

KOAA Facebook Advertisement Project Report

1. Introduction

Social media marketing is the use of social media platforms and websites to promote a product or service. Although the terms e-marketing and digital marketing are still dominant in academia, social media marketing is becoming more popular for both practitioners and researchers. Some examples of popular social media platforms over the years are Facebook, Instagram, TikTok, Twitter and Snapchat. Facebook is the leading social media platform where users can hyper-target their ads. Our sponsor KOAA, an online guide to everything such as where to go for kids, teenagers and families in different regions, also uses Facebook as their advertising platform to obtain more new clicks on their website. Therefore, our team wants to help KOAA to understand the influence of Facebook advertising on its website visitor growth, and to obtain more market shares in a shorter period of time. In this project, we collected data provided by our sponsor Debra Ross and generated some exploratory analysis to identify the influence of Facebook advertising on different times and dates. After visualizing the data, We applied analytical techniques to help KOAA determine if Facebook advertising accelerates the revenue curve and how, for example, when is the optimal time to spend money on Facebook advertising, and what are the optimal investment strategies in Facebook advertising in different months and regions. Finally, based on the results, we generated our final conclusion which will be presented at the end of this report.

2. Data

2.1 Dataset Description

As for our original dataset, the data includes eighteen new sites launched in 2017 with Facebook advertisement and two mature sites Rochester and St. Louis to compare. And for each month and region, there are 6 different types of data: unique visitors in month, Facebook expenditure, readers who click through Facebook, number of sessions in month, number of page views in month and % of total readers in month that came from Facebook. In order to align with our main objective, we focused on 4 types including Facebook expenditure, readers who click through Facebook, unique visitors in month and % of total readers in month to generate our visualization and model development

2.2 Data Normalization

As the dataset given contains data from different cities, we first divided the unique visitors per month of each site by the total urban population of each site to eliminate the regional population difference when conducting the analysis. Then, we used the Min-max normalization method to reshape all the data values into the same scale between 0 to 1. Below shows the formula of the Min-max normalization method:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

3. Exploratory Analysis

3.1 Correlation Matrix

3.1.1 Milwaukee

Our sponsor informed us that she intentionally changed the Facebook Advertising expenditure to see the effect on their visitors in KOAA's Milwaukee site, so we first plotted the correlation matrix of Milwaukee along. (Fig 3.1) The matrix contains the correlations between Facebook Expenditure, Unique Visitors, Click Through Facebook Rate and the Change of Unique Visitors and these values are all input as per month unit.

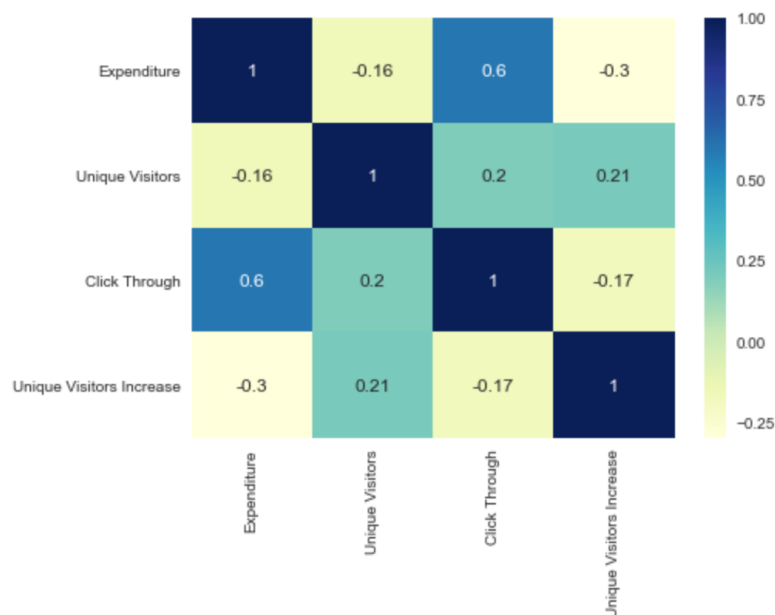


Fig 3.1

From the matrix, we found that there is little correlation between Expenditure and Unique Visitors(-0.16) or Unique Visitors Increase(-0.3). We reasoned that there are other factors having more influence on unique visitors such as seasonality or regionality. However, there is a significant correlation between Expenditure and Click-Through Facebook Rate(0.6) and Click Through Facebook Rate is where the increase of Facebook advertising expenditure has the most influence on, which means it should most directly reflect the effectiveness of expenditure of Facebook advertising.

3.1.2 2018 vs 2019

We also plotted the same correlation matrix for sites applied to Facebook advertising in 2018(Fig 3.2) and in 2019(Fig 3.3) to see the different effects when Facebook advertising is deployed one year earlier and one year later.

