Rossmann Sales Data Exploration

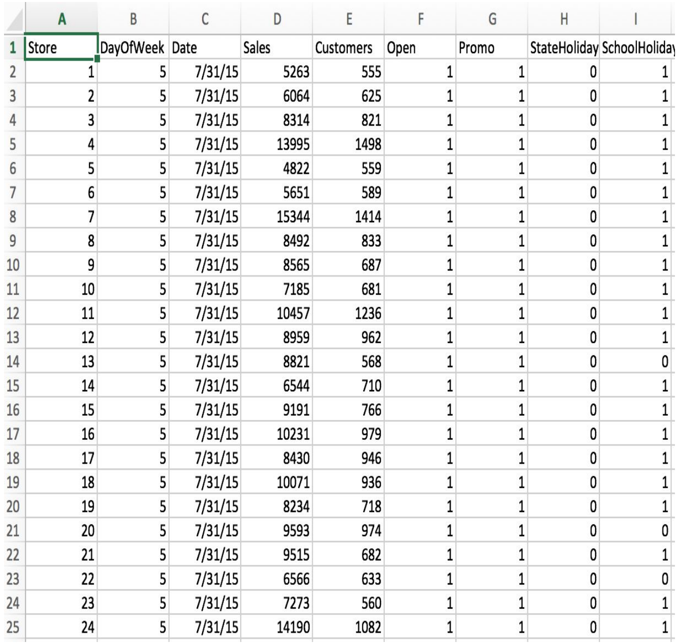
Introduction

- Second-largest drug store chain in Europe
- 3600 stores, more than a million visit on a daily-basis
- Average sales area of 500 m^2, with a product range of more than 17,500 different articles.
- Consolidated sales of 9 billion euros in 2017
- Fast expansion: 230 new stores were opened in 2013
- World's 50 fastest growing trading companies

- **Store types data exploration**
    - Data Processing
    - Store type analysis
- **Forecasting**
- **Store 1 sales data exploration**
    - Correlation between variables
    - Multiple regression analysis
    - Regression with interactions of variables
- **Conclusion**
    - Suggestions for Rossmann based on our findings
- **Future improvement**

Introduction

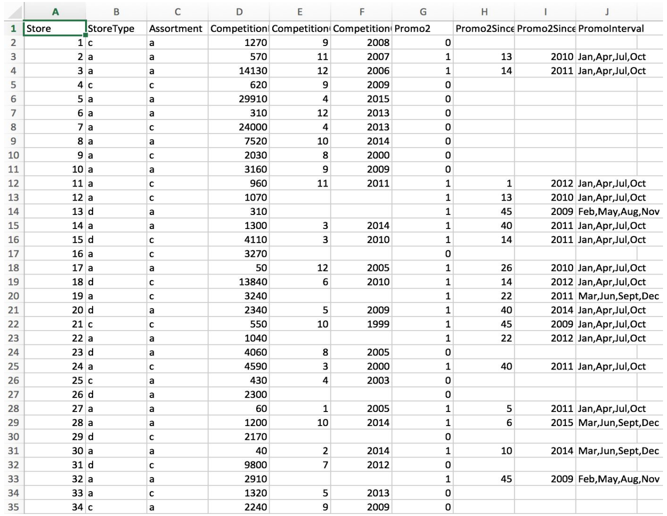
- Data description - Historical sales dataset (Train.csv)
    - Store - a unique Id for each store
    - Day of Week - an indicator for the day of week: 1 = Monday, 2 = Tuesday … 7 = Sunday
    - Date - MM/DD/YY
    - Sales - the turnover on a given day
    - Customers - the number of customers on a given day
    - Open - an indicator for whether the store was open: 0 = closed, 1 = open
    - Promo - indicates whether a store is running a promo on that day: 0 = no promo, 1 = have promo
    - StateHoliday - indicates a state holiday: a = public holiday, b = Easter holiday, c = Christmas, 0 = None
    - SchoolHoliday - indicates if the (Store, Date) was affected by the closure of public schools: 0 = No, 1= Yes



	A	B	C	D	E	F	G	H	I
1	Store	DayOfWeek	Date	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday
2	1	5	7/31/15	5263	555	1	1	0	1
3	2	5	7/31/15	6064	625	1	1	0	1
4	3	5	7/31/15	8314	821	1	1	0	1
5	4	5	7/31/15	13995	1498	1	1	0	1
6	5	5	7/31/15	4822	559	1	1	0	1
7	6	5	7/31/15	5651	589	1	1	0	1
8	7	5	7/31/15	15344	1414	1	1	0	1
9	8	5	7/31/15	8492	833	1	1	0	1
10	9	5	7/31/15	8565	687	1	1	0	1
11	10	5	7/31/15	7185	681	1	1	0	1
12	11	5	7/31/15	10457	1236	1	1	0	1
13	12	5	7/31/15	8959	962	1	1	0	1
14	13	5	7/31/15	8821	568	1	1	0	0
15	14	5	7/31/15	6544	710	1	1	0	1
16	15	5	7/31/15	9191	766	1	1	0	1
17	16	5	7/31/15	10231	979	1	1	0	1
18	17	5	7/31/15	8430	946	1	1	0	1
19	18	5	7/31/15	10071	936	1	1	0	1
20	19	5	7/31/15	8234	718	1	1	0	1
21	20	5	7/31/15	9593	974	1	1	0	0
22	21	5	7/31/15	9515	682	1	1	0	1
23	22	5	7/31/15	6566	633	1	1	0	0
24	23	5	7/31/15	7273	560	1	1	0	1
25	24	5	7/31/15	14190	1082	1	1	0	1

Introduction

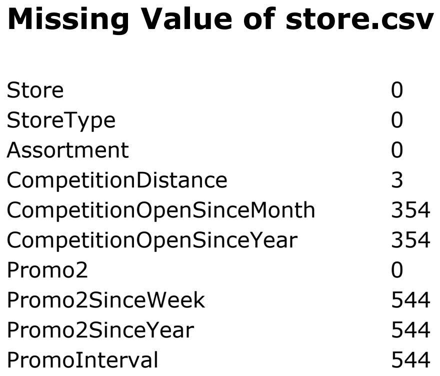
- Data description - Store Information (Store.csv)
  - StoreType - 4 different store models: a, b, c, d
  - Assortment - describes an assortment level: a = basic, b = extra, c = extended
  - CompetitionDistance - distance in meters to the nearest competitor store
  - CompetitionOpenSince[Month/Year] - gives the approximate year and month of the time the nearest competitor was opened
  - Promo2 - a continuing and consecutive promotion for some stores: 0 = store is not participating, 1 = store is participating
  - Promo2Since[Year/Week] - describes the year and calendar week when the store started participating in Promo2
  - PromoInterval - describes the consecutive intervals Promo2 is started, naming the months the promotion is started anew.



	A	B	C	D	E	F	G	H	I	J
1	Store	StoreType	Assortment	Competition	Competition	Competition	Promo2	Promo2Since	Promo2Since	PromoInterval
2	1	c	a	1270	9	2008	0			
3	2	a	a	570	11	2007	1	13	2010	Jan,Apr,Jul,Oct
4	3	a	a	14130	12	2006	1	14	2011	Jan,Apr,Jul,Oct
5	4	c	c	620	9	2009	0			
6	5	a	a	29910	4	2015	0			
7	6	a	a	310	12	2013	0			
8	7	a	c	24000	4	2013	0			
9	8	a	a	7520	10	2014	0			
10	9	a	c	2030	8	2000	0			
11	10	a	a	3160	9	2009	0			
12	11	a	c	960	11	2011	1	1	2012	Jan,Apr,Jul,Oct
13	12	a	c	1070			1	13	2010	Jan,Apr,Jul,Oct
14	13	d	a	310			1	45	2009	Feb,May,Aug,Nov
15	14	a	a	1300	3	2014	1	40	2011	Jan,Apr,Jul,Oct
16	15	d	c	4110	3	2010	1	14	2011	Jan,Apr,Jul,Oct
17	16	a	c	3270			0			
18	17	a	a	50	12	2005	1	26	2010	Jan,Apr,Jul,Oct
19	18	d	c	13840	6	2010	1	14	2012	Jan,Apr,Jul,Oct
20	19	a	c	3240			1	22	2011	Mar,Jun,Sept,Dec
21	20	d	a	2340	5	2009	1	40	2014	Jan,Apr,Jul,Oct
22	21	c	c	550	10	1999	1	45	2009	Jan,Apr,Jul,Oct
23	22	a	a	1040			1	22	2012	Jan,Apr,Jul,Oct
24	23	d	a	4060	8	2005	0			
25	24	a	c	4590	3	2000	1	40	2011	Jan,Apr,Jul,Oct
26	25	c	a	430	4	2003	0			
27	26	d	a	2300			0			
28	27	a	a	60	1	2005	1	5	2011	Jan,Apr,Jul,Oct
29	28	a	a	1200	10	2014	1	6	2015	Mar,Jun,Sept,Dec
30	29	d	c	2170			0			
31	30	a	a	40	2	2014	1	10	2014	Mar,Jun,Sept,Dec
32	31	d	c	9800	7	2012	0			
33	32	a	a	2910			1	45	2009	Feb,May,Aug,Nov
34	33	a	c	1320	5	2013	0			
35	34	c	a	2240	9	2009	0			

Store type data exploration

- Data Processing
  - Remove all closed stores in train.csv
  - Add new variables in train.csv: 'Year', 'Month', 'Day', 'WeekOfYear' and ‘SalePerCustomer’
  - Replace missing value in ‘CompetitionDistance’ with median value
  - Others will be replaced by 0
  - Merge train.csv and store.csv based on Store Id

