WALMART RECRUITING - STORE SALES ANALYSIS & FORECASTING

- HAOTIAN YU

- FINAL PROJECT
- EMES 6992
- GWU


Walmart
Supercentre

BACKGROUND

Data Analysis and Prediction is very important in the Selling industry

Walmart is a very good data resource for the analysis. It's useful for the application in business.

INTRODUCTION

- Analyze what factors can decide the sale of a Walmart store.
- prediction and forecasting of sales in 45 Walmart stores
- The data analysis is done based on the sale data for different departments in different weeks for each location in the past time

~ What can significantly decide the sales for a Walmart store?

~ How will be sale in holidays?

~ What will the sale for each store be like in future?

- What's your guess for these questions?

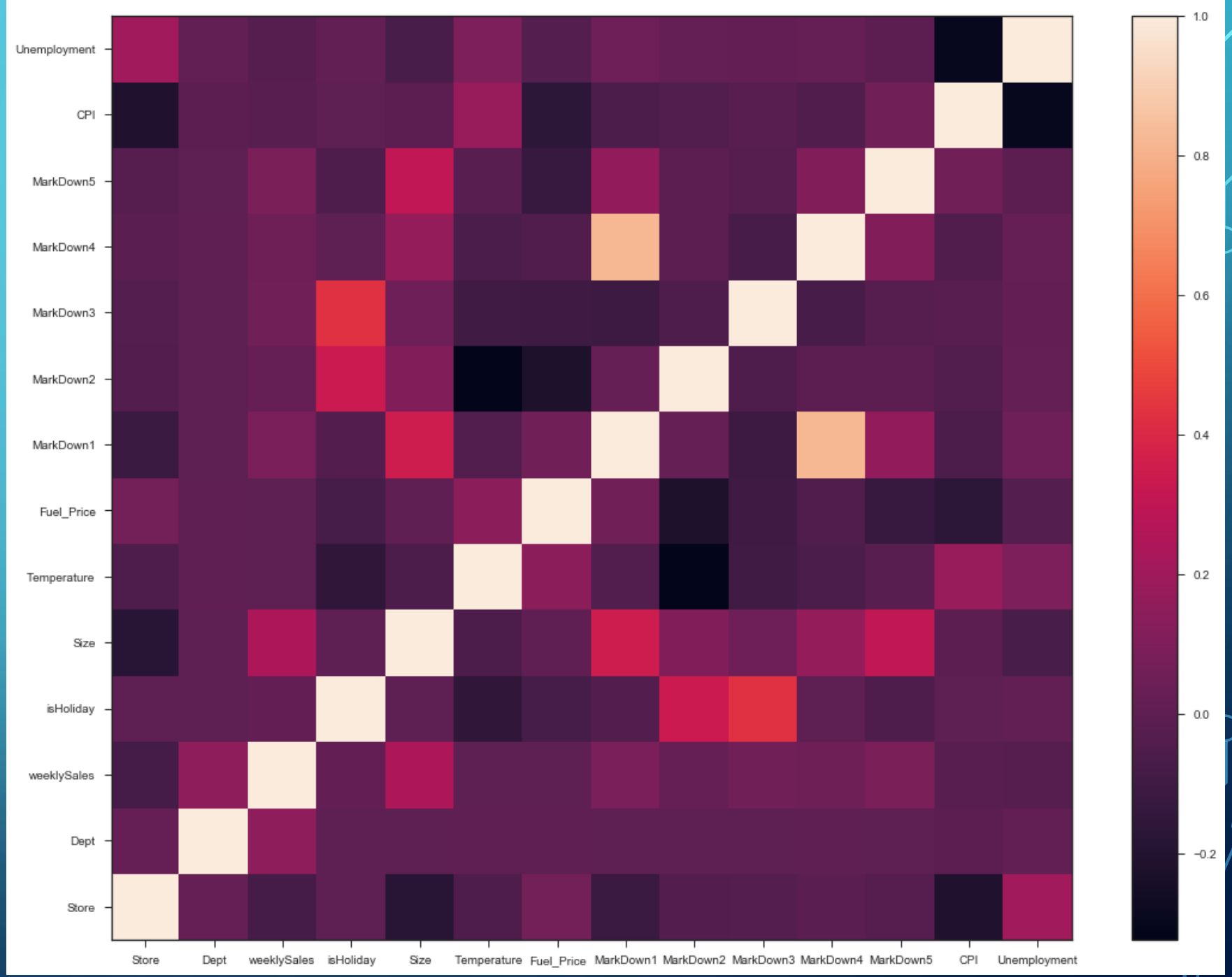
DATA VARIABLES

- historical sales data for 45 Walmart stores in different areas.
- Store - the store number
- Date - the week
- Temperature - average temperature in the region
- Fuel_Price - cost of fuel in the region
- CPI - the consumer price index
- Unemployment - the unemployment rate
- IsHoliday - whether the week is a special holiday week

DATA STRUCTURE

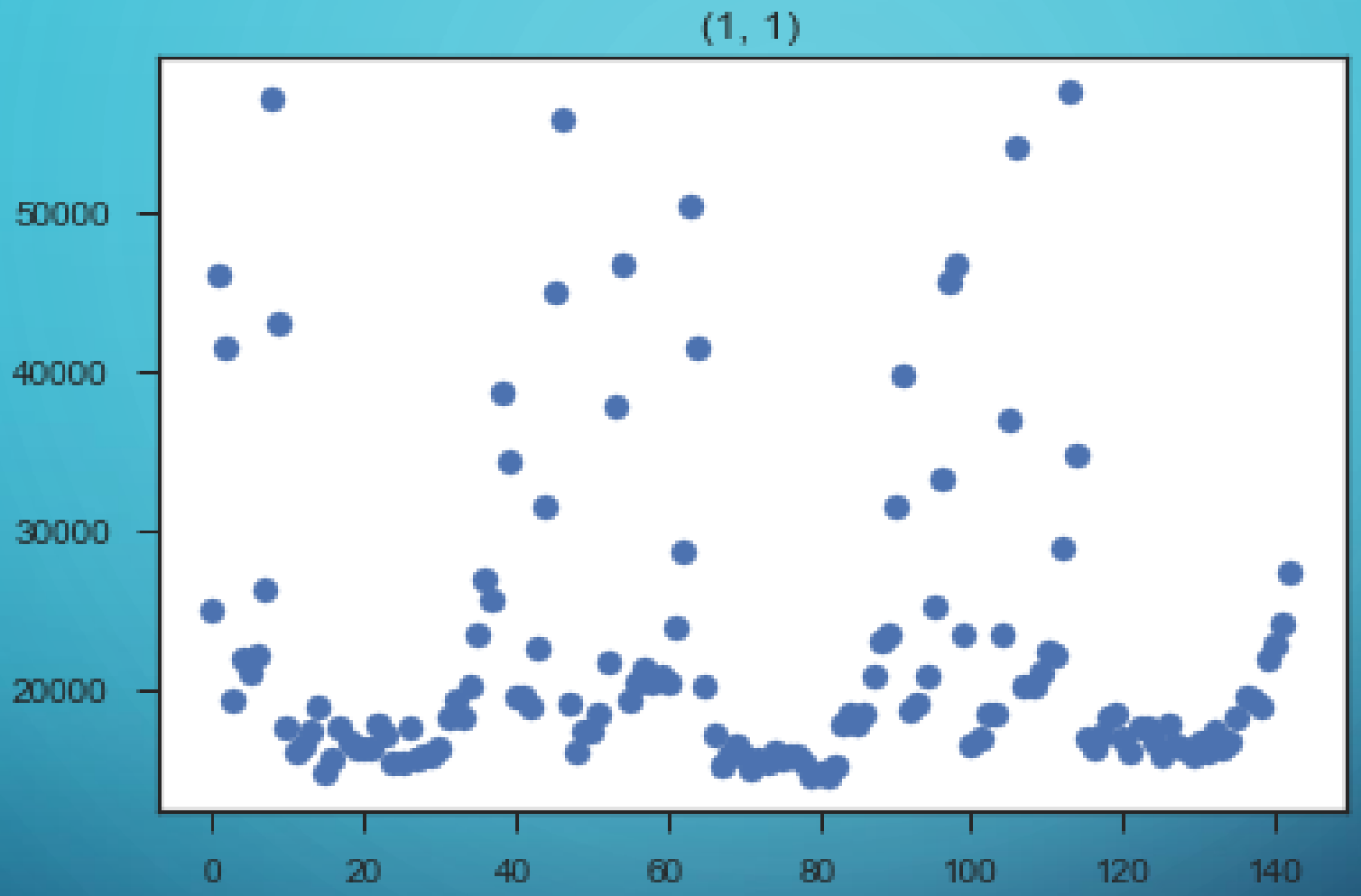
	Store	Dept	Date	Weekly Sales	isHoliday	Type	Size	Temperature	Fuel_ Price	CPI	Unemployment
0	1	1	2010-02-05	24924.50	False	A	151315	42.31	2.572	211.096358	8.106
1	1	1	2010-02-12	46039.49	True	A	151315	38.51	2.548	211.242170	8.106
2	1	1	2010-02-19	41595.55	False	A	151315	39.93	2.514	211.289143	8.106
3	1	1	2010-02-26	19403.54	False	A	151315	46.63	2.561	211.319643	8.106
4	1	1	2010-03-05	21827.90	False	A	151315	46.50	2.625	211.350143	8.106

