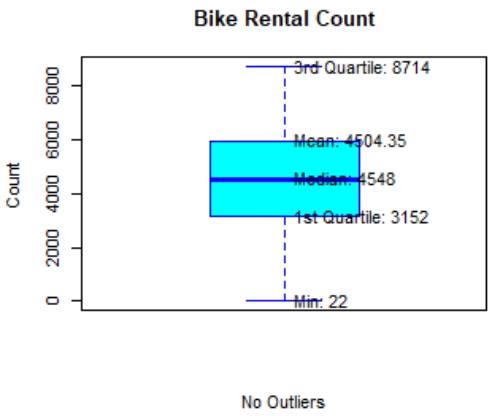
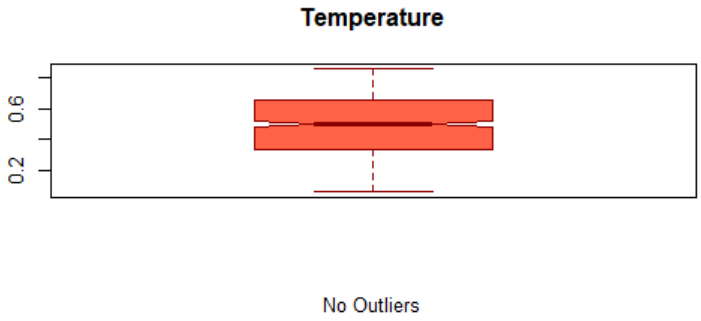
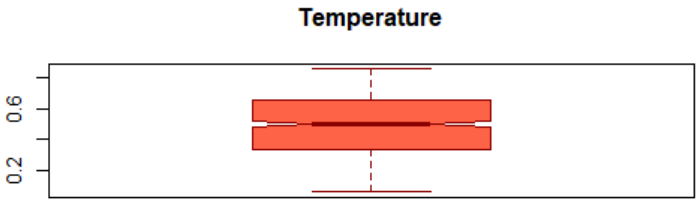
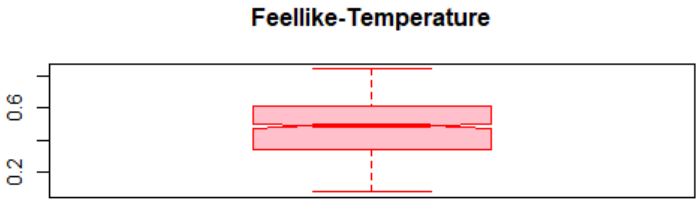

Bike Rental Count
Count
0
2000
4000
6000
8000
3rd Quartile: 8714
Mean: 4504.35
Median: 4548
1st Quartile: 3152
Min: 22
No Outliers



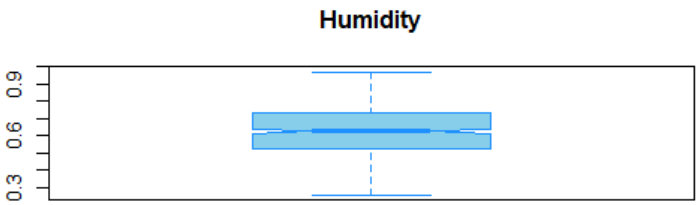
No Outliers



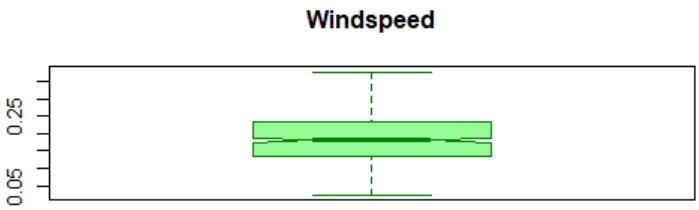
No Outliers



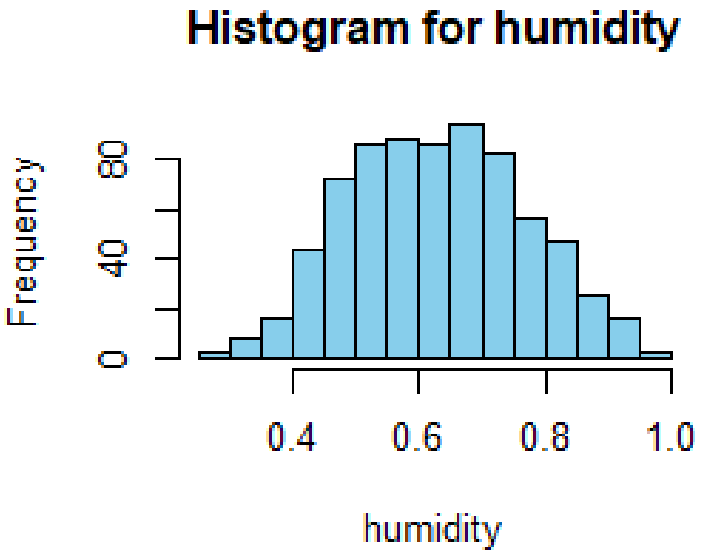
No Outliers

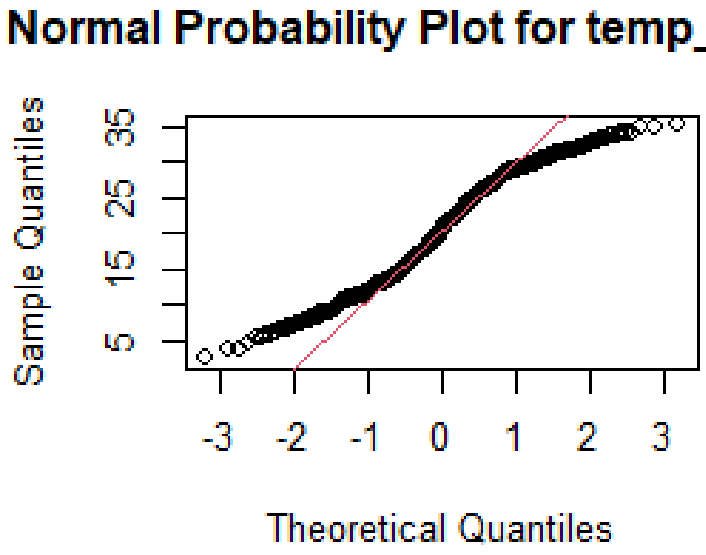


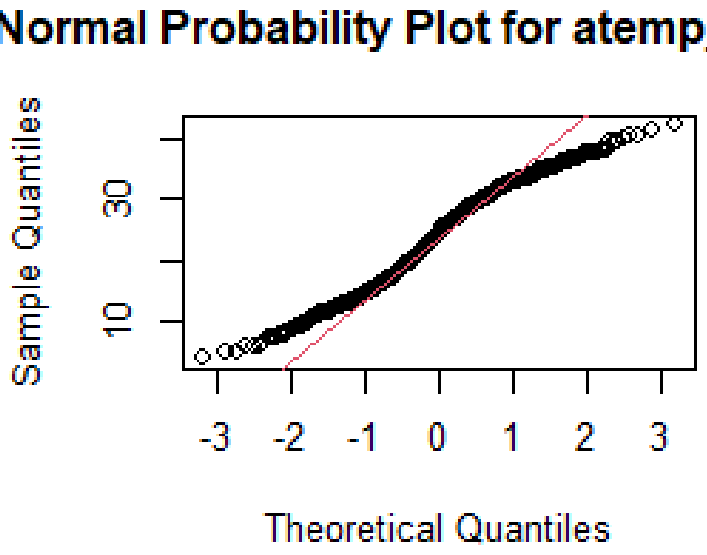
Outliers: 0.187917

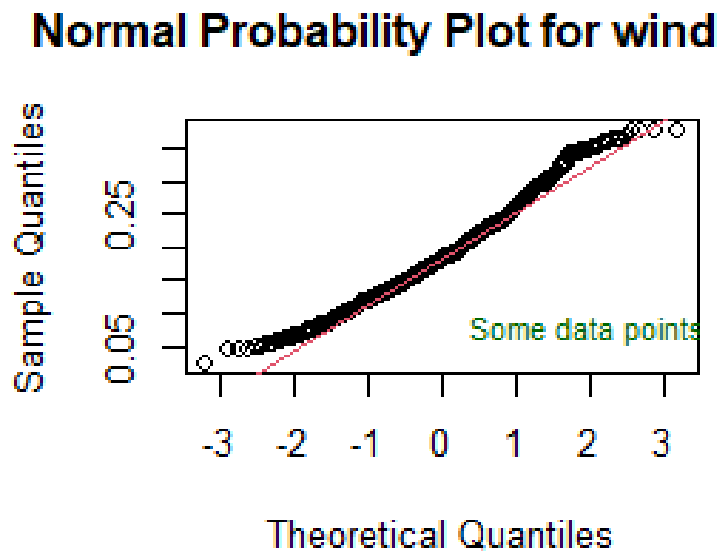


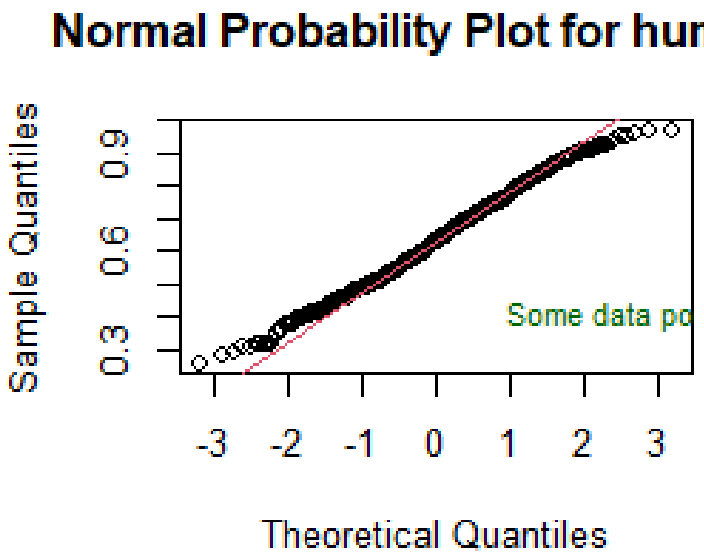
Outliers: 0.417908


Histogram for humidity
Frequency
0
40
80
0.4
0.6
0.8
1.0
humidity

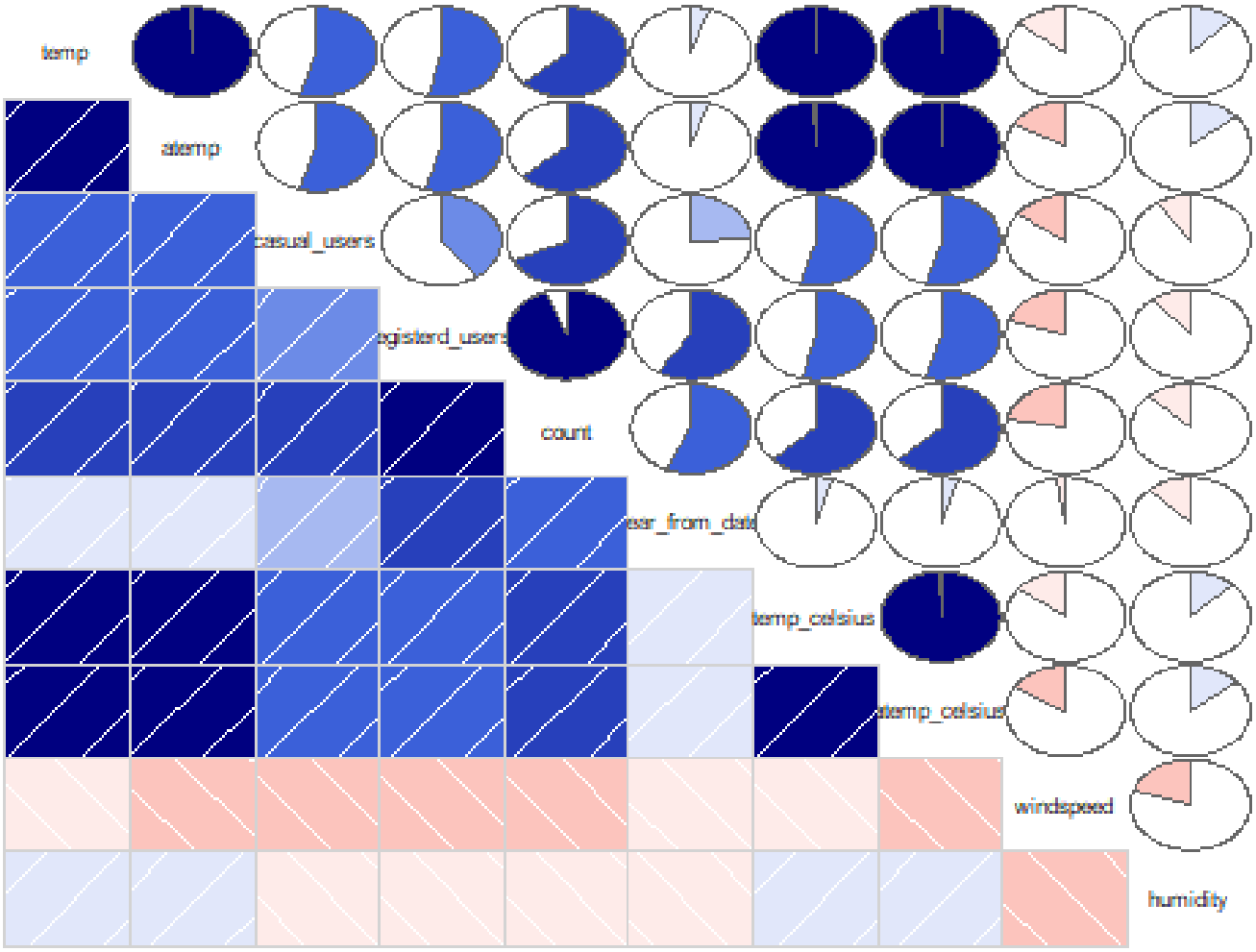

Normal Probability Plot for temp_cel
Sample Quantiles
5
15
25
35
-3
-2
-1
0
1
2
3
Theoretical Quantiles


Normal Probability Plot for atemp_cel
Sample Quantiles
10
30
-3
-2
-1
0
1
2
3
Theoretical Quantiles

