Facebook Page Growth Modeling

Report by James Fulford

# Introduction and Motivating Question

Facebook marketing is essential for small organizations and businesses. Whether one is arousing public awareness, engaging potential clients, or attracting donations, posting to Facebook is a great place to start. However, effective Facebook marketing involves paying to promote a post. Before measuring the effectiveness of paid advertising, it would be prudent to establish a baseline for comparison.

This report aims to establish a baseline model to answer this question: how do small Facebook pages grow organically over time?

# Overview

Eight sample datasets were collected, each with metrics pertaining to the Facebook page of a different small non-profit organization all in the same region. Data was gathered for thirteen weeks prior to collection date (numbered 0 through 12). That is, week 0 ended on Tuesday, 11/15/2016. The week increments on Tuesdays to avoid splitting potentially active weekends into separate weeks.

The Lifetime Total Likes (LTL) on a page, grouped by week, was used as a proxy to the growth of a page’s audience. The value is cumulative, growing every time a new person likes a page. Alternative measures such as post engagements and weekly page reach depend too much on the content and frequency of posts and is a better measure of the interest level of a page’s audience, not the number of people interested in the page.

In general, the LTL of a page goes up very slowly. Exceptions to this rule are when a page is first created or when something significant occurs, such as an outreach or someone influential mentioning the page. For the sake of clarity, an event is when something significant occurs on a page that causes a considerable increase in LTL. In these datasets, the events are mainly the creation of a page.

# Approach

Two different models were fit to each dataset: a simple linear model and a logistic model. If a Facebook Page started during the period of study, the model was fit to the data after the page was started.

(Note that in all models, R2 is a number between 0 and 1, where higher numbers indicate a better-fitting model.)

## Linear Model

A linear model can be used to describe and predict short-term behavior. It can account for lengths of time without any events. It assumes continual and constant growth.

The slope of each line describes how many new likes a page can expect in an uneventful week. Small slope values, as found in these datasets, indicates slow organic page growth. Lines for new pages will not accurately predict LTL prior to the creation of the page.

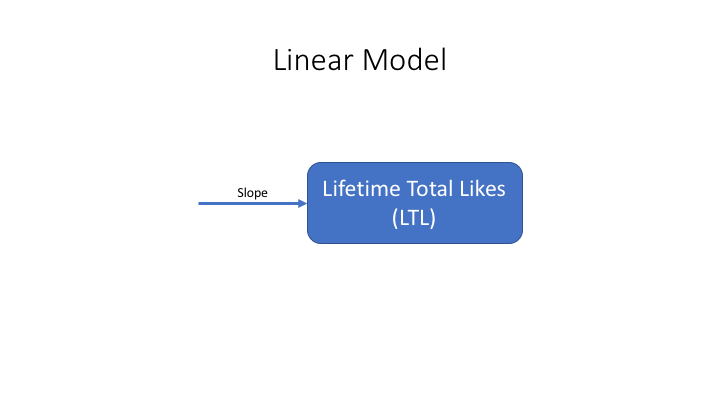
The starting value states how many likes a page would have had by the end of week 0 according to the model. For a line, this value is called the y-intercept, since it meets the y-axis at this height.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dataset** | **Slope** | **Starting Value** | **R2** |
| 1 | 0.32 | 816.93 | 0.3442\* |
| 2 (week 6) | 2.25 | 195.75 | 0.9709 |
| 3 | 0.09 | 43.52 | 0.5432 |
| 4 | 0.86 | 168.7 | 0.7279 |
| 5 | 5.21 | 21.64 | 0.8486 |
| 6 (week 1) | 9.01 | 431.74 | 0.808 |
| 7 | 0.77 | 887.62 | 0.625 |
| 8 | 0.37 | 392.5 | 0.9491 |

