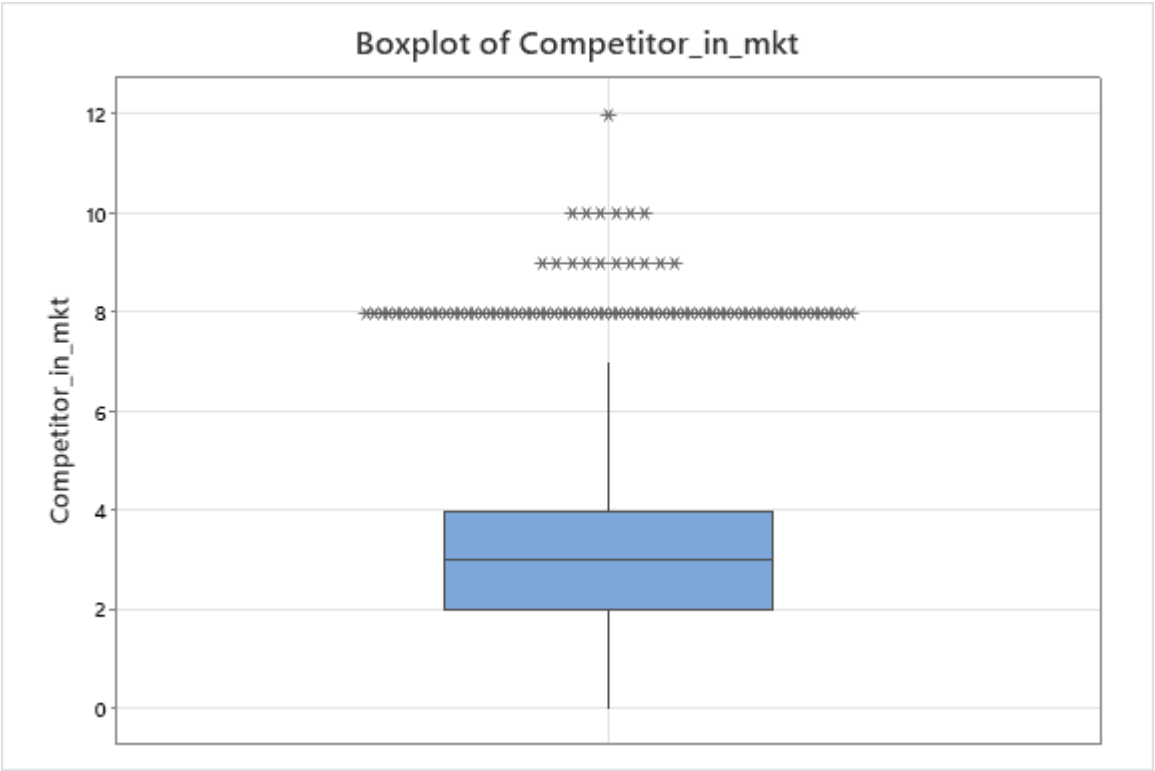

Histogram (with Normal Curve) of wh_breakdown_l3m
Mean 3.488
StDev 1.692
N 22150
Frequency
5000
4000
3000
2000
1000
0
0
1
2
3
4
5
6
7
wh_breakdown_l3m


Histogram (with Normal Curve) of govt_check_l3m
Mean 18.77
StDev 8.645
N 22150
Frequency
govt_check_l3m
0
500
1000
1500
2000
2500
0
6
12
18
24
30
36

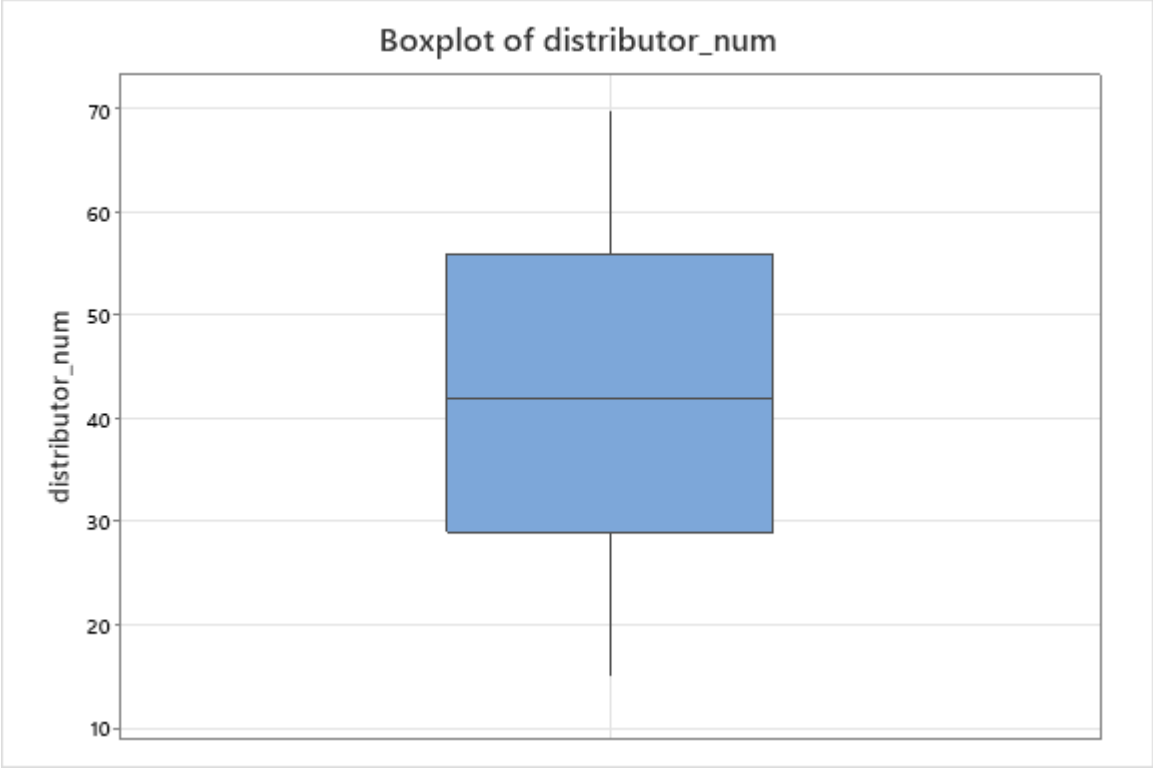

Histogram (with Normal Curve) of product_wg_ton
Mean 22087
StDev 11626
N 22150
Frequency
product_wg_ton
0
200
400
600
800
1000
1200
0
8000
16000
24000
32000
40000
48000
56000

