Retail Store Performance Analytics

-- Data Solution to Reveal Deeper Relationships

between 45 Store Sales and Holiday Mark Down

Lin (Lily) Han

August 29th, 2020

Agenda

I. Project Background

II. Performance Variables Analysis

III. Data manipulation

IV. Correlation Analysis & Business Insights

V. Further Consideration

VI. Q & A

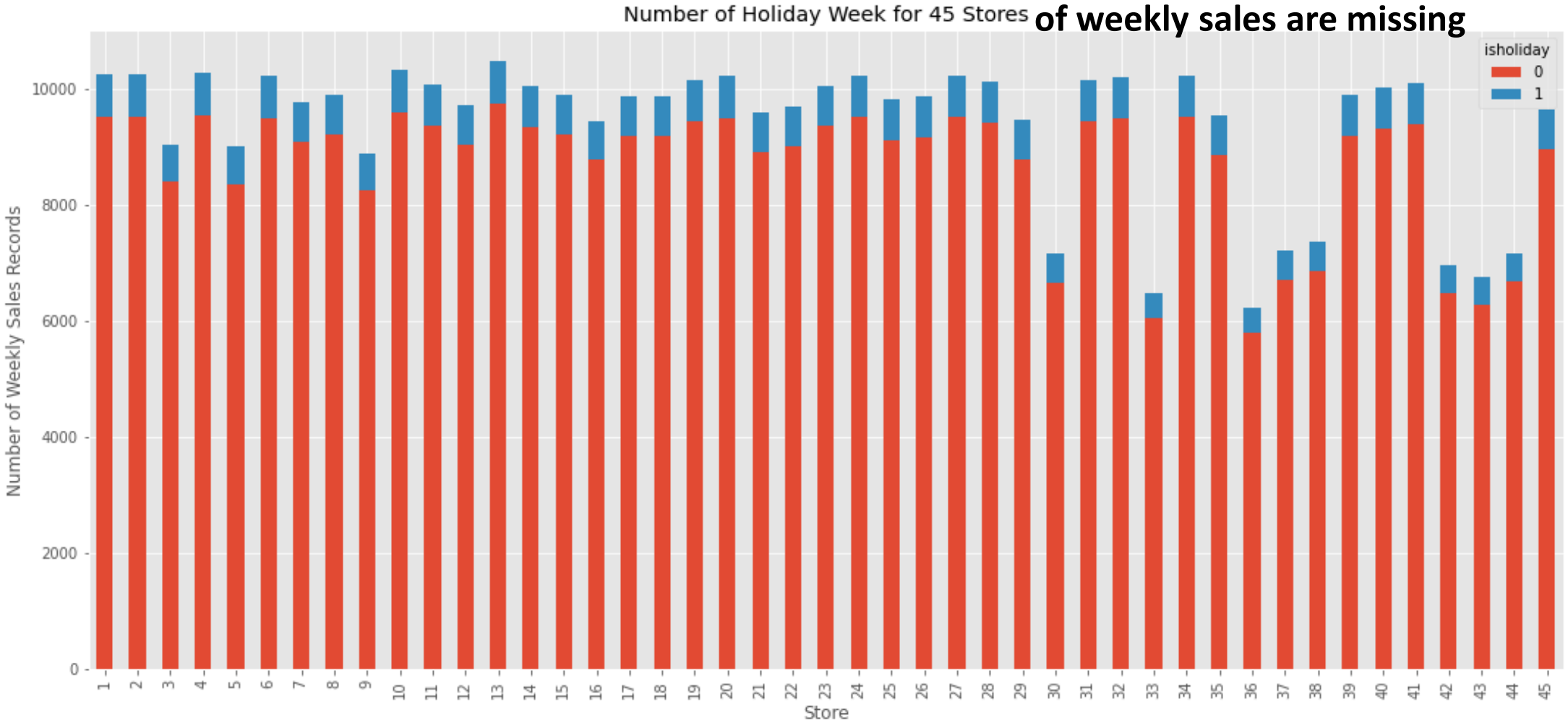
Project Background

- The dataset I chose is “Retail Data Analytics-Historical sales data from 45 stores” downloaded from Kaggle website (Link).

