DATA SCIENCE WITH SAS

PROJECT 2:

RETAIL ANALYSIS WITH WALMART DATA

AUTHOR: Manasa Devikoppa Basavarajappa

SUBMISSION DATE: 04 April, 2021

TABLE OF INDEX

CHAPTER 1: PROBLEM STATEMENT

DESCRIPTION

One of the leading retail stores in the US, Walmart, would like to predict the sales and demand accurately. There are certain events and holidays which impact sales on each day. There are sales data available for 45 stores of Walmart. The business is facing a challenge due to unforeseen demands and runs out of stock some times, due to the inappropriate machine learning algorithm. An ideal ML algorithm will predict demand at different points of time covering seasonality and ingest factors like economic conditions including CPI, Unemployment Index, etc.

Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of all, which are the Super Bowl, Labour Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks. Part of the challenge presented by this competition is modeling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data. Historical sales data for 45 Walmart stores located in different regions are available.

Dataset Description

This is the historical data which covers sales from 2010-02-05 to 2012-11-01, in the file Walmart_Store_sales. Within this file you will find the following fields:

- Store - the store number
- Date - the week of sales
- Weekly_Sales - sales for the given store
- Holiday_Flag - whether the week is a special holiday week 1 – Holiday week 0 – Non-holiday week
- Temperature - Temperature on the day of sale
- Fuel_Price - Cost of fuel in the region
- CPI – Prevailing consumer price index
- Unemployment - Prevailing unemployment rate

Holiday Events

Super Bowl: 12-Feb-10, 11-Feb-11, 10-Feb-12, 8-Feb-13
Labour Day: 10-Sep-10, 9-Sep-11, 7-Sep-12, 6-Sep-13
Thanksgiving: 26-Nov-10, 25-Nov-11, 23-Nov-12, 29-Nov-13
Christmas: 31-Dec-10, 30-Dec-11, 28-Dec-12, 27-Dec-13

Analysis Tasks

Basic Statistics tasks

- Which store has maximum sales
- Which store has maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation
- Which store/s has good quarterly growth rate 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
- Provide a monthly and semester view of sales in units and give insights

Statistical Model

For Store 1 – Build prediction models to forecast demand

- Linear Regression – Utilize variables like date and restructure dates as 1 for 5 Feb 2010(starting from the earliest date in order). Hypothesize if CPI, unemployment, and fuel price have any impact on sales.
- Time series forecasting model –
    - Hypothesize if the data is fit for time series analysis – check for white noise probability test
    - Make adjustments in historical data for events like holidays, if applicable
    - Build ARIMA model to forecast 6 months i.e., input utilize only till April 2012.

Predict next 6 months i.e., June to Oct 2010. Check for MAPE.

Select the model which gives best accuracy.

Click here to download the datasets

CHAPTER 2: WRITEUP

The project is related to the data analysis of the Historical sales data for 45 Walmart stores located in different regions including holiday events.

The goal of the project is to predict the sales and demand at different points of time covering seasonality and ingest factors like economic conditions including CPI, Unemployment Index, etc.

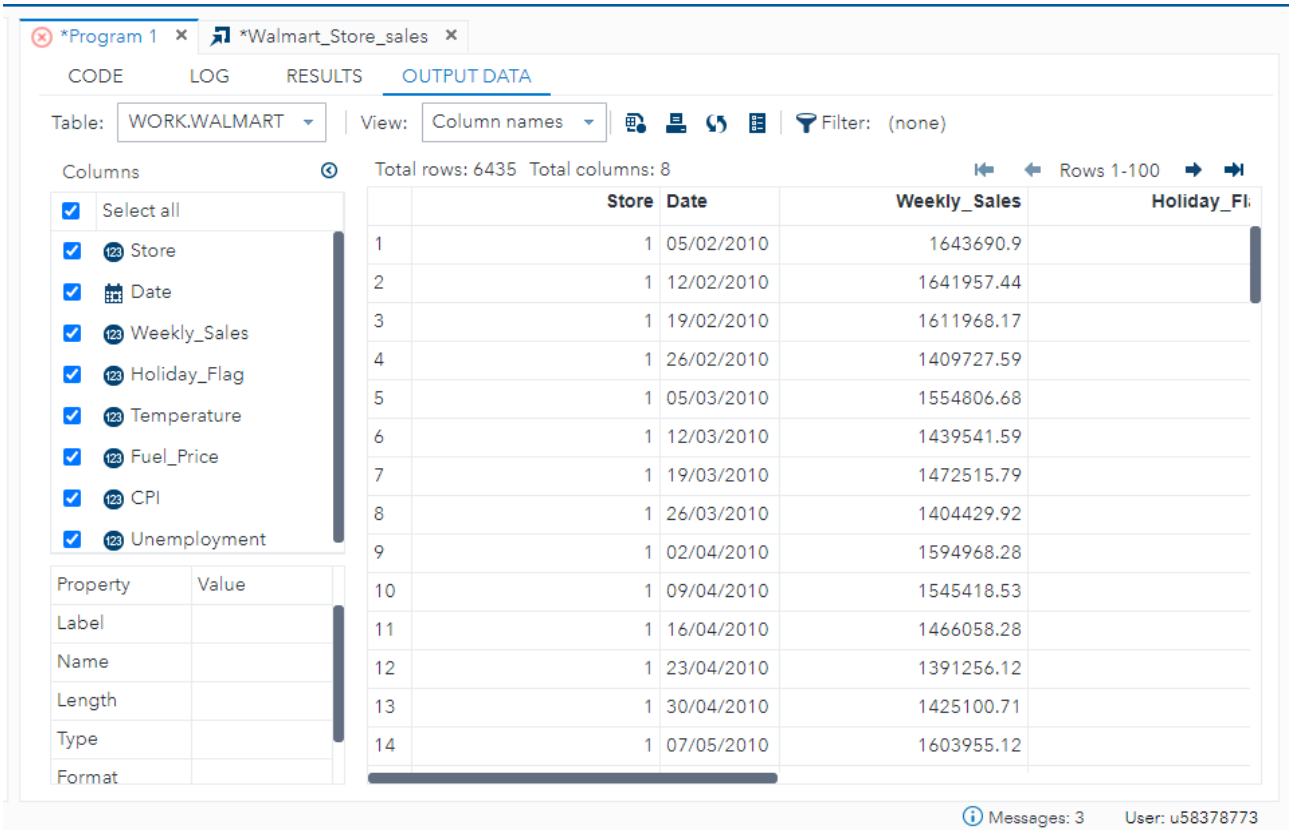
Part of the challenge presented by this competition is modeling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data.

The project uses SAS Studio to run the code and uses SAS procedures such as MEANS, SUMMARY, SORT, SQL, CORR, SGPLOT, TIMESERIES, ARIMA to achieve the results.

CHAPTER 3: SOURCE CODE WITH OUTPUT

```
/* Import the Dataset */

FILENAME REFFILE '/home/u58378773/SAS Project/Walmart_Store_sales.csv';

PROC IMPORT DATAFILE=REFFILE
        DBMS=CSV
        OUT=WORK.walmart;
        replace;
        GETNAMES=YES;
RUN;
```

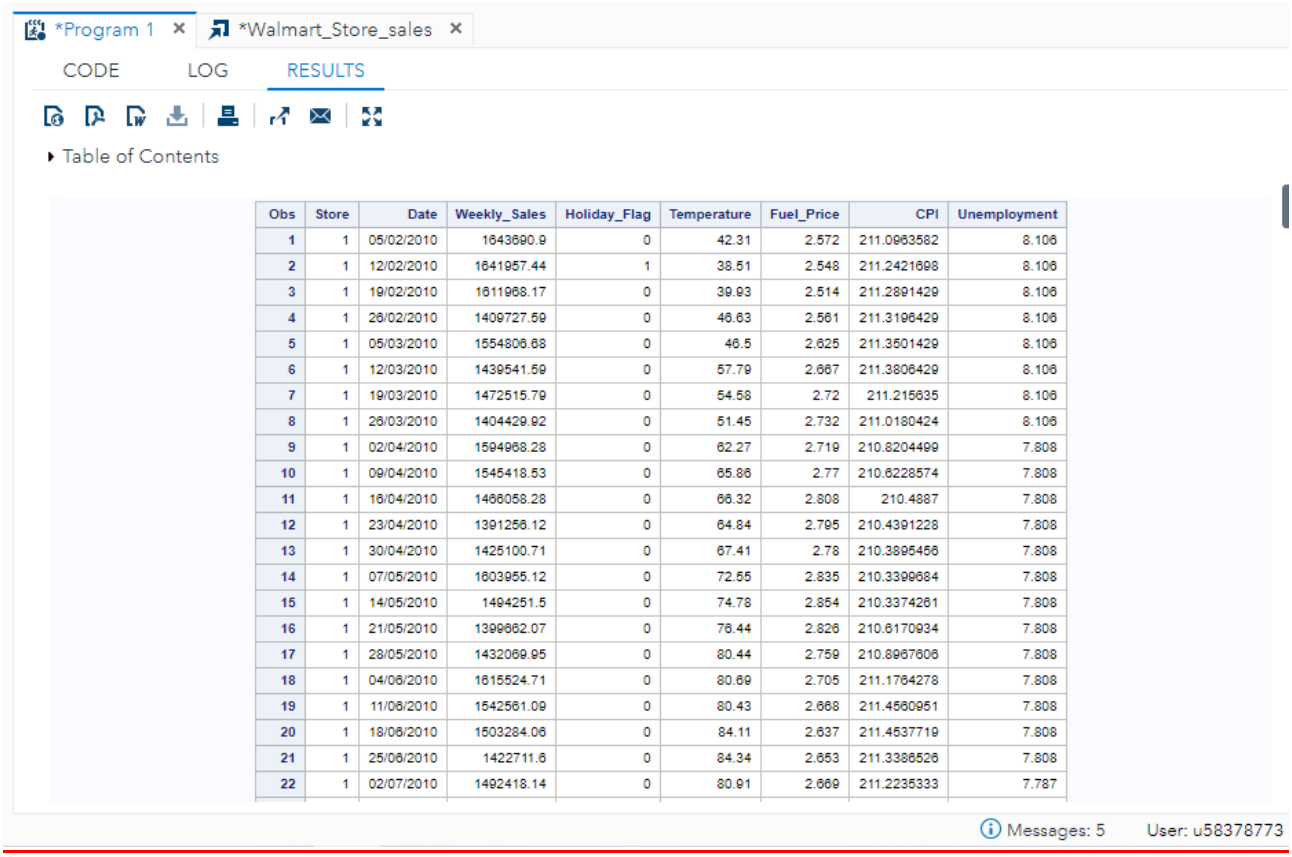
Output:



	Store	Date	Weekly_Sales	Holiday_Fl
1	1	05/02/2010	1643690.9	
2	1	12/02/2010	1641957.44	
3	1	19/02/2010	1611968.17	
4	1	26/02/2010	1409727.59	
5	1	05/03/2010	1554806.68	
6	1	12/03/2010	1439541.59	
7	1	19/03/2010	1472515.79	
8	1	26/03/2010	1404429.92	
9	1	02/04/2010	1594968.28	
10	1	09/04/2010	1545418.53	
11	1	16/04/2010	1466058.28	
12	1	23/04/2010	1391256.12	
13	1	30/04/2010	1425100.71	
14	1	07/05/2010	1603955.12	

```
PROC PRINT DATA=walmart;
RUN;
```

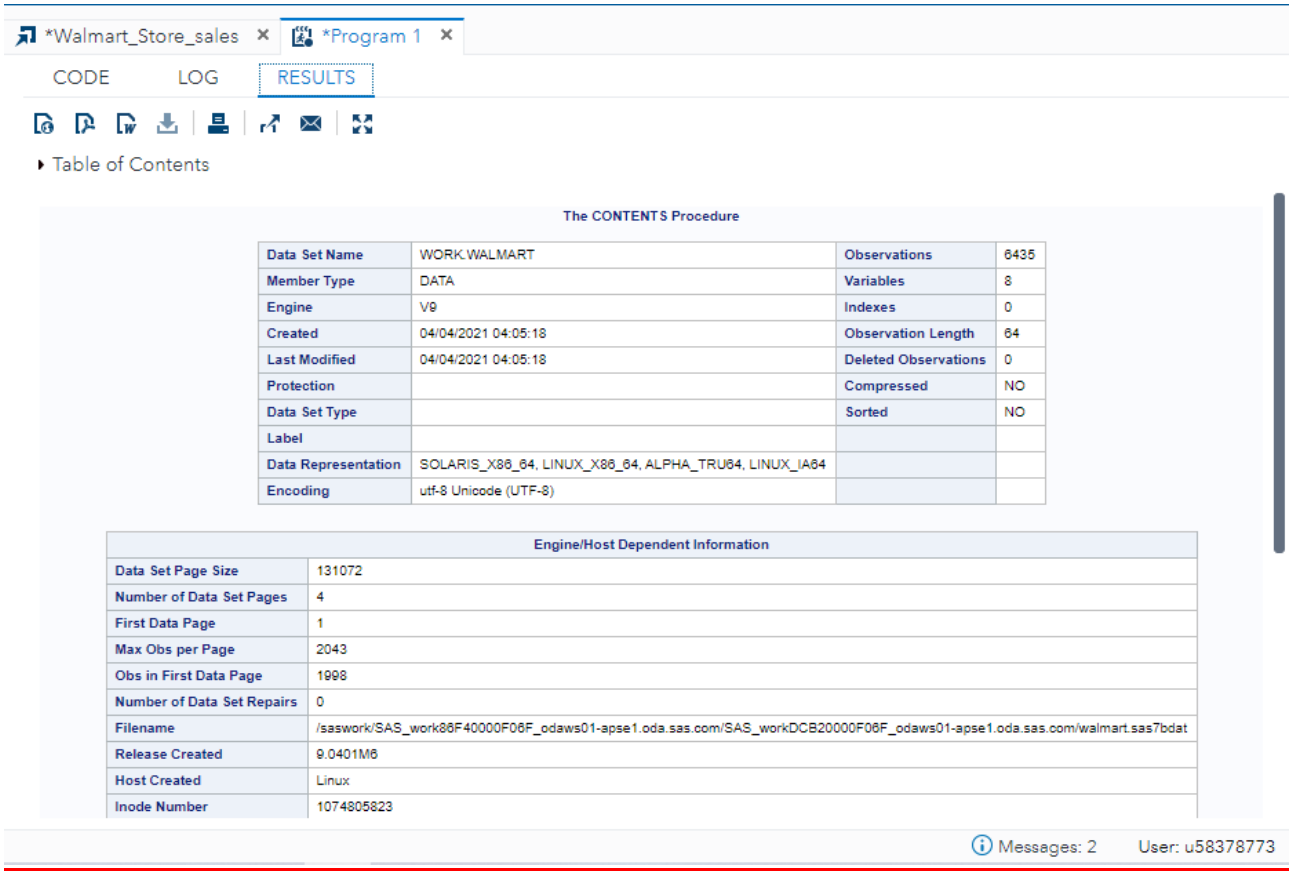
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	05/02/2010	1643690.9	0	42.31	2.572	211.0963582	8.106
2	1	12/02/2010	1641957.44	1	38.51	2.548	211.2421698	8.106
3	1	19/02/2010	1611968.17	0	39.93	2.514	211.2891429	8.106
4	1	26/02/2010	1409727.59	0	46.63	2.561	211.3196429	8.106
5	1	05/03/2010	1554806.68	0	46.5	2.625	211.3501429	8.106
6	1	12/03/2010	1439541.59	0	57.79	2.667	211.3806429	8.106
7	1	19/03/2010	1472515.79	0	54.58	2.72	211.215635	8.106
8	1	26/03/2010	1404429.92	0	51.45	2.732	211.0180424	8.106
9	1	02/04/2010	1594968.28	0	62.27	2.719	210.8204499	7.808
10	1	09/04/2010	1545418.53	0	65.86	2.77	210.6228574	7.808
11	1	16/04/2010	1466058.28	0	66.32	2.808	210.4887	7.808
12	1	23/04/2010	1391256.12	0	64.84	2.795	210.4391228	7.808
13	1	30/04/2010	1425100.71	0	67.41	2.78	210.3895456	7.808
14	1	07/05/2010	1603955.12	0	72.55	2.835	210.3399684	7.808
15	1	14/05/2010	1494251.5	0	74.78	2.854	210.3374261	7.808
16	1	21/05/2010	1399662.07	0	76.44	2.826	210.6170934	7.808
17	1	28/05/2010	1432069.95	0	80.44	2.759	210.8967606	7.808
18	1	04/06/2010	1615524.71	0	80.69	2.705	211.1764278	7.808
19	1	11/06/2010	1542561.09	0	80.43	2.668	211.4560951	7.808
20	1	18/06/2010	1503284.06	0	84.11	2.637	211.4537719	7.808
21	1	25/06/2010	1422711.6	0	84.34	2.653	211.3386526	7.808
22	1	02/07/2010	1492418.14	0	80.91	2.669	211.2235333	7.787

```
/* Check the content of the data */

PROC CONTENTS DATA=WORK.walmart;
RUN;
```

Output:



The CONTENTS Procedure

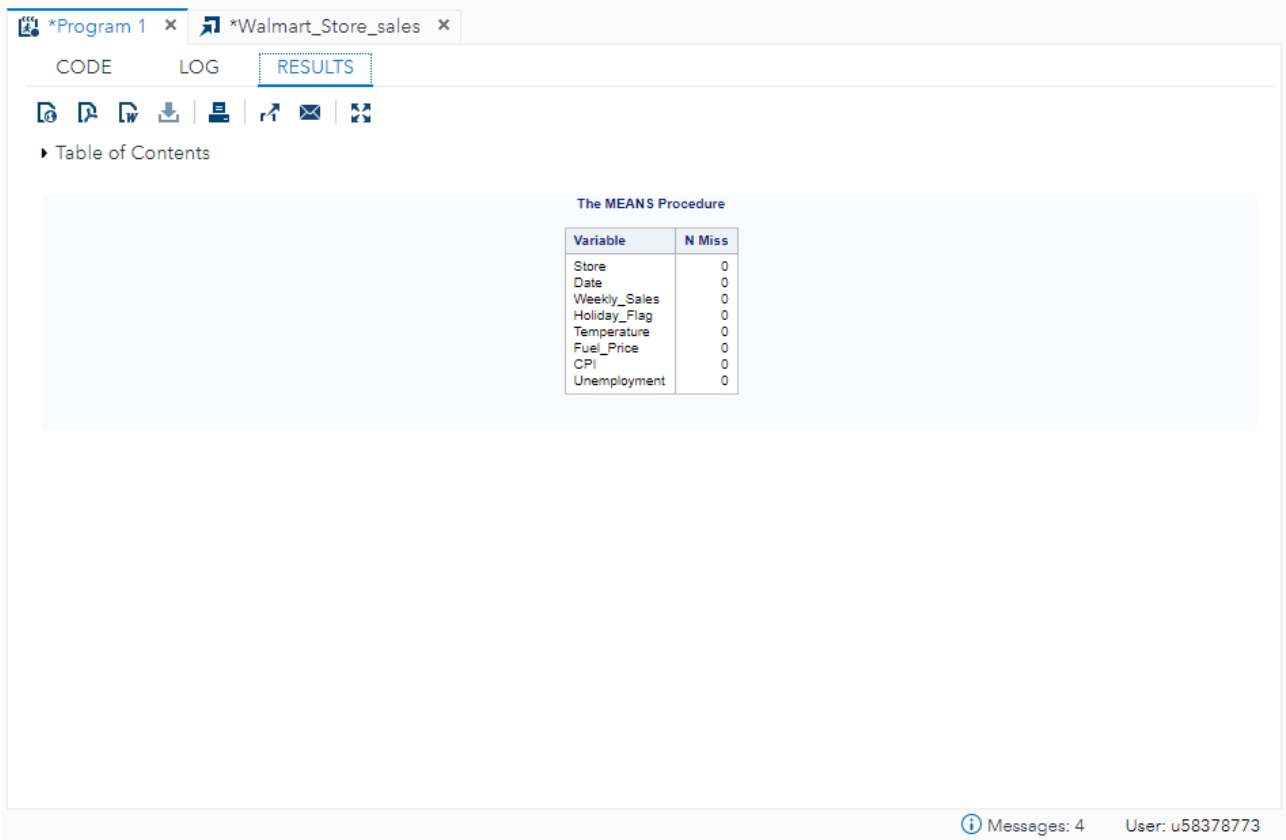
Data Set Name	WORK.WALMART	Observations	6435
Member Type	DATA	Variables	8
Engine	V9	Indexes	0
Created	04/04/2021 04:05:18	Observation Length	64
Last Modified	04/04/2021 04:05:18	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	4
First Data Page	1
Max Obs per Page	2043
Obs in First Data Page	1998
Number of Data Set Repairs	0
Filename	/saswork/SAS_work86F40000F06F_odaws01-apse1.oda.sas.com/SAS_workDCB20000F06F_odaws01-apse1.oda.sas.com/walmart.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	1074805823

Messages: 2 User: u58378773

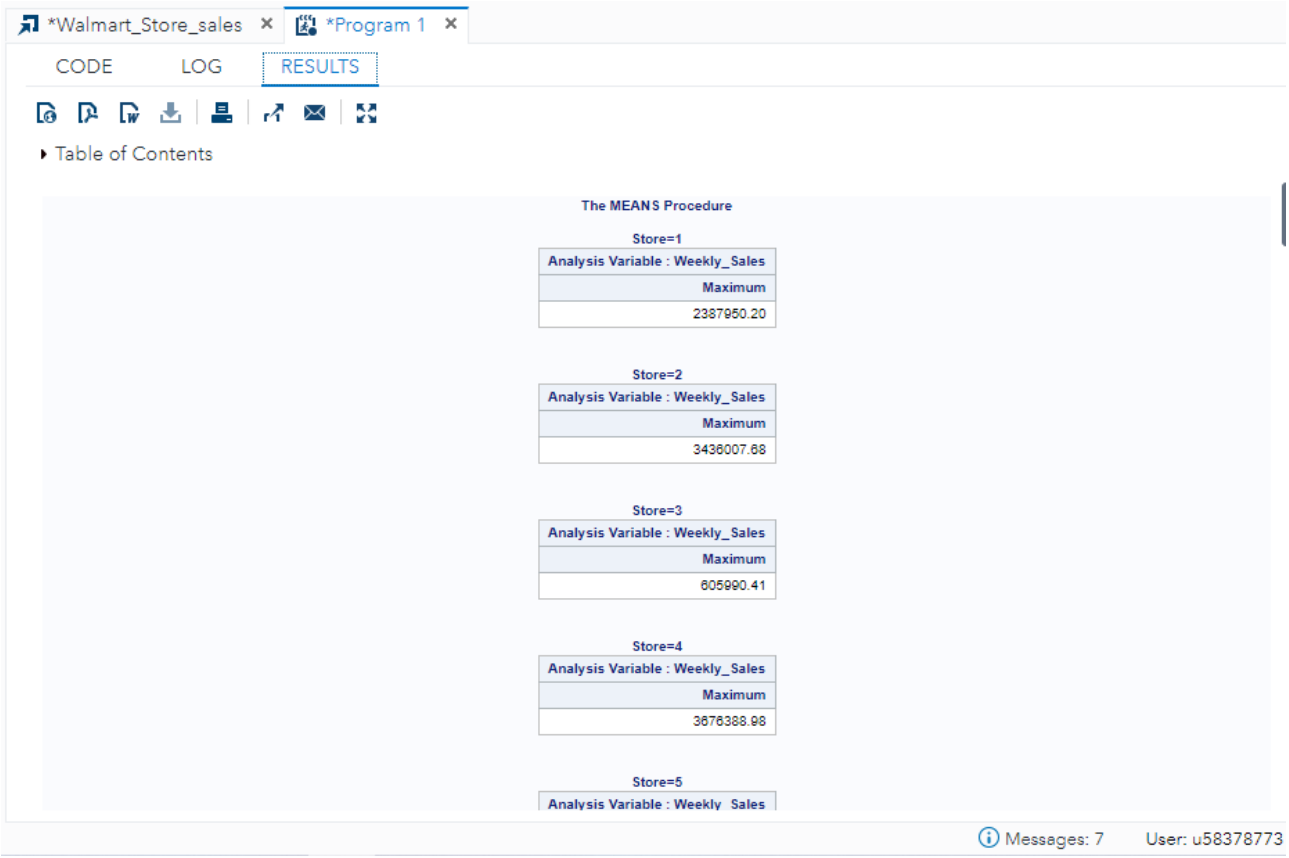
```
/* Check the missing value */

PROC MEANS DATA=work.walmart nmiss;
RUN;
```

Output:



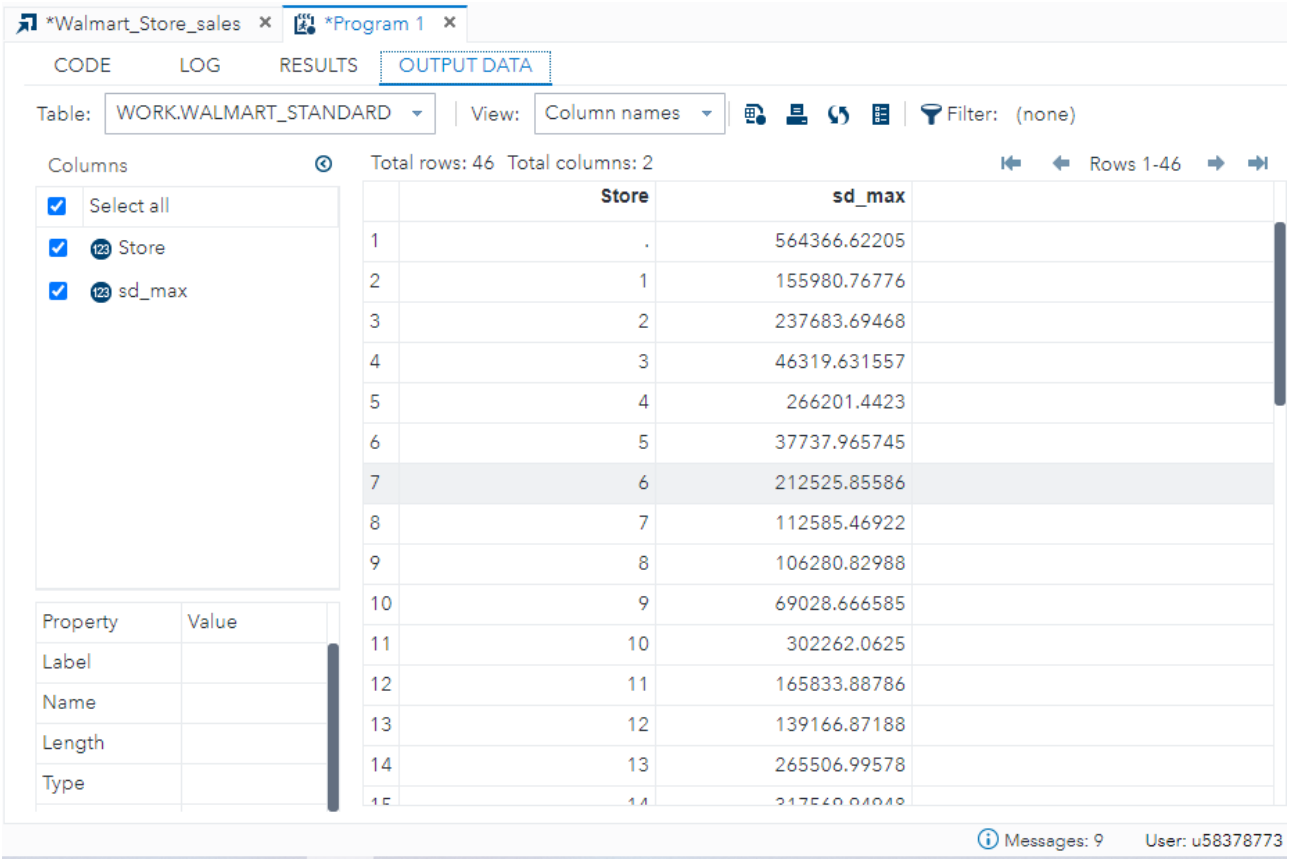
```
/* Which store has maximum sales */

PROC MEANS DATA=work.walmart  max;
        by store;
        var weekly_sales;
RUN;
```

Output:



```
/* Which store has maximum standard deviation */

PROC SUMMARY DATA=work.walmart;
        class store;
        output  out=  walmart_standard(drop=  _type_  _freq_)
        std(weekly_sales)=sd_max;
RUN;
```

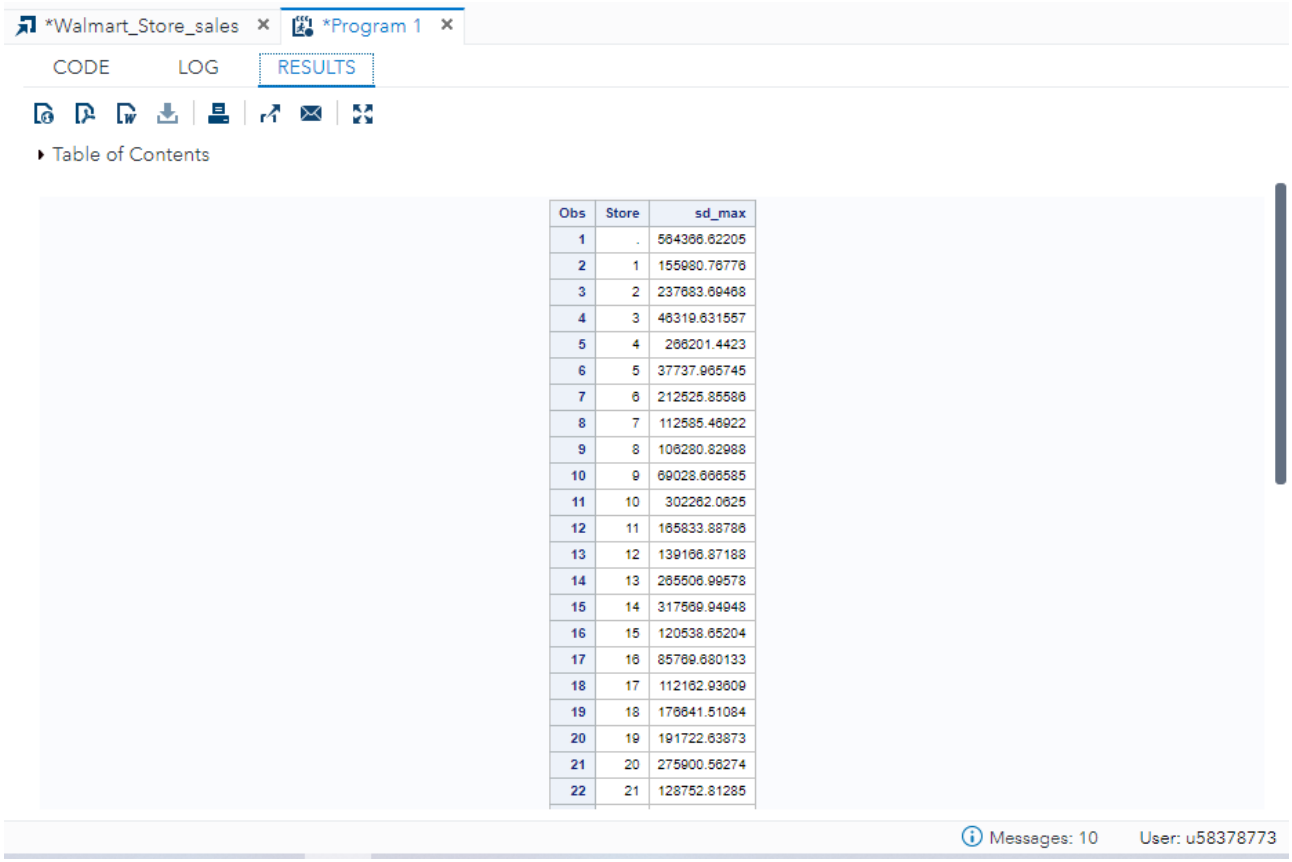
Output:



	Store	sd_max
1	.	564366.62205
2	1	155980.76776
3	2	237683.69468
4	3	46319.631557
5	4	266201.4423
6	5	37737.965745
7	6	212525.85586
8	7	112585.46922
9	8	106280.82988
10	9	69028.666585
11	10	302262.0625
12	11	165833.88786
13	12	139166.87188
14	13	265506.99578

```
PROC PRINT DATA=work.walmart_standard;
RUN;
```

