Director: Harshita

Producer: I-ting

Starring:

Harshita (Business Analyst)

Saran (Database Administrator)

I-Ting (Data Scientist)

Ghazal (Daughter of the owner of the drug store chain)

&

(Business Analyst)

Dialogues: Ghazal

Graphics: Saran

Few hours later..

DRUG STORE SALES

Business Analytics with R

Group – 9

PROJECT SCOPE

DRUG STORE

- Unique to Europe
- Operating over 3000 stores across 7 countries

Predicting Sales for 1115 Locations

Increased productivity and motivation

DATA:

Training Set : Store, Date , Promotional Periods, Sales, Customers, State and School Holidays

Store Data: Competition Data, Store Type, Assortment.

1.
Understanding the Data

Let’s try to understand the entire data

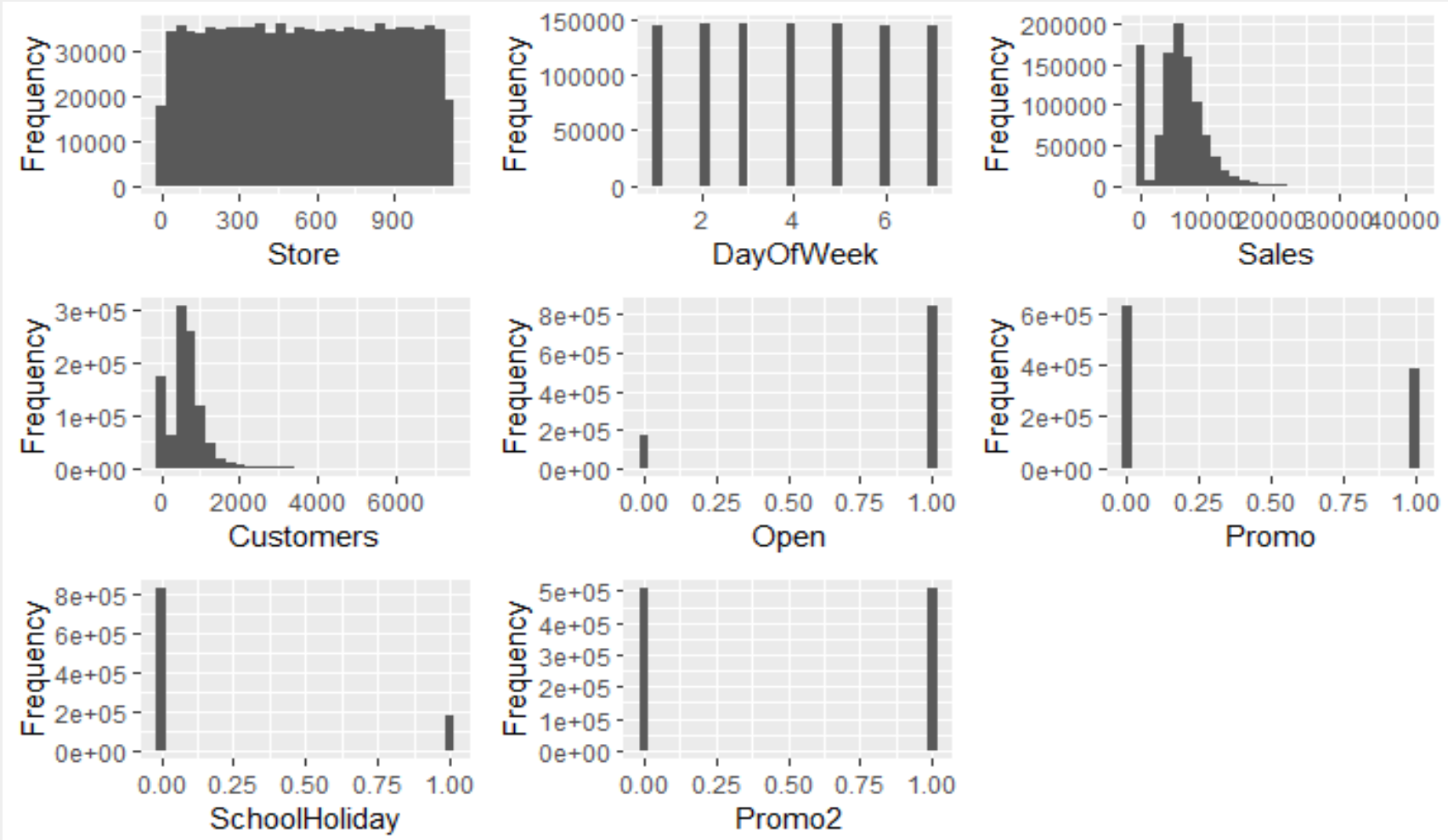
Store	0
DayOfWeek	0
Date	0
Sales	0
Customers	0
Open	0
Promo	0
StateHoliday	0
SchoolHoliday	0
StoreType	0
Assortment	0
CompetitionDistance	2642
CompetitionOpenSin...	323348
CompetitionOpenSin...	323348
Promo2	0
Promo2SinceWeek	508031
Promo2SinceYear	508031
PromoInterval	508031

Checking Data Quality:

Null values check, outlier detection and/or anomalies

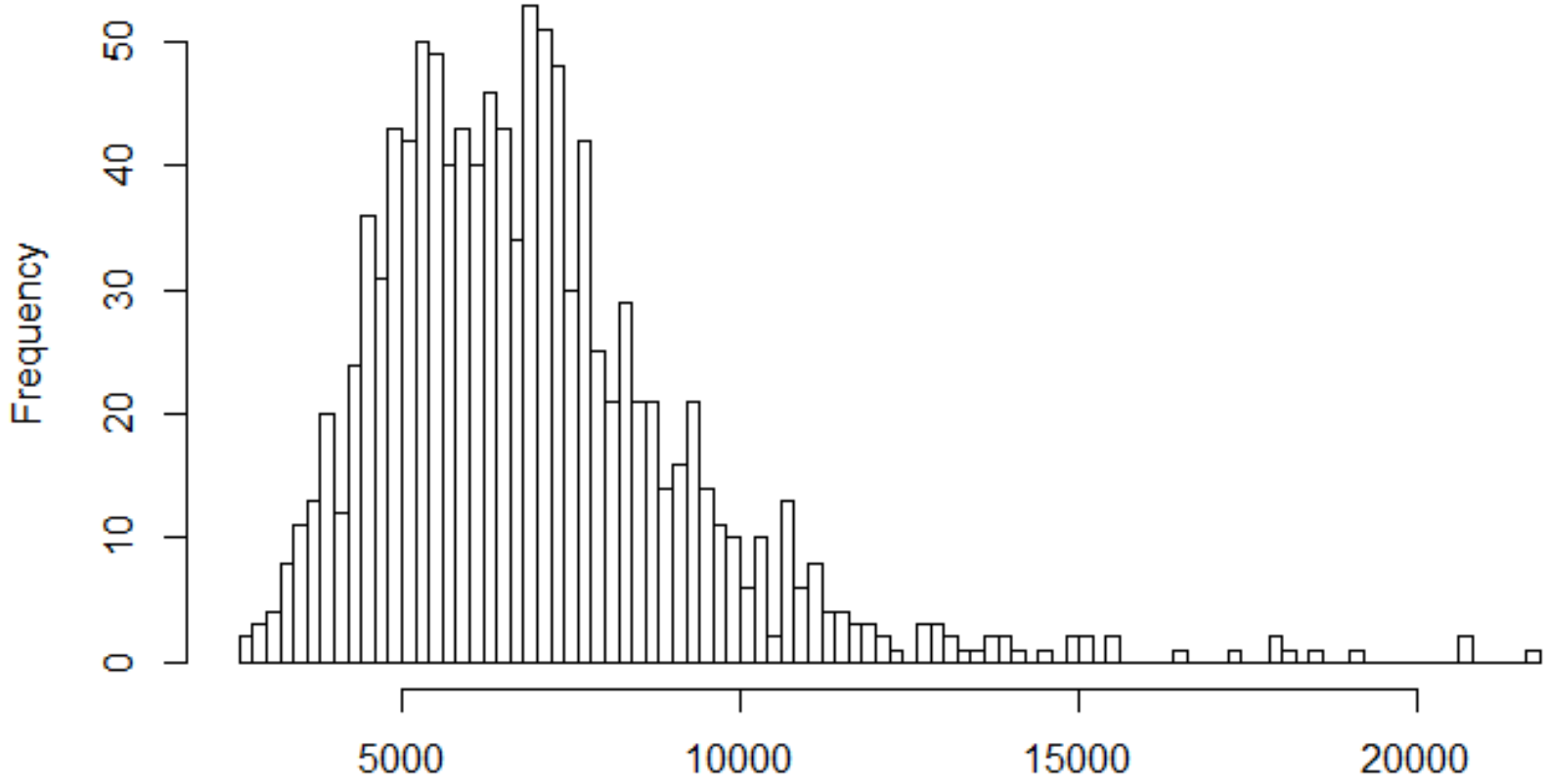
Does the type of store, Promos and Holidays have anything to do with SALES??

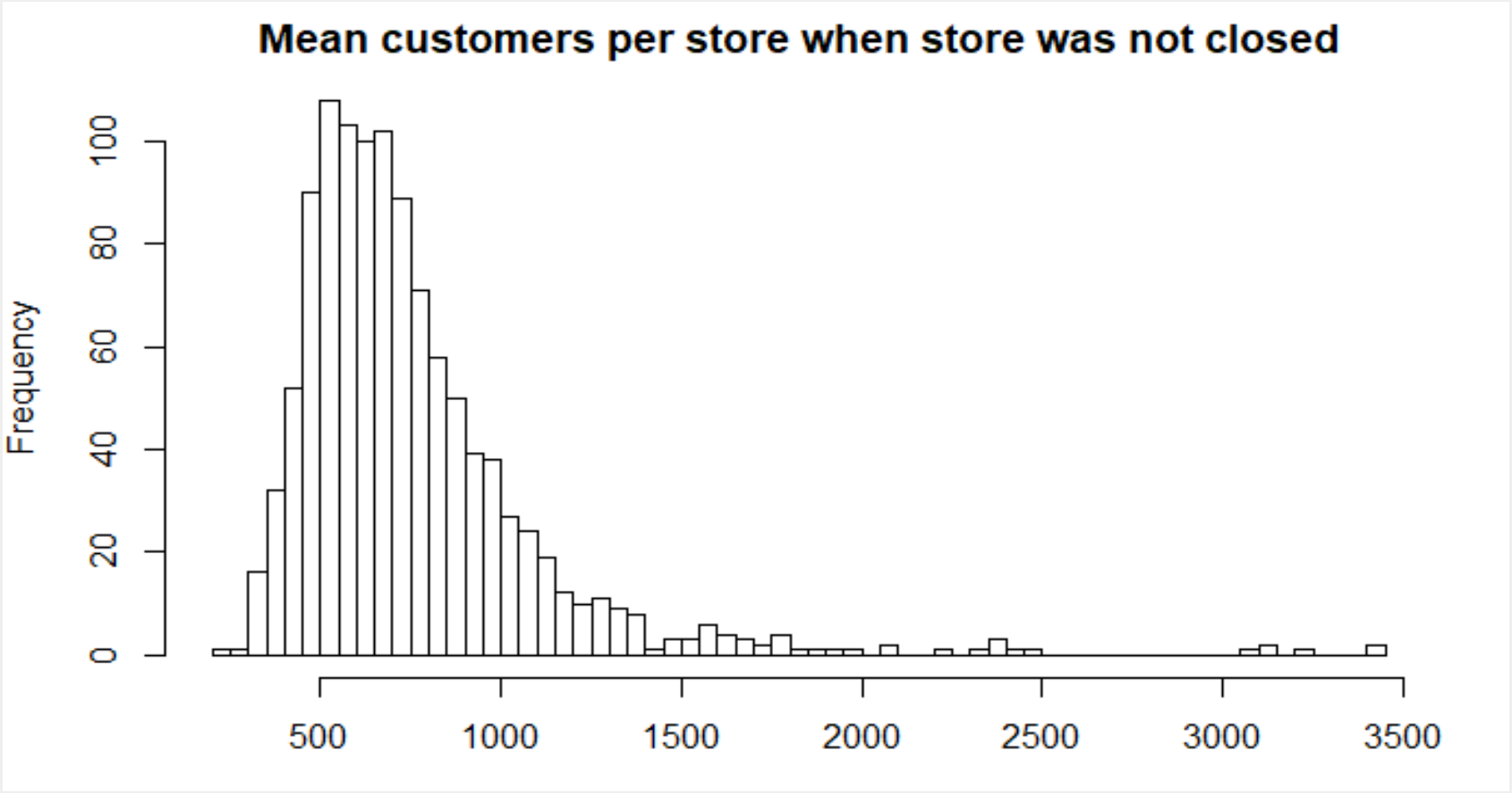
We will try to see some relations now!

