Capstone Project Report

Airbnb Selection: Social Media Analysis

by

Sin Yee Ran
(17027186)

482 Information System (Business Analytics)

:

Supervisor : Prof. Dr Teh Phoey Lee

Date : 13 November, 2020

Project title : Airbnb Selection: Social Media Analysis
Date : 13 November 2020
Student : Sin Yee Ran
Supervisor : Prof Dr Teh Phoey Lee

Abstract

There is no doubt that the rapid growth of Airbnb has changed the lodging industry and tourists' behaviours dramatically since the advent of the sharing economy. Since its first appearance in 2008, Airbnb has had a fast growth that raises questions about its current and future impacts on the traditional accommodation sector. The sharing platform enables people to rent out their private living space for those who requiring short-term accommodation. In 2017, international tourist arrivals grew by a remarkable 7 percent, recording a total of 1.3 billion. This is a great chance for Airbnb to understand their demand and attract more customers.

The main purpose of this study is to examine the motivations of Asians and Westerners in choosing their accommodation form Airbnb when they are travelling domestic or abroad. Text data were obtained from Twitter because Twitter is one of the most popular microblogging sites. Tweet texts were scrapped from Asian and Western countries include UK, USA, Canada, Mexico, Hong Kong, Malaysia, Indonesia, Japan, Korea, China, and so on. In all, there are 21 countries where 15 are Asian countries and 6 are Western countries. In total, there are 5507 Tweet text were scrapped. The Tweet texts are visualized using word frequency, word association to investigate the motivations of Asians and Westerners which lay behind the Tweet text. In addition, 4 motivations are – price, host, amenities and location are chosen to investigate the emotion of Asians and Westerners towards these aspects. sentiment analysis and emotion classification are implemented as the performance evaluation by Airbnb customers.

The conclusion of this study offers implications and suggestion for future studies. It also provides useful insights on Airbnb customers favourable attributes that drive them to select accommodation. The study is useful in assisting Airbnb host in designing strategies and make enhancement to satisfy their customers.

Keywords: Airbnb, Tweet text, Motivations, Text Mining, Sentiment Analysis

Table of Contents

List of Figures

List of Tables

List of Abbreviations

B2C	Business-to-Consumer
C2C	Consumer-to -Consumer
P2P	Peer-to-Peer
UK	United Kingdom
USA	United States of America

1 Introduction

In today's fast flowing age of advancement, technologies have enabled several service industry businesses to develop innovative ways to reach potential customers and expand their customer base (Varma et al., 2016). The advent of the sharing economy has changed consumer behaviours dramatically in recent years. The popularized "sharing economy" term has been used frequently to describe different organizations that connect users or rentals with owners or providers through consumer-to-consumer (C2C) or business-to-consumer (B2C) platforms (Parente, Geleitla, & Rong, 2018). Due to the sharing economy marketplace, some businesses within tourism field developed fast by providing local services to tourists (Ert, Fleischer, & Magen, 2016). One of the on-going success stories has been Airbnb, which was initially created by founder Brain Chesky and Joe Gebbia in 2008 as a short-term solution to cover rent in San Francisco.

Airbnb used to be a little-know start-up has now served over 35 million guests worldwide (Raymond & Charlie, 2019). As a popular home-sharing platform, Airbnb is globally operating and offering different kind of host services to match their customers' preference. Airbnb offers short-term rental service to people from all around the world. The local hosts rent out their personal properties such as apartments, entire houses, bungalows, house boats and more to their guests. According to Güçlü, Roche, & Marimon (2019), Airbnb customers hosted in 33,000 cities across 192 countries, and posted millions of reviews about Airbnb quality and rating their services. There are so many data can be extracted publicly and privately to assess the main criteria of Airbnb customers in accommodation rentals.

Based on U.S. Travel Association (2015), travellers are relying on reviews and feedbacks from social media to make their itinerary. Social media is a unique platform for web application that focus on user participation, most online businesses have begun to set up social media account to attract customers. Social media is one of the outside forces in tourism industry that affecting how the travellers make their decisions on their lodging choice during their vacation (McCarthy, Stock, & Verma, 2010). There is vast amount of information about tourism can be found from social media like Facebook, Twitter, Instagram and so on for making decisions and improvements (Ćurlin, Jaković, & Jaković, 2019).

