WALMART 2010-12 WEEKLY SALES

Path = '*data/train.csv***'**

Train.csv

Fields:

-- Store: 45
-- Dept: 81 (1-99)
-- Weekly_Sales: includes negative values [-4988.94, 693099.36]
-- Date:
-- IsHoliday

Shape:

(421570, 5)

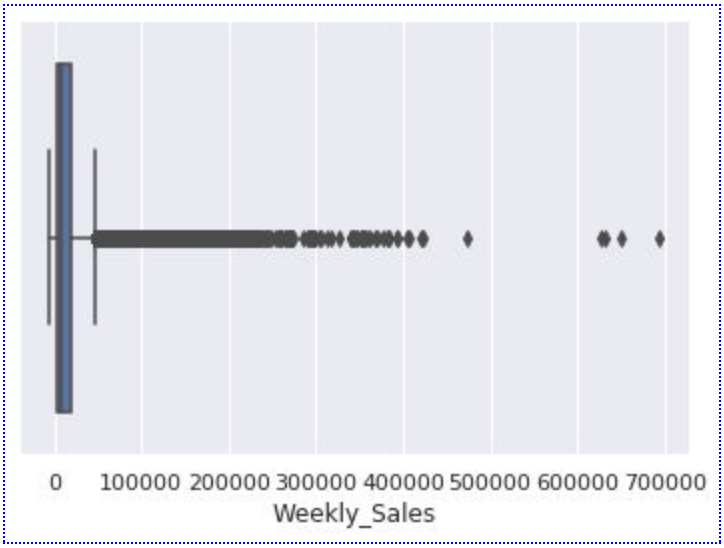
Description

	Store	Dept	Weekly_Sales
count	421570.000000	421570.000000	421570.000000
mean	22.200546	44.260317	15981.258123
std	12.785297	30.492054	22711.183519
min	1.000000	1.000000	-4988.940000
25%	11.000000	18.000000	2079.650000
50%	22.000000	37.000000	7612.030000
75%	33.000000	74.000000	20205.852500
max	45.000000	99.000000	693099.360000

Fig: 1.0

Missing depts in the dataset :

[15, 53, 57, 61, 62, 63, 64, 66, 68, 69, 70, 73, 75, 76, 84, 86, 88, 89]



“There seems to be a lot of outliers in Sales data”

Fig: 1.1

Impact of Holidays

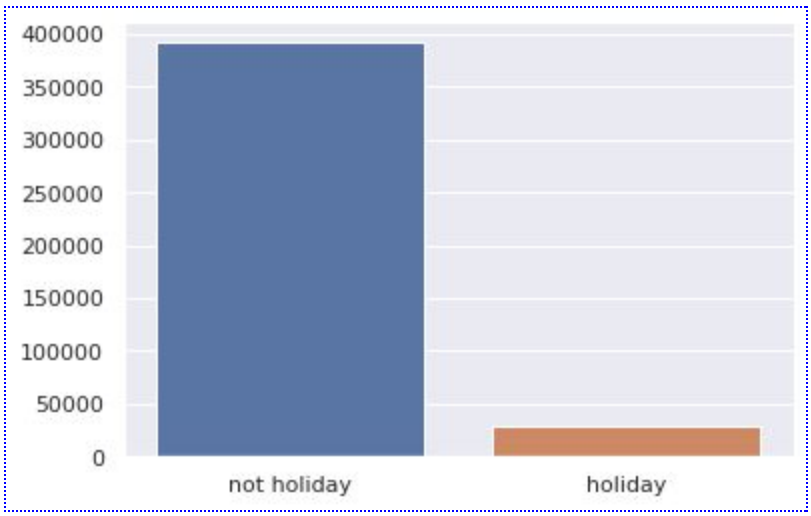


Fig: 1.2.0

Overall

Relative holidays for each store

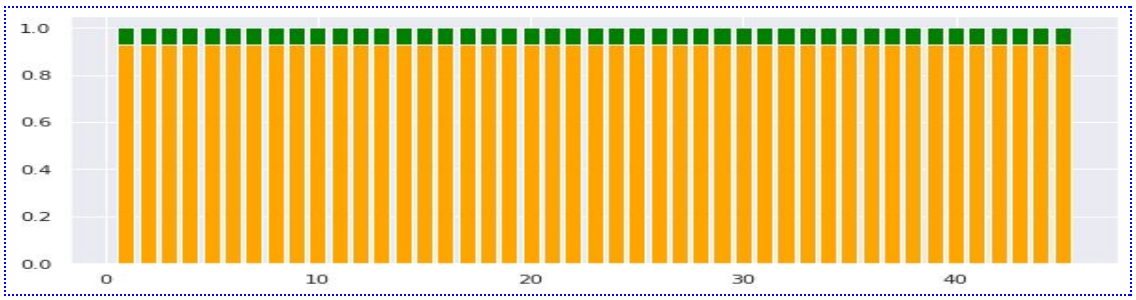
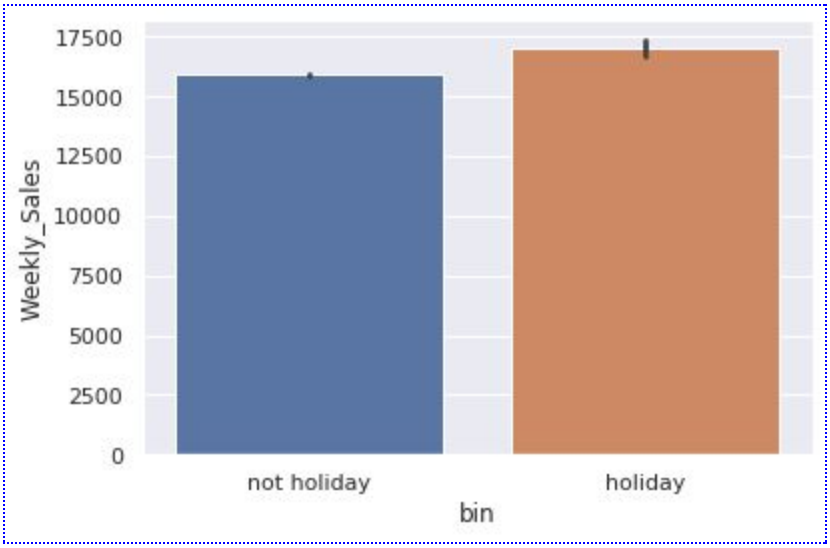


