
Program Summary - Data_Prep_Outlier.sas

Execution Environment

Author: u63876948
File: /home/u63876948/Portfolio/Outliers/Data_Prep_Outlier.sas
SAS Platform: Linux LIN X64 5.14.0-284.30.1.el9_2.x86_64
SAS Host: ODAWS02-USW2-2.ODA.SAS.COM
SAS Version: 9.04.01M7P08062020
SAS Locale: en_US
Submission Time: 11/10/2024, 12:57:54 PM
Browser Host: 135.0.146.25
User Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/130.0.0.0 Safari/537.36
Application Server: ODAMID00-USW2-2.ODA.SAS.COM

Code: Data_Prep_Outlier.sas

```
/*1. load the csv bike rental data into a sas dataset*/

libname A4 '/home/u63876948/Portfolio/Outliers';

proc import datafile='/home/u63876948/Portfolio/Outliers/bike_rental.csv'
out=A4.bike_rental
dbms=csv;
getnames=yes;
run;

proc print data=a4.bike_rental(obs=5);
run;

/*2. use proc means to examine the list of variables, their minimum and maximum*/

proc means data= a4.bike_rental min mean median max;
run;

title 'Distribution of hum';
proc sgplot data=a4.bike_rental;
    histogram hum;
    density hum;
run;

title "Box plot of humidity from bike dataset";
proc sgplot data=a4.bike_rental;
    vbox hum;
run;

title 'Scatter plot of humidity and count of rental bike';
proc sgplot data=a4.bike_rental;
    scatter x=hum y=cnt;
    xaxis grid;
    yaxis grid;
run;

/*3. Output the outliers for the variable hum based on 2 standard deviation*/

proc means data=a4.bike_rental;
    var hum;
run;

%let hum_mean = 0.63;
%let hum_std = 0.14;

title "Outliers of hum based on 2 Standard Deviations";
data _NULL_;
    set a4.bike_rental (keep=instant hum cnt);
    file print;
    if hum < &hum_mean - 2*&hum_std and not missing(hum)
    or hum > &hum_mean + 2*&hum_std then put
    instant= hum= cnt=;
run;

/*4. Output the outliers for the variable hum based on the interquantile range method.*/

title "Outliers of hum based on interquantile range";

proc univariate data= a4.bike_rental;
    var hum;
run;

%let Q1 = 0.52;
%let Q3 = 0.73;
%let IQR = 0.21;
```

```

data _null_;
  set a4.bike_rental (keep=instant hum cnt);
  file print;
  if hum < &Q1 -1.5*&IQR and not missing(hum)
  or hum > &Q3 +1.5*&IQR then put
  instant= hum= cnt=;
run;

/*5. As it can be seen from the box plot and scatter plot that there are 2 dominant outliers,
which are not in +,- the range of 1.5*IQR.
I will get rid of these strong outliers from the dataset first using the following code*/

data a4.bike_OutliersRemoved;
  set a4.bike_rental;
  if &Q1 -1.5*&IQR <= hum <= &Q3 +1.5*&IQR;
run;

/*6. Box plot humidity from bike_OutliersRemoved*/

Title 'Distribution of hum after remove outlier using IQR';
proc sgplot data=a4.bike_OutliersRemoved;
  histogram hum;
  density hum;
run;

/*Histogram of the variable windspeed*/

title 'Distribution of windspeed';
proc sgplot data=a4.bike_rental;
  histogram windspeed;
  density windspeed;
run;

/* Box plot of the variable windspeed */
title "Box plot of windspeed from bike dataset";
proc sgplot data= a4.bike_rental;
  vbox windspeed;
run;

title 'Scatter plot of windspeed and count of rental bike';
proc sgplot data=a4.bike_rental;
  scatter x=windspeed y=cnt;
  xaxis grid;
  yaxis grid;
run;

/*7. Output the outliers for the variable windspeed based on the interquartile range method.*/
title "Outliers of windspeed based on Interquartile Range";
proc univariate data=a4.bike_rental;
  var windspeed;
run;

%let Q1 = 0.135;
%let Q3 = 0.233;
%let IQR = 0.098;

data _NULL_;
  set a4.bike_rental;
  file print;
  if windspeed < &Q1-1.5*&IQR and not missing(windspeed)
  or windspeed > &Q3+1.5*&IQR then put
  instant= windspeed= cnt=;
run;

/*8. Using correlation function to see the association*/

title 'Correlation';
proc corr data=a4.bike_rental plots=matrix(hist);
  var windspeed cnt;
run;

proc corr data=a4.bike_rental spearman plots=matrix(hist);
  var windspeed cnt;
run;

```

Log: Data_Prep_Outlier.sas

Notes (45)

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69      /*1. load the csv bike rental data into a sas dataset*/
70
71      libname A4 '/home/u63876948/Portfolio/Outliers';
NOTE: Libref A4 was successfully assigned as follows:
      Engine:          V9
      Physical Name: /home/u63876948/Portfolio/Outliers
72
73      proc import datafile='/home/u63876948/Portfolio/Outliers/bike_rental.csv'
74      out=A4.bike_rental
75      dbms=csv;
76      getnames=yes;
77      run;

