

## Question 1

Code:

```

libname pg2 base "E:\SAS\PG2\data";

data np_monuments;
  set pg2.np_final;
  putlog "START DATA STEP ITERATION";
  keep Region ParkName AvgMonthlyVisitors Acres Size;
  where Type="MONUMENT";
  length Size $ 6;
  format AvgMonthlyVisitors Acres comma10.;
  putlog Type=;
  Type=propcase(Type);
  putlog Type=;
  AvgMonthlyVisitors=sum(DayVisits,Campers,OtherLodging)/12;
  if Acres<1000 then Size="Small";
  else if Acres<100000 then Size="Medium";
  else Size="Large";
  putlog _ALL_;
run;

```

Screenshot of the log showing the last four iterations:

```

START DATA STEP ITERATION
Type=MONUMENT
Type=Monument
Region=Southeast Type=Monument ParkName=Fort Pulaski National Monument DayVisits=344,921 Campers=0
OtherLodging=0 Acres=5,623 Size=Medium AvgMonthlyVisitors=28,743 _ERROR_=0 _N_=60
START DATA STEP ITERATION
Type=MONUMENT
Type=Monument
Region=Southeast Type=Monument ParkName=Fort Sumter National Monument DayVisits=888,330 Campers=0
OtherLodging=0 Acres=235 Size=Small AvgMonthlyVisitors=74,028 _ERROR_=0 _N_=61
START DATA STEP ITERATION
Type=MONUMENT
Type=Monument
Region=Southeast Type=Monument ParkName=Ocmulgee National Monument DayVisits=162,260 Campers=0
OtherLodging=0 Acres=704 Size=Small AvgMonthlyVisitors=13,522 _ERROR_=0 _N_=62
START DATA STEP ITERATION
Type=MONUMENT
Type=Monument
Region=Southeast Type=Monument ParkName=Russell Cave National Monument DayVisits=24,649 Campers=0
OtherLodging=0 Acres=310 Size=Small AvgMonthlyVisitors=2,054 _ERROR_=0 _N_=63
NOTE: There were 63 observations read from the data set PG2.NP_FINAL.
      WHERE Type='MONUMENT';
NOTE: The data set WORK.NP_MONUMENTS has 63 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.18 seconds
      cpu time            0.14 seconds

```

## Question 2

Code:

```
❏ data monument(drop = ParkType) park(drop = ParkType) other;  
   set pg2.np_yearlytraffic (keep = ParkName ParkType Location);  
   select;  
   when (ParkType="National Monument") output monument;  
   when (ParkType="National Park") output park;  
   otherwise output other;  
   end;  
   run;
```

Log notes:

```
137 data monument(drop = ParkType) park(drop = ParkType) other;  
138 set pg2.np_yearlytraffic (keep = ParkName ParkType Location);  
139 select;  
140 when (ParkType="National Monument") output monument;  
141 when (ParkType="National Park") output park;  
142 otherwise output other;  
143 end;  
144 run;  
  
NOTE: There were 478 observations read from the data set PG2.NP_YEARLYTRAFFIC.  
NOTE: The data set WORK.MONUMENT has 84 observations and 2 variables.  
NOTE: The data set WORK.PARK has 246 observations and 2 variables.  
NOTE: The data set WORK.OTHER has 148 observations and 3 variables.  
NOTE: DATA statement used (Total process time):  
      real time          0.01 seconds  
      cpu time           0.00 seconds
```

### Question 3

Code:

```
❑ proc sort data = pg2.np_yearlytraffic out = traffic_sort;  
  where ParkType in ("National Preserve" "National River" "National  
  Seashore");  
  by ParkType ParkName;  
  run;  
  
❑ data total_traffic;  
  set traffic_sort;  
  by ParkType ParkName;  
  if first.ParkType then TypeCount = 0;  
  TypeCount+Count;  
  if first.ParkName then NameCount = 0;  
  NameCount+Count;  
  format TypeCount NameCount COMMA.;  
  keep ParkType ParkName TypeCount NameCount;  
  if last.ParkName ;  
  run;  
  
❑ proc print data = total_traffic;  
  run;
```

Output:

#### The SAS System

Obs	ParkName	ParkType	TypeCount	NameCount
1	Big Cypress NPRES	National Preserve	239769	239769
2	Big Thicket NPRES	National Preserve	263913	24,144
3	Little River Canyon NPRES	National Preserve	473044	209131
4	Mojave NPRES	National Preserve	771629	298585
5	Tallgrass Prairie NPRES	National Preserve	773468	1,839
6	Timucuan EHP	National Preserve	1.07E6	293848
7	Big South Fork NRRRA	National River	292943	292943
8	Buffalo NR	National River	774615	481672
9	New River Gorge NR	National River	1.14E6	365073
10	Ozark NSR	National River	1.5E6	359809

## Question 4

Code:

```
libname pg2 base "E:\SAS\PG2\data";

data rainsummary;
  set pg2.np_hourlyrain;
  by Month;
  keep StationName MonthlyRainTotal MonthlyRainTotalMM Date MonthEnd;
  Date = datepart(DateTime);
  if first.Month=1 then MonthlyRainTotal=0;
  MonthlyRainTotal+Rain;
  MonthlyRainTotalMM = round(MonthlyRainTotal*25.4, .5);
  MonthEnd=intnx('month',Date,0,'end');
  format Date MonthEnd date.;
  if last.Month=1;
run;

proc print data=rainsummary;
run;
```

Output:

### The SAS System

Obs	StationName	Date	MonthlyRainTotal	MonthlyRainTotalMM	MonthEnd
1	PANTHER JUNCTION TX	24JAN17	0.3	7.5	31JAN17
2	PANTHER JUNCTION TX	01FEB17	0.0	0.0	28FEB17
3	PANTHER JUNCTION TX	01MAR17	0.0	0.0	31MAR17
4	PANTHER JUNCTION TX	16APR17	0.0	0.0	30APR17
5	PANTHER JUNCTION TX	27MAY17	2.0	51.0	31MAY17
6	PANTHER JUNCTION TX	30JUN17	1.3	33.0	30JUN17
7	PANTHER JUNCTION TX	30JUL17	4.8	122.0	31JUL17
8	PANTHER JUNCTION TX	29AUG17	2.4	61.0	31AUG17
9	PANTHER JUNCTION TX	25SEP17	3.4	86.5	30SEP17
10	PANTHER JUNCTION TX	22OCT17	0.3	7.5	31OCT17
11	PANTHER JUNCTION TX	12NOV17	0.1	2.5	30NOV17
12	PANTHER JUNCTION TX	31DEC17	0.1	2.5	31DEC17

## Question 5

Code:

```
libname pg2 base "E:\SAS\PG2\data";

data clean_traffic;
set pg2.np_monthlytraffic;
length Type $ 5;
Type = scan(ParkName,-1);
Park = substr(ParkName, 1, Find(ParkName, strip(Type))-1);
Region = upcase(compress(Region,' '));
Location = propcase(compbl(Location));
Gate = tranwrd(Location,'Traffic Count At', ' ');
GateCode = cats(ParkCode, ':', Gate);
run;

title "clean_traffic details";

proc print data = clean_traffic (obs = 10);
var Type Park Region Location GateCode Count;
where Month = 3;
run;
```

Output:

clean_traffic details						
Obs	Type	Park	Region	Location	GateCode	Count
3	NP	Acadia	NORTHEAST	Traffic Count At Sand Beach	ACAD:Sand Beach	3,849
15	NP	Acadia	NORTHEAST	Traffic Count At Schoodic	ACAD:Schoodic	2,600
27	NP	Arches	INTERMOUNTAIN	Total Vehicles Entering Park	ARCH:Total Vehicles Entering Park	47,326
39	NS	Assateague Island	NORTHEAST	Traffic Count At Bayberry Drive	ASIS:Bayberry Drive	9,670
51	NS	Assateague Island	NORTHEAST	Traffic Count At Fws Entrance	ASIS:Fws Entrance	14,848
63	NP	Badlands	MIDWEST	Total Traffic Count At Interior Entrance (2602)	BADL:Total Interior Entrance (2602)	4,425
75	NP	Badlands	MIDWEST	Total Traffic Count At Northeast Entrance (2601)	BADL:Total Northeast Entrance (2601)	4,808
87	NP	Badlands	MIDWEST	Total Traffic Count At Pinnacles Entrance (2603)	BADL:Total Pinnacles Entrance (2603)	2,673
99	NM	Bandelier	INTERMOUNTAIN	Traffic Count At Entrance	BAND:Entrance	0
111	NP	Big Bend	INTERMOUNTAIN	Traffic Count At Route 11-Pers.Gap	BIBE:Route 11-Pers.Gap	10,009

## Question 6

Code:

```
data parklookup;  
  set pg2.np_unstructured_codes end=LastRow;  
  length ParkCode $ 4;  
  ParkCode=scan(Column1, 2, ' {}:', "()-");  
  ParkName=scan(Column1, 4, ' {}:', "()");  
  NameLength = length(ParkName);  
  Retain MaxLength 0;  
  MaxLength = max(NameLength, MaxLength);  
  if LastRow = 1 then putlog MaxLength= ;  
run;
```

Log output:

```
80  data parklookup;  
81  set pg2.np_unstructured_codes end=LastRow;  
82  length ParkCode $ 4;  
83  ParkCode=scan(Column1, 2, ' {}:', "()-");  
84  ParkName=scan(Column1, 4, ' {}:', "()");  
85  NameLength = length(ParkName);  
86  Retain MaxLength 0;  
87  MaxLength = max(NameLength, MaxLength);  
88  if LastRow = 1 then putlog MaxLength= ;  
89  run;  
  
MaxLength=83  
NOTE: There were 755 observations read from the data set PG2.NP_UNSTRUCTURED_CODES.  
NOTE: The data set WORK.PARKLOOKUP has 755 observations and 5 variables.  
NOTE: DATA statement used (Total process time):  
      real time           0.01 seconds  
      cpu time            0.01 seconds
```