```
*** Karan Singal SAS Project:
*** Dataset Provided by Instructor Ar Kar Min ;
*** Number of datasets =2;
*** Retail sales analysis;
*** Date of submition ;

* Dataset no 1 Main Dataset ;

PROC IMPORT OUT = WORK.project
DATAFILE='C:\Users\Admin\OneDrive\Desktop\Project\Final SAS project
\Karan Singal SAS Project\transactionhistoryforcurrentcustomers.csv'
DBMS= csv replace;
GETNAMES=yes;
datarow=2;
RUN;

proc contents data = work.project;
run;


* Dataset no 2 EC90 ;

PROC IMPORT OUT = WORK.project2
DATAFILE= 'C:\Users\Admin\OneDrive\Desktop\Project\Final SAS project
\Karan Singal SAS Project\ec90 data.csv'
DBMS= csv replace;
GETNAMES=yes;
datarow=2;
RUN;

proc contents data = work.project2;
run;

*Distinct values in Main Dataset;
TITLE "Count of Distinct Customer IDs in Project1";
    PROC SQL;
    SELECT COUNT(Customer_ID)AS TOTAL_COUNT, COUNT(DISTINCT
Customer_ID) AS UNIQUE_COUNT
    FROM WORK.project
    ;
    QUIT;

*Distinct values in Dataset EC90;
TITLE "Count of Distinct Customer IDs in Project2";
    PROC SQL;
    SELECT COUNT(Customer_Number)AS TOTAL_COUNT, COUNT(DISTINCT
Customer_Number) AS UNIQUE_COUNT
    FROM WORK.project2
    ;
    QUIT;


*INNER JOIN;
```

```
proc sql ;
create table work.Innerjoin as
select * from WORK.project as a, WORK.project2 as b
      where a.Customer_ID = b.Customer_Number;
      run;



*DELETE TABLES;

data work.Final_data(drop = Order_Number Order_First_Time Sales_amount
Item_Num Category_code);
set WORK.Innerjoin;
run;



*REMOVING DUP RECORDS ;

proc sort data = work.Final_data noduprecs;
      by _all_ ;
Run;

*MISSING VALUES;

proc format;
 value $missfmt ' '='Missing' other='Not Missing';
 value  missfmt  . ='Missing' other='Not Missing';
run;

proc freq data= WORK.Final_data;
format _CHAR_ $missfmt.;
tables _CHAR_ / missing missprint nocum nopercent;
format _NUMERIC_ missfmt.;
tables _NUMERIC_ / missing missprint nocum nopercent;
run;


proc contents data = work.Final_data;
run;

proc means data = work.Final_data;
run;


*Data Preparation Scatter plot matrix;

proc sgscatter data=WORK.Final_data;
title 'Order Date vs Price vs Quantity';
matrix Order_Date price quantity / group= prov diagonal=(histogram
kernel);
 run;
```

```
*Q1 Histogram of sales by month ;

proc sgplot data=WORK.Final_data;
title 'Order By month';
    histogram Order_Date / scale=count fillattrs=(transparency=0.25)
        dataskin=sheen;
    density Order_Date / type=Kernel;
run;
quit;

*Q2 Bar Chart Sales by Province;

proc sgplot data=WORK.Final_data;
title 'Sales by Province';
    vbar Prov / fillattrs=(color=CXe40c0c transparency=0.25) datalabel
        dataskin=sheen;
run;


*Q 3 PIE cart by Sourse;

proc template;
define statgraph simplepie;
  begingraph ;
    entrytitle "Sale by Sourse";
    layout REGION;
      piechart category=Source /stat=pct dataskin=gloss ;
    endlayout;
  endgraph;
end;
run;

proc sgrender data=work.Final_data
              template=simplepie;
run;

        *2********************;

proc gchart data=WORK.Final_data;
pie3d Source/pct=Outside;
run;



* Q 4 Province wise sale first time or not;


proc sgplot data=WORK.project2;
    vbar Prov / group='Order_First_Time'n groupdisplay=cluster;
    yaxis grid;
```

```
run;

* Graph 2 % for first time buyer or not ;
proc sgplot data=WORK.project2;
    vbar 'Order_First_Time'n / fillattrs=(color=CX3cd888 transparency=
0.25)
        datalabel stat=percent dataskin=sheen;
run;



*Q 5 Bubble plot by quantity;


proc sgplot data=WORK.Final_data;
title'Order date vs price';
    bubble x=Order_Date y=price size=price/ fillattrs=(color=CX911a32
        transparency=0.25) dataskin=gloss bradiusmin=7 bradiusmax=14;
run;


* Q 6 Box Plot for Quantity;


proc sgplot data=WORK.Final_data;

title 'Box plot of Quantity';
    vbox Quantity /fillattrs=(transparency=0.25) notches
dataskin=sheen;
run;
quit;

    *Box plot for Price;
proc sgplot data=WORK.Final_data;
title 'Box plot of Price';
    vbox price / fillattrs=(transparency=0.25) notches dataskin=sheen;
    yaxis max=1000 grid;
run;
quit;



*Q 7 chi square;
proc freq data = work.Project2;
title 'Chi-Square Statistics for Source and Order first time or not ';
tables Source*Order_First_Time
/chisq
;
run;



*Heat map orderdate quantity;
```

```
proc sgplot data=WORK.Final_data;
    heatmap x=price y=quantity  / name='HeatMap';
    gradlegend 'HeatMap';
run;



* which city of ontario is spending more than 1000 $;

ods noproctitle;

proc sort data=WORK.Final_data out=work.temp_On_Price;
    by Prov;
run;

 proc freq data=WORK.temp_On_Price;
    where Prov ="ON" and price >1000;
    tables City / plots=(freqplot);
run;

proc delete data=work.temp_On_Price;
run;


* which City of ontario is buying more than 10  items;

ods noproctitle;

proc sort data=WORK.Final_data out=work._chardata_sorted;
    by Prov;
run;

 proc freq data=WORK._chardata_sorted;
    where Prov ="ON" and quantity >10;
    tables City / plots=(freqplot);
run;

proc delete data=work._chardata_sorted;
run;



**************Testing***************;


* correlation trying;
proc corr data=sashelp.iris plots=matrix(histogram);
run;
```