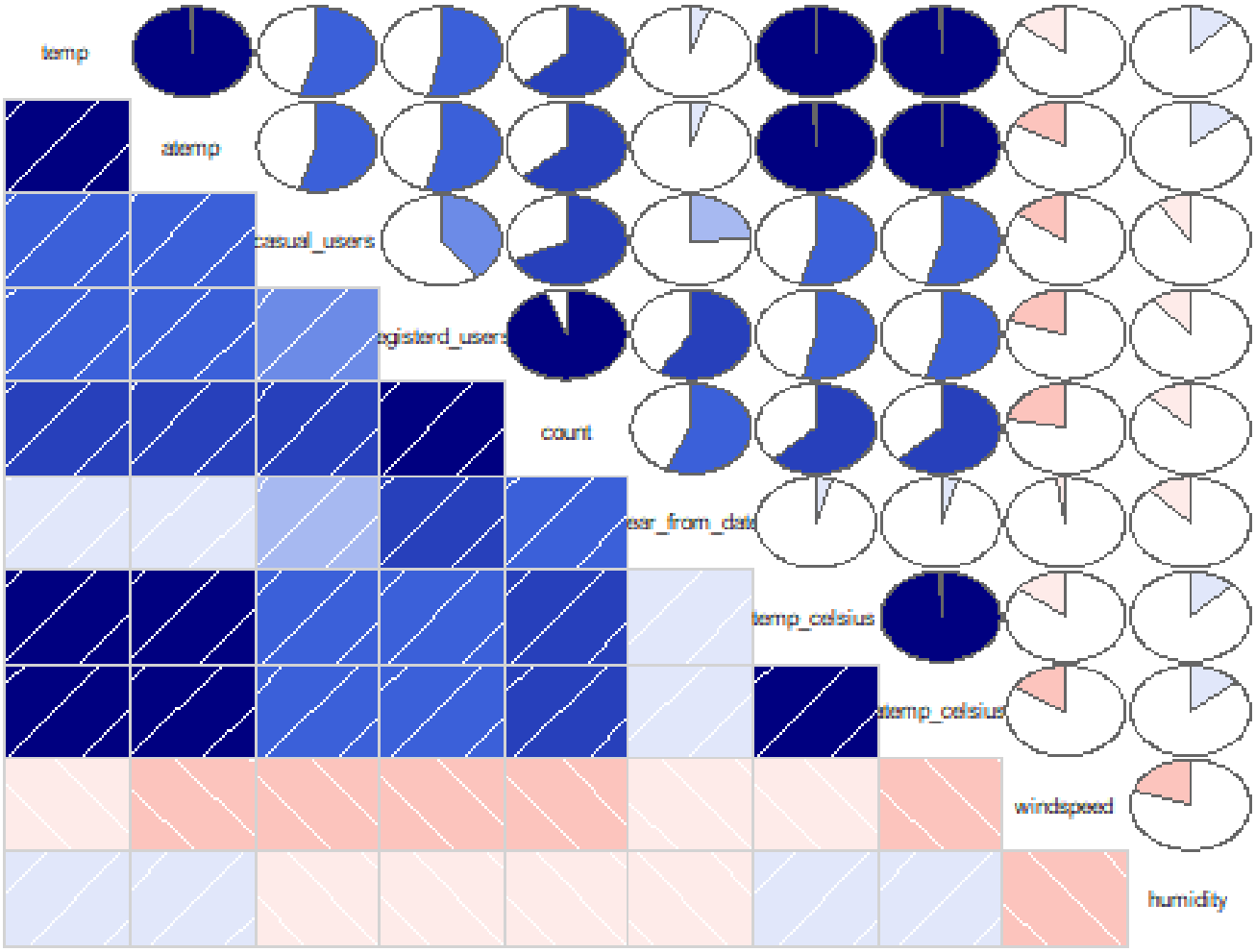

Normal Probability Plot for windspe
Sample Quantiles
0.05
0.25
Some data points
-3
-2
-1
0
1
2
3
Theoretical Quantiles

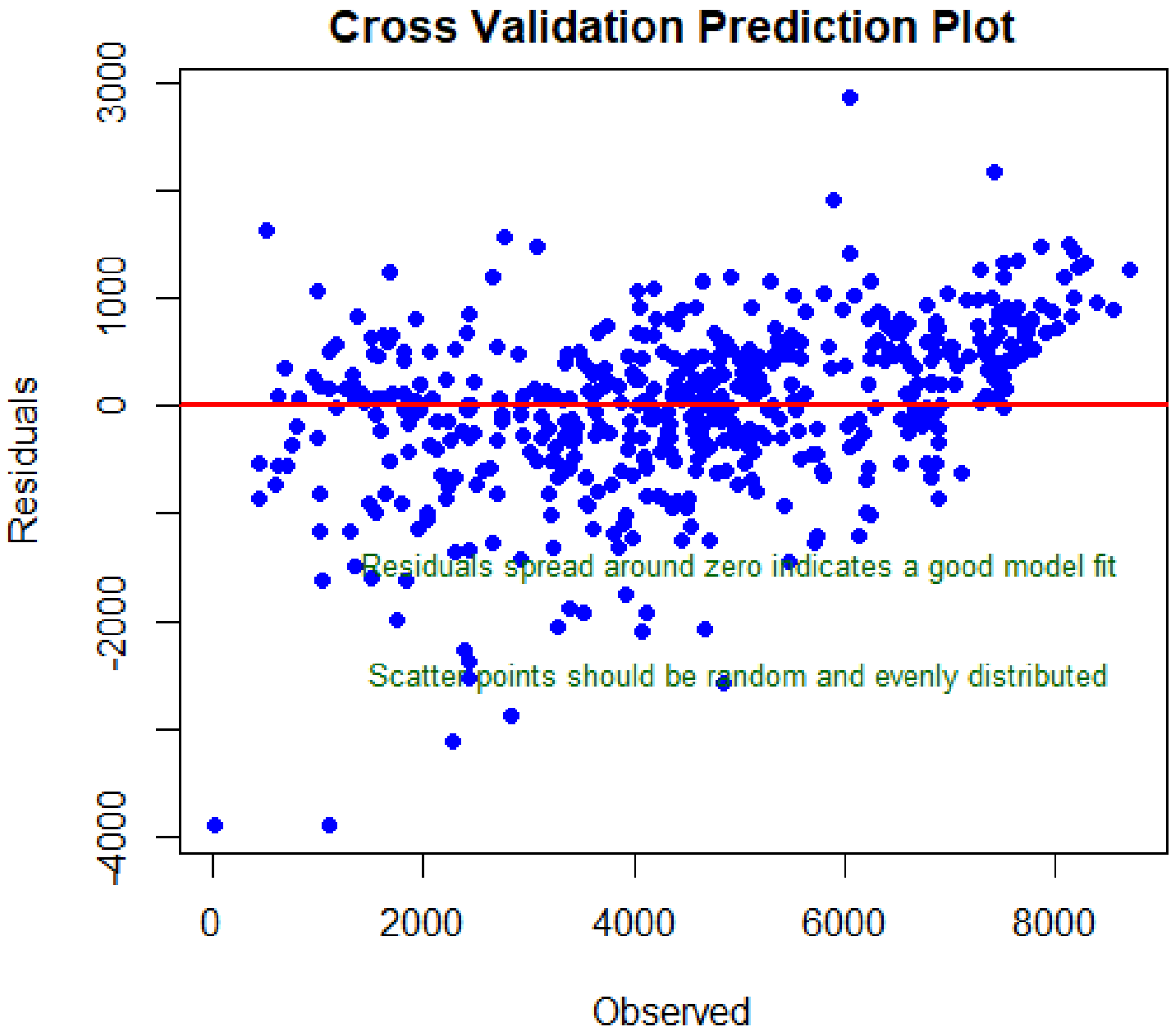

Normal Probability Plot for humidit
Sample Quantiles
0.3
0.6
0.9
Some data po
-3
-2
-1
0
1
2
3
Theoretical Quantiles