As one of the top ten most visited websites and apps, Twitter is considered as one of the most well-known microblogging sites all around the world (Antoniadis, Zafiropoulos, & Varna, 2015). Aslam (2020) stated that Twitter has an average of 220 million monthly active users in the first quarter of 2019. There are about 350,000 new tweets posted on Twitter every single minute per day. The total number of tweets sent is about 500 million and roughly 42% of Twitter users used this platform every day. Hence, Twitter is influential and useful for tourism industry to gather customers' opinion and feedback to establish competitive benchmarks.

Tourism has grown virtually uninterrupted over the past decades, becoming one of the largest and fast-growing economic sectors in the world. International tourist arrivals have grown form 25 million globally in 1950 to 669 million in 2000 and reached 1 billion in 2012. In 2017, international tourist arrivals grew by a remarkable 7 percent, recording a total of 1.3 billion. This strong momentum is expected to continue in 2018 at a rate of 4 to 5 percent. This is above the 3.8 percent average increase projected for the period 2010-2020 by UNWTO. By 2030, the world's total international tourist arrivals are expected to reach 1.8 billion. Since the rapid growth of international tourists, it is important to understand their demand for selecting accommodation.

There is limited research has investigated the difference of motivation to choose Airbnb between Asians and Westerners. There is a big gap between eh Asians' and Westerners' culture difference. Hence, this project aims to explore the difference of motivation that drives Asians and Westerners to choose Airbnb as their accommodation when they travel abroad or in domestic. The motivations chosen in this study are Hosts, Price, Location and Amenities. This study wants to explore the do these aspects have any effect on customers' lodging select intention. In addition, difference of motivation in Asians and Westerner will be identified. It will identify the most significant factors in choosing Airbnb through Tweet text. A better understanding of guests' motivations for using Airbnb can offer valuable marketing insights for Airbnb, the host, and competing accommodation firms. Only with a clear understanding of consumers' reason for choosing Airbnb can make informed decisions regarding how best to market toward Airbnb's users. So, the objectives of this project are:

1. Discover what Asians and Westerners care about Airbnb:

RQ1: What is the motivation that drives Asians to choose Airbnb?
RQ2: What is the motivation that drives Westerners to choose Airbnb?

2. Discover what is the attitude of Asians and Westerners towards different motivations:

RQ3: What is the attitude of Asians and Westerners towards price?
RQ4: What is the attitude of Asians and Westerners towards host?

RQ5: What is the attitude of Asians and Westerners towards amenities?

RQ6: What is the attitude of Asians and Westerners towards location?

2 Literature review

2.1 Airbnb

Believe between consumers and providers had successfully made sharing economy forms the basis of a successful resource-sharing transaction because of the believe between consumers and providers (Zhang, Yan, & Zhang, 2018). As explained by Guttentag (2015), sharing economy has brought a disruptive change. Airbnb is a successful example and large rise of sharing economy. Airbnb is a famous online marketplace for people to rent out their private spaces for low budget guests who are seeking for short-term P2P accommodation on Airbnb (Guttentag, 2015). Airbnb is an company that has become widely known around the world for its online home sharing site (www.airbnb.com). The Airbnb website is straightforward for people to book accommodation style they preferred. User can search their accommodation conveniently based on location, travel dates, number of guests. Besides that, they filter and refine the list of accommodations showed by attributes such as price, environment, and amenities. Then, users able to check for the details of the accommodation they choose by looking into the description, photographs provided by host or owners and reviews from previous guests (Guttentag, 2015).

According to Tam (2019), Airbnb started out in 2008 when founders Brian Chesky, and Joe Gebbia realised that they could rent out an air mattress in their living room for people visiting San Francisco to cover their rent, then idea evolved into a website. Airbnb started off with customer innovation by creating a new market of people who will rent rooms in other people's homes. Most people sleeping over at their destination would need a "home away from home". Their mission is to create a world where anyone can belong anywhere through magical travel that provides unique, authentic and local experiences (Airbnb, 2019).