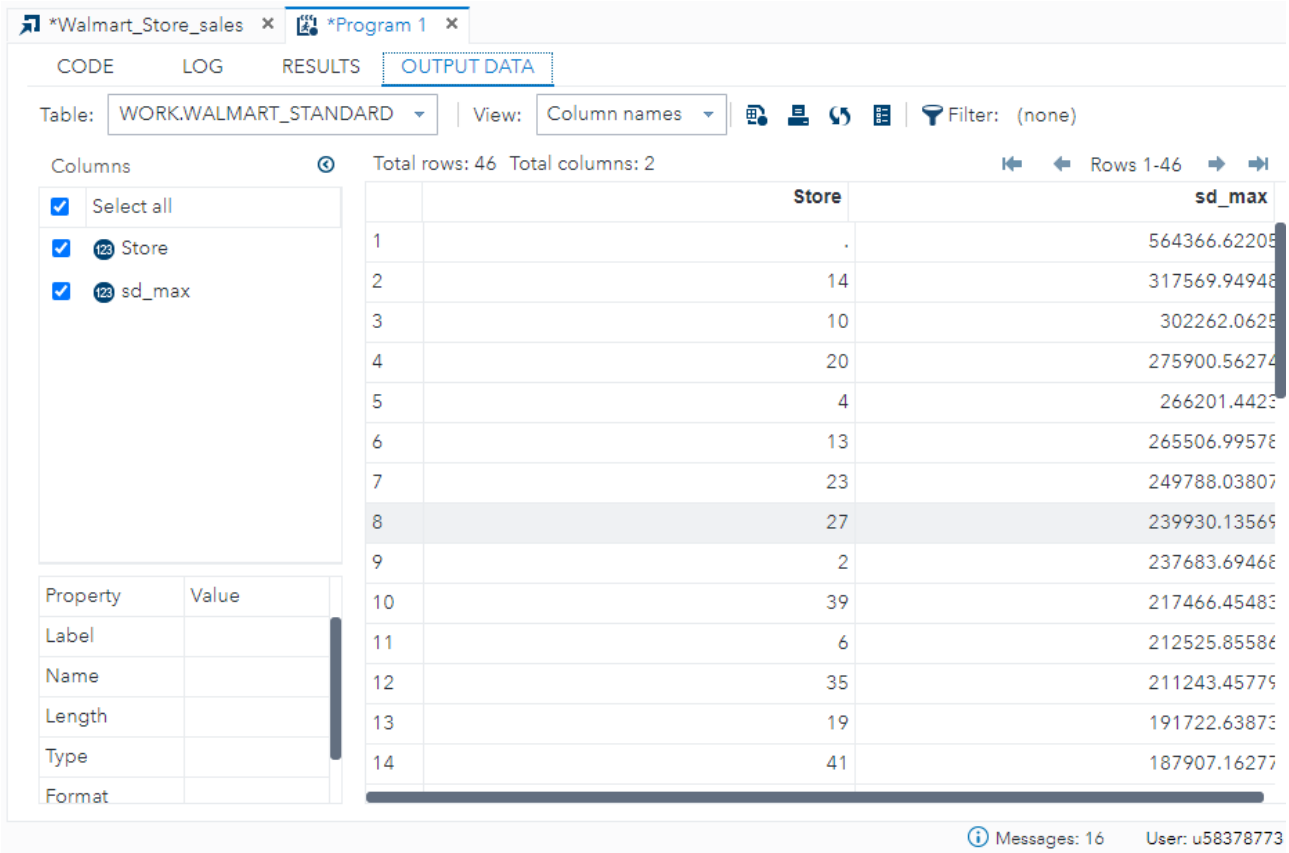
Output:



Obs	Store	sd_max
1	.	564366.62205
2	1	155980.76776
3	2	237683.69468
4	3	46319.631557
5	4	266201.4423
6	5	37737.965745
7	6	212525.85586
8	7	112585.46922
9	8	106280.82988
10	9	69028.666585
11	10	302262.0625
12	11	165833.88786
13	12	139166.87188
14	13	265506.99578
15	14	317569.94948
16	15	120538.65204
17	16	85769.680133
18	17	112162.93609
19	18	176641.51084
20	19	191722.63873
21	20	275900.56274
22	21	128752.81285

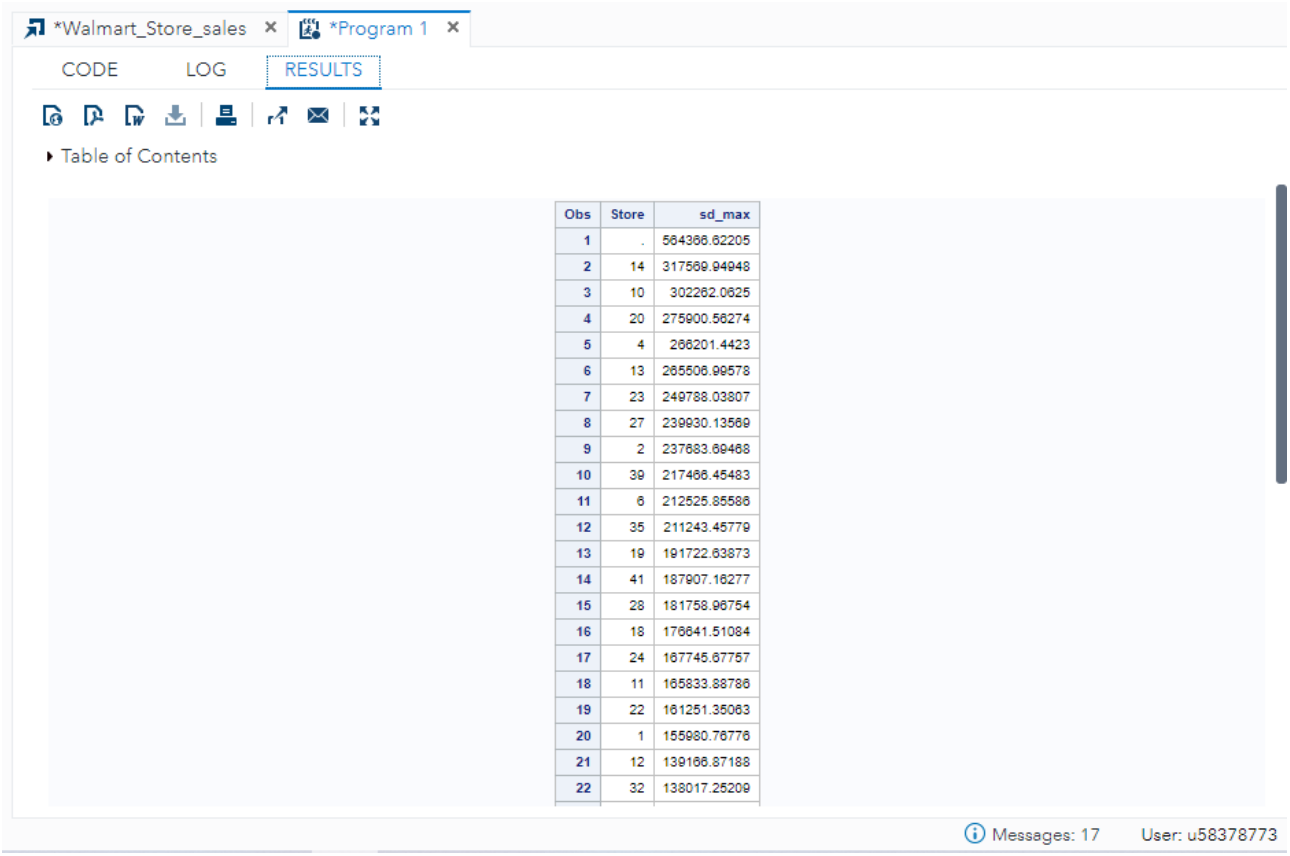
```
PROC SORT DATA=work.walmart_standard;
                by  descending  sd_max;
RUN;
```

Output:



```
PROC PRINT DATA=work.walmart_standard;
RUN;
```

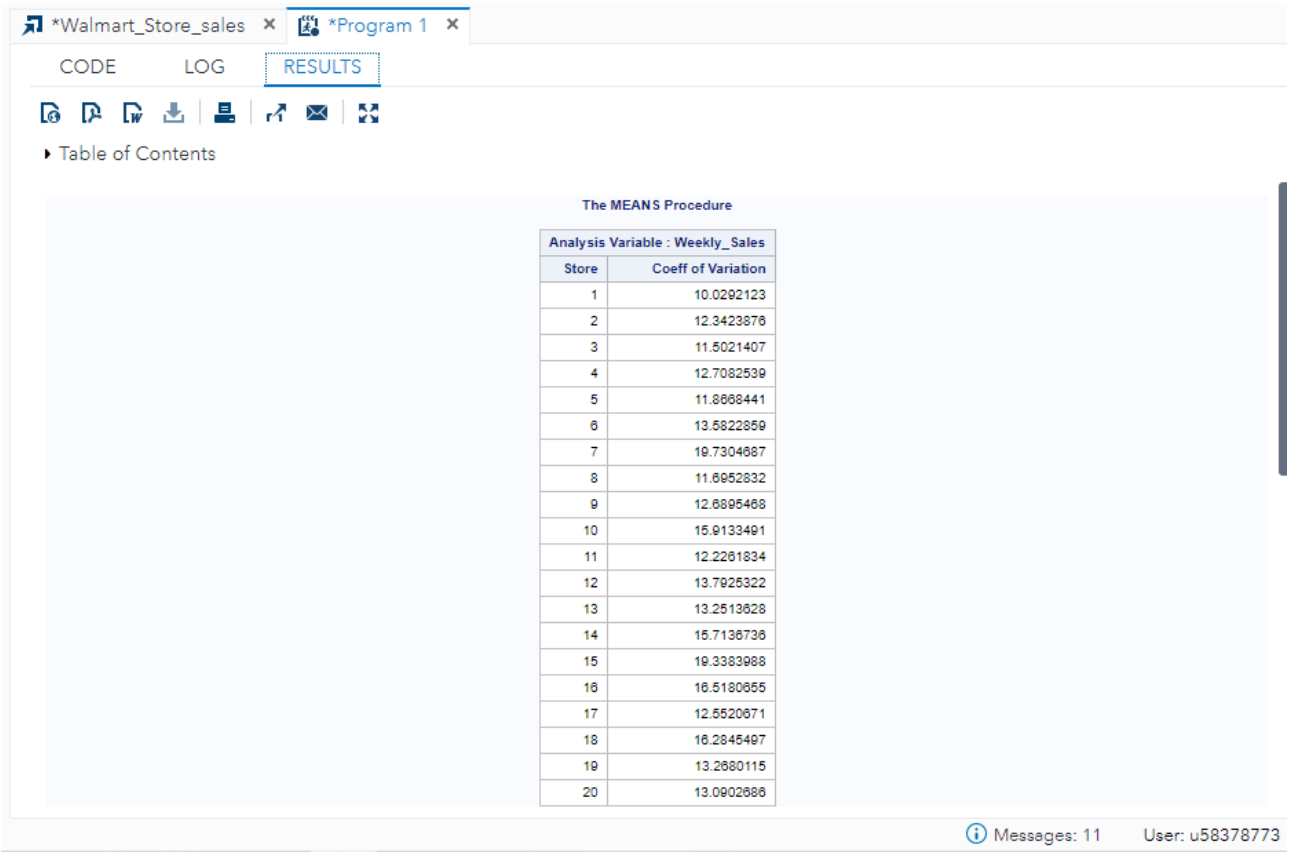
Output:



Obs	Store	sd_max
1	.	564366.62205
2	14	317569.94948
3	10	302262.0625
4	20	275900.56274
5	4	266201.4423
6	13	265506.99578
7	23	249788.03807
8	27	239930.13569
9	2	237683.69468
10	39	217466.45483
11	6	212525.85586
12	35	211243.45779
13	19	191722.63873
14	41	187907.16277
15	28	181758.96754
16	18	176641.51084
17	24	167745.67757
18	11	165833.88786
19	22	161251.35063
20	1	155980.76776
21	12	139166.87188
22	32	138017.25209

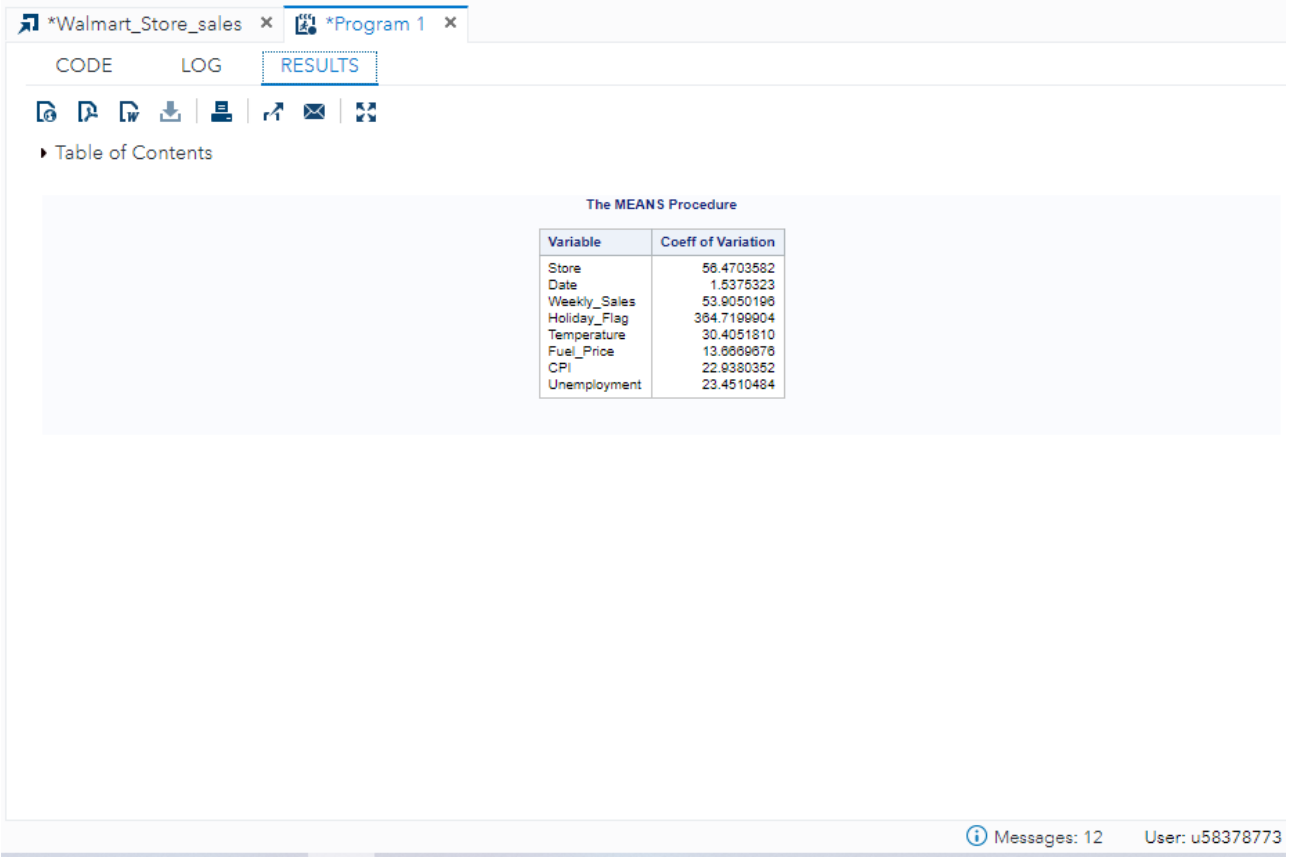
```
/* Find out the coefficient of mean to standard deviation */

PROC MEANS DATA=work.walmart nonobs cv;
        class store;
        var weekly_sales;
RUN;
```

Output:



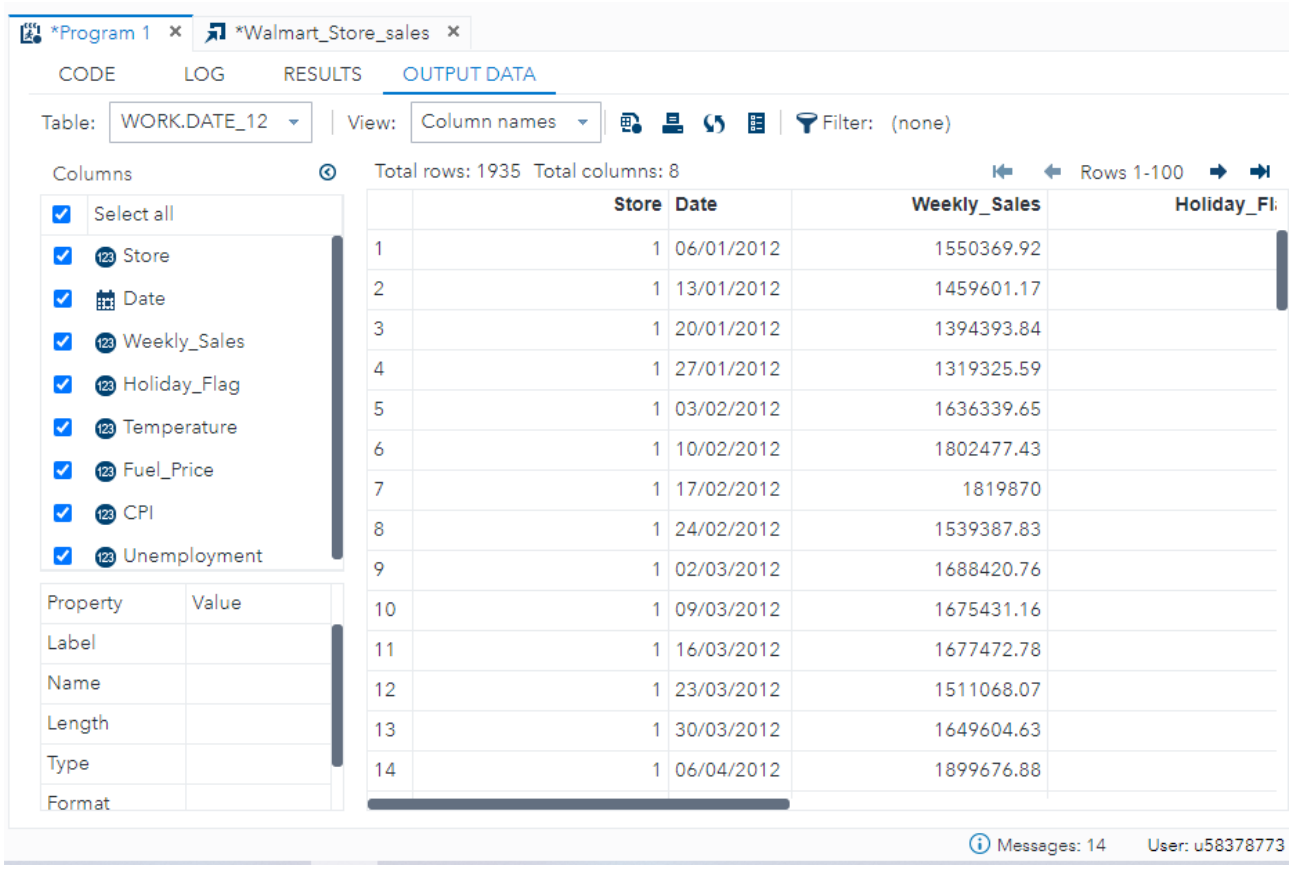
```
/* OR */

PROC MEANS DATA=work.walmart nonobs cv;
RUN;
```

Output:



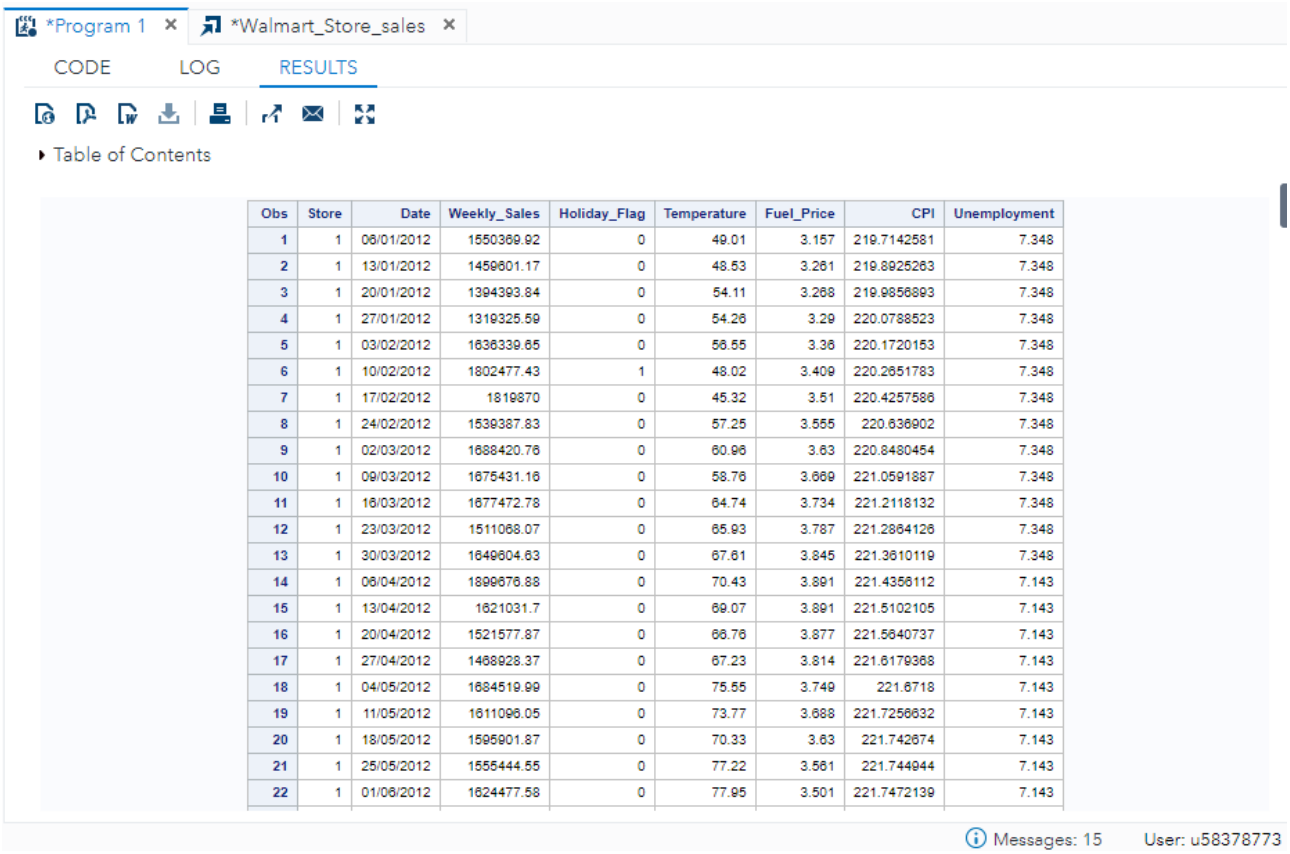
```
/* Which store/s has good quarterly growth rate in Q3'2012 */

/* Filter year(2012) */

DATA date_12;
set work.walmart;

where  year(date)=2012;
RUN;
```

Output:



```
PROC PRINT DATA= date_12;
RUN;
```

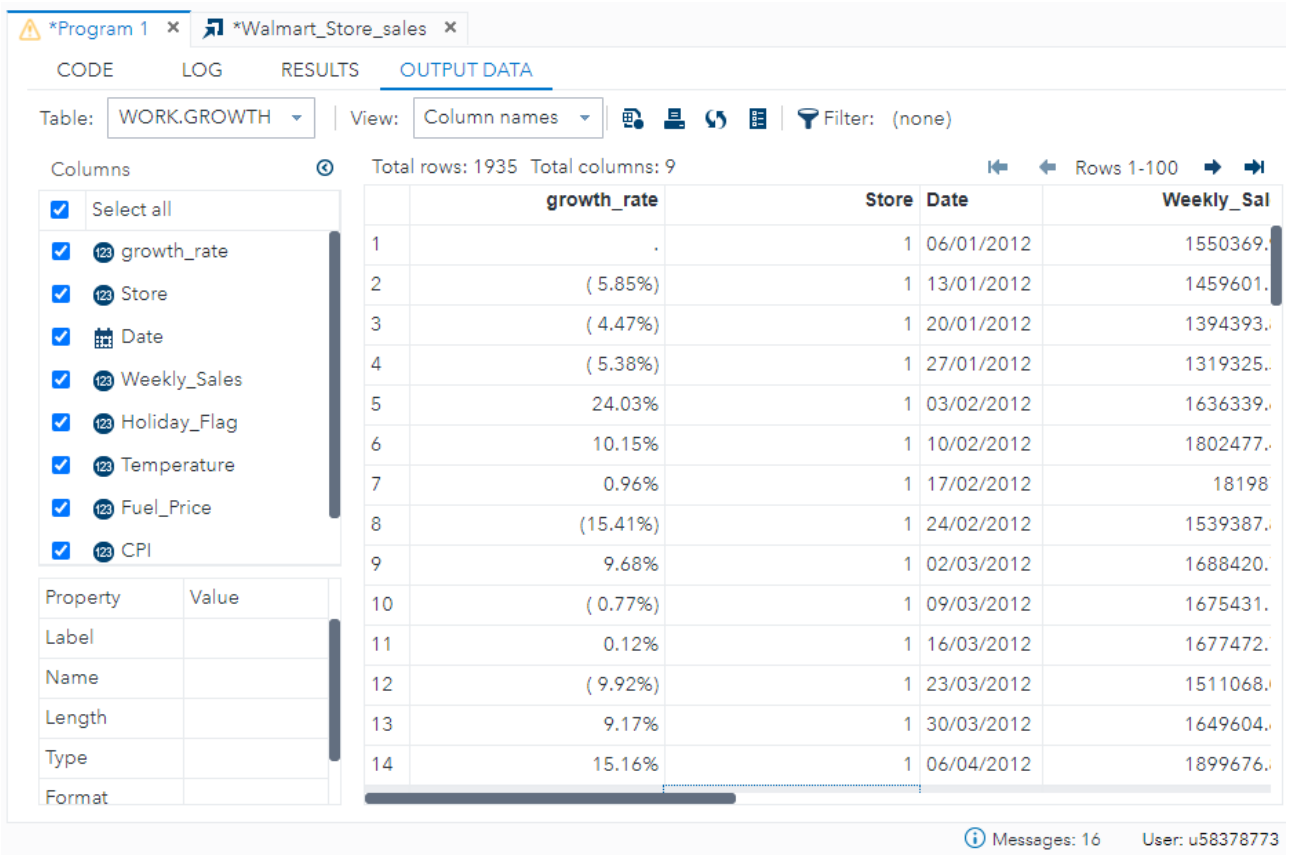
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	05/01/2012	1550369.92	0	49.01	3.157	219.7142581	7.348
2	1	13/01/2012	1459601.17	0	48.53	3.261	219.8925263	7.348
3	1	20/01/2012	1394393.84	0	54.11	3.268	219.9856893	7.348
4	1	27/01/2012	1319325.59	0	54.26	3.29	220.0788523	7.348
5	1	03/02/2012	1636339.65	0	56.55	3.36	220.1720153	7.348
6	1	10/02/2012	1802477.43	1	48.02	3.409	220.2651783	7.348
7	1	17/02/2012	1819870	0	45.32	3.51	220.4257586	7.348
8	1	24/02/2012	1539387.83	0	57.25	3.555	220.636902	7.348
9	1	02/03/2012	1688420.76	0	60.96	3.63	220.8480454	7.348
10	1	09/03/2012	1675431.16	0	58.76	3.669	221.0591887	7.348
11	1	16/03/2012	1677472.78	0	64.74	3.734	221.2118132	7.348
12	1	23/03/2012	1511068.07	0	65.93	3.787	221.2864126	7.348
13	1	30/03/2012	1649604.63	0	67.61	3.845	221.3610119	7.348
14	1	06/04/2012	1899676.88	0	70.43	3.891	221.4356112	7.143
15	1	13/04/2012	1621031.7	0	69.07	3.891	221.5102105	7.143
16	1	20/04/2012	1521577.87	0	66.76	3.877	221.5640737	7.143
17	1	27/04/2012	1468928.37	0	67.23	3.814	221.6179368	7.143
18	1	04/05/2012	1684519.99	0	75.55	3.749	221.6718	7.143
19	1	11/05/2012	1611096.05	0	73.77	3.688	221.7256632	7.143
20	1	18/05/2012	1595901.87	0	70.33	3.63	221.742674	7.143
21	1	25/05/2012	1555444.55	0	77.22	3.561	221.744944	7.143
22	1	01/06/2012	1624477.58	0	77.95	3.501	221.7472139	7.143

```
/* Calculate growth rate */

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;
```

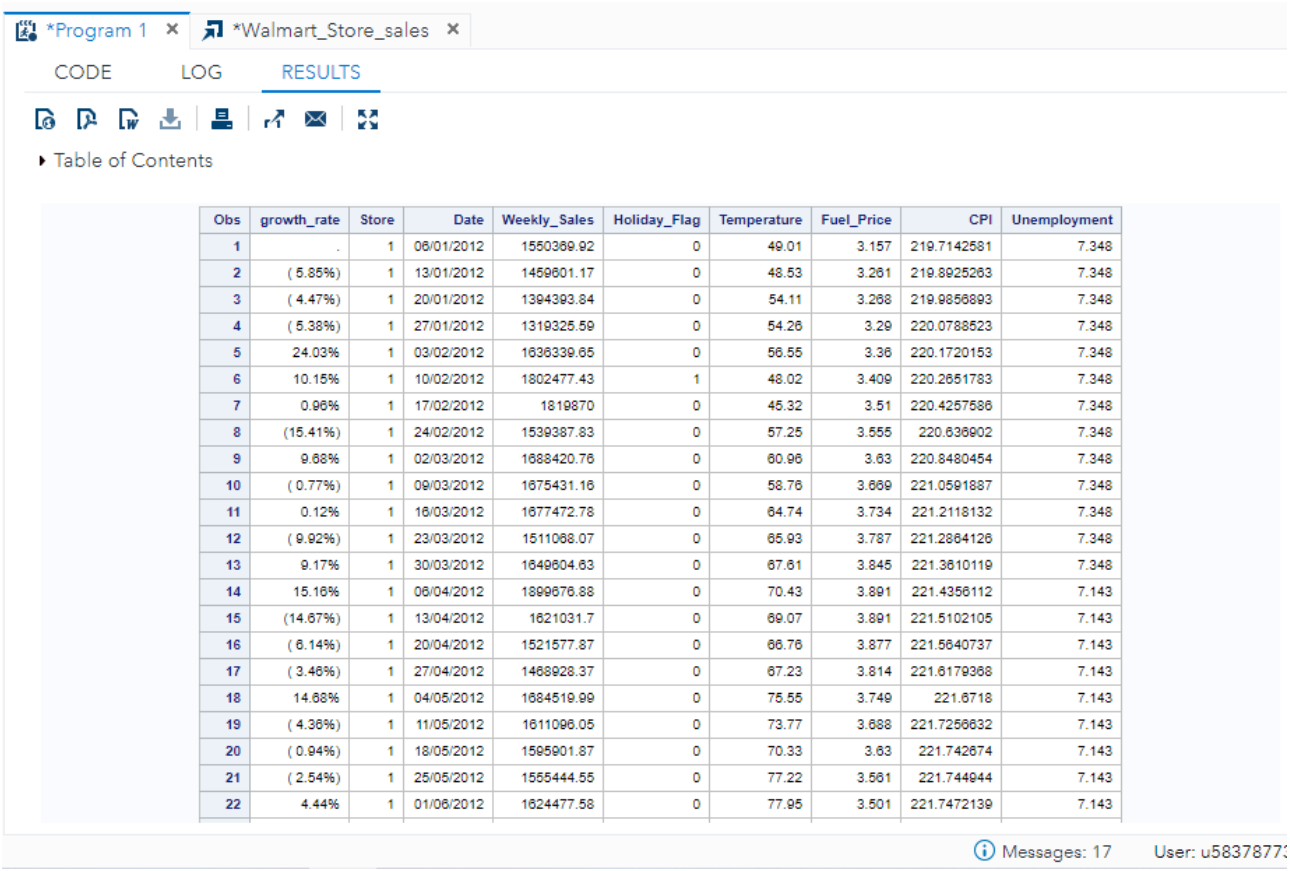
Output:



	growth_rate	Store	Date	Weekly_Sal
1	.	1	06/01/2012	1550369.
2	(5.85%)	1	13/01/2012	1459601.
3	(4.47%)	1	20/01/2012	1394393.
4	(5.38%)	1	27/01/2012	1319325.
5	24.03%	1	03/02/2012	1636339.
6	10.15%	1	10/02/2012	1802477.
7	0.96%	1	17/02/2012	18198
8	(15.41%)	1	24/02/2012	1539387.
9	9.68%	1	02/03/2012	1688420.
10	(0.77%)	1	09/03/2012	1675431.
11	0.12%	1	16/03/2012	1677472.
12	(9.92%)	1	23/03/2012	1511068.
13	9.17%	1	30/03/2012	1649604.
14	15.16%	1	06/04/2012	1899676.

```
PROC PRINT DATA=growth;
RUN;
```

