



# A Quantitative Analysis on Customer Feedback in Hotel Industry

A Personal Project by  
Yusi Chen  
June 19<sup>th</sup>, 2020

# The Project at a Glance

## Project Purpose

Quantify customer feedbacks and extract business insights for hotel industry based on a large sample through advanced data analytics.

## Modules Used

Natural Language Toolkit (NLTK), Numpy, Pandas, Wordcloud, Scipy, data visualization (Tableau, Seaborn, Matplotlib )

## Dataset Description

This dataset is a list of 879 hotels in the US and 35,000 reviews collected mainly between 2013 and 2017 from Datafiniti's Business Database.

## Question:

When a customer is unsatisfied with the hotel experience, will she/he write a longer feedback ?

## Hypotheses:

Independent Sample T-tests: Test the differences in the average length of comments between customers who rated the hotel low and those who didn't.

## Workflow:

Data acquiring → data cleaning → data analysing → data visualizing → data reporting  
Dataframe shape (35912 X 19) → (34181 X 4538)

[https://www.kaggle.com/datafiniti/hotel-reviews?select=7282\\_1.csv](https://www.kaggle.com/datafiniti/hotel-reviews?select=7282_1.csv)



# Customer Experience is What Hotels All About



# Customer Review Matters



---

## 01 MARKET SIZE

The market size of the global hotel industry was more than **\$ 600 billion** in 2018.

---

## 03 MINIMUM RATING TO GO

87% of American-based customers need a business to have a **minimum star rating of 3** to use it

---

## 02 REVIEW MEANS MONEY

A 1-star increase in ratings equals a **5-9% increase in revenue**. 1-1.5 star rating reports 33% less revenue than the average ones

---

## 04 PEOPLE TRUST REVIEWS

93% of the people use online reviews when determining which hotel they want to stay at







## QUESTION

---

When a customer is unsatisfied with the hotel experience, will she/he write a longer feedback ?

# Data Overview

- 36k reviews for 879 hotels located in 761 cities across the US.
- On average 500 review per hotel.
- Highly imbalanced rating distribution, only 17.7% were bad ratings.