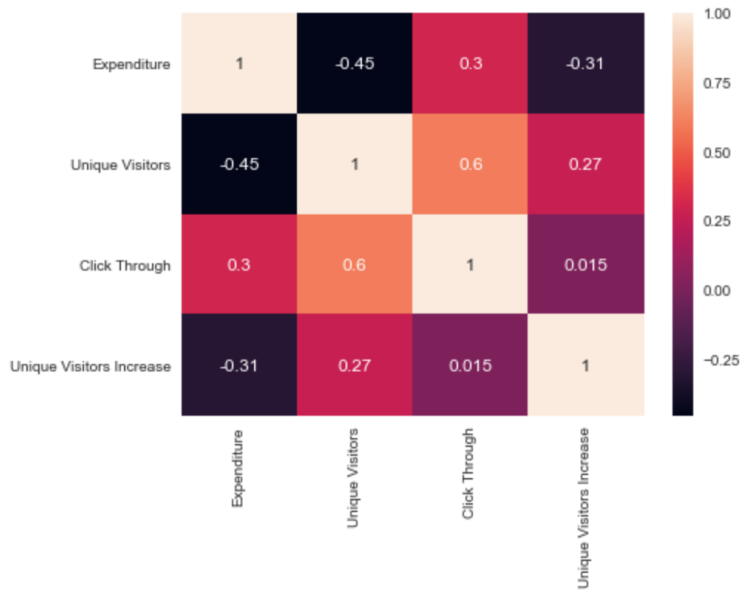


Fig3.2

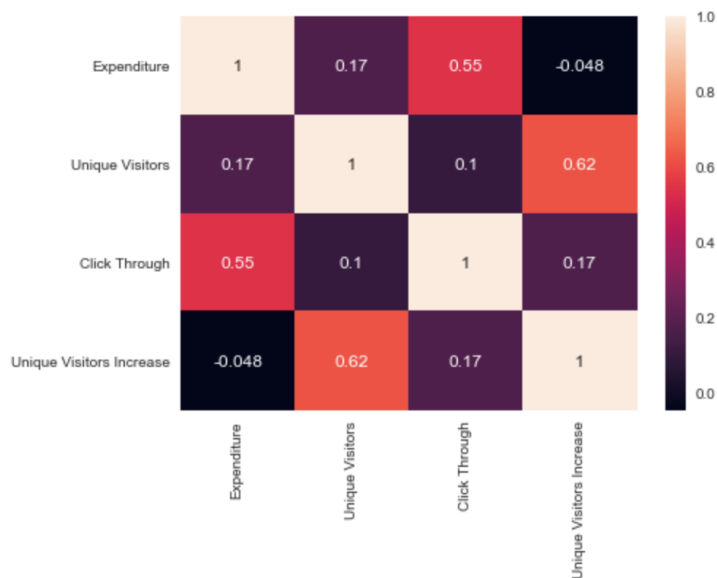


Fig 3.3

There two interesting observations:

1. The correlation between Unique Visitors and Click-Through Facebook Rate is higher in 2018(0.6) than in 2019(0.1). We reasoned that due to the presence of other search engines like Google, part of the potential customers would be recommended to KOAA in 2018 so in 2019, they were already using KOAA and therefore would not access the website through Facebook advertising.
2. The correlation between Expenditure and Click-Through Facebook Rate is higher in 2019(0.55) than in 2018(0.3). As we discovered from Milwaukee's correlation matrix, Click Through Facebook Rate is the most direct indicator of the effectiveness of Facebook advertising. It indicates the Facebook advertising is doing a better job in 2019 than in 2018 and the reason

why it is doing better in 2019 maybe that in 2019, with the information gathered in 2018, the sponsor could better specify the range of objects she wished Facebook to advertise to and Facebook too, can more accurately spot the potential customers that will click into the advertisement based on the 2018's data.

Another aspect we intend to compare is the effect of Facebook advertisements on each month of the two years. From Fig 3.4, we could conclude that spending more money during the early stage (2018 curve in graph) could bring more new clicks. There is also some observable seasonality in this graph, from March to May, the percentage decreases gradually. This may result from the school times where most of the kids are in school. After May or June, there is a rebound in the percentage of people who click Facebook Ads, which we propose is due to the beginning of summer break. Also, we can observe that there is a sharp increase in November, where Thanksgiving lies. This is an intuitive and informative graph for analyzing seasonality and trends. This result also corresponds to our further ANOVA analysis on number of clicks per month.

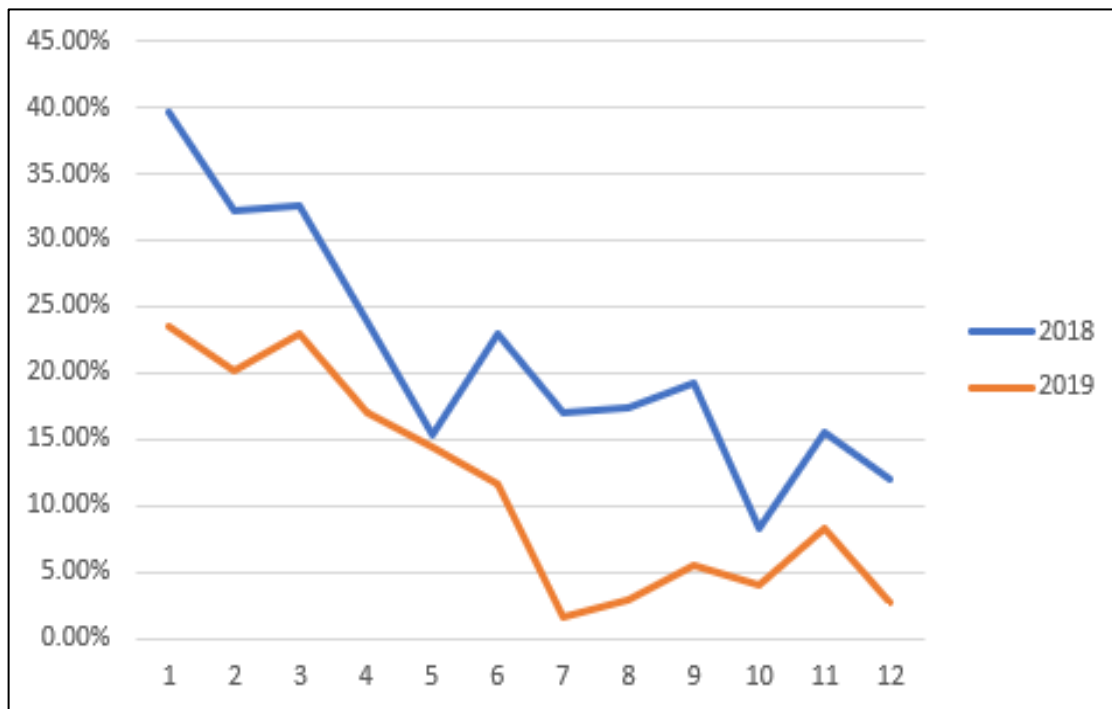


Fig 3.4

3.2 Line charts

3.2.1 data preparation and preprocessing

Knowing that there are 18 sites that are conducting advertisements on Facebook, we would like to learn about the effectiveness of such investment. To put it another word, do these advertisements help KOAA realize its initial objective, which is accelerating the website's visitor growth. In order to draw such a conclusion, we need to know what the visitor's natural growth curve looks like when there's no advertisement involved, and by comparing it with the growth curve of our advertising sites, it would be possible to observe the distinction between those curves and provide some insights. The sites that we selected to generate the natural growth curve are Atlanta, Phoenix and Minneapolis, and the reason we

