

Program Summary - BAN130 Project Code Group_6.sas

Execution Environment

Author: u60685886
 File: /home/u60685886/BAN130 Workshops and Quiz/BAN130 Project Code Group_6.sas
 SAS Platform: Linux LIN X64 3.10.0-1062.9.1.el7.x86_64
 SAS Host: ODAWS02-USW2.ODA.SAS.COM
 SAS Version: 9.04.01M6P11072018
 SAS Locale: en_US
 Submission Time: 4/14/2022, 12:05:39 PM
 Browser Host: 157.49.244.211
 User Agent: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/100.0.4896.88 Safari/537.36
 Application Server: ODAMID02-USW2.ODA.SAS.COM

Code: BAN130 Project Code Group_6.sas

```

/* BAN130 Final Project - Group 6
   Arunadevi Krishnan
   Dhananjay Kumar
   Manish Kumar
   Pratik Harshad Gherwada */

/* importing the flight delay file */

FILENAME REFFILE '/home/u60685886/BAN130 Workshops and Quiz/FlightDelaysMain.xlsx';

PROC IMPORT DATAFILE=REFFILE DBMS=XLSX OUT=flight;
RUN;

title 'First 100 observation of FlightsDelays dataset';
proc print data=flight(obs=100);
run;

title 'Identifying the missing values using mean proc';
proc means data=flight nmiss n;
run;

*Fixing the time variable dep_time crs_dep_time;
data flight_new;
set flight;
departure_time = input(put(dep_time,z4.),hhmmss4.) ;
format departure_time time5. ;
crs_departure_time = input(put(crs_dep_time,z4.),hhmmss4.) ;
format crs_departure_time time5.;
drop dep_time crs_dep_time;
rename departure_time=dep_time;
rename crs_departure_time=crs_dep_time;
run;

*Section A;

*Step 1;
proc format;
  value missingcount
  .='missing' other='notmissing';
  value $Missingchar ' '='Missing' other='NonMissing';
run;

proc freq data=flight_new;
  table CRS_DEP_TIME DEP_TIME DISTANCE FL_DATE FL_NUM carrier dest flight_status
  origin tail_num Weather DAY_WEEK DAY_OF_MONTH /missing;
  format CRS_DEP_TIME missingcount. DEP_TIME missingcount. DISTANCE
  missingcount. FL_DATE missingcount. FL_NUM missingcount. Weather
  missingcount. DAY_WEEK missingcount. DAY_OF_MONTH
  missingcount. carrier $Missingchar. dest $Missingchar.
  flight_status $Missingchar. origin $Missingchar. tail_num $Missingchar.;
run;

data flight_new2;
set flight_new;
if missing(crs_dep_time) or missing(dep_time) or missing(distance) or
  missing(fl_date) or missing(fl_num) or missing(carrier) or missing(dest) or

```

```
missing(dest) or missing(flight_status) or missing(origin) or
missing(tail_num) then
    delete;

run;

title 'Checking missing values after handling';
proc means data=flight_new2 nmiss n;
run;

*Step 2;

data flightdelays replace;
set flight_new2;
where Origin="DCA";
if Flight_status="delayed" then
    DelayedFlight=1;
else if Flight_status="ontime" then
    DelayedFlight=0;
if Flight_status="delayed" then
    delay_time_minutes=(dep_time-crs_dep_time)/60;
else delay_time_minutes=0;
run;

*Step 3;

data flight_delay replace;
set flightdelays;
where fl_date<01/09/2004;
where flight_status="delayed";
run;

proc sql;
create table delayed_avg as
select avg(delay_time_minutes) as delayed_avg_minutes, dest from flight_delay group by dest;
run;

proc sgplot data=delayed_avg;
yaxis label="delayed_avg_minutes" max=60;
vbar dest / response=delayed_avg_minutes;
run;

*Step 4;

proc sql;
create table flight_count as
select count(distinct fl_num) as flight_count, carrier, fl_date
from flightdelays group by carrier,fl_date;

create table mean_flights as
select round(avg(flight_count)) as mean_flights, carrier from flight_count group by carrier;
run;

title "Flights per day for Carrier RU";
proc sgplot data=flight_count ;
where carrier="RU";
scatter y=flight_count x=fl_date ;
run;

title "Average flights per day by Carrier";
proc sgplot data=mean_flights ;
yaxis label="mean_flights" max=25;
vbar carrier / response=mean_flights;
run;

*Step 5;

proc univariate data=flightdelays;
var distance delay_time_minutes;
histogram ;
run;

*Step 6;

ods noproctitle;
ods graphics / imagemap=on;

title "Flight status by Origin";
proc means data=flight_new2 nonobs chartype n vardef=df;
var FL_NUM;
class ORIGIN Flight_status;
run;
```

```

title "Flight Status for different carrier";
proc means data=flight_new2 nonobs chartype n vardef=df;
    var FL_NUM;
    class Flight_status CARRIER;
run;

title "Destination and distance affecting flight status ";
proc means data=flight_new2 nonobs chartype n vardef=df;
    var FL_NUM;
    class Flight_status DEST DISTANCE;
run;

title "Weather affecting flight status";
proc means data=flight_new2 nonobs chartype n vardef=df;
    var FL_NUM;
    class Flight_status Weather ORIGIN;
run;

*Section B;

*step-1;
data FlightDelaysTrainingData replace;
set flightdelays (drop= DAY_WEEK DAY_OF_MONTH TAIL_NUM flight_status );
run;

proc export data=FlightDelaysTrainingData
outfile="/home/u60685886/BAN130 Workshops and Quiz/SAS output/FlightDelaysTrainingData" dbms=csv;
run;

title 'Contents on the data reduction data set - FlightDelaysTrainingData';
proc contents data= FlightDelaysTrainingData varnum;
run;

*Step 2;

data FlightDelaysTrainingData;
set FlightDelaysTrainingData;
if carrier="CO" then carrier_num=1;
else if carrier='DH' then carrier_num=2;
else if carrier='DL' then carrier_num=3;
else if carrier='MQ' then carrier_num=4;
else if carrier='OH' then carrier_num=5;
else if carrier='RU' then carrier_num=6;
else if carrier='US' then carrier_num=7;
else if carrier='UA' then carrier_num=9;
if origin="BWI" then origin_num=1;
else if origin="DCA" then origin_num=2;
else if origin="IAD" then origin_num=3;
if dest="EWR" then dest_num=1;
else if dest="JFK" then dest_num=2;
else if dest="LGA" then dest_num=3;
drop carrier origin dest;
rename origin_num=origin;
rename carrier_num=carrier;
rename dest_num=dest;
run;

title 'Data after conversion';
proc print data=FlightDelaysTrainingData (obs=5);
run;

*Creating a reference tables for data conversion;
proc sql;
create table referencetable
(carrier_no num, carrier char(20),
origin_no num, origin char(20),
dest_no num, dest char(20));
quit;

Proc sql;
Insert into referencetable values(1, 'CO',1,'BWI',1,'EWR');
Insert into referencetable values(2, 'DH',2,'DCA',2,'JFK');
Insert into referencetable values(3, 'DL',3,'IAD',3,'LGA');
Insert into referencetable (carrier_no,carrier)values (4, 'MQ');
Insert into referencetable (carrier_no,carrier) values(5, 'OH');
Insert into referencetable (carrier_no,carrier) values(6, 'RU');
Insert into referencetable (carrier_no,carrier) values(7, 'US');
Insert into referencetable (carrier_no,carrier) values(8, 'UA');

```

```

title 'Data after conversion reference table';
proc sql;
select origin_no,origin from referencetable
where origin_no=2;

```

Log: BAN130 Project Code Group_6.sas

Notes (93)

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
NOTE: ODS statements in the SAS Studio environment may disable some output features.
69
70      /* BAN130 Final Project - Group 6
71      Arunadevi Krishnan
72      Dhananjay Kumar
73      Manish Kumar
74      Pratik Harshad Gherwada*/
75
76      /* importing the flight delay file */
77
78      FILENAME REFFILE '/home/u60685886/BAN130 Workshops and Quiz/FlightDelaysMain.xlsx';
79
80      PROC IMPORT DATAFILE=REFFILE DBMS=XLSX OUT=flight;
81      RUN;

NOTE: Import cancelled. Output dataset WORK.FLIGHT already exists. Specify REPLACE option to overwrite it.
NOTE: The SAS System stopped processing this step because of errors.
NOTE: PROCEDURE IMPORT used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              689.34k
      OS Memory           36244.00k
      Timestamp           04/14/2022 06:35:36 AM
      Step Count          92  Switch Count  0
      Page Faults         0
      Page Reclaims       138
      Page Swaps          0
      Voluntary Context Switches 0
      Involuntary Context Switches 0
      Block Input Operations 0
      Block Output Operations 0

82
83      title 'First 100 observation of FlightsDelays dataset';

84
85      proc print data=flight(obs=100);
      run;

NOTE: There were 100 observations read from the data set WORK.FLIGHT.
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.21 seconds
      user cpu time       0.22 seconds
      system cpu time     0.00 seconds
      memory              1591.40k
      OS Memory           36008.00k
      Timestamp           04/14/2022 06:35:36 AM
      Step Count          93  Switch Count  0
      Page Faults         0
      Page Reclaims       61
      Page Swaps          0
      Voluntary Context Switches 0
      Involuntary Context Switches 1
      Block Input Operations 0
      Block Output Operations 88

86
87      title 'Identifying the missing values using mean proc';
88      proc means data=flight nmiss n;
89      run;

NOTE: There were 2201 observations read from the data set WORK.FLIGHT.
NOTE: PROCEDURE MEANS used (Total process time):
      real time           0.03 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              6513.40k
      OS Memory           41148.00k
      Timestamp           04/14/2022 06:35:36 AM
      Step Count          94  Switch Count  1
      Page Faults         0
      Page Reclaims       1347

```

```

Page Swaps                0
Voluntary Context Switches 19
Involuntary Context Switches 0
Block Input Operations     0
Block Output Operations     0

```

```

90
91      *Fixing the time variable dep_time crs_dep_time;
92      data flight_new;
93      set flight;
94      departure_time = input(put(dep_time,z4.),hhmmss4.) ;
95      format departure_time time5. ;
96      crs_departure_time = input(put(crs_dep_time,z4.),hhmmss4.) ;
97      format crs_departure_time time5.;
98      drop dep_time crs_dep_time;
99      rename departure_time=dep_time;
100     rename crs_departure_time=crs_dep_time;
101     run;

```

NOTE: There were 2201 observations read from the data set WORK.FLIGHT.
 NOTE: The data set WORK.FLIGHT_NEW has 2201 observations and 13 variables.
 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            1220.84k
OS Memory          36268.00k
Timestamp          04/14/2022 06:35:36 AM
Step Count         95  Switch Count  2
Page Faults        0
Page Reclaims      95
Page Swaps         0
Voluntary Context Switches 11
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 528

```

```

102
103
104      *Section A;
105
106      *Step 1;
107      proc format;
108      ! value missingcount
109      .='missing' other='notmissing';
NOTE: Format MISSINGCOUNT is already on the library WORK.FORMATS.
NOTE: Format MISSINGCOUNT has been output.
110
111      ! value $Missingchar ' '= 'Missing' other='NonMissing';
NOTE: Format $MISSINGCHAR is already on the library WORK.FORMATS.
NOTE: Format $MISSINGCHAR has been output.
112      run;

```

NOTE: PROCEDURE FORMAT used (Total process time):

```

real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory            247.59k
OS Memory          35748.00k
Timestamp          04/14/2022 06:35:36 AM
Step Count         96  Switch Count  0
Page Faults        0
Page Reclaims      14
Page Swaps         0
Voluntary Context Switches 0
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 32

```

```

112
113     proc freq data=flight_new;
114     table CRS_DEP_TIME DEP_TIME DISTANCE FL_DATE FL_NUM carrier dest flight_status
115     origin tail_num Weather DAY_WEEK DAY_OF_MONTH /missing;
116     format CRS_DEP_TIME missingcount. DEP_TIME missingcount. DISTANCE
117     missingcount. FL_DATE missingcount. FL_NUM missingcount. Weather
118     missingcount. DAY_WEEK missingcount. DAY_OF_MONTH
119     missingcount. carrier $missingchar. dest $missingchar.
120     flight_status $missingchar. origin $missingchar. tail_num $missingchar.;
121     run;

```

NOTE: There were 2201 observations read from the data set WORK.FLIGHT_NEW.
 NOTE: PROCEDURE FREQ used (Total process time):
 real time 0.09 seconds

```

user cpu time      0.09 seconds
system cpu time    0.00 seconds
memory             1389.68k
OS Memory          36268.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         97  Switch Count  2
Page Faults        0
Page Reclaims      127
Page Swaps         0
Voluntary Context Switches  9
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 296

```

```

122
123     data flight_new2;
124     set flight_new;
125     if missing(crs_dep_time) or missing(dep_time) or missing(distance) or
126     missing(fl_date) or missing(fl_num) or missing(carrier) or missing(dest) or
127     missing(dest) or missing(flight_status) or missing(origin) or
128     missing(tail_num) then
129     delete;
130     run;

```

NOTE: There were 2201 observations read from the data set WORK.FLIGHT_NEW.

