Predicting Well-Being and Depression using Instagram Data Group 2 - Use case 2: Business

Xavier Moreno, Dimitris Michailidis, Dimitris Papadimas, Rafael Pierre, and Jeroen Schmidt

University of Amsterdam, Amsterdam, Netherlands
xavi.morenoliceras1,dimitris.michailidis,dimitris.papadimas,
rafael.vellosopierre,jeroen.schmidt@student.uva.nl

Abstract. Social media is becoming a major part of social interactions, with billions of posts being created and sent daily. Researchers have leveraged this data and used social media to predict individual well-being, based on text and media. This paper presents a research on predicting individual well being based on their posts on Instagram.

Keywords: social media · well being · depression · instagram · machine learning

1 Introduction

Instagram is a social media platform used by millions of people every day. Billions of pictures are shared every hour expressing a sentiment or status. Today, this can easily be computed and analysed in order to determine someone's state based on their social media content. Social media content has been used to analyse the relation between an individual and their measure of depression [5] or their well-being [4]. This last one was based on text based on Facebook and Twitter, but a research about the relation between visual social media content and the well-being has not been made yet, and Instagram is the perfect platform to perform this analysis, since it only allows you to post pictures and videos.

Well-being is measured by a psychological score called PERMA − Positive emotion, Engagement, Relationships, Meaning, Accomplishment. This score can be collected by taking a survey which several questions. The aggregation of this questions gives a PERMA score. A group of people was willing to take this survey logging in with their Instagram account. Consented access to the subjects social media content and the outcomes from the survey was provided, indicating their well-being.

The main goal of this paper is to analyse the visual social media content provided in order to make sense of it and perform analyses to find a relation between these data and their well-being. The data provided were already processed in the level of image, therefore, an aggregation to user level was needed. This paper will discuss the methodology employed and results obtained in determining the indicators of PERMA defined well being.

2 Methodology

The data consists of seven dataframes; all of which are joined on either a user or image ID. There are three levels of (aggregation) detail in the data; *User Level*, *Image Level* and *Image Detail Level*. The data's complexity, due to its multiple levels of aggregation. necessitated the use of the Fractal Analytic Design model by M. Mark et al.[2]. This design methodology allowed for the project to be iterative worked on in isolation by dividing it into 4 distinct areas such that each level was built on top of the last level while simultaneously feeding back information to the previous level to make it better, this methodology made it easier for the 5 authors of this paper to more easily contribute discoveries and insights with one another. This iterative methodology allowed for the creation of 310 different features, however a large portion of these features were later deemed to be irrelevant with only a small subset actually contributing towards predicting the PERMA scores. The methodology was as a result organised as follows:

1. *Generation of Features at the Image Level*
2. *Generation of Features at the User Level*
3. *User Level Feature Selection and Exploration*
4. *Building and Testing of Model*

The iterative steps were also informed by the research conducted by Reece et al. [5] and D. Choudhury et al.[1] whom each respectively looked at the effect of Instagram on Depression and Twitter on emotional states of pregnant women. The insights these papers contributed will be discussed in the subsections that used them.

2.1 Exploratory Data Analysis

The demographics of the survey takers were explored, considering income, employment status, gender, age and education. Although the survey dataset was not big enough to show a real trend, it is observed on Figure 2 that a higher education level tends to lead to a higher PERMA score and therefore more happiness. It is also observe in Figure 1 that people who are unable to work or are looking for a job tend to be less happy than the employed ones or those who do not actively seek employment.

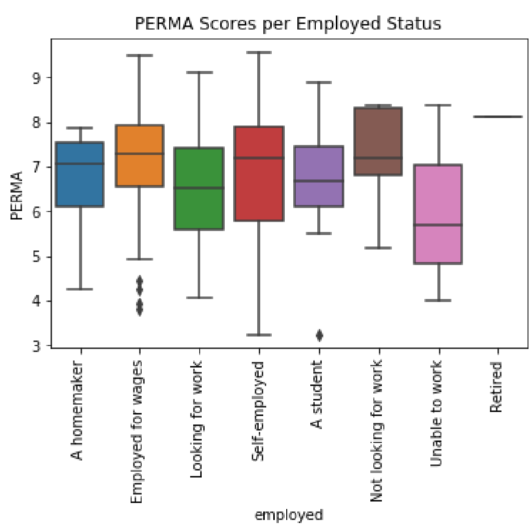

Fig. 1: PERMA Scores versus Employment Status

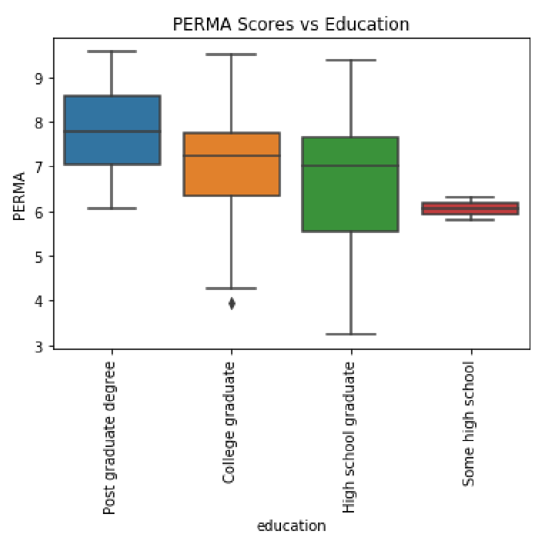

Fig. 2: PERMA Scores versus Education Level

2.2 Data Cleaning and Preparation

Data cleaning is an essential step for this project. Data should be coherent before are processed, therefore, the points that were considered are as follows:

Removing rows with no response variables: It was found that one survey (out of 161) did not have any value for *HAP* and *PERMA*. Hence, it was decided to eliminate that survey.

