BIG BOX. INC.

Lower prices, guaranteed.

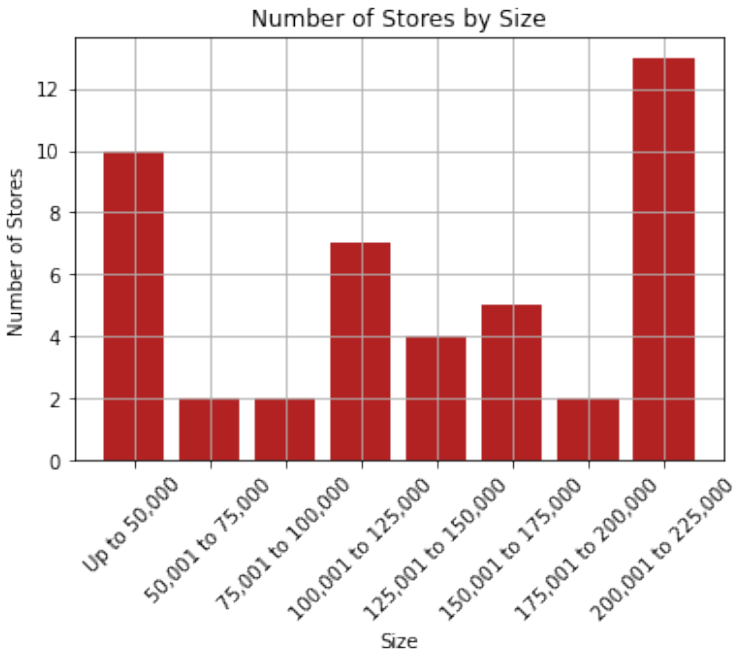
SALES FORECASTING USING TIME SERIES MODELING

WE WILL DESCRIBE THE DATA, OUR EXPLORATORY ANALYSIS, AND OUR MODELING AND FORECASTING OF WEEKLY SALES

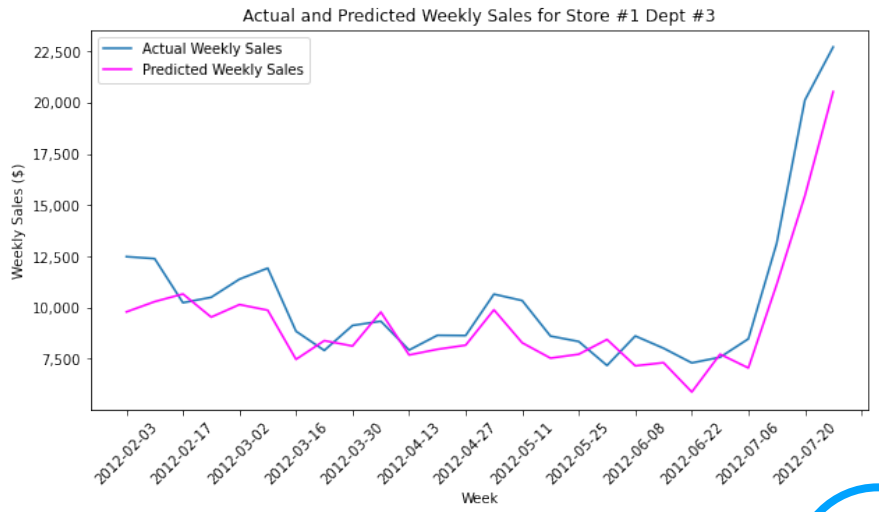
The Data

	Store	Dept	Date	Weekly_Sales	IsHoliday
1	1	1	2010-02-05	24924.50	False
2	1	1	2010-02-12	46039.49	True
3	1	1	2010-02-19	41595.55	False
4	1	1	2010-02-26	19403.54	False
5	1	1	2010-03-05	21827.90	False

Exploration



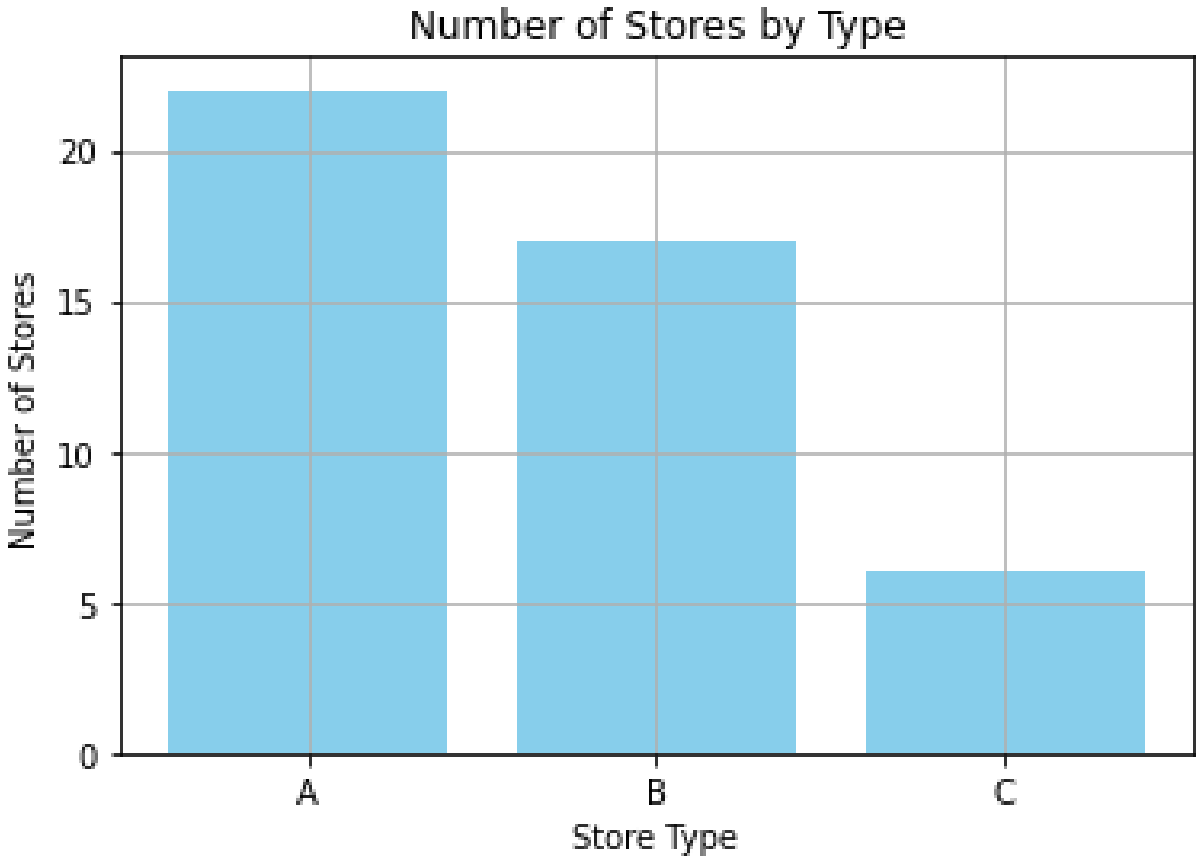
Modeling and Forecasting



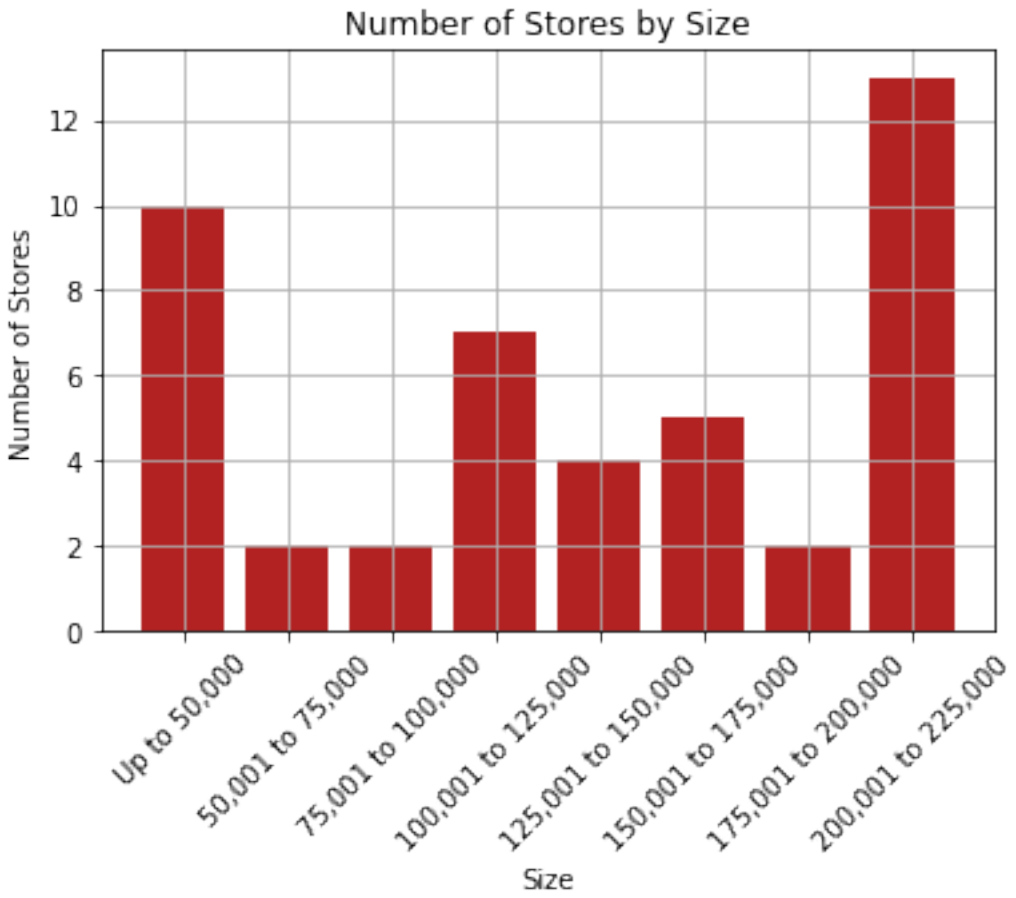
THE TEAM WAS PROVIDED WITH DATA RELATED TO THE STORES AND THEIR SURROUNDING AREAS, THE DEPTS, AND HOLIDAYS

Store	• Type, Size • Markdowns
Department	• Weekly sales (Feb 2010 to July 2012) • Holiday weeks
Surrounding Area	• Weekly avg temperature and fuel price • Weekly avg CPI and unemployment rate

