

Social Media Influence and Brand Engagement

Beth Hilbert, Gabrielle Peters, & Katelynn Blalock

Purpose

Analyze social media's role in influencing customers by measuring the impact of status updates and advertisement.

- ▶ **Dataset**

- ▶ Performance metrics of renowned cosmetic's brand Facebook page.

- ▶ **Response Variable**

- ▶ “Page Total Likes”
- ▶ Assuming this is a good representation of brand reputation and social media engagement allowing for analysis of other metrics' impact on customer engagement via social media.

Exploratory Data Analysis

- ▶ 500 observations of 19 variables:

Page Total Likes
Type
Category
Post Month
Post Weekday
Post Hour
Paid
Total Reach
Total Impressions
Engaged Users

Consumers
Consumptions
Impressions for Users with Likes
Reach by Users with Likes
Users with Likes and Engagement
Comment
Like
Share
Total Interactions

Missing Data Values

- ▶ Percentage of missing values:
 - ▶ Paid .2%
 - ▶ Like .2%
 - ▶ Share .2%
- ▶ Impact of missing values:
 - ▶ Biased estimates
 - ▶ Loss of information
 - ▶ Decreased statistical power
 - ▶ Increased standard errors
 - ▶ Weakened interpretation of findings

Cleaning Data

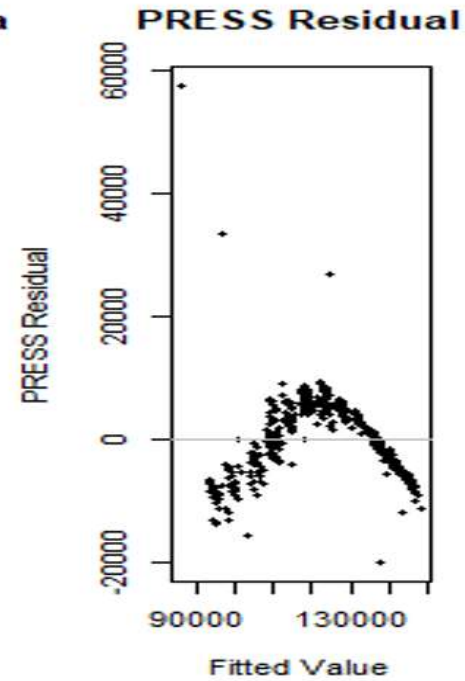
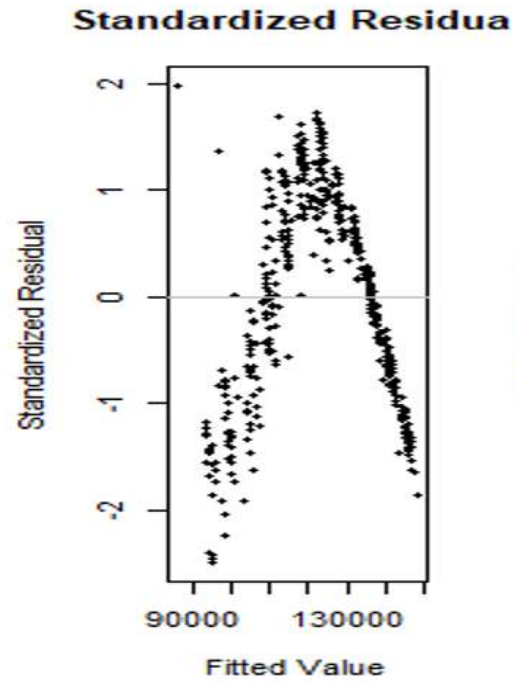
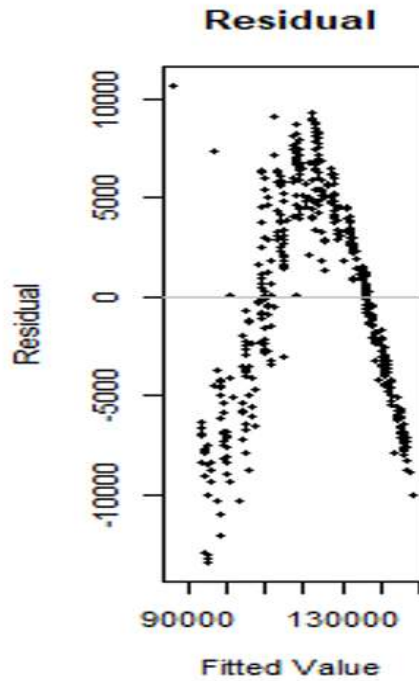
- ▶ Evaluated two methods for the missing data values:
 - ▶ Replacing with column median
 - ▶ Replacing with 0's
- ▶ Pros for replacing missing data values
 - ▶ Reduce bias
 - ▶ Maintain sample size
 - ▶ High efficiency
- ▶ Cons to replacing missing data values
 - ▶ Risk of multicollinearity
 - ▶ Inconsistent bias
 - ▶ Underestimation of standard errors