Instagram users taking the survey twice: Two of the surveys taken were repeated for two people, i.e. two Instagram users took the same survey twice, indicating different values for PERMA scores. The decision was to remove the oldest survey and keep the recent one.

Time Frame of Observations: Well-being is an state of being comfortable, healthy, or happy at some point in time. People may have different states in the stages of their life, their well-being could change over time. This means that a same Instagram user could go through different stages in their life and therefore, they would have pictures expressing different PERMA Scores. This begs the question, *when do PERMA characteristics express them selves in some ones life?*. Reece et al. [5] and D. Choudhury et al.[1] only looked at social media posts from 6 months before and up till the date of filling in the study questionnaires, and they were able to obtain favourable results by doing so. This study opted to use a time span of 12 months because at 6 months many of the features generated became to spares to be useful.

2.3 Features Creation

The aggregation of features from different levels to a single, user level, was paramount because the PERMA scores were expressed as single observation, response variables, at the user level. Over 150 different features at the user level were created during the Fractal Analytic Design process. The vast majority of these features were count, average and percentage aggregations but a handful of features were created using either advanced methods or domain knowledge, they will now be discussed.

Object labels The objects that appear in each image were present in one of the data set provided. The labels were first selected with the highest frequencies on the images and then a manual and cooperative selection was done determining which objects were related to well-being. Ultimately, a one-hot encoding was applied to this categorical data and then aggrigated up as a count and ratio value to the user level, thus a new set of features were created in image level indicating whether an image contained that object or not.

ANP Dataset Clustering: Gaussian clustering and principal component analysis were performed was performed on the ANP emotional scores data. Each image had 4 ANP emotions (out of 24 emotions) associated with it. *Boxcox* transformation was used on the data in order to better normalise the data (due to PCAs sensitivity to outliers). PCA was able to reduce the 24 dimensions into 2 dimensions with an 81% variance preservation. Guassian mixed clustering (*GMC* was used because it was one of the few clustering algorithms that could handle both circular and extreme elongated elliptical clusters. Three distinct clusters were identified by *GMC* with the best silhouette score of 48.8%.

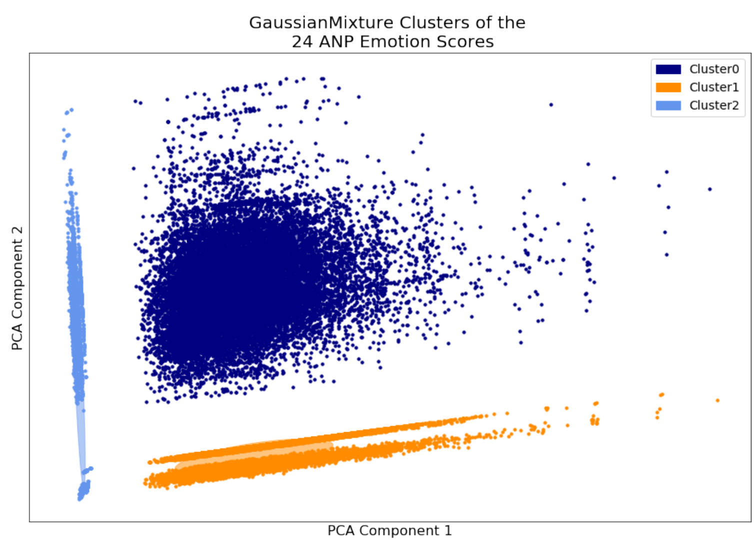

Fig. 3: ANP Emotions - PCA GMC Clusters

Associating Depression with Instagram Filters: Reece et al. showed that depressed users are less likely to use any filters at all in their Instagram posts[5]. When they do use filters, they prefer "Inkwell", which converts the picture to black and white, contrary to happy users who tend to prefer more lightening filters, like "Valencia". In their research, Reese et al. ranked and classified the filters based on the differences between observed and expected usage frequencies, into "Healthy" and "Depressed". The distribution of the filters in the dataset can be seen in Fig. 4. Based on these classifications, the number of filters associated with health and depression was then aggregated per user.

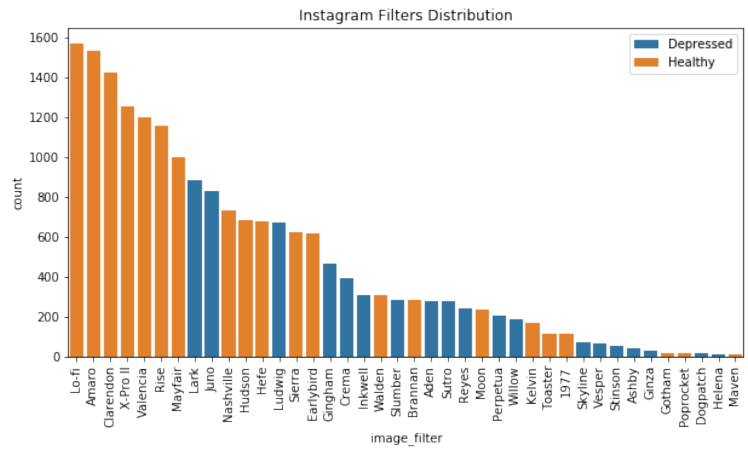

Fig. 4: Distributions of Filters in the Data Set

3 Regression Results

One of the main objective in this project is determine which features extracted from images from social media are related to the well-being of an individual. Hence, only models that could supply information about the feature importance were considered, such as Linear Regressor, Gradient Boosting Regressor, Support Vector Machine with a linear kernel and so on. It was found that the Linear Models performed the best when predicting for the different *PERMA* scores.

