Walmart Sales Analysis

Swathi Badicole
Information Visualization,Department of Data Analytics
San Jose State University
ORCID:015258360
California, United States of America
Email ID :swathi.badicole@sjsu.edu

Mounica Bachu
Information Visualization, Department of Data Analytics
San Jose State University
ORCID:015262520
California, United States of America
Email ID : mounica.bachu@sjsu.edu

Aishwarya Mohan Iyengar
Information Visualization, Department of Data Analytics
San Jose State University
ORCID:015269371
California, United States of America
aishwarya.mohaniyengar@sjsu.edu

Abstract—**Visualization helps to understand the results of data analysis quickly and concisely in the entire process of collecting, cleaning and analyzing data. Starting with Tableau installation, our project describes the process of understanding and visualizing Walmart sales data. The aim of our project is to load, transform and analyze the Walmart sales data by creating an interactive dashboard on tableau which provides insights on yearly sales, monthly sales, weekly sales, department wise sales of Walmart stores. This helps to give cognizance on overall performance of the stores. This can help management on decision making to improve their sales.**

Key words: Data visualization, Tableau, Analysis, Walmart Sales.

I. INTRODUCTION

We are all very curious about our future! Very excited to know what will happen with us the very next moment, tomorrow, similarly retailers will also be curious in knowing their business and its progress. Visualization helps to understand the results of data analysis quickly and concisely in the entire process of collecting, cleaning and analyzing data. By adopting certain aspects retailers can avoid losses in sales. The efficiency of understanding profits and losses in various departments through common steps is not particularly possible because data will be huge.

According to [8], Walmart was first opened in 1962 as a single discount store by Rogers over the intention of selling more items with less price as their tagline "Everyday low prices on a broad assortment- anytime, anywhere" and their brand of "Great Value" suggests. Walmart launched its initial public offer in 1970 at a price of 16.5$ per share which slowly helped in the stores expansion of Walmart. In the process of expansion, Walmart started to open new format stores such as Walmart Super Centre and Sam's Club and established the stores and few acquisitions in Mexico, Canada, United Kingdom, Chile, India, South Africa, Columbia and China apart from the U.S. In 2000, Walmart added the online format shopping available to the customers. In 2002 Walmart was at the apex of fortune 500 ranking. As per 2015 statistics employment power crossed 2 Million around the world along with serving 200 million customers each week in 27 countries with a magnanimous amount of 11,433 stores.

Walmart[9] claims being the largest onsite green power generator in the United States. Some of the notable brands by Walmart include Great Value, Hayneedle, Moosejaw, Bonobos, Eloquii and Art.com. [10] states that Walmart is expecting 559 billion dollars as revenue of 2021. As per [11] walmart has 4,743 stores in the U.S. along with 599 Stores of Sam's Club and 6,101 Walmart International stores which concludes upto 11,443 stores overall.

Walmart stores sell the household products and obtain profit by that. This store has several subsidiaries in various geographical locations which is difficult for retailers to track the business profits and understand the customer needs. It is highly complex to analyze and drill down the sales for each store with respect to the department during a period with the help of traditional methods using excel at that particular location and during special occasions the rate of sales might go high or even low so, in that particular store the communication between analytics insights and team might help in understanding where the problem is and why the sales are going down, and can be fixed in that store. Based on the data we can observe the sales pattern of all products from each department of a store which can be used to predict sales in future to help the management to keep track and to initialize the key performance indicators to increase the profits and to analyze supply and demand for certain departments.

During the holiday season the sales in particular departments like clothing, footwear etc. and during the black friday sale, sales for electronics may be high. It is observed that sales of products from different departments is not linear, based on the analysis of historical data we can give insights regarding the future sales, to help the management to keep track of their supply and demand based on the profits each product provides.

In our project we collected Walmart sales data of 45 stores for a period of 3 years. We have explored the factors contributing these sales by analyzing the dataset and to provide related graphs using Tableau as a visualization tool. We implemented an interactive dashboard and created visuals of sales among different departments in Walmart stores. This Project provides insights on yearly sales, monthly sales, weekly sales and department wise sales of Walmart stores.

II. RELATED WORK

In 2015, Harsoor and Patil [2] worked on forecasting sales of Walmart stores using big data applications: Hadoop, MapReduce and Hive so that resources are managed efficiently. This paper used the same sales data set that we utilized for analysis, however they forecasted the sales for the upcoming 39 weeks Their strategy included the collection of huge Sales data and transferred on HDFS [3] and performed Map Reduce which later due to enormous data size, proved difficult to draw conclusion. Thus Hive processing was done to calculate average sales features for all 45 stores and 99 departments. Machine Learning algorithm, R programming was used for statistical computing. Henceforth, Holt winters [2] was used for training a dataset provided by Walmart and then sales prediction was done. Subsequently the predicted sales were given graphical representation using Tableau interactive data visualization.