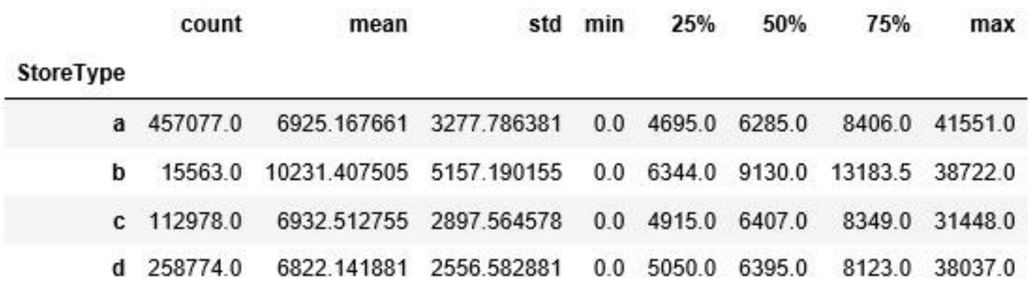


Missing Value of store.csv

```
Store                        0
StoreType                    0
Assortment                   0
CompetitionDistance          3
CompetitionOpenSinceMonth  354
CompetitionOpenSinceYear   354
Promo2                       0
Promo2SinceWeek            544
Promo2SinceYear            544
PromoInterval              544
```

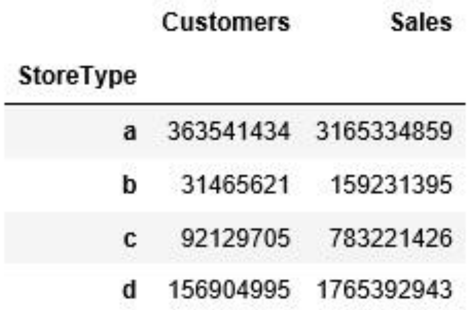
Store type data exploration

- Store type analysis



StoreType	count	mean	std	min	25%	50%	75%	max
a	457077.0	6925.167661	3277.786381	0.0	4695.0	6285.0	8406.0	41551.0
b	15563.0	10231.407505	5157.190155	0.0	6344.0	9130.0	13183.5	38722.0
c	112978.0	6932.512755	2897.564578	0.0	4915.0	6407.0	8349.0	31448.0
d	258774.0	6822.141881	2556.582881	0.0	5050.0	6395.0	8123.0	38037.0

Description of Sales on each StoreType

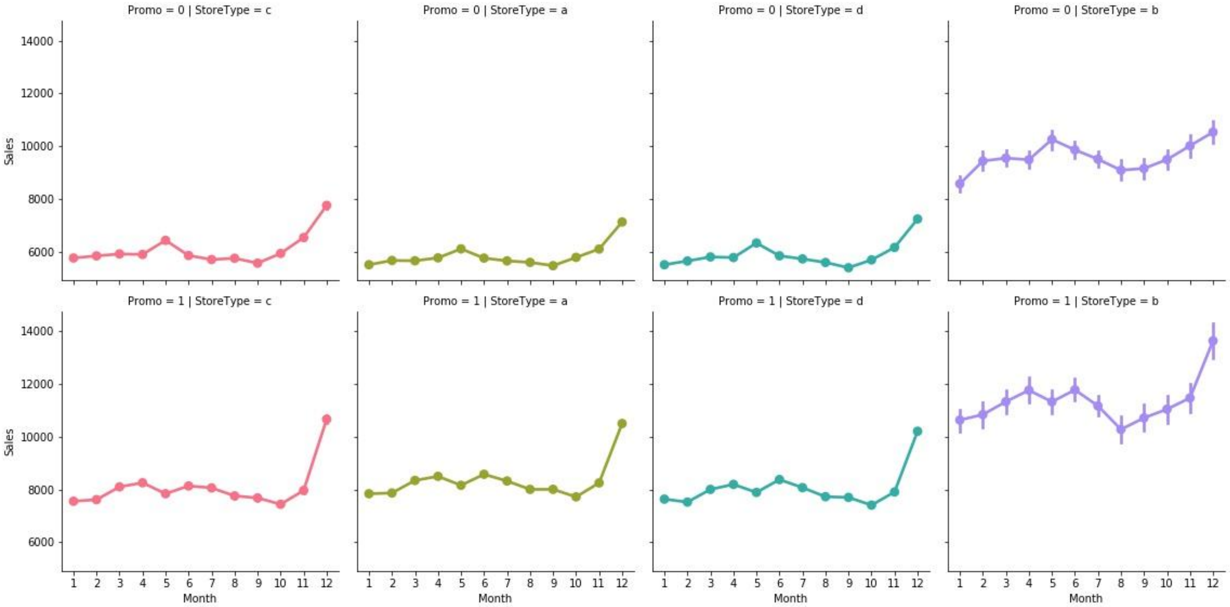


StoreType	Customers	Sales
a	363541434	3165334859
b	31465621	159231395
c	92129705	783221426
d	156904995	1765392943

