PROGRAMMING FOR ANALYTICS (BAN 130)

PROJECT REPORT Group 4

Seneca

Academic Year	2020 – 2021		
Semester	☐ Fall	☒ Winter	☐ Summer
Course Code - Name	BAN13NAA.05716.2211		
Instructor	Dr. Adeel Javed		
Assessment	Projects		
Due Date	Friday, April 02, 2021		

Student ID	Student Name	Role
165680190	Ronn Kurien Kollamana	Group Lead
106873201	Laher Gregory Pereira	Member
106837206	Pratik Andrade	Member
103240206	Anushka Niteen Sankpal	Member
126904200	Tanviben Becharbhai Patel	Member
126184209	Yuvraj Singh Chauhan	Member

For this project, we were provided with four csv files which includes the raw data from the Minard's Napolean graphic. We will use these four csv files to produce three datasets, troops1, troops2, and troops3 using SAS code. Each troop file is based on a different division, but they follow the same format.

The common variables used in the datasets are as follows:

- **Longitude, Latitude**: Captures the position of the army

- **Troop_Size**: Size of the army

- **CharDirection**: Shows the direction in which the troop is marching. A is for advance and R for retreat. We convert this to +1 and -1 for convenience and to perform mathematical calculations easily.

- **Division**: Shows the current division of army out of the 3 divisions.

- **Date**: Shows the exact day. Each observation is one day in the russian campaign. This goes from June 1812 to December 1812.

- **Region**: We create an additional variable Region which is the nearest city. The formula for creating the region is Region = Round (Longitude,5) *100 + Round (Latitude,5).

- **Temperature**: Temperature at a particular city on a particular day.

- **Event**: Battles and events taking place on a particular date.

Code snippets and Explanation with output:

1. We start by creating the dataset called cities by importing the cities.csv file. This dataset contains the longitude and latitude and the name of cities. We instruct SAS how to read the city variable by using an informat statement. A new variable called Region is used to locate the nearest city. The longitude and latitudes are rounded off to the nearest whole number since the coordinates listed are at the centre of the city and can't be exactly matched with the troops.csv file which we will be looking at later. After the dataset has been created, we sort it by Region and print the output.

```
data cities;

        infile '/folders/myfolders/BAN_130_Project/cities.csv' delimiter=',';

        informat City $32.;

        input Longitude Latitude City $;

        Region = Round(Longitude,5)*100+Round(Latitude,5);

run;


proc sort data=cities;

by Region;

run;


proc print data=cities;

run;
```

The following screenshot shows the output of the **Cities** dataset. This has been sorted by **Region**.

Cities

Obs	City	Longitude	Latitude	Region
1	Kowno	24.0	55.0	2555
2	Wilna	25.3	54.7	2555
3	Smorgoni	26.4	54.4	2555
4	Molodexno	26.8	54.3	2555
5	Gloubokoe	27.7	55.2	3055
6	Minsk	27.6	53.9	3055
7	Studienska	28.5	54.3	3055
8	Polotzk	28.7	55.5	3055
9	Bobr	29.2	54.4	3055
10	Witebsk	30.2	55.3	3055
11	Orscha	30.4	54.5	3055
12	Mohilow	30.4	53.9	3055
13	Smolensk	32.0	54.8	3055
14	Dorogobouge	33.2	54.9	3555
15	Wixma	34.3	55.2	3555
16	Chjat	34.4	55.5	3555
17	Mojaisk	36.0	55.5	3555
18	Tarantino	36.6	55.3	3555
19	Malo-jarosewli	36.5	55.0	3555
20	Moscou	37.6	55.8	4055
21	Krasnoi	54.6	31.4	5530

2. As a part of the second step, we will work on the battles.csv file. The file contains dates and events of significant events. The battles and other significant events are saved in the events column. The battles are named based on the closest city name. Again, we instruct SAS to read the Event variable by using an informat statement. We then create a new variable called City which will hold the same values as Events in our case since the names of the battles are based on the closest city. We use the anydtdte12. format to account for the data value '18November1812'. This creates our battles dataset.

```
data battles;                              /* City, Date */

   infile '/folders/myfolders/BAN_130_Project/battles.csv' delimiter=',';

   informat Event $32.

              Date anydtdte12.;      /* informat for Date should be date9. */

   input Event $ Date;

     City = Event;                      /* create new variable City with the same values as Event (since
                                            events are named after their city) */

        format Date date9.;
run;


proc print data=battles;
run;
```

The following screenshot shows the output of the Battles dataset.

