Unveil the Mysterious Impact of Negative Reviews and Latitude on Helpfulness of Reviews : A Quantitative Analysis Based on Tripadvisor

MSMK7025A - “Hello World”

Tripadvisor

TABLE OF CONTENTS

Research Abstract

We started our research with a strong motivation derived from abundant outside research on the influence of negative reviews and latitude. Based on our research interest and our data set, we developed **5 hypothesis**.

Our analysis focuses on looking for significant relationships between relevant variables and helpfulness. Our results show that **rating and subjectivity are positively correlated with helpfulness, and latitude is negatively correlated with helpfulness.**

Based on our observations and findings stated above, we tailored our suggestions to be aligned with Tripadvisor's value, and we provided **8 suggestions** for them.

1.1 Motivation – "Negative" are overlooked in online marketplaces

Compared with positive reviews, negative reviews may be …

More Attention-Grabbing

People notice and care negative reviews more than positive ones — despite their questionable credibility

More Informative and Valuable

People may see **negative reviews as more informative**, and therefore more valuable, than positive ones because **they highlight defects** — even if they're not actually more accurate

More Helpful To Businesses

Businesses want to feel secure in their decision-making processes, and **they can utilize negative reviews to understand their risks and reduce losses**

More Likely to be Written Intentionally

"Super Contributors" On Tripadvisor may write more critically to appear more professional, Nevertheless, consumers disproportionately value and trust reviews professing expertise

References:
[1] Melián-González, Santiago, et al, "Online Customer Reviews of Hotels: As Participation Increases, Better Evaluation Is Obtained.", *Cornell Hospitality Quarterly*, March 29, 2013, doi:10.1177/1938965513481498.
[2] Sangwon Park, Juan L. Nicolau, "Asymmetric effects of online consumer reviews.", *Annals of Tourism Research*, January 2015, https://doi.org/10.1016/j.annals.2014.10.007.
[3] Beaton, Caroline"Why You Can't Really Trust Negative Online Reviews.", June 13, 2018, *The New York Times*, https://www.nytimes.com/2018/06/13/smarter-living/trust-negative-product-reviews.html

1.2 Motivation – "Latitude" influences human beings in various ways

If someone lives farther away from the equator, he/she is more likely to ...

Have Better Wellness

Globally, **happiness, creativity, life satisfaction, and individualism is higher** in countries farther away from the equator

Be Open and Extraverted

North America (eg. Canada), in comparison with Southeast (eg. Indonesia), is **higher in extraversion and openness to experience**

Have Lower Neuroticism

Polar workers **scored higher than a normative group on all factors except neuroticism**, which results in higher happiness and satisfaction

Live in A Richer Country

The world's most polar (far from the equator) population account for **21%** of the world population but produces **69%** of the world's GDP

References:
[1] Chen, Nao et al. "Where You Are Is Who You Are? The Geographical Account Of Psychological Phenomena.", *Frontiers In Psychology*, 24 March, 2020, https://doi.org/10.3389/fpsyg.2020.00536.
[2] Steel GD, Suedfeld P, Peri A, Palinkas LA., "People in high latitudes: the "Big Five" personality characteristics of the circumpolar sojourner.", *Environ Behav*., May 1997, doi: 10.1177/001391659702900302.
[3] Rentfrow, Peter J. et al., "Happy States Of America: A State-Level Analysis Of Psychological, Economic, And Social Well-Being", *Journal Of Research In Personality*, 25 August, 2009, https://doi.org/10.1016/j.jrp.2009.08.005.
[4] Tanzer, Sam, "Why Is There Such A Strong Correlation Between Geographic Distance From The Equator And Prosperity?", March 20, 2012, *Forbes*, https://www.forbes.com/sites/quora/2012/03/20/why-is-there-such-a-strong-correlation-between-geographic-distance-from-the-equator-and-prosperity/?sh=50ca304aab6d

1.3 Motivation – The Logic Chain (Hypothesis)

Research Questions

Q1: Does rating and subjectivity affect helpfulness?

Hypothesis 1: The review with lower Rating will be more helpful

Hypothesis 2: The more subjective of low rating reviews, the more helpful it will be

Fixed Effect

Q2: Does latitude affect helpfulness?