In 2013,[4] A.Katal, Wazid, and Goudar performed thorough studies about handling Big Data; their issues, challenges, various tools and good practices. Technical challenges like scalability, fault tolerance, data quality and heterogeneous data processing was mentioned. They have proposed a Parallel Programming Model Like Distributed file system. Mapreduce[5] and spark as a good tool for Big Data.[4]

A study conducted by Orbis Research[6] claims that tableau along with Zoomdata, JOS, Sisense, Periscope Data would still be considered as top players in the Global Data Visualization Platform even till 2026.

M.Diamond,A.Mattia from their study of "Data Visualization: an exploratory study into software tools used by businesses"[6], where they compared key tools used for data visualization such as Microsoft Excel, Microsoft Power BI, Tableau and IBM Watson Analytics concluded that it all the business schools should focus on how data visualization can be helpful in taking strategic business decisions in their business education curriculum. They claimed that when they were creating the dashboard using Tableau, it was similar to the excel functionality, but tableau was easier to work with on the graphs.

In the year 2018, 5 authors namely Alper Sarikaya, Michael Correll, Lyn Bartram,Melanie Tory, Danyel Fisher joined and did research on dashboard in tableau. This literature review suggests a various number of fruitful research directions to better support the dashboard design, implementation and also, they have explained the proper way for the end user to use it. We mainly got an idea of the spacing that should be done in a standard dashboard creation. Visual features are explained for the new dashboard creators to know the standards to be maintained like construction and Composition, Multipage, Interactive Interface, Highlighting and annotation, Modify data or world. They have also discussed a lot about the representation of different data in different ways like social data as context, sales data as context etc. Each type of data has its way of representing it making it understandable to the end users who are using the dashboard.

Recently during April 2020, Michael Correll, Enrico Bertini and Steven Franconeri [5] worked on research which included finding an answer to the question. Truncate the Y axis, threat or menace? As we have the aim of doing the dashboard with many visuals, We have referred this paper. The Bar chart is more highlighted in this research paper.Bar charts with y axes that don't begin at zero can visually exaggerate effect sizes. Hence this paper helped us to get more examples of visuals where this y-axis truncation can be beneficial as well as harmful, depending on the communicative and analytic intent. It also gave ideas by showing the results of a series of crowd-sourced experiments in which we examine how y-axis truncation impacts subjective effect size across visualization types with alternative designs that more directly alert viewers to this truncation.

III. DATA SET

As we all know, in a data visualization project data plays a vital role. One of our findings while referring to the related research papers that if the scale of the dataset is small their analysis will not be meaningful and much applicable. Hence, to find a relevant dataset for our study.

As we have chosen to do the interactive dashboard for the Walmart sales analysis, we wanted a dataset which contains the details of the sales happening.

We were able to locate an appropriate data from kaggle https://www.kaggle.com/c/walmart-recruiting-store-sales-fo recasting/data

Dataset consists of information about weekly sales of 45 stores along with 99 departments for a period of 3 years.

The departments definition was missing in the dataset, but the department 1-99 are grouped as below[13]

In addition, store and geography specific information such as store size, unemployment rate, temperature and promotional markdowns as present in the data.

Column	Definition
Store	Store Number (1 - 45)
Dept	Departments of the store (1- 99)
Date	Date of the week and All the sales for the week were reported on friday of each week. (yyyy-mm-dd) Range: 2010-02-05 to 2012-11-01
Weekly sales	Sales for each store was reported cumulatively once in a week
isHoliday	It is a boolean value which specifies if the week has any special holiday in it. Holidays included in the data set are: Super Bowl, Labour Day, Thanksgiving, and Christmas
Temperature	Average Temperature of the region is reported per store.
Fuel_price	Price of the fuel in region
CPI	Consumer Price Index
Unemployment	unemployment rate of the region
Type	Type of store. (A (Super centre), B (Discount store), and, C(Express store))
Size	size of the store
MarkDown 1-5	Data of the promotional markdowns. (data is available for only few dates after 2011 November)

IV. METHODS

A. Data Collection:

Dataset obtained from kaggle contains Walmart sales data divided into 5 different files namely, Features, Stores, Training data, Testing data and Sample Submission. Data which is important for data visualization has been considered and joined such as features, Stores and Training

data file to create a dataframe using pyspark in jupyter notebook.

B. Data PreProcessing:

The data frame has been checked for duplicate values and we found that none of the data has been duplicated. Missing values for each column was calculated, based on the results we observed that there were no missing values present in any of the columns. During the processing of the data, we have observed that some of the sales values were negative and zero in nature, these values were filtered out using spark-sql. All the pre-processed data was exported to excel using openpyxl and the same file has been loaded into Tableau.

Once the data has been loaded into Tableau, the following Dimensions and Measures was formed.