Fig:1.2.1

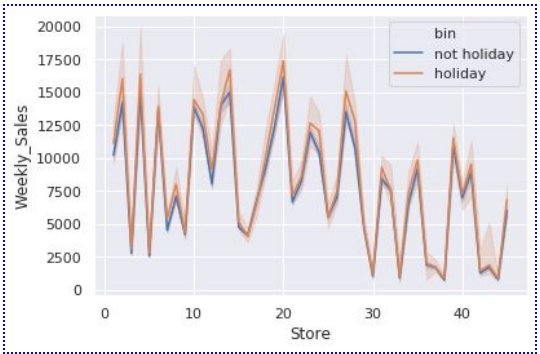
Relative holidays vs non holidays is almost same across each store [0.93, 0.07]

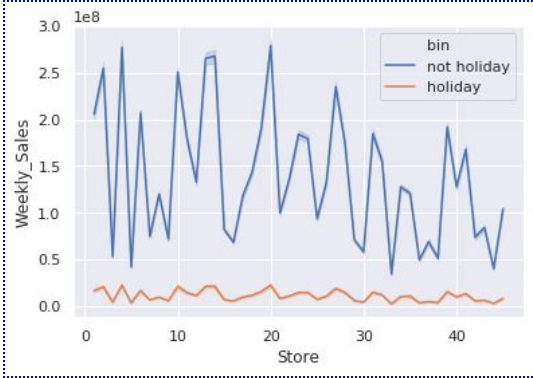


“Weekly_sales”

Fig: 1.2.2

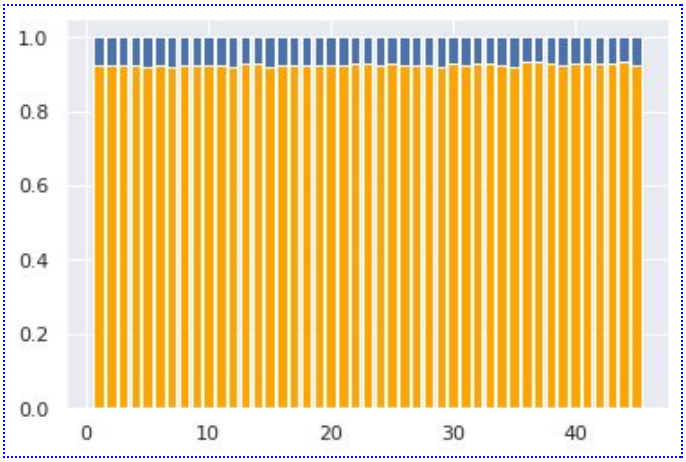
Weekly_sales vs stores





Estimator: ‘median’ **‘summation’**

Fig: 1.2.2a **1.2.2b**



Relative weekly sales across stores

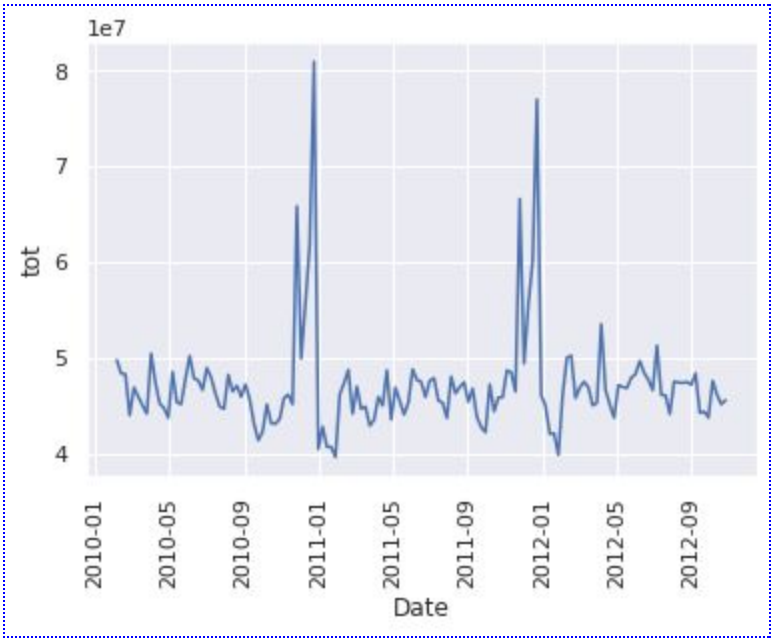
“Relative weekly sales across stores for non holidays, holidays: [0.925, 0.075]”

