r4e

Springboard Capstone Project

Website Traffic Analysis & Predictive Model

Flashlight

Springboard Capstone Project

Attn: Amit Dingare

Date: December 2, 2018

Submitted by: Byron Richards, MBA
                       Digital Marketing Strategist

                       Flashlight Infomediary Solutions
                        611-38 Masthead Court
                       Halifax, NS   B3M 0C7
                       Office: (902) 407-3336
                       Mobile: (902) 237-0945
                        Email: byron@flashlightis.com

Website: www.flashlightis.com

Social Media: www.facebook.com/flashlightis
                       twitter.com/flashlightis
                       www.linkedin.com/in/byronrichards/

Business No.   3274890

Table of Contents

Introduction:

Companies do not know if a digital marketing campaign is successful and/or know the true results of the endeavor since they are unaware of what the traffic to their website would've have been without any PPC campaigns.

I will deliver a predictive model of what the web traffic would've been over the same time period that my client engaged in two digital marketing campaigns. From this, we will be able to determine how successful (or unsuccessful) each campaign was by examining the results of the statistical model.

The client is a family run business with multiple locations over a wide geographical area. They offer home-based services for those who are unable to physically or mentally support themselves. Because their marketing budget is not robust & since they are trying to reach a mass audience, we decided that a PPC and a digital display campaign would be a good fit for them.

Gaining deeper insights into these digital marketing campaigns will give them a more accurate assessment of the marketing ROI.

Client Relevance:

The client is Compassion Home Care, who is a family run business that offers several home-based services for seniors or injured persons needing everything from physiotherapy to personal care. The company has multiple locations in Newfoundland (Canada) and they want to do a Google AdWords campaign to generate awareness & increase traffic to their website.

  o   www.compassionhomecare.ca

Without this information, they have no idea how to determine what kind of results to expect, thereby lack the ability to make informed marketing decisions.

Data:

Data Limitations:

The data is not telling the entire story of customer engagement, since it is too difficult to define "Quality Traffic" and to differentiate quality traffic from mistaken traffic brought by ineffective AdWords.

Also, the lack of any type of online sale makes it very difficult to determine a monetary value of web traffic. Lastly, since the client does not track how their customers are acquired, it's uncertain to know the importance of the website traffic.

However, the client believes that potential customers and their families do research Home Care providers online and this is evidenced by the key word research as well. Therefore, digital marketing is important to grow Compassion Home Care's business.

Data Wrangling Steps:

In order to build an effective predictive model of the website traffic, 6 years of Google Analytics data was gathered from the client's company website. This would also provide a visual of the site traffic to determine if any seasonal trends were present.

1) The original data set was comprised of daily web traffic starting on March 12, 2012 until March 12, 2018. The Google Analytics platform allowed the following metrics to be exported as a CSV file individually:
   - Sessions
   - Number of Sessions per User
   - Average Session Duration
   - Pageviews
   - Pageviews per Session
   - Users
   - New Users
   - Bounce Rate

Each export included the date and the value of the metric, so all the data was combined into a single dataset "web_data3.csv" (https://github.com/Flashlightis/Capstone-Project/tree/master/Data-Sets).

The next step was to examine the data to determine the classifications of each variable, look for missing values and any outliers. Since Google Analytics data is very reliable, only 1 outlier and 1 missing value were discovered. The outlier was found in the Sessions variable and was

700% larger than the largest value. After looking at that record in the Google Analytics data, this large value was due to bot traffic, so it was removed along with the missing value since it was contained in the same observation as the outlier.

- The code to do this is labelled "Data Cleaning.rmd" which is found:
    - https://github.com/Flashlightis/Capstone-Project/tree/master/Code

2) To gain better understanding and to help visualize the data, the following subsets were created:
- Mean Monthly Values
- Mean Yearly Values
- All 2012 Data
- All 2013 Data
- All 2014 Data
- All 2015 Data
- All 2016 Data
- All 2017 Data
- All 2018 Data
- The code to do this is labelled "Subsetting Data.rmd" which is found:
    - https://github.com/Flashlightis/Capstone-Project/tree/master/Code

After conferring with the client, it was decided that focusing on the Sessions was most important to his business. Therefore, the focus moving forward was to examine the Session data.

