

# Text analysis and Sentiment analysis of AirBnb Users' reviews using SAS Enterprise Miner

## ABSTRACT:

The objective of this paper is to increase credibility of website by doing text analysis of users' reviews to understand what facilities guest wants and what are the things/events which caused inconvenience during their stay.

In this paper, my goal is to analyze the overall sentiment prevailing in AirBnB reviews. In order to achieve this, I have extracted real time data from AirBnB studied the responses of the people. I have used SAS® Enterprise Miner and SAS Sentimental Analysis Studio to evaluate key questions regarding the do's and don'ts for a host such as understanding the needs and expectations of guest and perception of guest about what they like and what they didn't like during their stay.

This analysis will allow AirbnB to plan further to improve their service by indicating their host about dos and don'ts.

The analysis will be done by leveraging the text analytics capability of SAS Enterprise Miner 14, using text parsing, text filter nodes to train model to classify positive and negative reviews. Further, this model will classify whether the review given by user is positive, negative or neutral. This will help website to efficiently understand users' needs and thus will help in increasing their revenue.

## INTRODUCTION:

Today tourism has become integrated part of our life and within past few years, this industry has gained a great deal of popularity. People these days want to see world at minimum expenses. Hotels are one of the main concerns while planning for a vacation, business trips etc. AirbnB is a well-known company which has provided a platform for hosts and visitors to communicate. They provide a platform where traveler don't have to worry about authenticity of host and the place. AirbnB is the world's largest home sharing company and has over 800,000 listings in more than 34,000 cities and 190 countries. Entire website is based upon the rating system, which is a both way process. Host who offer the place also gets rating on different parameters and guests are also rated based up their behavior. This review system allows users' to decide place based upon their budget and requirements.

Reviews plays an important role in capturing people's sentiments and opinions about a wide range of products and services. Reviews provides a platform which allows people to share or express their unbiased opinions. Attention to the opinions and feedback which guests provide about the products and services via reviews is a critical factor to the success of the hosts in the market place.

I have collected reviews from guests. I have used these reviews to study the sentiment of the guests regarding their stay. The main objective of this analysis is to help AirBnB hosts to improve the quality of

their listings to meet the expectations of the guest and to generate maximum revenue. This paper also focuses on understanding what guests liked and disliked which may be one of the pioneer contributors to the revenue generated by the company and also the host.

## TIMELINE:

Timeline	Action Items	Timeframe
	<b>Text Analysis and Sentimental Analysis</b>	<b>May-June 18</b>
a.	Data preparation and Cleaning	4 <sup>th</sup> May – 7 <sup>th</sup> May
b.	Descriptive Analysis	21 <sup>st</sup> May – 30 <sup>th</sup> May
c.	Text analysis to find out most frequent terms, concept links etc.	2 <sup>nd</sup> Jun – 20 <sup>th</sup> June
d.	Sentiment Analysis	21 <sup>st</sup> June – 30 <sup>th</sup> June

**Figure 1. Timeline of the project**

## DATA COLLECTION, CLEANING AND CONSOLIDATION:

I scraped data from Airbnb website on December 3rd 2016 for New York City and our dataset has all the listings reported as of that day on the Airbnb site. I captured around 40,228 listings and following information related to them.

- Information related to listing description given by host in text format such as location, house rules and transit and also latitude and longitude.
- Information related to location, response time and response rate of the host.
- Information about property type, room type, bathrooms, bedrooms, amenities, minimum nights, maximum nights, allowed number of persons, cancellation policy and calendar availability.
- Information about daily price, weekly price, monthly price, cleaning fee, extra person fee, security deposit.
- Information about review ratings on cleanliness, location, communication, check-in and all reviews of a listing.

