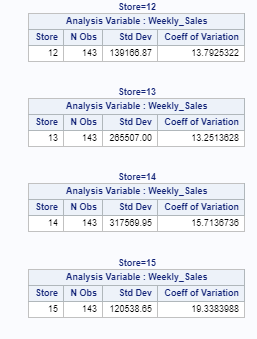
**Project 2: Walmart Sales**

Which store has maximum sales?



Store 20 has the maximum total sales than any other store.

Which store has maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation?



The store with the maximum standard deviation is store 14, having a standard deviation of 317569.95 and coefficient of variation is 15.7136783.

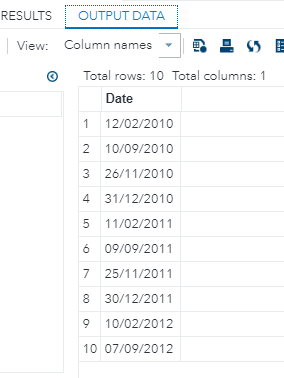
Which store/s has good quarterly growth rate in Q3’2012



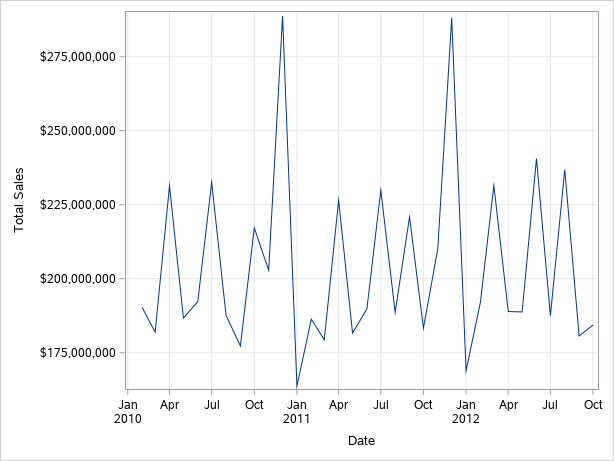
Store 4 had the maximum total sales in q3 2012.

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?

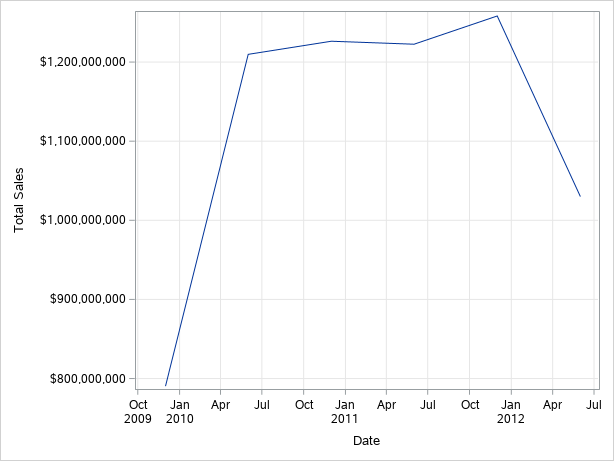
The table below shows the holiday dates that impact sales mostly, on these holidays the mean sales of store is greater than the mean sales of all stores across non holiday season.



Provide a monthly and semester view of sales in units and give insights?

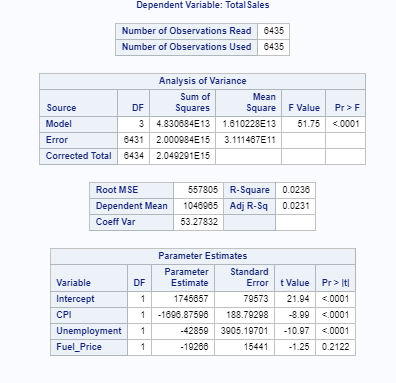


The sales for the store seem to exhibit a cyclic patterns. Sales are always lowest in January and rise from October to December to drop again in January.



**Statistical Model: multiple linear regression:**

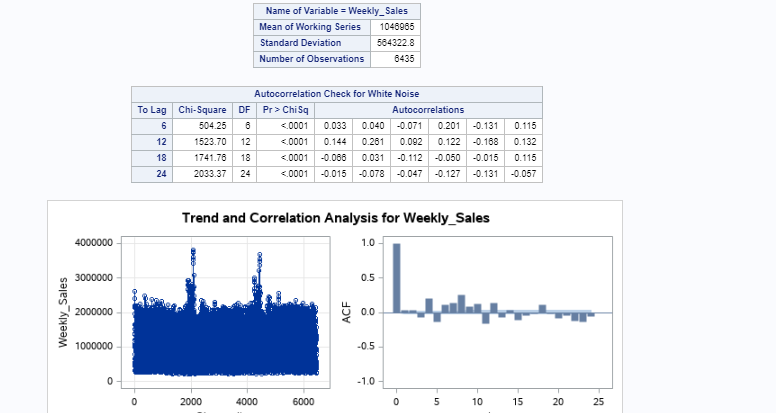
Hypothesize if CPI, unemployment, and fuel price have any impact on sales.



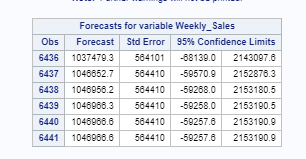
For all the independent variables the p values generated by the model (<0.0001) were less than alpha (0.05). this means that the null hypothesis of no effect is rejected. Unemployment, CPI and Fuel price all have an effect on store sales. R square is 0.0236, this suggest that this effect is not very strong.