Step-up p-value feature selection was used in order to produce the optimal models for each score. This method was settled on after grid search and iterative hold out feature selection methods fielded subpar results(often producing models with performance worse then random chance). The usage of feature selection was particularly important because well over 300 features were investigated[1]. Table 1 shows the R^2 scores for each *PERMA* score for the training and testing data sets. Detailed information of the models for each score can be found in Appendix A. The following subsections elaborate more on the scores and provide the names of the features that were used for each model. Discussions of the results can be found in the next section.

A brief note to the reader before they proceede to the next subsections, the feature names follow the following conventions;

- *count_*: These features refer t othe count of a topic/object for the past 12 months of user activity from the date that the data was collected.
- *ratio_*: These features refer to the ratio of a topic/object with respect to the rest of the users total image count.
- *ALL-CAPS*: These refer to the ratio of ANP emotions that each image had, eg CONFUSED, ANGERY etc

[1] Learning curves could also be employed in future research to investigate and justify the requirment for more data to be collected

Predictor	R^2 Training Score	R^2 Testing Score
PERMA	0.65	0.059
P	0.41	0.005
E	0.91	0.0007
R	0.54	0.06
M	0.48	0.005
A	0.8	0.004

Table 1: R^2 Scores for Training and Test Data Sets

3.1 (P)ositive Emotion

This score relates to an individuals felling of positivity, such as happiness. Not only is this metric an indicator if somone is displaying happy emotion (eg smiling in a picture), but also about being in an optimistic state of mind and living the life in a positive way.

The following features were associated with predicting positivist: *count_Sunlight ratio_Vegetable*, *ratio_Art*, *count_Handwriting*, *count_CONFUSED*, *avg_posts_late_night*, *ratio_Bedroom*, *cluster_0*, *count_Head*, *count_Manx*, *count_Hair*, *count_Child*, *count_Beverage*, *count_Dusk*, *count_Selfie*.

3.2 (E)ngagement

Engagement is the element that is related to have an activity or activities that makes people feel engaged. This element is important and part of PERMA since being participate of an activity it will make absorbs someone's present moment and give a blissful immersion into the activity involved.

There were 86 features selected which helped predict for the accomplishment score[2]. Of the 37, the following 29 features are worth mentioning: *count_Handwriting*, *ratio_Shoe*, *avg_comments*, *born*, *ratio_Electronics*, *ratio_Collage*, *ratio_Cream*, *count_Beverage*, *ratio_Chocolate*, *ratio_Bread*, *ratio_Outdoors*, *ratio_Goggles*, *education*, *count_Drink*, *age*, *ANGRY*, *count_Jar*, *count_Shirt*, *ratio_Tattoo*, *ratio_Vegetable*, *ratio_Carrot*, *ratio_Handwriting*, *count_Chair*, *count_Musical.Instrument*, *employed*, *avg_posts_early_night*, *count_Boat*, *count_Abyssinian*, *count_Sport*, *count_Electronics*, *avg_posts_night*, *count_Bicycle*, *ratio_Smile*, *ratio_Potted.Plant*, *cluster_2*, *count_Drawing*, *ratio_Light*,

3.3 (R)elationships

Relationships and social life is essential and strongly connected to someone's well-being. Having someone in someone's life is important to avoid loneliness since loneliness does not relate with well-being positively.

The following features were associated with predicting relationships: *count_Book*, *ratio_Vegetable*, *count_Water*, *ratio_Percussion*, *count_Female*, *ratio_Art*, *count_Screen*, *count_Clothing*, *ratio_Beverage*, *count_Bicycle*, *count_Suit*, *count_Appliance*, *count_Shoe*, *count_Canine*, *ratio_Tree*, *count_Child*, *count_Bread*, *count_Kid*, *ratio_Road*, *SURPRISED_y*, *ratio_Red*,.Sky *ratio_Wood*, *count_Blonde*, *ratio_Flyer*, *count_Coast*.

3.4 (M)eaning

Giving a meaning to our lives is making sense of it. Meaning gives to people a reason to live and not finding it would be a factor of depression[3], the opposite of well-being.

The following features were determined to be the best predictors for the meaning score: *count_Handwriting*, *ratio_Produce*, *count_Child*, *ratio_male*, *ratio_Luggage*, *ratio_Bedroom*, *count_Cup*, *count_Creme*, *count_Canine*, *ratio_Bottle*, *count_Kid*, *ratio_Terrier*, *ratio_Canine*, *count_Clothing*, *ratio_Appliance*, *count_Head*, *ratio_Blanket*, *count_Bedroom*, *ratio_Jar*, *count_Sunlight*, *ratio_Manx*, *count_Dawn*,

[2] refer to figure A.5 for all the features

3.5 (A)ccomplishments

This score refers to an individuals sense of accomplishment with how they are reaching their ambitions.

There were 55 features selected which helped predict for the accomplishment score[3]. Of the 55, the following 29 features are worth mentioning: *imagecount*, *ratio_Beverage*, *count_Canine*, *count_Glasses*, *ratio_Beard*, *ratio_Head*, *count_Hair*, *ratio_Shoe*, *ratio_Label*, *ratio_Guitar*, *count_Musical_Instrument*, *ratio_Rose*, *CONFUSED*, *count_Female*, *count_Rose*, *ratio_Terrier*, *ratio_Dog*, *ratio_Television*, *count_Clothing*, *count_Paper*, *HAPPY*, *avg_number_of_faces_over_images_with_faces*, *count_Kid*, *count_Handwriting*, *count_Bikini*, *count_Collage*, *ratio_Musician*, *ratio_TV*, *ratio_Chocolate*

3.6 PERMA Score

The PERMA score is calculated by taking a mean from the individual P, E, R, M and A scores. The PERMA score could be though of as a score of general well being. For the aggregate score, 34 features were selected as part of the step-up p-value method.