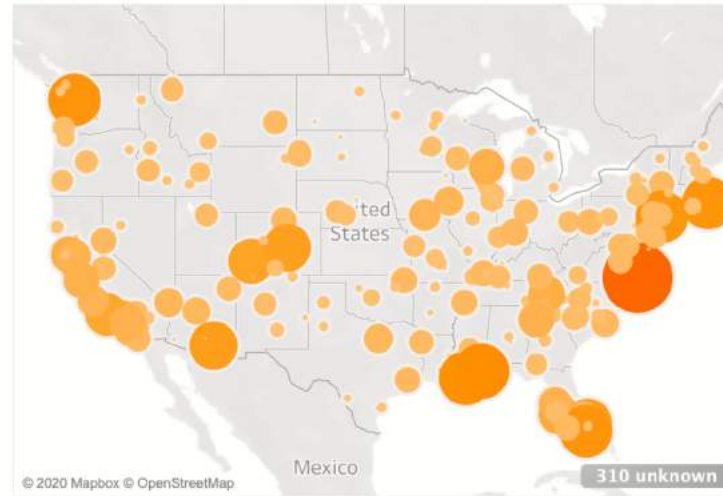
Review Distribution



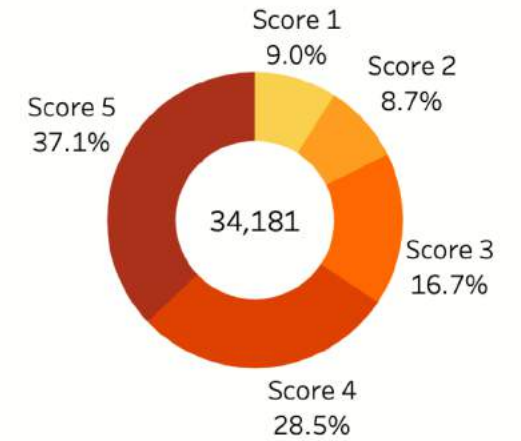
Rating Distribution #



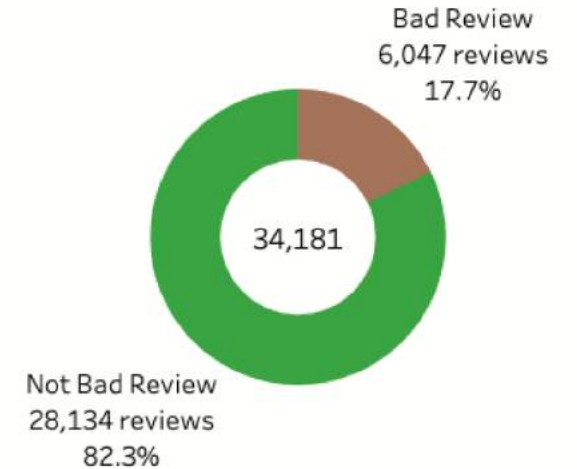
Rating Distribution by City



Rating Distribution %



Bad Rating Distribution %



# Sentiment Analysis – Quantify Customer Feedback Emotions

The Natural Language Toolkit (NLTK)



work with human language data for applying in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning.



Vader

a part of the NLTK module designed for sentiment analysis. for each text, Vader returns 4 values.

Example:

‘ Very nice for the price.’

a negativity score

0

a neutrality score

0.564

a positivity score

0.436

a compound score

0.475



## Highest positive sentiment reviews



Pos	Customer Comment
1	Nice - friendly - helpful
0.932	Super friendly, helpful, gracious staff!
0.926	great value. Clean, comfortable room
0.91	Great value! SUpEr friendly, super comfortable, super clean! Totally worth the price :
0.896	Comfortable and a good value.
0.884	great room value, comfortable, clean, nice room.
0.865	awesome hotel! great staff, super friendly yummy breakfast!
0.865	Awesome hotel! Great Staff, super friendly yummy breakfast!
0.864	Friendly, great service, comfortable, great location, all great!
0.862	great value quiet clean and comfortable

## Highest negative sentiment reviews



Neg	Customer Comment
0.732	Dirty, roaches, dirty linens, terrible breakfast, bad smell, terrible stay !!!
0.7	Bad staff and bad stay
0.699	No securiry cam. Bad service, nasty room
0.681	Front desk dirty overall bad
0.676	Horrible horrible horrible staff. Valet parking was terrible temperature of room was terrible
0.67	It was nasty and stinky
0.662	Bugs broken tiles broken tub missing towels missing light bulbs broken door
0.651	terrible staff. he over charged me terrible f....cking arrogant manager
0.62	No problems and not exceptional.
0.612	Worst place ever! Bad service, worst shuttle service! The stay overall is horrible ;(





# Hypothesis Testing: Independent Sample T-tests



**H0** There is no difference in the average length of comments between customers who rate the hotel low ( only give 1-2 star ) and those who don't.



**H1** There is a difference in the average length of comments between customers who rate the hotel low ( only give 1-2 star ) and those who don't.



Python

```
from scipy.stats import ttest_ind

Bad=reviews_df[reviews_df['is_bad_review']=='Bad Review']['nb_words']
NotBad=reviews_df[reviews_df['is_bad_review']=='Not Bad Review']['nb_words']

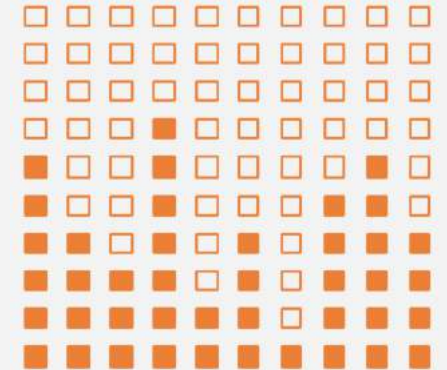
stat,p=ttest_ind(Bad,NotBad,equal_var=False)
```