Sum of Customers and Sales

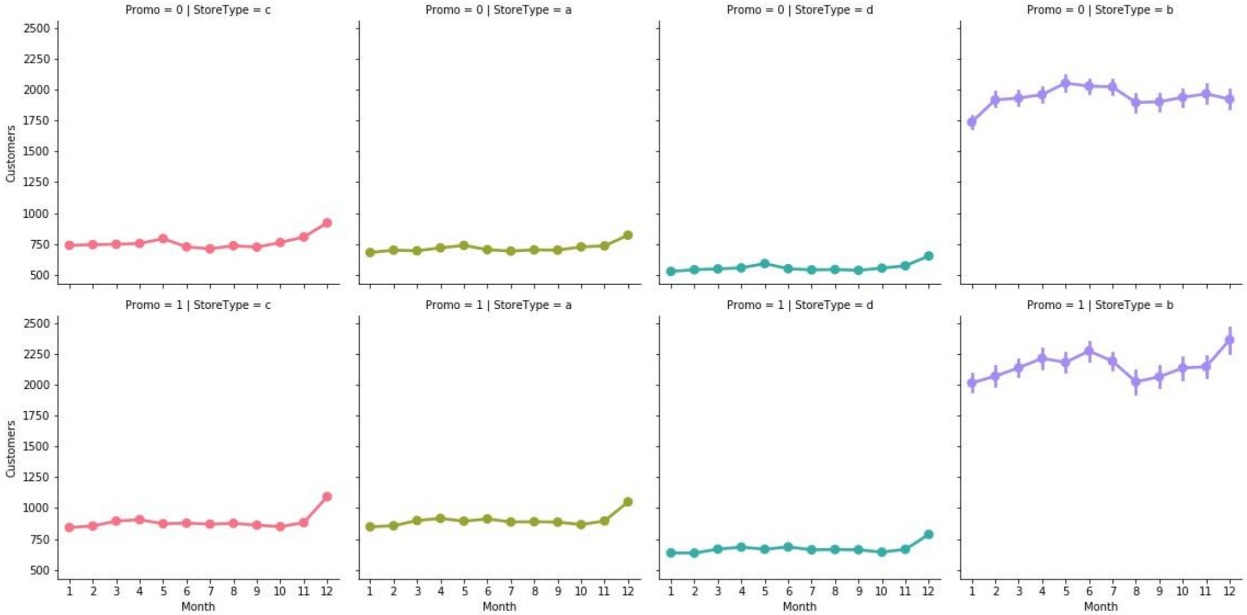
Store type data exploration

- Store type analysis - Sales Trend



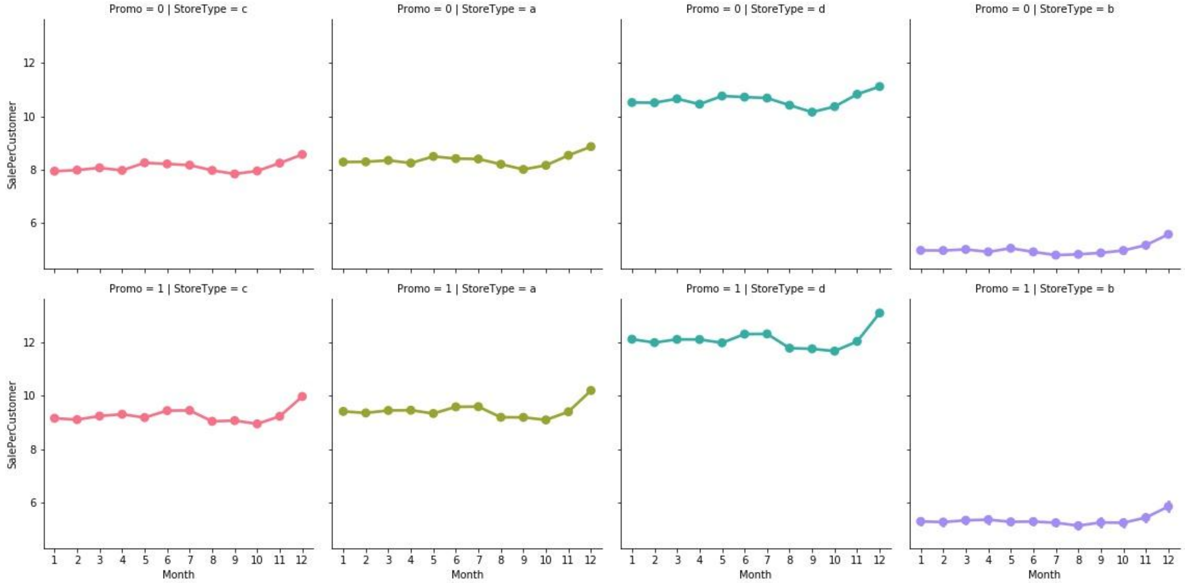
Store type data exploration

- Store type analysis - Customers Trend



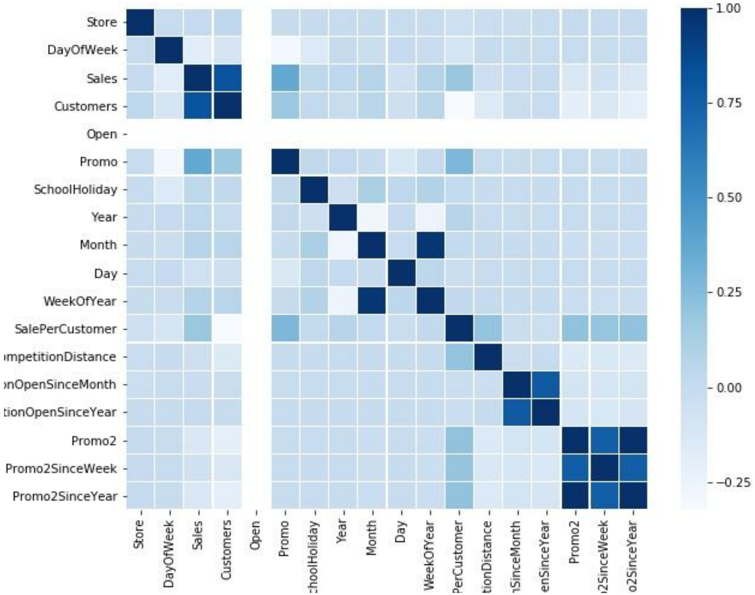
Store type data exploration

- Store type analysis - Sales per Customer Trend



Store type data exploration

- Correlation Matrix



- Strong positive correlation between Sales and Customers

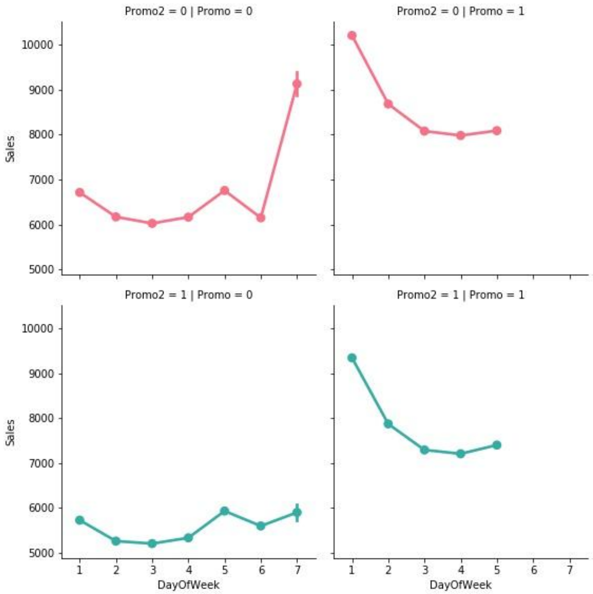
- Positive correlation between promotion (Promo) and Customers

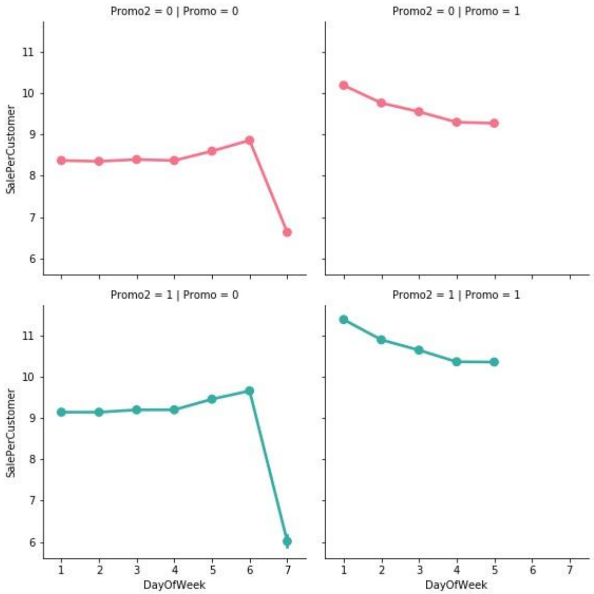
- If the store continues a consecutive promotion (Promo2), sales per customer will increase

- Negative correlation between promotion (Promo) and the day of a week

Store type data exploration

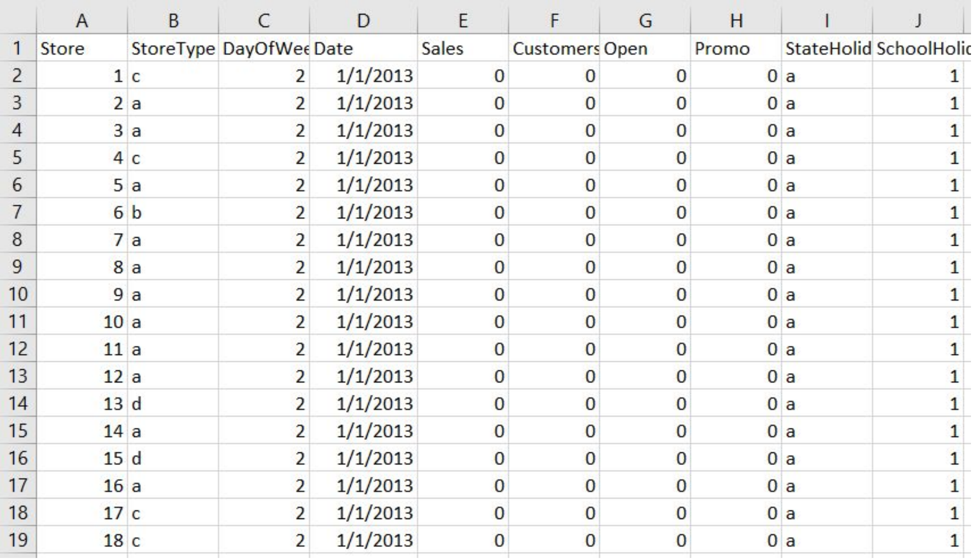
- Promotion and DayOfWeek





Forecasting

Part of the raw data from Rossmann store



	A	B	C	D	E	F	G	H	I	J
1	Store	StoreType	DayOfWee	Date	Sales	Customers	Open	Promo	StateHolid	SchoolHolic
2	1	c	2	1/1/2013	0	0	0	0	a	1
3	2	a	2	1/1/2013	0	0	0	0	a	1
4	3	a	2	1/1/2013	0	0	0	0	a	1
5	4	c	2	1/1/2013	0	0	0	0	a	1
6	5	a	2	1/1/2013	0	0	0	0	a	1
7	6	b	2	1/1/2013	0	0	0	0	a	1
8	7	a	2	1/1/2013	0	0	0	0	a	1
9	8	a	2	1/1/2013	0	0	0	0	a	1
10	9	a	2	1/1/2013	0	0	0	0	a	1
11	10	a	2	1/1/2013	0	0	0	0	a	1
12	11	a	2	1/1/2013	0	0	0	0	a	1
13	12	a	2	1/1/2013	0	0	0	0	a	1
14	13	d	2	1/1/2013	0	0	0	0	a	1
15	14	a	2	1/1/2013	0	0	0	0	a	1
16	15	d	2	1/1/2013	0	0	0	0	a	1
17	16	a	2	1/1/2013	0	0	0	0	a	1
18	17	c	2	1/1/2013	0	0	0	0	a	1
19	18	c	2	1/1/2013	0	0	0	0	a	1

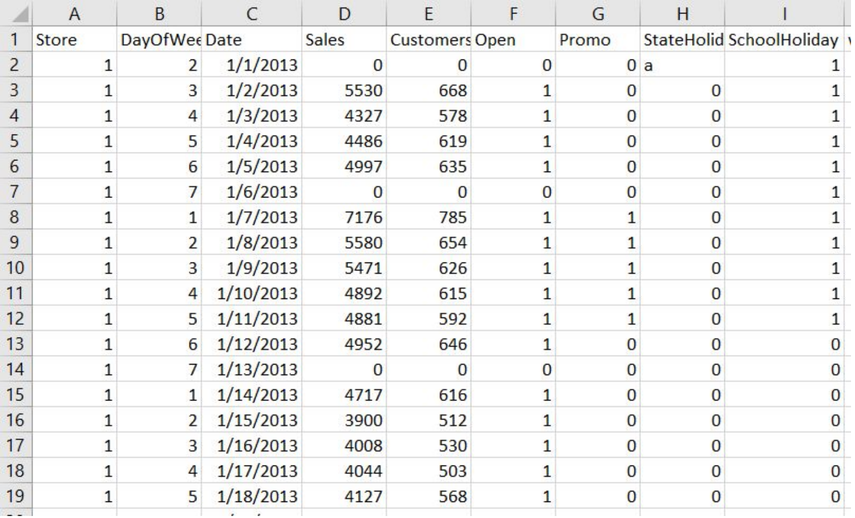
There are sales data for 1,115 Rossmann stores. These store have 4 types - a,b,c,d.