Tables

- 🗓 Date
- T|F Is Holiday
- Abc Type
- Abc *Measure Names*

Fig 1   Dimensions

- \# CPI
- \# Dept
- \# Fuel Price
- \# Size
- \# Store
- \# Temperature
- \# Unemployment
- \# Weekly Sales
- \# *Sheet1 (Count)*
- \# *Measure Values*

Fig 2   Measures

```
root
 |-- Store: integer (nullable = true)
 |-- Dept: integer (nullable = true)
 |-- Type: string (nullable = true)
 |-- Size: integer (nullable = true)
 |-- Date: date (nullable = true)
 |-- Weekly_Sales: double (nullable = true)
 |-- Temperature: double (nullable = true)
 |-- Fuel_Price: double (nullable = true)
 |-- CPI: double (nullable = true)
 |-- Unemployment: double (nullable = true)
 |-- IsHoliday: boolean (nullable = true)
```

C. Visualization Design:

Retailers plan to insure success or maximum profit by learning about the factors that affect their sales. Big organizations and retailers around the world such as the one this work is based on Walmart Stores, Inc., try to maximize the profit by providing maximum customer satisfaction in all geographical locations to maintain the standards of the stores.

Walmart Sales data is considered for this work since most of the challenges faced by the company are universal or that all other big retailers are facing similar problems that is to maintain, manage and organize their retail shops data in a way that it provides useful insights on the company as an overall retailer. The retailers have to overcome the challenges to stay on the top. So, by knowing the issues that are affecting their business and to avoid losses we are aiming to analyse the data of Walmart sales and create an interactive dashboard which provides insights on Yearly, Monthly, Weekly and Department wise sales. Tableau has a function which can divide Date automatically into hierarchy. As this function gives time scales such as year, month, quarters etc. This will be a very useful feature in tableau that helps us to analyse data of Walmart sales and create visuals.

⊟ YEAR(Date)	⊟ QUARTER(Date)	⊞ MONTH(Date)

For Walmart Sales dashboard, we have decided to use this hierarchy and analyze monthly, yearly, weekly and quarterly sales. We have data visualized to comprehend the pattern of weekly sales for all 45 stores across different locations observing the years from 2010 to 2012, fuel price and temperature respectively.

D. Tableau operation:

Our objective in this data visualization project is to build an interactive dashboard. Initially we thought it was not easy to design a single dashboard with multiple analysis. However, with all the factors that need to be considered we successfully met our needs. We have created an interactive dashboard using Tableau.

The first step is loading our cleaned and filtered dataset into Tableau.

The second step is to create a worksheet. We created multiple worksheets of different data analysis like monthly sales, yearly sales, department wise sales etc.

The third step is grouping all worksheets into a dashboard. We have grouped all our worksheets into a single dashboard and created an interactive dashboard where we can filter yearly sales, department wise sales, and holiday vs non holiday sales. We have carefully designed graphs for better visualization.
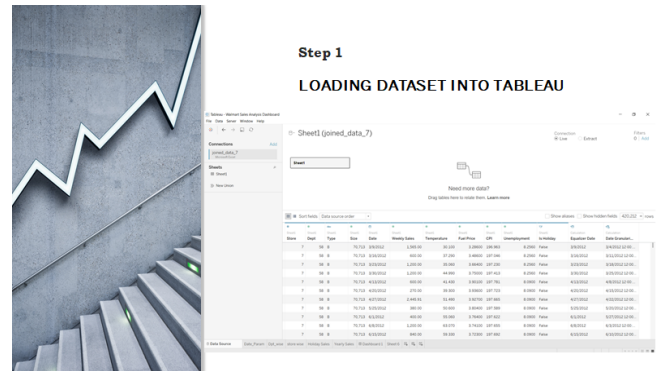
The Steps are shown below :

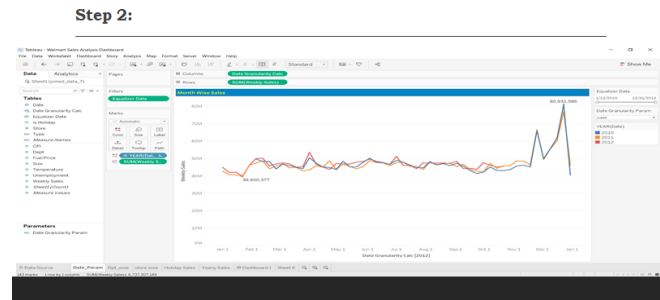Fig 3 Loading Data Set into Tableau

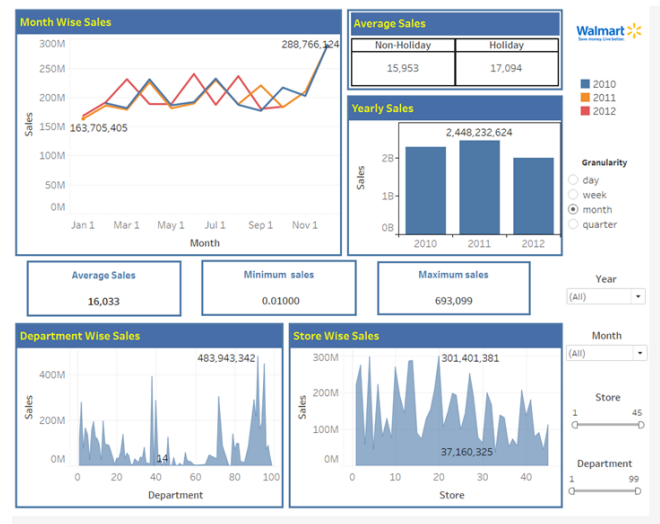Fig 4 Creating Worksheet -- Sample 1

Fig 5 Grouping all Worksheets into Dashboard

V. RESULTS

We have come up with an interactive dashboard which contains 6 different visuals with different filters in it. We have done the Walmart sales analysis using the data from the year 2010 to 2012. The data has the sales data from 45 stores having 99 departments each.