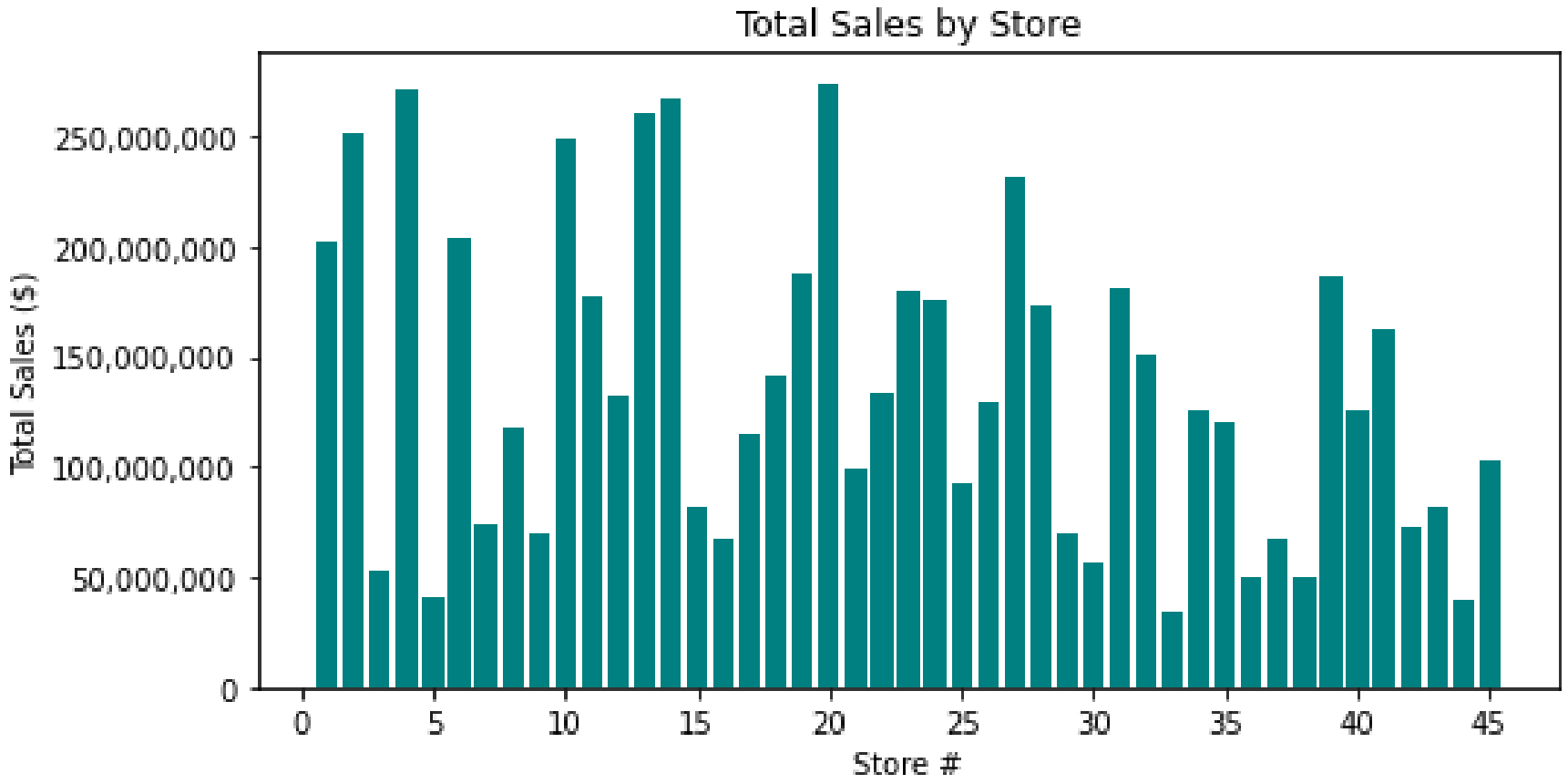
MOST OF THE STORES ARE TYPE A AND TYPE B, WHILE MUCH FEWER ARE TYPE C



THE SIZES OF THE STORES VARY WHERE THE SMALLEST AND LARGEST GROUPS MAKE UP 50% OF ALL STORES



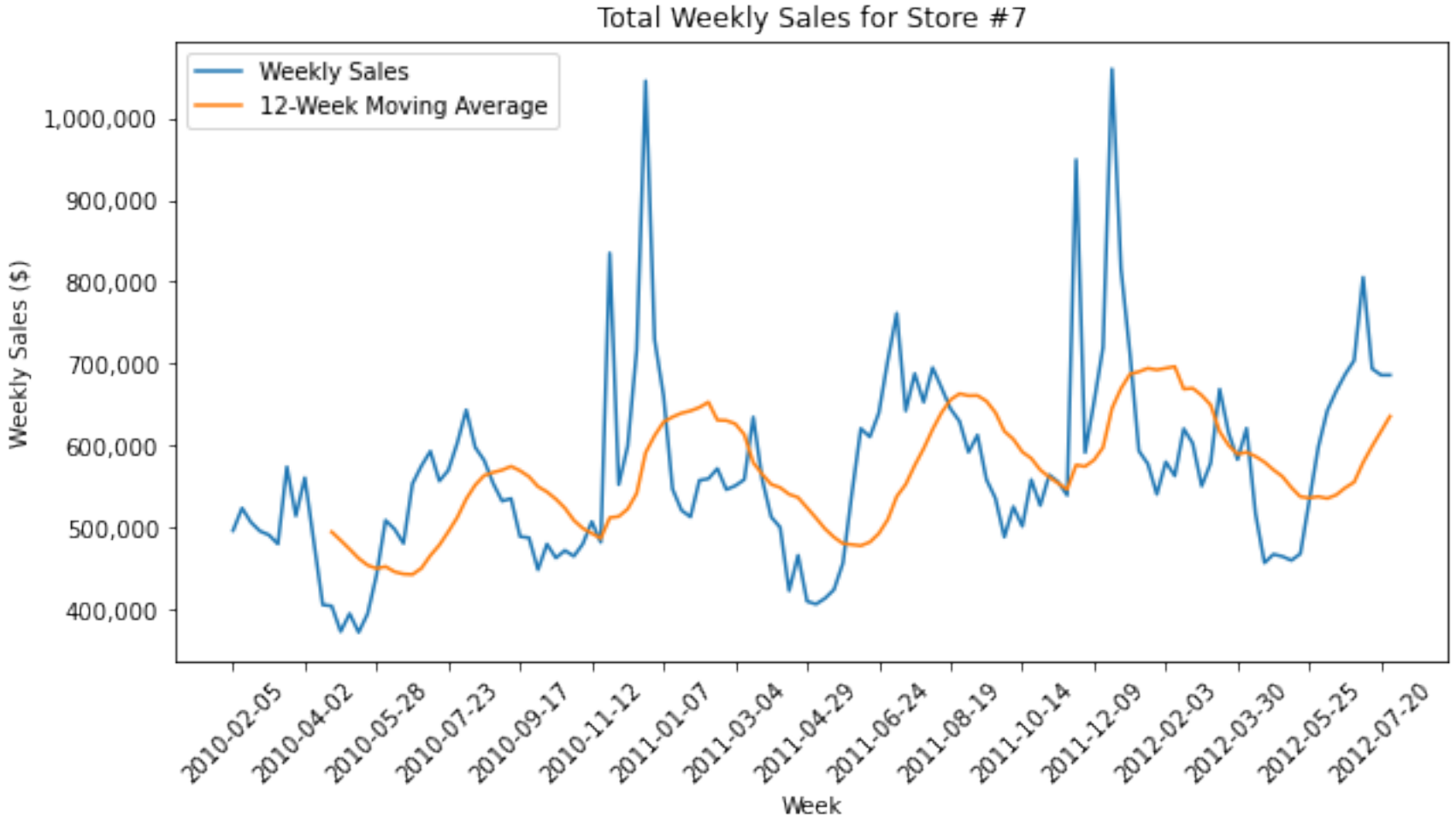
TOTAL STORE SALES VARY WHERE STORE #20 HAS THE HIGHEST SALES AND STORE #33 HAS THE LOWEST SALES



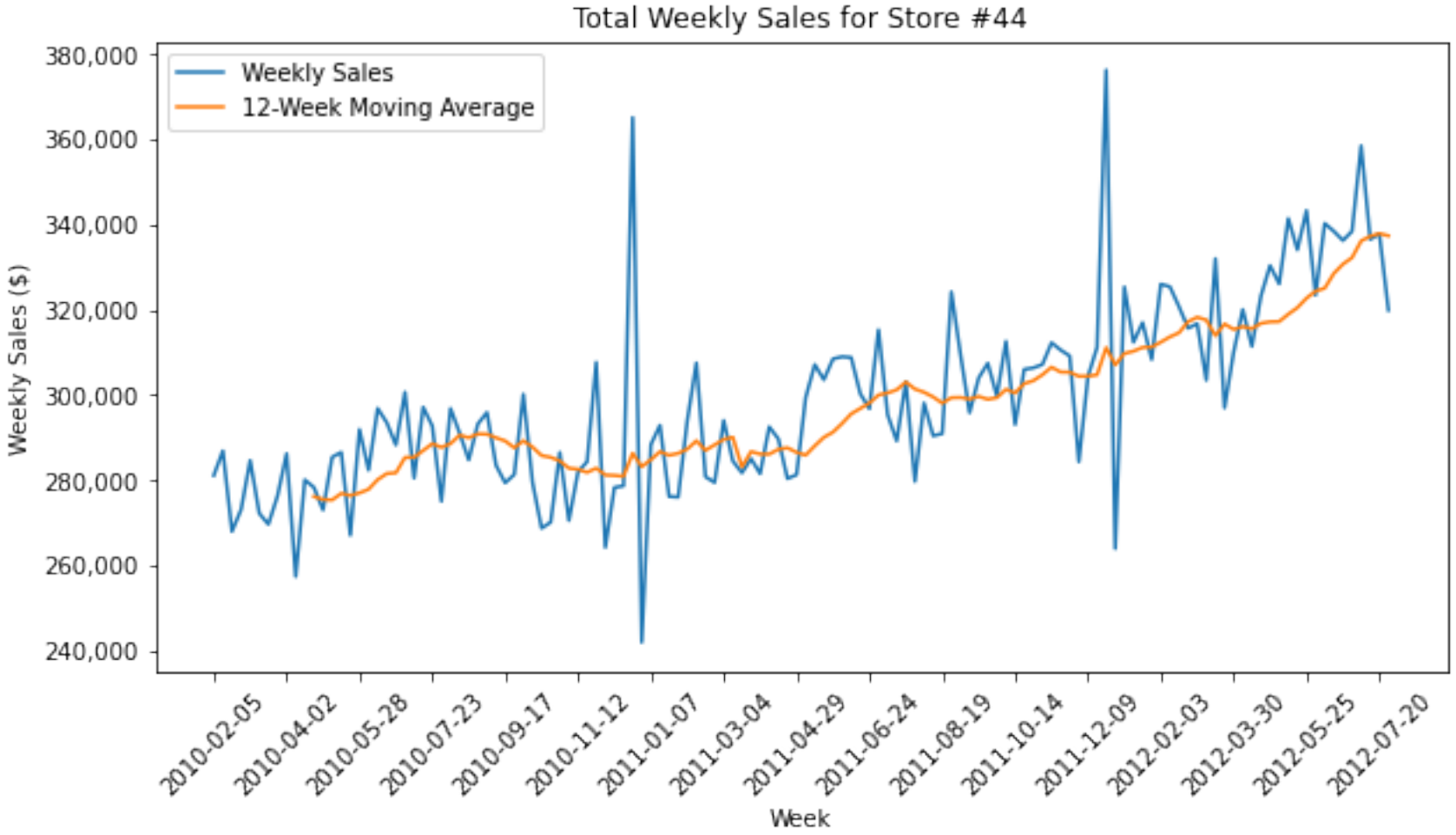
TOTAL STORE SALES VARY WHERE STORE #20 HAS THE HIGHEST SALES AND STORE #33 HAS THE LOWEST

Store #	Total Sales (million of dollars)
20	274.4
4	271.7
14	267.6
13	260.1
...	...
5	41.3
44	38.9
33	33.7

