**Business Statistics – Using Daily Traffic and Sales Information to Identify Trends and Forecast ADS**

Christopher Kozeny

CIST 2500

**Abstract:**

The concept and use of business statistics to forecast and predict future events has long been in use. Analysis of different variables that result from a day of sales can easily help a company better prepare for the next step. It allows companies to predict sales later in the year, when peak sales times will be, as well as what sales tactics may or not be working. I’ve collected information from a local retail outlet. I’ve recorded daily sales metrics from July 2017 until September 2017. The intended use is to be able to analyze whether our Average Dollars per Sale (ADS) is completely reliant on the amount of traffic we receive. Furthermore, do we drive higher sales per capita with a higher traffic volume, or does spending more time with a smaller traffic sample warrant higher ADS per transaction? Many companies rely on this type of information, and it could provide them a more comfortable look into the future.

**Introduction:**

There are many different types of analysis that can be done in a retail setting. Many managers and decision makers would like to know everything about their customers - how they prefer to shop, browse and buy. What drives these customers to make the purchase, and how we can build on and accelerate the process is always a topic of concern that constantly requires improvement. We live in a world where almost every American can browse Amazon or eBay at their discretion and have nearly any item on their doorstep in two days. That puts retailers and small shops in a bind – they must provide service and quality that outweighs the convenience and ease of online shopping. Statistics in the business world are exponentially useful. With statistics, retailers can (relatively) predict sales for the following year, project when the most high volume times will be, as well as receive insight to providing the most efficient and cost-effective service to the consumer. All of these statistics can help drive a better sales team, provide more sales, and squeeze the most out of their bottom line. Companies that don’t analyze their store data - even simple calculations to better staff the location - are losing time and operational efficiency nearly every day. No day of sales can be perfectly predicted, but the more information and inferential statistics you’re armed with, the more prepared you become. Time is money, and companies can’t afford to lose either.

**Data:**

Monthly data compiled in this report have been pulled from the Gretna, Nebraska Oakley store. The data set provided by the company system contains information related to daily and hourly traffic, transactions, average dollars per sale (ADS), total sales of the day and month in total, as well as the amount of units sold (in all categories, sunglasses, clothes, accessories, etc.). For this report and for further analysis, I’ve made use of the daily traffic numbers, ADS, and total transaction counts. The scope of the information used relates to the back end of the summer quarter of 2017. Namely, the months of July, August and September 2017 are included. The entirety of my data has retrieved from the Retail Pro 9 Point of Sale software found on the company owned computers. Retail Pro is a trusted retail management software used in over 100 countries. This system lies at the core of our sales, collecting all of our daily information - sales trends, employee productivity, and even sales by hour and department. The information collected has been aggregated into a single Excel file where most of my calculations take place. I’ve taken the liberty of providing a brief section of descriptive statistics relating to the standard deviations, means, and correlations of said data. Central tendency is also analyzed with the use of the mean, median, and mode. The resulting file will be included in conjunction with this report. Additionally, inferential statistics have been implemented in order to better understand the future outcomes that this data can help provide.