Hypothesis 3: Users in different cities have different preferences for reviews

Hypothesis 4: Latitude impacts helpful reviews in different cities

Hypothesis 5: In higher latitude cities, sentiment has less impact on whether reviews are helpful

2.1 Description Analysis

Summary Statistics

VarName	Obs	Mean	SD	Min	Max
helpful	149912	0.57	1.59	0.00	206.00
rating	149912	4.51	0.83	1.00	5.00
sentiment	149912	0.35	0.22	-1.00	1.00
subjectivity	149912	0.58	0.17	0.00	1.00
local	149912	0.20	0.40	0.00	1.00

Table1. Descriptive statistics of the main variables

In the sample data:

- The data distribution of Helpful is very **scattered** and the **value varies greatly**.
- The mean of Rating is high, **most reviewers like the restaurants they went to**.
- The mean of Subjectivity is 0.58 and the Standard deviation number is 0.17, which means **reviewsl include both objective and subjective descriptions and feelings**
- Most of the reviews are written by **nonlocal people**.

2.2 Model Design

MODEL	Dependent Variable	Independent Variable	Interaction variable
MODEL 1 - Basic linear model	Helpful	Rating	/
MODEL 2 - Interaction model		Rating	Rating*Subjectivity
MODEL 3 - Polynomial model		Rating + Subjectivity + Subjectivity^2	/
MODEL 4 - Fixed effects		Rating + Factor(cities)	/
MODEL 5 - Linear regression		Latitude	/
MODEL 6 - Interaction model		Latitude	Latitude*Sentiment

Table2. Variable Setting Instruction

Additional Notes:

*We searched the **exact number of latitude of cities** mentioned in the dataset on Google Map and **take their "absolute values"** to **create a variable called "latitude"**(eg: NewYork0=NewYork1=...=NewYork29=40.71; the special one is Sydney, which is 33.87°S, and we kept it as 33.87).*

2.3 Model 1 - The relationship between Helpful and rating

	Dependent variable:
	Helpful
Rating negatively related	−0.269***
	(0.005)
Constant	1.780***
	(0.022)
Observations	149,912
R^2	0.020
Adjusted R^2	0.020
Residual Std. Error	1.570 (df = 149910)
F Statistic	3,050.481*** (df = 1; 149910)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table3. Linear regression of Helpful and Rating

#1 Data Processing

```
result <- lm(Helpful ~ Rating, data = mydata)
summary(result)
```

#2 Data Finding

Helpful=1.78-0.269Rating

#3 Result

- Helpful and Rating is significantly negatively correlated.
- **People like reviews with low rating.**

2.3 Model 2 - Interaction effect of Rating and Subjectivity

#1 Data Processing

```
# linear regression
result <- lm(Helpful ~ Rating*Subjectivity, data = mydata)
summary(result)
```

	Dependent variable:
	Helpful
Rating	−0.446***
	(0.016)
Subjectivity	−1.629***
	(0.127)
Rating:Subjectivity	0.330***
	(0.028)
Constant	2.654***
	(0.072)
Observations	149,912
R^2	0.021
Adjusted R^2	0.021
Residual Std. Error	1.569 (df = 149908)
F Statistic	1,076.863*** (df = 3; 149908)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table4. Interaction effect of Rating and Subjectivity

#2 Data Finding

Helpful=2.654-0.446Rating-1.629Subjectivity+0.330Rating*Subjectivity

Rating	Subjectivity	Helpful
0	0	2.654
0	1	1.025
1	0	2.208
1	1	0.909

Table5. Interaction effect Matrix

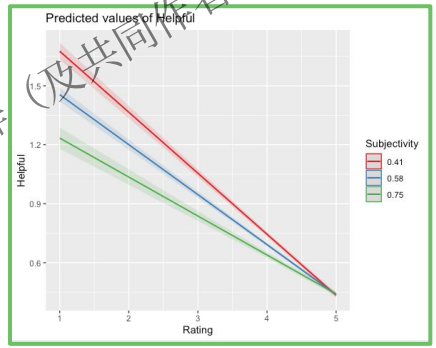


Figure1. Interaction plot

#3 Results

- Rating and Helpful is significantly negatively correlated(-0.446***). People dislike high rating reviews.
- Subjectivity and Helpful is significantly negatively correlated(-1.629***). **People dislike subjective reviews.**
- Subjectivity increases the {lower rating is more useful} effect(0.330***). (主观性会增强“打分越高越有用”效应)Although people dislike subjective reviews, **while given it is already a subjective review, people would prefer it if it is a high rating review.**

2.3 Model 3 – The quadratic relationship between Subjectivity and Helpful

	Dependent variable:
	Helpful
Rating	−0.267***
	(0.005)
Subjectivity	−0.449***
	(0.080)
I(Subjectivity2)	0.298***
	(0.073)
Constant	1.920***
	(0.031)
Observations	149,912
R^2	0.020
Adjusted R^2	0.020
Residual Std. Error	1.569 (df = 149908)
F Statistic	1,034.348*** (df = 3; 149908)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table5. The quadratic relationship between Subjectivity and Helpful

#1 Data Processing

```
result <- lm(Helpful ~ Rating + Subjectivity
+I(Subjectivity^2), data = mydata)
summary(result)
```

#2 Data Finding

Helpful=0.298Subjectivity^2-0.449Subjectivity-0.267Rating+1.92

#3 Results

- The coefficient of the quadratic term is positive. Therefore, people like very subjective and very objective reviews.