NOTE: Unable to open parameter catalog: SASUSER.PARMS.PARMS.SLIST in update mode. Temporary parameter values will be saved to
WORK.PARMS.PARMS.SLIST.
78      /*****
79      *   PRODUCT:    SAS
80      *   VERSION:    9.4
81      *   CREATOR:    External File Interface
82      *   DATE:       10NOV24
83      *   DESC:       Generated SAS Daststep Code
84      *   TEMPLATE SOURCE: (None Specified.)
85      *****/
86      data A4.BIKE_RENTAL ;
87      %let _EFIERR_ = 0; /* set the ERROR detection macro variable */
88      infile '/home/u63876948/Portfolio/Outliers/bike_rental.csv' delimiter = ',' MISSOVER DSD lrecl=32767 firstobs=2
89      informat instant best32. ;
90      informat dteday yymmdd10. ;
91      informat season best32. ;
92      informat yr best32. ;
93      informat mnth best32. ;
94      informat holiday best32. ;
95      informat weekday best32. ;
96      informat workingday best32. ;
97      informat weathersit best32. ;
98      informat temp best32. ;
99      informat atemp best32. ;
100     informat hum best32. ;
101     informat windspeed best32. ;
102     informat casual best32. ;
103     informat registered best32. ;
104     informat cnt best32. ;
105     format instant best12. ;
106     format dteday yymmdd10. ;
107     format season best12. ;
108     format yr best12. ;
109     format mnth best12. ;
110     format holiday best12. ;
111     format weekday best12. ;
112     format workingday best12. ;
113     format weathersit best12. ;
114     format temp best12. ;
115     format atemp best12. ;
116     format hum best12. ;
117     format windspeed best12. ;
118     format casual best12. ;
119     format registered best12. ;
120     format cnt best12. ;
121     input
122         instant
123         dteday
124         season
125         yr
126         mnth
127         holiday
128         weekday
129         workingday
130         weathersit
131         temp
132         atemp
133         hum
134         windspeed
135         casual
136         registered
137         cnt
138     ;
139     if _ERROR_ then call symputx('_EFIERR_',1); /* set ERROR detection macro variable */
140     run;

```

NOTE: The infile '/home/u63876948/Portfolio/Outliers/bike_rental.csv' is:
 Filename=/home/u63876948/Portfolio/Outliers/bike_rental.csv,
 Owner Name=u63876948,Group Name=oda,
 Access Permission=-rw-r--r--,
 Last Modified=10Nov2024:12:55:49,
 File Size (bytes)=57569

```
NOTE: 731 records were read from the infile '/home/u63876948/Portfolio/Outliers/bike_rental.csv'.
      The minimum record length was 61.
      The maximum record length was 81.
NOTE: The data set A4.BIKE_RENTAL has 731 observations and 16 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              10477.06k
      OS Memory           32284.00k
      Timestamp           11/10/2024 05:57:53 PM
      Step Count          48  Switch Count  1
      Page Faults         0
      Page Reclaims       161
      Page Swaps           0
      Voluntary Context Switches  29
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 272
```

731 rows created in A4.BIKE_RENTAL from /home/u63876948/Portfolio/Outliers/bike_rental.csv.

```
NOTE: A4.BIKE_RENTAL data set was successfully created.
NOTE: The data set A4.BIKE_RENTAL has 731 observations and 16 variables.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.08 seconds
      user cpu time       0.03 seconds
      system cpu time     0.01 seconds
      memory              10477.06k
      OS Memory           32544.00k
      Timestamp           11/10/2024 05:57:53 PM
      Step Count          48  Switch Count  8
      Page Faults         0
      Page Reclaims       2535
      Page Swaps           0
      Voluntary Context Switches 133
      Involuntary Context Switches 2
      Block Input Operations 0
      Block Output Operations 360
```

```
141
142      proc print data=a4.bike_rental(obs=5);
143      run;
```

```
NOTE: There were 5 observations read from the data set A4.BIKE_RENTAL.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.01 seconds
      user cpu time       0.02 seconds
      system cpu time     0.00 seconds
      memory              1224.53k
      OS Memory           26792.00k
      Timestamp           11/10/2024 05:57:53 PM
      Step Count          49  Switch Count  0
      Page Faults         0
      Page Reclaims       70
      Page Swaps           0
      Voluntary Context Switches 6
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 8
```

```
144
145      /*2. use proc means to examine the list of variables, their minimum and maximum*/
146
147      proc means data= a4.bike_rental min mean median max;
148      run;
```

```
NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.04 seconds
      user cpu time       0.04 seconds
      system cpu time     0.01 seconds
      memory              6980.43k
      OS Memory           31932.00k
      Timestamp           11/10/2024 05:57:53 PM
      Step Count          50  Switch Count  1
      Page Faults         0
      Page Reclaims       1352
      Page Swaps           0
      Voluntary Context Switches 28
      Involuntary Context Switches 3
      Block Input Operations 0
      Block Output Operations 0
```

```
149
150      title 'Distribution of hum';
151      proc sgplot data=a4.bike_rental;
152          histogram hum;
153          density hum;
154      run;
```

```
NOTE: PROCEDURE SGPLLOT used (Total process time):
```

```

real time          0.11 seconds
user cpu time      0.04 seconds
system cpu time    0.00 seconds
memory             8317.32k
OS Memory          32816.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         51  Switch Count  1
Page Faults        0
Page Reclaims      1666
Page Swaps         0
Voluntary Context Switches 177
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 768

```

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

```

155
156     title "Box plot of humidity from bike dataset";
157     proc sgplot data=a4.bike_rental;
158         vbox hum;
159     run;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

```

real time          0.07 seconds
user cpu time      0.02 seconds
system cpu time    0.00 seconds
memory             2142.28k
OS Memory          32816.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         52  Switch Count  1
Page Faults        0
Page Reclaims      340
Page Swaps         0
Voluntary Context Switches 175
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 416