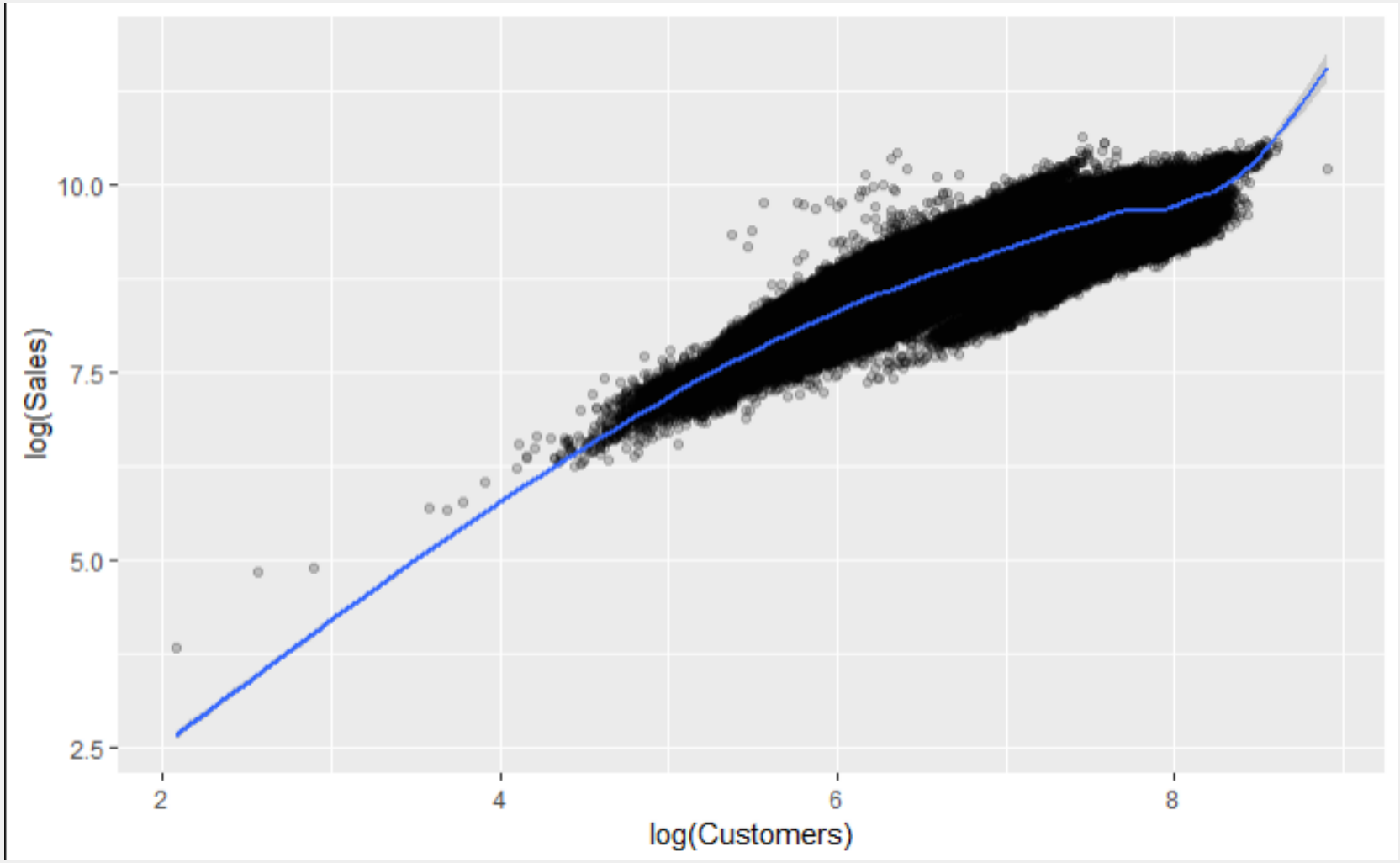


We can see the type of each variable here.

MEAN SALES PER STORE WHEN STORE WAS NOT CLOSED

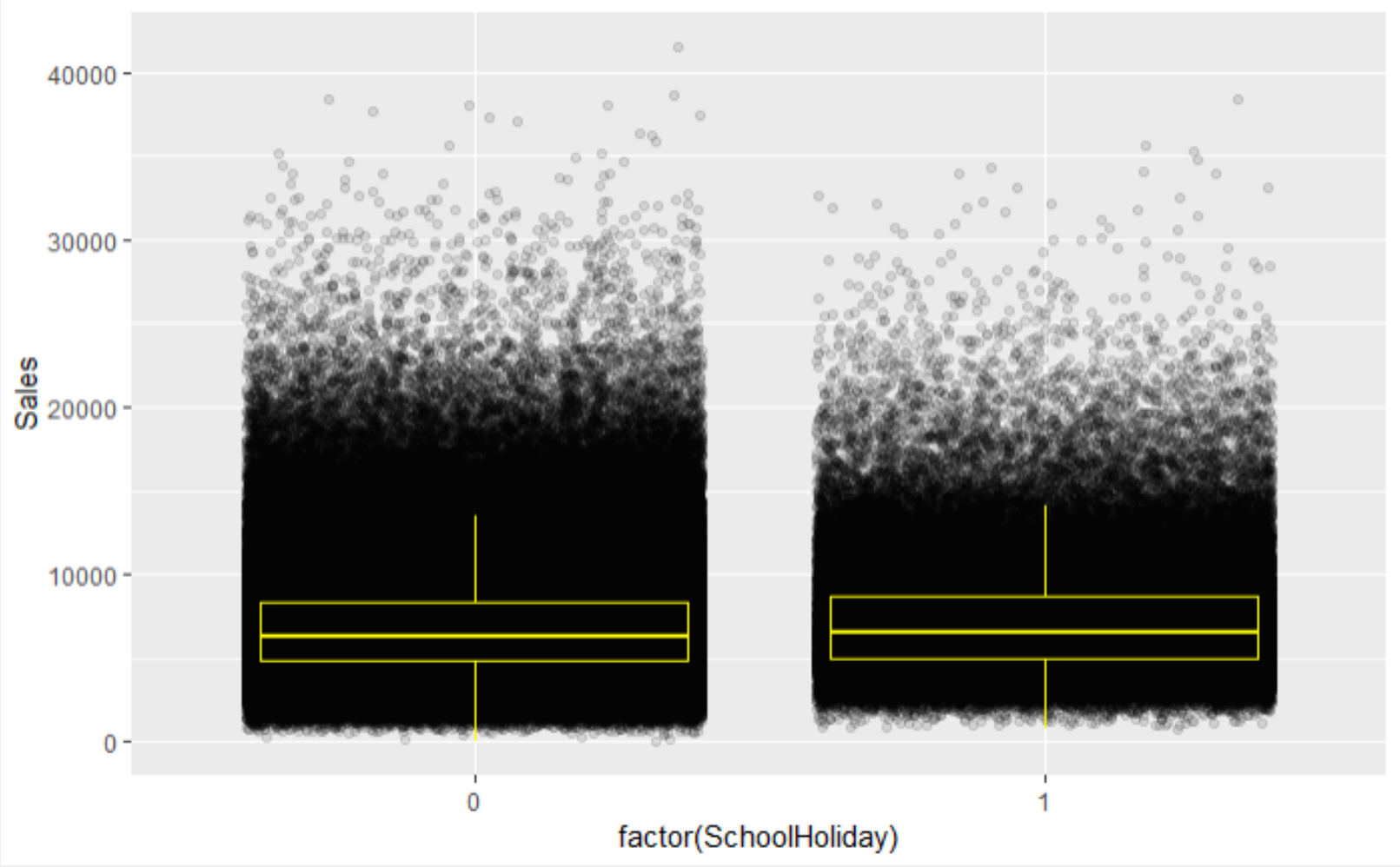



Mean customers per store when store was not closed
Frequency
0
20
40
60
80
100
500
1000
1500
2000
2500
3000
3500

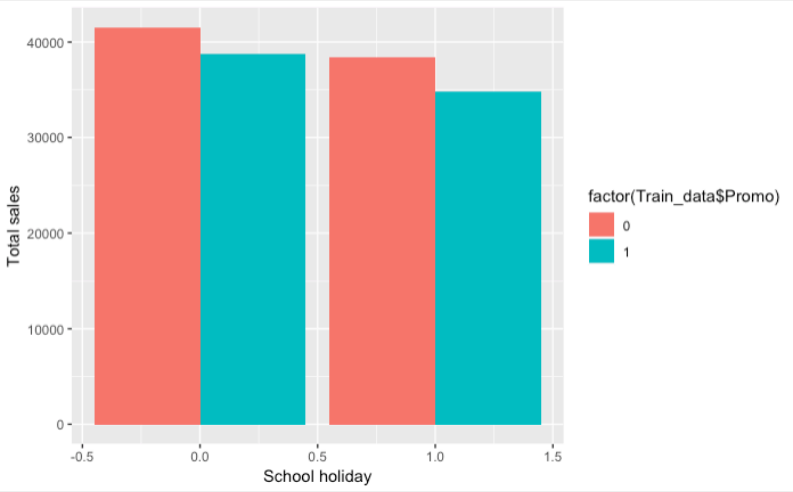


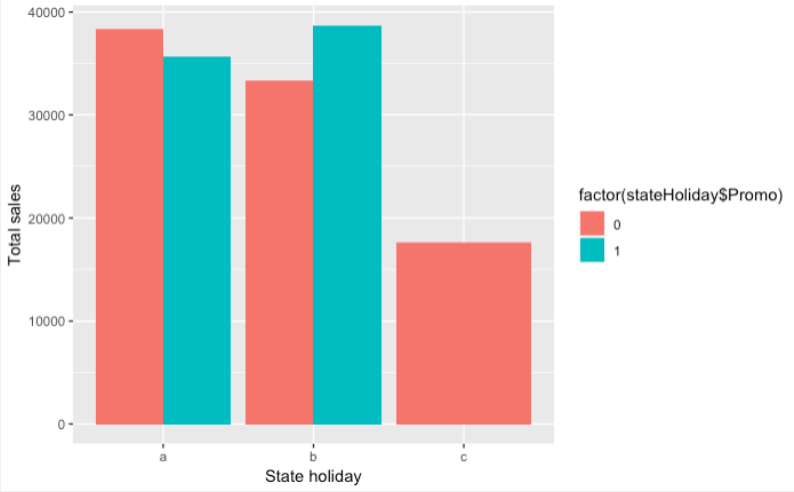
As expected Sales is strongly correlated to no of Customers!!

Check how sales are affected when there is a School Holiday.



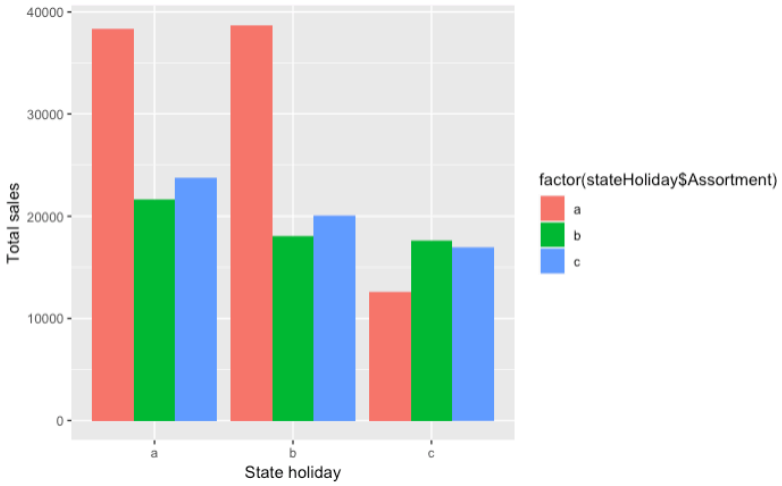
How sales are getting affected during Sales and State Holiday

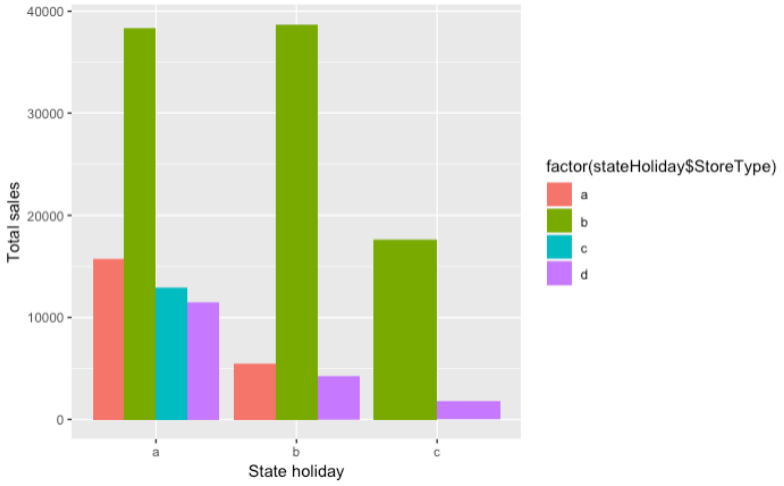




No promotion was running on any stores on Christmas (Type c)

What effect does State Holiday have on different Assortments and Store Types?





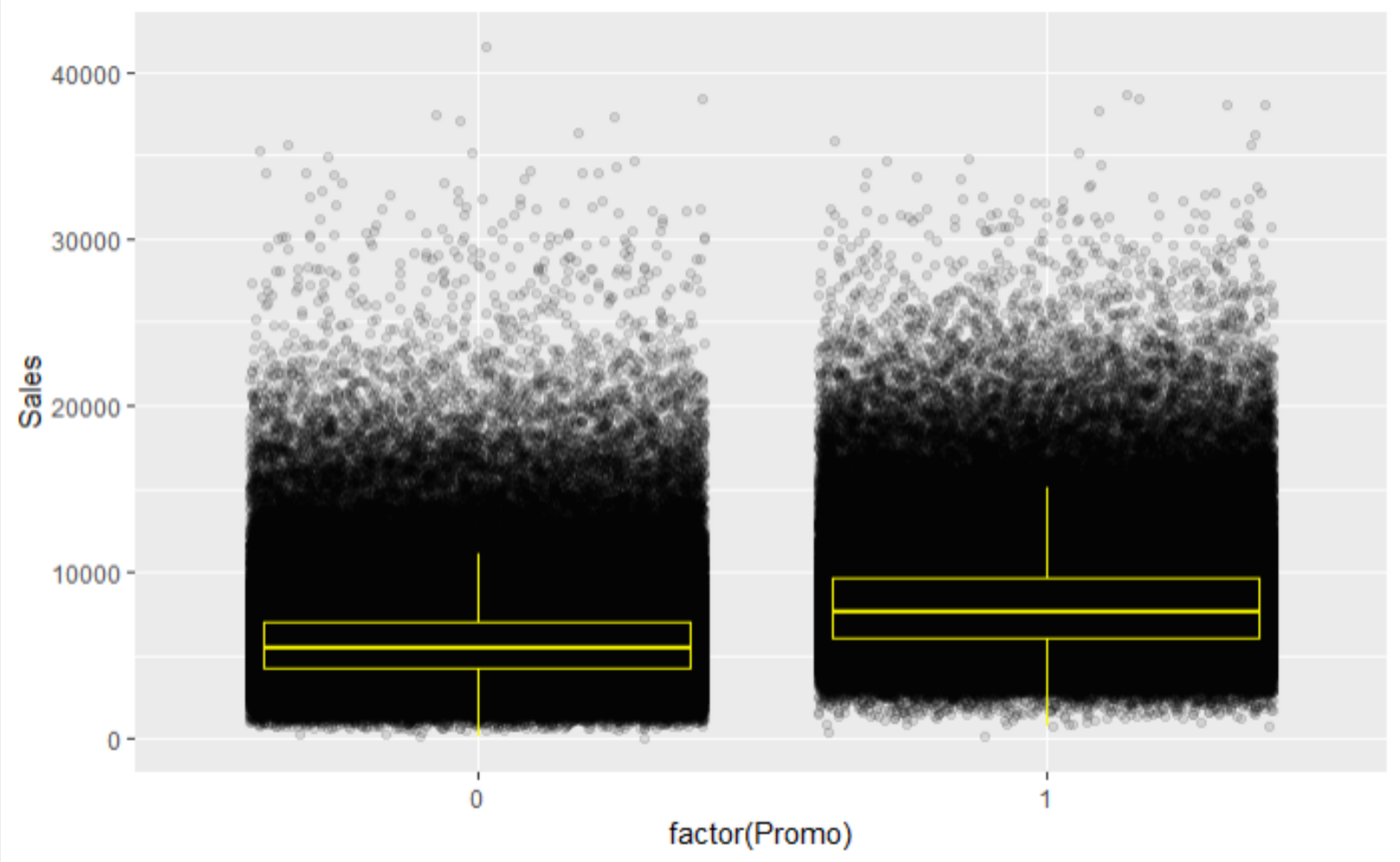
A – Public holiday
B – Easter
C - Christmas