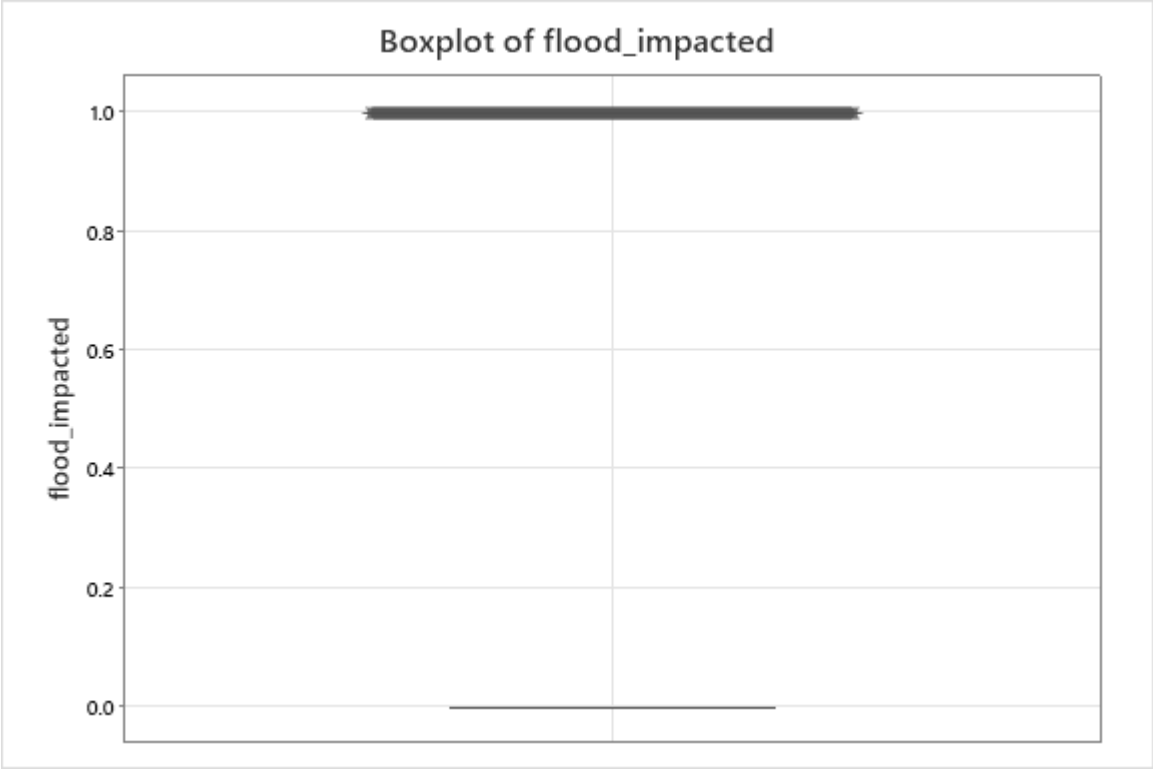

Boxplot of num_refill_req_l3m
num_refill_req_l3m
0
1
2
3
4
5
6
7
8
9

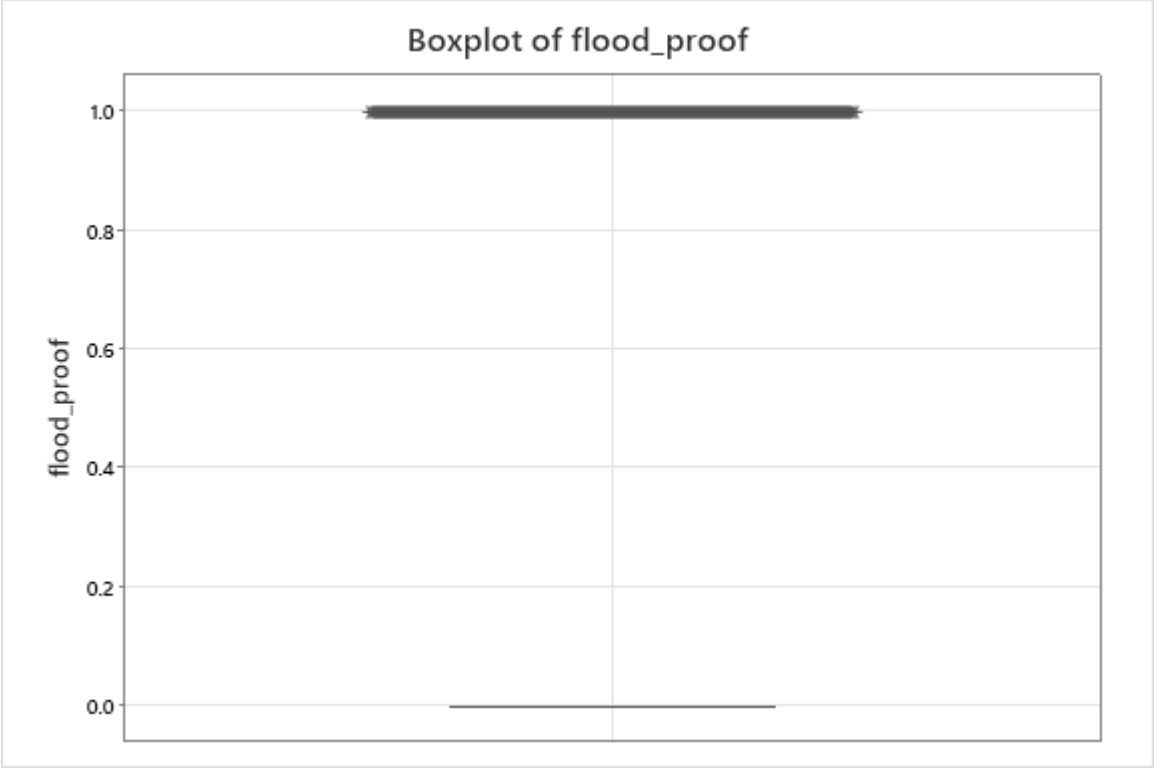

Boxplot of transport_issue_l1y
transport_issue_l1y
0
1
2
3
4
5

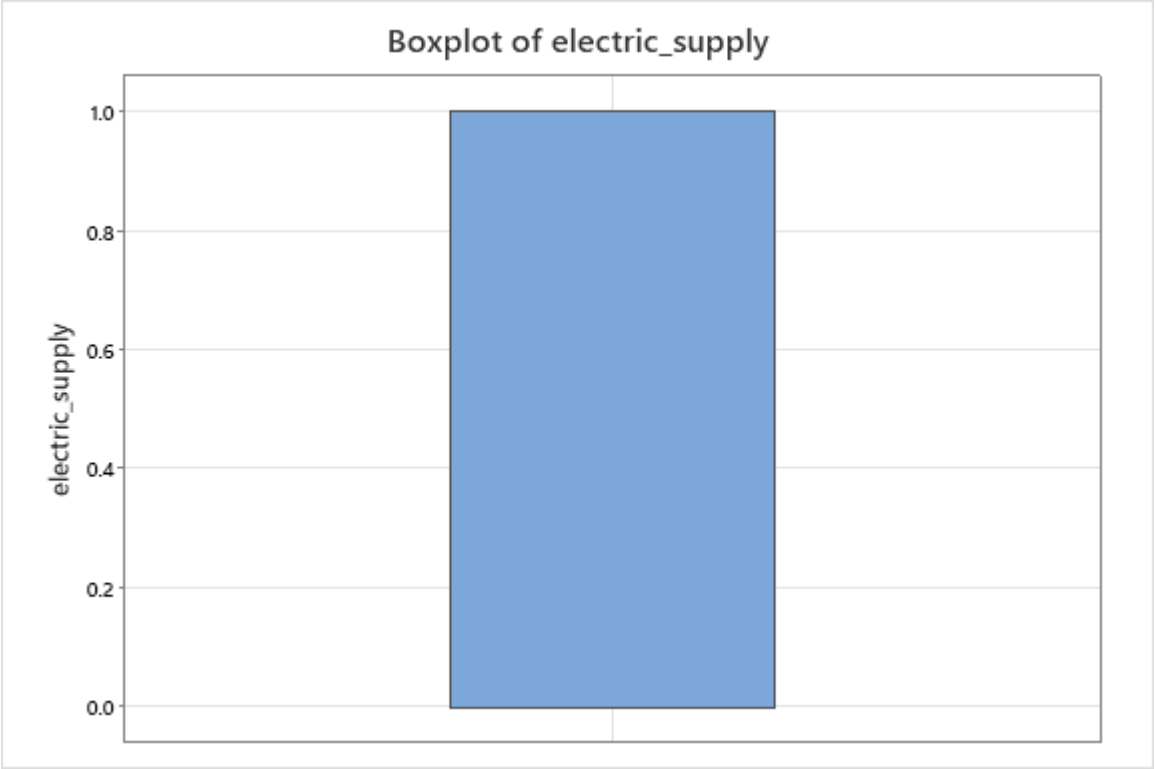

Boxplot of Competitor_in_mkt
Competitor_in_mkt
12
10
8
6
4
2
0

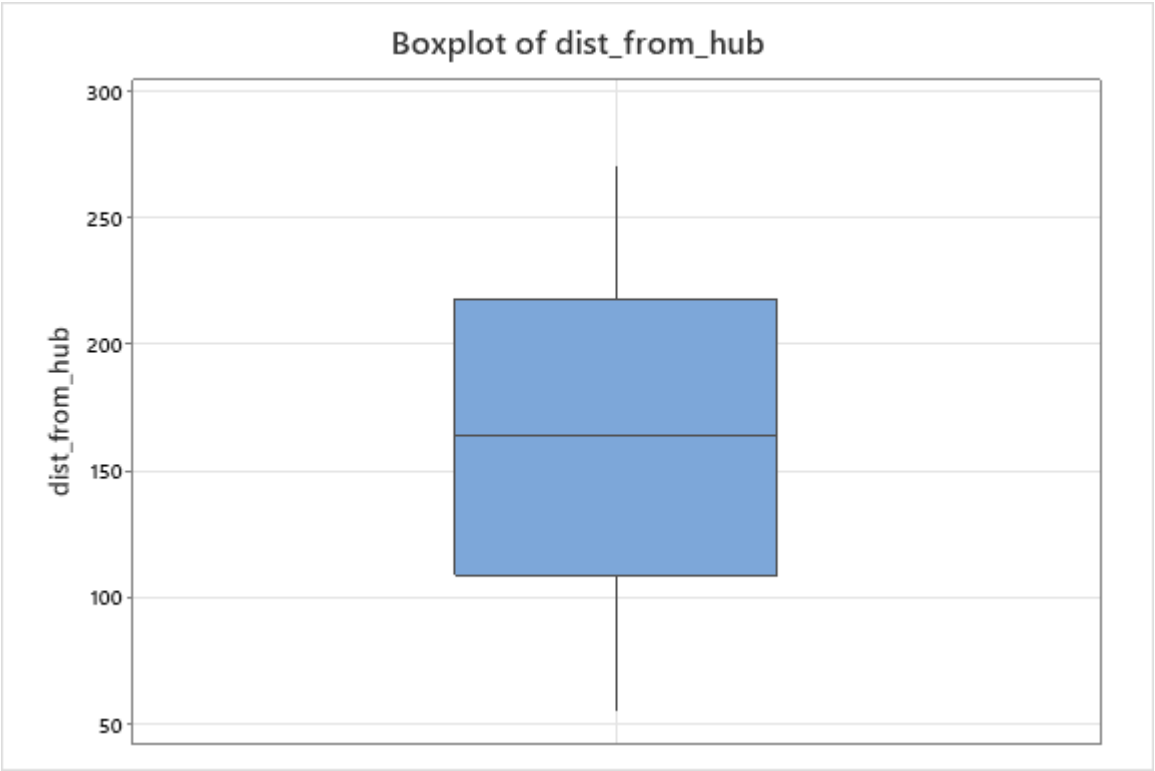
Boxplot of retail_shop_num
retail_shop_num
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000