Exploring Data Analysis:

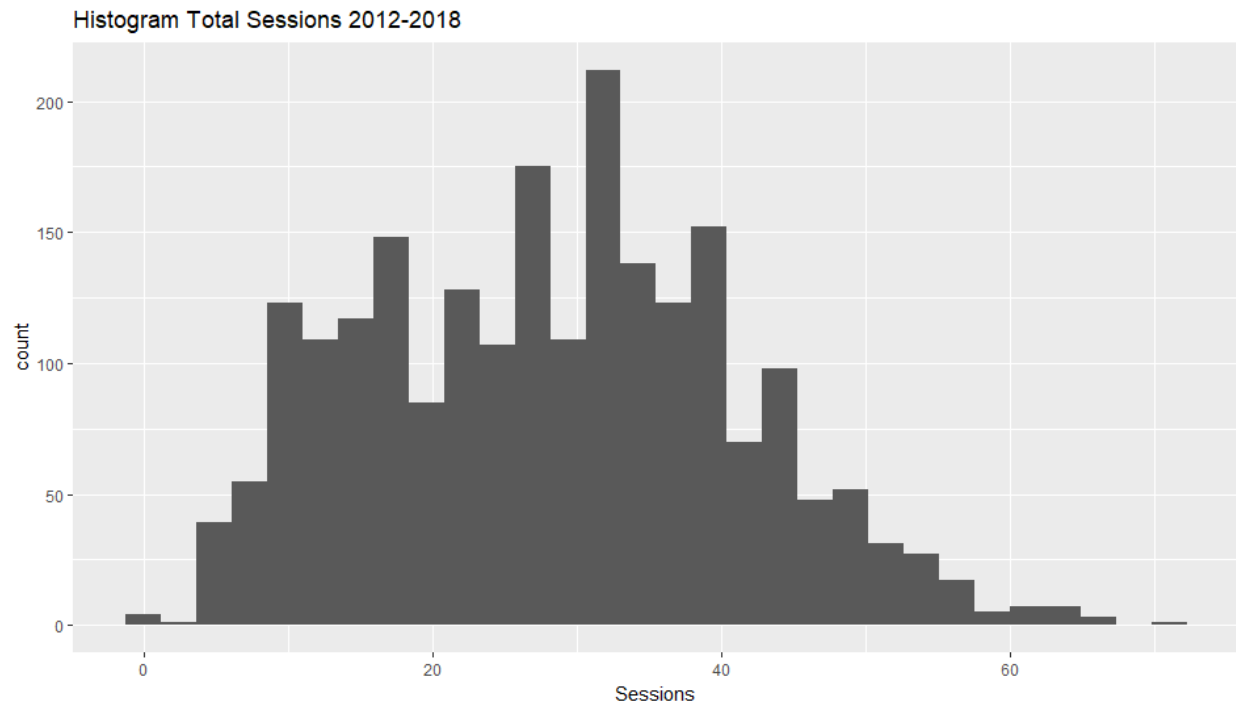**Bar Chart of Total Sessions by Year**
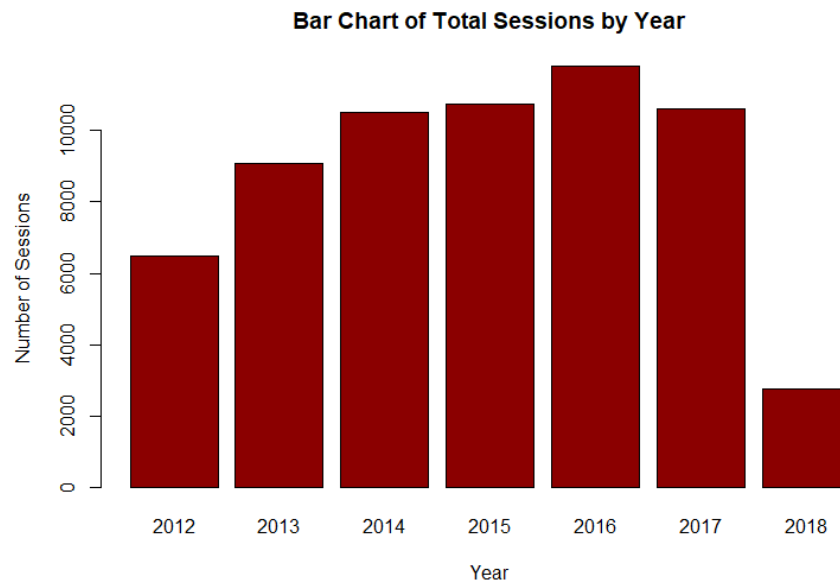
Looking at the Sessions over time, there is a solid increase in the number of Sessions year over year with a slight dip in 2017. The low amount of sessions in 2018 is because at the time, data was only taken until March of that year.

