Visualization Project: Walmart store sales analysis

Sneha KK

Introduction:

Sales analysis is an important aspect of running a store as we need to understand the factors that influence sales if we want to increase profit and customer satisfaction. It provides insights into the past, present, and future performance and can be used to help forecast trends, identify opportunities for growth, and develop a strategic action plan for the stores. The aim of this project is to analyze the sales of Walmart stores to find how different features impact the sales in a Walmart supermarket using exploratory data analysis and R visualizations.

Data set description:

The data set was downloaded from kaggle: https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting It contains the historical sales data for 45 Walmart stores located in different regions, and each store contains a number of departments. The data is distributed in four csv files:

1. stores.csv : type and size of stores

2. train.csv: This is the historical training data, which covers to 2010-02-05 to 2012-11-01.The features represent store number,department number, date , weekly sales(sales for the given department in the given store) and isholiday(whether the week is a special holiday or not)

3. test.csv: This file is identical to train.csv, except weekly sales column is absent.

4.features.csv: This file contains additional data related to the store, department, and regional activity for the given dates. It contains features like store number,date, temperature, fuel price, markdown, CPI, unemployment rate,holidays

Project methodology:

The dataset was downloaded and checked for null values:

```
##       Store        Dept        Date Weekly_Sales    IsHoliday
##           0           0           0            0            0


##     Store     Dept      Date IsHoliday
##         0        0         0         0


##       Store        Date Temperature  Fuel_Price    MarkDown1    MarkDown2
##  0.00000000  0.00000000  0.00000000  0.00000000  0.50769231  0.64334554
##    MarkDown3    MarkDown4    MarkDown5         CPI Unemployment    IsHoliday
##  0.55885226  0.57704518  0.50549451  0.07142857  0.07142857  0.00000000


## Store  Type  Size
##     0     0     0
```
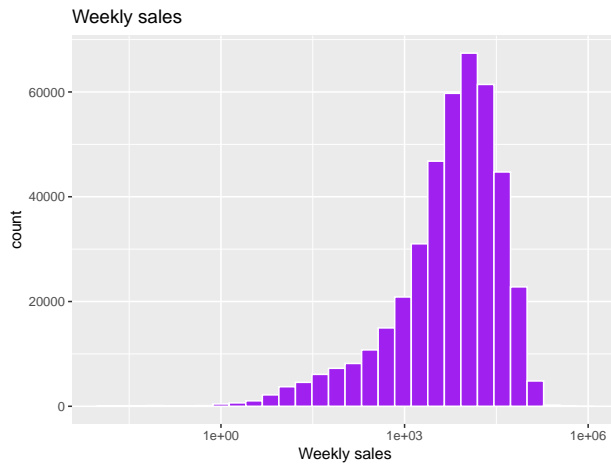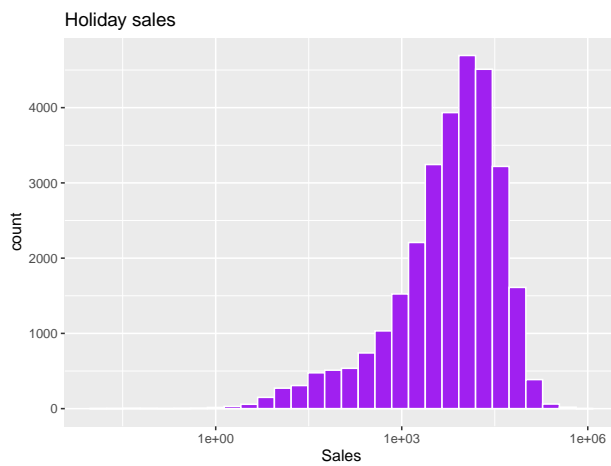
Only markdown columns have many null values, because the stores offer promotional markdowns only on some days of the year.

