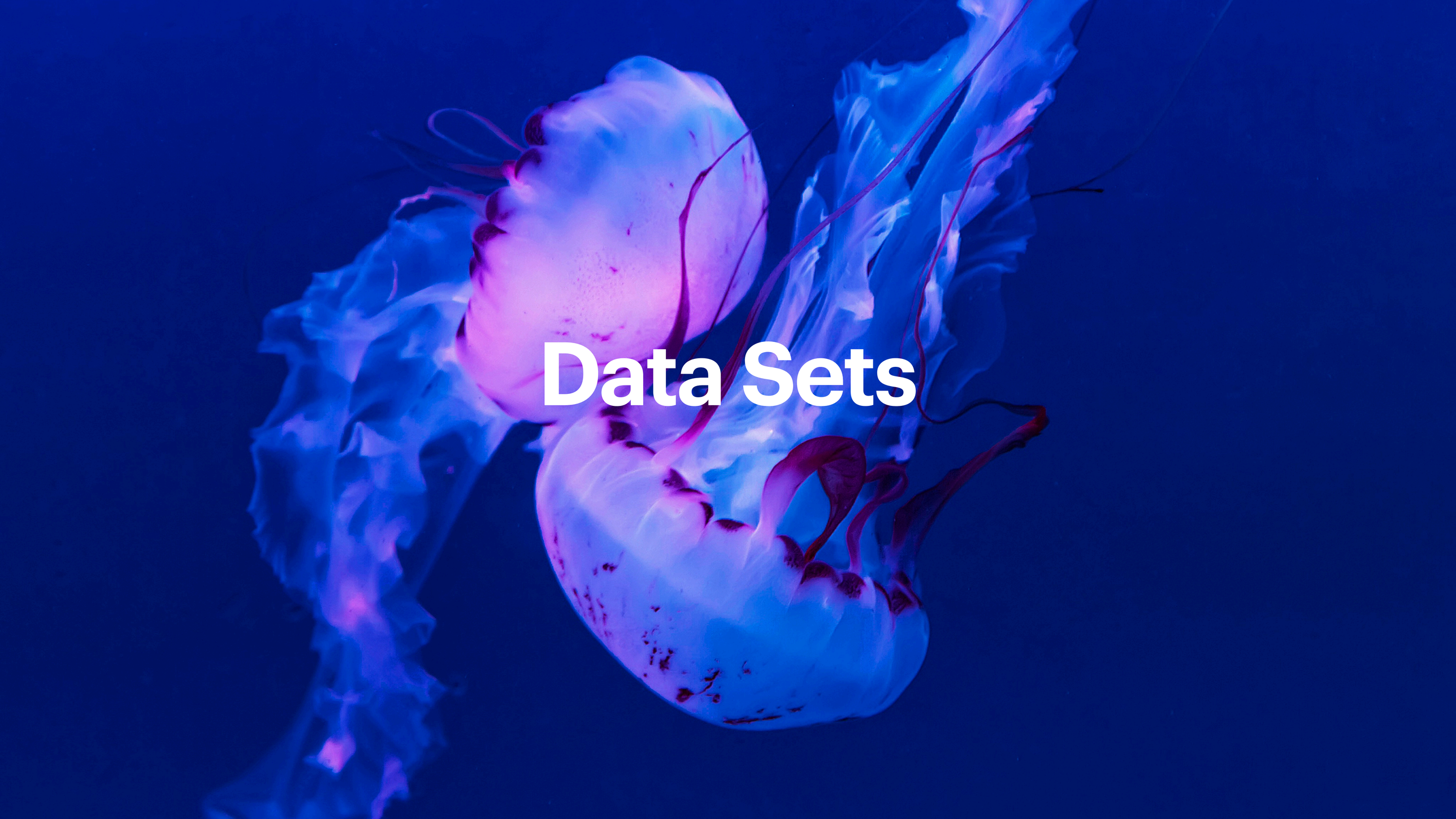
Walmart Store Sales

Revenue Forecasting

26 JUN 2023


Data Sets

Data Sets

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	NA	NA	NA	NA	NA	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	NA	NA	NA	NA	NA	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	NA	NA	NA	NA	NA	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	NA	NA	NA	NA	NA	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	NA	NA	NA	NA	NA	211.350143	8.106	FALSE

Data Sets

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE

Data Sets

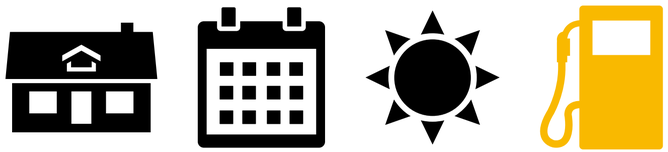
Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE

Data Sets

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE

Data Sets

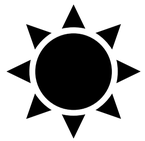
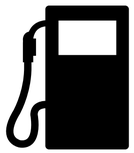
Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE



Data Sets

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE



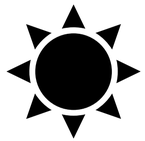
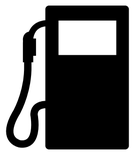



? ? ? ? ?

Data Sets

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE

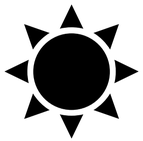
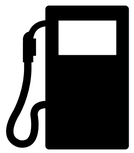
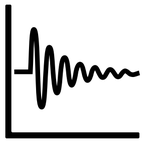




? ? ? ? ?


Data Sets

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE





? ? ? ? ?



