Walmart Recruiting: Store Sales Forecasting

CRISELYN R. SANTIAGO

MoTIS_9

AGENDA

Walmart

1. Goal
2. Dataset Exploration
3. Benefits
4. Deliverables
5. Approach
6. Ethical Implications

Goal

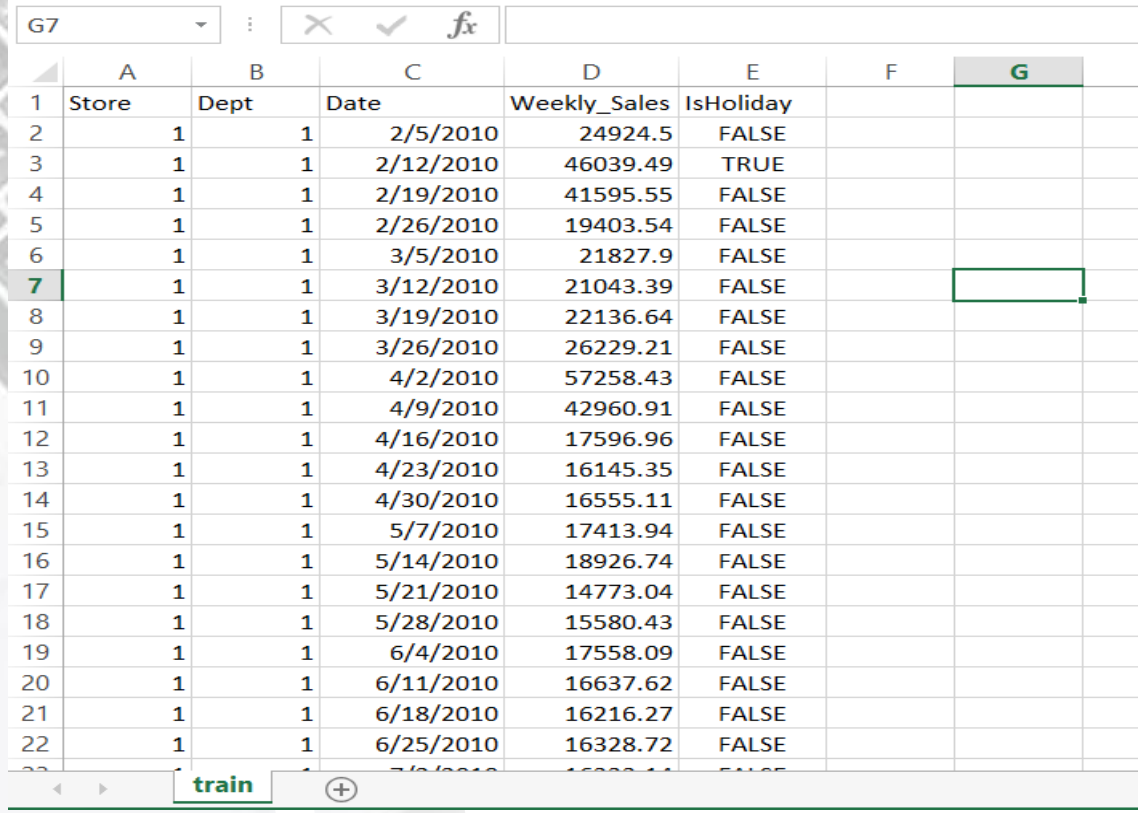
"To create a predictive model to predict the weekly sales of 45 retail stores of Walmart."

