



Selected Papers of #AoIR2021:  
The 22nd Annual Conference of the  
Association of Internet Researchers  
Virtual Event / 13-16 Oct 2021

## WHY MAKING MATTERS ONLINE: THE PINTEREST-DIY DATA SET

M. Sylvia Weintraub  
Sam Houston State University

David R.W. Sears  
Texas Tech University

### Abstract

The Do-It-Yourself (DIY) community is currently one of the largest creative content communities on Pinterest (Hall et al., 2018), a social networking service (SNS) that encourages users to both share information about creative processes and attempt projects in real life (IRL). Pinterest users share ongoing projects by creating Project “Pins”, which consist of images, videos, and text descriptions of creative content. And yet, while several studies have investigated user behavior in relation to everyday ideation and creativity on the site (Linder et al., 2014, Hu et al., 2018, Mull and Lee, 2014), little is known about the characteristics that lead users to prefer some DIY projects over others. Thus, this paper introduces the *Pinterest-DIY* data set, which consists of text data mined from 500 DIY project Pins on Pinterest. Using a custom sampling approach, we created a taxonomy of DIY characteristics related to each Pin’s project type, function, materials, and complexity. To measure user preferences on the site, we also conducted a sentiment analysis on user comments for each DIY project Pin. This paper introduces the data set and presents two use cases for the internet research community using both exploratory and confirmatory statistical methods. In our view, the Pinterest-DIY data set will provide further opportunities to examine whether, and to what degree, participation in online DIY communities promotes everyday creativity and increases engagement with physical matter.

### Introduction

Relative to other animals, humans are uniquely adept at manipulating matter, yet modern commodity culture discourages individuals from attempting to fix, adjust, construct, or re-fashion materials (Kuznetsov & Paulos, 2010). Moreover, it has been argued that virtual platforms distract and disengage users from their surrounding

Suggested Citation (APA): Weintraub, M.S., and Sears, D.R.W. (2021, October). *Why Making Matters Online: The Pinterest-DIY Data Set*. Paper presented at AoIR 2021: The 22nd Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from <http://spir.aoir.org>.

environment (Sears & Jacko, 2009). Thus, although humans exhibit creative and constructive making behaviors (e.g. creative ideation and everyday creativity), engagement in social media, social networking services (SNSs), and other online activities can discourage behaviors related to manipulating material entities. And yet, scholarship on SNSs like Pinterest, which affords users opportunities to learn about, and participate in, creative DIY practices, has yet to examine the impact of these online platforms on the DIY community. To address this issue, this paper introduces the *Pinterest-DIY* data set, which features text data mined from 500 DIY project Pins. We first describe the taxonomy of coded categories (i.e. Project Type, Project Theme, Material(s) Type, First Image, and Task Complexity) included in the data set, and then present two potential applications of the data set that feature both descriptive (exploratory) and inferential (confirmatory) statistical methods.

## Methods

The data for this study have been mined from DIY project Pins beginning prior to the Covid-19 pandemic. Pinterest's live feed showcases Pins as images, and DIY Pins were identified as aesthetic projects that may or may not also perform a utilitarian function. A two-tier systematic sampling approach was conducted that first searched for the keyword 'DIY' using Pinterest's search engine, and then filtered results to only include Pins that contain videos. Pins containing a minimum of five comments and a maximum of twenty-five comments were selected, and all reply comments appearing in threads were included. Self-promoting comments (e.g., "Pinboard" follow requests and advertisements) were excluded. For each Pin, the DIY Project Title, Description, number of comments at the time of data entry, whether videos begin with by showing in-process shots or completed projects, and the length of the videos were encoded. Finally, the total number of reactions (i.e. 'love', 'haha', 'thanks', 'good idea', and 'wow') to DIY Project Pins was also entered.

