Paper Replication of

Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures[1]

Submitted by:

Aoyi Li, Yuan Liu, and Xuewen Yao

Supervised by:

Prof. Jacob Eisenstein

Abstract

We identified the hourly and daily changes of personal moods using Twitter from English-speaking individuals. We found out that positive mood had a peak early in the morning and started to decreased throughout the day, then reaching a not obvious peak around midnight. Negative mood is lowest in the morning, and then started to increase and then reached its peak near midnight.

1. Introduction

Big data analysis is the emerging trend for social computing and twitter, a popular social networking service, has been used in many researches. For example, there are previous studies in sentiment analysis, crime prediction, trend prediction and so on.

The main findings of the original paper [1] are they identified individual-level diurnal seasonal mood rhythms in cultures across the globe, using 509 million Twitter messages from 2.4 million individuals across two years. They found that individuals awaken in a relatively good mood that deteriorates as the day progresses, which is also consistent with the effect of sleep, circadian rhythm, or even seasonal change. In addition, it is also found out that people are happier on weekends but their morning peak is delayed by 2 hours which indicate people awaken later on weekends.

Though this paper is short in length, we think it offers a more complete analysis of sentiment across cultures and across time, and the authors tried to explain their observations using psychological knowledge and geographical difference, and that's what makes this paper stands out to us. We decide to replicate this analysis by using similar method to validate the correctness of results given in the original paper , as well as consolidating our understanding of the subject.

2. Goal

2.1 Main goal

Twitter's framing tends to yield in-the-moment expressions that reflect users' current experiences [2], and this feature makes tweet an ideal indicator of user's real-time mood. We would like to identify the individual-level hourly and daily mood changes using Twitter data sent from English-speaking individuals across the globe. Using Twitter as our data source, we are able to explore mood changes in terms of time, location, and network structure, with time being our focus here.

2.2 Stretch goals

Since the main findings were relatively straightforward, the team decide to do few extensions. With the amount of data and the nature of the dataset (see section 3.1, 3.2), we should able to draw more connections between mood change and other attributes in the dataset. First of all, we decide to look for whether one's friend count and follower count have any impact on their tweeting mood. Secondly, we are going to check whether the mood vary with different devices and platforms that user were tweeting from. Last but not the least, we will find the top PMI words for all the tweets from morning (7am-10am) and night (8-11pm). They are expected to be different and represent the characteristic of that time period.

3. Replication Method

3.1 Data Source

We get our raw data from Prof. Jacob Eisenstein. Based on the size of data, along with our computing power and time constraints, we decide to collect 4 weeks' of twitter data for this replication. In order to avoid potential bias, we spread out those 4 weeks throughout the year. Eventually, we got twitter data from 1st to 7th from January, April, July and October in 2016. There are approximately 137 million tweets in the dataset with a compressed size around 64GB.

3.2 Data filtering

In order to analysis the data, we have to filter it first. The first thing we have to filter is language. Similar to the original paper, our sentiment analysis is only targeting tweets that send out in English. After limit the 'language' attribute to be 'eng' or 'eng-BG', we have approximately 35 millions of tweets left. We are still considering the size of data is sufficient for this replication.

In the original dataset, all the timestamp is based on UTC+0. Since our analysis is heavily depend on the local time of the tweets, we have to convert time to be local time. It requires us to know the user time zone. So we filtered our data to only include tweets that send from a user who has a non-empty 'timezone_offset' attribute. It reduce our data to half from the previous step.

To prevent meaningless tweets from bot, we set a limit that each users can have a maximum of 35 tweets within this 28-days of dataset, tweets that come from a user with

35 or more tweets will be ignored. The amount of tweets that eliminate from this step is very small. Final step of filtering is to construct each tweet object to only contains useful attributes that we may need in this replication. In the original dataset, each tweet message is represented in a complex nested object, the size of each object is reduced significantly after this step.

Eventually we have 17 millions of tweets left after filtering, which we still considering to be a sufficient number in order to replicate the analysis. The compressed size of data is also reduced from 64 GB to 1GB which is more feasible to analysis based on our current computing power.

3.3 Sentiment Analysis

Valence Aware Dictionary and sEntiment Reasoner (VADER) [3] is a lexicon and rule-based sentiment analysis tool, which is open-sourced and works well for analyzing sentiments expressed on social media. We use VADER to compute positive, negative, neutral and compound score of each tweet. The positive, negative and neutral scores are the ratios for proportions of text falling into each dimension. And compound score, a weighted composite score of positive and negative score, is normalized to [-1, +1].

