```
PROC PRINT DATA=WORK.inc;
RUN;

/* changing values of one column into lowercase */
data work.income;
    set work.inc;
    Occupation = lowcase(occupation);
run;


/* changing names of columns */
data work.income(RENAME=(All_workers=Total
M_workers=Males
F_workers=Females
 ));
    set work.income;
run;

proc sql;
 select count(*) as initail_count from work.income;
quit;

/* removing rows with null values */
data work.income2;
    SET work.income;
    IF cmiss(of _all_) gt 0 then
     DELETE;
run;

proc sql;
 select count(*) as final_count from work.income2;
quit;

/* dropping columns */
data work.income2(drop=All_weekly);
set work.income2;
run;

/* finding yearly salary, gender wise */
data work.income2;
set work.income2;
M_yearly=M_weekly*52;
F_yearly=F_weekly*52;
run;

/* finding tax percentage on yearly salary */
data work.income2;
set work.income2;
if M_yearly gt 0 AND M_yearly le 10275 then MTP=10;
else if M_yearly ge 10276 AND M_yearly le 41775 then MTP=12;
else if M_yearly ge 41776 AND M_yearly le 50000 then MTP=22;
else if M_yearly ge 51000 AND M_yearly le 170050 then MTP=24;
else if M_yearly ge 170051 AND M_yearly le 215950 then MTP=32;
else if M_yearly ge 215951 AND M_yearly le 539900 then MTP=50;
else if M_yearly ge 539901 then MTP=37;

if F_yearly gt 0 AND F_yearly le 10275 then FTP=10;
else if F_yearly ge 10276 AND F_yearly le 41775 then FTP=12;
else if F_yearly ge 41776 AND F_yearly le 45000 then FTP=22;
else if F_yearly ge 45001 AND F_yearly le 170050 then FTP=24;
else if F_yearly ge 170051 AND F_yearly le 215950 then FTP=32;
else if F_yearly ge 215951 AND F_yearly le 539900 then FTP=50;
else if F_yearly ge 539901 then FTP=37;

run;
```

```
PROC PRINT DATA=WORK.income2;
RUN;


/* finding means */
proc means data=work.income2 mean;
    var M_yearly;
    var F_yearly;
    var MTP;
    var FTP;
run;

/* finding female percentage in various occupation */
data work.income2;
set work.income2;
F_percentage=(Females/Total)*100;
run;

/* if female percentage more than 50 then labelling as more femals , vice a versa for femals */
data work.income2;
set work.income2;
if F_percentage gt 50 then inference='More_Females';
else inference='More_Males';
run;


PROC PRINT DATA=WORK.income2;
RUN;


data work.income3;
set work.income2;
where Occupation='business' or
Occupation='education administrators' or
Occupation='purchasing magers' or
Occupation='food service magers' or
Occupation='engineering';
run;

/* PROC PRINT DATA=WORK.income3; */
/* RUN; */

/* histogram plot of infernce which tells where more femals or more males */
/* inference= more pofessions with more males than more females */
proc gchart data=work.income3;
/* Hbar Males; */
vBAR inference / type=percent;
run;

# histogram plot of number of femals in which tax bracket
proc gchart data=work.income3;
vbar FTP / discrete type=percent;
 subgroup=inference;
 run;

/* segregating male income into very low,low,normal,high,very high */
data work.income2;
set work.income2;
if M_yearly le 10000 then col= ' very_Low';
else if M_yearly le 30000 then col= 'Low';
else if M_yearly le 40000 then col= 'normal';
else if M_yearly ge 50001 then col= 'high';
else if M_yearly ge 60000 then col= 'very high';
else col='outlier';
```

```
run;

/* histogram plot of various tags on male salary */
proc gchart data=work.income2;
/* Hbar Males; */
vBAR col / type=percent;
run;


/* inference= majority of females have their salary between 30000-45000 */
proc gplot data=work.income2;
 title ‘Scatter Plot of females and their yearly salary’;
plot Females* F_yearly;
/* plot F_yearly*Females; */
 run;
```