

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
70
71      /* STAT 382 HOMEWORK 5 */
72
73
74
75      /* Task 1: Data Frame Creation */
76      TITLE1 "Task 1";
77      TITLE2 "Q1 Part a) Import Datasets";
78
79      %web_drop_table(WORK.IMPORT);
80
81
82      FILENAME REFFILE '/home/u61397358/sasuser.v94/stocks_nasdaq2.csv';
83
84      PROC IMPORT DATAFILE=REFFILE
85      DBMS=CSV
86      OUT=NASDAQ2;
87      GETNAMES=YES;
88      RUN;

```

NOTE: Import cancelled. Output dataset WORK.NASDAQ2 already exists. Specify REPLACE option to overwrite it.

NOTE: The SAS System stopped processing this step because of errors.

NOTE: PROCEDURE IMPORT used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	194.21k
OS Memory	32168.00k
Timestamp	04/28/2022 02:58:50 AM
Step Count	72 Switch Count 0
Page Faults	0
Page Reclaims	58
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	0

89

```

90      PROC CONTENTS DATA=NASDAQ2; RUN;

```

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.06 seconds
user cpu time	0.07 seconds
system cpu time	0.01 seconds
memory	3032.71k
OS Memory	33712.00k
Timestamp	04/28/2022 02:58:50 AM
Step Count	73 Switch Count 0
Page Faults	0
Page Reclaims	291
Page Swaps	0
Voluntary Context Switches	2
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	24

```

91
92      %web_drop_table(WORK.IMPORT);
93
94
95      FILENAME REFFILE '/home/u61397358/sasuser.v94/stocks_nyse2.csv';
96
97      PROC IMPORT DATAFILE=REFFILE
98      DBMS=CSV
99      OUT=NYSE2;
100     GETNAMES=YES;
101     RUN;

```

NOTE: Import cancelled. Output dataset WORK.NYSE2 already exists. Specify REPLACE option to overwrite it.

NOTE: The SAS System stopped processing this step because of errors.

NOTE: PROCEDURE IMPORT used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	195.15k
OS Memory	33192.00k

```

Timestamp          04/28/2022 02:58:50 AM
Step Count          74  Switch Count  0
Page Faults         0
Page Reclaims       15
Page Swaps          0
Voluntary Context Switches  1
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  0

```

102

```
103      PROC CONTENTS DATA=NYSE2; RUN;
```

```
NOTE: PROCEDURE CONTENTS used (Total process time):
```

```

real time          0.05 seconds
user cpu time       0.05 seconds
system cpu time     0.00 seconds
memory             1034.25k
OS Memory          33968.00k
Timestamp          04/28/2022 02:58:50 AM
Step Count          75  Switch Count  0
Page Faults         0
Page Reclaims       115
Page Swaps          0
Voluntary Context Switches  0
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  32

```

104

105

```
106      TITLE2 "Q1 Part b) Modify column formats";
```

107

```
108      DATA NYSE2;
```

```
109      LENGTH Exchange $30 SYMBOL $30;
```

```
110      FORMAT Exchange $30. SYMBOL $30.;
```

```
111      SET NYSE2;
```

```
112      RUN;
```

```
NOTE: There were 359 observations read from the data set WORK.NYSE2.
```

```
NOTE: The data set WORK.NYSE2 has 359 observations and 14 variables.
```

```
NOTE: DATA statement used (Total process time):
```

```

real time          0.00 seconds
user cpu time       0.01 seconds
system cpu time     0.00 seconds
memory             972.96k
OS Memory          33968.00k
Timestamp          04/28/2022 02:58:50 AM
Step Count          76  Switch Count  2
Page Faults         0
Page Reclaims       110
Page Swaps          0
Voluntary Context Switches  13
Involuntary Context Switches 0
Block Input Operations  0
Block Output Operations  264

```

```
113      DATA NASDAQ2;
```

```
114      LENGTH Exchange $30 Symbol $30 Name $35;
```

```
115      FORMAT Exchange $30. Symbol $30. Name $35.;
```

```
116      SET NASDAQ2;
```

```
117      RUN;
```

```
NOTE: There were 72 observations read from the data set WORK.NASDAQ2.
```

```
NOTE: The data set WORK.NASDAQ2 has 72 observations and 14 variables.
```

```
NOTE: DATA statement used (Total process time):
```

```

real time          0.00 seconds
user cpu time       0.00 seconds
system cpu time     0.00 seconds
memory             965.81k
OS Memory          33968.00k
Timestamp          04/28/2022 02:58:50 AM
Step Count          77  Switch Count  2
Page Faults         0
Page Reclaims       120
Page Swaps          0
Voluntary Context Switches  11
Involuntary Context Switches 0
Block Input Operations  0

```

Block Output Operations 272