In the Airbnb's early stage start-up, trust issues between host and guests is the main problem of this market. Most people were worried about the idea of letting someone else live inside their home either with them or alone. Many people refuse to invest because this is the normal fear that many people still have even with Airbnb. It was a big challenge for establishing trust between hosts and guests. It was the main challenge that Airbnb had to overcome. Airbnb then overcame these obstacles with Web 2.0 internet technologies for users generate the content published on websites

(Guttentag, 2015). Airbnb using a robust system of identity verification and reviews for both hosts and guests. Guttentag et al. (2018) also proved that Airbnb’s key mechanism for facilitating trust between hosts and guests is its public review system, where both parties can post reviews.

P2P accommodation Airbnb managed to create a community of people that strive to behave nice to each other. Thus, Airbnb customers are likely to build effective commitments when they consider Airbnb as reliable and trustworthy P2P accommodation (Guttentag et al., 2018). Trust is the main piece of Airbnb's business model which allowed it to grow so fast. Now, Airbnb offering about 7 million listings in over 100k cities and have average 2 million of guests staying on Airbnb per night in 2019 (Airbnb, 2020).

2.2 Reviews on Social Media

Many hospitality companies started their business on online analytical platforms from the beginning of the Internet in 1990. Moon et al. (2019) stated that different form traditional accommodation sector, Peer-to-peer (P2P) accommodation brands rely heavily on the engagement of users on social media in their marketing and branding strategies. Online reviews started becoming an influential information for travellers to make their decisions while planning their itinerary (Fang et al., 2016). Fang et al. (2016) also suggested that the text readability and reviewer characteristics influence the perceived value of reviews.

There are some researches show that most people like to evaluate products and services on online media platform. Based on Lee, Park and Han (2008), word-of-mouth very significant in impacting, forming customers’ attitudes and buying intentions. Many people believed that products with positive contents have good quality while negative reviews always associated with low-quality products (Lee, Park, & Han, 2008). They also stated that products or services will probably have high ratings if positive feedbacks were obtained from customers and vice versa. The negative feedback negative relationship in affecting customers’ buying intention. Cui, Niu, and Guo (2015) also agreed that negative feedback will directly reduce customers’ buying intentions. A higher perception of negative feedback corresponds to lower customer purchase attitude. Besides that, there is a relationship between customers satisfaction with product quality (Suchánek, Richter, & Králová, 2015).

Gretzel and Yoo (2008) believed that many travellers prefer to read online feedback and experience posted by other travellers rather than look through the travel advices written by travel service platforms. Social media platforms provide opportunities for tourists to express their feelings, share their opinions, reviews, experiences and exchanging ideas with other tourists. Hence, many travellers will choose to use those opinions and experiences as reference in their travel planning (Ráthonyi, 2013). A survey carried out by Gretzel and Yoo (2008) has proved that almost 98% of people plan their itinerary by referencing the online reviews and 84% were affected. The posts and reviews on social media really play an important role in affecting customers' choices on accommodation, destinations, restaurants they want to visit (Ráthonyi, 2013). In Elrhoul (2014) opinion, one in two Twitter users said that Twitter content can influence their consideration of a travel brand.

2.3 Twitter

Twitter has become one of the most influential microblogging websites that allows people to receive and post short post which named tweets since 2006 (Thelwall, Buckley, & Paltoglou, 2011). As a microblogging site, Permission is given to Twitter users to post messages within 140 words which named "tweets". Tweets able to convey the primary and important meaning that users want to express by excluding the irrelevant content (Philander & Zhong, 2016). The unique symbol '#' in Twitter was named hashtag which is used to highlight keywords that users think are important in their messages. "Tweets" are written from a variety of fixed locations and mobile computing platforms, thus offering insights into the day-to-day discourse of personal and geopolitical events which offer a representative look at individual customers' personal and unique lodging experiences (Tighe et al., 2015). Hence, many industries would like to analyse tweet text to gain information thus understand what their customer desire for. Luo (2018) said that semantically analysing the tweets is helpful in gaining insights because it is short but full of meanings.

According to Aslam (2020), There is an average of 220 million of Twitter active users monthly in the first quarter of 2019. Every day, there will be about 350,000 new tweets posted on Twitter in a minute and 500 million of tweets will be sent out. Moreover, there are roughly 42% of Twitter users are on the platform every day. Based on Erlhoul (2014), 51% of Twitter users stated that Twitter contents have impaction on their consideration of travel brand. Twitter is a valuable tool for travellers to seek for travel tips and advices before they visit a country.