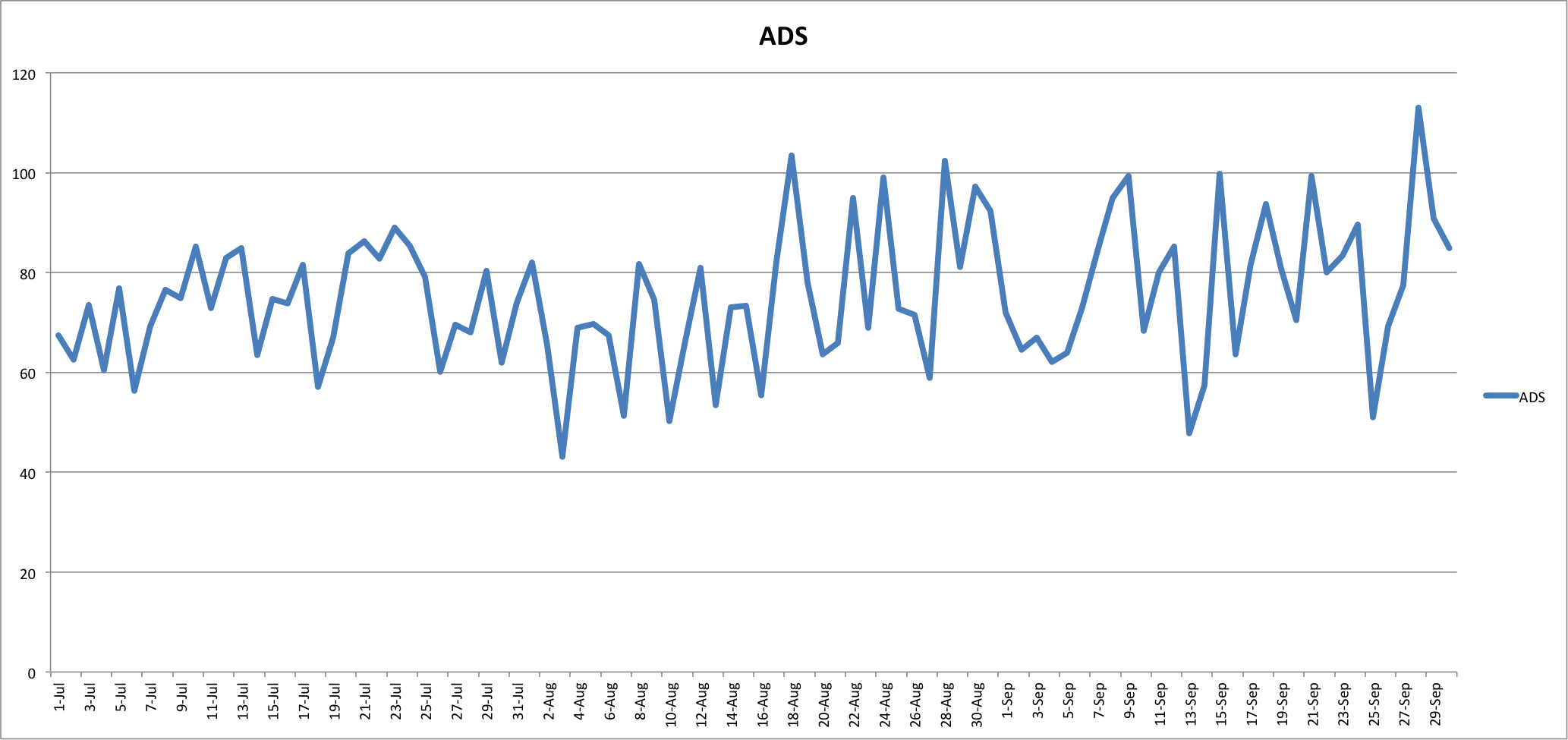
**Assumptions:**

There are a few assumptions that are inherently made here. Regarding the collection and projection of these statistics:

1. It’s assumed that external factors relating to traffic volume and other variables have no affect on the collected numbers. For example, a snowstorm comes through the area, effectively blocking everyone into his or her house for the day. This would clearly affect our numbers. Considering popular in-town events and weather changes can’t be predicted, we rely on and forecast the traffic numbers regardless of any externalities.
2. It’s assumed that all customers act and consume in the same manner. When making predictions regarding the traffic volume vs. ADS, all customers are treated equally. In reality, every customer requires different needs. Some have previously decided on their purchase and only need a recommendation, while others require a half-hour pitch to close the sale.
3. A “transaction” denotes that an item was purchased and a receipt was printed to complete the sale. These summations and calculations do not include returns, exchanges, or any other use of our Point of Sale system aside from a physical purchase with a monetary payment in exchange for our inventory.
4. It’s assumed that we have sufficient inventory to provide sales to any customer that wishes to purchase. Often, the store doesn’t have the correct sizes, the wrong lenses, etc. to fit a customers needs. These are sales that are foregone because of stocking issues within the store. For the purpose of this report and any future calculations, stocking issues are considered negligible.