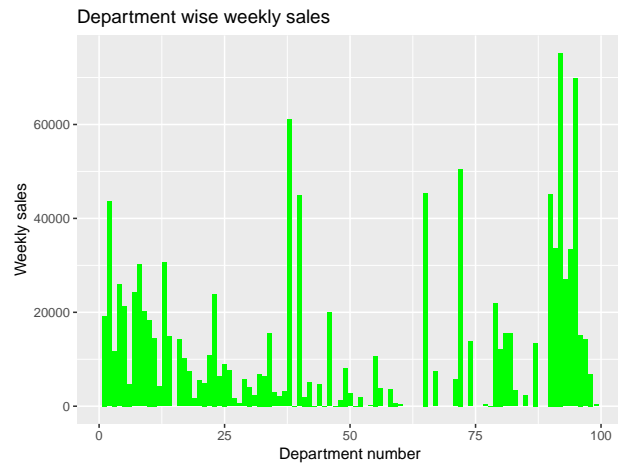Distribution of the weekly sales, log transformed:

Weekly sales

Mean weekly sales is approximately 16000 dollars.
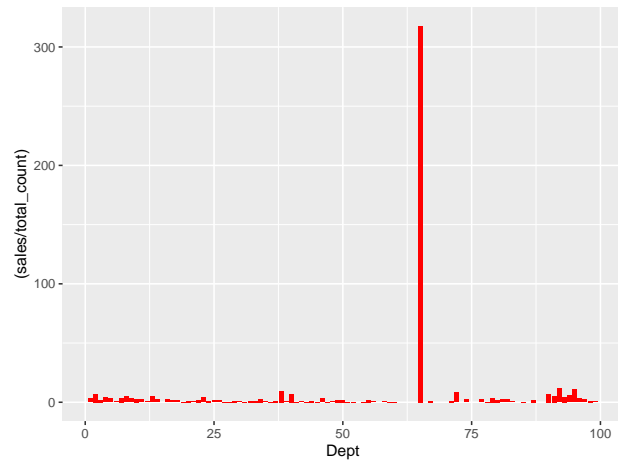Number of holidays among the 115064 days in dataset:

```
## [1] 29661
```


Holiday sales
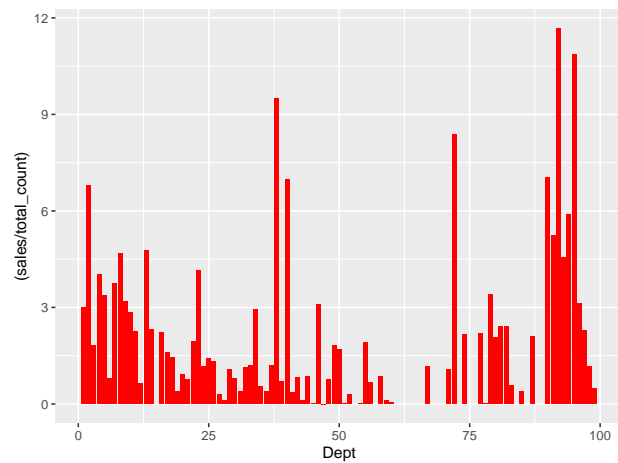
Mean holiday sales is approximately 17000 dollars.
Department wise weekly sales:
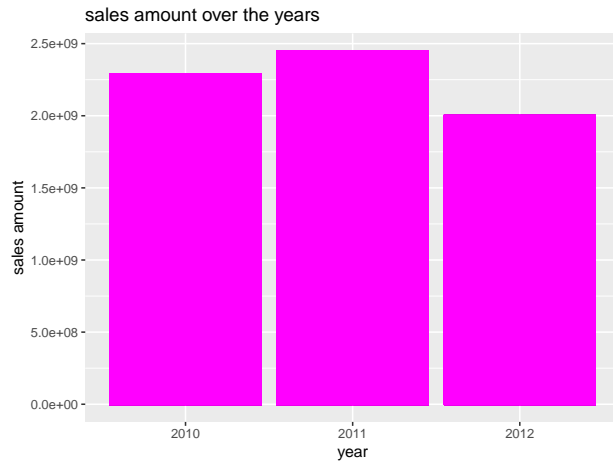
Department wise weekly sales

Departments 92,95 and 38 have the highest weekly sales.
Department wise mean sale amount:

Department 65 has unusually high values, so let's remove it from our graph:/


Departments 92,95,38 have the highest mean sales also.
Let's view the mean sales amount for the three years:

sales amount over the years

Sales in 2011 is the highest and 2012 is the least.
Variation of mean CPI over the years:

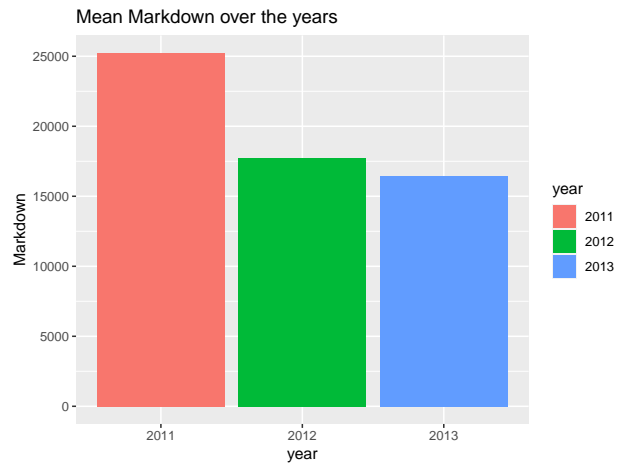Mean Consumer Price Index over the years

Mean Consumer Price index has an upward trend.
Mean unemployment rate:
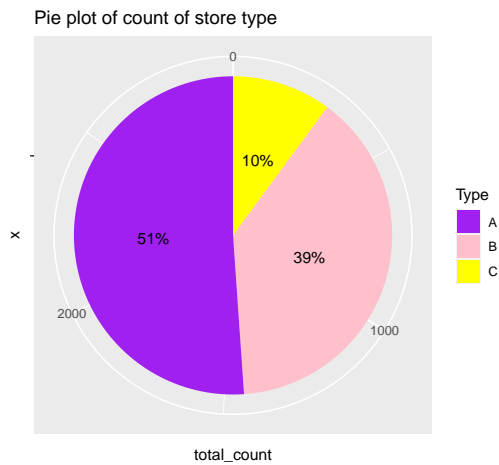

Mean Unemployment rate over the years
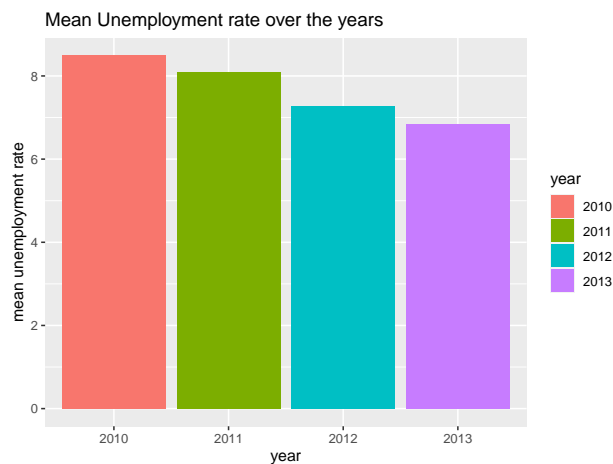
Mean unemployment rate is decreasing over the years.

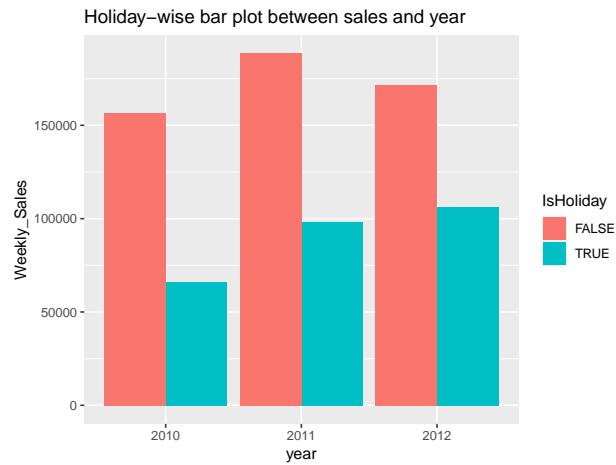Mean of the promotional markdowns offered:

Mean markdown was highest in 2011, almost same in 2012 and 2013.

Store type wise sales:

Type A stores contributed to 51% of the sales, type B stores contributed to 39% of the sales and type C stores contributed to only 10% of the sales.