Evaluating LINE Assumptions of All 19 Variables

- ▶ Evaluate residuals to verify LINE assumptions not violated:
 - ▶ Linear relationship
 - ▶ Independent errors
 - ▶ Normally distributed errors
 - ▶ Equal variance in errors
- ▶ Residual analysis checks the fit of model to available data
- ▶ Unstable models, different sample with totally different conclusion.

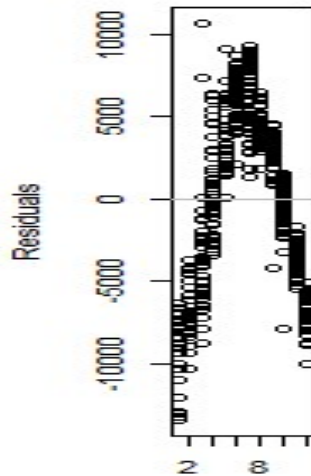
Residual Types

- ▶ Result: Non-Linearity



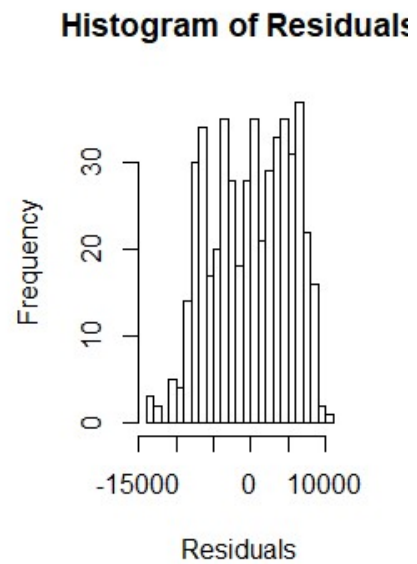
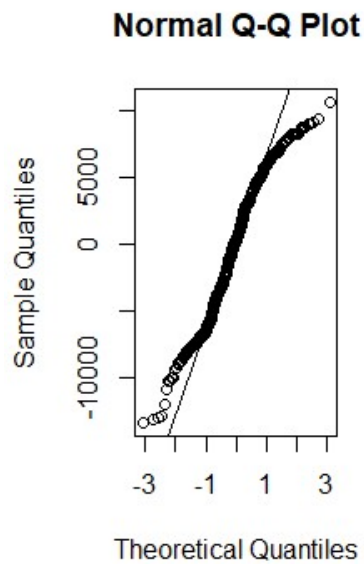
Check Linearity and Equal Variance

- ▶ Plot each covariate against the residuals
 - ▶ Post.Month variable was non-linear



Check Normality

- ▶ Plot residuals using QQ plot
 - ▶ Light tailed distribution



Transformation

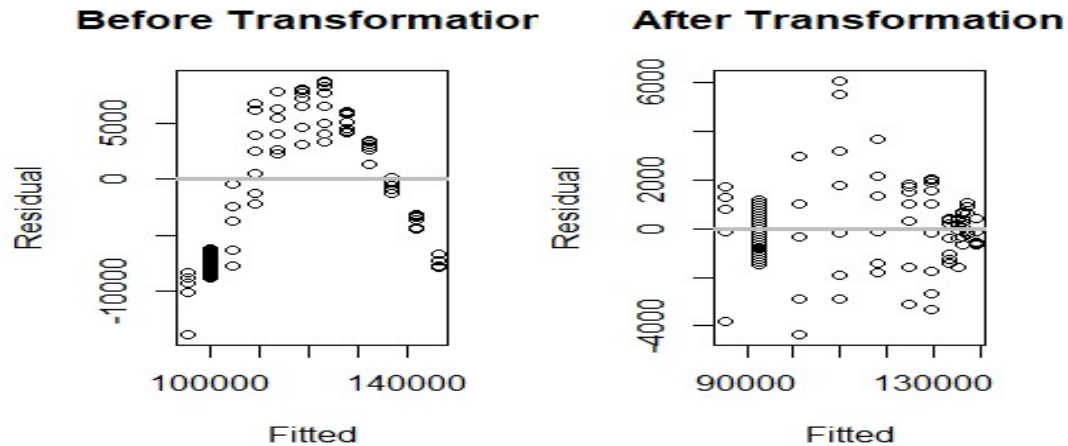
- ▶ Higher order polynomial transformation for Post.Month