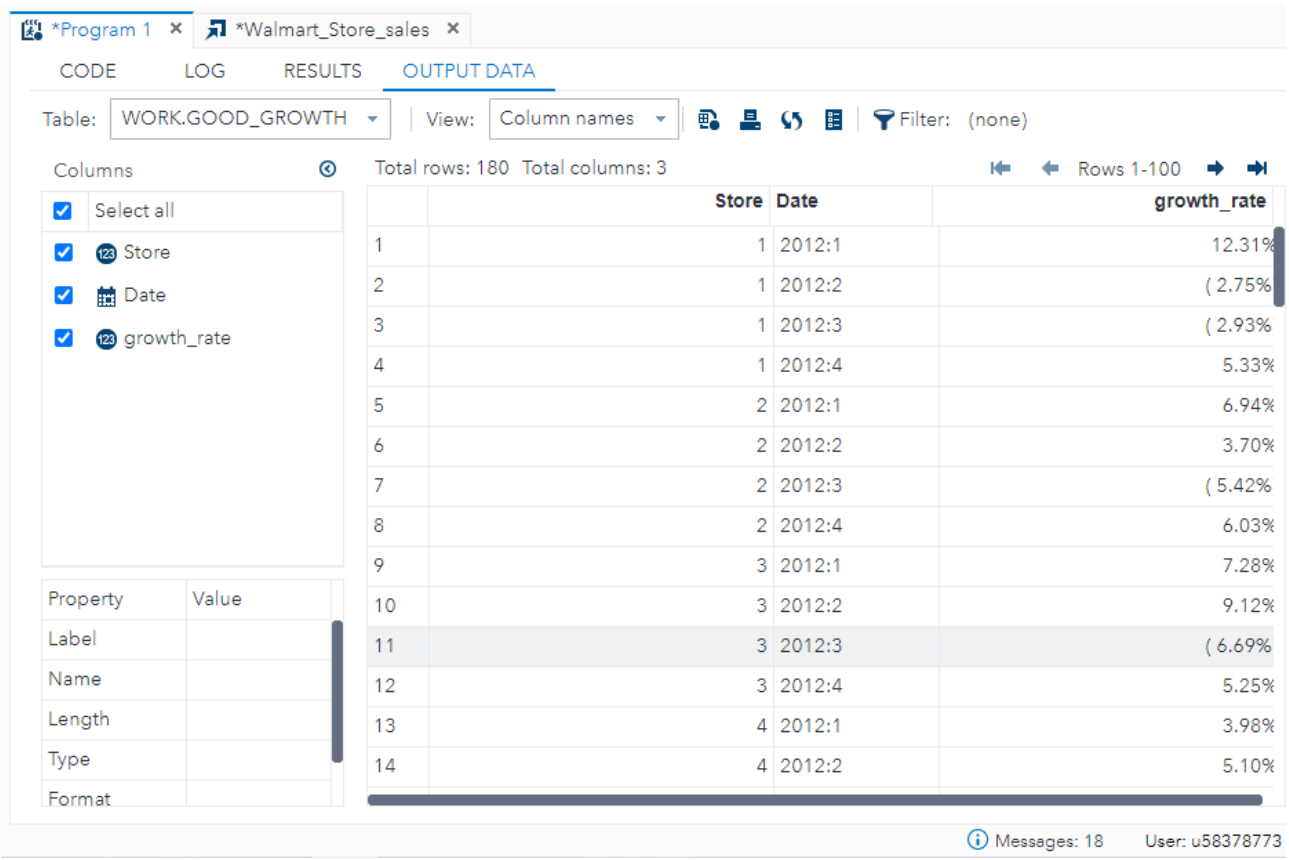
Output:



Obs	growth_rate	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	.	1	06/01/2012	1550369.92	0	49.01	3.157	219.7142581	7.348
2	(5.85%)	1	13/01/2012	1459601.17	0	48.53	3.261	219.8925263	7.348
3	(4.47%)	1	20/01/2012	1394393.84	0	54.11	3.268	219.9856893	7.348
4	(5.38%)	1	27/01/2012	1319325.59	0	54.26	3.29	220.0788523	7.348
5	24.03%	1	03/02/2012	1636339.65	0	56.55	3.36	220.1720153	7.348
6	10.15%	1	10/02/2012	1802477.43	1	48.02	3.409	220.2651783	7.348
7	0.96%	1	17/02/2012	1819870	0	45.32	3.51	220.4257586	7.348
8	(15.41%)	1	24/02/2012	1539387.83	0	57.25	3.555	220.636902	7.348
9	9.68%	1	02/03/2012	1688420.76	0	60.96	3.63	220.8480454	7.348
10	(0.77%)	1	09/03/2012	1675431.16	0	58.76	3.669	221.0591887	7.348
11	0.12%	1	16/03/2012	1677472.78	0	64.74	3.734	221.2118132	7.348
12	(9.92%)	1	23/03/2012	1511068.07	0	65.93	3.787	221.2864126	7.348
13	9.17%	1	30/03/2012	1649604.63	0	67.61	3.845	221.3610119	7.348
14	15.16%	1	06/04/2012	1899676.88	0	70.43	3.891	221.4356112	7.143
15	(14.67%)	1	13/04/2012	1621031.7	0	69.07	3.891	221.5102105	7.143
16	(6.14%)	1	20/04/2012	1521577.87	0	66.76	3.877	221.5640737	7.143
17	(3.46%)	1	27/04/2012	1468928.37	0	67.23	3.814	221.6179368	7.143
18	14.68%	1	04/05/2012	1684519.99	0	75.55	3.749	221.6718	7.143
19	(4.36%)	1	11/05/2012	1611096.05	0	73.77	3.688	221.7256632	7.143
20	(0.94%)	1	18/05/2012	1595901.87	0	70.33	3.63	221.742674	7.143
21	(2.54%)	1	25/05/2012	1555444.55	0	77.22	3.561	221.744944	7.143
22	4.44%	1	01/06/2012	1624477.58	0	77.95	3.501	221.7472139	7.143

```
/* 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;
```

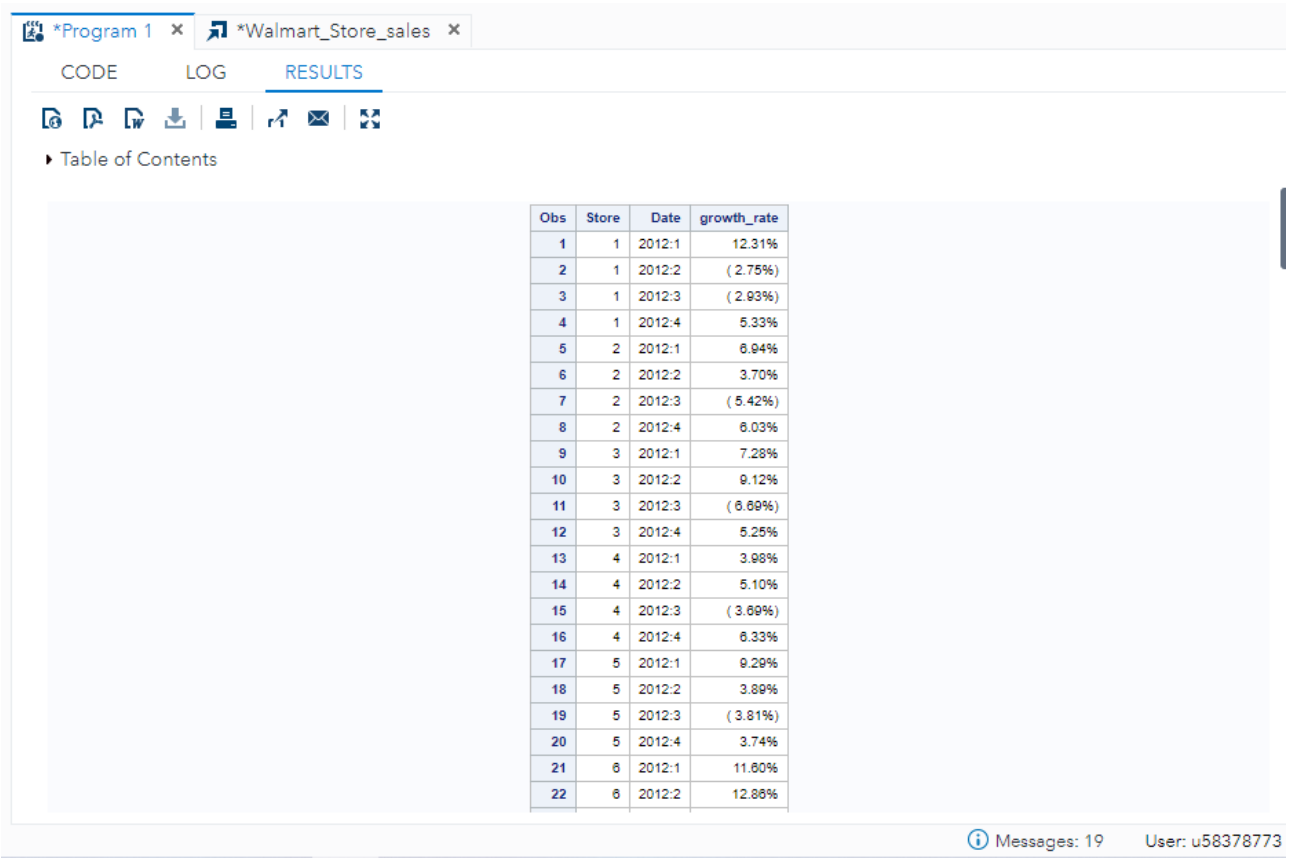
Output:



	Store	Date	growth_rate
1	1	2012:1	12.31%
2	1	2012:2	(2.75%
3	1	2012:3	(2.93%
4	1	2012:4	5.33%
5	2	2012:1	6.94%
6	2	2012:2	3.70%
7	2	2012:3	(5.42%
8	2	2012:4	6.03%
9	3	2012:1	7.28%
10	3	2012:2	9.12%
11	3	2012:3	(6.69%
12	3	2012:4	5.25%
13	4	2012:1	3.98%
14	4	2012:2	5.10%

```
PROC PRINT DATA=good_growth;
RUN;
```

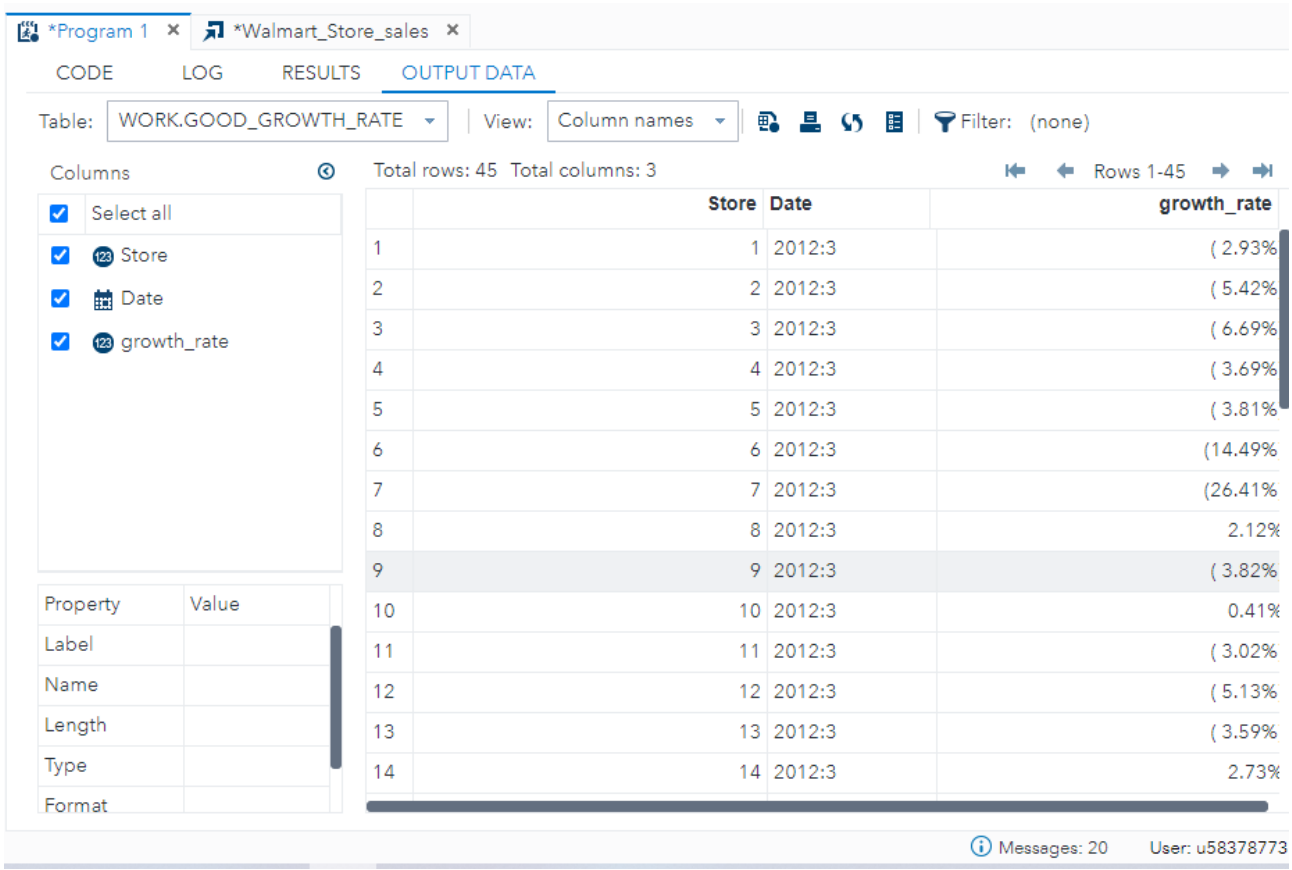
Output:



Obs	Store	Date	growth_rate
1	1	2012:1	12.31%
2	1	2012:2	(2.75%)
3	1	2012:3	(2.93%)
4	1	2012:4	5.33%
5	2	2012:1	6.94%
6	2	2012:2	3.70%
7	2	2012:3	(5.42%)
8	2	2012:4	6.03%
9	3	2012:1	7.28%
10	3	2012:2	9.12%
11	3	2012:3	(6.69%)
12	3	2012:4	5.25%
13	4	2012:1	3.98%
14	4	2012:2	5.10%
15	4	2012:3	(3.69%)
16	4	2012:4	6.33%
17	5	2012:1	9.29%
18	5	2012:2	3.89%
19	5	2012:3	(3.81%)
20	5	2012:4	3.74%
21	6	2012:1	11.60%
22	6	2012:2	12.86%

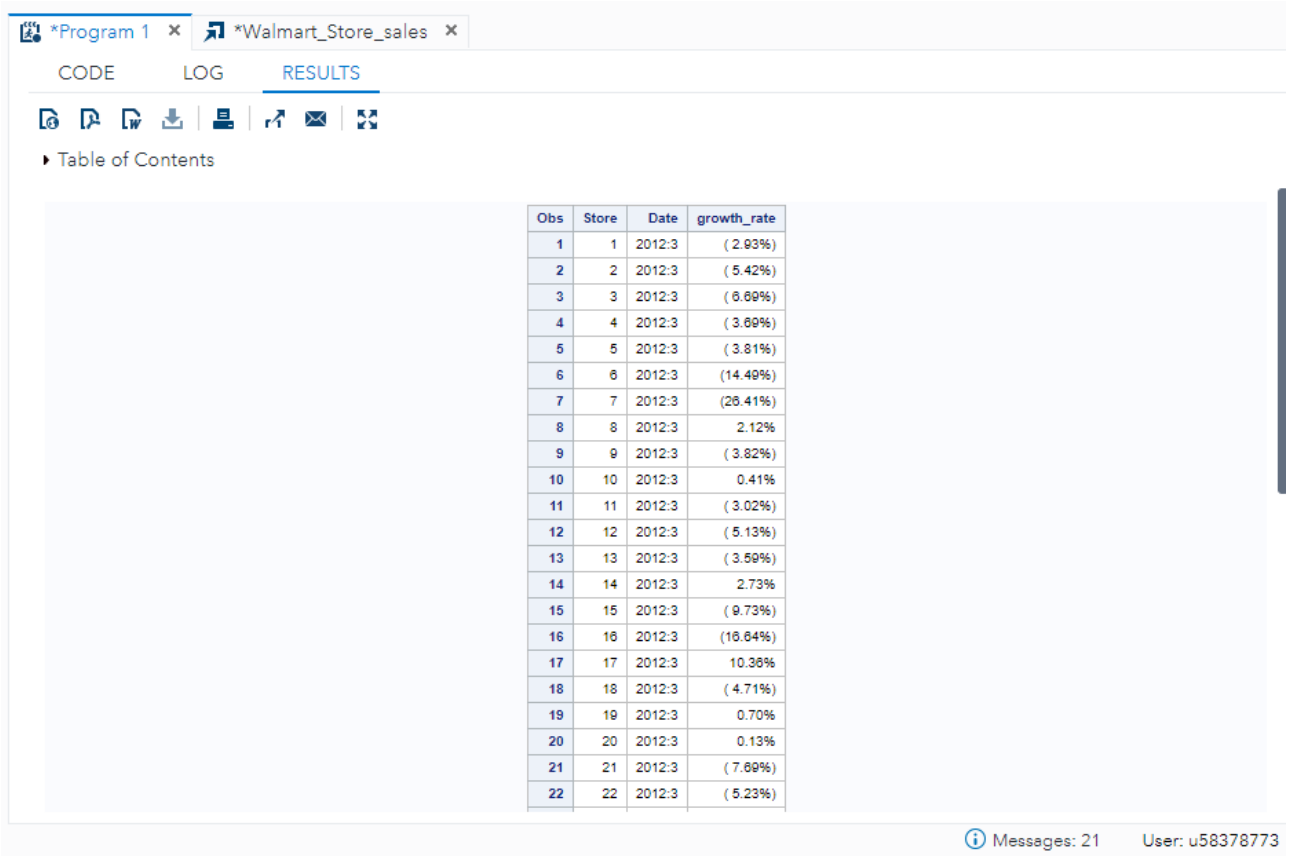
```
/* From timeseries data filtered only Q3 observations */

DATA good_growth_rate;
set good_growth;
where  qtr(date)=  3;
RUN;
```

Output:



```
PROC PRINT DATA=good_growth_rate;
RUN;
```

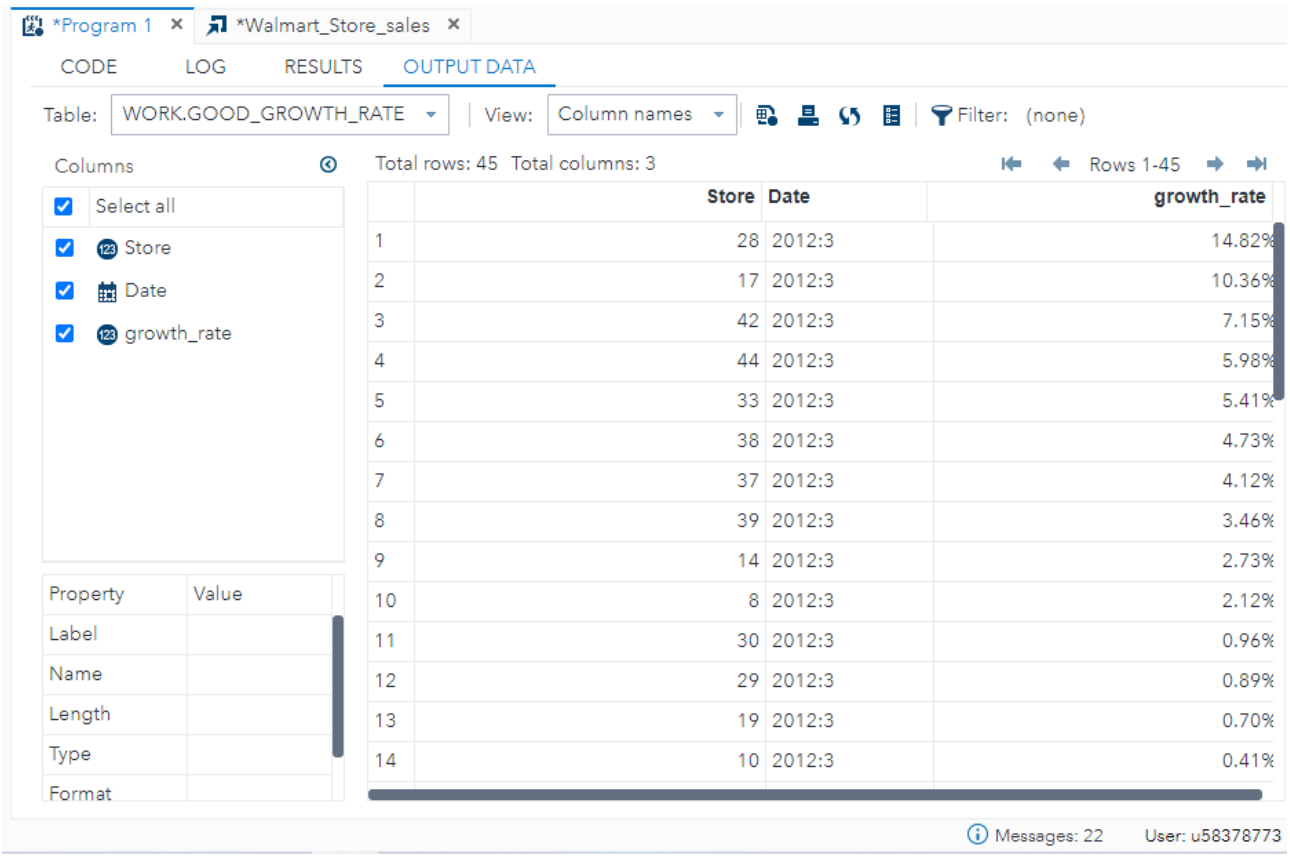
Output:



Obs	Store	Date	growth_rate
1	1	2012:3	(2.93%)
2	2	2012:3	(5.42%)
3	3	2012:3	(6.69%)
4	4	2012:3	(3.69%)
5	5	2012:3	(3.81%)
6	6	2012:3	(14.49%)
7	7	2012:3	(26.41%)
8	8	2012:3	2.12%
9	9	2012:3	(3.82%)
10	10	2012:3	0.41%
11	11	2012:3	(3.02%)
12	12	2012:3	(5.13%)
13	13	2012:3	(3.59%)
14	14	2012:3	2.73%
15	15	2012:3	(9.73%)
16	16	2012:3	(16.64%)
17	17	2012:3	10.36%
18	18	2012:3	(4.71%)
19	19	2012:3	0.70%
20	20	2012:3	0.13%
21	21	2012:3	(7.69%)
22	22	2012:3	(5.23%)

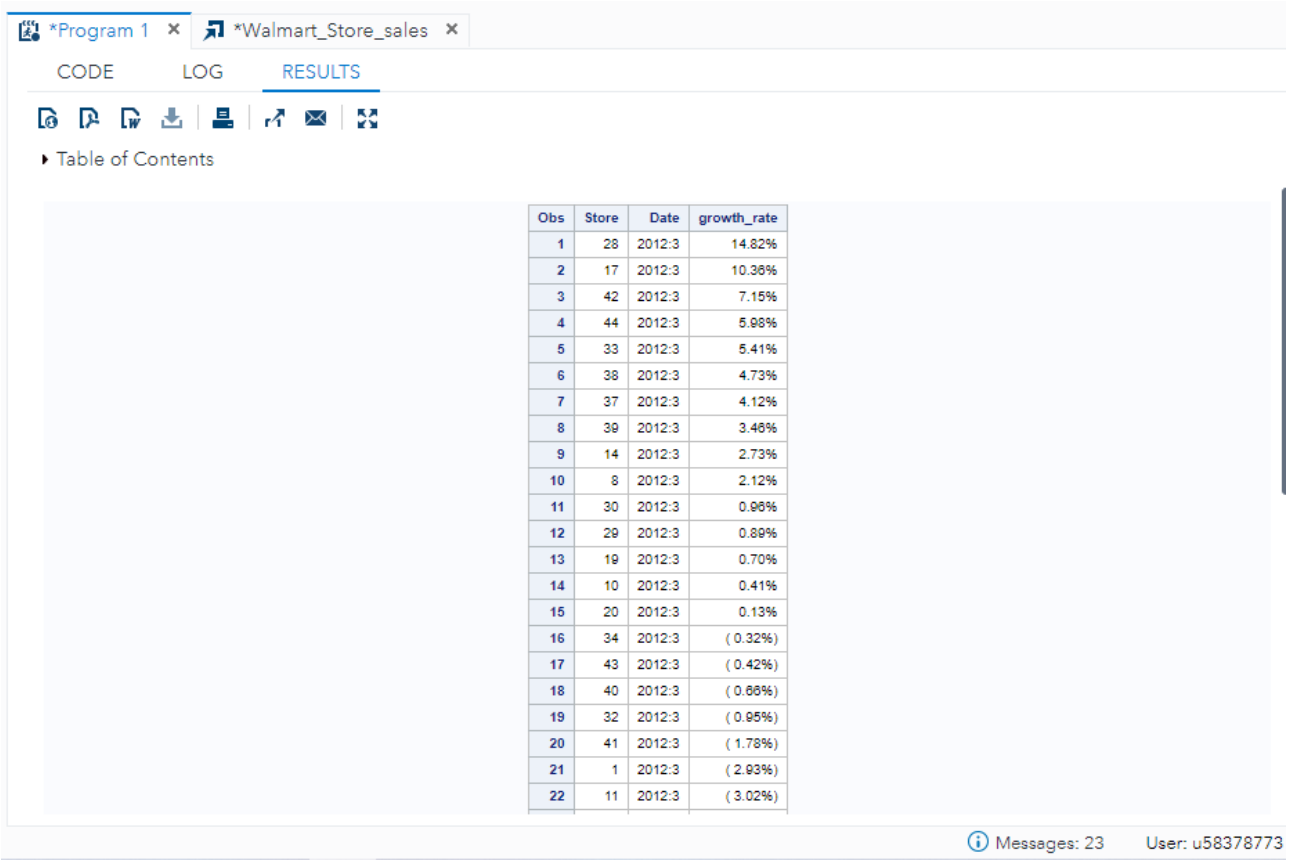
```
/* Now Sort the data to see the good growth rate store wise */

PROC SORT DATA= good_growth_rate;
        by descending growth_rate;
RUN;
```

Output:



```
PROC PRINT DATA= good_growth_rate;
RUN;
```

Output:



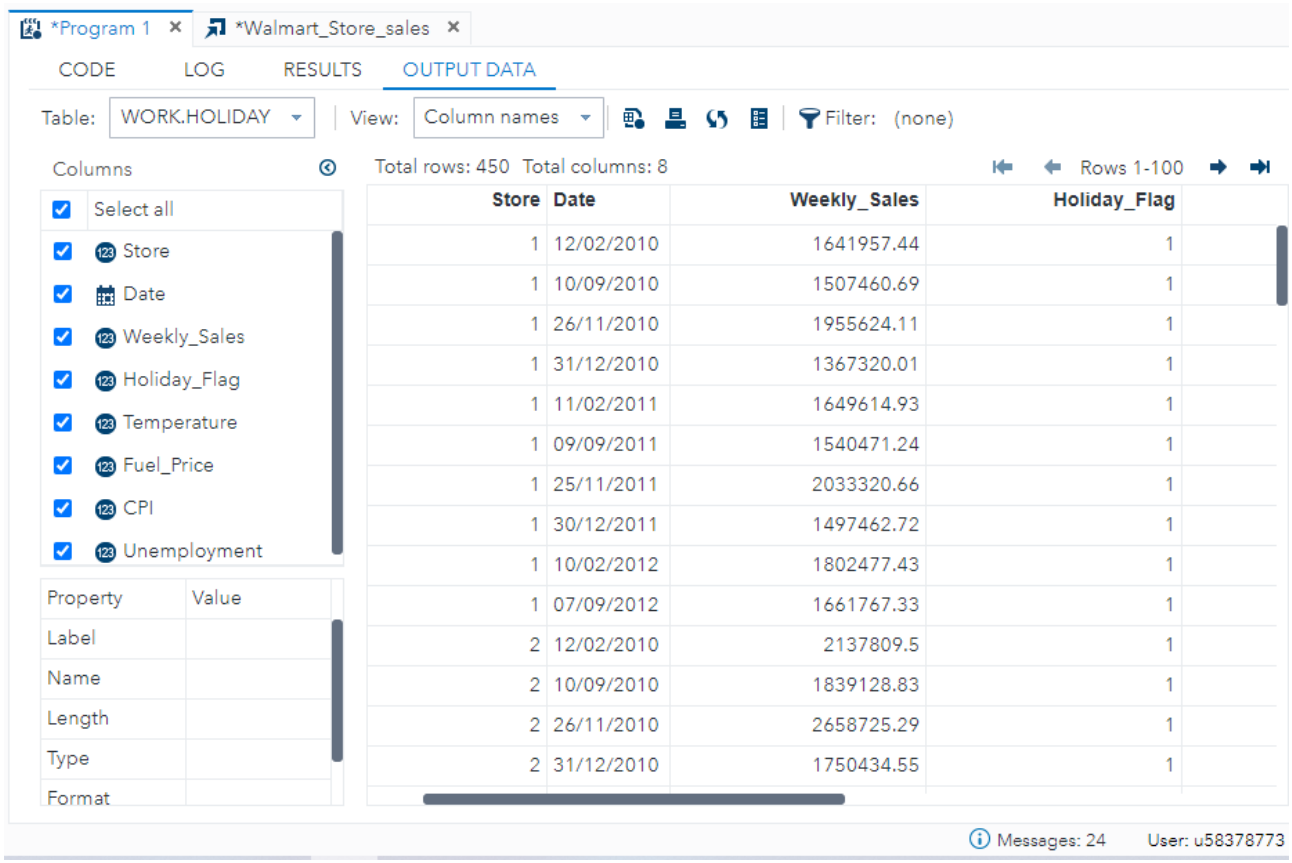
Obs	Store	Date	growth_rate
1	28	2012:3	14.82%
2	17	2012:3	10.36%
3	42	2012:3	7.15%
4	44	2012:3	5.98%
5	33	2012:3	5.41%
6	38	2012:3	4.73%
7	37	2012:3	4.12%
8	39	2012:3	3.46%
9	14	2012:3	2.73%
10	8	2012:3	2.12%
11	30	2012:3	0.96%
12	29	2012:3	0.89%
13	19	2012:3	0.70%
14	10	2012:3	0.41%
15	20	2012:3	0.13%
16	34	2012:3	(0.32%)
17	43	2012:3	(0.42%)
18	40	2012:3	(0.66%)
19	32	2012:3	(0.95%)
20	41	2012:3	(1.78%)
21	1	2012:3	(2.93%)
22	11	2012:3	(3.02%)

/* 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.walmart;
where  holiday_flag=1;
RUN;
```

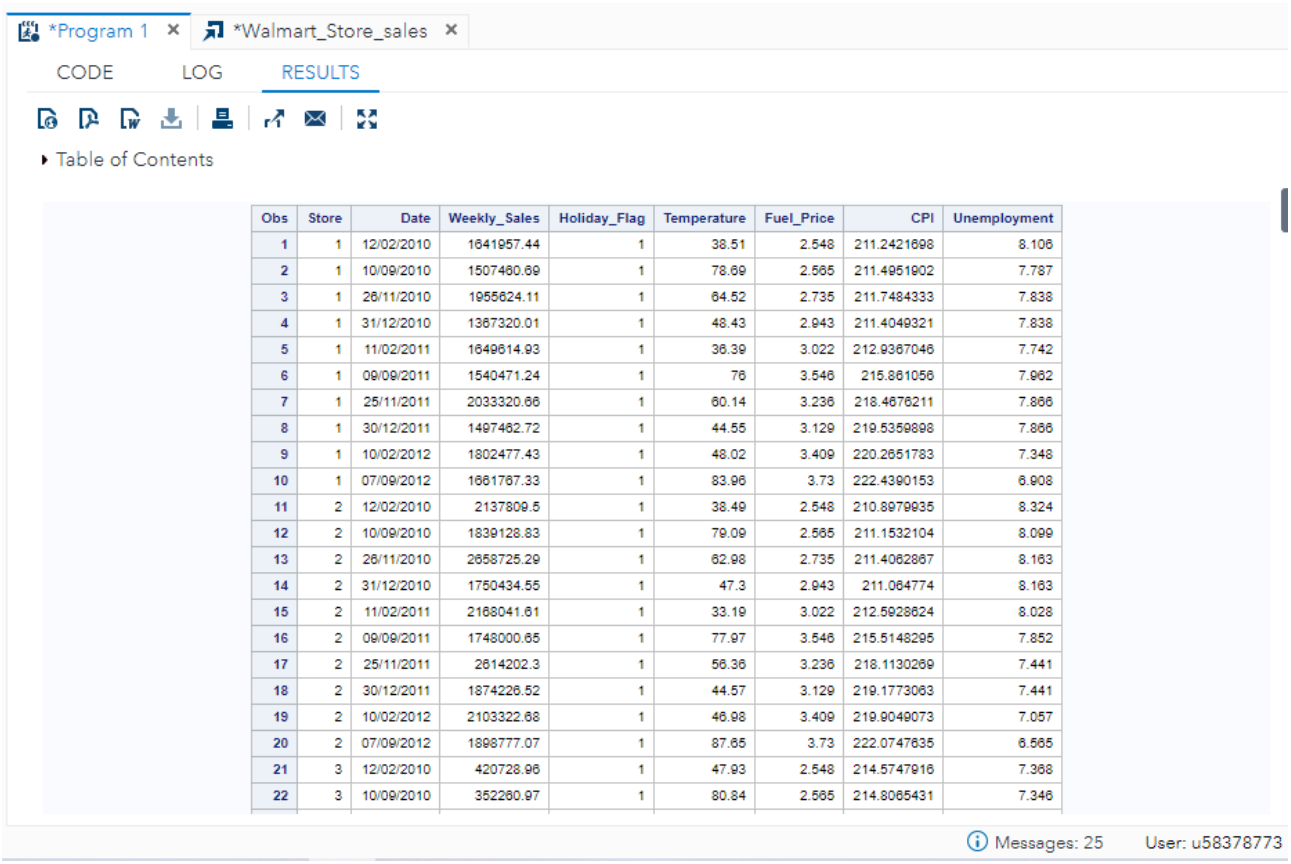
Output:



Store	Date	Weekly_Sales	Holiday_Flag
1	12/02/2010	1641957.44	1
1	10/09/2010	1507460.69	1
1	26/11/2010	1955624.11	1
1	31/12/2010	1367320.01	1
1	11/02/2011	1649614.93	1
1	09/09/2011	1540471.24	1
1	25/11/2011	2033320.66	1
1	30/12/2011	1497462.72	1
1	10/02/2012	1802477.43	1
1	07/09/2012	1661767.33	1
2	12/02/2010	2137809.5	1
2	10/09/2010	1839128.83	1
2	26/11/2010	2658725.29	1
2	31/12/2010	1750434.55	1

