```
1   /*************************************************************************
2    Program Name: jblubau1_hw15_script
3    Date Created: 11/29/2016
4    Author: Joseph Blubaugh
5    Purpose: Homework Assignment 15
6   ************************************************************************* */
7
8   libname datadb 'C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Materials'
8  ! access=readonly;
NOTE: Libref DATADB was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Materials
9   libname output 'C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Data';
NOTE: Libref OUTPUT was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Data
10
11   filename outpdf
11 ! 'C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Homework\jblubau1_hw15_output.pdf';
12
13   * 2) Import and parse data;
14   data peg;
15      infile 'C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Materials\pegasus.dat';
16      input @1 all $115. @;
17         if substr(all,1,8) = '(Level1)' then
18      input @7 Level $1.
19            @10 Details $96.
20            @106 Salary dollar10.;
21         if substr(all,1,8) = '(Level2)' then
22      input @16 Level $1.
23            @19 Details $87.
24            @106 Salary dollar10.;
25         if substr(all,1,8) = '(Level3)' then
26      input @25 Level $1.
27            @28 Details $74.
28            @106 Salary dollar10.;
29         if substr(all,1,8) = '(Level4)' then
30      input @34 Level $1.
31            @37 Details $69.
32            @106 Salary dollar10.;
33         if substr(all,1,8) = '(Level5)' then
34      input @43 Level $1.
35            @46 Details $60.
36            @106 Salary dollar10.;
37         if substr(all,1,8) = '(Level6)' then
38      input @52 Level $1.
39            @55 Details $51.
40            @106 Salary dollar10.;
41      Job_Title = substr(Details, 1, find(Details, "(")-1);
42      Employee_Name = compress(substr(Details, find(Details, "(")+1, find(Details, " )")),")");
43      drop all details;
44   run;

NOTE: The infile 'C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Materials\pegasus.dat' is:
      Filename=C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Materials\pegasus.dat,
      RECFM=V,LRECL=32767,File Size (bytes)=49608,
      Last Modified=29Nov2016:15:12:35,
```

```
      Create Time=29Nov2016:15:12:35

NOTE: 424 records were read from the infile
      'C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Materials\pegasus.dat'.
      The minimum record length was 115.
      The maximum record length was 115.
NOTE: The data set WORK.PEG has 424 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds


45
46    ods pdf file=outpdf;
NOTE: Writing ODS PDF output to DISK destination "OUTPDF", printer "PDF".
47
48    * 3) Use FREQ to identify inconsistent job titles with one employee;
49    proc freq data=peg;
50        tables Job_Title;
51        title Analysis of Pegasus Employee Data for Clean Up;
52        title3 Frequency Report of Job Title;
53    run;

NOTE: Writing HTML Body file: sashtml.htm
NOTE: There were 424 observations read from the data set WORK.PEG.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.29 seconds
      cpu time            0.22 seconds


54
55    * 4) Use UNIVARIATE to validate salaries;
56    proc univariate data=peg;
57        var salary;
58        title2 Analysis of Salary Values;
59        title3;
60    run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.06 seconds
      cpu time            0.03 seconds


61
62    * 5) Salaries that are suspicious;
63    proc print data=peg noobs;
64        where salary > 500000 or salary < 1000;
65        title2 Salary Values to be Investigated;
66    run;

NOTE: There were 6 observations read from the data set WORK.PEG.
      WHERE  not (salary>=1000 and salary<=500000);
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.04 seconds
      cpu time            0.00 seconds
```

```
67
68   * 6) Clean up Job Titles;
69   data output.pegasus;
70       set peg;
71       if job_title = 'Accountant i' then job_title = 'Accountant I';
72       else if job_title = 'Accountant ii' then job_title = 'Accountant II';
73       else if job_title = 'Accountant iii' then job_title = 'Accountant III';
74       else if job_title = 'Warehouse Assistant i' then job_title = 'Warehouse Assistant I';
75       else if job_title = 'Warehouse Assistant ii' then job_title = 'Warehouse Assistant II';
76       else job_title = job_title;
77   run;

NOTE: There were 424 observations read from the data set WORK.PEG.
NOTE: The data set OUTPUT.PEGASUS has 424 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


78
79   * 7) Verify Job Count;
80   proc freq data=output.pegasus nlevels;
81       tables Job_Title / noprint;
82       title Number of Different Jobs in Cleaned Data;
83       title3;
84       title3;
85   run;

NOTE: There were 424 observations read from the data set OUTPUT.PEGASUS.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.06 seconds
      cpu time            0.01 seconds


86
87   * 8) Print listing of employees with Chief, Director, or Temp;
88   proc sort data=output.pegasus;
89       by Level Job_Title Employee_Name;
90   run;

NOTE: There were 424 observations read from the data set OUTPUT.PEGASUS.
NOTE: The data set OUTPUT.PEGASUS has 424 observations and 4 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


91
92   proc print data=output.pegasus;
93       var Job_Title Employee_Name;
94       by Level;
95       id Level;
96       where Job_Title like '%Chief%' or
97             Job_Title like '%President%' or
98             Job_Title like '%Director%' or
99             Job_Title like '%Temp%';
100      title List of Pegasus Employees to be Reviewed for Orion Positions;
```

```
101  run;

NOTE: There were 34 observations read from the data set OUTPUT.PEGASUS.
      WHERE Job_Title like '%Chief%' or Job_Title like '%President%' or Job_Title like
      '%Director%' or Job_Title like '%Temp%';
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.06 seconds
      cpu time            0.01 seconds


102
103  title;
104
105  ods pdf close;
NOTE: ODS PDF printed 11 pages to
      C:\Users\Joseph\Projects\learning\Statistics\STAT_604\Homework\jblubau1_hw15_output.pdf.
```