A preliminary exploratory study was conducted to create a taxonomy of relevant categories based on a subset of 100 Pins from the data set. Based on the initial pilot study, four categories were established to code the properties of each DIY project Pin in the final data set: *Project Type*, *Project Theme*, *Material(s) Type*, *First Image*, and *Task Complexity*. The category *Project Type* is divided into two sub-categories: 'Repurposed/Upcycled' projects, which refers to objects that are transformed to perform new functions or objects that are given new surface treatments; and 'New Builds', which are constructed from raw materials. Next, *Project Theme* includes sub-categories for 'décor', 'utilitarian', 'craft', 'holiday', and 'wearable'. The *Material(s) Type* category contains 9 possible sub-categories for materials, including 'paper', 'glass', 'wood', 'fibers', 'metal', 'ceramic', 'natural', 'synthetic polymers' and 'mixed', which consist of at least a 50/50 mixture of two or more materials. *First Image* includes the subcategories 'final product', 'work in progress', and 'both' to indicate that images show the project both in process and as a completed project. Finally, the *Task Complexity* category classifies Pins based on the number of steps required to complete the project using a 5-point Likert scale from low to high. Projects classified as 'low' require approximately 2-3

steps to complete; 'low medium' require approximately 3-5 steps; 'medium' require approximately 5-7 steps; 'medium high' require approximately 7-9 steps; and 'high' require approximately 9 or more steps to complete.

To quantify user preferences for each of the project Pins, the data set also includes estimates of the emotional valence of the sentiments expressed by users in the comments section. To that end, we conducted a sentiment analysis on the user comments using the Valence Aware Dictionary and sEntiment Reasoner (VADER; Hutto & Gilbert, 2014). The tool computes a compound valenced sentiment score in a range from -1 (most extreme negative) to 1 (most extreme positive) for all of the text comments in each Pin.

## Applications

This paper presents two use cases for the Pinterest-DIY data set, which will be made available in an online open-access repository on Github before the conference. First, we present the findings from a descriptive statistical analysis of the data set for each of the categories in the taxonomy. We found, for example, that DIY projects typically employed synthetic polymers as materials, followed by wood, fibers, glass, and metals. Second, the taxonomy can be used to predict various aspects of user behavior on Pinterest, including the user sentiments estimated by VADER. To predict the valenced sentiment scores, we conducted a multiple linear regression (MLR) analysis using the encoded categories from the taxonomy as predictors. The mean compound sentiment score for each Pin was entered into a stepwise MLR model with forward selection using the predictors *Project Type*, *Project Function*, *Material(s) Type*, and *Task Complexity*, as well as first-order interaction terms (e.g., *Project Type* × *Task Complexity*), using 5-fold cross-validation to reduce overfitting. The MLR model revealed significant parameter estimates for *Project Type* and *Project Function*, indicating users were more likely to provide positive comments for Repurposed/Upcycled projects as well Utilitarian and Décor projects.

## Discussion

This data set offers a glimpse into a massive social system of creative endeavors found in DIY communities on Pinterest by looking at a broad spectrum of projects and user-generated content. Indeed, by studying trending characteristics and the rich textual content found in user comments, it is our hope that future studies employing this data set will shed light on how users share information about DIY processes, how they value and evaluate creative projects, and which materials and processes are more desirable than others.

## Citations

- Hall, C., & Zarro, M. (2012). Social curation on the website Pinterest. com. *Proceedings of the American Society for Information Science and Technology*, 49(1): 1-9. doi: 10.1002/meet.145049011
- Hu, S., Torphy, K. T., Opperman, A., Jansen, K., & Lo, Y. J. (2018). What do teachers share within socialized knowledge communities: A case of Pinterest. *Journal of Professional Capital and Community* 3(2): 97-122. doi: 10.1108/JPCC-11-2017-0025
- Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In E. Adar & P. Resnick (Eds.), *Proceedings of the eighth international AAAI conference on weblogs and social media* (pp. 216–225). Palo Alto: AAAI Press.
- Linder, Rhema, Snodgrass, Clair, & Kerne, Andruid. (2014). Everyday ideation. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2411-2420.
- Mull, I., & Lee, S. E. (2014). “PIN” pointing the motivational dimensions behind Pinterest. *Computers in Human Behavior*, 33, 192–200. doi: 10.31274/itaa\_proceedings-180814-661
- Sears, A., & Jacko, J.A. (Eds.). (2009). *Human-Computer Interaction: Design Issues, Solutions, and Applications* (1st ed.). CRC Press.  
<https://doi.org/10.1201/9781420088861>