

```

PROC IMPORT DATAFILE ="/folders/myfolders/sasuser.v94/Walmart.csv"
    OUT=WORK.Walmart1
    DBMS=CSV
    REPLACE;
    Getnames=YES;

RUN;

PROC PRINT DATA=WORK.Walmart1; RUN;

PROC CONTENTS DATA = WORK.Walmart1;
RUN;

PROC MEANS DATA =WORK.Walmart1 nmiss;
RUN;

PROC MEANS DATA = WORK.Walmart1 max;
    by Store;
    var Weekly_Sales;
RUN;

PROC SORT DATA=WORK.walmart1;
    by descending Weekly_Sales;
RUN;
/* Which store has maximum standard deviation */
PROC SUMMARY DATA =WORK.Walmart1;
    class store;
    output out= walmart_standard1(drop= _type_ _freq_) std(weekly_sales)=sd_max;
RUN;
PROC PRINT DATA=work.walmart_standard1;
RUN;
PROC SORT DATA=work.walmart_standard1;
    by descending sd_max;
RUN;
PROC PRINT DATA=work.walmart_standard1;
RUN;
/* Find out the coefficient of mean to standard deviation */
PROC MEANS DATA=work.walmart1 nonobs cv;
class store;
var weekly_sales;
RUN;
/* OR */
PROC MEANS DATA=work.walmart1 nonobs cv;
RUN;
/* Which store/s has good quarterly growth rate in Q3'2012 */
/* Filter year(2012) */
DATA date_12;
set work.walmart1;
where year(Date)=2012;
RUN;

PROC PRINT DATA= date_12;
RUN;
/* Calculate growth rate */

PROC SORT DATA= work.date_12;
    BY Store Date Weekly_Sales ;
RUN ;

DATA growth;
    format growth_rate percent8.2;
    set work.date_12;
    by store date weekly_sales;
    lag_sales = ifn(first.store,0,lag(weekly_sales));
    growth_rate = (weekly_sales/lag_sales)-1;
    drop lag_sales;
RUN;
PROC PRINT DATA=growth;
RUN;
/* Convert the normat DATA into timeseries DATA */
PROC TIMESERIES DATA= growth out= good_growth;
    by store;

```

```

        id date interval= qtr accumulate=total;
        var growth_rate;

RUN;
PROC PRINT DATA= good_growth;
RUN;
/* From timeseries DATA filtered only Q3 observations */
DATA good_growth_rate;
    set good_growth;
    where qtr(date)= 3;

RUN;
PROC PRINT DATA=good_growth_rate;
RUN;
/* Now Sort the DATA to see the good growth rate store wise */
PROC SORT DATA= good_growth_rate;
    by descending growth_rate;

RUN;
PROC PRINT DATA= good_growth_rate;
RUN;
/* Some holidays have a negative impact on sales.
Find out holidays which have higher sales than
the mean sales in non-holiday season for all stores together */
/* Separate the holiday dates from main DATASET's date */
DATA holiday;
    set work.walmart1;
    where holiday_flag=1;

RUN;
PROC PRINT DATA=holiday;
RUN;
/* Separate the non-holiday dates from main DATASET's date */
DATA non_holiday;
    set work.walmart1;
    where holiday_flag=0;

RUN;
PROC PRINT DATA=non_holiday;
RUN;
/* Calculate the mean weekly_sales of the non-holiday DATA */
PROC MEANS DATA= non_holiday mean nonobs;
output out= mean_sales;
var weekly_sales;

RUN;
/* Compare the mean weekly_sales of the non-holiday DATA with weekly_sales of the holiday DATA */
PROC SQL;
create table holiday_sales as
select store, weekly_sales, date, holiday_flag as holiday,
case
when weekly_sales > 1041256.38 then 'Higher'
when weekly_sales < 1041256.38 then 'Lower'
end
as higher_sales
from holiday;
quit;
PROC PRINT DATA= holiday_sales;
RUN;
/* Finally found out holidays which have higher sales than
the mean sales in non-holiday season for all stores together */
DATA higher_holiday_sales;
set work.holiday_sales;
where higher_sales = 'Higher';
drop higher_sales;
title 'Higher Sales during Holidays';
RUN;
PROC PRINT DATA= higher_holiday_sales;
RUN;
/* Provide a monthly and semester view of sales in units and give insights */
/* Monthly view of sales in units */
/* Convert walmart DATA into timeseries DATA */

PROC SORT DATA= work.WALMART1
    out= work.WALMART;
    BY Store ;

RUN ;
PROC PRINT DATA= WORK.WALMART;
RUN;
DATA WALMART1 (rename=(var Sn= var ID));data prices(rename=(Price=RATE_PRICE_TBA Coupon=pct_cpn PriceAsofDateTime=DT_MKT)
keep= PRICE_TBA pct_cpn DT_MKT);