Fig: 1.2.3

Impact across weeks

Num_weeks : *143*

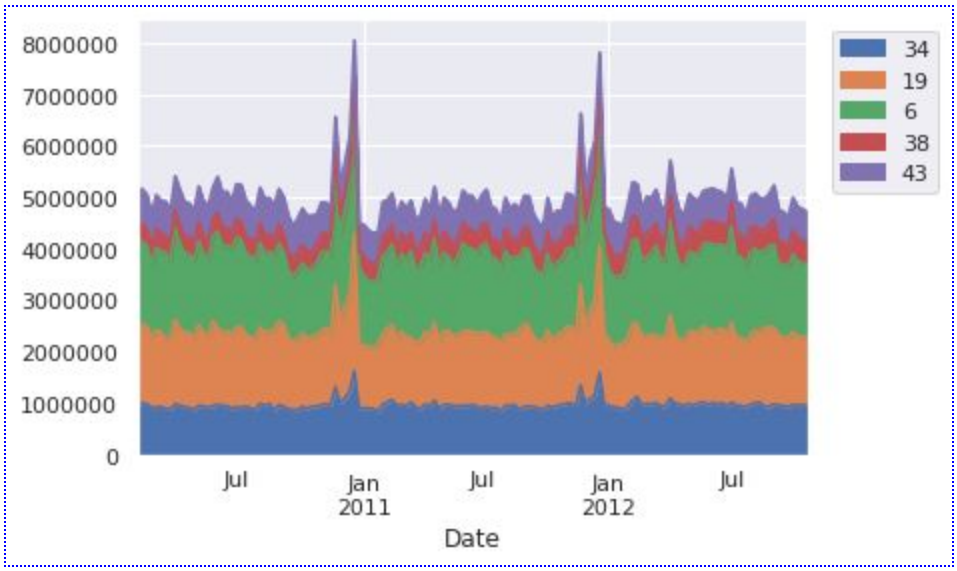
Weekly_Sales



"Peaks around Dec 2010 and Dec2011"

Fig: 1.3.0

Stores



Stores = [34, 19, 6, 38, 43]

"Dec2010 and Dec2011 both have peaks around them"

Fig: 1.3.1

Maximum sales week

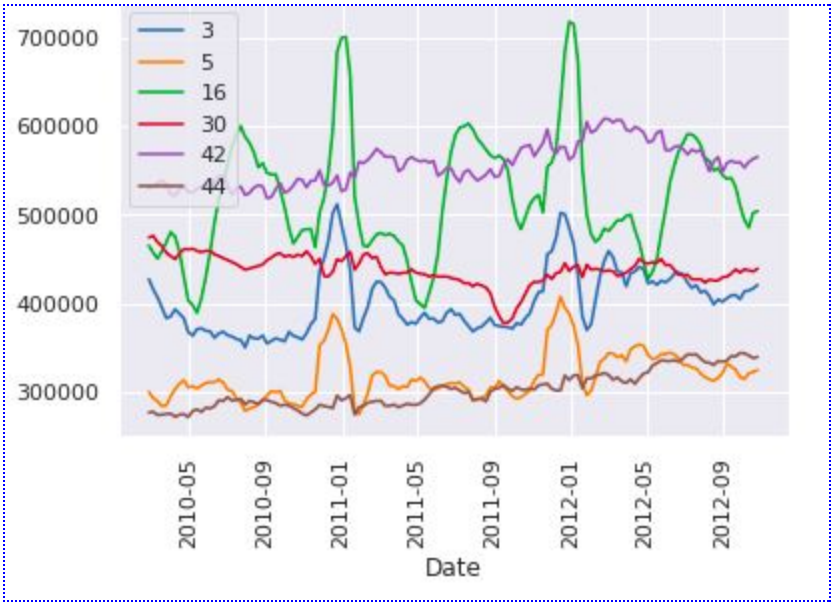
Stores have their peak in *Dec 24-12-2010 (Christmas eve)*

[4,7,9,32,37,39,41,44] : *Dec 23-12-2011 (Christmas eve)*

[5, 33, 35, 36, 38, 42, 43] : have local minima on Christmas eve
*stores: [36, 42, 43]

Fig: 1.3.2a

Minimum sales



"Most of the stores have their minimum sales in Jan 2010, some others in Jan 2011"

Fig: 1.3.2b

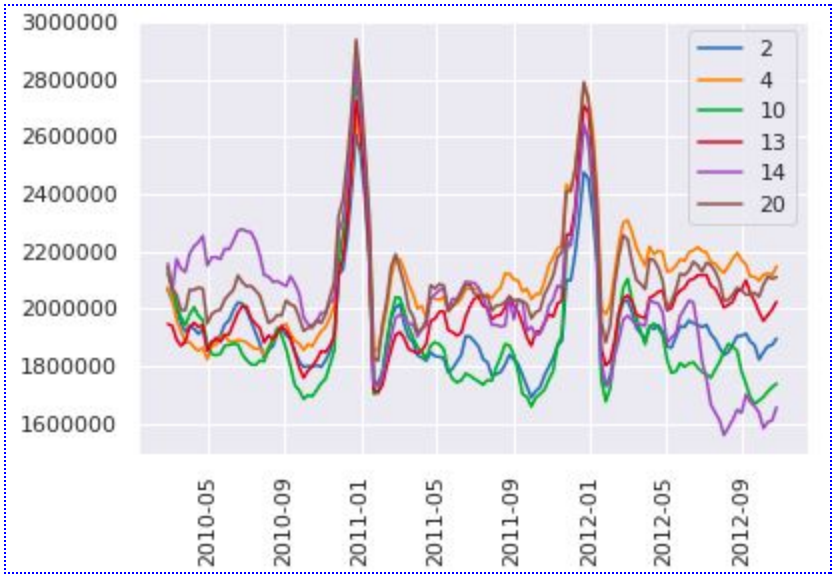
Store wise trends across weeks

There is a clear gradient among stores in terms of sales;
stores have a relative weekly sales pattern

Store	1	2	3	4	5	6	7	8	9
Date									
2010-02-05 00:00:00	1643690.900000	2136989.460000	461622.220000	2135143.870000	317173.100000	1652635.100000	496725.440000	1004137.090000	549505.550000
2010-02-12 00:00:00	1641957.440000	2137809.500000	420728.960000	2188307.390000	311825.700000	1606283.860000	524104.920000	994801.400000	552677.480000
2010-02-19 00:00:00	1611968.170000	2124451.540000	421642.190000	2049860.260000	303447.570000	1567138.070000	506760.540000	963960.370000	511327.900000
2010-02-26 00:00:00	1409727.590000	1865097.270000	407204.860000	1925728.840000	270281.630000	1432953.210000	496083.240000	847592.110000	473773.270000
2010-03-05	1554806.680000	1991013.130000	415202.040000	1971057.440000	288855.710000	1601348.820000	491419.550000	881503.950000	507297.880000