selected these three sites is that their launching years are between 2015-2016, which is close to 2017, the launching year of our advertising sites. As we all know that the Internet and the access to it has been expanding and changing rapidly, for example, the internet usage penetration among adults never stops growing [1] and more internet users go online from their mobile devices instead of sitting home and opening browsers from their laptops [2]. Therefore, we would like to rule out the effect of those factors and try to keep Facebook advertising as the only changing variable. Imagine If we include sites that were launched in 2013 or 2012, some elder adults did not have access to the Internet or mobile broadband subscription at that time. Then if there's a slow growing curve of visitors, it would be attributed to the lower availability of Internet instead of the missing Facebook advertisement. That's why Atlanta. Phoenix and Minneapolis are the choices that we made to draw the natural growing curve. Besides that, from the 18 advertising sites, we decided to remove MidCities, which is a small region located between Dallas and Ft Worth. The reasons behind this dropout have two folds. Firstly, no population information for Midcities was given because of its geographic location, and secondly, the volume of visitors in Midcities is extremely low, which is negligible compared to all the other sites.

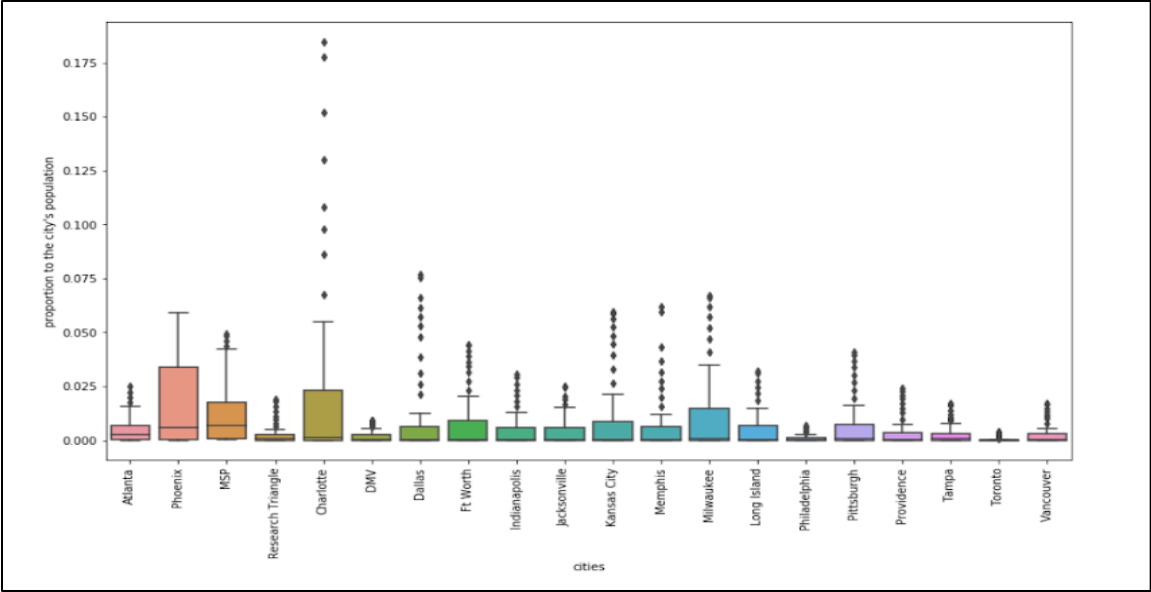


Fig 3.5

After including three non-advertising sites and dropping one trivial site, we first would like to have a general picture about the proportion of one-year cumulative visitors in each city to that city's total population. We adopted a boxplot to explore that information, which is shown in Fig 3.5. It's not hard to see that Charlotte is an outlier in this case. Most of our sites have a ratio no bigger than 0.075, but Charlotte goes far beyond that range, which could significantly affect our generation of growth curves. Therefore, we ultimately remove this outlier from our site collection, which leaves us with 19 sites including 3 sites without any advertisement, 8 sites advertised from Mar 2018 to Feb 2019, and 8 sites advertised from Mar 2019 to Feb 2020. Table-1 summarizes these three groups on which we will generate our growth curves.

Advertisement Categories	Collection of sites
--------------------------	---------------------

No ad	Atlanta, Phoenix, Minneapolis
Mar 2018-Feb 2019	DMV, Ft Worth, Indianapolis, Jacksonville, Milwaukee, Long Island, Pittsburgh
Mar 2019-Feb 2020	Research Triangle, Dallas, Kansas City, Memphis, Philadelphia, Providence, Tampa, Toronto

Table-1

3.2.2 No Ad vs With Ads

To show the comparison in an intuitive way, we calculated the average for each group and adopted a line chart to visualize the visitor growth pattern for these three groups of sites, which is shown in Fig 3.6. The x-axis is the time after the site launching date, and the y-axis refers to the normalized value after applying the min-max normalization techniques. It can be clearly seen that those sites with Facebook advertisements generally grow faster than those without advertisement.