3.4 Outcome Plotting

After filtering, each tweet has a unix timestamp which represents the local time when the tweet was sent out. Then datetime library from python was used to extract day and time information. Tweets are grouped together first based on the day then the hour.

Scores associated with each tweet are obtained from sentiment analysis. Positive affect (PA) uses positive scores while negative affect (NA) uses negative scores. The mean of PA/NA is calculated for each hour. And the plots are shown below (Figure 1 & Figure 2).

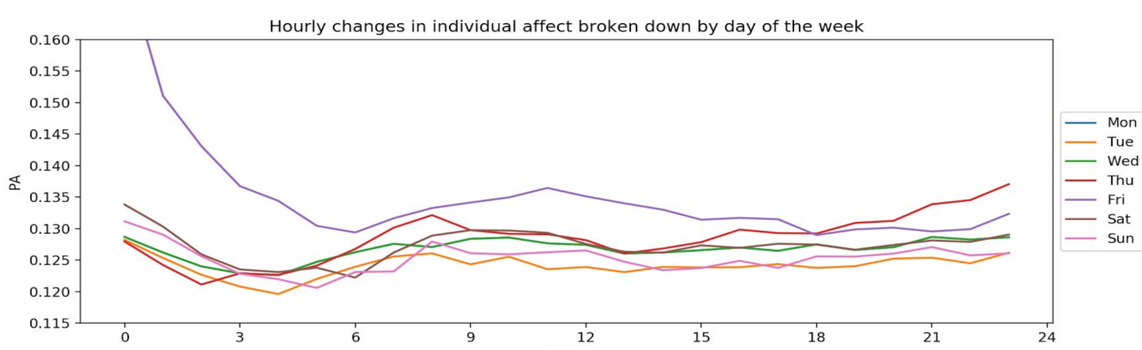

Figure 1. Hourly Changes of Positive Affect

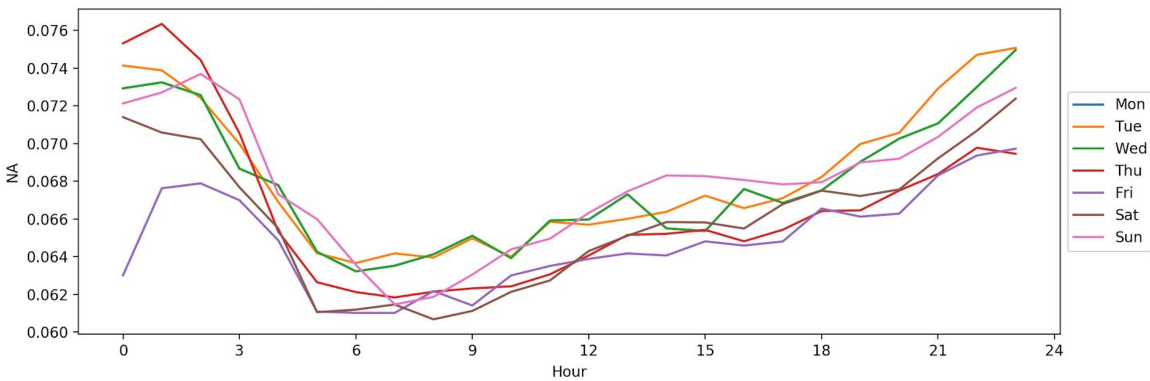

Figure 2. Hourly Changes of Negative Affect

4. Result

4.1 Main findings

We find that PA has a peak early in the morning then it starts to decrease throughout the day. However, it starts to increase again in the evening reaching a midnight peak though this trend is not obvious. PA decreases again after midnight. For NA, it is lowest in the morning, then it starts to increase till it peaks at midnight. These findings are the same as what the paper has found.

In addition, we plotted the compound score vs. hour for days of the week below (Figure 3). The trend of the curve resembles PA, and all scores are slightly higher than 0 meaning that overall PA dominates people's mood.