```
PROC PRINT DATA=holiday;
RUN;
```

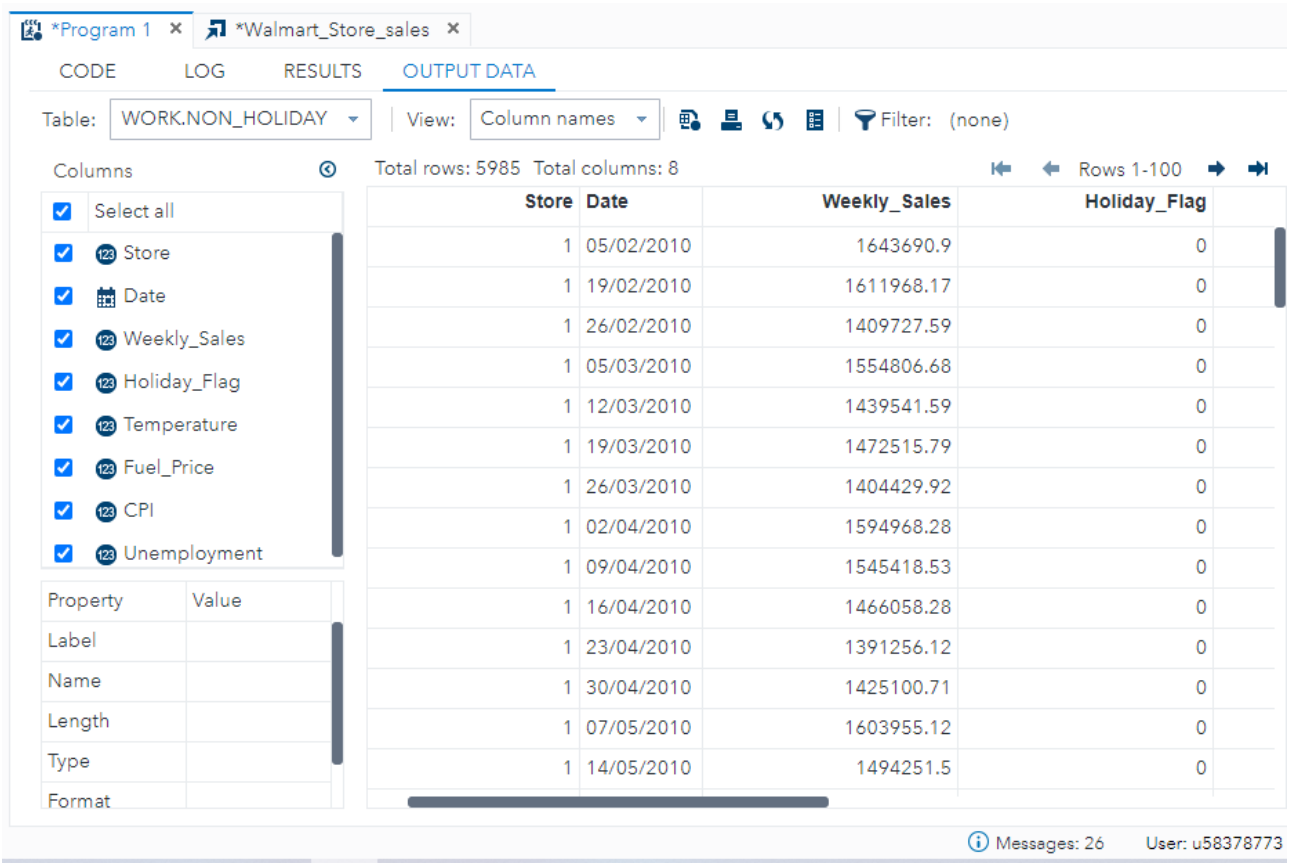
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	12/02/2010	1641957.44	1	38.51	2.548	211.2421698	8.106
2	1	10/09/2010	1507460.69	1	78.69	2.565	211.4951902	7.787
3	1	26/11/2010	1955624.11	1	64.52	2.735	211.7484333	7.838
4	1	31/12/2010	1367320.01	1	48.43	2.943	211.4049321	7.838
5	1	11/02/2011	1649614.93	1	36.39	3.022	212.9367046	7.742
6	1	09/09/2011	1540471.24	1	76	3.546	215.861056	7.962
7	1	25/11/2011	2033320.66	1	60.14	3.236	218.4676211	7.866
8	1	30/12/2011	1497462.72	1	44.55	3.129	219.5359898	7.866
9	1	10/02/2012	1802477.43	1	48.02	3.409	220.2651783	7.348
10	1	07/09/2012	1661767.33	1	83.96	3.73	222.4390153	6.908
11	2	12/02/2010	2137809.5	1	38.49	2.548	210.8979935	8.324
12	2	10/09/2010	1839128.83	1	79.09	2.565	211.1532104	8.099
13	2	26/11/2010	2658725.29	1	62.98	2.735	211.4062867	8.163
14	2	31/12/2010	1750434.55	1	47.3	2.943	211.064774	8.163
15	2	11/02/2011	2168041.61	1	33.19	3.022	212.5928624	8.028
16	2	09/09/2011	1748000.65	1	77.97	3.546	215.5148295	7.852
17	2	25/11/2011	2614202.3	1	56.36	3.236	218.1130269	7.441
18	2	30/12/2011	1874226.52	1	44.57	3.129	219.1773063	7.441
19	2	10/02/2012	2103322.68	1	46.98	3.409	219.9049073	7.057
20	2	07/09/2012	1898777.07	1	87.65	3.73	222.0747635	6.565
21	3	12/02/2010	420728.96	1	47.93	2.548	214.5747916	7.368
22	3	10/09/2010	352260.97	1	80.84	2.565	214.8065431	7.346

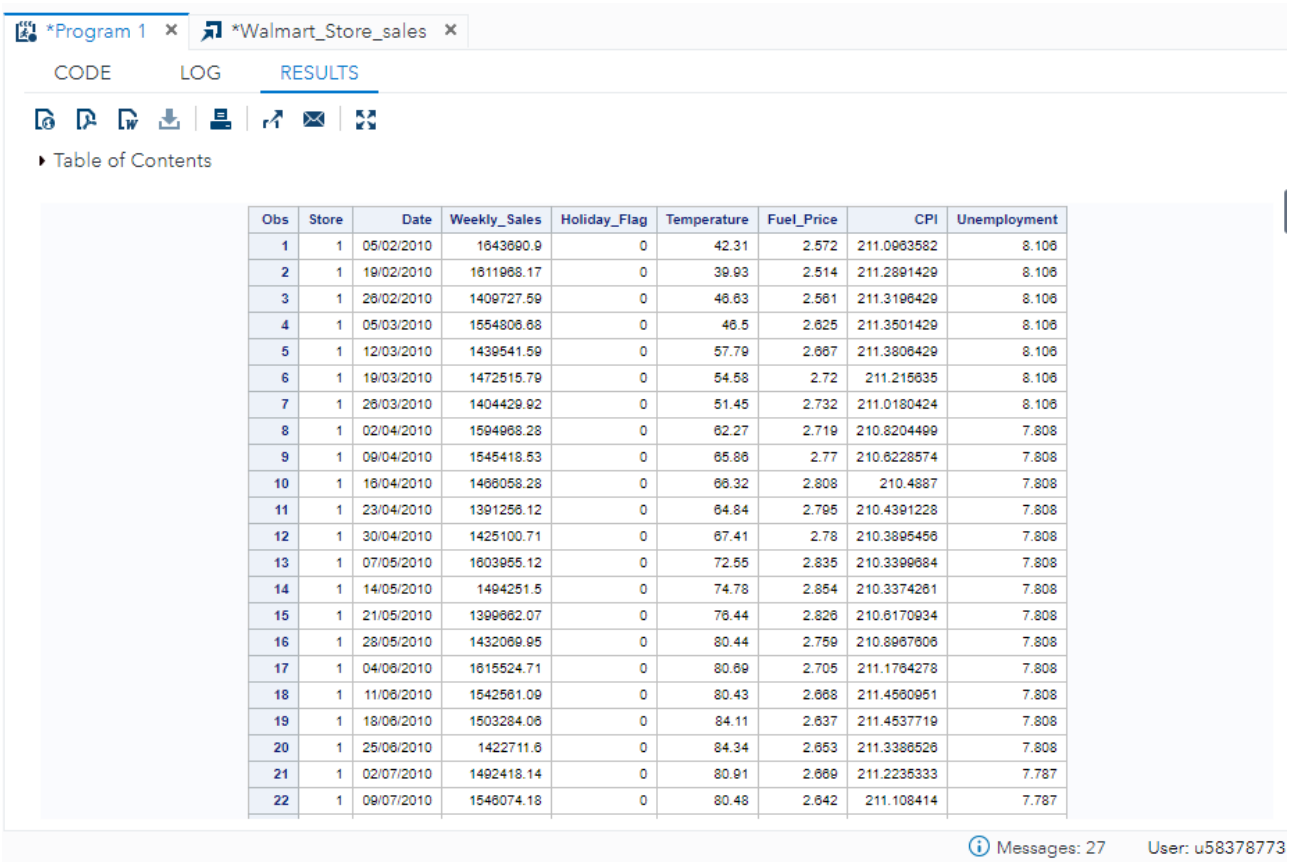
```
/* Separate the non-holiday dates from main dataset's date */

DATA non_holiday;
set work.walmart;
where  holiday_flag=0;
RUN;
```

Output:



```
PROC PRINT DATA=non_holiday;
RUN;
```

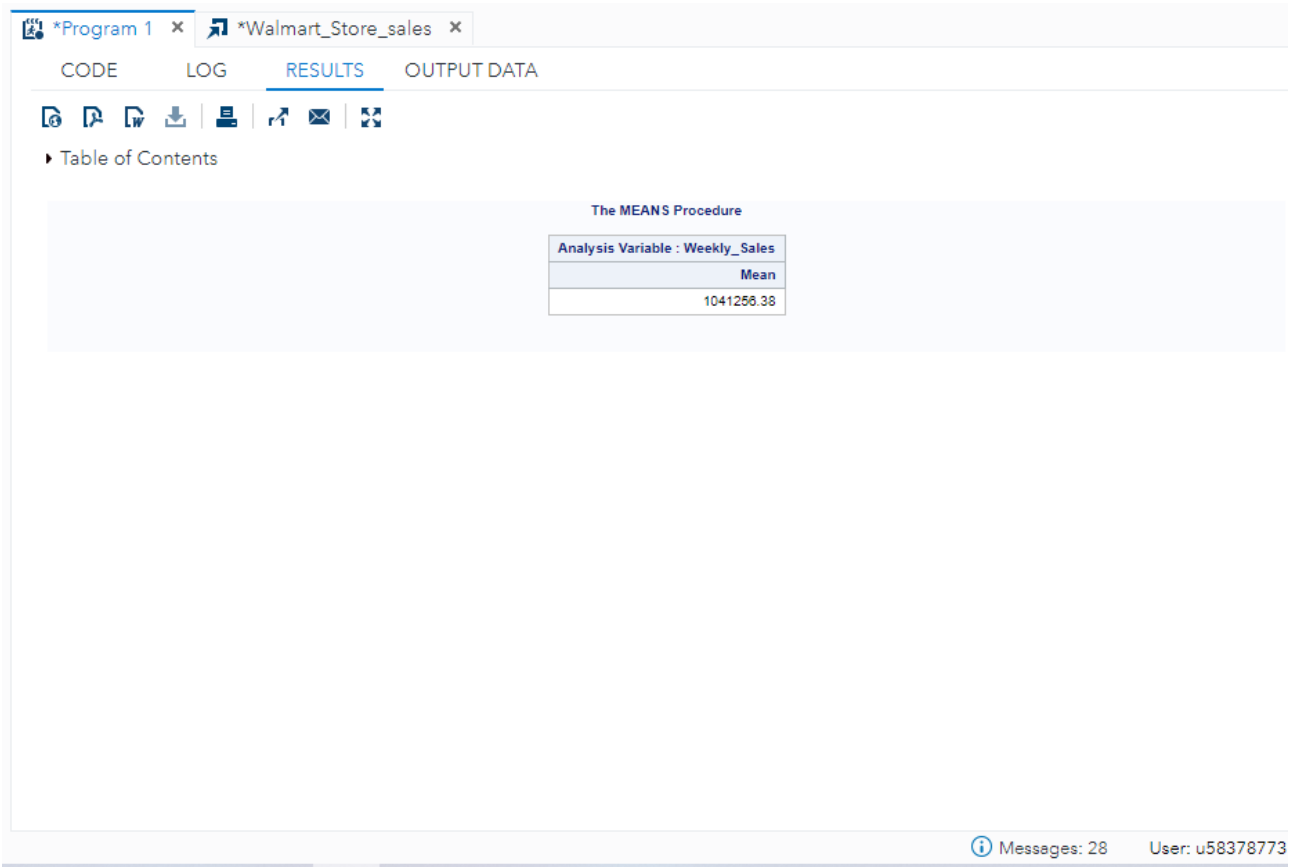
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	05/02/2010	1643690.9	0	42.31	2.572	211.0963582	8.106
2	1	19/02/2010	1611968.17	0	38.93	2.514	211.2891429	8.106
3	1	26/02/2010	1409727.59	0	46.63	2.561	211.3196429	8.106
4	1	05/03/2010	1554806.68	0	46.5	2.625	211.3501429	8.106
5	1	12/03/2010	1439541.59	0	57.79	2.667	211.3806429	8.106
6	1	19/03/2010	1472515.79	0	54.58	2.72	211.215635	8.106
7	1	26/03/2010	1404429.92	0	51.45	2.732	211.0180424	8.106
8	1	02/04/2010	1594968.28	0	62.27	2.719	210.8204499	7.808
9	1	09/04/2010	1545418.53	0	65.86	2.77	210.6228574	7.808
10	1	16/04/2010	1466058.28	0	66.32	2.808	210.4887	7.808
11	1	23/04/2010	1391256.12	0	64.84	2.795	210.4391228	7.808
12	1	30/04/2010	1425100.71	0	67.41	2.78	210.3895456	7.808
13	1	07/05/2010	1603955.12	0	72.55	2.835	210.3399684	7.808
14	1	14/05/2010	1494251.5	0	74.78	2.854	210.3374261	7.808
15	1	21/05/2010	1399662.07	0	76.44	2.826	210.6170934	7.808
16	1	28/05/2010	1432069.95	0	80.44	2.759	210.8967606	7.808
17	1	04/06/2010	1615524.71	0	80.69	2.705	211.1764278	7.808
18	1	11/06/2010	1542561.09	0	80.43	2.668	211.4560951	7.808
19	1	18/06/2010	1503284.06	0	84.11	2.637	211.4537719	7.808
20	1	25/06/2010	1422711.6	0	84.34	2.653	211.3386526	7.808
21	1	02/07/2010	1492418.14	0	80.91	2.669	211.2235333	7.787
22	1	09/07/2010	1546074.18	0	80.48	2.642	211.108414	7.787

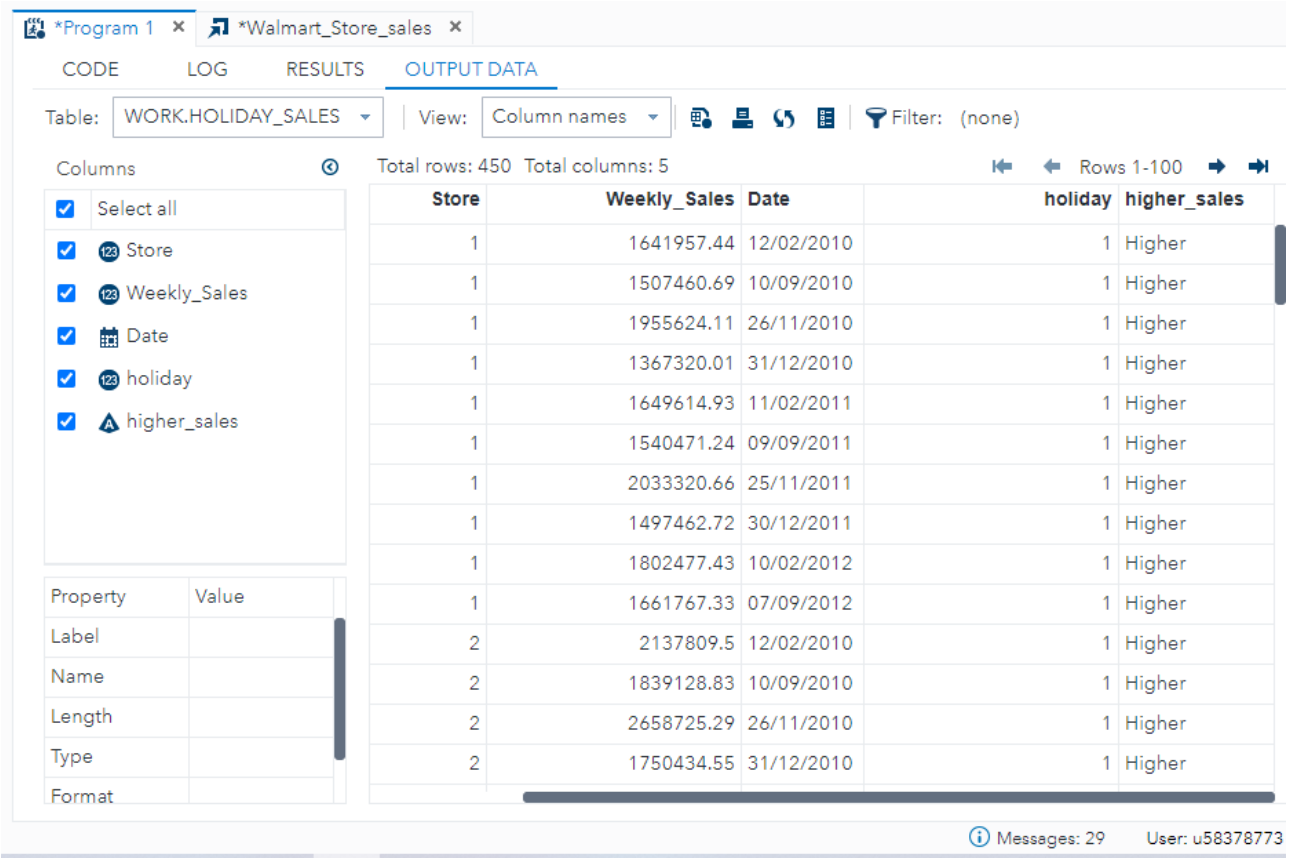
```
/* Calculate the mean weekly_sales of the non-holiday data */

PROC MEANS DATA= non_holiday mean nonobs;
        output out= mean_sales;
        var weekly_sales;
RUN;
```

Output:



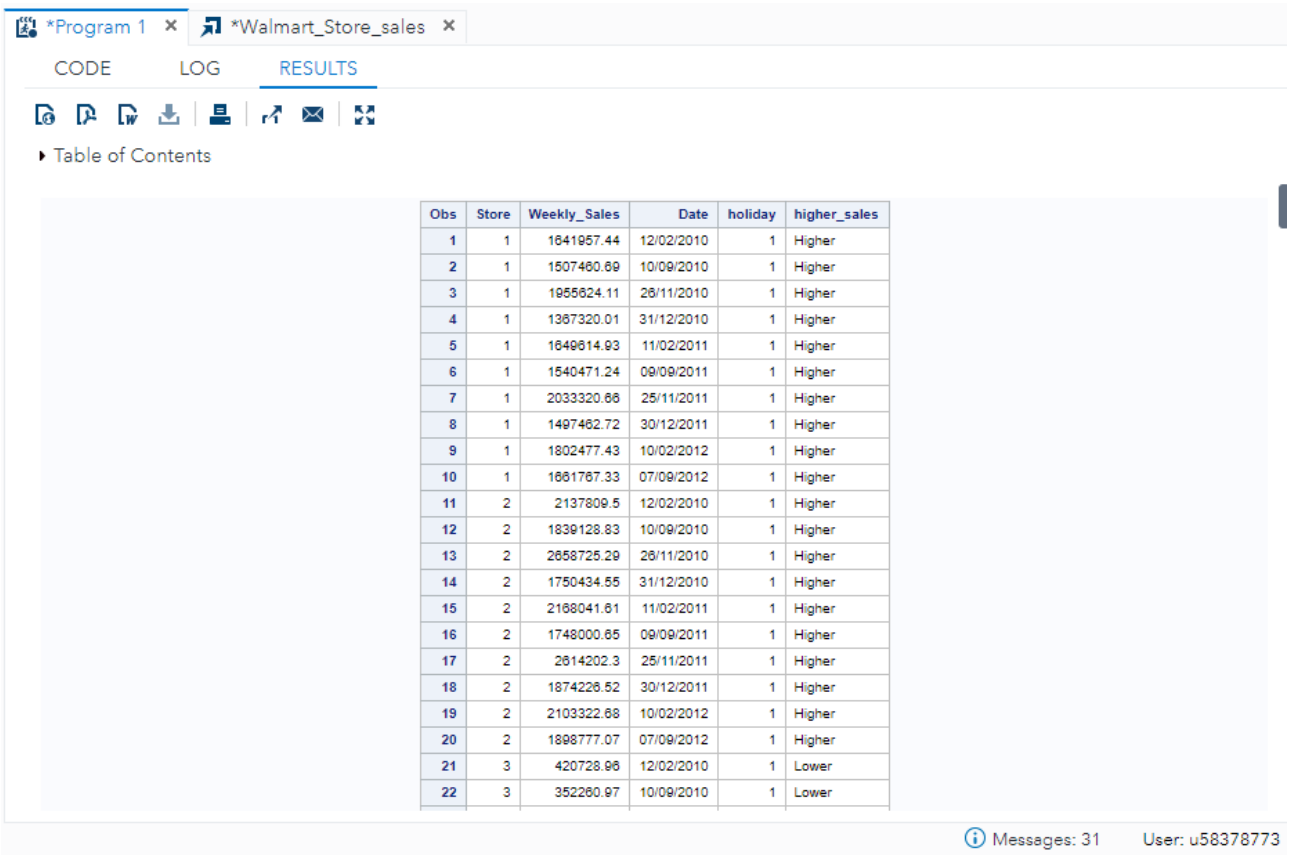
```
/* Compare the mean weekly_sales of the non-holiday data with weekly
sales of the holiday data */

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;
```

Output:



```
PROC PRINT DATA= holiday_sales;
RUN;
```

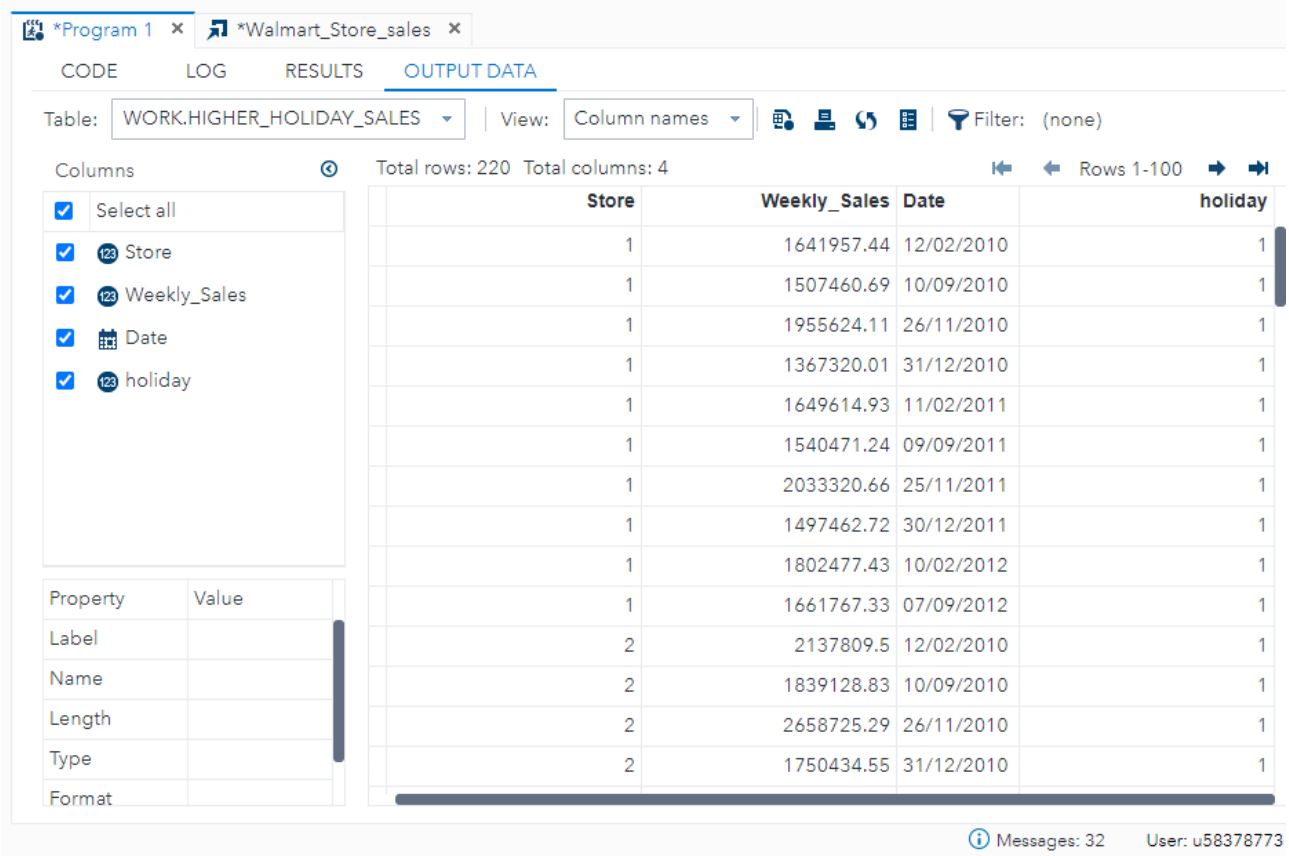
Output:



Obs	Store	Weekly_Sales	Date	holiday	higher_sales
1	1	1641957.44	12/02/2010	1	Higher
2	1	1507460.69	10/09/2010	1	Higher
3	1	1955624.11	26/11/2010	1	Higher
4	1	1367320.01	31/12/2010	1	Higher
5	1	1649614.93	11/02/2011	1	Higher
6	1	1540471.24	09/09/2011	1	Higher
7	1	2033320.66	25/11/2011	1	Higher
8	1	1497462.72	30/12/2011	1	Higher
9	1	1802477.43	10/02/2012	1	Higher
10	1	1661767.33	07/09/2012	1	Higher
11	2	2137809.5	12/02/2010	1	Higher
12	2	1839128.83	10/09/2010	1	Higher
13	2	2658725.29	26/11/2010	1	Higher
14	2	1750434.55	31/12/2010	1	Higher
15	2	2168041.61	11/02/2011	1	Higher
16	2	1748000.65	09/09/2011	1	Higher
17	2	2614202.3	25/11/2011	1	Higher
18	2	1874226.52	30/12/2011	1	Higher
19	2	2103322.68	10/02/2012	1	Higher
20	2	1898777.07	07/09/2012	1	Higher
21	3	420728.96	12/02/2010	1	Lower
22	3	352260.97	10/09/2010	1	Lower

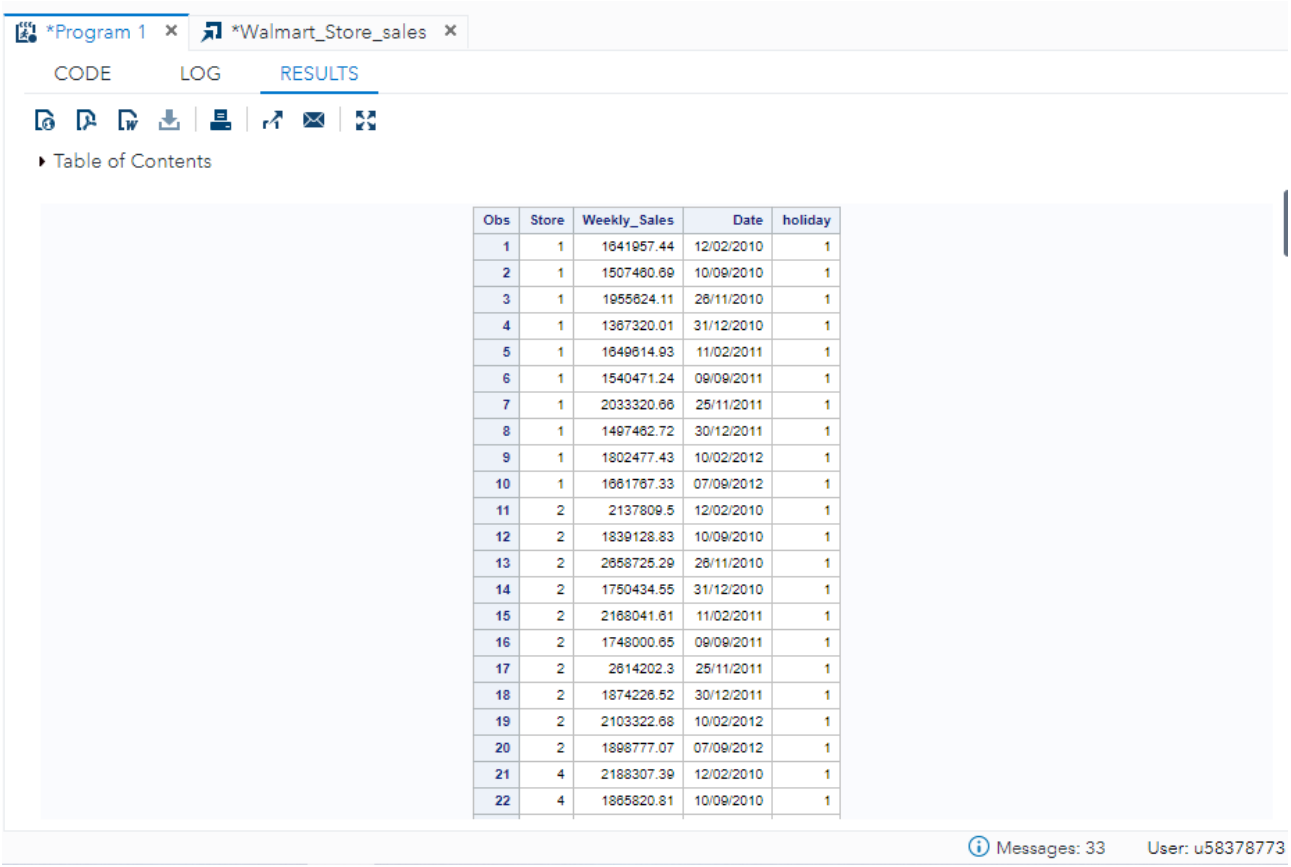
```
/* Finally found out holidays which have higher sales than the mean sales
in non-holiday season for all stores together */

DATA higher_holiday_sales;
set work.holiday_sales;
where higher_sales = 'Higher';
drop higher_sales;
title 'Higher Sales during Holidays';
RUN;
```

Output:



```
PROC PRINT DATA= higher_holiday_sales;
RUN;
```

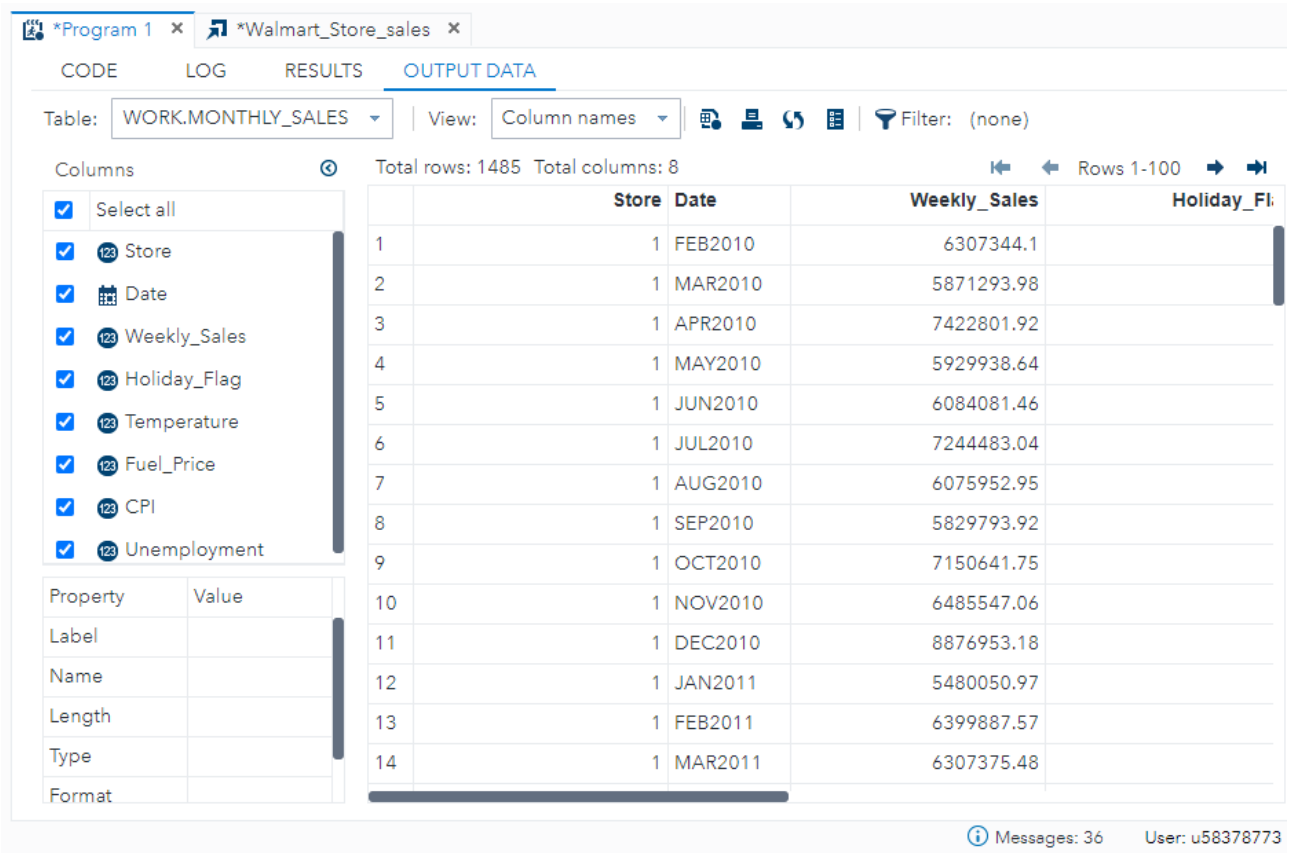
Output:



Obs	Store	Weekly_Sales	Date	holiday
1	1	1641957.44	12/02/2010	1
2	1	1507460.69	10/09/2010	1
3	1	1955624.11	26/11/2010	1
4	1	1367320.01	31/12/2010	1
5	1	1649614.93	11/02/2011	1
6	1	1540471.24	09/09/2011	1
7	1	2033320.66	25/11/2011	1
8	1	1497462.72	30/12/2011	1
9	1	1802477.43	10/02/2012	1
10	1	1661767.33	07/09/2012	1
11	2	2137809.5	12/02/2010	1
12	2	1839128.83	10/09/2010	1
13	2	2658725.29	26/11/2010	1
14	2	1750434.55	31/12/2010	1
15	2	2168041.61	11/02/2011	1
16	2	1748000.65	09/09/2011	1
17	2	2614202.3	25/11/2011	1
18	2	1874226.52	30/12/2011	1
19	2	2103322.68	10/02/2012	1
20	2	1898777.07	07/09/2012	1
21	4	2188307.39	12/02/2010	1
22	4	1865820.81	10/09/2010	1

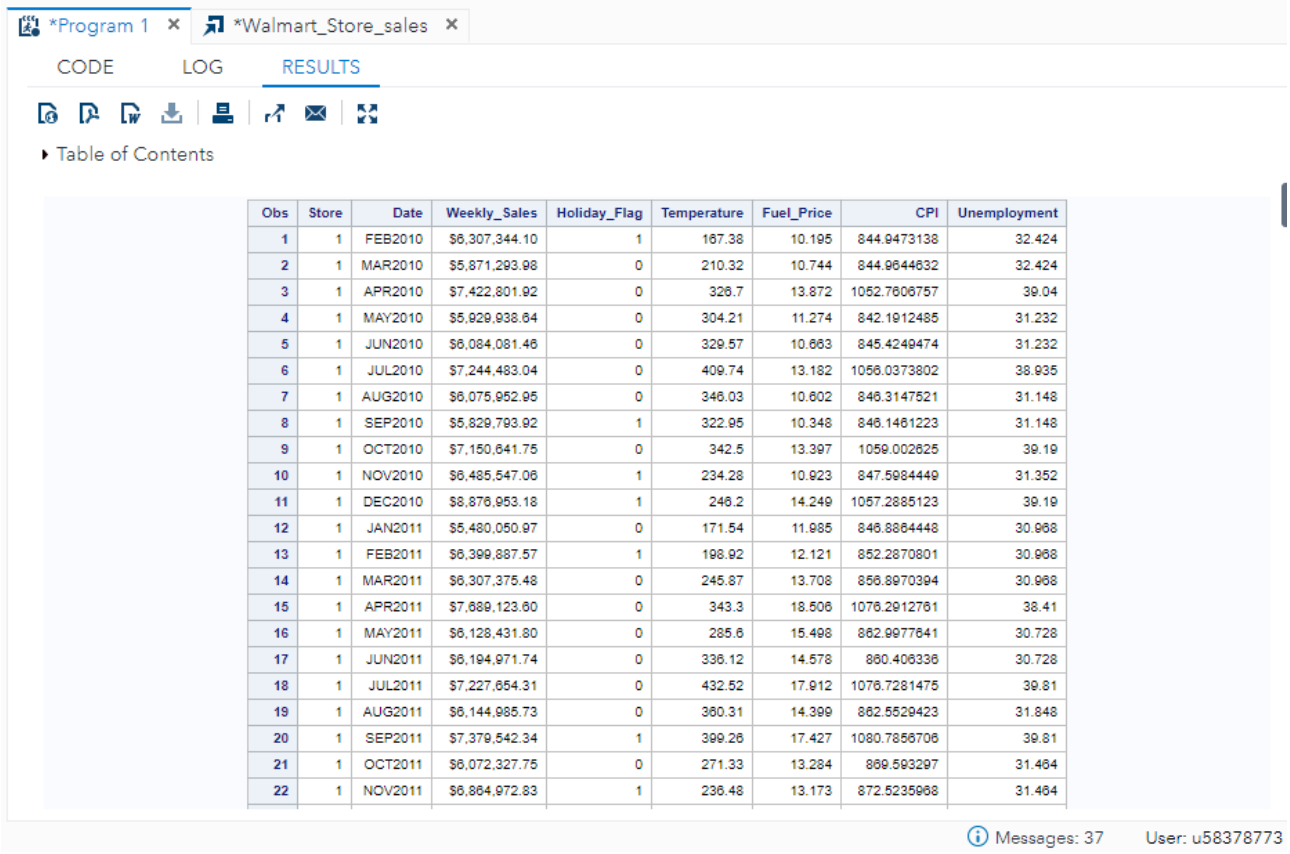
```
/* Provide a monthly and semester view of sales in units and give insights
*/

/* Monthly view of sales in units */

/* Convert walmart data into timeseries data */

PROC TIMESERIES DATA= work.walmart out= monthly_sales;
by store;
id  date  interval=month  accumulate=total;
var weekly_sales holiday_flag temperature fuel_price cpi unemployment;
RUN;
```

Output:



```
PROC PRINT DATA=work.monthly_sales;
format weekly_sales dollar16.2;
RUN;
```

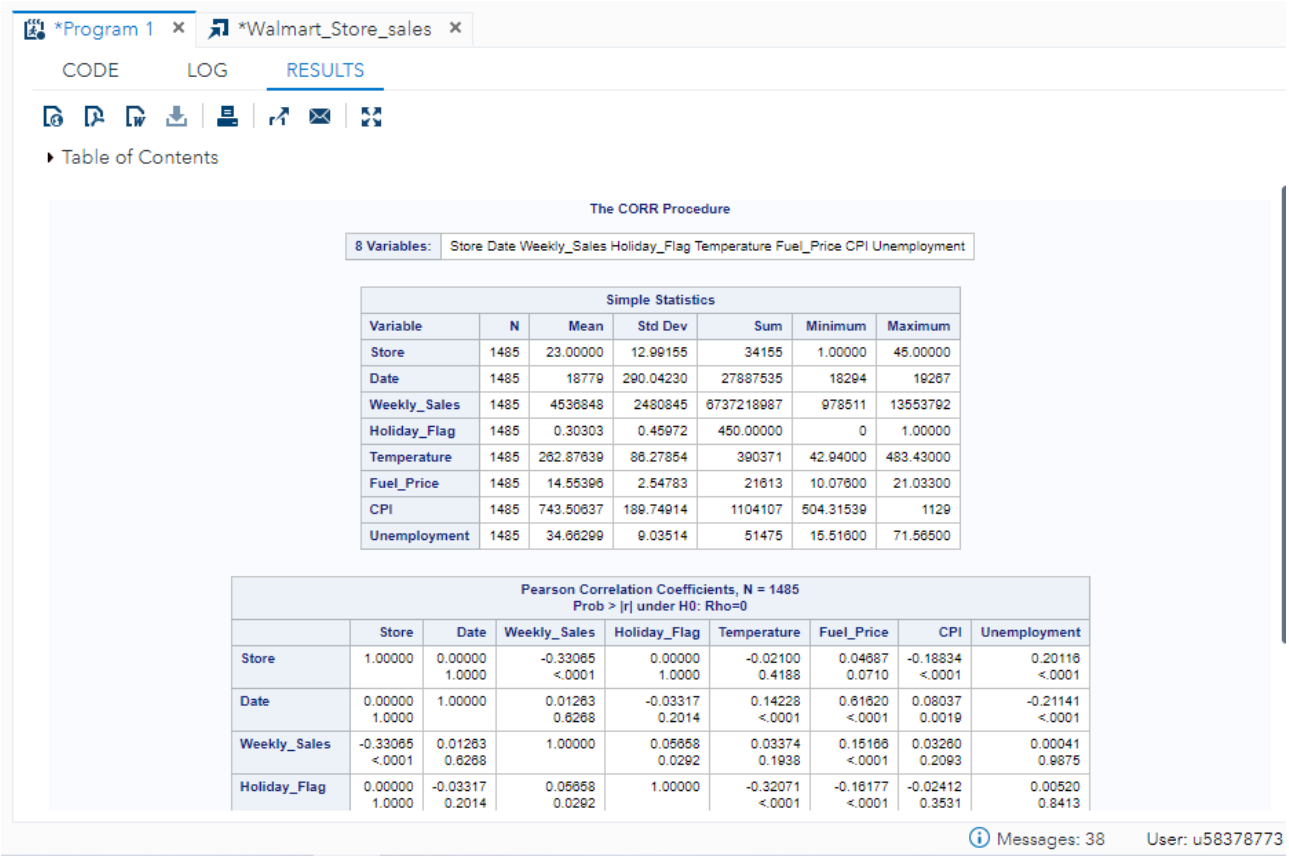
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	FEB2010	$6,307,344.10	1	167.38	10.195	844.9473138	32.424
2	1	MAR2010	$5,871,293.98	0	210.32	10.744	844.9644632	32.424
3	1	APR2010	$7,422,801.92	0	326.7	13.872	1052.7606757	39.04
4	1	MAY2010	$5,929,938.64	0	304.21	11.274	842.1912485	31.232
5	1	JUN2010	$6,084,081.46	0	329.57	10.663	845.4249474	31.232
6	1	JUL2010	$7,244,483.04	0	409.74	13.182	1056.0373802	38.935
7	1	AUG2010	$6,075,952.95	0	346.03	10.602	846.3147521	31.148
8	1	SEP2010	$5,829,793.92	1	322.95	10.348	846.1461223	31.148
9	1	OCT2010	$7,150,641.75	0	342.5	13.397	1059.002625	39.19
10	1	NOV2010	$6,485,547.06	1	234.28	10.923	847.5984449	31.352
11	1	DEC2010	$8,876,953.18	1	246.2	14.249	1057.2885123	39.19
12	1	JAN2011	$5,480,050.97	0	171.54	11.985	846.8864448	30.968
13	1	FEB2011	$6,399,887.57	1	198.92	12.121	852.2870801	30.968
14	1	MAR2011	$6,307,375.48	0	245.87	13.708	856.8970394	30.968
15	1	APR2011	$7,689,123.60	0	343.3	18.506	1076.2912761	38.41
16	1	MAY2011	$6,128,431.80	0	285.6	15.498	862.9977641	30.728
17	1	JUN2011	$6,194,971.74	0	336.12	14.578	860.406336	30.728
18	1	JUL2011	$7,227,654.31	0	432.52	17.912	1076.7281475	39.81
19	1	AUG2011	$6,144,985.73	0	360.31	14.399	862.5529423	31.848
20	1	SEP2011	$7,379,542.34	1	399.26	17.427	1080.7856706	39.81
21	1	OCT2011	$6,072,327.75	0	271.33	13.284	869.593297	31.464
22	1	NOV2011	$6,864,972.83	1	236.48	13.173	872.5235968	31.464

```
/* Giving insights */

/* Checking the correlation */

PROC CORR DATA= work.monthly_sales;
RUN;
```

Output:



*Program 1 | *Walmart_Store_sales

CODE | LOG | RESULTS

Table of Contents

The CORR Procedure

8 Variables: Store Date Weekly_Sales Holiday_Flag Temperature Fuel_Price CPI Unemployment

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Store	1485	23.00000	12.99155	34155	1.00000	45.00000
Date	1485	18779	290.04230	27887535	18294	19267
Weekly_Sales	1485	4536848	2480845	6737218987	978511	13553792
Holiday_Flag	1485	0.30303	0.45972	450.00000	0	1.00000
Temperature	1485	262.87639	86.27854	390371	42.94000	483.43000
Fuel_Price	1485	14.55396	2.54783	21613	10.07600	21.03300
CPI	1485	743.50637	189.74914	1104107	504.31539	1129
Unemployment	1485	34.66299	9.03514	51475	15.51600	71.56500

Pearson Correlation Coefficients, N = 1485
Prob > |r| under H0: Rho=0

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
Store	1.00000	0.00000 1.0000	-0.33065 <.0001	0.00000 1.0000	-0.02100 0.4188	0.04687 0.0710	-0.18834 <.0001	0.20116 <.0001
Date	0.00000 1.0000	1.00000	0.01263 0.6268	-0.03317 0.2014	0.14228 <.0001	0.61620 <.0001	0.08037 0.0019	-0.21141 <.0001
Weekly_Sales	-0.33065 <.0001	0.01263 0.6268	1.00000	0.05658 0.0292	0.03374 0.1938	0.15166 <.0001	0.03260 0.2093	0.00041 0.9875
Holiday_Flag	0.00000 1.0000	-0.03317 0.2014	0.05658 0.0292	1.00000	-0.32071 <.0001	-0.16177 <.0001	-0.02412 0.3531	0.00520 0.8413

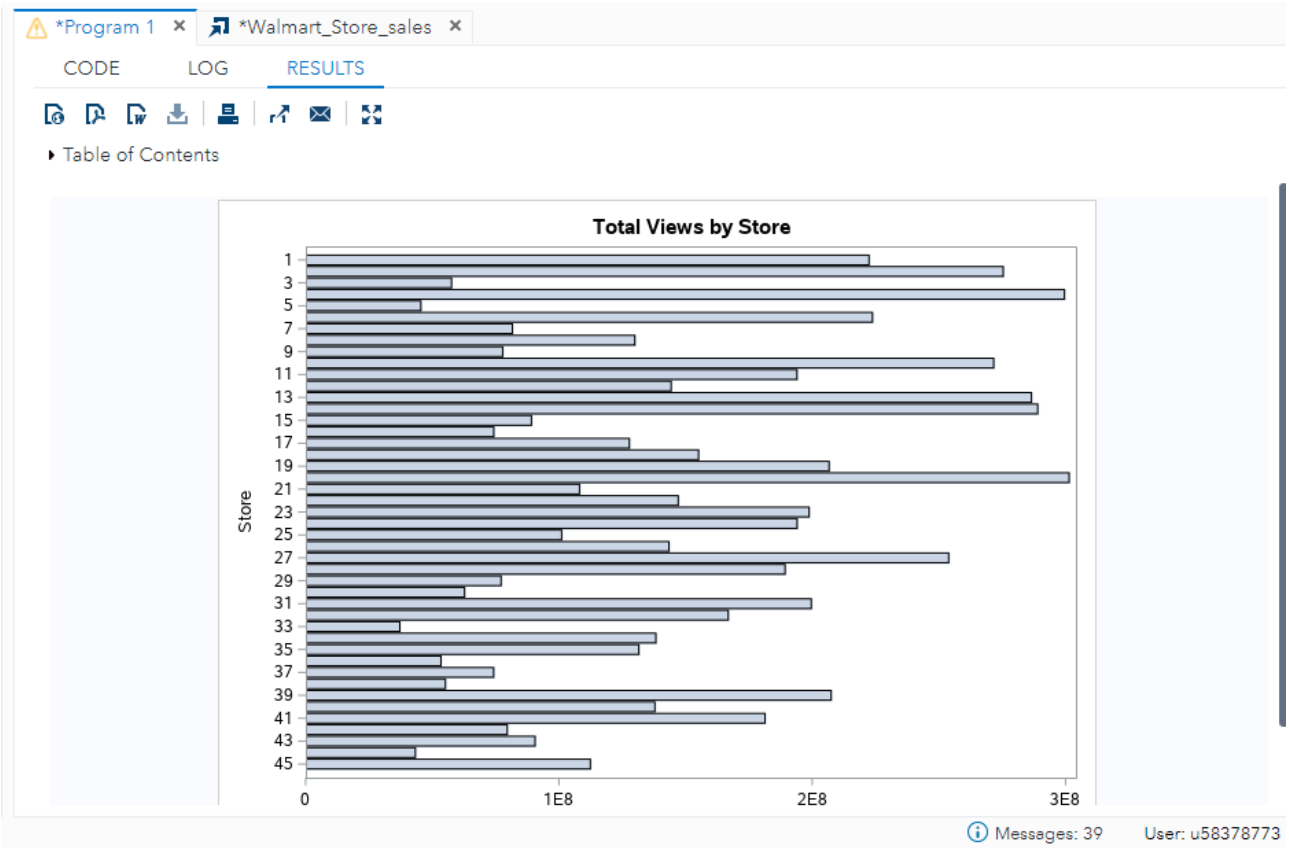
Messages: 38 User: u58378773

/* 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;
```

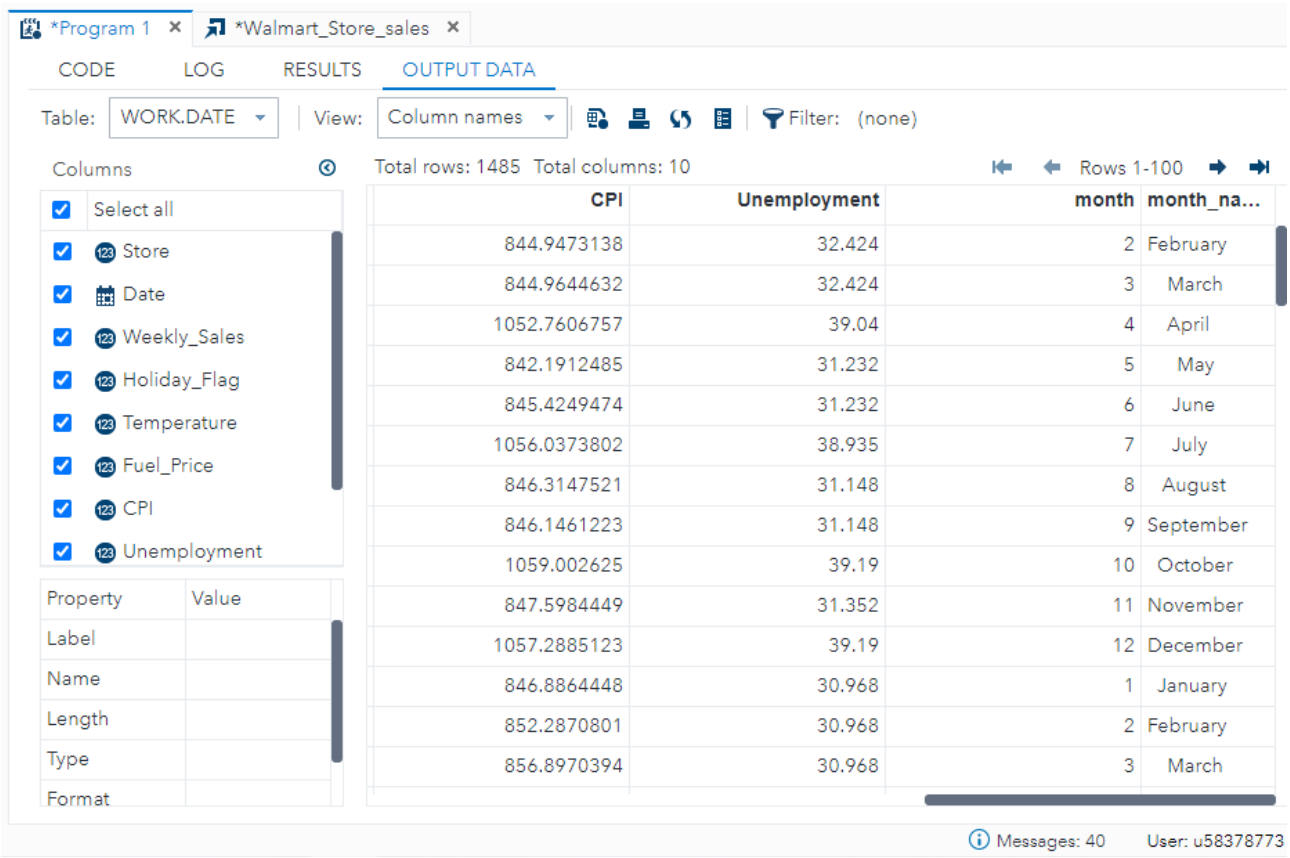
Output:



/* b) Clustered Bar Chart / Column Chart */

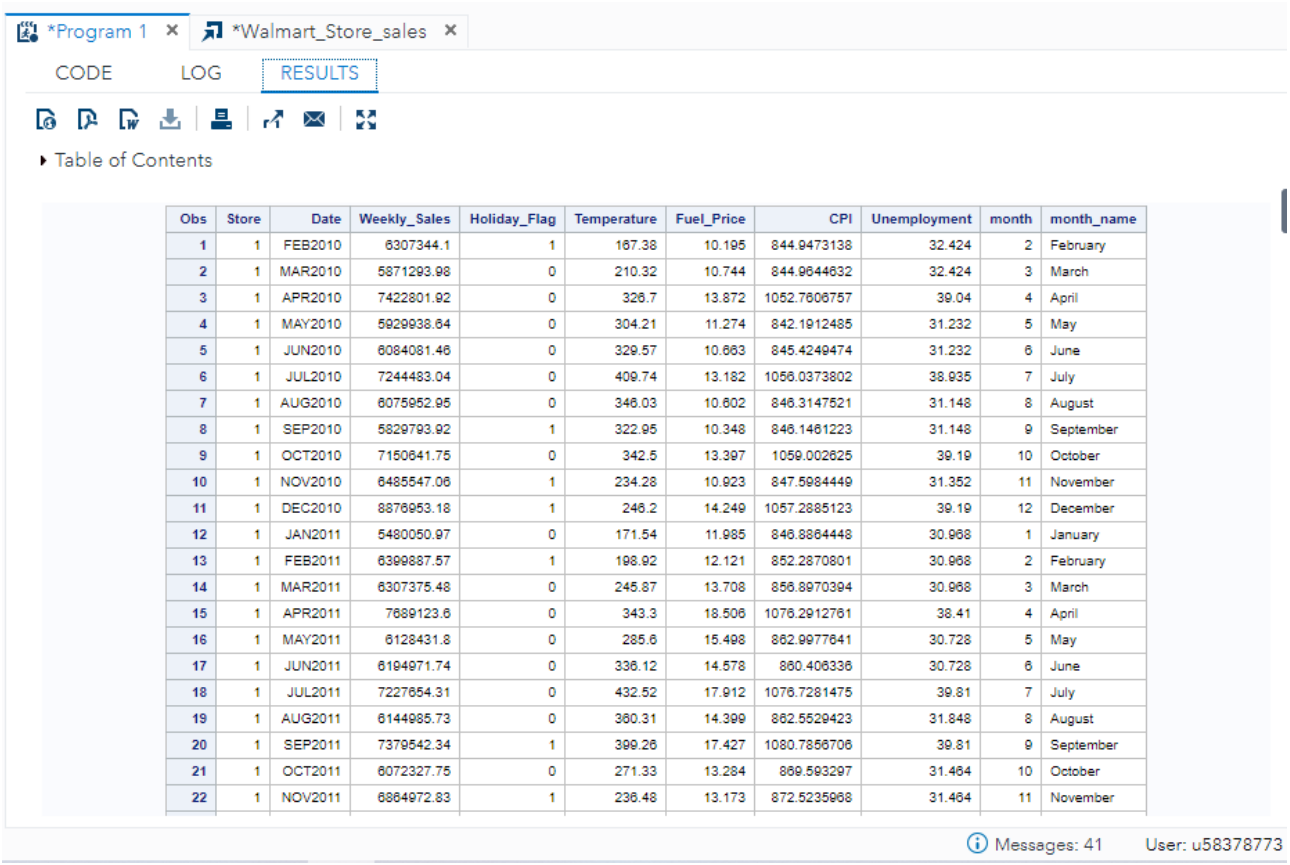
```
DATA date;
set work.monthly_sales;
month = month(Date);
month_name=PUT(Date,monname.);
put month_name= @;
RUN;
```

Output:



```
PROC PRINT DATA= date;
RUN;
```

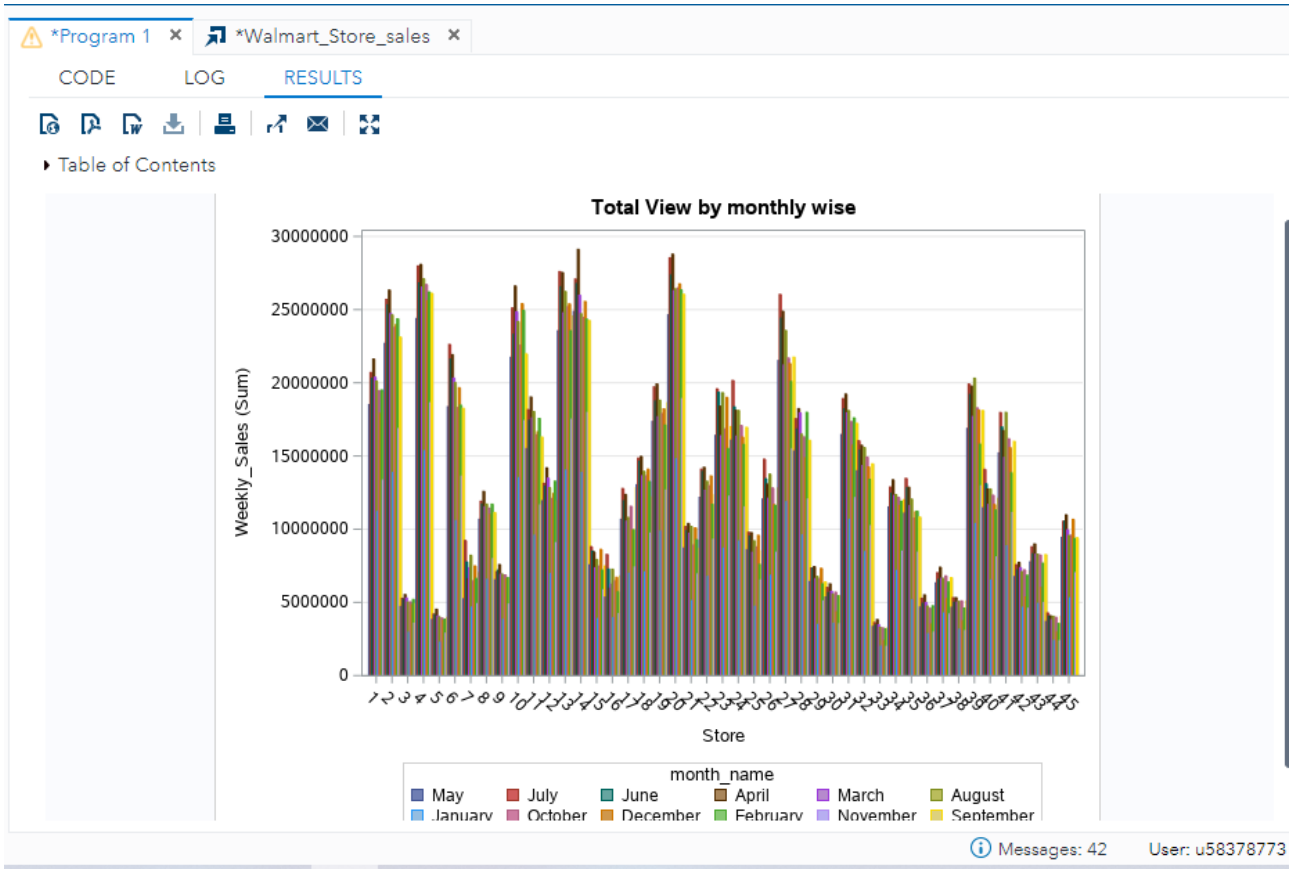
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	month	month_name
1	1	FEB2010	6307344.1	1	167.38	10.195	844.9473138	32.424	2	February
2	1	MAR2010	5871293.98	0	210.32	10.744	844.9644632	32.424	3	March
3	1	APR2010	7422801.92	0	326.7	13.872	1052.7606757	39.04	4	April
4	1	MAY2010	5929938.64	0	304.21	11.274	842.1912485	31.232	5	May
5	1	JUN2010	6084081.46	0	329.57	10.663	845.4249474	31.232	6	June
6	1	JUL2010	7244483.04	0	409.74	13.182	1056.0373802	38.935	7	July
7	1	AUG2010	6075952.95	0	346.03	10.602	846.3147521	31.148	8	August
8	1	SEP2010	5829793.92	1	322.95	10.348	846.1461223	31.148	9	September
9	1	OCT2010	7150641.75	0	342.5	13.397	1059.002625	39.19	10	October
10	1	NOV2010	6485547.06	1	234.28	10.923	847.5984449	31.352	11	November
11	1	DEC2010	8876953.18	1	246.2	14.249	1057.2885123	39.19	12	December
12	1	JAN2011	5480050.97	0	171.54	11.985	846.8864448	30.968	1	January
13	1	FEB2011	6399887.57	1	198.92	12.121	852.2870801	30.968	2	February
14	1	MAR2011	6307375.48	0	245.87	13.708	856.8970394	30.968	3	March
15	1	APR2011	7689123.6	0	343.3	18.506	1076.2912761	38.41	4	April
16	1	MAY2011	6128431.8	0	285.6	15.498	862.9977641	30.728	5	May
17	1	JUN2011	6194971.74	0	336.12	14.578	860.406336	30.728	6	June
18	1	JUL2011	7227654.31	0	432.52	17.912	1076.7281475	39.81	7	July
19	1	AUG2011	6144985.73	0	360.31	14.399	862.5529423	31.848	8	August
20	1	SEP2011	7379542.34	1	399.26	17.427	1080.7856706	39.81	9	September
21	1	OCT2011	6072327.75	0	271.33	13.284	869.593297	31.464	10	October
22	1	NOV2011	6864972.83	1	236.48	13.173	872.5235968	31.464	11	November

```
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;
```

Output:

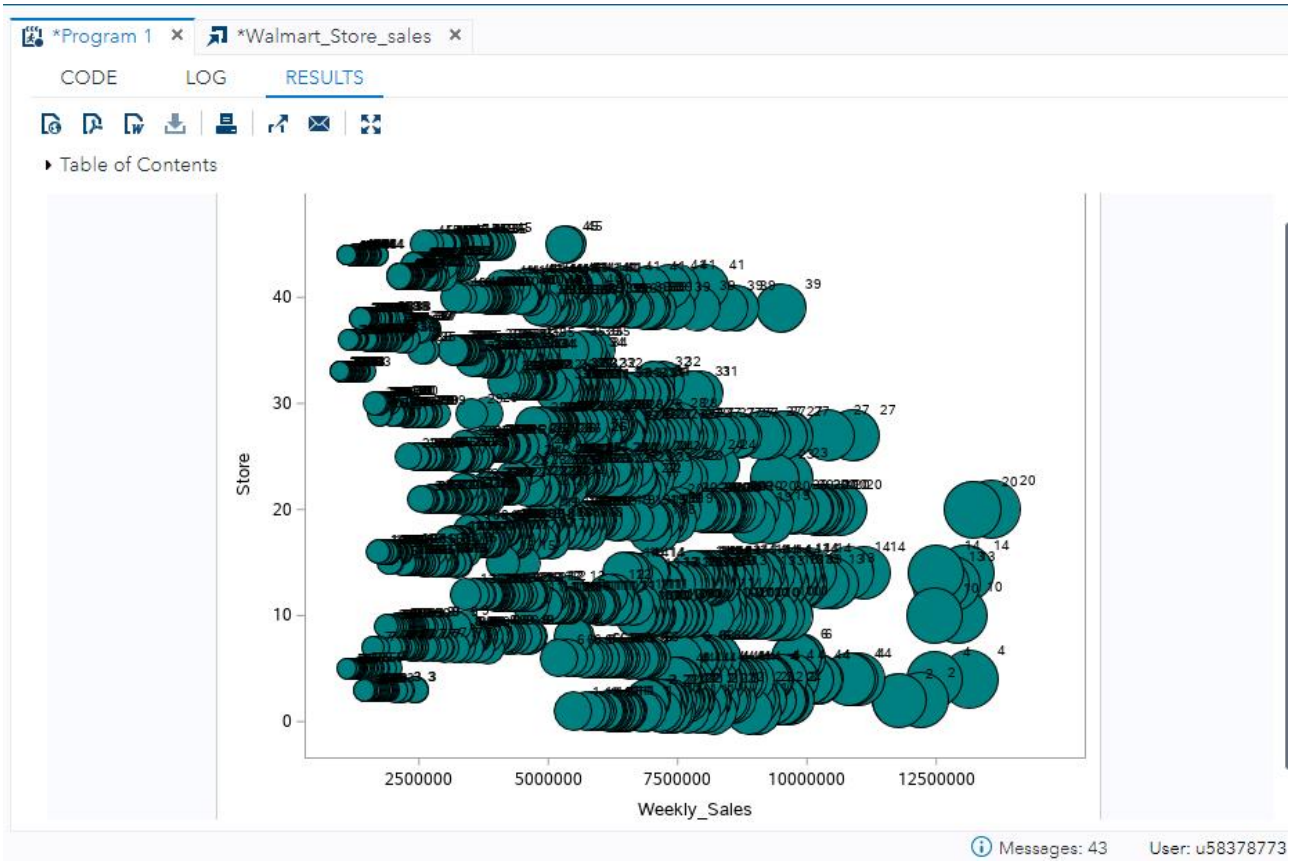


/* 2) Studying relationship */

/* a) Bubble Chart */

```
PROC SGPLOT DATA = work.monthly_sales;
bubble X=weekly_sales Y=store size= weekly_sales
/fillattrs=(color = teal) datalabel = store;
RUN;
```

