ABC Company Report

Alex Lopes

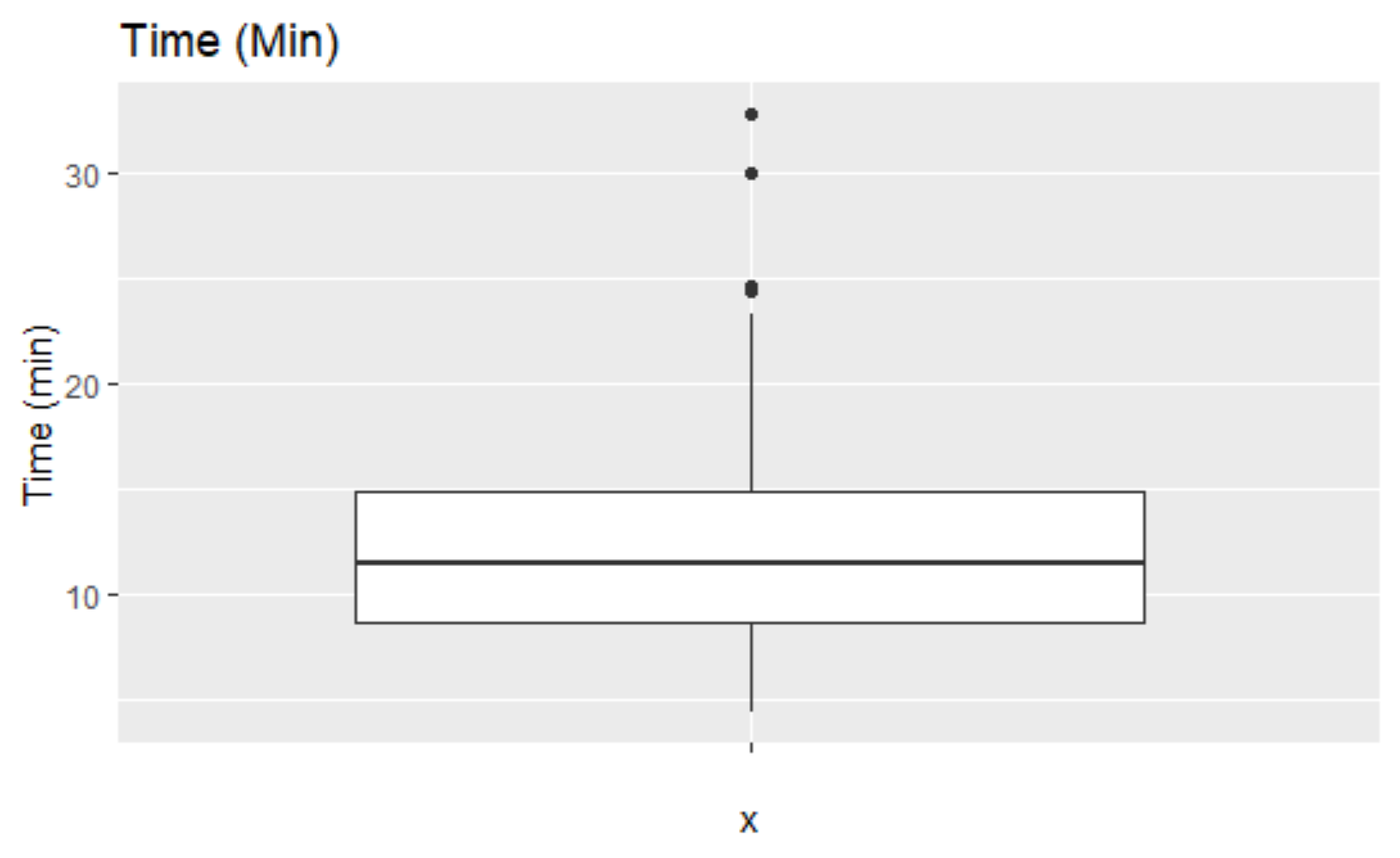
DATA 5400 Predictive Modeling

Michelle Rego

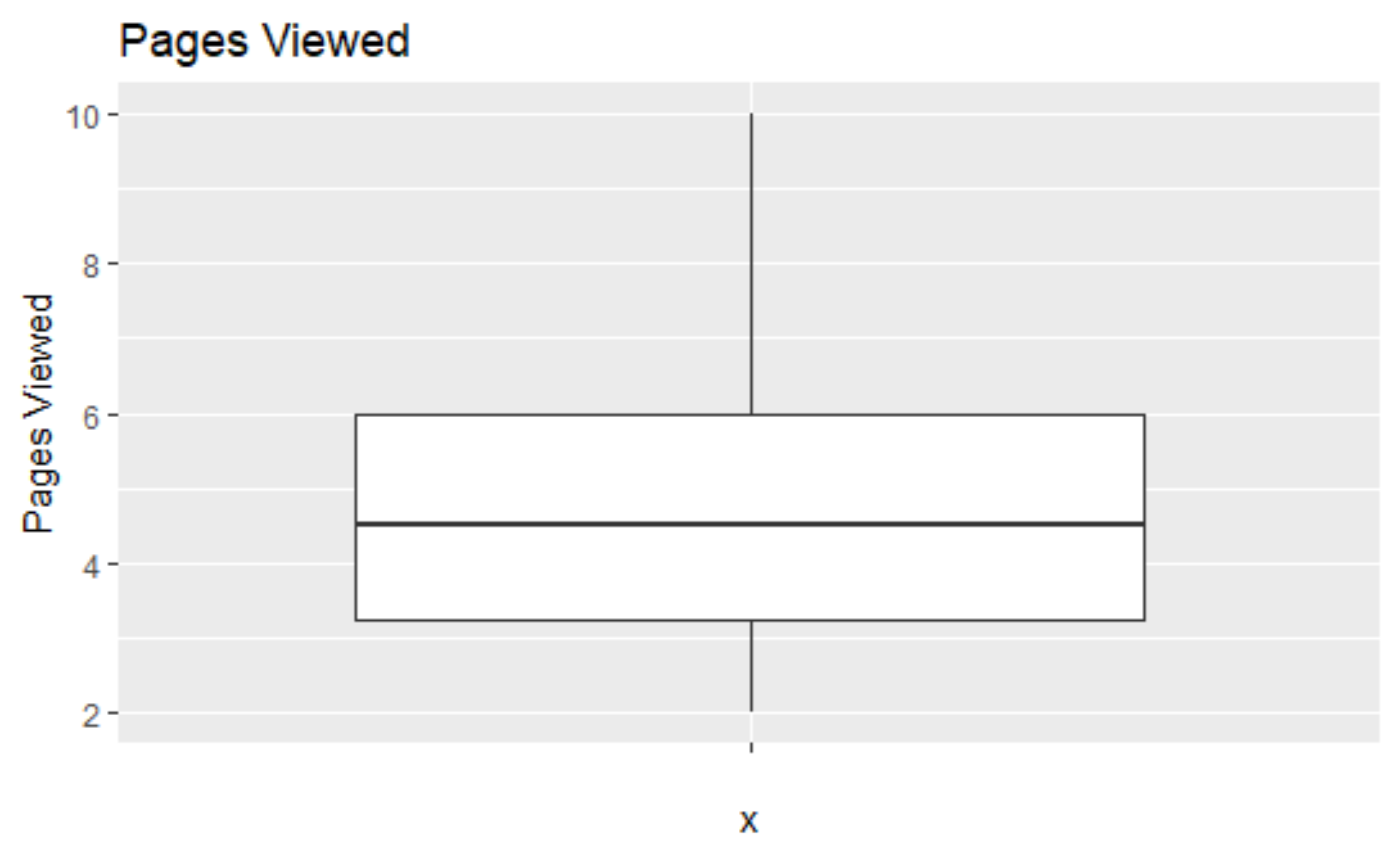
Abstract

ABC Company has recently opened their online sales and have recorded their previous months transactions in a dataset. Due to web site sales exceeding expectations management at the ABC Company in Bronx, New York is looking into the data from the previous month’s sales to determine if more time spent and more pages viewed on the website would translate into more money spent.