## Metadata of the data:

Variable Name	Type	Format	Length	Description
Listing id	Numeric	BEST4.	4	Unique id given to every host
Date	Date	DATE9.	8	Date on which review was given
ID	Numeric	BEST6.	6	Unique id given to every review
Reviewer Name	Character	CHAR30.	30	Name of the reviewer

Review	Character	CHAR300.	300	Review given by the guest
--------	-----------	----------	-----	---------------------------

Figure 2. Metadata of the data

## DATA PROCESSING USING SAS ENTERPRISE MINER:

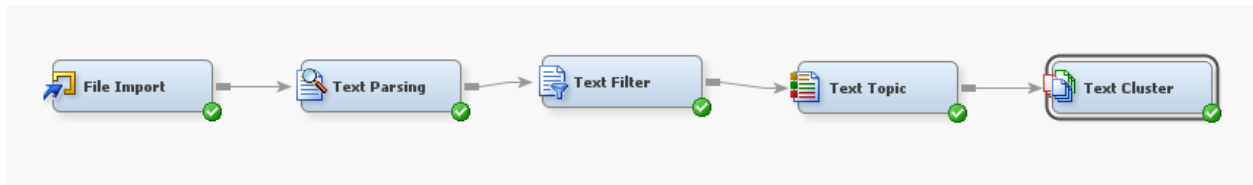
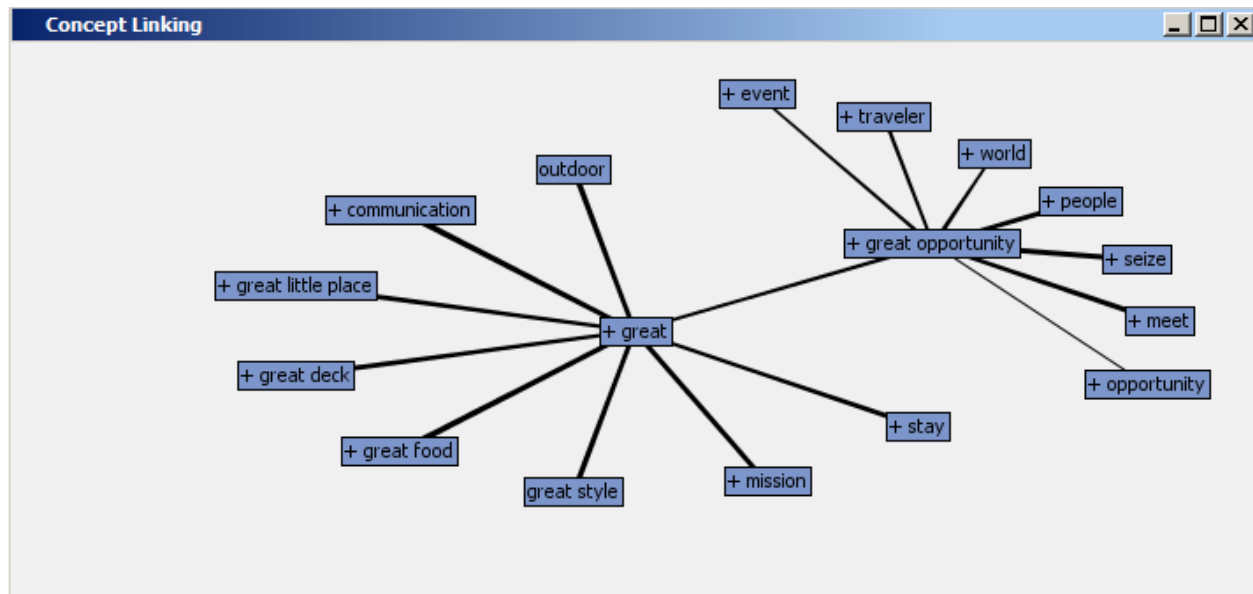


Figure 3. ER Diagram of the project

This is the methodology used for importing the data, parsing the data using online updated dictionary. Then data is filtered, and customized text clustering and text topic building is done. The modelling approach followed for the project is SEMMA (Sample, Explore, Modify, Model, Access). The data was partitioned into training & validation dataset. The training data is used to build the model. Validation data is used to test the accuracy of the model.