Time series forecasting model:

The probablity of white noise is rejected as the mean of the working series is not equal to zero, this means that this data is not fit for timeseries analysisi. The series is not stationary because the correlation trends are not similar.



Sales forcast for the next six months:



**Code:**

FILENAME REFFILE '/folders/myshortcuts/MyFolders/projects/Project2/Walmart\_Store\_sales.csv';  
  
PROC IMPORT DATAFILE=REFFILE  
DBMS=CSV  
OUT=WORK.walmart;  
GETNAMES=YES;  
RUN;  
  
data walmart2;  
set walmart;  
proc sort data=walmart;  
by store weekly\_sales;  
run;  
  
proc sql;  
create table wal as  
select store,weekly\_sales  
from walmart;  
quit;  
run;  
  
/\* question1 \*/  
proc sql;  
create table new222 as  
 select store, sum(weekly\_sales) as TotalSales  
 from walmart  
 where store is not missing  
 group by store;  
 run;  
   
proc sql;  
select store, totalSales  
from new222  
having totalsales=max(totalsales);  
run;  
  
/\* question2 \*/  
  
proc means stddev cv data=walmart;  
 class store;  
 var weekly\_sales;  
 by store;  
run;  
  
/\* question 3 q3 \*/  
  
\*RETURN YEAR;  
  
DATA q3;  
set walmart;  
new\_year=year(date);  
run;  
  
proc sql;  
create table QQ3 as  
select store,weekly\_sales,New\_year   
from q3  
where new\_year=2012;  
quit;  
run;  
  
data qtrly;  
 set walmart;  
 where date >= '1jul2012'd &  
 date < '30sep2012'd ;  
 keep store weekly\_sales date ;  
run;  
  
  
proc sql;  
create table question3 as  
 select store, sum(weekly\_sales) as TotalSales  
 from qtrly  
 where store is not missing  
 group by store;  
 run;  
  
proc sql;  
select store, totalSales  
from question3  
having totalsales=max(totalsales);  
run;  
  
/\* question 4 \*/  
/\* to get mean sales when ther are no holidays for all stores \*/  
  
proc sql;  
create table No\_Holiday as  
select store,weekly\_sales,holiday\_flag  
from walmart  
where holiday\_flag=0;  
quit;  
run;  
  
proc means data=no\_holiday;  
run;  
  
/\* the mean sale for all stores in No holiday season is 1041256.38 \*/  
/\* to find holiday dates where sales were greater than the mean sales of non holiday day \*/  
  
proc sql;  
create table hod as  
select unique(date) from walmart   
having holiday\_flag=1 and mean(weekly\_sales)>1041256.38  
order by date  
;  
quit;  
run;  
  
  
/\* question 5 \*/  
   
/\* sorting data for timeseries step \*/  
data newWalmat;  
set walmart;  
proc sort data=walmart;  
by Date;  
run;  
  
data NEW3;  
SET NEWWALMAT;  
format weekly\_sales dollar9.;  
RUN;  
  
  
/\* perfoming time series at a monthly granularity level \*/  
  
proc timeseries data=NEW3  
out=Timedset;  
Id date interval=month accumulate=total;  
var weekly\_sales;  
run;  
  
/\* ploting monthly sales \*/  
ods graphics / reset width=6.4in height=4.8in imagemap;  
  
proc sort data=WORK.TIMEDSET out=\_SeriesPlotTaskData;  
by Date;  
run;  
  
proc sgplot data=\_SeriesPlotTaskData;  
series x=Date y=Weekly\_Sales /;  
xaxis grid label="Date";  
yaxis grid label="Total Sales";  
run;  
  
ods graphics / reset;  
  
proc datasets library=WORK noprint;  
delete \_SeriesPlotTaskData;  
run;  
  
/\* perfoming time series at a semester granularity level \*/  
  
proc timeseries data=NEW3  
out=Timsemester;  
Id date interval=semiyear.6 accumulate=total;  
var weekly\_sales;  
run;  
  
/\* to plot semester sales \*/  
  
ods graphics / reset width=6.4in height=4.8in imagemap;  
  
proc sort data=WORK.TIMSEMESTER out=\_SeriesPlotTaskData;  
by Date;  
run;  
  
proc sgplot data=\_SeriesPlotTaskData;  
series x=Date y=Weekly\_Sales /;  
xaxis grid label="Date";  
yaxis grid label="Total Sales";  
run;  
  
ods graphics / reset;  
  
proc datasets library=WORK noprint;  
delete \_SeriesPlotTaskData;  
run;  
  
/\* Models Models Models Models Models Models Models Models Models Models Models Models Models \*/  
proc sort data= walmart;  
by date;  
run;  
  
proc sql;  
create table regression as  
 select store, date,cpi,unemployment,fuel\_price,sum(weekly\_sales) as TotalSales  
 from walmart  
 where store is not missing  
 group by store,date;  
 run;  
  
proc reg data=regression PLOTS(MAXPOINTS=10000);  
model totalsales=cpi unemployment fuel\_price;  
run;  
  
/\* Checking for white noise \*/  
  
proc arima data= walmart;  
identify var=weekly\_sales nlag=24;  
run;  
  
/\* differentiation \*/  
  
proc arima data= walmart;  
identify var=weekly\_sales nlag=24;  
estimate p=1;  
forecast lead=6 interval=month id=date out=walmart20;  
run;