```

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

```

160
161     title 'Scatter plot of humidity and count of rental bike';
162     proc sgplot data=a4.bike_rental;
163         scatter x=hum y=cnt;
164         xaxis grid;
165         yaxis grid;
166     run;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

```

real time          0.09 seconds
user cpu time      0.03 seconds
system cpu time    0.00 seconds
memory             1889.37k
OS Memory          32816.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         53  Switch Count  1
Page Faults        0
Page Reclaims      294
Page Swaps         0
Voluntary Context Switches 140
Involuntary Context Switches 2
Block Input Operations 0
Block Output Operations 488

```

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

```

167
168     /*3. Output the outliers for the variable hum based on 2 standard deviation*/
169
170     proc means data=a4.bike_rental;
171         var hum;
172     run;

```

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time          0.01 seconds
user cpu time      0.01 seconds
system cpu time    0.01 seconds
memory             7270.32k
OS Memory          38092.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         54  Switch Count  1
Page Faults        0
Page Reclaims      1618
Page Swaps         0
Voluntary Context Switches 26
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 8

```

```

173
174     %let hum_mean = 0.63;
175     %let hum_std = 0.14;
176
177     title "Outliers of hum based on 2 Standard Deviations";
178     data _NULL_;

```

```

179      set a4.bike_rental (keep=instant hum cnt);
180      file print;
181      if hum < &hum_mean - 2*&hum_std and not missing(hum)
182      or hum > &hum_mean + 2*&hum_std then put
183      instant= hum= cnt=;
184      run;

```

NOTE: 27 lines were written to file PRINT.

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

NOTE: DATA statement used (Total process time):

```

real time          0.00 seconds
user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory             756.93k
OS Memory          31912.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         55  Switch Count  0
Page Faults        0
Page Reclaims      63
Page Swaps         0
Voluntary Context Switches  6
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

```

185
186      /*4. Output the outliers for the variable hum based on the interquantile range method.*/
187
188      title "Outliers of hum based on interquantile range";
189
190      proc univariate data= a4.bike_rental;
191      var hum;
192      run;

```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

```

real time          0.03 seconds
user cpu time      0.03 seconds
system cpu time    0.00 seconds
memory             961.87k
OS Memory          31912.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         56  Switch Count  0
Page Faults        0
Page Reclaims      66
Page Swaps         0
Voluntary Context Switches  8
Involuntary Context Switches 3
Block Input Operations  0
Block Output Operations 16

```

```

193
194      %let Q1 = 0.52;
195      %let Q3 = 0.73;
196      %let IQR = 0.21;
197
198      data _null_;
199      set a4.bike_rental (keep=instant hum cnt);
200      file print;
201      if hum < &Q1 -1.5*&IQR and not missing(hum)
202      or hum > &Q3 +1.5*&IQR then put
203      instant= hum= cnt=;
204      run;

```

NOTE: Variable hum is uninitialized.

NOTE: 2 lines were written to file PRINT.

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

NOTE: DATA statement used (Total process time):

```

real time          0.00 seconds
user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory             758.15k
OS Memory          31912.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         57  Switch Count  0
Page Faults        0
Page Reclaims      59
Page Swaps         0
Voluntary Context Switches  6
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

```

205
206      /*5. As it can be seen from the box plot and scatter plot that there are 2 dominant outliers,
207      which are not in +/- the range of 1.5*IQR.
208      I will get rid of these strong outliers from the dataset first using the following code*/
209
210      data a4.bike_OutliersRemoved;
211      set a4.bike_rental;
212      if &Q1 -1.5*&IQR <= hum <= &Q3 +1.5*&IQR;
213      run;

```

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

NOTE: The data set A4.BIKE_OUTLIERSREMOVED has 729 observations and 16 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.01 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory             965.00k
OS Memory          32172.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         58  Switch Count  1
Page Faults        0
Page Reclaims      97
Page Swaps         0
Voluntary Context Switches 33
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 264

```

```

214
215      /*6. Box plot humidity from bike_OutliersRemoved*/
216
217      Title 'Distribution of hum after remove outlier using IQR';
218      proc sgplot data=a4.bike_OutliersRemoved;
219      histogram hum;
220      density hum;
221      run;

```

NOTE: PROCEDURE SGPLLOT used (Total process time):

```

real time          0.08 seconds
user cpu time      0.03 seconds
system cpu time    0.01 seconds
memory             2184.01k
OS Memory          32816.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         59  Switch Count  1
Page Faults        0
Page Reclaims      299
Page Swaps         0
Voluntary Context Switches 175
Involuntary Context Switches 2
Block Input Operations 0
Block Output Operations 496

```

NOTE: There were 729 observations read from the data set A4.BIKE_OUTLIERSREMOVED.

