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 from kaggle 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
```
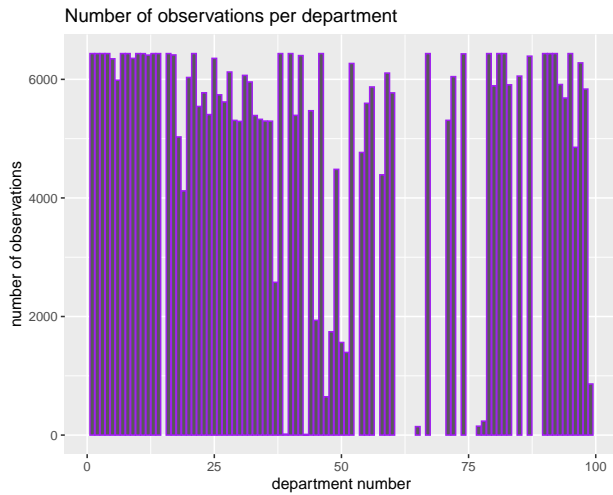
Only markdown columns have many null values, because the stores offer promotional markdowns only on some days of the year.

Some departments did not sell any items in some weeks, and the department numbers do not over all integer values:

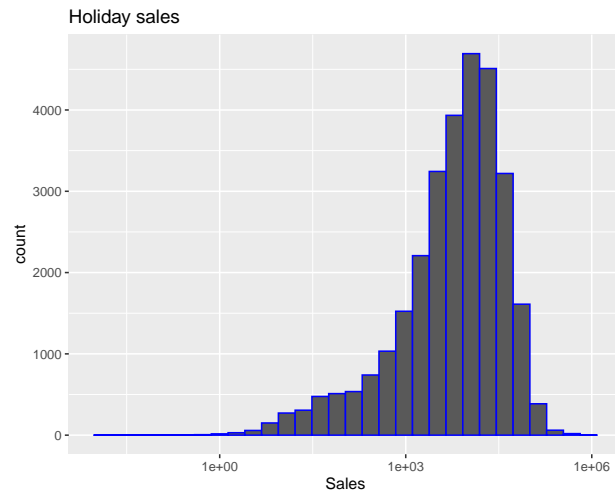Number of observations per department

Most of the departments have more than 6000 observations while a few departments like 47,76,77,99 rarely made any sales.
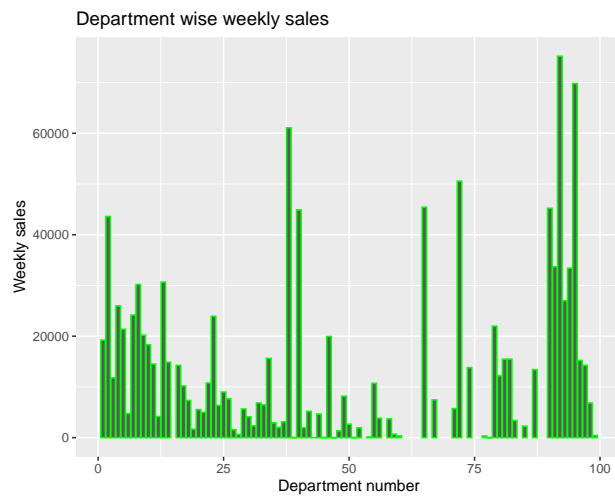
Distribution of the weekly sales, log transformed:
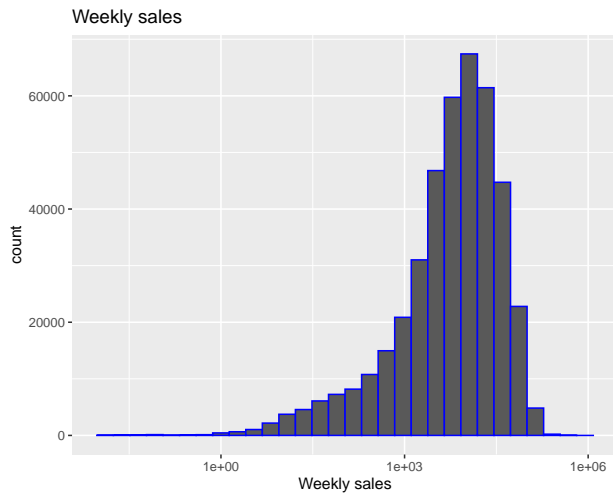
Weekly sales

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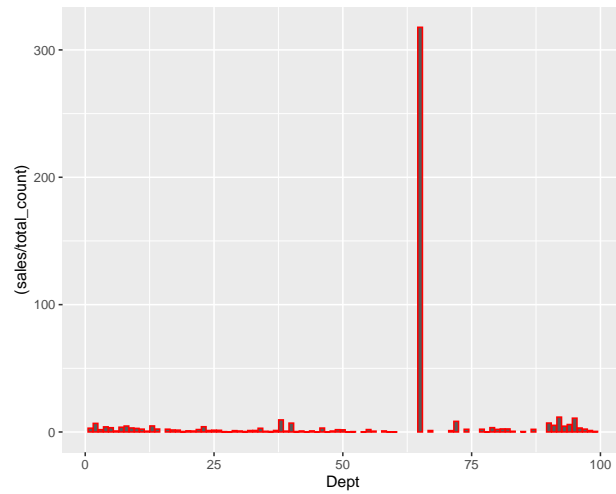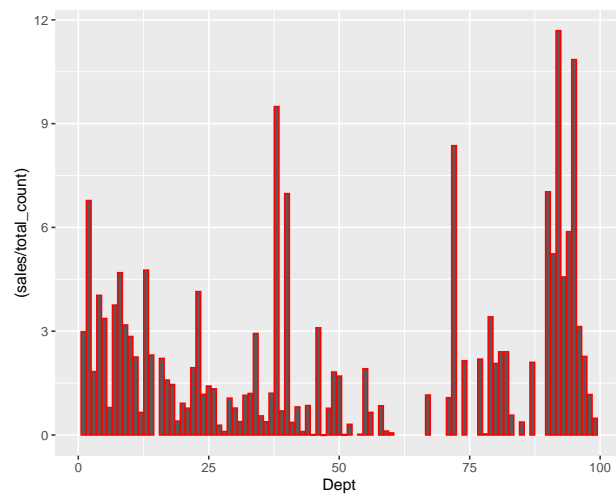Number of holidays among the 115064 days in dataset:

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

