SOCIAL MEDIA USAGE AND EMOTIONAL WELL BEING

Madison Klinefelter & Sofia Mucci Dr. Khyam Paneru (Mentor)
Department of Mathematics, The University of Tampa

INTRODUCTION

Technology is a part of most individual's everyday life, but the use of social media is their choice. A data set named Social Media Usage and Emotional Well Being found on Kaggle.com showcases individual use of social media and the emotions they felt while using these sites. The dataset contains 1000 values with 14 primary variables made up of 5 quantitative and 3 categorical variables, which were converted into 9 dummy variables. These variables help address the question of how daily usage of social media and certain social media platforms affect emotions in adolescents and young adults.

Table 1 : Variables

Variables	Category
Daily usage Time	Integer value, representing minutes
Age	Integer value
Posts Per Day	Integer value
Likes Per Day	Integer value
Comments Received Per Day	Integer value
Messages Received Per day	Integer value
Gender	Male = 0
	Female = 1
Platform	**Facebook** = base variable
	Twitter : 1 = yes
	Instagram : 1 = yes
	Snapchat : 1 = yes
Emotion	**Neutral** = base emotion
	Happy : 1 = dominant
	Sad : 1 = dominant
	Anger : 1 = dominant
	Anxiety : 1 = dominant
	Boredom : 1 = dominant

Figure 1: Distribution of Age

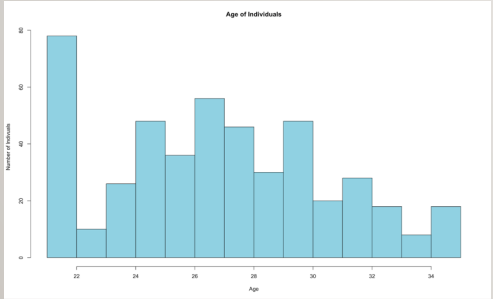
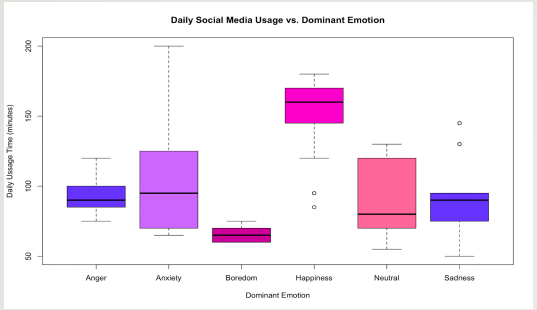


Figure 2: Daily Social Media Usage By Dominant Emotion



The graph above links daily social media usage to dominant emotions. High usage is associated with emotions like anxiety and happiness, whereas boredom and sadness are linked with lower usage. Emotions such as anxiety have a few outliers, for there are individuals with notably high usage above the whiskers. The graph below compares social media usage across platforms. Instagram has the highest median daily usage of about 150 minutes spent on it per day. Snapchat has a noticeable outlier with a user spending significantly less time than others.

Figure 3: The Time Spent on Different Platforms

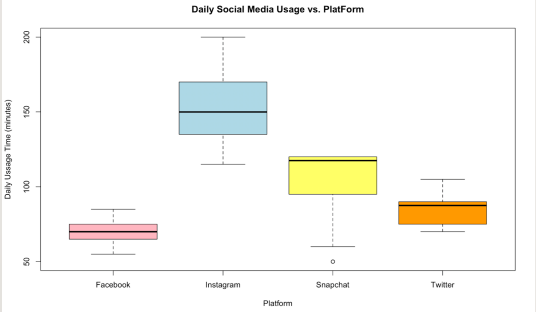
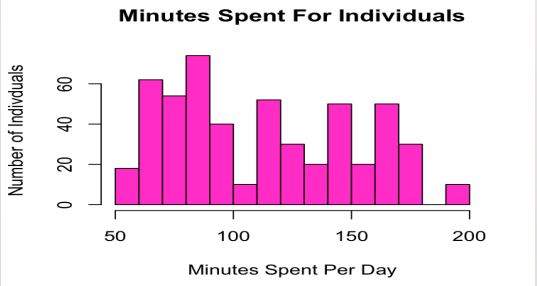


