```
1 /* Problem Statement 2 SAS Code*/
3
 4
  /* Pre-Processing */
5 /* import data 'crashes' */
6
7 filename MVCc
8
      '/folders/myfolders/group project/Motor Vehicle Collisions - Crashes.json';
9
  libname MVCc JSON fileref=MVCc;
  libname gp '/folders/myfolders/group project';
10
11
12 data gp.crash; set MVCc.data;
13 run:
14
15
16
17
   /* run every time open the file */
18
19 libname gp '/folders/myfolders/group project';
20 data gp.preprocess_crash; set gp.crash;
21 run;
22
23
24
   data gp.try_crash; set gp.preprocess_crash;
25
  run:
26
27
28
   /* drop the columns*/
29
  data gp.try crash; set gp.try crash;
30
       drop ordinal_root ordinal_data element1 element2 element3 element4 element5 element6 element7 element8
31
       element12 element15 element16 element17 element18;
32
   run:
33
34
   /* change the form of the colomns and rename */
   data gp.try_crash; set gp.try_crash;
35
       format time time.;
36
       time = input(element10, time.);
37
38
       drop element10;
39
40
       format collision ID 8.;
       collision ID = input(element32,8.);
41
       drop element32;
42
43
       format ppl_injured ppl_killed pedes_injured pedes_killed cyc_injured
44
45
       cyc killed motor injured motor killed 4.;
46
       ppl injured = input(element19, 4.);
47
       ppl_killed = input(element20, 4.);
48
       pedes injured = input(element21, 4.);
       pedes_killed= input(element22, 4.);
49
50
       cyc_injured= input(element23, 4.);
51
       cyc killed = input(element24, 4.);
52
       motor_injured= input(element25, 4.);
53
       motor killed= input(element26, 4.);
54
       drop element19 element20 element21 element22 element23 element24 element25 element26 ;
55
56
       rename element9=date element11=BOROUGH element13=LATITUDE element14=LONGITUDE
       element 27 = contributing\_vehicle\_1 \ element 28 = contributing\_vehicle\_2 \ element 29 = contributing\_vehicle\_3
57
58
       element30=contributing_vehicle_4 element31=contributing_vehicle_5
59
       element33=vehicle_type_1 element34=vehicle_type_2 element35=vehicle_type_3
60
       element36=vehicle_type_4 element37=vehicle_type_5;
61
   run:
62
63
   /* form the date and time variables */
   data gp.try_crash; set gp.try_crash;
64
65
       date year = substr(date,1,4);
       date_month = substr(date,6,2);
66
67
       date_day = substr(date,9,2);
68
       drop date;
69
       format date2 MMDDYY10.;
70
       date2 = input(cats(date month, date day, date year), MMDDYY10.);
71
       drop date_month date_day date_year;
72
       rename date2 = date;
73
  run:
74
75
76
   /* count the number of vehicles involved based on the contributing_vehicle */
77
78
79
   data gp.try crash count; set gp.try crash;
       do:
80
```

run;

/* count the number of vehicles involved based on the contributing_vehicle */

data gp.try_crash_count; set gp.try_crash_count;

do;

if vehicle_type_1='' then vehicle_type_1_2=0;
else vehicle_type_1_2 = 1;
end;
drop vehicle_type_1;
rename vehicle_type_1_2 = vehicle_type_1;

140 else vehicle_type_2_2 = 1; 141 end; 142 drop vehicle_type_2; 143 rename vehicle type 2 2 = vehicle type 2; 144 | run ; 145 data gp.try_crash_count; set gp.try_crash_count; 146 do; if vehicle_type_3='' then vehicle_type_3_2=0; 147 148 else vehicle type 3 2 = 1; end; 149 150 drop vehicle_type_3; 151 rename vehicle_type_3_2 = vehicle_type_3; 152 **run** ;

153 data gp.try crash count; set gp.try crash count;

137 data gp.try_crash_count; set gp.try_crash_count;

if vehicle_type_2='' then vehicle_type_2_2=0;

136 **run** ;

do;

138

139

154

155 156

157

158

159

160 run ;