2.3 Model 4 – Does different cities have an fixed effect on Usefulness?

	Dependent variable:
	Helpful
Rating	−0.316***
	(0.005)
factor(City)Chicago	0.028*
	(0.016)
factor(City)LasVegas	0.407***
	(0.015)
factor(City)London	0.952***
	(0.017)
factor(City)LosAngeles	−0.125***
	(0.019)
factor(City)NewYork	0.526***
	(0.016)
factor(City)Toronto	0.018
	(0.017)
factor(City)Vancouver	0.069***
	(0.017)
Constant	1.735***
	(0.024)
Observations	149,912
R^2	0.059
Adjusted R^2	0.059
Residual Std. Error	1.538 (df = 149903)
F Statistic	1,174.386*** (df = 8; 149903)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table6. Use city to do fixed effect

#1 Data Processing

```
result <- lm(Helpful ~ Rating + factor(City), data = mydata)
summary(result)
```

*Here we reset Sydney(with the lowest latitude) as our benchmark.

#2 Data Finding

Helpful = -0.316Rating+

0	If Sydney
-0.125	If Los Angeles
0.407	If Las Vegas
0.526	If New York
0.069	If Vancouver
0.952	If London

Low → Latitude → High

#3 Results

- There are **regional differences in the usefulness of the reviews.**
- The regional differences between Chicago, Toronto and Sydney are not significant, while the differences in other cities is strongly significant.

Generally, we find that people's preference of reviews roughly follows the ascending trend of latitude.

Then, base on our motivation, we conjecture that **whether there is correlation between latitude and people's preference of reviews?** We conducted further analysis in the following slides.

2.3 Model 5 – Linear regression of Helpful and Latitude

	Dependent variable:		
	Helpful		
	all	Local=1	Local=0
	(1)	(2)	(3)
Latitude Positive correlated	0.018***	0.032***	0.013***
	(0.001)	(0.002)	(0.001)
Constant	−0.138***	−0.762***	0.079**
	(0.034)	(0.068)	(0.040)
Observations	124,236	26,563	97,673
R^2	0.004	0.015	0.002
Adjusted R^2	0.004	0.015	0.002
Residual Std. Error	1.626 (df = 124234)	1.658 (df = 26561)	1.616 (df = 97671)
F Statistic	455.411*** (df = 1; 124234)	392.052*** (df = 1; 26561)	166.883*** (df = 1; 97671)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table7. Linear regression of Helpful and Latitude

Interpretation of the larger value appeared at "Local=1": latitude are likely to have a larger impact on people who are born or live in the area for a long time,

#1 Data Processing

```
result <- lm(Helpful ~ Latitude, data = mydata)
summary(result)
```

#2 Data Finding

【All】Helpful=-0.138+0.018Latitude
【Local=1】Helpful=-0.762+0.032Latitude
【Local=0】Helpful=0.079+0.013Latitude

#3 Results

- Latitude is positively correlated with Helpful. (0.018***)People like reviews written in higher latitudes.
- The impact of latitude on helpfulness is larger on local people than non-local people. (0.032***>0.013***)

2.3 Model 6 – Interaction effect of Latitude and Sentiment

#1 Data Processing

Step1: basic model

```
result <- lm(Helpful ~ Latitude*Sentiment, data = mydata)
summary(result)
```

Step2: use Local=0/1 as conditional variable

All variable
significantly related

	Dependent variable:		
		Helpful	
	all	Local=1	Local=0
	(1)	(2)	(3)
Latitude	0.030***	0.047***	0.023***
	(0.002)	(0.003)	(0.002)
Sentiment	0.778***	1.130***	0.589***
	(0.152)	(0.298)	(0.179)
Latitude:Sentiment Negative correlated	−0.032***	−0.041***	−0.028***
	(0.004)	(0.007)	(0.004)
Constant	−0.442***	−1.181***	−0.163**
	(0.062)	(0.121)	(0.074)
Observations	124,236	26,563	97,673
R^2	0.010	0.022	0.008
Adjusted R^2	0.010	0.022	0.008
Residual Std. Error	1.621 (df = 124232)	1.652 (df = 26559)	1.611 (df = 97669)
F Statistic	415.156*** (df = 3; 124232)	202.930*** (df = 3; 26559)	247.465*** (df = 3; 97669)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table8. Interaction Effect of Latitude and Sentiment

