```sas
1  /*  STAT 382 HOMEWORK 5 */
2
3
4
5  /* Task 1: Data Frame Creation */
6  TITLE1 "Task 1";
7  TITLE2 "Q1 Part a) Import Datasets";
8
9  %web_drop_table(WORK.IMPORT);
10
11
12 FILENAME REFFILE '/home/u61397358/sasuser.v94/stocks_nasdaq2.csv';
13
14 PROC IMPORT DATAFILE=REFFILE
15     DBMS=CSV
16     OUT=NASDAQ2;
17     GETNAMES=YES;
18 RUN;
19
20 PROC CONTENTS DATA=NASDAQ2; RUN;
21
22 %web_drop_table(WORK.IMPORT);
23
24
25 FILENAME REFFILE '/home/u61397358/sasuser.v94/stocks_nyse2.csv';
26
27 PROC IMPORT DATAFILE=REFFILE
28     DBMS=CSV
29     OUT=NYSE2;
30     GETNAMES=YES;
31 RUN;
32
33 PROC CONTENTS DATA=NYSE2; RUN;
34
35
36 TITLE2 "Q1 Part b) Modify column formats";
37
38 DATA NYSE2;
39 LENGTH Exchange $30 SYMBOL $30;
40 FORMAT Exchange $30. SYMBOL $30.;
41 SET NYSE2;
42 RUN;
43 DATA NASDAQ2;
44 LENGTH Exchange $30 Symbol $30 Name $35;
45 FORMAT Exchange $30. Symbol $30. Name $35.;
46 SET NASDAQ2;
47 RUN;
48
49
50 TITLE2 "Q1 Part c) Combine NASDAQ and NYSE Datasets";
51
52 DATA stocks;
53 MERGE NYSE2 NASDAQ2;
54
55 PROC PRINT DATA = stocks;
56 RUN;
57
58
59 TITLE2 "Q1 Part d) Sort stocks by ascending Sector";
60
61 PROC SORT DATA = stocks out=stocks;
62 by sector;
63 RUN;
64
65
66 TITLE2 "Q1 Part e) Create the dataset SectorVolatility";
67
68 DATA SectorVolatility;
69 INFILE datalines dlm=',' dsd;
70 LENGTH Sector $22.;
71 LENGTH Sector_Volatility $30.;
72 INPUT Sector $ Sector_Volatility $ @@;
73 datalines;
74 Materials, Low
75 Information Technology, High
76 Consumer Discretionary, High
```

```sas
77  Health Care, Moderate
78  Consumer Staples, Moderate
79  Telecomm Services, High
80  Financials, High
81  Industrials, Low
82  Energy, High
83  Utilities, Moderate
84  Real Estate, Low
85  ;
86  RUN;
87
88  PROC PRINT DATA = SectorVolatility;
89  RUN;
90
91
92  TITLE2 "Q1 Part f) Sort SectorVolatility by ascending Sector";
93
94  PROC SORT DATA = SectorVolatility out=SectorVolatility;
95  by sector;
96  RUN;
97
98
99  TITLE2 "Q1 Part g) Merge the stocks dataset and the SectorVolatility dataset to create stocks2";
100
101 DATA stocks2;
102 MERGE stocks SectorVolatility;
103
104 PROC PRINT DATA = stocks2;
105 RUN;
106
107
108 /* Task 2: Cleaning and Formatting */
109 TITLE1 "Task 2";
110 TITLE2 "Q2) Replace NA values for DividendYield with 0s in new dataset stocks_divfix";
111
112 DATA stocks_divfix;
113 SET stocks2;
114 IF DividendYield = . THEN DividendYield = 0;
115 RUN;
116
117
118 TITLE2 "Q3) Remove rows for PricetoEarnings and PricetoBook that are NA. Do this in a new dataset called stocks_clean";
119
120 DATA stocks_clean;
121 SET STOCKS_DIVFIX;
122 IF PricetoEarnings = . THEN DELETE;
123 IF PricetoBook = . THEN DELETE;
124 RUN;
125
126 /* Task 3: Variable Creation and Subsetting */
127 TITLE1 "Task 3";
128 TITLE2 "Q4) Create a new variable called spread = FiftytwoWeekHigh - FiftytwoWeekLow";
129
130 DATA stocks_clean;
131 SET STOCKS_DIVFIX;
132 IF PricetoEarnings = . THEN DELETE;
133 IF PricetoBook = . THEN DELETE;
134 Spread=FiftytwoWeekHigh-FiftytwoWeekLow;
135 RUN;
136
137
138 TITLE2 "Q5) Create a new variable called Earnings_category";
139
140 DATA stocks_clean;
141 SET STOCKS_DIVFIX;
142 IF PricetoEarnings = . THEN DELETE;
143 IF PricetoBook = . THEN DELETE;
144 Spread=FiftytwoWeekHigh-FiftytwoWeekLow;
145 IF EarningsperShare<0 THEN EarningsCategory = "Loss";
146 IF EarningsperShare>=0 AND EarningsperShare<3 THEN EarningsCategory = "Small";
147 IF EarningsperShare>=3 AND EarningsperShare<10 THEN EarningsCategory = "Good";
148 IF EarningsperShare>=10 THEN EarningsCategory = "Strong";
149 RUN;
150
151
152 TITLE2 "Q6) Create dividends dataset";
153
```

```
154 DATA dividends;
155 SET stocks_clean;
156 dividends_mean = mean(DividendYield);
157 RUN;
```