CORRELATION



PAIR SCATTER PLOTS

- Weekly sale
- Fuel Price
- Size
- CPI
- Department
- Temperature
- Unemployment



WEEKLY SALE IN DEPARTMENTS OF STORES

GROUP BY STORE AND DEPT

DATA ANALYSIS

- Relationship between Weekly sale and

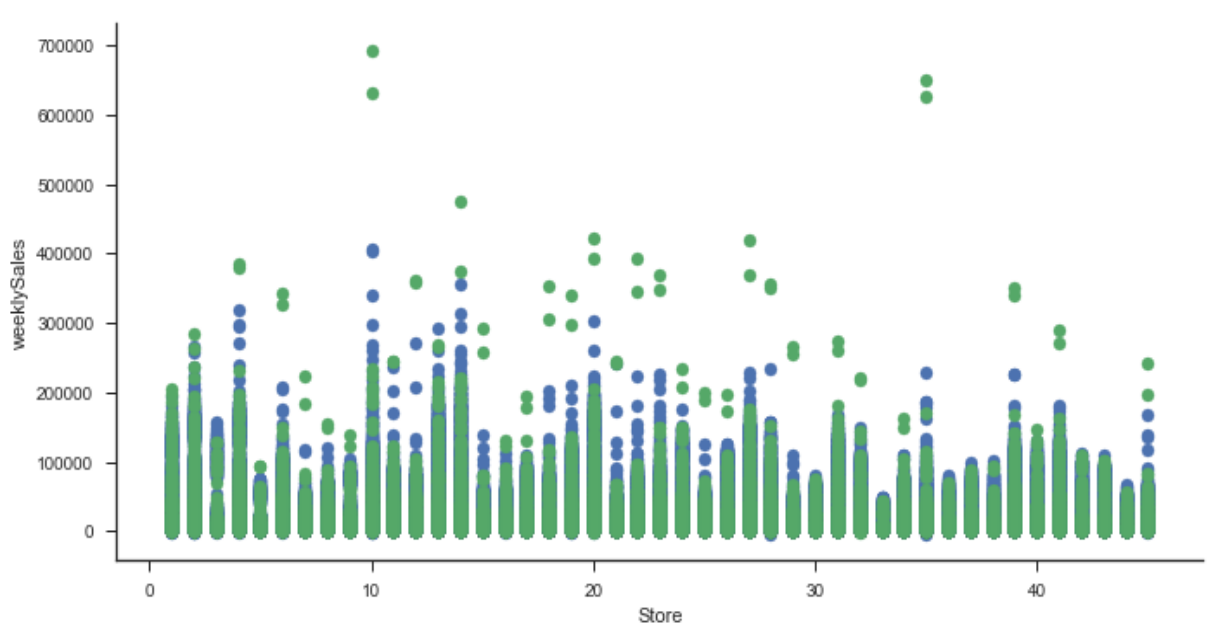
'Store', 'Date', 'Dept', 'Fuel_Price', 'CPI', 'Unemployment'

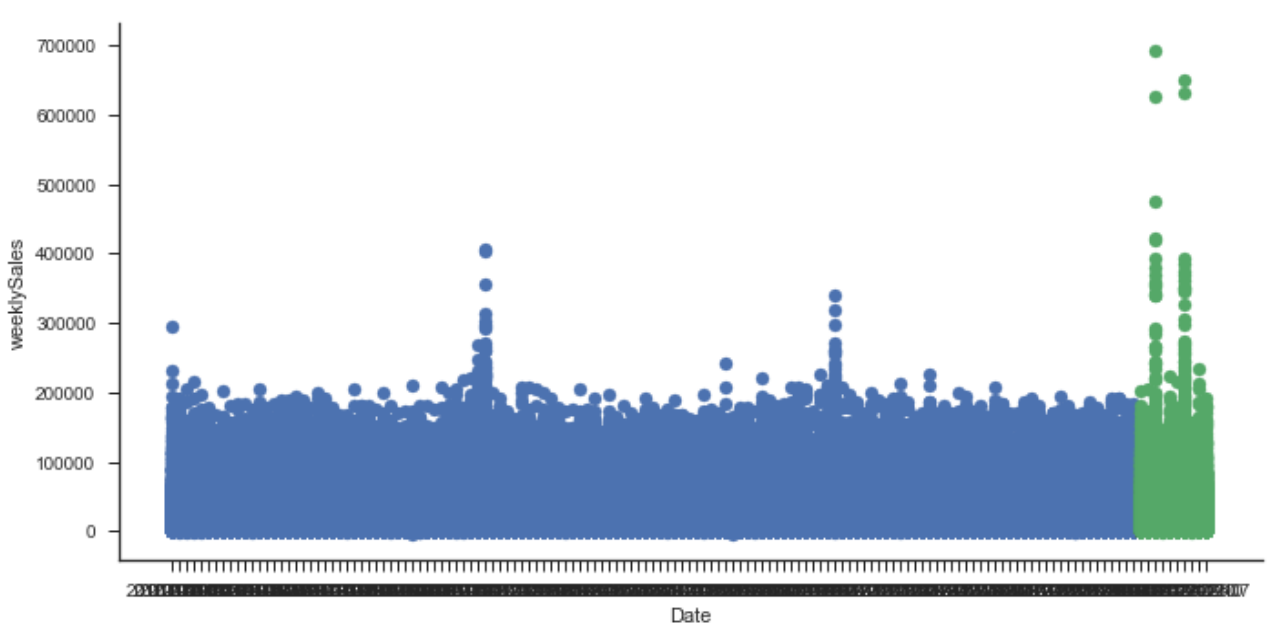
in two different isHoliday Group(True or False).

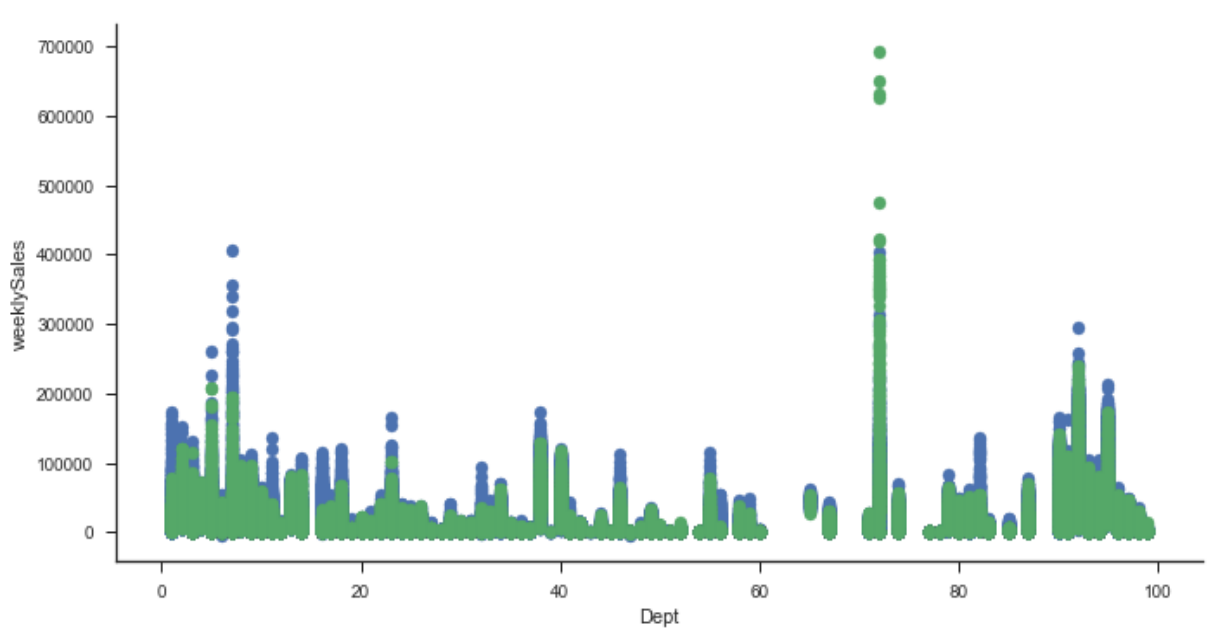
- Relationship between Weekly sale and

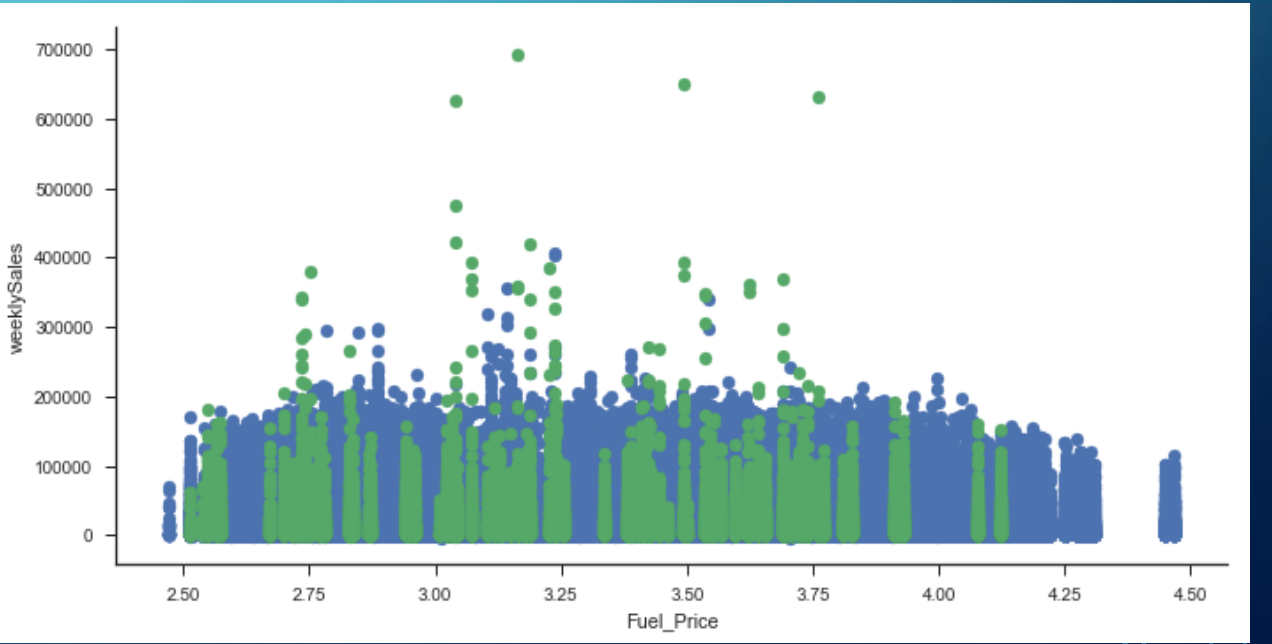
'Store', 'Date', 'Dept', 'Fuel_Price', 'CPI', 'Unemployment'

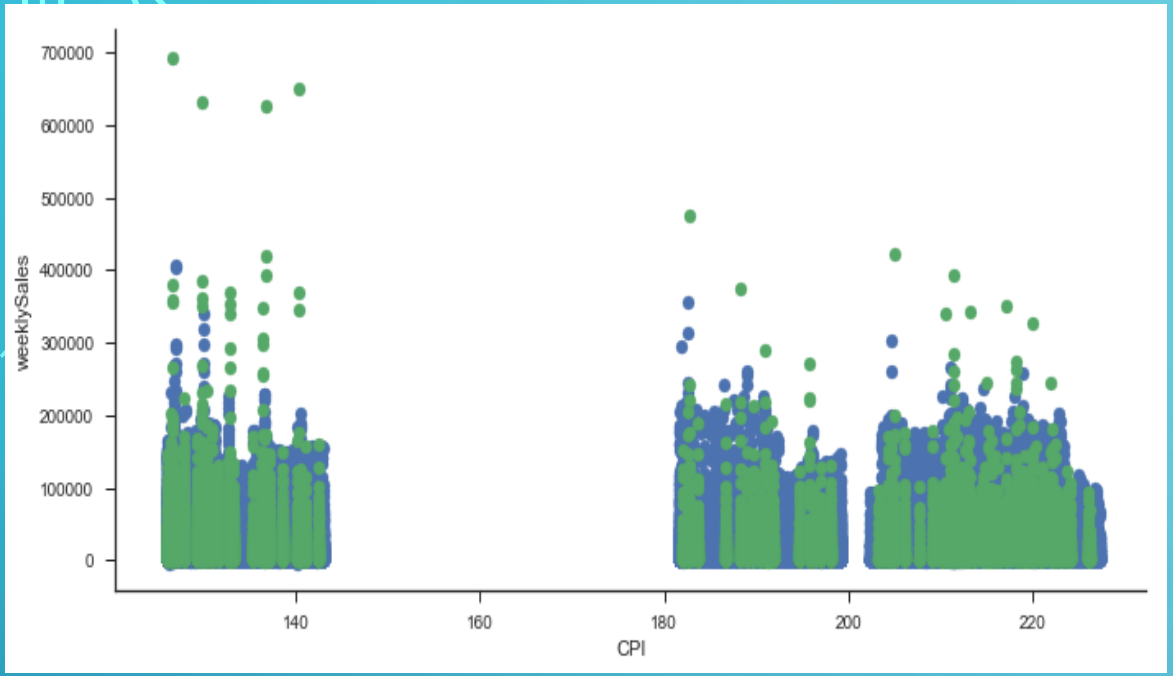
in different Size Group.

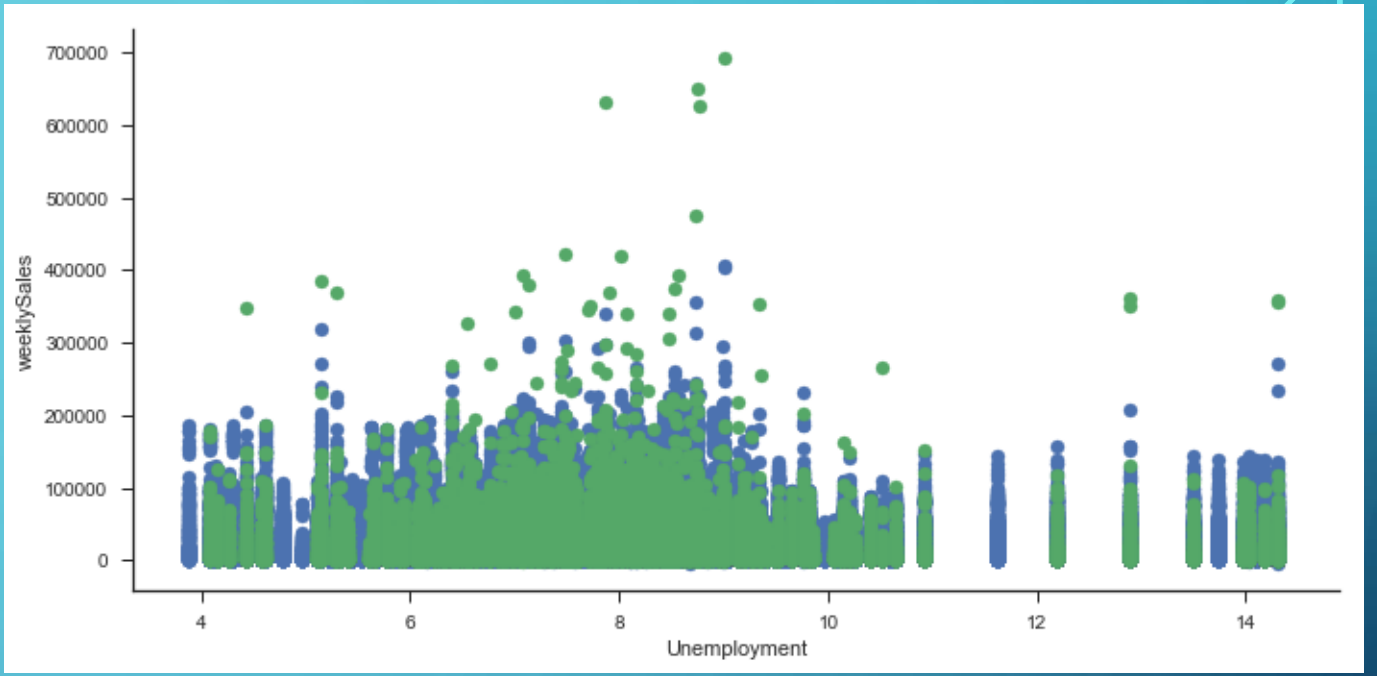

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
0
10
20
30
40
Store

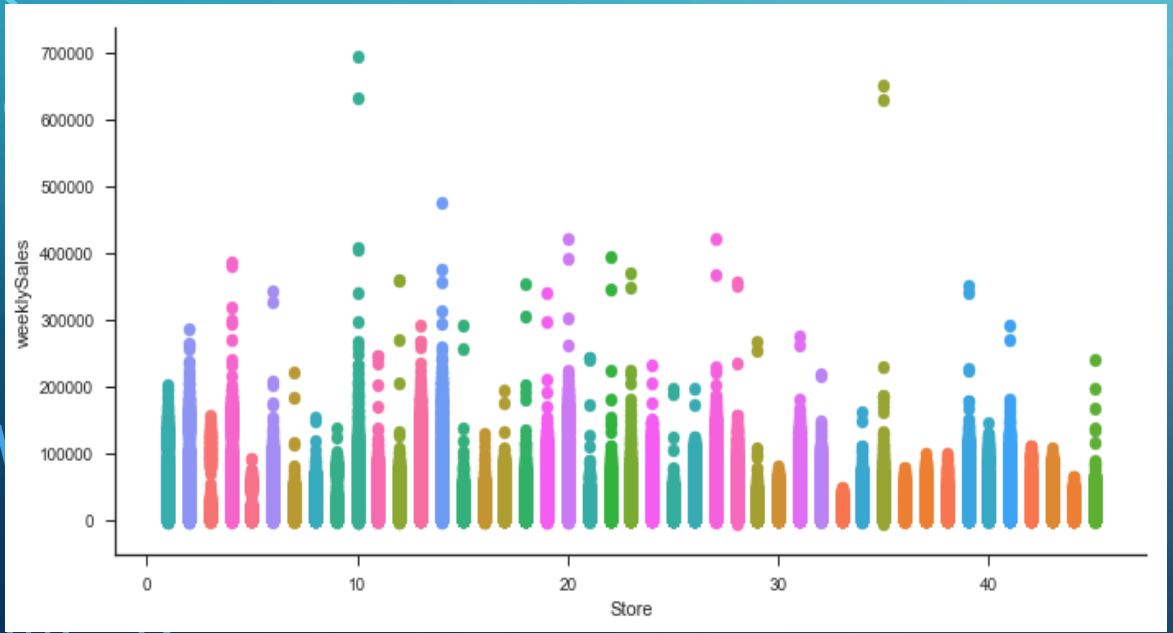

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
Date

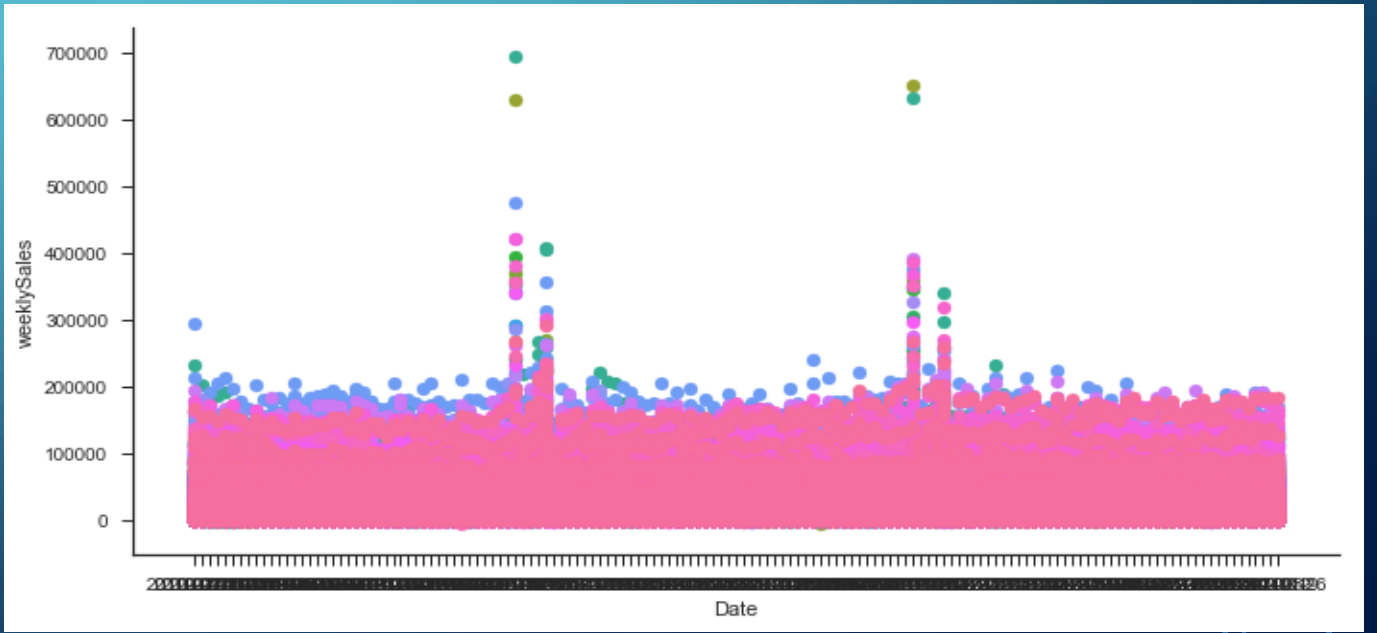

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
0
20
40
60
80
100
Dept

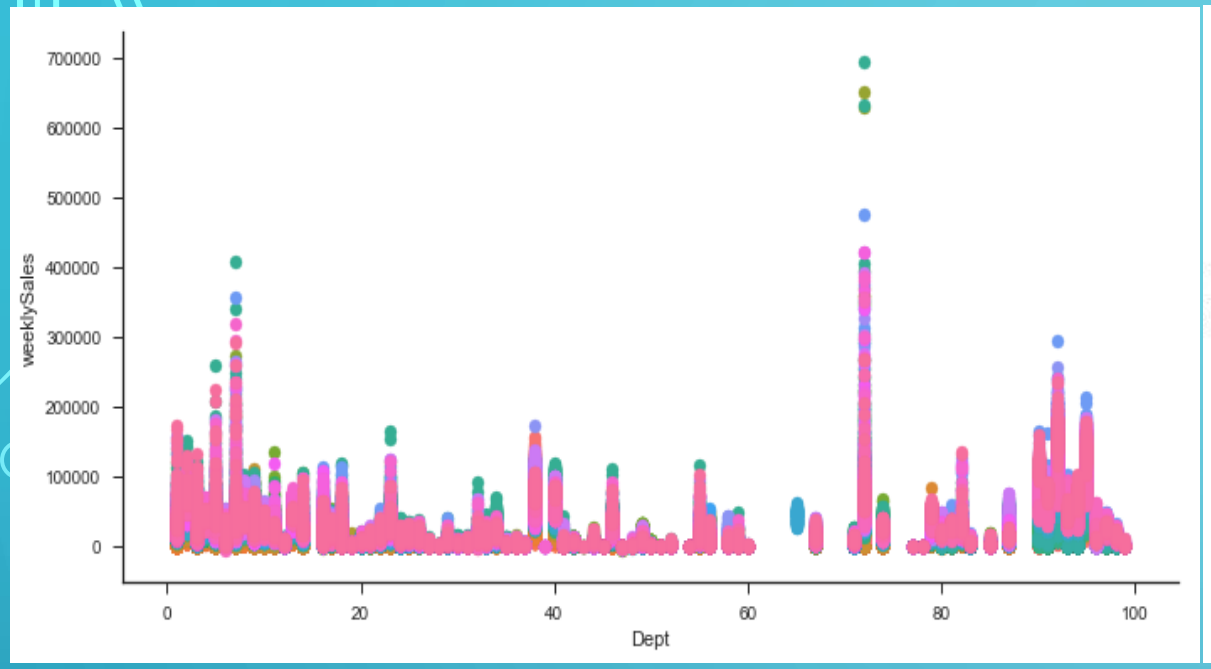

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
2.50
2.75
3.00
3.25
3.50
3.75
4.00
4.25
4.50
Fuel_Price

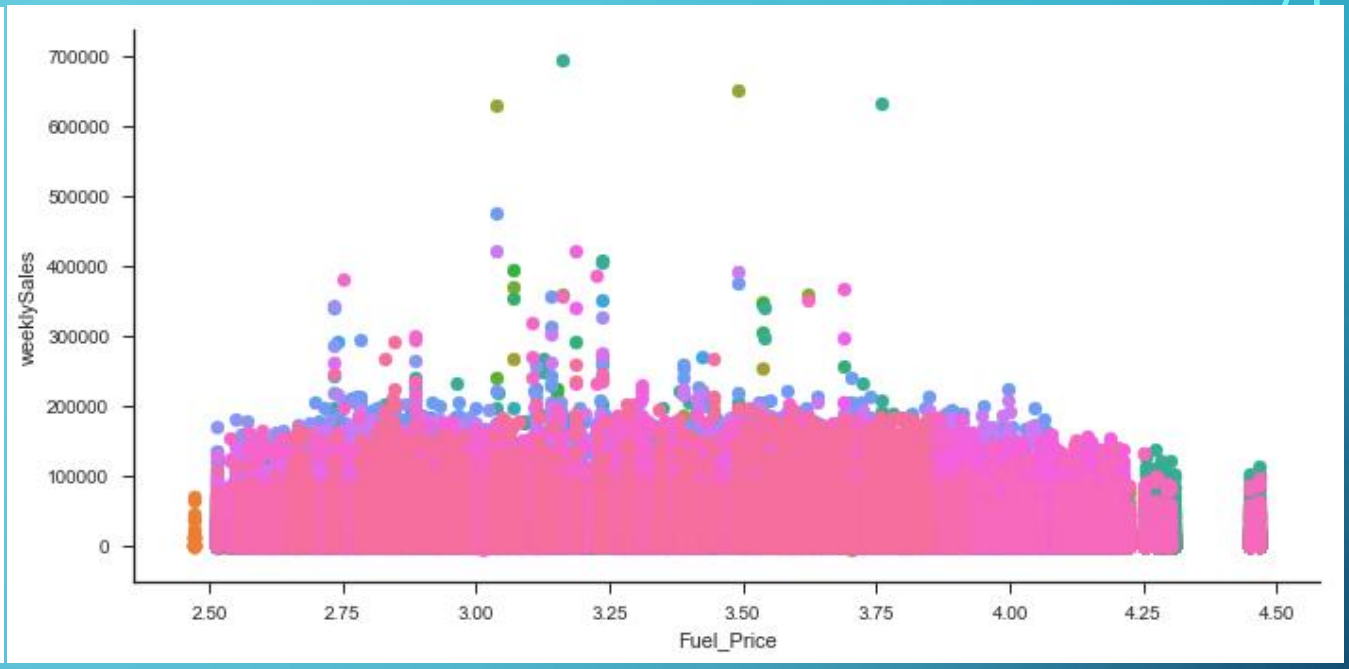

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
140
160
180
200
220
CPI

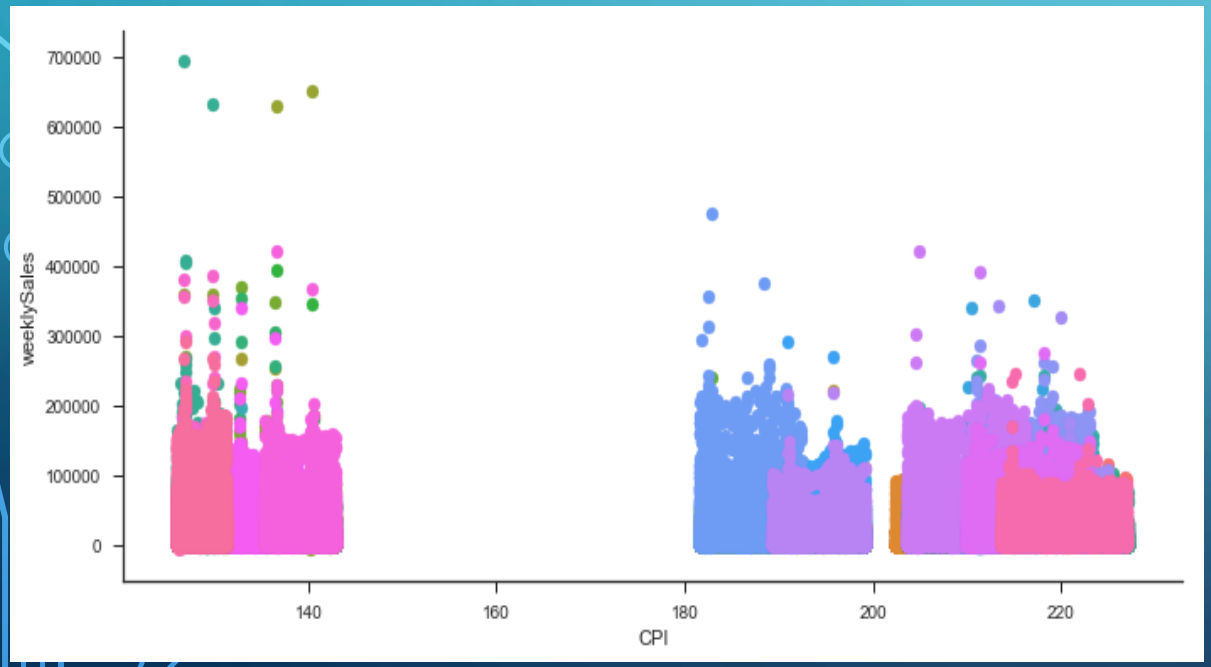

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
4
6
8
10
12
14
Unemployment

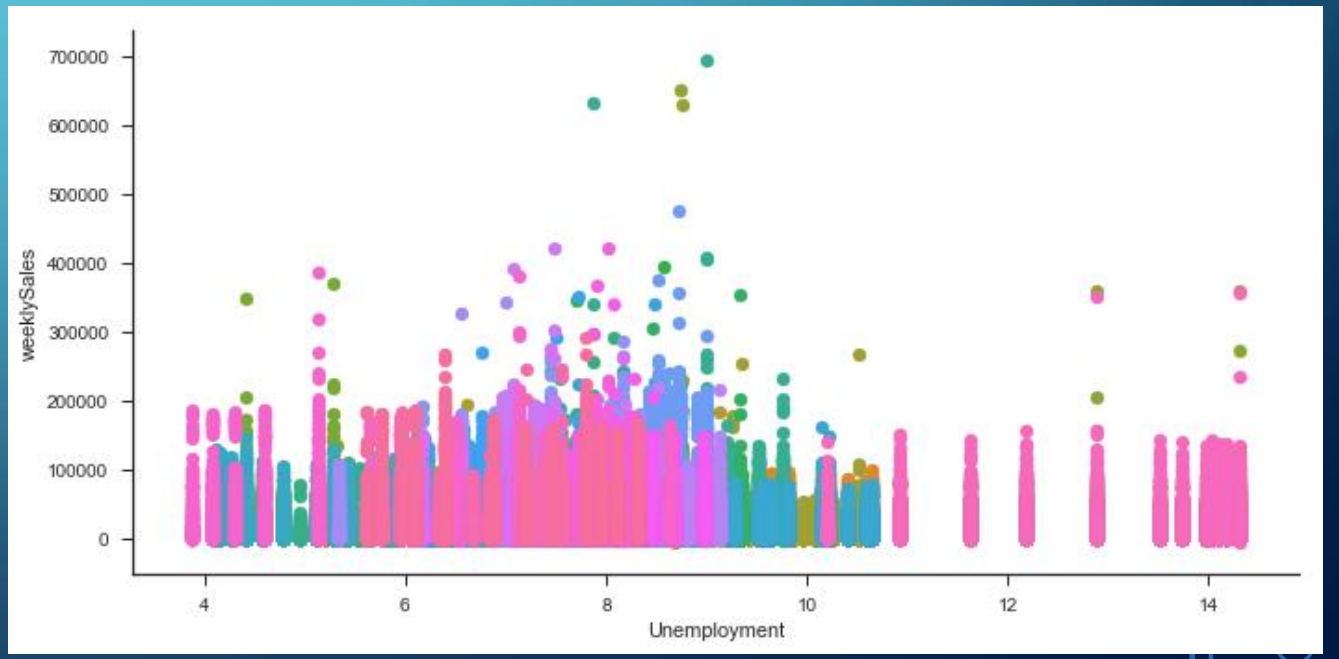

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
0
10
20
30
40
Store


weeklySales
700000
600000
500000
400000
300000
200000
100000
0
Date


weeklySales
700000
600000
500000
400000
300000
200000
100000
0
0
20
40
60
80
100
Dept

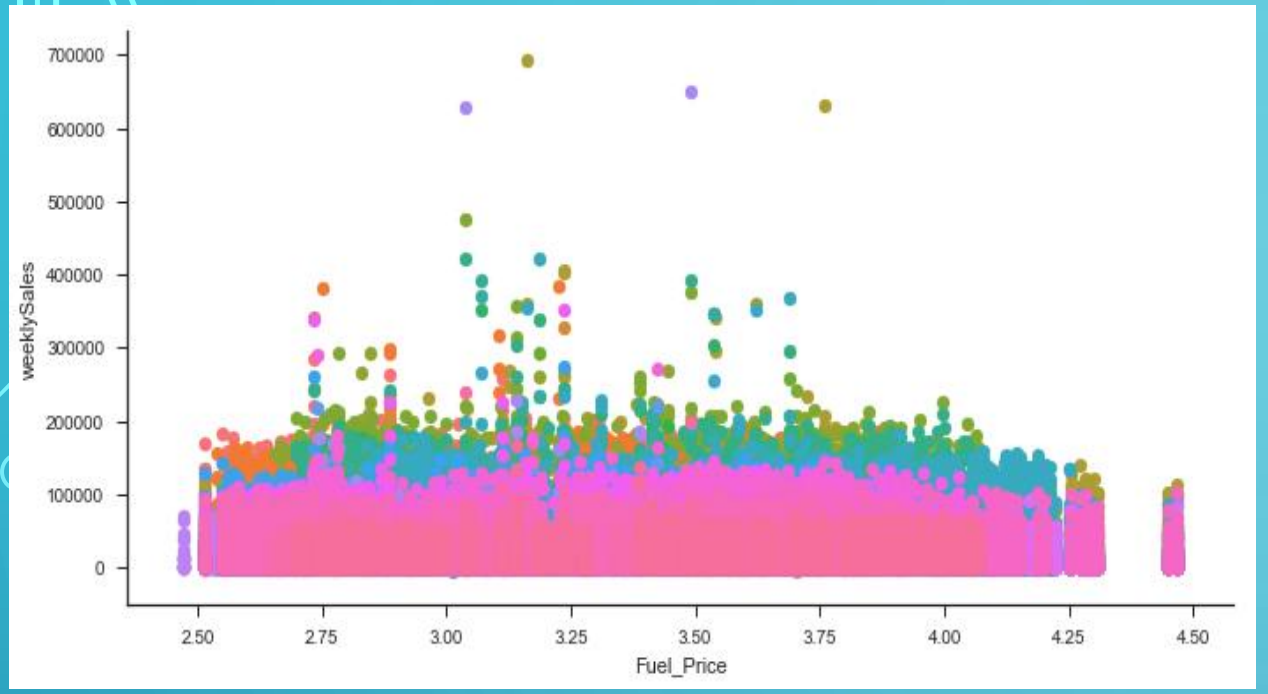

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
2.50
2.75
3.00
3.25
3.50
3.75
4.00
4.25
4.50
Fuel_Price

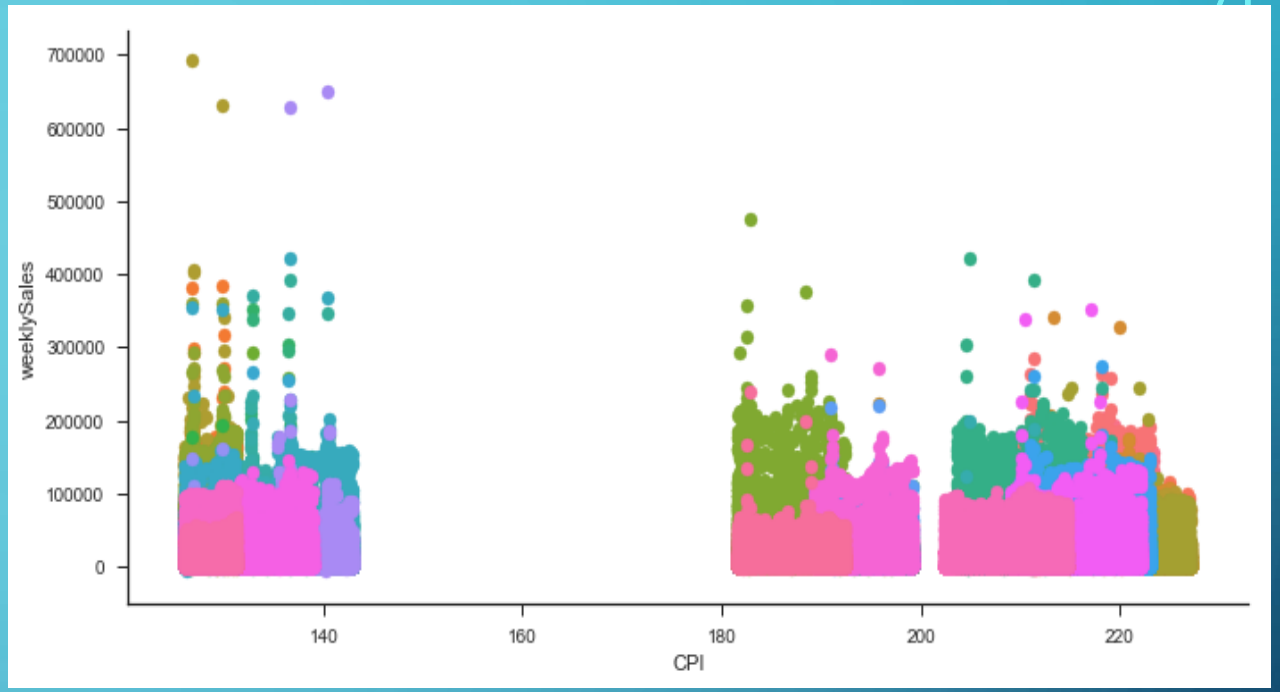

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
140
160
180
200
220
CPI

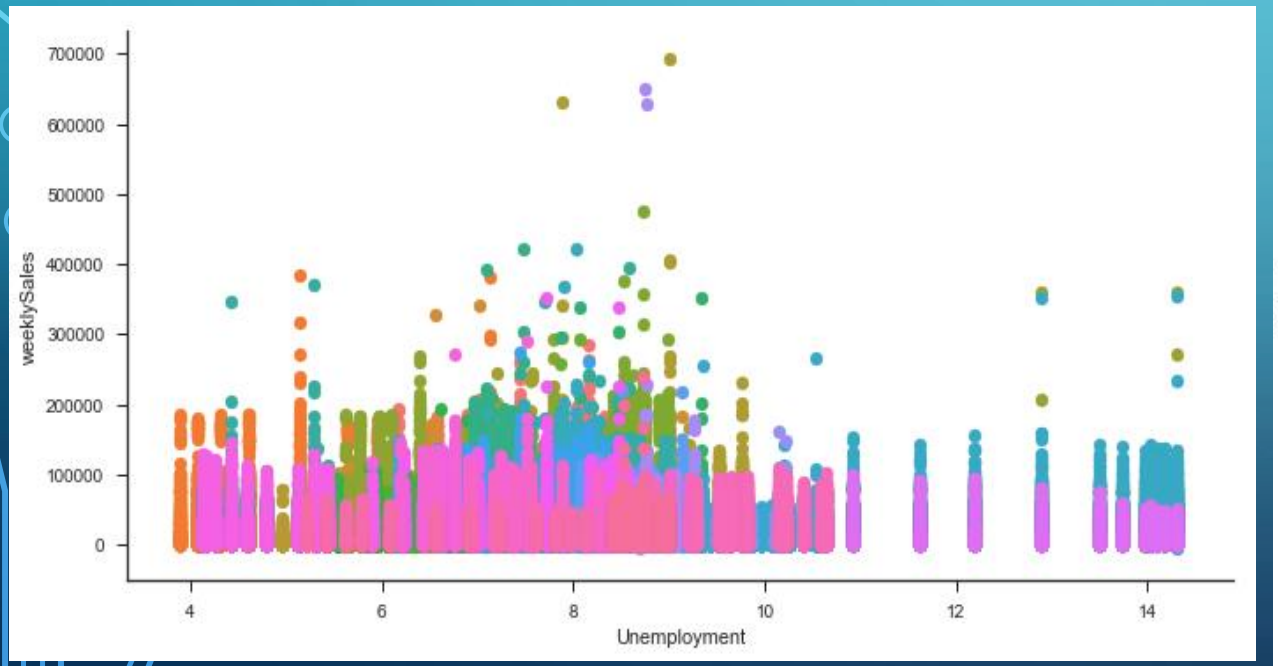

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
4
6
8
10
12
14
Unemployment