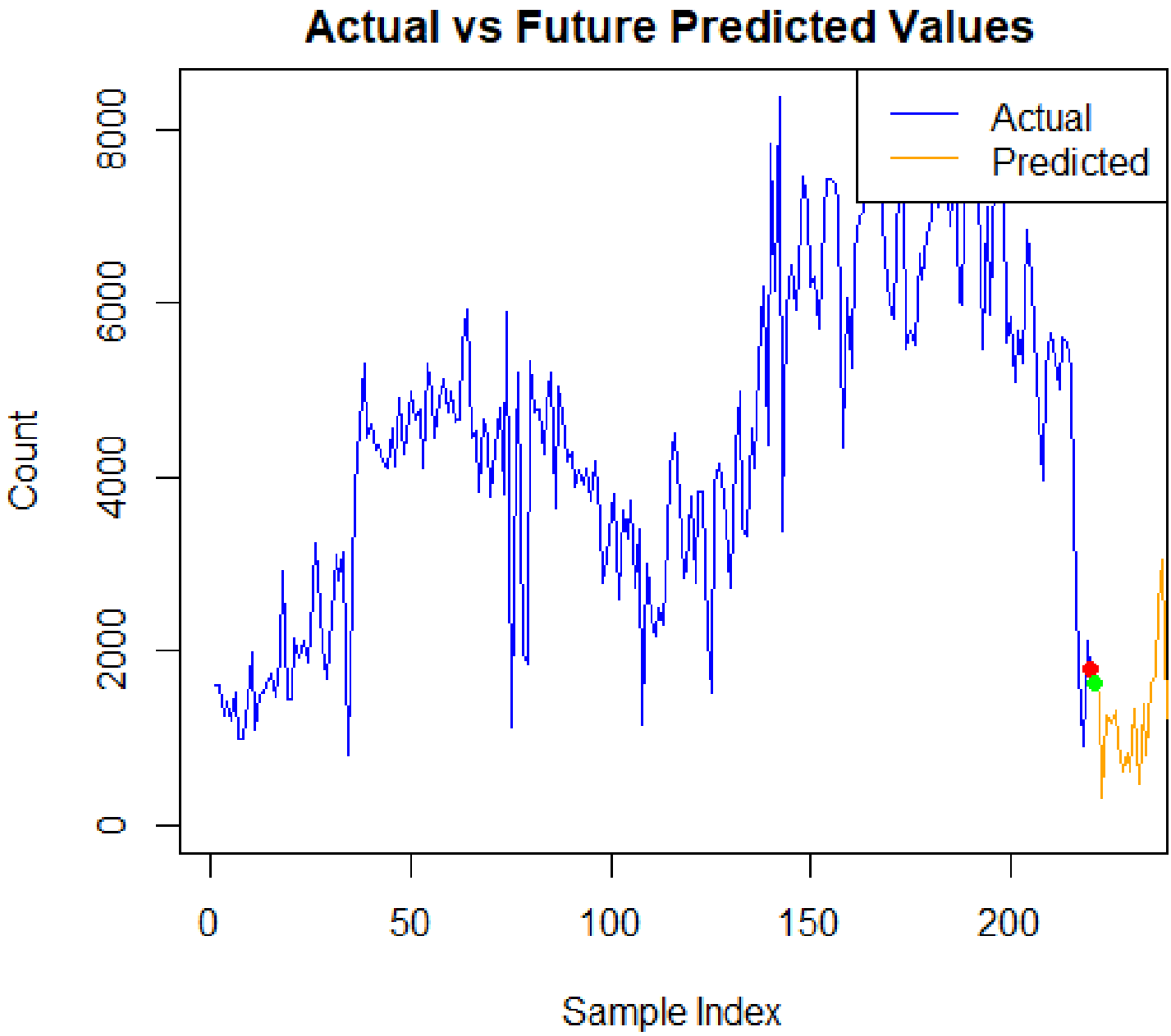
Correlation Plot

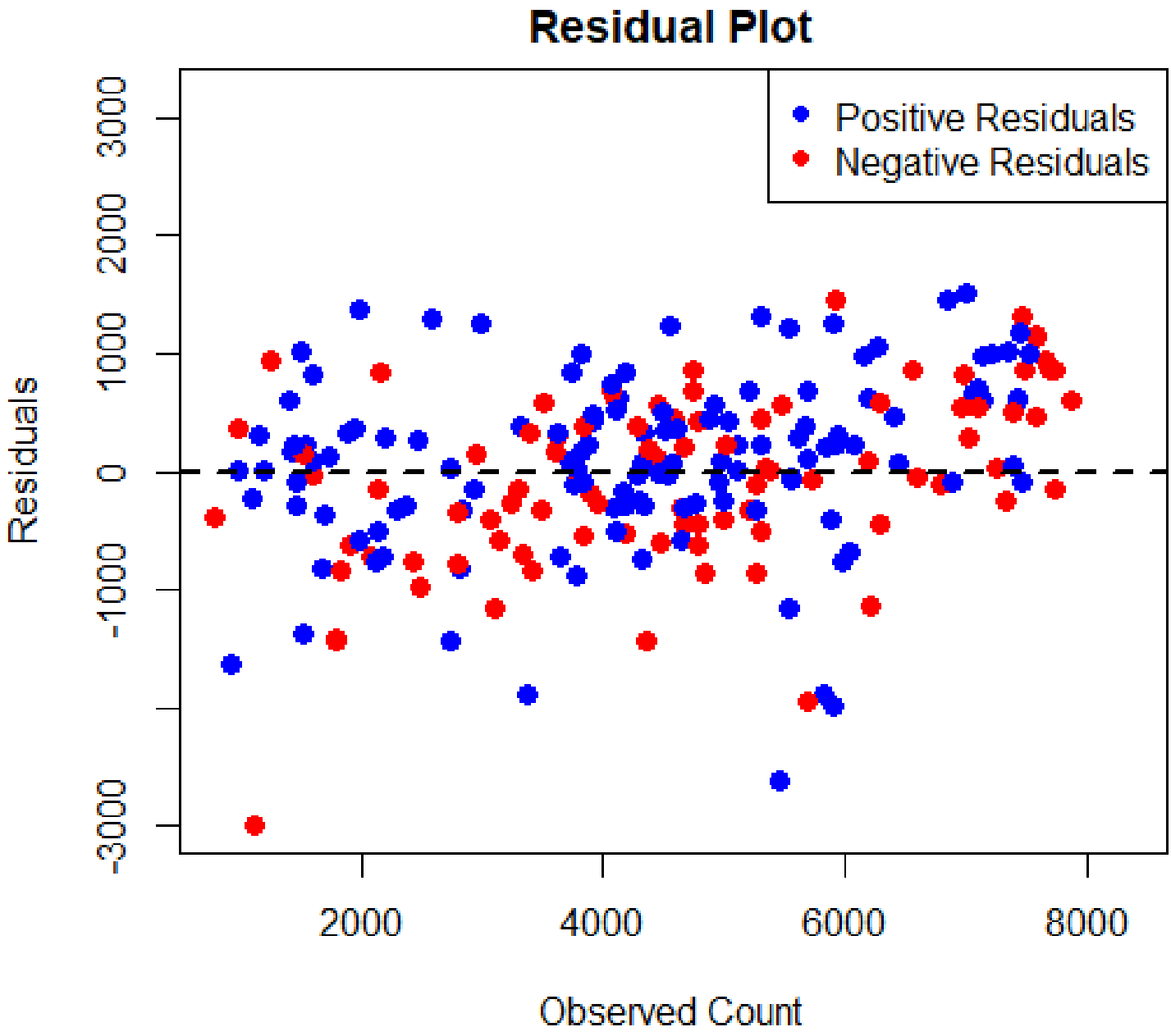


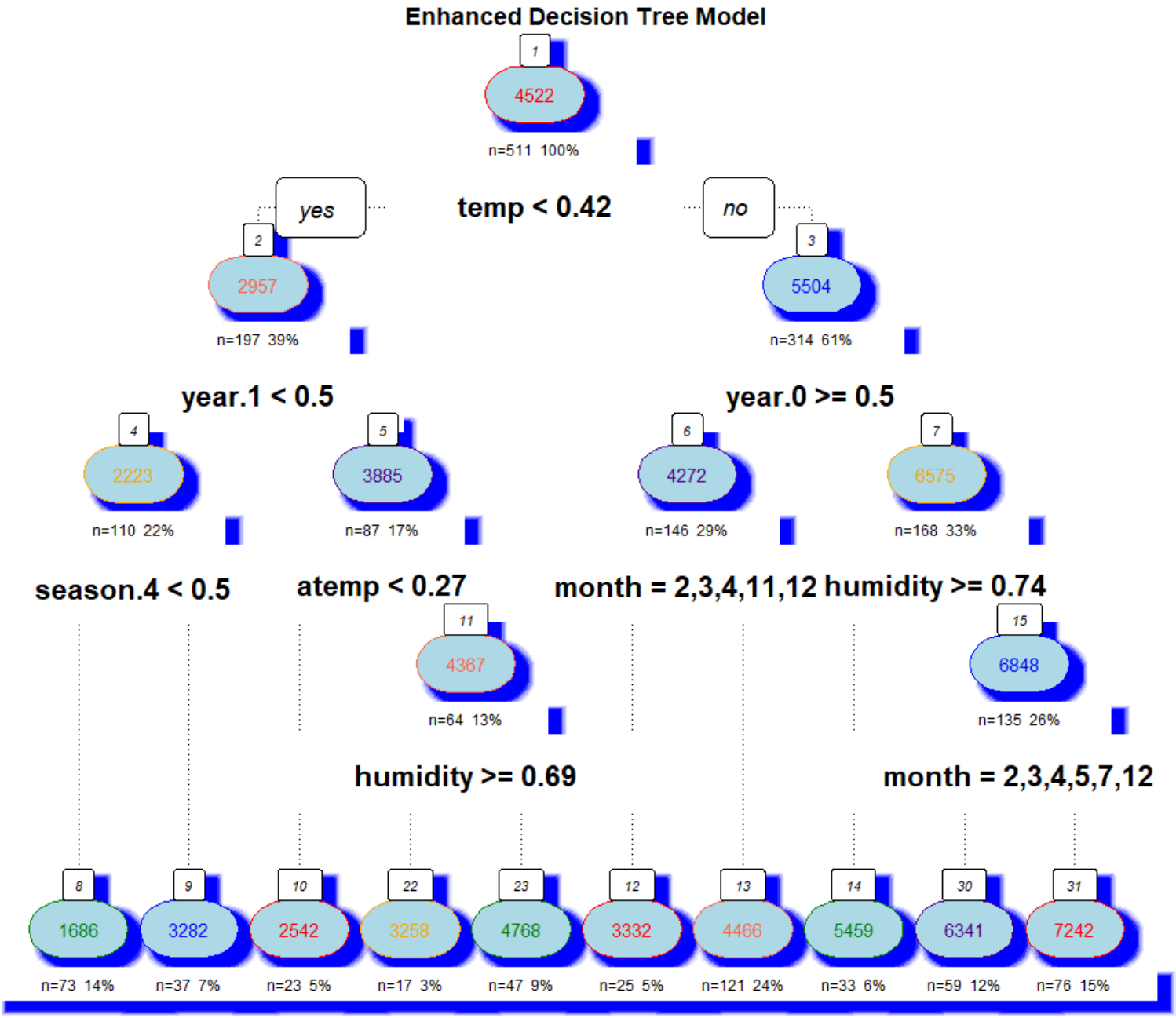
Correlation Plot

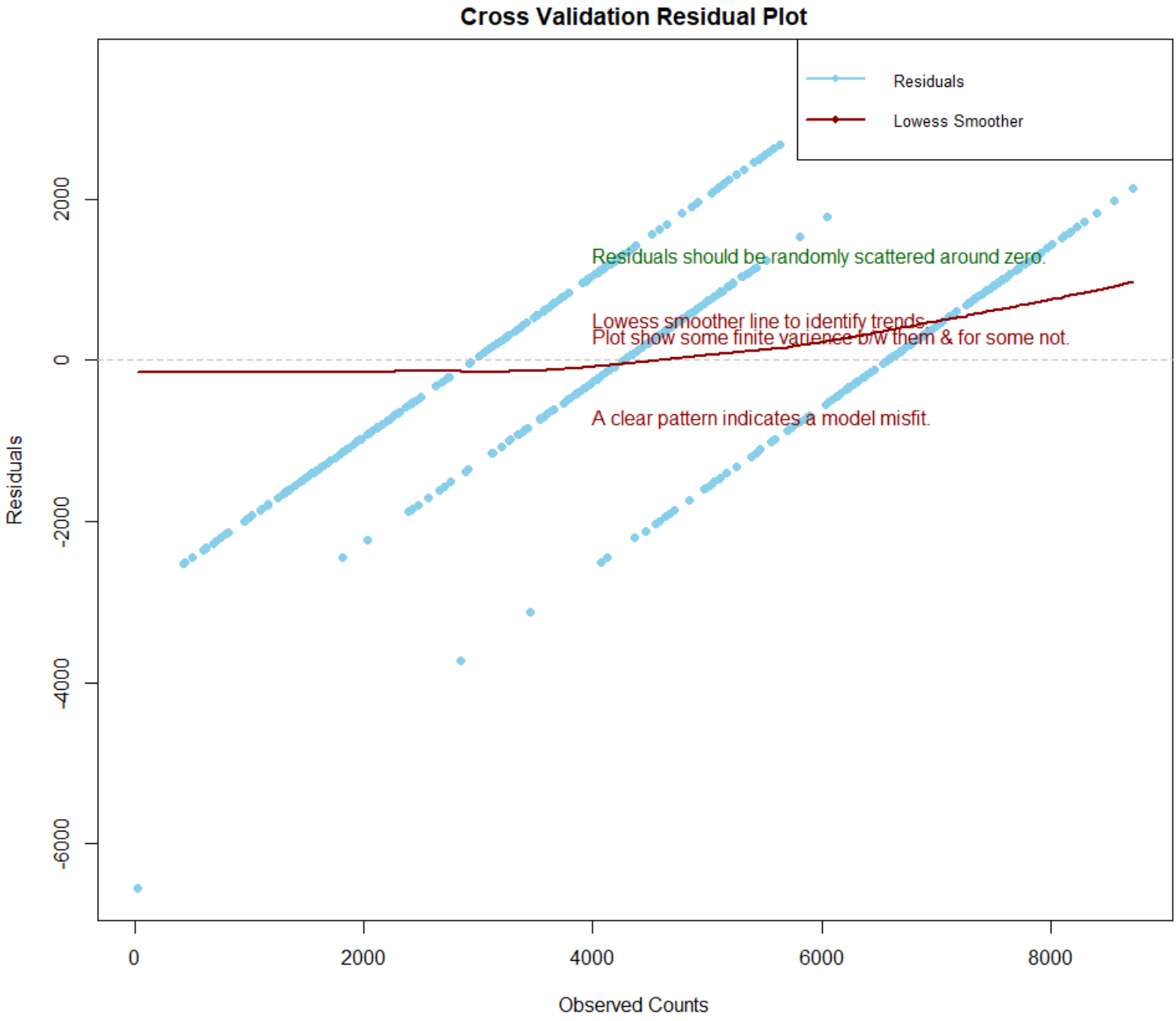


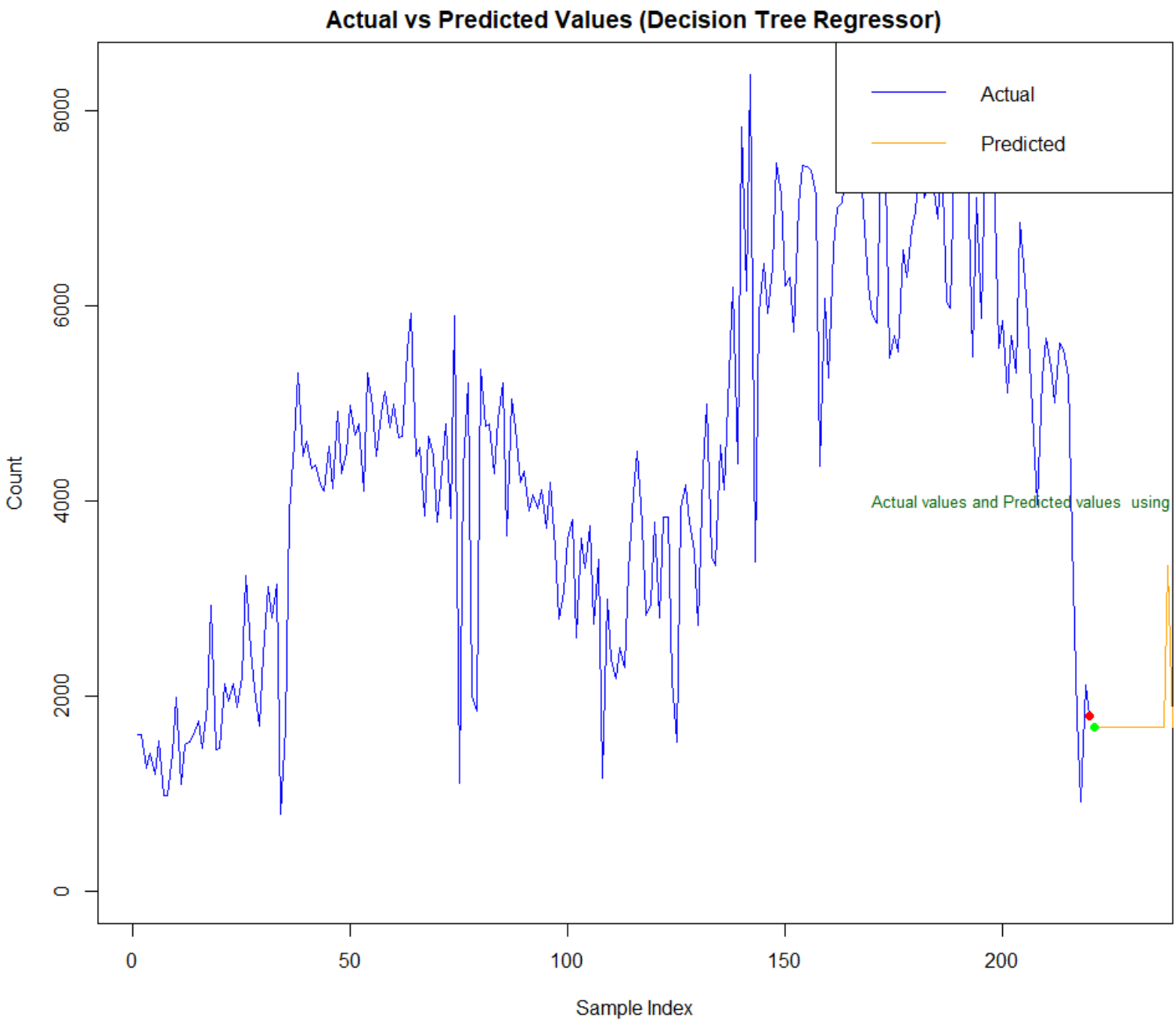

Cross Validation Prediction Plot
Residuals
Observed
3000
1000
0
-2000
-4000
0
2000
4000
6000
8000
Residuals spread around zero indicates a good model fit
Scatter points should be random and evenly distributed

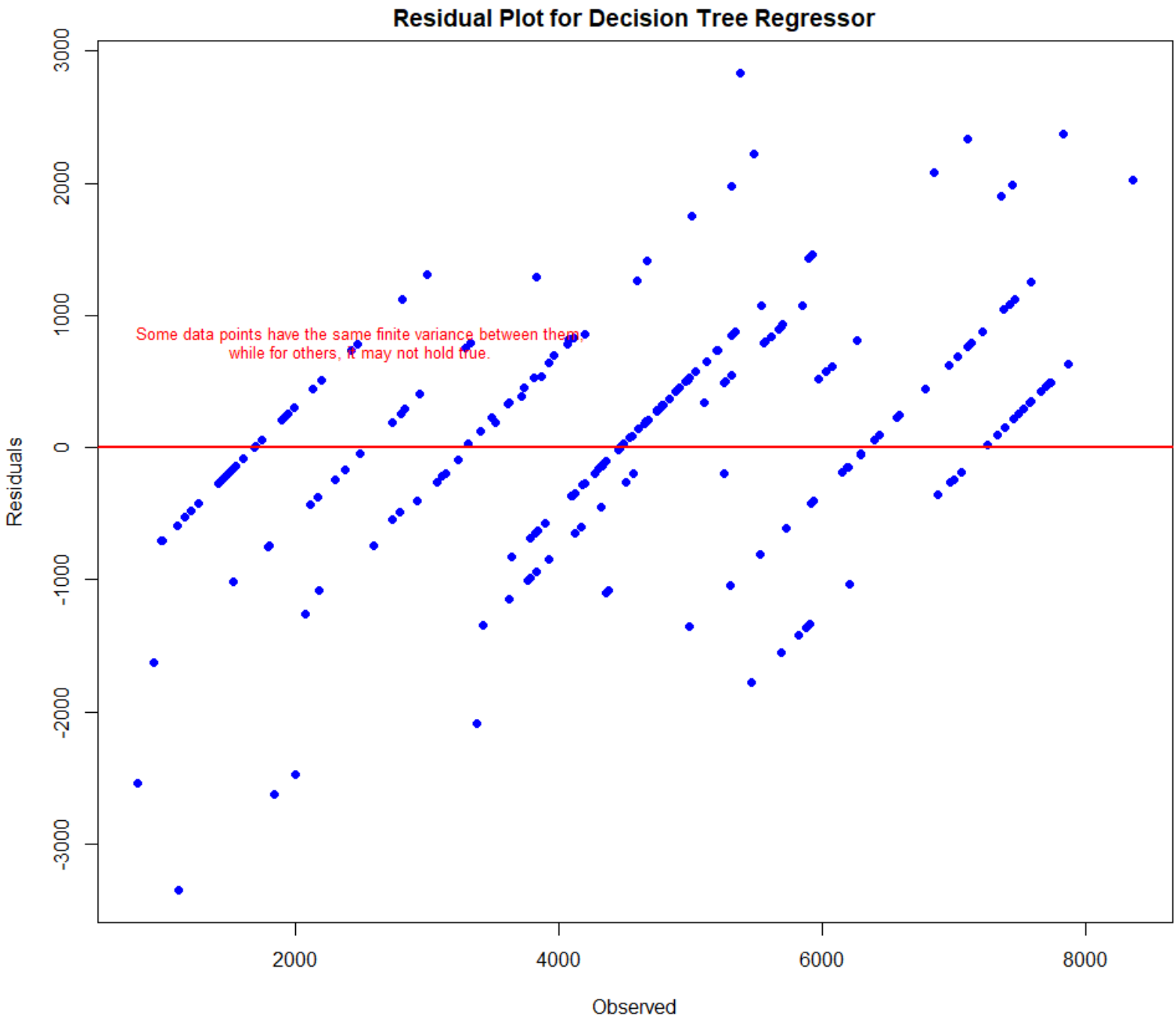

Actual vs Future Predicted Values
Actual
Predicted
Count
0
2000
4000
6000
8000
0
50
100
150
200
Sample Index

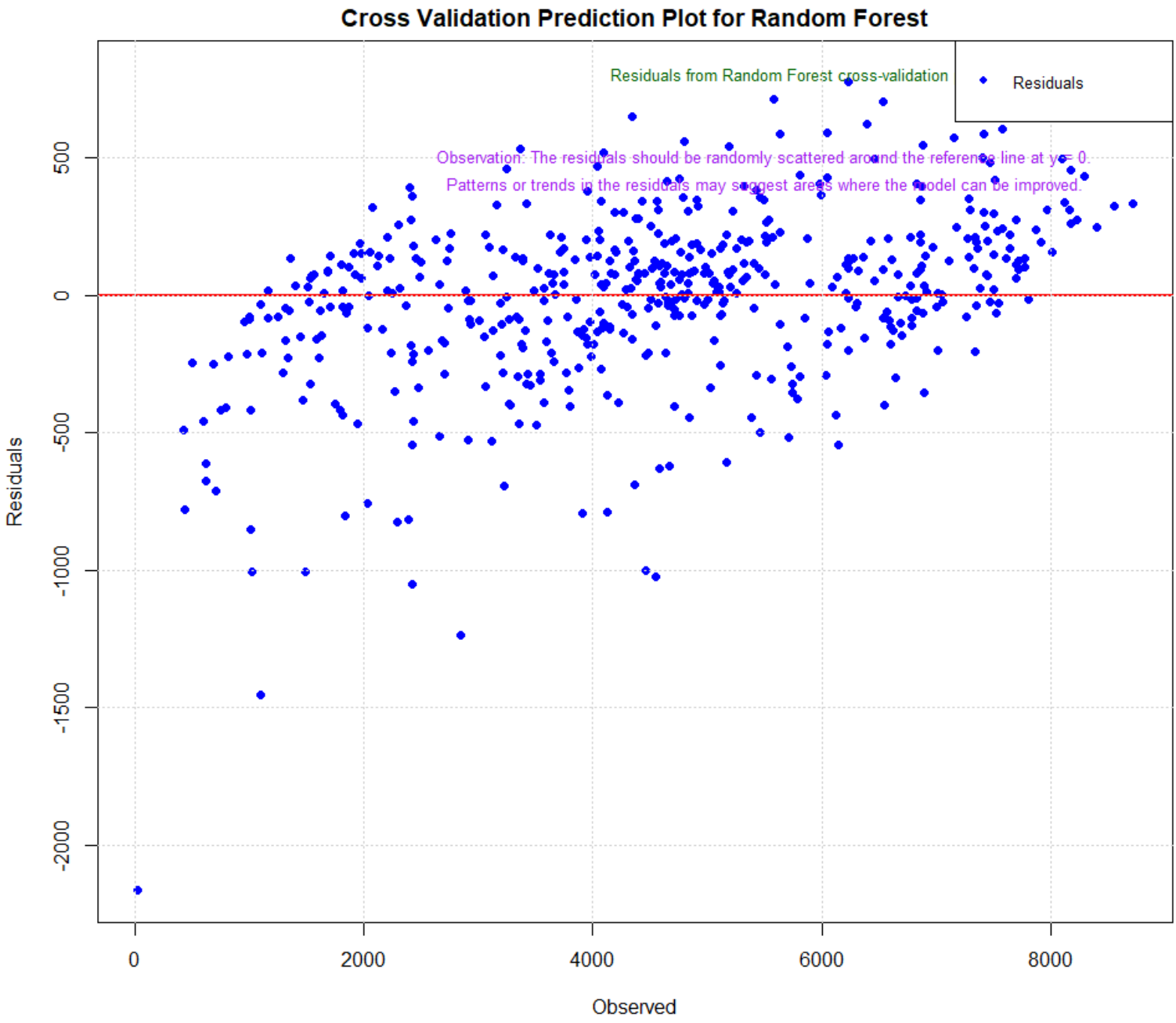

Residual Plot
Positive Residuals
Negative Residuals
Residuals
3000
2000
1000
0
-1000
-3000
2000
4000
6000
8000
Observed Count