NOTE: The data set WORK.FLIGHT_NEW2 has 2190 observations and 13 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory             1262.25k
OS Memory          36268.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         98  Switch Count  2
Page Faults        0
Page Reclaims      95
Page Swaps         0
Voluntary Context Switches 16
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 528

```

```

131
132     title 'Checking missing values after handling';
133     proc means data=flight_new2 nmiss n;
134     run;

```

NOTE: There were 2190 observations read from the data set WORK.FLIGHT_NEW2.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time          0.03 seconds
user cpu time      0.03 seconds
system cpu time    0.01 seconds
memory             6304.50k
OS Memory          41148.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         99  Switch Count  1
Page Faults        0
Page Reclaims      1346
Page Swaps         0
Voluntary Context Switches 20
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

```

135
136     *Step 2;
137
138     data flightdelays replace;
139     set flight_new2;
140     where Origin="DCA";
141     if Flight_status="delayed" then
142     DelayedFlight=1;
143     else if Flight_status="ontime" then
144     DelayedFlight=0;
145     if Flight_status="delayed" then
146     delay_time_minutes=(dep_time-crs_dep_time)/60;
147     else delay_time_minutes=0;
148     run;

```

NOTE: There were 1363 observations read from the data set WORK.FLIGHT_NEW2.

WHERE Origin='DCA';

NOTE: The data set WORK.FLIGHTDELAYS has 1363 observations and 15 variables.

NOTE: The data set WORK.REPLACE has 1363 observations and 15 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory            1553.84k
OS Memory          36528.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         100   Switch Count  15
Page Faults        0
Page Reclaims      128
Page Swaps         0
Voluntary Context Switches  63
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 1040

```

```

149
150      *Step 3;
151      data flight_delay replace;
152      set flightdelays;
153      where fl_date<01/09/2004;
154      where flight_status="delayed";
NOTE: WHERE clause has been replaced.
155      run;

```

NOTE: There were 221 observations read from the data set WORK.FLIGHTDELAYS.

WHERE flight_status='delayed';

NOTE: The data set WORK.FLIGHT_DELAY has 221 observations and 15 variables.

NOTE: The data set WORK.REPLACE has 221 observations and 15 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory            1489.84k
OS Memory          36528.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         101   Switch Count  6
Page Faults        0
Page Reclaims      177
Page Swaps         0
Voluntary Context Switches  35
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 528

```

```

156
157      proc sql;
158      create table delayed_avg as
159      select avg(delay_time_minutes) as delayed_avg_minutes, dest from flight_delay group by dest;
NOTE: Table WORK.DELAYED_AVG created, with 3 rows and 2 columns.

```

```

160      run;
NOTE: PROC SQL statements are executed immediately; The RUN statement has no effect.
161

```

NOTE: PROCEDURE SQL used (Total process time):

```

real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory            5601.71k
OS Memory          41132.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         102   Switch Count  2
Page Faults        0
Page Reclaims      150
Page Swaps         0
Voluntary Context Switches  15
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations 272

```

```

162      proc sgplot data=delayed_avg;
163      yaxis label="delayed_avg_minutes" max=60;
164      vbar dest / response=delayed_avg_minutes;
165      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            7340.03k
OS Memory          38448.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         103   Switch Count  2
Page Faults        0
Page Reclaims      914

```

Page Swaps	0
Voluntary Context Switches	160
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	608

NOTE: There were 3 observations read from the data set WORK.DELAYED_AVG.

```

166
167      *Step 4;
168      proc sql;
169      create table flight_count as
170      select count(distinct fl_num) as flight_count, carrier, fl_date
171      from flightdelays group by carrier,fl_date;

```

NOTE: Table WORK.FLIGHT_COUNT created, with 182 rows and 3 columns.

```

172
173      create table mean_flights as
174      select round(avg(flight_count)) as mean_flights, carrier from flight_count group by carrier;

```

NOTE: Table WORK.MEAN_FLIGHTS created, with 6 rows and 2 columns.

```

175      run;
NOTE: PROC SQL statements are executed immediately; The RUN statement has no effect.

```

```

176
177      title "Flights per day for Carrier RU";

```

NOTE: PROCEDURE SQL used (Total process time):

real time	0.00 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	5764.18k
OS Memory	42668.00k
Timestamp	04/14/2022 06:35:37 AM
Step Count	104
Page Faults	0
Page Reclaims	260
Page Swaps	0
Voluntary Context Switches	46
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	544

```

178      proc sgplot data=flight_count ;
179      where carrier="RU";
180      scatter y=flight_count x=fl_date ;
181      run;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

real time	0.08 seconds
user cpu time	0.03 seconds
system cpu time	0.00 seconds
memory	2048.50k
OS Memory	38708.00k
Timestamp	04/14/2022 06:35:37 AM
Step Count	105
Page Faults	0
Page Reclaims	334
Page Swaps	0
Voluntary Context Switches	149
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	392

NOTE: The column format MMDDYY10 is replaced by an auto-generated format on the axis.

NOTE: There were 31 observations read from the data set WORK.FLIGHT_COUNT.
WHERE carrier='RU';

```

182
183      title "Average flights per day by Carrier";
184      proc sgplot data=mean_flights ;
185      yaxis label="mean_flights" max=25;
186      vbar carrier / response=mean_flights;
187      run;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

real time	0.08 seconds
user cpu time	0.04 seconds
system cpu time	0.00 seconds
memory	2444.50k
OS Memory	38572.00k
Timestamp	04/14/2022 06:35:37 AM
Step Count	106
Page Faults	0
Page Reclaims	440
Page Swaps	0
Voluntary Context Switches	163
Involuntary Context Switches	0


```

Block Input Operations      0
Block Output Operations     368

```

NOTE: There were 6 observations read from the data set WORK.MEAN_FLIGHTS.

```

188
189      *Step 5;
190      proc univariate data=flightdelays;
191      var distance delay_time_minutes;
192      histogram ;
193      run;

```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

```

real time      0.24 seconds
user cpu time  0.17 seconds
system cpu time 0.01 seconds
memory         7540.68k
OS Memory      43524.00k
Timestamp      04/14/2022 06:35:37 AM
Step Count     107  Switch Count  0
Page Faults    0
Page Reclaims  2003
Page Swaps     0
Voluntary Context Switches 239
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 504

```

```

194
195      *Step 6;
196      ods noproctitle;
197      ods graphics / imagemap=on;
198
199      title "Flight status by Origin";
200      proc means data=flight_new2 nonobs chartype n vardef=df;
201      var FL_NUM;
202      class ORIGIN Flight_status;
203      run;

```

NOTE: There were 2190 observations read from the data set WORK.FLIGHT_NEW2.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time      0.02 seconds
user cpu time  0.02 seconds
system cpu time 0.01 seconds
memory         8757.06k
OS Memory      47036.00k
Timestamp      04/14/2022 06:35:37 AM
Step Count     108  Switch Count  1
Page Faults    0
Page Reclaims  1934
Page Swaps     0
Voluntary Context Switches 29
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 0

```

```

204
205      title "Flight Status for different carrier";
206      proc means data=flight_new2 nonobs chartype n vardef=df;
207      var FL_NUM;
208      class Flight_status CARRIER;
209      run;

```

NOTE: There were 2190 observations read from the data set WORK.FLIGHT_NEW2.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time      0.02 seconds
user cpu time  0.02 seconds
system cpu time 0.01 seconds
memory         8755.62k
OS Memory      47036.00k
Timestamp      04/14/2022 06:35:37 AM
Step Count     109  Switch Count  1
Page Faults    0
Page Reclaims  1930
Page Swaps     0
Voluntary Context Switches 29
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 0

```

```

210
211      title "Destination and distance affecting flight status ";
212      proc means data=flight_new2 nonobs chartype n vardef=df;
213      var FL_NUM;

```

```

214      class Flight_status DEST DISTANCE;
215      run;

```

NOTE: There were 2190 observations read from the data set WORK.FLIGHT_NEW2.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time          0.02 seconds
user cpu time      0.03 seconds
system cpu time    0.00 seconds
memory            8472.28k
OS Memory          45500.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         110  Switch Count  1
Page Faults        0
Page Reclaims      1820
Page Swaps         0
Voluntary Context Switches  26
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

```

216
217      title "Weather affecting flight status";
218      proc means data=flight_new2 nonobs chartype n vardef=df;
219      var FL_NUM;
220      class Flight_status Weather ORIGIN;
221      run;

```

NOTE: There were 2190 observations read from the data set WORK.FLIGHT_NEW2.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time          0.02 seconds
user cpu time      0.02 seconds
system cpu time    0.01 seconds
memory            8472.96k
OS Memory          45500.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         111  Switch Count  1
Page Faults        0
Page Reclaims      1807
Page Swaps         0
Voluntary Context Switches  27
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  16

```

```

222
223      *Section B;
224
225      *step-1;
226      data FlightDelaysTrainingData replace;
227      set flightdelays (drop= DAY_WEEK DAY_OF_MONTH TAIL_NUM flight_status );
228      run;

```

NOTE: There were 1363 observations read from the data set WORK.FLIGHTDELAYS.

NOTE: The data set WORK.FLIGHTDELAYSTRAININGDATA has 1363 observations and 11 variables.