- ▶ *# centering*

- `dataset$Post.Month.Centered = dataset$Post.Month - mean(dataset$Post.Month)`

- ▶ *# transforming*

- `lm(dataset$Page.Total.Likes ~ dataset$Post.Month.Centered + I(dataset$Post.Month.Centered^2) + I(dataset$Post.Month.Centered^3) + I(dataset$Post.Month.Centered^4))`



Remodeling - Variable Selection

- ▶ Variance inflation factors (VIF) used to test multicollinearity.
- ▶ 12 variables with a VIF score higher than 10.
- ▶ Backwards elimination of variables, eliminating the highest VIF score each time, resulted in eliminating Post.Month, Total.Impressions, Engaged.Users and Total.Interactions.

dataset\$Post.Hour	dataset\$Paid	dataset\$Total.Reach
1.035778	1.042742	3.193144
dataset\$Consumers	dataset\$Consumptions	dataset\$Impressions.for.Users.with.Likes
4.996422	2.087754	1.722414
dataset\$Reach.by.Users.with.Likes	dataset\$Users.with.Likes.and.Engagement	dataset\$Comment
6.001376	5.170489	4.605599
dataset\$Like	dataset\$Share	
8.382640	7.791800	

Remodeling - Standardized Coefficient

- ▶ Compare the relative strength of the coefficients by converting the coefficient to a standard unit of measure.
- ▶ Backward Elimination using partial F tests to eliminate insignificant variables from the model.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.807e-15	4.115e-02	0.000	1.000000	
Post.Hour	-1.543e-01	4.192e-02	-3.681	0.000258	***
Paid	1.779e-02	4.206e-02	0.423	0.672609	
Total.Reach	1.470e-01	7.361e-02	1.997	0.046413	*
Consumers	-5.451e-01	9.208e-02	-5.920	6.07e-09	***
Consumptions	-6.009e-02	5.952e-02	-1.010	0.313201	
Impressions.for.Users.with.Likes	-5.297e-02	5.406e-02	-0.980	0.327623	
Reach.by.Users.with.Likes	-3.350e-01	1.009e-01	-3.320	0.000969	***
Users.with.Likes.and.Engagement	6.487e-01	9.367e-02	6.925	1.38e-11	***
Comment	1.532e-01	8.840e-02	1.733	0.083669	.
Like	3.500e-01	1.193e-01	2.935	0.003493	**
Share	-4.444e-01	1.150e-01	-3.865	0.000126	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Remodeling - Elimination Results

- ▶ We eliminated the following as the partial F-test showed that the p-values of the following are greater than .05 and were thus eliminated from the model: Comment (0.08367) Paid (0.2045) Impressions.for.Users.with.Likes (0.2938) Consumptions (0.3682) Total.Reach (0.09598)
- ▶ After eliminating these variables from the standardized model, we can see from the results of the summary that all remaining variables are significant.

