

Financial Data Management by Prof. Xing Gao

Homework 2

Submitted by- Vineeth Kotala

/*1.1. Downloading the data */

```
libname HW2 "/home/u63116105/sasuser.v94/FIN557/HW2"; proc contents  
data=hw2.ownership;  
run;
```

The CONTENTS Procedure

| | | | |
|---------------------|-----------------------|----------------------|-------|
| Data Set Name | HW2.OWNERSHIP | Observations | 13070 |
| Member Type | DATA | Variables | 5 |
| Engine | V9 | Indexes | 0 |
| Created | 22/02/2023 12:50:36 | Observation Length | 40 |
| Last Modified | 22/02/2023 12:50:36 | Deleted Observations | 0 |
| Protection | | Compressed | NO |
| Data Set Type | | Sorted | NO |
| Label | | | |
| Data Representation | WINDOWS_64 | | |
| Encoding | us-ascii ASCII (ANSI) | | |

Engine/Host Dependent Information

| | |
|----------------------------|---|
| Data Set Page Size | 65536 |
| Number of Data Set Pages | 9 |
| First Data Page | 1 |
| Max Obs per Page | 1632 |
| Obs in First Data Page | 1594 |
| Number of Data Set Repairs | 0 |
| Filename | /home/u63116105/sasuser.v94/FIN557/HW2/ownership.sas7bdat |
| Release Created | 9.0401M7 |
| Host Created | Linux |
| Inode Number | 10207136234 |
| Access Permission | rw-r--r-- |
| Owner Name | u63116105 |
| File Size | 640KB |
| File Size (bytes) | 655360 |

Alphabetic List of Variables and Attributes

| # | Variable | Type | Len | Format | Label |
|---|----------|------|-----|-----------|--|
| 3 | FDATE | Num | 8 | MMDDYY10. | File Date |
| 1 | MGRNO | Num | 8 | | Manager Number |
| 4 | SHARES | Num | 8 | | Shares Held at End of Qtr |
| 5 | SHROUT2 | Num | 8 | | Shares Outstanding in 1000s, as of FDATE |
| 2 | TICKER | Char | 6 | | Ticker Symbol |

/*1.2 Creating new variables */

```
data own1;
set hw2.ownership;
SHROUT=SHROUT2*1000; PERCENT=(SHARES/SHROUT)*100;
keep FDATE TICKER SHARES SHROUT SHROUT2 PERCENT;
run;
proc print data=own1 (obs=10);
run;
```

| Obs | TICKER | FDATE | SHARES | SHROUT2 | SHROUT | PERCENT |
|-----|--------|------------|---------|---------|------------|----------|
| 1 | AAPL | 12/31/2018 | 19927 | 4745398 | 4745398000 | 0.000420 |
| 2 | AAPL | 12/31/2018 | 7322 | 4745398 | 4745398000 | 0.000154 |
| 3 | AAPL | 12/31/2018 | 996 | 4745398 | 4745398000 | 0.000021 |
| 4 | AAPL | 12/31/2018 | 1038011 | 4745398 | 4745398000 | 0.021874 |
| 5 | AAPL | 12/31/2018 | 8235 | 4745398 | 4745398000 | 0.000174 |
| 6 | AAPL | 12/31/2018 | 67680 | 4745398 | 4745398000 | 0.001426 |
| 7 | AAPL | 12/31/2018 | 29611 | 4745398 | 4745398000 | 0.000624 |
| 8 | AAPL | 12/31/2018 | 5966 | 4745398 | 4745398000 | 0.000126 |
| 9 | AAPL | 12/31/2018 | 17126 | 4745398 | 4745398000 | 0.000361 |
| 10 | AAPL | 12/31/2018 | 1212 | 4745398 | 4745398000 | 0.000026 |