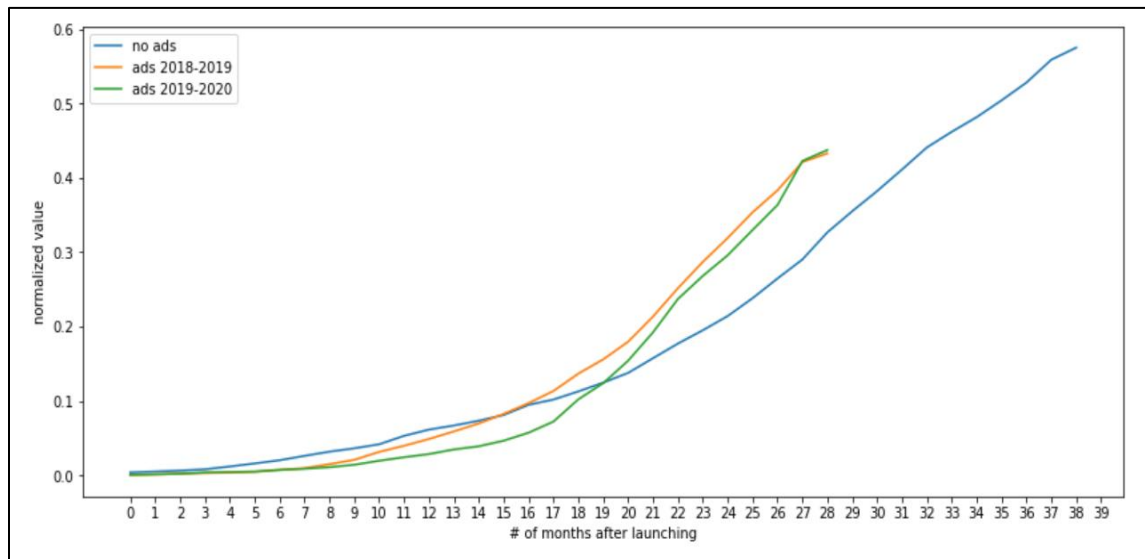


Fig 3.6

3.2.3 Growing curve with 2018 ad vs Growing curve with 2019 ads

If we take a closer look at these two advertising groups, as shown in Fig 3.7, we can even see that those sites started advertising from 2018 have a faster growth than other sites started advertising from 2019, but these two lines finally meet each other in Jan 2020.

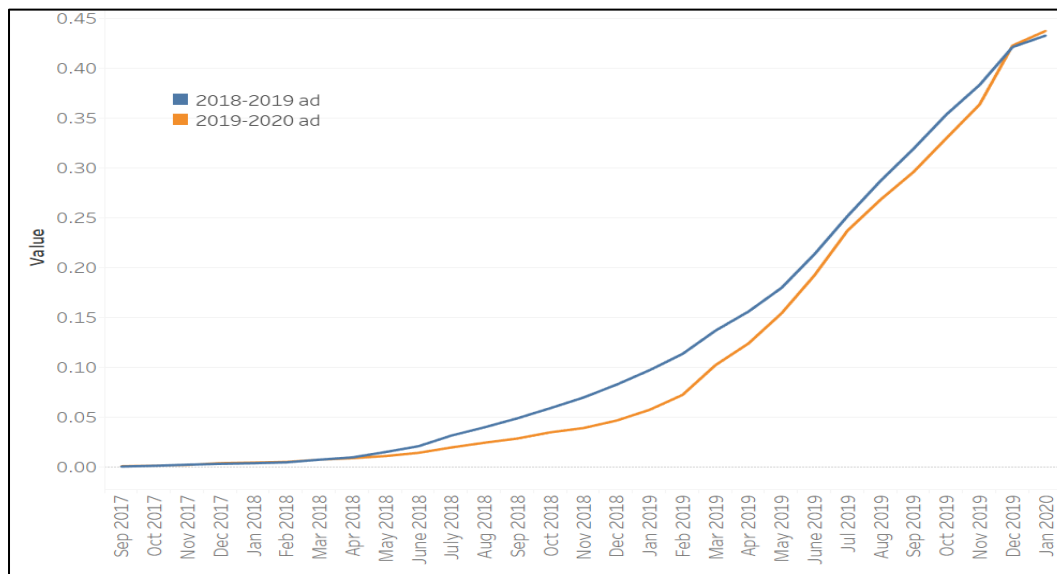


Fig 3.7

If we divide the time interval into two parts, more insightful details can be revealed. According to the left plot of Fig 3.8, which shows the growing trends of two advertising groups from March 2018 to Feb 2019, the gap between the first group who had advertising from 2018-2019 and the second group who didn't is gradually increasing throughout the entire advertising period. However, when KOAA started advertising the second group and ended the first group's advertisements from March 2019, as shown in the Fig 3.9, the growth of the visitors in the first group slows down, while the growth of the second group is boosted and accelerated by the advertisement. The gap between these two lines gets smaller and finally vanishes, which means the second group gets almost the same volume of visitors as the first group in the end. However, if we keep the first group's advertising, it will be extremely hard for the second group to catch up with. Therefore, we consolidated the sponsor's guess that Facebook advertising can definitely accelerate the visitor growth of KOAA and earlier action can help the website obtain more market shares in a shorter period of time.