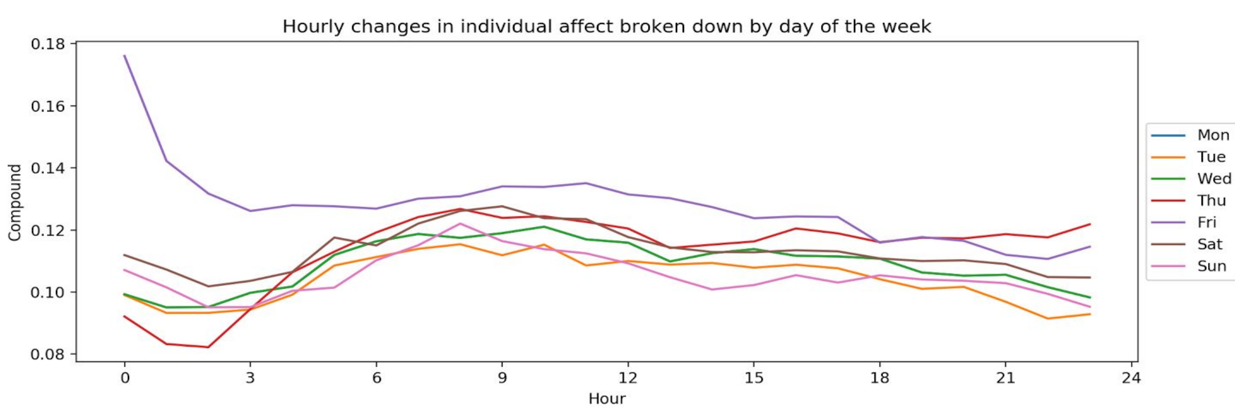

Figure 3. Hourly Changes of Compound Score

It can also be noticed that in our finding, the curve of Friday seems to be rather abnormal, thus we did more observations on the data and plotted more graphs for it. It is found that three of the four Fridays we have are holidays listed below, meaning that people have reasons to be happy and celebrate, therefore they are more likely to send out messages with positive affects, which explains why PA for Friday is way above the other days especially just past midnight and NA is low at midnight for Friday.

January 1	New Year
April 1	April Fool's Day
July 1	Canada Day

To confirm our speculation, we plotted PA/NA vs. Hour for all Fridays (Figure 4 & Figure 5). From the plot, we can see that people have more positive affects and fewer negative affects on Jan. 1, which is expected. But for other Fridays, the effects are not obvious. Also, PA for April 1, July 1 and Oct 7 are quite flat throughout the day, and we are not clear why it is so.

