proc print data=class.interviewdata (obs=20);

run;

proc contents data=class.interviewdata;

run;

proc sql;

create table branch as select count(distinct Branch) AS identifiervariable

from class.interviewdata;

proc sql;

/\* Total Sales by year \*/

create table TotalSales AS select year, sum(Sales) AS Total\_Sales

from class.interviewdata

group By Year;

/\* Average Customers per month, month of Apri\*/

proc sql;

create table AverageCustomers AS select avg(customers), month

from class.interviewdata

group By month;

proc means data=class.interviewdata nmiss;

var Year Customers Sales;

run;

/\* Since mean is less than median, data is left-skewed because of the negative sales \*/

proc means data=class.interviewdata mean std max min median q1 q3 p5 p10 p25

p50 p75 p90 p95;

var Sales;

run;

/\* Logistically speaking, it is not reasonable to have negative customers in one month and 226000 customers \*/

/\* in another month. These numbers must be caused by human errors.\*/

proc means data=class.interviewdata mean std max min median q1 q3 p5 p10 p25

p50 p75 p90 p95;

var Customers;

run;

/\* In this case, year of 20144 is not reasonable and must be caused by human errors \*/

proc univariate data=class.interviewdata;

var Year;

run;

proc freq data=class.interviewdata;

table Customers \* Sales /norow nocol;

run;