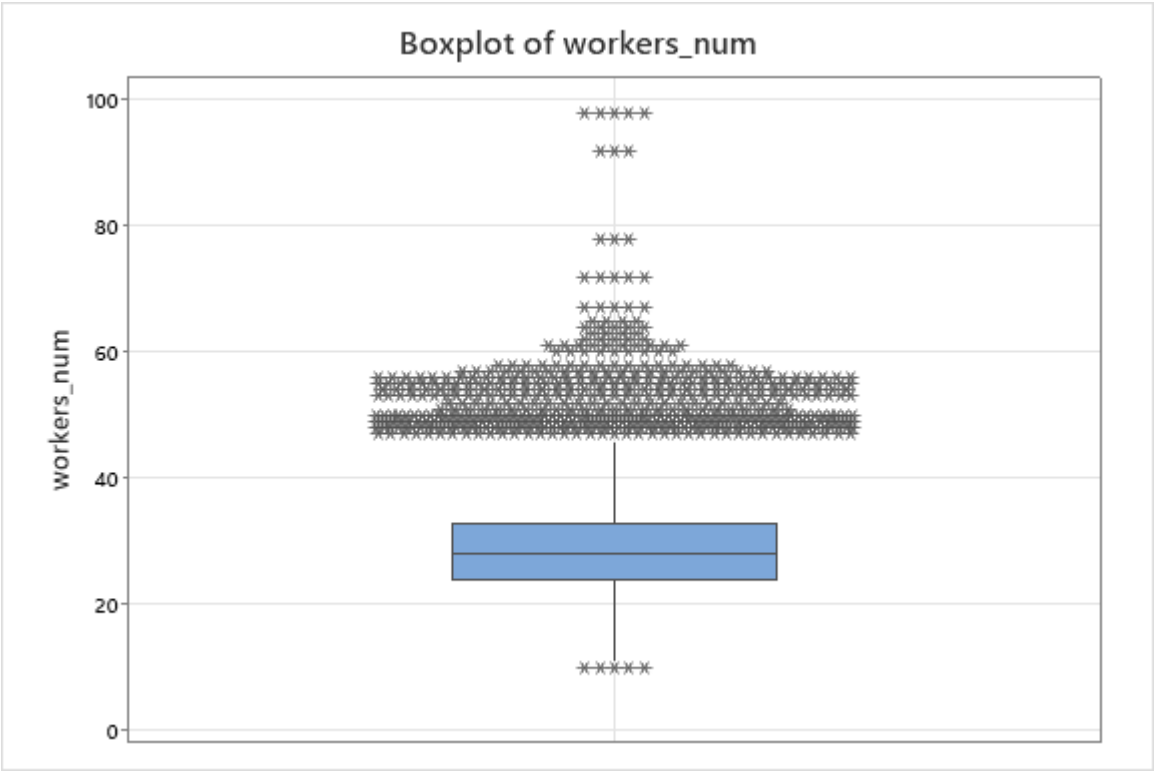

Boxplot of distributor_num
distributor_num
70
60
50
40
30
20
10

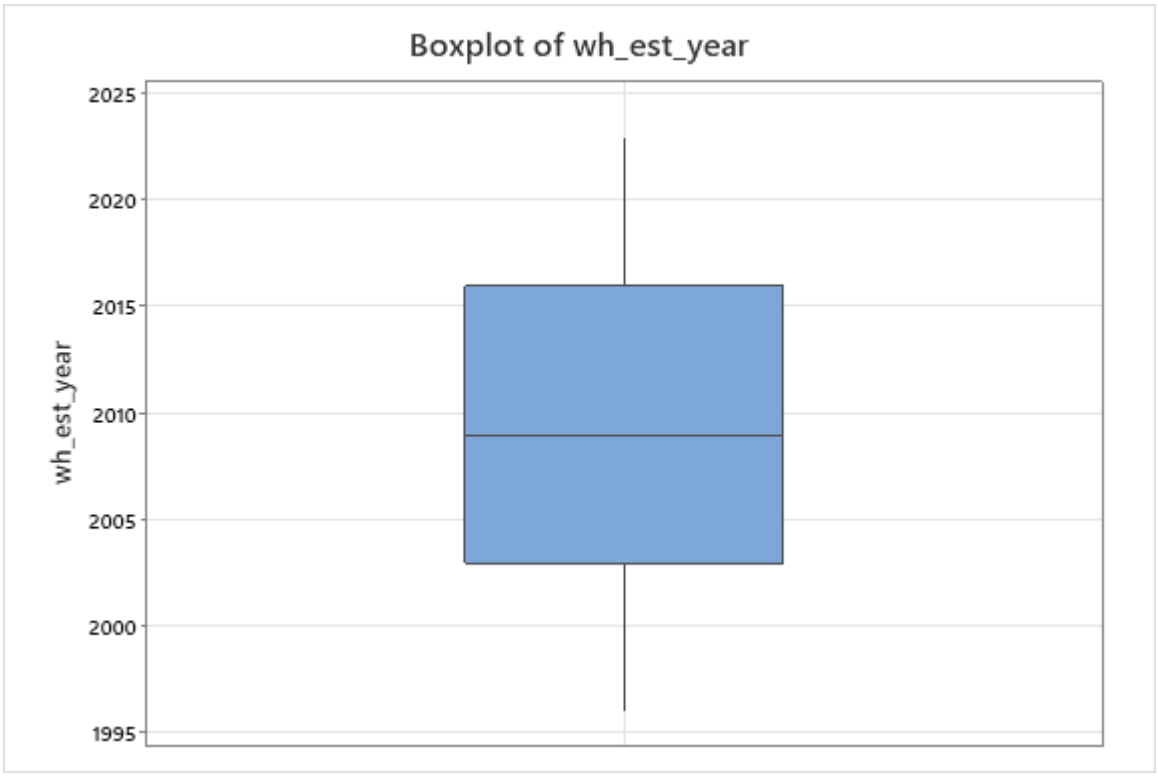

Boxplot of flood_impacted
flood_impacted
1.0
0.8
0.6
0.4
0.2
0.0


Boxplot of flood_proof
flood_proof
1.0
0.8
0.6
0.4
0.2
0.0

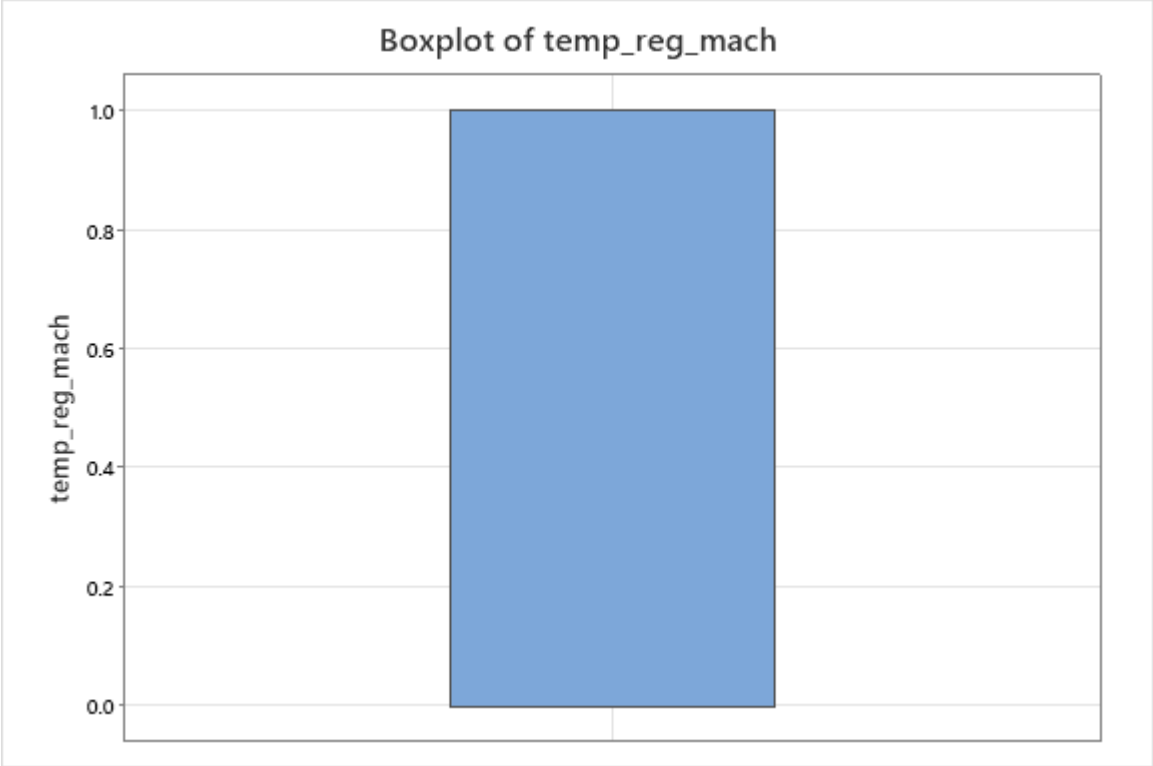

Boxplot of electric_supply
electric_supply
1.0
0.8
0.6
0.4
0.2
0.0

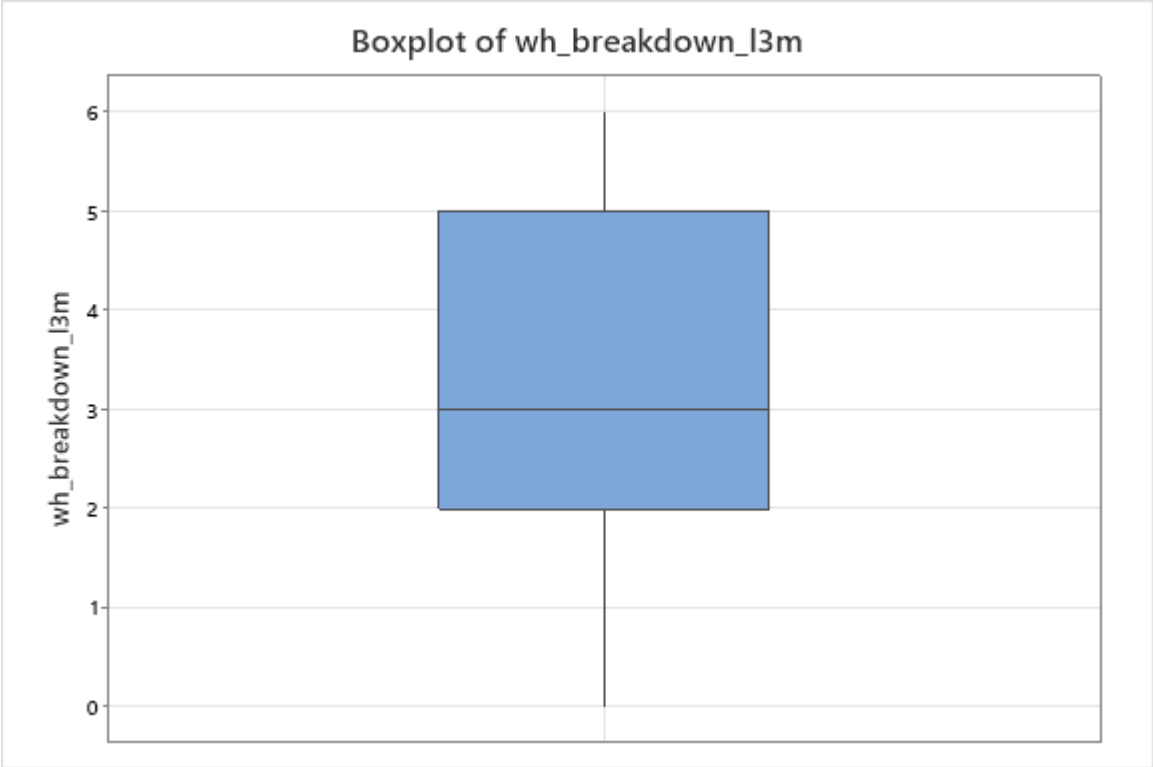