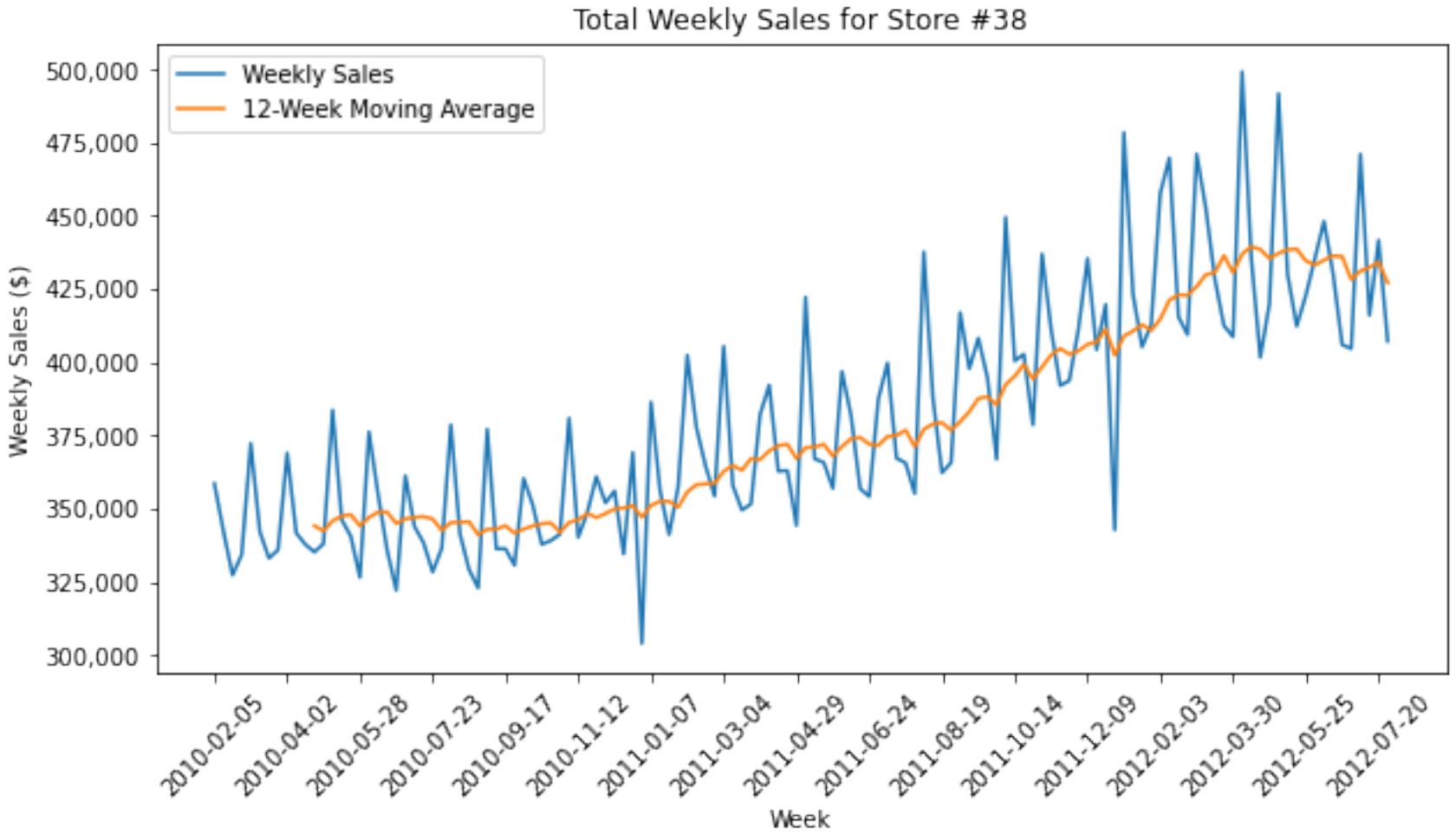
WEEKLY STORE SALES VARY WHERE SOME HAVE A SEASONAL PATTERN



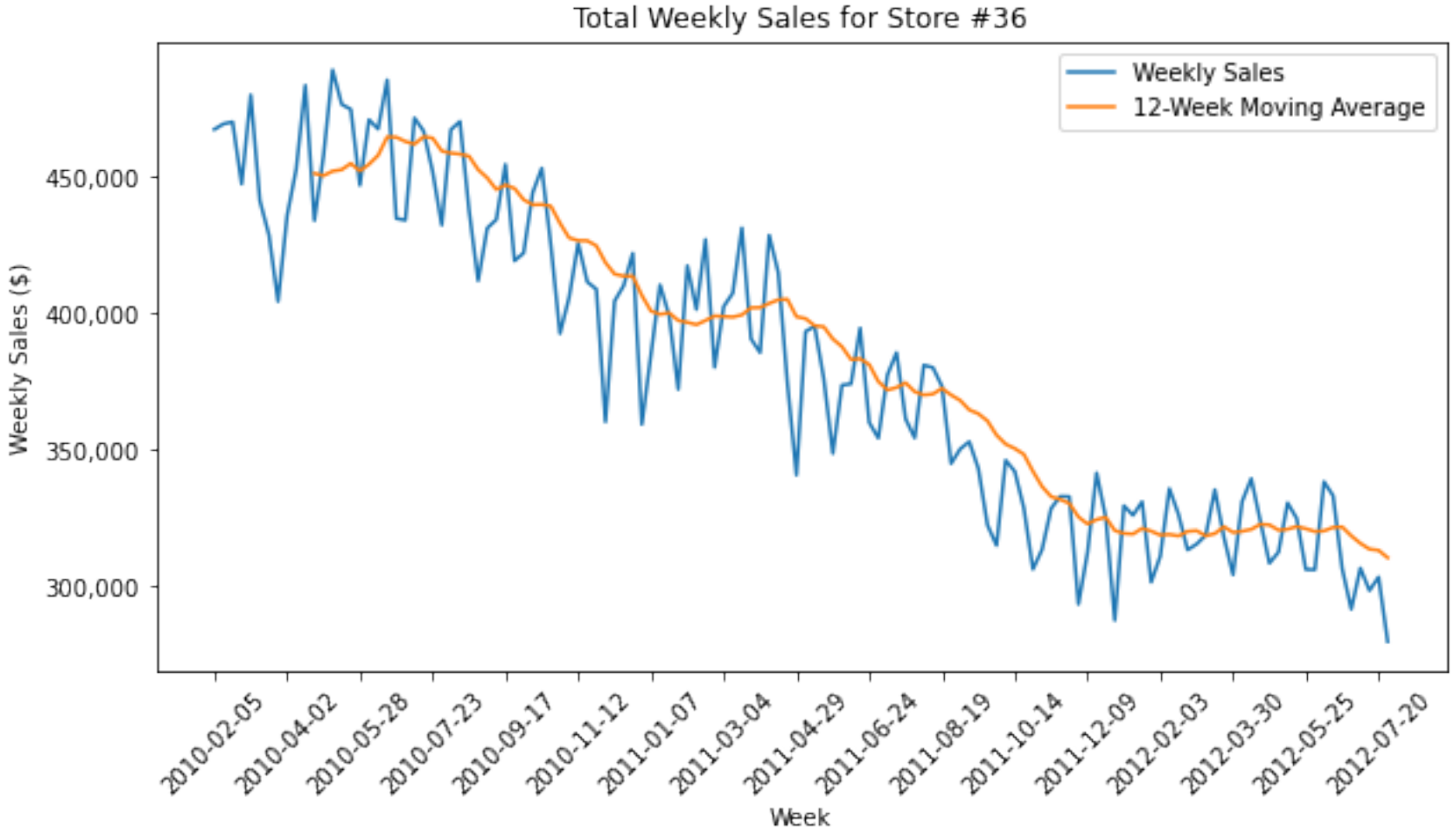
WEEKLY STORE SALES VARY WHERE SOME HAVE PEAKS



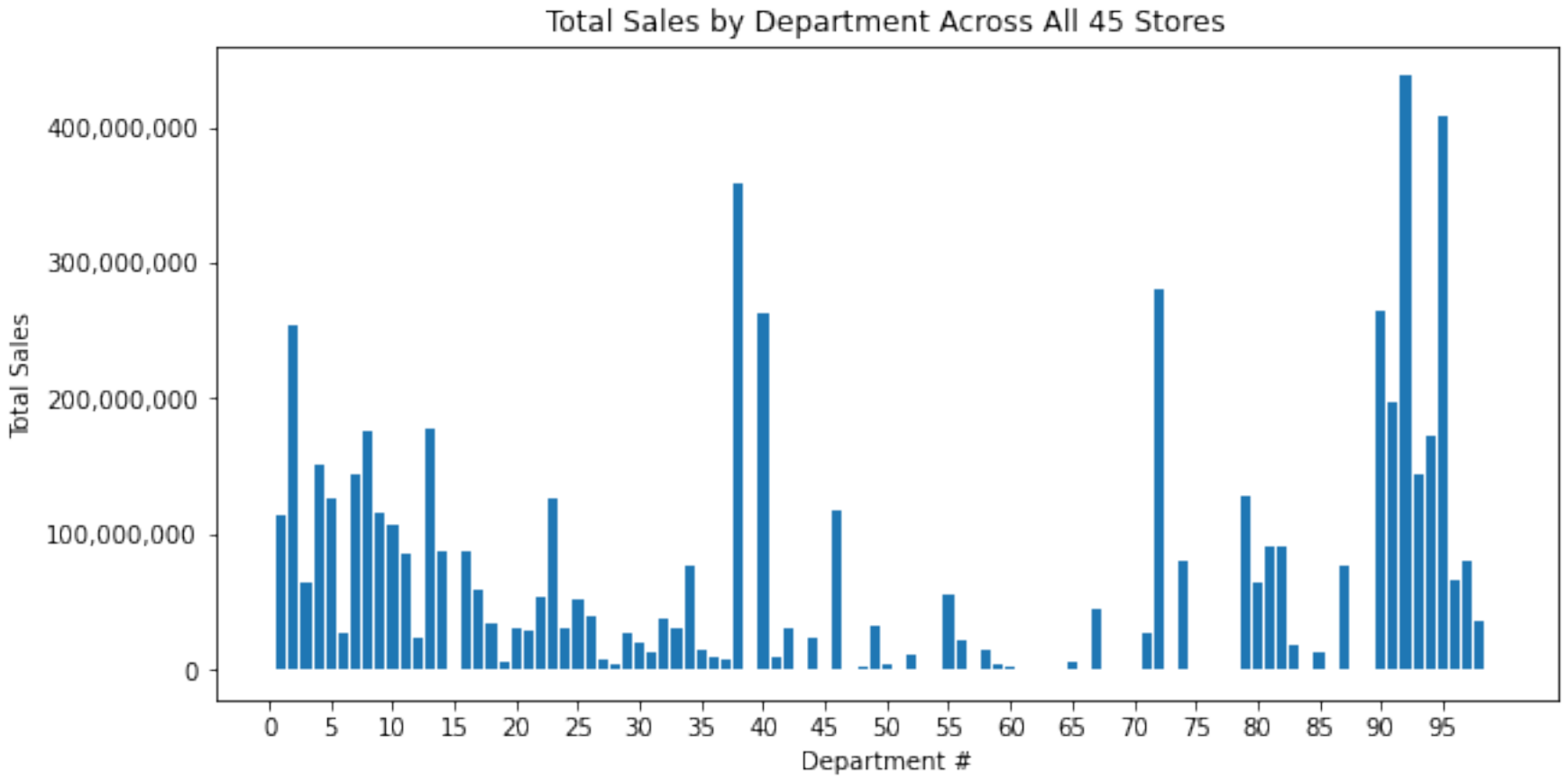
WEEKLY STORE SALES VARY WHERE SOME HAVE AN INCREASING TREND