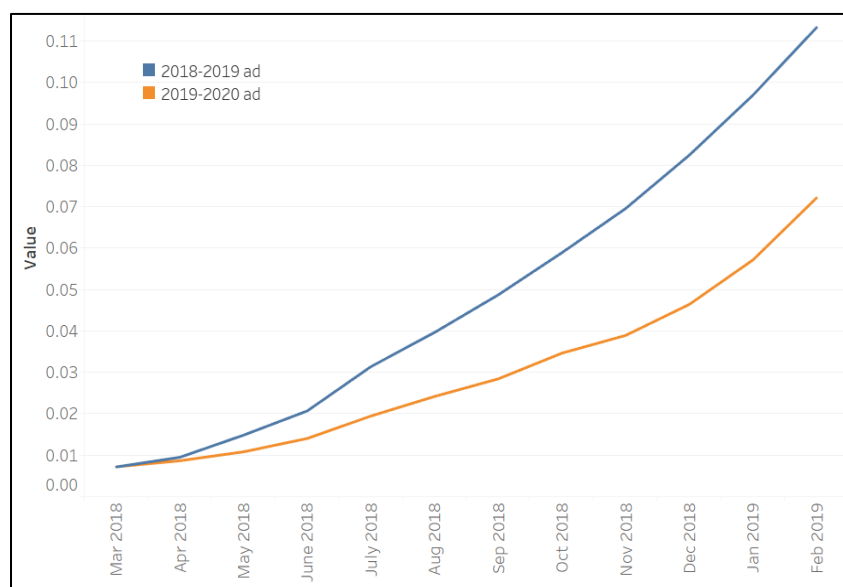


Fig 3.8

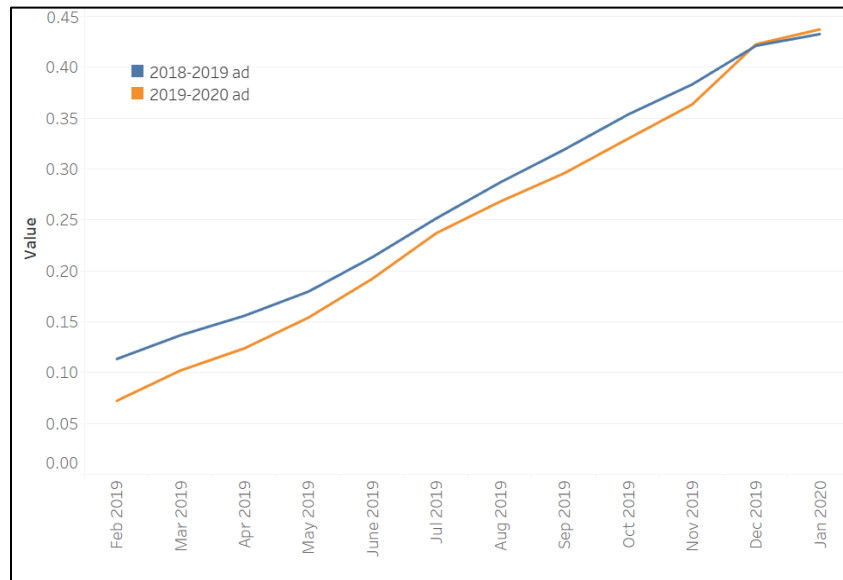


Fig 3.9

4. Statistical Model and Analysis

One-way analysis of variance (ANOVA) is used to test if there are any significant differences between means of three or more independent groups. If it returns a statistically significant result, we can conclude that there are at least two group means that are statistically significantly different from each other. However, it cannot tell which specific groups were statistically significantly different from each other, which requires further analysis through post-hoc tests [4]. Tukey's test is the type of post-hoc test that we are going to apply to our ANOVA result. It compares the means of all treatments to the mean of every other treatment, performs comparison between specific groups and finds out whether these groups are significantly different or similar [5].

Seasonality is definitely an important factor to consider when developing the optimal investing strategy because of the nature of KOAA. Therefore, we conducted a one-way ANOVA analysis to explore the impact of different months on Facebook clicks. We want to know if there's any difference in average Facebook clicks among 12 months, and if there is, which ones have higher clicks, and which ones have lower. To accomplish this, we only need to worry about the month dimension instead of year, which leaves us with 12 rows showing all months and 18 regional columns. Each cell in our data would be the number of unique readers who click through to KOAA from Facebook advertisement in that month, as shown in Table-2. Based on this, we did some exploration work and created a boxplot to get an intuitive look at the Facebook clicks in all 12 months.

	Research 1	Charlotte	DMV	Dallas	Ft Worth	Indianapoli	Jacksonvill	Kansas Cit	Memphis	Milwaukee
Mar	255	123	121	371	150	199	142	284	455	174
Apr	348	156	85	402	90	126	114	345	487	151
May	574	295	186	641	348	251	392	687	930	378
June	682	304	184	644	349	236	342	506	772	336
July	790	271	165	640	277	146	286	758	701	267
Aug	501	417	203	353	329	175	347	433	403	353
Sep	100	325	195	83	274	154	282	79	47	455
Oct	253	302	245	269	284	321	389	75	71	337
Nov	315	355	306	370	295	234	332	239	266	331
Dec	206	179	215	228	167	132	180	226	336	131
Jan	336	491	437	209	365	227	336	54	77	427
Feb	278	377	362	41	355	257	380	97	86	275

Table-2

In accordance with Fig 4.1, it's not hard to see that May, June and July appear to have more Facebook clicks, while it looks like things do not go well in Sept, Apr and Dec which seemingly have lower clicks. However, it's all about our guesses, and some statistical analysis would be inevitable if we want to prove that is the case.

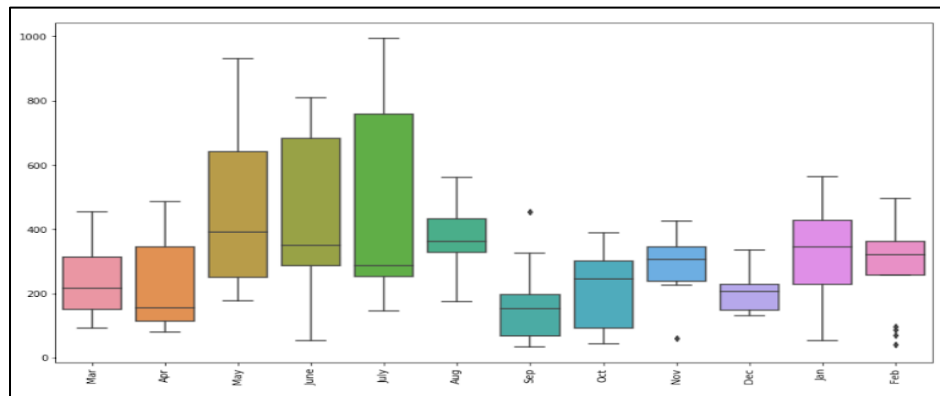


Fig 4.1

Our null hypothesis in this case is that every month has the same number of readers coming from Facebook, and our alternative hypothesis is that at least one month has a different number of readers coming from Facebook than the rest months. Table-3 summarizes the result of our ANOVA test.

	sum_sq	df	F	PR(>F)
C(months)	2.477508e+06	11.0	8.115895	1.375404e-11
Residual	5.328282e+06	192.0	NaN	NaN

Table-3

The p-value is small enough ($1.375e-11$) to reject a null hypothesis ($p\text{-value} < 0.05$), so we can conclude that there is at least one month having a different number of Facebook clicks than the other months. With that in mind, we can perform the Tukey's test (Post-hoc test) to determine the specific months in which we have higher clicks and lower clicks. In Fig-4.2&4.3, each line represents the 95% confidence interval centered around the average clicks of that month.