- According to sales for these store between 1/1/2013 and 7/31/2015, we use "Exponential smoothing", "Double Exponential Smoothing" and "Seasonal Exponential Smoothing" and the data for the store to **forecast the sales for the 6 weeks after 7/31/2015.**

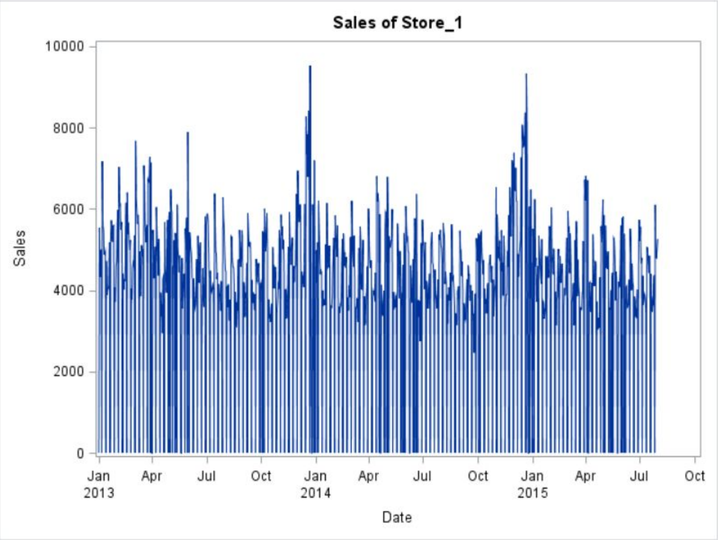
- Among 1115 Rossmann stores, we randomly chose 4 store with 4 different type.

Forecasting

Data of store1, sorting by day



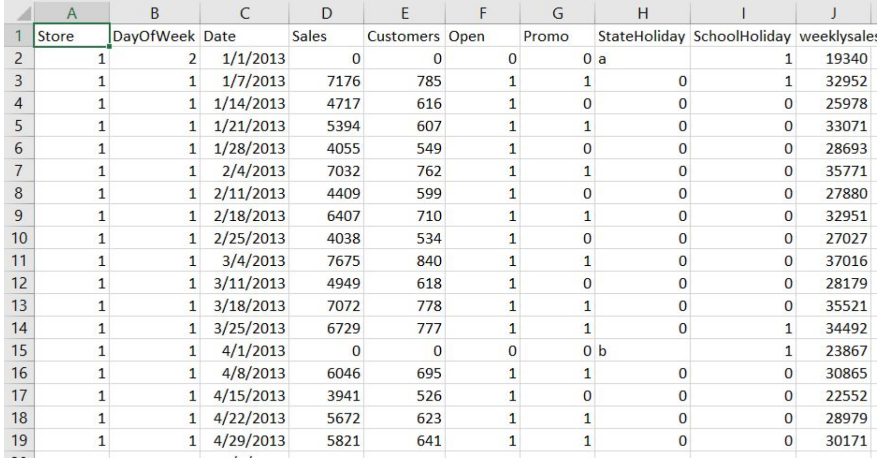
	A	B	C	D	E	F	G	H	I
1	Store	DayOfWee	Date	Sales	Customers	Open	Promo	StateHolid	SchoolHoliday
2	1	2	1/1/2013	0	0	0	0	a	1
3	1	3	1/2/2013	5530	668	1	0	0	1
4	1	4	1/3/2013	4327	578	1	0	0	1
5	1	5	1/4/2013	4486	619	1	0	0	1
6	1	6	1/5/2013	4997	635	1	0	0	1
7	1	7	1/6/2013	0	0	0	0	0	1
8	1	1	1/7/2013	7176	785	1	1	0	1
9	1	2	1/8/2013	5580	654	1	1	0	1
10	1	3	1/9/2013	5471	626	1	1	0	1
11	1	4	1/10/2013	4892	615	1	1	0	1
12	1	5	1/11/2013	4881	592	1	1	0	1
13	1	6	1/12/2013	4952	646	1	0	0	0
14	1	7	1/13/2013	0	0	0	0	0	0
15	1	1	1/14/2013	4717	616	1	0	0	0
16	1	2	1/15/2013	3900	512	1	0	0	0
17	1	3	1/16/2013	4008	530	1	0	0	0
18	1	4	1/17/2013	4044	503	1	0	0	0
19	1	5	1/18/2013	4127	568	1	0	0	0



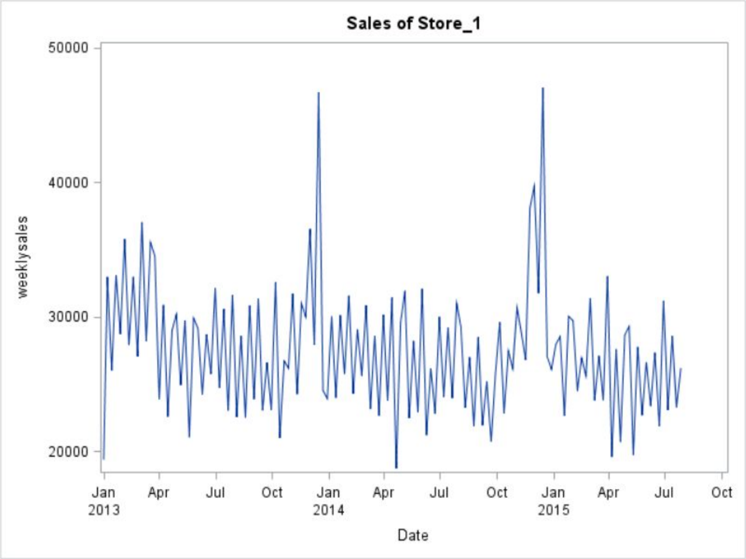
- x-axis=day
  y-axis=daily sales
- **Too many variables**
- **Hard to clearly show the forecasting line**

Forecasting

Data of store1, sorting by week



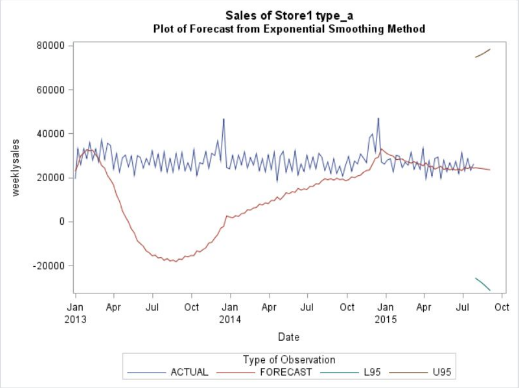
	A	B	C	D	E	F	G	H	I	J
1	Store	DayOfWeek	Date	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday	weeklysale
2	1	2	1/1/2013	0	0	0	0	a	1	19340
3	1	1	1/7/2013	7176	785	1	1	0	1	32952
4	1	1	1/14/2013	4717	616	1	0	0	0	25978
5	1	1	1/21/2013	5394	607	1	1	0	0	33071
6	1	1	1/28/2013	4055	549	1	0	0	0	28693
7	1	1	2/4/2013	7032	762	1	1	0	0	35771
8	1	1	2/11/2013	4409	599	1	0	0	0	27880
9	1	1	2/18/2013	6407	710	1	1	0	0	32951
10	1	1	2/25/2013	4038	534	1	0	0	0	27027
11	1	1	3/4/2013	7675	840	1	1	0	0	37016
12	1	1	3/11/2013	4949	618	1	0	0	0	28179
13	1	1	3/18/2013	7072	778	1	1	0	0	35521
14	1	1	3/25/2013	6729	777	1	1	0	1	34492
15	1	1	4/1/2013	0	0	0	0	b	1	23867
16	1	1	4/8/2013	6046	695	1	1	0	0	30865
17	1	1	4/15/2013	3941	526	1	0	0	0	22552
18	1	1	4/22/2013	5672	623	1	1	0	0	28979
19	1	1	4/29/2013	5821	641	1	1	0	0	30171

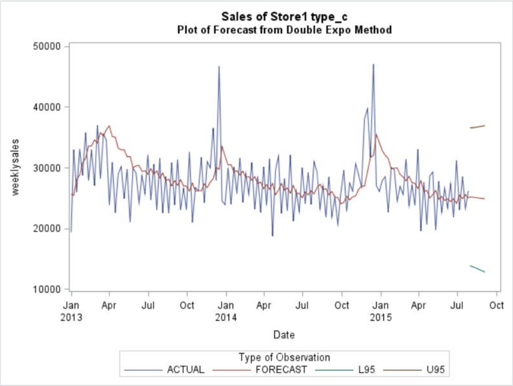


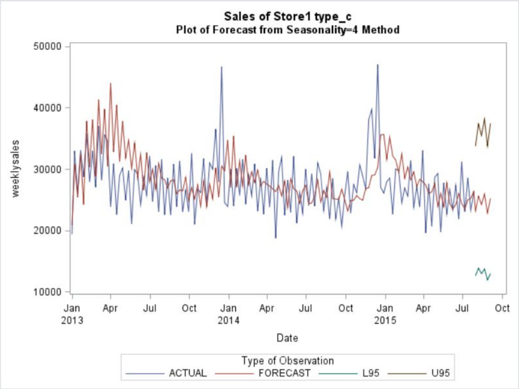
- x-axis=week
  y-axis=weekly sales

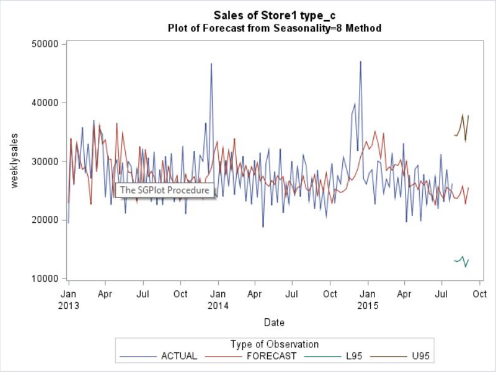
- **Weekly sales**
  - **Could show the forecasting line more clearly**