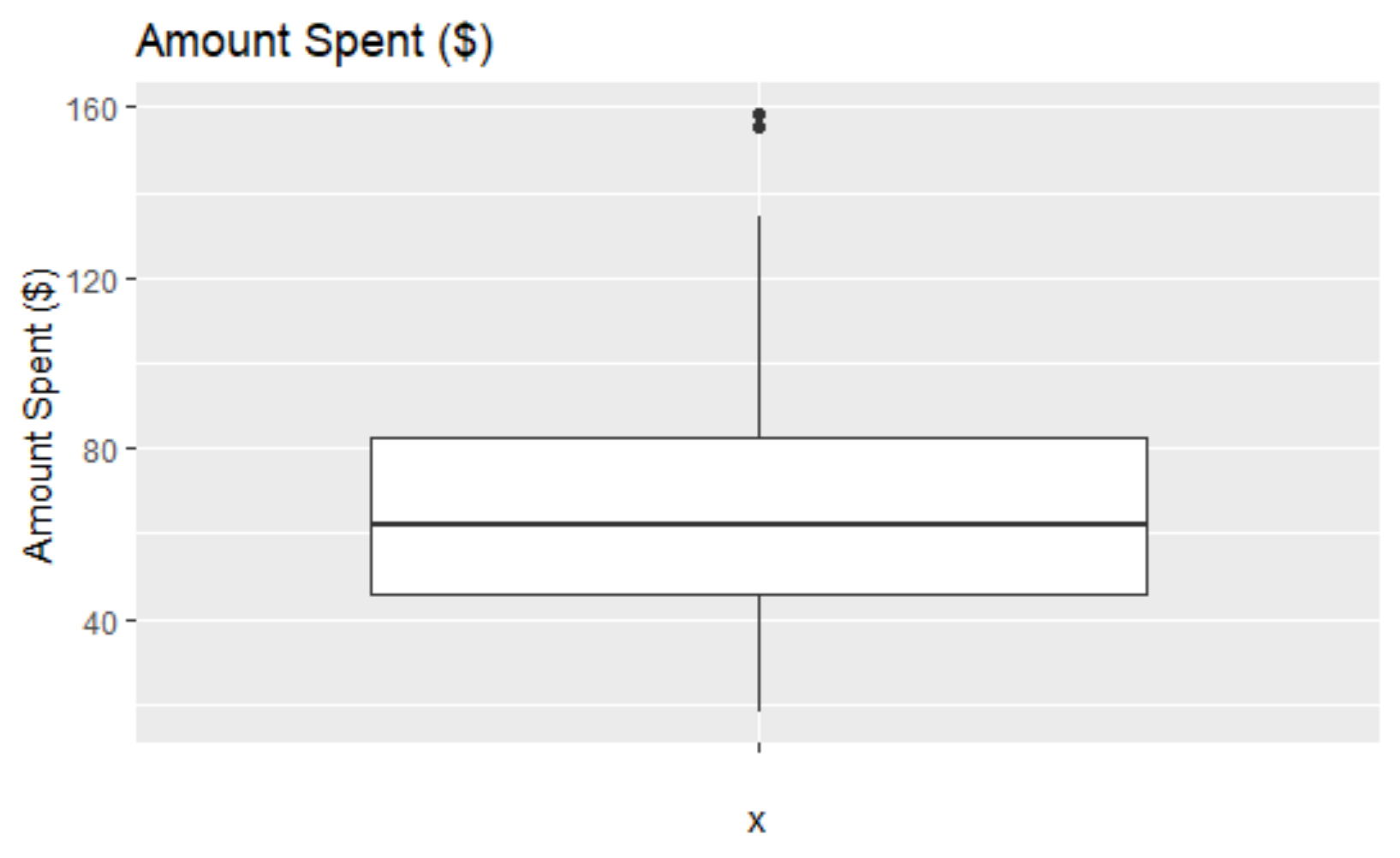
Descriptive Analysis:

Length of time was calculated in minutes with a range of 4.3 to 32.9 minutes. The first and third quartiles are 8.65 and 14.90 with the mean being 12.81 and a median of 11.40. There was a standard deviation of 6.06 with a variance of 36.76. Most customers spend around 8 to 10 minutes on the website with some as little as 4 minutes and no one on the site longer than 33 minutes. 

The number of pages viewed had a variance of 4.15 and a standard deviation of 2.03 with a range of 2 to 10. The first quartile was 3.25 and the 3rd quartile is 6.00 with a mean of 4.82 and a median of 4.5. Customers appear to be viewing around 3-6 pages during their transactions.



