Thesis

# Introduction

#### Sharing Economy

* The sharing economy has become a popular alternative to traditional services across a range of sectors
* What is the sharing economy? – The sharing economy has been described as economic opportunities, where instead of owning goods/services, participants would share them with one another
* Transforming the way in which we interact with one another. Examples:
  + ‘Sharing’ a meal with strangers
  + ‘Sharing’ a car ride
  + ‘Sharing’ your apartment
* The nature of this platform has fast tracked the number of personal interactions which we have with “strangers”
* This platform has become quite a controversial topic – especially with platforms like Uber, Airbnb
  + Use of profile pictures, and personal information has been proven to foster certain biases
  + Competition with existing industries
  + Biases based on names and countries of origin
* It has been shown that similarity between host and guests have been found to have a positive relationship with trust.
* For the purposes of this paper, however, we focus on whether user similarity impacts the satisfaction of the interaction
* i.e. are people who are similar more likely to have favourable interactions

**Airbnb**

In this study I focus on AirBnb as the sharing economy platform.

* Airbnb allows property owners to rent a part of/whole property to people looking for a place to stay. The accommodation types advertised include
  + shared rooms: guests share a room with either the host, or other parties
  + private rooms: guests have the room to themselves, however, they share the rest of the apartment/house with other parties
  + entire apartments: Guests have the entire apartment to themselves.
* Its initial selling point was that in addition to offering accommodation, users had the opportunity to social interact with each other. There have been many studies challenging whether that is the main motivation for participants using it.
* We distinguish between full (entire property) vs shared accommodation (shared and private rooms), as the interactions between these usually differ, as hosts renting full accommodations usually have a less interaction with their guests.

**Aim and contributions of this Study**

* A major part of this study consists of defining various types of similarity that may impact satisfaction between hosts and guests.
* For the purposes of this study we focus on engaged hosts in Manhattan, and their guests’ experience
* Manhattan
  + most densely populated neighbourhood within New York city.
  + Cultural and financial capital of the world
  + High level of income inequality
  + Sizeable tourism industry
* Overall, we make X main contributions:
  + Define similarity measures: define ways in which similarity can be measured based on user profiles and their interactions on the Airbnb platform
  + Analyse the relationship between similarity and satisfaction
    - Do people who are demographically similar result in greater satisfaction with their stays?
    - Do similar personalities result in higher satisfaction of stays?

**Motivation/Importance**

* A major concern within the sharing community is the potential for conflict between service users and providers: by being able to assess what drives satisfaction, we might be able to contribute to users having more highly positive experiences.
* Possibility to guide future larger scale studies on the relationship between similarity and satisfaction on the Airbnb platform.
* Airbnb has direct access to data, hence they could explore recommending listings based on the similarity relationship between host and guest.

**Report Structure**

In Chapter 2 I focus on giving the reader a relevant background of previous research done within the sharing economy and previous studies on similarity, satisfaction and linking them. Chapter 3 provides a descriptive analysis of our data and discusses the various restrictions that we made on the data. In Chapter 4, I delve into the specifics of how I measured the various similarity and satisfaction metrics, as well as, exploring their relationship. Following that, I discuss the results and analysis, in Chapter 5 and Chapter 6, respectively . Finally, in Chapter 7, I conclude the study by reviewing the various limitations, possible future work and summarizing the dissertation in its entirety.

# Related Works

Jundall and Liu 2006

Liu 2010

## Sharing Economy – what studies have been done on it

## Airbnb – what are the major topics on Airbnb – similarity + bias

## Methods of assessing similarity

## Methods of assessing satisfaction and sentiment

## Relationship between similarity and satisfaction - humans (high level, self report) + using Data only

# Dataset

## Data Collection

**[Diagram of where the data came from ]**

#### Inside Airbnb

Majority of the data was obtained from Inside Airbnb, which contains snapshots of city level data. We used it to collect:

* **Listing information**: This contained information about the listing such as accommodation type, as well as where it was located, and information about the host.
* **Host Information:** Information about the host was extracted from the Listing data. This included the host’s description of themselves, how long they have been a member for and how many listings they had.
* **Guest-to-Host Reviews**: We used the guest’s review as our main source of data in determining the satisfaction of a stay

#### Airbnb

While majority of our data came from InsideAirbnb, it did not contain all the information for us to complete our study (e.g. the profile of the guests). I made use of an Airbnb crawler in order to gain additional information. I worked with the developers in order to expand the crawler’s capabilities for additional information required. The original tool was programmed in java using Jsoup library which is designed to parse, extract, and manipulate data stored in HTML documents.