Fig: 1.3.3

Stores: 2, 4, 10, 13, 14, 20 have relatively high sales throughout

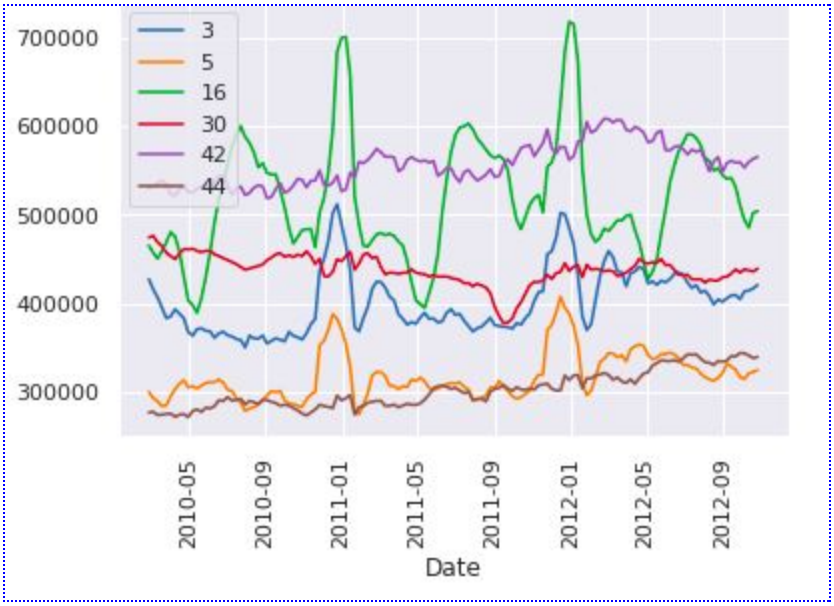


"Stores have their peak in Dec 2010 and another in Dec 2011"

Fig: 1.3.4a

"Stores: 10 and 14 have a negative trend"

FIg: 1.3.4b



3, 5, 7, 9, 16, 29, 30, 33, 38, 44

Fig: 1.3.5

Weekly Sales across each dept for each store

80	81	82	83	85	87	90	91	92	93	94	
00	747310.800000	577369.200000	166944.950000	87214.760000	609066.670000	2199107.520000	1755822.240000	3781162.870000	1165221.800000	1301383.110000	3170529.9
00	680040.170000	591428.530000	171428.950000	88086.990000	541000.950000	2031157.150000	1603013.190000	3443837.300000	1077422.700000	1317069.860000	2976149.3
00	657129.600000	599874.140000	203166.200000	88368.190000	518886.950000	1976215.640000	1546117.670000	3312991.710000	1005051.420000	1513229.840000	2814037.7
00	664440.150000	543934.640000	175047.900000	88094.860000	490205.080000	1935114.730000	1486244.190000	3204210.060000	950067.100000	1212218.700000	2789412.1
00	697920.180000	571815.620000	191223.450000	89731.270000	549397.900000	2005670.830000	1607815.570000	3480725.210000	1052044.390000	1287397.560000	2994339.2

Depts: [38, 40, 72, 92, 95] seem to have marginally high sales throughout for a randomly selected sample stores [22, 17] and for overall weekly sales across depts

Fig: 1.4.0

Depts: [19, 37, 39, 43, 45, 47, 48, 50, 51, 65, 77, 78, 99] have less than 4500 entries out of 6435 for [Date,Store] vs Dept

Depts with maximum sales

Depts: [38, 40, 72, 92, 95]

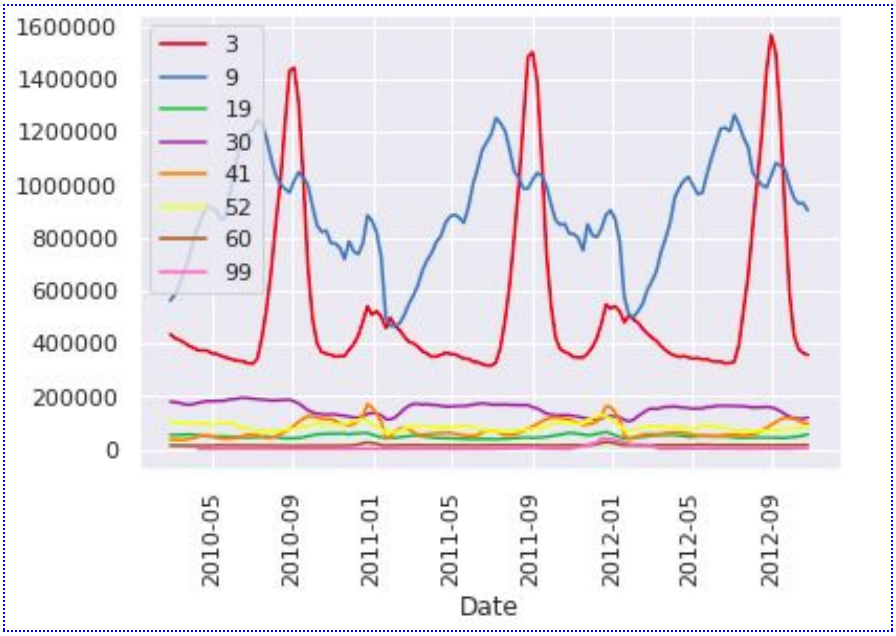
"Dept 72 has two large peaks around Dec2010 and Dec2011"

"While there are two smaller peaks around Apr2011 and Apr2012; it has a symmetry in sales"

"Dept 95 have reverse trends with lower sales around Christmas eve"