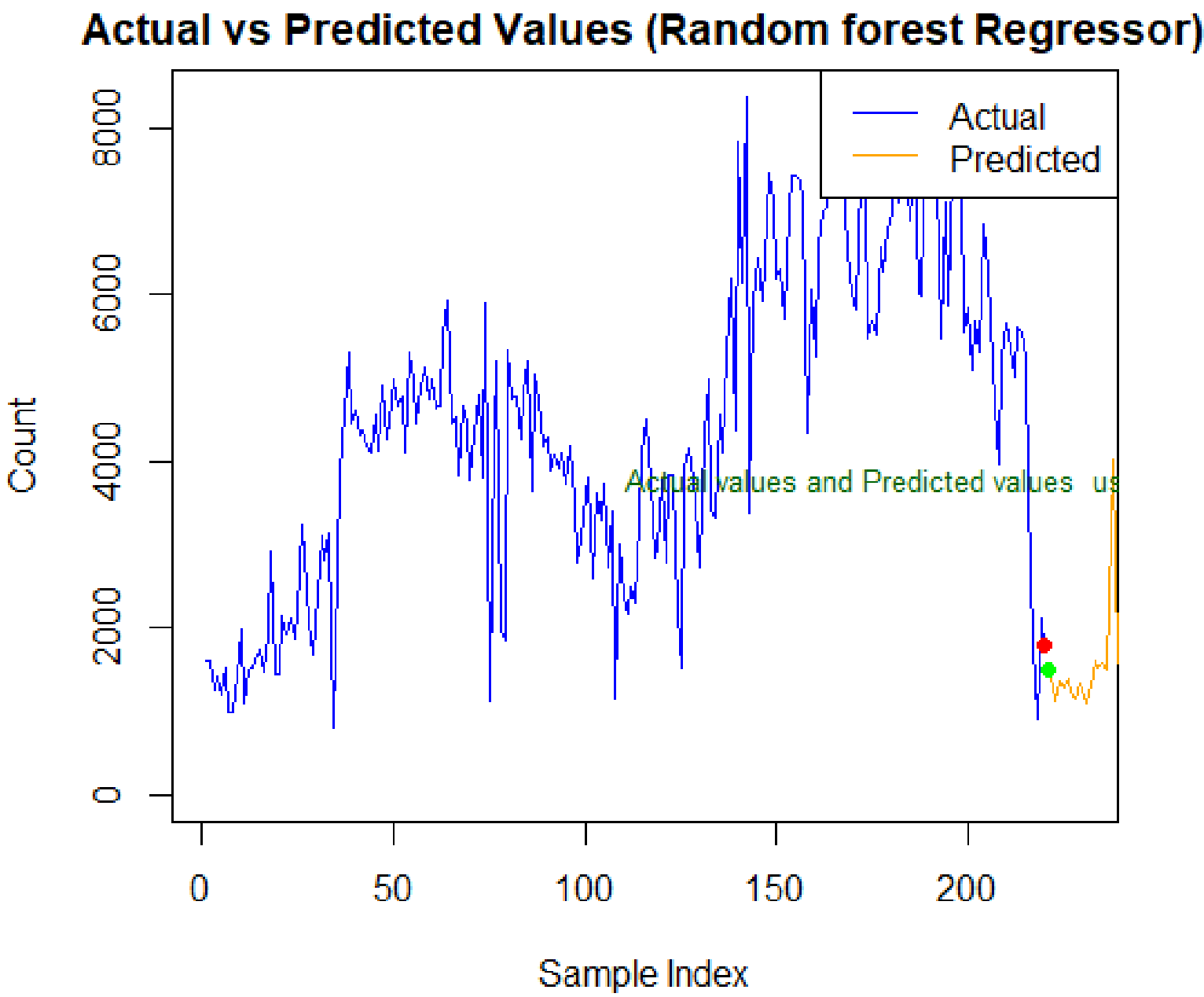

Enhanced Decision Tree Model
1
4522
n=511 100%
yes
temp < 0.42
no
2
2957
n=197 39%
3
5504
n=314 61%
year.1 < 0.5
year.0 >= 0.5
4
2223
n=110 22%
5
3885
n=87 17%
6
4272
n=146 29%
7
6575
n=168 33%
season.4 < 0.5
atemp < 0.27
month = 2,3,4,11,12
humidity >= 0.74
11
4367
n=64 13%
15
6848
n=135 26%
humidity >= 0.69
month = 2,3,4,5,7,12
8
1686
n=73 14%
9
3282
n=37 7%
10
2542
n=23 5%
22
3258
n=17 3%
23
4768
n=47 9%
12
3332
n=25 5%
13
4466
n=121 24%
14
5459
n=33 6%
30
6341
n=59 12%
31
7242
n=76 15%

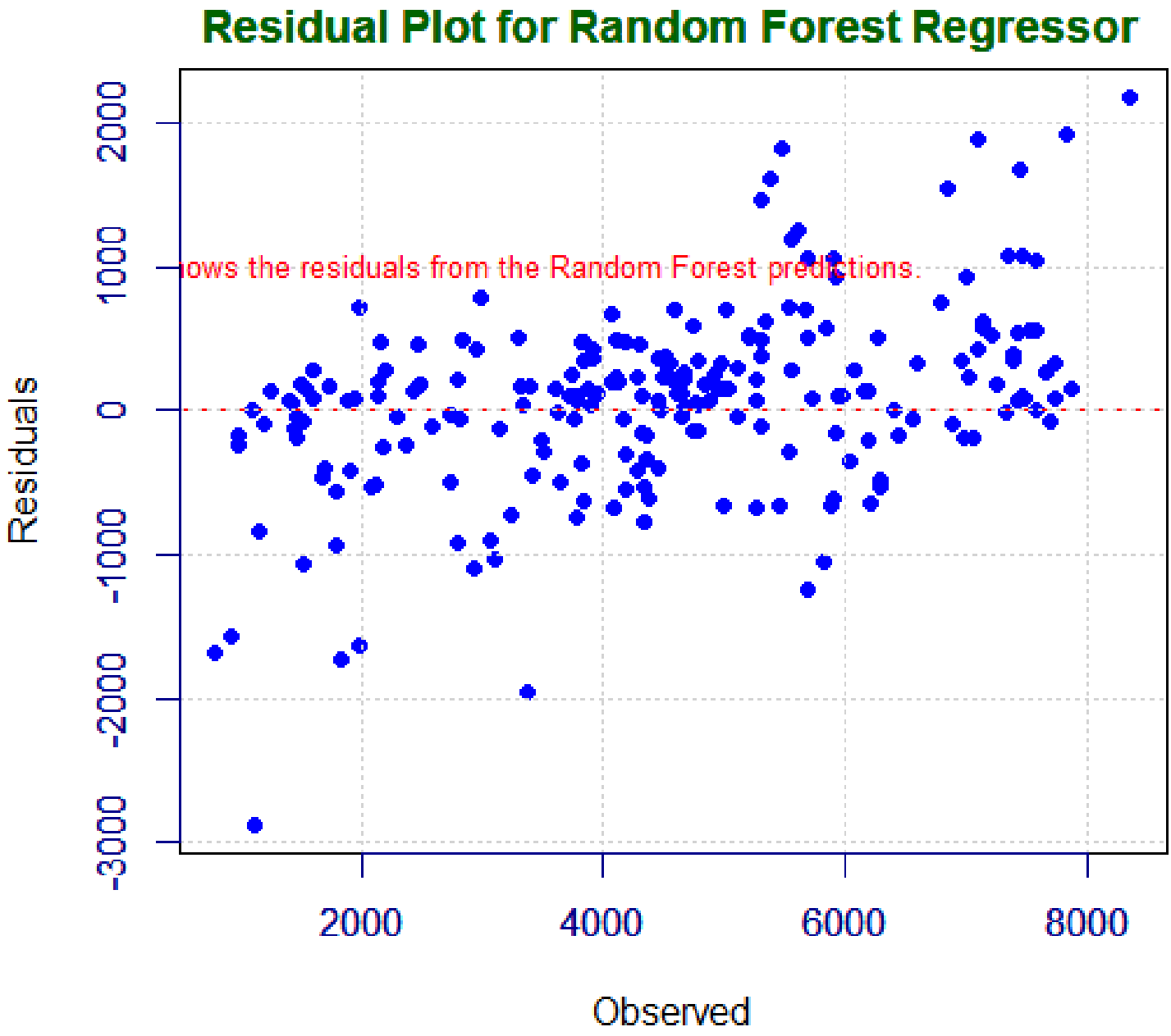

Cross Validation Residual Plot
Residuals
Lowess Smoother
Residuals should be randomly scattered around zero.
Lowess smoother line to identify trends.
Plot show some finite variance b/w them & for some not.
A clear pattern indicates a model misfit.
Residuals
2000
0
-2000
-4000
-6000
0
2000
4000
6000
8000
Observed Counts

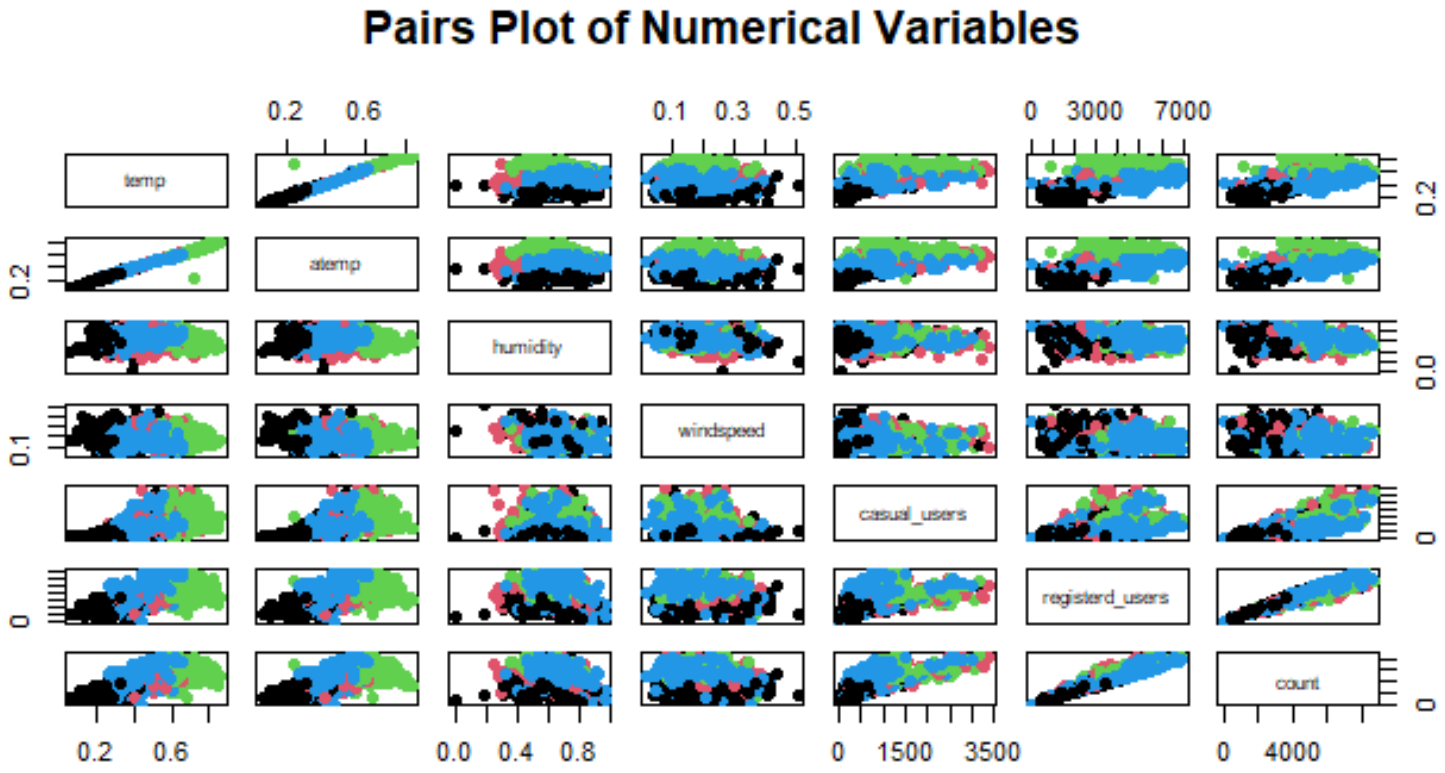

Actual vs Predicted Values (Decision Tree Regressor)
Actual
Predicted
Actual values and Predicted values using
Count
Sample Index
0
2000
4000
6000
8000
0
50
100
150
200