Battles

Obs	Event	Date	City
1	Kowno	24JUN1812	Kowno
2	Wilna	28JUN1812	Wilna
3	Witebsk	12JUL1812	Witebsk
4	Witebsk	12AUG1812	Witebsk
5	Krasnoi	15AUG1812	Krasnoi
6	Smolensk	16AUG1812	Smolensk
7	Smolensk	18AUG1812	Smolensk
8	Mojaisk	08SEP1812	Mojaisk
9	Moscou	14SEP1812	Moscou
10	Moscou	08OCT1812	Moscou
11	Malo-jarosewli	13OCT1812	Malo-jarosewli
12	Wixma	03NOV1812	Wixma
13	Krasnoi	15NOV1812	Krasnoi
14	Krasnoi	18NOV1812	Krasnoi
15	Studienska	26NOV1812	Studienska
16	Studienska	29NOV1812	Studienska
17	Molodexno	05DEC1812	Molodexno
18	Kowno	14DEC1812	Kowno

3. We first create a dataset called temps by importing the temp.csv file. There are variables like temperature, days and Longitude. As this CSV file is in retreat, a new variable called direction is initialized to -1. Another variable called Date is created to store the new start_date once we take cumulative total of it by adding days and the new start date. Here we have initialized the start date to be 18 OCT 1812. So, the final dataset which is created gives us the temperature, and its corresponding start date.

```
data temps;
        infile '/folders/myfolders/BAN_130_Project/temps.csv' delimiter=',';
        input Longitude
                Temperature
                Days;
    Direction=-1;       /* add new variable Direction, equal to -1, since this file is 100% retreat*/

    start_date='18OCT1812'd;                          /*    first one is 18OCT1812 */
      if _n_=1 then Date = start_date;                /* Create new variable Date: */


        else Date+Days;                 /* then add Days for Date of each further observation */
        drop start_date;
    format Date date9.;
run;
proc print data=temps;
run;
```

The following screenshot shows the output of the Temperatures dataset.

Temperatures

Obs	Longitude	Temperature	Days	Direction	Date
1	37.6	0	6	-1	18OCT1812
2	36.0	0	6	-1	24OCT1812
3	33.2	-9	16	-1	09NOV1812
4	32.0	-21	5	-1	14NOV1812
5	29.2	-11	10	-1	24NOV1812
6	28.5	-20	4	-1	28NOV1812
7	27.2	-24	3	-1	01DEC1812
8	26.7	-30	5	-1	06DEC1812
9	25.3	-26	1	-1	07DEC1812

4. We start by creating the dataset called troops by importing the troops.csv file. This dataset contains the longitude, latitude, troop size, direction, division, and date variables. We instruct SAS to input the longitude, latitude, and other variables by using an input statement. A new variable called Region is used. Region is used to locate the nearest city by using the same formula we used in the dataset called cities. In the dataset, direction is a character variable so first we convert it to a numeric variable. After the dataset has been created, we sort it by Region and print the output.

```
data troops;
    infile '/folders/myfolders/BAN_130_Project/troops.csv' delimiter=',';
    input Longitude
        Latitude
        Troop_Size
        CharDirection $
        Division
        Date date9.;
    Region = Round(Longitude,5)*100+Round(Latitude,5);    /* set Region same as in cities */
    if CharDirection="A" then NumDirection=1;             /* create numeric Direction variable: */
    else if CharDirection="R" then NumDirection=-1;
        format Date date9.;
run;
proc print data=troops;
run;
```

The following screenshot shows the output of the Troops dataset.