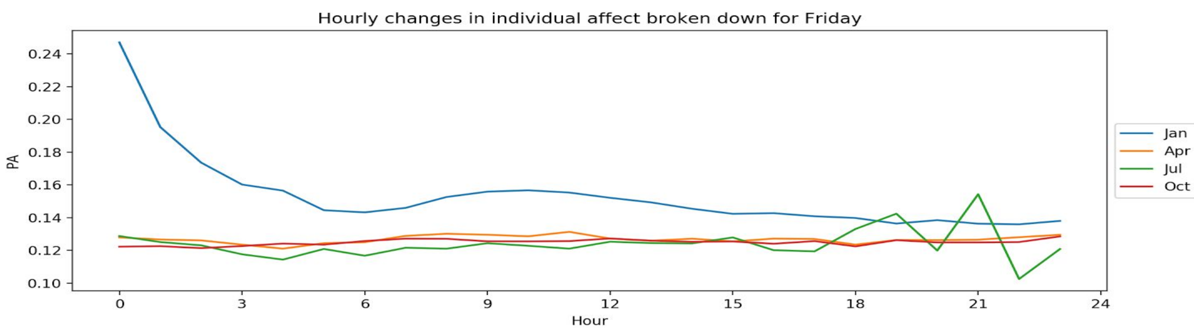

Figure 4. Hourly changes of PA for Fridays

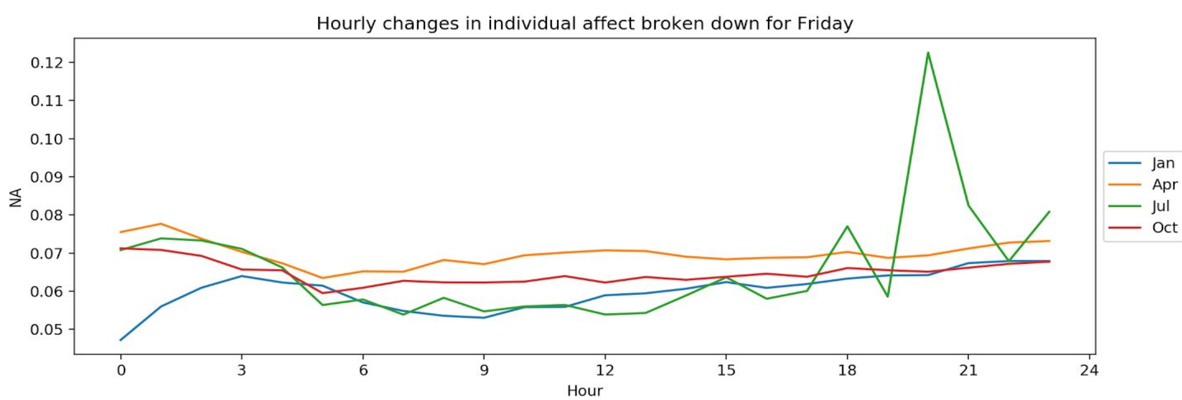

Figure 5. Hourly changes of NA for Fridays

4.2 Comparisons

Below are the graphs (Figure 6 & 7) cropped from the original paper. Compared with our findings above, it can be seen that both findings have the same trend as the original paper, but there are still some differences.

For PA, it peaked at 6 originally for weekdays and 8 for weekends in the paper. Nevertheless, it peaks at 8 for both weekdays and weekends in our case. It might be because different countries have different working hours [4] and in our replication, we limited the tweets to be in English and British English only, meaning that we probably limited our data to English-speaking countries where people usually work from 9 to 5 [4]. Also, the effects of weekends on PA cannot be seen from our replication. We think that's due to our lack of data. If enough data were given. The finding can be replicated completely.

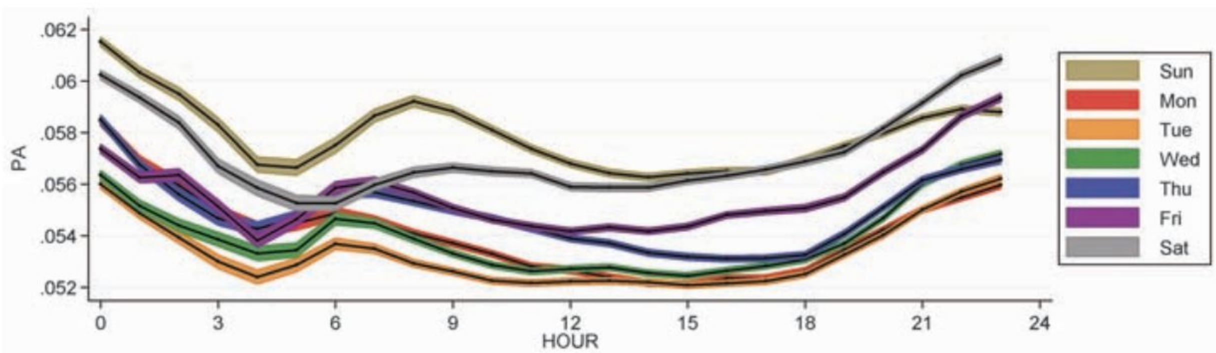

Figure 6. Hourly Changes of Positive Affect of the paper [1]

For NA, the graphs look very similar to each other, but the effects of weekends cannot be observed in this replication.

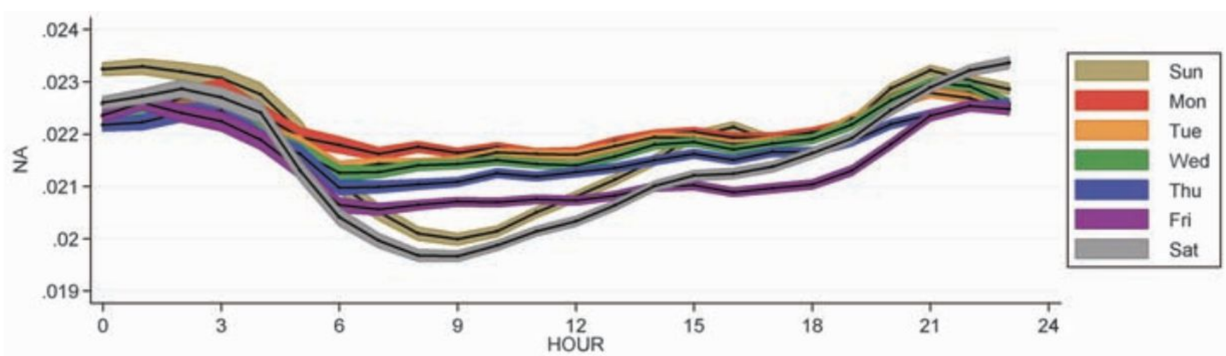

Figure 7. Hourly Changes of Negative Affect of the paper [1]

The findings of the paper can definitely be replicated. In our case, with only 4 weeks of data, we generated plots similar enough to the original one. And with more data given, weekend effects and seasonal changes can also be replicated.