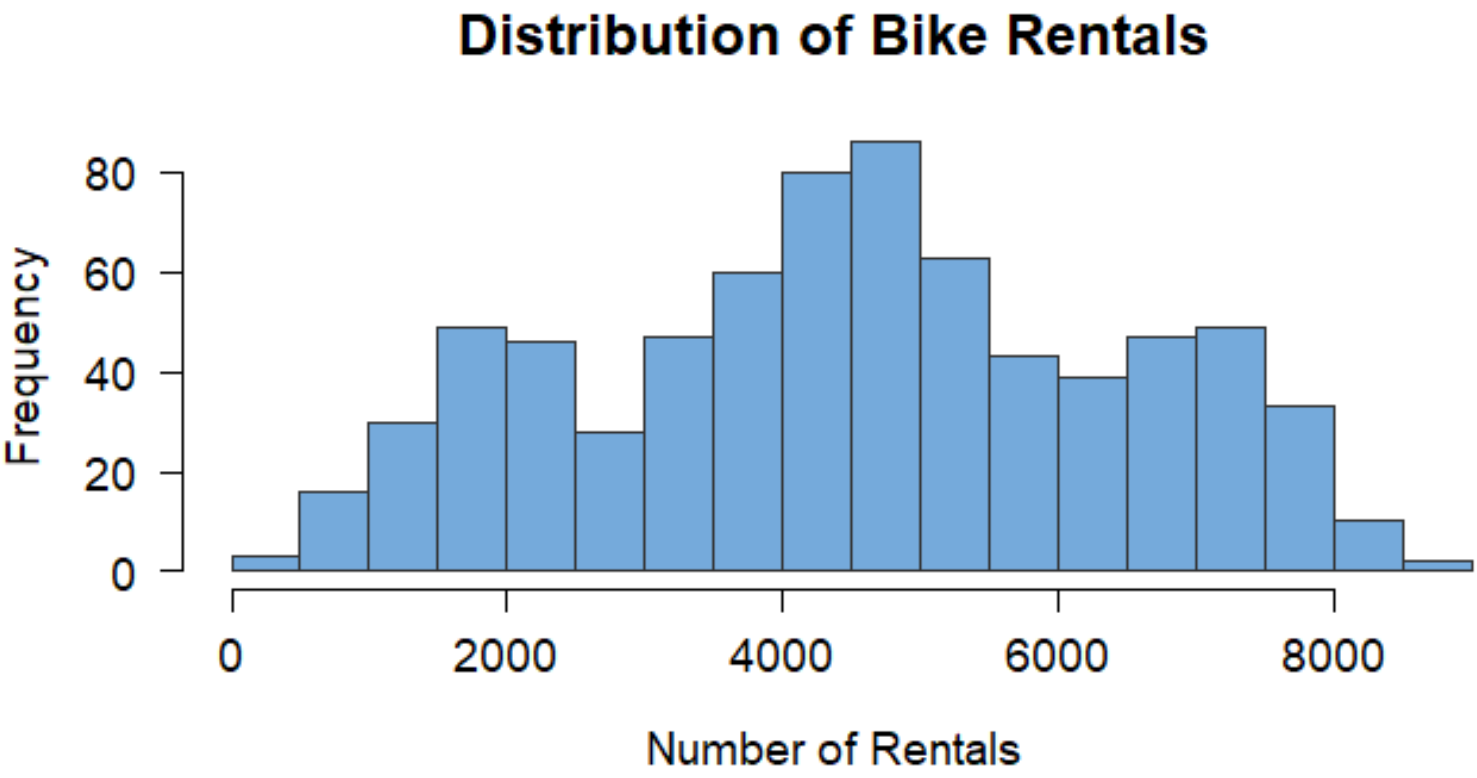

Residual Plot for Decision Tree Regressor
Some data points have the same finite variance between them,
while for others, it may not hold true.
Residuals
Observed
3000
2000
1000
0
-1000
-2000
-3000
2000
4000
6000
8000

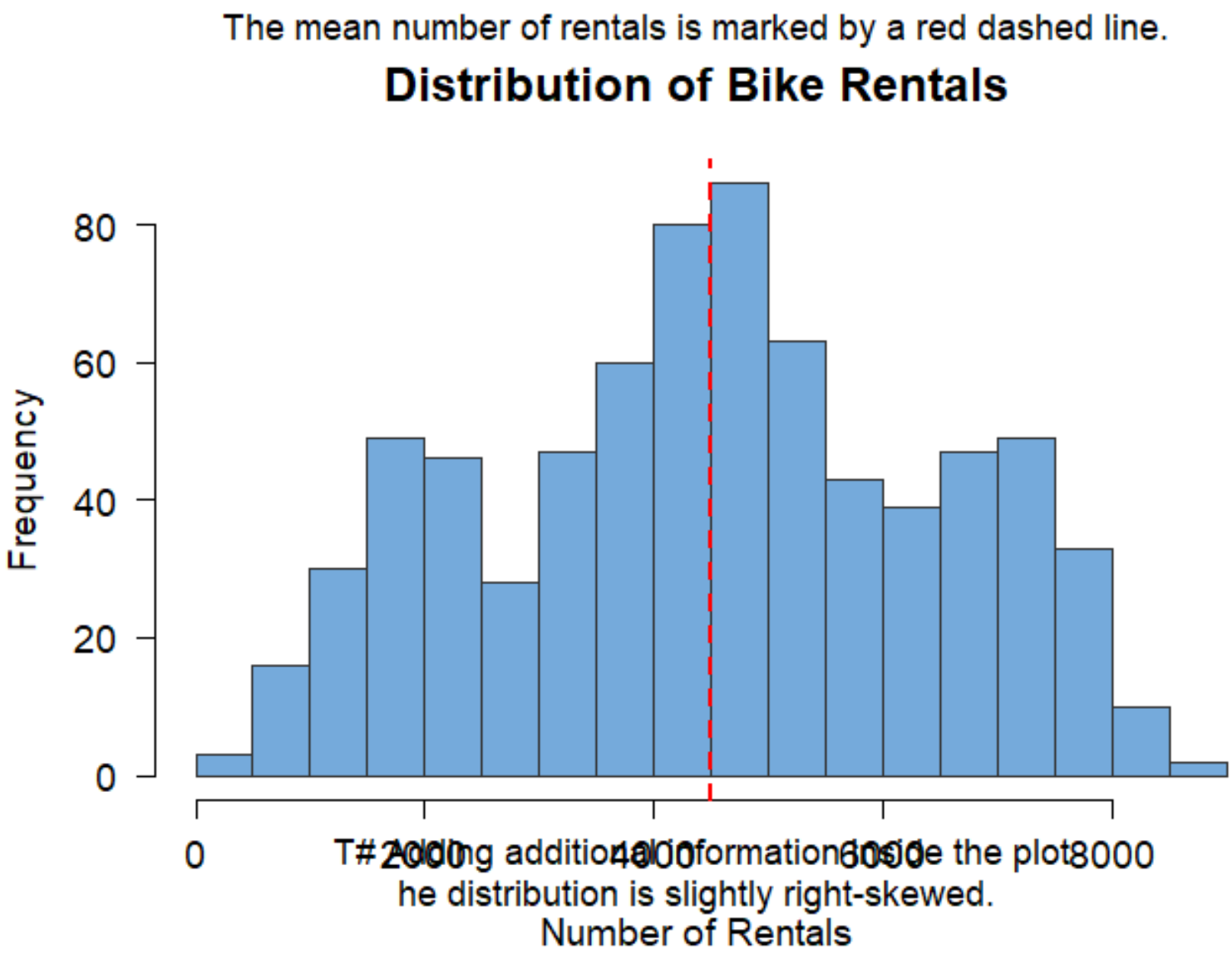

Cross Validation Prediction Plot for Random Forest
Residuals from Random Forest cross-validation
Residuals
Observation: The residuals should be randomly scattered around the reference line at y = 0.
Patterns or trends in the residuals may suggest areas where the model can be improved.
Residuals
500
0
-500
-1000
-1500
-2000
0
2000
4000
6000
8000
Observed


Actual vs Predicted Values (Random forest Regressor)
Actual
Predicted
Count
0
2000
4000
6000
8000
0
50
100
150
200
Sample Index
Actual values and Predicted values us

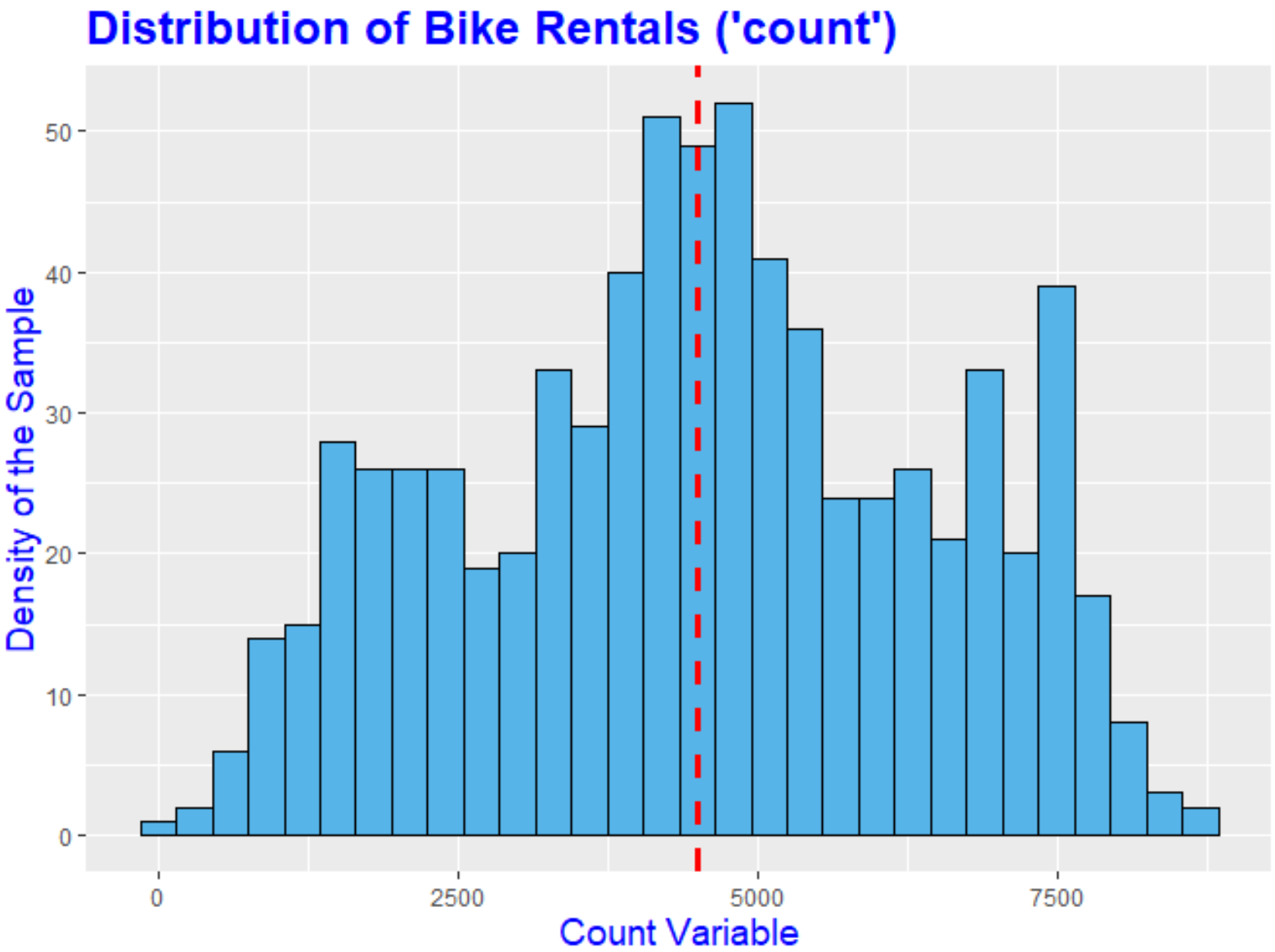

Residual Plot for Random Forest Regressor
ows the residuals from the Random Forest predictions.
Residuals
2000
1000
0
-1000
-2000
-3000
2000
4000
6000
8000
Observed

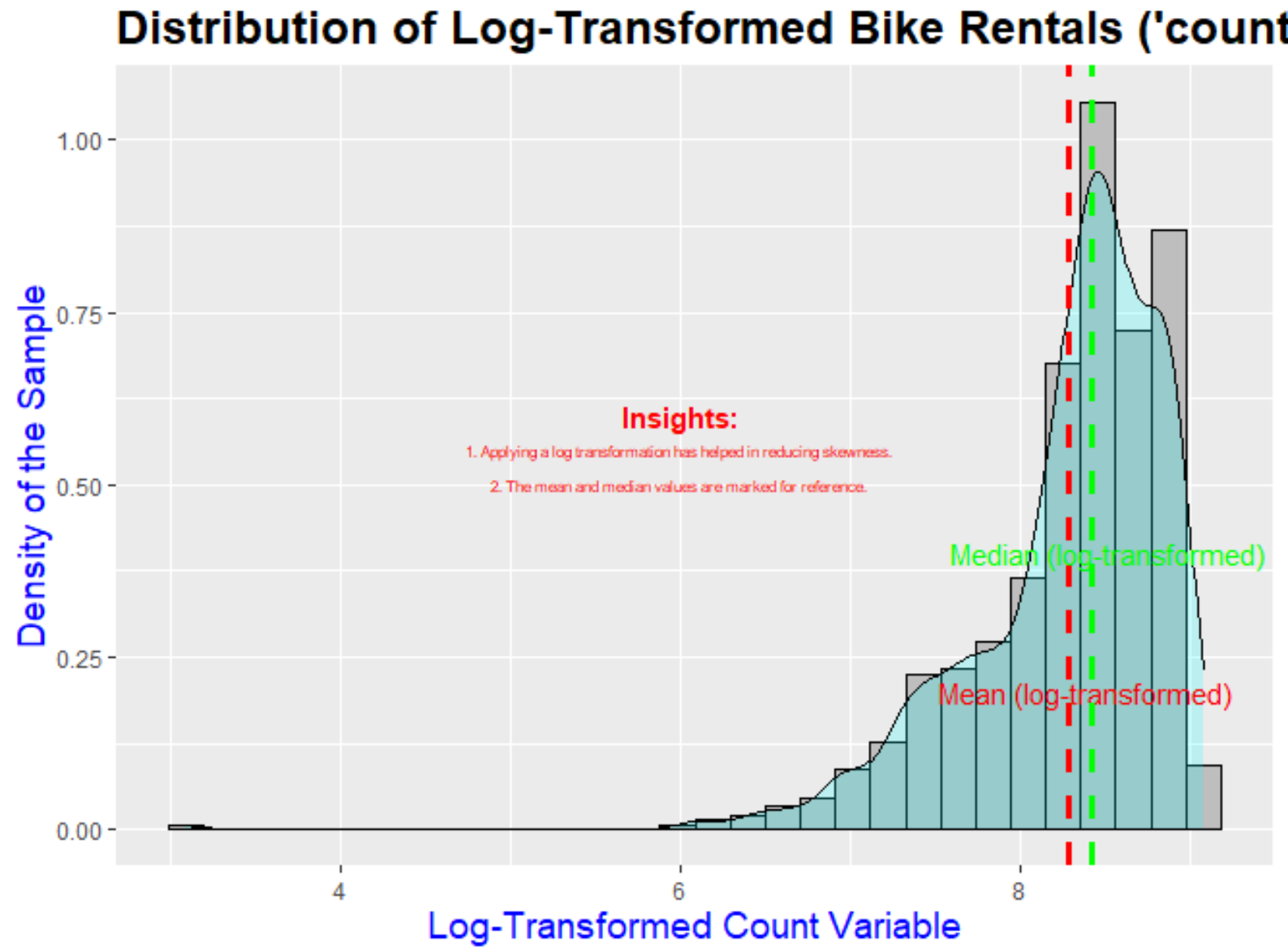

Pairs Plot of Numerical Variables
temp
atemp
humidity
windspeed
casual_users
registerd_users
count
0.2
0.6
0.1
0.3
0.5
0
3000
7000
0.0
0.4
0.8
1500
3500
4000