NOTE: The data set WORK.REPLACE has 1363 observations and 11 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory            1390.43k
OS Memory          38576.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         112  Switch Count  4
Page Faults        0
Page Reclaims      138
Page Swaps         0
Voluntary Context Switches  21
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  528

```

```

229
230      proc export data=FlightDelaysTrainingData
231      outfile="/home/u60685886/BAN130 Workshops and Quiz/SAS output/FlightDelaysTrainingData" dbms=csv;
232      run;

```

NOTE: Export cancelled. Output file /home/u60685886/BAN130 Workshops and Quiz/SAS output/FlightDelaysTrainingData already exists. Specify REPLACE option to overwrite it.

NOTE: The SAS System stopped processing this step because of errors.

NOTE: PROCEDURE EXPORT used (Total process time):

```

real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory            251.56k

```

```

OS Memory          37796.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         113  Switch Count  0
Page Faults        0
Page Reclaims      15
Page Swaps         0
Voluntary Context Switches  2
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  8

```

```

233
234      title 'Contents on the data reduction data set - FlightDelaysTrainingData';

```

```

235      proc contents data= FlightDelaysTrainingData varnum;
236      run;

```

NOTE: PROCEDURE CONTENTS used (Total process time):

```

real time          0.04 seconds
user cpu time      0.04 seconds
system cpu time    0.00 seconds
memory            924.84k
OS Memory          38316.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         114  Switch Count  0
Page Faults        0
Page Reclaims      93
Page Swaps         0
Voluntary Context Switches  2
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  24

```

```

237
238      *Step 2;
239
240      data FlightDelaysTrainingData;
241      set FlightDelaysTrainingData;
242      if carrier="CO" then carrier_num=1;
243      else if carrier="DH" then carrier_num=2;
244      else if carrier="DL" then carrier_num=3;
245      else if carrier="MQ" then carrier_num=4;
246      else if carrier="OH" then carrier_num=5;
247      else if carrier="RU" then carrier_num=6;
248      else if carrier="US" then carrier_num=7;
249      else if carrier="UA" then carrier_num=9;
250      if origin="BWI" then origin_num=1;
251      else if origin="DCA" then origin_num=2;
252      else if origin="IAD" then origin_num=3;
253      if dest="EWR" then dest_num=1;
254      else if dest="JFK" then dest_num=2;
255      else if dest="LGA" then dest_num=3;
256      drop carrier origin dest;
257      rename origin_num=origin;
258      rename carrier_num=carrier;
259      rename dest_num=dest;
260      run;

```

NOTE: There were 1363 observations read from the data set WORK.FLIGHTDELAYSTRAININGDATA.

NOTE: The data set WORK.FLIGHTDELAYSTRAININGDATA has 1363 observations and 11 variables.

NOTE: DATA statement used (Total process time):

```

real time          0.00 seconds
user cpu time      0.00 seconds
system cpu time    0.00 seconds
memory            1020.34k
OS Memory          38316.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         115  Switch Count  2
Page Faults        0
Page Reclaims      106
Page Swaps         0
Voluntary Context Switches  15
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  264

```

```

261
262      title 'Data after conversion';
263      proc print data=FlightDelaysTrainingData (obs=5);
264      run;

```

NOTE: There were 5 observations read from the data set WORK.FLIGHTDELAYSTRAININGDATA.

NOTE: PROCEDURE PRINT used (Total process time):

```

real time          0.01 seconds

```

```

user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory             686.46k
OS Memory          38056.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         116  Switch Count  0
Page Faults        0
Page Reclaims      61
Page Swaps         0
Voluntary Context Switches  0
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

```

265
266      *Creating a reference tables for data conversion;
267      proc sql;
268      create table referencetable
269      (carrier_no num,carrier char(20),
270      origin_no num, origin char(20),
271      dest_no num, dest char(20));
NOTE: Table WORK.REFERENCETABLE created, with 0 rows and 6 columns.
272      quit;

```

```

NOTE: PROCEDURE SQL used (Total process time):
real time          0.00 seconds
user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory             629.34k
OS Memory          38056.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         117  Switch Count  2
Page Faults        0
Page Reclaims      78
Page Swaps         0
Voluntary Context Switches  14
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  264

```

```

273
274      Proc sql;
275      Insert into referencetable values(1, 'CO',1,'BWI',1,'EWR');
NOTE: 1 row was inserted into WORK.REFERENCETABLE.

```

```

276      Insert into referencetable values(2, 'DH',2,'DCA',2,'JFK');
NOTE: 1 row was inserted into WORK.REFERENCETABLE.

```

```

277      Insert into referencetable values(3, 'DL',3,'IAD',3,'LGA');
NOTE: 1 row was inserted into WORK.REFERENCETABLE.

```

```

278      Insert into referencetable (carrier_no,carrier)values (4, 'MQ');
NOTE: 1 row was inserted into WORK.REFERENCETABLE.

```

```

279      Insert into referencetable (carrier_no,carrier) values(5, 'OH');
NOTE: 1 row was inserted into WORK.REFERENCETABLE.

```

```

280      Insert into referencetable (carrier_no,carrier) values(6, 'RU');
NOTE: 1 row was inserted into WORK.REFERENCETABLE.

```

```

281      Insert into referencetable (carrier_no,carrier) values(7, 'US');
NOTE: 1 row was inserted into WORK.REFERENCETABLE.

```

```

282      Insert into referencetable (carrier_no,carrier) values(8, 'UA');
NOTE: 1 row was inserted into WORK.REFERENCETABLE.

```

```

283
284      title 'Data after conversion reference table';
NOTE: PROCEDURE SQL used (Total process time):
real time          0.00 seconds
user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory             5781.78k
OS Memory          43180.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         118  Switch Count  1
Page Faults        0
Page Reclaims      302
Page Swaps         0
Voluntary Context Switches  9
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

```

285      proc sql;

```

```

286      select origin_no,origin from referencetable
287      where origin_no=2;
288
289      *Step-3;
290      title "Fitting a Prediction Model for Flight Delay";
291      *Removed dest, origin and distance field as they did not add value to the model,
292      removing those variables increased the R-squared value by approx 15%;

```

NOTE: PROCEDURE SQL used (Total process time):

```

real time          0.00 seconds
user cpu time      0.01 seconds
system cpu time    0.00 seconds
memory            5577.28k
OS Memory          43180.00k
Timestamp          04/14/2022 06:35:37 AM
Step Count         119   Switch Count   1
Page Faults        0
Page Reclaims      54
Page Swaps         0
Voluntary Context Switches 10
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 0

```

```

293      proc pls data=FlightDelaysTrainingData plots=all;
294          model delayedflight=weather|carrier;
295      run;

```

NOTE: Maximum number of factors to extract set to 3. This may be too high; you should consider using either the NFAC= option or the CV= option to set the number of factors.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: HTML data tips have been disabled for at least one plot because the threshold has been reached. You can set TIPMAX=1400 in the ODS GRAPHICS statement to produce data tips for all plots.

NOTE: There were 1363 observations read from the data set WORK.FLIGHTDELAYSTRAININGDATA.

NOTE: PROCEDURE PLS used (Total process time):

```

real time          2.60 seconds
user cpu time      0.89 seconds
system cpu time    0.15 seconds
memory            15292.59k
OS Memory          50640.00k
Timestamp          04/14/2022 06:35:40 AM
Step Count         120   Switch Count   42
Page Faults        0
Page Reclaims      21989
Page Swaps         0
Voluntary Context Switches 12246
Involuntary Context Switches 3
Block Input Operations 0
Block Output Operations 8080

```