Troops

Obs	Longitude	Latitude	Troop_Size	CharDirection	Division	Date	Region	NumDirection
1	24.0	54.9	340000	A	1	24JUN1812	2555	1
2	24.5	55.0	340000	A	1	25JUN1812	2555	1
3	25.5	54.5	340000	A	1	26JUN1812	2555	1
4	26.0	54.7	320000	A	1	28JUN1812	2555	1
5	27.0	54.8	300000	A	1	05JUL1812	2555	1
6	28.0	54.9	280000	A	1	12JUL1812	3055	1
7	28.5	55.0	240000	A	1	12AUG1812	3055	1
8	29.0	55.1	210000	A	1	16AUG1812	3055	1
9	30.0	55.2	180000	A	1	17JUN1812	3055	1
10	30.3	55.3	175000	A	1	18AUG1812	3055	1
11	32.0	54.8	145000	A	1	24AUG1812	3055	1
12	33.2	54.9	140000	A	1	27AUG1812	3555	1
13	34.4	55.5	127100	A	1	01SEP1812	3555	1
14	35.5	55.4	100000	A	1	08SEP1812	3555	1
15	36.0	55.5	100000	A	1	14SEP1812	3555	1
16	37.6	55.8	100000	R	1	18OCT1812	4055	-1
17	37.5	55.7	98000	R	1	19OCT1812	4055	-1
18	37.0	55.0	97000	R	1	20OCT1812	3555	-1
19	36.8	55.0	96000	R	1	22OCT1812	3555	-1
20	35.4	55.3	87000	R	1	24OCT1812	3555	-1
21	34.3	55.2	55000	R	1	05NOV1812	3555	-1
22	33.3	54.8	37000	R	1	09NOV1812	3555	-1
23	32.0	54.6	24000	R	1	14NOV1812	3055	-1
24	30.4	54.4	20000	R	1	19NOV1812	3055	-1
25	29.2	54.4	20000	R	1	24NOV1812	3055	-1
26	28.5	54.3	20000	R	1	28NOV1812	3055	-1
27	28.3	54.4	20000	R	1	30NOV1812	3055	-1
28	24.0	55.1	60000	A	2	24JUN1812	2555	1
29	24.5	55.2	60000	A	2	13AUG1812	2555	1
30	25.5	54.7	60000	A	2	12SEP1812	2555	1
31	26.6	55.7	40000	A	2	18OCT1812	2555	1
32	27.4	55.6	33000	A	2	19OCT1812	2555	1
33	28.7	55.5	30000	R	2	20OCT1812	3055	-1
34	29.2	54.3	30000	R	2	24NOV1812	3055	-1
35	28.5	54.2	30000	R	2	26NOV1812	3055	-1
36	28.3	54.3	28000	R	2	30NOV1812	3055	-1
37	27.5	54.5	20000	R	2	01DEC1812	3055	-1
38	26.8	54.3	12000	R	2	05DEC1812	2555	-1
39	26.4	54.4	14000	R	2	06DEC1812	2555	-1
40	24.6	54.5	8000	R	2	07DEC1812	2555	-1
41	24.4	54.4	4000	R	2	09DEC1812	2555	-1
42	24.2	54.4	4000	R	2	12DEC1812	2555	-1
43	24.1	54.3	4000	R	2	14DEC1812	2555	-1
44	24.0	55.2	22000	A	3	24JUN1812	2555	1
45	24.5	55.3	22000	A	3	13AUG1812	2555	1
46	24.6	55.8	6000	R	3	18OCT1812	2555	-1
47	24.2	54.4	6000	R	3	10NOV1812	2555	-1
48	24.1	54.3	6000	R	3	10DEC1812	2555	-1

5. In this program, we have created a dataset called **troops_citites** which contains the merged data from troops and cities dataset which we have used earlier in the same project. Before merging the dataset we need to sort them by a common variable Region and then merge the two data sets by Region. We perform a left outer join to keep everything in troops. To sort the data of **troops_cities**, we use the sort keyword with division in the descending order of Troop Size.

```
proc sort data=troops;

by Region;

run;

proc sort data=cities;

by Region;

run;


data troops_cities;

  merge troops(in=T) cities(in=C);     /* merge troops & cities into troops_cities */

  by Region;                           /* by region */

  if (T=1);                            /* keeping everything in troops */

run;


proc sort data=troops_cities;          /* sort */

by Division DESCENDING Troop_Size;

/* a convenient sort order is first by Division and then Troop_Size (descending) */

run;


proc print data=troops_cities;

run;
```

The following screenshot shows the output of the troops_cities dataset.