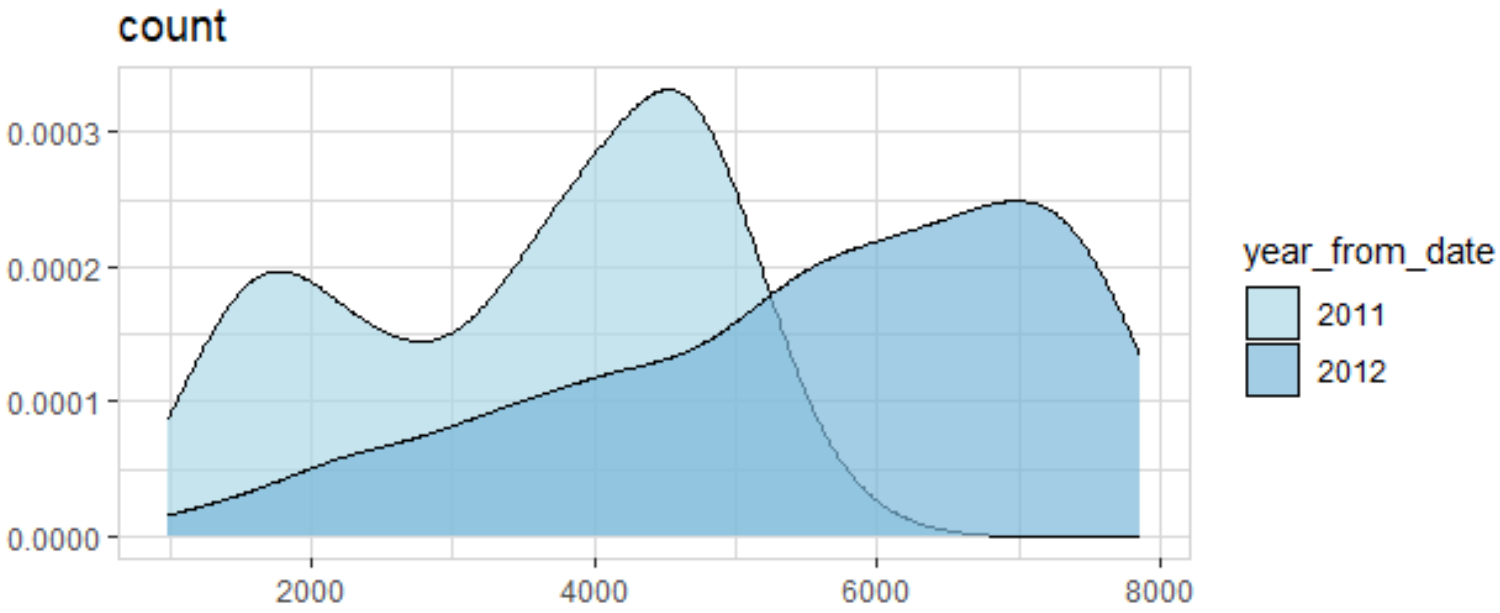

Distribution of Bike Rentals
Frequency
0
20
40
60
80
0
2000
4000
6000
8000
Number of Rentals

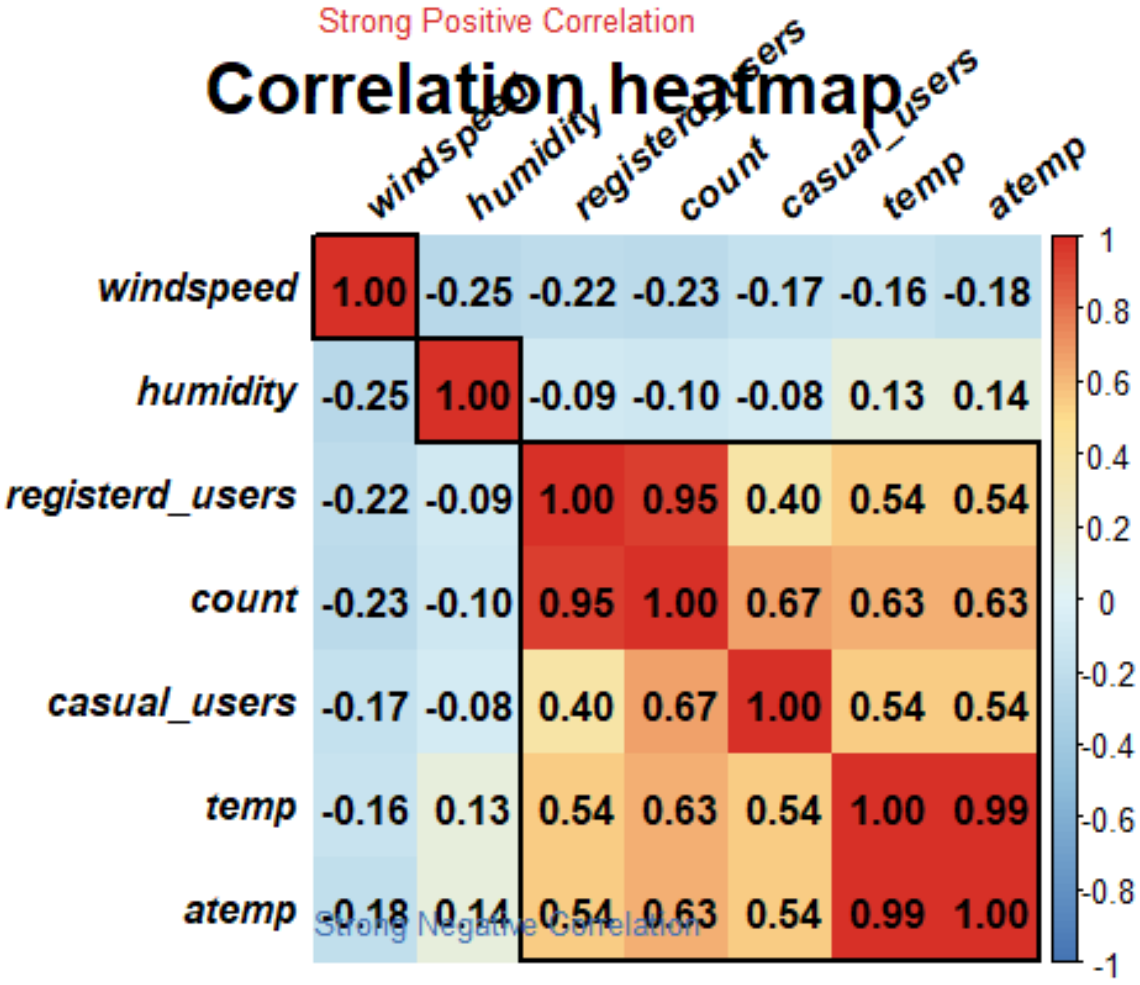


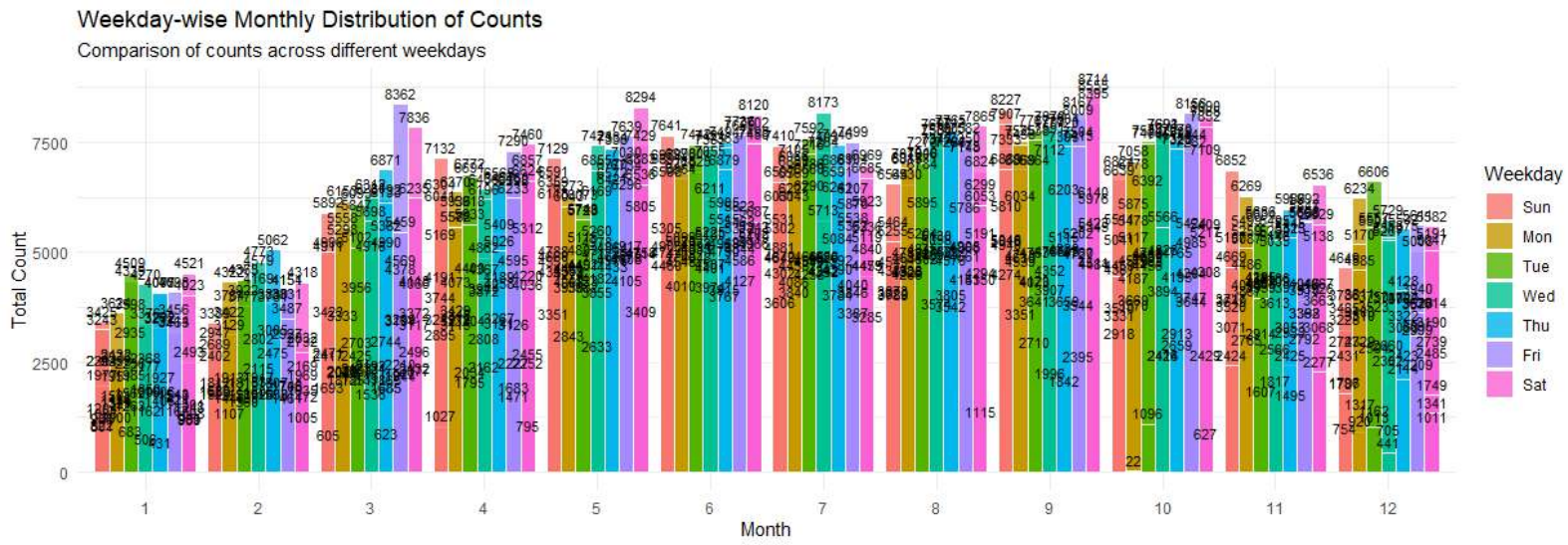
T# Adding additional information inside the plot
he distribution is slightly right-skewed.

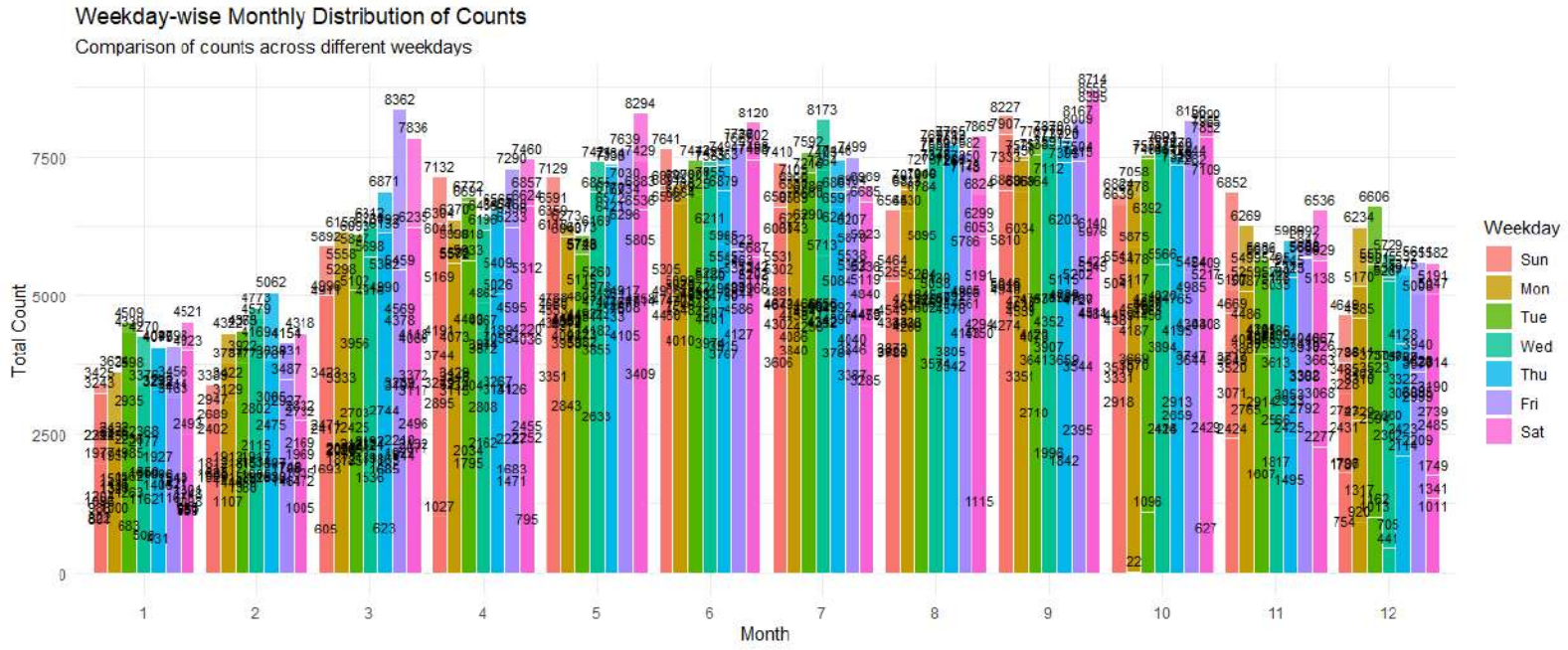

Distribution of Bike Rentals ('count')
Density of the Sample
0
10
20
30
40
50
0
2500
5000
7500
Count Variable

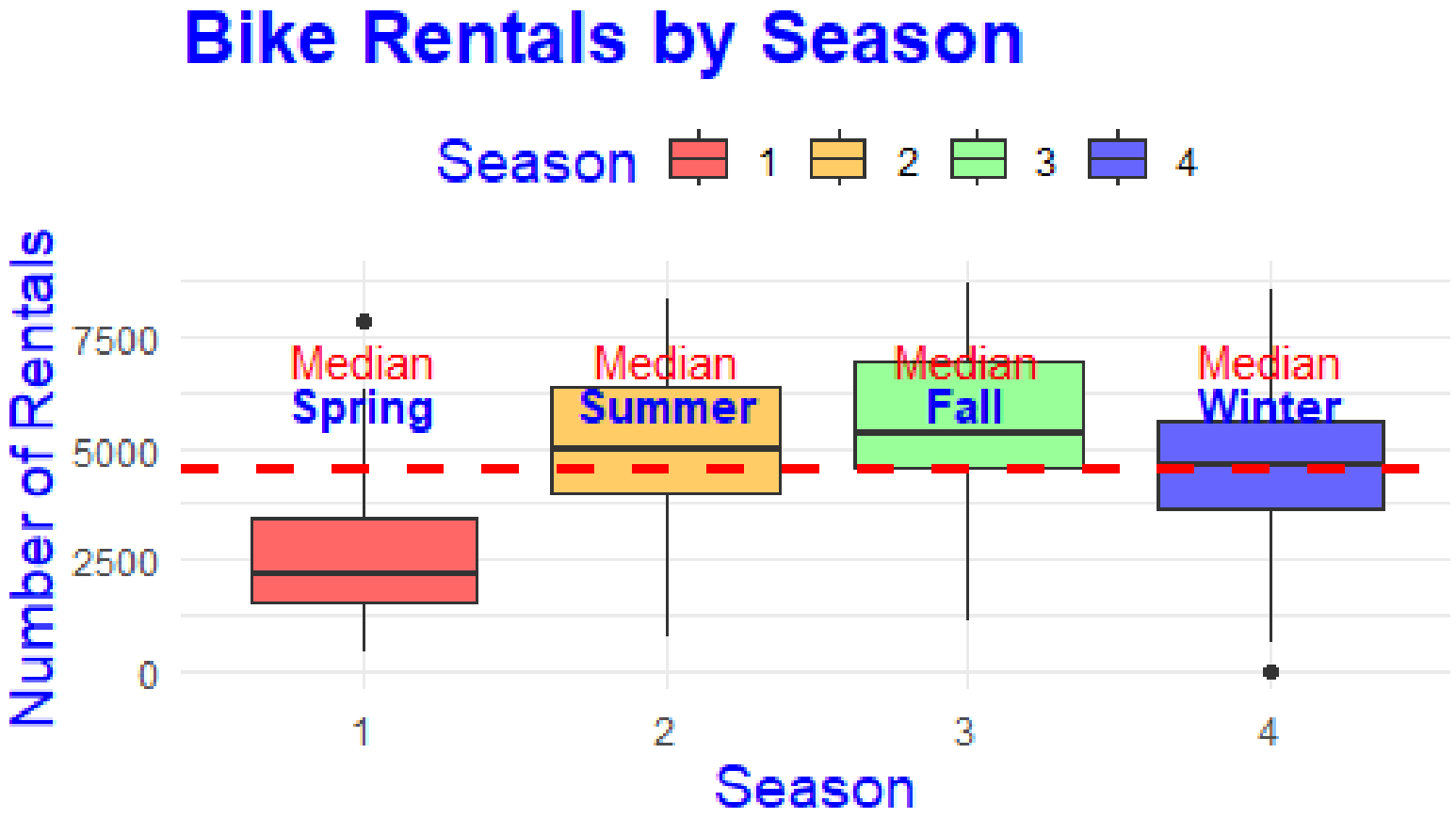

Distribution of Log-Transformed Bike Rentals ('count
Density of the Sample
1.00
0.75
0.50
0.25
0.00
Insights:
1. Applying a log transformation has helped in reducing skewness.
2. The mean and median values are marked for reference.
Median (log-transformed)
Mean (log-transformed)
4
6
8
Log-Transformed Count Variable