```

222
223      /*Histogram of the variable windspeed*/
224
225      title 'Distribution of windspeed';
226      proc sgplot data=a4.bike_rental;
227      histogram windspeed;
228      density windspeed;
229      run;

```

NOTE: PROCEDURE SGPLLOT used (Total process time):

```

real time          0.07 seconds
user cpu time      0.03 seconds
system cpu time    0.00 seconds
memory             2333.62k
OS Memory          32816.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         60  Switch Count  1
Page Faults        0
Page Reclaims      300
Page Swaps         0
Voluntary Context Switches 169
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 456

```

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

```

230
231      /* Box plot of the variable windspeed */
232      title "Box plot of windspeed from bike dataset";
233      proc sgplot data= a4.bike_rental;
234      vbox windspeed;
235      run;

```

NOTE: PROCEDURE SGPLLOT used (Total process time):

```

real time          0.08 seconds
user cpu time      0.03 seconds
system cpu time    0.00 seconds
memory             2277.93k
OS Memory          33072.00k
Timestamp          11/10/2024 05:57:53 PM
Step Count         61  Switch Count  1
Page Faults        0
Page Reclaims      308
Page Swaps         0
Voluntary Context Switches 210
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 424

```

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

```

236
237      title 'Scatter plot of windspeed and count of rental bike';
238      proc sgplot data=a4.bike_rental;

```

```

239      scatter x=windspeed y=cnt;
240      xaxis grid;
241      yaxis grid;
242      run;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

```

real time      0.09 seconds
user cpu time   0.04 seconds
system cpu time 0.01 seconds
memory         1857.46k
OS Memory      33072.00k
Timestamp      11/10/2024 05:57:53 PM
Step Count     62  Switch Count  1
Page Faults    0
Page Reclaims  307
Page Swaps     0
Voluntary Context Switches 139
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 488

```

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

```

243
244      /*7. Output the outliers for the variable windspeed based on the interquantile range method.*/
245      title "Outliers of windspeed based on Interquartile Range";
246      proc univariate data=a4.bike_rental;
247      var windspeed;
248      run;

```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

```

real time      0.03 seconds
user cpu time   0.04 seconds
system cpu time 0.00 seconds
memory         1059.43k
OS Memory      32168.00k
Timestamp      11/10/2024 05:57:53 PM
Step Count     63  Switch Count  0
Page Faults    0
Page Reclaims  63
Page Swaps     0
Voluntary Context Switches 7
Involuntary Context Switches 3
Block Input Operations 0
Block Output Operations 0

```

```

249
250      %let Q1 = 0.135;
251      %let Q3 = 0.233;
252      %let IQR = 0.098;
253
254      data _NULL_;
255      set a4.bike_rental;
256      file print;
257      if windspeed < &Q1-1.5*&IQR and not missing(windspeed)
258      or windspeed > &Q3+1.5*&IQR then put
259      instant= windspeed= cnt=;
260      run;

```

NOTE: 13 lines were written to file PRINT.

NOTE: There were 731 observations read from the data set A4.BIKE_RENTAL.

NOTE: DATA statement used (Total process time):

```

real time      0.01 seconds
user cpu time   0.00 seconds
system cpu time 0.00 seconds
memory         763.21k
OS Memory      32168.00k
Timestamp      11/10/2024 05:57:53 PM
Step Count     64  Switch Count  0
Page Faults    0
Page Reclaims  61
Page Swaps     0
Voluntary Context Switches 6
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 16

```

```

261
262      /*8. Using correlation function to see the association*/
263
264      title 'Correlation';
265      proc corr data=a4.bike_rental plots=matrix(hist);
266      var windspeed cnt;
267      run;

```

NOTE: PROCEDURE CORR used (Total process time):

```

real time      0.13 seconds
user cpu time   0.06 seconds
system cpu time 0.01 seconds
memory         4027.81k
OS Memory      34140.00k
Timestamp      11/10/2024 05:57:54 PM
Step Count     65  Switch Count  0
Page Faults    0
Page Reclaims  528
Page Swaps     0