WEEKLY STORE SALES VARY WHERE SOME HAVE A DECREASING TREND



TOTAL DEPARTMENT SALES* VARY WHERE DEPT #92 HAS THE HIGHEST SALES AND DEPT #47 HAS NEGATIVE SALES



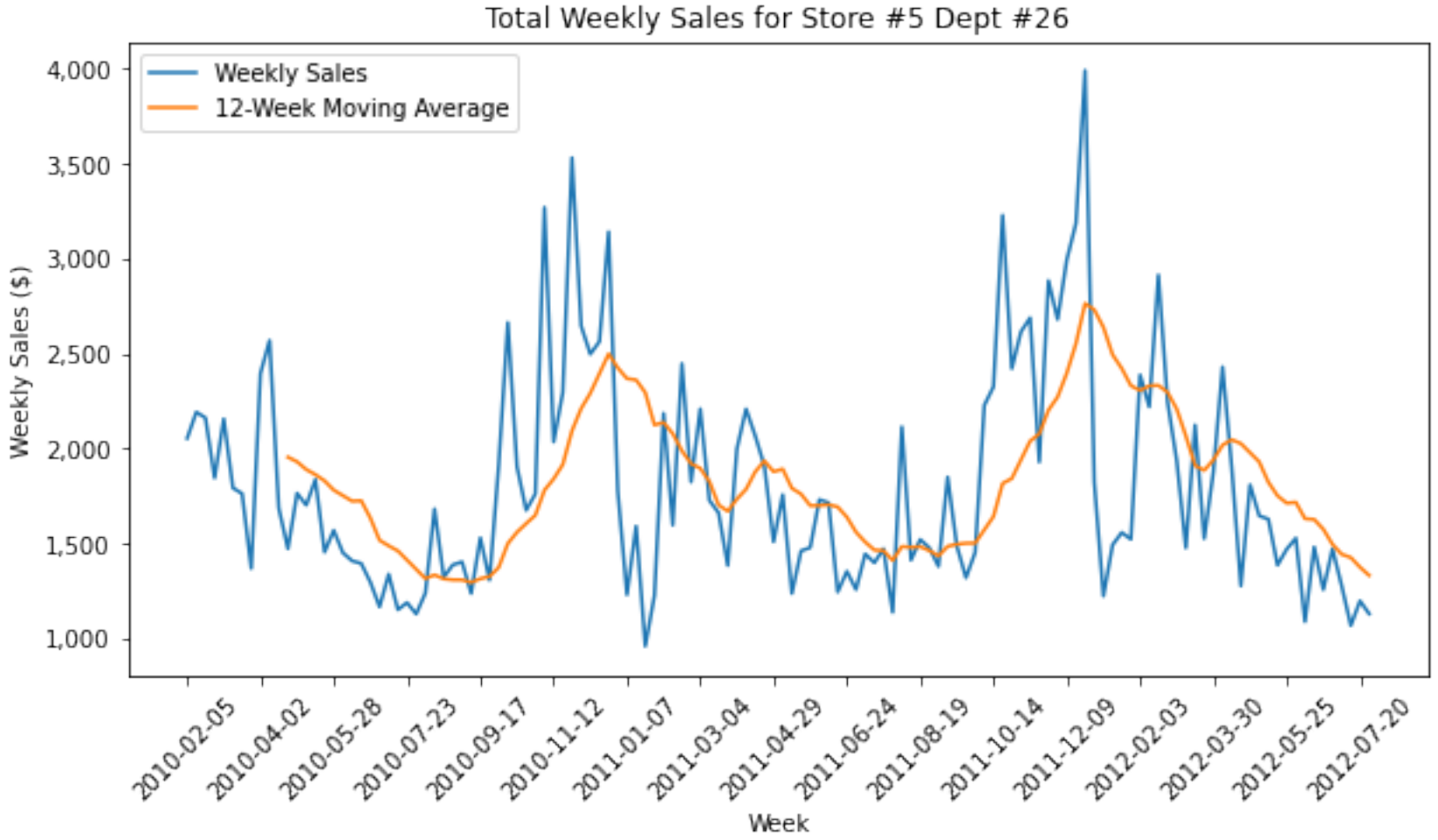
* Sales for one department across all stores

TOTAL DEPARTMENT SALES* VARY WHERE DEPT #92 HAS THE HIGHEST SALES AND DEPT #47 HAS NEGATIVE SALES

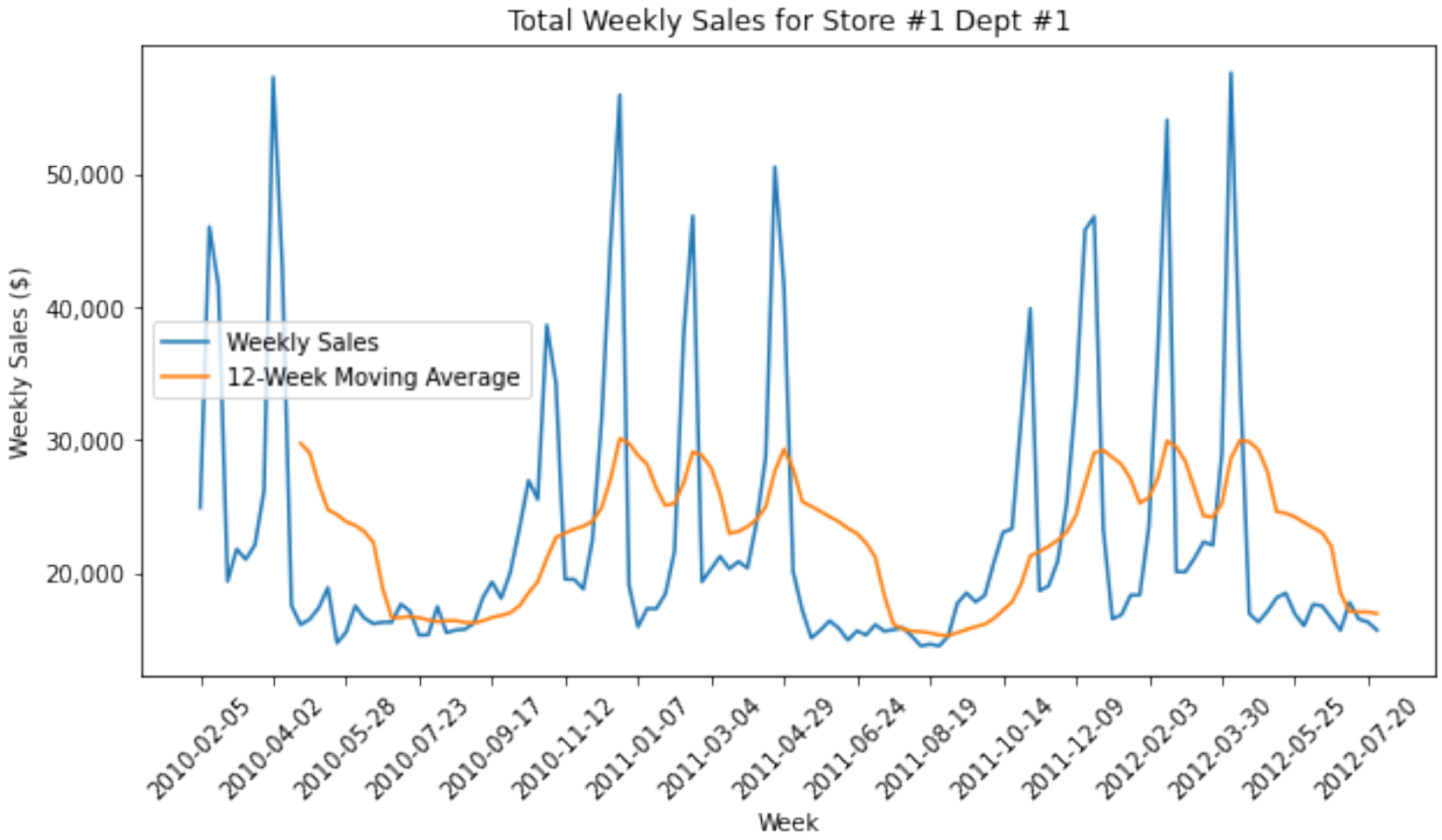
Department #	Total Sales (Dollars)
92	437,492,300
95	407,511,100
38	358,290,900
...	...
78	1,172
39	41
43	14
47	-4,654