- The dataset contains 3 Tables – Stores, Features and Sales.
  I. “Stores” table contains the 45 stores’ type and sizes.
  II. “Features” table contains additional data related to the store, department, and regional activity for the given dates, having 12 variables and 8,190 observations.
  III. “Sales” dataset contains historical sales data, which covers to 2010-02-05 to 2012-11-01, having 5 columns and 421,570 records.

Store Performance Variables Analysis

Some store performance recodes of weekly sales are missing



Data Manipulation

1. Join 3 tables using SQLite 3, including 16 variables and 421,570 records.
2. Remove irrelevant variables that not significant for holiday week or not:

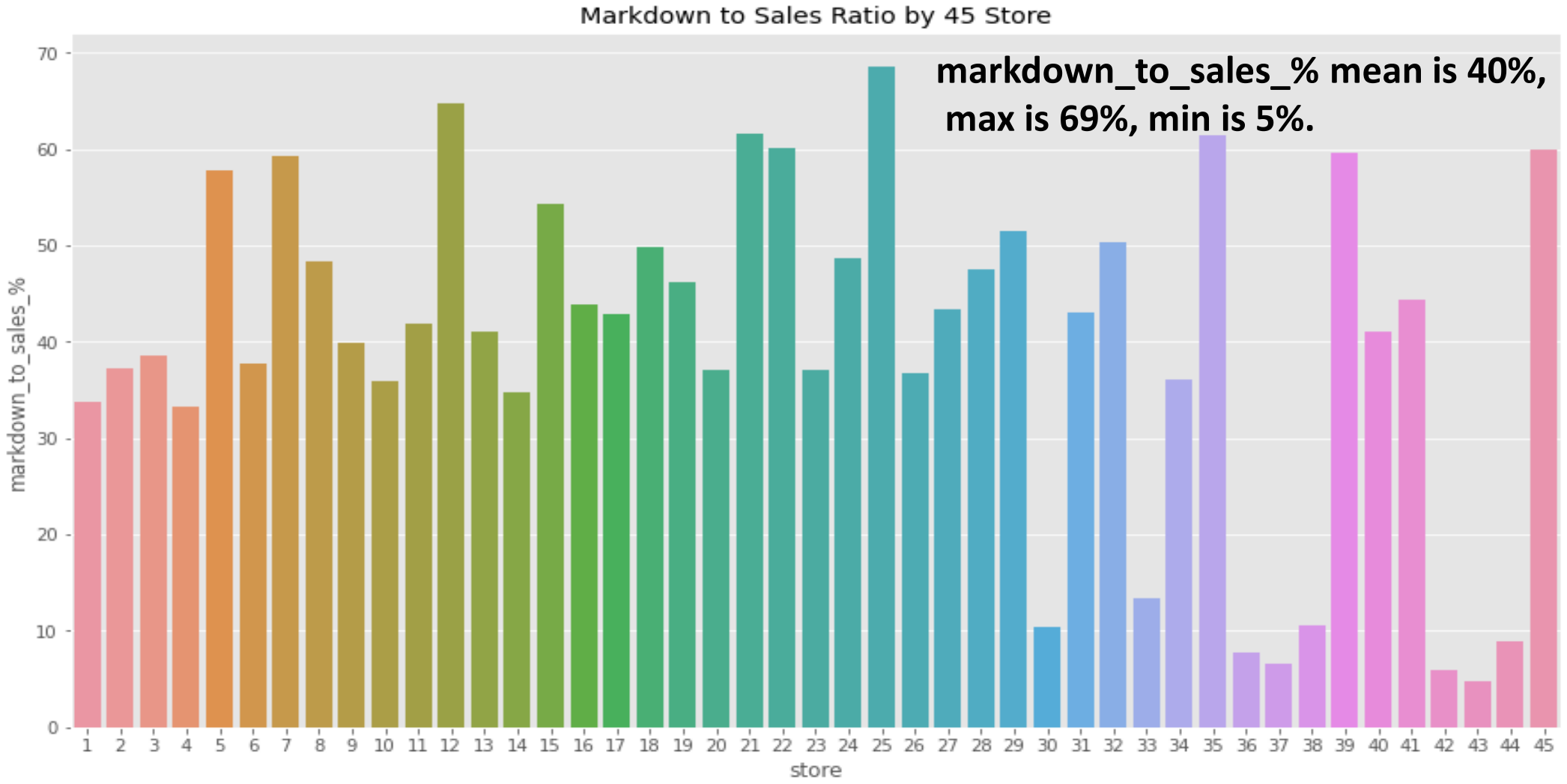
   'temperature’, 'fuel_price’, 'cpi’, 'unemployment’, 'size’, ‘department’, ‘type’, ect.