Fig: 1.4.1

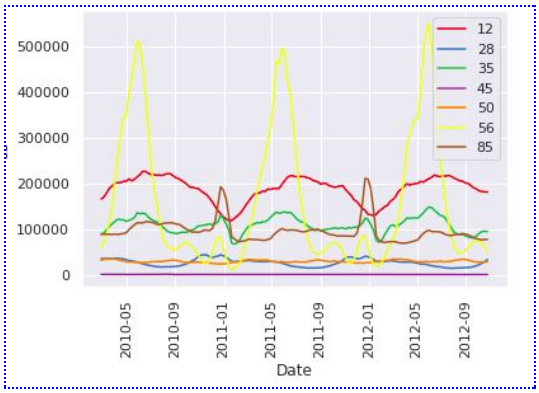
Depts with minimum sales

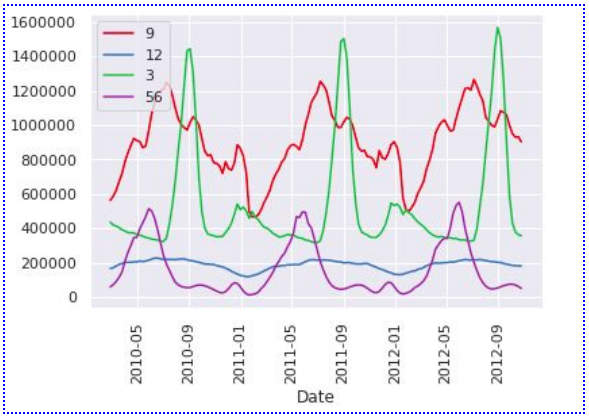


"Dept 3 has 3 large peaks and 2 smaller peaks around Aug2010, Aug2011, Aug2012 and Dec2010, Dec2011 respectively"

Fig: 1.4.2

Depts:[3,6,9,12,18,19,27,28,29,30,31,35,36,37
,41,42,45,47,48,50,51,52,53,54,56,58,59,60,61,85,99]`



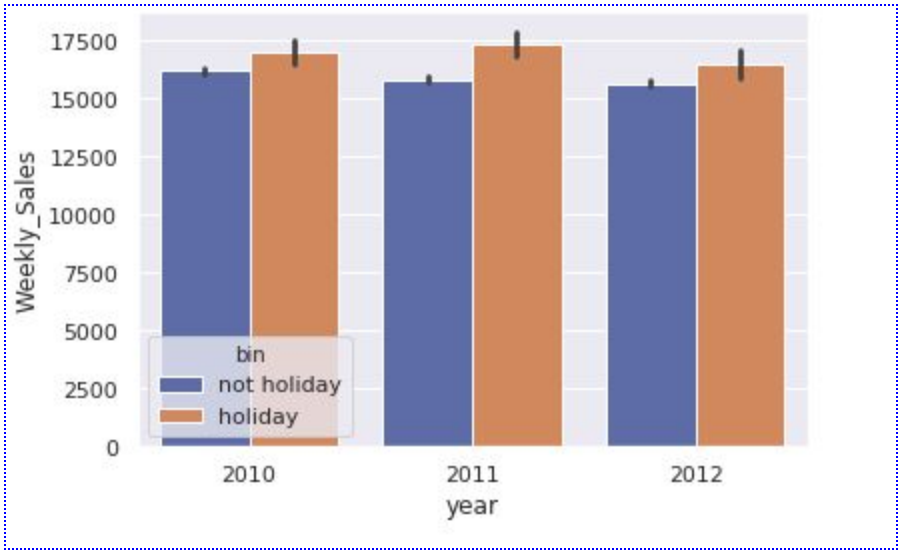


Dept 56 has trends similar to Dept 3
Dept 12 has trends similar to Dept 9

Yearly

Years: *2010, 2011, 2012*

Total sales



"Average sales across years for **Holiday** is greater than **non holidays"**

"It is same across all the years for non holidays"