* Sales for one department across all stores

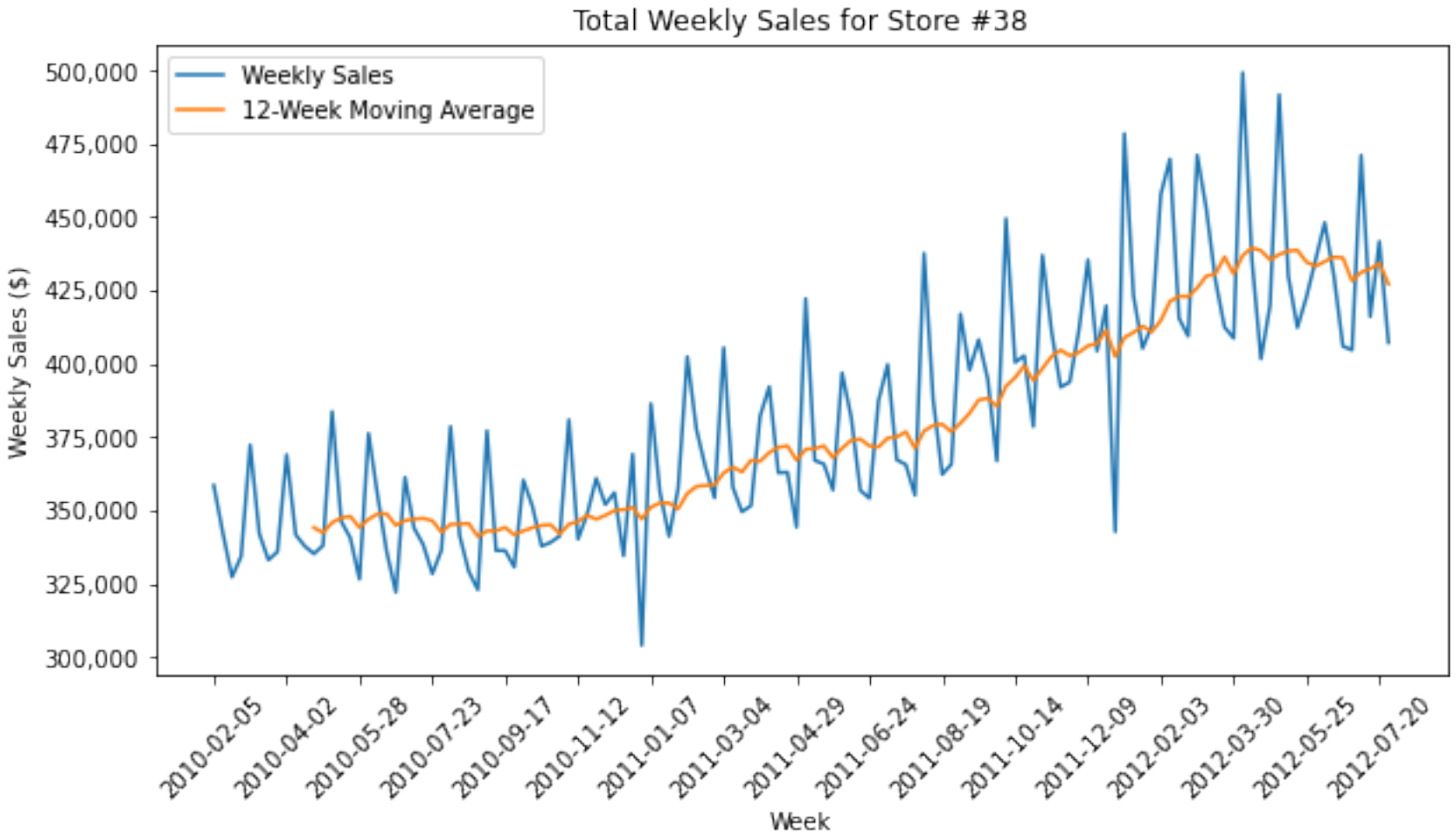
WEEKLY SALES OF INDIVIDUAL DEPARTMENTS VARY WHERE SOME HAVE A SEASONAL PATTERN



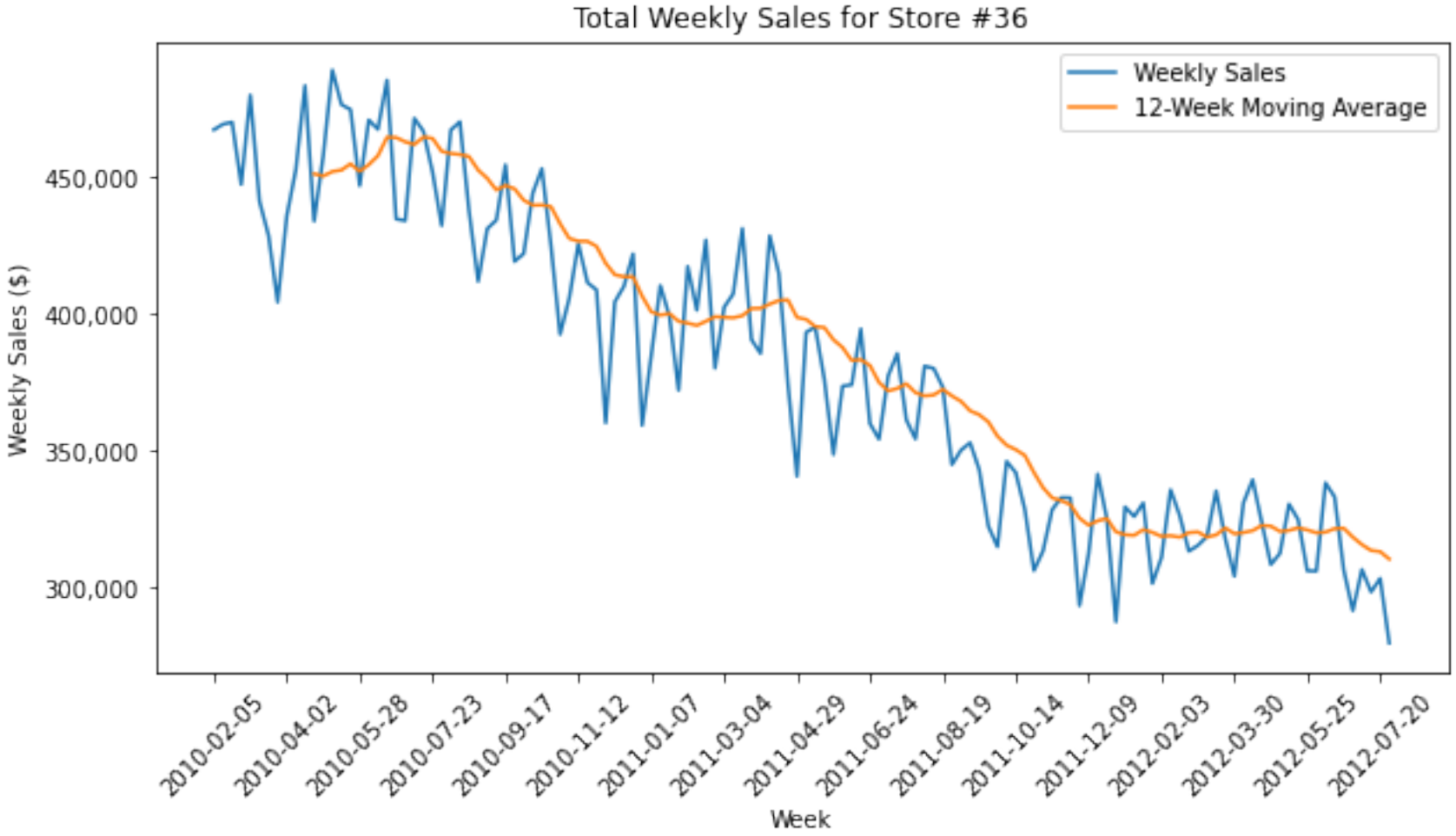
WEEKLY SALES OF INDIVIDUAL DEPARTMENTS VARY WHERE SOME HAVE PEAKS



WEEKLY SALES OF INDIVIDUAL DEPARTMENTS VARY WHERE SOME HAVE AN INCREASING TREND



WEEKLY SALES OF INDIVIDUAL DEPARTMENTS VARY WHERE SOME HAVE A DECREASING TREND



MODELING REQUIRES DATA PREPROCESSING SUCH AS HANDLING MISSING DATA

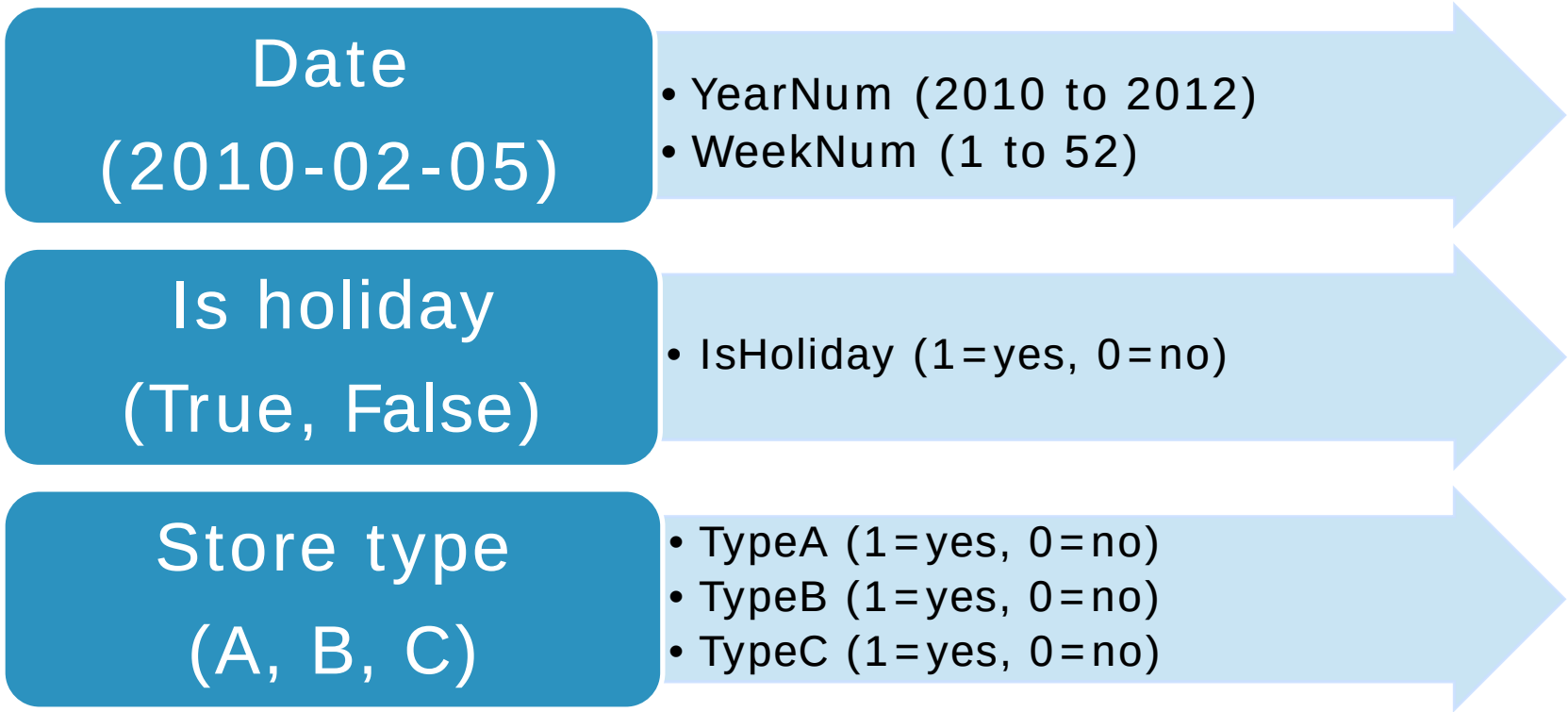
	Store	Dept	Date	Weekly_Sales	IsHoliday
140592	16	58	2010-02-05	346.0	False
140593	16	58	2010-02-12	170.0	True
140594	16	58	2010-05-21	538.0	False
140595	16	58	2010-08-06	340.0	False
140596	16	58	2010-09-03	413.0	False

Only 16 out of 130 weeks of sales available for store #16 department #58

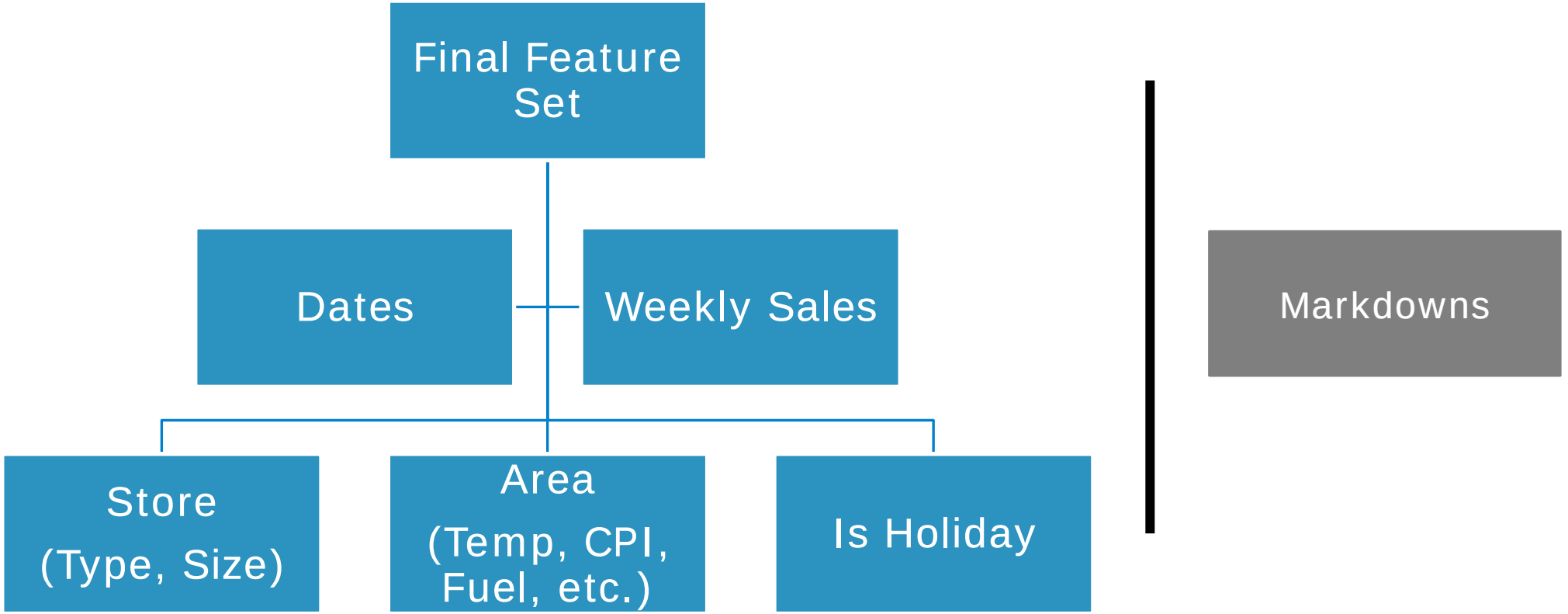
Markdowns are only available Starting November 2011

	Store	Date	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	IsHoliday
1	1	2010-02-05	42.31	2.572	NaN	NaN	NaN	NaN	NaN	211.096358	8.106	False
2	1	2010-02-12	38.51	2.548	NaN	NaN	NaN	NaN	NaN	211.242170	8.106	True
3	1	2010-02-19	39.93	2.514	NaN	NaN	NaN	NaN	NaN	211.289143	8.106	False
4	1	2010-02-26	46.63	2.561	NaN	NaN	NaN	NaN	NaN	211.319643	8.106	False
5	1	2010-03-05	46.50	2.625	NaN	NaN	NaN	NaN	NaN	211.350143	8.106	False

MODELING REQUIRES DATA PREPROCESSING SUCH AS CONVERTING NON-NUMERICAL VALUES TO NUMERICAL



MODELING REQUIRES DATA PREPROCESSING SUCH AS FEATURE SELECTION



MODELING REQUIRES TRAINING AND TESTING THE MODEL

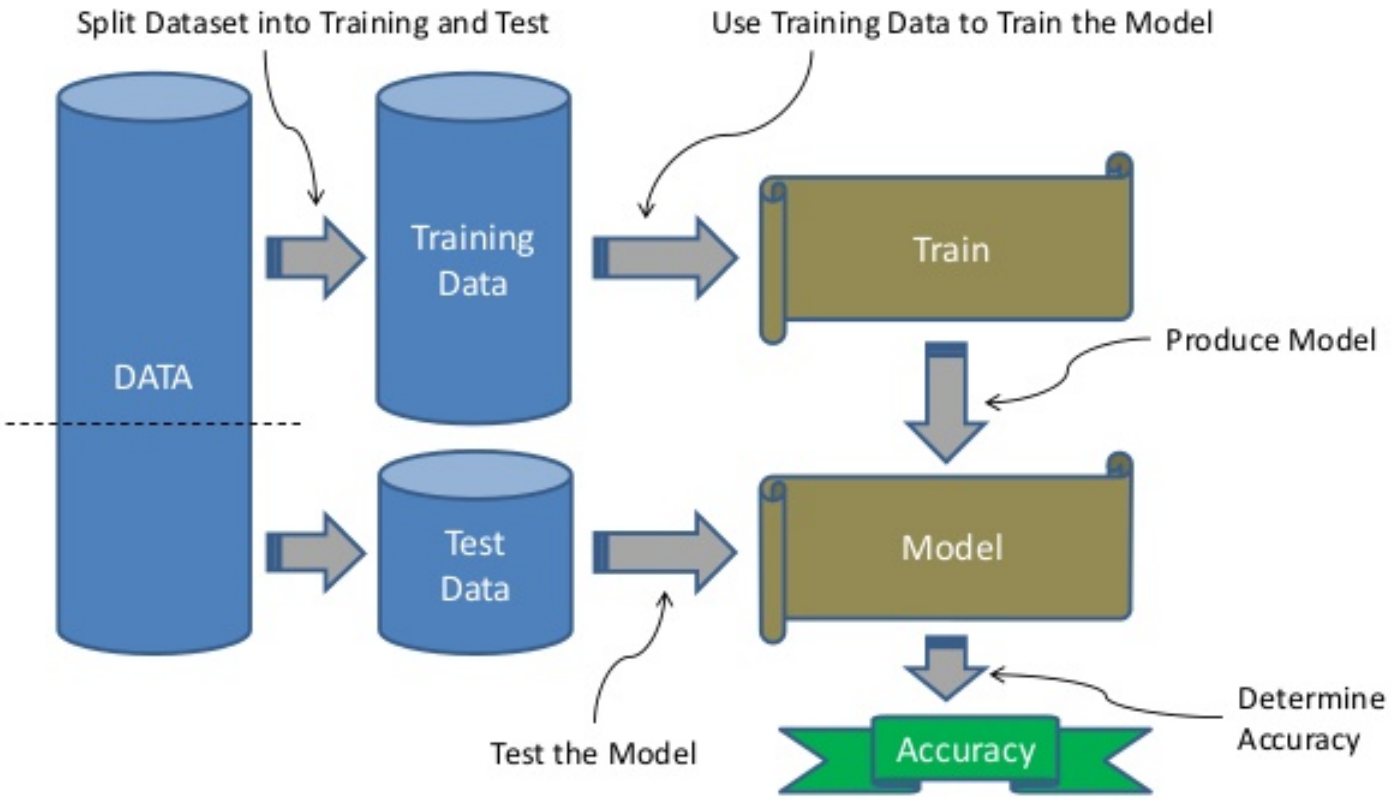
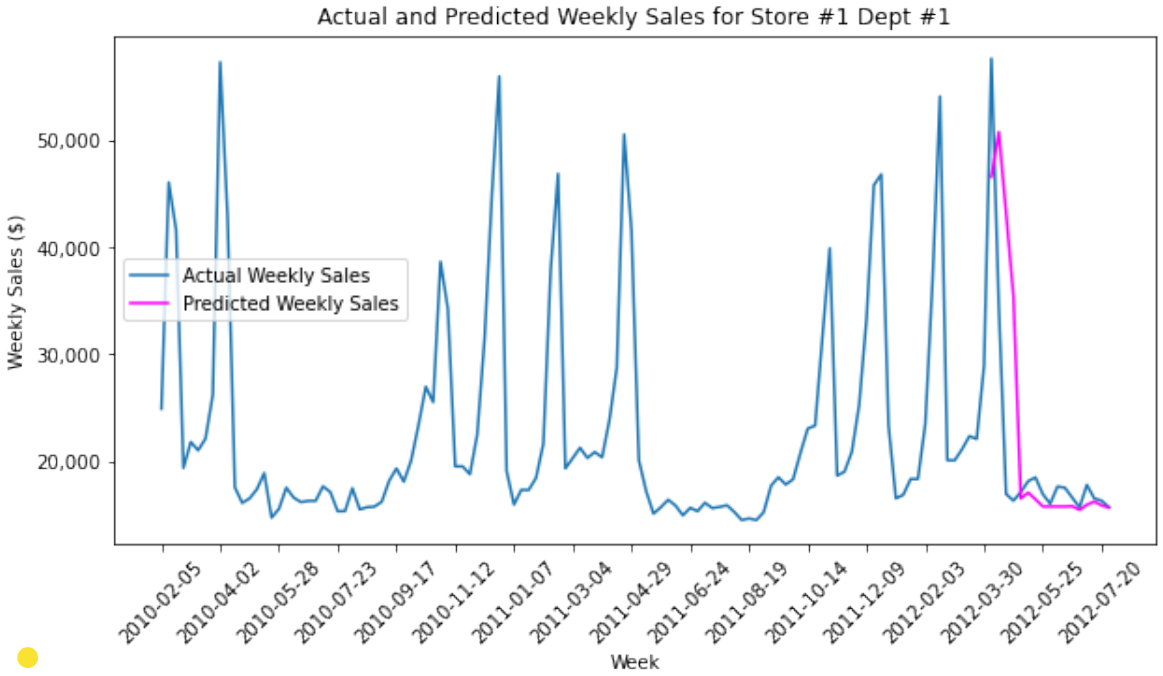
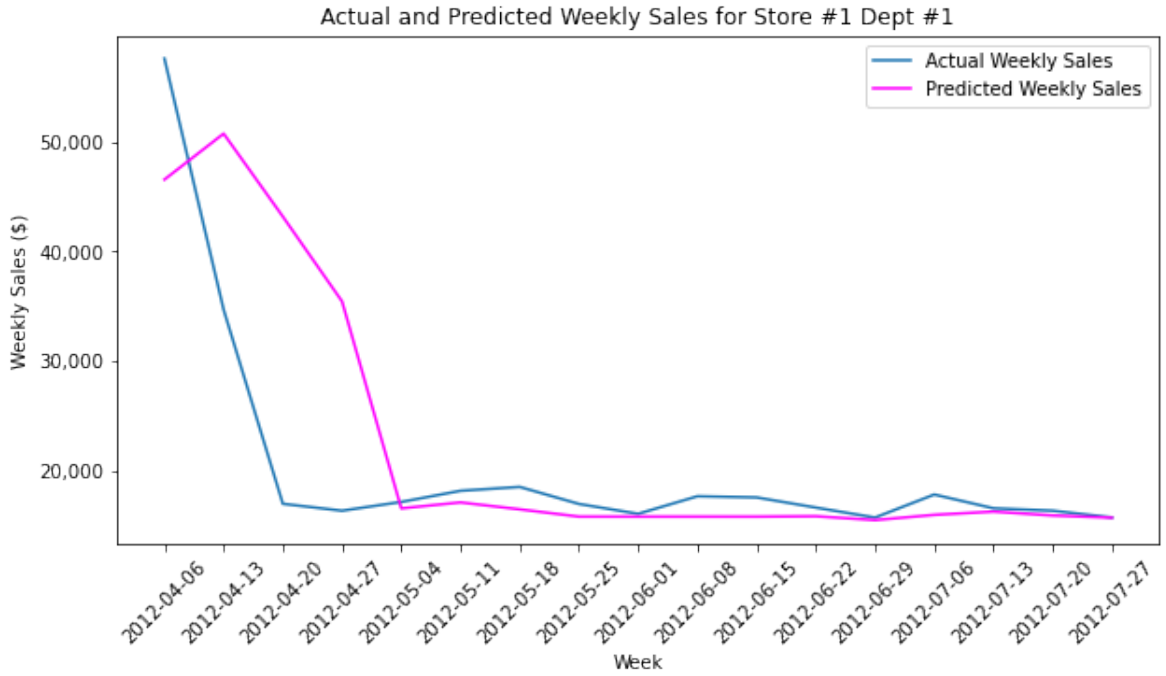


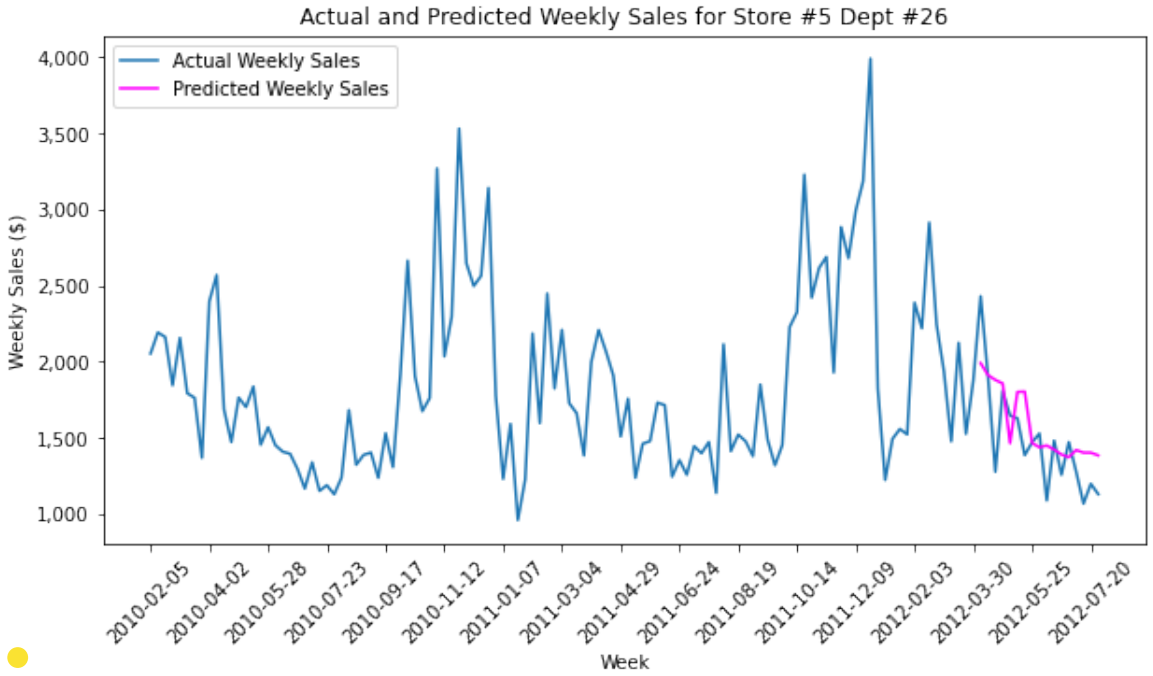
Fig. 1. From *It's About Training* by Ferlitsch, A (2017, September 14). Retrieved from https://www.slideshare.net/AndrewFerlitsch/machine-learning-splitting-datasets

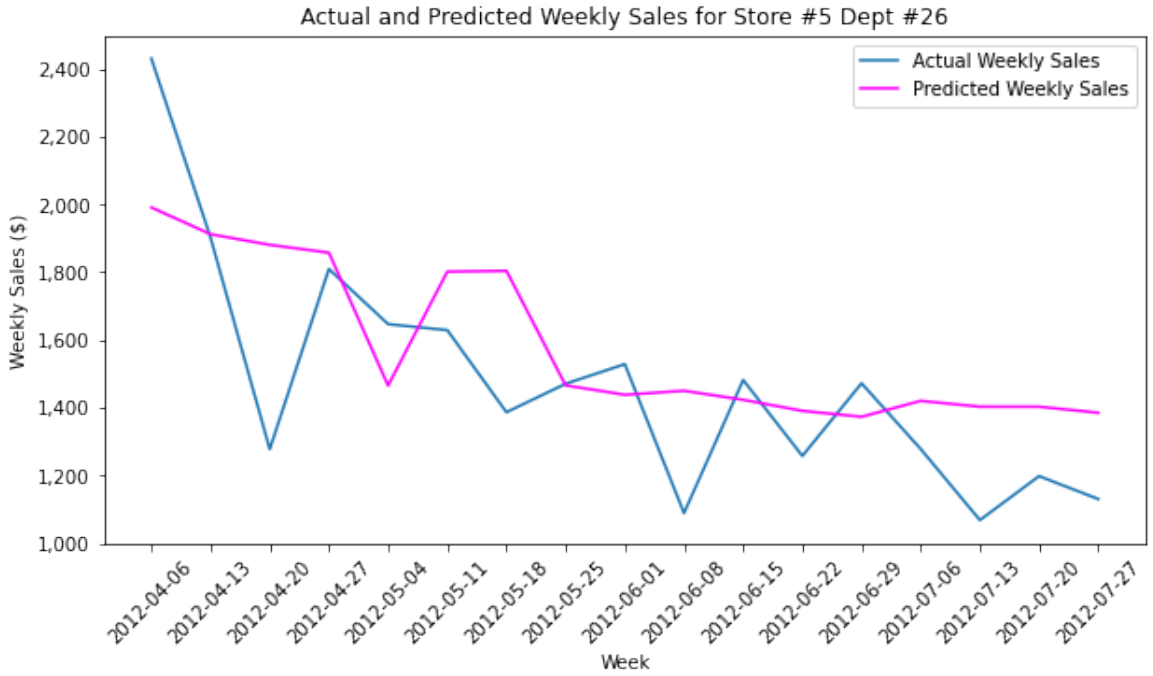
WE USED A RANDOM FOREST REGRESSION MODEL TO FORECAST SALES FOR STORE #1 DEPARTMENT #1 (LARGER SALES)



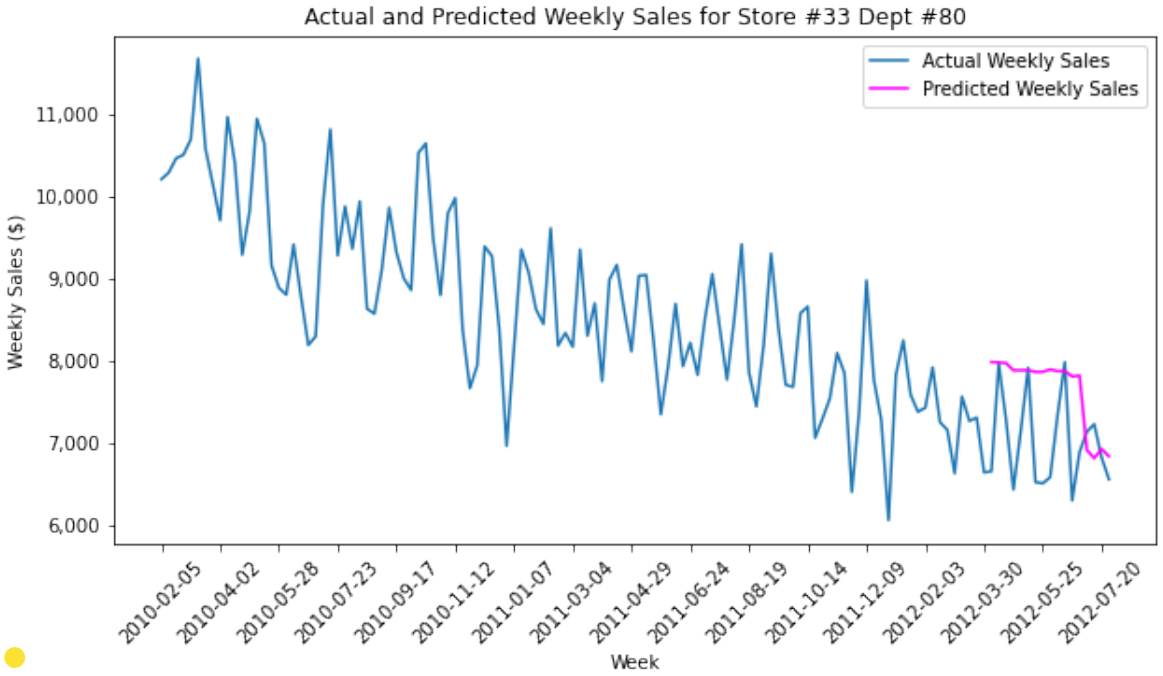


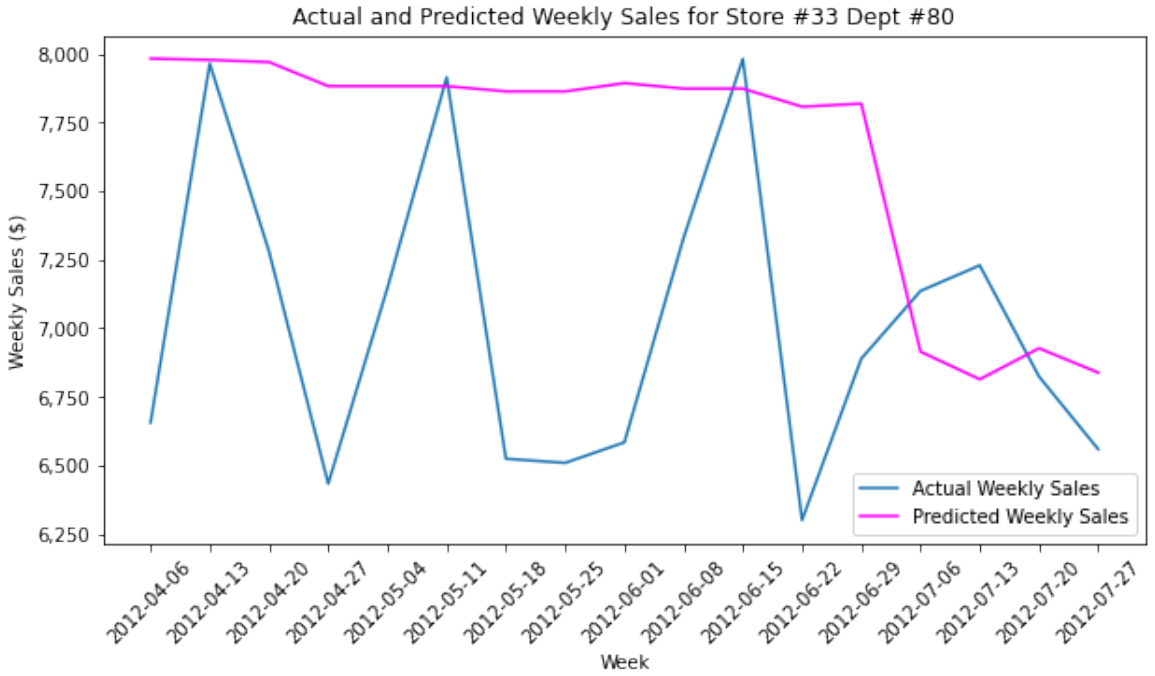
WE USED A RANDOM FOREST REGRESSION MODEL TO FORECAST SALES FOR STORE #5 DEPARTMENT #26 (SMALLER SALES)