```
118
119
120     TITLE2 "Q1 Part c) Combine NASDAQ and NYSE Datasets";
121
122     DATA stocks;
123     MERGE NYSE2 NASDAQ2;
124
```

NOTE: There were 359 observations read from the data set WORK.NYSE2.
 NOTE: There were 72 observations read from the data set WORK.NASDAQ2.
 NOTE: The data set WORK.STOCKS has 359 observations and 14 variables.
 NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	1304.25k
OS Memory	34228.00k
Timestamp	04/28/2022 02:58:50 AM
Step Count	78 Switch Count 2
Page Faults	0
Page Reclaims	141
Page Swaps	0
Voluntary Context Switches	14
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```
125     PROC PRINT DATA = stocks;
126     RUN;
```

NOTE: There were 359 observations read from the data set WORK.STOCKS.
 NOTE: PROCEDURE PRINT used (Total process time):

real time	1.59 seconds
user cpu time	1.59 seconds
system cpu time	0.01 seconds
memory	2188.81k
OS Memory	35244.00k
Timestamp	04/28/2022 02:58:52 AM
Step Count	79 Switch Count 0
Page Faults	0
Page Reclaims	414
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	1
Block Input Operations	0
Block Output Operations	656

```
127
128
129     TITLE2 "Q1 Part d) Sort stocks by ascending Sector";
130
131     PROC SORT DATA = stocks out=stocks;
132     by sector;
133     RUN;
```

NOTE: There were 359 observations read from the data set WORK.STOCKS.
 NOTE: The data set WORK.STOCKS has 359 observations and 14 variables.
 NOTE: PROCEDURE SORT used (Total process time):

real time	0.00 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	1196.84k
OS Memory	35764.00k
Timestamp	04/28/2022 02:58:52 AM
Step Count	80 Switch Count 2
Page Faults	0
Page Reclaims	132
Page Swaps	0
Voluntary Context Switches	13
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	272

```
134
135
136     TITLE2 "Q1 Part e) Create the dataset SectorVolatility";
137
```

```

138      DATA SectorVolatility;
139      INFILE datalines dlm=', ' dsd;
140      LENGTH Sector $22.;
141      LENGTH Sector_Volatility $30.;
142      INPUT Sector $ Sector_Volatility $ @@;
143      datalines;

```

NOTE: SAS went to a new line when INPUT statement reached past the end of a line.

NOTE: The data set WORK.SECTORVOLATILITY has 11 observations and 2 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	663.90k
OS Memory	35500.00k
Timestamp	04/28/2022 02:58:52 AM
Step Count	81
Page Faults	0
Page Reclaims	129
Page Swaps	0
Voluntary Context Switches	15
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

155      ;
156      RUN;
157
158      PROC PRINT DATA = SectorVolatility;
159      RUN;

```

NOTE: There were 11 observations read from the data set WORK.SECTORVOLATILITY.

NOTE: PROCEDURE PRINT used (Total process time):

real time	0.01 seconds
user cpu time	0.02 seconds
system cpu time	0.00 seconds
memory	628.21k
OS Memory	35500.00k
Timestamp	04/28/2022 02:58:52 AM
Step Count	82
Page Faults	0
Page Reclaims	65
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	24

```

160
161
162      TITLE2 "Q1 Part f) Sort SectorVolatility by ascending Sector";
163
164      PROC SORT DATA = SectorVolatility out=SectorVolatility;
165      by sector;
166      RUN;

```

NOTE: There were 11 observations read from the data set WORK.SECTORVOLATILITY.

NOTE: The data set WORK.SECTORVOLATILITY has 11 observations and 2 variables.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.00 seconds
user cpu time	0.01 seconds
system cpu time	0.00 seconds
memory	1195.56k
OS Memory	36020.00k
Timestamp	04/28/2022 02:58:52 AM
Step Count	83
Page Faults	0
Page Reclaims	147
Page Swaps	0
Voluntary Context Switches	14
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	272

```

167
168
169      TITLE2 "Q1 Part g) Merge the stocks dataset and the SectorVolatility dataset to create stocks2";
170
171      DATA stocks2;
172      MERGE stocks SectorVolatility;

```

173

NOTE: There were 359 observations read from the data set WORK.STOCKS.
 NOTE: There were 11 observations read from the data set WORK.SECTORVOLATILITY.
 NOTE: The data set WORK.STOCKS2 has 359 observations and 15 variables.
 NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.01 seconds
memory	1341.09k
OS Memory	36020.00k
Timestamp	04/28/2022 02:58:52 AM
Step Count	84
Page Faults	0
Page Reclaims	141
Page Swaps	0
Voluntary Context Switches	9
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```
174      PROC PRINT DATA = stocks2;
175      RUN;
```