The dashboard will be explained from top to down and they are :

- ❏ Month wise sales
- ❏ Average Sales (Holiday vs Non-Holiday)
- ❏ Yearly sales
- ❏ Three containers having average sales, maximum sales, minimum sales
- ❏ Department wise sales
- ❏ Store wise sales

A. Month wise sales:

In the first visual, we have done the line graph. We wanted to show the month wise on the line chart to show the different values across the year. Hence, we used multiple line charts, which indicated the different years. In the data we had the sales data available for 2010,2011,2012. This line chart has three different colours to differentiate.


The above figure shows how the line chart looks. 2010 is represented by blue line,2011 by orange and 2012 in red colour. We have two filters which are associated with our visuals. As we wanted to make the dashboard interactive, we have used many filters which can be selected by the user like a data granularity parameter.

Data granularity parameter means the parameter which can be used when we want to dynamically change the categories. In our parameter we have used the 4 options namely day, week, month, quarter.

Based on the selection on the data granularity parameter selection in the dashboard the line chart will be displayed. This is the first interactive activity we built. If we select the month, we see the data based on month among all three years. Similarly, we can see the difference when we select the week and quarter. We have also added a filter called Date. This is a slider which can select the dates in range and we get to see the data for the range selected in the line chart. From this we were successfully able to implement

interactive line chart visuals. Below is the parameter details shown with the way it reacts to month and quarter selection.
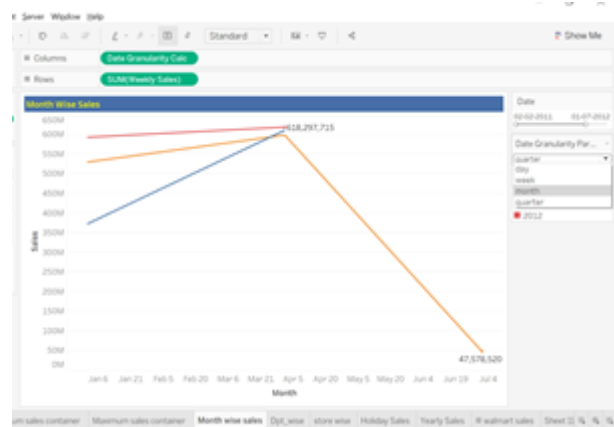
B. Average Sales (Holiday Vs Non Holiday)

We all know that, for the sales to take place, the day also matters like if we can compare the holiday vs non holiday. People tend to shop during the Holiday seasons more than Non holiday. We wanted to do analysis of the Average sales with the consideration of Holiday and non holiday. The festival season like thanksgiving in the time of November shows that there are more sales compared to any other days. This comparison will help the manager to defend himself from the leadership to show the performance of the store has not gone down even during the holiday seasons.

There is a filter of month, so this filter helps to filter the data flowing into the table of non holiday and holiday average sales. If we do a selection in the month , it filters the data for that month and then displays the value. For example, if I select the month jan and there is no holiday seen in that month, we will be just able to see average sales under the Non Holiday section of the table.

Below two visuals shows how the visual acts when there is holiday in that month and no holiday at all. The first image shows the holiday and non holiday but the second figure we can see we have just selected the month January which shows just the Non holiday average sales as there is no holiday seen in that month.

C. Yearly sales

Bar chart is one of the common visuals used by everyone as it is easy when we will have to do a comparison. As we have to plainly show the yearly sales between 2010,2011,2012 we choose the bar chart which helps us to do simple calculations. As we have done the dashboard keeping in mind the manager of the Walmart store. This will easily help him to show the yearly sales between the years and analyse for himself by comparing it with a selection of the year he needs. We have the data across 45 stores with 99 departments each. So if the manager wants to check the yearly sales for just store number 1 and department 1-10, we can do this using visuals. The visual has three filters like year, store and department. The year has the selection between three years 2010,2011,2012 , the store and department are the sliders which help us to select the required filter to get displayed on the bar chart for comparison. The below figure shows the bar chart visual with its filters implemented and how it displayed the yearly sales happened for the particular filter selected. Here we have shown a filter where there is all department in just store number 1 for three years.

D.Containers

Containers are the type of visuals which can be used in tableau to just display the KPI values. Here the KPI we are considering to display in our dashboard are the maximum sales, minimum sales and average sales happened. We have created three containers containing these values in them. We have created interactive containers which change dynamically as we do the filter selection. Containers can also be static which contains just one value and does not change. But here we have chosen to do a dynamic container which will be helpful for the manager who is our targeted end user.