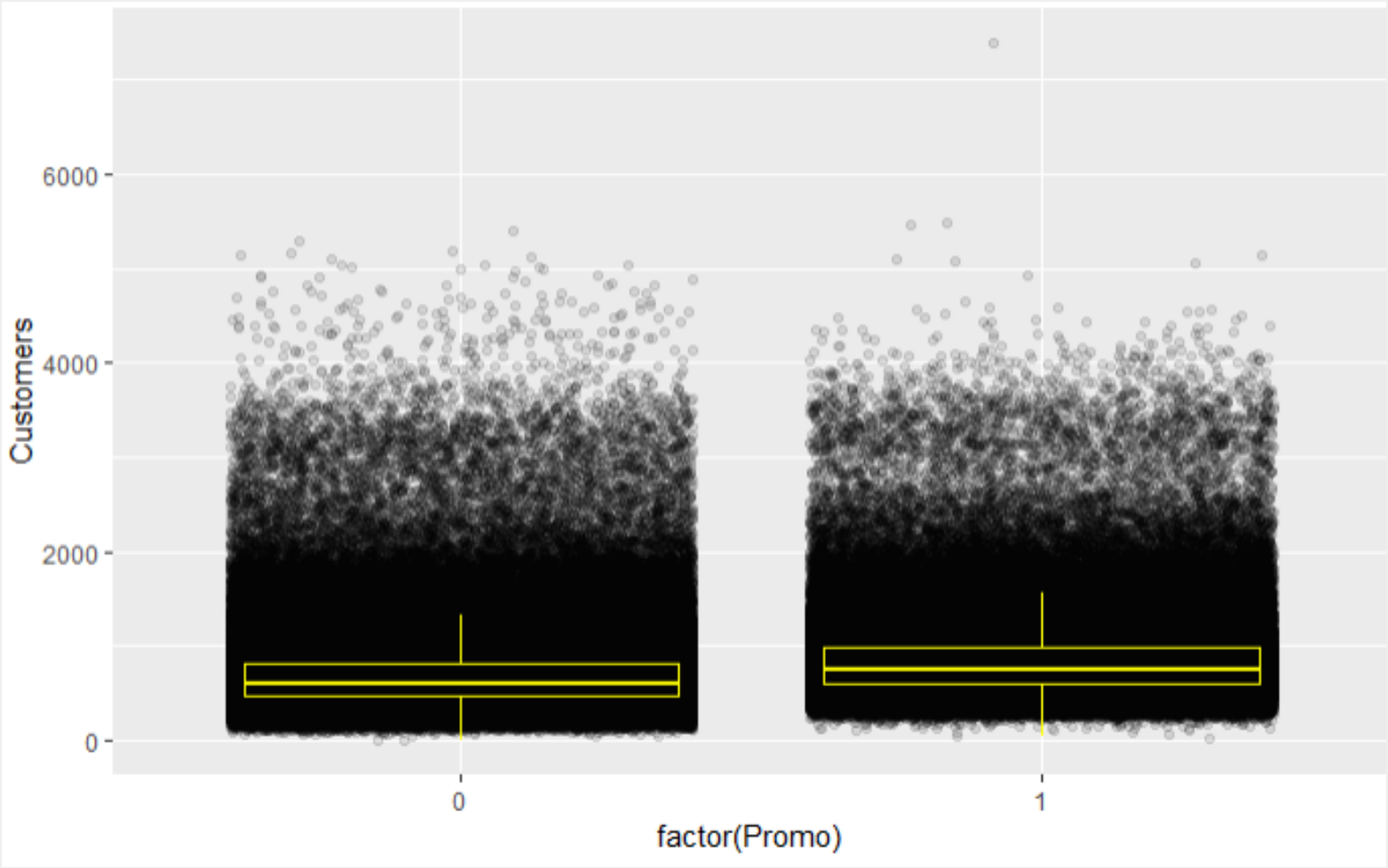
Only small parts of stores open on state holiday, so we zoom in to. Check

All three kinds of Assortments sell on all the State Holidays!

Type C not opened on Christmas and Easter day

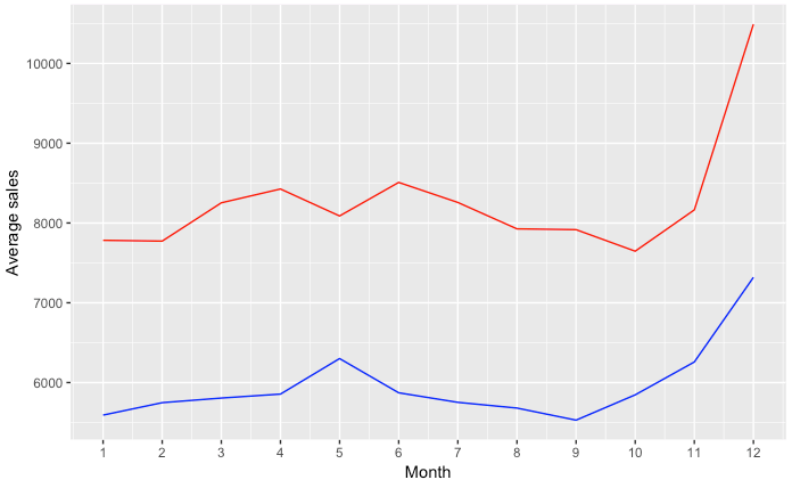
Check how sales are affected when there is a Promotional Period.



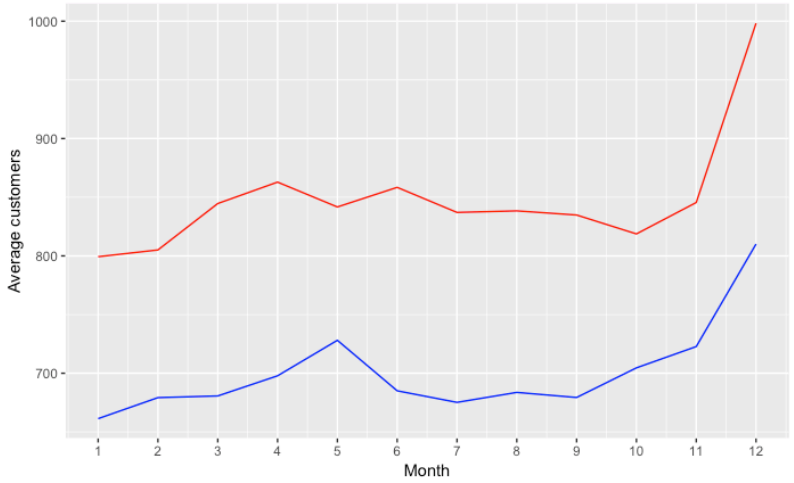


It looks like Boxplots of customers overlap a little more than the boxplots of sales. This would mean that the promos are not mainly attracting more customers but make customer spend ore. The mean amount spent per customer is about one Euro Higher.

Average sales/customers - promotion

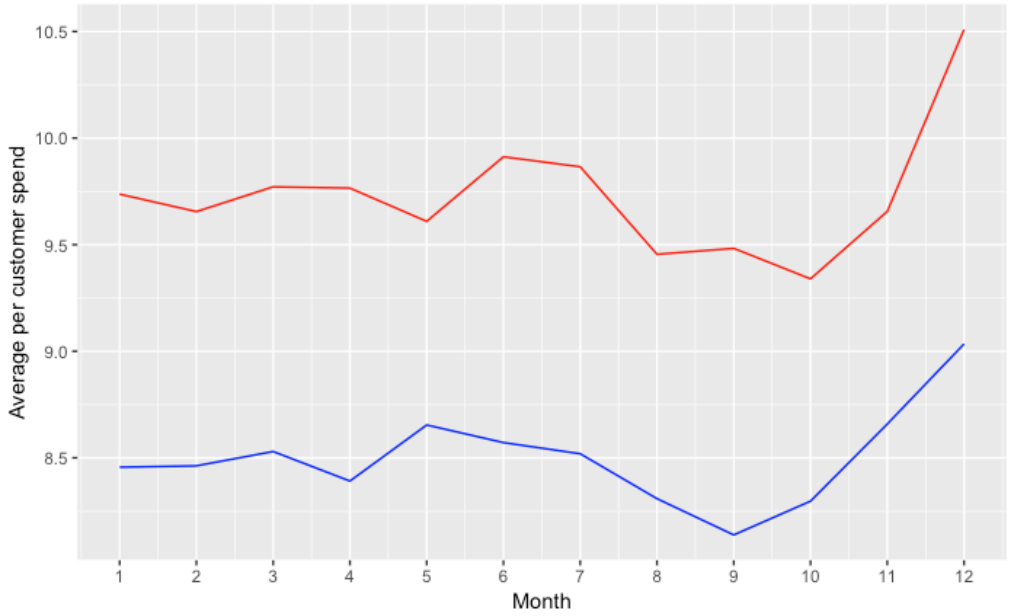


Red: participate promotion
Blue: do not participate promotion



Sales and customer with promotion is more higher than no promotion, especially in December.

Average per customer spend in store - promotion



Red: participate promotion
Blue: do not participate promotion

1. Per customer spend with promotion is slightly higher than without promotion.
2. But having promotion in Dec. seems a good idea.

There are no sales when the store is closed but there are some stores that are open but have no sales even if they had some customers. These observations may be error in the data or outliers.

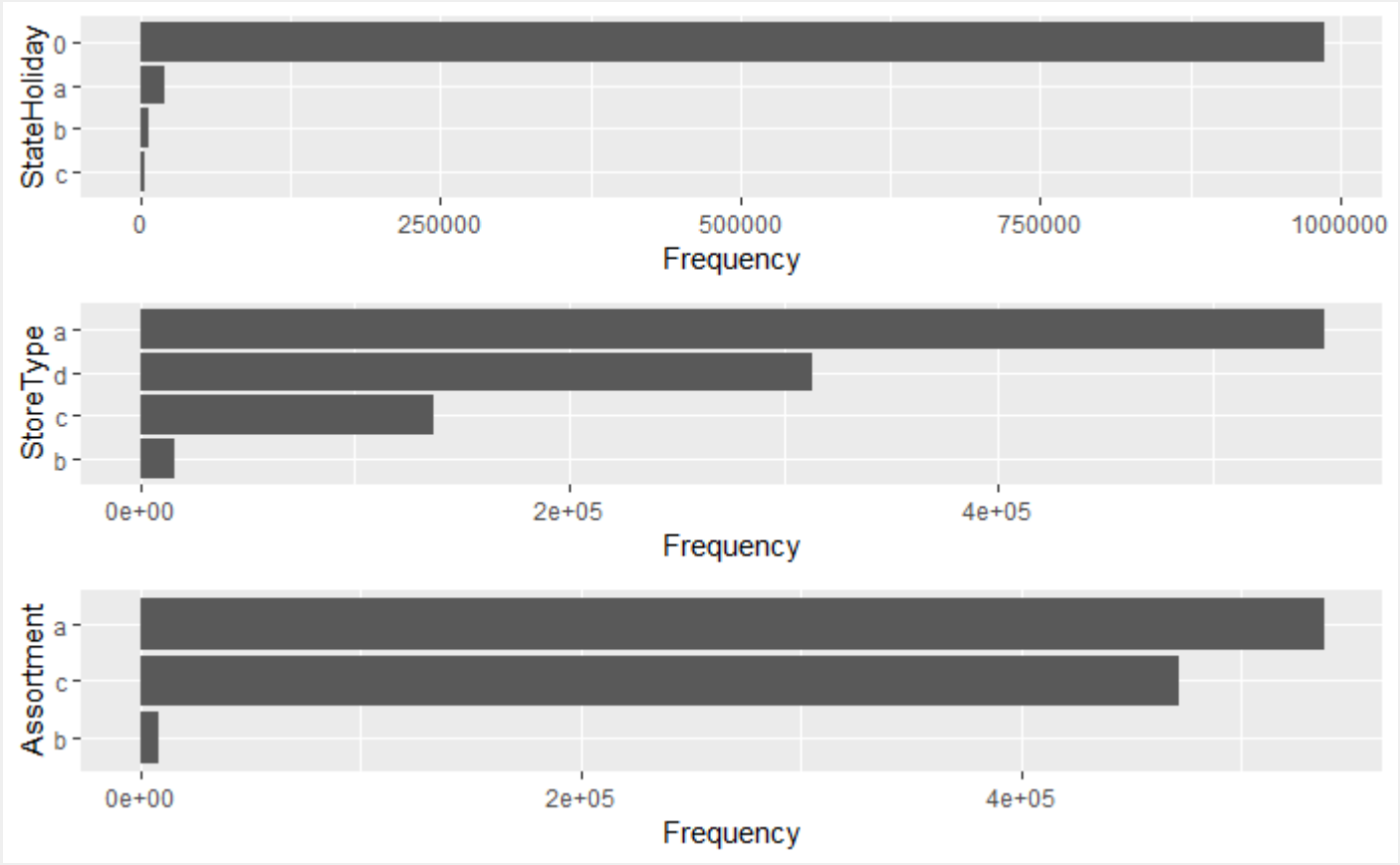
	No Promo	Promo	
Sales = 0	161666	11205	
Saes>0	467463	376875	