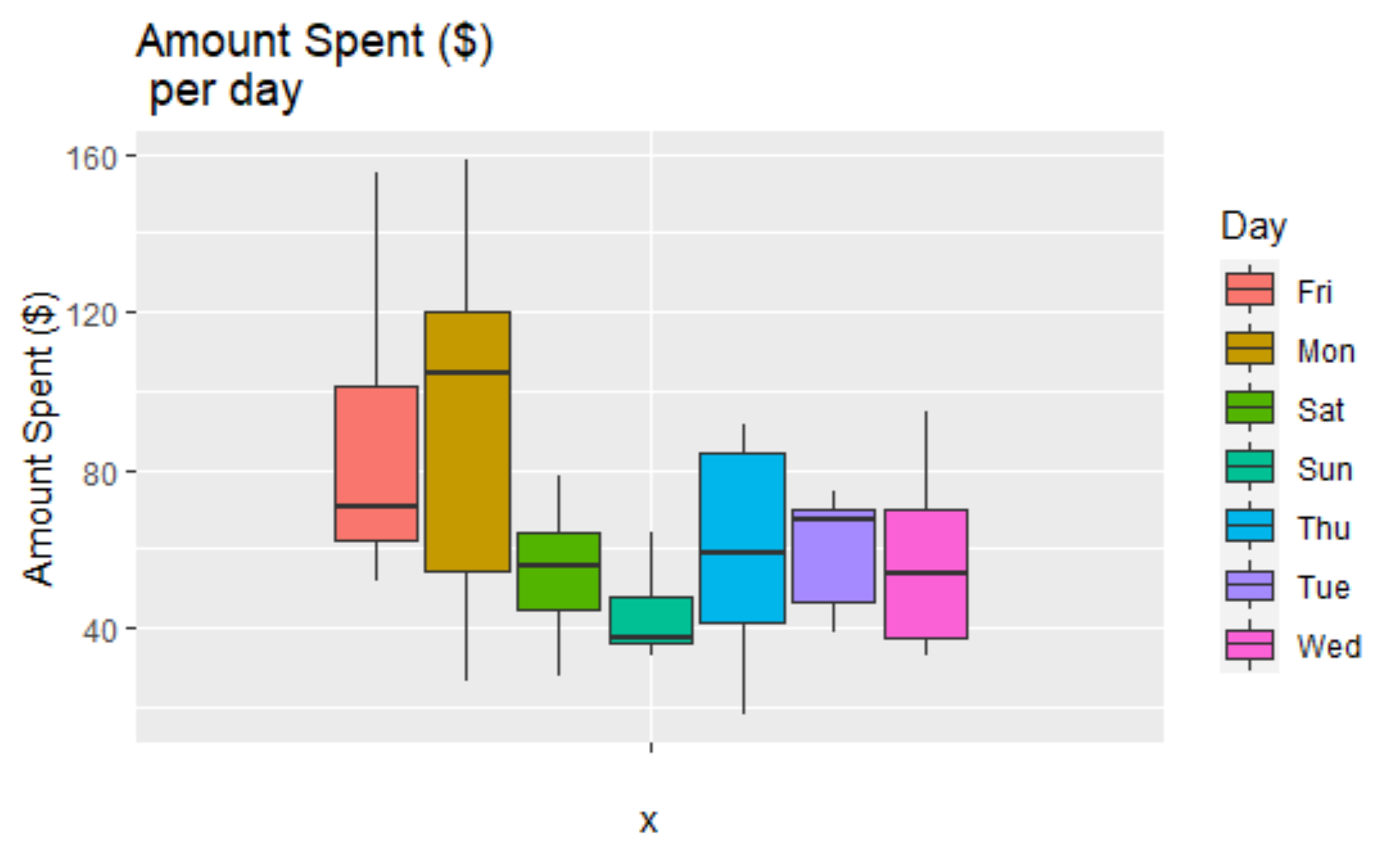
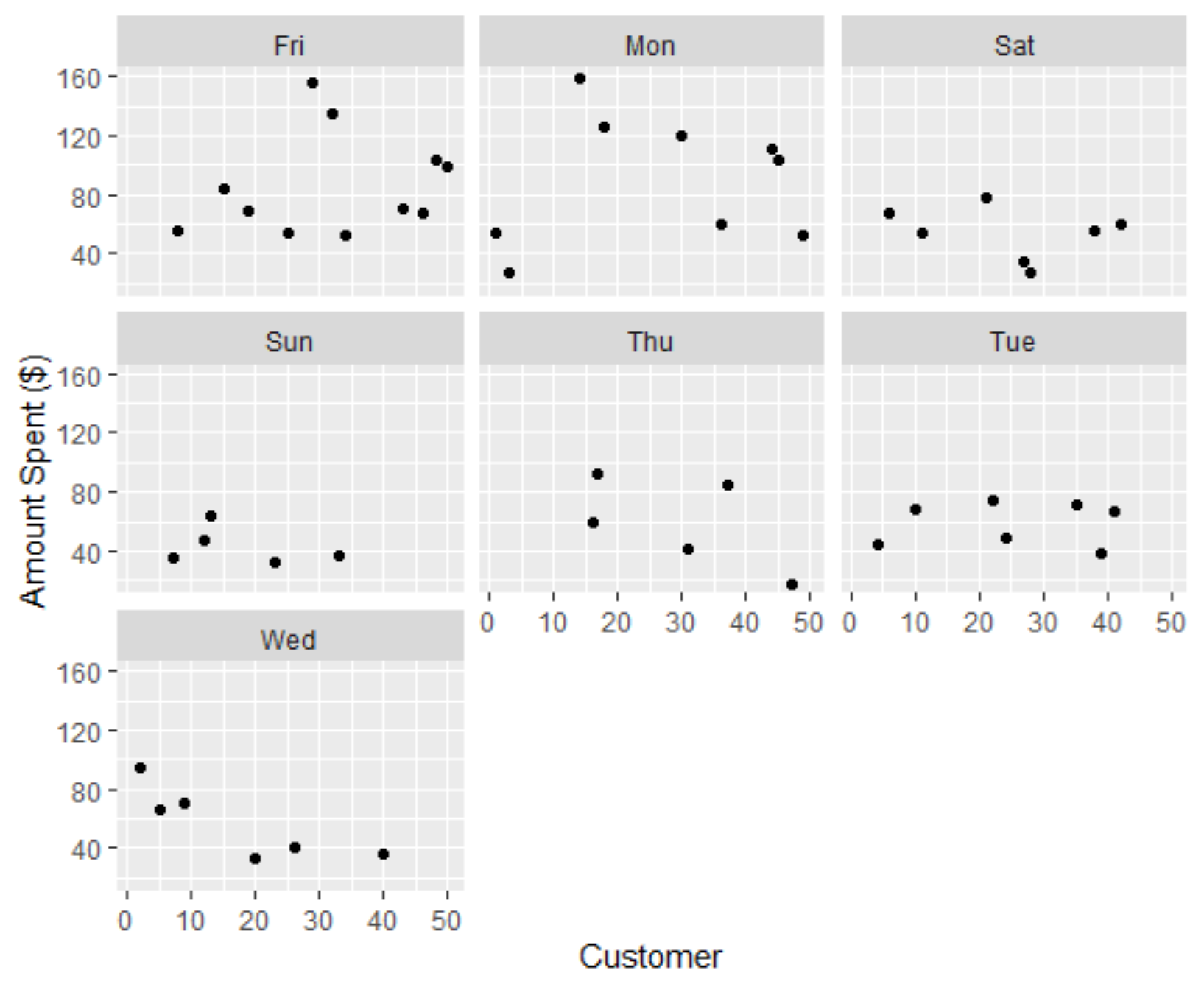
When reviewing the amount spent in US dollars there was a range of 17.84 to 158.51 with a first quartile of 45.56 and a third quartile of 82.73 and a mean of 68.13 with a median of 63.15. There was a standard deviation of 3.34 with a variance of 1046.12.