Benefits

- Predictable sales during holidays
- Analysis of the effect of markdown
- This will help the Walmart for taking better decision on which inventory product they will focus and what kind of product they will need for that holiday.
- Also, to avoid overstocking in a store which is a waste money for them

Dataset Exploration

- Train.csv → historical training data which covers to 2010-02-05 to 2012-11-01



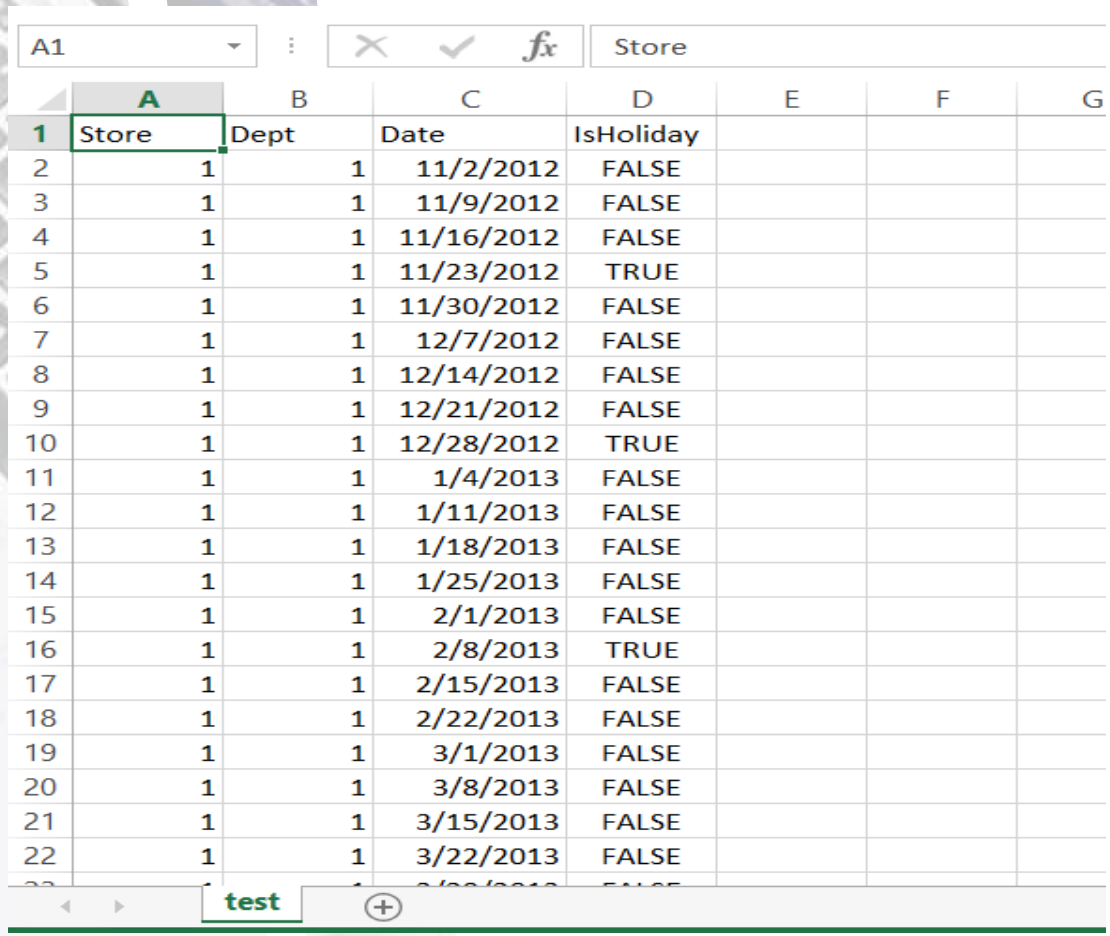
G7

	A	B	C	D	E	F	G
1	Store	Dept	Date	Weekly_Sales	IsHoliday		
2	1	1	2/5/2010	24924.5	FALSE		
3	1	1	2/12/2010	46039.49	TRUE		
4	1	1	2/19/2010	41595.55	FALSE		
5	1	1	2/26/2010	19403.54	FALSE		
6	1	1	3/5/2010	21827.9	FALSE		
7	1	1	3/12/2010	21043.39	FALSE		
8	1	1	3/19/2010	22136.64	FALSE		
9	1	1	3/26/2010	26229.21	FALSE		
10	1	1	4/2/2010	57258.43	FALSE		
11	1	1	4/9/2010	42960.91	FALSE		
12	1	1	4/16/2010	17596.96	FALSE		
13	1	1	4/23/2010	16145.35	FALSE		
14	1	1	4/30/2010	16555.11	FALSE		
15	1	1	5/7/2010	17413.94	FALSE		
16	1	1	5/14/2010	18926.74	FALSE		
17	1	1	5/21/2010	14773.04	FALSE		
18	1	1	5/28/2010	15580.43	FALSE		
19	1	1	6/4/2010	17558.09	FALSE		
20	1	1	6/11/2010	16637.62	FALSE		
21	1	1	6/18/2010	16216.27	FALSE		
22	1	1	6/25/2010	16328.72	FALSE		