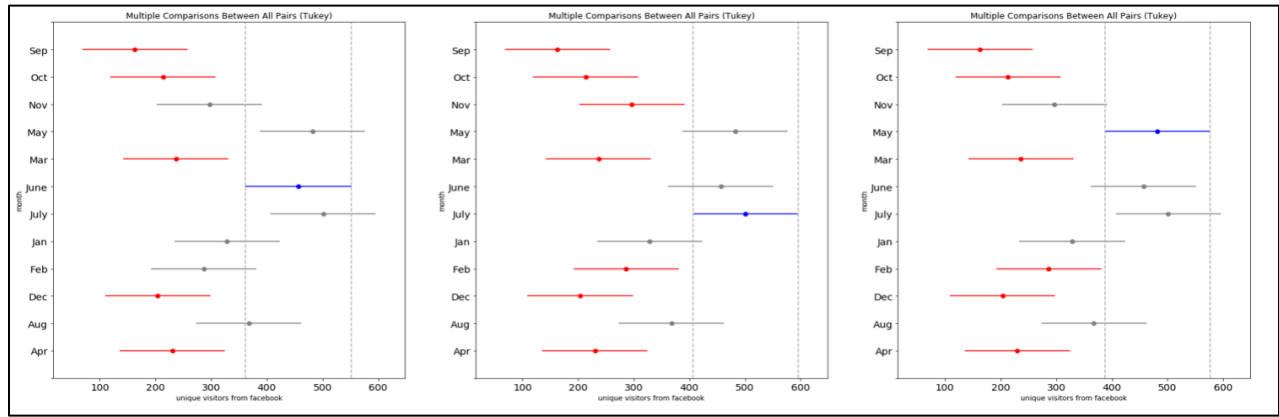


Fig 4.2

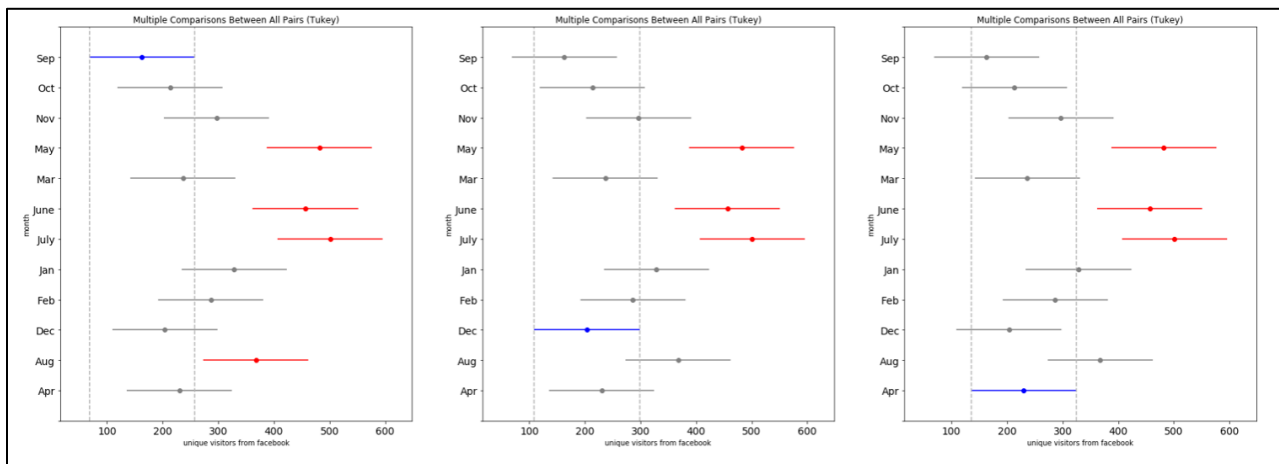


Fig 4.3

As a result, the three months that have the highest clicks are exactly what we expected, which are May, June and July, and we highlighted them in Figure-4.2 using blue lines. The red lines represent all the months that have significant lower clicks than the blue ones, and we are 95% confident about that. For example, based on the left plot in Figure-4.2, we can say that we are 95% confident that Sept, Oct, Mar, Dec and Apr have lower clicks than June. The reason behind these three months having the highest clicks is possibly because they are in the middle of summer times, and summer is usually the prime time for families or friends to enjoy the sunlight and join in some outdoor activities together, especially when kids and teens are having their summer vacation. As we know that KOAA is the place to find the comprehensive calendar and activities in local communities for kids, teens, and families [3], higher clicks to the website during such times make a lot of sense. We would suggest that KOAA could invest more on Facebook advertising during these months as they tend to bring a higher click through rate from Facebook and therefore direct more traffic and readers to the websites. Yet, things do not go well during Sept, Dec and Apr as we expected because they are the three months that have the lowest clicks, marked as blue in Figure-4.3. Red lines show the months which have significant higher clicks. December is where Christmas is, and people would like to stay home and spend time with their family. April is usually the period of midterms and finals, so kids and teens are busy reviewing and studying, while September is when the semester begins. Above reasons contribute to the inefficiency of Facebook

advertisements during these three months, and we would suggest that KOAA could cut their spending to some extent since the conversion rate from Facebook users to the website’s reader will be low in such time.

5. Further Exploration

5.1 Cost Efficiency Analysis

We have already conducted a significant amount of analysis regarding the effect of Facebook on KOAA itself, but it is not clear how is the performance of Facebook Ads on KOAA compared to the benchmark of its industry. Therefore, we approached this problem by analyzing and comparing the cost efficiency of Facebook Ads.

The cost efficiency of the advertisement is calculated through a variable called Cost Per Click (CPC). It determines the average cost for KOAA to spend on Facebook Ads in order to get a click.

Cost Per Click = Total Costs on Advertisements/ Total Number of Clicks

Based on the data from WordStream, we found that the benchmark Cost Per Click for the “Travel and Hospitality” industry is \$0.63 per click. And we conducted the same analysis on the Facebook Ads of KOAA.