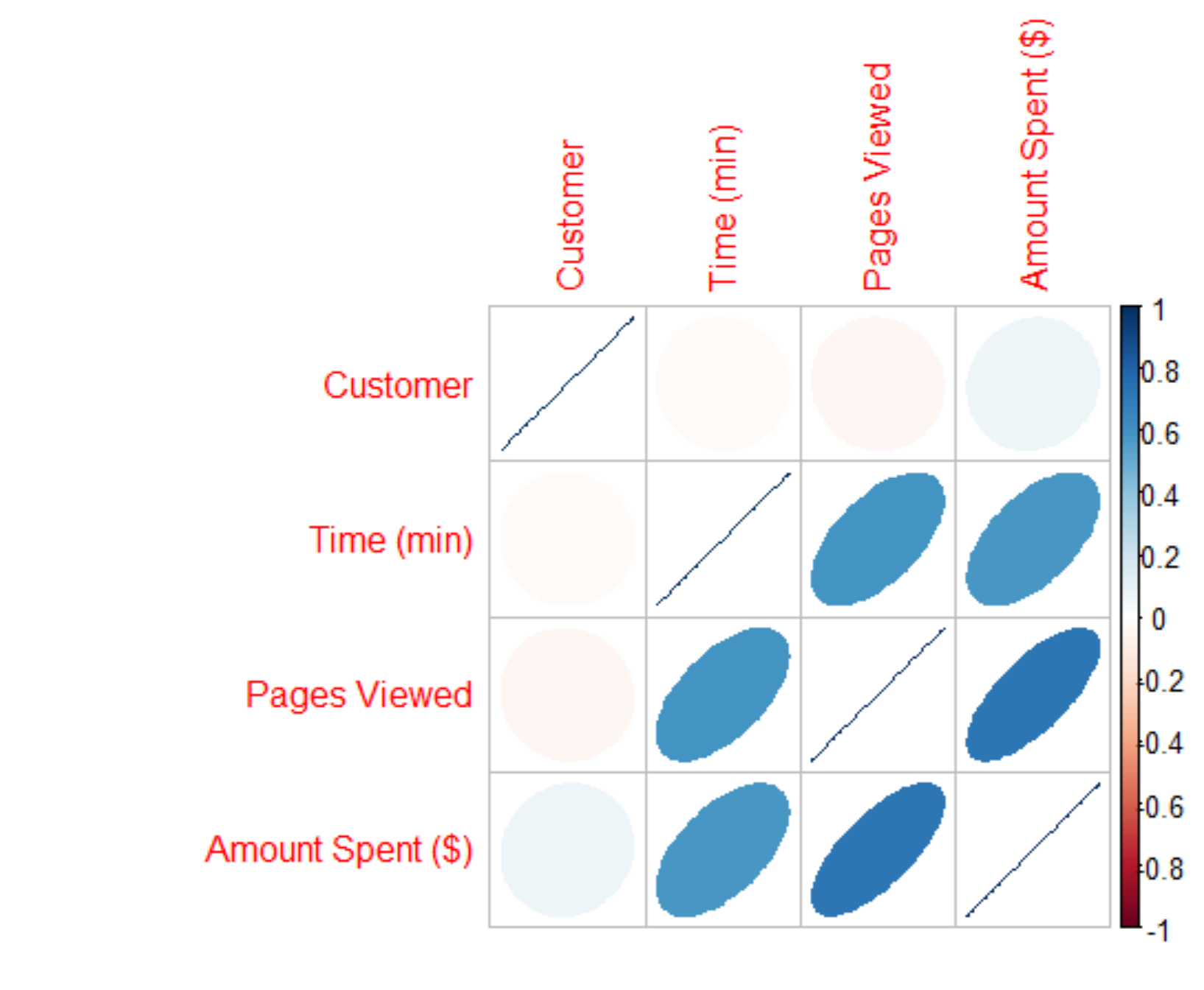
The minimum value is 17.84 and the first quartile is 45.56 however the median 62.15. From this I assume there is a marketing push for $50 dollar shipping, a common practice in in order to impact customer retention, acquisition, and cart abandonment (LEWIS, 2006). If there is not a shipping benefit in place ABC could add in regards shipping program to help bolster purchases especially when seeing how much a minimum amount spent is relative to the quartile values.

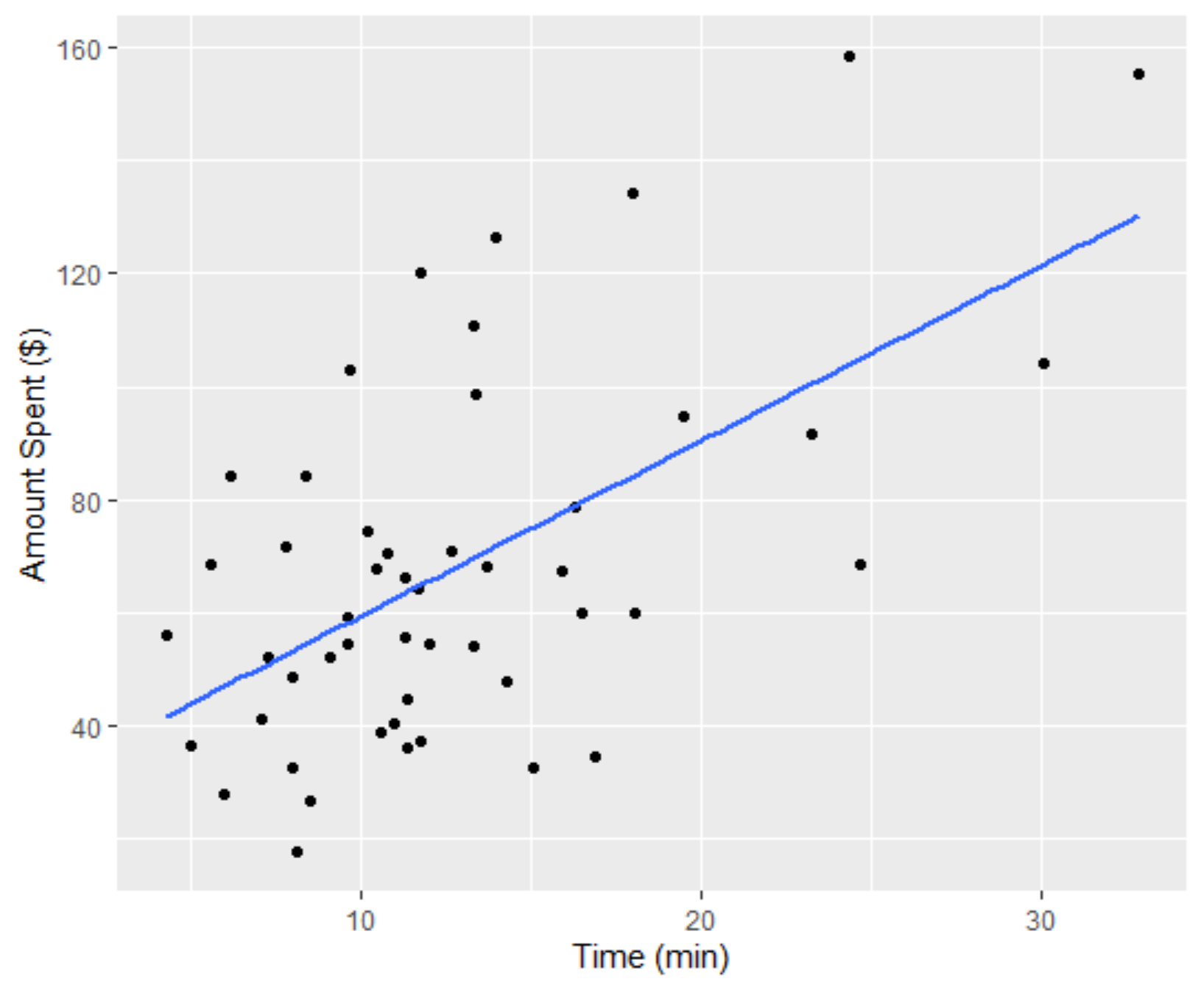
In reference to the mean value of 68.13 and a minimum value of 17.84 we can assume the most purchases do not have more than 3-5 items when dividing the minimum value from the mean, first and third quartiles. Through what is presented from the data most customers view 3 – 6 pages on the website within 8 – 10 minutes leading to ~$45 - $82 in amount spent with no more than 3 – 5 items assumed in the transaction.

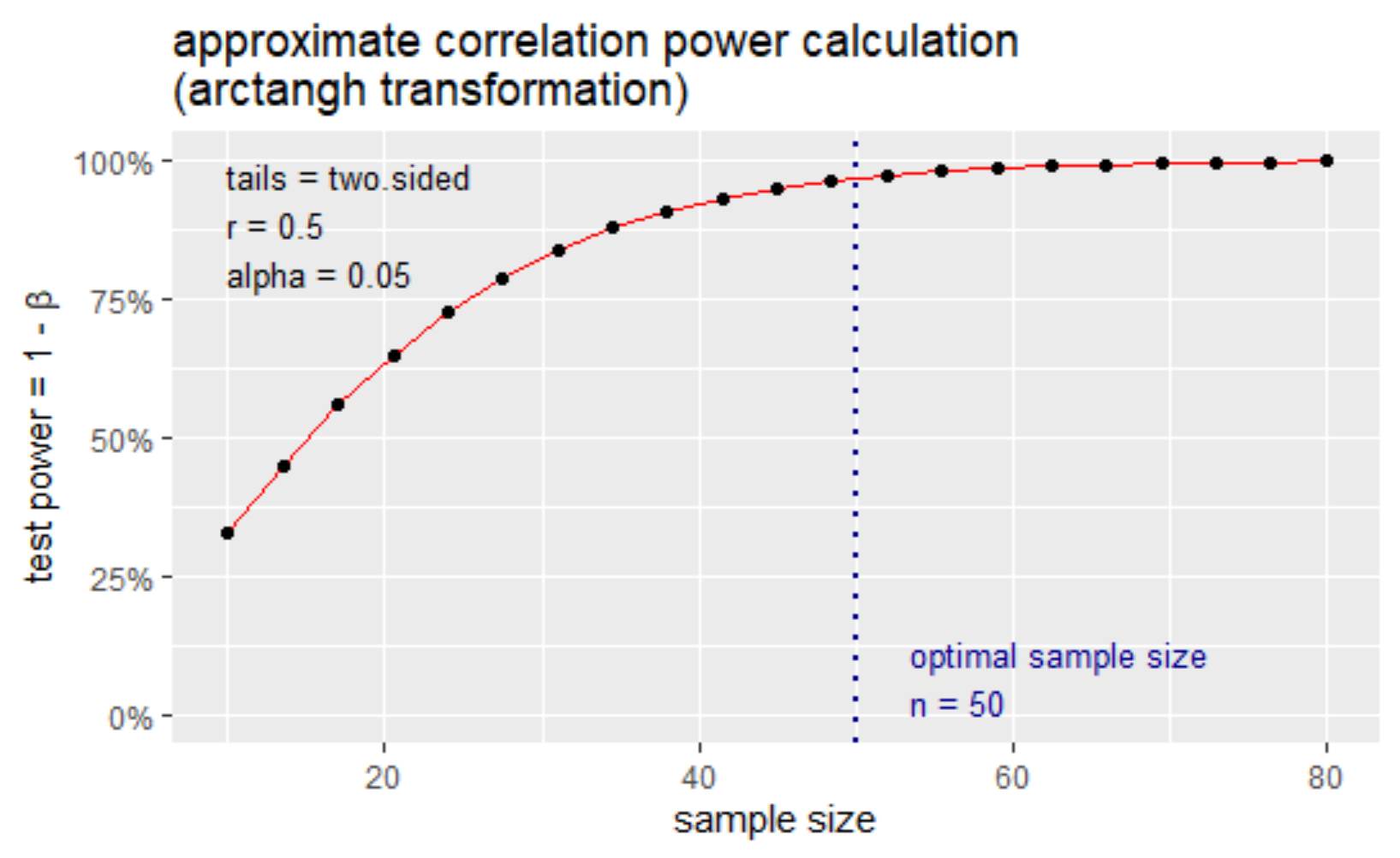
The total dollars spent on Fridays are $945.43, Mondays with $813.38 and Tuesday at $414.86. The day with the least amount of money spent is Sunday with $218.15 in online order amount spent.



Method & Results:

The dataset of ABC online shoppers is an excel file so we start by reading in the file and examining the structure of the dataset. When reviewing the correlation plot, amount spent and time on the website appear to be highly positively correlated. One other correlation was number of pages viewed and time spent on website. Although most of the focus of the study has been on time and amount spent on the website and we do see that time and number of pages viewed were also highly correlated. Amount of pages viewed were more positively correlated with amount spent on the website from online shoppers relative to time spent on the website. The main difference can be that with more pages there ae more products being considered as with time the customer could be contemplating a purchase on their website. 





Limitations/Assumptions:

What are the limitations of the analysis performed?

Within the last scatterplot in the code there I a linear model line added. When looking at the plot you will notice a few outliers especially on the ends of the line. These outliers could be distracting from the actual relationship as the are skewing the ends on the slop. Running a sample that more accurately represents the population with the removal of outliers could lead to different conclusions. Also a larger sample would lead to greater insight with the law of large numbers. Having another year’s worth of data

One assumption within this research is assuming each data points quality. If someone made a good amount of purchases but disfavored the service and left multiple negative reviews on the products they purchased, their purchase behavior may skew the sample away from a true representation of the population.

We also assumed that the minimum value for amount sol was one item. If this was a multiple item purchase or if there is a minimum spend for online orders that we are not privy to this will greatly impact how many items we believe to be in the cart.

Recommendations:

Knowledge of the holidays and marketing events are a couple simple common indicators that can add information to the sales data. Since the sample size was so low if there was a holiday or if a new marketing campaign was initiated then that could lead to an anomaly. Another topic for research could be a customer journey to gather more information. The dataset we worked with was only completed transaction but there is plenty of information in abandoned sales in figuring out consumer behavior.

Having a way to speed up customer journeys as well. There was a complication with time and number of pages having different results and correlations with the amount spent. This leads to an issue of data integrity since as a variable we do not know if it is truly positive if someone is spending more time on a website or not.

When looking at our regression line in the last scatterplot we notice a few outliers. When looking at the individual variables such as time spent on a page and amount spent there are some obvious outliers. In such a small dataset we may want to remove those outliers when running a regression model as that can skew the results and slope of the regression line.

References:

LEWIS, M. (2006). The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. *Journal of Retailing*, *82*(1), 13–23. https://doi.org/10.1016/j.jretai.2005.11.005

‌