train

Dataset Exploration

- Test.csv → identical to train except the weekly sales



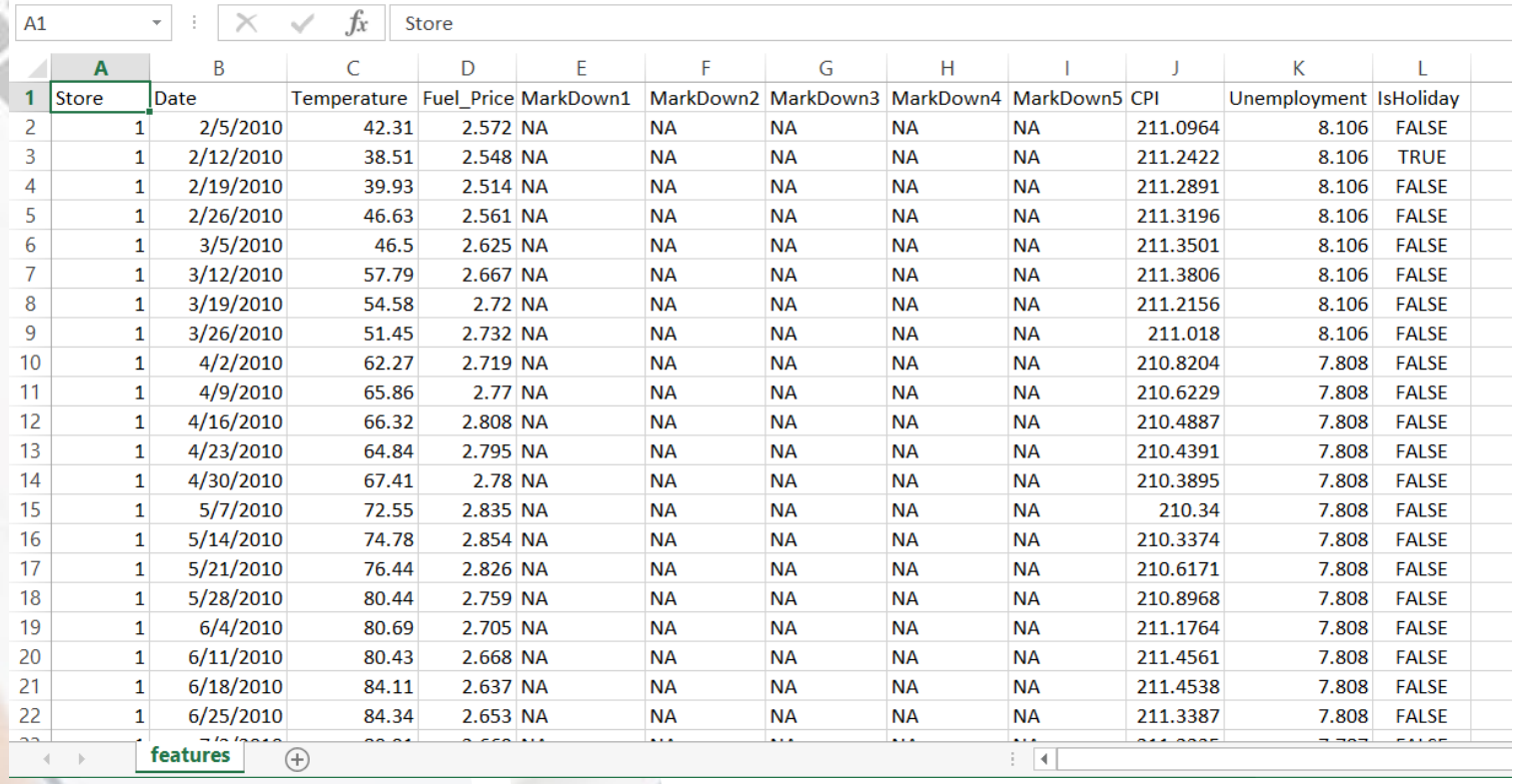
A1 | Store

	A	B	C	D	E	F	G
1	Store	Dept	Date	IsHoliday			
2	1	1	11/2/2012	FALSE			
3	1	1	11/9/2012	FALSE			
4	1	1	11/16/2012	FALSE			
5	1	1	11/23/2012	TRUE			
6	1	1	11/30/2012	FALSE			
7	1	1	12/7/2012	FALSE			
8	1	1	12/14/2012	FALSE			
9	1	1	12/21/2012	FALSE			
10	1	1	12/28/2012	TRUE			
11	1	1	1/4/2013	FALSE			
12	1	1	1/11/2013	FALSE			
13	1	1	1/18/2013	FALSE			
14	1	1	1/25/2013	FALSE			
15	1	1	2/1/2013	FALSE			
16	1	1	2/8/2013	TRUE			
17	1	1	2/15/2013	FALSE			
18	1	1	2/22/2013	FALSE			
19	1	1	3/1/2013	FALSE			
20	1	1	3/8/2013	FALSE			
21	1	1	3/15/2013	FALSE			
22	1	1	3/22/2013	FALSE			

test

Dataset Exploration

- Features.csv → additional data related to the store, department and regional activity for the given dates



