-- 1. Which store has maximum sales?

*This store, Store #20 is the ideal image of an exceptional Walmart retail store in terms of overall performance. In terms of sales, this store has the highest sales over the calculated timeframe - which includes data from May 5, 2010 to December 26, 2012. In terms of the standard deviation, this store is competitive in having the most consistent sales performance all year round, regardless if it is a holiday season or it is the off season. Thus, when we evaluate the overall performance of any other store in this dataset, we can use this store as a clear reference point. Furthermore, we can explain the best conditions for any given store and relate that information to store 20, in order to address any discrepancies which can factor into a given store’s performance. Finally, when we look at store 20, and draw conclusions based on these various factors which make it exceptional, we can apply this information to aid in improving the performance of other stores. In other words, we can match up a store to store #20, and try to replicate the conditions of that store to store #20 in order to identify how we can improve performance.*

*Of course, there are many other important factors which go into this store’s success, which are not present in our table such as customer service, stock and shelving, business and customer relations, customer feedback etc. Regardless, this is still very useful and relevant information to have, provided that we have store #20 as a given reference point in this business example.*

-- 2. Which store/s has a good quarterly growth rate in Q3’2012?

*Best quarterly growth is an indicator of positive change and improvement in performance. Using this data, we can comparatively assess which specific factors go into demonstrating growth. Here, we see that Store#4 has the best quarterly growth rate in Quarter 3 of 2012. However, the question still remains: What can growth specifically tell us about the business? For one, growth can indicate a turn-around in terms of a specific business timeframe as it is relevant to a store’s performance. Two, we can examine this data in order to understand what factors are needed in order to keep a given store’s performance consistent in terms of exceptional sales. Three, we can apply this data in order to understand how we can translate a specific period of growth to overall performance. Thus, we can look at these factors and apply the knowledge we have gained from this information in practical terms. For example, we can specifically point out a column, such as temperature and anticipate the likelihood of higher sales on warmer days, but also, expect a small decrease in sales on days that are above 80 degrees fahrenheit. This information allows us as a business to prepare ahead of time for a pattern of events which can directly relate to sales performance. With the information we have gained from this data, we are able to maximize our ability to efficiently organize and prepare our stores for a number of given specific conditions.*

*Of course, the downside of this dataset is that it includes external factors which cannot be controlled on the business side - such as a rise in unemployment. However, we can again anticipate a decrease in specific stores given this information and act accordingly.*

-- 3. Provide a monthly view of sales for any chosen particular store in any given year.

*A monthly view provides a very specific image of how a store is performing over a shortened time frame. This dataset can even provide us a weekly or even daily view, which can be incredibly useful when specifically looking at a store’s performance during a holiday season, or in the time leading up to a major holiday. Additionally, this data can also identify a drop-off in sales post-holiday at the beginning of the holiday off-season. For our project, we have opted to go for a 12 month view of the average of total sales for store#1 in the years 2010, 2011, and 2012. Now, let’s run through some business insights. For example, we can see that Store one’s performance drops off significantly in January, after Christmas 3 years in a row. In February, March, and April, we can see that the total sales go back up, averaging out to approximately 20503037.9267 per month. That’s nearly a 54% increase in total sales when comparing total sales in January to February, March and April! Here’s what we can gain from this knowledge. Firstly, we can identify an issue: that being that total sales performance in Store#1 in January is consistently lower than average. Secondly, we can look at our other columns and rows and determine what factors go into Store 1’s performance in January, given that January follows two major holidays, Christmas and New Year’s. Thirdly, by using this data, we can try to find the answer as to why this happens, in order to provide our customers the solution that they need.*

*While the data won’t necessarily give us the solution directly, it does help us identify both a specific business need, and a specific customer need, and the information gained from this data allows us to accomplish both. Thus, we are able to better answer the more important questions, such as: What are our customers needs and expectations in January? What can we provide them during this month, specifically, that will answer their needs as a customer? How can we best address our needs as a business, and our customer needs as a client? Having this data is invaluable in terms of being able to understand our customers, and identify what we can do for them.*

-- 4. Which store has the maximum standard deviation? What does the standard deviation tell us, and what insights can we gain from it?

*Standard deviation is a direct measurement of risk, which indicates volatility in the market. The average amount by which individual data points differ from the mean (the central point, or figure on our dataset) helps us determine the spread - in other words, the numerical distance that a value will space itself away from our average calculated value. So, by measuring the spread of values over our dataset, we can point out which stores not only are the most unpredictable in terms of performance, but are also most inconsistent in terms of performance. These stores are our outliers. For example, we can see that store 14, has a significant standard deviation in relation to the mean. In terms of the holidays, store 14’s performance sometimes does well during the holidays (specifically during easter), but also on average, does not do well in November, which is the month of Thanksgiving. Off season performance is equally as unpredictable. For example, there is a 15% drop off in total sales between the months of April and May, but then from May to June, there is an 8% increase in total sales performance, and increase from June to July, and then another 9% drop off in total sales performance from July to August.*

*On the other hand, store 37 has a standard deviation which is consistently in range of the mean sales, and rarely performs any better or worse regardless of whether it is the holiday season or the off season. We can see now why measuring the standard deviation is arguably one of the most important mathematical calculations we can perform with this data set. It allows us to pinpoint the individual strengths and weaknesses in terms of performance of each store over any selected amount of time. We can now determine that we need to use further data in order to improve store 14’s consistency both during the holiday season and the off-season, and on the other hand, we can now try to understand why store 37 does not peak in performance in comparison to other stores during seasonal holidays. This data is invaluable in assessing specific performance needs of any store - and provides the information we need to specifically address consistency, or lack thereof in terms of performance in any given month.*

-- 5. Some holidays have a negative impact on sales. Find out holidays which have higher sales than the mean sales in non-holiday seasons for all stores together.