The additional information that we gained by using the crawler included the following:

* **Guest Profiles:** This allowed us to gain information about the guests’ description of themselves, where they were from, and how long they have been a user for. In order to obtain this information we made use of existing code for scraping this data, using reviews as input.
* **Images:** As Airbnb does not provide demographic information about its users we made use of online API’s that could derive this information, using the images as input. Thus , we had to crawl Airbnb in order to obtain these images.
* **Host’s Host Locations:** I built on additional functionality in order to crawl the hosts’ profile pages in order to retrieve location’s of their host’s reviews (i.e. when they were a guest). This was used in order to give us an idea of whether the host has made use of the platform as a guest, and where they travelled to. This was limited to reviews where the reviewer had allowed their location to be seen.
* **Host-to-Guest reviews** - > Additional functionality was built in order to retrieve reviews that host’s have left for the guest.

In **Table X**, you can observe the total number of records obtained initially.

Table of Final Numbers that were collected of each (independently): Number of Reviews, listings, reviews, host trips, host-guestreviews, guesthostreviews, host images, guest images,

#### Country Data

Information regarding such as Region, subregions of countries was fetched from <https://restcountries.eu/rest/v2/all>. A total of 13 fields was retrieved for all countries in the world.

#### Cultural Data

In order to explore cultural similarity between host and guest locations, country Individualism, and Power indices were obtained from [mindmap] study.

[include further explanation]

#### Airbnb Dictionary

In order to explore user motivations, the Airbnb dictionary constructed in [AirbnbDic] study was obtained.

[include further explanation]

## Data Cleansing

Before pushing all the data into a SQL database, the data was cleansed.

**Guest Profiles**

* Trimming white spaces.
* Removing duplicate records.
* Removing erroneous records (IDs with non-numerical data).
* Translations of data - as the crawler used -> crawled from the Italian version of the Airbnb site, some of the information was in Italian. I translated using google translate to translate the Verifications, Months, and country Data to English and stored them to a json file, in order to be used in the future.
* Used the country data in order to extract country information from the guest profiles – convert country to an Alpha2code.

**Host Profile**

* Host Profile was extracted from the listing Data.

**Reviews**

* The Host table was used to attach the host ID to the guest review.
* White Spaces trimmed.
* Line Breaks were removed and replaced with white spaces in order to make text processing easier later.
* Some reviews were cancellation notices. We placed a cancellation marker on all reviews that started with:
  + 'The host cancelled’
  + 'The host canceled'
  + 'The reservation was cancelled'
  + 'The reservation was canceled'

**Host Trips**

* The country table was used again in order to convert the information we had obtained about the Host’s previous trips: extracting information about the city, country and state (if from America) of previous trips.

**Host-to-guest Reviews**

* Trimmed white spaces, and line breaks, and removed records with erroneous IDs.

Once all data was cleansed we pushed this to a local sql database.

## Restricting the Dataset

Due to certain limitations such as pace of data crawling and cost of certain online tools, we decided to reduce our study by making certain restrictions.

### Restricting Data to Manhattan

We restricted our dataset to listings only within the Manhattan district, hence we only had to crawl the data for hosts of these listings and guests that have visited (and reviewed these properties).

### Restricting Data to only hosts with single listings

In order to reduce the possibility of confounding variable of hosts which are mainly business focused, we limited the hosts to only those with one listing. When hosts have multiple listings it makes it more likely that they may not be overseeing the transaction themselves and so it would be less likely to extract personality data.

### Restricting the data English

For the purposes of this study we wanted to work with reviews that were detected as English reviews, as we wanted to work with reviews in their natural state (at least mostly). In order to limit the reviews to those only in English, we made use of two tools, namely, langdetect, and google-translate.