Fig: 1.5.0

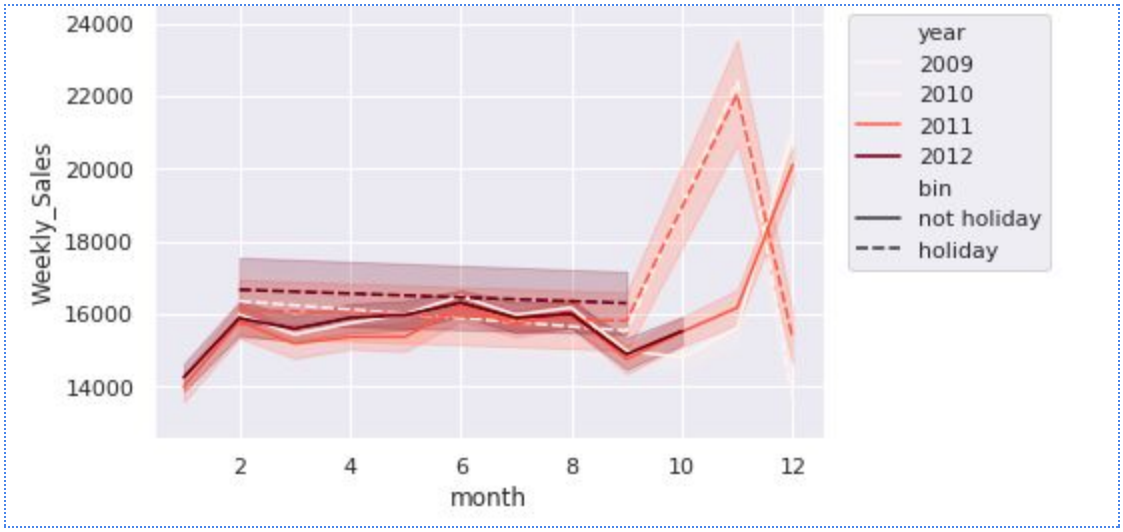


Fig: 1.5.1

Average Sale across months for all 3 years are equivalent

Sales holidays for 2011 increases marginally around months 9-12

Average sales across months for randomly chosen stores

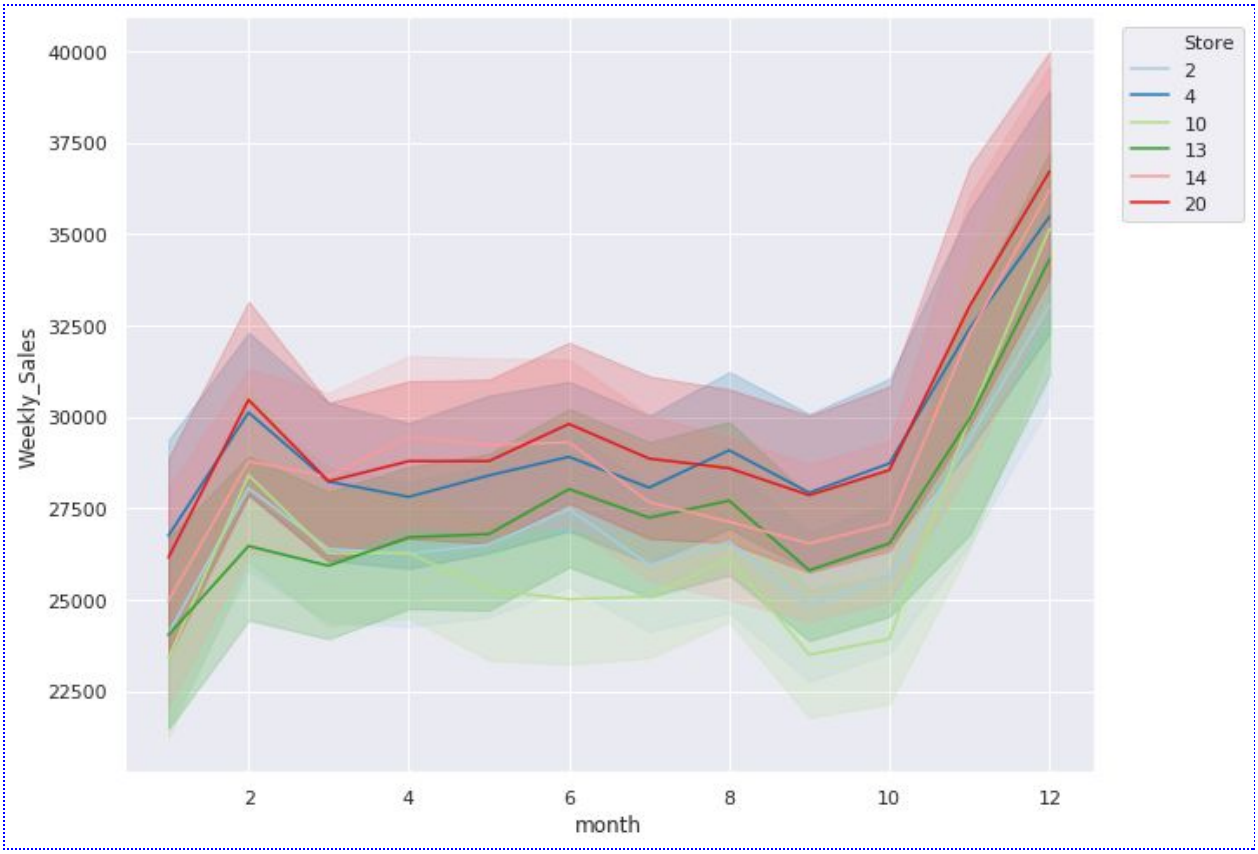


Fig: 1.5.2a

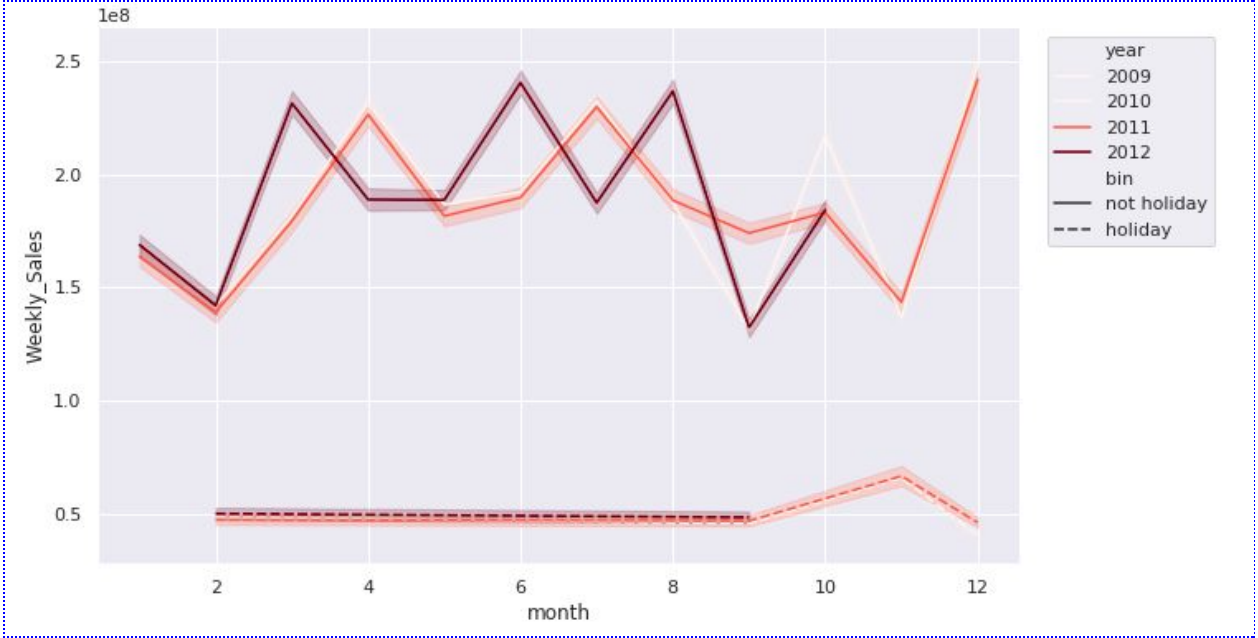
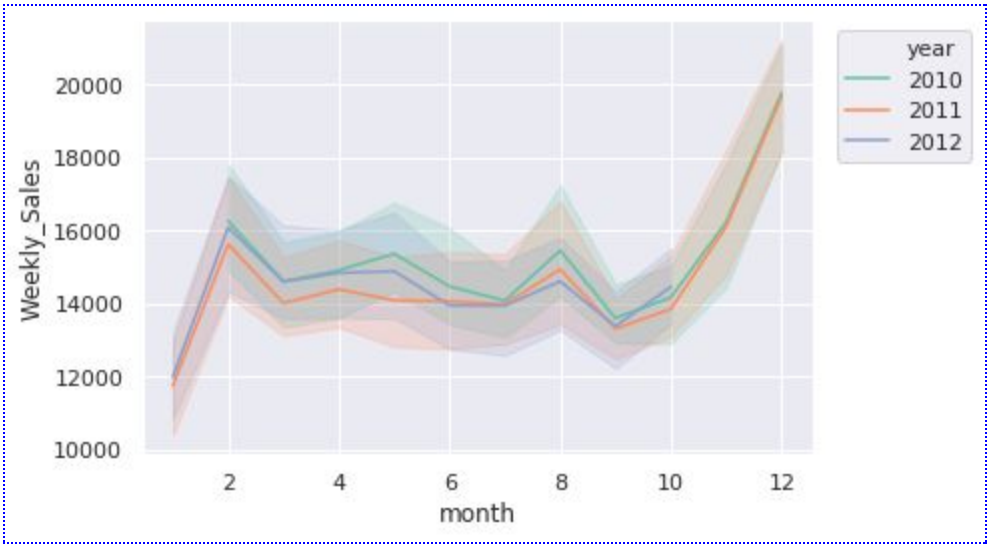


Fig: 1.5.2b