```



```
Voluntary Context Switches      209
Involuntary Context Switches    4
Block Input Operations           0
Block Output Operations         600

268
269      proc corr data=a4.bike_rental spearman plots=matrix(hist);
270      var windspeed cnt;
271      run;

NOTE: PROCEDURE CORR used (Total process time):
      real time      0.13 seconds
      user cpu time  0.06 seconds
      system cpu time 0.00 seconds
      memory         3618.34k
      OS Memory      34652.00k
      Timestamp      11/10/2024 05:57:54 PM
      Step Count     66  Switch Count  0
      Page Faults    0
      Page Reclaims  483
      Page Swaps     0
      Voluntary Context Switches 209
      Involuntary Context Switches 2
      Block Input Operations      0
      Block Output Operations    592

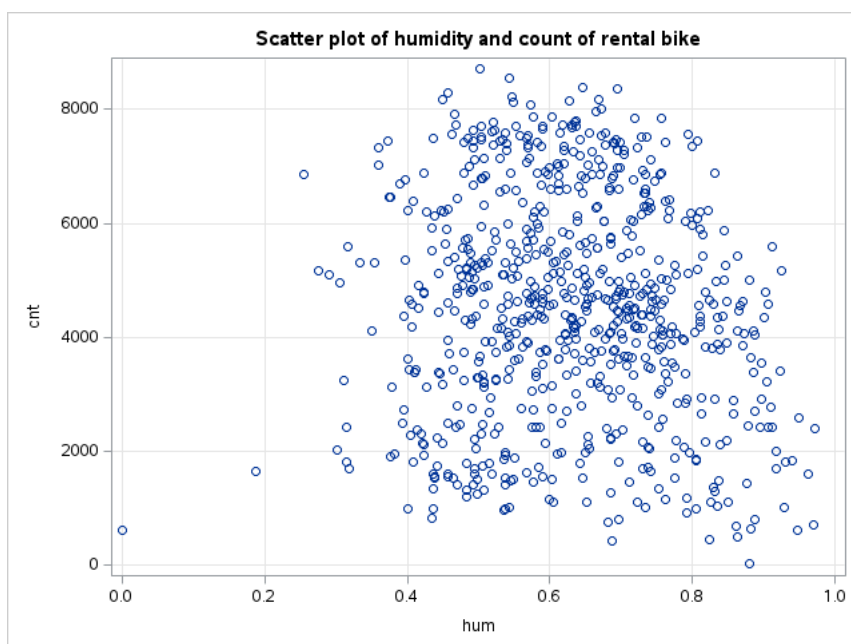
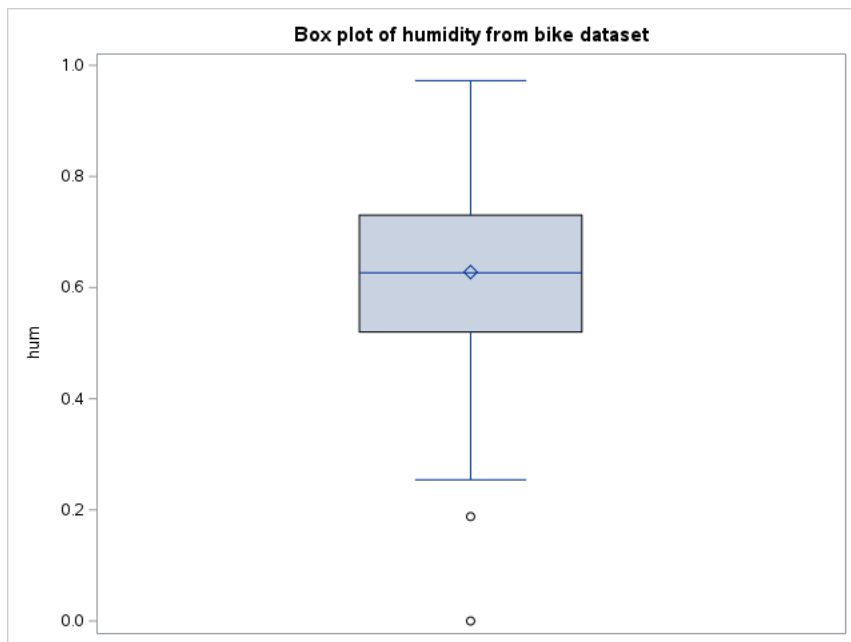
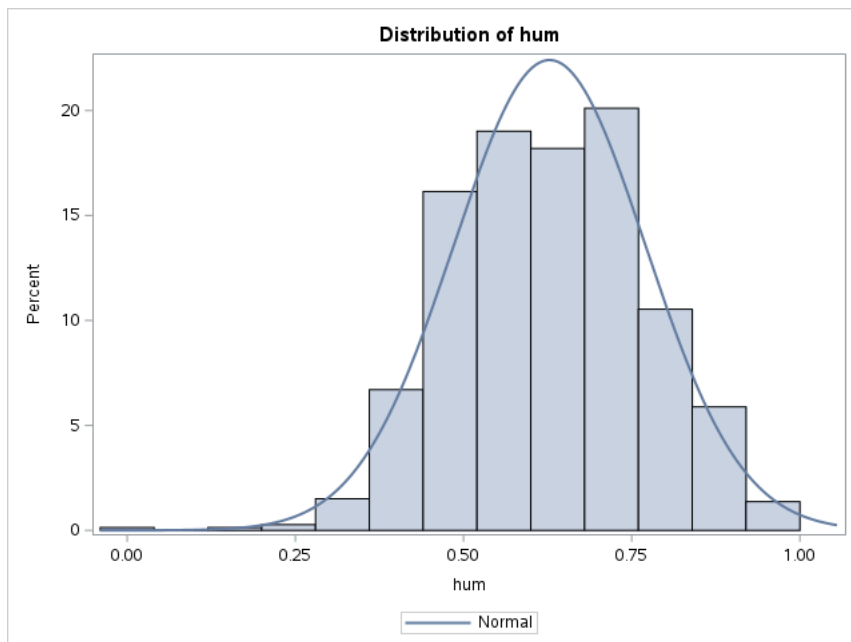
272
273
274
275
276
277
278
279
280
281
282
283      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
293
```

Results: Data_Prep_Outlier.sas

Obs	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
1	1	2011-01-01	1	0	1	0	6	0	2	0.344167	0.363625	0.805833	0.160446	331	654	985
2	2	2011-01-02	1	0	1	0	0	0	2	0.363478	0.353739	0.696087	0.248539	131	670	801
3	3	2011-01-03	1	0	1	0	1	1	1	0.196364	0.189405	0.437273	0.248309	120	1229	1349
4	4	2011-01-04	1	0	1	0	2	1	1	0.2	0.212122	0.590435	0.160296	108	1454	1562
5	5	2011-01-05	1	0	1	0	3	1	1	0.226957	0.22927	0.436957	0.1869	82	1518	1600

The MEANS Procedure

Variable	Minimum	Mean	Median	Maximum
instant	1.0000000	366.0000000	366.0000000	731.0000000
dteday	18628.00	18993.00	18993.00	19358.00
season	1.0000000	2.4965800	3.0000000	4.0000000
yr	0	0.5006840	1.0000000	1.0000000
mnth	1.0000000	6.5198358	7.0000000	12.0000000
holiday	0	0.0287278	0	1.0000000
weekday	0	2.9972640	3.0000000	6.0000000
workingday	0	0.6839945	1.0000000	1.0000000
weathersit	1.0000000	1.3853488	1.0000000	3.0000000
temp	0.0591304	0.4953848	0.4983330	0.8616670
atemp	0.0790696	0.4743540	0.4867330	0.8408960
hum	0	0.6278941	0.6266670	0.9725000
windspeed	0.0223917	0.1904862	0.1809750	0.5074630
casual	2.0000000	848.1764706	713.0000000	3410.00
registered	20.0000000	3656.17	3662.00	6946.00
cnt	22.0000000	4504.35	4548.00	8714.00