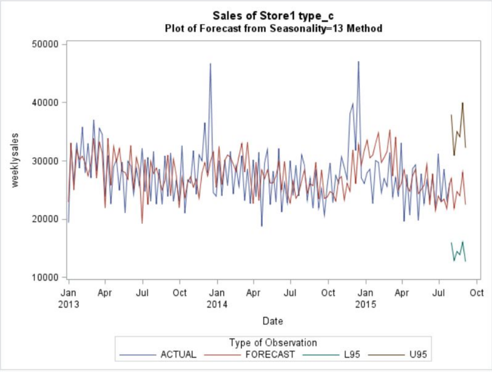
Forecasting of Store_1 (type_c)


Sales of Store1 type_a
Plot of Forecast from Exponential Smoothing Method


Sales of Store1 type_c
Plot of Forecast from Double Expo Method

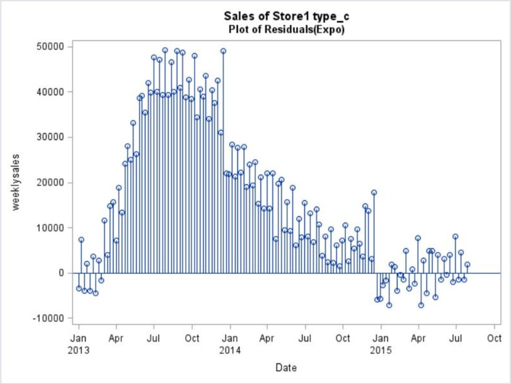

Sales of Store1 type_c
Plot of Forecast from Seasonality=4 Method

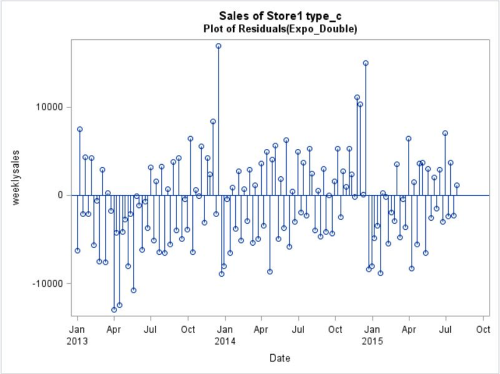

Sales of Store1 type_c
Plot of Forecast from Seasonality=8 Method

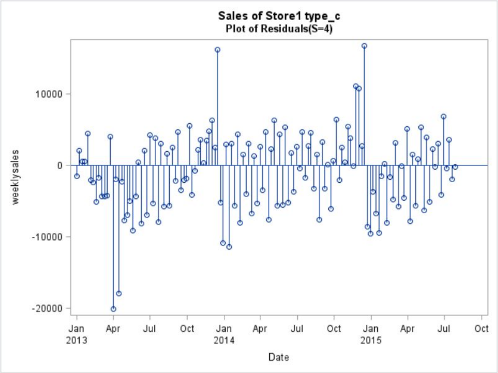

Sales of Store1 type_c
Plot of Forecast from Seasonality=13 Method

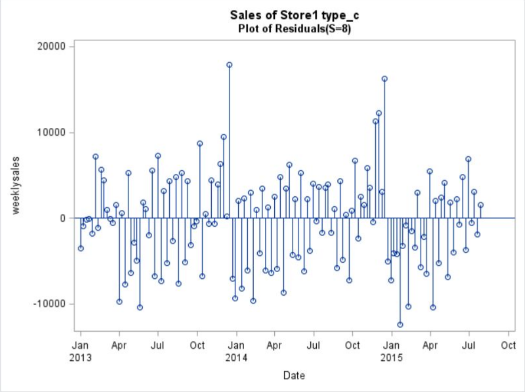
Sales of Store1 (type_c)

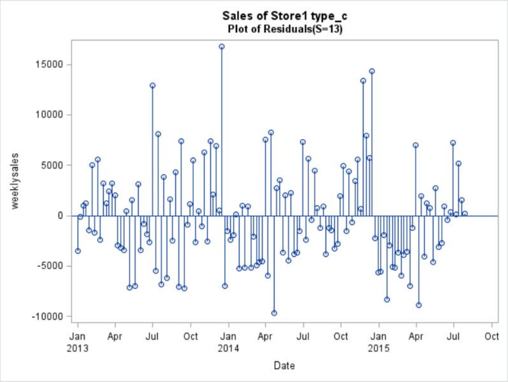
- Plot of Residuals



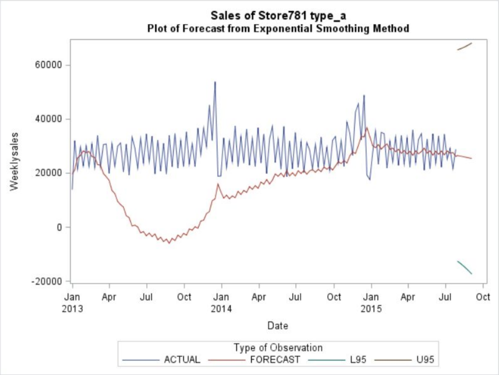


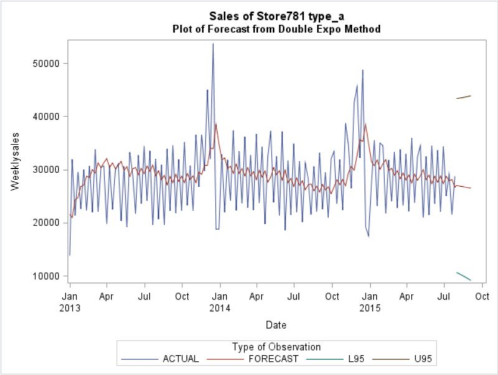


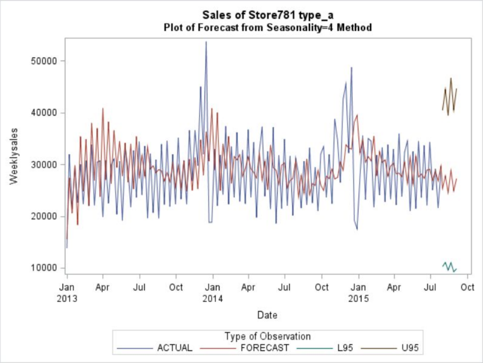


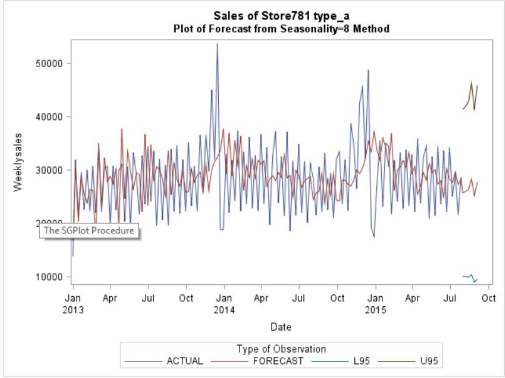


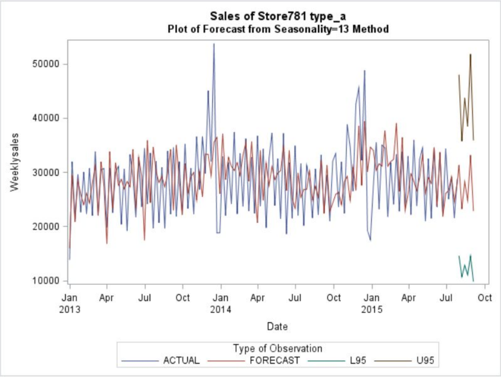
Forecasting of Store_781 (type_a)



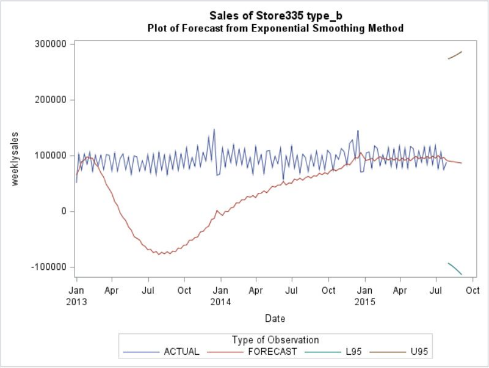


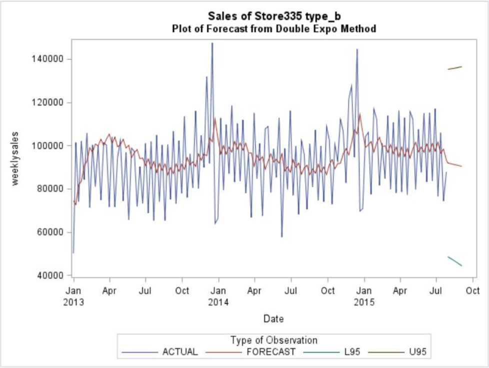


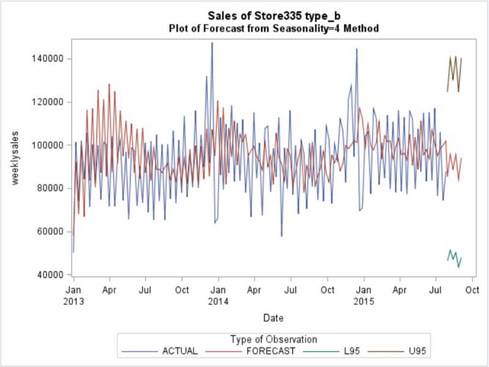


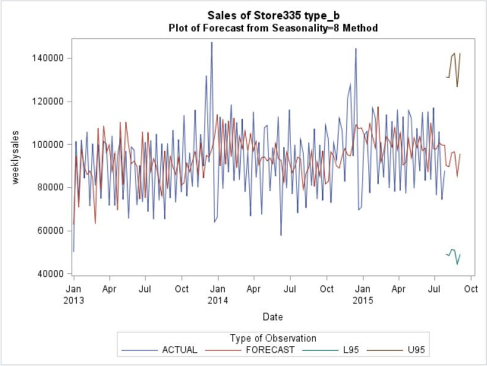


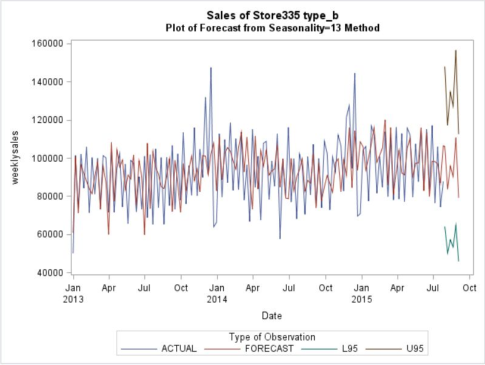
Forecasting of Store_335 (type_b)