Language prediction was done using both tools, we then checked for agreement between them, and the various errors they made.

Langdetect

* Bad at predicting short text

Google-translate

* Erroneous output when emoticons were present within the review.

When the above instances occurred, we made use of the alternative tool’s prediction (given that its confidence was above 90%).

[include number of languages detected - when both tools were in agreement]

### Restricting by the number of reviews that host had - gain an average

In order to only consider hosts, where guests were engaged, we further retricted the dataset to hosts with 5 and more reviews, in order to be able to make comparisons with the guests average satisfaction at that particular listing.

Methodology

# Defining Similarity

[introduction sentence]

Similarity between people can manifest itself in many various ways. In this study we focus on:

1. Demographics: age, gender, ethnicity, and Country
2. Personality – using linguistic cues, host traveller metric, emotion from images, and user’s motivation and interests

Ways of measuring similarity:

* closeness through some distance metric

Demographics

One of the most common ways to measure user similarity, is through demographics.

[include more background]

Focus: Do similarities or differences between guests and hosts in terms of demographics impact the satisfaction?

This study focused on the following demographics:

#### Age

We wanted to assess whether hosts and guests which are similar in age tend to have more satisfying stays. If there was an age gap, did it matter whether the host was older or younger than the guest.

In order to assess the similarity, we tested:

* Absolute age difference
* Directional age difference
* Age quadrants – divided age groups into quadrants

#### Gender

H: Host’s and guests that are of the same gender have more favourable stays.

For this we tested:

* 4 classes – HG - MM , FF, MF , FM

#### Race

A couple studies have studied that hosts and guests have mostly gravitated towards those of similar ethnic backgrounds on sharing economy sites. We wanted to investigate whether a guest was more satisfied with their stay if there was a racial similarity with the host.

In order to calculate the racial difference, we took the output from Sightcorp API and converted it to a vector. Using these vectors we computed the Euclidean distance between the host and guest.

\*\* Caution on results: based on skin colour and facial features only.

### Demographic extraction

* (1) Tool localisation: In order to (a) get an idea of how many people were in the picture, and (b) get a bounding box (if required) we used Indico.io. This tool has been noted to provide results with an accuracy of X% [offlinebiasstudy ].
* (2) An assessment was done of 3 possible tools. We assessed a set of 50 manually selected images (in order to provide a **variety of images**). The three tools assed were BetaFace , Cognitive Face and Sightcorp F.A.C.E. We decided to use Cognitive Face for age and gender, and Sightcorp for Ethnicity. The reason for not choosing BetaFace, was that it gave results similar to Sightcorp, however, it came at a much higher price. Cognitive Api performed better than both of these tools on Age.
* (3) Once we had all our images and had decided on our tools, we then proceeded to processing only host and guest images where only 1 person was detected. We decided to do this so that we could focus on 1-on-1 relationships rather than comparing similarity between users where we may be dealing with couples/groups.
* [show the proportions of how many people were detected in the pictures for hosts and guests].

#### Are they living in similar places?

We then wanted to assess whether the host and guests living in similar places had an impact on satisfaction.

In order to do this we extracted the country of the host and guest from their profiles and made use of the country information that we obtained using [country table x].

Initially, we compared the distance between representative longitude and latitude of the respective countries. We then also computed an indicator of whether they were from the same country or not (1/0). We also compared whether they were from the same regions or not, which can be seen in [table X].

Following this, we wanted to assess whether hosts and guests coming from countries with similar culture indices result in higher satisfaction. For this part of the study we used the values obtained by [culture study] on individualism index and power distance.

Definition of Individualism: The extent to which individuals are integrated into groups.

Definition of Power distance. Extent to which a country expects unequal power distributions

In order to compute a similarity metric for these cultural differences we experimented with taking an absolute difference versus a relative difference.

## Personality

It has been found that people respond to stimuli based on the situation they are in, as well as their underlying personality traits.

In the nature of booking a place online, the main method that users can assess someone’s personality is through their profile (description, motivation for using Airbnb, their past experiences, and their profile photo). We aimed to systematically assess these elements and determine how personality similarity affects satisfaction.