The following features were determined to be the best predictors for the PERMA score: *count_Handwriting*, *ratio_Vegetable*, *ratio_Sunlight*, *ratio_Art*, *count_CONFUSED*, *ratio_Beverage*, *count_Selfie*, *ratio_Red.Sky*, *ratio_Toy*, *ratio_Fruit*, *count_Suit*, *ratio_Percussion*, *count_Kid*, *count_Electronics*, *count_Clothing*, *count_Suitcase*, *cluster_0*, *count_Mammal*, *count_Text*, *count_Alcohol*, *count_Appliance*, *count_Bike*, *count_Oven*, *count_Overcoat*, *ratio_Tree*, *income*, *happy_flt_pct* (% of photos with happy filters), *num_images_no_comments*, *count_Baby*, *ratio_HAPPY*, *ratio_CONFUSED*, *ratio_Bicycle*, *avg_number_of_faces_over_images_with_faces*, *ratio_CALM.*

4 Regression Analysis and Discussions

From table 1, it is observed that large differences in the test and training R^2 scores exist, this is indicative that over-fitting is occurring and that more could be done to improve the model.

The two best PERMA scores with the highest R^2 scores were *PERMA* and *(R)elationships* with R^2 scores of 0.059 and 0.06 respectively. Looking at the features that were used in the models to predict each score it becomes apparent that there are overlaps in what helps to determine the PERMA and (R)elationship scores. Both models use features that either have a count or ratio of a feature that indicates the presence of people, such as children, people with blond hair, clothing (people with clothing) these observations make sense due to the fact that the presence of other people is a clear indicator of the social activity of the subject. Features are also present that are indicative of indicative of healthy and social gathering through the presence of vegetables (which indicate that they might take pictures of food with vegetables when eating out) and picture of drinks, being indicative of social drinking. Another noteworthy observation that is in line with research on social well being is that both include models use features that show the presence of animals (eg mammals and dogs).

Figures 5 & 6 show predicted values vs true values from test data. As expected for R^2 values that are this low, very little linear correlation exists between the true values and the predicted values would be more prevalent in if the model was able to better predict the PERMA and (R)elationship scores.

The remaining PERMA scores *(Positive Emotion)*, *(Engagement)*, *(M)eaning* and *(A)ccomplishments* yielded R^2 scores even closer to 0. R^2 scores so close to 0 indicates a marginal better performance over random chance, they will as such be only briefly discussed.

The clustering of images according to the PCA reduction of ANP emotion score also proved to be useful. The models that predicted for PERMA and (P)ositive Emotion Scores both depended on the count of images that belonged to cluster 0 (with a significance value of 0.05) while the model that predicted for (E)engagement used the count of images that belonged to cluster 2 (with a p-value significance of 0.01). PCA

[3] refer to figure A.8 for all the features

dimensional reduction makes it difficult to interpret what each cluster represents, while not impossible, it's interpretation was decided to be left for further work[4].

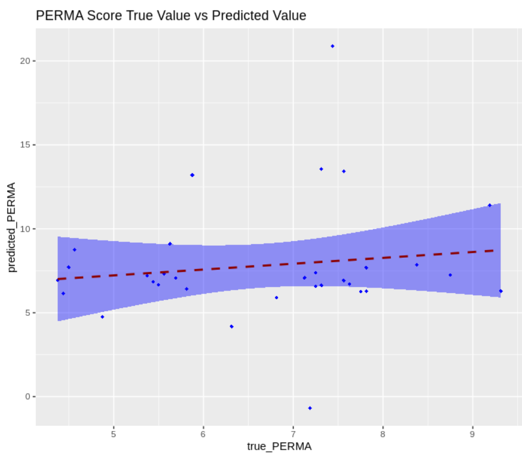

Fig. 5: PERMA Model: True vs Predicted

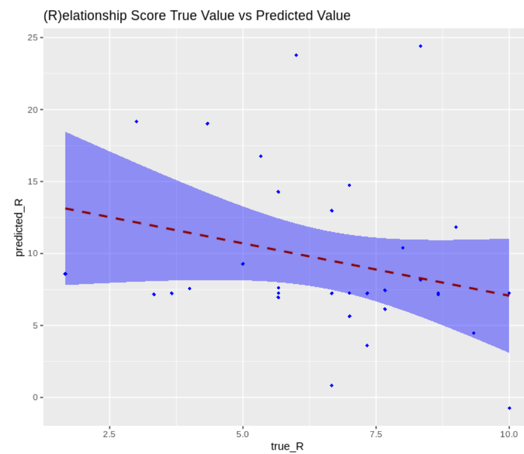

Fig. 6: (R)elationship Model: True vs Predicted

None of the models besides the PERMA model used features that indicated filter usage. The PERMA model used *happy_flt_pct* which captured the percentage of photos that user had which used with happy filters as was defined by Reece et al.[5]. This result was encouraging because while the work by Reece et al.[5] looked into indicators for depression, this paper was able to show that quality of life can be determined partially by the type of filter used.

5 Conclusion

Through an near exhaustive search of different features using Fractal Analytic Design this paper was unable to find a good relation between visual social media content and the well-being of individuals. All the models that were attempted could not adequately describe the reality, howeve, even so this paper was able to define feature importance for each score; particularly in aiding a better understanding of what features best describe the PERMA score and (R)elationship score. Potential value was also found in using PCA dimensional reduction and Gaussian clustering in helping to predict some of the scores. The types of filters used also proved to have been a valuable insight that also helped. This paper would therefor recommend that these avenues be further explored.