Data Sets

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE

Data Sets

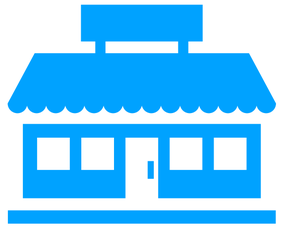
Store	Type	Size
1	A	151315
2	A	202307
3	B	37392
4	A	205863
5	B	34875
6	A	202505
7	B	70713
8	A	155078
9	B	125833
10	B	126512
11	A	207499
12	B	112238
13	A	219622
14	A	200898
15	B	123737
16	B	57197
17	B	93188
18	B	120653
19	A	203819
20	A	203742
21	B	140167
22	B	119557
23	B	114533
24	A	203819
25	B	128107
26	A	152513
27	A	204184
28	A	206302

Data Sets

Store	Type	Size
1	A	151315
2	A	202307
3	B	37392
4	A	205863
5	B	34875
6	A	202505
7	B	70713
8	A	155078
9	B	125833
10	B	126512
11	A	207499
12	B	112238
13	A	219622
14	A	200898
15	B	123737
16	B	57197
17	B	93188
18	B	120653
19	A	203819
20	A	203742
21	B	140167
22	B	119557
23	B	114533
24	A	203819
25	B	128107
26	A	152513
27	A	204184
28	A	206302

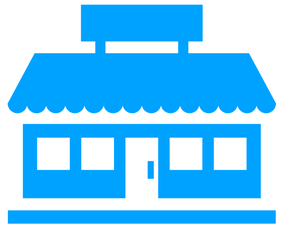
Data Sets

Store	Type	Size
1	A	151315
2	A	202307
3	B	37392
4	A	205863
5	B	34875
6	A	202505
7	B	70713
8	A	155078
9	B	125833
10	B	126512
11	A	207499
12	B	112238
13	A	219622
14	A	200898
15	B	123737
16	B	57197
17	B	93188
18	B	120653
19	A	203819
20	A	203742
21	B	140167
22	B	119557
23	B	114533
24	A	203819
25	B	128107
26	A	152513
27	A	204184
28	A	206302



Data Sets

Store	Type	Size
1	A	151315
2	A	202307
3	B	37392
4	A	205863
5	B	34875
6	A	202505
7	B	70713
8	A	155078
9	B	125833
10	B	126512
11	A	207499
12	B	112238
13	A	219622
14	A	200898
15	B	123737
16	B	57197
17	B	93188
18	B	120653
19	A	203819
20	A	203742
21	B	140167
22	B	119557
23	B	114533
24	A	203819
25	B	128107
26	A	152513
27	A	204184
28	A	206302



Data Sets

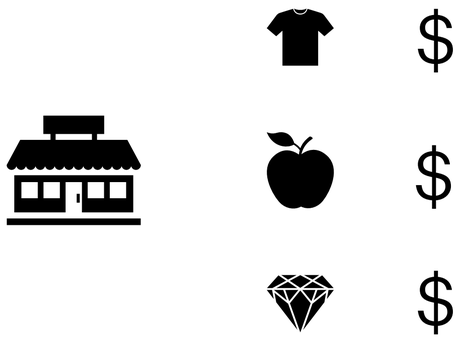
Store	Dept	Date	Weekly_S	IsHoliday
1	1	2010/2/5	24924.5	FALSE
1	1	2010/2/12	46039.49	TRUE
1	1	2010/2/19	41595.55	FALSE
1	1	2010/2/2	19403.54	FALSE
1	1	2010/3/5	21827.9	FALSE
1	1	2010/3/12	21043.39	FALSE
1	1	2010/3/1	22136.64	FALSE
1	1	2010/3/2	26229.21	FALSE
1	1	2010/4/2	57258.43	FALSE
1	1	2010/4/9	42960.91	FALSE
1	1	2010/4/1	17596.96	FALSE
1	1	2010/4/2	16145.35	FALSE
1	1	2010/4/3	16555.11	FALSE
1	1	2010/5/7	17413.94	FALSE
1	1	2010/5/14	18926.74	FALSE
1	1	2010/5/21	14773.04	FALSE
1	1	2010/5/2	15580.43	FALSE
1	1	2010/6/4	17558.09	FALSE
1	1	2010/6/11	16637.62	FALSE
1	1	2010/6/1	16216.27	FALSE
1	1	2010/6/2	16328.72	FALSE
1	1	2010/7/2	16333.14	FALSE
1	1	2010/7/9	17688.76	FALSE
1	1	2010/7/16	17150.84	FALSE
1	1	2010/7/2	15360.45	FALSE
1	1	2010/7/3	15381.82	FALSE
1	1	2010/8/6	17508.41	FALSE

Data Sets

Store	Dept	Date	Weekly_S	IsHoliday
1	1	2010/2/5	24924.5	FALSE
1	1	2010/2/12	46039.49	TRUE
1	1	2010/2/19	41595.55	FALSE
1	1	2010/2/2	19403.54	FALSE
1	1	2010/3/5	21827.9	FALSE
1	1	2010/3/12	21043.39	FALSE
1	1	2010/3/1	22136.64	FALSE
1	1	2010/3/2	26229.21	FALSE
1	1	2010/4/2	57258.43	FALSE
1	1	2010/4/9	42960.91	FALSE
1	1	2010/4/1	17596.96	FALSE
1	1	2010/4/2	16145.35	FALSE
1	1	2010/4/3	16555.11	FALSE
1	1	2010/5/7	17413.94	FALSE
1	1	2010/5/14	18926.74	FALSE
1	1	2010/5/21	14773.04	FALSE
1	1	2010/5/2	15580.43	FALSE
1	1	2010/6/4	17558.09	FALSE
1	1	2010/6/11	16637.62	FALSE
1	1	2010/6/1	16216.27	FALSE
1	1	2010/6/2	16328.72	FALSE
1	1	2010/7/2	16333.14	FALSE
1	1	2010/7/9	17688.76	FALSE
1	1	2010/7/16	17150.84	FALSE
1	1	2010/7/2	15360.45	FALSE
1	1	2010/7/3	15381.82	FALSE
1	1	2010/8/6	17508.41	FALSE