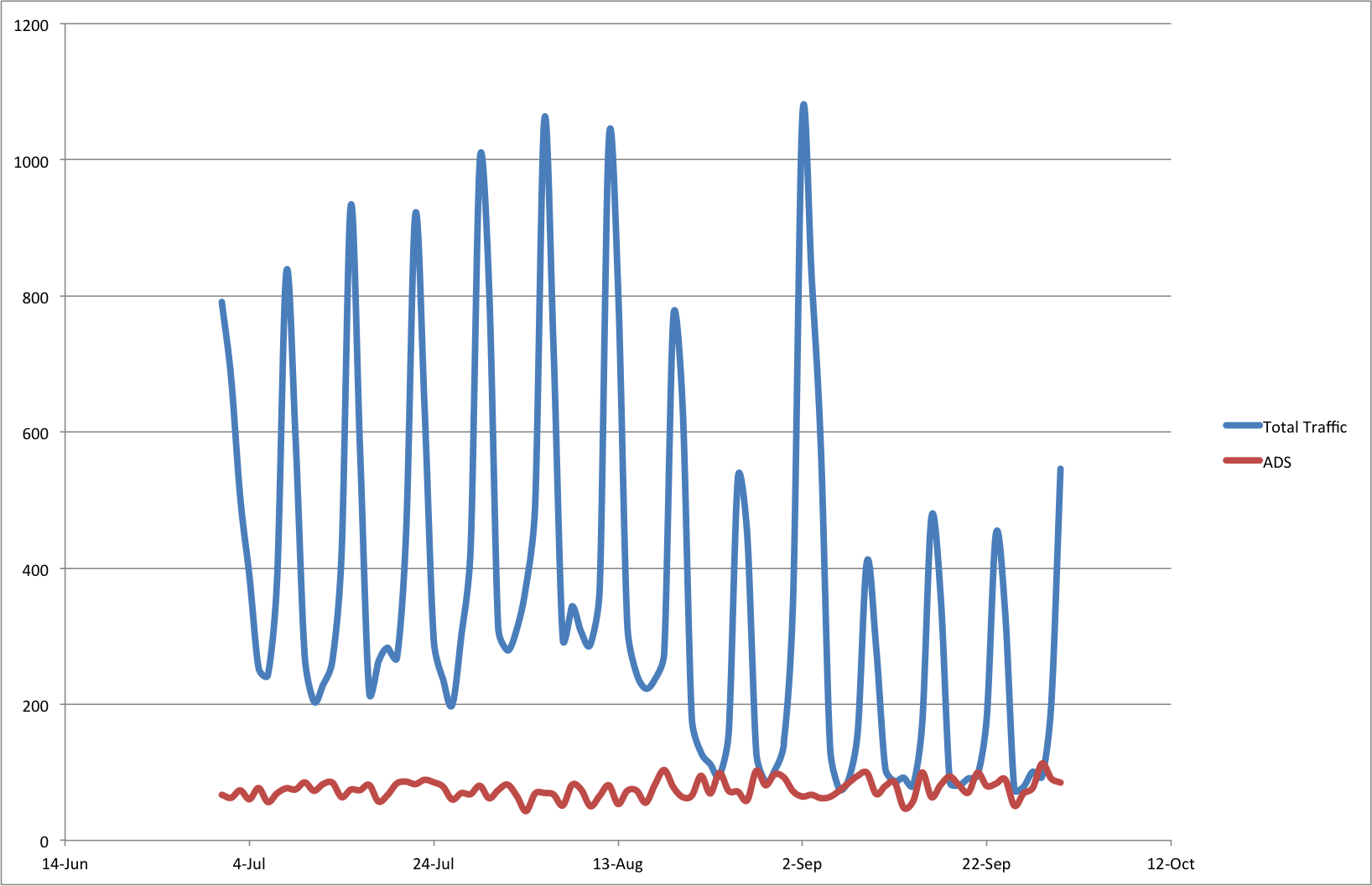
**Data Analysis:**

This plot depicts the average dollars per sale over the quarter. As you can see, ADS is a bit of a volatile variable and will vary immensely over time.



ADS is dependent on many different things – Ideally, traffic being one of them. Traffic may or may not drive a higher ADS, but it’s used to calculate the conversion rate of browsers to customers once they’ve walked through the door. The higher the conversion, the more sales we make in a day. On the contrary, that doesn’t necessarily imply “big” sales that influence our ADS to climb. This could be as simple as 20 people all bought key chains for $3. Sure we made a good amount of sales, but none of them were of quality.

ADS can be described in many ways. Although sporadic, we come upon some interesting conclusions among the data set. Many of the averages across the board are within range of each other. July, one of the big kick-off months of the summer, comes in with an average ADS of $73.59, along with an average traffic volume of 458 customers. This is a perfectly acceptable number for the organization, especially as an overall average over the month. But that brings up the topic of August numbers, and whether these will line up with those of the previous month. August shows some slight variations regarding traffic and average ADS. Over the month of August, the average reported ADS for the month was $73.84. A trivial difference from that of the month before, but August had an average traffic rate at only 371 customers. This implies that even with a lower volume of traffic, ADS is still able to remain constant throughout the month. Furthermore, September’s numbers present the same conclusion. A lower average traffic volume of 259, with an even higher ADS of $78.27. As pictured, traffic varies immensely. Over the weekends traffic is usually 300+, while weekday traffic usually bottoms out on a Wednesday or Thursday and lingers around 20% of what a busy Saturday would be.



Plotting traffic and ADS together provide an easy to digest solution to visualizing the store data over time and observing the relationship between the two. There are a few outliers within the data, but considering 90+ days are considered, these outliers aren’t extreme enough to cause a statistical irregularity. Traffic is pretty consistent from week to week on how it varies. Although a trend is beginning to form – As the graph suggests, ADS predominantly goes up as daily traffic begins to fall.

Correlation between these variables also plays a big factor in how they influence each other. A positive correlation between two variables suggests that the two are related. As one variable climbs, such as traffic, our other variable, ADS, would follow it. A negative correlation between variables suggests just the opposite.

Untitled:Users:Cj:Desktop:BDC:Correlation Traffic vs. ADS.tiffProvided on the following page are some correlation calculations regarding total traffic volume and ADS.

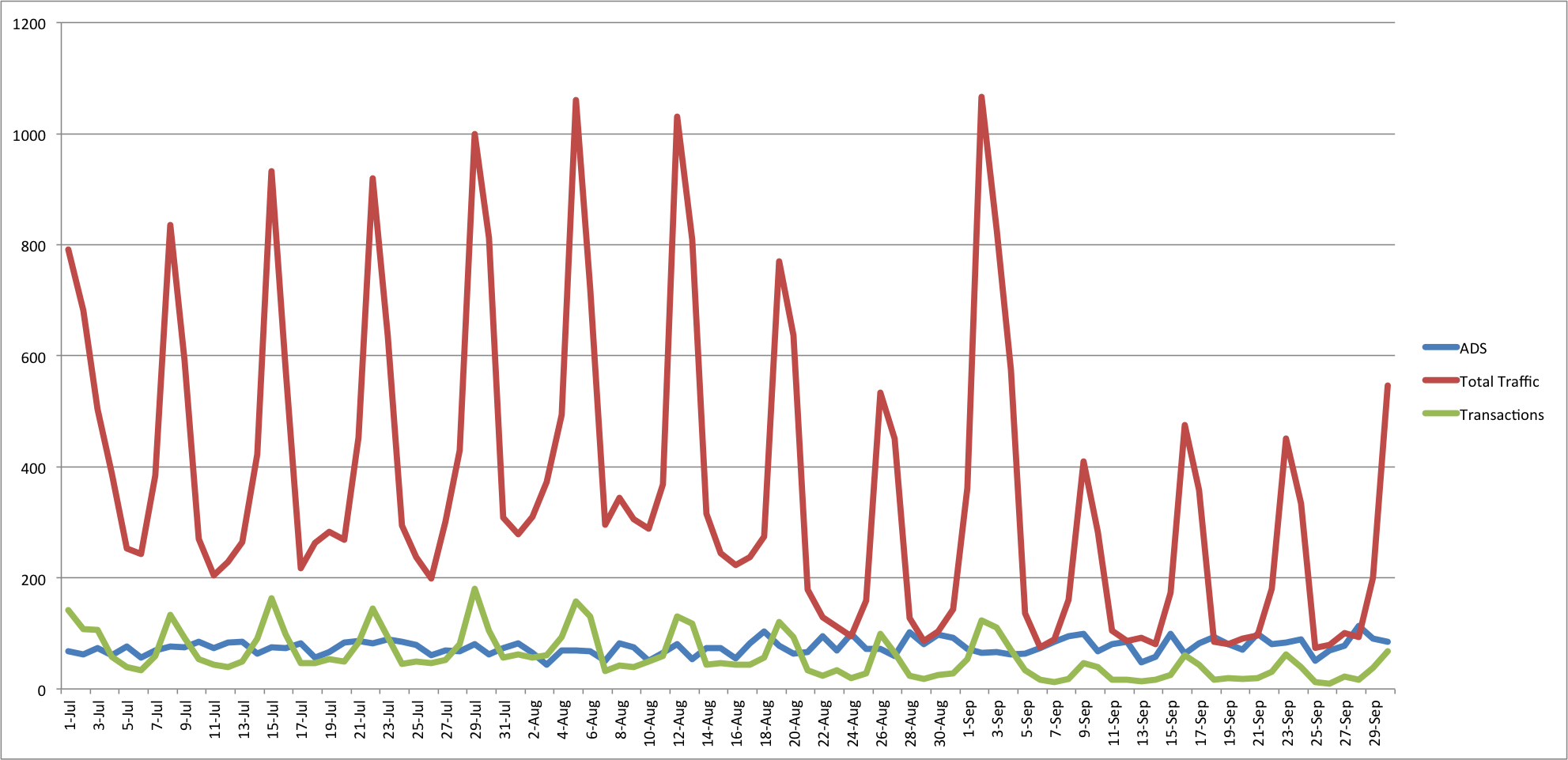
Untitled:Users:Cj:Desktop:BDC:Total Correlation.tiffAlthough it’s tempting, we must remember that correlation does not imply causation. Here, we have a little bit of a mix of data. July shows a slight correlation between higher traffic and ADS, but it’s marginal. While both August and September show a negative correlation between the two. When we contrast the 3-month total correlation to the individual months, a few observations can be made. Here, I’ve also included the correlation of different categories in reference to each other. Some are more direct than others.

The comparison between total traffic and the number of transactions is expected to have a very high correlation. Without more customers, we can’t make more sales. Again, a transaction does not denote that the sale was a quality or high-dollar sale. A transaction is recorded for any item purchased by a customer, big or small. Next is our comparison of the transaction count vs. ADS. Based on my calculations, for roughly every five transactions, ADS will drop $1, respectively. Now for the most important factor, traffic against ADS. Month by month, ADS seems to be inversely related to traffic. Likewise to transactions, (roughly) for every five additional units of traffic, we will see a $1 decrease in our daily ADS. This seems like a small feat, and at only a dollar at a time, it seems to be a non-issue. But on days where traffic hovers around 600+, this could present a sizable issue with keeping the store ADS healthy.

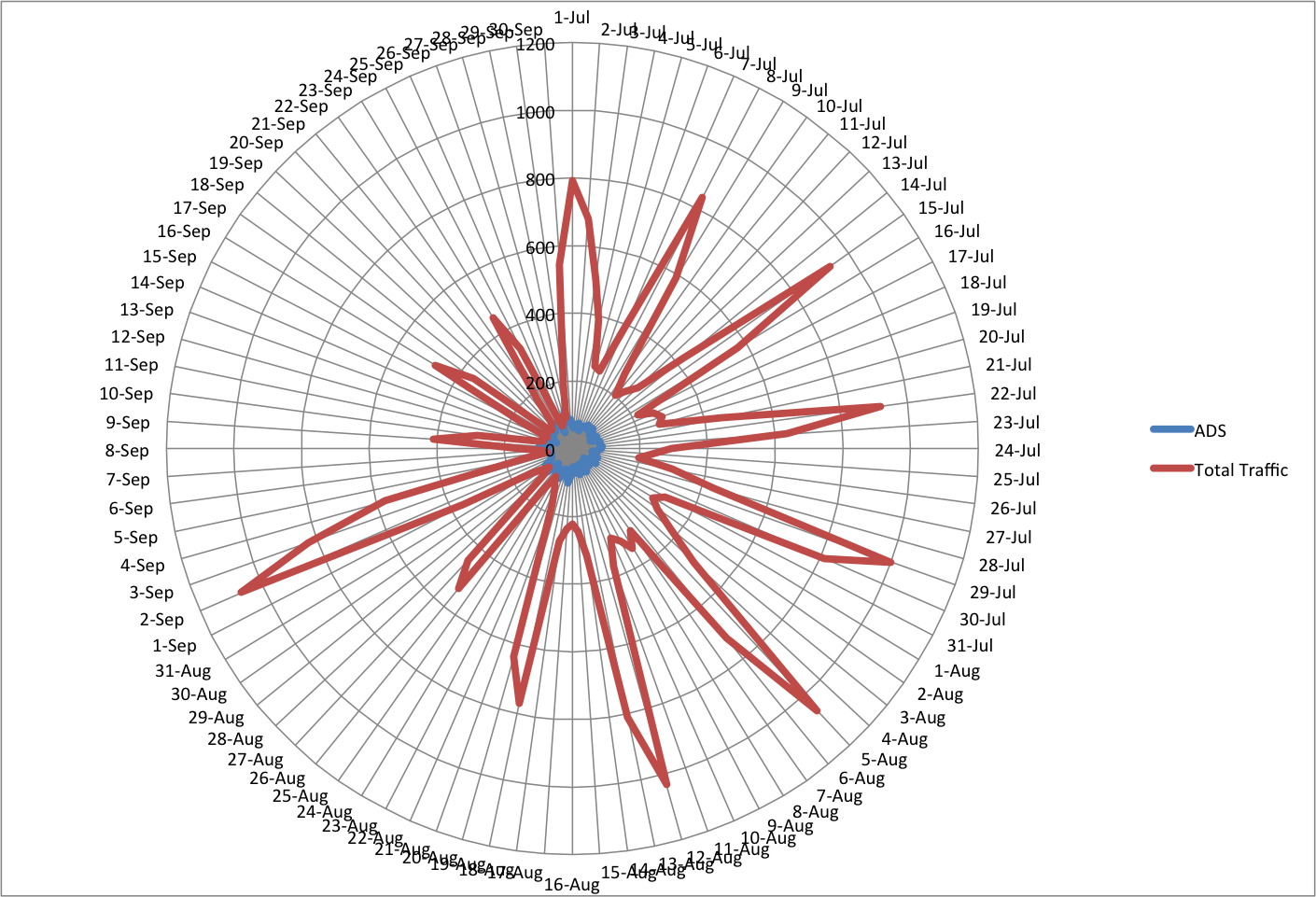
Traffic numbers throughout the week for the store vary quite a bit. But on an average July day, we easily see around 450 customers. At 95% confidence, the total traffic can be expressed within a confidence interval for each month.