The data provided was too small as the response variables at the user level only had 161 observations. Even though enough data was available per user at the image level, alot of information loss occurred through the aggregation process. It would be recommended that more data be collected to further investigate this area of research and allow for deeper relations in the data to be explored.

References

1. De Choudhury, M., Counts, S., Horvitz, E.: Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 3267–3276. ACM (2013)

[4] The variables that contribute the most information that helps capture the variance within PCA components could be looked at to determine an intuitive reasoning of what each cluster represents

2. Herman, M., Rivera, S., Mills, S., Sullivan, J., Guerra, P., Cosmas, A., Farris, D., Kohlwey, E., Yacci, P., Keller, B., Kherlopian, A., Kim, M.: The Field Guide to Data Science (11 2013)
3. Joaqun G., Jos H. M., S.P.: Predicting role of the meaning in life on depression, hopelessness, and suicide risk among borderline personality disorder patients (2014), http://www.scielo.org.co/pdf/rups/v13n4/v13n4a25.pdf
4. Park, G., D.S., Kosinski, M.: Predicting individual well-being through the language of social media (2016), https://www.ncbi.nlm.nih.gov/pubmed/26776214
5. Reece, A.G., Danforth, C.M.: Instagram photos reveal predictive markers of depression. EPJ Data Science **6**(1), 15 (2017), http://arxiv.org/abs/1608.03282

A Appendix - PERMA Score Results

```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              7.379e+00  2.236e-01  32.992  < 2e-16 ***
count_Handwriting                       -5.651e-01  1.826e-01  -3.096 0.002610 **
ratio_Vegetable                          3.315e+01  2.542e+01   1.304 0.195384
ratio_Sunlight                          -1.247e+02  9.668e+01  -1.290 0.200310
ratio_Art                               -4.913e+01  2.180e+01  -2.254 0.026617 *
count_CONFUSED                          -4.632e-02  2.417e-01  -0.192 0.848428
ratio_Beverage                           3.100e+01  1.365e+01   2.270 0.025544 *
count_Selfie                            -2.485e-02  9.257e-02  -0.268 0.788995
ratio_Red.Sky                            4.448e+02  1.381e+02   3.220 0.001780 **
ratio_Toy                               -2.613e+02  5.792e+01  -4.512 1.92e-05 ***
ratio_Fruit                              2.882e+02  5.160e+01   5.584 2.42e-07 ***
count_Suit                               2.854e+00  7.306e-01   3.906 0.000180 ***
ratio_Percussion                        -8.333e+01  1.905e+01  -4.374 3.25e-05 ***
count_Kid                               -6.030e-01  5.083e-01  -1.186 0.238546
count_Electronics                        5.686e-01  1.314e-01   4.328 3.86e-05 ***
count_Clothing                          -6.311e-01  1.234e-01  -5.113 1.74e-06 ***
count_Suitcase                           1.628e-01  2.038e-01   0.799 0.426490
cluster_0                                1.148e+00  6.153e-01   1.866 0.065282 .
count_Mammal                            -1.031e-01  2.375e-02  -4.340 3.69e-05 ***
count_Text                              -5.892e-02  4.506e-02  -1.308 0.194246
count_Alcohol                           -2.135e-01  1.001e-01  -2.133 0.035623 *
count_Appliance                          4.996e+00  1.287e+00   3.883 0.000196 ***
count_Bike                               3.673e-01  1.825e-01   2.012 0.047162 *
count_Oven                              -5.307e+00  1.437e+00  -3.694 0.000377 ***
count_Overcoat                          -2.045e+00  7.655e-01  -2.671 0.008958 **
ratio_Tree                               1.015e+02  5.578e+01   1.819 0.072125 .
income                                  -5.962e-02  2.795e-02  -2.133 0.035593 *
happy_flt_pct                           -1.325e+00  1.040e+00  -1.273 0.206157
num_images_no_comments                   8.303e-03  4.460e-03   1.862 0.065881 .
count_Baby                               9.328e-01  5.261e-01   1.773 0.079544 .
HAPPY                                    1.147e+00  4.575e-01   2.506 0.013975 *
avg_number_of_faces_over_images_with_faces -7.855e-01  3.648e-01  -2.153 0.033966 *
CONFUSED                                 8.312e+00  5.751e+00   1.445 0.151794
ratio_Bicycle                           -8.219e+01  6.289e+01  -1.307 0.194556
CALM                                     7.971e+00  6.474e+00   1.231 0.221444
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9786 on 91 degrees of freedom
Multiple R-squared:  0.6574,    Adjusted R-squared:  0.5294
F-statistic: 5.137 on 34 and 91 DF,  p-value: 2.36e-10
```

Fig. A.1: Summary Stats of Linear Model to Predict *PERMA*

```
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)           6.89176    0.21559  31.966  < 2e-16 ***
count_Sunlight       -1.13730    0.37440  -3.038 0.002978 **
ratio_Vegetable      73.42617   31.26937   2.348 0.020652 *
ratio_Art           -58.96032   24.96665  -2.362 0.019957 *
count_Handwriting    -0.64484    0.16613  -3.882 0.000177 ***
count_CONFUSED       -0.04791    0.32473  -0.148 0.882968
avg_posts_late_night -0.80784    0.34421  -2.347 0.020716 *
ratio_Bedroom      -363.97677  124.61487  -2.921 0.004237 **
cluster_0             1.68127    0.80998   2.076 0.040254 *
count_Head            0.26760    0.12444   2.150 0.033718 *
count_Manx            0.29946    0.12521   2.392 0.018470 *
count_Hair            0.40419    0.16645   2.428 0.016793 *
count_Child           0.19669    0.09882   1.990 0.049037 *
count_Beverage        0.11326    0.05389   2.102 0.037878 *
count_Dusk            0.64649    0.37570   1.721 0.088107 .
count_Selfie         -0.15253    0.10373  -1.470 0.144293
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.384 on 110 degrees of freedom
Multiple R-squared:  0.4179,    Adjusted R-squared:  0.3385
F-statistic: 5.265 on 15 and 110 DF,  p-value: 8.133e-08
```