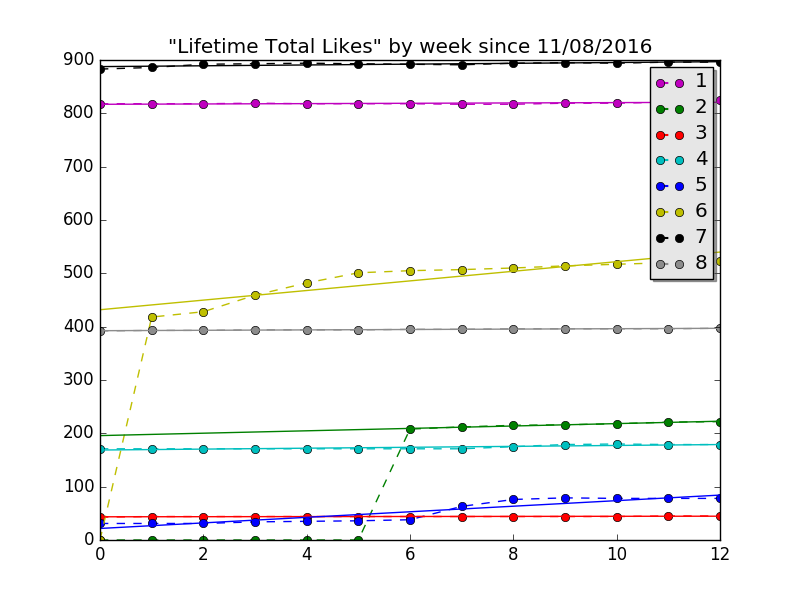
\* Week 12 has 4.25 more likes than model predicts. Graph of model and points hides this point behind legend.

If the page doesn’t have any events in the period of study, it seems the R2-value is higher. So, expected growth is usually quite low if nothing new happens, which matches intuition. Furthermore, if a page undergoes an event, the R2-value is lower. This second relationship can be used to detect events on a page. After detecting this event, a linear model loses its meaning because of the outliers, so a different model will be needed.

Sometimes, it proves useful to think of how this model would behave going from one week to the next. A valid way of interpreting slope is -as the number of likes added in a non-interesting week. So, starting at week 0 with the starting value, a recursive model can be fit to the data by the following relationship and initial condition, visualized with the following compartmental diagram:



A plot of each dataset’s data points (connected by dotted lines) along with its linear model (solid line) is provided below.



## Logistic Model

The logistic model is useful to describe medium-term behavior. It can be used to account for a single event.

A logistic model predicts quick growth around the single event as well as an upper limit on LTL until another event occurs. Intuitively, if no serious marketing has been done for a while, this value is likely close to the current value. This value can raise awareness for the need for marketing, if used properly.

The model is as follows:

Where Minimum is the smallest value LTL has taken in the dataset and Offset is the lowest valued non-zero week used mainly for the late start pages. The parameters of regression were not Minimum nor Offset, as those were trivial to find. The parameters of regression were C and r, which describes how strong the event was for LTL levels and C is relates to how the center of the logistic curve is shifted from Offset. More specifically, is the effective shift right from Offset, essentially giving when the single event occurred or is anticipated to occur.

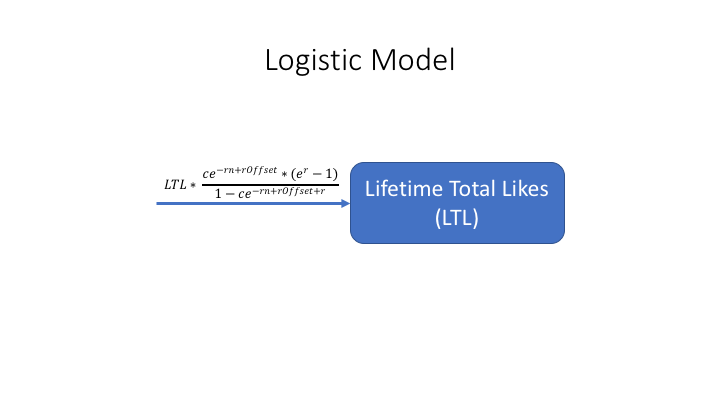
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Projected Maximum Likes** | **r** |  | **R2** |
| 1 | 462264 | 0.593 | 30.518 | 0.805 |
| 2 (week 6) | 223 | 0.746 | 2.646 | 0.955 |
| 3 | 45 | 6.959 | 0.756 | 0.44 |
| 4 | 180 | 4.053 | 8.019 | 0.996 |
| 5 | 79 | 2.24 | 6.69 | 0.993 |
| 6 (week 1) | 515 | 1.174 | 2.451 | 0.982 |
| 7 | 894 | 2.517 | 1.388 | 0.864 |
| 8 | 397 | 0.389 | 5.127 | 0.946 |

In the situation where a high C value is reported, the model envisions the page increasing in size a while into the future. If the upturn is predicted to be a while into the future (prophesied), chances are the model is locally linear, so a linear model may be better. This model is not designed to prophesy the strength and time of the next event, only to detect and account for very soon or already passed events.

