****

Walmart Company

**Mahmoud Essam Fathy GABR**

****

**Walmart Main Details:**

* **Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of all, which are the Super Bowl, Labour Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks. Part of the challenge presented by this assignment is modeling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data. Historical sales data for 45 Walmart stores located in different regions are available.**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Walmart focus**

**Walmart focus sure is how to increase Revenue based on her historical data that covers sales from 2010-02-05 to 2012-11-01**

**As her attributes were**

**• Store •**

**• Date •**

**• Weekly Sales •**

**• Holiday Flag (0 for work, one for special holiday) •**

**• Temperature •**

**• Fuel Price •**

**• CPI •**

**• Unemployment •**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**To solve Walmart Problem, we had to solve some questions:**

**a) Which store has maximum sales?**

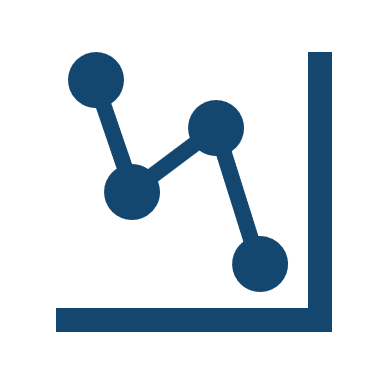
**b) Which store has maximum standard deviation i.e., the sales vary a lot.**

**c) Some holidays have a negative impact on sales. Find out holidays**

**that have higher sales than the mean sales in the non-holiday season for all stores together.**

**d) Provide a monthly and semester view of sales in units and give insights.**

**e) Plot the relations between weekly sales vs. other numeric features and give insights.**

****

**Explaining Work of the project:**

1. **Text

   Description automatically generatedImporting Data and the important modules used in the project:**

**Table

Description automatically generated**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

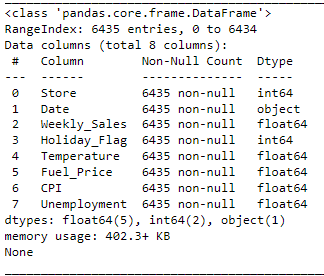
1. Text

   Description automatically generated**Second Step is the to check data from “nullity-duplication-and important statistical data”:**

**Table

Description automatically generated with low confidenceTable

Description automatically generated**

****Table

Description automatically generated

Shape

Description automatically generatedText

Description automatically generated

**☑ Cleaned Data ☑  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

1. **Text

   Description automatically generated with medium confidenceThird step visualize quantitative variables distributions:**

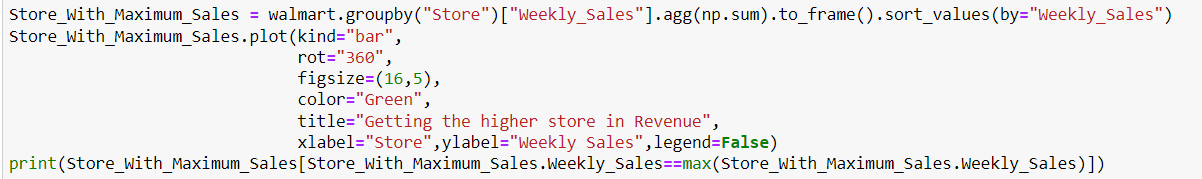
**Chart, histogram

Description automatically generated**

**• Getting Insights based on answering questions •**

**a) Which store has maximum sales?**

**to perform this topic, we need to make pivoting between stores and sum of weekly sales to know who has the most sales between all weeks**

****

Chart, bar chart, histogram

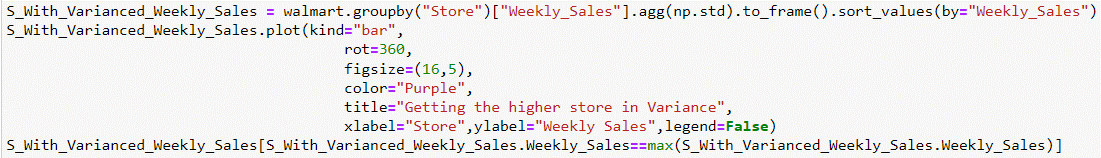
Description automatically generated

**Based on help of Visualization to make sure that result is true**

**### Then the highest in Weekly sales was Store number “20” with weekly sales = 3.013978e+08 ###**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**b) Which store has maximum standard deviation i.e., the sales vary a lot:**

**To perform this topic we need to pivot between Store and the variance in sales**

Chart, bar chart, histogram

Description automatically generated

**Sure as same as store with highly revenue style, You have noticed the store with Highest Variance Based on the Bar photo**

**### Then the store who has a great vary in the sales is store Number “14” with SD = 317569.949476 ###**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**c) Some holidays have a negative impact on sales. Find out holidays that have higher sales than the mean sales in the non-holiday season for all stores together:**

**I can consider this question as the weirdest one in our data based on Holidays and the questions itself**

**So I solved this question in 2 ways to at least get the important insight from my point of view**

1. **To perform that one then we need for Holiday flag, as 0 tends to non-holiday and 1 tends to holiday , and will make comparison between it and the weekly sales, Then get the stores with Negative Impact with Holidays**
2. **Get Real Date from Kaggle and make Compare by equaling it with the same date in data and extract mean of each one then gains the insight easily**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Before Starting about this topic**

**Here is a little simple comparison between**

**Holidays vs non-Holidays**

**Chart, pie chart

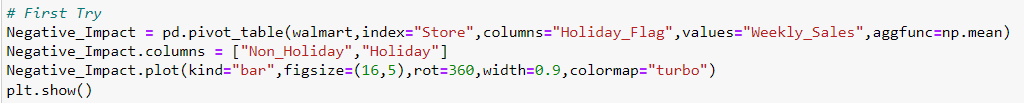
Description automatically generatedText

Description automatically generated**

**So Based in this comparison will see that the difference between revenue in Holidays and non-holidays is about 3.8%!**

**Performing first Try in Holidays Question:**

Chart, bar chart

Description automatically generated 

**Simple view of this Viz has resulted a great insight, that he stores**

**"30,36,37,38 and 44" Have a badly impact in revenue**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Performing Second Try in Holidays Question:**

**Chart, bar chart

Description automatically generated**

**Then Notice again that the "Thanksgiving day" Was the best day in revenue, then "Super Bowl" then "Labour”, and Christmas in last rank**