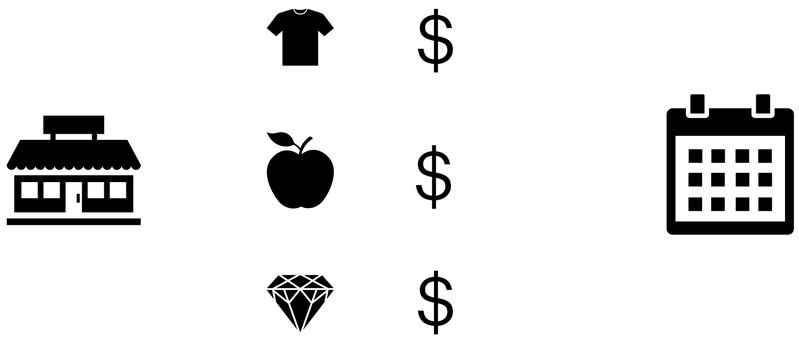
Data Sets

Store	Dept	Date	Weekly_S	IsHoliday
1	1	2010/2/5	24924.5	FALSE
1	1	2010/2/12	46039.49	TRUE
1	1	2010/2/19	41595.55	FALSE
1	1	2010/2/2	19403.54	FALSE
1	1	2010/3/5	21827.9	FALSE
1	1	2010/3/12	21043.39	FALSE
1	1	2010/3/1	22136.64	FALSE
1	1	2010/3/2	26229.21	FALSE
1	1	2010/4/2	57258.43	FALSE
1	1	2010/4/9	42960.91	FALSE
1	1	2010/4/1	17596.96	FALSE
1	1	2010/4/2	16145.35	FALSE
1	1	2010/4/3	16555.11	FALSE
1	1	2010/5/7	17413.94	FALSE
1	1	2010/5/14	18926.74	FALSE
1	1	2010/5/21	14773.04	FALSE
1	1	2010/5/2	15580.43	FALSE
1	1	2010/6/4	17558.09	FALSE
1	1	2010/6/11	16637.62	FALSE
1	1	2010/6/1	16216.27	FALSE
1	1	2010/6/2	16328.72	FALSE
1	1	2010/7/2	16333.14	FALSE
1	1	2010/7/9	17688.76	FALSE
1	1	2010/7/16	17150.84	FALSE
1	1	2010/7/2	15360.45	FALSE
1	1	2010/7/3	15381.82	FALSE
1	1	2010/8/6	17508.41	FALSE



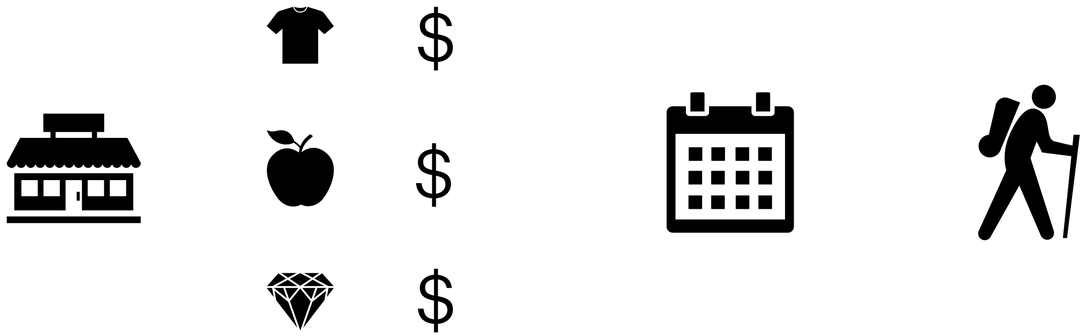
Data Sets

Store	Dept	Date	Weekly_S	IsHoliday
1	1	2010/2/5	24924.5	FALSE
1	1	2010/2/12	46039.49	TRUE
1	1	2010/2/19	41595.55	FALSE
1	1	2010/2/2	19403.54	FALSE
1	1	2010/3/5	21827.9	FALSE
1	1	2010/3/12	21043.39	FALSE
1	1	2010/3/1	22136.64	FALSE
1	1	2010/3/2	26229.21	FALSE
1	1	2010/4/2	57258.43	FALSE
1	1	2010/4/9	42960.91	FALSE
1	1	2010/4/1	17596.96	FALSE
1	1	2010/4/2	16145.35	FALSE
1	1	2010/4/3	16555.11	FALSE
1	1	2010/5/7	17413.94	FALSE
1	1	2010/5/14	18926.74	FALSE
1	1	2010/5/21	14773.04	FALSE
1	1	2010/5/2	15580.43	FALSE
1	1	2010/6/4	17558.09	FALSE
1	1	2010/6/11	16637.62	FALSE
1	1	2010/6/1	16216.27	FALSE
1	1	2010/6/2	16328.72	FALSE
1	1	2010/7/2	16333.14	FALSE
1	1	2010/7/9	17688.76	FALSE
1	1	2010/7/16	17150.84	FALSE
1	1	2010/7/2	15360.45	FALSE
1	1	2010/7/3	15381.82	FALSE
1	1	2010/8/6	17508.41	FALSE



Data Sets

Store	Dept	Date	Weekly_S	IsHoliday
1	1	2010/2/5	24924.5	FALSE
1	1	2010/2/12	46039.49	TRUE
1	1	2010/2/19	41595.55	FALSE
1	1	2010/2/2	19403.54	FALSE
1	1	2010/3/5	21827.9	FALSE
1	1	2010/3/12	21043.39	FALSE
1	1	2010/3/1	22136.64	FALSE
1	1	2010/3/2	26229.21	FALSE
1	1	2010/4/2	57258.43	FALSE
1	1	2010/4/9	42960.91	FALSE
1	1	2010/4/1	17596.96	FALSE
1	1	2010/4/2	16145.35	FALSE
1	1	2010/4/3	16555.11	FALSE
1	1	2010/5/7	17413.94	FALSE
1	1	2010/5/14	18926.74	FALSE
1	1	2010/5/21	14773.04	FALSE
1	1	2010/5/2	15580.43	FALSE
1	1	2010/6/4	17558.09	FALSE
1	1	2010/6/11	16637.62	FALSE
1	1	2010/6/1	16216.27	FALSE
1	1	2010/6/2	16328.72	FALSE
1	1	2010/7/2	16333.14	FALSE
1	1	2010/7/9	17688.76	FALSE
1	1	2010/7/16	17150.84	FALSE
1	1	2010/7/2	15360.45	FALSE
1	1	2010/7/3	15381.82	FALSE
1	1	2010/8/6	17508.41	FALSE



EDA / Data Preparation

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE

Store	Type	Size
1	A	151315
2	A	202307
3	B	37392
4	A	205863
5	B	34875
6	A	202505
7	B	70713
8	A	155078
9	B	125833

Merge

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployme	IsHoliday	Type	Size
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE	A	151315
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE	A	151315
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE	A	151315
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE	A	151315
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE	A	151315

Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployme	IsHoliday	Type	Size
1	2010/2/5	42.31	2.572	6074.12	254.39	51.98	427.39	5988.57	211.096358	8.106	FALSE	A	151315
1	2010/2/12	38.51	2.548	410.31	98	55805.51	8	554.92	211.24217	8.106	TRUE	A	151315
1	2010/2/19	39.93	2.514	5629.51	68	1398.11	2084.64	20475.32	211.289143	8.106	FALSE	A	151315
1	2010/2/26	46.63	2.561	4640.65	19	105.02	3639.42	14461.82	211.319643	8.106	FALSE	A	151315
1	2010/3/5	46.5	2.625	5011.32	67	347.37	225.79	4011.37	211.350143	8.106	FALSE	A	151315

Store	Dept	Date	Weekly_S	IsHoliday
1	1	2010/2/5	24924.5	FALSE
1	1	2010/2/12	46039.49	TRUE
1	1	2010/2/19	41595.55	FALSE
1	1	2010/2/2	19403.54	FALSE
1	1	2010/3/5	21827.9	FALSE
1	1	2010/3/12	21043.39	FALSE
1	1	2010/3/1	22136.64	FALSE
1	1	2010/3/2	26229.21	FALSE
1	1	2010/4/2	57258.43	FALSE
1	1	2010/4/9	42960.91	FALSE
1	1	2010/4/1	17596.96	FALSE
1	1	2010/4/2	16145.35	FALSE
1	1	2010/4/3	16555.11	FALSE
1	1	2010/5/7	17413.94	FALSE
1	1	2010/5/14	18926.74	FALSE
1	1	2010/5/21	14773.04	FALSE
1	1	2010/5/2	15580.43	FALSE
1	1	2010/6/4	17558.09	FALSE
1	1	2010/6/11	16637.62	FALSE

Merge

Store	Dept	Date	IsHoliday	data_type	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployme	Type	Size	Weekly_Sales
1	1	2011/11/11	0	train	59.11	3.297	10382.9	6115.67	215.07	2406.62	6551.42	217.9981	7.866	1	151315	18689.54
1	1	2011/11/18	0	train	62.25	3.308	6074.12	254.39	51.98	427.39	5988.57	218.2205	7.866	1	151315	19050.66
1	1	2011/11/25	1	train	60.14	3.236	410.31	98	55805.51	8	554.92	218.4676	7.866	1	151315	20911.25
1	1	2011/12/2	0	train	48.91	3.172	5629.51	68	1398.11	2084.64	20475.32	218.7147	7.866	1	151315	25293.49
1	1	2011/12/9	0	train	43.93	3.158	4640.65	19	105.02	3639.42	14461.82	218.9618	7.866	1	151315	33305.92
1	1	2011/12/16	0	train	51.63	3.159	5011.32	67	347.37	225.79	4011.37	219.1795	7.866	1	151315	45773.03
1	1	2011/12/23	0	train	47.96	3.112	2725.36	40.48	634.7	24.9	2739.43	219.3577	7.866	1	151315	46788.75
1	1	2011/12/30	1	train	44.55	3.129	5762.1	46011.38	260.36	983.65	4735.78	219.536	7.866	1	151315	23350.88
1	1	2012/1/6	0	train	49.01	3.157	6277.39	21813.16	143.1	1450.13	8483	219.7143	7.348	1	151315	16567.69
1	1	2012/1/13	0	train	48.53	3.261	5183.29	8025.87	42.24	453.08	3719.38	219.8925	7.348	1	151315	16894.4
1	1	2012/1/20	0	train	54.11	3.268	4139.87	2807.19	33.88	500.62	3400.21	219.9857	7.348	1	151315	18365.1
1	1	2012/1/27	0	train	54.26	3.29	1164.46	1082.74	44	11	1222.19	220.0789	7.348	1	151315	18378.16

Store	Dept	Date	IsHoliday	data_type	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployme	Type	Size	Weekly_Sales
1	1	2011/11/11	0	train	59.11	3.297	10382.9	6115.67	215.07	2406.62	6551.42	217.9981	7.866	1	151315	18689.54
1	1	2011/11/18	0	train	62.25	3.308	6074.12	254.39	51.98	427.39	5988.57	218.2205	7.866	1	151315	19050.66
1	1	2011/11/25	1	train	60.14	3.236	410.31	98	55805.51	8	554.92	218.4676	7.866	1	151315	20911.25
1	1	2011/12/2	0	train	48.91	3.172	5629.51	68	1398.11	2084.64	20475.32	218.7147	7.866	1	151315	25293.49
1	1	2011/12/9	0	train	43.93	3.158	4640.65	19	105.02	3639.42	14461.82	218.9618	7.866	1	151315	33305.92
1	1	2011/12/16	0	train	51.63	3.159	5011.32	67	347.37	225.79	4011.37	219.1795	7.866	1	151315	45773.03
1	1	2011/12/23	0	train	47.96	3.112	2725.36	40.48	634.7	24.9	2739.43	219.3577	7.866	1	151315	46788.75
1	1	2011/12/30	1	train	44.55	3.129	5762.1	46011.38	260.36	983.65	4735.78	219.536	7.866	1	151315	23350.88
1	1	2012/1/6	0	train	49.01	3.157	6277.39	21813.16	143.1	1450.13	8483	219.7143	7.348	1	151315	16567.69
1	1	2012/1/13	0	train	48.53	3.261	5183.29	8025.87	42.24	453.08	3719.38	219.8925	7.348	1	151315	16894.4
1	1	2012/1/20	0	train	54.11	3.268	4139.87	2807.19	33.88	500.62	3400.21	219.9857	7.348	1	151315	18365.1
1	1	2012/1/27	0	train	54.26	3.29	1164.46	1082.74	44	11	1222.19	220.0789	7.348	1	151315	18378.16