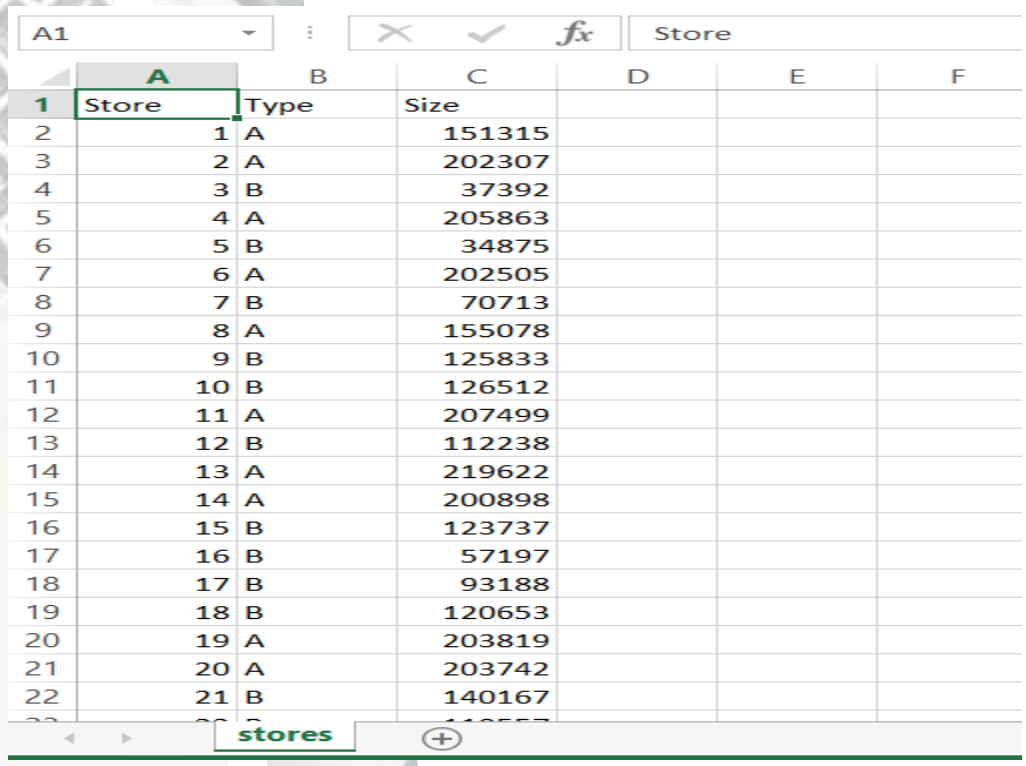
A1 | Store

	A	B	C	D	E	F	G	H	I	J	K	L
1	Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
2	1	2/5/2010	42.31	2.572	NA	NA	NA	NA	NA	211.0964	8.106	FALSE
3	1	2/12/2010	38.51	2.548	NA	NA	NA	NA	NA	211.2422	8.106	TRUE
4	1	2/19/2010	39.93	2.514	NA	NA	NA	NA	NA	211.2891	8.106	FALSE
5	1	2/26/2010	46.63	2.561	NA	NA	NA	NA	NA	211.3196	8.106	FALSE
6	1	3/5/2010	46.5	2.625	NA	NA	NA	NA	NA	211.3501	8.106	FALSE
7	1	3/12/2010	57.79	2.667	NA	NA	NA	NA	NA	211.3806	8.106	FALSE
8	1	3/19/2010	54.58	2.72	NA	NA	NA	NA	NA	211.2156	8.106	FALSE
9	1	3/26/2010	51.45	2.732	NA	NA	NA	NA	NA	211.018	8.106	FALSE
10	1	4/2/2010	62.27	2.719	NA	NA	NA	NA	NA	210.8204	7.808	FALSE
11	1	4/9/2010	65.86	2.77	NA	NA	NA	NA	NA	210.6229	7.808	FALSE
12	1	4/16/2010	66.32	2.808	NA	NA	NA	NA	NA	210.4887	7.808	FALSE
13	1	4/23/2010	64.84	2.795	NA	NA	NA	NA	NA	210.4391	7.808	FALSE
14	1	4/30/2010	67.41	2.78	NA	NA	NA	NA	NA	210.3895	7.808	FALSE
15	1	5/7/2010	72.55	2.835	NA	NA	NA	NA	NA	210.34	7.808	FALSE
16	1	5/14/2010	74.78	2.854	NA	NA	NA	NA	NA	210.3374	7.808	FALSE
17	1	5/21/2010	76.44	2.826	NA	NA	NA	NA	NA	210.6171	7.808	FALSE
18	1	5/28/2010	80.44	2.759	NA	NA	NA	NA	NA	210.8968	7.808	FALSE