```

```

set QUERY_FOR_PROCESS_REQUEST;
run;
data prices(rename=(Price=RATE_PRICE_TBA Coupon=pct_cpn PriceAsofDateTime=DT_MKT)
keep=ID_PRODT PRICE_TBA pct_cpn DT_MKT);
set QUERY_FOR_PROCESS_REQUEST;
run;
RUN;
PROC TIMESERIES DATA= work.walmart
out= work.monthly_sales;
by store;
id date interval=month accumulate=total;
var weekly_sales holiday_flag temperature fuel_price cpi unemployment;
RUN;
PROC PRINT DATA=work.monthly_sales;
format weekly_sales dollar16.2;
RUN;
/* Giving insights */
/* Checking the correlation */
PROC CORR DATA= work.monthly_sales;
RUN;
/* 1. Doing Comparison */
/* a) Bar Chart */
PROC SGPLOT DATA = work.monthly_sales;
hbar store/response = weekly_sales stat= sum
DATAlabel DATAlabelattrs=(weight=bold);
title 'Total Views by Store';
RUN;
/* b) Clustered Bar Chart / Column Chart */
DATA date;
set work.monthly_sales;
month = month(Date);
month_name=PUT(Date,monname.);
put month_name= @;
RUN;
PROC PRINT DATA= date;
RUN;
PROC SGPLOT DATA= date;
vbar store/ response= weekly_sales group=month_name groupdisplay=cluster
DATAlabel DATAlabelattrs = (weight = bold) DATAskin=gloss; yaxis grid;
title 'Total View by monthly wise';
RUN;
/* 2) Studying relationship */
/* a) Bubble Chart */
PROC SGPLOT DATA = work.monthly_sales;
bubble X=weekly_sales Y=store size= weekly_sales
/fillattrs=(color = teal) DATAlabel = store;
RUN;
/* b) Scatter Plot for Relationship */
PROC SGPLOT DATA = work.monthly_sales;
title 'Relationship of Store with Weekly_sales';
scatter X= weekly_sales Y = store/
markerattrs=(symbol=circlefilled size=15);
RUN;
/* 3. Studying Distribution */
/* a) Histogram */
PROC SGPLOT DATA = work.monthly_sales;
histogram weekly_sales/fillattrs=(color = steel)scale = proportion;
density weekly_sales;
RUN;
/* b) Scatter Plot */
PROC SGPLOT DATA = work.monthly_sales;
scatter X= date Y = weekly_sales/group= store groupdisplay=cluster
markerattrs=(symbol=circlefilled size=15);
RUN;
/* 4) Composition */
/* a) Stacked Column Chart: */
PROC SGPLOT DATA= work.monthly_sales;
title 'Weekly_sales by Store and date';
vbar date / response= weekly_sales group= store stat=percent DATAlabel;
xaxis display=(nolabel);
yaxis grid label='Weekly_sales';
RUN;
/* Semester view of sales in units */
/* Convert walmart DATA into timeseries DATA */
PROC TIMESERIES DATA= work.walmart

```

```

out= semester_sales;
by store;
id date interval= semiyear accumulate= total;
var weekly_sales holiday_flag temperature fuel_price cpi unemployment;
RUN;
PROC PRINT DATA= work.semester_sales;
RUN;
/* Giving insights */
/* Checking the correlation */
PROC CORR DATA= work.semester_sales;
RUN;
/* 1. Doing Comparison */
/* a) Bar Chart */
PROC SGPLOT DATA = work.semester_sales;
hbar store/response = weekly_sales stat= sum
DATAlabel DATAlabelattrs=(weight=bold);
title 'Total Views by Store';
RUN;
/* b) Clustered Bar Chart / Column Chart */
DATA date;
set work.semester_sales;
month = month(Date);
month_name=PUT(Date,monname.);
put month_name= @;
RUN;
PROC PRINT DATA= date1;
RUN;
PROC SGPLOT DATA= date;
vbar store/ response= weekly_sales group=month_name groupdisplay=cluster
DATAlabel DATAlabelattrs = (weight = bold) DATAskin=gloss; yaxis grid;
title 'Total View by monthly wise';
RUN;
/* 2) Studying relationship */
/* a) Bubble Chart */
PROC SGPLOT DATA = work.semester_sales;
bubble X=weekly_sales Y=store size= weekly_sales
/ fillattrs=(color = teal) DATAlabel = store;
RUN;
/* b) Scatter Plot for Relationship */
PROC SGPLOT DATA = work.semester_sales;
title 'Relationship of Store with Weekly_sales';
scatter X= weekly_sales Y = store/
markerattrs=(symbol=circlefilled size=15);
RUN;
/* 3. Studying Distribution */
/* a) Histogram */
PROC SGPLOT DATA = work.semester_sales;
    histogram weekly_sales/ fillattrs=(color = steel) scale = proportion;
    density weekly_sales;
RUN;
/* b) Scatter Plot */
PROC SGPLOT DATA = work.semester_sales;
    scatter X= date Y = weekly_sales/group= store groupdisplay=cluster
    markerattrs=(symbol=circlefilled size=15);
RUN;
/* 4) Composition */
/* a) Stacked Column Chart: */
PROC SGPLOT DATA= work.semester_sales;
    title 'Weekly_sales by Store and date';
    vbar date / response= weekly_sales group= store stat=percent DATAlabel;
    xaxis display=(nolabel);
    yaxis grid label='Weekly_sales';
RUN;
/* For Store 1 - Build prediction models to forecast demand */
/* Store-1 DATA */
DATA store1;
    set work.walmart;
    where store = 1;
RUN;
PROC PRINT DATA= store1;
RUN;
/* Convert store-1 DATA into timeseries DATA */
PROC TIMESERIES DATA= store1
    out= store_1;
    by store;

```

```
id date interval= month accumulate= total;
var weekly_sales holiday_flag temperature fuel_price cpi unemployment;

RUN;
PROC PRINT DATA= work.store_1;
RUN;
/* Build Model */
ods noproctitle;
ods graphics / imagemap=on;
PROC ARIMA DATA=WORK.STORE_1 plots
  (only)=(series(corr crosscorr) residual(corr normal)
  forecast(forecast forecastonly) );
  identify var=Weekly_Sales(1);
  estimate p=(1 2 3) q=(1) method=ML;
  forecast lead=4 back=0 alpha=0.05;
  outlier;
```