	Sales = 0	Sales>0	
Closed	172817	0	
Open	54	844338	

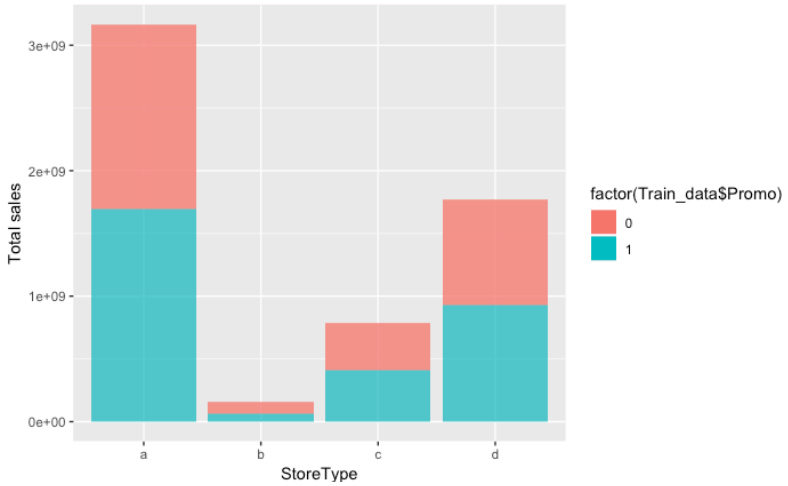
	Customers = 0	Customers >0	
Closed	172817	0	
Open	52	844340	

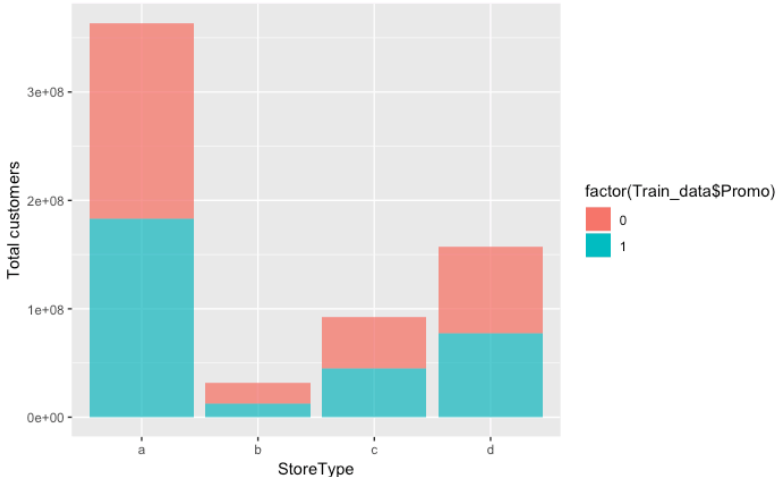
	Sales = 0	Sales>0	
Easter	142731	843428	
Public	30140	910	

	Customers = 0	Customers>0	
Sales = 0	172869	2	
Sales>0	0	844338	


StateHoliday
0
a
b
c
0
250000
500000
750000
1000000
Frequency
StoreType
a
d
c
b
0e+00
2e+05
4e+05
Frequency
Assortment
a
c
b
0e+00
2e+05
4e+05
Frequency

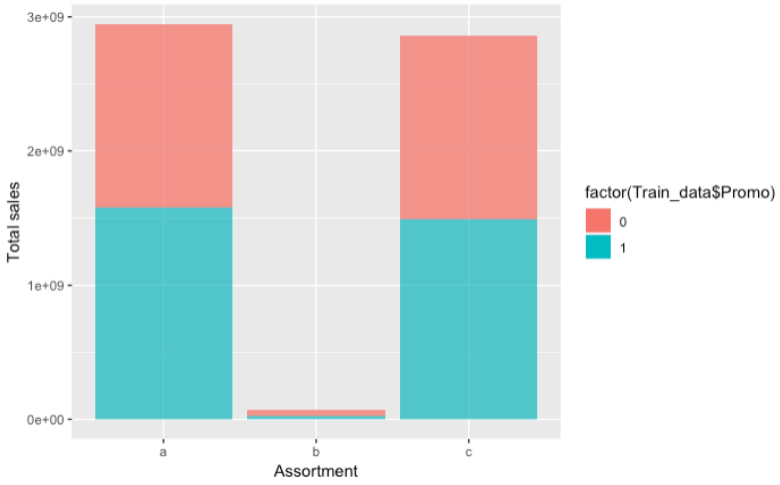
Total sales/customers by store types

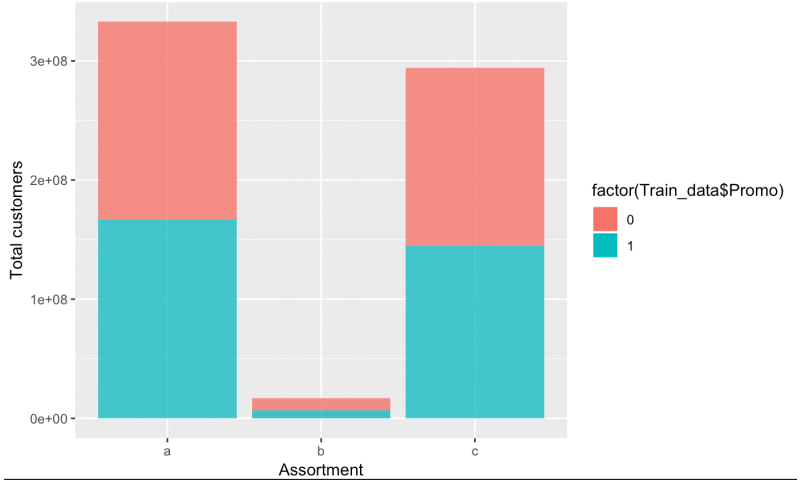




Sales and sales per customer spend of store type B are low, but the average is higher than others, because number of type B store is very few.

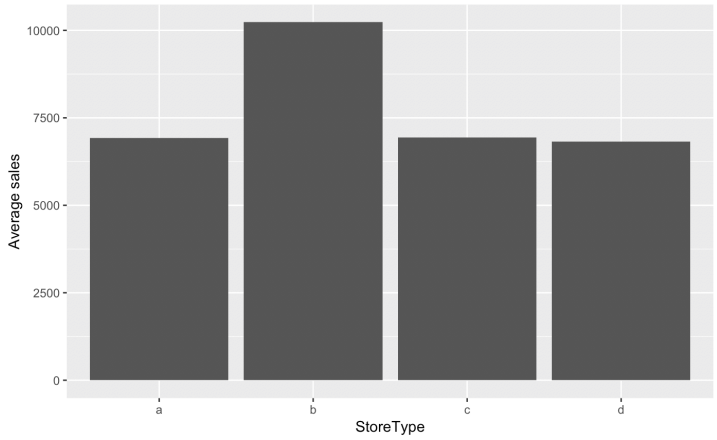
Total sales/customers by assortment

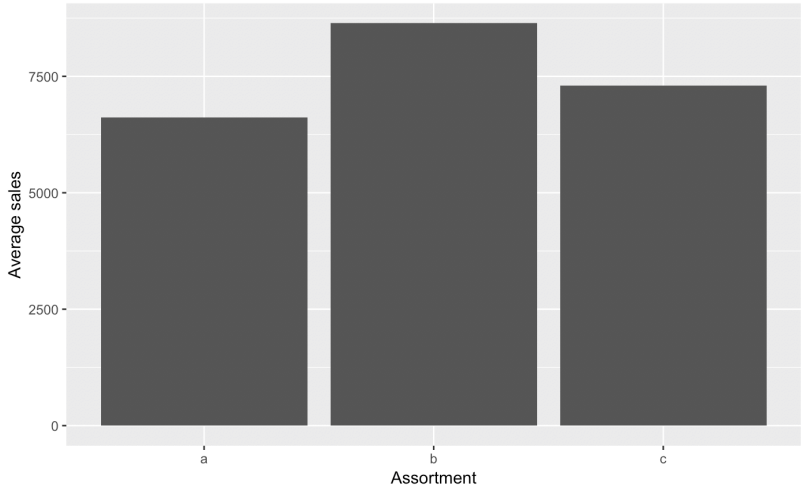




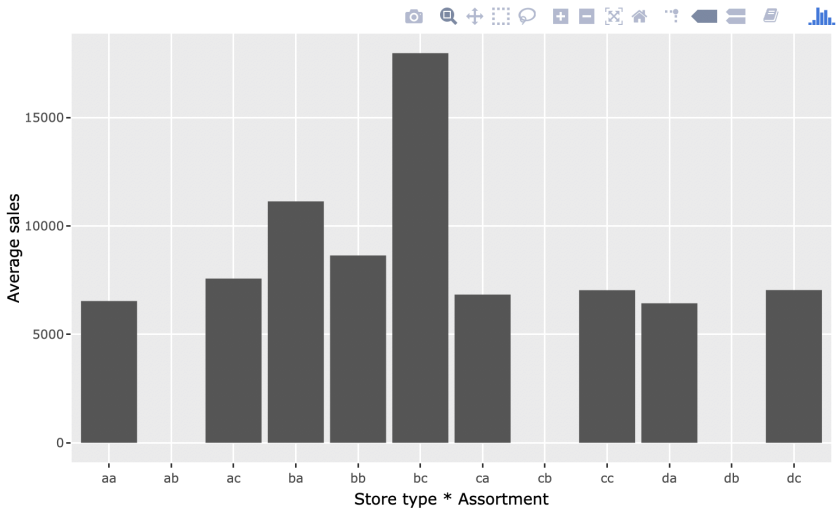
Sales of assortment B has the same pattern as sales of store type B.

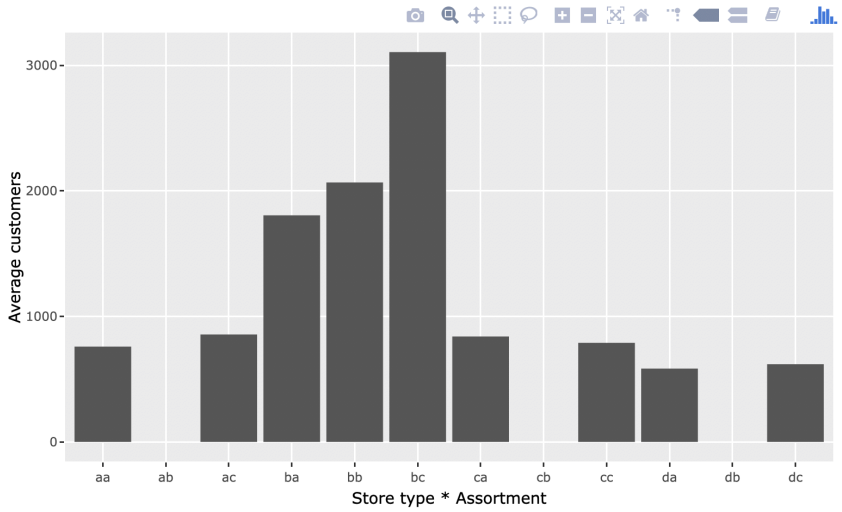
Total sales/customers by store types/assortment