NOTE: There were 359 observations read from the data set WORK.STOCKS2.
 NOTE: PROCEDURE PRINT used (Total process time):

real time	1.62 seconds
user cpu time	1.63 seconds
system cpu time	0.00 seconds
memory	822.09k
OS Memory	35500.00k
Timestamp	04/28/2022 02:58:53 AM
Step Count	85
Page Faults	0
Page Reclaims	90
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	4
Block Input Operations	0
Block Output Operations	680

```
176
177
178      /* Task 2: Cleaning and Formatting */
179      TITLE1 "Task 2";
180      TITLE2 "Q2) Replace NA values for DividendYield with 0s in new dataset stocks_divfix";
181
182      DATA stocks_divfix;
183      SET stocks2;
184      IF DividendYield = . THEN DividendYield = 0;
185      RUN;
```

NOTE: There were 359 observations read from the data set WORK.STOCKS2.
 NOTE: The data set WORK.STOCKS_DIVFIX has 359 observations and 15 variables.
 NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	968.21k
OS Memory	35760.00k
Timestamp	04/28/2022 02:58:53 AM
Step Count	86
Page Faults	0
Page Reclaims	102
Page Swaps	0
Voluntary Context Switches	17
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```
186
187
188      TITLE2 "Q3) Remove rows for PricetoEarnings and PricetoBook that are NA. Do this in a new dataset called stocks_clean";
189
190      DATA stocks_clean;
191      SET STOCKS_DIVFIX;
192      IF PricetoEarnings = . THEN DELETE;
193      IF PricetoBook = . THEN DELETE;
194      RUN;
```

NOTE: There were 359 observations read from the data set WORK.STOCKS_DIVFIX.

NOTE: The data set WORK.STOCKS_CLEAN has 351 observations and 15 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	969.21k
OS Memory	35760.00k
Timestamp	04/28/2022 02:58:53 AM
Step Count	87 Switch Count 2
Page Faults	0
Page Reclaims	103
Page Swaps	0
Voluntary Context Switches	12
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

195
196      /* Task 3: Variable Creation and Subsetting */
197      TITLE1 "Task 3";
198      TITLE2 "Q4) Create a new variable called spread = FiftytwoWeekHigh - FiftytwoWeekLow";
199
200      DATA stocks_clean;
201      SET STOCKS_DIVFIX;
202      IF PricetoEarnings = . THEN DELETE;
203      IF PricetoBook = . THEN DELETE;
204      Spread=FiftytwoWeekHigh-FiftytwoWeekLow;
205      RUN;

```

NOTE: There were 359 observations read from the data set WORK.STOCKS_DIVFIX.

NOTE: The data set WORK.STOCKS_CLEAN has 351 observations and 16 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	972.59k
OS Memory	35760.00k
Timestamp	04/28/2022 02:58:53 AM
Step Count	88 Switch Count 2
Page Faults	0
Page Reclaims	104
Page Swaps	0
Voluntary Context Switches	9
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

206
207
208      TITLE2 "Q5) Create a new variable called Earnings_category";
209
210      DATA stocks_clean;
211      SET STOCKS_DIVFIX;
212      IF PricetoEarnings = . THEN DELETE;
213      IF PricetoBook = . THEN DELETE;
214      Spread=FiftytwoWeekHigh-FiftytwoWeekLow;
215      IF EarningsperShare<0 THEN EarningsCategory = "Loss";
216      IF EarningsperShare>=0 AND EarningsperShare<3 THEN EarningsCategory = "Small";
217      IF EarningsperShare>=3 AND EarningsperShare<10 THEN EarningsCategory = "Good";
218      IF EarningsperShare>=10 THEN EarningsCategory = "Strong";
219      RUN;

```

NOTE: There were 359 observations read from the data set WORK.STOCKS_DIVFIX.

NOTE: The data set WORK.STOCKS_CLEAN has 351 observations and 17 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	990.71k
OS Memory	35760.00k
Timestamp	04/28/2022 02:58:53 AM
Step Count	89 Switch Count 2
Page Faults	0
Page Reclaims	104
Page Swaps	0
Voluntary Context Switches	10
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```
220
221
222     TITLE2 "Q6) Create dividends dataset";
223
224     DATA dividends;
225     SET stocks_clean;
226     dividends_mean = mean(DividendYield);
227     RUN;
```

NOTE: There were 351 observations read from the data set WORK.STOCKS_CLEAN.

NOTE: The data set WORK.DIVIDENDS has 351 observations and 18 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	988.56k
OS Memory	35760.00k
Timestamp	04/28/2022 02:58:53 AM
Step Count	90
Page Faults	0
Page Reclaims	101
Page Swaps	0
Voluntary Context Switches	10
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```
228
229     OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
240
```