/*1.3 Sorting the data */

```
proc sort data=own1 out=own2;
by TICKER descending PERCENT;
run;
proc print data=own2 (obs=10);
run;
```

| Obs | TICKER | FDATE | SHARES | SHROUT2 | SHROUT | PERCENT |
|-----|--------|------------|-----------|---------|------------|---------|
| 1 | AAPL | 12/31/2018 | 338533988 | 4745398 | 4745398000 | 7.13394 |
| 2 | AAPL | 12/31/2018 | 295605914 | 4745398 | 4745398000 | 6.22932 |
| 3 | AAPL | 12/31/2018 | 249589329 | 4745398 | 4745398000 | 5.25961 |
| 4 | AAPL | 12/31/2018 | 185419773 | 4745398 | 4745398000 | 3.90736 |
| 5 | AAPL | 12/31/2018 | 108036612 | 4745398 | 4745398000 | 2.27666 |
| 6 | AAPL | 12/31/2018 | 59311465 | 4745398 | 4745398000 | 1.24987 |
| 7 | AAPL | 12/31/2018 | 58414412 | 4745398 | 4745398000 | 1.23097 |
| 8 | AAPL | 12/31/2018 | 47548838 | 4745398 | 4745398000 | 1.00200 |
| 9 | AAPL | 12/31/2018 | 43621951 | 4745398 | 4745398000 | 0.91925 |
| 10 | AAPL | 12/31/2018 | 37252585 | 4745398 | 4745398000 | 0.78503 |

/*1.4 total percentage of institutional ownership and the total number of institutional investors for each firm */

data own3; set own2;

by TICKER descending PERCENT; if first.TICKER=1 then SUM=0; SUM=SUM+PERCENT;

if first.TICKER=1 then N=0;

N=N+1;

if last.TICKER=1 then output;

run;

proc print data=own3;

run;

| Obs | TICKER | FDATE | SHARES | SHROUT2 | SHROUT | PERCENT | SUM | N |
|-----|--------|------------|--------|---------|------------|------------|-----|---|
| 1 | AAPL | 12/31/2018 | 6 | 4745398 | 4745398000 | .000000126 | . | . |
| 2 | FB | 12/31/2018 | 4 | 2402466 | 2402466000 | .000000166 | . | . |
| 3 | IBM | 12/31/2018 | 2 | 892479 | 892479000 | .000000224 | . | . |
| 4 | INTC | 12/31/2018 | 11 | 4564000 | 4564000000 | .000000241 | . | . |
| 5 | MSFT | 12/31/2018 | 16 | 7676219 | 7676219000 | .000000208 | . | . |
| 6 | ORCL | 12/31/2018 | 11 | 3588919 | 3588919000 | .000000306 | . | . |

/*1.5 Finding the top 10 investors */

```
data own4; set own2;
by TICKER descending PERCENT; if first.TICKER=1 then SUM=0; SUM=SUM+PERCENT;
if first.TICKER=1 then N=0;
N=N+1;
if N<=10 then output;
run;
proc print data=own4 (obs=10);
run;
```

| Obs | TICKER | FDATE | SHARES | SHROUT2 | SHROUT | PERCENT | SUM | N |
|-----|--------|------------|-----------|---------|------------|---------|---------|---|
| 1 | AAPL | 12/31/2018 | 338533988 | 4745398 | 4745398000 | 7.13394 | 7.13394 | 1 |
| 2 | AAPL | 12/31/2018 | 295605914 | 4745398 | 4745398000 | 6.22932 | . | . |
| 3 | AAPL | 12/31/2018 | 249589329 | 4745398 | 4745398000 | 5.25961 | . | . |
| 4 | AAPL | 12/31/2018 | 185419773 | 4745398 | 4745398000 | 3.90736 | . | . |
| 5 | AAPL | 12/31/2018 | 108036612 | 4745398 | 4745398000 | 2.27666 | . | . |
| 6 | AAPL | 12/31/2018 | 59311465 | 4745398 | 4745398000 | 1.24987 | . | . |
| 7 | AAPL | 12/31/2018 | 58414412 | 4745398 | 4745398000 | 1.23097 | . | . |
| 8 | AAPL | 12/31/2018 | 47548838 | 4745398 | 4745398000 | 1.00200 | . | . |
| 9 | AAPL | 12/31/2018 | 43621951 | 4745398 | 4745398000 | 0.91925 | . | . |
| 10 | AAPL | 12/31/2018 | 37252585 | 4745398 | 4745398000 | 0.78503 | . | . |