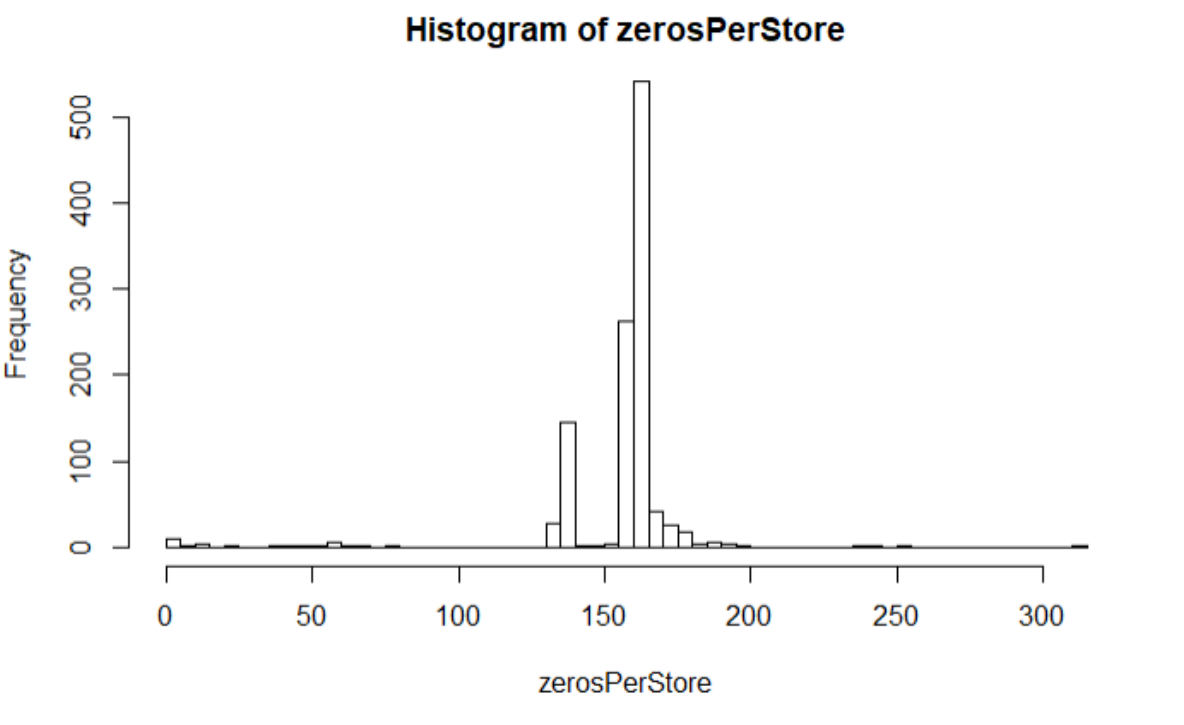


Store type * assortment

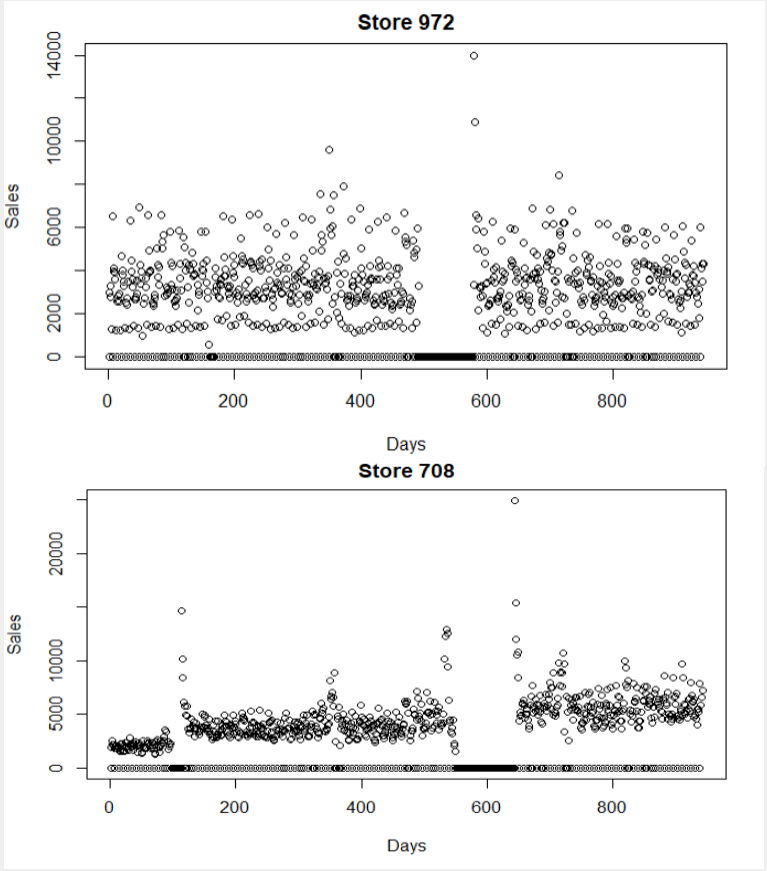


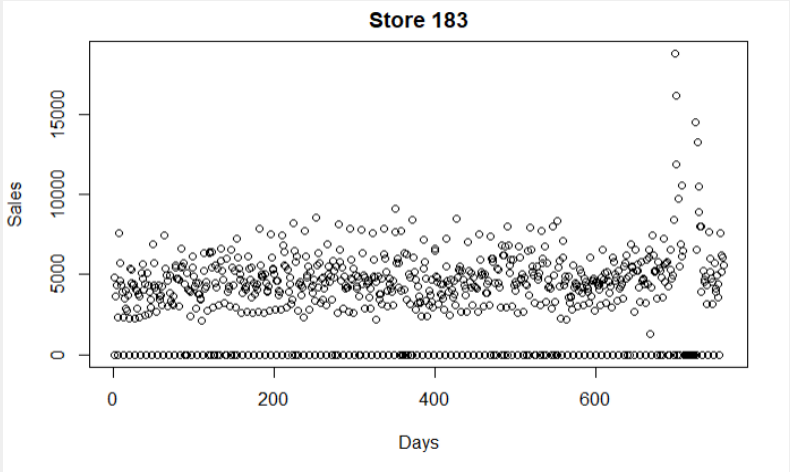


1. It seems like the combination of store type B and assortment C has the highest average sales and customers.
2. Only store type B has assortment B.

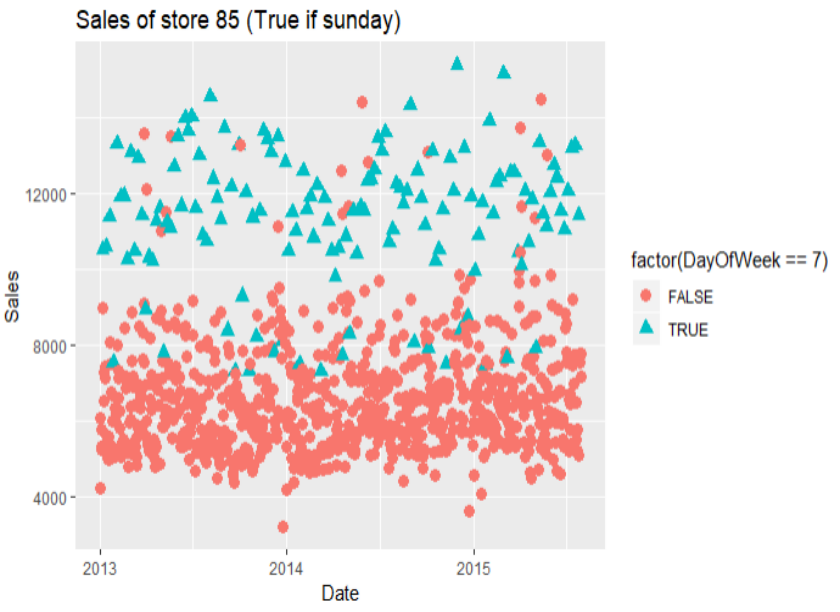


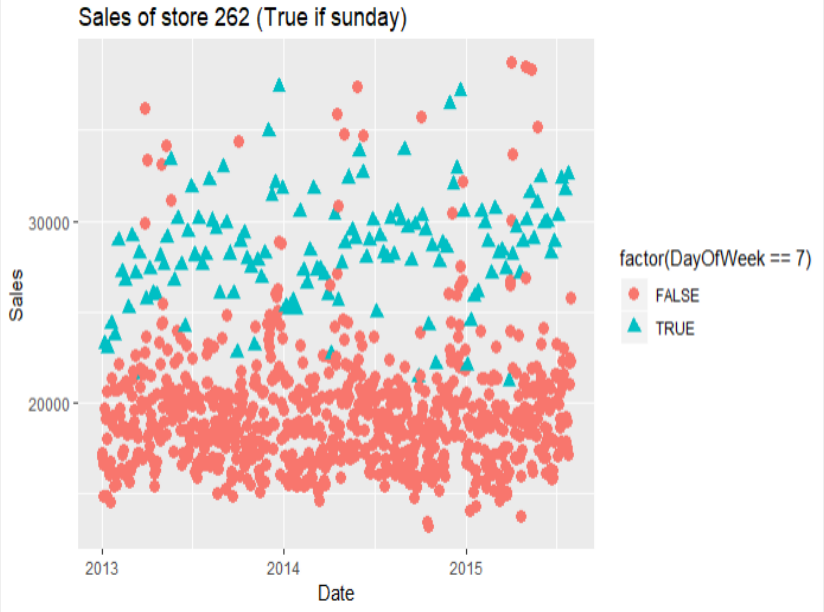
Top 10 stores with most number of zero sales:
105, 339, 837, 25, 560, 674, 972, 349, 708, 103, 188, 188, 191, 192, 195, 197, 240, 242, 255, 311



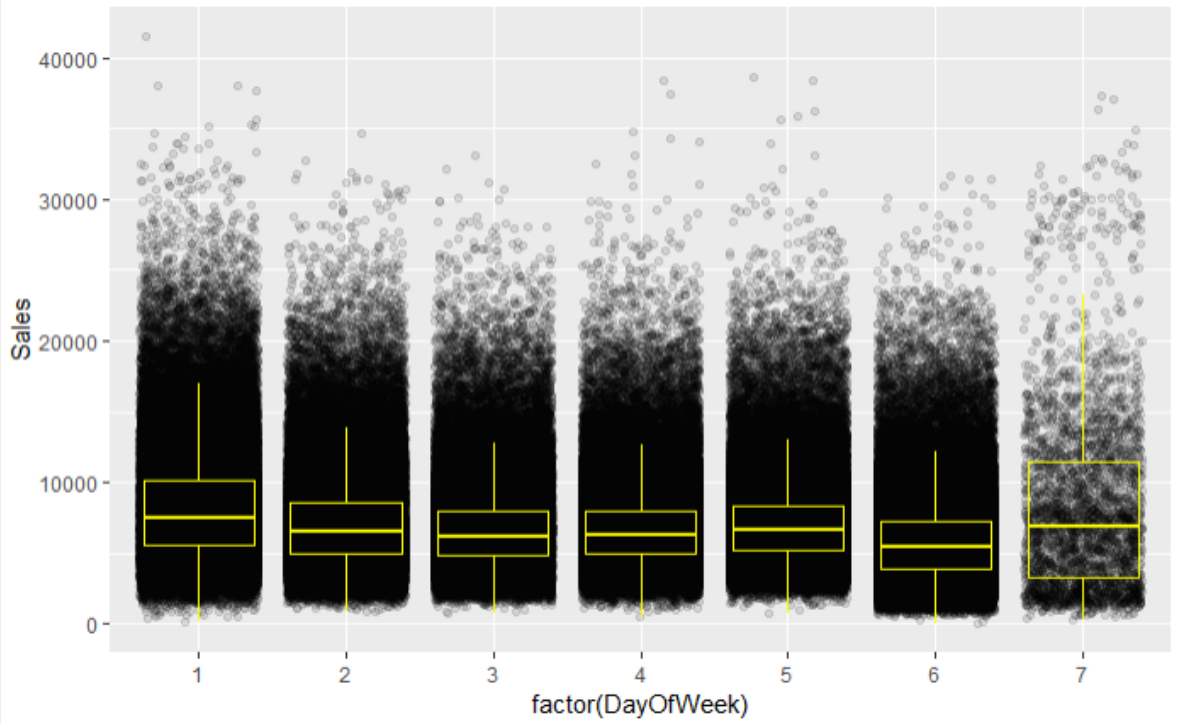


The store have different amounts of days with zero sales. There are spikes in the sales before the stores close and after the stores reopen.



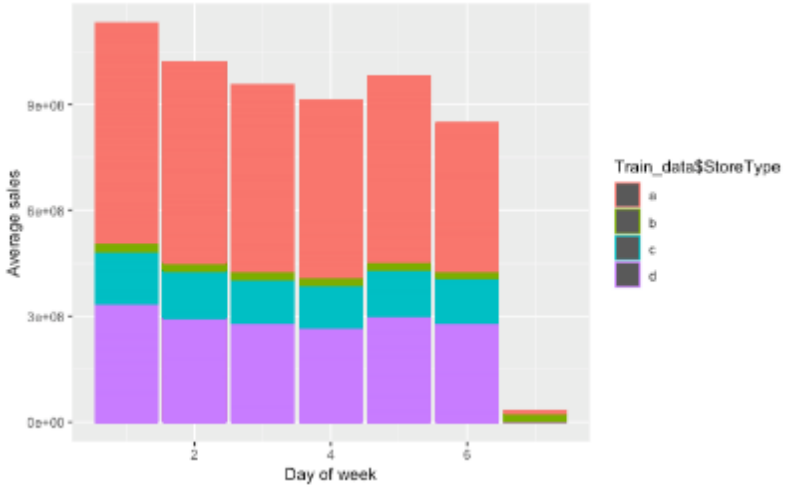


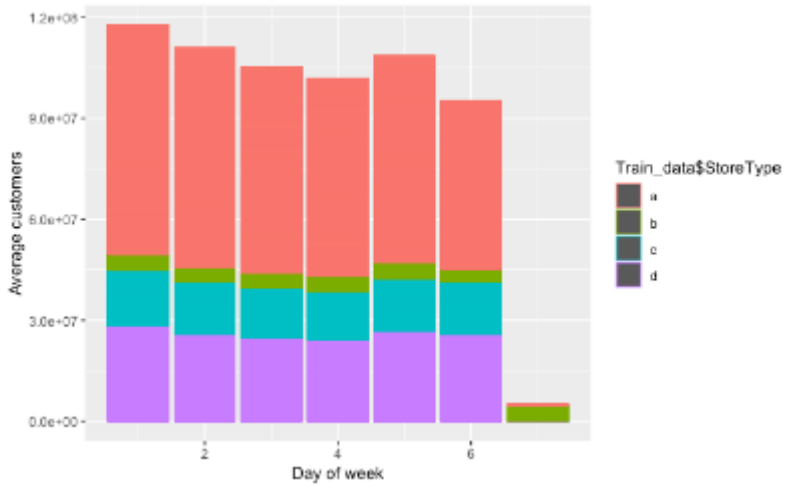
There are stores that have no zeros in their sales. these are exceptions because they are opened also on Sundays/ holidays.



There is a lot more variability on Sundays than on other days, despite median being almost the same.

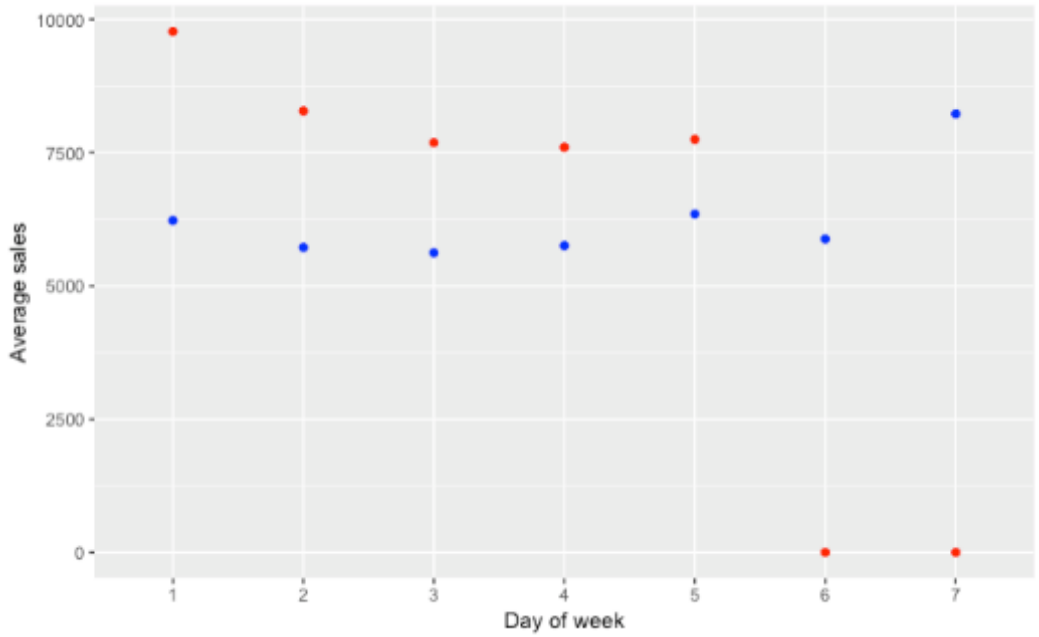
Total sales/customers by day of week





1. While Sunday has the lowest sales and customer, its sales and customer mean are the highest. It may because there's only few stores opened.
2. It seems like Monday has the highest sales and customer, and its sales and customer mean are also very high. Also, per customer spend more money on Monday when stores have promotion.