Boxplot of Total Sessions
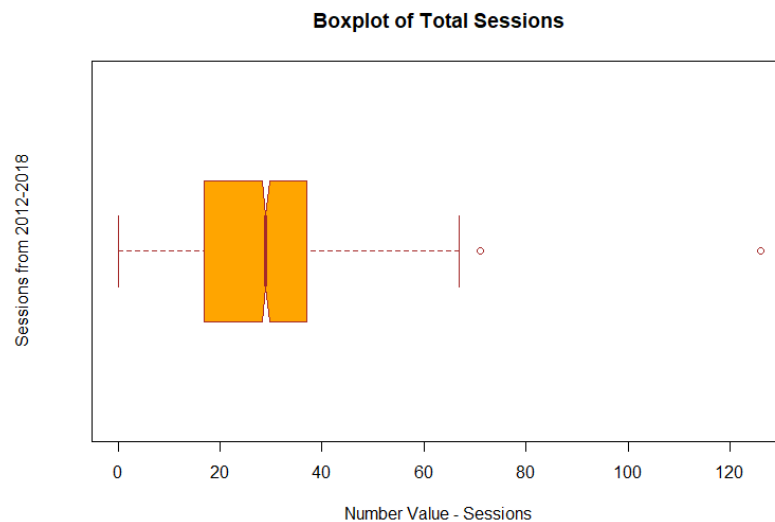
The boxplot of all the Session data indicates:
- Median is roughly 30
- 2 outliers – 1 is about 75 and the other is roughly 125
- 50% of Sessions land roughly between 20 & 40
- Range is approximately 65

- Code & other boxplots can be found:
    - https://github.com/Flashlightis/Capstone-Project/tree/master/Box%20Plots

Histogram Total Sessions 2012-2018

The histogram of all the Session data shows:
- Multimodal distribution
- Most Sessions occur roughly at 30
- Slightly skewed to the left
- As the boxplot indicated, there's a few outliers past 60
- Data is widely distributed
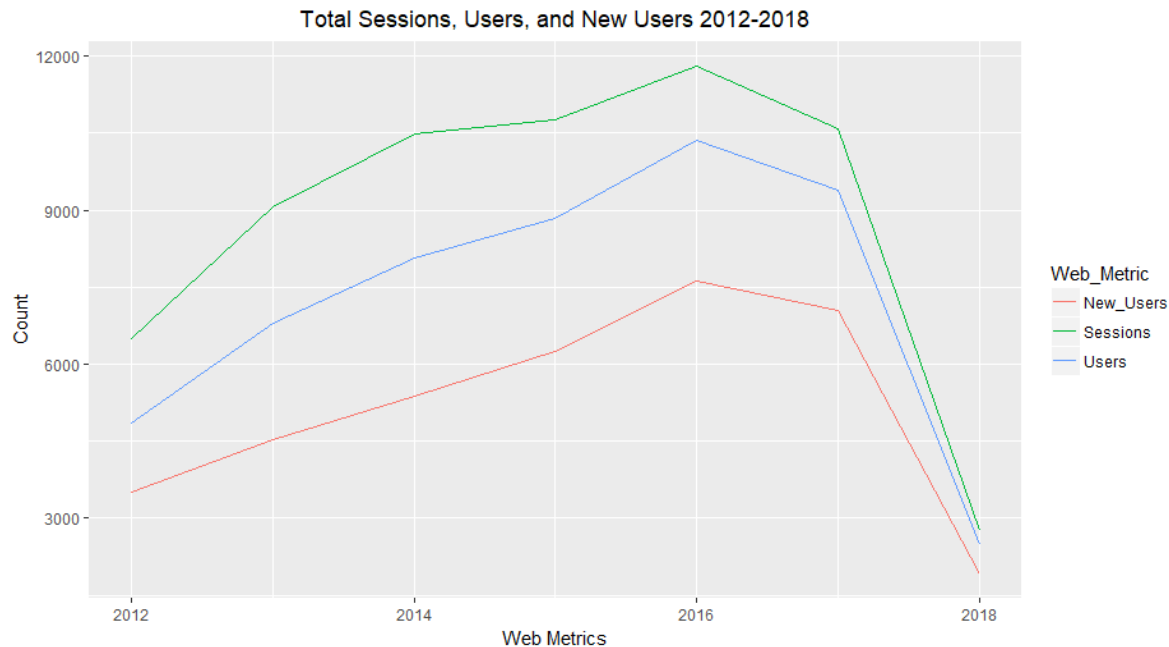- Code & other histograms can be found:
    - https://github.com/Flashlightis/Capstone-Project/tree/master/Histograms

A statistical summary of all the sessions shows:
- Median is 29
- Mean is 28.21
- 1st Quartile is 17
- 3rd Quartile is 37

Creating Tidy Data:

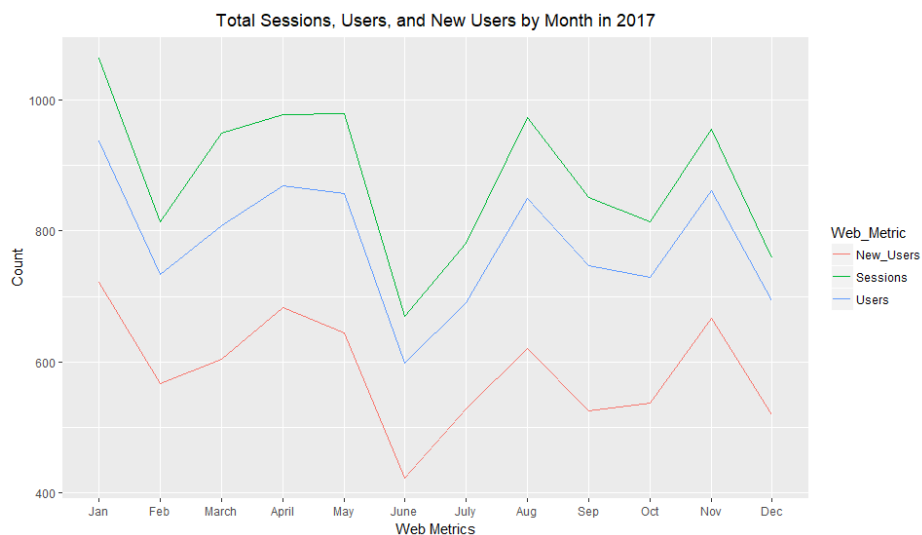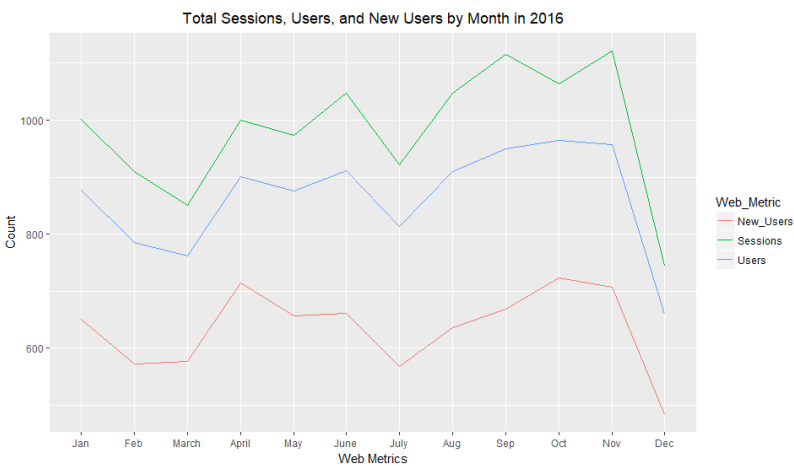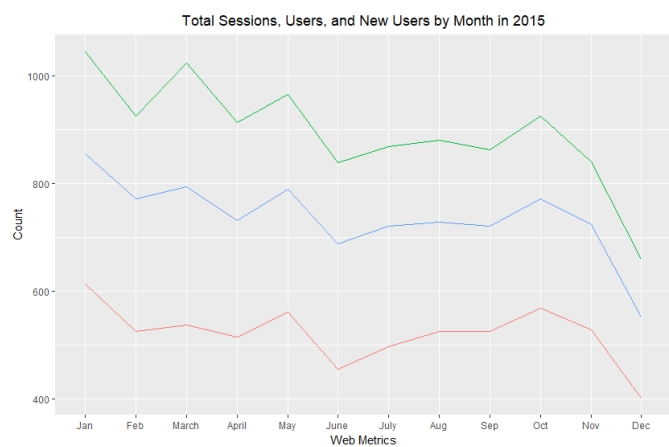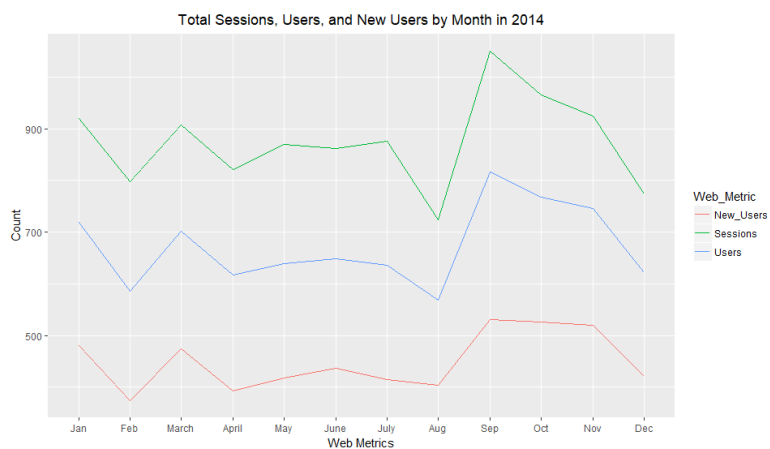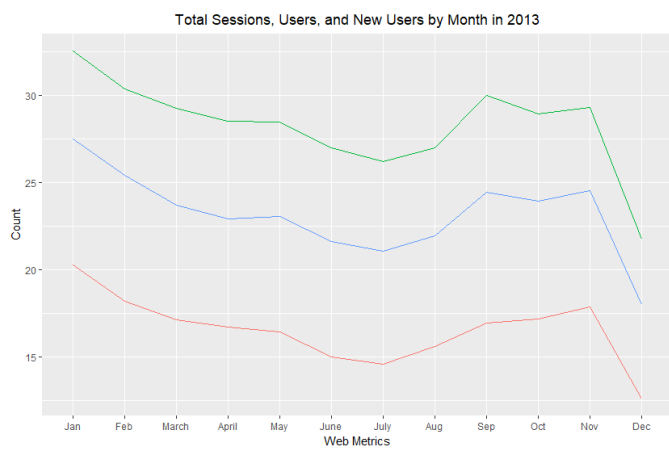Before doing any kind of analysis, the data set must be made tidy. A new variable "Web_Metric" needs to be created that contains Sessions in order to visualize and analyse the data. The code to make the entire dataset tidy is labelled "Make Tidy" and can be found:

- https://github.com/Flashlightis/Capstone-Project/tree/master/Code

Total Sessions, Users, and New Users 2012-2018

The total sessions by year doesn't show anything significant. It only shows that web traffic (Sessions), along with New Users and Users are growing year over year, but then begin to decline in 2017.

Taking a deeper dive into the dataset might reveal more insights. Therefore, the data was broken down by year to gain a better understanding of the existence of any seasonality or trends in the website traffic.

Total Sessions, Users, and New Users by Month in 2013

Total Sessions, Users, and New Users by Month in 2014

Total Sessions, Users, and New Users by Month in 2015

Total Sessions, Users, and New Users by Month in 2016

Total Sessions, Users, and New Users by Month in 2017

When the data is broken out by year, the visuals show seasonality does exist that is also fairly consistent throughout 2013-2017. We observe:

- Sessions (green line) declines from approx. January to February, May to June and then from November to December

- Sessions tend to increase from December to January, March to May, June to August, and then from October to November.
- Code & charts can be found:
  - https://github.com/Flashlightis/Capstone-Project/tree/master/Monthly%20Line%20Chart%20by%20Year

After conferring with the client concerning these traffic trends, he said that the website traffic trends coincided with his sales cycles and feels that the seasonality is typical in the industry. While Compassion Home Care's range of services are diverse, most of the company's customers are the children (typically 35+ years old) of aging parents. They will find care for their parents or relatives who do not yet require around the clock care but help to remain independent in their own homes.

The website traffic cycles tend to revolve around holidays. The adult children will visit their aging relatives and will realize that they need help after spending time with them, so website traffic spikes after holidays and dips beforehand. Popular holidays like Christmas, Easter, and Canadian Thanksgiving (October) seems to fit this trend.

This major dip from roughly May to June and sharp rise in traffic from July to August can be explained by a similar rationale. As the weather warms and summer gets closer these Adult Children are focused on their own families, making vacation plans and finishing school rather than on aging relatives.

When summer arrives and vacations begin, Adult Children visit with their aging relatives and probably even help them around the house. But as the summer progresses, the Adult Children begin to look for Home Care help for the winter.

Predictive Modelling – Facebook Prophet Analysis:

Using Facebook Prophet, a very sophisticated predictive modelling tool that specializes in time series forecasts that incorporates many different techniques (i.e. ARIMA, exponential smoothing, etc.), it's possible to determine what Compassion Home Care's web traffic would have been had without any marketing campaigns. Fore more information on Facebook Prophet and how it works: https://research.fb.com/prophet-forecasting-at-scale/

After the modeling is complete, it's possible to ascertain the success (or lack thereof) of the two marketing campaigns that were deployed. The first digital marketing campaign was AdWords only running from March 1, 2018 – May 26th, 2018. Another campaign ran from August 14th, 2018 – September 1st, 2018, consisting of AdWords and Google Display.

Facebook Prophet Insights:

Using the tidy data from March 2012 until December 31, 2017, Facebook Prophet forecasted the website traffic (Sessions) for 2018. Datasets and code can be found:

- https://github.com/Flashlightis/Capstone-Project/tree/master/Facebook%20Prophet

The initial results provided some interesting insights into Compassion Home Care's website traffic:
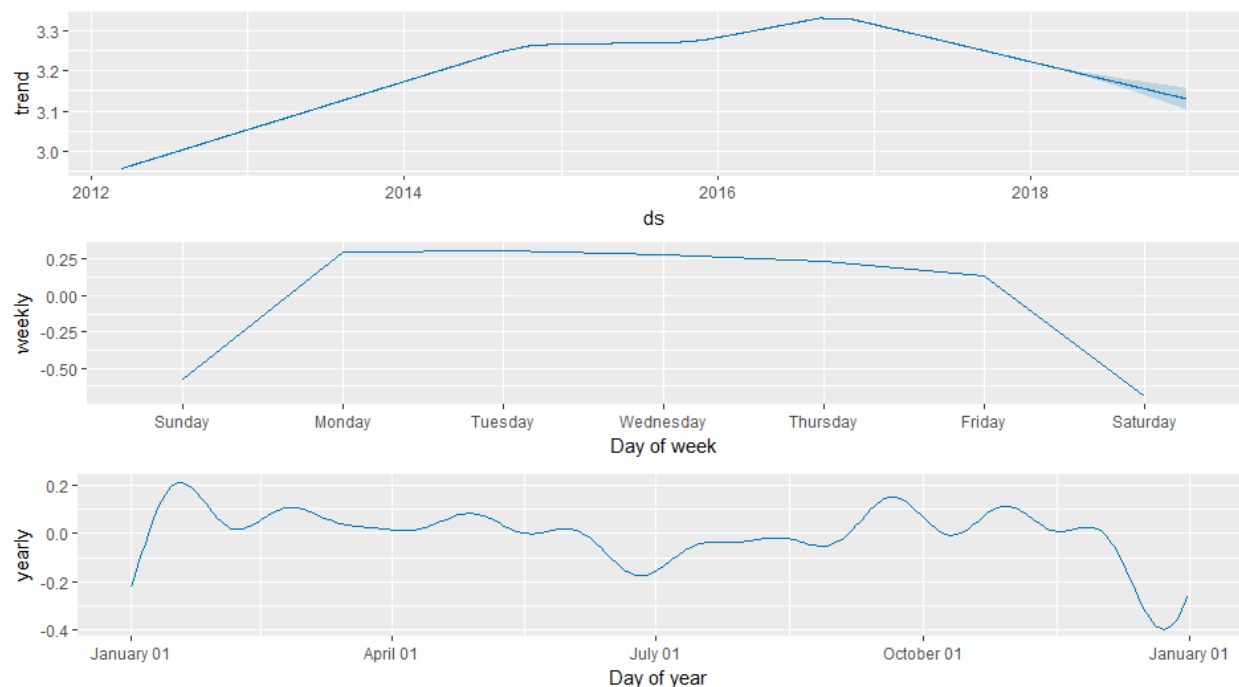

Facebook Prophet reveals similar patterns that were discussed earlier when reviewing the individual yearly website traffic. Seasonality is confirmed, downswings prior to major holidays and upswings afterwards.

However, the weekly website traffic reveals a very interesting trend. There is very little traffic on the weekends, almost all the traffic occurs throughout the week. It's possible that weekends act as "mini holidays", in that since most of the Adult Children spend time with their aging relatives over the weekend and then search for help throughout the week. The slight traffic peak on Monday with the slow decline throughout the rest of the week is consistent with the major holiday trend.

Predictive Model:

Before comparing the 2018 predicted versus actual website traffic, a test needs to be completed to test the model's accuracy. Using traffic from March 2012 – December 31st, 2016 was used to predict 2017's website traffic. If the model accurately predicts the traffic, then the model can be used to evaluate the 2018 marketing campaigns.
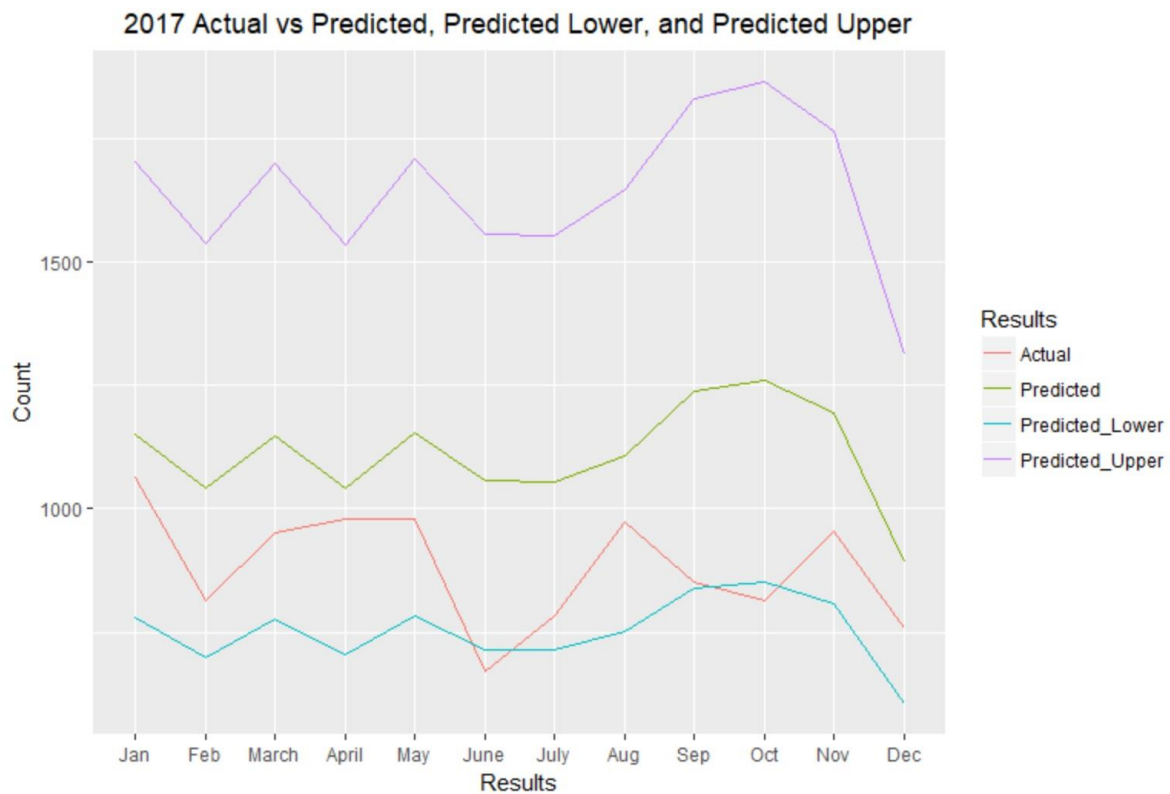
The model passes the test since all but two slight dips in June and October fits in-between the lower and upper prediction parameters. The actual traffic is skewed to the lower prediction parameters, but the model still proves to be accurate enough to evaluate the 2018 campaigns.

It's important to not that the MAPE for the test dataset is 28.5%. While it would be better to see a lower MAPE, factors such as the amount of traffic being small inflates the percentage. For instance, there are days when the actual is within 4 of the predicted, which on the face of it sounds very close. However, due to the lower than ideal traffic, 4 sessions skews the percentage when there were 11 sessions during that instance.

Since the predictive model works, it's time to evaluate the marketing campaigns effectiveness.

2018 Actual vs Predicted, Predicted Lower, and Predicted Upper
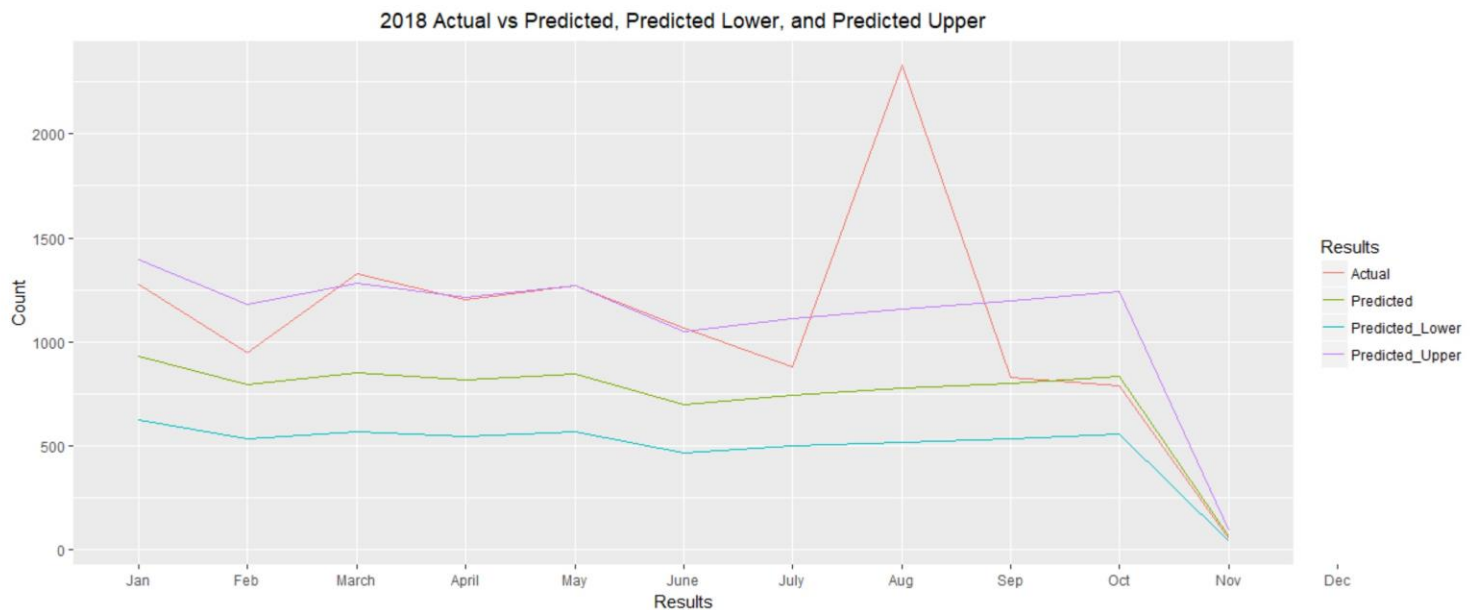
This demonstrates that both marketing campaigns were successful and made a significant difference to the websites traffic. The actual traffic that was straddling the lower predicted values, is now overtaking the upper predicted values. It is very evident during the Google Display campaign was by far the most effective.

The Mean for 2018 traffic during the first marketing campaign is 41.23 and for 2017 it's 31.95, a year-to-year mean increase of 9.28, which is significant. Regarding evaluating the effectiveness of the first digital marketing campaign, the predicted mean was 27.46 and the actual mean is 41.23.

Therefore, the first marketing campaign increased mean traffic by 13.77. So, it's safe to conclude that it was quite successful.

It's easy to see the success of the second marketing campaign, since the actual traffic is much greater than what was predicted. The mean traffic between August 14th, 2018 and September 1st, 2018 is 99.17. There's a mean increase of 65.39 over the same time period of 2017 and an increase of 73.68 between predicted and actual traffic in 2018. An increase of almost 300%.

It's quite evident that the second marketing campaign that included Google Display was much more successful that AdWords alone.

Client Recommendations:

It's clear that online marketing campaigns are effective for Compassion Home Care. The first campaign saw an average increase of 13.77 Sessions a day over predicted traffic and the second marketing endeavour was responsible for an average of 73.68 more daily Sessions. Therefore, the first recommendation would be to do more online advertising.

Digital marketing is often less expensive and more targeted than other mass advertising mediums like, TV, Radio or Print. Knowing the success rate of online marketing, along with the other benefits, should make it clear to the client to do more of it.

The second recommendation would be to advise the client to always incorporate some form of display advertising given the overwhelming success. There were an average increase of 73.68 Sessions a day when display ads were used in conjunction with paid search.

Future Considerations:

The client should consider doing more analysis on other website metrics that's important to their business. For instance, examining time on site or pageviews per session could provide indicators to user engagement and how customers, along with potential customers interact with their website. Analyzing more metrics could help Compassion Home Care capture more business, instead of just focusing on website traffic.

By having a greater understanding of how customers make their purchase decisions and knowing what's important in their decision making would be extremely beneficial to increasing Compassion Home Care's market share.