#2 Data Finding

Helpful=-0.442+0.03Latitude+0.778Sentiment-0.032Latitude*Sentiment

Latitude	Sentiment	Helpful
0	0	-0.442
0	1	0.336
1	0	-0.412
1	1	0.334

Table10. Interaction Effect Matrix

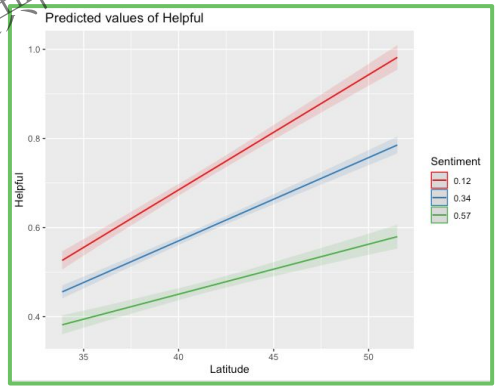


Figure2. Interaction Plot

#3 Results

- People prefer positive reviews. But the higher the latitude, the impact of positive sentiment on helpful is decreasing.(高纬度会削弱“越积极越有用”效应)
- Given a specific sentiment value, its positive impact on helpfulness is not as strong in a higher latitude.

3.1 Implications for TripAdvisor – Research Question #1

Encourage users to write reviews & Improve the quality of reviews

Observations/Findings

Finding 1: The review with lower rating will be more helpful

Finding 2: The review with higher objectivity will be more helpful

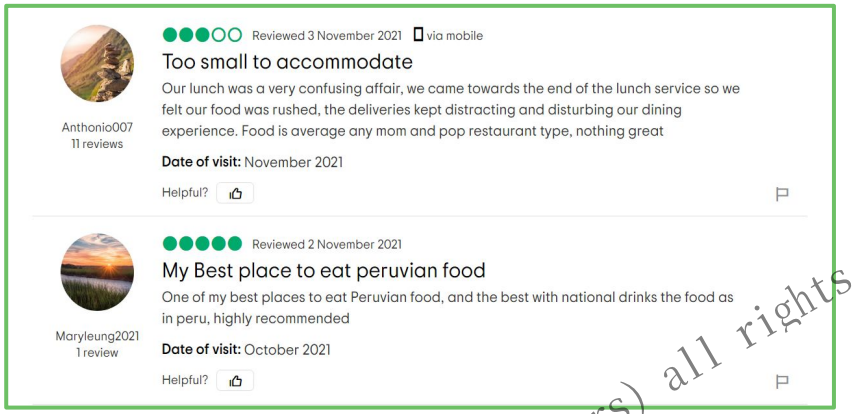
Finding 3: The more subjective of low rating reviews, the more helpful it will be

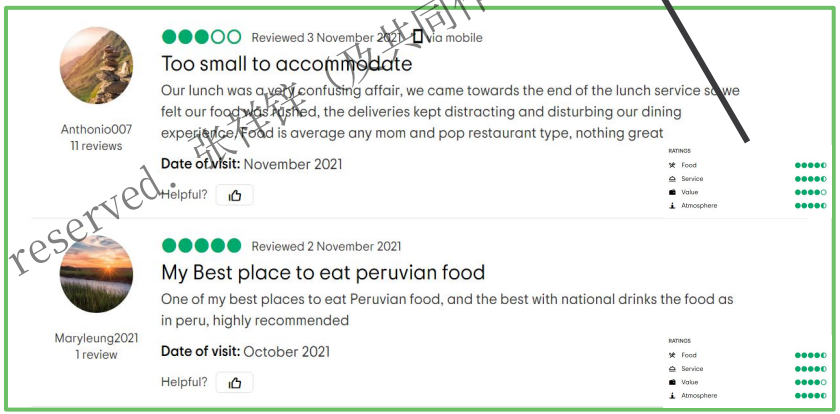
Observation 1: Most of the reviews are written by nonlocal people

Suggestions

1. Advance the algorithm to identify **intentional negative reviews** and adjust their impact
2. To enhance the credibility of ratings by **showing more evaluating dimensions of every review,** such as food, service, value, and atmosphere
   (A redesigned layout is shown on the next page)
3. **Help objective reviews to be popular** and get more attention by displaying them on the landing page and ranking them higher in the comment section
4. **Encourage local people to post more reviews**, and increase the weight of local people's reviews in the ranking

Showcase - Implications 3.1 - Suggestion #2





Original Layout

It shows general reviews, which only include rating, helpful, date of review, and data of visit, etc

Redesigned Layout

The multi-dimensional reviews that cover food, service, value, and atmosphere.