```

296
297      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
307

```

Results: BAN130 Project Code Group_6.sas

First 100 observation of FlightsDelays dataset

Obs	CRS_DEP_TIME	CARRIER	DEP_TIME	DEST	DISTANCE	FL_DATE	FL_NUM	ORIGIN	Weather	DAY_WEEK	DAY_OF_MONTH	TAIL_NUM	Flight_status
1	1455	OH	1455	JFK	184	.	5935	BWI	0	4	1	N940CA	ontime
2	1640	DH	1640	JFK	213	01/01/2004	6155	DCA	0	4	1	N405FJ	ontime
3	1245	DH	1245	LGA	229	01/01/2004	7208	IAD	0	4	1	N695BR	ontime
4	1715	DH	1709	LGA	229	01/01/2004	7215	IAD	0	4	1	N662BR	ontime
5	1039	DH	1035	LGA	229	01/01/2004	7792	IAD	0	4	1	N698BR	ontime
6	840	DH	839	JFK	228	01/01/2004	7800	IAD	0	4	1	N687BR	ontime
7	1240	DH	1243	JFK	228	01/01/2004	7806	IAD	0	4	1	N321UE	ontime
8	1645	DH	1644	JFK	228	01/01/2004	7810	IAD	0	4	1	N301UE	ontime
9	1715	DH	1710	JFK	228	01/01/2004	7812	IAD	0	4	1	N328UE	ontime

Obs	CRS_DEP_TIME	CARRIER	DEP_TIME	DEST	DISTANCE	FL_DATE	FL_NUM	ORIGIN	Weather	DAY_WEEK	DAY_OF_MONTH	TAIL_NUM	Flight_status
10	2120	DH	2129	JFK	228	01/01/2004	7814	IAD	0	4	1	N685BR	ontime
11	2120	DH	2114	LGA	229	01/01/2004	7924	IAD	0	4	1	N645BR	ontime
12	1455	DL	1458		213	01/01/2004	746	DCA	0	4	1	N918DE	ontime
13	930	DL	932	LGA	214	01/01/2004	1746	DCA	0	4	1	N242DL	ontime
14	1230	DL	1228	LGA	214	01/01/2004	1752	DCA	0	4	1	N241DL	ontime
15	1430	DL	1429	LGA	214	01/01/2004	1756	DCA	0	4	1	N242DL	ontime
16	1730	DL	1728	LGA	214	01/01/2004	1762	DCA	0	4	1	N241DL	ontime
17	2030	DL	2029	LGA	214	01/01/2004	1768	DCA	0	4	1	N242DL	ontime
18	1530	MQ	1525	JFK	213	01/01/2004	4752	DCA	0	4	1	N709MQ	ontime
19	600	MQ	556	JFK	213	01/01/2004	4760	DCA	0	4	1	N717MQ	ontime
20	1830	MQ	1822	JFK	213	01/01/2004	4784	DCA	0	4	1	N707MQ	ontime
21	900	MQ	853	LGA	214	01/01/2004	4956	DCA	0	4	1	N737MQ	ontime
22	1300	MQ	1254	LGA	214	01/01/2004	4964	DCA	0	4	1	N717MQ	ontime
23	1400	MQ	1356	LGA	214	01/01/2004	4966	DCA	0	4	1	N726MQ	ontime
24	1500	MQ	1452	LGA	214	01/01/2004	4968	DCA	0	4	1	N724MQ	ontime
25	1900	MQ	1853	LGA	214	01/01/2004	4976	DCA	0	4	1	N724MQ	ontime
26	850	UA	841	LGA	229	01/01/2004	846	IAD	0	4	1	N513UA	ontime
27	900	US	858	LGA	214	01/01/2004	2164	DCA	0	4	1	N709UW	ontime
28	1100	US	1056	LGA	214	01/01/2004	2168	DCA	0	4	1	N748UW	ontime
29	1300	US	1253	LGA	214	01/01/2004	2172	DCA	0	4	1	N709UW	ontime
30	1500	US	1458	LGA	214	01/01/2004	2176	DCA	0	4	1	N748UW	ontime
31	1700	US	1655	LGA	214	01/01/2004	2180	DCA	0	4	1	N709UW	ontime
32	2100	US	2055	LGA	214	01/01/2004	2188	DCA	0	4	1	N709UW	ontime
33	1455	RU	1452	EWB	169	01/01/2004	2403	BWI	0	4	1	N14916	ontime
34	1720	RU	1710	EWB	169	01/01/2004	2675	BWI	0	4	1	N16954	ontime
35	1030	RU	1030	EWB	169	01/01/2004	2303	BWI	0	4	1	N26549	ontime
36	700	RU	656	EWB	169	01/01/2004	2703	BWI	0	4	1	N16954	ontime
37	1300	CO	1256	EWB	199	01/01/2004	808	DCA	0	4	1	N18611	ontime
38	1730	CO	1726	EWB	199	01/01/2004	814	DCA	0	4	1	N19357	ontime
39	840	DH	840	EWB	213	01/01/2004	7299	IAD	0	4	1	N691BR	ontime
40	1710	DH	1704	EWB	213	01/01/2004	7302	IAD	0	4	1	N691BR	ontime
41	1245	DH	1245	EWB	213	01/01/2004	7303	IAD	0	4	1	N697BR	ontime
42	2120	DH	2118	EWB	213	01/01/2004	7304	IAD	0	4	1	N699BR	ontime