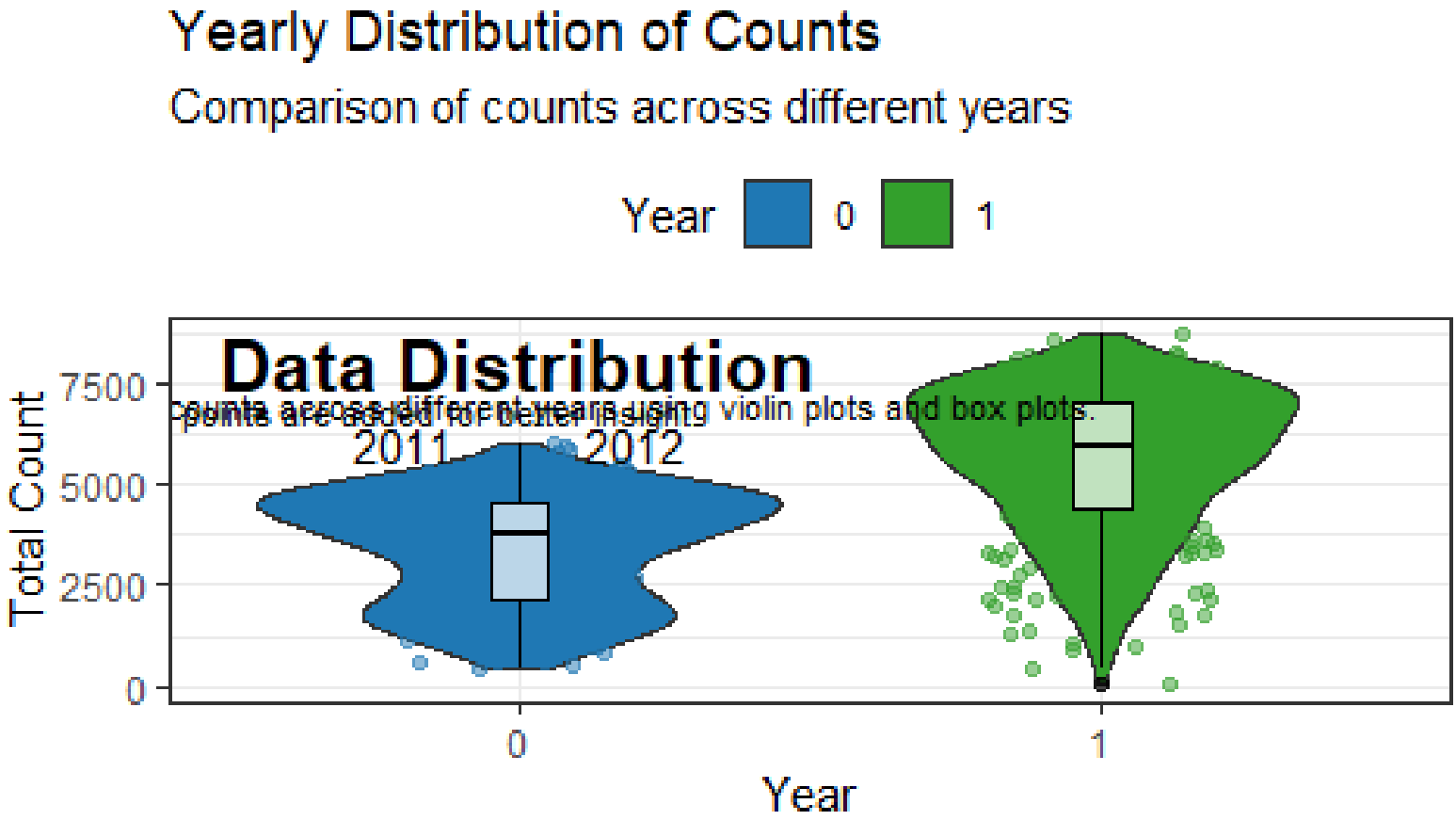

count
0.0003
0.0002
0.0001
0.0000
2000
4000
6000
8000
year_from_date
2011
2012

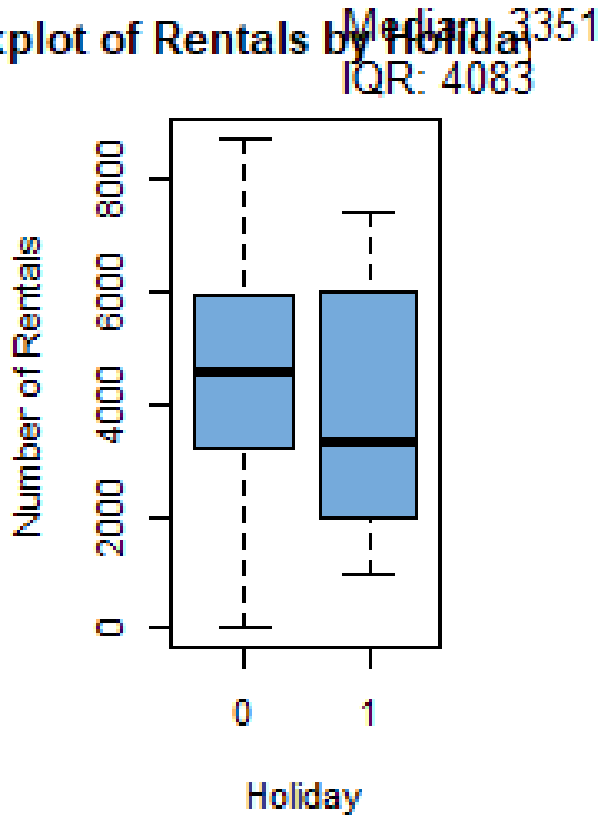

Strong Positive Correlation
Correlation heatmap
windspeed
humidity
registerd_users
count
casual_users
temp
atemp
windspeed 1.00 -0.25 -0.22 -0.23 -0.17 -0.16 -0.18
humidity -0.25 1.00 -0.09 -0.10 -0.08 0.13 0.14
registerd_users -0.22 -0.09 1.00 0.95 0.40 0.54 0.54
count -0.23 -0.10 0.95 1.00 0.67 0.63 0.63
casual_users -0.17 -0.08 0.40 0.67 1.00 0.54 0.54
temp -0.16 0.13 0.54 0.63 0.54 1.00 0.99
atemp -0.18 0.14 0.54 0.63 0.54 0.99 1.00
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
Strong Negative Correlation

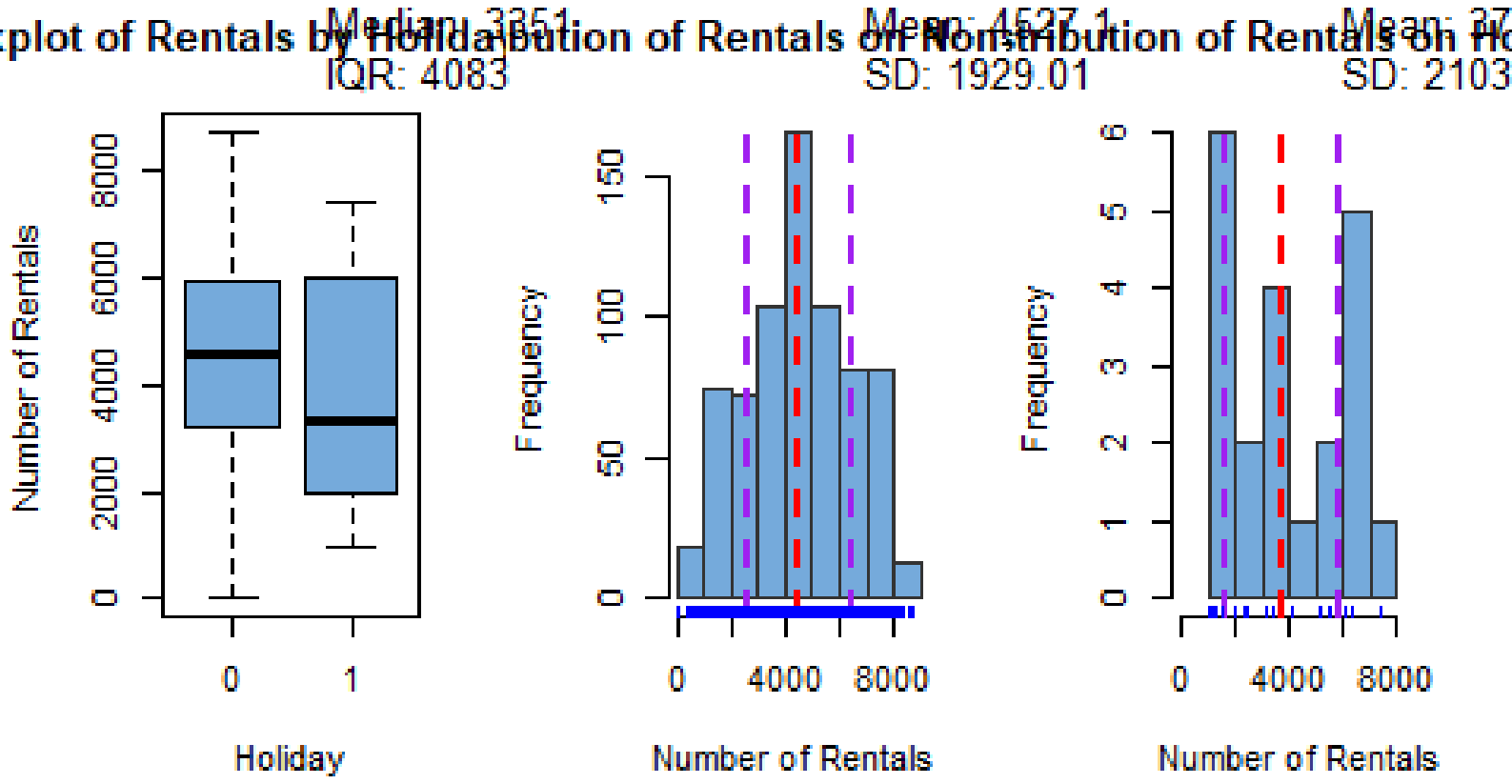

Weekday-wise Monthly Distribution of Counts
Comparison of counts across different weekdays
Total Count
Month
Weekday
Sun
Mon
Tue
Wed
Thu
Fri
Sat

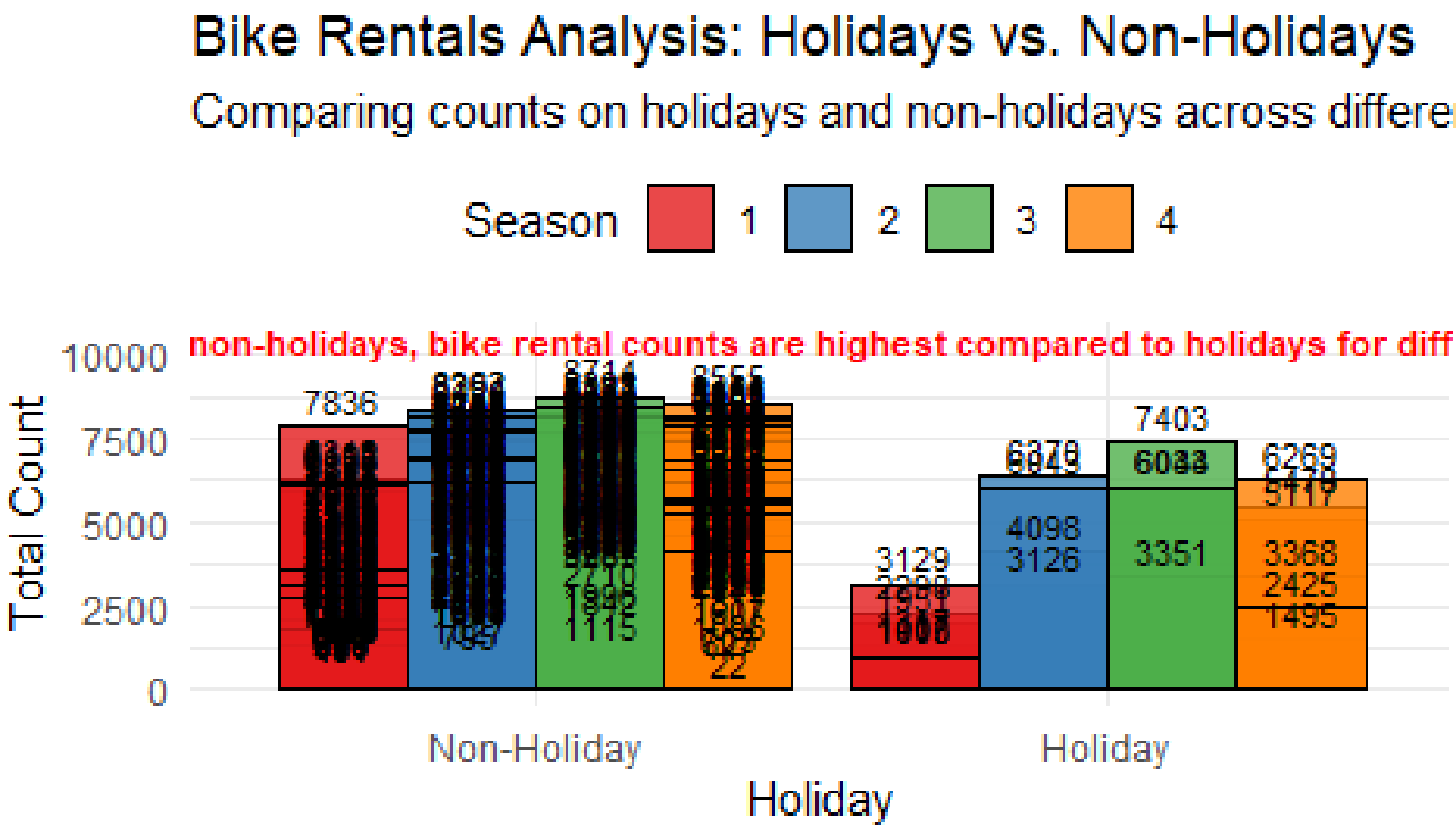

Weekday-wise Monthly Distribution of Counts
Comparison of counts across different weekdays
Total Count
Month
Weekday
Sun
Mon
Tue
Wed
Thu
Fri
Sat