Fig. A.2: Summary Stats of Linear Model to Predict P

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.956e+03	6.132e+02	4.820	2.10e-05	***
count_Handwriting	-1.517e+00	3.177e-01	-4.774	2.42e-05	***
ratio_Bike	-4.267e+01	8.791e+01	-0.485	0.630063	
ratio_Dessert	5.787e+01	3.846e+01	1.505	0.140260	
count_Dog	1.608e-01	1.130e-01	1.422	0.162723	
count_Terrier	9.345e-02	3.608e-01	0.259	0.796978	
ratio_Shoe	1.158e+01	2.594e+00	4.467	6.36e-05	***
avg_comments	4.594e-01	1.478e-01	3.108	0.003458	**
count_Outdoors	9.371e-02	1.301e-01	0.720	0.475422	
ratio_Dog	-1.400e+02	7.321e+01	-1.912	0.063016	.
born	-1.463e+00	3.042e-01	-4.809	2.17e-05	***
ratio_Electronics	2.234e+02	4.947e+01	4.517	5.44e-05	***
ratio_Collage	-6.197e+01	1.091e+01	-5.679	1.34e-06	***
count_Baby	4.390e-01	1.833e-01	2.395	0.021411	*
depressed_flt_pct	1.269e+00	7.648e-01	1.660	0.104837	
count_Tattoo	5.693e-01	7.035e-01	0.809	0.423193	
user_posted_photos	-7.889e-04	7.007e-04	-1.126	0.266942	
ratio_Cream	-5.655e+02	1.836e+02	-3.080	0.003734	**
count_Beverage	-1.200e+00	3.300e-01	-3.637	0.000779	***
ratio_Chocolate	-4.369e+02	1.167e+02	-3.745	0.000569	***
ratio_Bread	4.862e+02	1.074e+02	4.528	5.25e-05	***
count_Asleep	6.958e-02	2.641e-01	0.263	0.793568	
ratio_Housing	4.758e+01	3.323e+01	1.432	0.159952	
ratio_Head	4.119e+01	3.516e+01	1.171	0.248444	
ratio_Outdoors	-1.474e+02	3.627e+01	-4.063	0.000220	***
ratio_Goggles	1.522e+02	7.467e+01	2.039	0.048102	*
education	-3.752e-01	1.738e-01	-2.158	0.036969	*
count_Drink	1.583e+00	3.727e-01	4.247	0.000126	***
count_Teddy.Bear	1.210e+00	5.985e-01	2.021	0.050021	.
count_Toy	-1.017e+00	5.907e-01	-1.721	0.092942	.
participate	-3.460e-01	4.977e-01	-0.695	0.490973	
age	-1.399e+00	3.049e-01	-4.587	4.37e-05	***
count_Female	-1.077e-01	9.283e-02	-1.161	0.252687	
happy_to_depressed_flt_ratio	-5.143e-02	7.731e-02	-0.665	0.509710	
ANGRY	-6.168e+01	2.234e+01	-2.762	0.008645	**
ratio_Mammal	2.465e+01	5.739e+01	0.430	0.669862	
ratio_Cat	3.583e+01	5.139e+01	0.697	0.489731	
ratio_Building	1.048e+02	5.451e+01	1.923	0.061648	.
num_images_no_comments	1.147e-02	7.991e-03	1.435	0.159089	
count_Jar	-1.570e+00	4.541e-01	-3.457	0.001310	**
count_Shirt	-2.510e+00	4.908e-01	-5.115	8.21e-06	***
count_Bedroom	6.369e-01	3.110e-01	2.048	0.047181	*
ratio_Tattoo	-1.148e+03	3.836e+02	-2.994	0.004704	**
ratio_Terrier	6.057e+00	9.406e+01	0.064	0.948973	
ratio_Vegetable	-2.405e+02	5.062e+01	-4.751	2.60e-05	***
ratio_Carrot	6.483e+02	1.525e+02	4.252	0.000123	***
CALM	9.258e+00	5.709e+00	1.622	0.112722	
ratio_Drawing	-6.202e+01	1.538e+02	-0.403	0.688983	
count_DISGUSTED	3.606e+00	1.397e+00	2.582	0.013606	*
ratio_Handwriting	5.391e+02	1.173e+02	4.597	4.24e-05	***