Boxplot of dist_from_hub
dist_from_hub
300
250
200
150
100
50

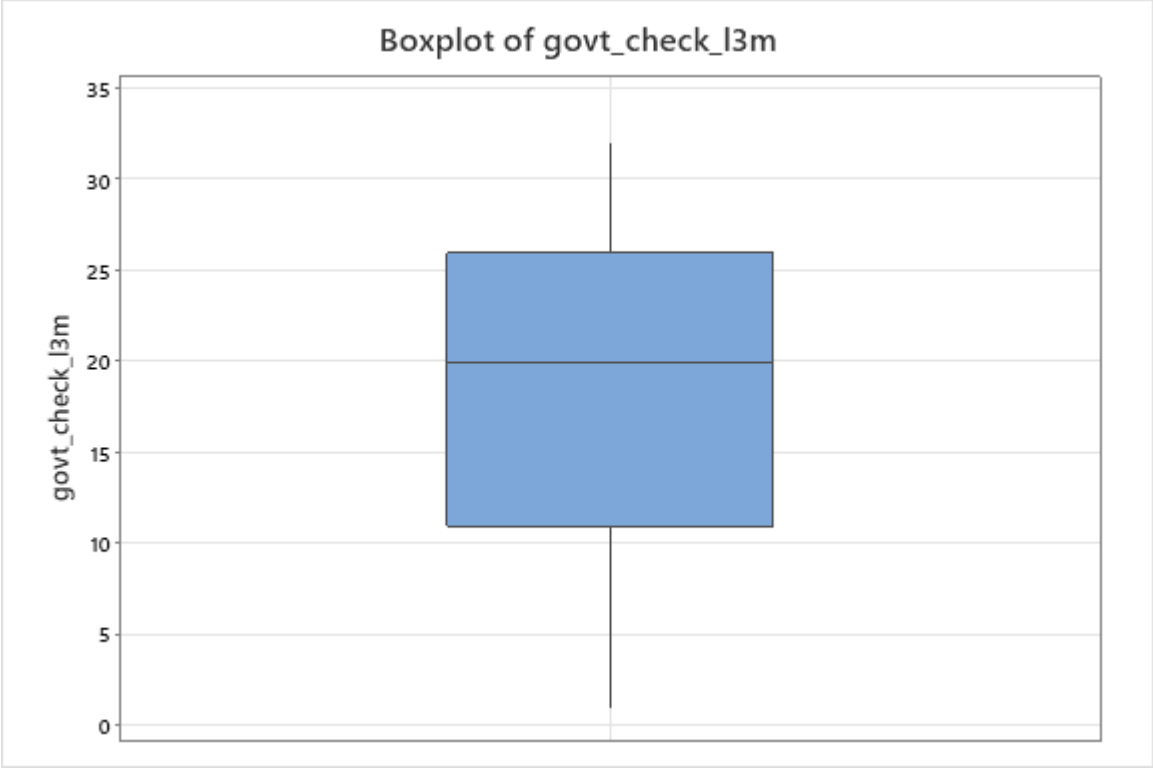

Boxplot of workers_num
workers_num
100
80
60
40
20
0

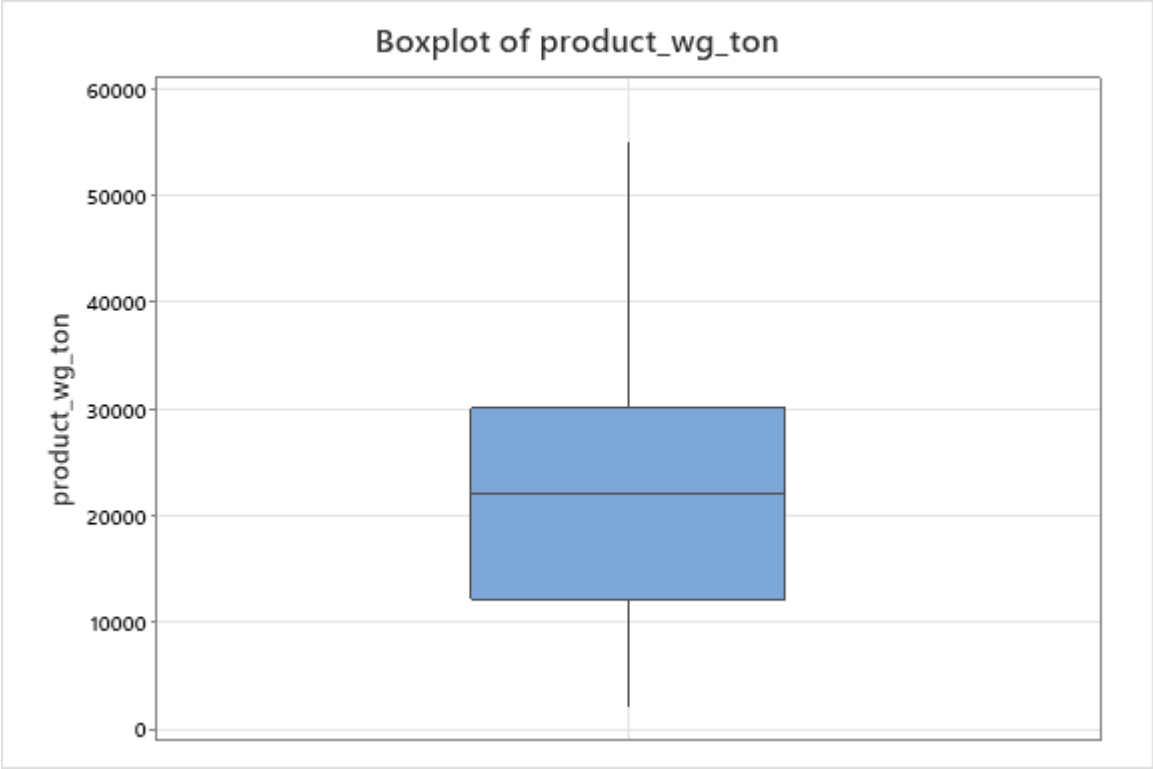

Boxplot of wh_est_year
wh_est_year
2025
2020
2015
2010
2005
2000
1995

Boxplot of storage_issue_reported_l3m
storage_issue_reported_l3m
40
30
20
10
0


Boxplot of temp_reg_mach
temp_reg_mach
1.0
0.8
0.6
0.4
0.2
0.0


Boxplot of wh_breakdown_l3m