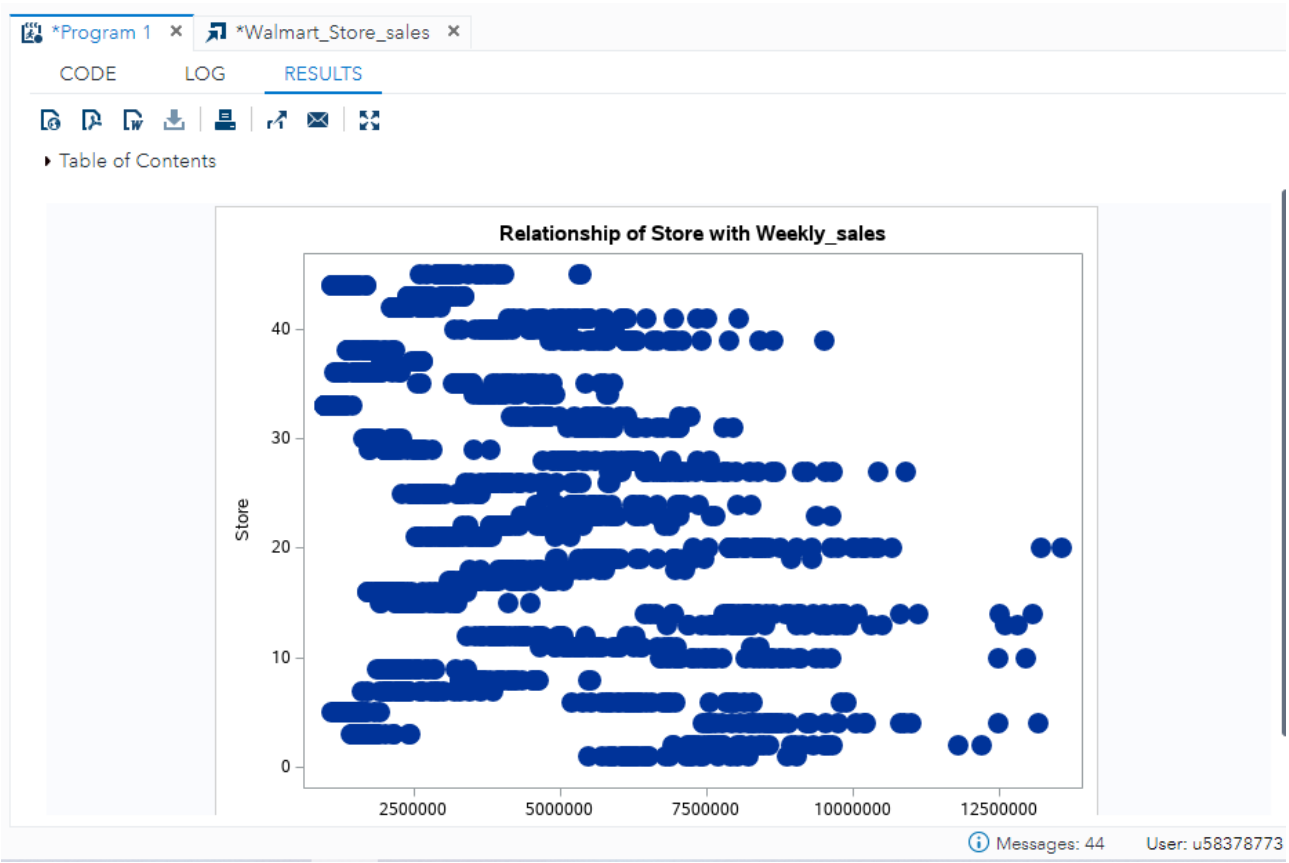
Output:



/* 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;
```

Output:

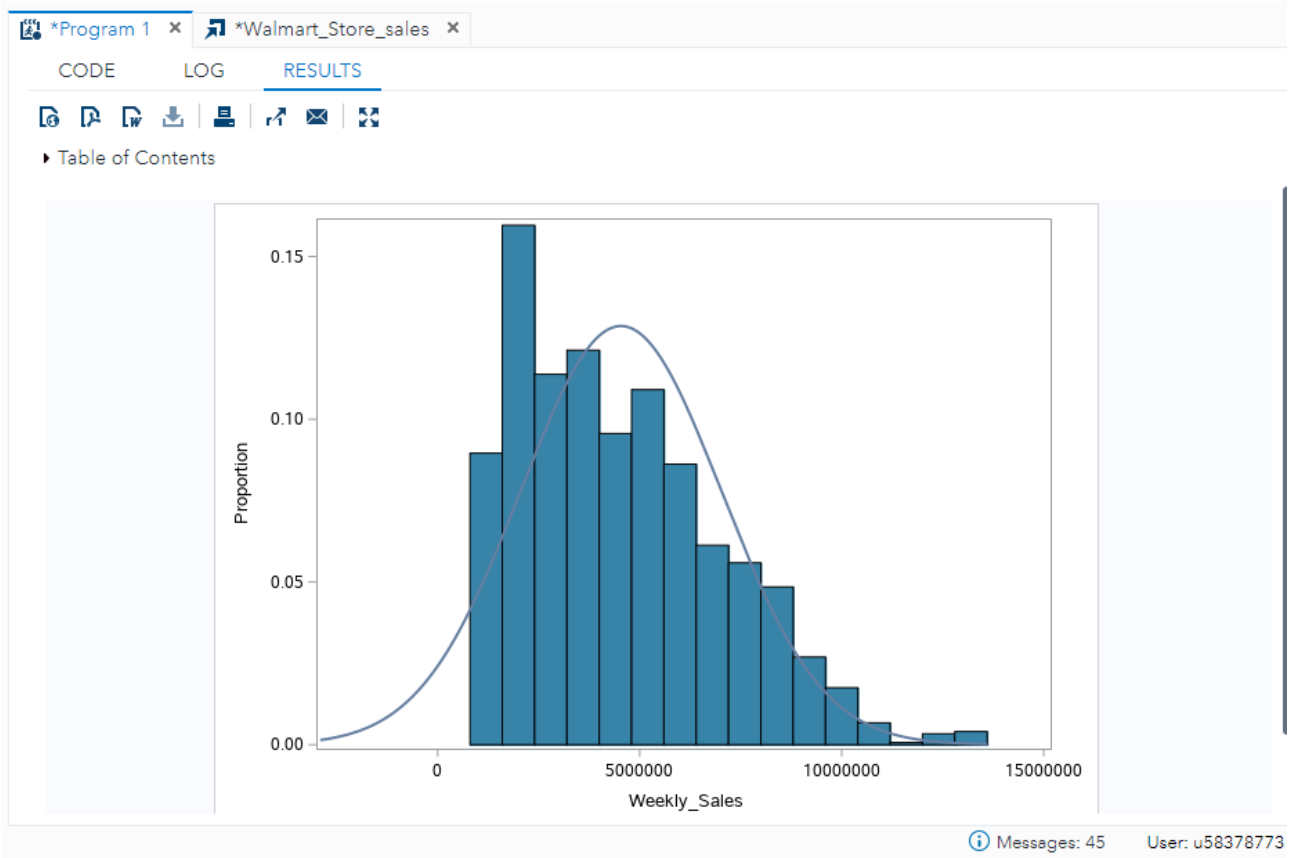


/* 3. Studying Distribution */

/* a) Histogram */

```
PROC SGPLOT DATA = work.monthly_sales;
histogram weekly_sales/fillattrs=(color = steel)scale = proportion;
density weekly_sales;
RUN;
```

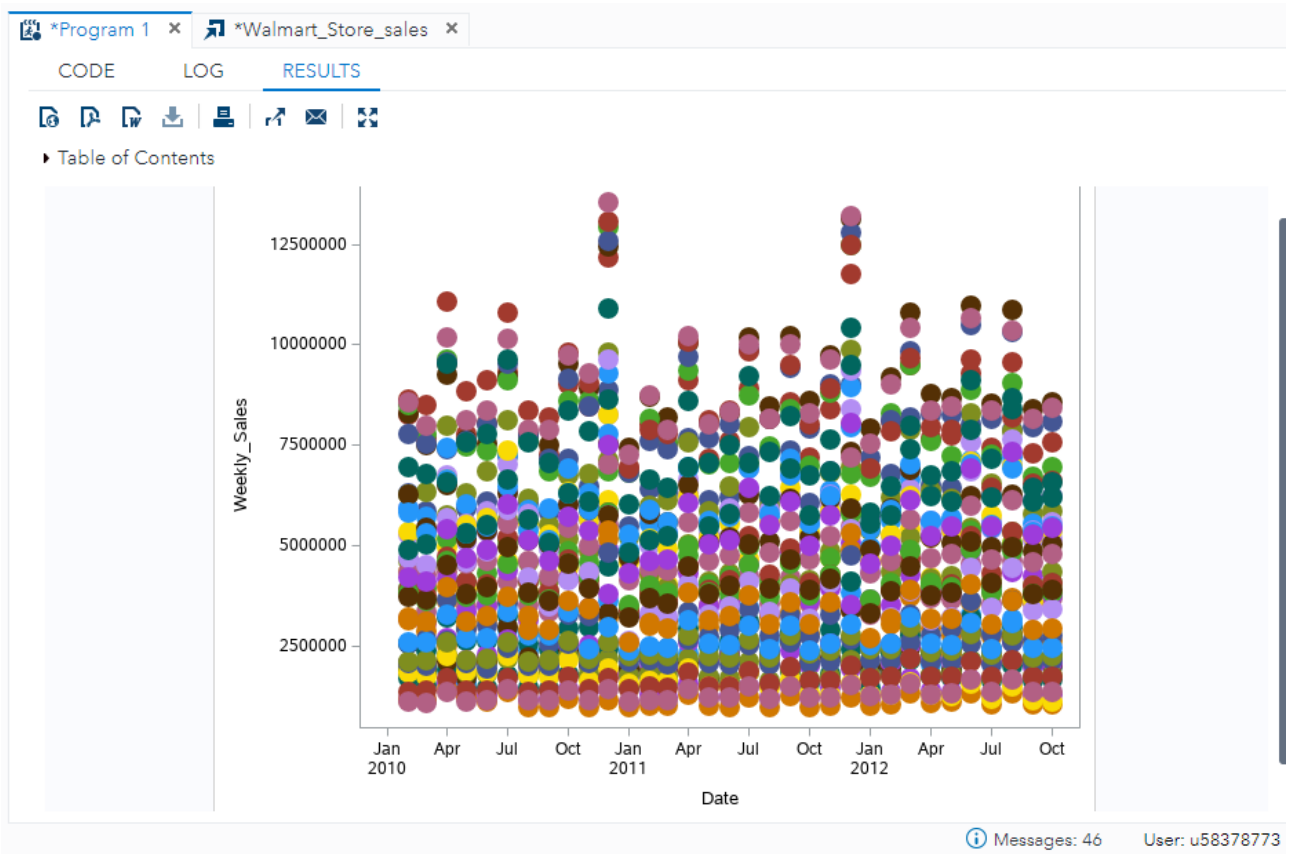
Output:



/* 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;
```

Output:



/* 4) Composition */

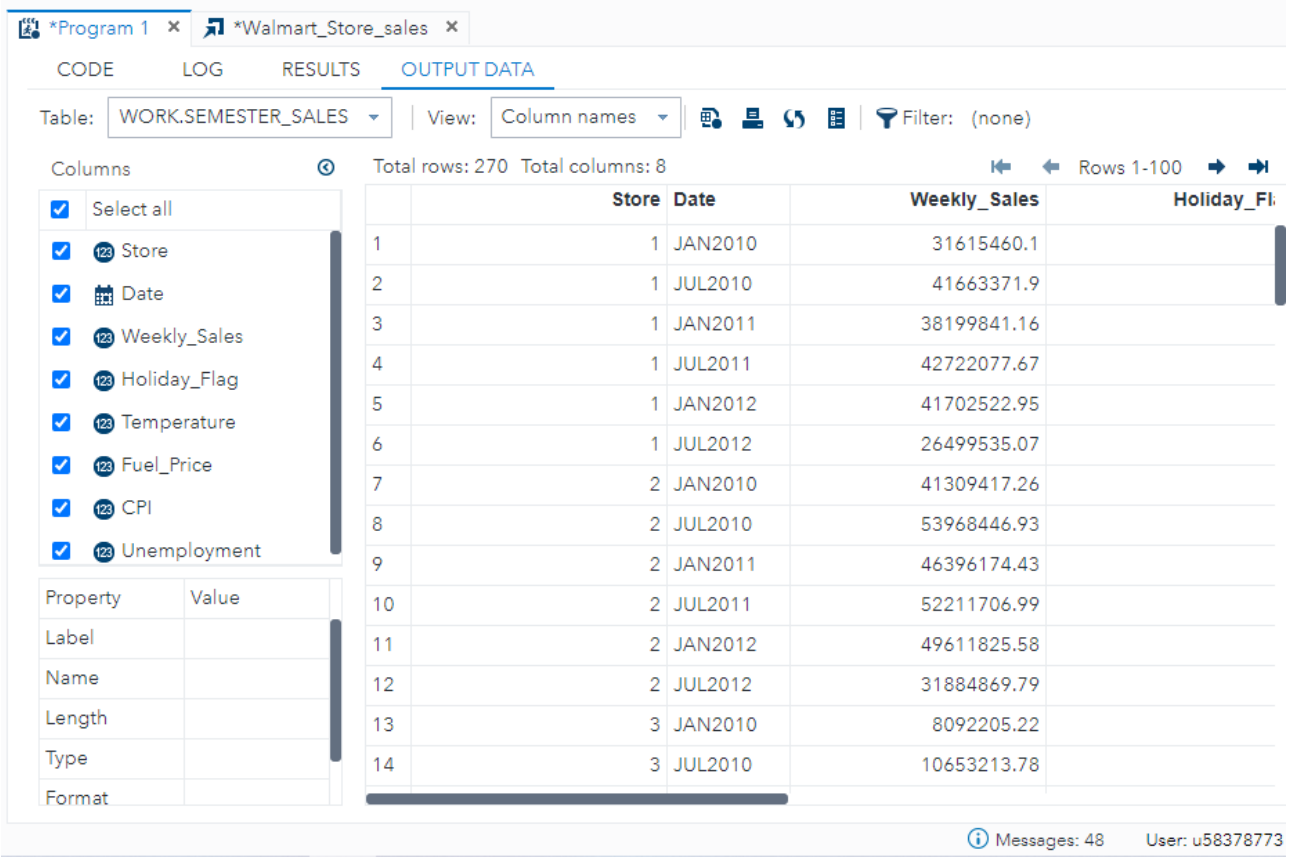
/* a) Stacked Column Chart: */

```
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;
```

Output:

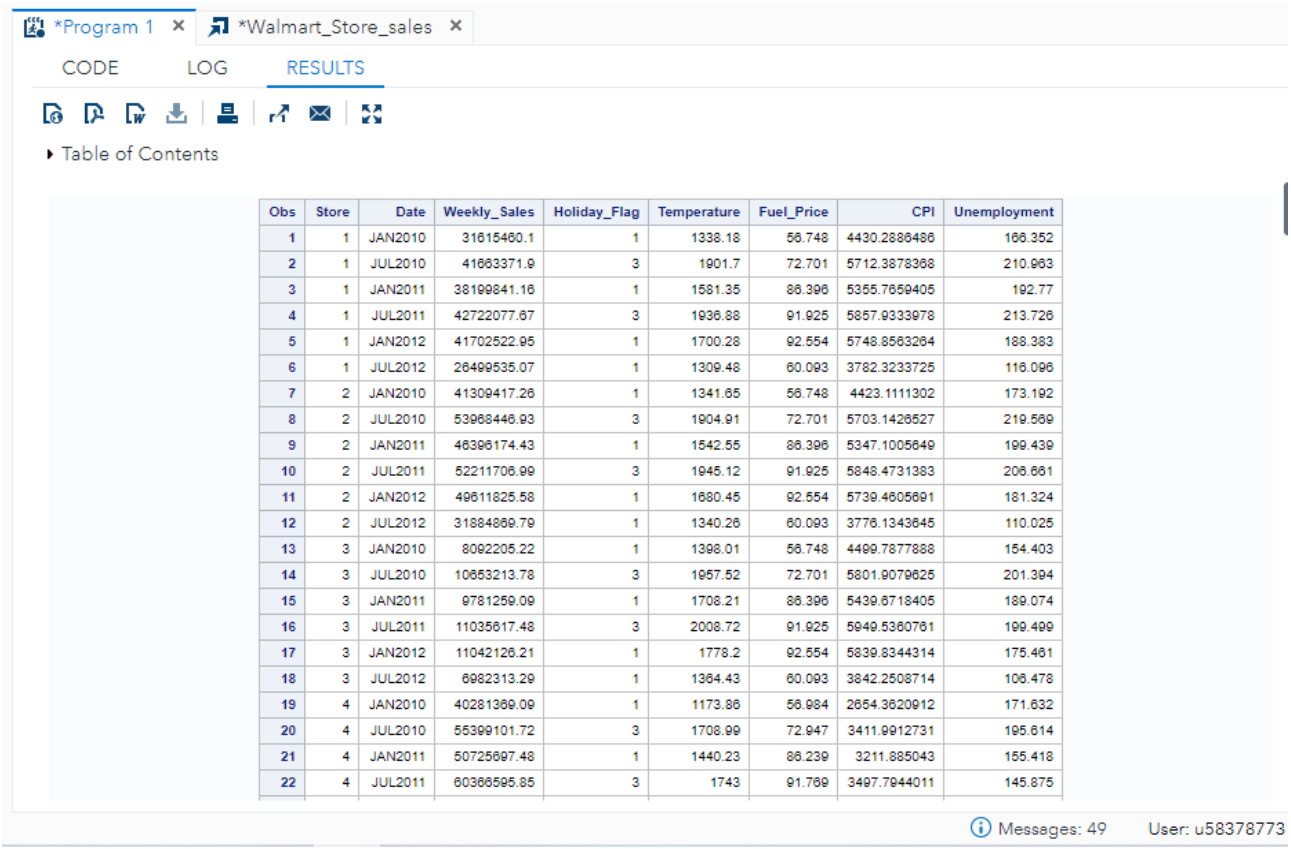
```
/* Semester view of sales in units */

/* 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;
```

Output:



```
PROC PRINT DATA= work.semester_sales;
RUN;
```

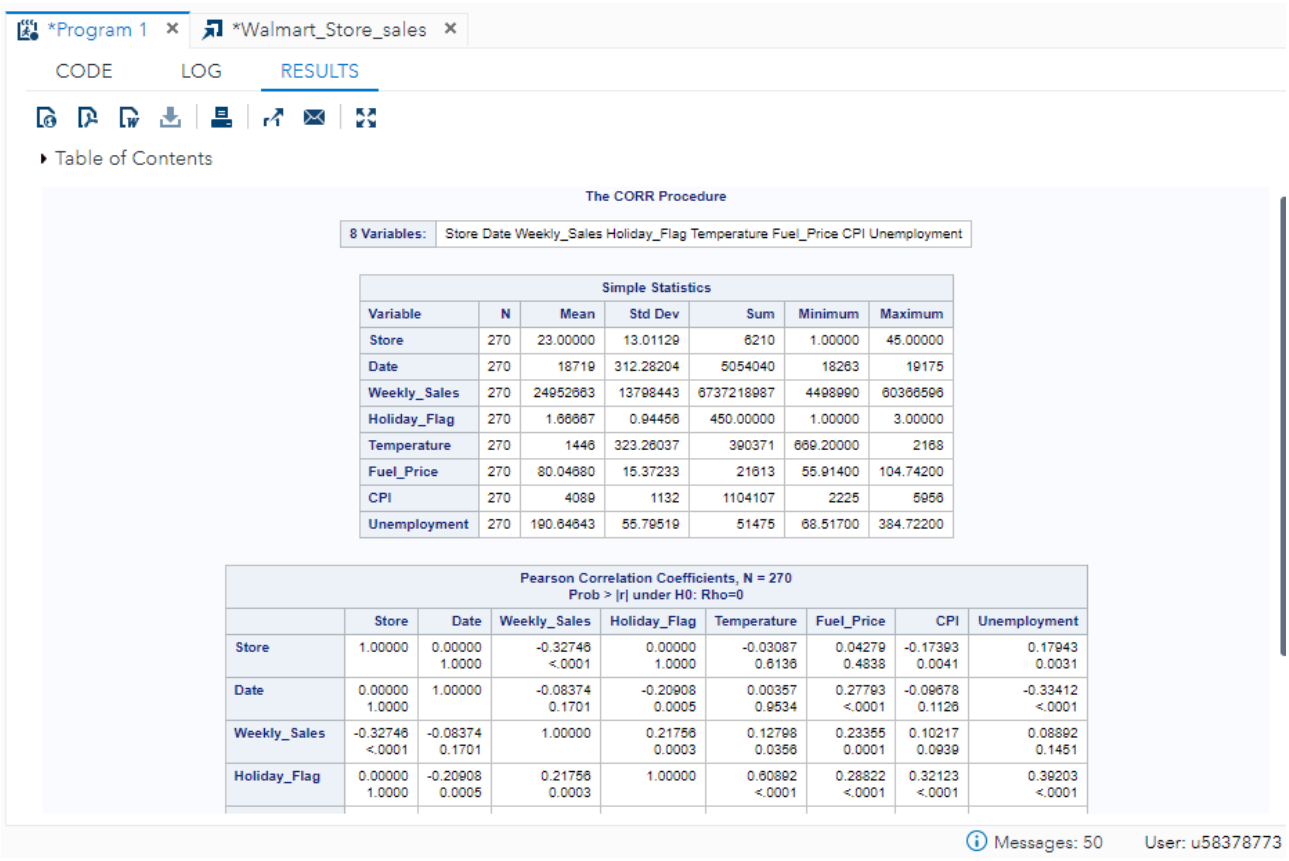
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	JAN2010	31615460.1	1	1338.18	56.748	4430.2886486	166.352
2	1	JUL2010	41663371.9	3	1901.7	72.701	5712.3878368	210.963
3	1	JAN2011	38199841.16	1	1581.35	86.396	5355.7659405	192.77
4	1	JUL2011	42722077.67	3	1936.88	91.925	5857.9333978	213.726
5	1	JAN2012	41702522.95	1	1700.28	92.554	5748.8563264	188.383
6	1	JUL2012	26499535.07	1	1309.48	60.093	3782.3233725	116.096
7	2	JAN2010	41309417.26	1	1341.65	56.748	4423.1111302	173.192
8	2	JUL2010	53968446.93	3	1904.91	72.701	5703.1426527	219.569
9	2	JAN2011	46396174.43	1	1542.55	86.396	5347.1005649	199.439
10	2	JUL2011	52211706.99	3	1945.12	91.925	5848.4731383	206.661
11	2	JAN2012	49611825.58	1	1680.45	92.554	5739.4605691	181.324
12	2	JUL2012	31884869.79	1	1340.26	60.093	3776.1343645	110.025
13	3	JAN2010	8092205.22	1	1398.01	56.748	4499.7877888	154.403
14	3	JUL2010	10653213.78	3	1957.52	72.701	5801.9079625	201.394
15	3	JAN2011	9781259.09	1	1708.21	86.396	5439.6718405	189.074
16	3	JUL2011	11035617.48	3	2008.72	91.925	5949.5360761	199.499
17	3	JAN2012	11042126.21	1	1778.2	92.554	5839.8344314	175.461
18	3	JUL2012	6982313.29	1	1364.43	60.093	3842.2508714	106.478
19	4	JAN2010	40281369.09	1	1173.86	56.984	2654.3620912	171.632
20	4	JUL2010	55399101.72	3	1708.99	72.947	3411.9912731	195.614
21	4	JAN2011	50725697.48	1	1440.23	86.239	3211.885043	155.418
22	4	JUL2011	60366595.85	3	1743	91.769	3497.7944011	145.875

```
/* Giving insights */

 /* Checking the correlation */

PROC CORR DATA= work.semester_sales;
RUN;
```

Output:



*Program 1 | *Walmart_Store_sales

CODE | LOG | RESULTS

Table of Contents

The CORR Procedure

8 Variables: Store Date Weekly_Sales Holiday_Flag Temperature Fuel_Price CPI Unemployment

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Store	270	23.00000	13.01129	6210	1.00000	45.00000
Date	270	18719	312.28204	5054040	18263	19175
Weekly_Sales	270	24952663	13798443	6737218987	4498990	60366596
Holiday_Flag	270	1.66667	0.94456	450.00000	1.00000	3.00000
Temperature	270	1446	323.26037	390371	669.20000	2168
Fuel_Price	270	80.04680	15.37233	21613	55.91400	104.74200
CPI	270	4089	1132	1104107	2225	5956
Unemployment	270	190.64643	55.79519	51475	68.51700	384.72200

Pearson Correlation Coefficients, N = 270
Prob > |r| under H0: Rho=0

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
Store	1.00000	0.00000 1.0000	-0.32746 <.0001	0.00000 1.0000	-0.03087 0.6136	0.04279 0.4838	-0.17393 0.0041	0.17943 0.0031
Date	0.00000 1.0000	1.00000	-0.08374 0.1701	-0.20908 0.0005	0.00357 0.9534	0.27793 <.0001	-0.09678 0.1126	-0.33412 <.0001
Weekly_Sales	-0.32746 <.0001	-0.08374 0.1701	1.00000	0.21756 0.0003	0.12798 0.0356	0.23355 0.0001	0.10217 0.0939	0.08892 0.1451
Holiday_Flag	0.00000 1.0000	-0.20908 0.0005	0.21756 0.0003	1.00000	0.60892 <.0001	0.28822 <.0001	0.32123 <.0001	0.39203 <.0001

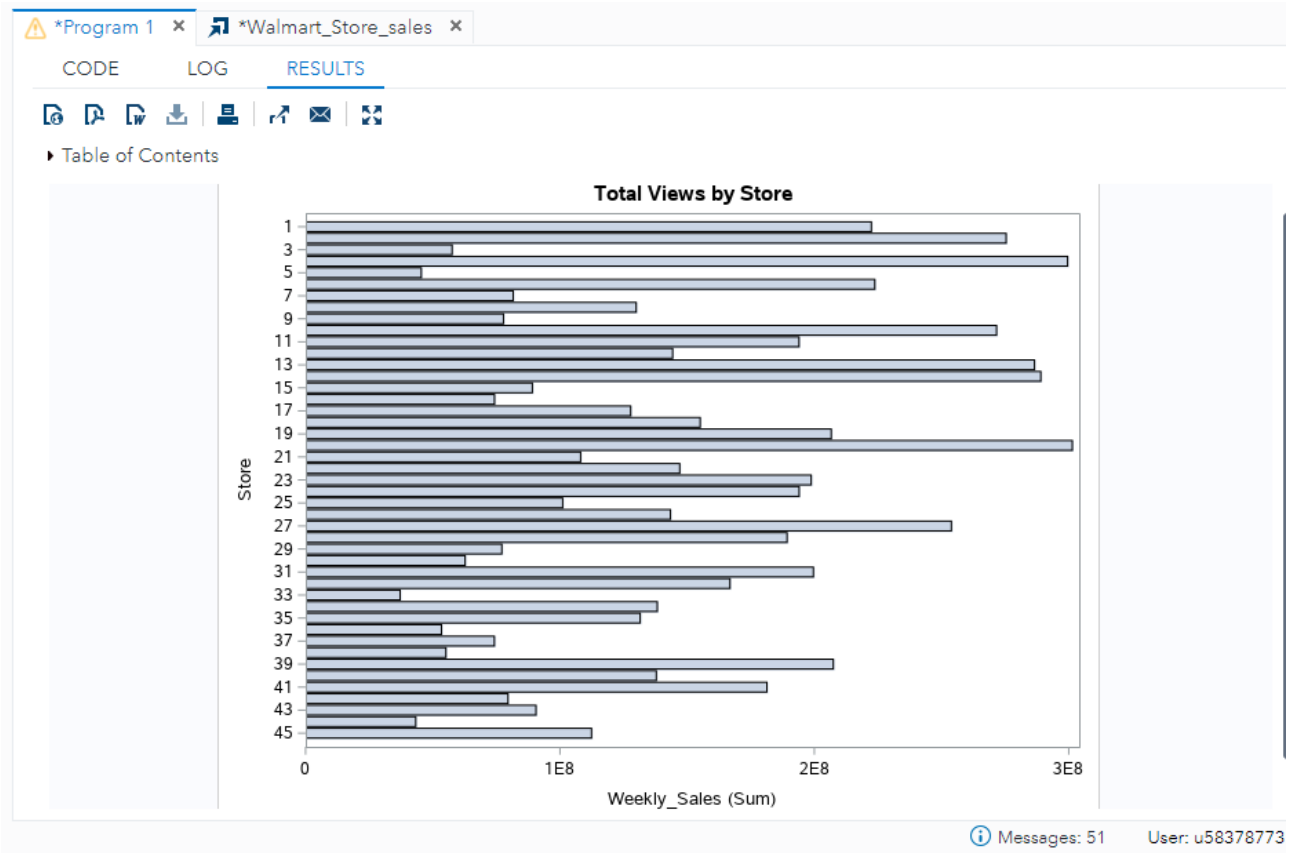
Messages: 50 User: u58378773

/* 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;
```

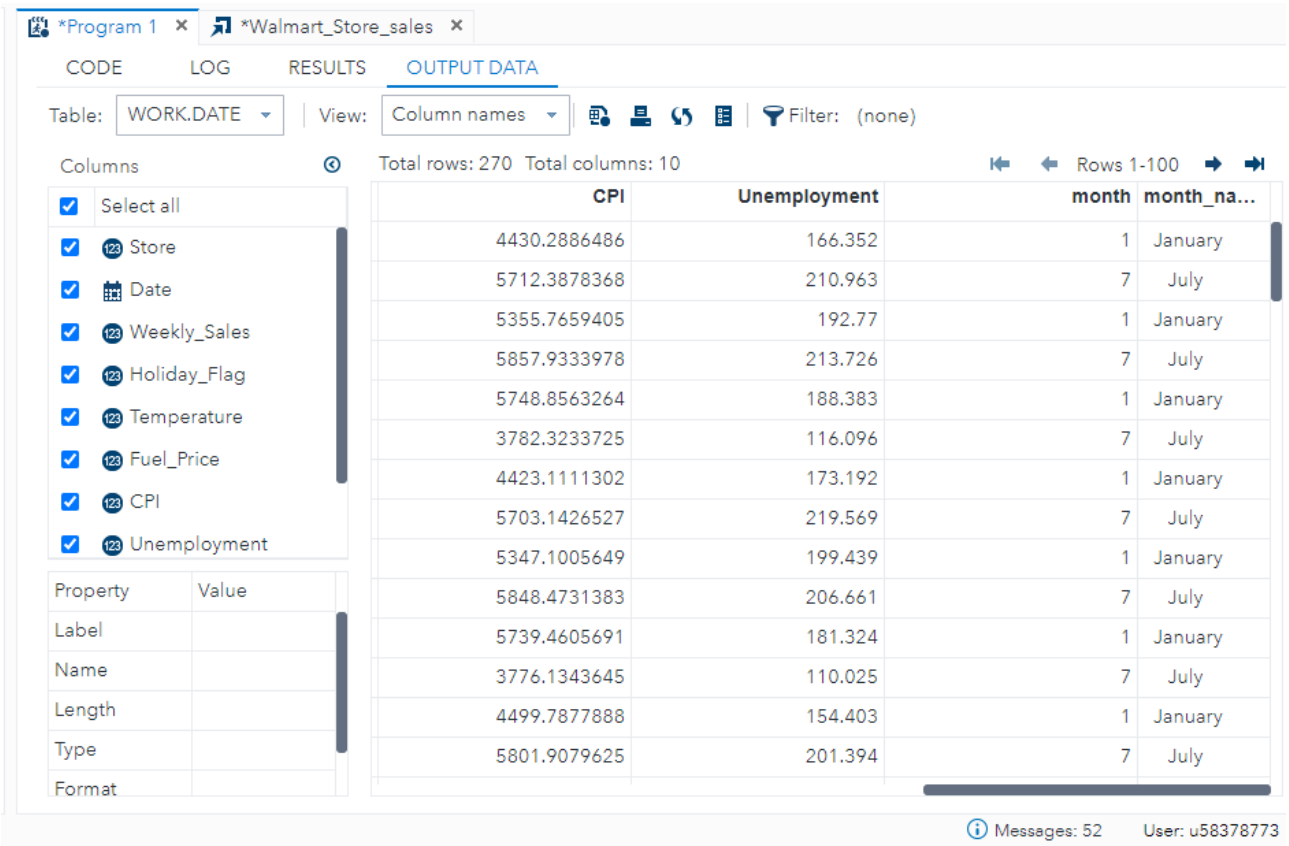
Output:



/* b) Clustered Bar Chart / Column Chart */

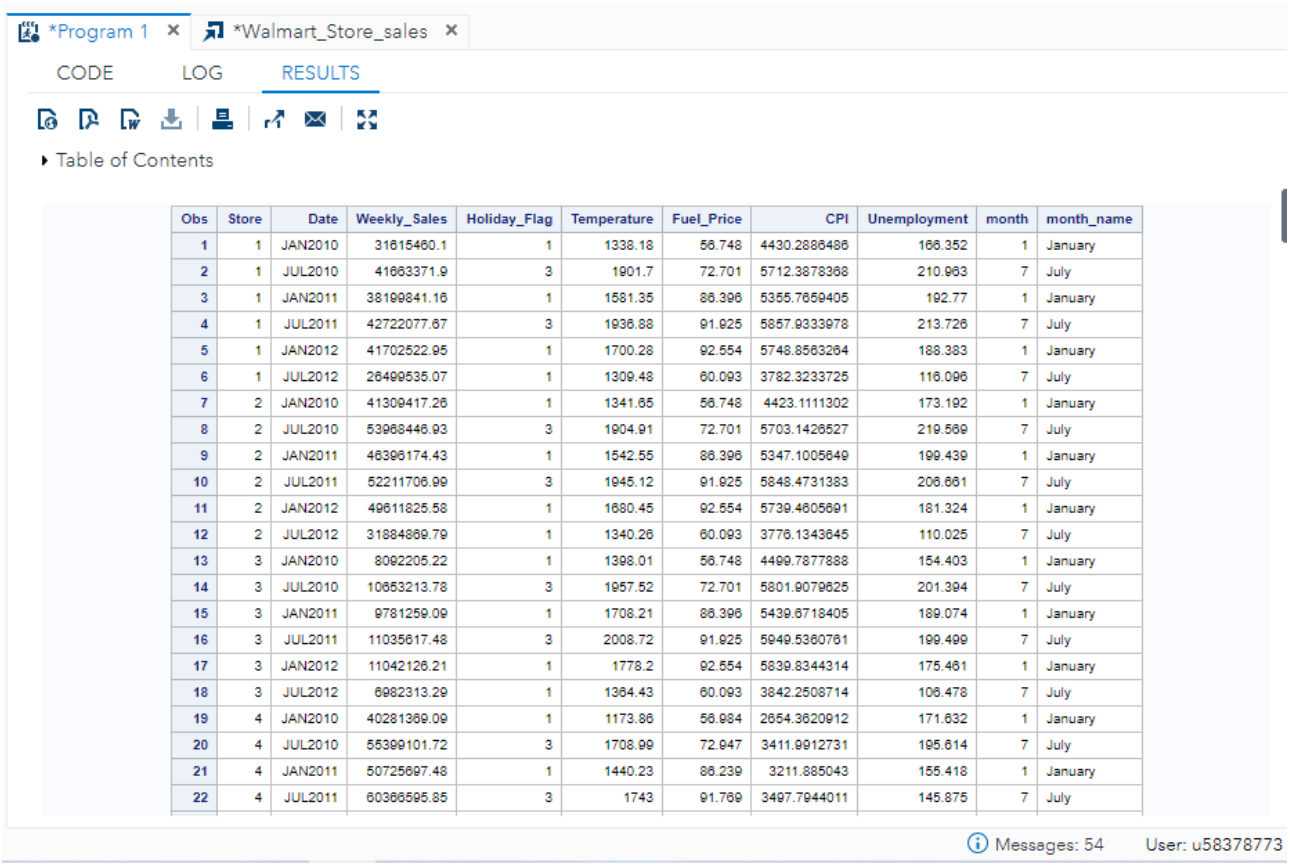
```
DATA date;
set work.semester_sales; month = month(Date);
month_name=PUT(Date,monname.);
put month_name= @;
RUN;
```

Output:



```
PROC PRINT DATA= date;
RUN;
```

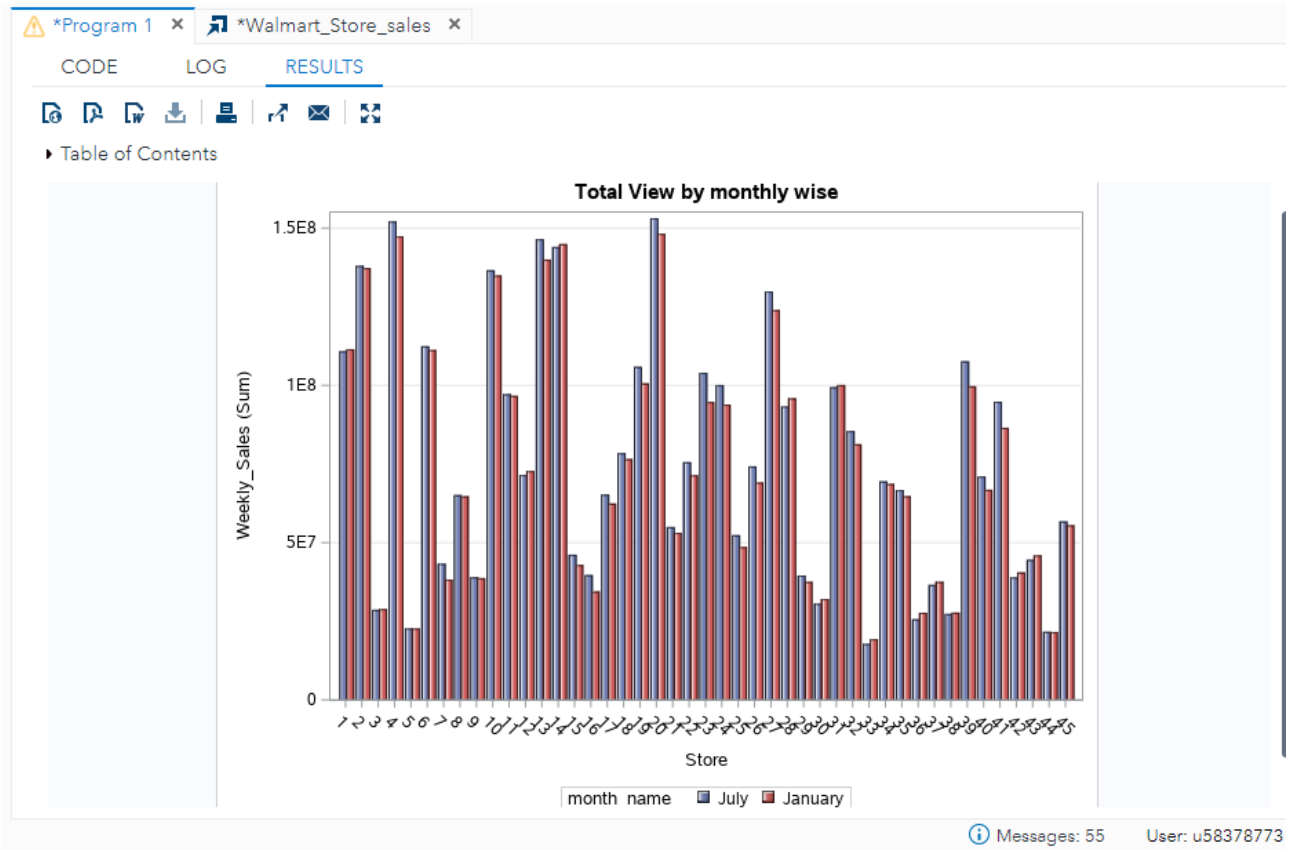
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	month	month_name
1	1	JAN2010	31615460.1	1	1338.18	56.748	4430.2886486	166.352	1	January
2	1	JUL2010	41663371.9	3	1901.7	72.701	5712.3878368	210.963	7	July
3	1	JAN2011	38199841.16	1	1581.35	86.396	5355.7659405	192.77	1	January
4	1	JUL2011	42722077.67	3	1936.88	91.925	5857.9333978	213.726	7	July
5	1	JAN2012	41702522.95	1	1700.28	92.554	5748.8563264	188.383	1	January
6	1	JUL2012	26499535.07	1	1309.48	60.093	3782.3233725	116.096	7	July
7	2	JAN2010	41309417.26	1	1341.65	56.748	4423.1111302	173.192	1	January
8	2	JUL2010	53968446.93	3	1904.91	72.701	5703.1426527	219.569	7	July
9	2	JAN2011	46396174.43	1	1542.55	86.396	5347.1005649	199.439	1	January
10	2	JUL2011	52211706.99	3	1945.12	91.925	5848.4731383	206.661	7	July
11	2	JAN2012	49611825.58	1	1680.45	92.554	5739.4605691	181.324	1	January
12	2	JUL2012	31884869.79	1	1340.26	60.093	3776.1343645	110.025	7	July
13	3	JAN2010	8092205.22	1	1398.01	56.748	4499.7877888	154.403	1	January
14	3	JUL2010	10653213.78	3	1957.52	72.701	5801.9079625	201.394	7	July
15	3	JAN2011	9781259.09	1	1708.21	86.396	5439.6718405	189.074	1	January
16	3	JUL2011	11035617.48	3	2008.72	91.925	5949.5360761	199.499	7	July
17	3	JAN2012	11042126.21	1	1778.2	92.554	5839.8344314	175.461	1	January
18	3	JUL2012	6982313.29	1	1364.43	60.093	3842.2508714	106.478	7	July
19	4	JAN2010	40281369.09	1	1173.86	56.984	2654.3620912	171.632	1	January
20	4	JUL2010	55399101.72	3	1708.99	72.947	3411.9912731	195.614	7	July
21	4	JAN2011	50725697.48	1	1440.23	86.239	3211.885043	155.418	1	January
22	4	JUL2011	60366595.85	3	1743	91.769	3497.7944011	145.875	7	July

```
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;
```

Output:

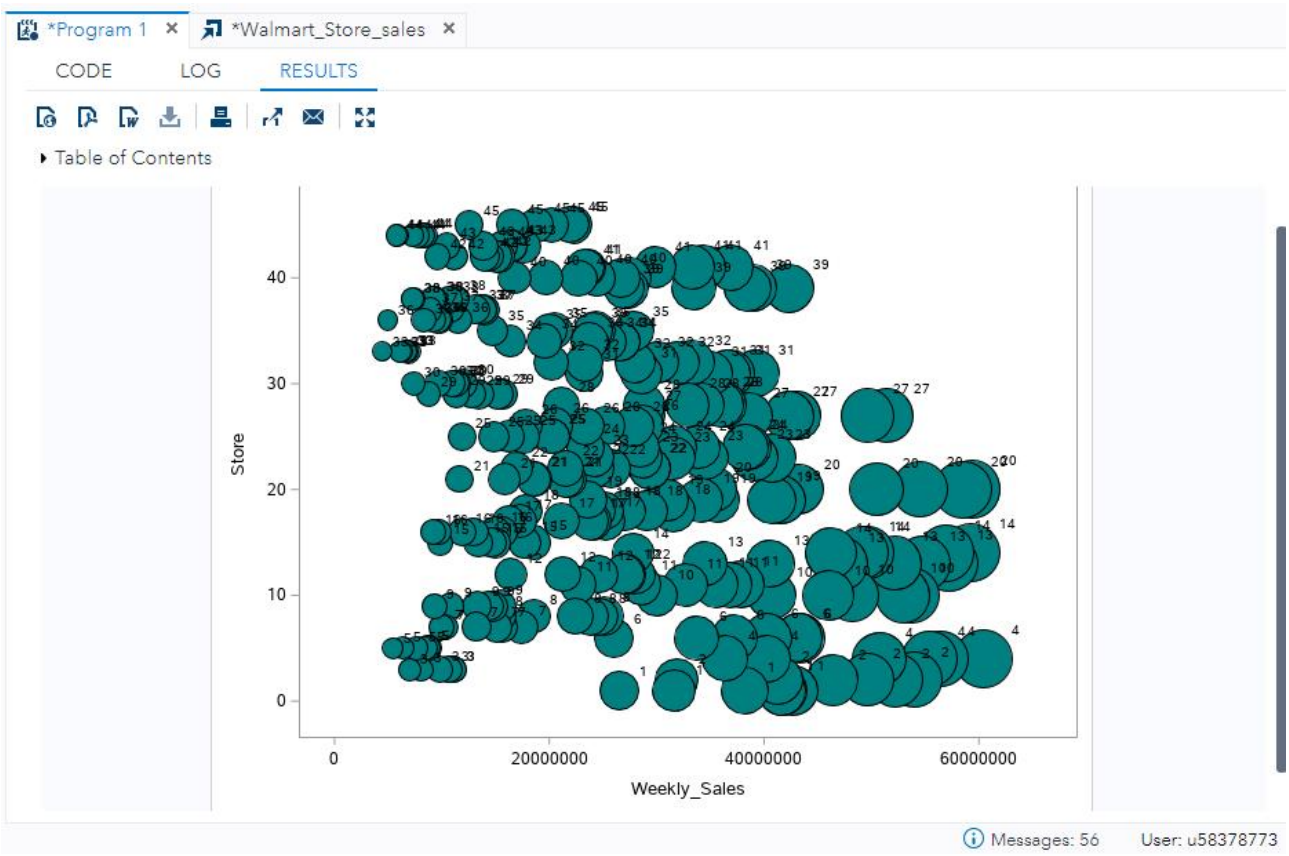


/* 2) Studying relationship */

/* a) Bubble Chart */

```
PROC SGPLOT DATA = work.semester_sales;
bubble X=weekly_sales Y=store size= weekly_sales
/fillattrs=(color = teal) datalabel = store;
RUN;
```

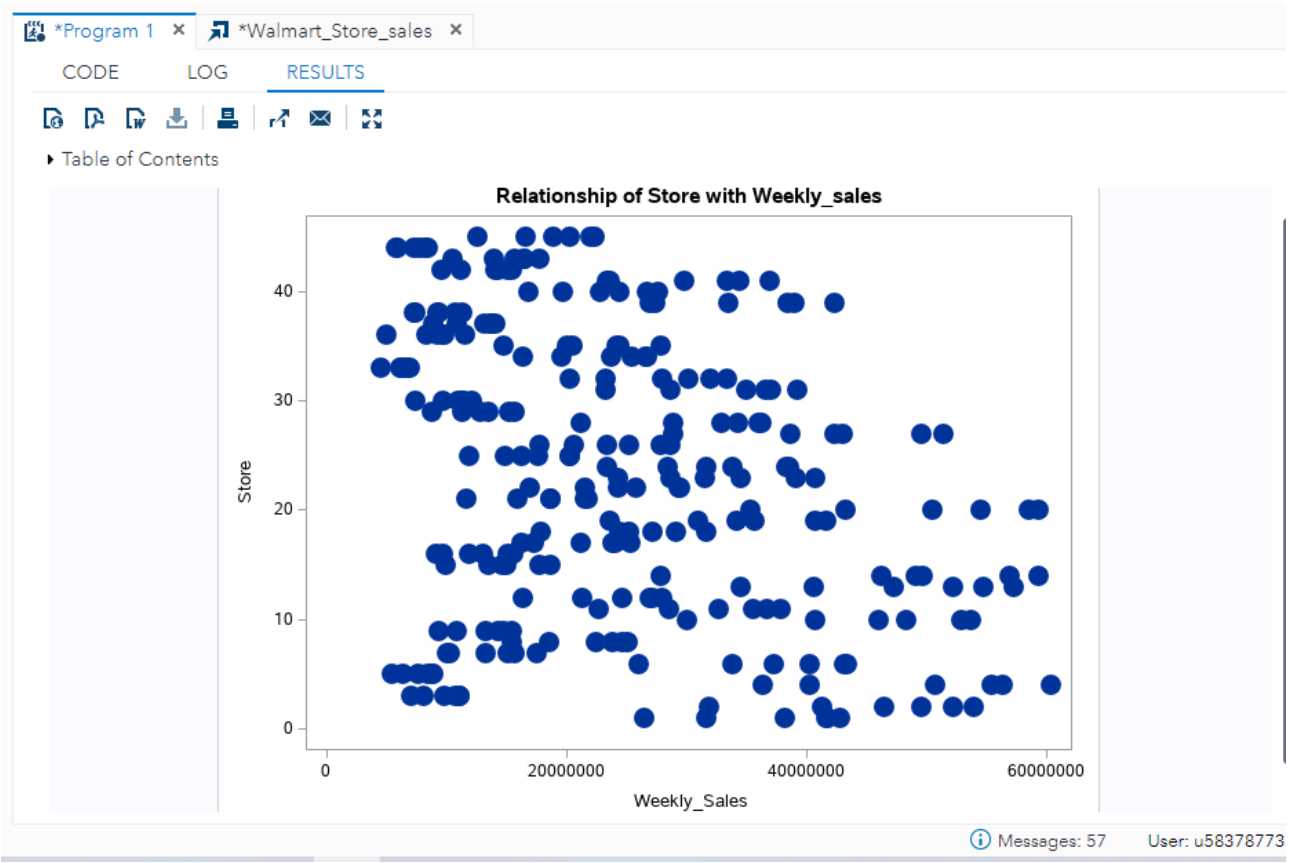
Output:



/* 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;
```

Output:

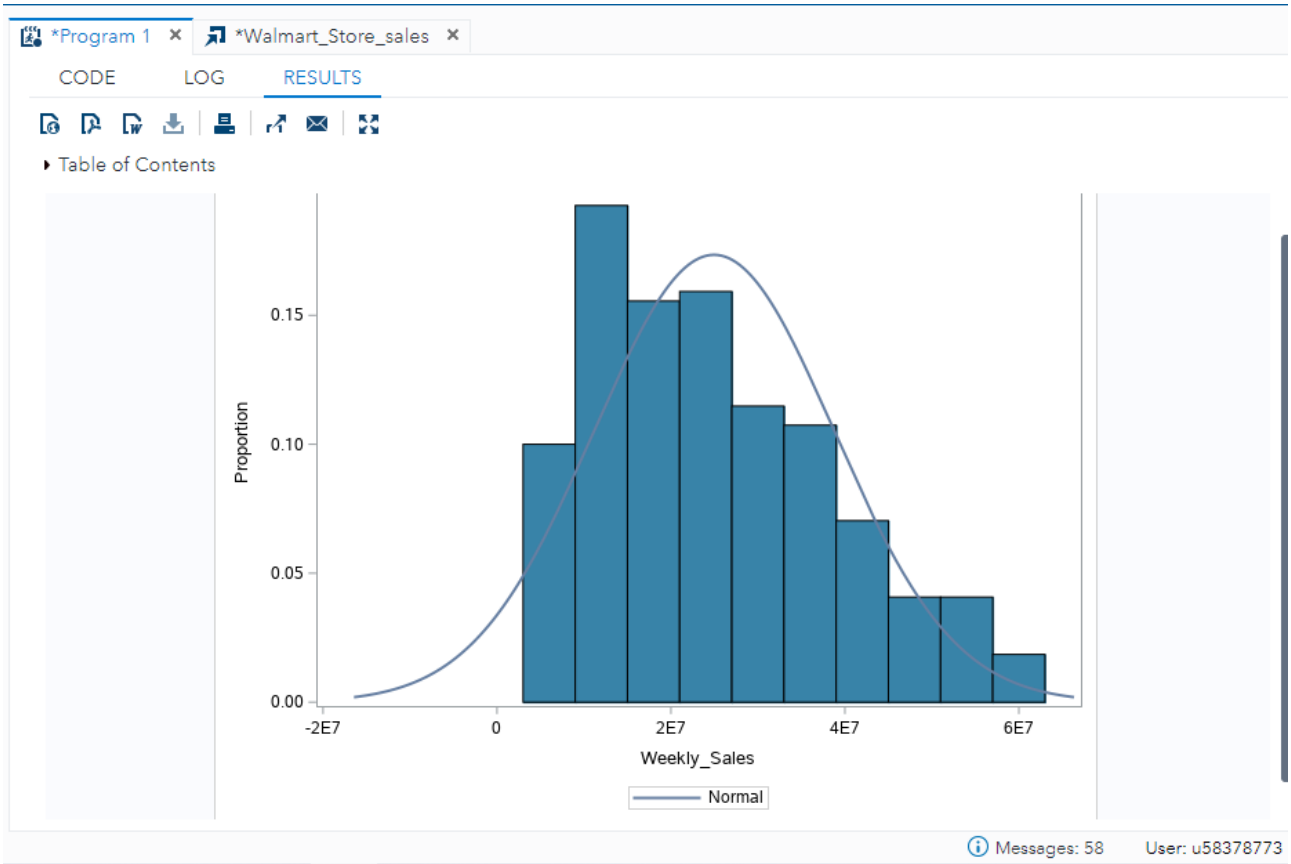


/* 3. Studying Distribution */

/* a) Histogram */

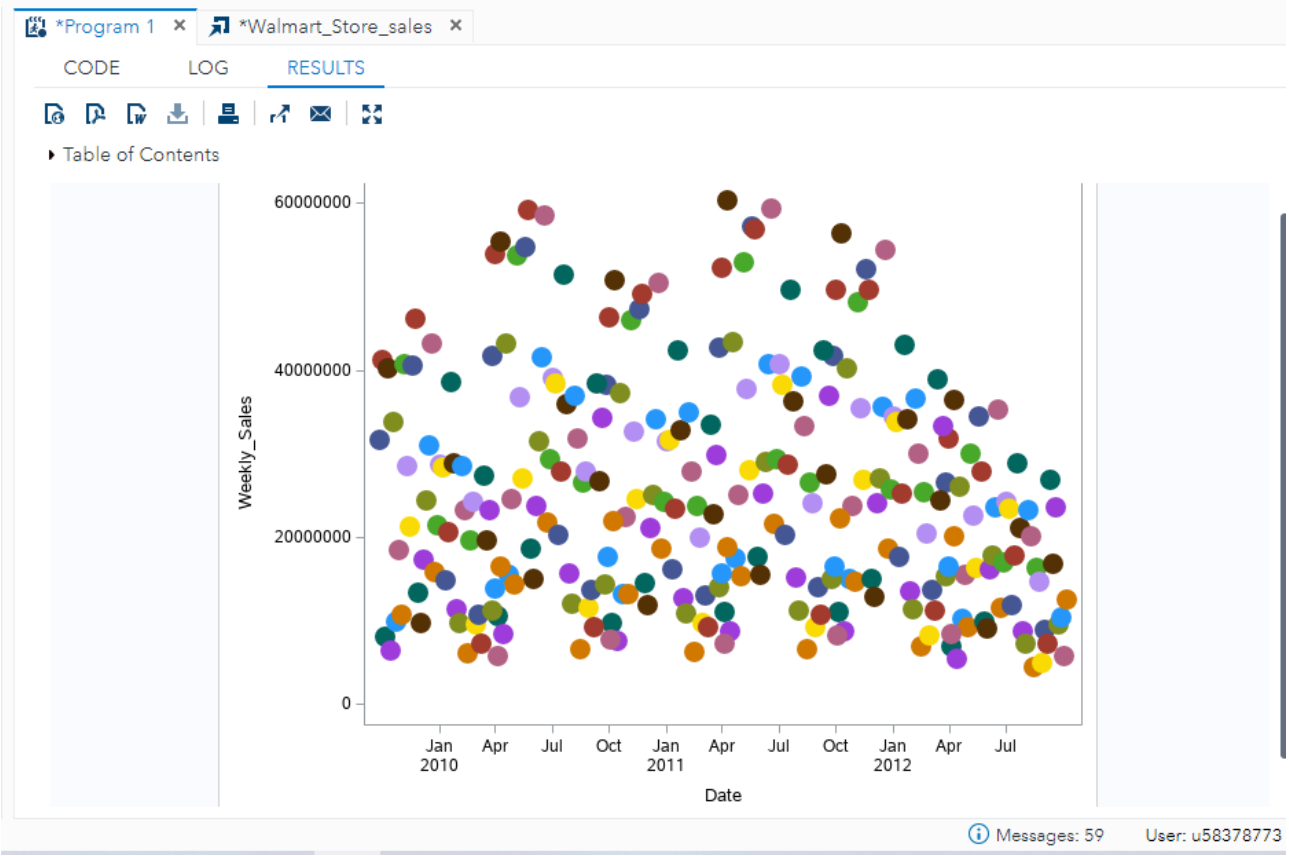
```
PROC SGPLOT DATA = work.semester_sales;
histogram weekly_sales/fillattrs=(color = steel)scale = proportion;
density weekly_sales;
RUN;
```

Output:



```
/* 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;
```

Output:

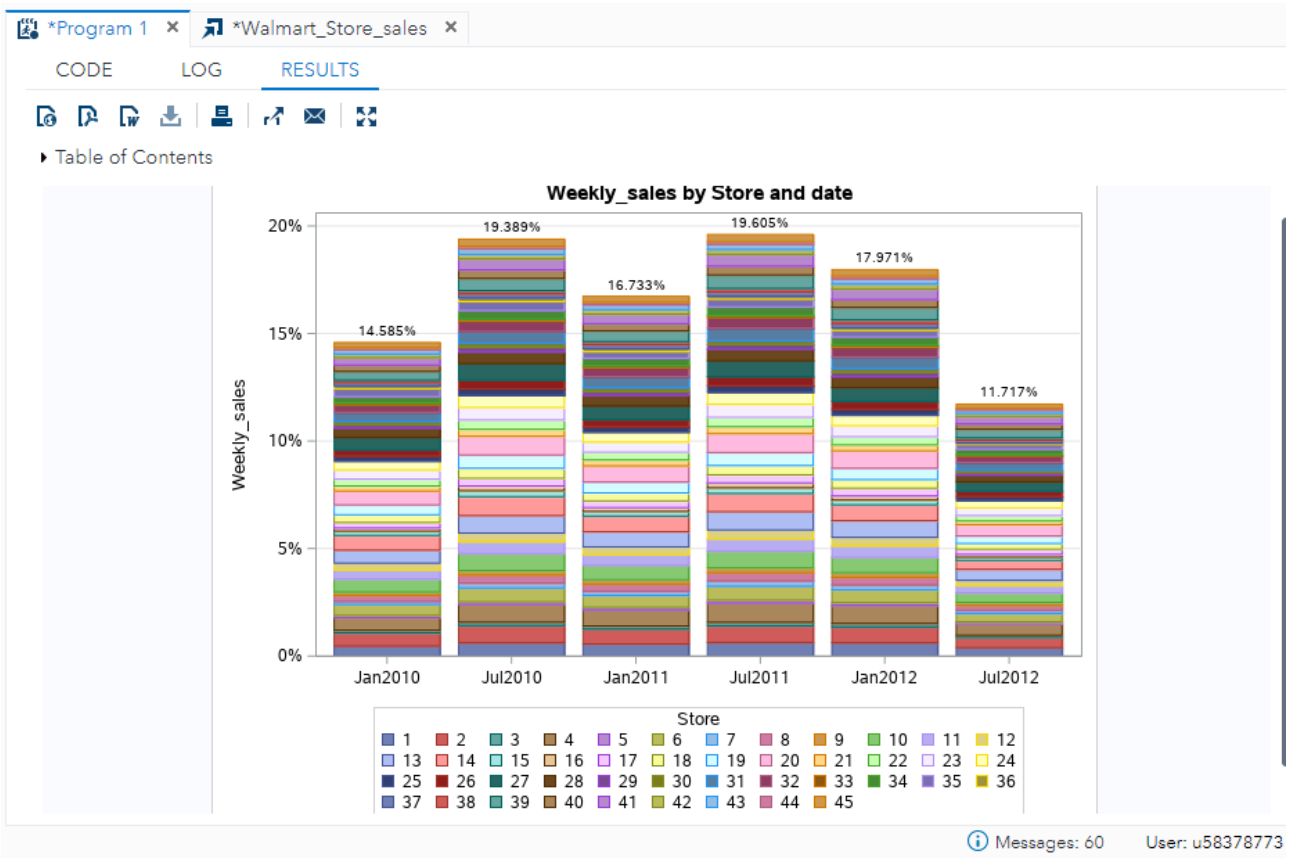


/* 4) Composition */

/* a) Stacked Column Chart: */

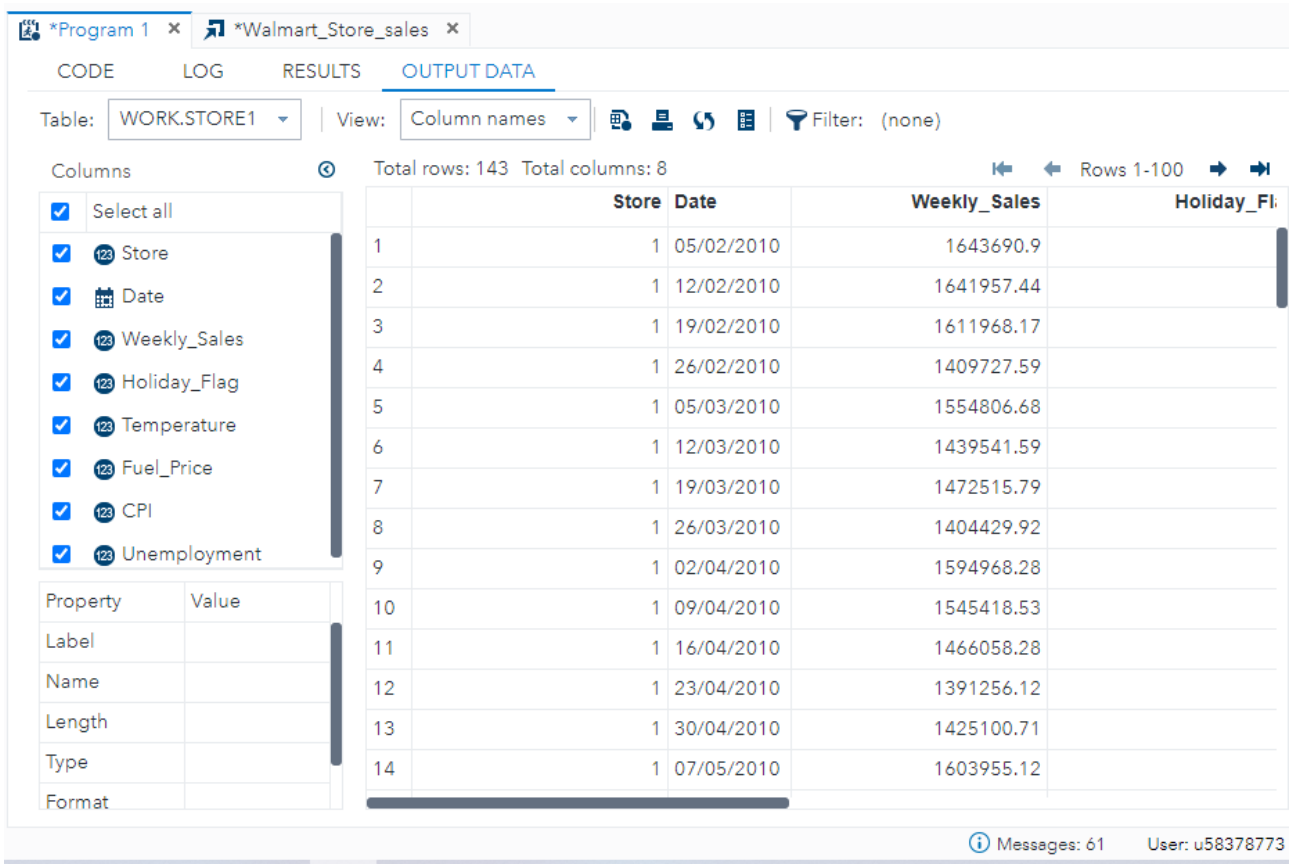
```
PROC SGPLOT DATA= work.semester_sales;
title 'Weekly_sales by Store and date';
vbar date / response= weekly_sales group= store stat=percent datalabel;
xaxis display=(nolabel);
yaxis  grid    label='Weekly_sales';
RUN;
```

Output:



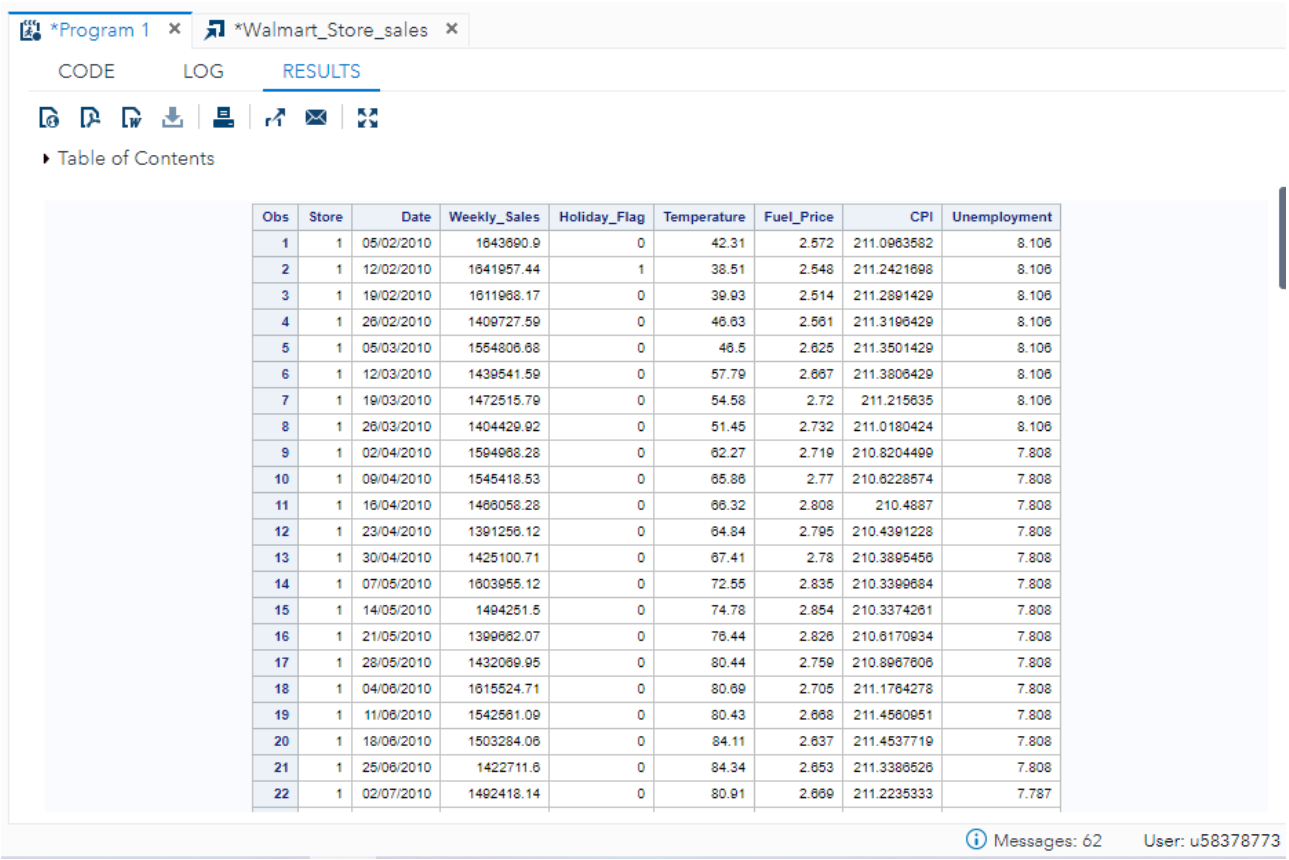
```
/* For Store 1 – Build prediction models to forecast demand */

/* Store-1 data */

DATA store1;
set work.walmart;
where  store  =  1;
RUN;
```