Obs	Longitude	Latitude	Troop_Size	CharDirection	Division	Date	Region	NumDirection	City
1	24.0	55.0	340000	A	1	24JUN1812	2555	1	Kowno
2	25.3	54.7	340000	A	1	25JUN1812	2555	1	Wilna
3	26.4	54.4	340000	A	1	26JUN1812	2555	1	Smorgoni
4	26.8	54.3	320000	A	1	28JUN1812	2555	1	Molodexno
5	27.0	54.8	300000	A	1	05JUL1812	2555	1	Molodexno
6	27.7	55.2	280000	A	1	12JUL1812	3055	1	Gloubokoe
7	27.6	53.9	240000	A	1	12AUG1812	3055	1	Minsk
8	28.5	54.3	210000	A	1	16AUG1812	3055	1	Studienska
9	28.7	55.5	180000	A	1	17JUN1812	3055	1	Polotzk
10	29.2	54.4	175000	A	1	18AUG1812	3055	1	Bobr
11	30.2	55.3	145000	A	1	24AUG1812	3055	1	Witebsk
12	33.2	54.9	140000	A	1	27AUG1812	3555	1	Dorogobouge
13	34.3	55.2	127100	A	1	01SEP1812	3555	1	Wixma
14	34.4	55.5	100000	A	1	08SEP1812	3555	1	Chjat
15	36.0	55.5	100000	A	1	14SEP1812	3555	1	Mojaisk
16	37.6	55.8	100000	R	1	18OCT1812	4055	-1	Moscou
17	37.5	55.7	98000	R	1	19OCT1812	4055	-1	Moscou
18	36.6	55.3	97000	R	1	20OCT1812	3555	-1	Tarantino
19	36.5	55.0	96000	R	1	22OCT1812	3555	-1	Malo-jarosewli
20	35.4	55.3	87000	R	1	24OCT1812	3555	-1	Malo-jarosewli
21	34.3	55.2	55000	R	1	05NOV1812	3555	-1	Malo-jarosewli
22	33.3	54.8	37000	R	1	09NOV1812	3555	-1	Malo-jarosewli
23	30.4	54.5	24000	R	1	14NOV1812	3055	-1	Orscha
24	30.4	53.9	20000	R	1	19NOV1812	3055	-1	Mohilow
25	32.0	54.8	20000	R	1	24NOV1812	3055	-1	Smolensk
26	28.5	54.3	20000	R	1	28NOV1812	3055	-1	Smolensk
27	28.3	54.4	20000	R	1	30NOV1812	3055	-1	Smolensk
28	24.0	55.1	60000	A	2	24JUN1812	2555	1	Molodexno
29	24.5	55.2	60000	A	2	13AUG1812	2555	1	Molodexno
30	25.5	54.7	60000	A	2	12SEP1812	2555	1	Molodexno
31	26.6	55.7	40000	A	2	18OCT1812	2555	1	Molodexno
32	27.4	55.6	33000	A	2	19OCT1812	2555	1	Molodexno
33	28.7	55.5	30000	R	2	20OCT1812	3055	-1	Smolensk
34	29.2	54.3	30000	R	2	24NOV1812	3055	-1	Smolensk
35	28.5	54.2	30000	R	2	26NOV1812	3055	-1	Smolensk
36	28.3	54.3	28000	R	2	30NOV1812	3055	-1	Smolensk
37	27.5	54.5	20000	R	2	01DEC1812	3055	-1	Smolensk
38	26.4	54.4	14000	R	2	06DEC1812	2555	-1	Molodexno
39	26.8	54.3	12000	R	2	05DEC1812	2555	-1	Molodexno
40	24.6	54.5	8000	R	2	07DEC1812	2555	-1	Molodexno
41	24.4	54.4	4000	R	2	09DEC1812	2555	-1	Molodexno
42	24.2	54.4	4000	R	2	12DEC1812	2555	-1	Molodexno
43	24.1	54.3	4000	R	2	14DEC1812	2555	-1	Molodexno
44	24.0	55.2	22000	A	3	24JUN1812	2555	1	Molodexno
45	24.5	55.3	22000	A	3	13AUG1812	2555	1	Molodexno
46	24.6	55.8	6000	R	3	18OCT1812	2555	-1	Molodexno
47	24.2	54.4	6000	R	3	10NOV1812	2555	-1	Molodexno
48	24.1	54.3	6000	R	3	10DEC1812	2555	-1	Molodexno

6. In this step, we create a dataset called **battle_cities** by merging two datasets, battle and cities with inner join pattern. Before merging the dataset we need to sort them by a common variable City and then merge the two data sets by City.

proc sort data=battles;

by City;

run;

proc sort data=cities;

by City;

run;

data battles_cities;

  merge battles(in=B) cities(in=C);   /* merge battles & cities by City, inner join */

  by City;   /* into new dataset battles_cities */

  if B=1 and C=1;

run;

The following screenshot shows the output of the battles_cities dataset.