Store	Dept	Date	IsHoliday	data_type	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployme	Type	Size	Weekly_Sales
1	1	2011/11/11	0	train	59.11	3.297	10382.9	6115.67	215.07	2406.62	6551.42	217.9981	7.866	1	151315	18689.54
1	1	2011/11/18	0	train	62.25	3.308	6074.12	254.39	51.98	427.39	5988.57	218.2205	7.866	1	151315	19050.66
1	1	2011/11/25	1	train	60.14	3.236	410.31	98	55805.51	8	554.92	218.4676	7.866	1	151315	20911.25
1	1	2011/12/2	0	train	48.91	3.172	5629.51	68	1398.11	2084.64	20475.32	218.7147	7.866	1	151315	25293.49
1	1	2011/12/9	0	train	43.93	3.158	4640.65	19	105.02	3639.42	14461.82	218.9618	7.866	1	151315	33305.92
1	1	2011/12/16	0	train	51.63	3.159	5011.32	67	347.37	225.79	4011.37	219.1795	7.866	1	151315	45773.03
1	1	2011/12/23	0	train	47.96	3.112	2725.36	40.48	634.7	24.9	2739.43	219.3577	7.866	1	151315	46788.75
1	1	2011/12/30	1	train	44.55	3.129	5762.1	46011.38	260.36	983.65	4735.78	219.536	7.866	1	151315	23350.88
1	1	2012/1/6	0	train	49.01	3.157	6277.39	21813.16	143.1	1450.13	8483	219.7143	7.348	1	151315	16567.69
1	1	2012/1/13	0	train	48.53	3.261	5183.29	8025.87	42.24	453.08	3719.38	219.8925	7.348	1	151315	16894.4
1	1	2012/1/20	0	train	54.11	3.268	4139.87	2807.19	33.88	500.62	3400.21	219.9857	7.348	1	151315	18365.1
1	1	2012/1/27	0	train	54.26	3.29	1164.46	1082.74	44	11	1222.19	220.0789	7.348	1	151315	18378.16

Day in a week?
Week in a year?
Year?
Month in a year?

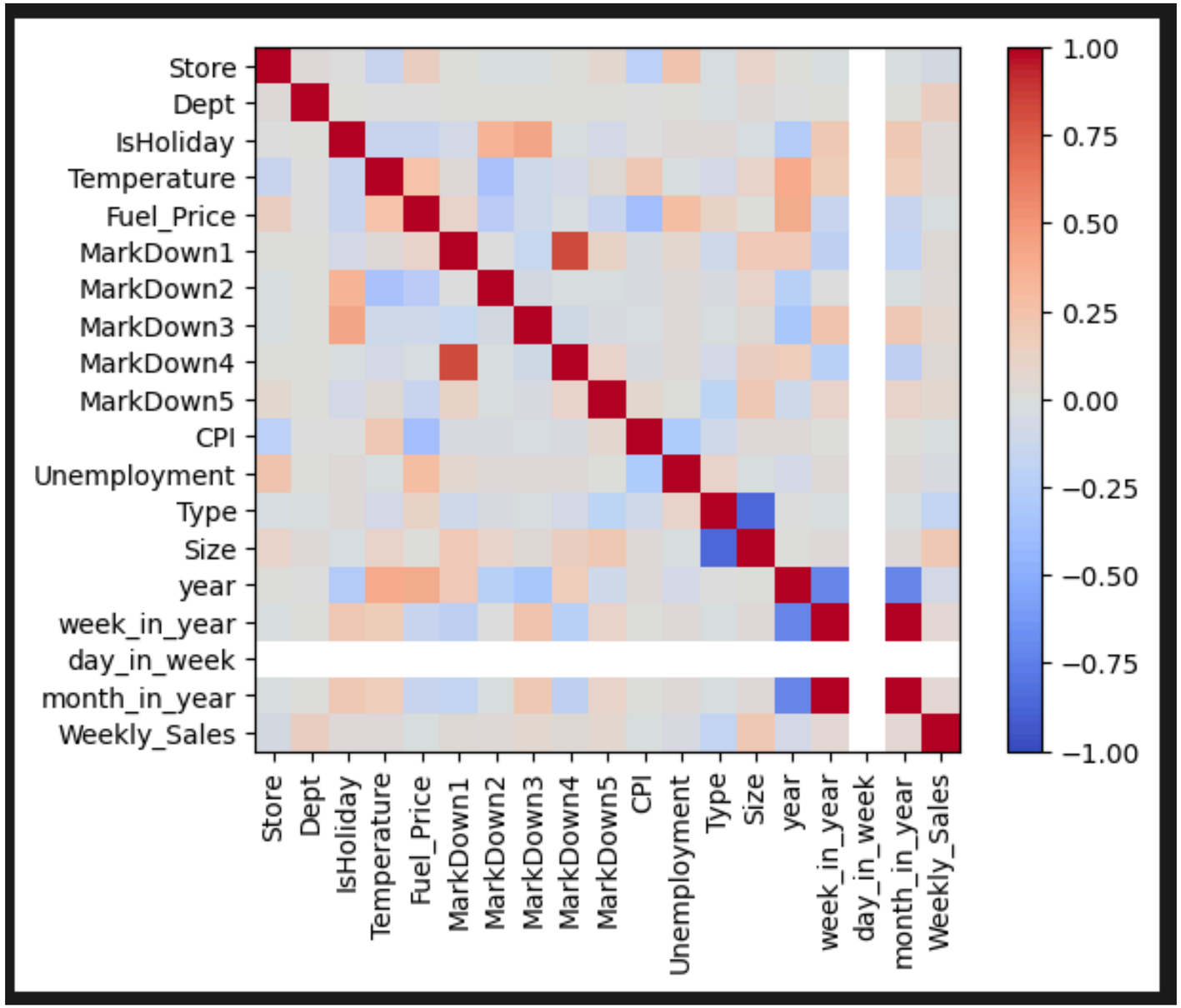
Store	Dept	Date	IsHoliday	data_type	Temperatu	Fuel_Price	MarkDown	MarkDown	MarkDown	MarkDown	MarkDown	CPI	Unemploy	Type	Size	year	week_in_y	day_in_we	month_in_	Weekly_Sa
1	1	2011/11/11	0	train	59.11	3.297	10382.9	6115.67	215.07	2406.62	6551.42	217.9981	7.866	1	151315	2011	45	5	11	18689.54
1	1	2011/11/18	0	train	62.25	3.308	6074.12	254.39	51.98	427.39	5988.57	218.2205	7.866	1	151315	2011	46	5	11	19050.66
1	1	2011/11/25	1	train	60.14	3.236	410.31	98	55805.51	8	554.92	218.4676	7.866	1	151315	2011	47	5	11	20911.25
1	1	2011/12/2	0	train	48.91	3.172	5629.51	68	1398.11	2084.64	20475.32	218.7147	7.866	1	151315	2011	48	5	12	25293.49
1	1	2011/12/9	0	train	43.93	3.158	4640.65	19	105.02	3639.42	14461.82	218.9618	7.866	1	151315	2011	49	5	12	33305.92
1	1	2011/12/16	0	train	51.63	3.159	5011.32	67	347.37	225.79	4011.37	219.1795	7.866	1	151315	2011	50	5	12	45773.03
1	1	2011/12/23	0	train	47.96	3.112	2725.36	40.48	634.7	24.9	2739.43	219.3577	7.866	1	151315	2011	51	5	12	46788.75
1	1	2011/12/30	1	train	44.55	3.129	5762.1	46011.38	260.36	983.65	4735.78	219.536	7.866	1	151315	2011	52	5	12	23350.88
1	1	2012/1/6	0	train	49.01	3.157	6277.39	21813.16	143.1	1450.13	8483	219.7143	7.348	1	151315	2012	1	5	1	16567.69
1	1	2012/1/13	0	train	48.53	3.261	5183.29	8025.87	42.24	453.08	3719.38	219.8925	7.348	1	151315	2012	2	5	1	16894.4
1	1	2012/1/20	0	train	54.11	3.268	4139.87	2807.19	33.88	500.62	3400.21	219.9857	7.348	1	151315	2012	3	5	1	18365.1
1	1	2012/1/27	0	train	54.26	3.29	1164.46	1082.74	44	11	1222.19	220.0789	7.348	1	151315	2012	4	5	1	18378.16