In textual personality, we can extract behavioural cues from their writing. For example, it may be that their writing about X topic can help them infer something about the personality of the author. In many studies about personality the big 5 model is used as the grounding research.

Previous studies have also shown that people form opinions based on profile pictures before they have even met someone.

**Big 5**

O(Openness):

C (Conscientiousness)

E (Extraversion):

A (Agreeableness)

N (Neurotism or emotional stability)

What method did we choose – mixture of linguistic cues & elements from profile, (besides using features previously discussed (important for similarity)

## Motivations

Previous studies have focused on whether a user is either primarily motivated by business or social interactions. [SocialStudy] generated a dictionary based on what users talk about in their reviews. This dictionary contains 4 main themes:

1. Property
2. Location
3. Professional conduct
4. Social Interaction

The first two are more focused on the listing alone, whereas the last two speak about the interactions they had with people.

[Talk about Dictionary construction]

We analysed the reviews of hosts and guests in order to explore how often they spoke about each theme. From this we could then assess how business minded/socially minded they are. We used this in order to compare the host and guest’s motivations

## How do they describe themselves?

Users of the Airbnb platform have an area in which they are able to tell other users about themselves, where they come from ,and what their interests are. Using this information, we extracted topics that hosts and guests talked about.

[Talk about empath tool ]

Using the Empath tool, which has a set of X categories, we formed an occurrence vector for both host and guest. Using these vectors we could compute a cosine similarity between the two profiles based on the content of their description entries.

It should, however, be noted that only X% of hosts had non-empty profiles and X% of guests. Merging the two, only 30% of host-guest interactions had both host and guest with non-empty profiles. From the non-empty profiles the distribution of profile length was positively skewed for both host and guest.

## Language Cues

There have been multiple studies using linguistic cues in order to predict personalities.

[Table Explaining the Cues and how they relate to the different big 5 metrics]

#### Formality

In Heylighen and Dewaele (2002) ‘s study on personality they propose a metric, which they have termed Formality, which has shown to correlate positively with introversion, education level and the femininity of a speaker

|  |  |
| --- | --- |
| Trait | Characteristics |
| Extraversion | High Average WC\*, High positive emotion words, high use of first person singular pronouns, High number of verbs,adverbs,pronouns, low number of negations, low words per sentence, low number of articles, low diversity of words, low diversity of POS, low formality, low number of negative words |
| Openness to experience | High number of unique words, high number of longer words (>6 letters), high tentative words, low first person singular pronouns, low number of present tense words, low number of articles |
| Neurotism | High number of qst person singular pronouns, high number of negative emotions, low number of positive words, low number of articles, high number of swear words |
| Agreeableness | High number of use of “to”, high number of really positive words, low number of negative words, High number of first person , high number of articles |
| Conscientiousness | Low number of article words, low number of negative emotion words, low number of tentative words, low number of unique words, high number of positive words |

* Average words count – averaged over each piece of text from user i.e. word count was calculated for each review and profile separately then averaged for each user.

Decided to calculate values for all users based on text from their profiles and reviews (that they have given). From these values we converted these values into categories -> high, average, low. Furthermore, for each trait we calculated how many of the characteristics that user had portrayed given the text. This resulted in a measure for each trait.

As many profiles were empty, we decided to concatenated user profiles with reviews they have given (and we have retrieved).

[TODO : sample study using online api]

[TODO May want to exclude current recipient review out ]

### Exploration metric

As the primary function of Airbnb is focused on offering accommodation to travellers, we wanted to investigate whether the hosts were travellers themselves. We did this by assessing:

* Whether the host has used Airbnb as a guest before. This could tell us whether they could possibly know what makes a guests stay positive.
* This is important ,because Airbnb is all about living in someone else’s place, respecting their culture and space. Simply being a host may not expose you to some of the troubles that may come with being a guest. Things that you may not notice. Experience of being a guest gives you the perspective to look inwardly and ask yourself what would I want if I was travelling.
* How many times has the host used Airbnb as guest, and where they have gone. Do they normally go to the same area or do they explore different places?
* If a host uses Airbnb a lot, but always goes to the same place, it could mean that they are using it for business instead.
* Does exposure to other countries make for a better stay, or experience as guest
* Focusing then on who they allow into their home – do they have a variation of guests based on age, gender, race, country, language? We use a sample of hosts with multiple guests and explore
  + Age – mean and variance of guests age, higher variance means that they are open to many ages
  + Gender – what is the ratio between male and female guests. We use the difference between these proportions to assess whether there is a big gap between the proportions of male vs female.
  + Country - How many different countries have their guests come from.
  + What are the languages spoken by their guests - using initial dataset?
* Is this different for full vs shared
* We then look further into whether the host has travelled to the guest’s country
* What percentage of the host have travelled (have reviews where guests have allowed location to be seen) – be cautious.
* Hosts that have been Hosted in Manhattan (160)

### Profile Pictures

There are multiple different studies that state that personality can be inferred by a picture.

From the demographic information we obtained from the Demographic Extraction tools, we were able to retrieve 2 different metrics that may give us further insight into the personality of a person .

1. Smile index – the degree to which a person is smiling in the picture
2. An emotion vector – the proportions of emotions experience in a single image.

We use the absolute difference between the smile index of the host and guest in order to determine how similar their expressions were.

We take the Euclidean distance between the host and guest’s emotion vector in order to obtain an emotion similarity distance. The emotions detected were Happiness, Sadness, fear, anger, surprise, and disgust.

[Talk more about the tools]

### What do other people say about you?

Using Host’s review of guests, as well as the Airbnb dictionary, we were able to extract sentences involving professional conduct and social interactions with the Guest/Host. From these sentences we used NLTK tool to extract all the adjectives. Once this was completed for the whole review database, we formed a vocabulary set from these adjectives and derived a vector for each host and guest.

Finally, we computed the cosine similarity and TFIDF metric to compute whether people say the same things about the user.

## Defining Satisfaction

**What is satisfaction?**

[definitition from paper] In this study we refer to satisfaction based on the reviews written by host/guest.

There are various ways in which we could look at satisfaction:

1. One Way: By only looking at the satisfaction based on the guest’s review.
2. Alternatively, we could also incorporate the Host’s review of the guest, and provide an average of their reviews.

For the purposes of this study we only considered the former. [May still include two way]

**Reviews transformation**

* Translating to English. How many reviews did you have to translate? **(Test)**
* Spelling correction **(test)**
  + Get dictionary of words
  + Analysed words that were unknown in the Word2Vec model.
  + Viewed the counts of these unknown words within the reviews
  + Only corrected those that were popular mistakes (count > 5)
  + Spelling correction using spelling corrector tool (expand on)
  + Derived a dictionary of corrections

**How do we compute it?**

* Normally satisfaction is measured by using ratings. Airbnb, however, does not provide the individual ratings for each review. We decided to use sentiment analysis tools in order to assess how satisfied the reviewer was based on their text/review.
* As we are not provided the guest rating for each review, we require some form of an estimate of it. I did an initial assessment of some of the possible tools that could be used to predict their sentiment.

**Tools Assessed**

* [iFeel paper] provides a benchmark comparison on a wide range of sentiment analysis tools on a variety of datasets. In this study they notice that no single tool has the best prediction performance across all datasets. i.e. some tools perform better on certain datasets. Polarity of short text (sentence level). From the iFeel Paper, I looked at the tools for Sentiment analysis on comments (3-point scale) and reviews (only 2-point scale available):
* I proceeded with the tools that were placed in the top 10 of both of these (AFINN, VADER, SO\_CAL, OPINION LEXICON, SEMANTRIA. However, I replaced Semantria (commercial product) with the top performing tool in reviews.

**[Table with Definitions, coverage, min, max, mean, std dev, distribution]**

AFINN - Builds a Twitter based sentiment Lexicon including Internet slangs and obscene words. AFINN can be considered as an expansion of ANEW [30], a dictionary created to provides emotional ratings for English words. ANEW dictionary rates words in terms of pleasure, arousal and dominance.

**(+- 70% Coverage, also takes emoticons into consideration)**

*Works directly with module*

VADER - It is a human-validated sentiment analysis method developed for Twitter and social media contexts. VADER was created from a generalizable, valence-based, human-curated gold standard sentiment lexicon.