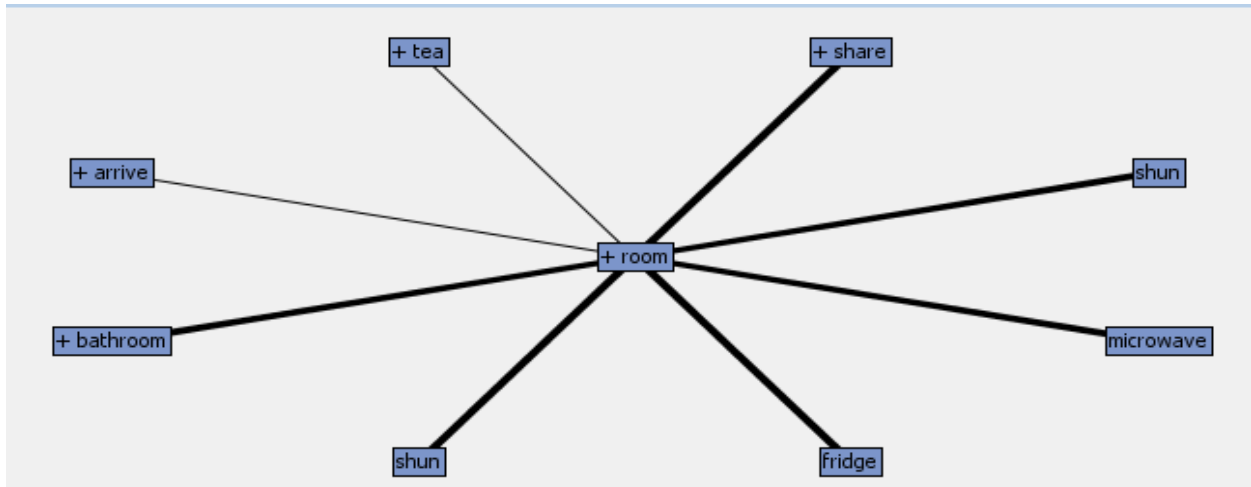
For the analysis, I have considered all the reviews and followed the general text analytic approach suggested by Chakraborty, Pagolu and Garla (2013). This involved using concepts like NLP techniques, lemmatization, concept linking, use of synonyms, etc.

Concept links help in understanding the relationship between words (terms) based on the co-occurrence of words (terms) in the document. It shows the terms that are strongly associated with the given term. Thick links indicate strong association between the terms.



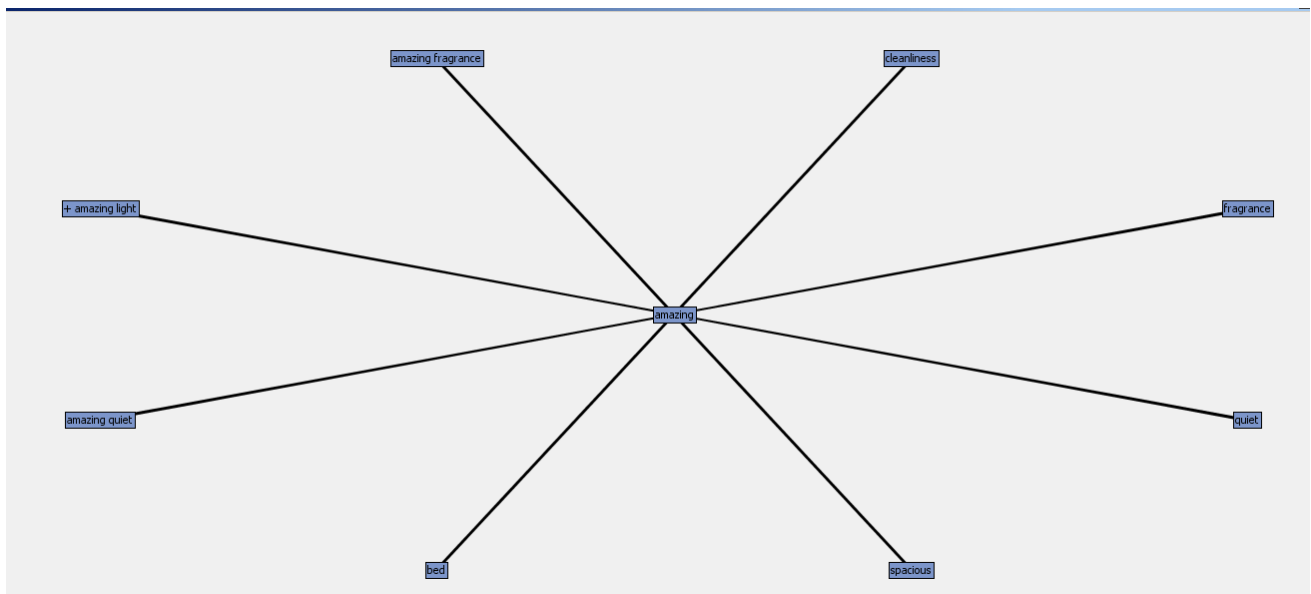
**Figure 4. Concept link of the word great**

The concept linking for the word great shows that people consider communication to be a very important factor during their stay at Air BnB. It will behoove a host to be a good communicator to get good reviews from guests.



