

# GOOGLE MY BUSINESS MACHINE LEARNING

Quantitative Analysis of AI Focus Operations on GMB

## Introduction and Overview

Thank you for downloading our newsletter. We are AI Focus, an artificial intelligence startup in Silicon Beach using machine learning to explore and develop new marketing technologies for our clients. The focus of this newsletter is the statistical analysis of the results of our Gbizbot, an online AI tool which drives customers to engage with brick and mortar businesses. You were invited to download our presentation because of your interest and expertise in statistics, big data and new business technologies. We welcome your thoughts, questions and critiques regarding our results and methodology. If you are interested in analyzing our raw data for an academic article or PhD project, please contact us at <https://www.aifocus> as we are happy to share our data.

### **We are AI Focus**

The primary goal of a business is to increase revenue. Revenue is an increasing function of the number of customers a business serves. Our company AI Focus leverages artificial intelligence and big data to drive customers towards our clients. Using machine learning, we determine how much value our clients are getting from their current digital marketing efforts and how their resources can produce better results. The most basic way clients measure the impact of digital marketing is through revenue.

To achieve increased revenue, we must manipulate the variables which have the

highest impact on revenue. For instance, increasing the number of people who discover our clients' website, increasing phone calls, or increasing the frequency of branded searches can all increase the revenue for a client. While there exists marketing industry standard methods to impact revenue, we leveraged our decades of marketing expertise with new technologies to develop advanced tools and methodologies capable of manipulating variables far beyond industry expectations.

### **Methodology**

To protect confidentiality, the data presented here has been anonymized and

codes will be used to represent various clients. The project presented herein is a basic comparison between 10 different businesses before the intervention of our Gbizbot and after our interventions, thus providing two sets of data for comparison. All of our client data is provided by Google for the Google My Business page and by Caller Insights for our client's tracked phone numbers. Here are the variables tracked on a daily basis.

## **Variables provided by Google for each client**

- 1) Total searches
- 2) Discovery searches
- 3) Total views
- 4) Search views
- 5) Maps views
- 6) Total actions
- 7) Direct searches
- 8) Website actions
- 9) Directions actions
- 10) Phone call actions
- 11) Total owner photos
- 12) Owner photo views
- 13) Total customer photos

We have found the following six variables have the most significant effect on revenue

and thus our analysis will focus on these. From past analysis, we know the other seven variables Google tracks are

functions of the six key variables stated below so we do not need to analyze all thirteen variables.

- 1) Direct searches
- 2) Website actions
- 3) Directions actions
- 4) Phone call actions
- 5) Owner photo views
- 6) Customer photo views

As Google provides data for these variables via their Google My Business page, for the purpose of this project we have sufficient, accurate, consistent data from a credible and universal source. As we have access to Google data recorded before our clients initiated our services, we can test the impact of our activities on our clients' trading position at different points in time.

## **Descriptive Statistics**

The table below shows the mean, standard deviations and confidence intervals for each of the variables 100 days before and after our interventions. For each of our clients, there are six variables with the last part of the variable name being the code assigned to the client for the purpose of anonymity. The 95% confidence intervals of the means shows that the bounds for all the variables before our intervention are lower than the bounds of the same variables after our intervention. In other words, the

confidence intervals do not intersect and thus the means of each of the variable was significantly lower before our intervention compared to the means of 100 days after our interventions.

Descriptives			
		95% Confidence Interval for Mean	
		Lower Bound	Upper Bound
Direct searche	Before	369.444	426.616
	After	556.655	692.751
	Total	471.528	551.205
Websiteaction	Before	24.38	29.818
	After	31.154	38.885
	Total	28.653	33.466
Directionsactions	Before	59.751	71.318
	After	75.163	90.758
	Total	69.285	79.21
GMB call actions	Before	15.066	18.974
	After	22.294	31.844
	Total	19.393	24.696
Ownerphotoviews	Before	1512.069	1917.495
	After	2307.086	2786.667
	Total	1967.891	2303.223
Customerphotoviews	Before	2415.916	2975.533
	After	3330.505	4285.023
	Total	2970.396	3545.729

## Analysis of Variance

The table shows the analysis of variance output for the six variables. We can easily

observe that in every case the p value is way below 0.05. In fact for this test we are able to test our hypothesis at 99% level of confidence. In all cases the null hypothesis that the means for these variables before our intervention and after are rejected with overwhelming evidence.