43	1700	RU	1651	EWB	213	01/01/2004	2497	IAD	0	4	1	N12540	ontime
44	1900	RU	1850	EWB	213	01/01/2004	2385	IAD	0	4	1	N16149	ontime
45	1525	RU	1521	EWB	199	01/01/2004	2261	DCA	0	4	1	N12564	ontime
46	1900	RU	1855	EWB	199	01/01/2004	2336	DCA	0	4	1	N21537	ontime
47	1400	RU	1357	EWB	199	01/01/2004	2216	DCA	0	4	1	N15983	ontime
48	1515	RU	1508	EWB		01/01/2004	2156	IAD	0	4	1	N16149	ontime
49	1300	RU	1255	EWB	213	01/01/2004	2664	IAD	0	4	1	N12519	ontime
50	1630	RU	1625	EWB	199	01/01/2004	2181	DCA	0	4	1	N19966	ontime
51	1455	OH	1455		184	01/02/2004		BWI	0	5	2	N995CA	ontime
52	1640	DH	1641	JFK	213	01/02/2004	6155	DCA	0	5	2	N415FJ	ontime
53	1245	DH	1249	LGA	229	01/02/2004	7208	IAD	0	5	2	N688BR	ontime
54	1455	DH	1531	LGA	229	01/02/2004	7211	IAD	0	5	2	N665BR	delayed
55	1715	DH	1712	LGA	229	01/02/2004	7215	IAD	0	5	2	N639BR	ontime
56	640	DH	645	LGA	229	01/02/2004	7790	IAD	0	5	2	N686BR	ontime
57	1039	DH	1236	LGA	229	01/02/2004	7792	IAD	0	5	2	N665BR	delayed
58	840	DH	859	JFK	228	01/02/2004	7800	IAD	0	5	2	N645BR	ontime
59	1240	DH	1232	JFK	228	01/02/2004	7806	IAD	0	5	2	N332UE	ontime
60	1455	DH	1455	JFK	228	01/02/2004	7808	IAD	0	5	2	N324UE	ontime
61	1645	DH	1645	JFK	228	01/02/2004	7810	IAD	0	5	2	N305UE	ontime
62	1715	DH	1716	JFK	228	01/02/2004	7812	IAD	0	5	2	N322UE	ontime
63	2120	DH	2305	JFK	228	01/02/2004	7814	IAD	0	5	2	N657BR	delayed
64	1610	DH	1605	JFK	228	01/02/2004	7816	IAD	0	5	2	N315UE	ontime
65	2120	DH	2118	LGA	229	01/02/2004	7924	IAD	0	5	2	N709BR	ontime
66	1455	DL	1458	JFK	213	01/02/2004	746	DCA	0	5	2	N964DL	ontime
67	930	DL	930	LGA	214	01/02/2004	1746	DCA	0	5	2	N241DL	ontime
68	1230	DL	1230	LGA	214	01/02/2004	1752	DCA	0	5	2	N225DL	ontime
69	1430	DL	1427	LGA	214	01/02/2004	1756	DCA	0	5	2	N241DL	ontime
70	1730	DL	1730	LGA	214	01/02/2004	1762	DCA	0	5	2	N225DL	ontime
71	2030	DL	2028	LGA	214	01/02/2004	1768	DCA	0	5	2	N241DL	ontime
72	1530	MQ	1522	JFK	213	01/02/2004	4752	DCA	0	5	2		ontime
73	600	MQ	552	JFK	213	01/02/2004		DCA	0	5	2	N736MQ	ontime
74	1830	MQ	1847	JFK	213	01/02/2004	4784	DCA	0	5	2	N727MQ	ontime
75	900	MQ	852	LGA	214	01/02/2004	4956	DCA	0	5	2	N713MQ	ontime
76	1100	MQ	1053	LGA	214	01/02/2004	4960	DCA	0	5	2	N708MQ	ontime
77	1300	MQ		LGA	214	01/02/2004	4964	DCA	0	5	2	N713MQ	ontime
78	1400	MQ	1402	LGA	214	01/02/2004	4966	DCA	0	5	2	N718MQ	ontime
79	1500	MQ	1456	LGA	214	01/02/2004	4968	DCA	0	5	2	N708MQ	ontime
80	850	UA	850	LGA	229	01/02/2004	846	IAD	0	5	2	N556UA	ontime
81	700	US	657	LGA	214	01/02/2004	2160	DCA	0	5	2	N710UW	ontime
82	900	US	857	LGA	214	01/02/2004	2164	DCA	0	5	2	N736UW	ontime
83	1100	US	1058	LGA	214	01/02/2004	2168	DCA	0	5	2	N710UW	ontime
84	1300	US	1258	LGA	214	01/02/2004	2172	DCA	0	5	2	N736UW	ontime
85	1500	US	1458	LGA	214	01/02/2004	2176	DCA	0	5	2	N710UW	ontime
86	1700	US	1655	LGA	214	01/02/2004	2180	DCA	0	5	2	N736UW	ontime
87	1900	US	1855	LGA	214	01/02/2004	2184	DCA	0	5	2	N710UW	ontime
88	2100	US	2056	LGA	214	01/02/2004	2188	DCA	0	5	2	N736UW	ontime
89	1720	RU	1715	EWB	169	01/02/2004	2675	BWI	0	5	2	N19966	ontime
90	1030	RU	1030	EWB	169	01/02/2004	2303	BWI	0	5	2	N12540	ontime