5. Extension

5.1 Friends and Follower number impact

We divide user into five groups based on their friends number on twitter and plot the trend of their compound mood score. The result is in [Figure 6] :

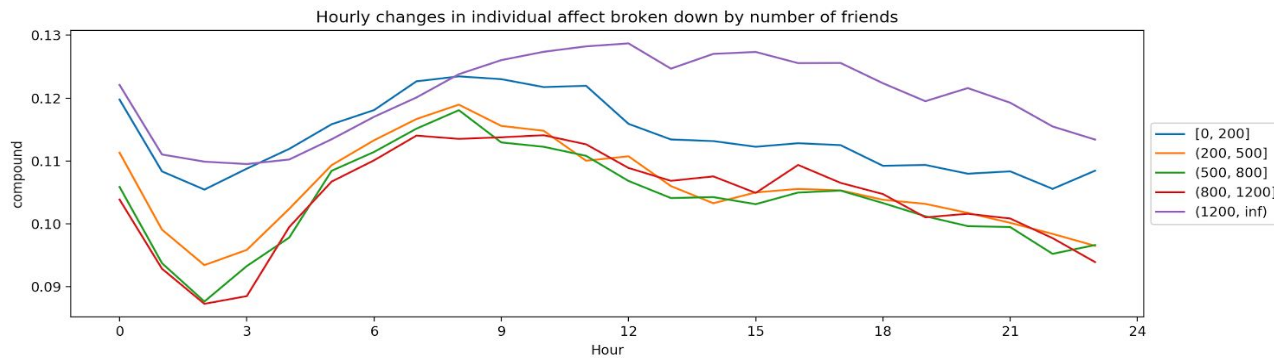

Figure 8.

And we did similar thing based on user's follower number [Figure 7] :

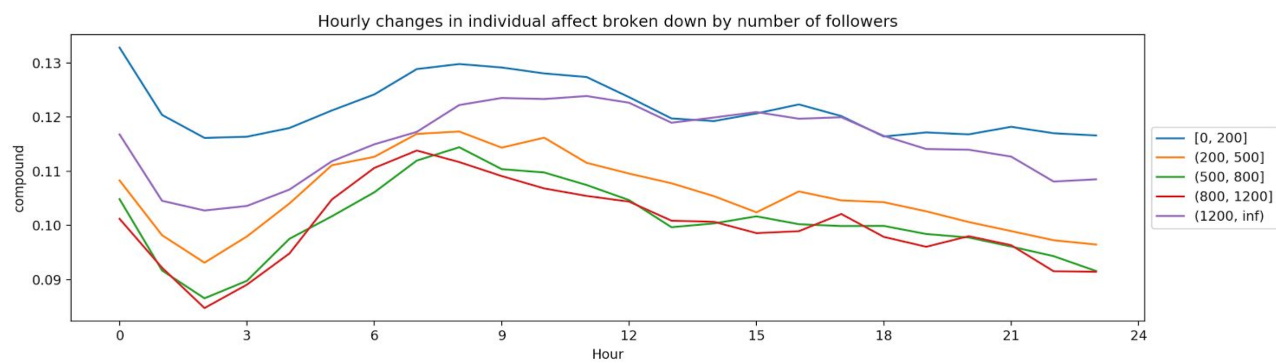

Figure 9.

The results of those two graph is similar and quite interesting: user's that have least number of followers/friends (0-200), along with user's that have most followers/friends (1200 and more) are likely to have slightly higher mood score compare to other user group.

5.2 Top PMI words

Pointwise mutual information (PMI), is a measure of association used in information theory and statistics[5]. Another extension we did is to measure the top PMI words morning and night, based on our dataset. Because we only have 28 days of data, to avoid potential bias (eg. one big news in a morning with huge number of retweet), we have to use only original tweet content to do our analysis.