Variable	F	Sig
DirectSearch	37.124	0
WebsiteAction	11.057	0.001
DirectionAction	12.683	0
GMBCallAction	14.934	0
OwnerPhotoViews	27.63	0
CustomerPhotoView	15.779	0

## K Means Clustering

From the cluster centers table below we can see that the clusters for before and after are quite different. If this table is compared to the raw data, we can observe that the cases for all the last half of the 100 days after our interventions were quite high. That therefore means that some cases for the 100 days after were included in the 100 days before cluster. All the six variables exhibits and strong increasing relationship from past research by our analytics department. This fact can also be observed from the clusters in that all the cluster means after were higher compared to the cluster means before.

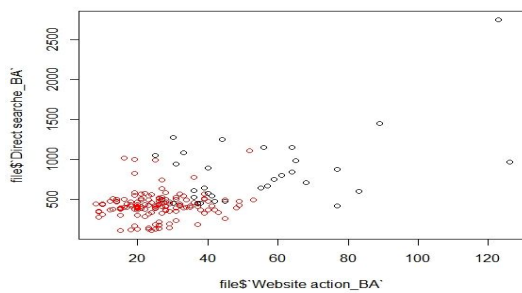
Final Cluster Centers
-----------------------

	Cluster	
	Before	After
Direct Search	489.7	2748
Websiteaction	30.4	123
Directionsactions	74	233
GMB collections	21.7	36
Ownerphotoviews	2132.8	2624
Customerphotoview	3153.6	21540

different implying that the data has been changing over time as the impact of our activity took effect.

### Three Means Cluster Analysis

#### Graph1



K-means clustering with 2 clusters			
Cluster means:			
Direct search	Website action	D_actions_BA	
1)	419.72	24.98	62.13
2)	736.56	47.65	110.86
GMBcalactions	Owner P_views	Cus_P_views	
1)	17.07	1621.45	2571.08
2)	35.17	3588.48	5199.54

From graph1 above we can clearly see that the web activity data can be clustered into two clusters. There is the first cluster that contains the web activity data for the period when we had not been signed up for marketing services. **The second cluster contains the data for the web activity after the impact of our interventions had taken full effect.**

The clustering means table 2 below shows the distribution of the means of the six variables across the two clusters. By observation we can see that the means for the two clusters of each variable are quite

### Percentage Changes

Percentage Change From Cluster 1 to Cluster 2		
Direct Search	Website Action	Direction Action
75.49%	90.73%	78.45%
Phone call Actions	Owners Photo Views	Get Directions
106.07%	121.31%	123.00%

The clustering vector below shows the distribution of all data points that did not have missing values for the 100 days before and after our intervention. We can easily observe that the data is divided into two clusters. The first cluster contains more 1's than the second cluster. The second cluster took most of the 2's. Given that the data is divided into two equal portions where we had 100 days before our interventions and 100 days after our interventions, we can easily observe that there are more 1's compared to 2's. This can be explained by the fact that there is a 20+/-5 days lag before the impact of our intervention is felt. From the tables above, we know that the 2's or cluster two represent the days when web activity was high. On the other hand, 1's represent the days when web activity was low. By observations we can see that 1's dominate the first half of the clustering vector which implies that most of the 100 days before our intervention had low web activity. For the 100 days after the 2's dominate which shows that most of the

days that had high web activity are in the 100 days after our interventions.

Clustering vector:

```

1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 1 1
1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 2
1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2
2 2 2 2 1 1 1 1 2 2 2 1 2 2 2 2 2 1 2 1 2 1 2 1 2 2 2 1 2 1
2 1 2 1 1 1 2 2 2 1 2 1 1 2 2 1 2 2 2 1 2 2 2 1 2 1 2 1 2 1
1 2 2 2 2

```

## CONCLUSION

Thank you for reviewing our data sets. We hope you found the presentation informative and interesting. The GbizBot, our artificial intelligence tool, is constantly learning and being upgraded, and we will update this newsletter in the future when we have 50 client data sets.

We also plan to release future newsletters with statistical analysis of the GbizBot results for the 12 month time frame and analysis of what befalls a client when the Gbizbot is turned off.

We welcome your thoughts on the results presented. If you are interested in the raw data, we are happy to share the anonymized data for your own analysis.

## References

- <https://docs.google.com/spreadsheets/d/1Lnxfy55nNyQ4HXrYVhCNpEQXIVqHf0jZPZzKfSFADOI/edit#gid=1506828419>
- <https://www.bigcommerce.com/e-commerce-answers/what-is-web>
- <https://blog.halfabubbleout.com/blog/why-is-website-traffic-important-to-my-business>
- <http://www.outofseo.com/website-traffic-data-monitoring-statistics-and-performance-analysis/>