“Holiday sales are primarily only during months 9 to 12 and is same across years”

Monthly

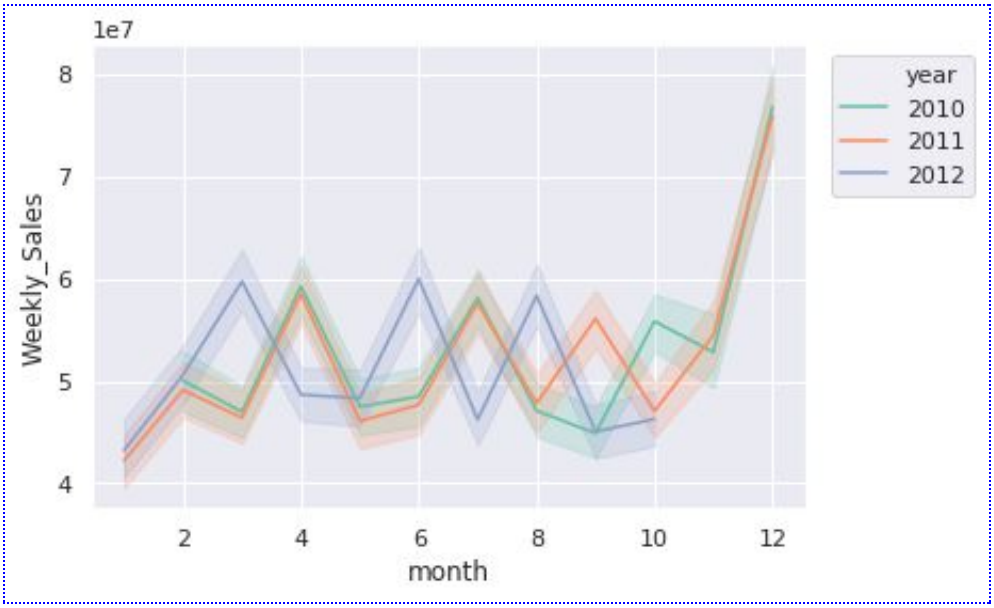
monthly sales over 3 year



Estimator: 'median'

"Sales increase marginally in Dec"

Fig: 1.6.1a



Estimator: 'Summation'

"Total sales across months for years 2010 and 2011 are isometric"

Fig: 1.6.1b

Monthly sales across stores

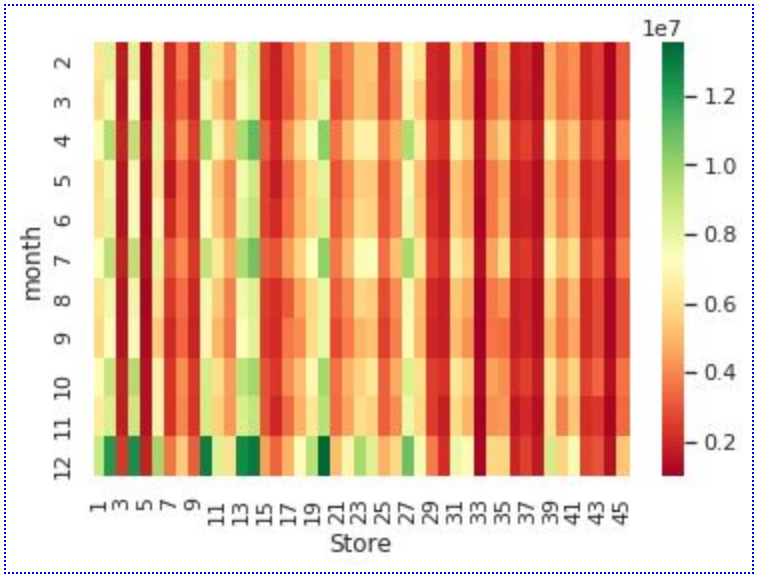
Year: 2010

Months with high sales throughout

Months: [4, 7, 10, 11, 12]

