### Finding the store with the maximum sales

Store 20 has the highest total sales with 301397792$

### Finding the store with the maximum standard deviation

Store 14 is the most unstable store with a standard deviation of 317569.9$

### Finding the Q3'2012 of each store to figure out which has a good quarterly growth rate

Store 4 has the highest sales throughout Q3 with up to 27796792$ in revenue.

However, in terms of growth, is considered a low performing considering its negative growth.

On a different note, Store 7 grew by 11.8% outgrowing all other stores by a large margin.

This is followed by stores 16 and 35 respectively. Additionally, only 10 stores experience a positive growth.

### Can some holidays have a negative impact on sales?

It is found that the average of weekly sales in normal days is 1041256$

Consequently, Christmas sales perform below average compared to the average sales on normal days.

This could be attributed to the overall spending mostly occurring during the first and last

weeks of Christmas, as well as most people spending time with family and friends on other days.

On the other hand, Thanksgiving had the highest average of 1471273$

Overall, the average weekly sales on holidays outperform the average weekly sales on normal days

### Provide a monthly and semester view of sales in units and give insight

Looking at the monthly sales, we can see that the average monthly sales across the stores are highest during July. Additionally, the lowest sales are on January.

The trend aligns with expectations, since July is often considered month of holidays for students.

On the other hand, January is preceded by celebrations and parties taken in December for New Year’s Eve and Christmas.

However, as noted previously we can see the standard deviation of December is also the greatest due to the imbalance spending throughout the month.

# Additional insight looking at the monthly sales of each store rather than of all combined. Interestingly, Store 14 holds the highest monthly sale and as found before is the store with highest deviation.

However, Store 20 holds more spots at the top, as expected considering its found to be the most lucrative store.

Additionally, July is seen multiple times as a top performing month.

Still, April and June are also months performing considerably well too.

# Semester View of sales

Semester two has a total of 95.8 million dollars more than semester one in revenue.

Additionally, it is also higher in terms of average revenue per sale and standard deviation of sales.

Hence, the trend follows the previous findings considering that July and December are both in semester 2.

Similarly, a low performing month like January is in the 1st semester.

### Building a linear regression model

Stores 20 and 4, the stores with the highest sales were chosen (since they are more likely to behave similarly).

It is clear the CPI and Unemployment affect the sales of Walmart.

A MAPE value of 7% is obtained, which is acceptable by industry standards as it is close to 5%

Removing extreme outliers (found after the next approach) improves the MAPE to 5.8%

# Another model is done. In this case, rather than grouping stores together the outliers shall be removed to reduce the number of errors.

Usually, 1.5x values away from the mean are removed. However, this dataset does not contain outliers below (considering its nature it will become negative and non-intuitive). Meanwhile, selecting 0.8\* of the mean kept many clustered outliers, while removing extreme valued ones.

A MAPE value of 65.8% is obtained. This indicates that the grouping of stores according to behaviour is necessary.

# First model is reworked, with extreme outliers removed

It is clear again that the CPI and Unemployment affect the sales of Walmart.

A MAPE value of 5.8% is obtained, which is acceptable