Average sales – day of week



Red: participate promotion
Blue: do not participate promotion

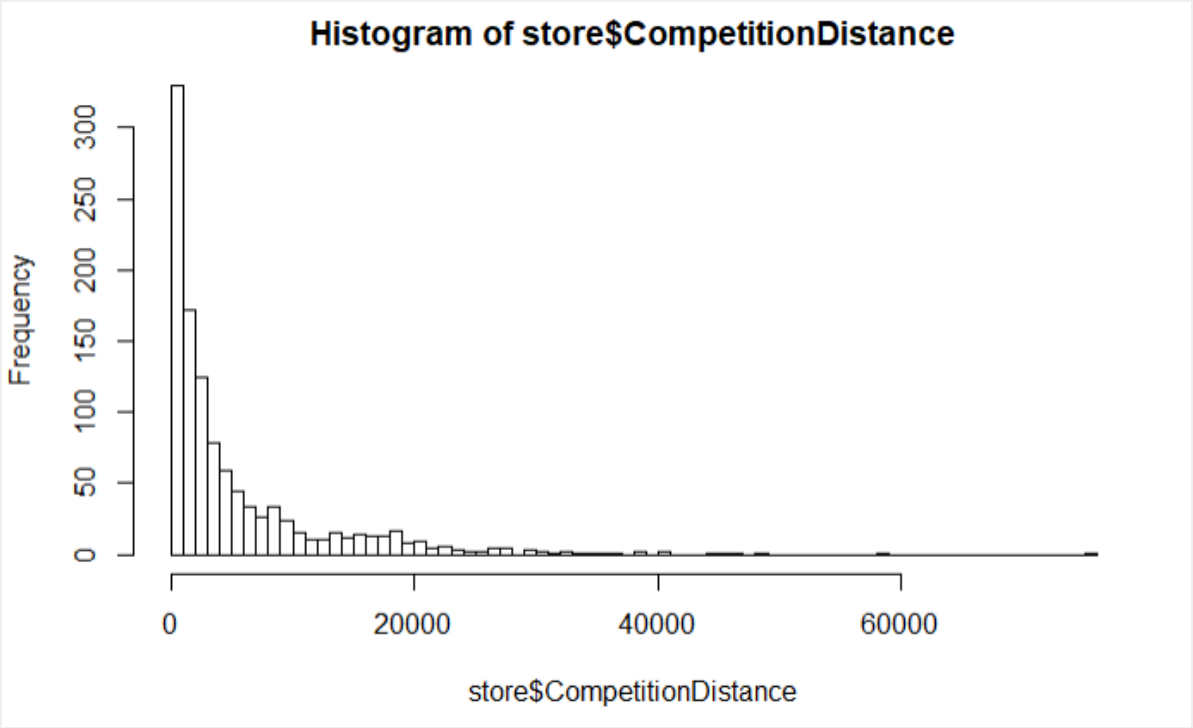
Promotion definitely affect sales and customers, surprisingly Saturday and Sunday don't have promotion.

Store Type			
a	b	c	d
602	17	148	348

Store Assortment		
a	b	c
593	9	513

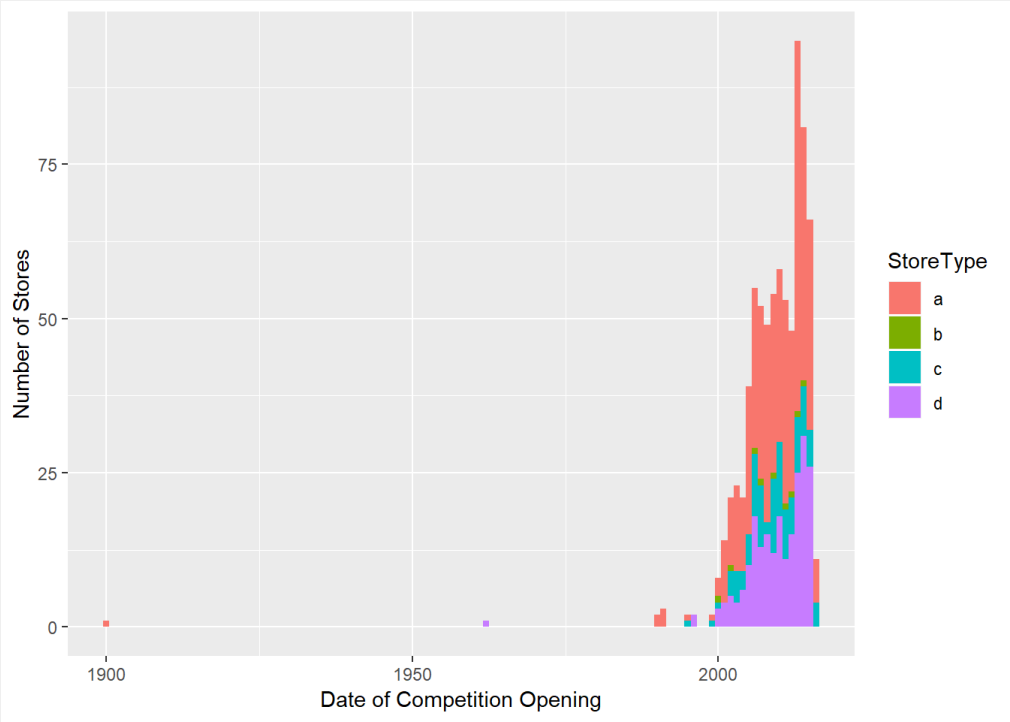
	Store Type				
Assortment		a	b	c	d
	a	381	7	77	128
	b	0	9	0	0
	c	221	1	71	220

There seems to be a connection between the StoreType and Assortment

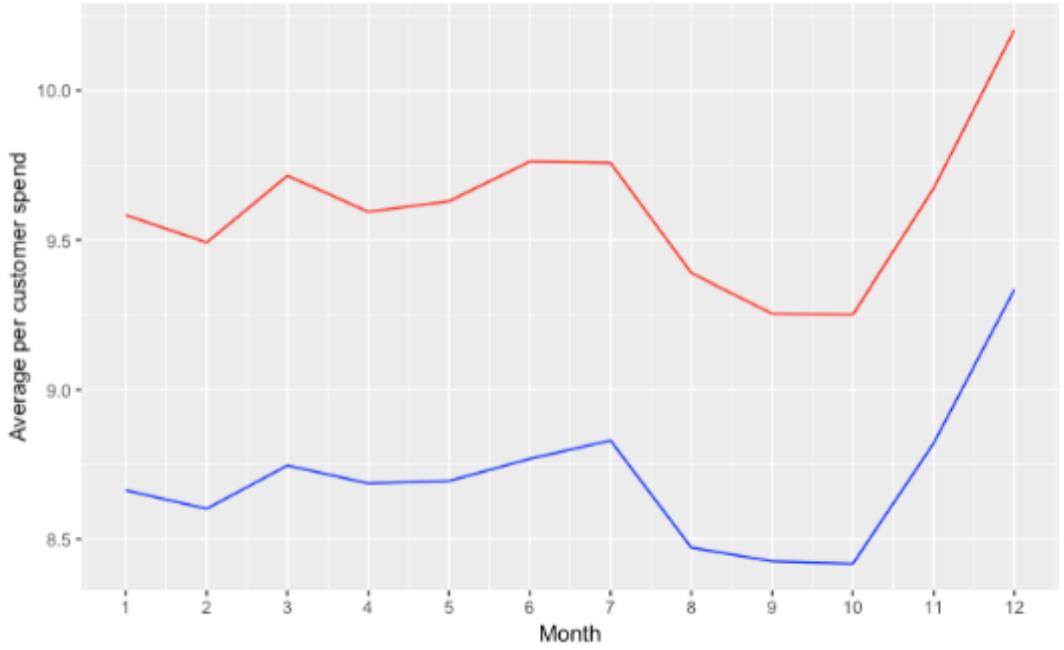


A huge number of stores are closely located

We can say that the company had a very few competition for a long period of time . Very few stores opened before 2000 and almost 90% of the competitors were opened after 2000 . Hence the Sales might have been better before the year 2000



Average per customer spend in store – continuing promotion



Red: participate continuing and consecutive promotion
Blue: do not participate continuing and consecutive promotion

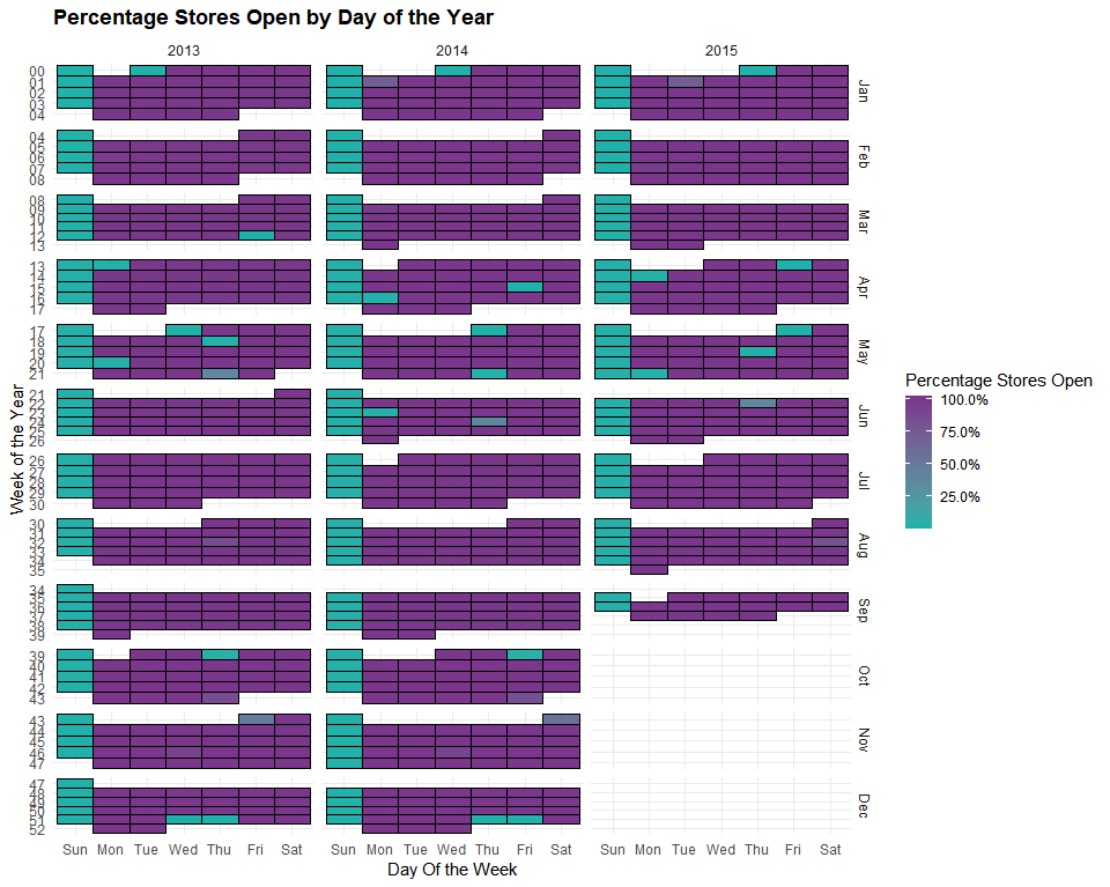
No promo	Feb, May, Aug, Nov	Jan, Apr, Jul, Oct	Mar, Jun, Sept, Dec
544	130	335	106

Seems like maximum number of the stores have no specific promo interval

Day-Level Patterns

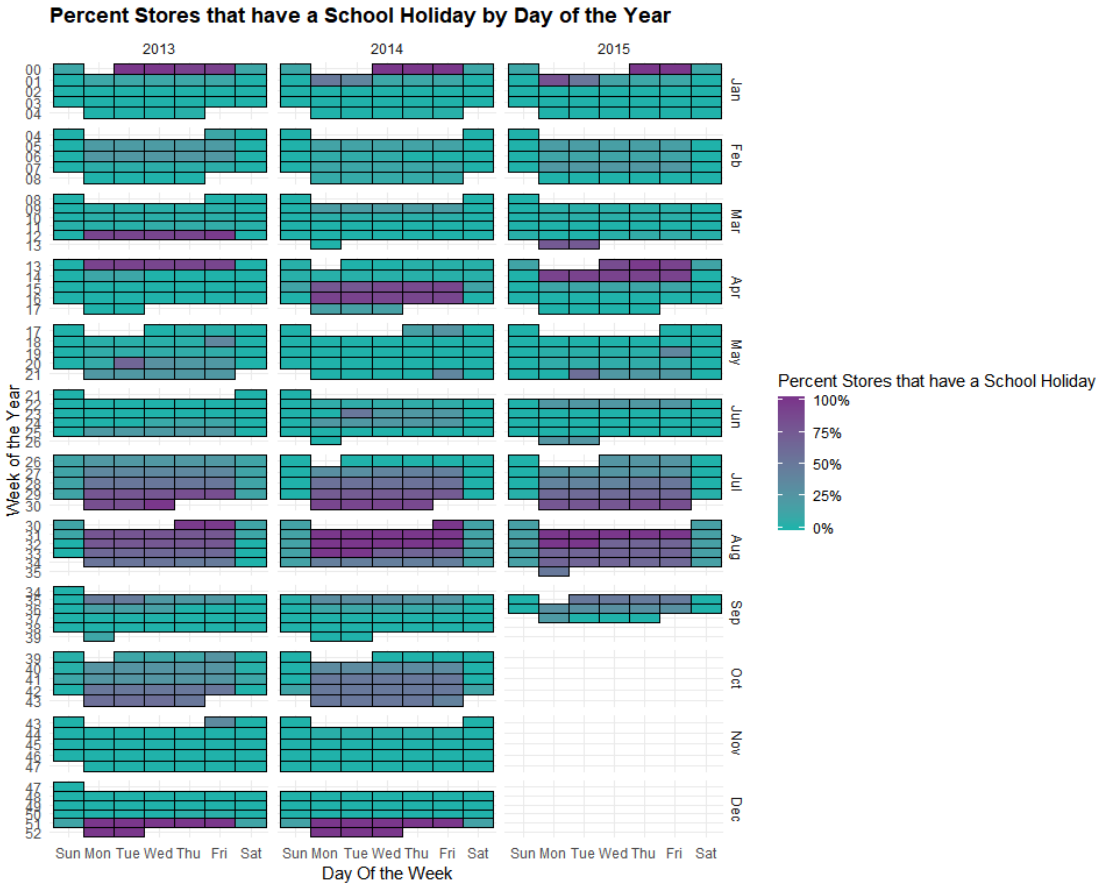
Let’s look deeper into the Data and find out some patterns and trends on a daily basis

Percentage of Stores that are open for each observation



- Most of the stores are closed on Sundays and State Holidays
- Almost 100% stores are open most days of the year
- There must be a reason why some stores are open on state holidays and Sundays when rest of them are closed!