Output:



```
PROC PRINT DATA= store1;
RUN;
```

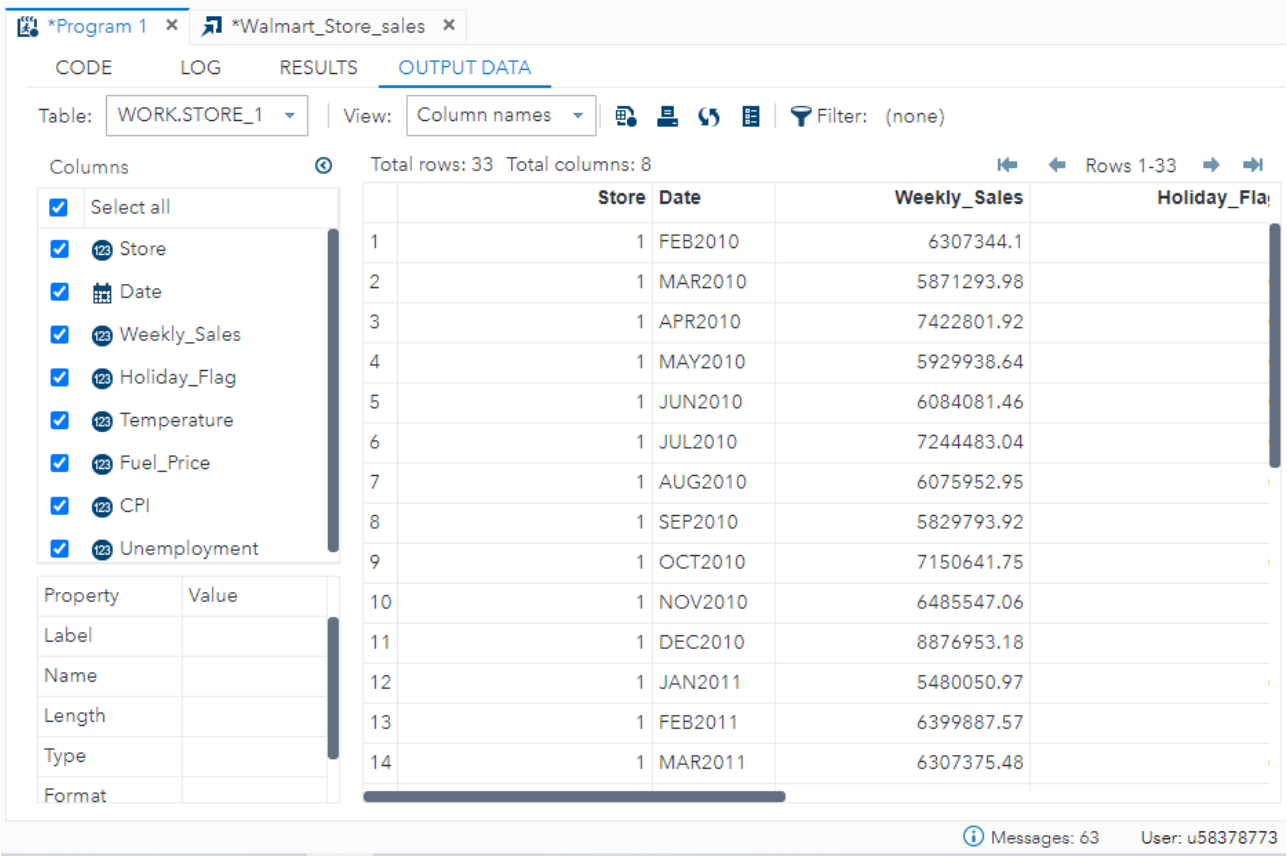
Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	05/02/2010	1643690.9	0	42.31	2.572	211.0963582	8.106
2	1	12/02/2010	1641957.44	1	38.51	2.548	211.2421698	8.106
3	1	19/02/2010	1611968.17	0	39.93	2.514	211.2891429	8.106
4	1	26/02/2010	1409727.59	0	46.63	2.561	211.3196429	8.106
5	1	05/03/2010	1554806.68	0	46.5	2.625	211.3501429	8.106
6	1	12/03/2010	1439541.59	0	57.79	2.667	211.3806429	8.106
7	1	19/03/2010	1472515.79	0	54.58	2.72	211.215635	8.106
8	1	26/03/2010	1404429.92	0	51.45	2.732	211.0180424	8.106
9	1	02/04/2010	1594968.28	0	62.27	2.719	210.8204499	7.808
10	1	09/04/2010	1545418.53	0	65.86	2.77	210.6228574	7.808
11	1	16/04/2010	1466058.28	0	66.32	2.808	210.4887	7.808
12	1	23/04/2010	1391256.12	0	64.84	2.795	210.4391228	7.808
13	1	30/04/2010	1425100.71	0	67.41	2.78	210.3895456	7.808
14	1	07/05/2010	1603955.12	0	72.55	2.835	210.3399684	7.808
15	1	14/05/2010	1494251.5	0	74.78	2.854	210.3374261	7.808
16	1	21/05/2010	1399662.07	0	76.44	2.826	210.6170934	7.808
17	1	28/05/2010	1432069.95	0	80.44	2.759	210.8967606	7.808
18	1	04/06/2010	1615524.71	0	80.69	2.705	211.1764278	7.808
19	1	11/06/2010	1542561.09	0	80.43	2.668	211.4560951	7.808
20	1	18/06/2010	1503284.06	0	84.11	2.637	211.4537719	7.808
21	1	25/06/2010	1422711.6	0	84.34	2.653	211.3386526	7.808
22	1	02/07/2010	1492418.14	0	80.91	2.669	211.2235333	7.787

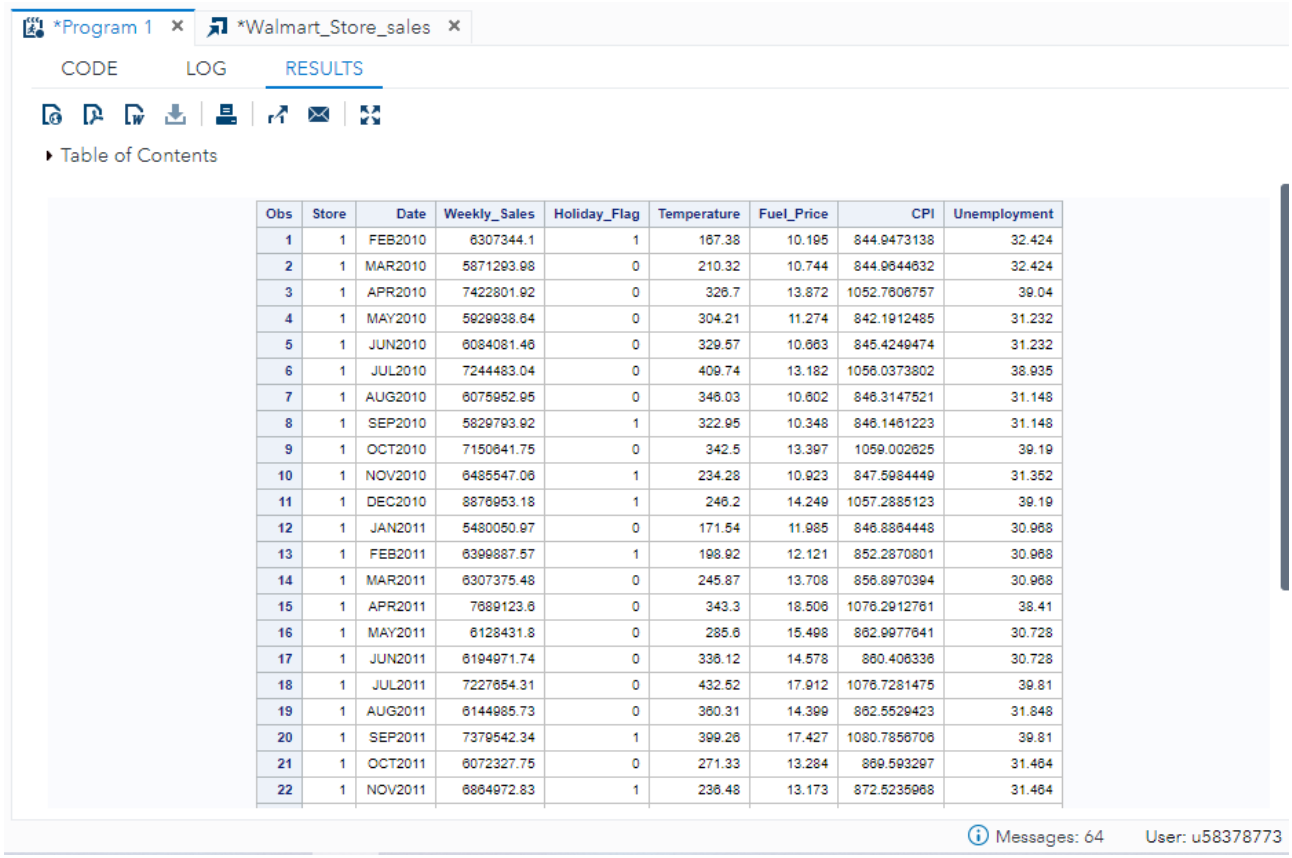
```
/* Convert store-1 data into timeseries data */

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;
```

Output:



```
PROC PRINT DATA= work.store_1;
RUN;
```

Output:



Obs	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment
1	1	FEB2010	6307344.1	1	167.38	10.195	844.9473138	32.424
2	1	MAR2010	5871293.98	0	210.32	10.744	844.9644632	32.424
3	1	APR2010	7422801.92	0	326.7	13.872	1052.7606757	39.04
4	1	MAY2010	5929938.64	0	304.21	11.274	842.1912485	31.232
5	1	JUN2010	6084081.46	0	329.57	10.663	845.4249474	31.232
6	1	JUL2010	7244483.04	0	409.74	13.182	1056.0373802	38.935
7	1	AUG2010	6075952.95	0	346.03	10.602	846.3147521	31.148
8	1	SEP2010	5829793.92	1	322.95	10.348	846.1461223	31.148
9	1	OCT2010	7150641.75	0	342.5	13.397	1059.002625	39.19
10	1	NOV2010	6485547.06	1	234.28	10.923	847.5984449	31.352
11	1	DEC2010	8876953.18	1	246.2	14.249	1057.2885123	39.19
12	1	JAN2011	5480050.97	0	171.54	11.985	846.8864448	30.968
13	1	FEB2011	6399887.57	1	198.92	12.121	852.2870801	30.968
14	1	MAR2011	6307375.48	0	245.87	13.708	856.8970394	30.968
15	1	APR2011	7689123.6	0	343.3	18.506	1076.2912761	38.41
16	1	MAY2011	6128431.8	0	285.6	15.498	862.9977641	30.728
17	1	JUN2011	6194971.74	0	336.12	14.578	860.406336	30.728
18	1	JUL2011	7227654.31	0	432.52	17.912	1076.7281475	39.81
19	1	AUG2011	6144985.73	0	360.31	14.399	862.5529423	31.848
20	1	SEP2011	7379542.34	1	399.26	17.427	1080.7856706	39.81
21	1	OCT2011	6072327.75	0	271.33	13.284	869.593297	31.464
22	1	NOV2011	6864972.83	1	236.48	13.173	872.5235968	31.464

```
/* Build Model */

/*ARIMA MODEL*/

PROC ARIMA DATA=WORK.STORE_1;
identify var=Weekly_Sales(1);
estimate  p=1 q=1;
forecast  lead=6 interval=month id=date;
RUN;
```

Output:

*Walmart_Store_sales *Program 1

CODE LOG RESULTS

Table of Contents

ARIMA Estimation Optimization Summary	
Estimation Method	Conditional Least Squares
Parameters Estimated	3
Termination Criteria	Maximum Relative Change in Estimates
Iteration Stopping Value	0.001
Criteria Value	3.332094
Maximum Absolute Value of Gradient	1.295E13
R-Square Change from Last Iteration	0.692209
Objective Function	Sum of Squared Residuals
Objective Function Value	2.678E13
Marquardt's Lambda Coefficient	0.00001
Numerical Derivative Perturbation Delta	0.001
Iterations	12
Warning Message	Estimates may not have converged.

Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MU	25013.5	32047.2	0.78	0.4414	0
MA1,1	0.99998	0.41405	2.42	0.0223	1
AR1,1	0.05932	0.29260	0.20	0.8408	1

Constant Estimate	23529.64
Variance Estimate	9.236E11
Std Error Estimate	961025.3
AIC	975.3104
SBC	979.7076
Number of Residuals	32

Messages: 6 User: u58378773

*Walmart_Store_sales *Program 1

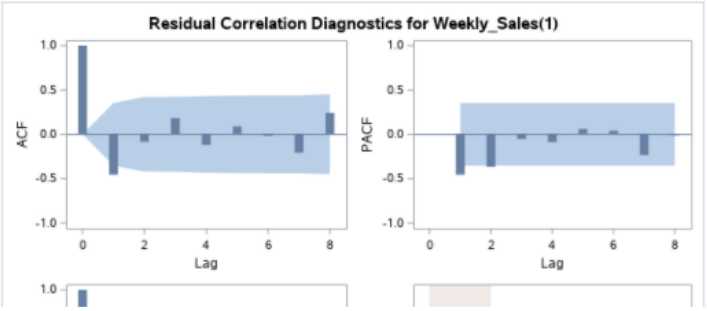
CODE LOG RESULTS

Table of Contents

* AIC and SBC do not include log determinant.

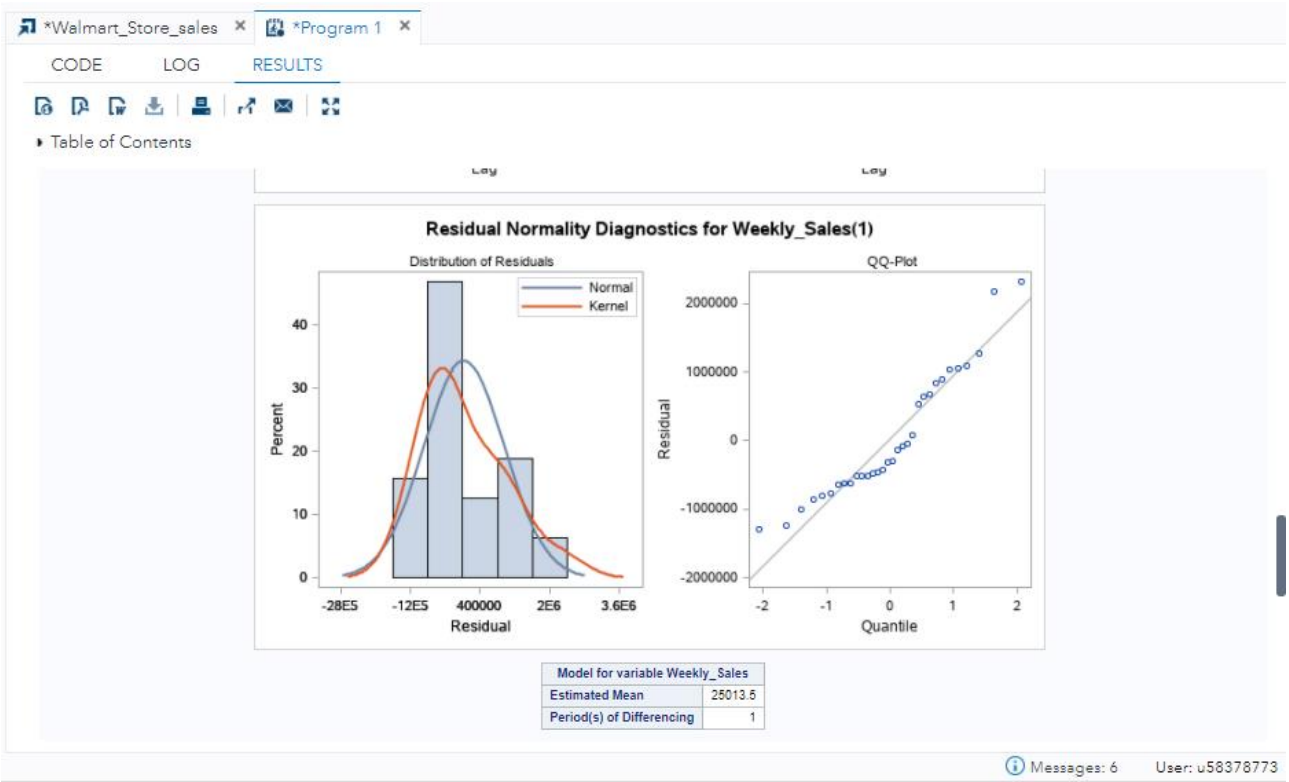
Correlations of Parameter Estimates			
Parameter	MU	MA1,1	AR1,1
MU	1.000	0.955	0.706
MA1,1	0.955	1.000	0.750
AR1,1	0.706	0.750	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9.57	4	0.0483	-0.453	-0.081	0.185	-0.117	0.090	-0.015
12	20.59	10	0.0241	-0.203	0.244	-0.012	-0.270	0.060	0.224
18	22.95	16	0.1150	-0.173	0.068	-0.000	0.009	0.055	-0.043
24	40.66	22	0.0090	-0.138	0.330	-0.211	-0.055	0.071	-0.083



Messages: 6 User: u58378773


*Walmart_Store_sales
*Program 1
CODE
LOG
RESULTS
Table of Contents
Residual Correlation Diagnostics for Weekly_Sales(1)
ACF
PACF
IACF
White Noise Prob
Lag
Messages: 6
User: u58378773


*Walmart_Store_sales
*Program 1
CODE
LOG
RESULTS
Table of Contents
Residual Normality Diagnostics for Weekly_Sales(1)
Distribution of Residuals
QQ-Plot
Normal
Kernel
Percent
Residual
Quantile
Model for variable Weekly_Sales
Estimated Mean 25013.5
Period(s) of Differencing 1
Messages: 6
User: u58378773

*Walmart_Store_sales × *Program 1 ×

CODE LOG RESULTS

Table of Contents

Model for variable Weekly_Sales	
Estimated Mean	25013.5
Period(s) of Differencing	1

Autoregressive Factors	
Factor 1:	1 - 0.05932 B**(1)

Moving Average Factors	
Factor 1:	1 - 0.99998 B**(1)

Forecasts for variable Weekly_Sales

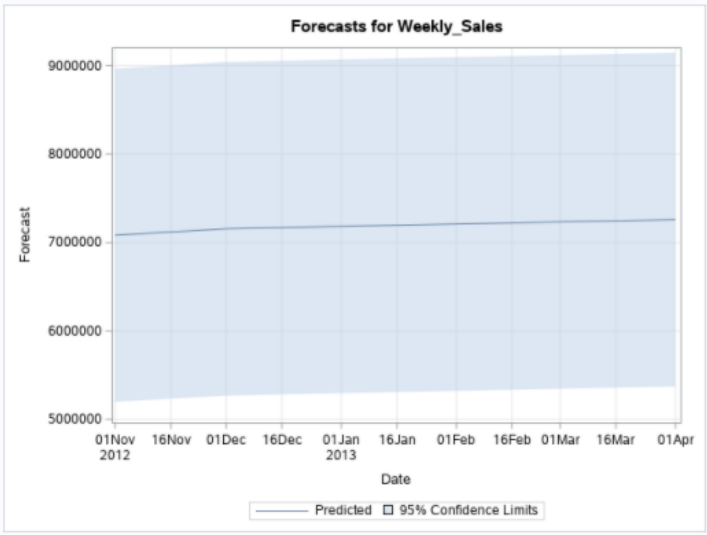
Obs	Forecast	Std Error	95% Confidence Limits	
34	7081658.0	961025	5198083.1	8965233.0
35	7154785.2	962716	5267897.1	9041673.4
36	7182653.0	962722	5295753.1	9069552.9
37	7207835.8	962722	5320935.8	9094735.7
38	7232859.3	962722	5345959.4	9119759.3
39	7257873.4	962722	5370973.5	9144773.4

Messages: 6 User: u58378773

*Walmart_Store_sales × *Program 1 ×

CODE LOG RESULTS

Table of Contents

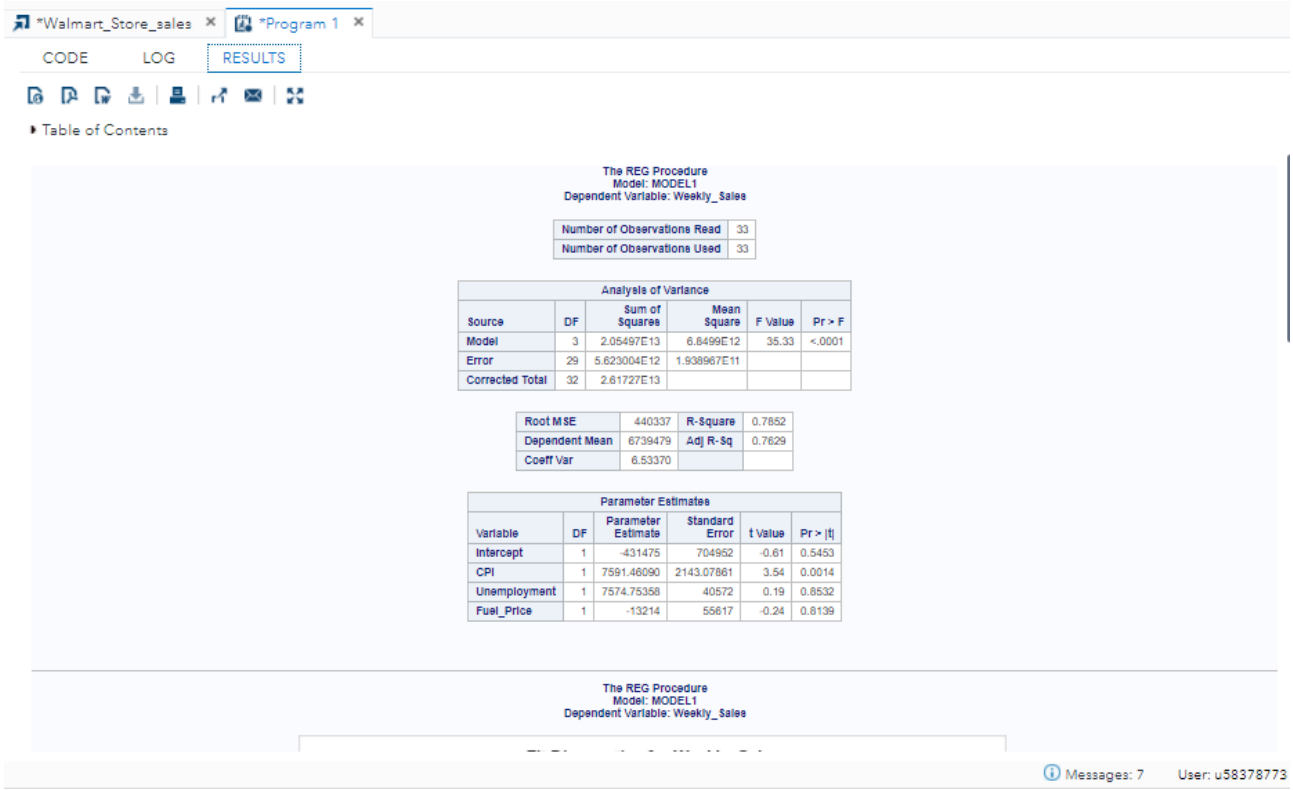


Messages: 6 User: u58378773

/* LINEAR REGRESSION MODEL*/

```
PROC REG DATA= work.store_1;
Model Weekly_Sales = CPI Unemployment Fuel_Price;
RUN;
```

Output:



*Walmart_Store_sales *Program 1

CODE LOG RESULTS

Table of Contents

The REG Procedure
Model: MODEL1
Dependent Variable: Weekly_Sales

Number of Observations Read	33
Number of Observations Used	33

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2.05497E13	6.8499E12	35.33	<.0001
Error	29	5.623004E12	1.938967E11		
Corrected Total	32	2.61727E13			

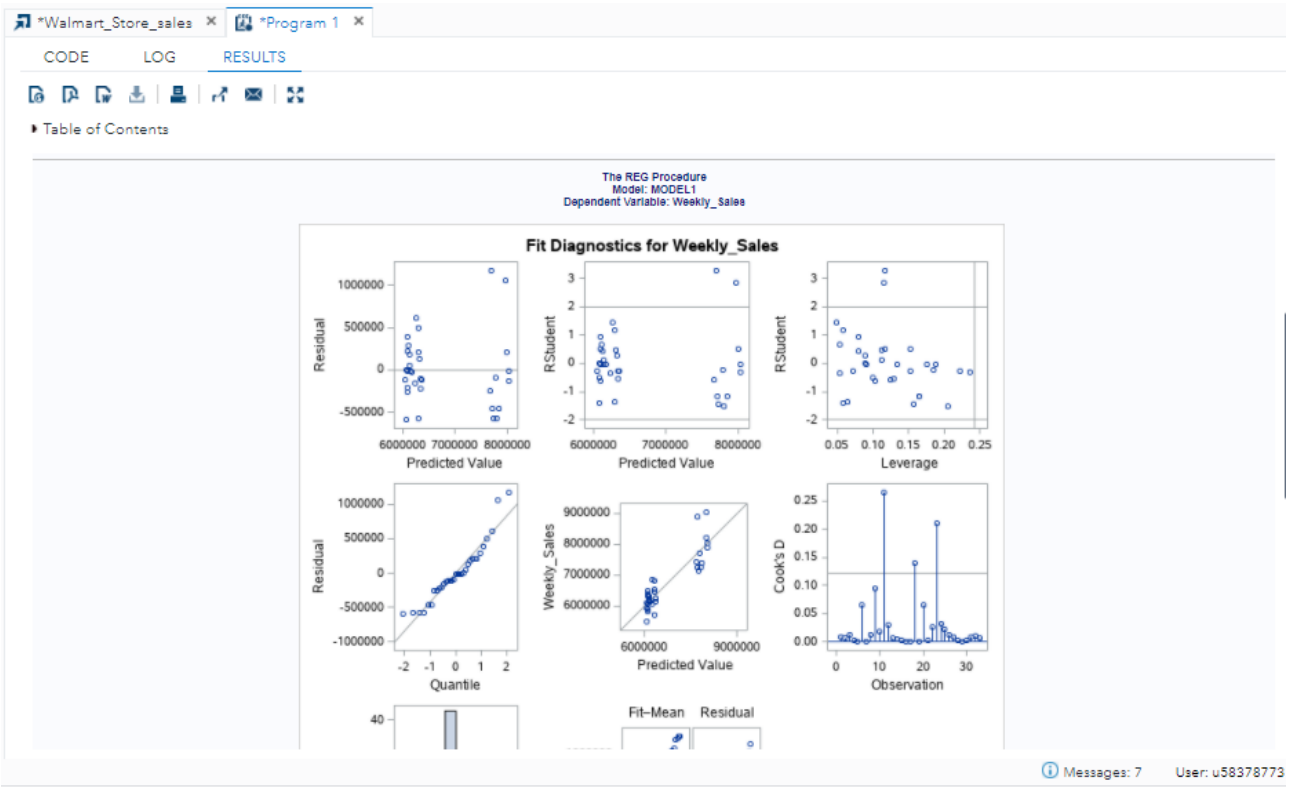
Root MSE	440337	R-Square	0.7852
Dependent Mean	6739479	Adj R-Sq	0.7629
Coeff Var	6.53370		

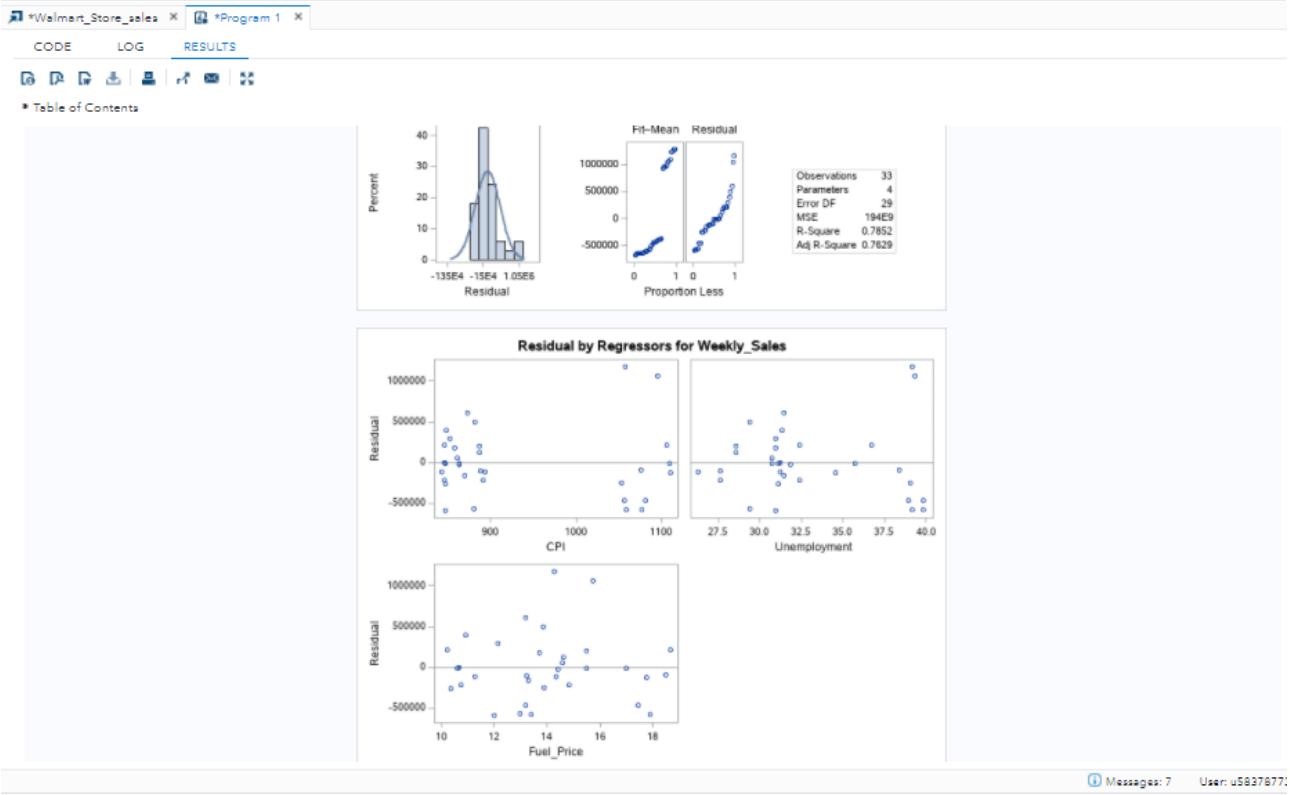
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-431475	704952	-0.61	0.5453
CPI	1	7591.46090	2143.07861	3.54	0.0014
Unemployment	1	7574.75358	40572	0.19	0.8532
Fuel_Price	1	-13214	55617	-0.24	0.8139

The REG Procedure
Model: MODEL1
Dependent Variable: Weekly_Sales

Messages: 7 User: u58378773


*Walmart_Store_sales
*Program 1
CODE
LOG
RESULTS
Table of Contents
The REG Procedure
Model: MODEL1
Dependent Variable: Weekly_Sales
Fit Diagnostics for Weekly_Sales
Residual
RStudent
Predicted Value
Leverage
Quantile
Weekly_Sales
Cook's D
Observation
Fit-Mean
Residual
Messages: 7
User: u58378773


*Walmart_Store_sales
*Program 1
CODE
LOG
RESULTS
Table of Contents
Percent
Residual
Fit-Mean
Proportion Less
Observations 33
Parameters 4
Error DF 29
MSE 194E9
R-Square 0.7852
Adj R-Square 0.7629
Residual by Regressors for Weekly_Sales
CPI
Unemployment
Fuel_Price
Messages: 7