Scatter plot of humidity and count of rental bike

The MEANS Procedure

Analysis Variable : hum				
N	Mean	Std Dev	Minimum	Maximum
731	0.6278941	0.1424291	0	0.9725000

Outliers of hum based on 2 Standard Deviations

instant=36	hum=0.929167	cnt=1005
instant=46	hum=0.314348	cnt=1815
instant=50	hum=0.187917	cnt=1635
instant=62	hum=0.318333	cnt=1685
instant=65	hum=0.948261	cnt=605
instant=69	hum=0	cnt=623
instant=87	hum=0.302174	cnt=2028
instant=88	hum=0.314167	cnt=2425
instant=90	hum=0.918333	cnt=1685
instant=134	hum=0.9225	cnt=3409
instant=153	hum=0.305	cnt=4968
instant=250	hum=0.917083	cnt=1996
instant=251	hum=0.939565	cnt=1842
instant=266	hum=0.9725	cnt=2395
instant=320	hum=0.93	cnt=1817
instant=326	hum=0.9625	cnt=1607
instant=340	hum=0.949583	cnt=2594
instant=341	hum=0.970417	cnt=705
instant=388	hum=0.91125	cnt=2432
instant=394	hum=0.31125	cnt=3243
instant=452	hum=0.29	cnt=5102
instant=463	hum=0.254167	cnt=6857
instant=464	hum=0.275833	cnt=5169
instant=465	hum=0.3175	cnt=5585
instant=678	hum=0.333478	cnt=5315
instant=708	hum=0.91125	cnt=5582
instant=710	hum=0.925	cnt=5170

Outliers of hum based on interquartile range

The UNIVARIATE Procedure
Variable: hum

Moments			
N	731	Sum Weights	731
Mean	0.62789406	Sum Observations	458.99056
Std Deviation	0.1424291	Variance	0.02028605
Skewness	-0.0697834	Kurtosis	-0.0645301
Uncorrected SS	303.006262	Corrected SS	14.8088144
Coeff Variation	22.6836187	Std Error Mean	0.00526793

Basic Statistical Measures			
Location		Variability	
Mean	0.627894	Std Deviation	0.14243
Median	0.626667	Variance	0.02029
Mode	0.613333	Range	0.97250
		Interquartile Range	0.21042

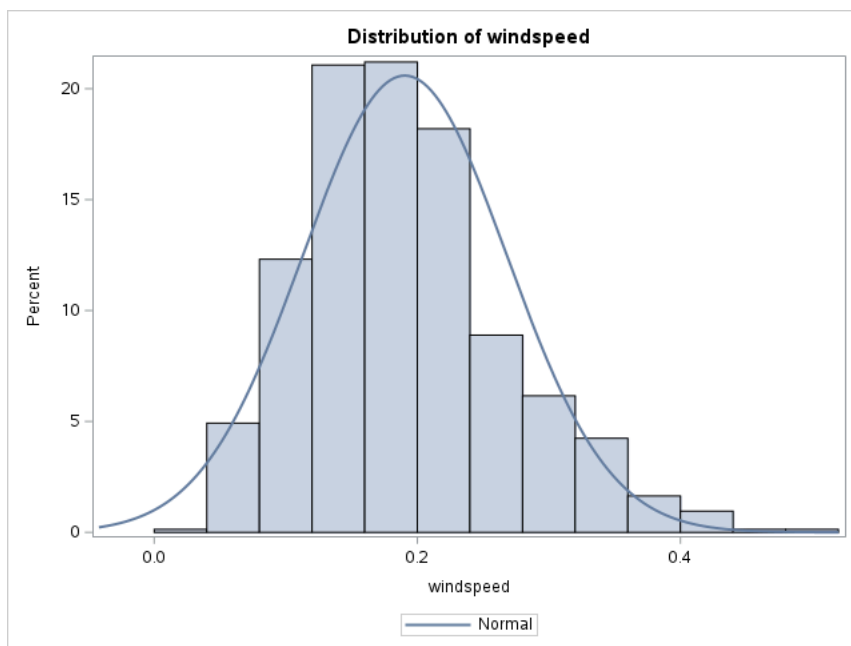
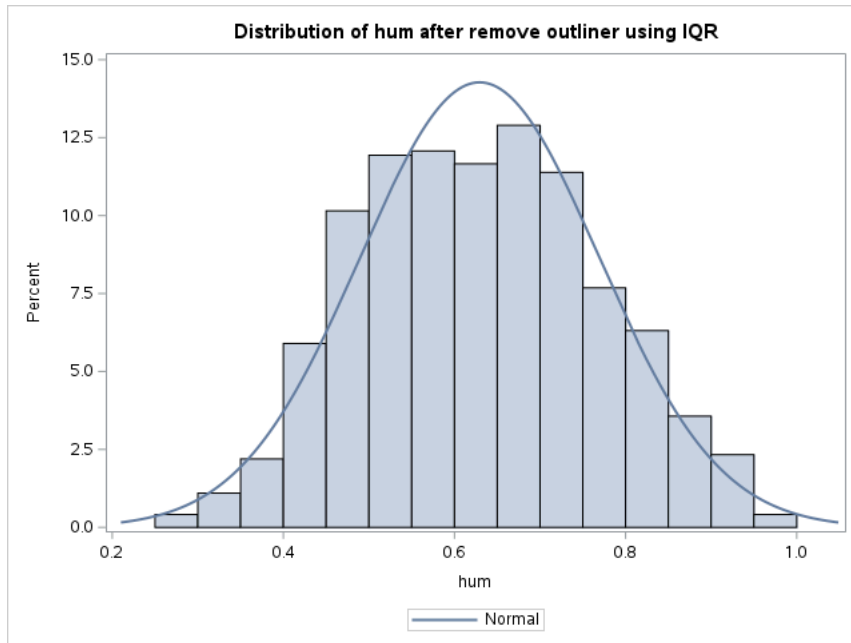
Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	119.1918	Pr > t	<.0001
Sign	M	365	Pr >= M	<.0001
Signed Rank	S	133407.5	Pr >= S	<.0001