Store	1	2	3	4	5	6	7	8	
month									
2	6307344.100000	8264347.770000	1711198.230000	8299040.360000	1202728.000000	6259010.240000	2023674.140000	3810490.970000	2087284.20
3	5871293.980000	7677765.600000	1534680.590000	7525350.190000	1141138.680000	6325062.470000	2061053.480000	3354290.230000	1972134.85
4	7422801.920000	9520276.090000	1895382.980000	9269170.710000	1553822.750000	7995878.230000	2230043.440000	4342564.850000	2540761.94
5	5929938.640000	7777385.060000	1477534.200000	7637509.720000	1241824.970000	6320388.540000	1606057.640000	3476533.570000	2083339.88
6	6084081.460000	8069642.740000	1473409.220000	7550298.110000	1243520.700000	6860962.560000	2042623.050000	3557902.940000	2131114.29
7	7244483.040000	9316686.920000	1783914.220000	9237037.480000	1422175.500000	8127765.740000	2899977.830000	4119065.060000	2419603.89
8	6075952.950000	7715686.750000	1456095.440000	7760274.340000	1165245.500000	6171593.190000	2378701.980000	3520812.200000	2010828.47
9	5829793.920000	7262197.740000	1437426.450000	7512322.680000	1206021.310000	5519280.350000	2045951.730000	3340894.600000	1912238.07
10	7150641.750000	9012420.110000	1796963.560000	9522604.020000	1417374.250000	6751407.940000	2330553.050000	4168178.210000	2430655.45
11	6485547.060000	8471338.610000	1747923.750000	8900188.900000	1412885.080000	6787272.460000	2305814.350000	3986106.520000	2326610.48
12	8876953.180000	12190116.800000	2430890.360000	12466674.300000	1829294.030000	9793698.970000	3643627.460000	5527635.690000	3214648.24

Fig: 1.6.2

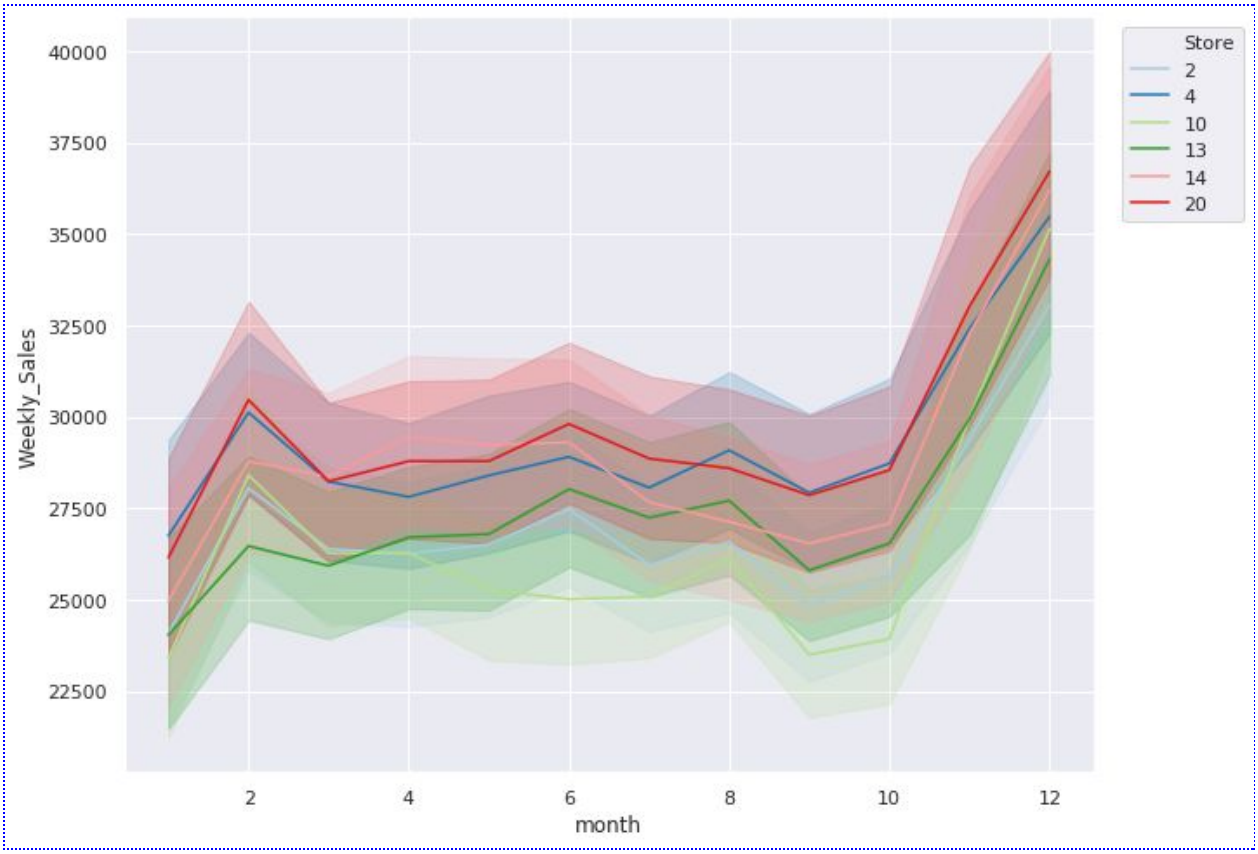


Stores with high sales throughout

“Sales are isometric across stores”

Fig: 1.6.3

Sales across stores over months



Estimator: 'mean'
Fig: 1.6.4a