Battles_Cities

Obs	Event	Date	City	Longitude	Latitude	Region
1	Kowno	24JUN1812	Kowno	24.0	55.0	2555
2	Kowno	14DEC1812	Kowno	24.0	55.0	2555
3	Krasnoi	15AUG1812	Krasnoi	54.6	31.4	5530
4	Krasnoi	15NOV1812	Krasnoi	54.6	31.4	5530
5	Krasnoi	18NOV1812	Krasnoi	54.6	31.4	5530
6	Malo-jarosewli	13OCT1812	Malo-jarosewli	36.5	55.0	3555
7	Mojaisk	08SEP1812	Mojaisk	36.0	55.5	3555
8	Molodexno	05DEC1812	Molodexno	26.8	54.3	2555
9	Moscou	14SEP1812	Moscou	37.6	55.8	4055
10	Moscou	08OCT1812	Moscou	37.6	55.8	4055
11	Smolensk	16AUG1812	Smolensk	32.0	54.8	3055
12	Smolensk	18AUG1812	Smolensk	32.0	54.8	3055
13	Studienska	26NOV1812	Studienska	28.5	54.3	3055
14	Studienska	29NOV1812	Studienska	28.5	54.3	3055
15	Wilna	28JUN1812	Wilna	25.3	54.7	2555
16	Witebsk	12JUL1812	Witebsk	30.2	55.3	3055
17	Witebsk	12AUG1812	Witebsk	30.2	55.3	3055
18	Wixma	03NOV1812	Wixma	34.3	55.2	3555

7. We are creating a derived dataset named as **Troop_battles** from **troop_cities** and **Troop_battles** which helps us to find the cities were troops had battles. We merge two datasets by date to get our new dataset **Troop_battles** by performing a left outer join. Before we perform the merging we sort the troops_cities dataset and the **battle_cities** dataset by date. After that we sort data by **troop_size** and run proc print to see the result.

```
proc sort data=troops_cities;

by Date;

run;

proc sort data=battles_cities;

by Date;

run;


data troops_battles;

  merge troops_cities(in=TC) battles_cities(in=BC);

/* merge troops_cities & battles_cities by Date, outer join, all troops_cities */

  by Date;                         /* into new dataset troops_battles */

  if (TC=1);

run;


proc sort data=troops_battles;    /* keep sorted as before for troops */

by Division DESCENDING Troop_Size;

run;
```

The following screenshot shows the output of the **troops_battles** dataset.

Obs	Longitude	Latitude	Troop_Size	CharDirection	Division	Date	Region	NumDirection	City	Event
1	24.0	55.0	340000	A	1	24JUN1812	2555	1	Kowno	Kowno
2	25.3	54.7	340000	A	1	25JUN1812	2555	1	Wilna	
3	26.4	54.4	340000	A	1	26JUN1812	2555	1	Smorgoni	
4	25.3	54.7	320000	A	1	28JUN1812	2555	1	Wilna	Wilna
5	27.0	54.8	300000	A	1	05JUL1812	2555	1	Molodexno	
6	30.2	55.3	280000	A	1	12JUL1812	3055	1	Witebsk	Witebsk