3. Group 45 stores’ sales, number of holiday weeks and other variables.
4. Add 2 new variables as store performance measurement:

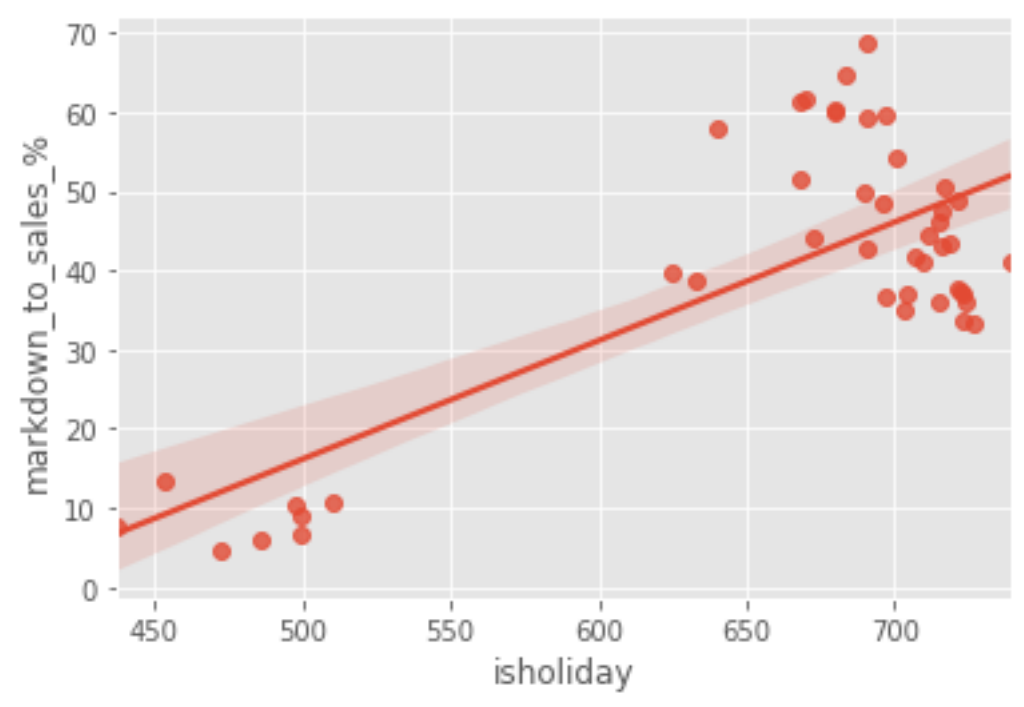
   1) ‘total_markdown’ = sum of 5 mark down