Percentage of Stores that have a School Holiday

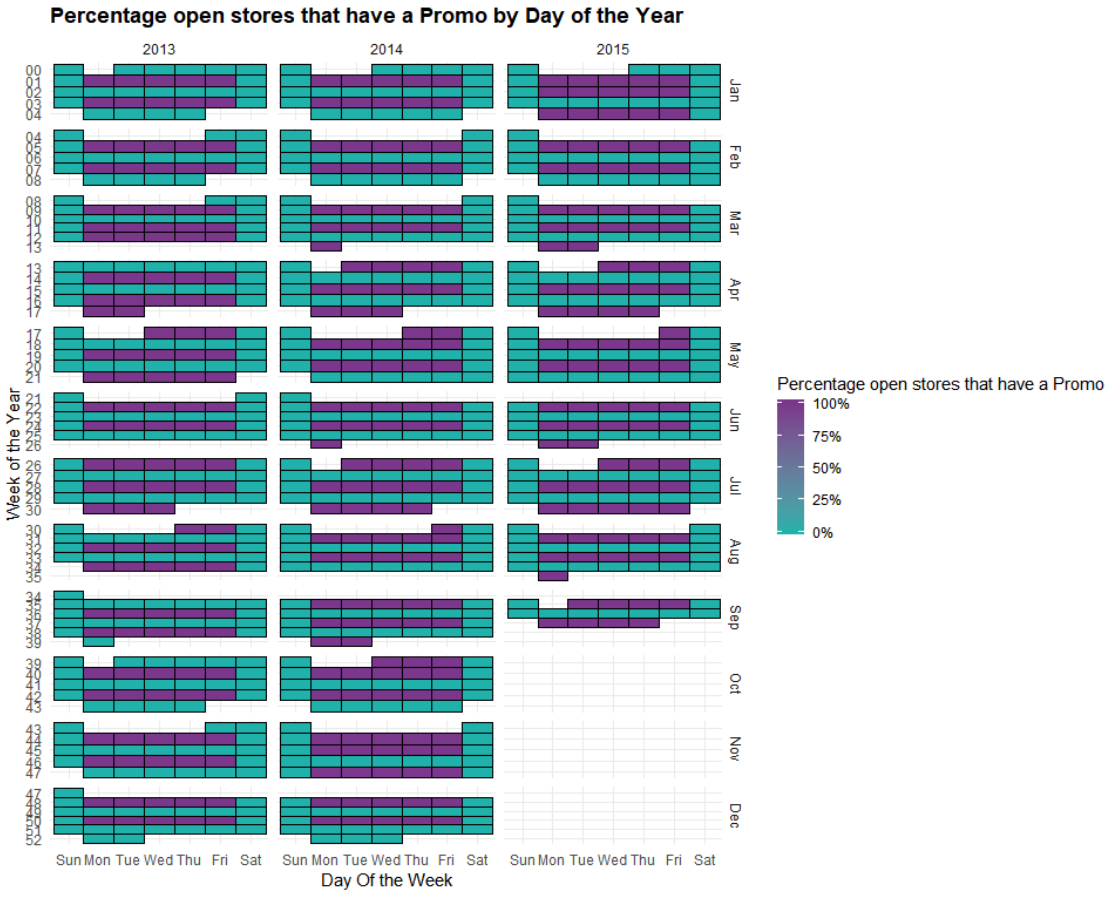


Most of the stores recognize school holidays between Late July and Early August, New Year's week and the week prior

Some of the stores recognize school holidays within October.

Timing of the school holidays shifts from last week of March and first week of April in 2013 to two weeks in April in 2014 and then back to last week of March and first week of April in 2015.

Percentage of Stores that have a promotion



Interestingly, on any given day, there are only
2 situations possible:

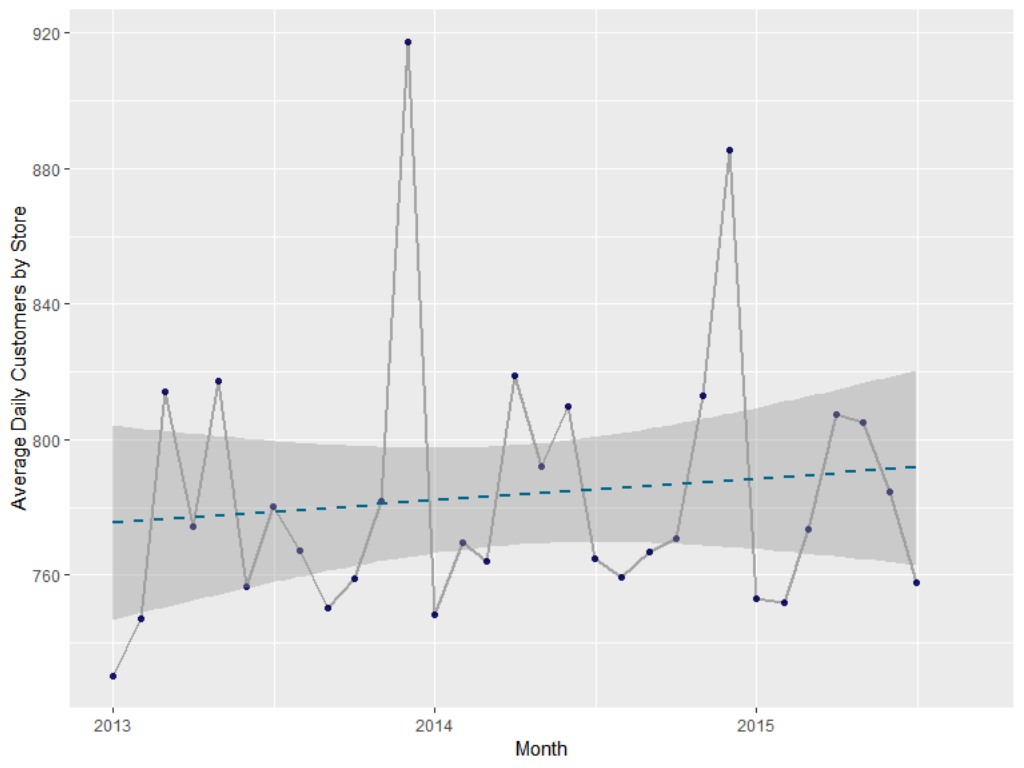
- the store either has a promotion
- the store does not have a promotion

There is a promotion every alternate week.
We can also see some back-to-back Promotions

Because of the back-to-back promotions, we suspect that the sales during the second
week might lack in sales!
A suspicion we will confirm later!

Monthly Customer Averages

We will try to see the average customers that visit the store on a monthly basis for 3 years



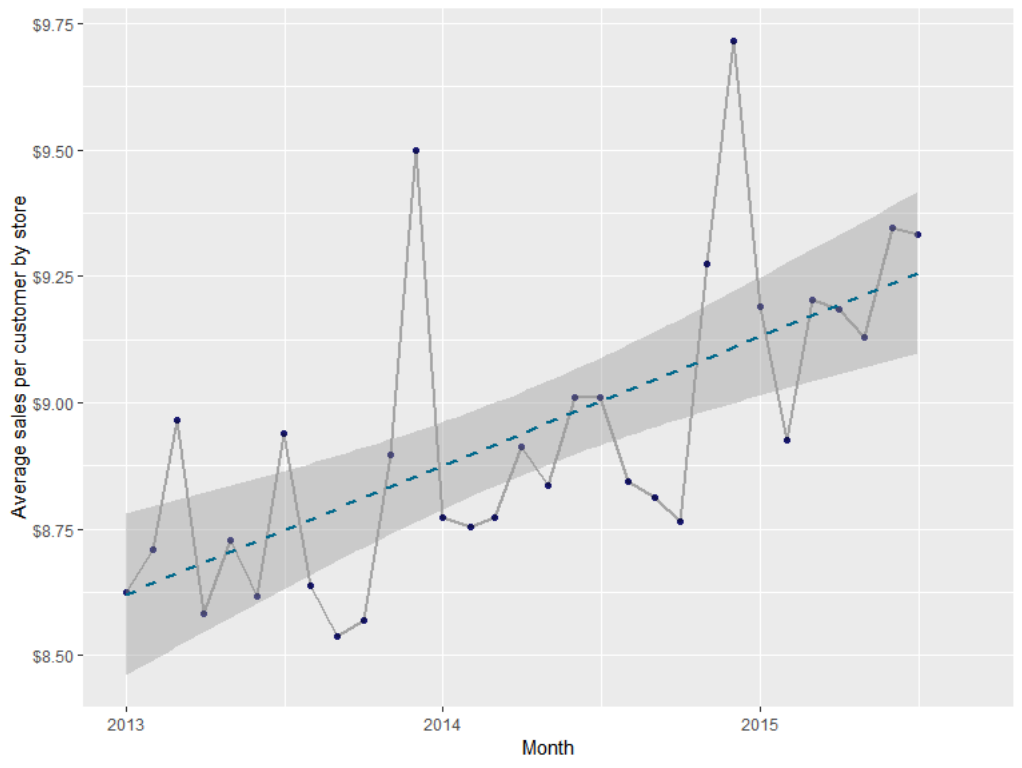
We clearly see the Seasonal Patterns. There is a surge in the customers during November – December.

March and April perform well also.

It is possible that the trend line gets affected by the significant increase in the customers during December-January period

Monthly Sales per Customer

We will try to see the average sales by customers that visit the store on a monthly basis for 3 years



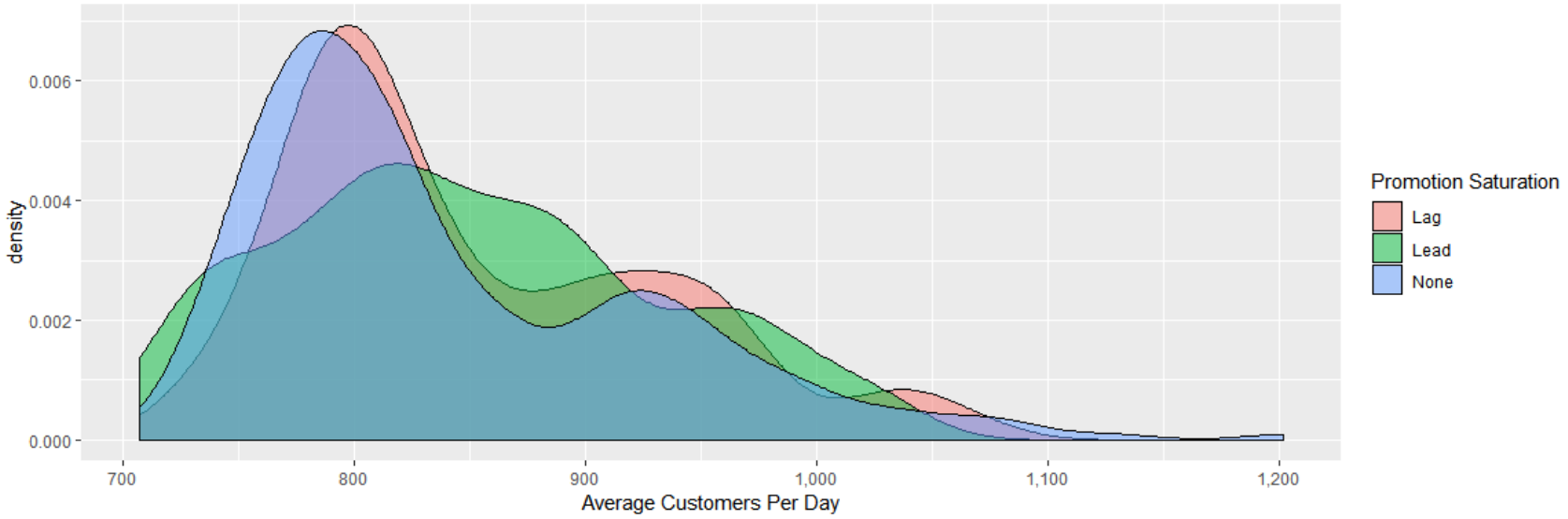
December is also the month with highest average sales per customer.

The sales is March, April and May are not in line with the previous plot

The trend is steeper and positive throughout the years

Promotion Saturation for Average Customers

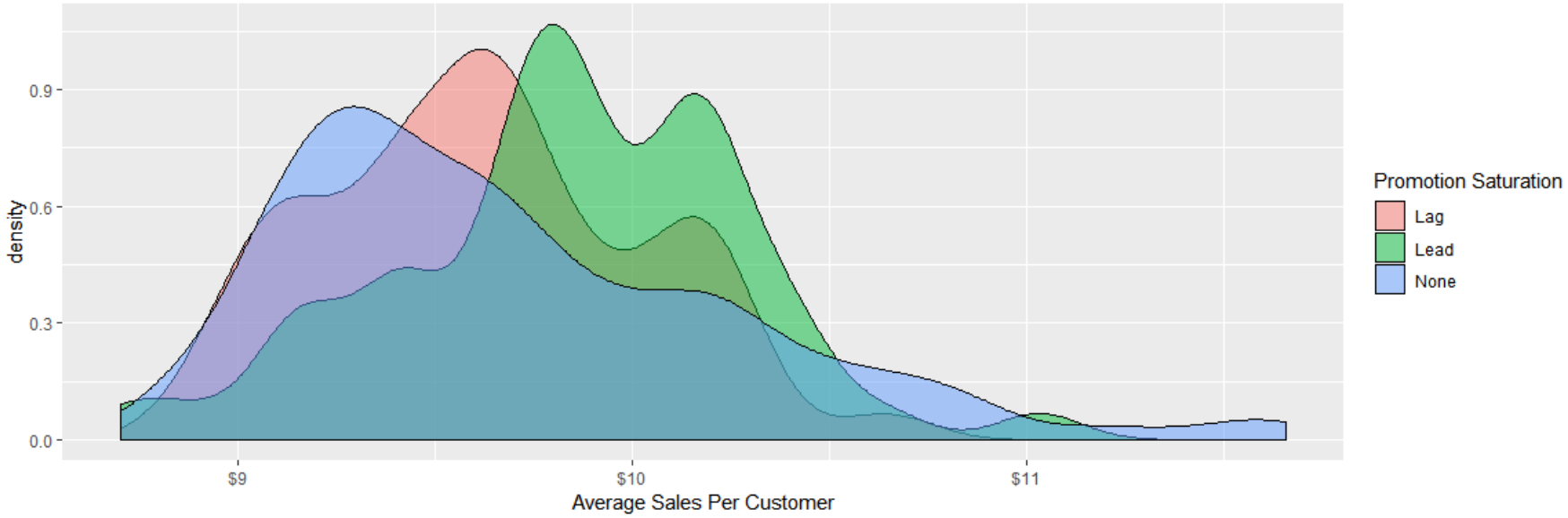
We suspected that during back-to-back promotions, we might have a decline in sales due to prolonged period. We can study it:



"**lead**" week would be the first week, the "**lag**" week would be the second week, and "**none**" corresponds to a normal one-week promotion schedule.