wh_breakdown_l3m
6
5
4
3
2
1
0


Boxplot of govt_check_l3m
govt_check_l3m
35
30
25
20
15
10
5
0


Boxplot of product_wg_ton
product_wg_ton
60000
50000
40000
30000
20000
10000
0

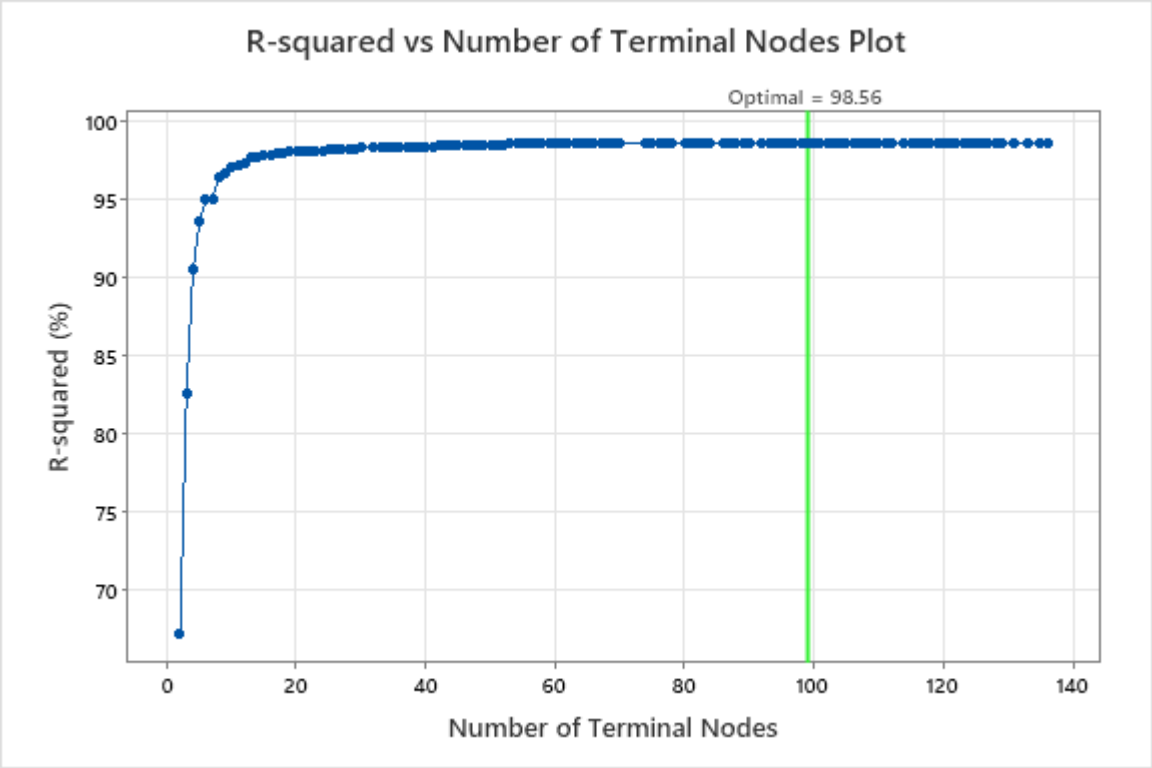
2. CART REGRESSION MODEL:

Method

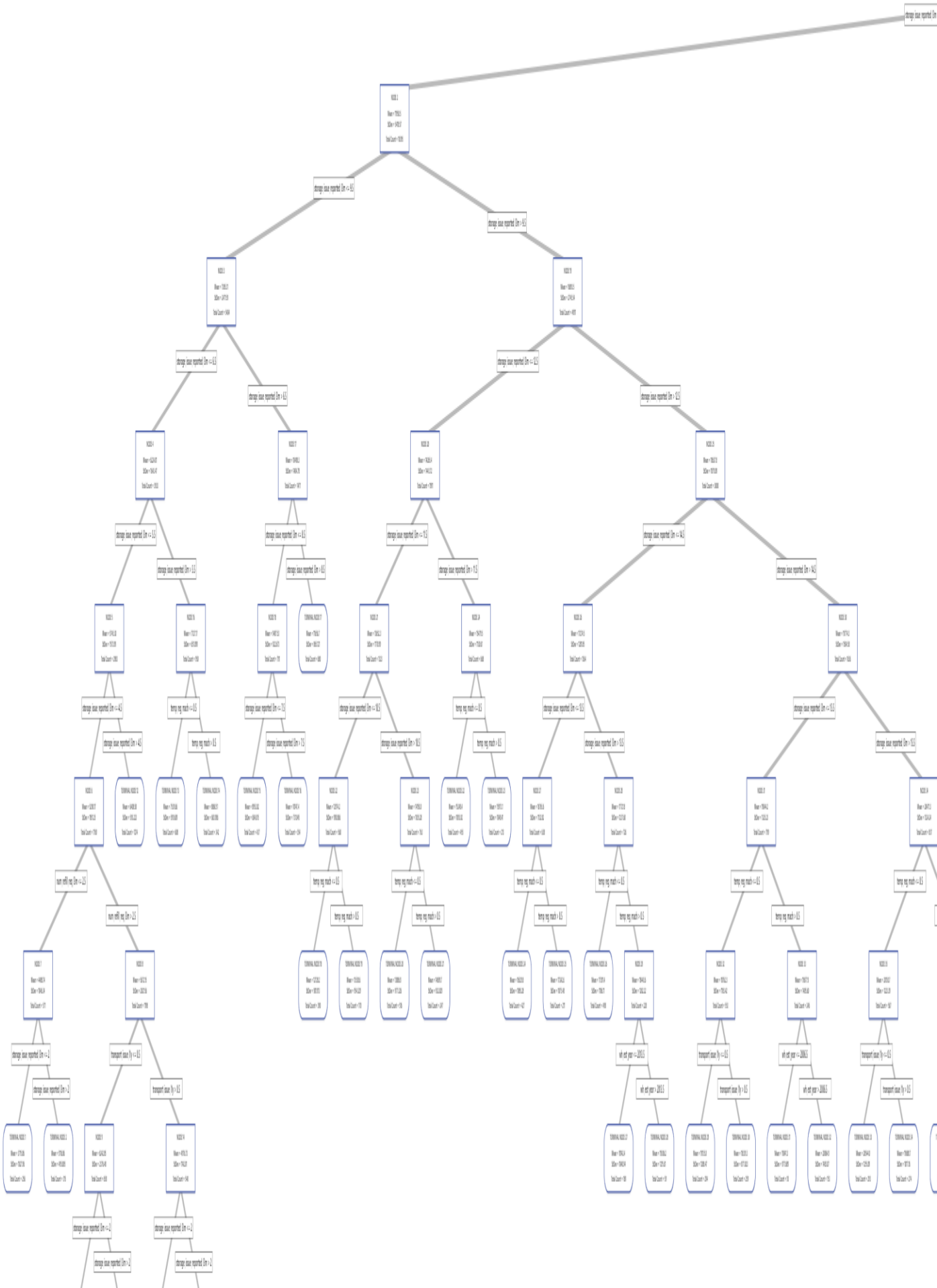
Node splitting	Least squared error
Optimal tree	Within 1 standard error of maximum R-squared
Model validation	10-fold cross-validation
Rows used	22150