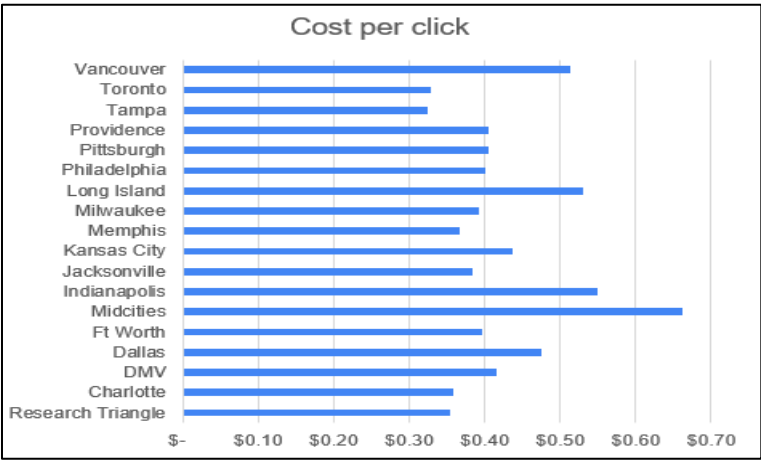


Fig 5.1.1

From Fig 5.1.1, the result reveals that the overall cost of KOAA is lower than the benchmark of the industry, except Mid Cities. Therefore, it can be concluded that Facebook Ads for KOAA is very cost efficient

5.2 Regional Analysis

We explored the regional differences of the unique visitors per month. All the sites with Facebook advertising can be divided into four main regions, Canada, Midwest, Northeast and South. We first plotted the general trend of Unique visitors for each region (Fig 5.2.1) and we found that Midwest and south are the two largest markets for KOAA.

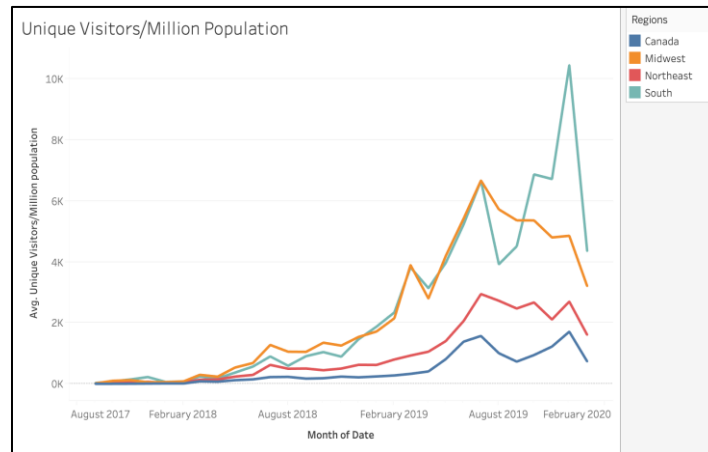


Fig 5.2.1

In order to find the seasonal difference of the unique visitors in each region, we also plotted the average unique visitors per month of each region (Fig 5.2.2) and we found that in winter times, October, November and December, all regions' unique visitors drop except the south region's. We think the main reason is that when the other three regions are experiencing cold and harsh weather during winter, the weather in the south is actually cool and comfortable and that is why more people are using KOAA in the south to take the kids out during winter.

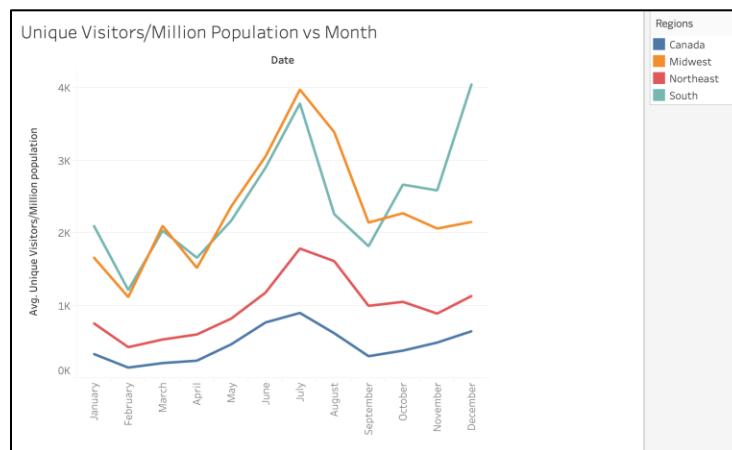


Fig 5.2.2

5.3 Maturity of growth

How many visitors can KOAA finally get, and should we keep advertising it persistently? With that question in mind, we would like to explore the limit of our website's visitor growth. We carefully selected six sites which should have already reached their maturity in their respective region, which are St. Louis,

Albany, Buffalo, Fairfield County, Houston, Westchester, and most of them were launched during 2011 - 2013. By plotting the ratio of the site's one-year cumulative visitors to its total population, as shown in Fig 5.3.1, we can easily tell that their one-year cumulative visitor's number fluctuated in recent years, which could be a signal of maturity.