Sales of Store335 type_b
Plot of Forecast from Exponential Smoothing Method

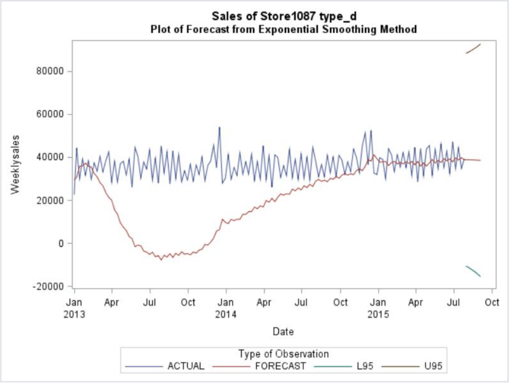

Sales of Store335 type_b
Plot of Forecast from Double Expo Method

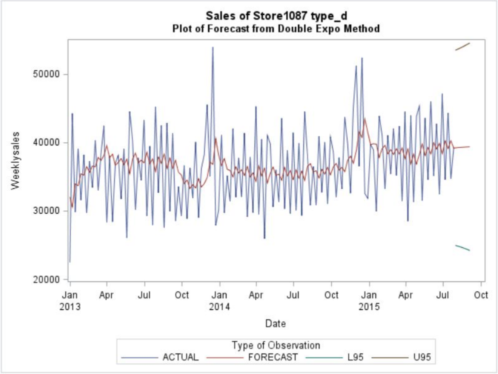

Sales of Store335 type_b
Plot of Forecast from Seasonality=4 Method

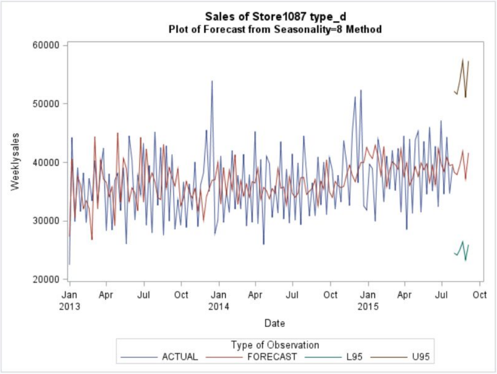

Sales of Store335 type_b
Plot of Forecast from Seasonality=8 Method

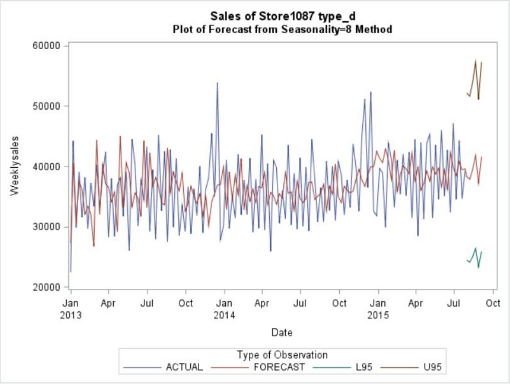

Sales of Store335 type_b
Plot of Forecast from Seasonality=13 Method

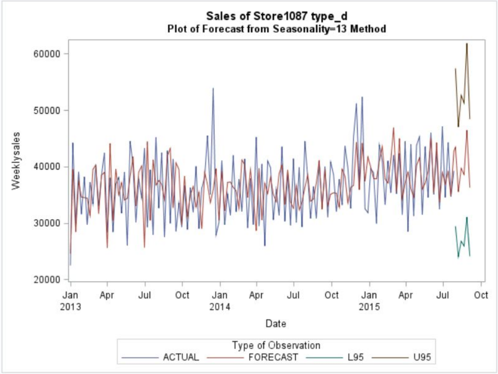
Forecasting of Store_1087 (type_d)



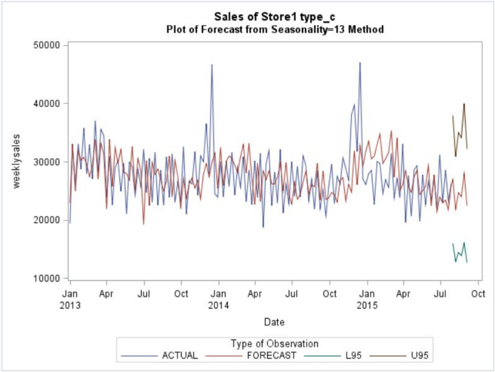




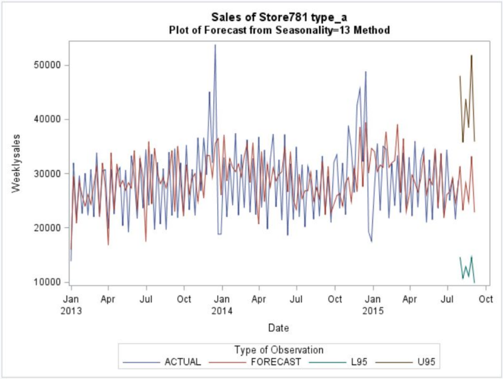




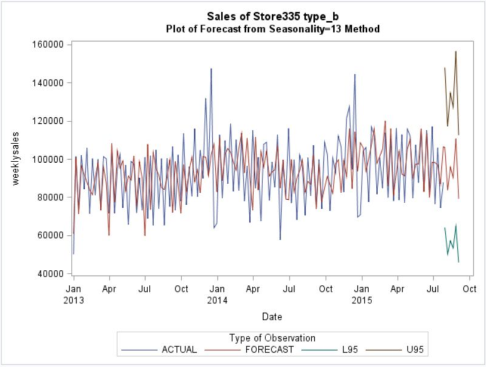
Store1_type_c



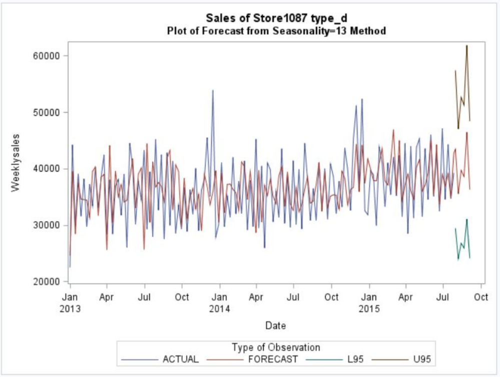
Store781_type_a



Store335_type_b



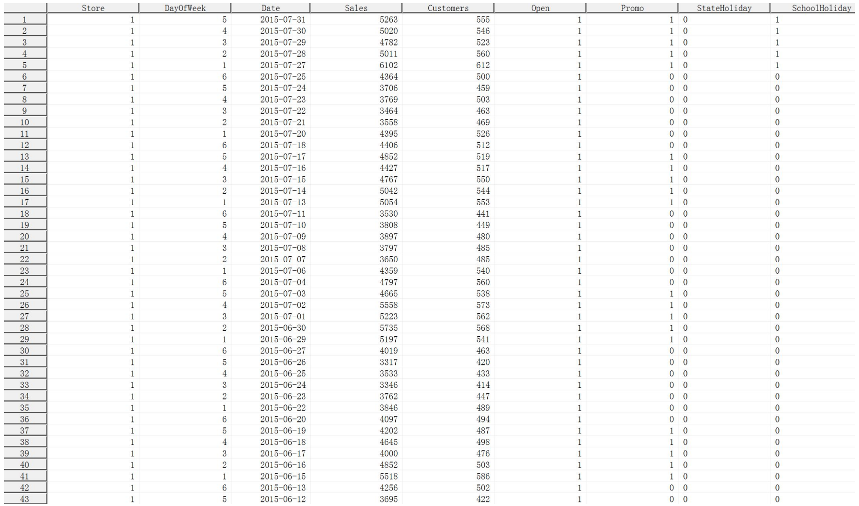
Store1087_type_d



- As we can see before, the **Seasonal Exponential Smoothing (seasonality=13)** is a better methodology to forecasts the store sales for all 4 types.

