Basic Statistics tasks

1. Which store has maximum sales

Store	Weekly_Sales
20	301397792.0

2. Which store has maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation

Store	Weekly_Sales
14	317569.949476

```
15.713673600948338
```

   - The co-efficient of mean to std deviation of store 14 is 15.713%

3. Which store/s has good quarterly growth rate in Q3'2012

Store	Q2 Sales	Q3 Sales	Difference	Growth Rate
16	6626133.0	6441311.0	-184822.0	-0.027893

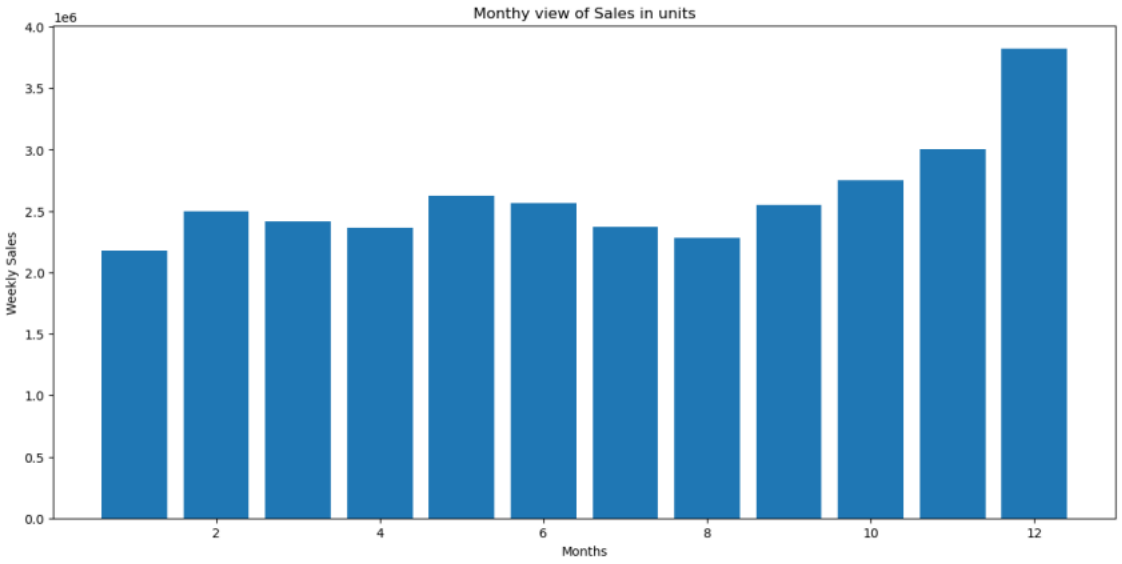
   - Here, the store 16 has good quarterly growth rate in Q3'2012 which has a Growth Rate of -0.027

4. Some holidays have a negative impact on sales. Find out holidays which have higher sales than the mean sales in non-holiday season for all stores together

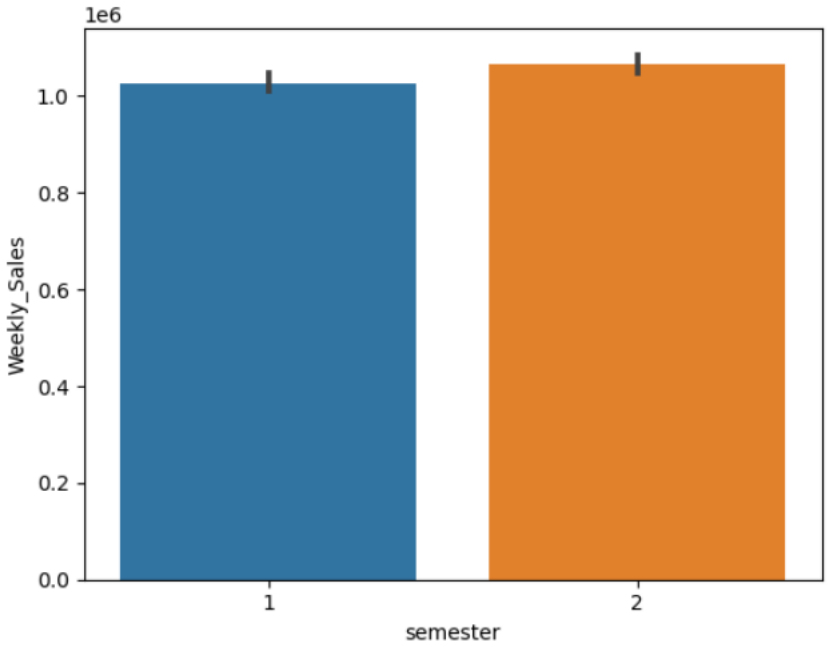
```
Super bowl: 1079127.99
Labour Day: 1039182.83
Thanksgiving: 1471273.43
Christmas: 960833.11
Non-Holiday sales: 1041256.38
```

   - Here, Thanksgiving has the highest sales than the mean sales in non-holiday season for all stores

5. Provide a monthly and semester view of sales in units and give insights



- Here, December has the highest sales



- Here, Semester 2 has the hightest sales in unit

Statistical Model

For Store 1 – Build prediction models to forecast demand

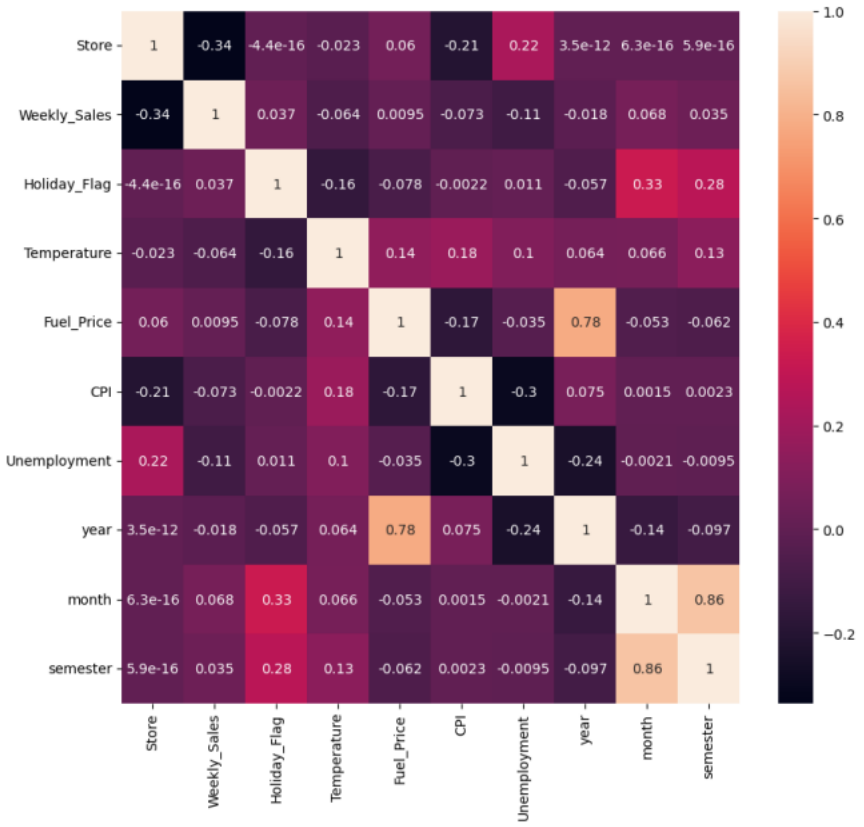
1. Linear Regression – Utilize variables like date and restructure dates as 1 for 5 Feb 2010 (starting from the earliest date in order). Hypothesize if CPI, unemployment, and fuel price have any impact on sales.

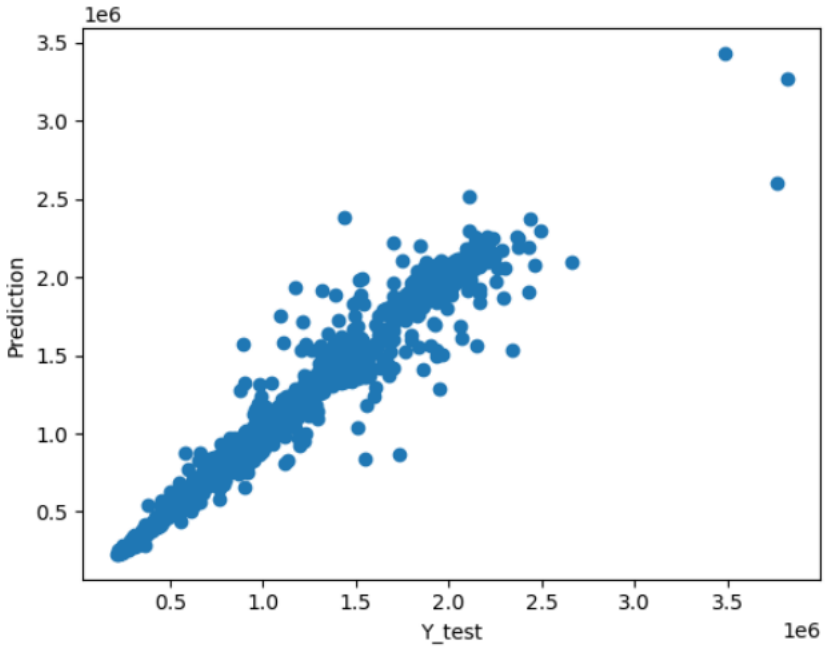
```
R2 score: 0.14894500845355385
Linear model accuracy: 0.14372803259754718
Mean Squared Error: 274171250281.01086
```

Select the model which gives best accuracy. – Random Forest model

```
R2 score: 0.9524234944111671
Root Mean Squared Error: 123802.21502366115
```

```
Mean Squared Error: 15326988444.76483
```





- For Random Forest Regressor R2 score: R2 score: 0.9524234944111671 Root Mean Squared Error: 123802.21502366115 Mean Squared Error: 15326988444.76483
- The Random Forest Regressor model would be the best fit for the outcome

2. Change dates into days by creating new variable.

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	year	month	day	semester
0	1	2010-05-02	1643690.90	0	42.31	2.572	211.096358	8.106	2010	5	Sunday	1
1	1	2010-12-02	1641957.44	1	38.51	2.548	211.242170	8.106	2010	12	Thursday	2
2	1	2010-02-19	1611968.17	0	39.93	2.514	211.289143	8.106	2010	2	Friday	1
3	1	2010-02-26	1409727.59	0	46.63	2.561	211.319643	8.106	2010	2	Friday	1
4	1	2010-05-03	1554806.68	0	46.50	2.625	211.350143	8.106	2010	5	Monday	1