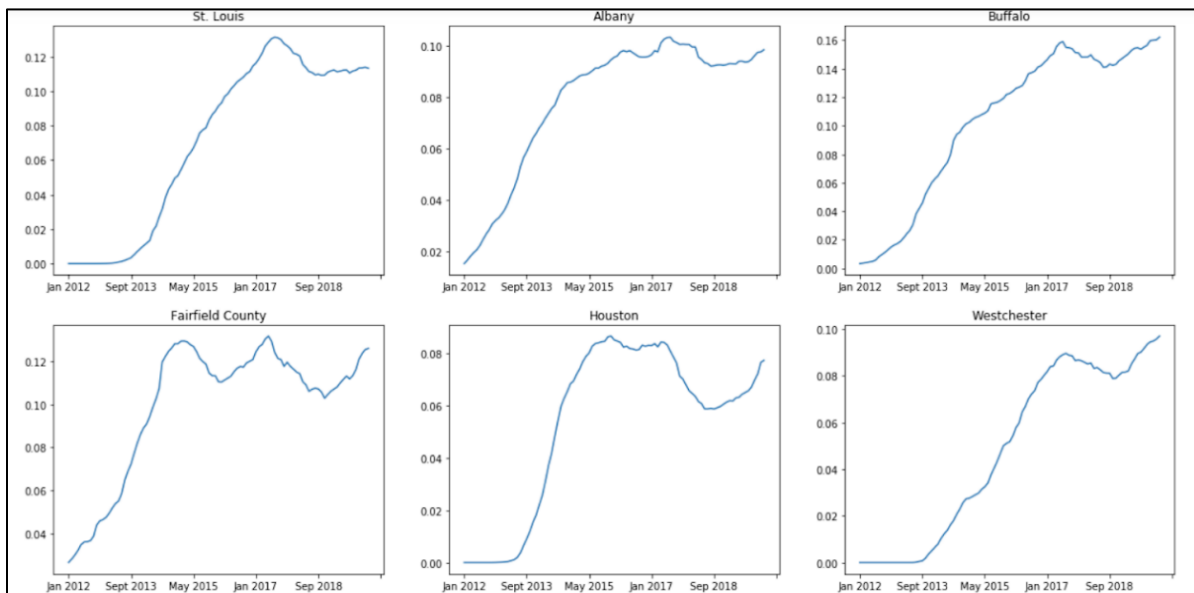


Fig 5.3.1

Rochester was not included here because KOAA started in Rochester and operated for nearly 20 years as a local site, so the number of readers in Rochester is not even close compared to all the other sites and it would be an outlier if we included it here. We also calculated the maximum ratio for all six sites, and Table-4 shows the result.

St. Louis	0.13
Albany	0.1
Buffalo	0.16
Fairfield County	0.13
Houston	0.09
Westchester	0.1

Table-4

The number in Buffalo is obviously higher than the rest 5 sites, and it's probably because of its vicinity to Rochester where KOAA originated. By ruling it out, we can say that 0.1-0.12 of the total population is a reasonable estimate of a site's maturity because the above numbers are the maximum, which means they fluctuate below that number and never go beyond that. Compared with those mature sites, we can visualize the same thing on our advertising sites. Fig 5.3.2 shows the growing curves for Dallas, Philadelphia, and Pittsburgh, which are randomly selected from our 18 advertising sites.

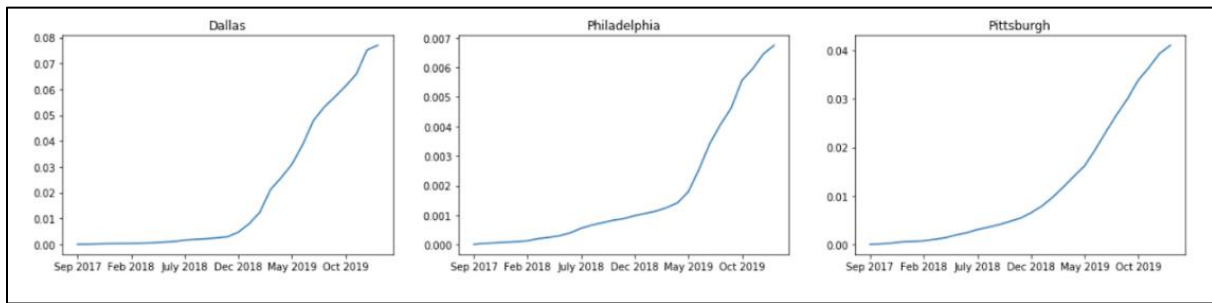


Fig 5.3.2

It's obvious that they are still growing rapidly and far away from the maturity range that we estimated, which means we can still employ Facebook advertisements to accelerate their growth. But once the number of one-year cumulative readers reach 0.1-0.12 of the site's total population and that ratio does not really change much for a few months or so, we can stop Facebook advertising to avoid unnecessary spending since it is very likely that this site has already reached its limit.

6. Conclusions

Our conclusion can be summarized into four parts:

1. We believe Facebook advertising is effective.
Facebook advertising will accelerate visitor growth and help sites gain a higher volume of visitors in a shorter period of time, especially in the early stage of those emerging sites.
2. Facebook advertising is also cost-efficient.
The "Cost Per Click" of KOAA Facebook Ads in most of the regions is below the benchmark of the industry, except Mid Cities
3. The best time to apply Facebook advertising is May, June and July.
Seasonality is a crucial factor to consider when spending money on Facebook advertising. we encourage KOAA to spend more during months like May, June and July, in which the click-through rates are high while reducing some spending during months where the click-through rates are low.
4. South and Midwest should be the two largest markets.
If KOAA plans to expand, cities in the South and Midwest regions should be considered as the South and Midwest have the largest potential market to deploy Facebook advertising, especially the southern region during wintertime.

7. Next Steps

During this hard time, it's crucial for KOAA to know the strike of COVID-19 on its business so we can do some analysis using data from March and April to show that impact on its visitor's growth. Furthermore, based on the geographic and sociodemographic similarities, we want to make predictions on the effects of Facebook advertising in areas that KOAA hasn't deployed Facebook advertising yet. Lastly, We have explored the effect of Facebook advertising on the growing sites, with more data of the

matured site with Facebook advertising, we would like to explore the effect of Facebook Advertising on the matured site as well.

8. References

- [1] Clement, J. "U.S.: Internet Penetration." *Statista*, 18 June 2019, www.statista.com/statistics/185700/percentage-of-adult-internet-users-in-the-united-states-since-2000/.
- [2] Clement, J. "U.S. Mobile Broadband Subscriptions per 100 Inhabitants 2017." *Statista*, 1 Oct. 2019, www.statista.com/statistics/187104/mobile-broadband-subscriptions-per-100-inhabitants-in-the-usa-since-2004/.
- [3] "Debra E. Ross, Publisher, KidsOutAndAbout.com." *Kids Out and About Rochester*, rochester.kidsoutandabout.com/content/debra-e-ross-publisher-kidsoutandaboutcom.
- [4] "One-Way ANOVA." *One-Way ANOVA - An Introduction to When You Should Run This Test and the Test Hypothesis | Laerd Statistics*, statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php.
- [5] "Tukey's Test for Post-Hoc Analysis." *Aaron Schlegel's Notebook of Interesting Things*, 7 Sept. 2018, aaronSchlegel.me/tukeys-test-post-hoc-analysis.html.
- [6] "Facebook Ad Benchmarks for YOUR Industry [Data]." *WordStream*, www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks.