**Project- Retail Analysis using Walmart Data.**

After getting brief understanding about python and data science concepts based on the flow from business objective to deployment. First tried to go through what is the data all about how many features does it have .

**First step-1** Business problem and data acquisition.

The objective of the analysis is to check weekly sales happening in Walmart also to check how does holiday causes an impact on sales and how well it does comparing with non holiday days.

**Step 2-** Data preprocessing and wrangling.

Data is imported jupyter notebook after importing checked the size and shape of data.

Shape: 8 columns and 6435 rows.

Features: Store ,Date, Unemployment, Holiday Flag, CPI Temperature and Fuel Price.

Store: values ranges from 1-45

Unemployment: values ranges from 8.16 to 41 (approx.)

Holiday Flag: 0 and 1

CPI: ranges from 1 to 8 (12 is outlier).

Temperature: -2 to 100 degree Fahrenheit.

Fuel Price: 2.42 to 4.

Date has been converted to datetime

From Date , Days, Months and Years are separated.

On applying wrangling techniques the data doesn’t have any null values.

**Step 3:** Exploration.

Explored data using pandas, numpy, matplotlib libraries.

**Step 4:** Model development.

Data has been divided into train test with ratio 80:20 i.e. 80% training 20% testing.

Using Linear Regression: accuracy is 20%

Using logistic regression: accuracy is 92%

Using random forest regression: accuracy is 87%

Comparing both two logistic regression gives better result.

For rest kindly refer source code.