**Capstone Milestone Report**

**May 12th, 2018**

Review of the Problem:

* Companies do not know if a Pay-Per-Click (PPC) campaign is successful and/or know the true results of the endeavour, because 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 a Google AdWord campaign. From this, we will be able to determine how successful (or unsuccessful) the AdWord campaign was by examining the relevance 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 campaign would be a good fit for them.
* Gaining deeper insights into this marketing endeavour will give them a more accurate assessment of the marketing ROI.

Data:

*Limitations of Data:*

* 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.
* The lack of any type of online sale, makes it impossible to determine a monetary value of web traffic.
* Because the client does not track how their customers are acquired, it’s difficult to know the importance of the website traffic.

*Fields:*

* I took 6 years of Google Analytics data gathered from their company website, to get a better picture of basic metrics of their web traffic.
* My original data set included daily web traffic starting on March 12, 2012 until March 12, 2018.
* The website metrics used were:
  + Sessions
  + Number of Sessions per User
  + Average Session Duration
  + Pageviews
  + Pageviews per Session
  + Users
  + New Users
  + Bounce Rate
* I was able to download each metric on its own from the Google Analytics platform, and then I merged them into a single dataset.

*Cleaning & Wrangling:*

* The next step was to examine the data to determine the classifications of each variable, to look for missing values and any outliners. I found only one missing value that also included an outliner in the Sessions field that was 700% larger than the largest value. After looking at that record in the Google Analytics data, I found that the large value was due to bot traffic, so I removed the missing value as well as the outliner record.
* I changed the classification of the Date variable from a factor to a Date & changed the rest to numerical.
* To gain better understanding and to help visualize the data, I then created the following subsets for each variable from the main dataset:
  + 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

*Findings:*

* There is a solid increase in the number of Sessions year over year.
* There also appears to be some seasonality in the data as well.
* More exanimation is required

*Next Steps:*

* Need to do a deeper dive into the data and look at all the variables to see if all the variables have a positive linear relationship and if seasonality does exist.
* Decide on what variables are most important to analyze.
* Create dataset of web traffic during the PPC campaign & compare to historical traffic, looking for statistical relevance & determine marketing success.
* Use tools such as Facebook Prophet to build predictive models on what the web traffic would have been without a PPC campaign.
* Use methods such as t-tests, z-scores, and a chai-square analysis to find the statistical relevance of the predictive models.
* Prepare the necessary visual components to tell the story to the client.