7	30.2	55.3	240000	A	1	12AUG1812	3055	1	Witebsk	Witebsk
8	32.0	54.8	210000	A	1	16AUG1812	3055	1	Smolensk	Smolensk
9	28.7	55.5	180000	A	1	17JUN1812	3055	1	Polotzk	
10	32.0	54.8	175000	A	1	18AUG1812	3055	1	Smolensk	Smolensk
11	30.2	55.3	145000	A	1	24AUG1812	3055	1	Witebsk	
12	33.2	54.9	140000	A	1	27AUG1812	3555	1	Dorogobouge	
13	34.3	55.2	127100	A	1	01SEP1812	3555	1	Wixma	
14	36.0	55.5	100000	A	1	08SEP1812	3555	1	Mojaisk	Mojaisk
15	37.6	55.8	100000	A	1	14SEP1812	4055	1	Moscou	Moscou
16	37.6	55.8	100000	R	1	18OCT1812	4055	-1	Moscou	
17	37.5	55.7	98000	R	1	19OCT1812	4055	-1	Moscou	
18	36.6	55.3	97000	R	1	20OCT1812	3555	-1	Tarantino	
19	36.5	55.0	96000	R	1	22OCT1812	3555	-1	Malo-jarosewli	
20	35.4	55.3	87000	R	1	24OCT1812	3555	-1	Malo-jarosewli	
21	34.3	55.2	55000	R	1	05NOV1812	3555	-1	Malo-jarosewli	
22	33.3	54.8	37000	R	1	09NOV1812	3555	-1	Malo-jarosewli	
23	30.4	54.5	24000	R	1	14NOV1812	3055	-1	Orscha	
24	30.4	53.9	20000	R	1	19NOV1812	3055	-1	Mohilow	
25	32.0	54.8	20000	R	1	24NOV1812	3055	-1	Smolensk	
26	28.5	54.3	20000	R	1	28NOV1812	3055	-1	Smolensk	
27	28.3	54.4	20000	R	1	30NOV1812	3055	-1	Smolensk	
28	24.0	55.1	60000	A	2	24JUN1812	2555	1	Molodexno	Kowno
29	24.5	55.2	60000	A	2	13AUG1812	2555	1	Molodexno	
30	25.5	54.7	60000	A	2	12SEP1812	2555	1	Molodexno	
31	26.6	55.7	40000	A	2	18OCT1812	2555	1	Molodexno	
32	27.4	55.6	33000	A	2	19OCT1812	2555	1	Molodexno	
33	28.7	55.5	30000	R	2	20OCT1812	3055	-1	Smolensk	
34	29.2	54.3	30000	R	2	24NOV1812	3055	-1	Smolensk	
35	28.5	54.3	30000	R	2	26NOV1812	3055	-1	Studienska	Studienska
36	28.3	54.3	28000	R	2	30NOV1812	3055	-1	Smolensk	
37	27.5	54.5	20000	R	2	01DEC1812	3055	-1	Smolensk	
38	26.4	54.4	14000	R	2	06DEC1812	2555	-1	Molodexno	
39	26.8	54.3	12000	R	2	05DEC1812	2555	-1	Molodexno	Molodexno
40	24.6	54.5	8000	R	2	07DEC1812	2555	-1	Molodexno	
41	24.4	54.4	4000	R	2	09DEC1812	2555	-1	Molodexno	
42	24.2	54.4	4000	R	2	12DEC1812	2555	-1	Molodexno	
43	24.0	55.0	4000	R	2	14DEC1812	2555	-1	Kowno	Kowno
44	24.0	55.2	22000	A	3	24JUN1812	2555	1	Molodexno	Kowno
45	24.5	55.3	22000	A	3	13AUG1812	2555	1	Molodexno	
46	24.6	55.8	6000	R	3	18OCT1812	2555	-1	Molodexno	
47	24.2	54.4	6000	R	3	10NOV1812	2555	-1	Molodexno	
48	24.1	54.3	6000	R	3	10DEC1812	2555	-1	Molodexno	