If the manager wants to check the minimum sales and maximum sales, it can be done easily by seeing the container card. The filters which we have considered for these visuals are store, department, year and month. So here we see that the manager can select the store and department he needs to check with the year and month. Now if the manager wants to see the maximum or minimum or average sales in the year 2010 and month march for some particular store or department, the container will check the filter selected and compare the values which are maximum and minimum and display inside the container card.

There are three different containers done with different sheets. The figure below shows how it looks in the dashboard.


In the above figure we can see the container displaying maximum sales in all three years (2010,2011,2012) January with store 1-45 and department 1-99. After selecting the

filter, it sees which is the maximum sales happening within the filters selected and displays on the container.


E. Department wise sales

We have the data for 99 different departments like kitchen, accessories ,clothing etc. We have done a visual called area chart. As there were 99 different departments to be compared with each other, we have used an area chart which is continuous in nature which shows how the graph is raised from one department to another. We have the sum of sales in the rows and the department(which contains 1-99) in the columns. So the area plot visual is created to do the same. We have also enabled the label in the visual for the maximum and minimum value. The chart is self explanatory and the user will be able to view which department has what sales. This visual also has a year and month filter where we can select which year and which month department sales the users need to check. The figure below shows us the are chart plotted for all the departments in the year 2010 with all the months across and we could observe that department 92 among the filter selected has the maximum sales and department 47 has the least.
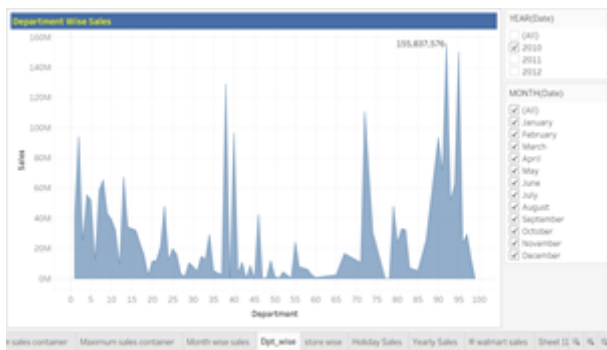

F. Store wise sales

As we have seen the department wise, similarly we have 45 stores across which we are going to compare. So the area chart visual is created similar to the department wise showing the stores(1-45) in that area plot and we will be able to compare between them. The Store wise area chart also has the same filters applied to it (year and month) if we select the year and month we will be able to check the sales happening across all the stores in a continuous area chart according to the selection made. All the filters used in our dashboard are multiple select dropdowns , where the user will be able to select himself the required selection and get the value. For eg the end user just wants to see the sales happening in year 2012 January, he will be able to select the same in the filter and he will be able to get the area chart visual which shows the sales across 45 stores in 2012 January. As we have seen an example. The figure below shows the store wise area chart visual when the store manager wants to check the sales between all the 45 stores in the year 2012 January.
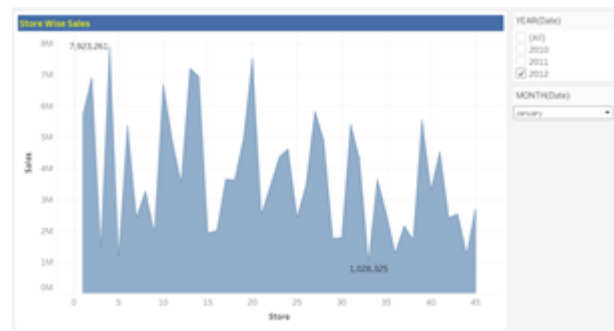
G. Additional Visuals

We have seen many visuals in the explanations above which we were able to build in the data set which was available. We have put up all the important visuals on the single dashboard named Walmart Sales Analysis. Apart from those visuals we also have additional visuals like type vs sales analysis. This is a store type which are A,B,C Vs the Sum of weekly sales. We considered checking the store type and its weekly sales, but we did not find it more productive or informative for the end user we are targeting, hence we did not add in our main dashboard.

In the below visual we see the bar graph which plots the sum of weekly sales versus the store types. Each store represents each colour. The visual is also self explanatory.
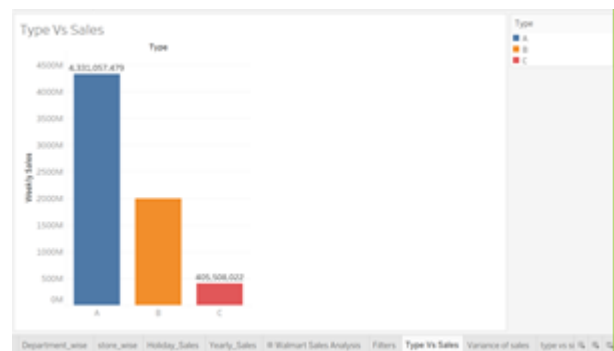
There is one more visual which says about the store type. The store type Vs size of it. We can see the size of the store with respect to the store type. Below figure shows the same.