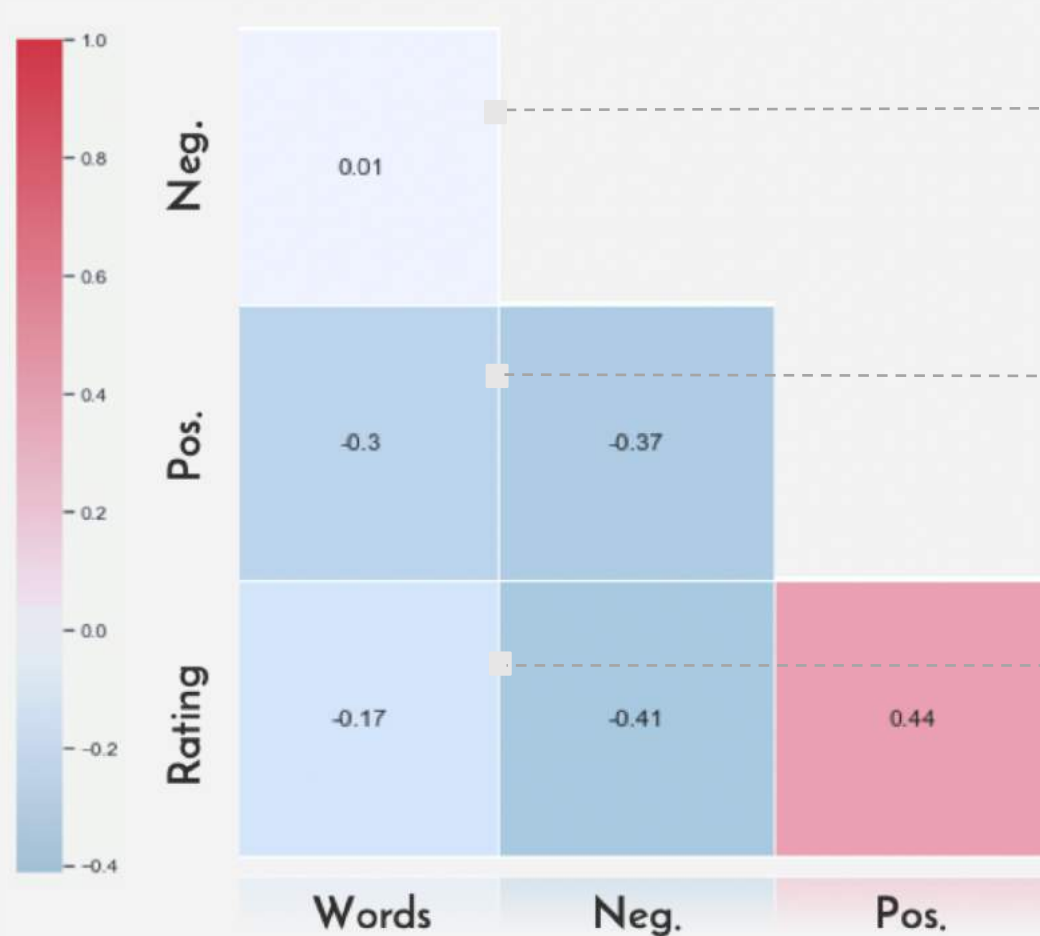
Result

P-value ( 3.562280119116915e-124 ) < Alpha ( 0.05 ). H0 can be rejected.

→ Significant differences in the mean number of words in customer comments between these 2 populations.



# Correlation Analysis: emotions VS length of comments



Almost no correlation between length of comments and negative emotions

A negative correlation.

When a customer feels positive about a hotel, a much shorter feedback story occurs.