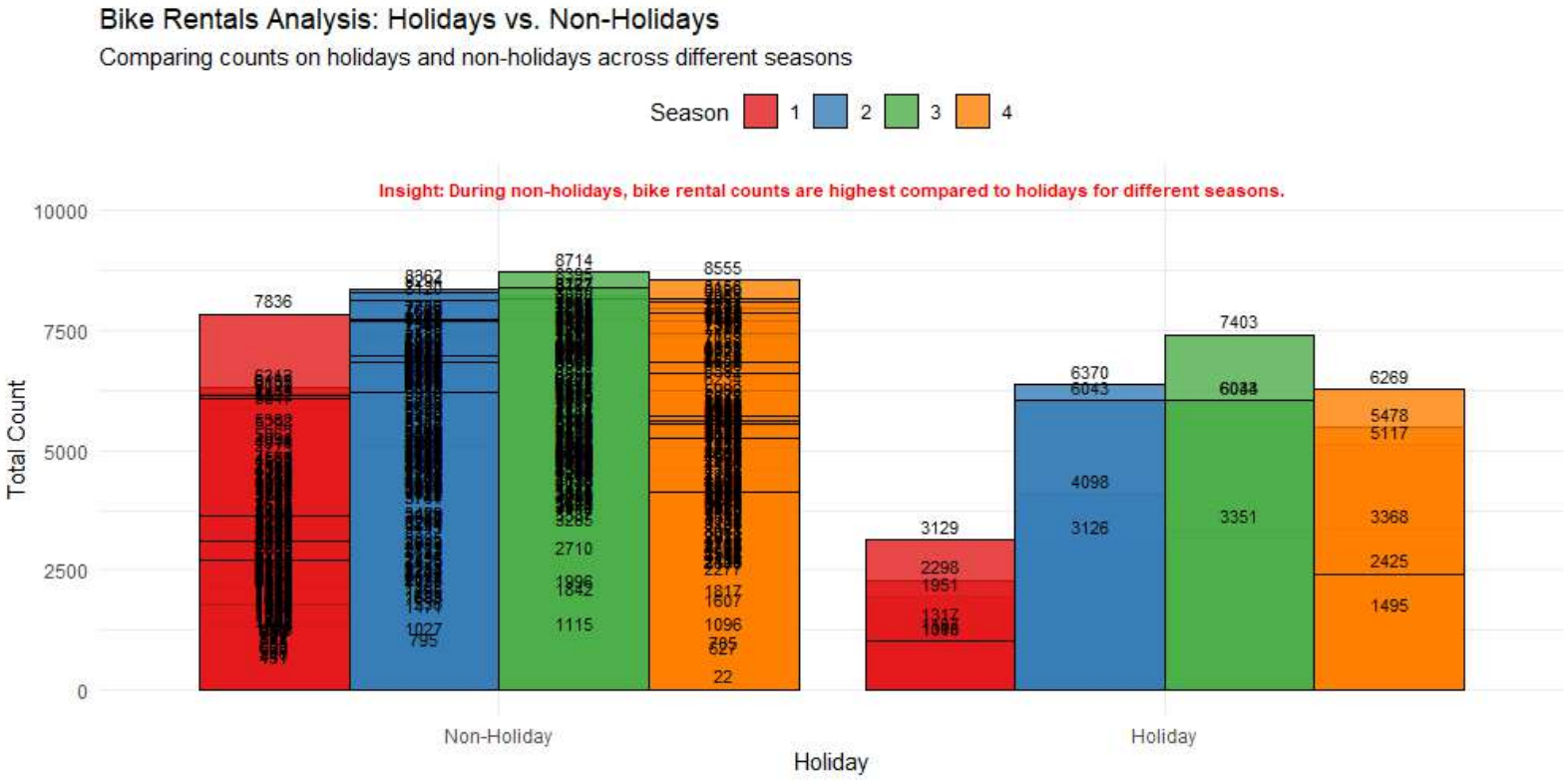

Bike Rentals by Season
Season
1
2
3
4
Number of Rentals
7500
5000
2500
0
Median
Spring
Median
Summer
Median
Fall
Median
Winter
1
2
3
4
Season


Yearly Distribution of Counts
Comparison of counts across different years
Year
0
1
Data Distribution
counts across different years using violin plots and box plots.
points are added for better insight
2011
2012
Total Count
7500
5000
2500
0
0
1
Year

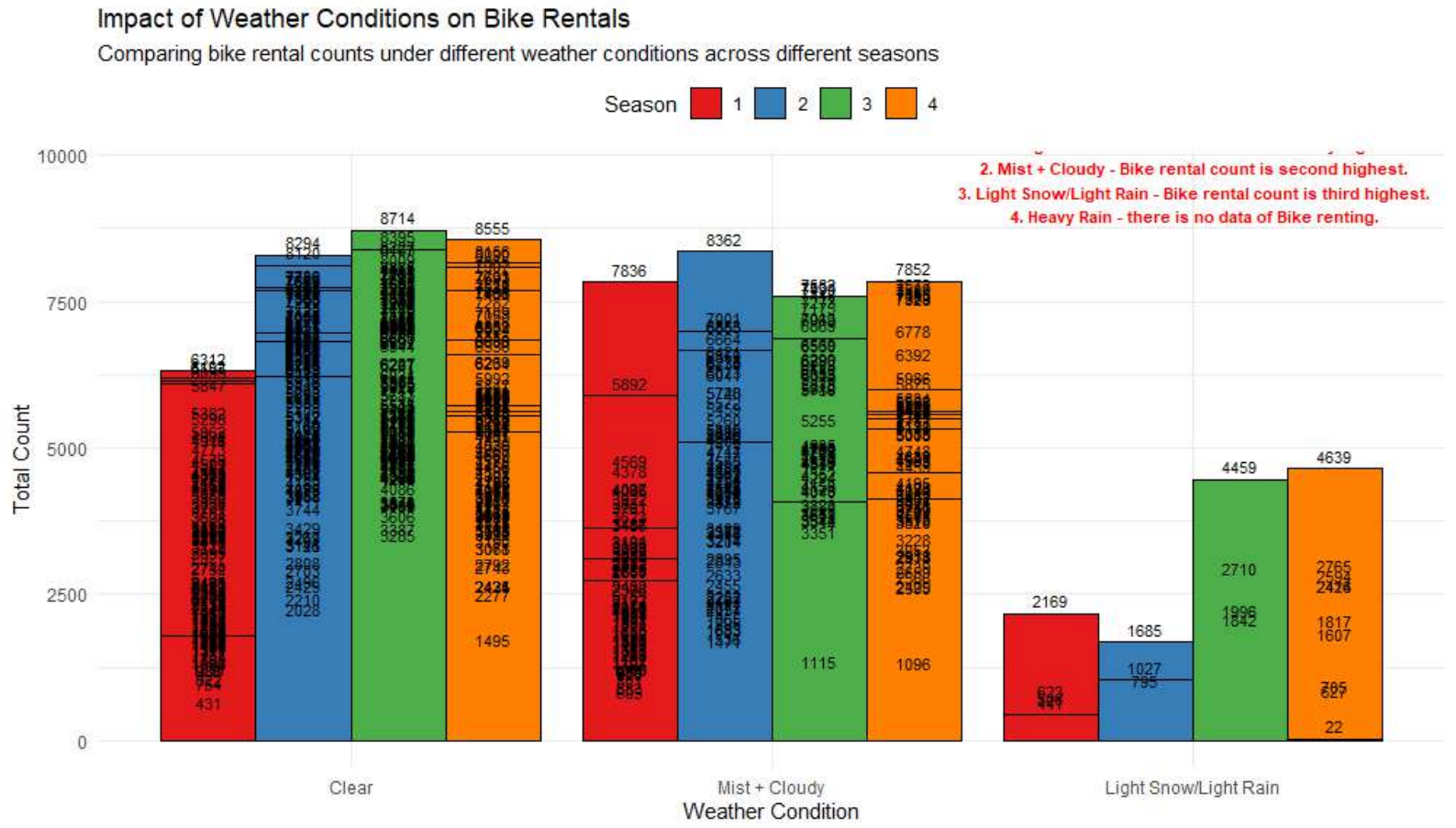

plot of Rentals by Holida
Median: 3351
IQR: 4083
Number of Rentals
0
2000
4000
6000
8000
0
1
Holiday

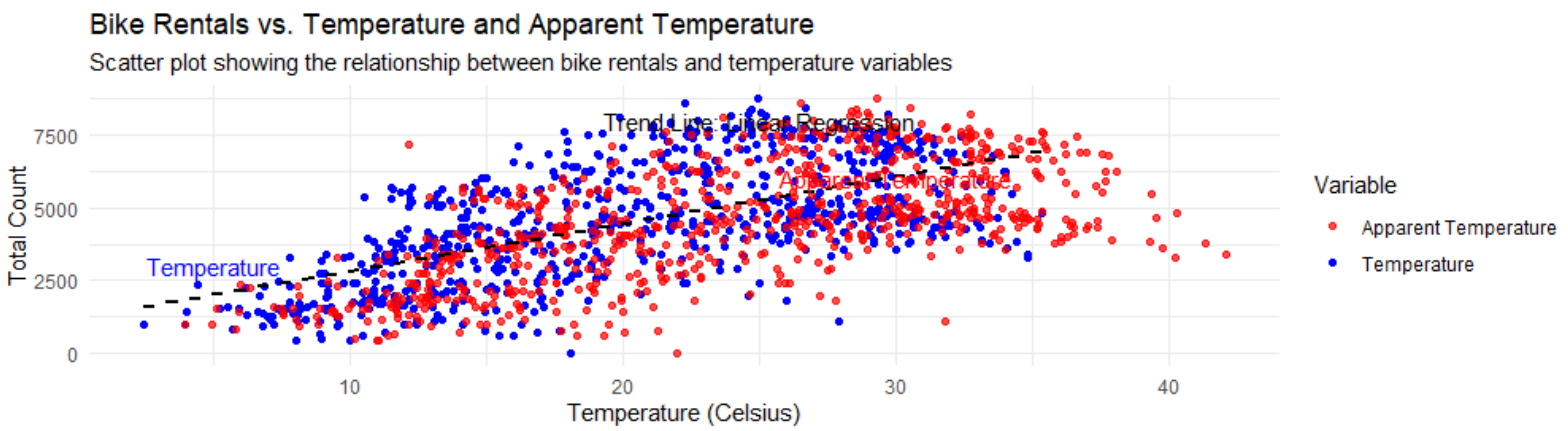

tplot of Rentals by Holida
Median: 3351
IQR: 4083
bution of Rentals on Nonholiday
Mean: 4527.1
SD: 1929.01
bution of Rentals on Ho
Mean: 37
SD: 2103
Number of Rentals
0
2000
4000
6000
8000
0
1
Holiday
Frequency
0
50
100
150
0
4000
8000
Number of Rentals
Frequency
0
1
2
3
4
5
6
0
4000
8000
Number of Rentals

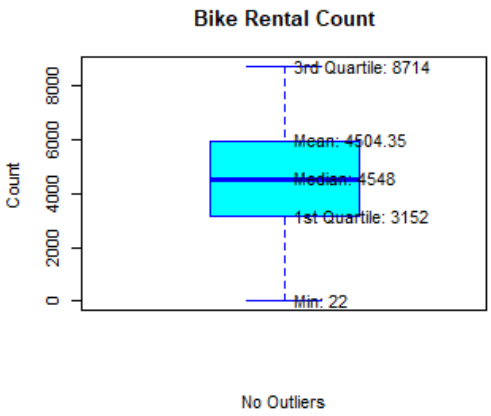
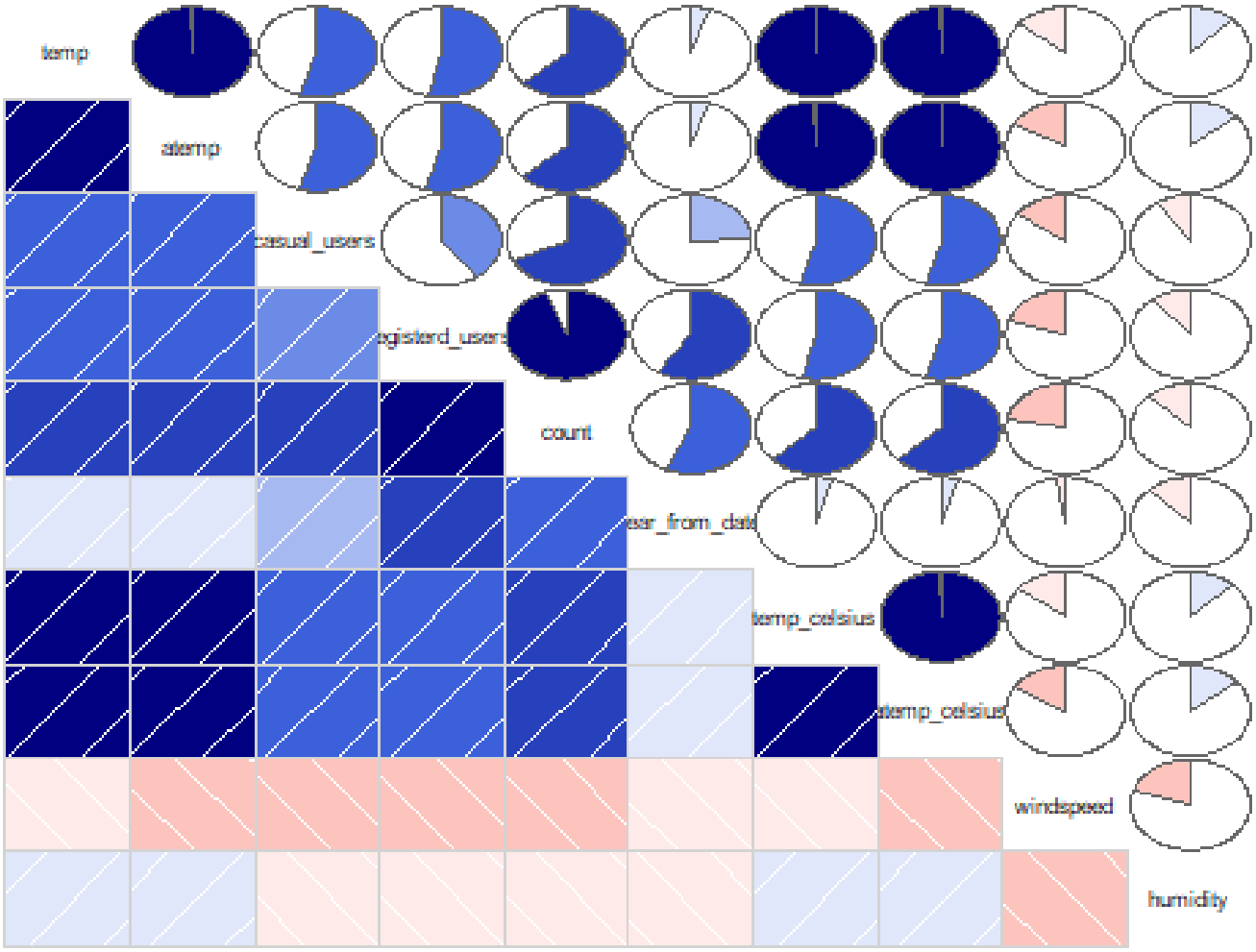

Bike Rentals Analysis: Holidays vs. Non-Holidays
Comparing counts on holidays and non-holidays across differen
Season
1
2
3
4
non-holidays, bike rental counts are highest compared to holidays for diff
Total Count
10000
7500
5000
2500
0
7836
8714
8555
3129
4098
3126
7403
3351
6269
3368
2425
1495
Non-Holiday
Holiday
Holiday


Bike Rentals Analysis: Holidays vs. Non-Holidays
Comparing counts on holidays and non-holidays across different seasons
Season 1 2 3 4
Insight: During non-holidays, bike rental counts are highest compared to holidays for different seasons.
Total Count
10000
7500
5000
2500
0
7836
8362
8714
8555
3129
6370
7403
6269
2710
1115
4098
3126
3351
5478
5117
3368
2425
1495
Non-Holiday
Holiday
Holiday

Bike Rentals Analysis: neiher weekend nor holiday vs. other days
Comparing counts on working days and non-working days across different seasons
Season 1 2 3 4
Insight: neiher weekend nor holiday - Bike rental count is higher.
other day - Bike rental count is lower
Total Count
10000
7500
5000
2500
0
neiher weekend nor holiday
other days
Working Day


Impact of Weather Conditions on Bike Rentals
Comparing bike rental counts under different weather conditions across different seasons
Season 1 2 3 4
2. Mist + Cloudy - Bike rental count is second highest.
3. Light Snow/Light Rain - Bike rental count is third highest.
4. Heavy Rain - there is no data of Bike renting.
Total Count
10000
7500
5000
2500
0
6312
8294
8714
8555
7836
8362
7852
2169
1685
4459
4639
Clear
Mist + Cloudy
Light Snow/Light Rain
Weather Condition


Bike Rentals vs. Temperature and Apparent Temperature
Scatter plot showing the relationship between bike rentals and temperature variables
Trend Line: Linear Regression
Apparent Temperature
Temperature
Total Count
7500
5000
2500
0
10
20
30
40
Temperature (Celsius)
Variable
Apparent Temperature
Temperature


Bike Rental Count
Count
0
2000
4000
6000
8000
3rd Quartile: 8714
Mean: 4504.35
Median: 4548
1st Quartile: 3152
Min: 22
No Outliers

Correlation Plot