Untitled:Users:Cj:Desktop:BDC:COnfidence interval total.tiff

Armed with this, the store can project daily sales volume for a certain day. For the most part, this would be useful for projecting volume for the following year around the same time period. This allows for better staffing and stocking of inventory, allowing us to be prepared and operate as efficiently as possible.



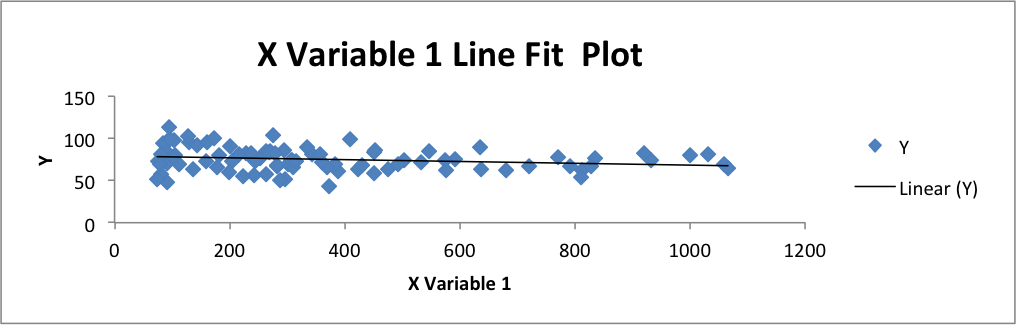
Provided above is a complete graph mapping out all provided variables together. This can help point to patterns in the data, and how they can be better optimized. As stated previously, it’s clear that transactions will rise as total traffic does. ADS is all over the board, but remains somewhat constant at around plus or minus $20. This points to the fact that neither traffic nor transactions positively affect ADS. ADS is it’s own entity, and seems to be completely separate from the daily workings of the store. To better illustrate the low-variance of the ADS, I’ve provided a radar graph depicting traffic and ADS.