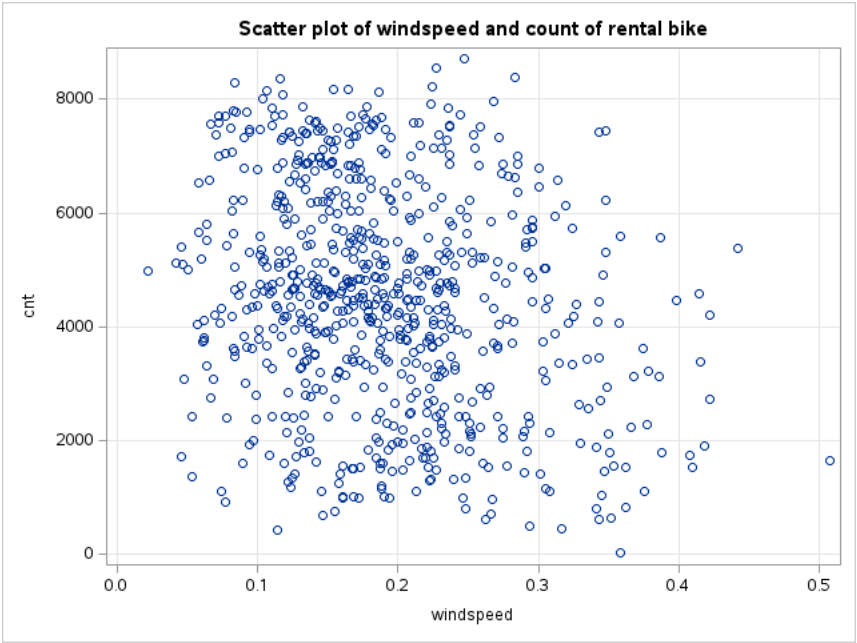
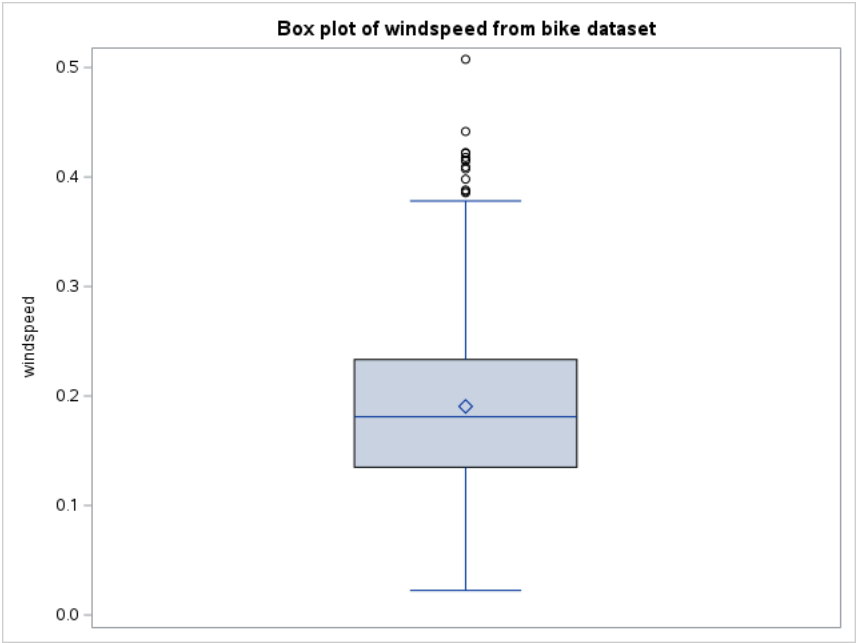
Quantiles (Definition 5)	
Level	Quantile
100% Max	0.972500
99%	0.929167
95%	0.870000
90%	0.817500
75% Q3	0.730417
50% Median	0.626667
25% Q1	0.520000
10%	0.450000
5%	0.407083
1%	0.311250
0% Min	0.000000

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.000000	69	0.948261	65
0.187917	50	0.949583	340
0.254167	463	0.962500	326
0.275833	464	0.970417	341
0.290000	452	0.972500	266

Outliers of hum based on interquartile range

```
instant=50 hum=0.187917 cnt=1635  
instant=69 hum=0 cnt=623
```





Outliers of windspeed based on Interquartile Range

The UNIVARIATE Procedure
Variable: windspeed

Moments			
N	731	Sum Weights	731
Mean	0.19048621	Sum Observations	139.245421
Std Deviation	0.07749787	Variance	0.00600592
Skewness	0.67734542	Kurtosis	0.41092227
Uncorrected SS	30.9086542	Corrected SS	4.38432157
Coeff Variation	40.6842417	Std Error Mean	0.00286636