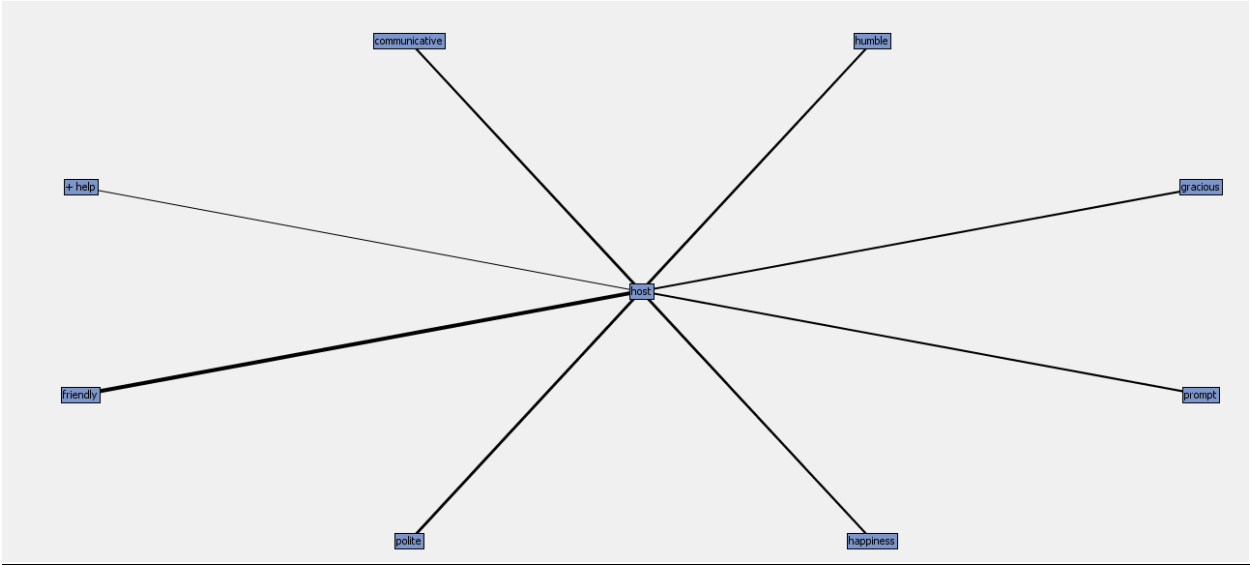
**Figure 5. Concept link of the word room**

From Figure 5, we can say that most people are interested in the specifications of the room like bathroom, microwave, fridge, tea etc. Most of the people really like a room if it has especially a bathroom.