Temperature was also one of the extra factors which we had found in the dataset. Even though it's not much to say about in our sales analysis, still he is a variation that we can find

in the sales based on the temperature. Practically, if the weather is good, people tend to come out for shopping. So here keeping the temperature in mind. We have tried the below visual which is the bar chart again which has temperature and sum of the weekly sales.


As there was a lot to show about the sum of weekly sales, we thought even the variance would play a vital role in the plotting of graphs. So the below bar graph was visualized which has store (1-45) vs the variance of the sales happened.
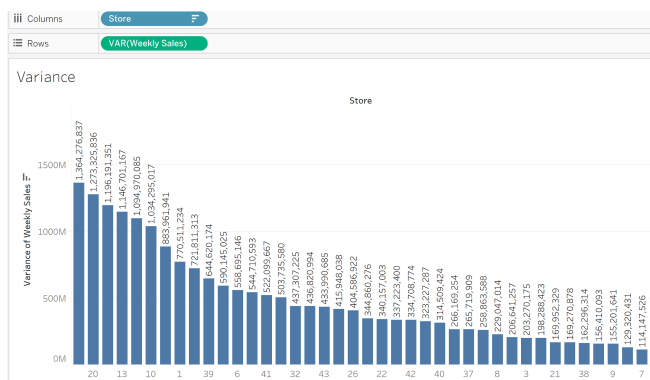
Considering all the visuals, we have chosen the best for our knowledge in order to get an interactive dashboard for the Walmart Sales analysis with the data for years 2010,2011,2012 for 45 stores and 99 departments.

As a part of future work we had considered to improvise this dashboard for all the Walmart stores and also increase the end users from just the store managers to the leadership level. At this point, we also thought to use a link for the online website of Walmart. If there is any issue happening about the price of a product and the cashier calls the manager, he will be able to directly go to the Walmart online website to check for the price of the product. Hence we have Implemented a Walmart logo on the dashboard which takes us to the Walmart website on the click.

H. Dashboard:
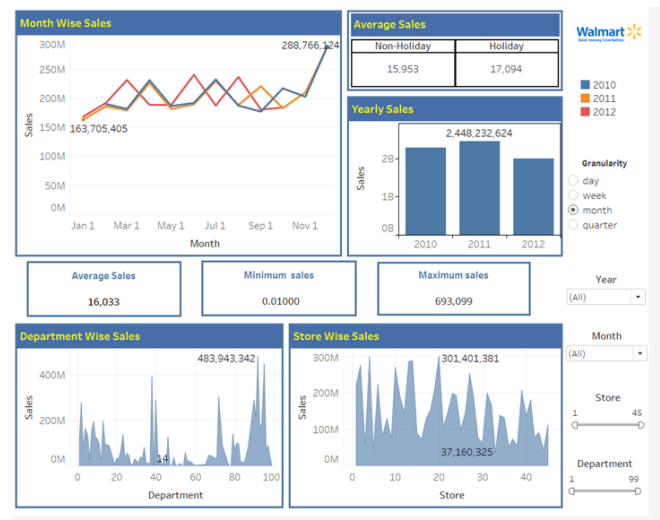
So, Finally the above image shows our interactive dashboard where you can select year wise and monthly wise as well as weekly sales. We have added filters on the side to make it interactive, so that retailers can select which year or monthly or week they want to analyse their sales.

VI. DATA ANALYSIS

- When we observed the pattern of sales between holiday weeks and non holiday weeks, we found that average weekly sales during thanksgiving and christmas were higher than the average weekly sales, the same can be observed from the average sales container in the dashboard.
- Being the only year with a complete data of all months, Walmart sales of 2011 reports more sales than 2010 and 2012. Within the sales, December month shows the highest sales among all months.
- During the analysis, Type A stores i.e super centres performed better than B and C.
- We cannot establish weekly sales to be in linear relationship with temperature, but higher sales were observed in the regions with temperature between 70 and 80 Fahrenheit.
- Based on cumulative sales, we observed that store 20 reported highest sales and 33 reported the minimum sales among all 45 stores.
- Based on the store wise analysis of variance, store 14 showed maximum variance compared to others and store 5 showed the least variance, we can say that store5 5 sales were more static in pattern.
- Type A stores were larger in size and type C stores were smaller compared to other two types of stores.
- November and December data for 2012 was not given, so quarter four sales of 2012 is inaccurate.
- Department number 92 i.e. grocery department which is used by everyone irrespective of the seasons, showed the maximum sales and department 43 showed the minimum sales compared to other 98 departments.

VII. DISCUSSION

There are many challenges in retail store network Planning. Retailers might fail in evaluation of the potential of the market. Sometimes they might ignore the competition in the market.

Our project on Walmart Sales Analysis expectation is to make retailers understand the root cause behind their sales and implement strategies accordingly to make profits out of their business. If the retailers understand the history of sales and cause that affects they can fix the issues and increase their sales. We hope our suggestions could help Walmart retailers in making profits and understanding customers interests through analysis of their sales.