Store_1 sales data exploration

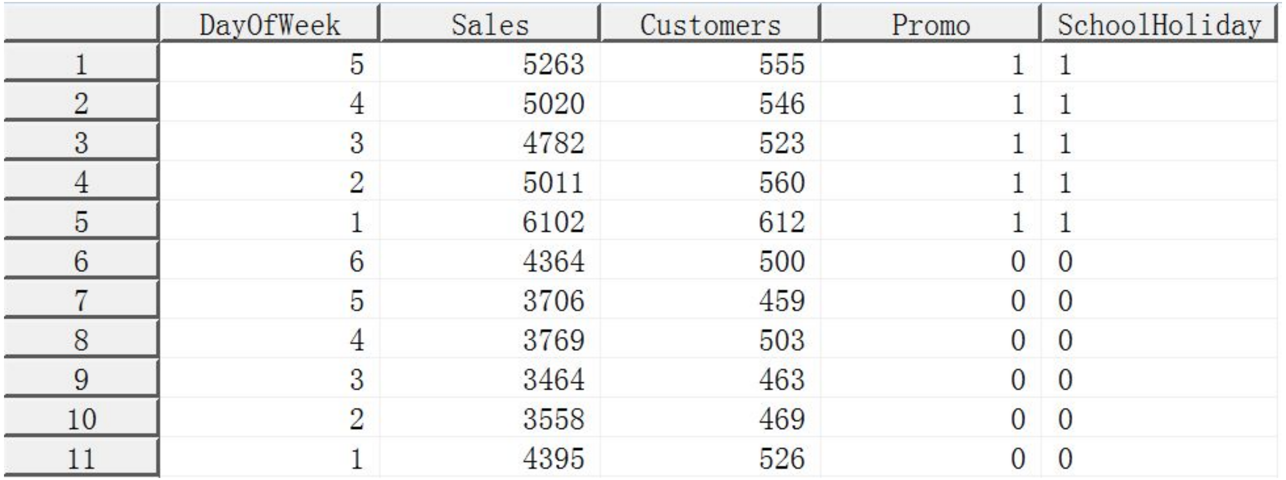
- Data preparation
  - Extract data for store_1 from Train.csv and keep open only



	Store	DayOfWeek	Date	Sales	Customers	Open	Promo	StateHoliday	SchoolHoliday
1	1	5	2015-07-31	5263	555	1	1	0	1
2	1	4	2015-07-30	5020	546	1	1	0	1
3	1	3	2015-07-29	4782	523	1	1	0	1
4	1	2	2015-07-28	5011	560	1	1	0	1
5	1	1	2015-07-27	6102	612	1	1	0	1
6	1	6	2015-07-25	4364	500	1	0	0	0
7	1	5	2015-07-24	3706	459	1	0	0	0
8	1	4	2015-07-23	3769	503	1	0	0	0
9	1	3	2015-07-22	3464	463	1	0	0	0
10	1	2	2015-07-21	3558	469	1	0	0	0
11	1	1	2015-07-20	4395	526	1	0	0	0
12	1	6	2015-07-18	4406	512	1	0	0	0
13	1	5	2015-07-17	4852	519	1	1	0	0
14	1	4	2015-07-16	4427	517	1	1	0	0
15	1	3	2015-07-15	4767	550	1	1	0	0
16	1	2	2015-07-14	5042	544	1	1	0	0
17	1	1	2015-07-13	5054	553	1	1	0	0
18	1	6	2015-07-11	3530	441	1	0	0	0
19	1	5	2015-07-10	3808	449	1	0	0	0
20	1	4	2015-07-09	3897	480	1	0	0	0
21	1	3	2015-07-08	3797	485	1	0	0	0
22	1	2	2015-07-07	3650	485	1	0	0	0
23	1	1	2015-07-06	4359	540	1	0	0	0
24	1	6	2015-07-04	4797	560	1	0	0	0
25	1	5	2015-07-03	4665	538	1	1	0	0
26	1	4	2015-07-02	5558	573	1	1	0	0
27	1	3	2015-07-01	5223	562	1	1	0	0
28	1	2	2015-06-30	5735	568	1	1	0	0
29	1	1	2015-06-29	5197	541	1	1	0	0
30	1	6	2015-06-27	4019	463	1	0	0	0
31	1	5	2015-06-26	3317	420	1	0	0	0
32	1	4	2015-06-25	3533	433	1	0	0	0
33	1	3	2015-06-24	3346	414	1	0	0	0
34	1	2	2015-06-23	3762	447	1	0	0	0
35	1	1	2015-06-22	3846	489	1	0	0	0
36	1	6	2015-06-20	4097	494	1	0	0	0
37	1	5	2015-06-19	4202	487	1	1	0	0
38	1	4	2015-06-18	4645	498	1	1	0	0
39	1	3	2015-06-17	4000	476	1	1	0	0
40	1	2	2015-06-16	4852	503	1	1	0	0
41	1	1	2015-06-15	5518	586	1	1	0	0
42	1	6	2015-06-13	4256	502	1	0	0	0
43	1	5	2015-06-12	3695	422	1	0	0	0

Store_1 sales data exploration

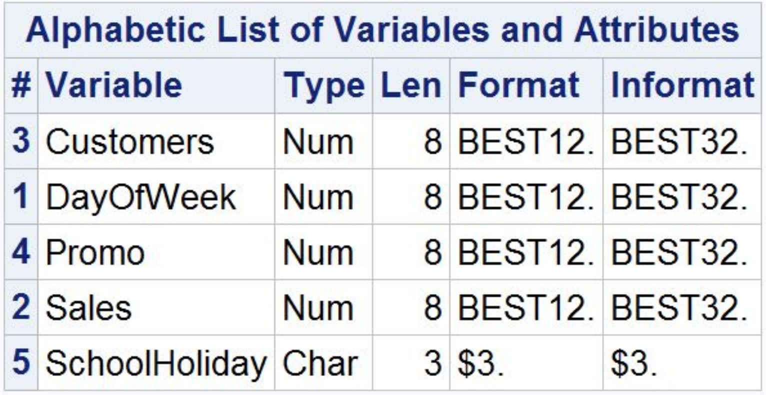
- Data preparation
  - Remove attributes - Store, Date, Open, StateHoliday (no StateHoliday in store_1 data)
  - Keep attributes - DayofWeek, Sales, Customers, Promo, SchoolHoliday



	DayOfWeek	Sales	Customers	Promo	SchoolHoliday
1	5	5263	555	1	1
2	4	5020	546	1	1
3	3	4782	523	1	1
4	2	5011	560	1	1
5	1	6102	612	1	1
6	6	4364	500	0	0
7	5	3706	459	0	0
8	4	3769	503	0	0
9	3	3464	463	0	0
10	2	3558	469	0	0
11	1	4395	526	0	0

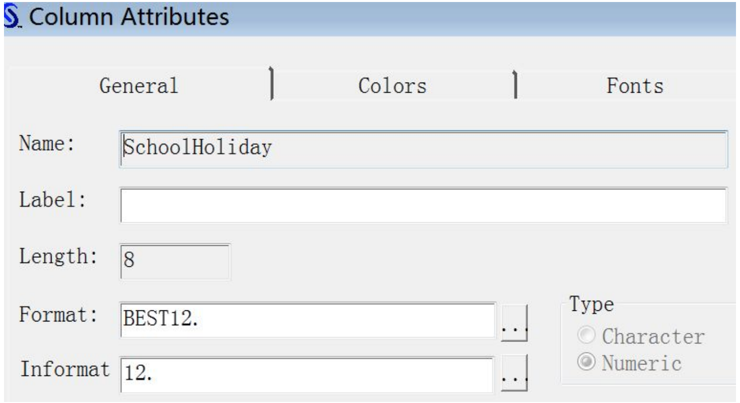
Store_1 sales data exploration

- Data preparation
  - Convert SchooloHoliday from Character to Numeric for future analysis



Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat
3	Customers	Num	8	BEST12.	BEST32.
1	DayOfWeek	Num	8	BEST12.	BEST32.
4	Promo	Num	8	BEST12.	BEST32.
2	Sales	Num	8	BEST12.	BEST32.
5	SchoolHoliday	Char	3	$3.	$3.



Column Attributes

Name: SchoolHoliday

Label:

Length: 8

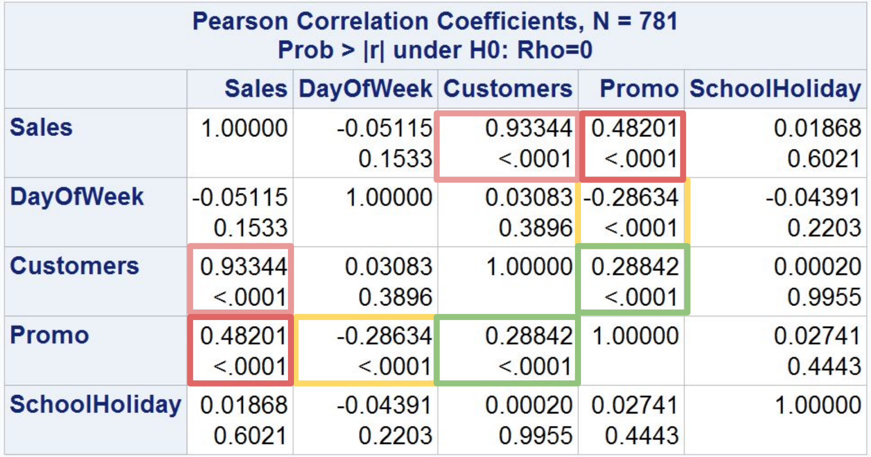
Format: BEST12.

Informat 12.