The result is based on morning (7am-10am) and night (8-11pm):

Morning: ['morning', 'job', 'hot', 'christmas', 'today']

Night : ['drunk', 'kiss', 'drink', 'safe', 'haha']

Most of the result is reflecting the characteristic of both time period.

5.3 Devices impact

We also curious that whether running on different devices or sharing from other sources may have a different in identifying emotional changes. We use box plot and mean value to compare the difference in moods across sources, and mean value of positive scores are similar across the platforms, but the mean value of negative scores are lower for tweets shared from Facebook/Instagram. And this phenomenon may indicate that people are more willing to share, but now there is no evidence to support this statement.

As shown in Appendix A., p-value is computed to test if every two independent samples have identical average values. Sharing tweet from Instagram and sending from Web showed the most significant difference, in terms of positive score, but they are showing to be identical, in terms of negative scores.

Table 1. Usage Percentage and Mean of Positive Score and Negative.

	iPhone	**iPad**	**Android**	**Web**	**Facebook**	**Instagram**	**Others**
Percentage	42.3%	1.8%	19.5%	14%	1.3%	1%	20%
Positive	0.131	0.131	0.134	0.134	0.125	0.124	0.109
Negative	0.078	0.067	0.067	0.063	0.039	0.025	0.052

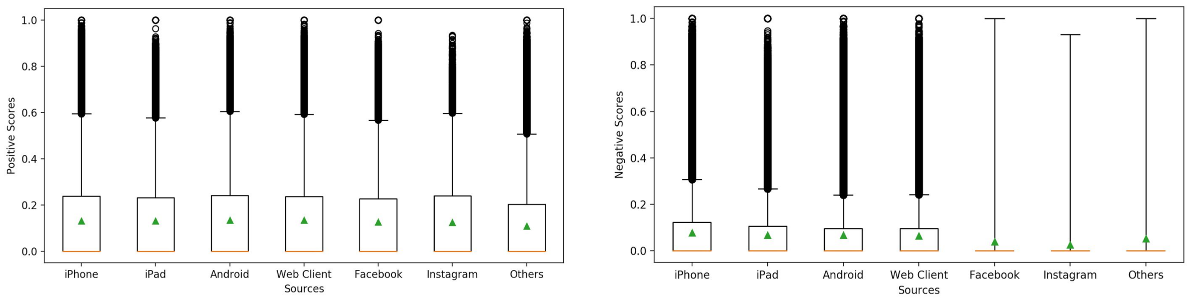

Figure 10. Box Plot of Positive/Negative Scores across devices

6. Conclusion

We successfully replicated the original paper [1], and examined how positive score and negative score vary with hours and days, based on individual-level. In addition, we also have few interesting findings in our extensions. However, there are also limitations in our research. First, lexical and sentiment analysis of tweets measure the textual expression of people, which may not reflect the real experience and feelings of people which is also the shortcoming of the original paper. Second, we are not able to replicate results focusing on seasonal changes and daylength, with the limited period of data we have.

Reference

1. S.A.Golder, M. W. Macy, Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures, *Science* 333, 6051 (2011). doi:10.1126/science.1202775

2. Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter." *PloS one* 6, no. 12 (2011): e26752.

3. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.

4. Working time. (2017, November 29). Retrieved December 06, 2017, from https://en.wikipedia.org/wiki/Working_time

5. Pointwise mutual information. (2017, December 02). Retrieved December 06, 2017, from https://en.wikipedia.org/wiki/Pointwise_mutual_information

Appendix A. P-Value - Device Impact Comparison

Sample A	Sample B	Positive	Negative
		p- value	p-value
iPhone	Facebook	1.14	0.00
iPhone	Instagram	1.96	0.00
iPhone	Others	0.00	0.00
iPhone	iPad	0.42	0.00
iPhone	Web	7.49	0.00
iPhone	Android	1.08	0.00
Facebook	Instagram	0.00	0.00
Facebook	Others	0.00	0.00
Facebook	iPad	1.89	0.00
Facebook	Web	3.72	0.00
Facebook	Android	3.76	0.00
Instagram	Others	1.35	0.00
Instagram	iPad	1.30	0.00
Instagram	Web	9.91	0.00
Instagram	Android	2.91	0.00
Others	iPad	0.00	0.00
Others	Web	0.00	0.00
Others	Android	0.00	0.00
iPad	Web	1.47	8.17
iPad	Android	1.80	0.48
Web	Android	0.04	5.51