8. In this step we are creating a new data set named as **troops_temps** which contains both the datasets troops_battles and temps sorted by date. Before we perform the merging we sort both datasets by Date and then sort the final dataset **troops_temps** by descending troop size.

```
proc sort data=troops_battles;

by Date;

run;


proc sort data=temps;

by Date;

run;


data troops_temps;

  merge troops_battles(in=TB) temps(in=TMP);

/* merge troops_battles & temps by Date, outer join, all troops_battles */

  by Date;                        /* into new dataset troops_temps */

  if (TB=1);                      /* keep sorted as before for troops */

run;


proc sort data=troops_temps;         /* keep sorted as before for troops */

by Division DESCENDING Troop_Size;

run;
```

The following screenshot shows the output of the **troops_temps** dataset.

Obs	Longitude	Latitude	Troop_Size	CharDirection	Division	Date	Region	NumDirection	City	Event	Temperature	Days	Direction
1	24.0	55.0	340000	A	1	24JUN1812	2555	1	Kowno	Kowno	.	.	.
2	25.3	54.7	340000	A	1	25JUN1812	2555	1	Wilna		.	.	.
3	26.4	54.4	340000	A	1	26JUN1812	2555	1	Smorgoni		.	.	.
4	25.3	54.7	320000	A	1	28JUN1812	2555	1	Wilna	Wilna	.	.	.
5	27.0	54.8	300000	A	1	05JUL1812	2555	1	Molodexno		.	.	.
6	30.2	55.3	280000	A	1	12JUL1812	3055	1	Witebsk	Witebsk	.	.	.
7	30.2	55.3	240000	A	1	12AUG1812	3055	1	Witebsk	Witebsk	.	.	.
8	32.0	54.8	210000	A	1	16AUG1812	3055	1	Smolensk	Smolensk	.	.	.
9	28.7	55.5	180000	A	1	17JUN1812	3055	1	Polotzk		.	.	.
10	32.0	54.8	175000	A	1	18AUG1812	3055	1	Smolensk	Smolensk	.	.	.
11	30.2	55.3	145000	A	1	24AUG1812	3055	1	Witebsk		.	.	.
12	33.2	54.9	140000	A	1	27AUG1812	3555	1	Dorogobouge		.	.	.
13	34.3	55.2	127100	A	1	01SEP1812	3555	1	Wixma		.	.	.
14	36.0	55.5	100000	A	1	08SEP1812	3555	1	Mojaisk	Mojaisk	.	.	.
15	37.6	55.8	100000	A	1	14SEP1812	4055	1	Moscou	Moscou	.	.	.
16	37.6	55.8	100000	R	1	18OCT1812	4055	-1	Moscou		0	6	-1
17	37.5	55.7	98000	R	1	19OCT1812	4055	-1	Moscou		.	.	.
18	36.6	55.3	97000	R	1	20OCT1812	3555	-1	Tarantino		.	.	.
19	36.5	55.0	96000	R	1	22OCT1812	3555	-1	Malo-jarosewli		.	.	.
20	36.0	55.3	87000	R	1	24OCT1812	3555	-1	Malo-jarosewli		0	6	-1
21	34.3	55.2	55000	R	1	05NOV1812	3555	-1	Malo-jarosewli		.	.	.
22	33.2	54.8	37000	R	1	09NOV1812	3555	-1	Malo-jarosewli		-9	16	-1
23	32.0	54.5	24000	R	1	14NOV1812	3055	-1	Orscha		-21	5	-1
24	30.4	53.9	20000	R	1	19NOV1812	3055	-1	Mohilow		.	.	.
25	29.2	54.8	20000	R	1	24NOV1812	3055	-1	Smolensk		-11	10	-1
26	28.5	54.3	20000	R	1	28NOV1812	3055	-1	Smolensk		-20	4	-1
27	28.3	54.4	20000	R	1	30NOV1812	3055	-1	Smolensk		.	.	.
28	24.0	55.1	60000	A	2	24JUN1812	2555	1	Molodexno	Kowno	.	.	.
29	24.5	55.2	60000	A	2	13AUG1812	2555	1	Molodexno		.	.	.
30	25.5	54.7	60000	A	2	12SEP1812	2555	1	Molodexno		.	.	.
31	26.6	55.7	40000	A	2	18OCT1812	2555	1	Molodexno		0	6	-1
32	27.4	55.6	33000	A	2	19OCT1812	2555	1	Molodexno		.	.	.
33	28.7	55.5	30000	R	2	20OCT1812	3055	-1	Smolensk		.	.	.
34	29.2	54.3	30000	R	2	24NOV1812	3055	-1	Smolensk		-11	10	-1
35	28.5	54.3	30000	R	2	26NOV1812	3055	-1	Studienska	Studienska	.	.	.
36	28.3	54.3	28000	R	2	30NOV1812	3055	-1	Smolensk		.	.	.
37	27.2	54.5	20000	R	2	01DEC1812	3055	-1	Smolensk		-24	3	-1
38	26.7	54.4	14000	R	2	06DEC1812	2555	-1	Molodexno		-30	5	-1
39	26.8	54.3	12000	R	2	05DEC1812	2555	-1	Molodexno	Molodexno	.	.	.
40	25.3	54.5	8000	R	2	07DEC1812	2555	-1	Molodexno		-26	1	-1
41	24.4	54.4	4000	R	2	09DEC1812	2555	-1	Molodexno		.	.	.
42	24.2	54.4	4000	R	2	12DEC1812	2555	-1	Molodexno		.	.	.
43	24.0	55.0	4000	R	2	14DEC1812	2555	-1	Kowno	Kowno	.	.	.
44	24.0	55.2	22000	A	3	24JUN1812	2555	1	Molodexno	Kowno	.	.	.
45	24.5	55.3	22000	A	3	13AUG1812	2555	1	Molodexno		.	.	.
46	24.6	55.8	6000	R	3	18OCT1812	2555	-1	Molodexno		0	6	-1
47	24.2	54.4	6000	R	3	10NOV1812	2555	-1	Molodexno		.	.	.
48	24.1	54.3	6000	R	3	10DEC1812	2555	-1	Molodexno		.	.	.

9. Now that we have a final merged dataset called troops_temps we output the data into 3 separate divisions 1,2 and 3 because they represent an independent trip through the same geographic area.