3.2 Implications for TripAdvisor – Research Question #1

Reduce the anchoring effect caused by extreme ratings

Observations/Findings → **Suggestions**

Finding 1: The review with lower Rating will be more helpful

Finding 3: The more subjective of low rating reviews, the more helpful it will be

1. Enable after-review response and **encourage restaurants to respond** to those extreme reviews. For example, they can offer refunds or future discounts to ask for deletion of extreme reviews, etc.

2. **Add a follow-up review function** ("追评功能"), which allows updated reviews to be seen. If a person gave a negative review at the first hand but gave a better review later, the follow-up review can help mitigate the anchoring effect caused by the prior extreme rating

3.3 Implications for TripAdvisor - Research Question #2

Leverage the impact of latitude on the helpfulness of reviews

Observations/Findings → **Suggestions**

Finding 4: Latitude impacts helpful reviews in different cities

Finding 5: In higher latitude cities, sentiment has less impact on whether reviews are helpful

To gain higher retention and loyalty, Tripadvisor should develop the algorithm (eg. utilize AI or machine learning) to show person-specific contents based on customers' geographic locations (基于纬度的千人千面)

One example: show more positive reviews on the recommendation page for high-latitude users than for low-latitude users

4. Limitations and Future Research

#1 Data Constraints on Latitude

- Our research is restricted by the fact that we only have **8 variations in the latitude data**, which hinders the strength of our findings.
- The majority of cities are in the relatively **higher latitudes**. (3 out of 8 cities are not considered high in latitudes, and only one city is considered to be in a low latitude.)
- All cities in the dataset locate in **developed countries**. We were unable to investigate how rating correlate differently with latitudes over developing countries/underdeveloped countries, therefore our findings may not necessarily apply to those regions.
- Future researchers can build on our research by utilizing a data set that has more diverse distributions across latitudes and regions.

#2 Weak R-Squared in the models

- In our analysis, the R-squared for our models are mostly around **0.01-0.02**, which implies that our independent variables only explains 1%-2% of the variations in dependent variable. Therefore, it is **not suggested to use our models for forecasting** votes received.
- Since there are many underlying elements affecting the complex result of helpful votes received, we admit that it's not easy to obtain a high R-squared in our models. The high significance of our estimates still implies that our findings are valid.

4. Limitations and Future Research

#3 Outliers Found in Quadratic Model

- In the third model of our analysis, we attempted to investigate the relationship between Subjectivity and Helpful, and received a **positive estimate on the quadratic term**. However, by observing the histogram, the majority of the data (Subjectivity:0.15-0.95) exhibit a flat and **downward curve**(see Figure. 3).
- We conclude that the positive estimate is affected by the **outliers** on both end of the dependent variable. Since Helpful is much higher when Subjectivity is extremely high extremely low, an upward U-shape curve is found by our model.
- In conclusion, the quadratic model is not a perfect fit for the data. By only considering the model, one might overlook the trend of the data points in the middle range of the independent variable.

#4 Omitted variables

- Due to the constraint of the dataset, we were **unable to consider certain features** of a location that can be potentially important to the helpfulness of a review, for example, culture, economical and political situation of a specific region.
- Future researchers can build on our research by incorporating such features into their analysis.

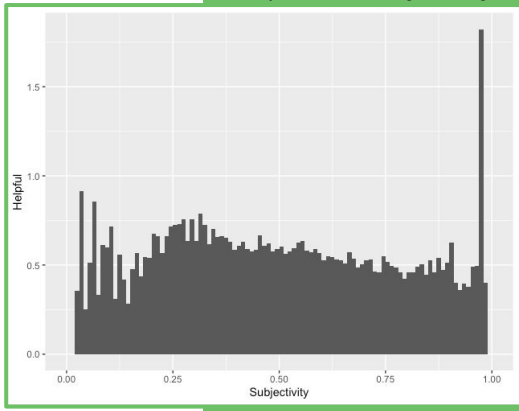


Figure.3 The Relationship Between Helpful and Subjectivity