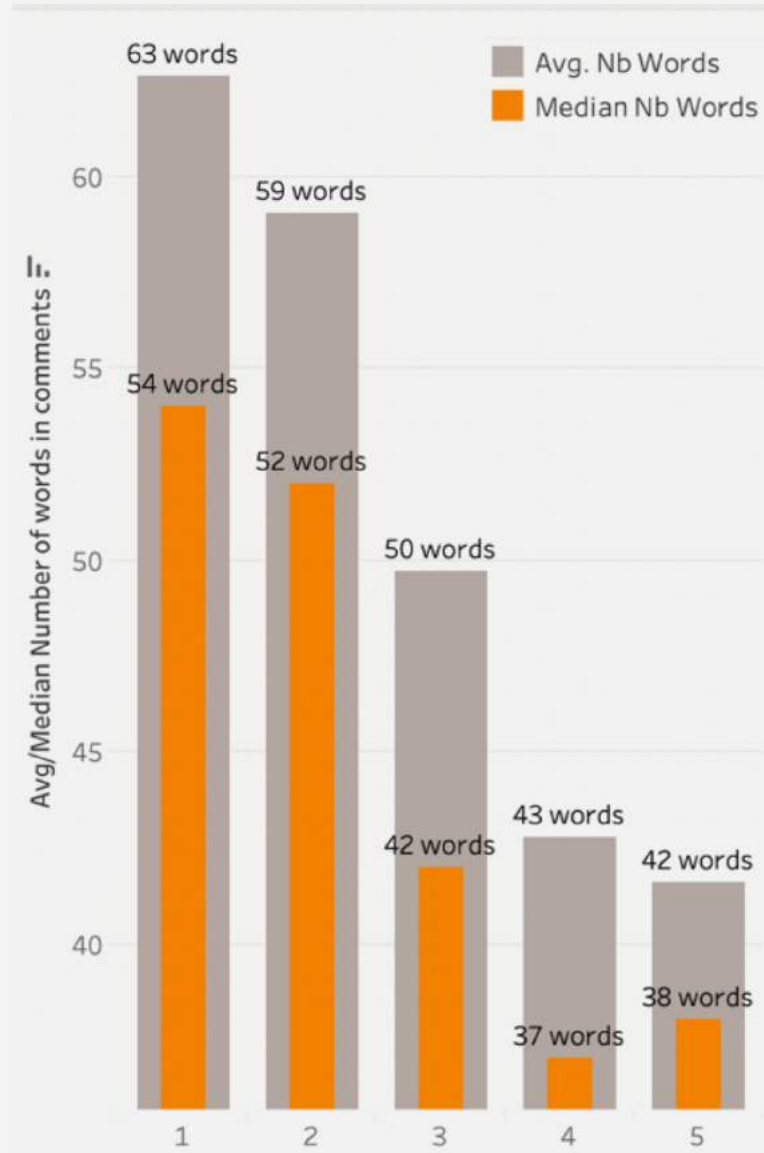
A slightly negative correlation.

The higher a customer rates a hotel, a shorter feedback is observed.

When a customer have a negative or positive feelings about the hotel experience, will she/he write a longer feedback ?

# Further Analysis:

## Feedback Word Length VS Review Rating



### Mean/Median length for each rating

- The lower the rating, the more words were written in customer feedback.
- Except for the rating of 4, the Mean and Median trend lines are highly parallel.
- People who give lowest rating write on average 1/3 longer than those give highest rating. That's 21 words in difference !

When a customer is unsatisfied with the hotel experience, will she/he write a longer feedback ?

YES



## Conclusion

unhappy → low rating → long feedback  
negative sentiment in feedback != long feedback  
negative sentiment in feedback != short feedback

happy → high rating → short feedback  
positive sentiment in feedback → short feedback



## Take-away

- Train your staff to invite customer to write a review
- Make it easy
- In exchange of benefits
- Response to feedbacks

Create Exceptional Customer Experiences To Get Positive Customer Reviews







**Empowering businesses  
through data analytics**



Keep in touch  
[yusi0928@gmail.com](mailto:yusi0928@gmail.com)

**Thank You**