**(Top performing on 3 point scale for comments, +- 80% Coverage)**

*Works with directly with module*

OPINION LEXICON - Focus on Product Reviews. Builds a Lexicon to predict polarity of product features phrases that are summarized to provide an overall score to that product feature.

**(+- 74% coverage)**

*Given a list of positive and negative words.*

*I take my text , tokenize it , remove stopwords, and then did two tallys:*

* *one using just the occurrence of +,- words, divide each occurrence count by the number of distinct words*
* *and another using the frequency,, divide each occurrence count by the number of words*

*Final Score = Positive score – Negative Score*

***IDEA:*** *use wordnet to add more similar words to each list by looking at top 3/5 similar words to each word.*

SENTIMENT-140 - A lexicon dictionary based on the same dataset used to train the Sentiment140 Method. The lexicon was built in a similar way to [33] but authors used the occurrence of emoticons to classify the tweet as positive or negative. Then, the n-gram score was calculated based on the frequency of occurrence in each class of tweets.

**(Top performing on reviews, +- 27% coverage)**

*Save reviews in text files, wrote a script that then feeds it to the online api, which returns scores*

SO\_CAL - Creates a new Lexicon with unigrams (verbs, adverbs, nouns and adjectives) and multi-grams (phrasal verbs and intensifiers) hand ranked with scale +5 (strongly positive) to –5 (strongly negative). Authors also included part of speech processing, negation and intensifiers.

**(+-83% coverage)**

*Save reviews as text files, Run a script which does preprocessing on text using StanfordCore-Nlp , Run another script which aggregates all of the scores and outputs a csv file with relevant scores.*

In order to assess these tools, I computed the score for each review and tool on the whole review. I then coded them into 5-point scale. Using this 5-point scale, I computed the majority vote for each tool for each review. Using this, I then aimed to see which tool predicted the majority classification most often.

Was sentence level better than on whole review?

I computed this on a sentence level too, where I split into sentences , ignore all average scores and computed the mean of the remaining valences. Finally comparing the sentence level scores with whole review scores.

[Graph showing comparisons]

Then Explored with combinations:

* + Average of all tools
  + Strategic average
  + Majority vote - on interrater

[Rank of tools that guessed correctly most of the time for each.]

#### Interrater comparison

For the comparison, I randomly selected 52 reviews with lengths in the upper quartile range (>77 tokens/words). In order to label these reviews, I created a survey, where humans could rate each of these reviews depending on what they thought the guest rated it (perceived).

I managed to get 4-5 people to rate this set and took the weighted average for each review.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Extremely Unsatisfied  (=<0.2) | Quite Unsatisfied  (>0.2 and <=0.4) | Average  (>0.4 and <=0.6) | Quite Satisfied  (>0.6 and <=0.8) | Extremely Satisfied  (0.8< ) |
| Count | 2 | 8 | 7 | 11 | 24 |

Raters were able to pick up on mixed reviews, and were mostly in agreement (same score / one point to either side).

[Alternative metric - Getting average scores for each host and getting a relative satisfaction]

**[Error Analysis]**

[insert results]

## Linking Similarity to Satisfaction

Having defined similarity and satisfaction in the previous two sections, we can now discuss how we will assess their relationship

#### Data Exploration

* Scatterplots
* Comparing means and averages within categorical comparisons

#### Collinearity

In order to reduce the features that we input into our model, we use a correlation coefficient matrix, and VIF metric to see which features are highly corelated with each other.

[VIF Metric and explanation]

#### Regression

We input the remaining features into a linear regression model in order to assess their relationship with satisfaction.

# Results & Analysis

## Demographic Similarity and Satisfaction

#### Results from exploraton

#### Results from linear regression (Only demographic)

## Personality Similarity and Satisfaction

#### Results from exploration

#### Results from linear regression (Only personality)

## Similarity and satisfaction

#### Collinearity + dimensionality reduction

#### Linear regression

#### Non-linear regression

# Conclusion

## Limitations

* + Data Crawling
  + Tool limitation
  + Limited to manhattan
  + Limited to English reviews
  + Satisfaction estimate – more accurate if we had actual satisfaction scores

## Future work

## Summary of study