Response Information

Mean	StDev	Minimum	Q1	Median	Q3	Maximum
22086.8	11626.2	2065	12151	22099	30102	55151



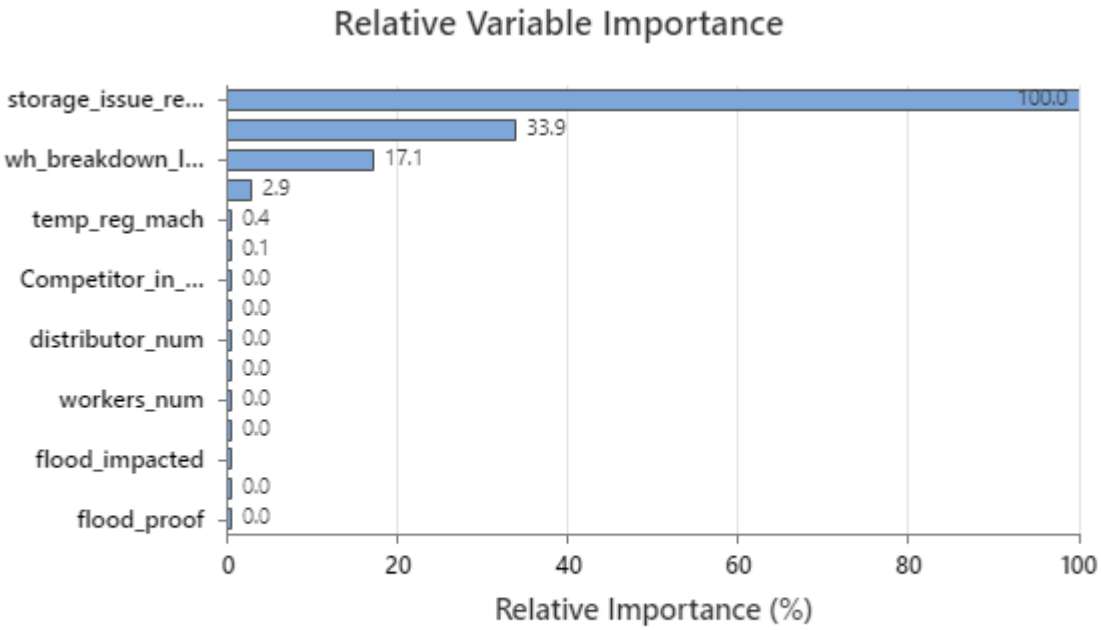
Alternative Tree Diagram


strap: sax reported 1m
X[22]: 1
Mean = 7706.1
SDm = 9181.7
Total Count = 10,315
strap: sax reported 1m <= 8.5
strap: sax reported 1m > 8.5
X[22]: 2
Mean = 7285.21
SDm = 3477.25
Total Count = 9464
X[22]: 19
Mean = 10355.6
SDm = 2741.9
Total Count = 851
strap: sax reported 1m <= 4.5
strap: sax reported 1m > 4.5
strap: sax reported 1m <= 8.5
strap: sax reported 1m > 8.5
X[22]: 4
Mean = 6224.07
SDm = 1641.47
Total Count = 1353
X[22]: 17
Mean = 7498.2
SDm = 1696.71
Total Count = 7611
X[22]: 20
Mean = 9284.4
SDm = 1442.2
Total Count = 791
X[22]: 21
Mean = 7877.6
SDm = 1773.19
Total Count = 1059
strap: sax reported 1m <= 3.5
strap: sax reported 1m > 3.5
strap: sax reported 1m <= 8.5
strap: sax reported 1m > 8.5
strap: sax reported 1m <= 11.5
strap: sax reported 1m > 11.5
strap: sax reported 1m <= 14.5
strap: sax reported 1m > 14.5
X[22]: 5
Mean = 5746.31
SDm = 1113.39
Total Count = 1303
X[22]: 16
Mean = 7727.7
SDm = 1411.09
Total Count = 1150
X[22]: 18
Mean = 9407.63
SDm = 1023.71
Total Count = 771
TERMINAL: X[22]: 17
Mean = 7936.7
SDm = 1661.33
Total Count = 680
X[22]: 23
Mean = 7285.3
SDm = 1778.99
Total Count = 1323
X[22]: 24
Mean = 9479.5
SDm = 1138.61
Total Count = 468
X[22]: 26
Mean = 7239.31
SDm = 1287.85
Total Count = 1184
X[22]: 28
Mean = 9774.3
SDm = 1344.39
Total Count = 1026
strap: sax reported 1m <= 4.5
strap: sax reported 1m > 4.5
temp: eg: mach <= 8.5
temp: eg: mach > 8.5
strap: sax reported 1m <= 7.5
strap: sax reported 1m > 7.5
strap: sax reported 1m <= 10.5
strap: sax reported 1m > 10.5
temp: eg: mach <= 8.5
temp: eg: mach > 8.5
strap: sax reported 1m <= 13.5
strap: sax reported 1m > 13.5
strap: sax reported 1m <= 15.5
strap: sax reported 1m > 15.5
X[22]: 6
Mean = 5238.37
SDm = 991.23
Total Count = 1198
TERMINAL: X[22]: 12
Mean = 6408.18
SDm = 1135.23
Total Count = 1214
TERMINAL: X[22]: 13
Mean = 7535.66
SDm = 1054.69
Total Count = 668
TERMINAL: X[22]: 14
Mean = 8866.31
SDm = 663.99
Total Count = 342
TERMINAL: X[22]: 15
Mean = 8551.66
SDm = 684.09
Total Count = 437
TERMINAL: X[22]: 9
Mean = 9041.4
SDm = 1134.46
Total Count = 334
X[22]: 22
Mean = 7294.2
SDm = 1186.86
Total Count = 1181
X[22]: 31
Mean = 9478.3
SDm = 1015.28
Total Count = 761
TERMINAL: X[22]: 32
Mean = 7534.84
SDm = 1005.42
Total Count = 443
TERMINAL: X[22]: 33
Mean = 9307.11
SDm = 1048.41
Total Count = 318
X[22]: 27
Mean = 8765.3
SDm = 1152.65
Total Count = 668
X[22]: 29
Mean = 7727.23
SDm = 1237.68
Total Count = 516
X[22]: 30
Mean = 7784.42
SDm = 1326.21
Total Count = 791
X[22]: 34
Mean = 10747.2
SDm = 1324.24
Total Count = 657
run: mfl: eg: 1m <= 2.2
run: mfl: eg: 1m > 2.5
temp: eg: mach <= 8.5
temp: eg: mach > 8.5
temp: eg: mach <= 8.5
temp: eg: mach > 8.5
temp: eg: mach <= 8.5
temp: eg: mach > 8.5
temp: eg: mach <= 8.5
temp: eg: mach > 8.5
temp: eg: mach <= 8.5
temp: eg: mach > 8.5
temp: eg: mach <= 8.5
temp: eg: mach > 8.5
X[22]: 7
Mean = 4481.9
SDm = 1146.31
Total Count = 177
X[22]: 8
Mean = 5812.9
SDm = 1207.16
Total Count = 1199
TERMINAL: X[22]: 11
Mean = 6772.61
SDm = 1017.11
Total Count = 1118
TERMINAL: X[22]: 10
Mean = 8181.2
SDm = 1134.46
Total Count = 1159
TERMINAL: X[22]: 35
Mean = 7139.5
SDm = 1146.71
Total Count = 668
TERMINAL: X[22]: 36
Mean = 7974.43
SDm = 1171.49
Total Count = 279
TERMINAL: X[22]: 37
Mean = 9128.59
SDm = 1095.28
Total Count = 427
TERMINAL: X[22]: 38
Mean = 9154.2
SDm = 1071.41
Total Count = 241
X[22]: 25
Mean = 7727.23
SDm = 1237.68
Total Count = 516
X[22]: 32
Mean = 7944.16
SDm = 1265.22
Total Count = 228
X[22]: 35
Mean = 7796.2
SDm = 1182.45
Total Count = 563
X[22]: 36
Mean = 7967.9
SDm = 1401.68
Total Count = 246
X[22]: 37
Mean = 10674.3
SDm = 1253.39
Total Count = 667
vh: est: year <= 2003.5
vh: est: year > 2003.5
transport: sae: fly <= 0.5
transport: sae: fly > 0.5
vh: est: year <= 2006.5
vh: est: year > 2006.5
transport: sae: fly <= 0.5
transport: sae: fly > 0.5
TERMINAL: X[22]: 1
Mean = 3776.06
SDm = 1567.06
Total Count = 236
TERMINAL: X[22]: 2
Mean = 4796.06
SDm = 1414.85
Total Count = 1135
X[22]: 3
Mean = 6242.95
SDm = 1274.45
Total Count = 1169
X[22]: 13
Mean = 4775.71
SDm = 1042.91
Total Count = 1146
TERMINAL: X[22]: 27
Mean = 8746.4
SDm = 1040.04
Total Count = 181
TERMINAL: X[22]: 28
Mean = 7938.2
SDm = 1295.01
Total Count = 131
TERMINAL: X[22]: 29
Mean = 7733.3
SDm = 1266.41
Total Count = 254
TERMINAL: X[22]: 31
Mean = 8530.1
SDm = 1072.82
Total Count = 239
TERMINAL: X[22]: 39
Mean = 7849.1
SDm = 1171.69
Total Count = 118
TERMINAL: X[22]: 40
Mean = 8584.6
SDm = 1430.07
Total Count = 112
TERMINAL: X[22]: 41
Mean = 8744.3
SDm = 1254.39
Total Count = 203
TERMINAL: X[22]: 42
Mean = 9168.7
SDm = 1071.91
Total Count = 474
strap: sax reported 1m <= 2
strap: sax reported 1m > 2
strap: sax reported 1m <= 2
strap: sax reported 1m > 2