By looking at the first graph in our dashboard retailers can easily understand how their monthly sales graph. We made this dashboard interactive so that retailers can select some particular year and analyse their monthly sales in that year. For example if they select year as 2010 their monthly sales analysis is different when compared to 2011. From our graph when I select year as 2011 the sum of sales in february month is less when compared to january but later in march it has gone high. So this is how retailers can analyze their sales every month in a particular year to know their business.

We have analysed Average sales, Maximum sales and Minimum sales of 45 stores and 99 departments. Retailers can check their average sales of any store among the 45 stores. This might help in understanding the average sales of different stores and analyse if any particular store's sales is going down.

From our second graph it is clear that sales in the holiday season versus non holiday season vary. So, retailers should understand which department sales are high in the holiday season and non holiday season. Our analysis helps retailers to compare their sales in both seasons.

We also analysed store wise analysis of 45 stores and we found that the sum of sales among different stores is varying. Likewise, there are 99 departments in the stores in which the sum of sales in every department is different. So, we are showing the difference and our analysis of department wise and store wise which might help retailers to understand the cause.

To summarize, we would like the retailers to notice the cause in their sales business. From our visualization, we have highlighted that there are ways to increase sales by analyzing the problem. There is a saying "For every problem there is a solution", So retailers can analyze the problem from our visualization to know the problem and solve it by understanding the customers needs and interest and increase their sales accordingly. For all the retail companies, customers are the god so, to increase their sales understanding customer interests and analyze their sales on a day to day basis will help them to grow their business.

VIII. FUTURE WORK

As seen in the dataset part mentioned above, the data available for the public is quite less. It has stale information which includes just 45 stores with 99 departments from 2010 to 2012. With data from more stores of recent times could yield accurate analysis information to predict sales for future. The geographical information has to be added to the dataset which can help Walmart higher management to decide in which locations they should establish new stores and to know which locations are majorly contributing to the profits and which locations are causing less profits and to focus on underlying issues.

Though having all the information of all the products will be huge, but with the information of products could lead to drill down of the sales per product in a department of a store in the given timeline. This not only will help the management to take decisions of either to increase the supply or to decrease the supply but it also helps the supervisor of the store to track and decide on the number of products should be present in the display as well as the number of products present in the warehouse including how frequently they should add the new stock to the shelves. The data we obtained was present in different excel sheets but for the future purpose storing the data of multiple stores, departments and products in databases like mongodb helps in decreasing the extraction, transform and retrieval time of the data. by applying tools like Tableau, PowerBI could help in giving insights from the data.

IX. WHAT WE LEARNED

As a first step, we have decided on working on forecasting sales of Walmart in our visualization project. Our target is to analyze the sales of Walmart stores and make retailers' lives easy in understanding their sales.

From the beginning of our Visualization Project, we are engaged in all the steps during our project work with utmost interest and enthusiasm in creating visuals that would help retailers in growing their business. As a first step in our project we found the dataset that best suits our plan of project. We acquired the dataset from kaggle. We found four files on kaggle which will be helpful for our project but here the tricky part we faced was we have to merge all four files so that it will be perfect fit for our visualization work. So taking into consideration merging files which are important to our project we used two tools that are pyspark in jupyter Notebook to create a data frame. We checked this dataframe for any duplicate values and found that there were no duplicate values in the dataset. While checking for any null values or negative values in the dataset we found that there are few negative and null values existing in the dataset. Using Spark-SQl we filtered out all those negative and null values. After all these cleaning and filtering we have a dataset that consists of 4,21,570 rows and 12 columns of 45 stores and 99 departments. This cleaning and preprocessing has been the most time consuming part of the project, Because any minor mistake in this step might affect

our visualization as we thought of analysing real time data of 45 stores sales.

The next step after cleaning and filtering out our dataset is loading the data into Tableau. So, the next critical and time consuming step in the process is extracting valuable information from our dataset and implementing graphs to demonstrate our analysis of Walmart Sales. This was the challenging task as it plays an important role in the whole project. Using Tableau for our project to create visuals is most convenient as we have to just drag and drop the dimensions and measures to create graphs. The time scales function also helped us a lot in filtering years, weeks and monthly to analyse our data.

We learnt from our class the usage of Tableau tools and to better display of our results using shapes, text and colours in our dashboard. We have learnt how to make an interactive dashboard. Also, creating sheets with different graphs and merging them into a single dashboard. All these things helped us in understanding this Visualization tool called Tableau in depth. This Tableau helped us to create some better visualizations for our project. We have implemented all our learnings from class and created an interactive dashboard for our project The important learning from this project is analysing sales to know and understand their business. We expect that we helped retailers to some extent in understanding and analyzing their sales.

X. NOVELTY OF THE IDEA

In our literature survey we found that there is no literature that is entirely concentrated on sales. We have created an interactive dashboard for our analysis focused on sales of Walmart which is entirely interactive in nature. We did not find any literature that created an interactive dashboard of sales analysis rather they were only focused on getting insights from the data such as, in the fourth quarter which store has more growth. With our visualization we made it easy for the retailers to analyse the sales of their store by making it interactive in nature where the number of insights we get is not fixed in nature and can keep changing depending on the user's requirement . Retail companies get profits by attracting customers with deals and offers, but when they are in loss to know the cause they need to know the history of sales like seasonality trends and randomness. So, our interactive dashboard makes this analysing easy for the retailers to understand their history of sales.

XI. CHARTING

- We made sure that our chart was easy to understand for color blind people as well. This feature was checked using color blindness simulator [12] in trichromatic, Anomalous Trichromacy, Dichromatic and monochromatic view.

- The graphs in the dashboard are created to trigger a long term memory in the brain as all of the visuals and containers are associated with meaning.
- Some of the features in our dashboard implement Gestalts laws of perceptual grouping such as,
    - Proximity: Sales analysis is performed by organizing into subsets, which give different insights about sales.
    - Closure: All the details of the figures are mentioned such as axis name, axis mark, title, filter of the data and what each color represents in the visuals e.t.c.
    - Similarity: All the visuals are designed in such a way that every visual looks as part of the family and not otherwise. All the visuals are mainly designed with the colors of Walmart logo.
    - Good Continuation: Visuals are placed in the dashboard in such a way that continuity exists between the visuals such as minimum sales is followed by average and maximum sales and department wise sales is followed by store wise sales.
    - Symmetry: We have used different types of visuals such as bar chart, area chart, line chart and containers. Symmetry exists in such a way that all containers have similar shape and dimensions. Similar kinds of visuals such as department wise sales and store wise sales are symmetrical in nature.
    - Periodicity: All the visuals based on timeline have moulded in such a way that periodicity is maintained in them.
- All the visuals are aids in parallel and serial processing.
- All the visuals are placed in the dashboard in such a way that we utilize the dashboard to optimum data ink ratio level. Special focus was given not to make the dashboard too clumsy and cluttered which would increase the user difficulty level.
- As we are dealing with huge amounts of data, if we mark the entire data it would get clumsy, so we have just marked maximum and minimum levels which can easily and quickly be perceived by the user.
- The dashboard's entire space is utilized to the maximum extent as more information would be available for the user to understand and it is ensured that no overlapping of the visuals takes place.
- Colors are chosen in such a way that it will not create any noise.
- Most of the data is conveyed using charts as it effectively conveys the information and all the related information of the charts is mentioned clearly along with it.
- High level of information is placed at the centre of the dashboard and detailed information is placed around it so that first focus should go towards the average, minimum and maximum sales and

supported information is placed in its surrounding for the chosen timeline, departments and stores.

- Dashboard is concised to a single screen and only important information is placed in the dashboard to help in decision making without exceeding the boundaries.
- This dashboard can be perceived by any user irrespective of their expertise i.e expert, intermediate or novice.

XII. REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,

[2] : A. S. Harsoor and A. Patil, "Forecast of sales of Walmart stores using Big Data application," International Journal of Research in Engineering and Technology, vol. 4, p. 6, June 2015.

[3] : M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. Mccauley, M. J. Franklin, S. Shenker, and I. Stoica, "Fast and interactive analytics over Hadoop data with Spark," Usenix - The Advanced Computing Systems Association, 2012.

[4] : A. Katal, M. Wazid, and R. H. Goudar, Big Data: Issues, Challenges, Tools and Good Practices, 2013.

[5] : J. Dean and S. Ghemawat, MapReduce: simplified data processing on large clusters. Association for Computing Machinery, 2008.

[6] : Anita, "Global Data Visualization Platform Market Top Players 2026: Zoomdata, Tableau, JOS, Sisense, Periscope DAta", Orbis2021. https://bisouv.com/energy/2663620/global-data-visualization-platform-market-top-players-2026-zoomdata-tableau-jos-sisense-periscope-data-etc

[7] : M.Diamond and A.Mattia, "Data Visualization: an exploratory study into the software tools used by businesses", Journal of Instructional Pedagogies.

[8]: Sam Walton, " Our History: a Walmart business history", Store opened in 1962. available at : https://corporate.walmart.com/our-story/our-history#sam-walton

[9]: Sam's Club, "Our business: Walmart business", available at: https://corporate.walmart.com/our-story/our-business

[10]: "Our story", Walmart Stores story about them, available at: https://corporate.walmart.com/our-story

[11]: "Our Location Facts", Walmart Stores Locations, available athttps://corporate.walmart.com/our-story/our-locations

[12]: Karen Levine, and Tom Van Beverine, "Color Blindness Simulator", available at : https://www.color-blindness.com/coblis-color-blindness-simulator/

[13]: Kevin Study, "Data Analytics - Walmart Sales Part", available at: https://medium.com/kevin-study-notes/data-analytics-walmart-sales-part1-3b9bb788fefc