Basic Statistical Measures			
Location		Variability	
Mean	0.190486	Std Deviation	0.07750
Median	0.180975	Variance	0.00601
Mode	0.106350	Range	0.48507
		Interquartile Range	0.09827

Note: The mode displayed is the smallest of 9 modes with a count of 3.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	66.45573	Pr > t	<.0001
Sign	M	365.5	Pr >= M	<.0001
Signed Rank	S	133773	Pr >= S	<.0001

Quantiles (Definition 5)	
Level	Quantile

Quantiles (Definition 5)	
Level	Quantile
100% Max	0.5074630
99%	0.4092120
95%	0.3432870
90%	0.2960290
75% Q3	0.2332210
50% Median	0.1809750
25% Q1	0.1349500
10%	0.1001330
5%	0.0783833
1%	0.0528708
0% Min	0.0223917

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.0223917	280	0.417908	45
0.0423042	283	0.421642	421
0.0454042	281	0.422275	293
0.0454083	38	0.441563	433
0.0466500	696	0.507463	50

Outliers of windspeed based on Interquartile Range

instant=45	windspeed=0.417908	cnt=1913
instant=50	windspeed=0.507463	cnt=1635
instant=94	windspeed=0.385571	cnt=3115
instant=95	windspeed=0.388067	cnt=1795
instant=293	windspeed=0.422275	cnt=4195
instant=383	windspeed=0.415429	cnt=3376
instant=408	windspeed=0.409212	cnt=1529
instant=421	windspeed=0.421642	cnt=2732
instant=433	windspeed=0.441563	cnt=5382
instant=434	windspeed=0.4148	cnt=4569
instant=451	windspeed=0.386821	cnt=5558
instant=667	windspeed=0.398008	cnt=4459
instant=722	windspeed=0.407346	cnt=1749

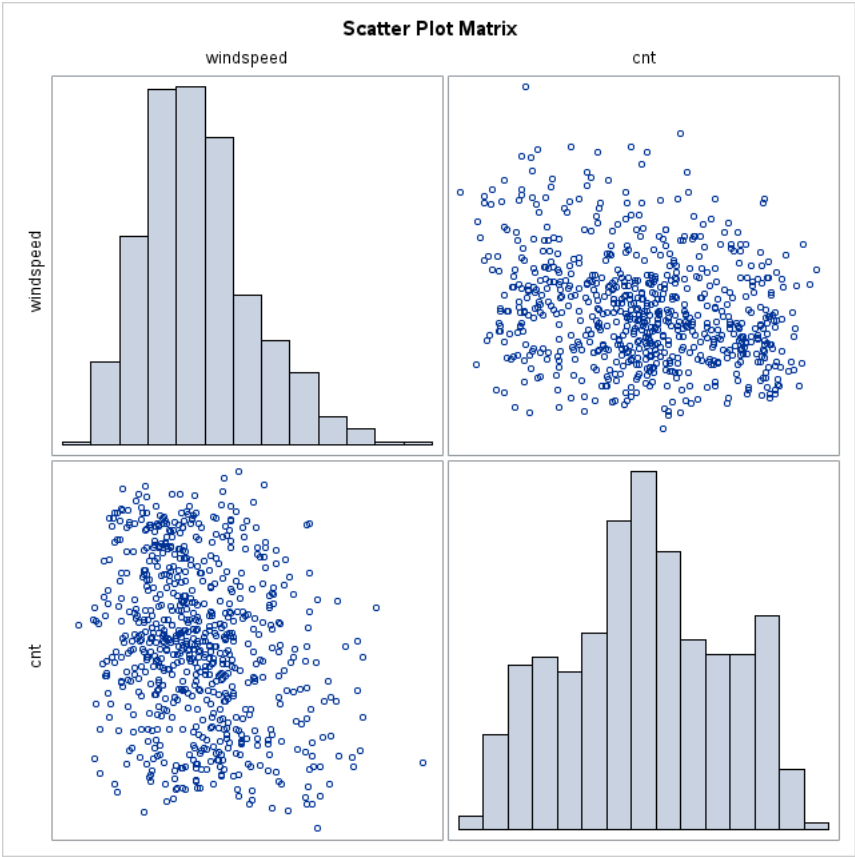
Correlation

The CORR Procedure

2 Variables:	windspeed cnt
--------------	---------------

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
windspeed	731	0.19049	0.07750	139.24542	0.02239	0.50746
cnt	731	4504	1937	3292679	22.00000	8714

Pearson Correlation Coefficients, N = 731 Prob > r under H0: Rho=0		
	windspeed	cnt
windspeed	1.00000	-0.23454 <.0001
cnt	-0.23454 <.0001	1.00000



Correlation

The CORR Procedure

2 Variables: windspeed cnt

Simple Statistics						
Variable	N	Mean	Std Dev	Median	Minimum	Maximum
windspeed	731	0.19049	0.07750	0.18098	0.02239	0.50746
cnt	731	4504	1937	4548	22.00000	8714

Pearson Correlation Coefficients, N = 731 Prob > r under H0: Rho=0		
	windspeed	cnt
windspeed	1.00000	-0.23454 <.0001
cnt	-0.23454 <.0001	1.00000

Spearman Correlation Coefficients, N = 731 Prob > r under H0: Rho=0		
	windspeed	cnt
windspeed	1.00000	-0.21720 <.0001
cnt	-0.21720 <.0001	1.00000