Figure 4 :Time Spent Daily by Individuals



REGRESSION MODEL

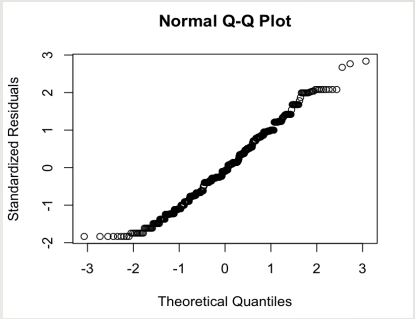


Figure 5: Q-Q Plot of Standardized Residuals

While creating our regression model, our data produced a very clean and positive linear model. Because of this result, we were able to conclude that it would not be necessary to transform our data. The Q-Q plot displays the residuals from the regression model closely follows a normal distribution. The points align well with the theoretical quantile line, with only slight deviations in the tails. So, no transformations are necessary as the residuals meet the assumption of normality.

FITTED MODEL

Our fitted model represents only our quantitative variables, but we still produced a higher R^2 value.

Table 2 : Fitted Model Estimates

	Est Std.	Error	T Value	P(>|t|)
(intercept)	20.37	4.69	4.34	0
Age	0.18	0.17	1.08	0.28
Posts Per Day	-0.03	0.99	-0.03	0.97
Likes Received Per Day	0.96	0.07	14.69	0
Comments Received Per Day	0.13	0.21	0.63	0.53
Messages Sent Per Day	1.36	0.17	7.79	0
R^2 : 88.33% , R^2_a : 88.21%				
F-statistic : 702.5 on 5 and 464 DF, p-value : < 0.00				

FULL MODEL WITH CATEGORICAL AND QUANTITATIVE VARIABLES

Our fitted model with all the included variables once again produces an impressive 93.8% for our R^2. This is extremely impressive as it shows the clear relationship between variables.

Table 4 : Full Model Estimates

	Est Std.	Error	T Value	P(>|t|)
(intercept)	41.49	4.15	9.99	0
Age	-0.17	0.14	-1.24	0.22
Posts Per Day	1.07	0.79	1.35	0.18
Likes Received Per Day	0.34	0.07	4.95	0
Comments Received Per Day	0.84	0.17	4.93	0
Messages Sent Per Day	0.99	0.14	6.93	0
Gender	0.07	1.04	0.07	0.95
Twitter	-2.91	1.98	-1.47	0.14
Instagram	22.96	2.83	8.10	0
Snapchat	20.81	2.70	7.71	0
Happy	8.01	1.90	4.22	0
Sad	-1.34	1.94	-0.69	0.49
Anger	-1.23	2.18	-0.57	0.57
Anxiety	-3.07	1.80	-1.71	0.09
Bored	7.36	3.00	2.45	0.01
R^2 : 93.8% , R^2_a : 93.5%				
F-statistic: 487.3 on 14 and 455 DF, p-value: < 0.00				

CONCLUSION

The fitted multiple linear regression model above exhibits a high R^2 value of 0.937 with an adjusted R^2 of 0.935. This result indicates that 93.5% of the variance in the dependent variable (Y) is explained form the predictors in the model. The most notable predictors were likes received per day (x3), comments received per day (x4), messages sent per day (x5), platform dummies, Instagram (d3) and Snapchat (d4), along with happiness dummy (d5), and boredom dummy (d9). Each of these variables has a p-value below 0.05, indicating it's significance. The residuals exhibit a standard error of 9.711, which also supports the model's accuracy. In conclusion, the model is statistically significant with $p < 0.0001$. So, the model effectively portrays the relationship between daily usage of social media (Y) and the quantitative and predictor variables