The plot above shows that the "lag" week follows a sales distribution extremely similar to a "none" one-week promotion, while the "lead" week may average more customers.

Promotion Saturation for Average Sales per Customer



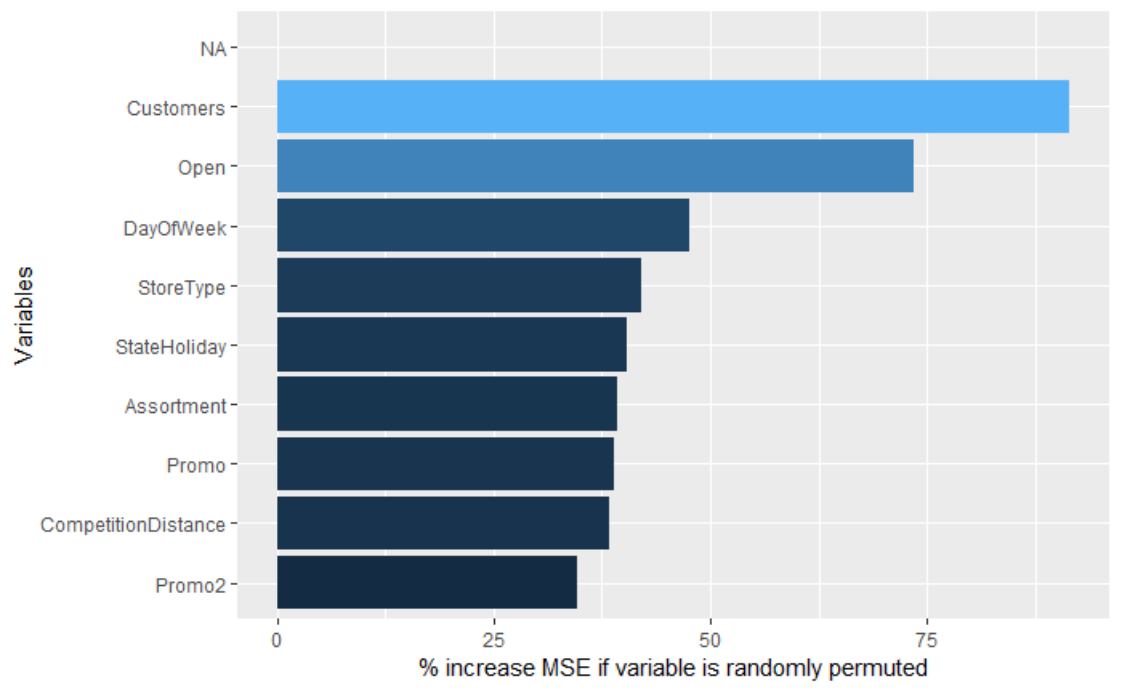
average sales are more for Lead and Lag than for None.

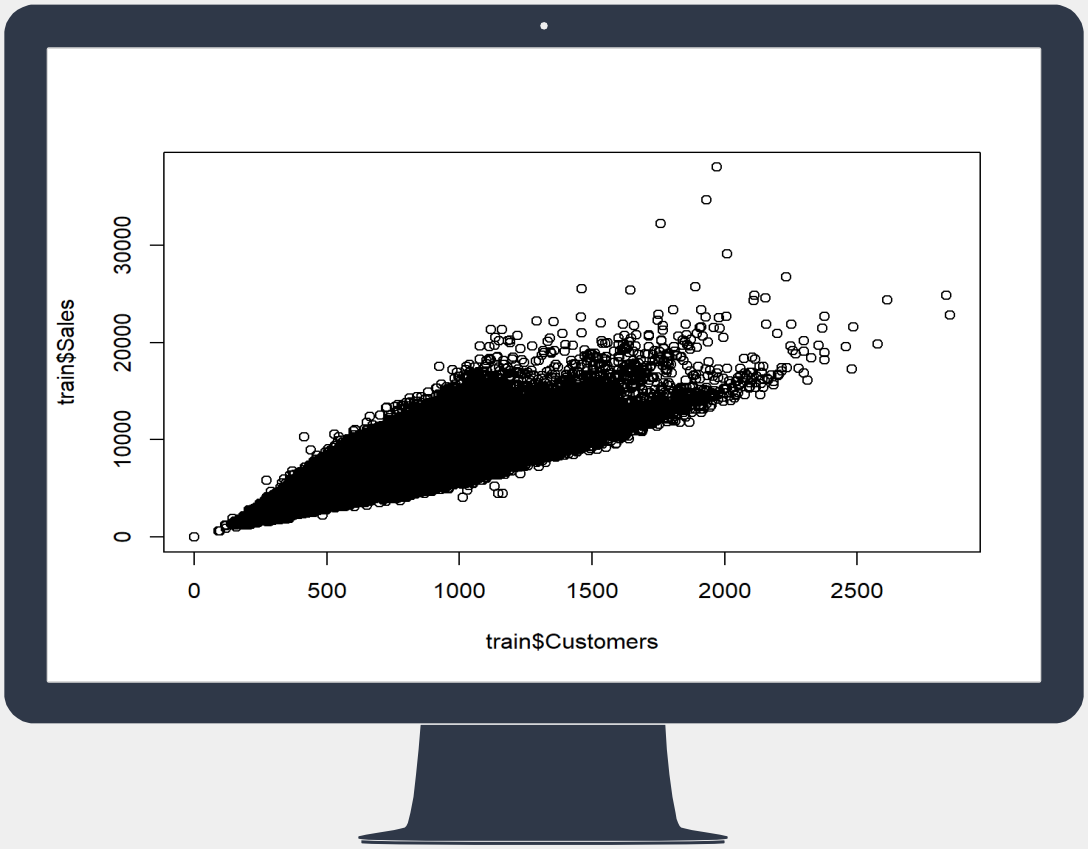
We can witness a business strategy here. We see that back-to-back promotions averages more customers and higher sales than non-promotion weeks. It is only logical to hold these promotions frequently after taking into consideration the related costs.

Predictions

We can now predict some sales and number of customers based on the studied trends

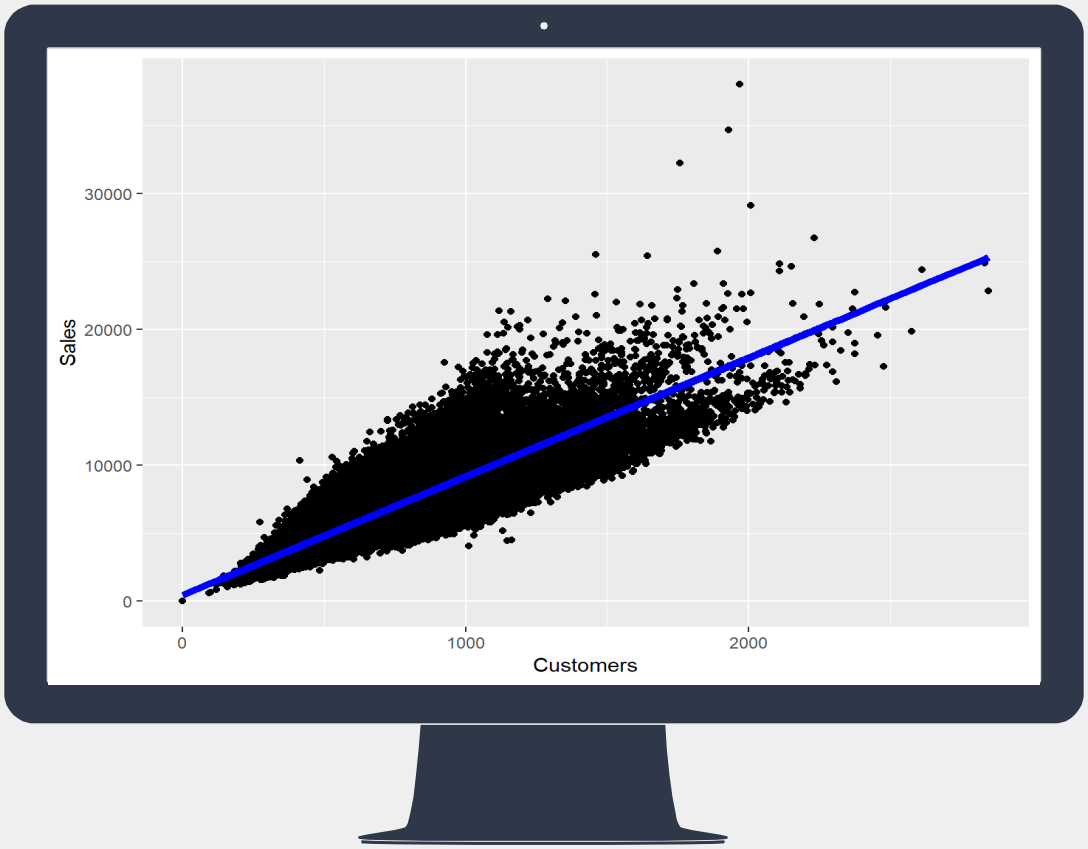
Though correlations give a good view of important numeric variables, we perform Random forest to get an overview of the most important variables including the categorical variables before visualisation. This shows that Customers,Open,DayofWeek have greater importance compared to the rest





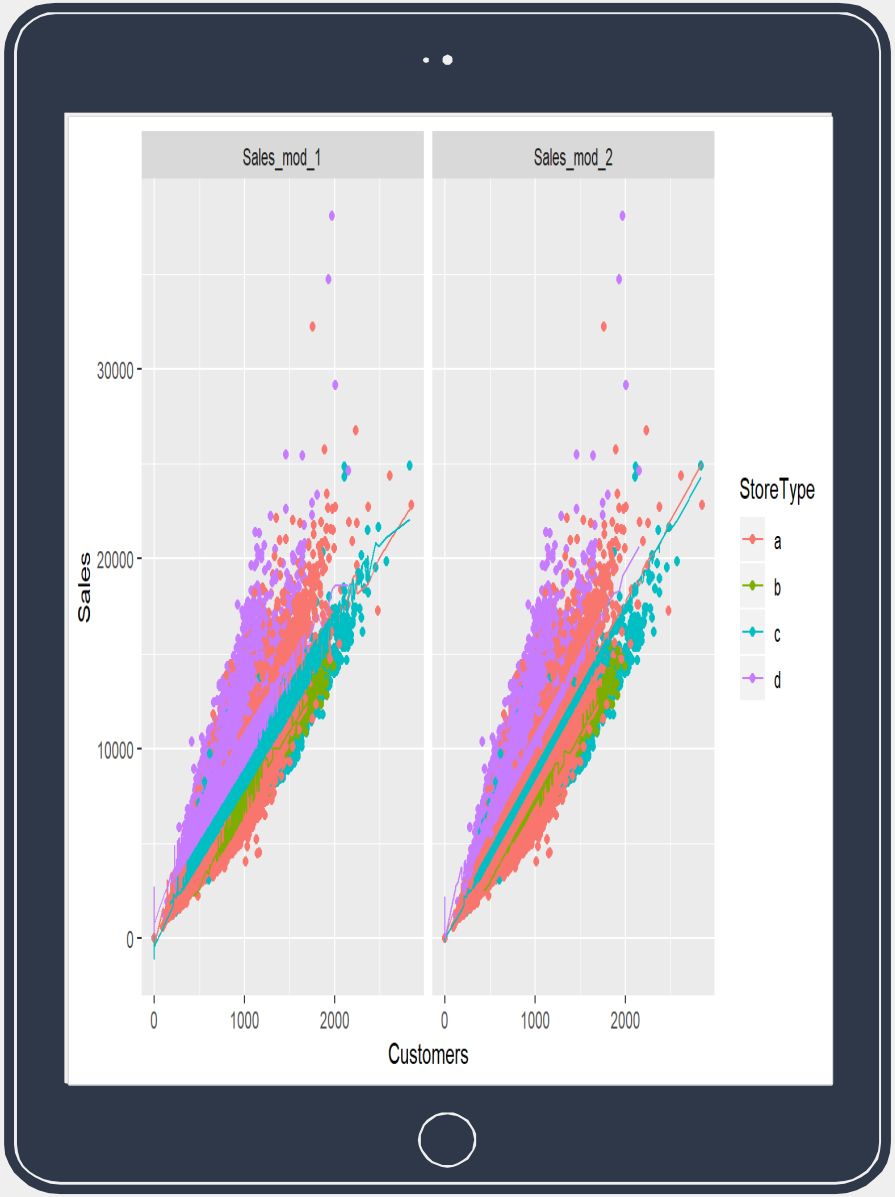
Linear Regression Between Sales and Customers

There is a positive relationship between Sales and number of customers.



The linear model is built for Sales(dependent) and customers(independent) variable. The R squared value is approximately 85 % which says that this model fits 85% of the data. The p value is 2.2e-16 which is extremely low and hence null hypothesis is rejected and it supports the alternate hypothesis.

Predicting future sales of different store type



Key Takeaway Points

Promotions

Have a very positive and consistent impact on the number of customers and the sales

Every

Store had a promotion every other week and 9 times, the promotions were back to back .

We

Did not see any saturation effect as suspected.

The

Few stores that opened on Sundays have a unique relation between Customers and Average sales per customer.

State

Holidays have a significant impact on the sales

The Average

Sales per customer and the number of customers is an increasing trend throughout the years.

Thanks!