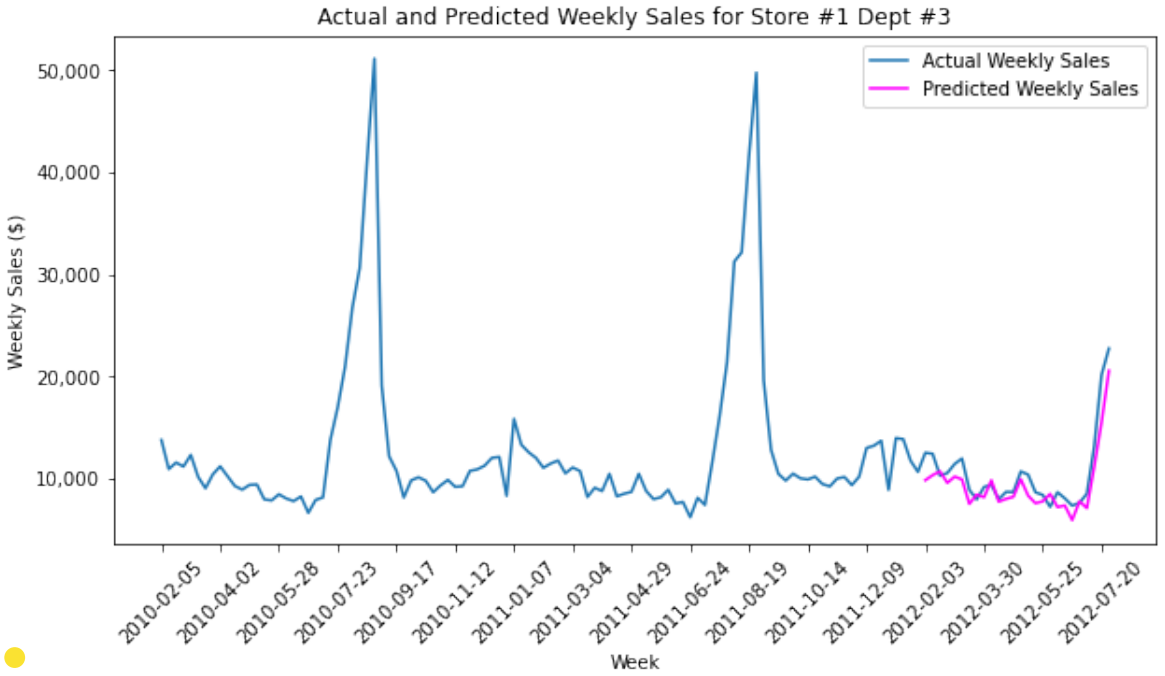


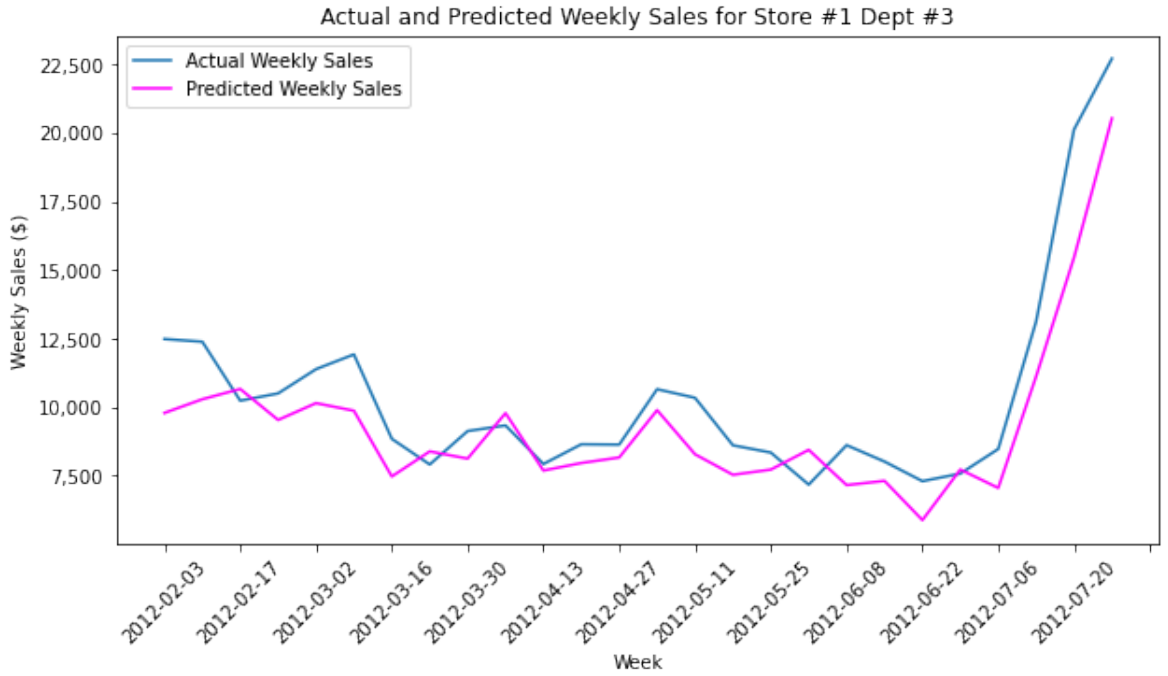
WE USED A RANDOM FOREST REGRESSION MODEL TO FORECAST SALES FOR STORE #33 DEPARTMENT #80 (SMALLER SALES)





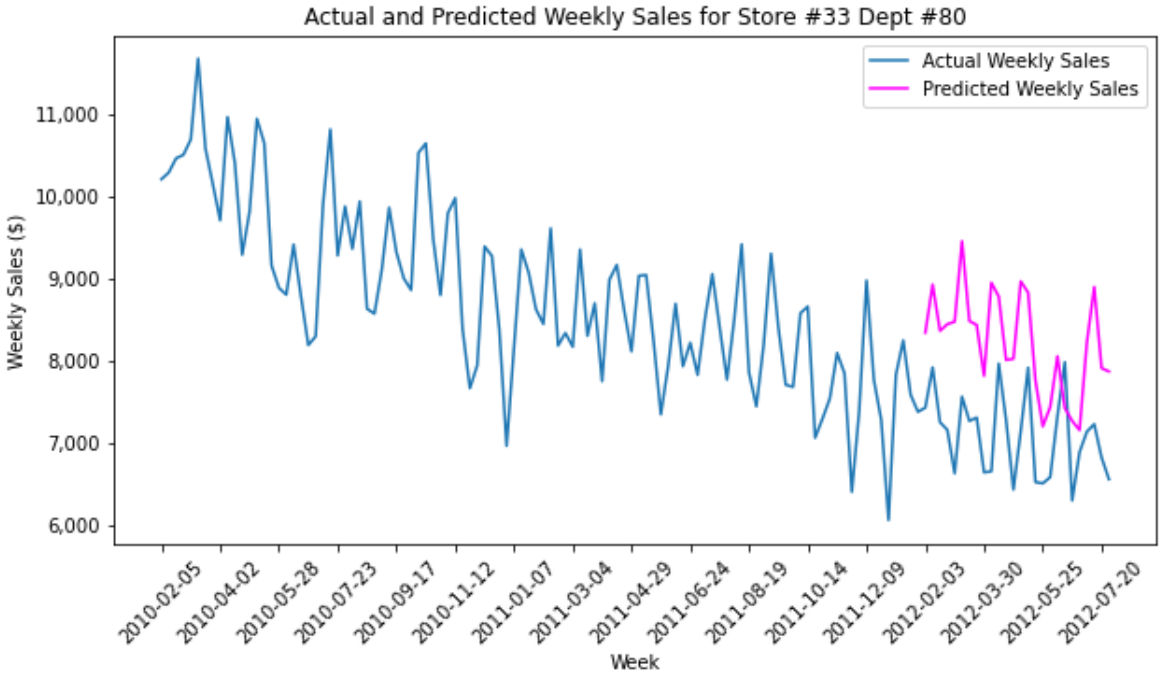
WE USED A SARIMA* MODEL TO FORECAST SALES FOR STORE #1 DEPARTMENT #3 (SMALLER SALES)

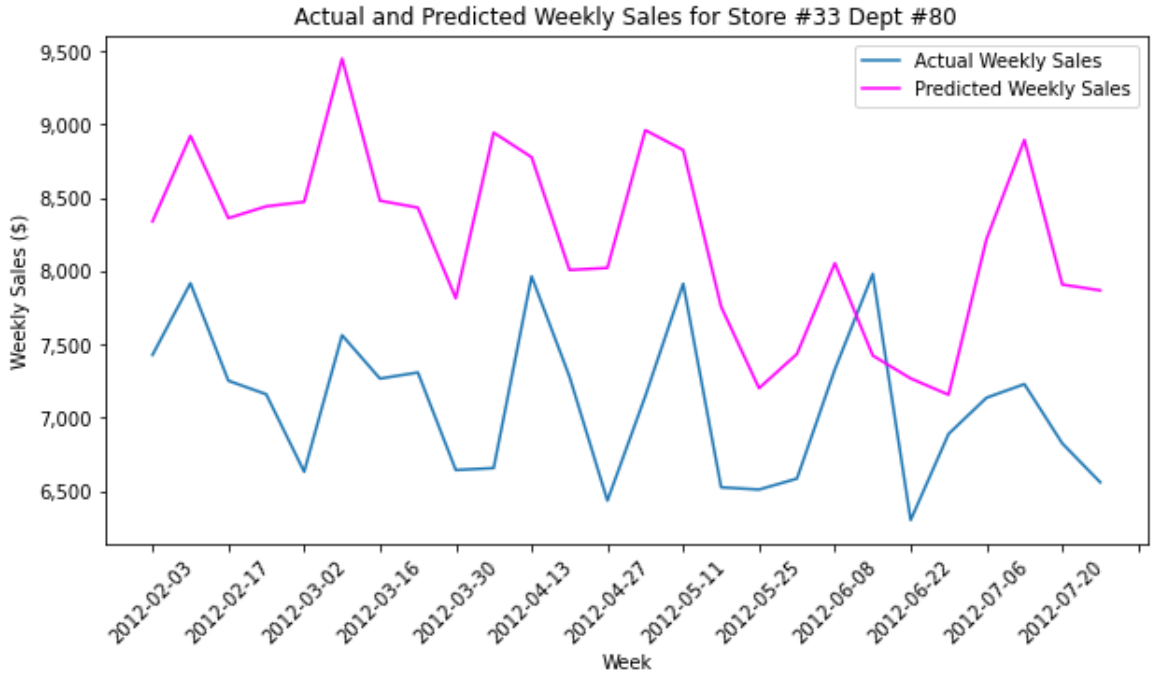




*SARIMA = Seasonal Autoregressive Integrated Moving Average

WE USED A SARIMA* MODEL TO FORECAST SALES FOR STORE #1 DEPARTMENT #3 (SMALLER SALES)





*SARIMA = Seasonal Autoregressive Integrated Moving Average

BOTH RANDOM FOREST REGRESSION AND SARIMA MODELS MAY BE IMPROVED

Random Forest Regression

- Create features for the number of weeks until each major holiday
- Create features for prior sales (1 week prior, 4 weeks prior)

SARIMA

- Remove the trend for departments where sales trend up or down

Both

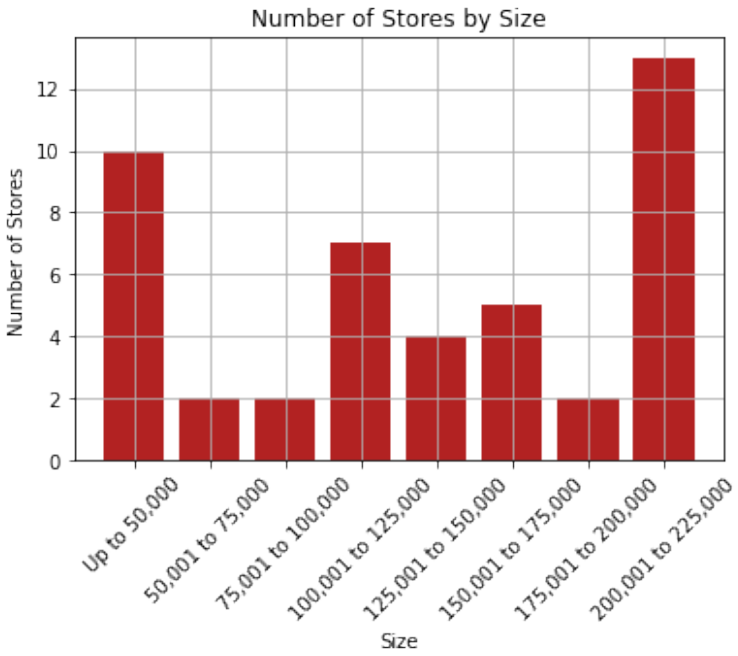
- Rerun the model with additional data when available

WE DESCRIBED THE DATA, OUR EXPLORATORY ANALYSIS, AND OUR MODELING AND FORECASTING OF WEEKLY SALES

The Data

	Store	Dept	Date	Weekly_Sales	IsHoliday
1	1	1	2010-02-05	24924.50	False
2	1	1	2010-02-12	46039.49	True
3	1	1	2010-02-19	41595.55	False
4	1	1	2010-02-26	19403.54	False
5	1	1	2010-03-05	21827.90	False

Exploration



Modeling and Forecasting