Holiday–wise bar plot between sales and year

Total holiday sales is increasing over the years while the total working day sales is highest in 2011.

Holidays vs sales
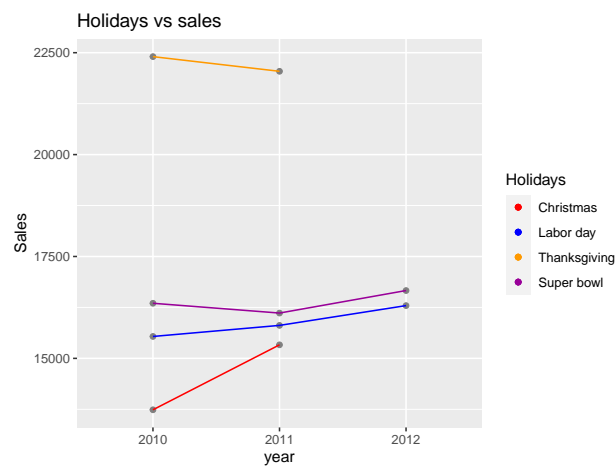
Sales during thanks giving is the highest and sales during christmas is the least.
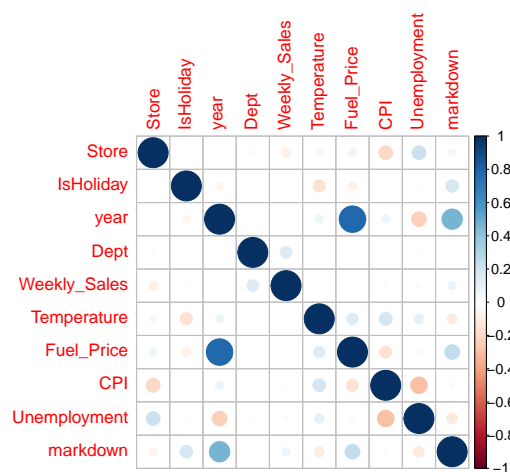
We can see that none of the variables have high correlation between them. Fuel price and year, markdown and year are positively correlated whereas Unemployment rate and CPI, CPI and stores and year and unemployment are negatively correlated.

Dashboard:

A dashboard of visuals was created using RShiny containing the following panels :
* Introduction
* Univariate analysis: Analysis of frequencies of different variables using bar plot, pie chart for categorical and histogram for continuous variables.
* Multivariate analysis: Analysing the relationship within factors using bar plots for categorical and categorical, categorical and continuous variables and scatter plot for analysis between continuous variables.
* Year wise analysis: Multi variate graphs combined and colored by year.
* Holiday wise analysis: This has two tab panels:
+ One panel contained multivariate graphs that were combined and colored by holiday/working day.
+ The second panel contained the line graphs of particular holidays.
* Store type wise analysis: Multivariate graphs combined and colored by store type.
Link to the dashboard: https://snehakk.shinyapps.io/visualization_project/

The dashboard helps us view our data in a consolidated format at one unified location. The interpretations drawn from the graphs helpd us understand the factors influencing sales amount and customer needs which in turn will lead to more profit and customer satisfaction.

Conclusions:
After detailed analysis and visualization of the Walmart sales data,we can infer the following :
* Mean weakly sales amount in Walmart super market is approximately 16000 dollars and is slightly decreasing over the years while mean holiday sales amount is approximately 17000 dollars.
* Departments 65,92,95,38 have the highest mean sales amount.
* Mean consumer price index is increasing over the years while mean unemployment rate and mean markdown amount are decreasing over the years.
* Mean markdown was highest in 2011, almost same in 2012 and 2013.
* Type A stores contributed to 51% of the sales, type B stores contributed to 39% of the sales and type C stores contributed to only 10% of the sales.
* Total holiday sales is increasing over the years while the total working day sales is highest in 2011.
* Sales during thanks giving is the highest and sales during christmas is the least.
* Fuel price and year, markdown and year are positively correlated whereas Unemployment rate and CPI, CPI and stores and year and unemployment are negatively correlated.

Hence, from Walmart sales data, we can conclude that lower CPI and higher markdown offers , investing in departments 65,92,95 and 38 would lead to an increase in sales profit.