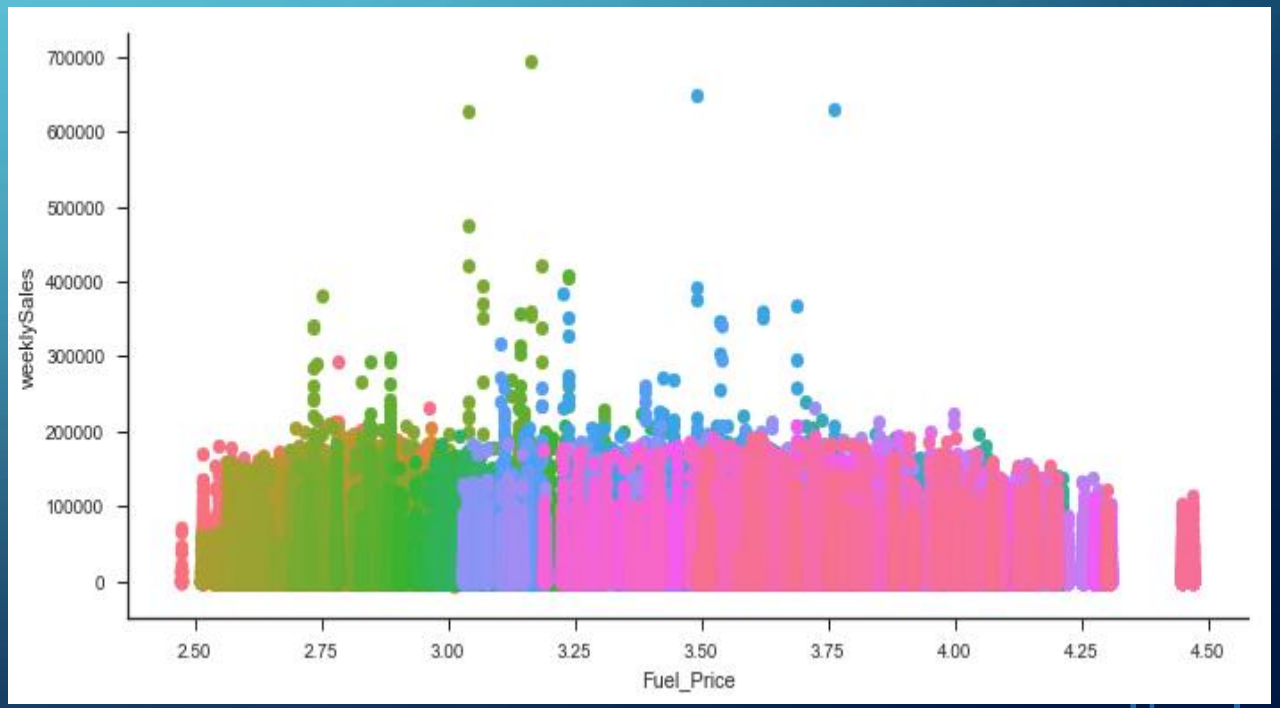
DATA ANALYSIS

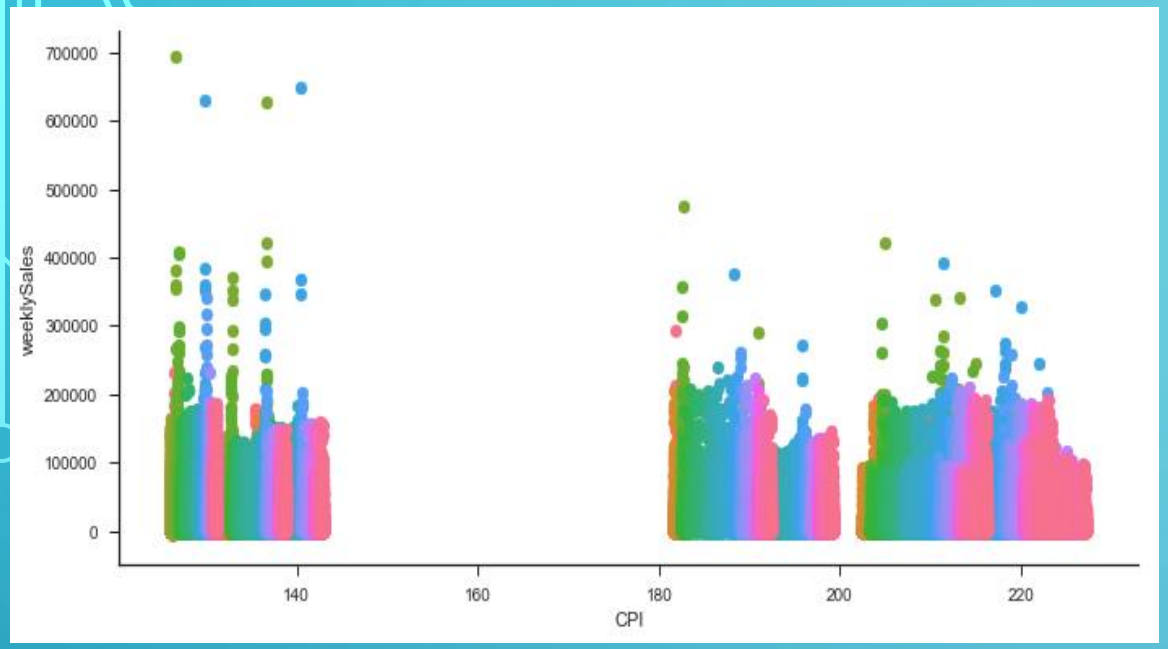
- Relationship between Weekly sale and 'Fuel_Price', 'CPI', 'Unemployment' in different **Stores.**
- Relationship between Weekly sale and 'Fuel_Price', 'CPI', 'Unemployment' in different **Date.**
- Relationship between Weekly sale and 'Fuel_Price', 'CPI', 'Unemployment' in different **Dapartrments.**

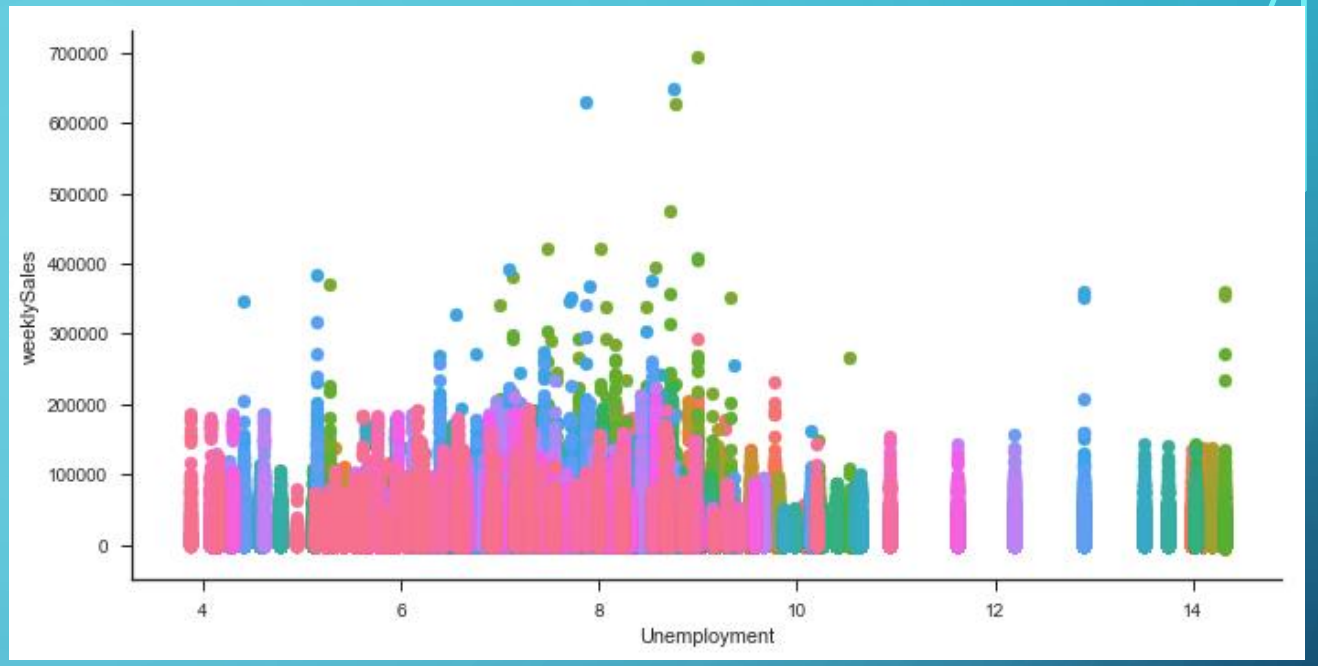

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
2.50
2.75
3.00
3.25
3.50
3.75
4.00
4.25
4.50
Fuel_Price

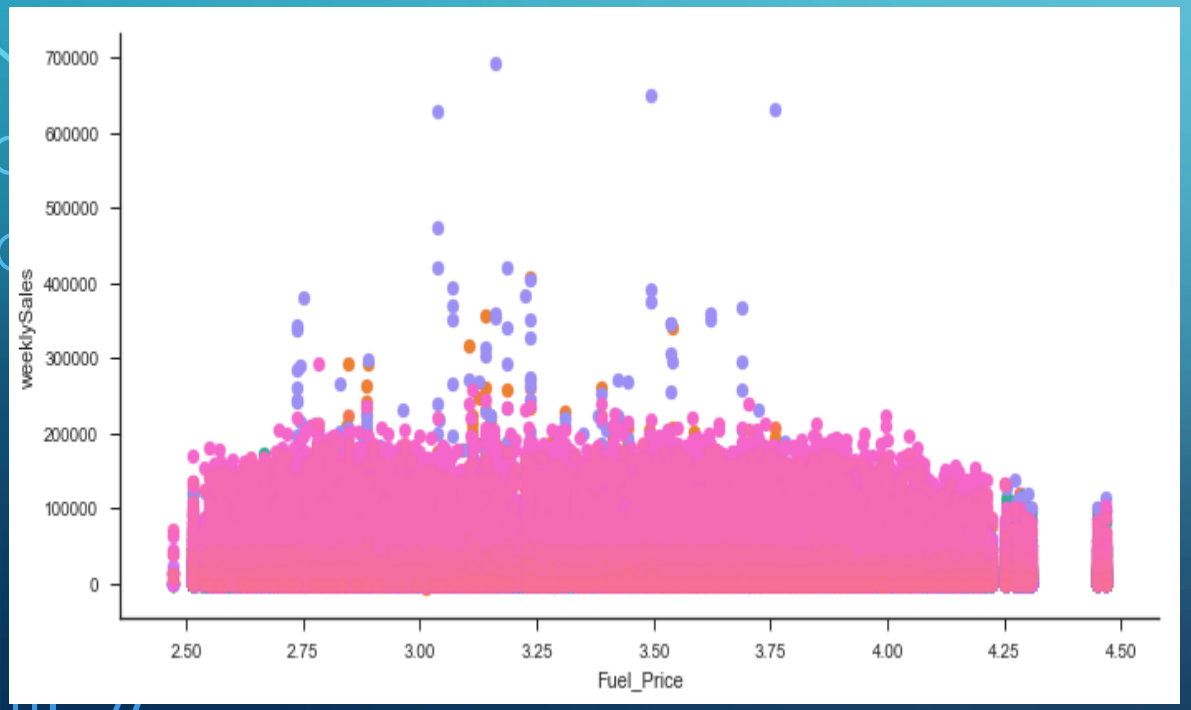

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
140
160
180
200
220
CPI

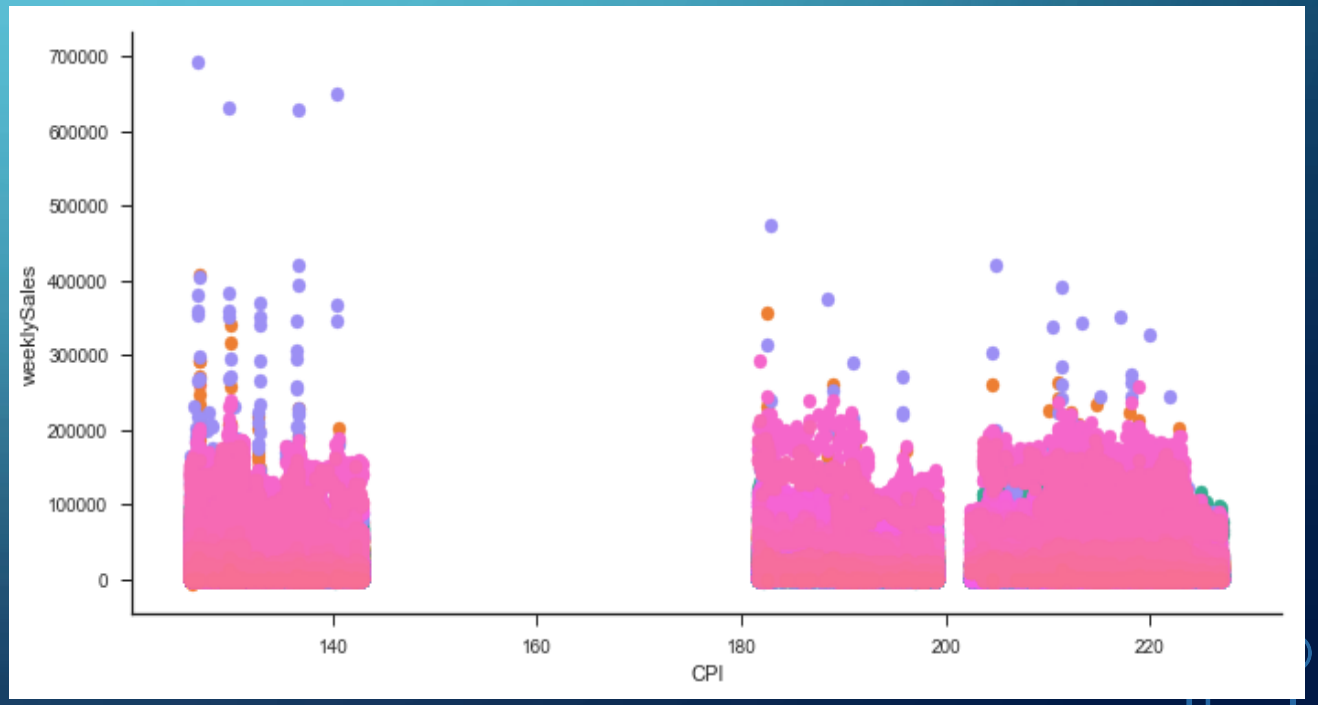

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
4
6
8
10
12
14
Unemployment

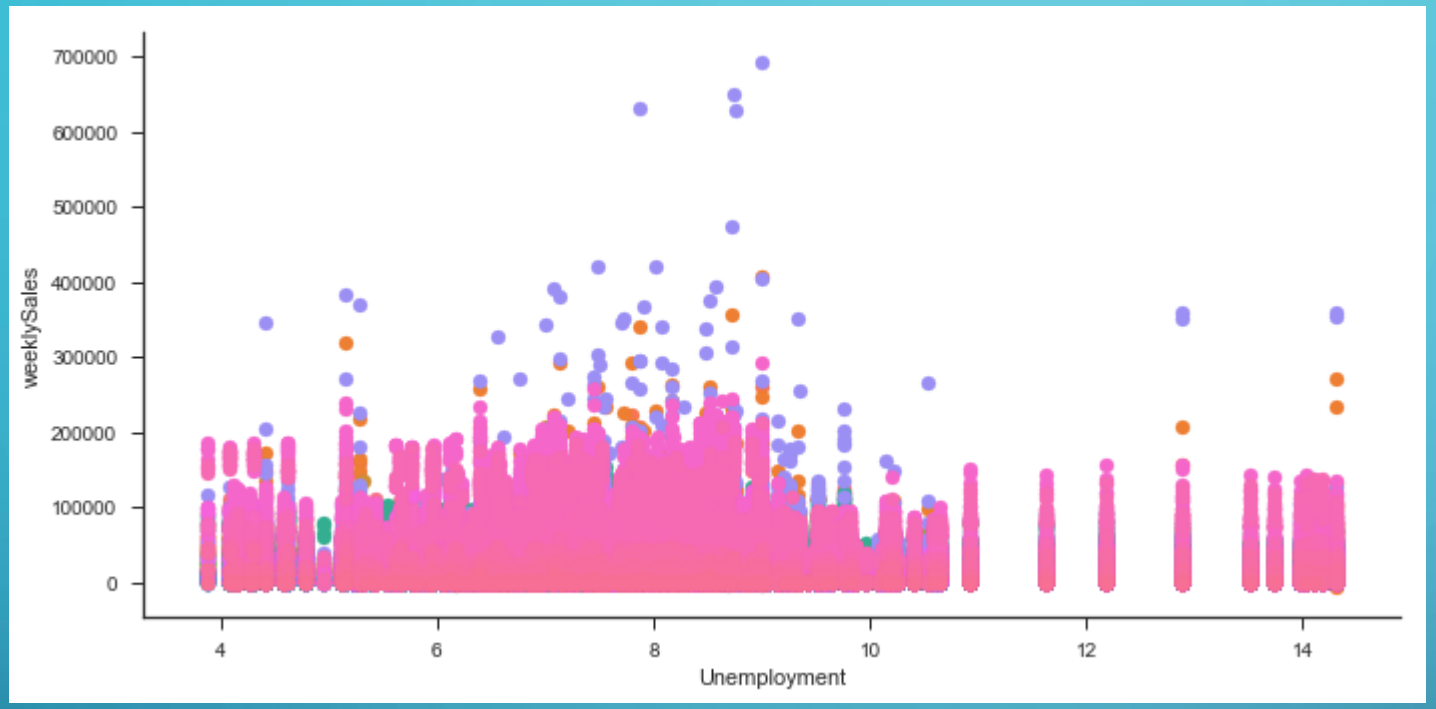

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
2.50
2.75
3.00
3.25
3.50
3.75
4.00
4.25
4.50
Fuel_Price

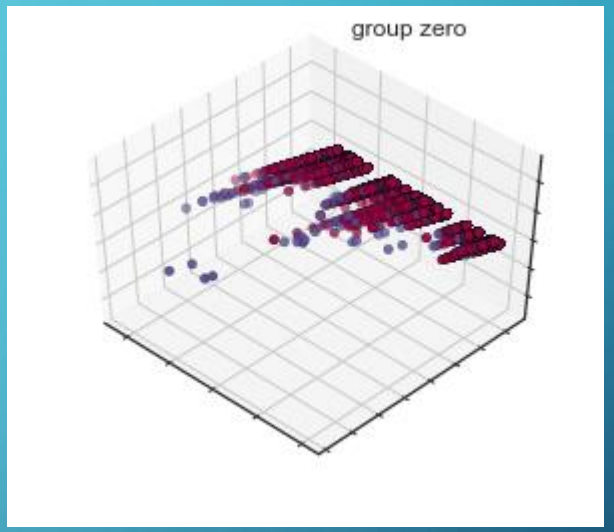

weeklySales
700000
600000
500000
400000
300000
200000
100000
0
140
160
180
200
220
CPI


weeklySales
700000
600000
500000
400000
300000
200000
100000
0
4
6
8
10
12
14
Unemployment


weeklySales
700000
600000
500000
400000
300000
200000
100000
0
2.50
2.75
3.00
3.25
3.50
3.75
4.00
4.25
4.50
Fuel_Price


weeklySales
700000
600000
500000
400000
300000
200000
100000
0
140
160
180
200
220
CPI





What We get from these graphs?

CONCLUSION

- 1. The situation of sale for different store are different. The sale of several specific holidays is much better than non-holidays. Some holidays even have less sale than regular non-holidays.
- 2. The end of the year has lots of holiday, and the sale of stores increase.
- 3. Different department have different sales. Most departments sell better in holiday than regular days.
- 4. For different stores (sizes is also different), the CPI is different.
- 5. Different sizes of stores have different unemployment situation.
- 6. For different Date, Fuel_Price and CPI are different.
- 7. If the Fuel price is very high in holiday, the weekly sale just decrease and be very little.

LINEAR REGRESSION

	Coefficients	Standard Errors	t values	P-values
0	6115.7007	400.665	15.264	0.0
1	-87.9700	2.764	-31.832	0.0
2	111.4940	1.098	101.565	0.0
3	0.0876	0.001	156.361	0.0
4	21.7148	1.912	11.358	0.0
5	-370.8796	75.879	-4.888	0.0
6	-22.1171	0.952	-23.242	0.0
7	-171.3561	19.494	-8.790	0.0

CONCLUSION& PREDICTION

The sale in the whole year is not with big up and down. Big increase of sale is in holiday.

I only get the prediction with continuous variables.

Weekly sale = 6115.7007- 87.9700*Store + 111.4940*Dept + 0.0876* Size+ 21.7148* Temperature -370.8796* Fuel_Price -22.1171*CPI - 171.3561* Unemployment

IMPROVE THE DATA AND MODEL

THANK YOU!!!