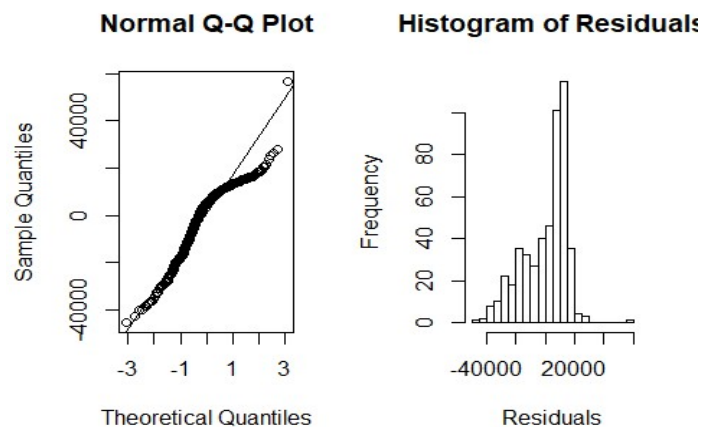
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-1.775e-15	4.133e-02	0.000	1.000000	
Post.Hour	-1.548e-01	4.167e-02	-3.715	0.000227	***
Consumers	-4.980e-01	7.223e-02	-6.895	1.65e-11	***
Reach.by.Users.with.Likes	-2.661e-01	6.436e-02	-4.134	4.19e-05	***
Users.with.Likes.and.Engagement	5.577e-01	8.198e-02	6.803	2.98e-11	***
Like	4.297e-01	1.140e-01	3.768	0.000184	***
Share	-3.468e-01	1.017e-01	-3.409	0.000704	***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

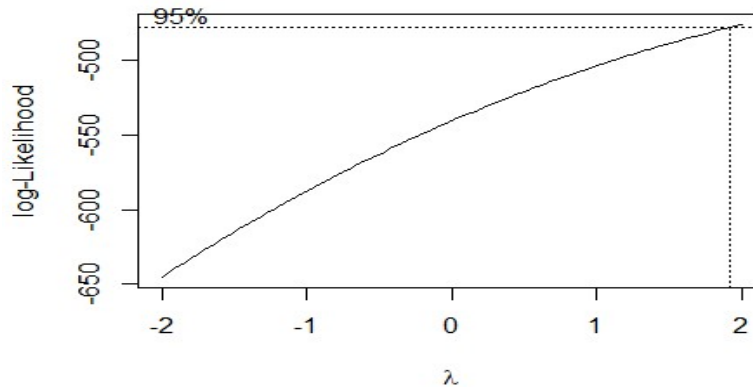
Model Adequacy Re-checking

- ▶ Ensure LINE assumptions have not been violated with new model.
 - ▶ Re-evaluate remaining variables in the dataset by plotting each covariate against the residuals to ensure linearity and have equal variance.
 - ▶ Rerun QQ plot to check normality assumption which appears to have a negative skew so we must transform predictor variable.



Transform Regressor to Adjust for Non-Normal Distribution

- ▶ We determined from the QQ plot that there was a slight negative skew and ran a boxcox power transformation on the new model and found that the fit would not be improved from a log transformation.



Final Model

Page.Total.Likes =

129500

- (0.05765 * Post.Hour)

- (9.184 * Consumers)

- (0.5636 * Reach.by.Users.with.Likes)

+ (14.81 * Users.with.Likes.and.Engagement)

+ (21.64 * Like)

- (0.01330 * Share)

Call:

lm(formula = Page.Total.Likes ~ Post.Hour + Consumers + Reach.by.Users.with.Likes + Users.with.Likes.and.Engagement + Like + Share, data = dataset)

Residuals:

Min	1Q	Median	3Q	Max
-45375	-10518	5267	11544	56702

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.295e+05	1.576e+03	82.153	< 2e-16 ***
Post.Hour	-5.765e+02	1.552e+02	-3.715	0.000227 ***
Consumers	-9.184e+00	1.332e+00	-6.895	1.65e-11 ***
Reach.by.Users.with.Likes	-5.636e-01	1.363e-01	-4.134	4.19e-05 ***
Users.with.Likes.and.Engagement	1.481e+01	2.177e+00	6.803	2.98e-11 ***
Like	2.164e+01	5.743e+00	3.768	0.000184 ***
Share	-1.330e+02	3.900e+01	-3.409	0.000704 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 15040 on 493 degrees of freedom

Multiple R-squared: 0.156, Adjusted R-squared: 0.1457

F-statistic: 15.19 on 6 and 493 DF, p-value: 5.479e-16

The slide features abstract blue geometric shapes on the left and right sides. On the left, there is a solid light blue trapezoidal shape. On the right, there is a complex arrangement of overlapping translucent triangles in various shades of blue, ranging from light to dark. The central text 'Questions...' is positioned in the middle of the slide.

Questions...