19	1	6/4/2010	80.69	2.705	NA	NA	NA	NA	NA	211.1764	7.808	FALSE
20	1	6/11/2010	80.43	2.668	NA	NA	NA	NA	NA	211.4561	7.808	FALSE
21	1	6/18/2010	84.11	2.637	NA	NA	NA	NA	NA	211.4538	7.808	FALSE
22	1	6/25/2010	84.34	2.653	NA	NA	NA	NA	NA	211.3387	7.808	FALSE

features

Dataset Exploration

- Store.csv → contains anonymized information about the 45 stores, indicating the type and size of store



A1 | Store

	A	B	C	D	E	F
1	Store	Type	Size			
2	1	A	151315			
3	2	A	202307			
4	3	B	37392			
5	4	A	205863			
6	5	B	34875			
7	6	A	202505			
8	7	B	70713			
9	8	A	155078			
10	9	B	125833			
11	10	B	126512			
12	11	A	207499			
13	12	B	112238			
14	13	A	219622			
15	14	A	200898			
16	15	B	123737			
17	16	B	57197			
18	17	B	93188			
19	18	B	120653			
20	19	A	203819			
21	20	A	203742			
22	21	B	140167			

stores

Deliverables

- Sales Analysis
- Forecasting
- A model that can predict the weekly sales of different 45 stores having at least 60% of accuracy

Approach

- Data Cleaning and build some features
- Predictive Modeling
- Visualization

NOTE: Further explanation in my report

Ethical Implications

DATA SECURITY

DATA PRIVACY

THANK YOU! ☺