/*1.6 Finding large investors within 20% of ownership */

```
data own5; set own2;
by TICKER descending PERCENT; if first.TICKER=1 then SUM=0; SUM=SUM+PERCENT;
if first.TICKER=1 then N=0;
N=N+1;
if SUM<=20 then output;
run;
proc print data=own5 (obs=10);
run;
```

| Obs | TICKER | FDATE | SHARES | SHROUT2 | SHROUT | PERCENT | SUM | N |
|-----|--------|------------|-----------|---------|------------|---------|---------|---|
| 1 | AAPL | 12/31/2018 | 338533988 | 4745398 | 4745398000 | 7.13394 | 7.13394 | 1 |
| 2 | AAPL | 12/31/2018 | 295605914 | 4745398 | 4745398000 | 6.22932 | . | . |
| 3 | AAPL | 12/31/2018 | 249589329 | 4745398 | 4745398000 | 5.25961 | . | . |
| 4 | AAPL | 12/31/2018 | 185419773 | 4745398 | 4745398000 | 3.90736 | . | . |
| 5 | AAPL | 12/31/2018 | 108036612 | 4745398 | 4745398000 | 2.27666 | . | . |
| 6 | AAPL | 12/31/2018 | 59311465 | 4745398 | 4745398000 | 1.24987 | . | . |
| 7 | AAPL | 12/31/2018 | 58414412 | 4745398 | 4745398000 | 1.23097 | . | . |
| 8 | AAPL | 12/31/2018 | 47548838 | 4745398 | 4745398000 | 1.00200 | . | . |
| 9 | AAPL | 12/31/2018 | 43621951 | 4745398 | 4745398000 | 0.91925 | . | . |
| 10 | AAPL | 12/31/2018 | 37252585 | 4745398 | 4745398000 | 0.78503 | . | . |

/* 2 Manipulating data with character functions*/

/*2.1 Removing characters from a string*/

```
proc contents data=hw2.location;
```

```
run;
```

```
data location1;
```

```
set hw2.location; location1=compbl(cat(LOCATION)); id1=compress(ID," "); id2=compress(ID," -");
```

```
run;
```

```
proc print data=location1 (obs=10);
```

```
run;
```


The CONTENTS Procedure

| | | | |
|----------------------------|---------------------------|-----------------------------|----|
| Data Set Name | HW2.LOCATION | Observations | 97 |
| Member Type | DATA | Variables | 2 |
| Engine | V9 | Indexes | 0 |
| Created | 19/03/2020 14:58:45 | Observation Length | 39 |
| Last Modified | 19/03/2020 14:58:45 | Deleted Observations | 0 |
| Protection | | Compressed | NO |
| Data Set Type | | Sorted | NO |
| Label | | | |
| Data Representation | WINDOWS_64 | | |
| Encoding | wlatin1 Western (Windows) | | |

| Engine/Host Dependent Information | |
|-----------------------------------|--|
| Data Set Page Size | 65536 |
| Number of Data Set Pages | 1 |
| First Data Page | 1 |
| Max Obs per Page | 1674 |
| Obs in First Data Page | 97 |
| Number of Data Set Repairs | 0 |
| ExtendObsCounter | YES |
| Filename | /home/u63116105/sasuser.v94/FIN557/HW2/location.sas7bdat |
| Release Created | 9.0401M3 |
| Host Created | X64_8HOME |
| Inode Number | 10207784587 |
| Access Permission | rw-r--r-- |
| Owner Name | u63116105 |
| File Size | 128KB |
| File Size (bytes) | 131072 |

| Alphabetic List of Variables and Attributes | | | | | |
|---|----------|------|-----|-----------|-----------|
| # | Variable | Type | Len | Format | Informat |
| 2 | ID | Char | 13 | | |
| 1 | Location | Char | 26 | \$CHAR26. | \$CHAR26. |

| Obs | Location | ID | location1 | id1 | id2 |
|-----|--------------------------|---------------|--------------------------|--------------|-------------|
| 1 | OWASE, Mie, JA | JA000047663 | OWASE, Mie, JA | JA000047663 | JA000047663 |
| 2 | TAKADA, Tokyo, JA | JA-000047612 | TAKADA, Tokyo, JA | JA-000047612 | JA000047612 |
| 3 | ABURATSU, Miyazaki, JA | JA000047835 | ABURATSU, Miyazaki, JA | JA000047835 | JA000047835 |
| 4 | NAZE, Kagoshima, JA | JA 0000 47909 | NAZE, Kagoshima, JA | JA000047909 | JA000047909 |
| 5 | TSURUGA, Fukui, JA | JA 0000 47631 | TSURUGA, Fukui, JA | JA000047631 | JA000047631 |
| 6 | TOYAMA, Toyama, JA | JA000047607 | TOYAMA, Toyama, JA | JA000047607 | JA000047607 |
| 7 | KANAZAWA, Ishikawa, JA | JA000047605 | KANAZAWA, Ishikawa, JA | JA000047605 | JA000047605 |
| 8 | MIYAKE-JIMA, Tokyo, JA | JA000047677 | MIYAKE-JIMA, Tokyo, JA | JA000047677 | JA000047677 |
| 9 | FUSHIKI, Toyama, JA | JA000047606 | FUSHIKI, Toyama, JA | JA000047606 | JA000047606 |
| 10 | MIYAKONOJO, Miyazaki, JA | JA000047829 | MIYAKONOJO, Miyazaki, JA | JA000047829 | JA000047829 |