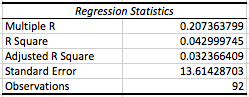
As the graph illustrates, ADS sticks around a central area and doesn’t stray much. Throughout the day, there are many factors that affect our traffic, our ability to complete transactions, etc. But all of these seemingly have no affect on ADS and it’s tendency to bounce back to a central area. There are days with an abnormally high ADS, and some with an extremely low one. But these are either outliers spread out over a distance of time or a trend over a miniscule time frame. Every now and again, the store will suffer a stretch of sales days where ADS simply won’t pick up. Maybe it’s the consumer’s willingness to spend, or maybe it’s our ability to properly sell the product.

Inferential statistics are extremely useful in the business world. They allow us to process the data we’ve received, and utilize it to better predict future events and sales trends. Inferences are made using only a sample of the population, manipulating and using that information to infer useful conclusions based on the data. Statistics such as correlation, regression and analysis of variance tests (ANOVA) are all considered inferential statistics. Regression analysis is primarily used for modeling and analyzing several variables, when the focus is on the relationship between these variables. One dependent variable, and one or more independent variables can be used for this analysis. For the sake of the following calculations, ADS is considered our dependent variable, while traffic on the other hand is used as our independent variable. Provided are the results regarding the regression analysis.

Untitled:Users:Cj:Desktop:BDC:regression.tiff

The regression analysis used on the data has provided some interesting conclusions. The standard error of both samples is quite low, which effectively measures the amount of noise within the variable that can’t be explained by the model. The coefficients provided by the analysis are very important here. These represent the values used in a linear regression. With these, we’re able to plot the data, and compare it to the expected values down the line. This method allows you to observe the data and create a “best fit” line through the graph.

For the above illustration, the “X Variable 1” represents the line that was created by the regression analysis. The points that are represented on the graphs are each of the points where ADS and traffic meet for that particular day. As you can see the regression line fits pretty well into the data. That being said, the line is nearly flat. This suggests that the slope coefficient is nearly zero and has almost no effect on ADS, the dependent variable. With the slope of the line almost negligible, this tells us that the ADS will for the most part always be within range of the Y-intercept value, at $79.14. In extreme cases where we see north of 800 customers, ADS slowly begins to decline. Using this data, we’re able to create an equation that could roughly predict the projected value of the ADS based on the traffic volume. This would be expressed as: Projected ADS = -.0108(Traffic) + $79.14. A best-fit line isn’t the only use of a regression, much more useful information can be derived from this analysis.



The resulting calculations of this regression can also be used for decision-making. The Multiple R calculation above represents the percentage of the response variable variation that can be explained by the linear model. Therefore, the predictor – traffic – can explain only about 21% of variance in ADS. Furthermore, the use of this regression shows us that the R Square value is quite low. The R Square value is the proportion of the variability in the ADS that can effectively be explained by the model. Coming in at only 4.29%, this means that the model is not doing very well at predicting the actual fluctuation in ADS.

**Conclusion:**

Retail stores face a multitude of different factors that affect their daily sales and efficiency threshold - many of these almost impossible to predict. They must use every resource available to them in order to stay ahead of the game and continue reaching goals and benchmarks for the company. Whether this includes more efficient staffing, ensuring sufficient quality and quantity of inventory, or all of the above. Many (including myself) would be led to believe that the more traffic that comes through the store, the higher the sales amount per transaction, or ADS. Given the descriptive and inferential statistics provided above, this doesn’t quite seem to be the case. The data points to the conclusion that traffic has little if any effect on ADS, until higher concentrations of traffic begin to flood the store. At that point, ADS begins to slowly fall. This could be due to many reasons. As the store becomes crowded, people may prefer to spend as little time as possible in a busy store with limited staff to help them. On days of 1000+, it’s impossible to tend to every customer with the best service possible. That being said, the quality of service is often what makes the difference between another unconverted customer and a $200+ sale. Like aforementioned, as online shopping becomes easier and shipping becomes cheaper and quicker, brick-and-mortar retailers need to be able to provide top-notch customer service in order to remain competitive. With a little statistics magic, retailers can make more informed decisions, and be as prepared as possible for the future.