Store	Dept	Date	IsHoliday	data_type	Temperatu	Fuel_Price	MarkDown	MarkDown	MarkDown	MarkDown	MarkDown	CPI	Unemploy	Type	Size	year	week_in_y	day_in_we	month_in_	Weekly_Sa
1	1	2011/11/11	0	train	59.11	3.297	10382.9	6115.67	215.07	2406.62	6551.42	217.9981	7.866	1	151315	2011	45	5	11	18689.54
1	1	2011/11/18	0	train	62.25	3.308	6074.12	254.39	51.98	427.39	5988.57	218.2205	7.866	1	151315	2011	46	5	11	19050.66
1	1	2011/11/25	1	train	60.14	3.236	410.31	98	55805.51	8	554.92	218.4676	7.866	1	151315	2011	47	5	11	20911.25
1	1	2011/12/2	0	train	48.91	3.172	5629.51	68	1398.11	2084.64	20475.32	218.7147	7.866	1	151315	2011	48	5	12	25293.49
1	1	2011/12/9	0	train	43.93	3.158	4640.65	19	105.02	3639.42	14461.82	218.9618	7.866	1	151315	2011	49	5	12	33305.92
1	1	2011/12/16	0	train	51.63	3.159	5011.32	67	347.37	225.79	4011.37	219.1795	7.866	1	151315	2011	50	5	12	45773.03
1	1	2011/12/23	0	train	47.96	3.112	2725.36	40.48	634.7	24.9	2739.43	219.3577	7.866	1	151315	2011	51	5	12	46788.75
1	1	2011/12/30	1	train	44.55	3.129	5762.1	46011.38	260.36	983.65	4735.78	219.536	7.866	1	151315	2011	52	5	12	23350.88
1	1	2012/1/6	0	train	49.01	3.157	6277.39	21813.16	143.1	1450.13	8483	219.7143	7.348	1	151315	2012	1	5	1	16567.69
1	1	2012/1/13	0	train	48.53	3.261	5183.29	8025.87	42.24	453.08	3719.38	219.8925	7.348	1	151315	2012	2	5	1	16894.4
1	1	2012/1/20	0	train	54.11	3.268	4139.87	2807.19	33.88	500.62	3400.21	219.9857	7.348	1	151315	2012	3	5	1	18365.1
1	1	2012/1/27	0	train	54.26	3.29	1164.46	1082.74	44	11	1222.19	220.0789	7.348	1	151315	2012	4	5	1	18378.16

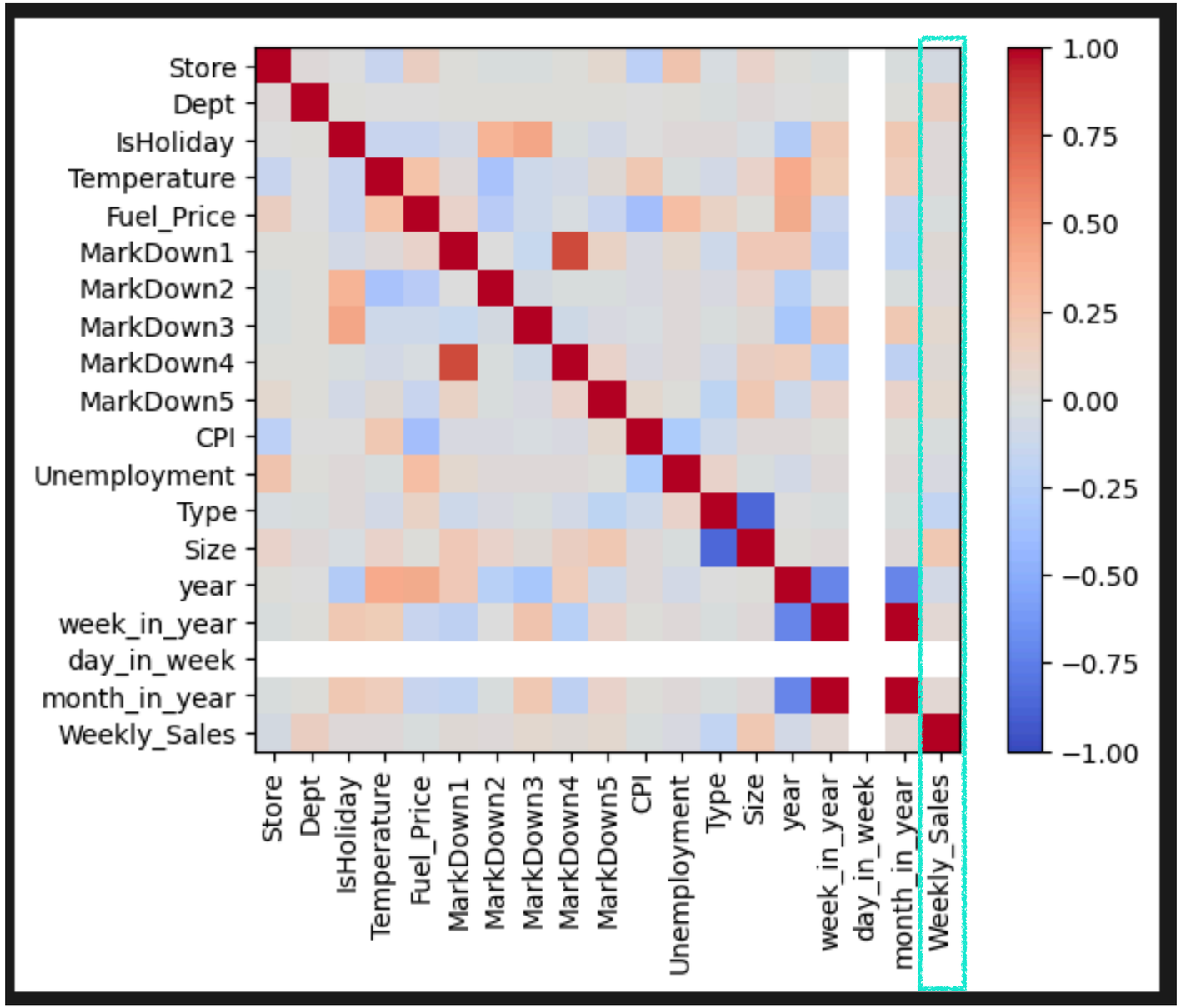
Correlation Analysis

Store	Dept	Date	IsHoliday	data_type	Temperatu	Fuel_Price	MarkDown	MarkDown	MarkDown	MarkDown	MarkDown	CPI	Unemploy	Type	Size	year	week_in_y	day_in_we	month_in_	Weekly_Sa
1	1	2011/11/11	0	train	59.11	3.297	10382.9	6115.67	215.07	2406.62	6551.42	217.9981	7.866	1	151315	2011	45	5	11	18689.54
1	1	2011/11/18	0	train	62.25	3.308	6074.12	254.39	51.98	427.39	5988.57	218.2205	7.866	1	151315	2011	46	5	11	19050.66
1	1	2011/11/25	1	train	60.14	3.236	410.31	98	55805.51	8	554.92	218.4676	7.866	1	151315	2011	47	5	11	20911.25
1	1	2011/12/2	0	train	48.91	3.172	5629.51	68	1398.11	2084.64	20475.32	218.7147	7.866	1	151315	2011	48	5	12	25293.49
1	1	2011/12/9	0	train	43.93	3.158	4640.65	19	105.02	3639.42	14461.82	218.9618	7.866	1	151315	2011	49	5	12	33305.92
1	1	2011/12/16	0	train	51.63	3.159	5011.32	67	347.37	225.79	4011.37	219.1795	7.866	1	151315	2011	50	5	12	45773.03
1	1	2011/12/23	0	train	47.96	3.112	2725.36	40.48	634.7	24.9	2739.43	219.3577	7.866	1	151315	2011	51	5	12	46788.75
1	1	2011/12/30	1	train	44.55	3.129	5762.1	46011.38	260.36	983.65	4735.78	219.536	7.866	1	151315	2011	52	5	12	23350.88
1	1	2012/1/6	0	train	49.01	3.157	6277.39	21813.16	143.1	1450.13	8483	219.7143	7.348	1	151315	2012	1	5	1	16567.69
1	1	2012/1/13	0	train	48.53	3.261	5183.29	8025.87	42.24	453.08	3719.38	219.8925	7.348	1	151315	2012	2	5	1	16894.4
1	1	2012/1/20	0	train	54.11	3.268	4139.87	2807.19	33.88	500.62	3400.21	219.9857	7.348	1	151315	2012	3	5	1	18365.1
1	1	2012/1/27	0	train	54.26	3.29	1164.46	1082.74	44	11	1222.19	220.0789	7.348	1	151315	2012	4	5	1	18378.16

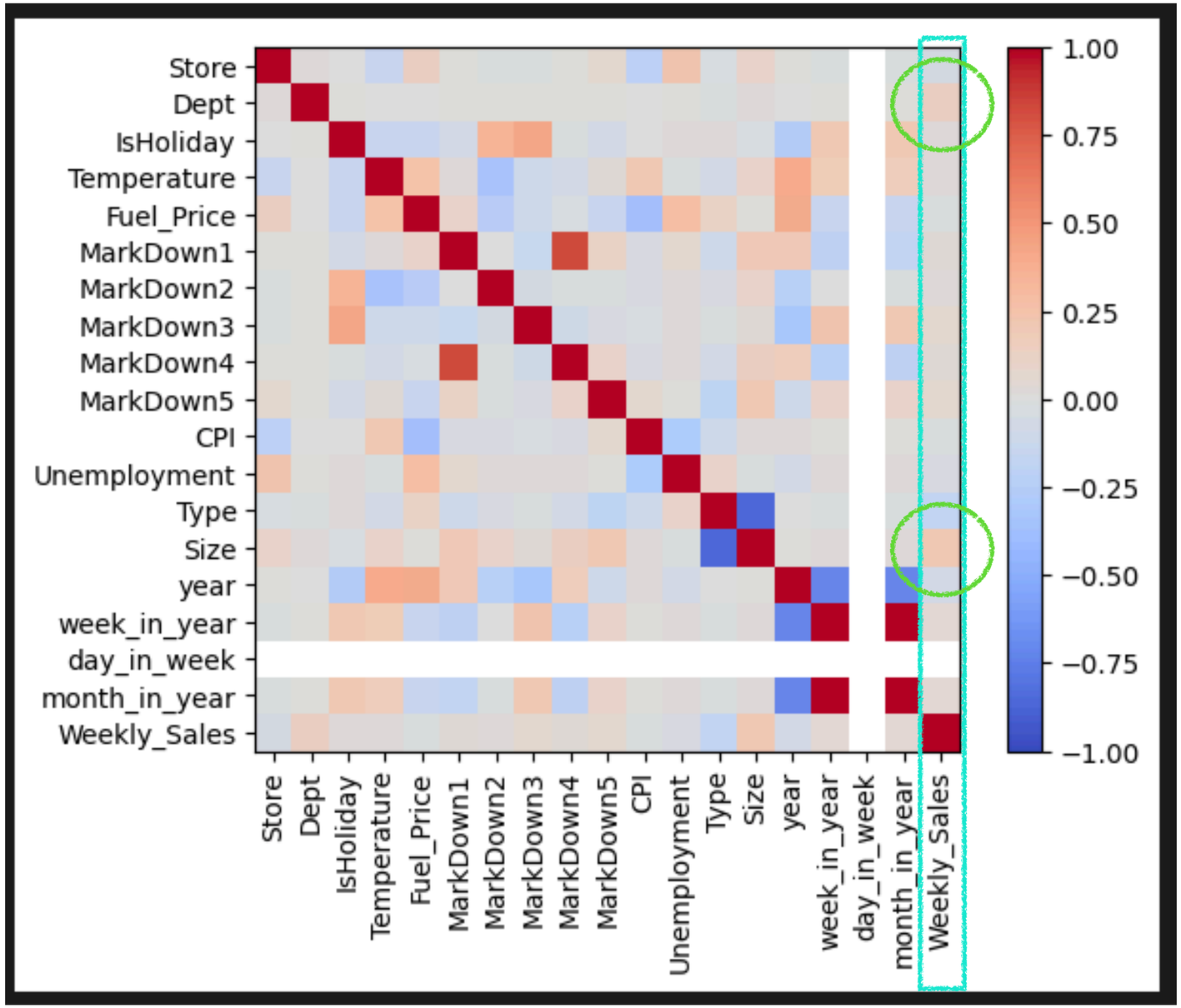
Correlation Analysis



Correlation Analysis



Correlation Analysis



Correlation Analysis

```
# Calculate the absolute correlation values
abs_corr = train_cor.abs()
# Extract the top n features based on absolute correlation
top_features = abs_corr.nlargest(6, 'Weekly_Sales')['Weekly_Sales'].index
# Print the top n features
print(top_features)
```

✓ 0.0s

```
Index(['Weekly_Sales', 'Size', 'Type', 'Dept', 'MarkDown3', 'MarkDown5'], dtype='object')
```

Correlation Analysis

```
# Calculate the absolute correlation values
abs_corr = train_cor.abs()
# Extract the top n features based on absolute correlation
top_features = abs_corr.nlargest(6, 'Weekly_Sales')['Weekly_Sales'].index
# Print the top n features
print(top_features)
```

✓ 0.0s

```
Index(['Weekly_Sales', 'Size', 'Type', 'Dept', 'MarkDown3', 'MarkDown5'], dtype='object')
```

Model Selection

```
models = {
    'LinearRegression' : LinearRegression(),
    'RandomForestRegressor' : RandomForestRegressor(),
}
```

✓ 0.0s

```
def try_model(model, X_train, y_train, X_valid, y_valid):
    model.fit(X_train, y_train)
    preds = model.predict(X_valid)
    rmse = mean_squared_error(y_valid, preds, squared=False)
    return rmse
```

✓ 0.0s

```
for model_name, model_object in models.items():
    rmse = try_model(model_object, X_train, y_train, X_valid, y_valid)
    print(f"{model_name} : {model_object}")
    print(rmse)
```

✓ 15.2s

```
LinearRegression : LinearRegression()
23848.217512448035
RandomForestRegressor : RandomForestRegressor()
6102.184544169955
```

Model Selection

```
LR = LinearRegression()
LR.fit(X_train, y_train)
preds = LR.predict(X_valid)
rmse = mean_squared_error(y_valid, preds, squared=False)
print(rmse)
```

✓ 0.0s

```
24357.816773671082
```

Model Selection

```
LR = LinearRegression()
LR.fit(X_train, y_train)
preds = LR.predict(X_valid)
rmse = mean_squared_error(y_valid, preds, squared=False)
print(rmse)
```

✓ 0.0s

```
24357.816773671082
```

```
RF = RandomForestRegressor()
RF.fit(X_train, y_train)
preds = RF.predict(X_valid)
rmse = mean_squared_error(y_valid, preds, squared=False)
print(rmse)
```

✓ 15.0s

```
6246.255231146499
```

RandomForest Fitting

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
```

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

✓ 18.3s

```
RMSE: 2302.3776215887024
```

RandomForest Fitting

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
```

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

✓ 18.3s

```
RMSE: 2302.3776215887024
```

RandomForest Fitting

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
```

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

✓ 18.3s

```
RMSE: 2302.3776215887024
```

Evaluation

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

18.3s

```
RMSE: 2302.3776215887024
```

Is 2261 good or bad?

Evaluation

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

18.3s

```
RMSE: 2302.3776215887024
```

Is 2261 good or bad?

Min: 500
Max: 3000

Evaluation

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

18.3s

```
RMSE: 2302.3776215887024
```

Is 2261 good or bad?

Min: 500
Max: 3000

Very Bad

Evaluation

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

18.3s

```
RMSE: 2302.3776215887024
```

Is 2261 good or bad?

Min: 500
Max: 3000

Very Bad

Min: 5000
Max: 30000

Evaluation

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

18.3s

```
RMSE: 2302.3776215887024
```

Is 2261 good or bad?

Min: 500
Max: 3000

Very Bad

Min: 5000
Max: 30000

Quite Good

Evaluation

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

✓ 18.3s

```
RMSE: 2302.3776215887024
```

```
#Normalize rmse with max-min
rmse / (test_y.max()-test_y.min())
```

✓ 0.0s

```
0.0060109657981701784
```

Evaluation

```
random_forest_model = RandomForestRegressor(random_state=5648)
random_forest_model.fit(X, y)
test_x = X_valid[features].copy()
test_y = y_valid
y_pred = random_forest_model.predict(test_x)

rmse = np.sqrt(mean_squared_error(test_y , y_pred))
print("RMSE:", rmse)
```

✓ 18.3s

```
RMSE: 2302.3776215887024
```

```
#Normalize rmse with max-min
rmse / (test_y.max()-test_y.min())
```

✓ 0.0s

```
0.0060109657981701784
```

Sounds Great!