/*2.2 Extracting words from a string */

```
data location2;
    set HW2.location;
    location=compbl(cat(LOCATION));
    city1=scan(LOCATION,1);
    prefecture1=scan(LOCATION,2);
    country1=scan(LOCATION,-1);
run;
proc print data=location2 (obs=10);
run;
```

| Obs | Location | ID | city1 | prefecture1 | country1 |
|-----|--------------------------|---------------|------------|-------------|----------|
| 1 | OWASE, Mie, JA | JA000047663 | OWASE | Mie | JA |
| 2 | TAKADA, Tokyo, JA | JA-000047612 | TAKADA | Tokyo | JA |
| 3 | ABURATSU, Miyazaki, JA | JA000047835 | ABURATSU | Miyazaki | JA |
| 4 | NAZE, Kagoshima, JA | JA 0000 47909 | NAZE | Kagoshima | JA |
| 5 | TSURUGA, Fukui, JA | JA 0000 47631 | TSURUGA | Fukui | JA |
| 6 | TOYAMA, Toyama, JA | JA000047607 | TOYAMA | Toyama | JA |
| 7 | KANAZAWA, Ishikawa, JA | JA000047605 | KANAZAWA | Ishikawa | JA |
| 8 | MIYAKE-JIMA, Tokyo, JA | JA000047677 | MIYAKE | JIMA | JA |
| 9 | FUSHIKI, Toyama, JA | JA000047606 | FUSHIKI | Toyama | JA |
| 10 | MIYAKONOJO, Miyazaki, JA | JA000047829 | MIYAKONOJO | Miyazaki | JA |

```

/*2.3 Correcting inconsistent character case */
data location3;
    set HW2.location;
    location=compbl(cat(LOCATION));
    city2=propcase(scan(LOCATION,1," "), "");
run;

proc print data=location3 (obs=10);
run;

```

| Obs | Location | ID | city2 |
|-----|--------------------------|---------------|-------------|
| 1 | OWASE, Mie, JA | JA000047663 | Owase |
| 2 | TAKADA, Tokyo, JA | JA-000047612 | Takada |
| 3 | ABURATSU, Miyazaki, JA | JA000047835 | Aburatsu |
| 4 | NAZE, Kagoshima, JA | JA 0000 47909 | Naze |
| 5 | TSURUGA, Fukui, JA | JA 0000 47631 | Tsuruga |
| 6 | TOYAMA, Toyama, JA | JA000047607 | Toyama |
| 7 | KANAZAWA, Ishikawa, JA | JA000047605 | Kanazawa |
| 8 | MIYAKE-JIMA, Tokyo, JA | JA000047677 | Miyake-jima |
| 9 | FUSHIKI, Toyama, JA | JA000047606 | Fushiki |
| 10 | MIYAKONOJO, Miyazaki, JA | JA000047829 | Miyakonojo |

```

/*2.4 Filtering the rows */
data location4;
    set HW2.location;
    location=compbl(cat(LOCATION));
    prefecture2=scan(LOCATION,2," ");
if find(prefecture2,"Tokyo") then output;

```

Total rows: 4 Total columns: 3



 Rows 1-4
 


| | Location | ID | prefecture2 |
|---|------------------------|--------------|-------------|
| 1 | TAKADA, Tokyo, JA | JA-000047612 | Tokyo |
| 2 | MIYAKE-JIMA, Tokyo, JA | JA000047677 | Tokyo |
| 3 | UENO, Tokyo, JA | JA-000047649 | Tokyo |
| 4 | TOKYO, Tokyo, JA | JA-000047662 | Tokyo |