Holiday sales

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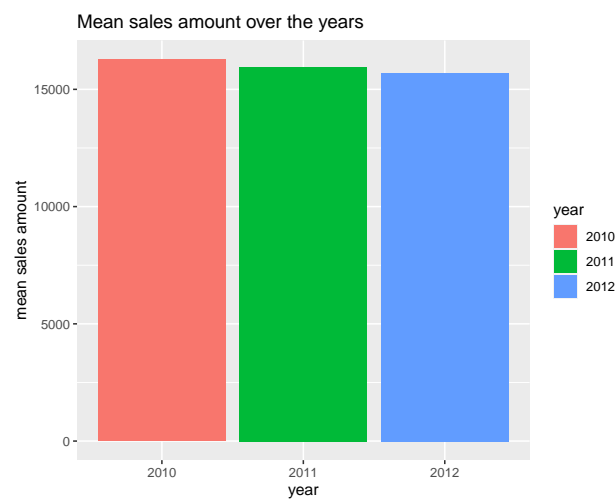Department wise weekly sales:

Department wise weekly sales

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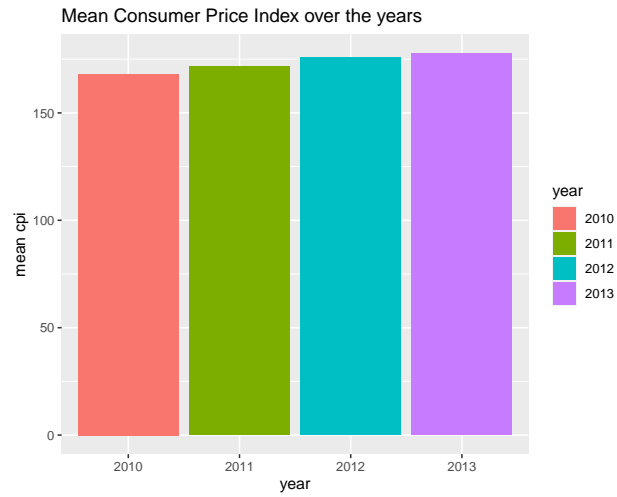Department wise mean sale amount:

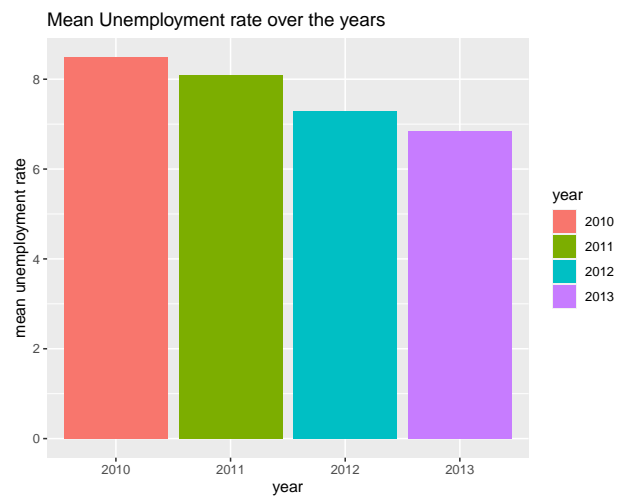Department 65 has unusually high values, so let's remove it from our graph:/

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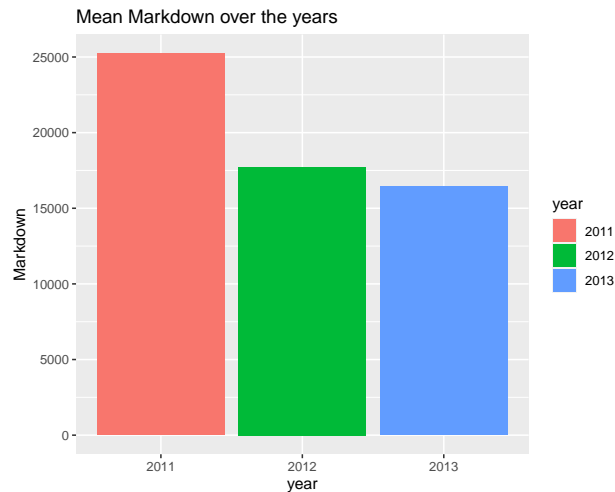Let's view the mean sales amount for the three years:

Mean sales amount is almost same but slightly decreasing over the years.
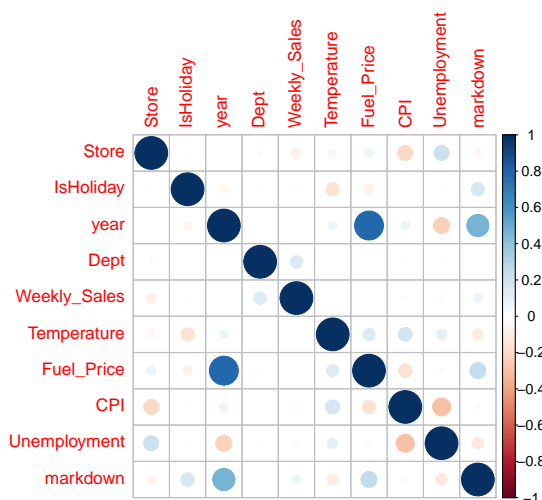Variation of mean CPI over the years:

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

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.

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.

A dashboard of visuals will be created using RShiny,and a tentative list of graphs to be plotted is provided below :

* A panel for plots of the four main holidays(Super bowl, labor day, thanksgiving, Christmas) vs weekly sales:

+ Bar graphs : year vs sales for each holiday.

+ Bar graphs: comparing total sales of all holidays.

+ Bar graphs: comparing mark downs of all holidays

* A panel of barplots: sales per year for each department.

* A pie chart showing sales distribution over the months.

* Histograms : to show the distribution of sale prices,holiday sales,markdowns.

* Bar plot: Number of observations per department.

* Bar plots: mean sales amount,unemployment rate,number of markdowns, mean markdown amount,mean consumer price per year.

* Line graphs: fuel price,markdown,consumer price,temperature and unemployment vs sales amount.

The dashboard will help us view our data in a consolidated format at one unified location. The interpretations drawn from the graphs will help 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.

* 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.