Fig. A.3: Part 1

ratio_Dusk	1.482e+02	1.456e+02	1.018	0.314781	
count_Chair	-1.121e+00	2.526e-01	-4.436	7.00e-05	***
count_Musical.Instrument	-1.913e-01	6.339e-02	-3.018	0.004409	**
ratio_Art	1.005e+01	4.820e+01	0.208	0.835954	
SURPRISED_x	-8.772e+00	4.161e+00	-2.108	0.041329	*
count_Building	-4.186e-01	2.935e-01	-1.426	0.161556	
employed	-3.136e-01	7.263e-02	-4.318	0.000101	***
avg_posts_early_night	-1.259e+00	4.631e-01	-2.718	0.009659	**
count_Boat	-2.010e+00	6.454e-01	-3.114	0.003400	**
count_Abyssinian	-2.438e+00	7.217e-01	-3.378	0.001637	**
count_Smile	-1.111e-02	5.551e-02	-0.200	0.842343	
ratio_Glasses	-1.100e+02	7.707e+01	-1.427	0.161422	
income	-1.300e-02	3.579e-02	-0.363	0.718347	
SURPRISED_y	1.763e-01	1.379e-01	1.278	0.208460	
count_Sport	4.198e-01	1.189e-01	3.532	0.001057	**
count_Electronics	1.190e+00	3.125e-01	3.809	0.000471	***
avg_posts_night	1.086e+00	3.663e-01	2.966	0.005066	**
ratio_Rose	5.125e+01	6.140e+01	0.835	0.408885	
count_Bicycle	1.116e+00	3.577e-01	3.121	0.003341	**
ratio_Shirt	8.658e+01	4.778e+01	1.812	0.077490	.
ratio_Smile	-7.348e+01	2.459e+01	-2.989	0.004774	**
ratio_Potted.Plant	7.222e+01	2.512e+01	2.875	0.006451	**
cluster_2	-4.588e+01	1.417e+01	-3.238	0.002424	**
count_Drawing	1.248e+00	4.662e-01	2.677	0.010730	*
ratio_Light	1.352e+02	4.898e+01	2.760	0.008688	**
count_Clothing	-2.706e-01	2.187e-01	-1.237	0.223295	
ratio_Portrait	5.734e+01	2.304e+01	2.489	0.017059	*
ratio_TV	-2.306e+02	1.415e+02	-1.629	0.111077	
ratio_Bottle	2.006e+01	9.400e+00	2.134	0.038994	*
count_Home.Decor	-6.980e-01	3.585e-01	-1.947	0.058594	.
count_Vegetable	4.921e-01	2.873e-01	1.713	0.094514	.
ratio_Flower	-1.201e+00	9.425e-01	-1.275	0.209789	
avg_number_of_faces_over_images_with_faces	3.495e-01	2.355e-01	1.484	0.145641	
count_Blanket	6.737e-01	6.170e-01	1.092	0.281438	
ratio_Trademark	1.376e+01	1.349e+01	1.021	0.313616	
count_Overcoat	-3.647e-01	4.133e-01	-0.882	0.382863	

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

Residual standard error: 0.7392 on 40 degrees of freedom
Multiple R-squared:  0.9159,    Adjusted R-squared:  0.7373
F-statistic: 5.126 on 85 and 40 DF,  p-value: 7.523e-08
```

Fig. A.4: Part 2

Fig. A.5: Summary Stats of Linear Model to Predict E

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        7.24493    0.17910  40.452  < 2e-16 ***
count_Book        -0.64213    0.12860  -4.993 2.52e-06 ***
ratio_Vegetable   68.04385   35.21273   1.932 0.056144 .
count_Water        0.57180    0.19857   2.880 0.004870 **
ratio_Percussion -109.48501  22.24830  -4.921 3.39e-06 ***
count_Female       0.69103    0.45772   1.510 0.134267
ratio_Art        -96.16574   32.21905  -2.985 0.003568 **
count_Screen       1.33822    0.26262   5.096 1.64e-06 ***
count_Clothing    -0.77655    0.19065  -4.073 9.30e-05 ***
ratio_Beverage    35.42604   16.33613   2.169 0.032487 *
count_Bicycle      0.71822    0.19019   3.776 0.000270 ***
count_Suit         1.09618    0.29794   3.679 0.000379 ***
count_Appliance    1.20366    0.28073   4.288 4.17e-05 ***
count_Shoe        -0.98488    0.23521  -4.187 6.08e-05 ***
count_Canine      -0.08487    0.03336  -2.544 0.012498 *
ratio_Tree       291.91503  103.99975   2.807 0.006014 **
count_Child        3.74100    1.82587   2.049 0.043092 *
count_Bread        0.21239    0.23609   0.900 0.370483
count_Kid         -3.38019    1.80581  -1.872 0.064150 .
ratio_Road      -100.89487   51.64776  -1.954 0.053552 .
SURPRISED_y       -0.29323    0.10033  -2.923 0.004293 **
ratio_Red.Sky    377.49688  158.32097   2.384 0.018994 *
ratio_Wood      -183.60238  155.79514  -1.178 0.241400
count_Blonde      -0.86237    0.46934  -1.837 0.069119 .
ratio_Flyer       15.76897    8.51756   1.851 0.067070 .
count_Coast       -0.62689    0.42643  -1.470 0.144673
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.403 on 100 degrees of freedom
Multiple R-squared:  0.5476,    Adjusted R-squared:  0.4346
F-statistic: 4.843 on 25 and 100 DF,  p-value: 6.85e-09
```

Fig. A.6: Summary Stats of Linear Model to Predict R

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         6.93242    0.19230  36.049  < 2e-16 ***
count_Handwriting  -0.65849    0.18107  -3.637 0.000433 ***
ratio_Produce     107.29084   42.51655   2.524 0.013147 *
count_Child         1.66088    1.83527   0.905 0.367590
ratio_male          1.82625    0.71619   2.550 0.012245 *
ratio_Luggage    -146.55736   65.16869  -2.249 0.026648 *
ratio_Bedroom    -348.61658  141.29462  -2.467 0.015262 *
count_Cup           0.26879    0.08319   3.231 0.001656 **
count_Creme        -1.29359    0.50408  -2.566 0.011719 *
count_Canine        0.25359    0.11278   2.248 0.026679 *
ratio_Bottle       17.25622   10.16616   1.697 0.092638 .
count_Kid          -1.28274    1.83626  -0.699 0.486398
ratio_Terrier     125.61303   42.29588   2.970 0.003707 **
ratio_Canine     -181.85903   58.28062  -3.120 0.002343 **
count_Clothing     -0.35230    0.11047  -3.189 0.001891 **
ratio_Appliance   167.92504   65.43257   2.566 0.011715 *
count_Head          0.20524    0.09577   2.143 0.034455 *
ratio_Blanket    -108.42571   66.12849  -1.640 0.104135
count_Bedroom       0.12027    0.22735   0.529 0.597931
ratio_Jar         209.54226  131.47585   1.594 0.114052
count_Sunlight     -0.45849    0.18846  -2.433 0.016706 *
ratio_Manx        324.86073  171.64527   1.893 0.061215 .
count_Dawn          0.57047    0.38694   1.474 0.143449
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.388 on 103 degrees of freedom
Multiple R-squared:  0.4882,    Adjusted R-squared:  0.3789
F-statistic: 4.466 on 22 and 103 DF,  p-value: 9.83e-08
```

Fig. A.7: Summary Stats of Linear Model to Predict M