We use the where clause to extract this information for the 3 divisions and print the results as follows:

data troops1;

set troops_temps;   /* create new dataset troops1 from troops_temps for Division #1 */

WHERE Division=1;   /* use WHERE clause */

drop division region days Direction CharDirection; /* Optional: drop Division and Region variables they are no longer needed */

run;

title "Troops1";

proc print data=troops1 noobs;

run;

The following screenshot shows the output of the **troops1** dataset:

Troops1

Longitude	Latitude	Troop_Size	Date	NumDirection	City	Event	Temperature
24.0	55.0	340000	24JUN1812	1	Kowno	Kowno	.
25.3	54.7	340000	25JUN1812	1	Wilna		.
26.4	54.4	340000	26JUN1812	1	Smorgoni		.
25.3	54.7	320000	28JUN1812	1	Wilna	Wilna	.
27.0	54.8	300000	05JUL1812	1	Molodexno		.
30.2	55.3	280000	12JUL1812	1	Witebsk	Witebsk	.
30.2	55.3	240000	12AUG1812	1	Witebsk	Witebsk	.
32.0	54.8	210000	16AUG1812	1	Smolensk	Smolensk	.
28.7	55.5	180000	17JUN1812	1	Polotzk		.
32.0	54.8	175000	18AUG1812	1	Smolensk	Smolensk	.
30.2	55.3	145000	24AUG1812	1	Witebsk		.
33.2	54.9	140000	27AUG1812	1	Dorogobouge		.
34.3	55.2	127100	01SEP1812	1	Wixma		.
36.0	55.5	100000	08SEP1812	1	Mojaisk	Mojaisk	.
37.6	55.8	100000	14SEP1812	1	Moscou	Moscou	.
37.6	55.8	100000	18OCT1812	-1	Moscou		0
37.5	55.7	98000	19OCT1812	-1	Moscou		.
36.6	55.3	97000	20OCT1812	-1	Tarantino		.
36.5	55.0	96000	22OCT1812	-1	Malo-jarosewli		.
36.0	55.3	87000	24OCT1812	-1	Malo-jarosewli		0
34.3	55.2	55000	05NOV1812	-1	Malo-jarosewli		.
33.2	54.8	37000	09NOV1812	-1	Malo-jarosewli		-9
32.0	54.5	24000	14NOV1812	-1	Orscha		-21
30.4	53.9	20000	19NOV1812	-1	Mohilow		.
29.2	54.8	20000	24NOV1812	-1	Smolensk		-11
28.5	54.3	20000	28NOV1812	-1	Smolensk		-20
28.3	54.4	20000	30NOV1812	-1	Smolensk		.

data troops2;

set troops_temps;

WHERE Division=2;

Drop division region days Direction CharDirection;

run;

title "Troops2";

proc print data=troops2 noobs;

run;   /* create new dataset troops2 from troops_temps for Division #2 */

The following screenshot shows the output of the **troops2** dataset.

Troops2

Longitude	Latitude	Troop_Size	Date	NumDirection	City	Event	Temperature
24.0	55.1	60000	24JUN1812	1	Molodexno	Kowno	.
24.5	55.2	60000	13AUG1812	1	Molodexno		.
25.5	54.7	60000	12SEP1812	1	Molodexno		.
26.6	55.7	40000	18OCT1812	1	Molodexno		0
27.4	55.6	33000	19OCT1812	1	Molodexno		.
28.7	55.5	30000	20OCT1812	-1	Smolensk		.
29.2	54.3	30000	24NOV1812	-1	Smolensk		-11
28.5	54.3	30000	26NOV1812	-1	Studienska	Studienska	.
28.3	54.3	28000	30NOV1812	-1	Smolensk		.
27.2	54.5	20000	01DEC1812	-1	Smolensk		-24
26.7	54.4	14000	06DEC1812	-1	Molodexno		-30
26.8	54.3	12000	05DEC1812	-1	Molodexno	Molodexno	.
25.3	54.5	8000	07DEC1812	-1	Molodexno		-26
24.4	54.4	4000	09DEC1812	-1	Molodexno		.
24.2	54.4	4000	12DEC1812	-1	Molodexno		.
24.0	55.0	4000	14DEC1812	-1	Kowno	Kowno	.

```
data troops3;

set troops_temps;

WHERE Division=3;

Drop division region days Direction CharDirection;

run;


title "Troops3";

proc print data=troops3 noobs;

run;                       /* create new dataset troops3 from troops_temps for Division #3 */
```

The following screenshot shows the output of the **troops3** dataset.

Troops3

Longitude	Latitude	Troop_Size	Date	NumDirection	City	Event	Temperature
24.0	55.2	22000	24JUN1812	1	Molodexno	Kowno	.
24.5	55.3	22000	13AUG1812	1	Molodexno		.
24.6	55.8	6000	18OCT1812	-1	Molodexno		0
24.2	54.4	6000	10NOV1812	-1	Molodexno		.
24.1	54.3	6000	10DEC1812	-1	Molodexno		.