Obs	CRS_DEP_TIME	CARRIER	DEP_TIME	DEST	DISTANCE	FL_DATE	FL_NUM	ORIGIN	Weather	DAY_WEEK	DAY_OF_MONTH	TAIL_NUM	Flight_status
91	700	RU	656	EWB	169	01/02/2004	2703	BWI	0	5	2		ontime
92	1455	RU	1456	EWB	169	01/02/2004	2403	BWI	0	5	2	N12946	ontime
93	1730	CO	1727	EWB	199	01/02/2004	814	DCA	0	5	2	N14342	ontime
94	1300	CO	1301	EWB	199	01/02/2004	808	DCA	0	5	2	N14664	ontime
95	759	CO	754	EWB	199	01/02/2004	806	DCA	0	5	2	N11641	ontime
96	840	DH	837	EWB	213	01/02/2004	7299	IAD	0	5	2	N679BR	ontime
97	1245	DH	1350	EWB	213	01/02/2004	7303	IAD	0	5	2	N686BR	delayed
98	1430	DH	1512	EWB	213	01/02/2004	7307	IAD	0	5	2	N309UE	delayed
99	630	DH	629	EWB	213	01/02/2004	7371	IAD	0	5	2	N312UE	ontime
100	1630	RU	1625	EWB	199	01/02/2004	2181	DCA	0	5	2	N14977	ontime

Identifying the missing values using mean proc

Variable	Label	N Miss	N
CRS_DEP_TIME	CRS_DEP_TIME	0	2201
DEP_TIME	DEP_TIME	2	2199
DISTANCE	DISTANCE	1	2200
FL_DATE	FL_DATE	1	2200
FL_NUM	FL_NUM	3	2198
Weather	Weather	0	2201
DAY_WEEK	DAY_WEEK	0	2201
DAY_OF_MONTH	DAY_OF_MONTH	0	2201

Identifying the missing values using mean proc

crs_dep_time	Frequency	Percent	Cumulative Frequency	Cumulative Percent
notmissing	2201	100.00	2201	100.00

dep_time	Frequency	Percent	Cumulative Frequency	Cumulative Percent
missing	2	0.09	2	0.09
notmissing	2199	99.91	2201	100.00

DISTANCE				
DISTANCE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
missing	1	0.05	1	0.05
notmissing	2200	99.95	2201	100.00

FL_DATE				
FL_DATE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
missing	1	0.05	1	0.05
notmissing	2200	99.95	2201	100.00

FL_NUM				
FL_NUM	Frequency	Percent	Cumulative Frequency	Cumulative Percent
missing	3	0.14	3	0.14
notmissing	2198	99.86	2201	100.00

CARRIER				
CARRIER	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NonMissing	2201	100.00	2201	100.00

DEST				
DEST	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	2	0.09	2	0.09
NonMissing	2199	99.91	2201	100.00

Flight_status				
Flight_status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NonMissing	2201	100.00	2201	100.00

ORIGIN				
ORIGIN	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing	1	0.05	1	0.05
NonMissing	2200	99.95	2201	100.00

TAIL_NUM				
----------	--	--	--	--

TAIL_NUM	Frequency	TAIL_NUM Percent	Cumulative Frequency	Cumulative Percent
Missing	2	0.09	2	0.09
NonMissing TAIL_NUM	2199	99.91	Cumulative Frequency	Cumulative Percent

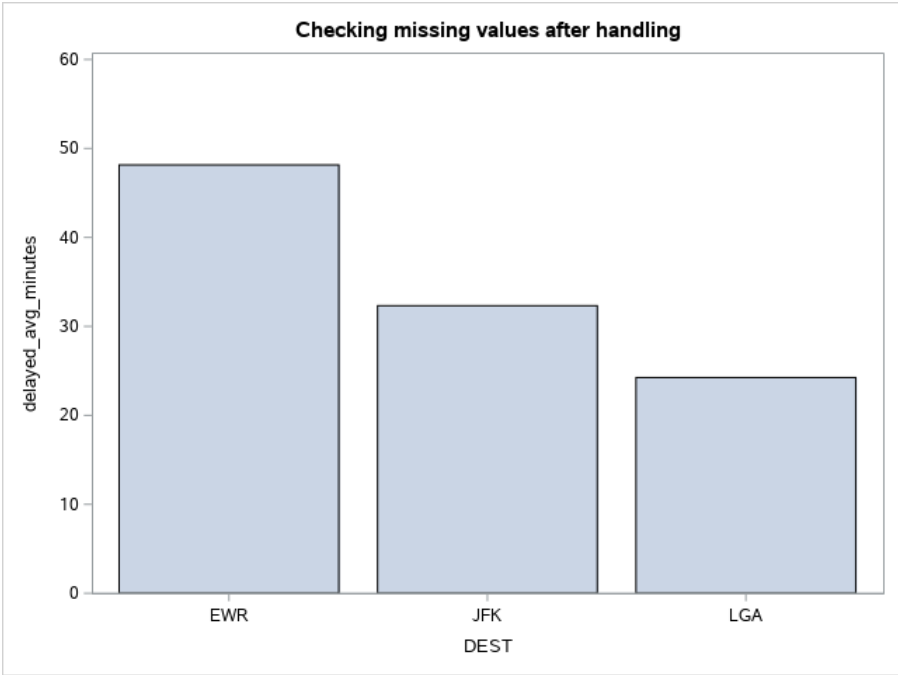
Weather				
Weather	Frequency	Percent	Cumulative Frequency	Cumulative Percent
notmissing	2201	100.00	2201	100.00

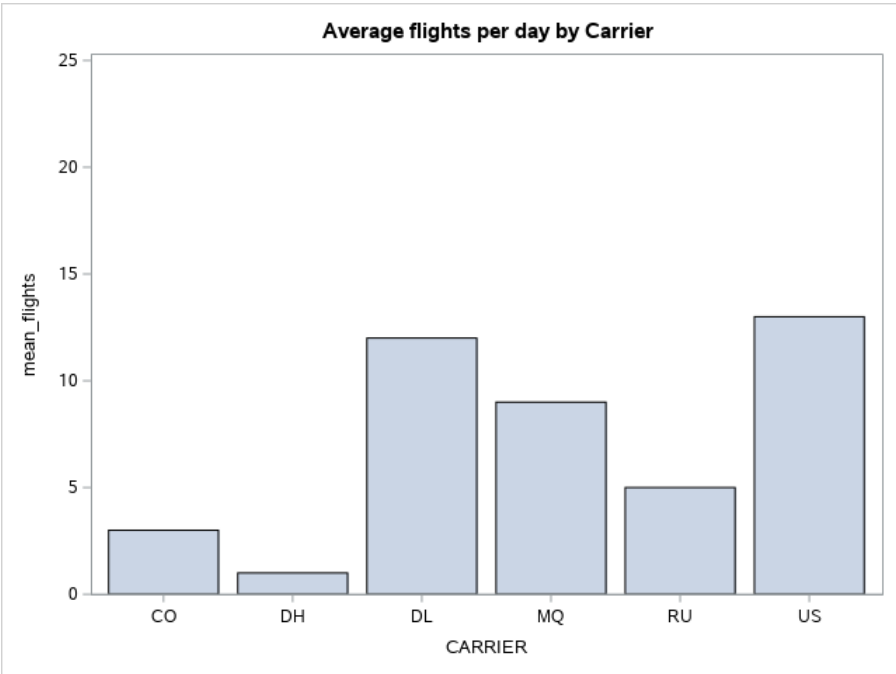
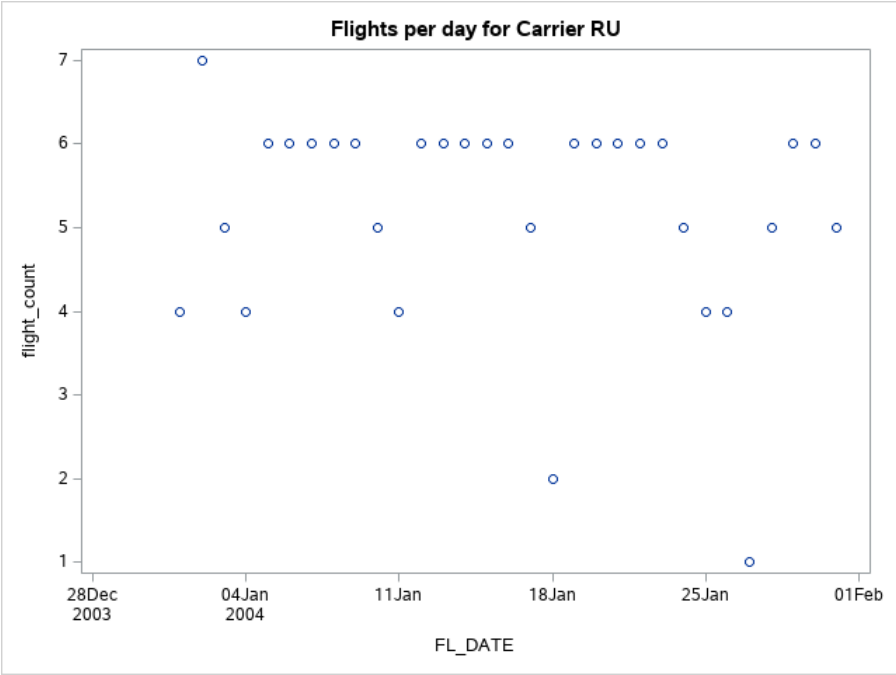
DAY_WEEK				
DAY_WEEK	Frequency	Percent	Cumulative Frequency	Cumulative Percent
notmissing	2201	100.00	2201	100.00

DAY_OF_MONTH				
DAY_OF_MONTH	Frequency	Percent	Cumulative Frequency	Cumulative Percent
notmissing	2201	100.00	2201	100.00

Checking missing values after handling

Variable	Label	N Miss	N
DISTANCE	DISTANCE	0	2190
FL_DATE	FL_DATE	0	2190
FL_NUM	FL_NUM	0	2190