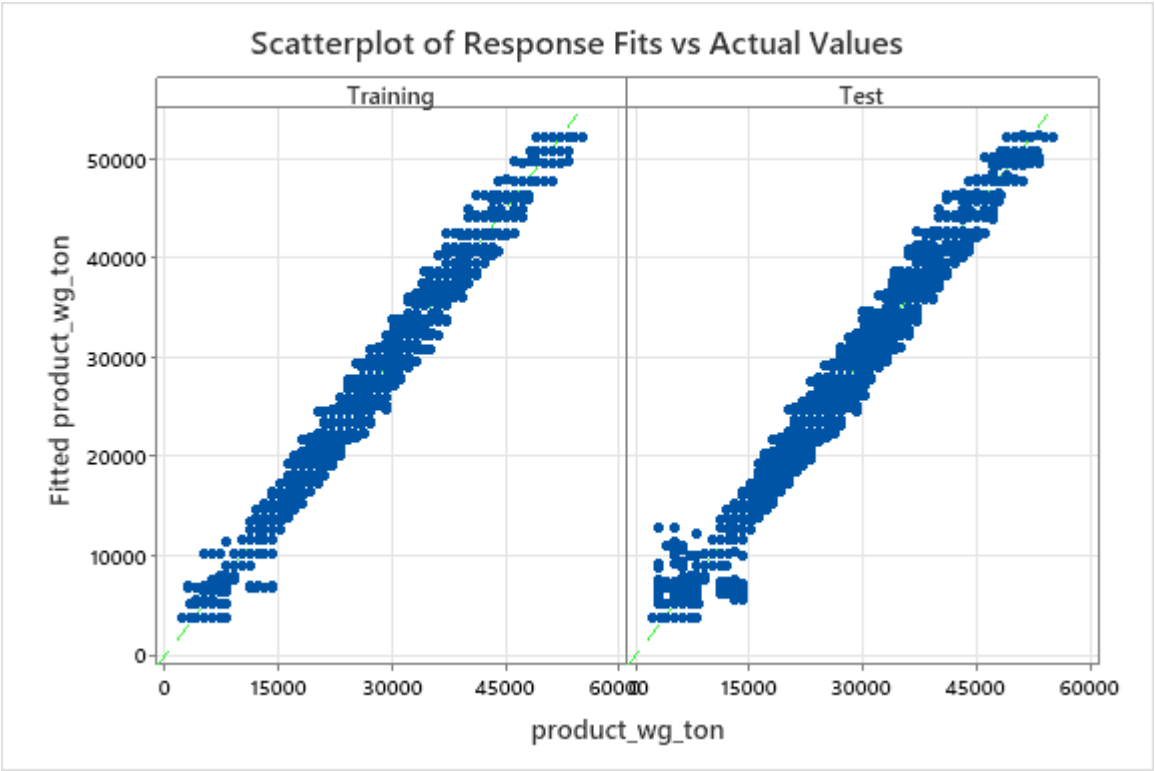
Model Summary

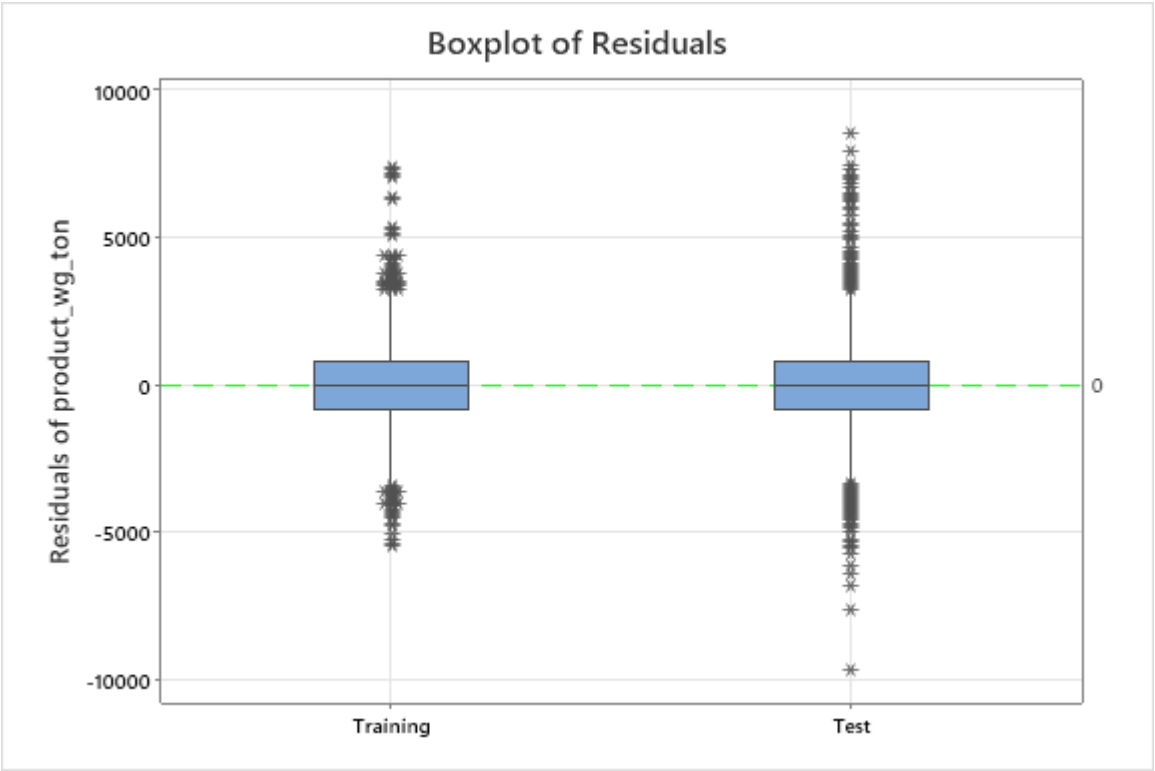
Total predictors	15
Important predictors	15
Number of terminal nodes	99
Minimum terminal node size	5