```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              8.458e+00  4.892e-01  17.290  < 2e-16 ***
imagecount                               2.046e-03  5.950e-04   3.439 0.000980 ***
ratio_Bike                               1.370e+02  1.181e+02   1.160 0.249950
ratio_Siamese                           -2.670e+02  1.363e+02  -1.959 0.054041 .
ratio_Beverage                           3.873e+01  1.300e+01   2.980 0.003944 **
count_Canine                             4.885e-01  1.215e-01   4.021 0.000143 ***
count_Terrier                           -8.019e-01  3.113e-01  -2.576 0.012086 *
count_Glasses                           -3.598e-01  8.667e-02  -4.151 9.06e-05 ***
ratio_Beard                              1.844e+02  4.353e+01   4.235 6.74e-05 ***
ratio_Hair                              -3.072e+01  2.459e+01  -1.249 0.215680
ratio_Collage                            4.974e+00  1.364e+01   0.365 0.716407
ratio_Head                               7.116e+01  2.281e+01   3.120 0.002617 **
count_Hair                              -2.253e-01  2.771e-01  -0.813 0.418868
ratio_Shoe                               7.533e+00  2.730e+00   2.759 0.007364 **
ratio_Canine                             2.125e+02  2.415e+02   0.880 0.381790
ratio_Road                              -1.245e+00  1.014e+02  -0.012 0.990237
ratio_Label                              9.801e+02  2.283e+02   4.292 5.50e-05 ***
ratio_Guitarist                         -4.123e+02  1.089e+03  -0.379 0.706092
ratio_Guitar                             2.855e+03  5.193e+02   5.499 5.68e-07 ***
ratio_Rose                               3.644e+02  8.342e+01   4.368 4.19e-05 ***
CONFUSED                                 2.209e+01  5.067e+00   4.360 4.31e-05 ***
ratio_Selfie                            -4.479e+00  6.205e+00  -0.722 0.472784
count_Female                            -2.521e-01  8.714e-02  -2.893 0.005064 **
ratio_Sunlight                          -1.802e+02  1.246e+02  -1.446 0.152546
count_Musical.Instrument                 7.988e-01  2.230e-01   3.582 0.000620 ***
count_Rose                              -4.788e-01  1.442e-01  -3.319 0.001426 **
ratio_Terrier                            3.568e+02  9.144e+01   3.902 0.000214 ***
ratio_Dog                               -5.088e+02  2.510e+02  -2.027 0.046410 *
ratio_Television                        -2.139e+03  8.073e+02  -2.649 0.009939 **
count_Clothing                          -4.050e-01  1.073e-01  -3.775 0.000329 ***
count_Paper                              6.289e-01  1.464e-01   4.296 5.44e-05 ***
HAPPY                                    2.405e+00  4.687e-01   5.132 2.40e-06 ***
avg_number_of_faces_over_images_with_faces -1.774e+00  4.274e-01  -4.152 9.04e-05 ***
count_Kid                                3.308e-01  1.110e-01   2.980 0.003944 **
CALM                                     7.901e+00  5.967e+00   1.324 0.189749
count_Handwriting                       -5.709e-01  1.680e-01  -3.399 0.001112 **
count_Bikini                             6.928e-01  1.665e-01   4.160 8.79e-05 ***
ratio_Overcoat                           4.780e+01  2.460e+01   1.943 0.055948 .
count_Collage                           -2.240e-01  7.286e-02  -3.075 0.002988 **
ratio_Musician                          -3.732e+03  9.946e+02  -3.752 0.000355 ***
count_Beard                             -5.129e-01  3.879e-01  -1.322 0.190364
ratio_TV                                 2.308e+03  8.132e+02   2.838 0.005912 **
ratio_Plant                              1.558e+00  9.031e-01   1.725 0.088827 .
employed                                -1.250e-01  5.535e-02  -2.259 0.026932 *
ratio_Child                              4.145e+01  2.121e+01   1.954 0.054600 .
count_Asleep                            -2.822e-01  1.382e-01  -2.042 0.044856 *
participate                             -7.730e-01  4.506e-01  -1.715 0.090624 .
ratio_Chocolate                          2.476e+02  9.132e+01   2.711 0.008407 **
ratio_Bowl                              -6.557e+01  2.813e+01  -2.331 0.022591 *
count_Bird                               3.992e-01  3.052e-01   1.308 0.195109
count_Alcohol                           -2.932e-01  1.205e-01  -2.434 0.017431 *
count_Shirt                             -4.143e-01  2.225e-01  -1.862 0.066758 .
count_Glass                              2.029e-01  9.244e-02   2.195 0.031399 *
avg_comments                            -2.489e-01  1.414e-01  -1.760 0.082749 .
user_followed_by                         2.313e-04  2.013e-04   1.149 0.254402
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.884 on 71 degrees of freedom
Multiple R-squared:  0.8078,    Adjusted R-squared:  0.6617
F-statistic: 5.527 on 54 and 71 DF,  p-value: 2.785e-11
```

Fig. A.8: Summary Stats of Linear Model to Predict A