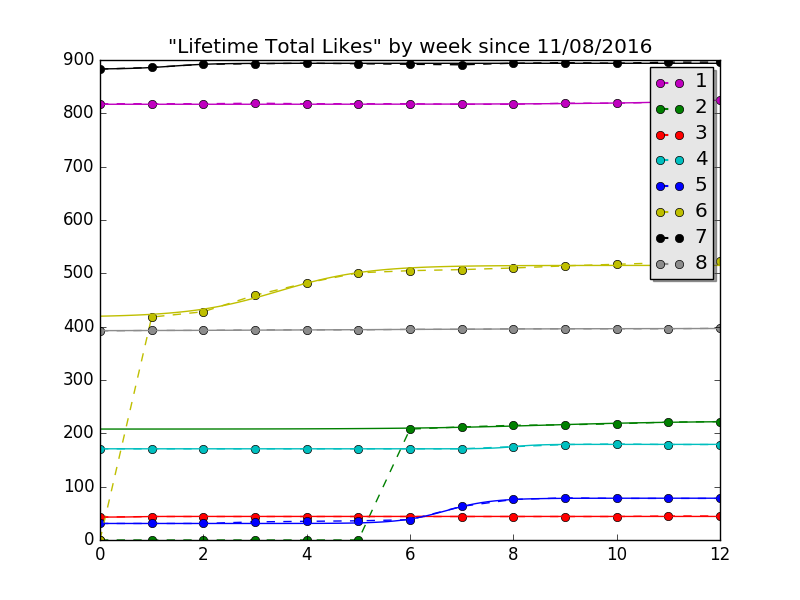
For example, dataset 1 anticipates an upturn over half a year later, which is well beyond the scope of the studied period. Referring to the linear model fit earlier, dataset 1 shows no indication of rapid growth anytime soon, so it is likely best to project dataset 1 with a linear model.

In the graph below, dotted lines connect points, which represent observations. Solid lines are the logistic models, plotted.

When the equation is forced into a recursive form at increments of 1 week, the difference equation can be viewed accordingly:



Let (on the right side of the equation) be called the multiplier. The long-term behavior of the multiplier is akin to after algebraic simplification, meaning that the increase from one state to the next always grows, yet the percent increase decays exponentially. An increase in r would accelerate decay. This recursive model also confirms that Offset and c act similar in that they simply delay the inevitable decay.



## Stochastic Logistic Model

This model would have been useful in medium-to-long-run predictions. It would have built on the logistic model by factoring in random events (environmental stochasticity), effectively playing out a Facebook page’s life if no organized marketing plan were to be put together.

To properly simulate random significant events, more data for each page would be required to roughly estimate the frequency and magnitude of significant events, assuming the frequency and magnitude were locally constant. As no dataset has many sample events to determine frequency from, this approach cannot be used for this dataset.

If one were to be formed, it would use the recursive formula outlined in the previous model. Every so often, according to some constant likelihood, the logistic curve would be rebased so that a great increase occurs again, followed by another plateau. Some way would have to be calculated to find a new upper limit and determine a new rate of increase r, both perhaps by some other random distribution modeling the reach and longevity (respectively) of the event.

This model has the weakness of loosely predicting the model, yet the strength of giving a general sense of where the page might go in the long term given the level of marketing used in the past.

# Conclusion

How do Facebook pages grow organically? The short answer is slowly. When there is no paid advertising, nor concentrated efforts on and off the platform, nor a new significant influence driving new growth to the page, it grows at almost a constant rate. If one of these events occurs, it grows suddenly, yet quickly approaches low linear growth again.

Means of measuring a campaign, advertisement, or other event were discussed under the logistic growth model, while means of detecting a random event were discussed in the linear model. Furthermore, the valuable results of average growth from the linear model and the growth cap from the logistic model can be effective in promoting the need for greater outreach in an institution, if it truly seeks to grow its audience in the modern world.

Datasets, images, python scripts (open with notepad if interested – just so it’s clear that I did use a computer to do this)