Anjoo Bharti

Departmet of CSE (computer science engineering) 3rd year

College Name – SR institute of management and technology Lucknow

Project- Analyse Walmart store sales

Submitted to:

Shaurya Shinha

Date- 02/10/2022

Abstract:

- The aim of this project is to enable category managers of Walmart to check the weekly and monthly sales of the departments. Analysis includes the effect of markdowns on the sales and the extent of effect on the sales by fuel prices, temperature, unemployment, CPI etc. Has been analysed using simple and multiple linear regression models. Analytical tools used in the project are studio and Shiny app. The user interface for this project is user friendly as the user has to use the drop down menus and slider inputs to see the variation in graph.

Keywords:

- weekly sales, monthly sales, temperature, markdowns, store type, store size, department, linear regression.

Business Problem:

- The decision makers of Walmart should be able to analyse the effect of various factors affecting the sales of the products in their 45 stores. The various factors include weather condition i.e., temperature, store size, fuel prices, markdown in prices, unemployment and CPI.

Analytics Problem :

- In this problem we have analysed sales across different departments by store type and created weekly and monthly dashboard. We have analysed the effect of various factors such as temperature, store size, fuel prices, markdown in prices, unemployment and CPI to determine which factors have a statistical significance in explaining sales in the stores by using simple and multiple linear regression.

Data:

- The data has been taken from the Cagle data analytics competition, it contains data of 45 Walmart stores and its various departments. The original data files used for our analysis were stores.csv, train.csv and features.csv which contained the below mentioned fields:

- Stores.csv: This file contains anonymized information about the 45 stores, indicating the type and size of store.
- Train.csv: This is the historical training data, which covers to 2010-02-05 to 2012-11-01. Within this file you will find the following fields.
- Store – the store number
- Dept. – the department number
- Date – the week
- Weekly Sales – sales for the given department in the given store
- Is Holiday – whether the week is a special holiday week.

Features.csv:

This file contains additional data related to the store, department, and regional activity for the given dates. It contains the following fields:

- Store – the store number
- Date – the week
- Temperature – average temperature in the region
- Fuel Price – cost of fuel in the region
- MarkDown1-5 – anonymized data related to promotional markdowns that Walmart is running. Markdown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA.
- CPI – the consumer price index
- Unemployment – the unemployment rate
- Is Holiday – whether the week is a special holiday week We merged the data files to our convenience for analysis which have been uploaded in our GitHub page.

Methodology Selection :

The methodologies which we used in this project are:

1. Merge different data sets: We had three data files – training, features and stores. We merged all the three data files to see effect of different variables on sale. Since this data was already cleaned, we didn't do any data cleaning work.
2. Study summary descriptive statistics: We have studied how different factors like week, month, store size, temperature effect sales by using plot function in R. An interesting observation during descriptive statistics came up where we realized markdowns were also impacting sales. But since markdown data wasn't big enough for linear regression model, we restricted ourselves with just descriptive statistics work.
3. Build linear regression models: We built linear regression in R to predict sales using week of the year, store size and temperature. We have used backward selection model to analyze the effects of various predictors on the sales.

Model Building:

- We have created separate dashboards to analyze variation of sales for departments with Week, Month, Temperature, Store Size, Markdowns. Next, we tried to get insights by comparing the factors like fuel prices, temperature, unemployment rate, CPI etc., with sales of the stores. We found that there were not much insights by comparing the sales with unemployment, CPI and that resulted in not much variation in the graphs. Store size, Temperature and Week of the year have shown some interesting findings and it affected the sales by month significantly.

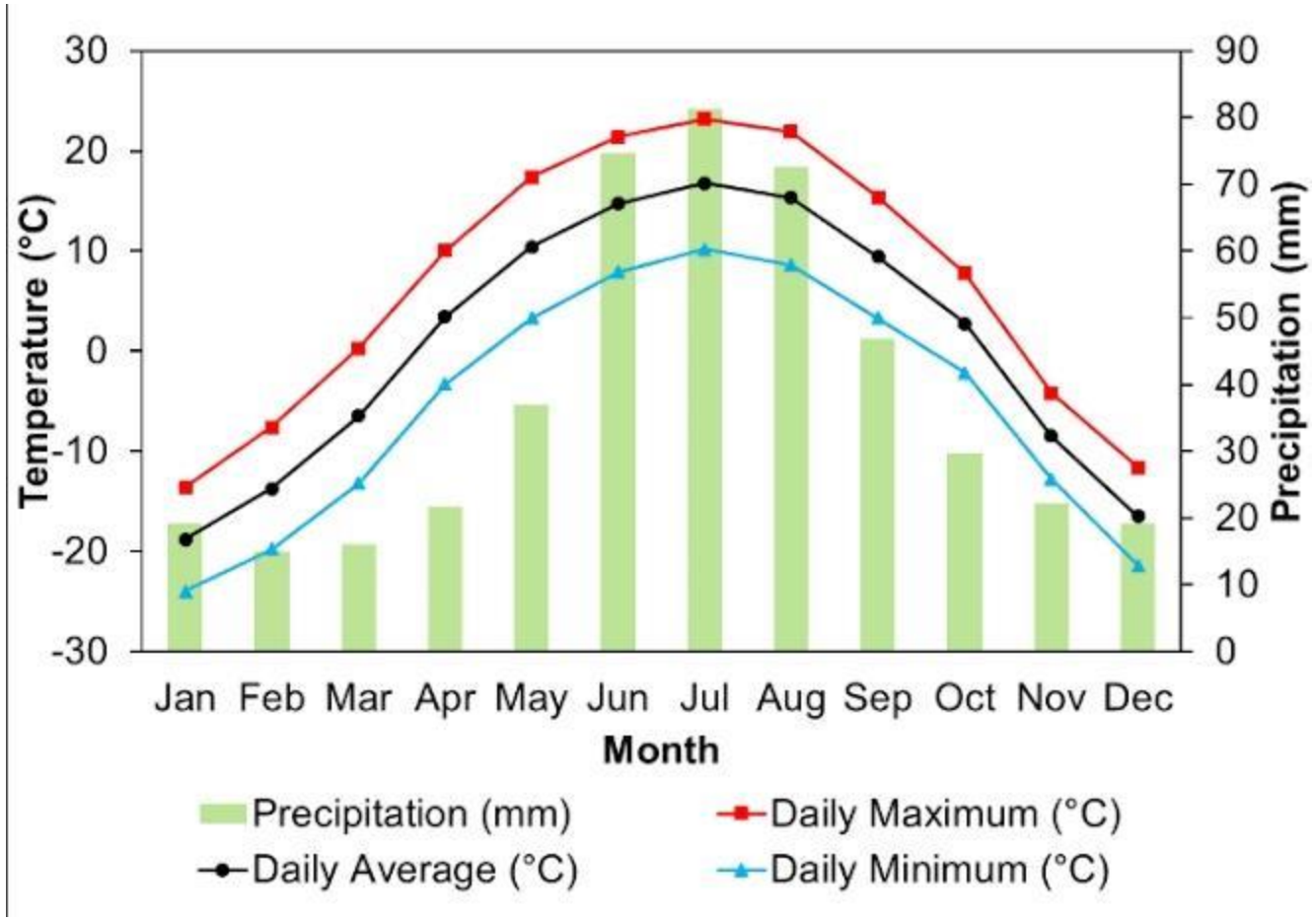
Functionality:

- The Shiny apps are built for easy use by a Category Manager to visually analyze the various factors affecting the sales in their stores and to determine which of these factors strongly affect the sales. The first app helps in descriptive analysis. Within the app, the first tab gives the leverage to check the monthly sales by store and its individual departments. Second tab gives the leverage to get to know the performance of the stores and we can also sort the stores by the store size. Next tab will give weekly sales versus the store type for different years. In third tab user can see how sales have been for his department by store and year. In fourth tab, user can see how markdowns impacting sales by store type for different departments. In fifth tab, user can see how sales vary with store sizes.

- The second app helps in linear regression analysis. Within this app, the left panel has an option for the user to input the portion of data to be included in training data set; an option to select the regression model to summarize and various other options to change the user preference. The first tab shows the summary of regression model and the statistical metrics of the trained data set such as R2, adjusted R2, f-value, t-value, p-value etc. The second tab shows the quality of the prediction of this trained model by showing the R2 and root mean square error of the training and test data sets.

GUI Design and Functionality :

- The first app helps in descriptive analysis of the data. The first tab (shown below) depicts the sales by stores. The user inputs the store number from 1 to 45 and the department number from 1 to 99. The corresponding sales vs month of the year, sales vs temperature etc. are displayed.

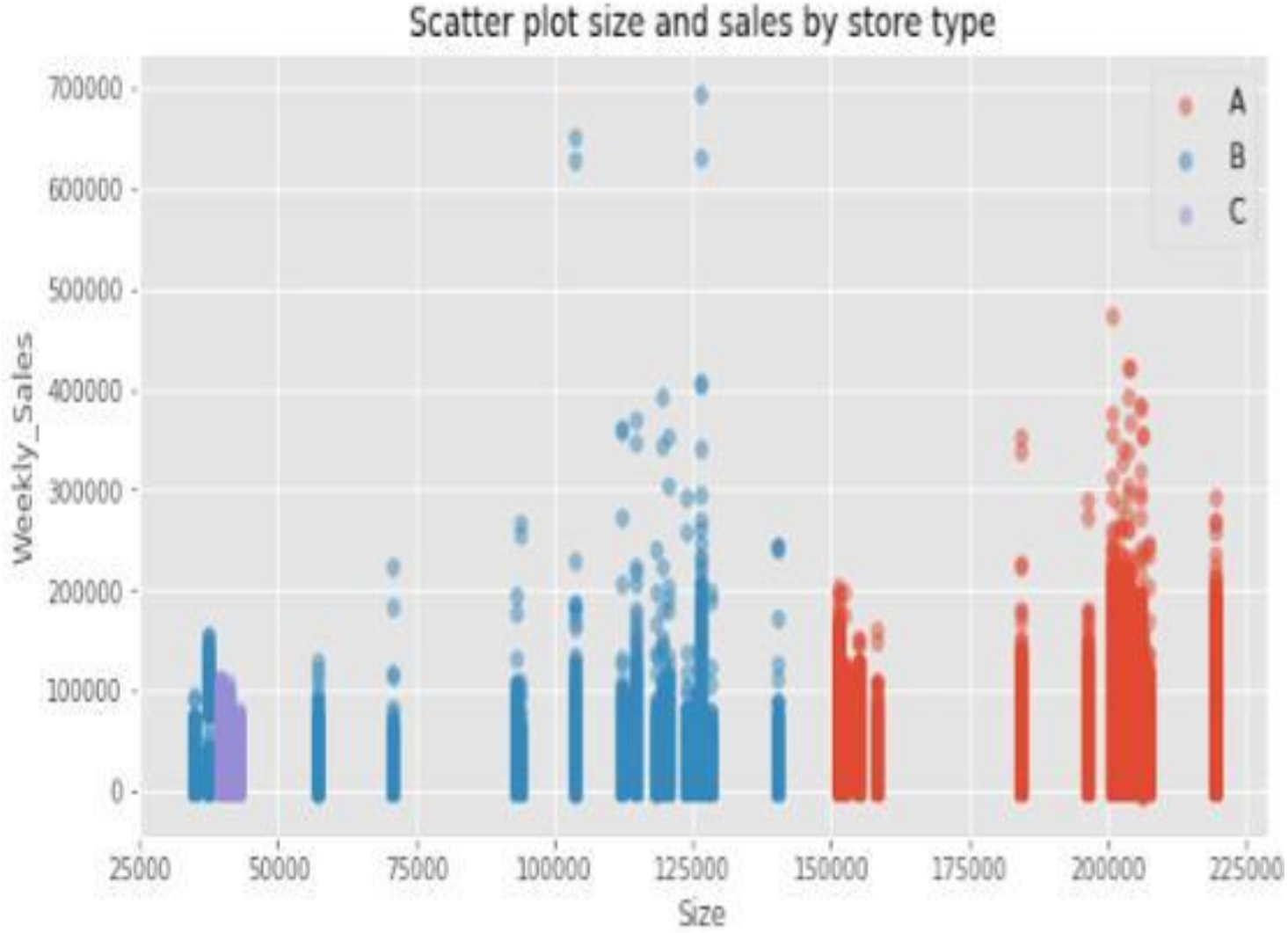


- The second Tab (below) shows the data with respect to the year and month. It takes the user inputs of month from 1 to 12 and year from 2010-2012

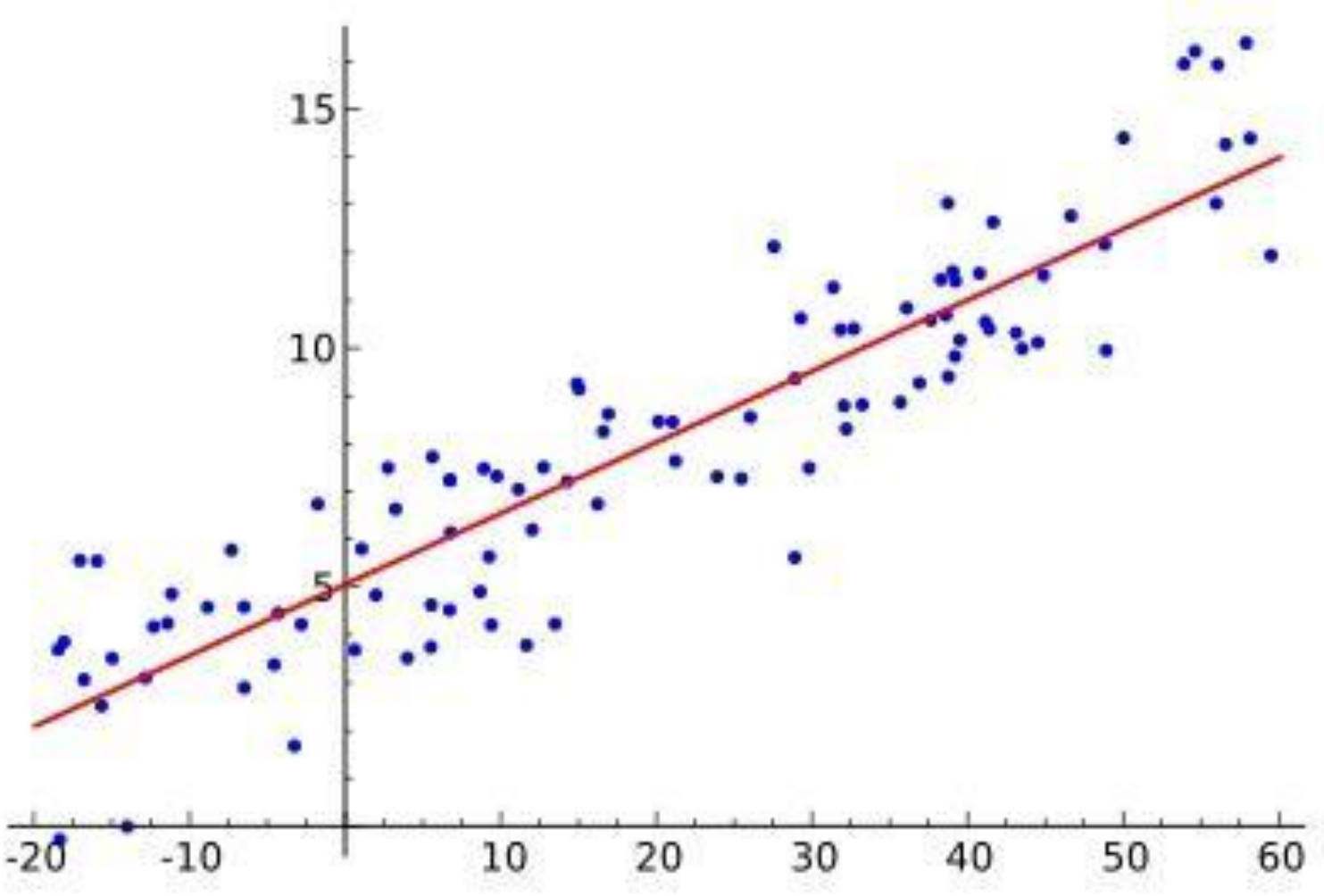
MONTHLY SALES PROJECTION TEMPLATE

YEAR	NEW BUSINESS			REORDERS			TOTALS		
20XX	GOAL	ACTUAL	VARIANCE	GOAL	ACTUAL	VARIANCE	GOAL	ACTUAL	VARIANCE
JANUARY	199	145	-27.1%	85	105	23.5%	284	250	-12.0%
FEBRUARY	166	238	43.4%	105	108	2.9%	271	346	27.7%
MARCH	204	152	-25.5%	90	71	-21.1%	294	223	-24.1%
APRIL	224	291	29.9%	89	78	-12.4%	313	369	17.9%
MAY	290	158	-45.5%	85	76	-10.6%	375	234	-37.6%
JUNE	138	272	97.1%	80	87	8.8%	218	359	64.7%
JULY	154	255	65.6%	73	74	1.4%	227	329	44.9%
AUGUST	268	207	-22.8%	73	108	47.9%	341	315	-7.6%
SEPTEMBER	294	284	-3.4%	78	86	10.3%	372	370	-0.5%
OCTOBER	249	213	-14.5%	102	72	-29.4%	351	285	-18.8%
NOVEMBER	153	278	81.7%	106	93	-12.3%	259	371	43.2%
DECEMBER	286	148	-48.3%	97	98	1.0%	383	246	-35.8%
	2,625	**2,641**	**0.6%**	**1,063**	**1,056**	**-0.7%**	**3,688**	**3,697**	**0.2%**

- The tab (below) shows how sales is affected by store size. The user input the year and The department number and output shows the graph of sales catagorized by store type A,B and C using different colours



- The second app show the linear regression analysis in effect. Within this app the left panel has the user input the portion he/she wants to be used for training the data and the user selects the regression model (sales –Week, Sales-Week+Size, Sales –Week+Size+temprature) he/she wants to analyse. The first tab (below) shows the summary of the regression analysis and the statistics and coefficient of the trained model.



Straight :

- Diversity in product and service
- Convenient prices and locations
- Strong market presence
- Custumer loyalty
- Strong financial performance
- Coat and pricing advantage over rivals

Weakness:

- Brand image weak reputation
- Low global presence
- Behind rivals in e-commerce

Opportunities:

- Global Expansion: new geographic areas
- Increasing online sales
- Strategic alliance
- Acquiring rival firms

Conclusion :

- Walmart has Business model that is already appealing to the Chinese consumer.
- The environment in China is one in which Walmart could replicate its success in the US.
- By fine tuning itself to the Chinese market and adjusting it's model with consideration for the resources in china, Walmart can grow market share with in the country and capitalize on the economic expression of China.

Thank you