Type
○ Character
● Numeric

Store_1 sales data exploration

- Correlation between variables



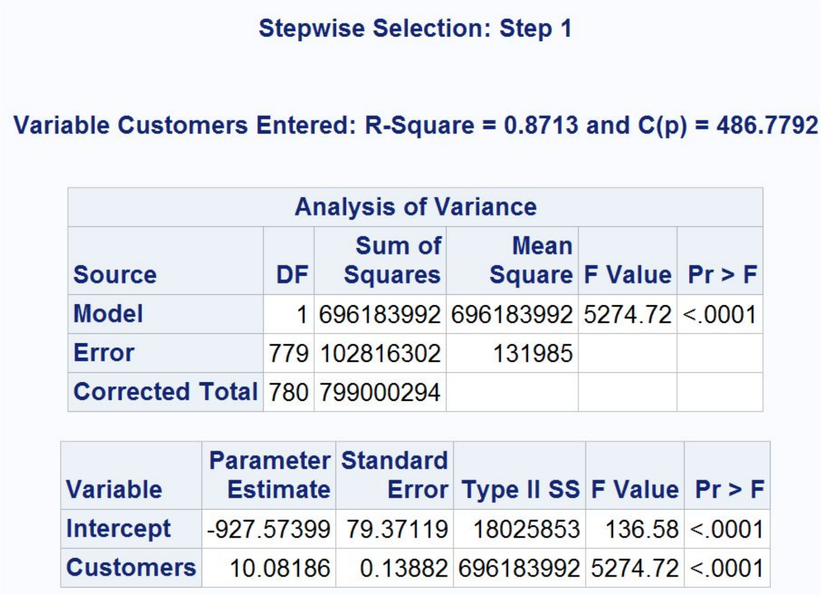
Pearson Correlation Coefficients, N = 781
Prob > |r| under H0: Rho=0

	Sales	DayOfWeek	Customers	Promo	SchoolHoliday
Sales	1.00000	-0.05115 0.1533	0.93344 <.0001	0.48201 <.0001	0.01868 0.6021
DayOfWeek	-0.05115 0.1533	1.00000	0.03083 0.3896	-0.28634 <.0001	-0.04391 0.2203
Customers	0.93344 <.0001	0.03083 0.3896	1.00000	0.28842 <.0001	0.00020 0.9955
Promo	0.48201 <.0001	-0.28634 <.0001	0.28842 <.0001	1.00000	0.02741 0.4443
SchoolHoliday	0.01868 0.6021	-0.04391 0.2203	0.00020 0.9955	0.02741 0.4443	1.00000

- Small strength of negative correlation between DayofWeek and Promo
- Small strength of positive correlation between Promo and Customers
- Large strength of positive correlation between Sales and Customers
- Medium strength of positive correlation between Sales and Promo

Store_1 sales data exploration

- Multiple regression
  - Stepwise regression
    - Combination of Forward and Backward
    - computationally efficient
- Step 1
  - Customers is the important predictor for sales prediction



Stepwise Selection: Step 1

Variable Customers Entered: R-Square = 0.8713 and C(p) = 486.7792

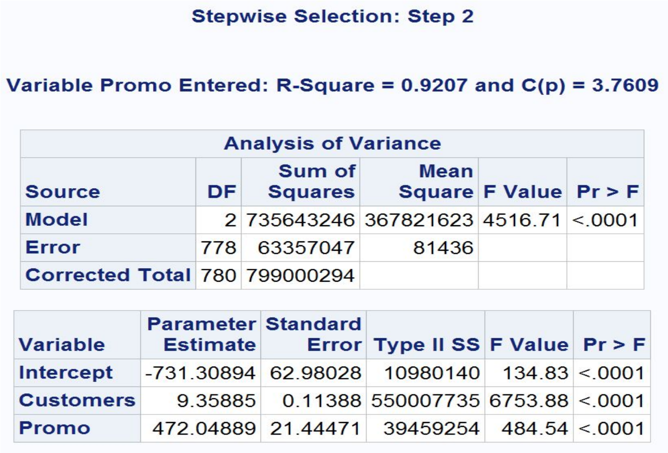
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	696183992	696183992	5274.72	<.0001
Error	779	102816302	131985		
Corrected Total	780	799000294			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-927.57399	79.37119	18025853	136.58	<.0001
Customers	10.08186	0.13882	696183992	5274.72	<.0001

Store_1 sales data exploration

- Step 2 - Customers and Promo are both the important predictors for sales prediction



Stepwise Selection: Step 2

Variable Promo Entered: R-Square = 0.9207 and C(p) = 3.7609

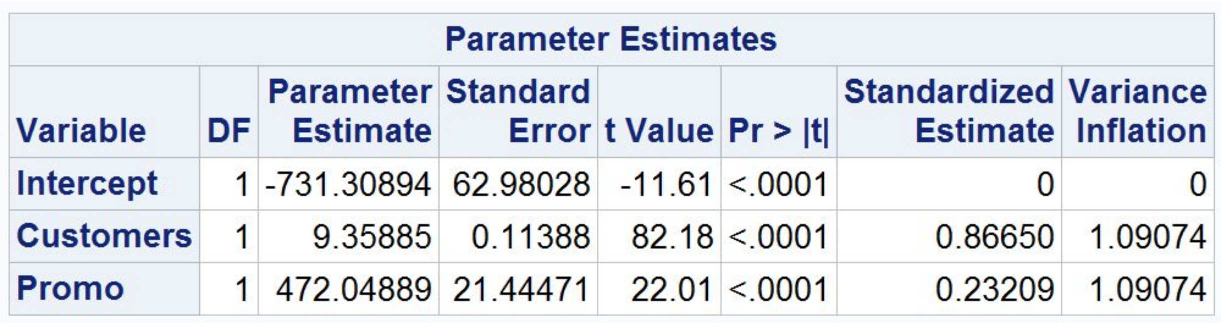
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	735643246	367821623	4516.71	<.0001
Error	778	63357047	81436		
Corrected Total	780	799000294			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-731.30894	62.98028	10980140	134.83	<.0001
Customers	9.35885	0.11388	550007735	6753.88	<.0001
Promo	472.04889	21.44471	39459254	484.54	<.0001

- No other variables met the significant level for entry into the model

Store_1 sales data exploration

- Multiple regression - Stepwise



Parameter Estimates

		Parameter	Standard			Standardized	Variance
Variable	DF	Estimate	Error	t Value	Pr > |t|	Estimate	Inflation
Intercept	1	-731.30894	62.98028	-11.61	<.0001	0	0
Customers	1	9.35885	0.11388	82.18	<.0001	0.86650	1.09074
Promo	1	472.04889	21.44471	22.01	<.0001	0.23209	1.09074

- Equation for calculating sales for store_1 from the model
  - Sales = -731.30894 + 9.35885*Customers + 472.04889*Promo

Store_1 sales data exploration

- Regression with Interactions of variables
    - manually computing the interaction among variables
        - Day_C=DayofWeek*Customers

          Day_P=DayofWeek*Promo

          Day_S=DayofWeek*SchoolHoliday

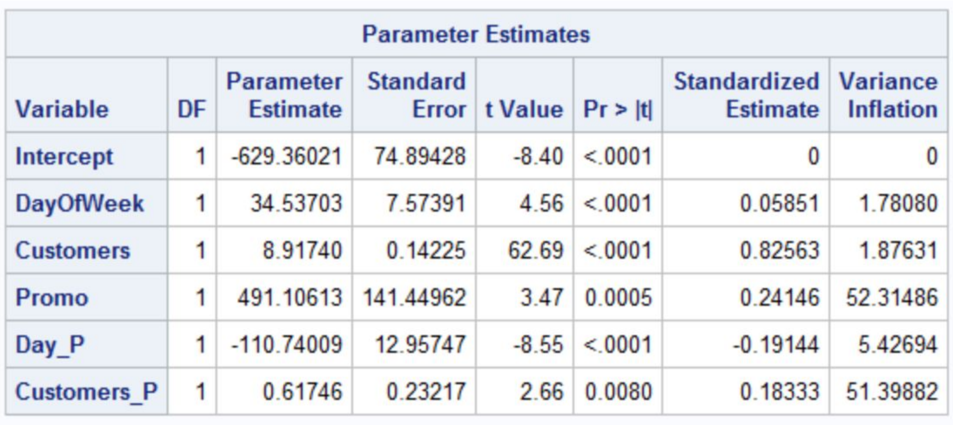
          Customers_P = Customers*Promo

          Customers_s = Customers*SchoolHoliday

          P_Sch=Promo*SchoolHoliday

Store_1 sales data exploration

- Regression with Interactions of variables -- Stepwise



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate	Variance Inflation
Intercept	1	-629.36021	74.89428	-8.40	<.0001	0	0
DayOfWeek	1	34.53703	7.57391	4.56	<.0001	0.05851	1.78080
Customers	1	8.91740	0.14225	62.69	<.0001	0.82563	1.87631
Promo	1	491.10613	141.44962	3.47	0.0005	0.24146	52.31486
Day_P	1	-110.74009	12.95747	-8.55	<.0001	-0.19144	5.42694
Customers_P	1	0.61746	0.23217	2.66	0.0080	0.18333	51.39882

- Equation with interactions of variables for predicting sales for store_1

Sales =

-629.36021+34.53703*DayofWeek+8.91740*Customers+491.10613*Promo-110.74009*DayofWeek*Promo+0.61746*Customers*Promo

Conclusion

- Suggestions for Rossman based on our findings
  - Store type analysis
    - Increasing marketing efforts on store type A to maximize sales
    - Provide most essentials in store type B to enlarge the sales amount
    - Stores type D should have a good stock of items to avoid any demand-supply gap
  - Store_1 sales
    - Create effective staff schedules that increase productivity and motivation based on predicted future sales for store_1
  - Forecasting
    - Seasonal Exponential Smoothing methodology recommended

Future improvement

- Predict sales for other individual stores to see if there are different patterns of sales
- Add more variables from store information to increase the accuracy
- Try different approaches to get the optimal model
- Build non-linear models such as tree-based model

Thank you!