Weather	Weather	0	2190
DAY_WEEK	DAY_WEEK	0	2190
DAY_OF_MONTH	DAY_OF_MONTH	0	2190
dep_time		0	2190
crs_dep_time		0	2190





Average flights per day by Carrier

Variable: DISTANCE (DISTANCE)

Moments			
N	1363	Sum Weights	1363
Mean	211.075569	Sum Observations	287696
Std Deviation	5.81716133	Variance	33.839366
Skewness	-1.5901324	Kurtosis	0.54553411
Uncorrected SS	60771686	Corrected SS	46089.2164
Coeff Variation	2.75596146	Std Error Mean	0.15756624

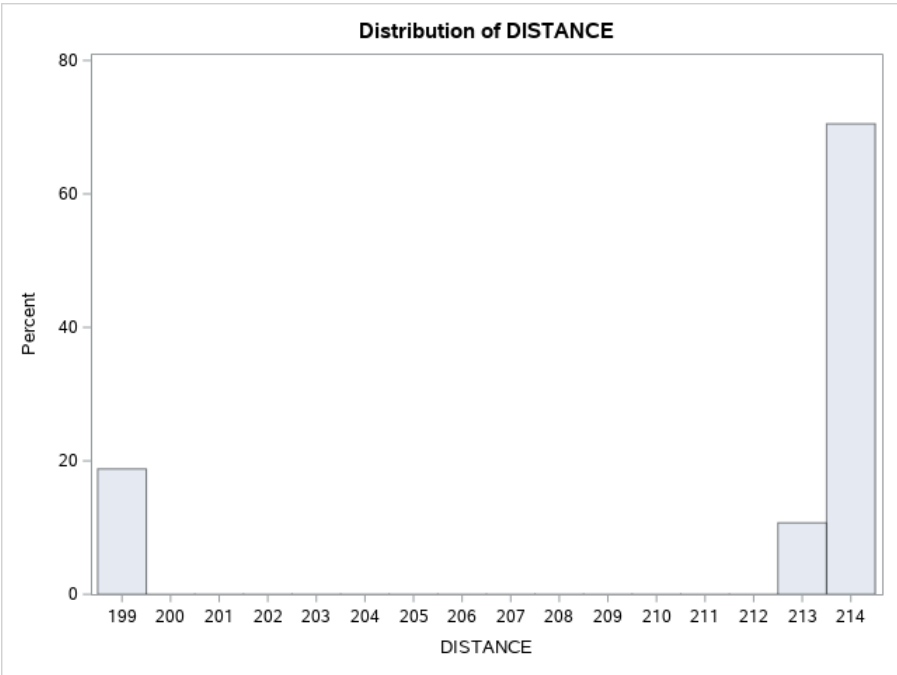
Basic Statistical Measures			
Location		Variability	
Mean	211.0756	Std Deviation	5.81716
Median	214.0000	Variance	33.83937
Mode	214.0000	Range	15.00000
		Interquartile Range	1.00000

Tests for Location: Mu0=0			
Test	Statistic		p Value
Student's t	t	1339.599	Pr > t <.0001
Sign	M	681.5	Pr >= M <.0001
Signed Rank	S	464783	Pr >= S <.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	214
99%	214
95%	214
90%	214
75% Q3	214
50% Median	214
25% Q1	213
10%	199
5%	199
1%	199
0% Min	199

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
199	1363	214	1353
199	1362	214	1354
199	1361	214	1355
199	1360	214	1356
199	1359	214	1357

Average flights per day by Carrier



Average flights per day by Carrier

Variable: delay_time_minutes

Moments			
N	1363	Sum Weights	1363
Mean	5.11592076	Sum Observations	6973
Std Deviation	19.7147366	Variance	388.67084
Skewness	5.04680237	Kurtosis	29.9125401
Uncorrected SS	565043	Corrected SS	529369.685
Coeff Variation	385.360476	Std Error Mean	0.53400221

Basic Statistical Measures			
Location		Variability	
Mean	5.115921	Std Deviation	19.71474
Median	0.000000	Variance	388.67084
Mode	0.000000	Range	200.00000
		Interquartile Range	0

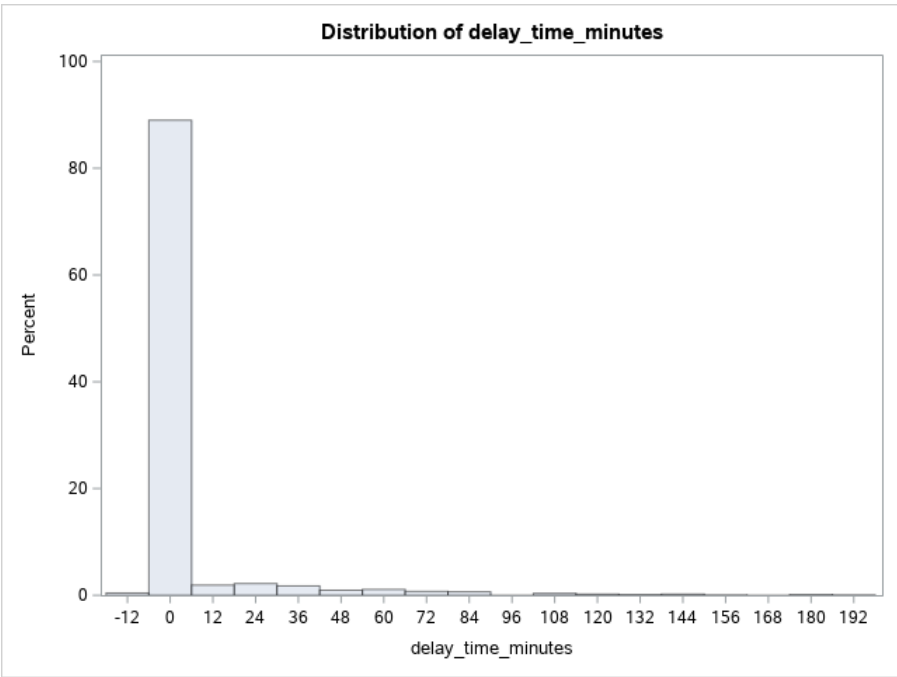
Tests for Location: Mu0=0			
Test	Statistic		p Value
Student's t	t	9.580336	Pr > t <.0001
Sign	M	54.5	Pr >= M <.0001
Signed Rank	S	9132	Pr >= S <.0001

Quantiles (Definition 5)	
Level	Quantile

Quantiles (Definition 5)	
Level	Quantile
100% Max	187
99%	110
95%	40
90%	10
75% Q3	0
50% Median	0
25% Q1	0
10%	0
5%	0
1%	-4
0% Min	-13

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
-13	687	143	778
-11	1042	150	1156
-10	1248	175	1173
-9	990	181	1157
-8	757	187	1135

Average flights per day by Carrier



Flight status by Origin

Analysis Variable : FL_NUM FL_NUM		
ORIGIN	Flight_status	N
BWI	delayed	37
	ontime	105
DCA	delayed	221
	ontime	1142
IAD	delayed	170
	ontime	515

Flight Status for different carrier

Analysis Variable : FL_NUM FL_NUM		
Flight_status	CARRIER	N
delayed	CO	26
	DH	137
	DL	47
	MQ	80
	OH	4
	RU	94
ontime	UA	5
	US	35
	CO	68
	DH	414

Analysis Variable : FL_NUM FL_NUM		
Flight_status	CARRIER	N
	DL	338
	MQ	212
	OH	24
	RU	312
	UA	26
	US	368

Destination and distance affecting flight status

Analysis Variable : FL_NUM FL_NUM			
Flight_status	DEST	DISTANCE	N
delayed	EWR	169	33
		199	54
		213	74
	JFK	184	4
		213	40
		228	40
ontime	LGA	214	127
		229	56
	EWR	169	81
		199	202
		213	219
	JFK	184	24
		213	106
		228	167
	LGA	214	834
		229	129

Weather affecting flight status

Analysis Variable : FL_NUM FL_NUM			