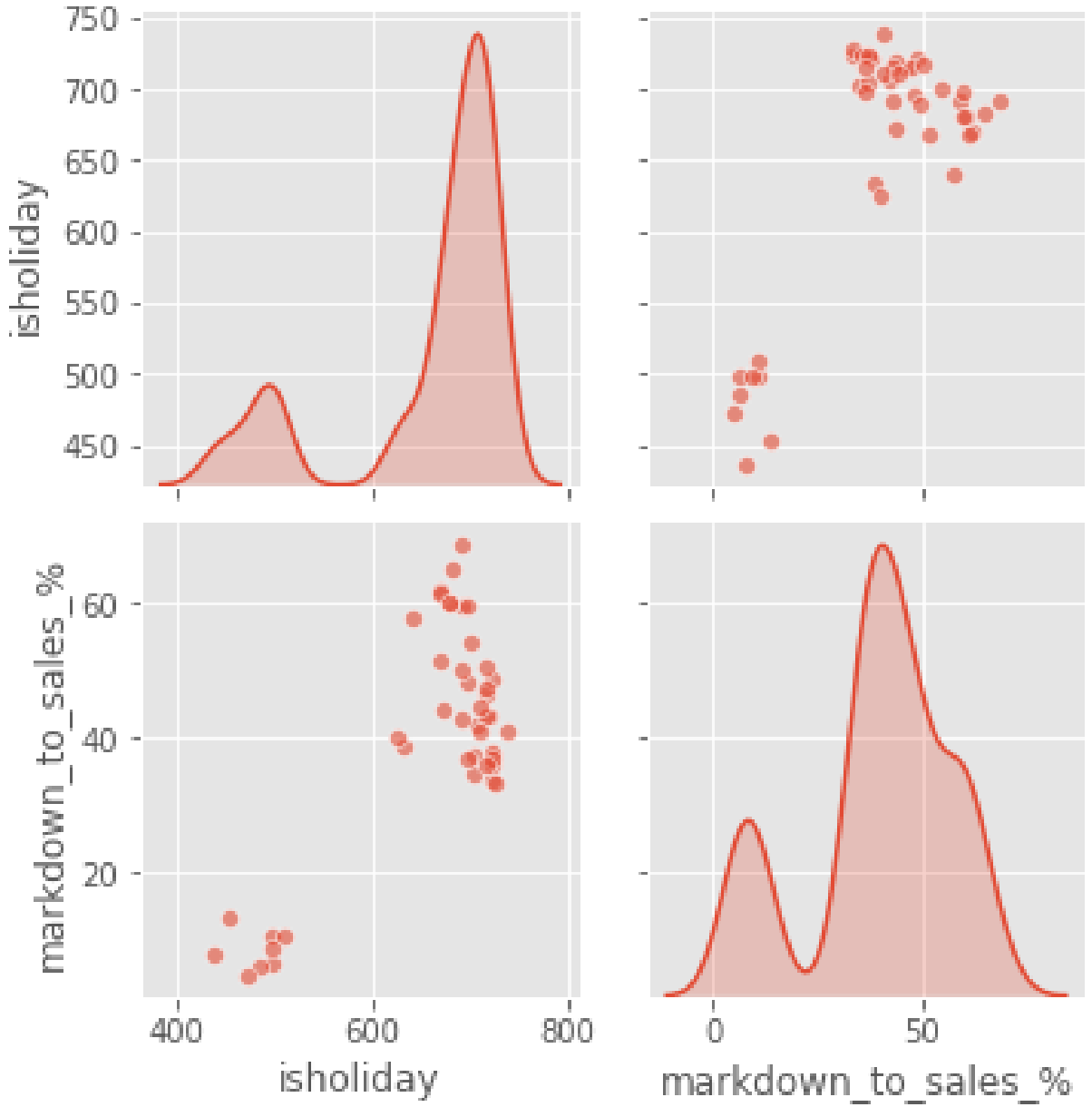
   2) 'markdown_to_sales_%’ = total markdown / total weekly sales

Mark Down to Sales Ratio 45 Stores Comparison



Correlation of Holiday Weeks Number & Mark Down to Sales Ratio





Pearson Correlation Test & Business Insights

- The 2 continuous variables ‘Holiday Week Numbers’ and ‘Mark Down to Sales Ratio’ are positively correlated (0.75). The test is significant based on p-value (2.33e-9).
- The correlation reveals the higher holiday week number, the higher Mark Down to Sales Ratio that means the profit ratio is lower.
- The retail company should pay more attention to the frequency of holiday mark down.

```
87] # Pearson's correlation coefficient is 0.75, Two-tailed p-value
    stats.pearsonr(arr_isholiday, arr_markdown_to_sales)

    (0.7534809268657268, 2.3291321242870816e-09)
```

Further Consideration

- Find more correlations of store performance and input variables
- Store historical performance comparison
- Sales prediction using Machine Learning models

Thank you for your listening!
Q & A

Reference

1. Dataset downloaded from Kaggle:
https://www.kaggle.com/manjeetsingh/retaildataset?select=stores+data-set.csv

2. Group variables by if the week has holiday or not:

```
[33] df.groupby('isholiday').mean()
```

isholiday	store	weekly_sales	temperature	fuel_price	markdown1	markdown2	markdown3	markdown4	markdown5	cpi	unemployment
0	22.202473	15901.445069	60.881518	3.370901	7332.984933	2298.528959	195.831724	3384.219730	4722.908034	171.222884	7.954927
1	22.175078	17035.823187	49.632550	3.230557	6240.733956	13048.763547	15077.548736	3371.283148	3536.647067	170.925303	8.031131