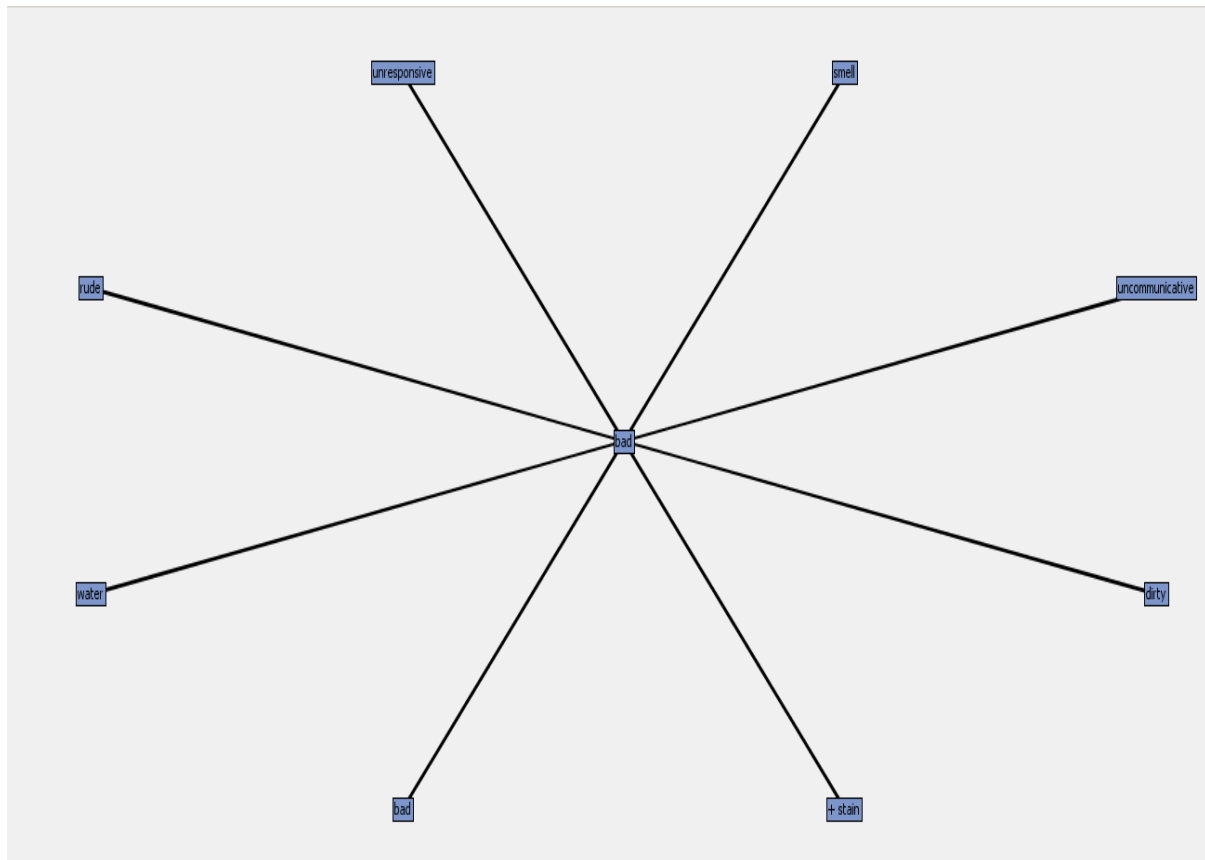
**Figure 6. Concept link of the word Amazing**

From Figure 6, Concept link diagram shows that guests are really amazed by properties or things like Cleanliness, Fragrance, if the place is really quiet, if the room is large enough and also by the quality of bed and mattress.



**Figure 7. Concept link of the word host**

From Figure 7, Concept link diagram shows that guests seeks the qualities such as friendly, polite , prompt, communicative, helping , humble and gracious in a host. For a host to increase his revenue he/she must possess these qualities to get good reviews from a guest which in turn will attract more guests.



**Figure 8. Concept link of the word bad**

From Figure 7, Concept link diagram shows the qualities that are considered bad by a guest. A host should never be rude to a guest, unresponsive and uncommunicative. A host should try to keep his listing free from smell, stains and also dirty.

## **CONCLUSIONS AND FUTURE WORK:**

To conclude this paper, I have performed text mining of AirBnB reviews about the host regarding their stay. I have used SAS Enterprise miner to clean and analyze the data. I have used the concept links to understand relationship between terms used in the reviews. This paper sets a stage to analyze the reviews that shows in general & people sentiments towards the host that will give us an idea of the success of the host. This paper not only deals with the sentiments of the people but give us a better understanding do's and don'ts by a host that will lead to better review. This paper was started with an idea to identify the people reaction towards unwanted events during their stay. This analysis can be extended to build a text predictive model wherein there is a scope of predicting the sentiments towards do's and don'ts and that will help us to set the stage better and prepared.

## **REFERENCES:**

1. <http://support.sas.com/publishing/pubcat/chaps/65646.pdf> - Text Mining and Analysis – Practical Methods, Examples & Case Studies using SAS@
2. <http://insideairbnb.com/get-the-data.html>
3. <https://support.sas.com/edu/schedules.html?id=2889&ctry=US&locationId=sf>
4. <https://en.wikipedia.org/wiki/Airbnb>

## **ACKNOWLEDGMENTS:**

I sincerely thank Dr. Goutam Chakraborty for his valuable guidance and motivation for accomplishing this paper. I also thank Dr. Miriam McGaugh for her constant support and suggestions throughout this study.

## **CONTACT INFORMATION :**

Prashant Kumar Gour  
Master's in Business Analytics  
Oklahoma State University, Stillwater  
405-385-3377  
[Prashant.gour@okstate.edu](mailto:Prashant.gour@okstate.edu)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.