do;
if vehicle_type_4='' then vehicle_type_4_2=0;
else vehicle_type_4_2 = 1;
end;
drop vehicle_type_4;
rename vehicle_type_4;

```
243 data GPVEHI.MULTI_VEHICLE_COLLISION_2019; set GPVEHI.MULTI_VEHICLE_COLLISION_2019;
244
        if VEHICLE OCCUPANTS = '' THEN DELETE;
245 RUN;
246
247 proc export
248
      data=GPVEHI.MULTI_VEHICLE_COLLISION_2019
249
      dbms=csv
      outfile="/folders/myfolders/group project/20200310/multi vehicle collision 2019.csv"
250
251
      replace;
252 run;
253
    /* count the number of crashes in each day in 2019 */
254
255 PROC SOL:
256 CREATE TABLE gpvehi. VehicleTimeLineByDay
257 AS
258 SELECT VEHICLE 2019.crash_date, COUNT(DISTINCT VEHICLE 2019.COLLISION_ID)
259 AS COLLISION_NUM
260 FROM GPVEHI. VEHICLE 2019 VEHICLE 2019
261 GROUP BY VEHICLE 2019.crash_date;
262 QUIT;
263
264
265
266
267 /* all plots used in problem statement 3 in the final report are produced by excel*/
268 /* plot the vehicle in 2019 by day*/
269 ods graphics / reset width=6.4in height=4.8in imagemap;
270
271 proc sort data=GPVEHI.VEHICLETIMELINEBYDAY out= SeriesPlotTaskData;
272
       by CRASH_DATE;
273 run;
274
275 proc sgplot data= SeriesPlotTaskData;
276
        title height=14pt "Vehicles involved in crashes in 2019 by day";
277
        series x=CRASH_DATE y=COLLISION_NUM /;
278
        xaxis grid;
279
        yaxis grid;
280 run:
281
282 ods graphics / reset;
283 title;
284
285 proc datasets library=WORK noprint;
286
        delete SeriesPlotTaskData;
287
        run:
288
289
290
291 /* export it to count the vehecles in each month */
292
293 proc export
      data=gpvehi.VehicleTimeLine
294
295
      dbms=csv
296
      outfile="/folders/myfolders/group project/20200310/multi-timeline-2019.csv"
297
298 run;
299
300 /* Used excel to adjust the date and calculate the number of crashes by month */
301
302
303
304
305 /* import the setup data and plot */
306
307 | FILENAME REFFILE '/folders/myfolders/group project/20200310/plot_vehecle2019.csv';
308
309 PROC IMPORT DATAFILE=REFFILE
310
        DBMS=CSV
        OUT=GPVEHI.PlotVehicle2019;
311
312
        GETNAMES=YES;
313 RUN;
314
315 PROC CONTENTS DATA=GPVEHI.PlotVehicle2019; RUN;
316
317
318 /* plot the vehicle in 2019 by month*/
319 ods graphics / reset width=6.4in height=4.8in imagemap;
320
    proc sort data=GPVEHI.PLOTVEHICLE2019 out=_SeriesPlotTaskData;
321
322
        by MMYY;
    run:
323
```

```
3/15/2020
  324
  325
      proc sgplot data= SeriesPlotTaskData;
          title height=14pt "Vehicles involved in crashes in 2019 by month";
  326
           series x=MMYY y=COLLISION_ID /;
  327
  328
           xaxis grid;
           yaxis grid label="Vehicles_involed";
  329
  330 run;
  331
  332
      ods graphics / reset;
  333 title;
  334
  335 proc datasets library=WORK noprint;
           delete _SeriesPlotTaskData;
  336
  337
  338
  339
  340
  341 /* plot bar chart to show different number of vehicles involved
  342 in one crash in different years*/
  343 data GP.TRY_CRASH_COUNT_plot; set GP.TRY_CRASH_COUNT;
  344
           year = year(date);
  345 run;
  346
  347 ods graphics / reset width=6.4in height=4.8in imagemap;
  348
  349 proc sgplot data=GP.TRY CRASH COUNT plot;
  350
           vbar year / group=vehicle_num groupdisplay=cluster datalabel;
  351
           yaxis grid;
  352 run;
  353
  354 ods graphics / reset;
  355
  356
  357 /* plot pie charts to show collision type in different year */
  358 data GP.TRY_CRASH_COUNT_plot; set GP.TRY_CRASH_COUNT;
  359
           year = year(date);
  360
           format collision type $8.;
           if vehicle_num <3 then collision_type = 'Simple';</pre>
  361
           else collision_type = 'Multi';
  362
  363 run;
  364
  365
           /* by changing the year to get percentage in different year*/
  366 proc template;
  367
           define statgraph SASStudio.Pie;
  368
               begingraph;
  369
               layout region;
  370
               piechart category=collision_type / stat=pct;
  371
               endlayout;
  372
               endgraph;
  373
           end;
  374 run;
  375
  376
       ods graphics / reset width=6.4in height=4.8in imagemap;
  377
  378 proc sgrender template=SASStudio.Pie
  379
               data=GP.TRY_CRASH_COUNT_PLOT (where=(year=2019));
  380 run:
  381
  382 ods graphics / reset;
  383
  384
  385
  386 /* plot grams in 2019 by different variables */
  387 /* plot the line gram to show the change of time */
  388 data GPVEHI.MULTI_VEHICLE_COLLISION_2019_p; set GPVEHI.MULTI_VEHICLE_COLLISION_2019;
  389
          hour = hour(time);
  390 run;
  391
  392 ods graphics / reset width=6.4in height=4.8in imagemap;
  393
  394 proc sgplot data=GPVEHI.MULTI VEHICLE COLLISION 2019 p;
  395
           vline hour /;
  396
           yaxis grid;
  397 run;
  398
  399 ods graphics / reset;
  400
  401
  402 /* plot the pie chart to show the percentage of driver's gender */
  403 proc template;
           define statgraph SASStudio.Pie;
   404
```

452 proc sgrender template=SASStudio.Pie data=GPVEHI.MULTI VEHICLE COLLISION 2019;