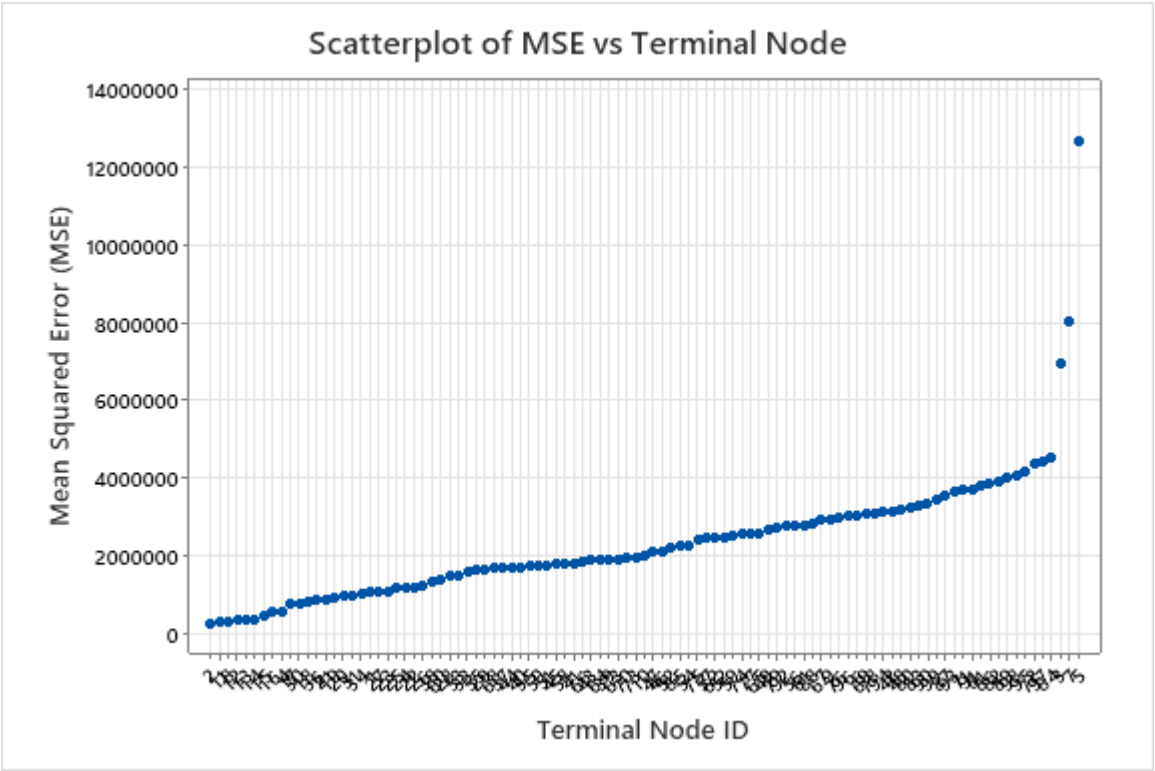
Statistics	Training	Test
R-squared	98.63%	98.56%
Root mean squared error (RMSE)	1360.1370	1396.7352
Mean squared error (MSE)	1.84997E+06	1.95087E+06
Mean absolute deviation (MAD)	1052.8104	1077.0763
Mean absolute percent error (MAPE)	0.0593	0.0611

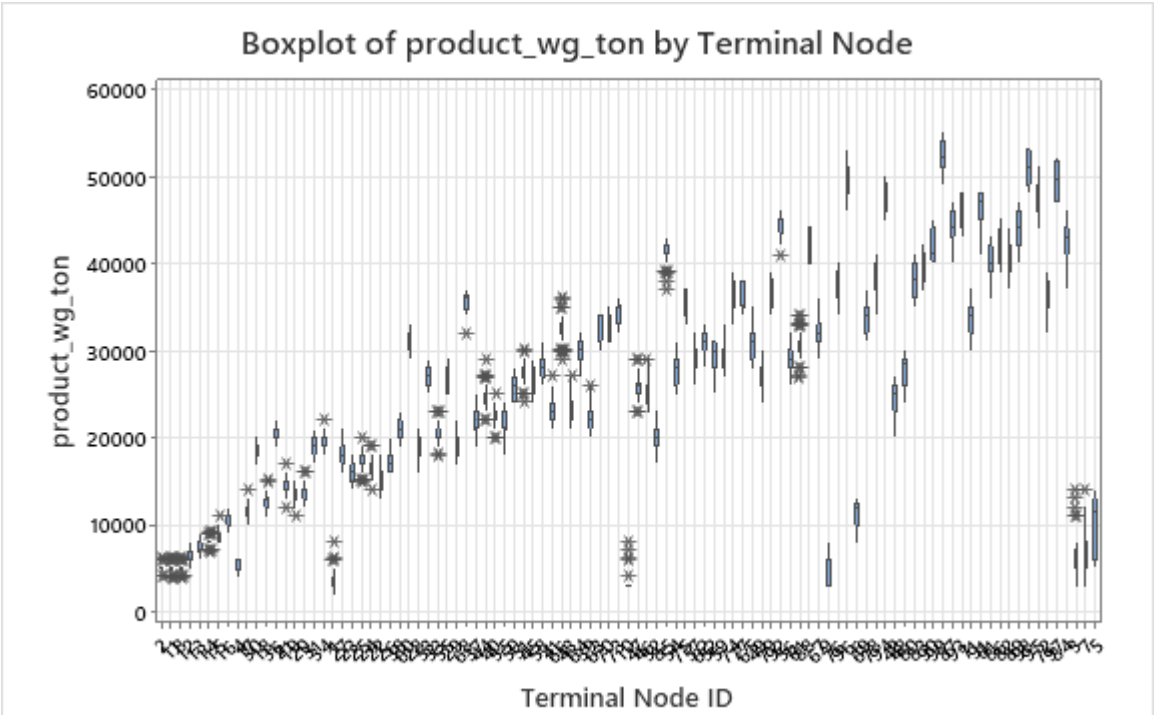


Variable importance measures model improvement when splits are made on a predictor. Relative importance is defined as % improvement with respect to the top predictor.

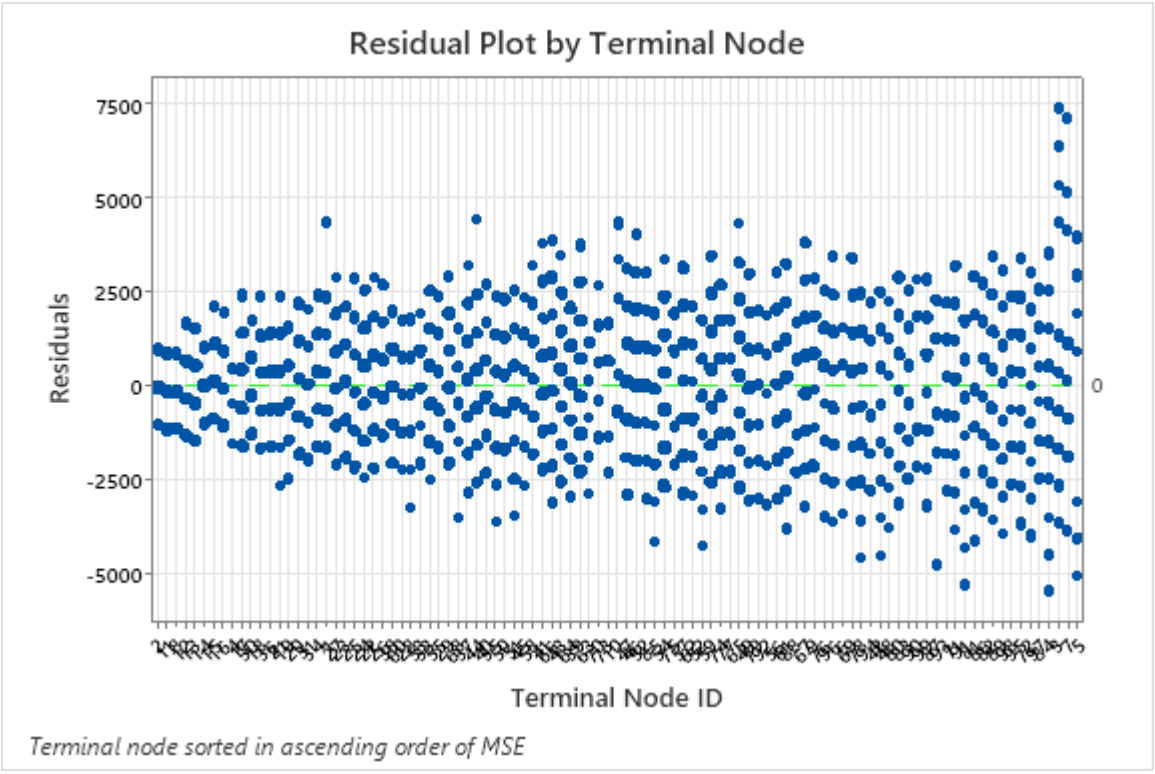

Scatterplot of Response Fits vs Actual Values
Training
Test
Fitted product_wg_ton
50000
40000
30000
20000
10000
0
0
15000
30000
45000
60000
product_wg_ton


Boxplot of Residuals
Residuals of product_wg_ton
10000
5000
0
-5000
-10000
0
Training
Test





Terminal node sorted in ascending order of MSE



Terminal node sorted in ascending order of MSE