Flight_status	Weather	ORIGIN	N
delayed	0	BWI	36
		DCA	204
		IAD	156
	1	BWI	1
		DCA	17
		IAD	14
ontime	0	BWI	105
		DCA	1142
		IAD	515

Contents on the data reduction data set - FlightDelaysTrainingData

Data Set Name	WORK.FLIGHTDELAYSTRAININGDATA	Observations	1363
Member Type	DATA	Variables	11
Engine	V9	Indexes	0
Created	04/14/2022 12:05:38	Observation Length	72
Last Modified	04/14/2022 12:05:38	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	1816
Obs in First Data Page	1363
Number of Data Set Repairs	0
Filename	/saswork/SAS_workDB4F00002C48_odaws02-usw2.oda.sas.com/SAS_work9BCB00002C48_odaws02-usw2.oda.sas.com/flightdelaystrainingdata.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	1765683
Access Permission	rw-r--r--
Owner Name	u60685886
File Size	256KB
File Size (bytes)	262144

Variables in Creation Order						
#	Variable	Type	Len	Format	Informat	Label
1	CARRIER	Char	2	\$2.	\$2.	CARRIER
2	DEST	Char	3	\$3.	\$3.	DEST

Variables in Creation Order						
#	Variable	Type	Len	Format	Informat	Label
3	DISTANCE	Num	8	BEST.		DISTANCE
4	FL_DATE	Num	8	MMDDYY10.		FL_DATE
5	FL_NUM	Num	8	BEST.		FL_NUM
6	ORIGIN	Char	3	\$3.		ORIGIN
7	Weather	Num	8	BEST.		Weather
8	dep_time	Num	8	TIME5.		
9	crs_dep_time	Num	8	TIME5.		
10	DelayedFlight	Num	8			
11	delay_time_minutes	Num	8			

Data after conversion

Obs	DISTANCE	FL_DATE	FL_NUM	Weather	dep_time	crs_dep_time	DelayedFlight	delay_time_minutes	carrier	origin	dest
1	213	01/01/2004	6155	0	16:40	16:40	0	0	2	2	2
2	214	01/01/2004	1746	0	9:32	9:30	0	0	3	2	3
3	214	01/01/2004	1752	0	12:28	12:30	0	0	3	2	3
4	214	01/01/2004	1756	0	14:29	14:30	0	0	3	2	3
5	214	01/01/2004	1762	0	17:28	17:30	0	0	3	2	3

Data after conversion reference table

origin_no	origin
2	DCA

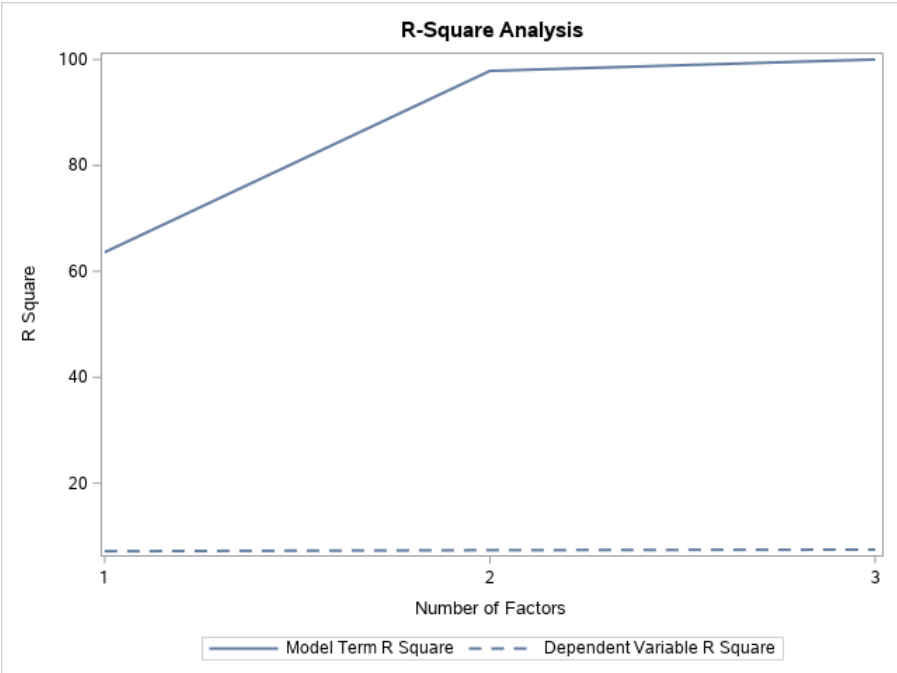
Fitting a Prediction Model for Flight Delay

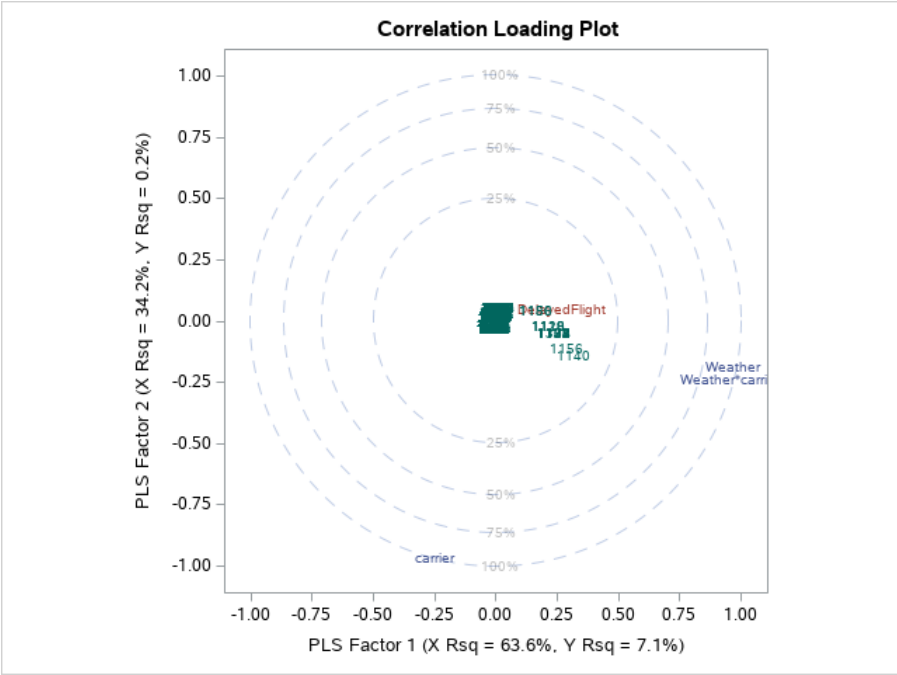
Data Set	WORK.FLIGHTDELAYSTRAININGDATA
Factor Extraction Method	Partial Least Squares
PLS Algorithm	NIPALS
Number of Response Variables	1
Number of Predictor Parameters	3
Missing Value Handling	Exclude
Number of Factors	3

Number of Observations Read	1363
Number of Observations Used	1363

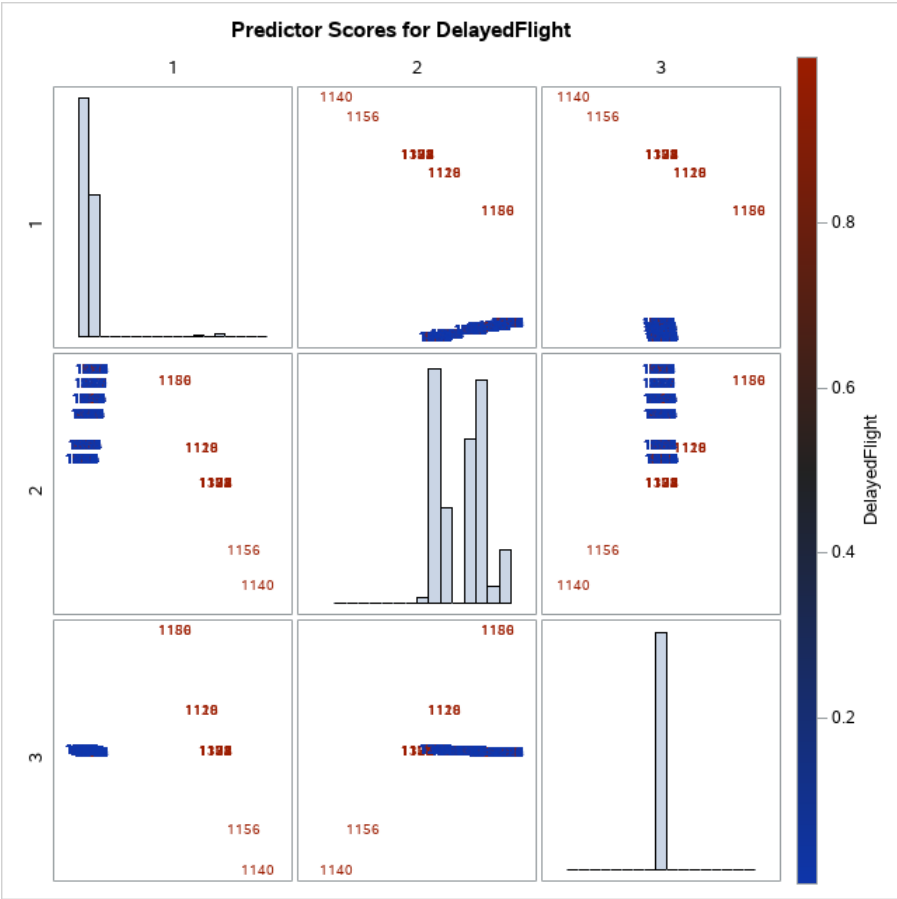
Fitting a Prediction Model for Flight Delay

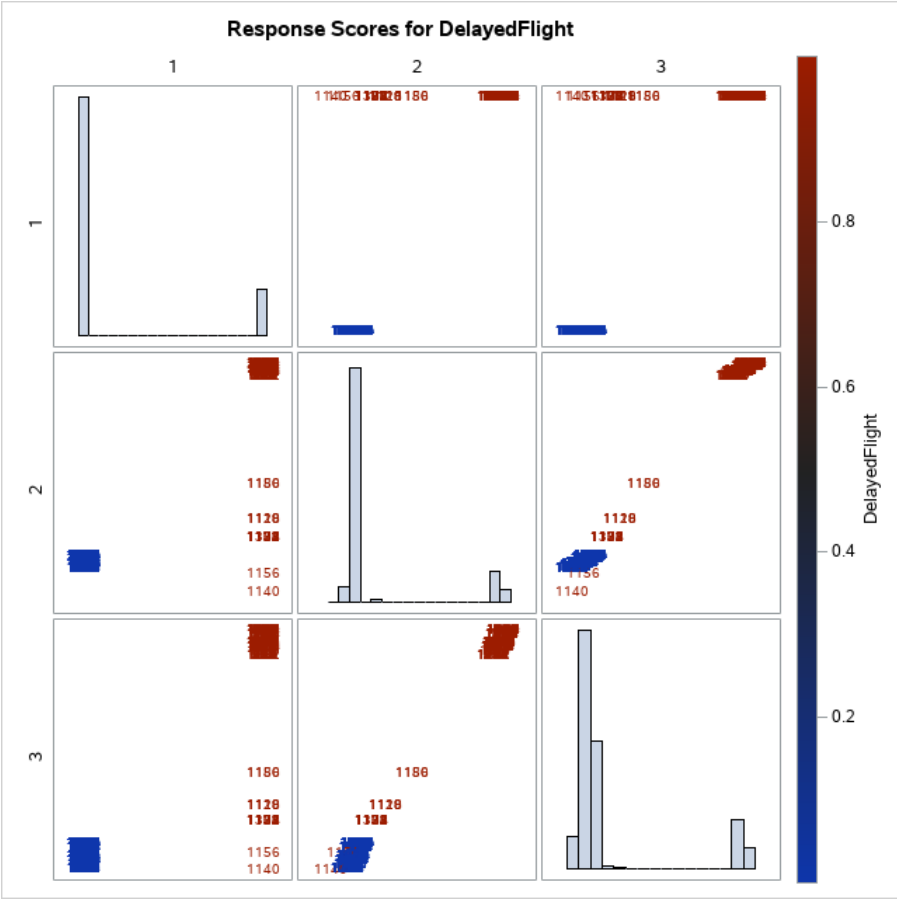
Percent Variation Accounted for by Partial Least Squares Factors				
Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	63.6066	63.6066	7.1459	7.1459
2	34.2343	97.8409	0.2052	7.3511
3	2.1591	100.0000	0.1072	7.4583



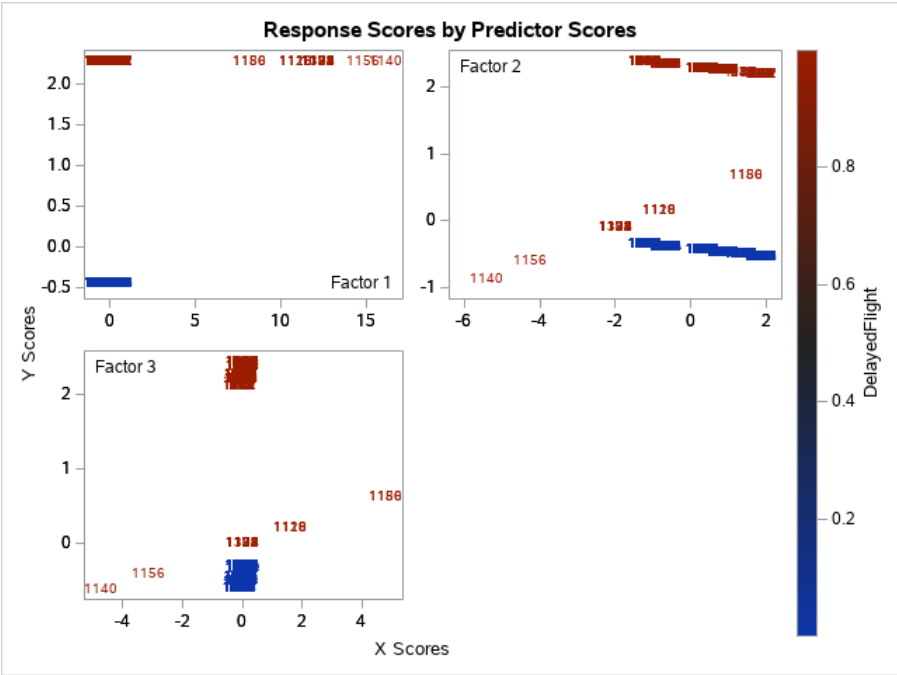


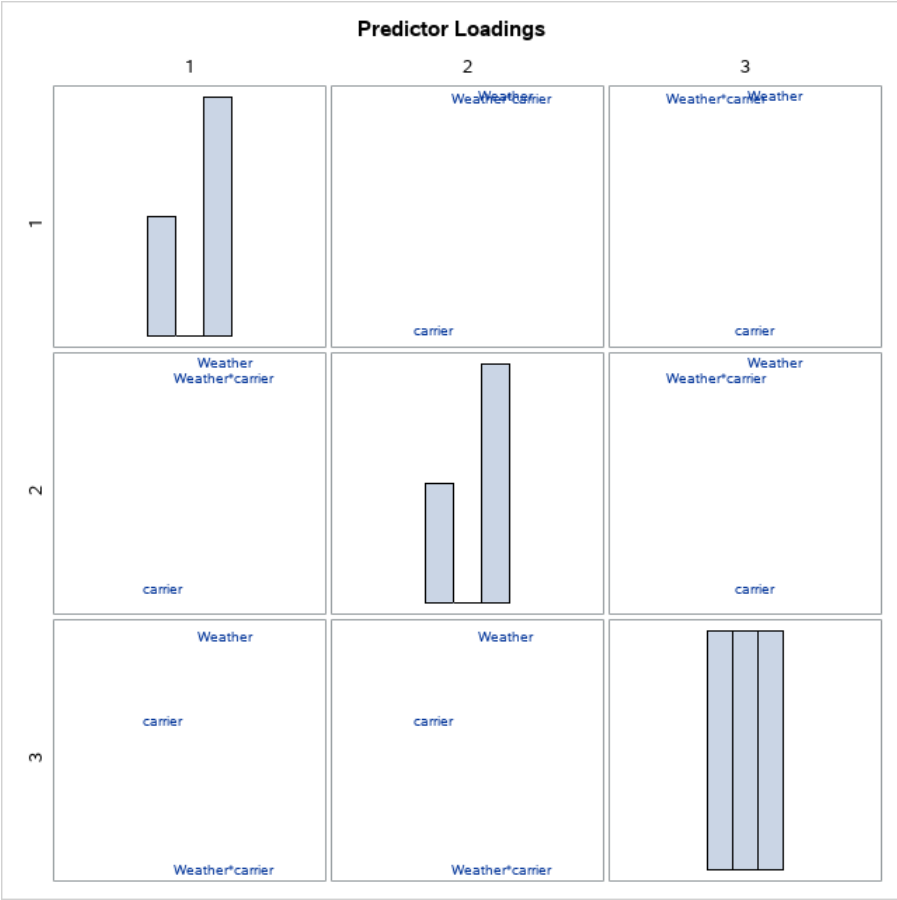
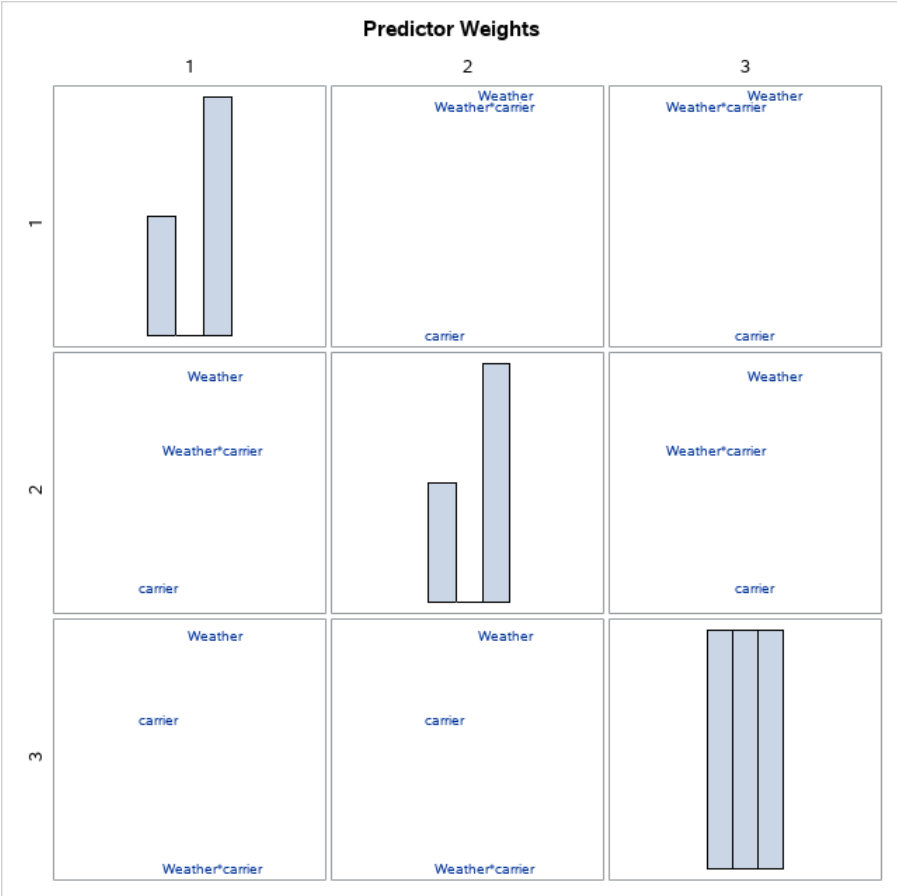
Fitting a Prediction Model for Flight Delay

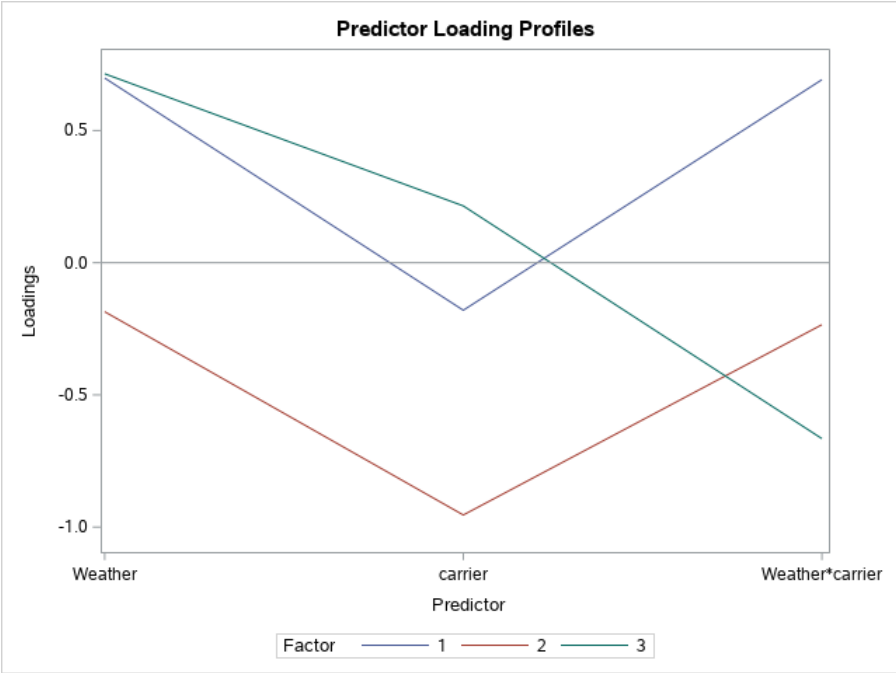
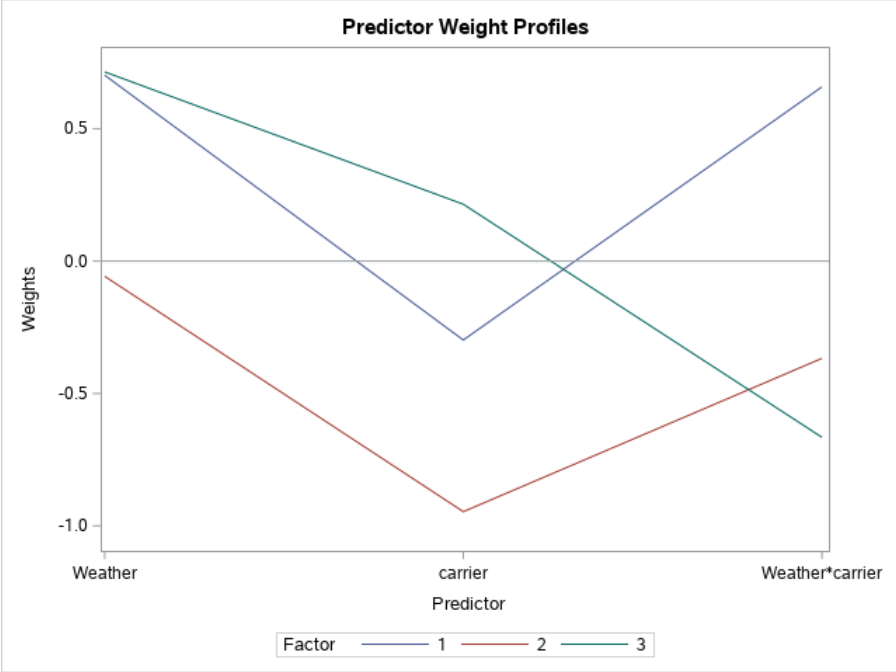
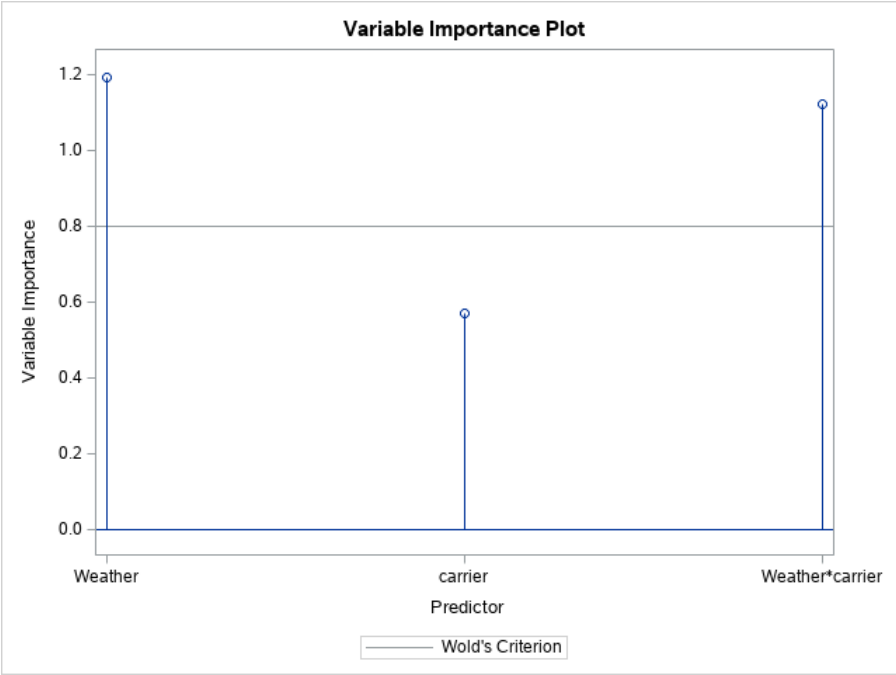


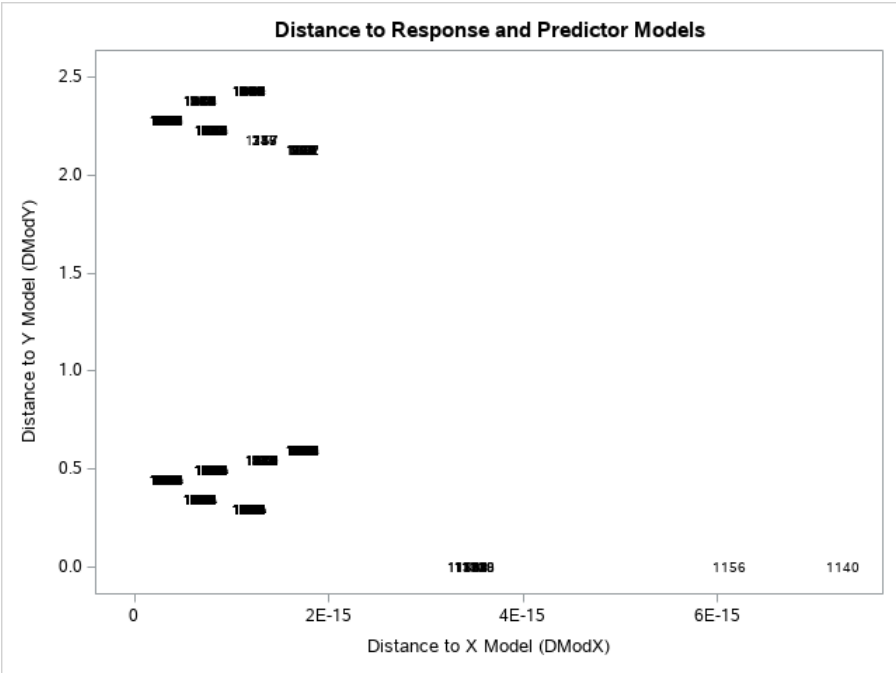
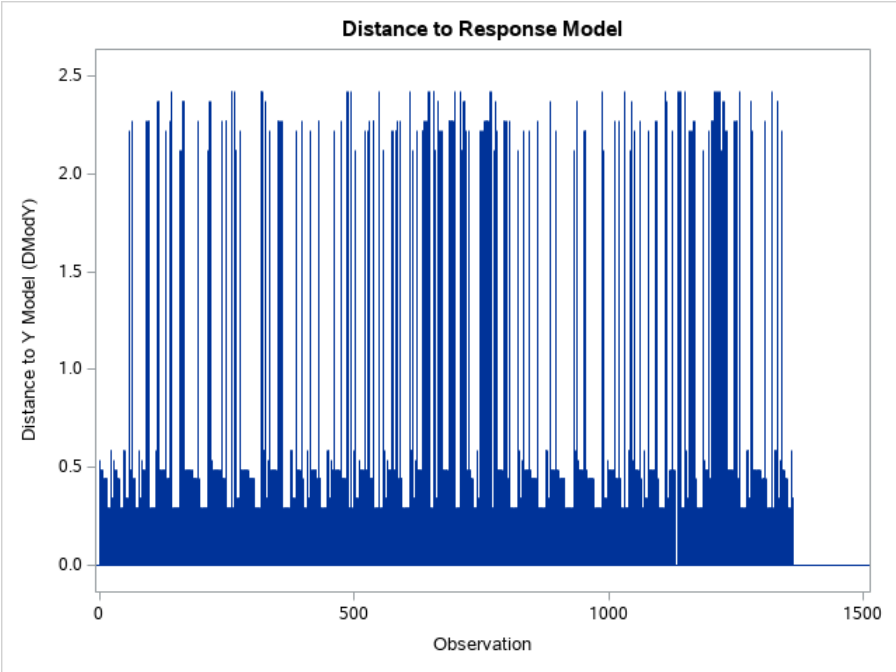
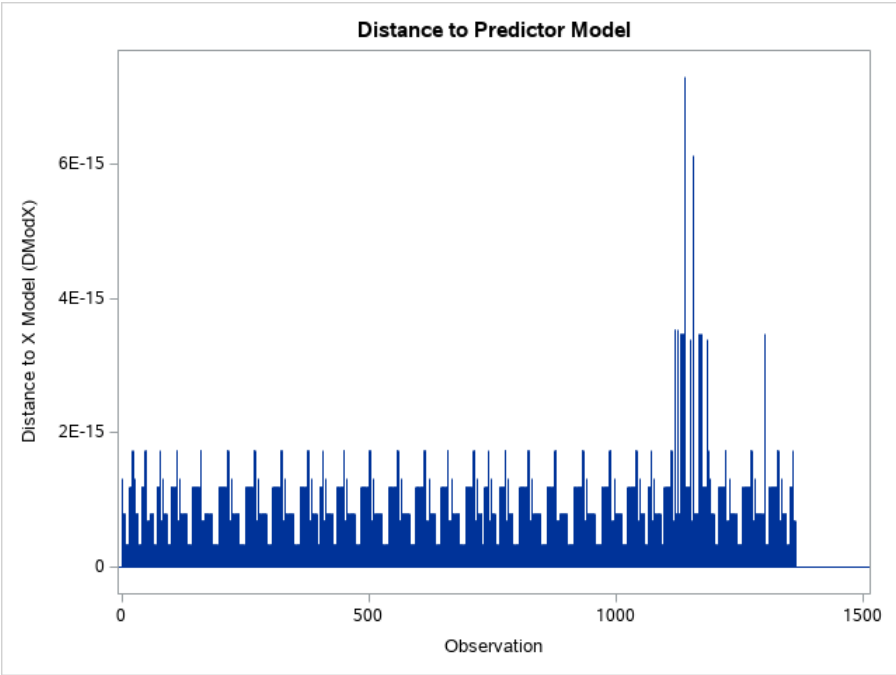


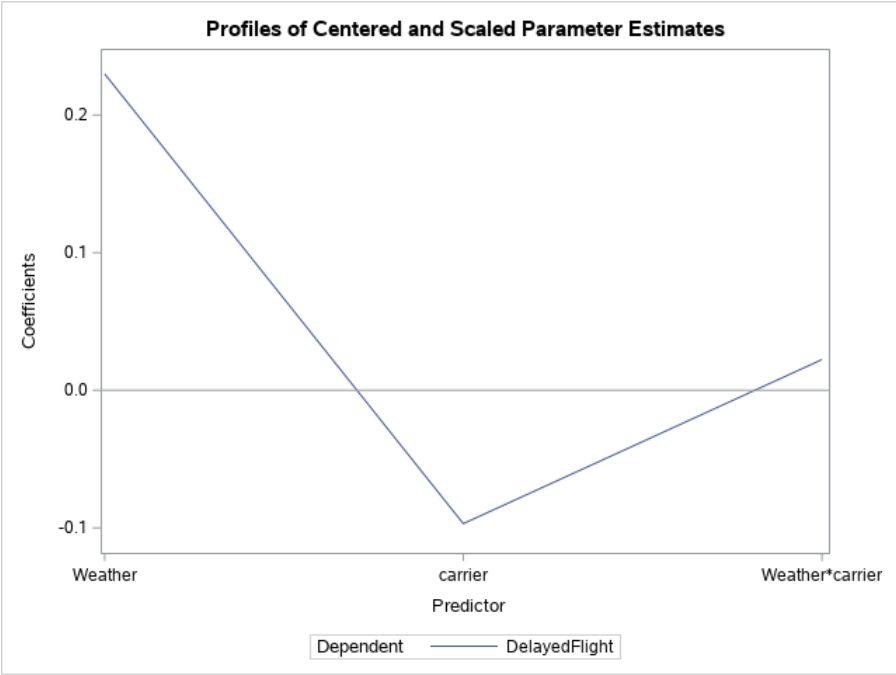
Fitting a Prediction Model for Flight Delay











Fitting a Prediction Model for Flight Delay