The large user base of Twitter has attracted hospitality and tourism industry to use it for promotion, distribute marketing management, communication and making market research (Leung et al., 2013). As claimed by Aslam (2020), Twitter has gained about $885 million in 2019 for advertising revenue which is an increase of 12% compared to 2018. Besides that, the advertisement engagements were up to 29% every year. Erlhoul (2014) believed that Twitter is a great platform for brands to gain their awareness because 44% of customers in his study are more likely to learn about travel brands on Twitter rather than average social network. It also always used as a platform for tourism industry to have communication with their customers and build rapport.

2.4 Text Mining

Text mining is a new area of computer science which foster strong connections with natural language processing, data mining, machine learning, information extraction and knowledge management (Radovanovic & Ivanovic, 2008). As stated by Dang and Ahmad (2015), text mining is a process of extracting hidden insights and valuable information from unstructured textual data through identification and exploration of patterns. For example, the unstructured data like emails, online reviews, tweets and other types of written feedback contain insights which is important for businesses to gain understanding on their customers. Due to the rapid change of technologies, there are different information and data can be scrapped online. Large volume of text data can be extracted form online, so it is impossible to read through such a large volume of document collections and discover patterns that were unknown beforehand. Hence, text mine the bunch of data not only can save time but also can reduce the useless information and navigate important features.

The work by Sinoara, Antunes, and Rezende (2017) stated that text mining process can be divided into 5 steps which are problem identification, pre-processing, pattern extraction, post-processing and knowledge usage. The process first starts with problem identification. The application objectives and scope must be specified. Then the specifications will guide the next steps of text mining process which is pre-processing. The pre-processing step is about preparing data for discovering trend. In this step, data will be transformed into some data representation format which is used as input for knowledge extraction algorithms. In the pattern extraction step, suitable algorithm will be applied depend on the data and pattern expected to discover hidden trends and patterns. Then, post-processing step will evaluate the extracted knowledge. If the information meets project scope the knowledge usage step can be performed.

Most text mining techniques are based on bag of words method and application of data mining techniques (Radovonic & Ivanovic, 2008; Sinoara, Antunes, & Rezende, 2017). On the report of Rodovanic and Ivanovic bag of words representation treats each term as single token in sentence no matter type or order. Words are just attributes of the document.

2.5 Sentiment Analysis

Sentiment analysis which also named opinion mining is a term that refers to the use of natural language processing, text analysis and computational linguistics in order to ascertain the attitude of a speaker or writer toward a specific topic (Liu, 2012; Schuller, Mousa, & Vryniotis, 2015). Sentiment analysis is a task that implies extracting opinions, emotions and sentiments form data (Liu, 2012). Hence, it is widely used to track the emotions and feelings from reviews and feedbacks to measure performance. Other research conducted by Schuller, Mousa, and Vryniotis (2015) suggests sentiment analysis involves in tokenization of document text to extract important features and classified the document into categories. It will identify the polarity and classified opinions into positive, neutral and negative (Liu, 2012; Schuller, Mousa, & Vryniotis, 2015). According to Liu (2012), the rapid growth of social media had caused individuals or businesses increasingly use contents from these media such as feedbacks, reviews and post to make important decision.

Sentiment analysis usually classified into three approaches including Machine Learning, Hybrid-based and Lexicon approaches (Feldman, 2013; Alaei, Becken, & Stantic, 2017). On the authority of Feldman (2013), Lexicon approaches also named rule-based approach. In this project, we only focus on using Lexicon-based approach. Sentiment lexicon is a list of sentiment words and phrases. This approach utilized a sentiment lexicon to describe the polarity (positive, negative and neutral). Lexicon-based approach can further divide into dictionary-based approach and corpus-based approach (Quratulain, Sajjad, & Sayeed, 2016). In the dictionary approach, a small set of opinion words is collected manually as a seed. Quratulain, Sajjad, and Sayeed (2016) realized that there is a main downside of dictionary approach which is unable to find a domain and context specific opinion words. Corpus-based approach able to go beyond the limits of dictionary approach by identifying specific opinion words using linguistic constraints. After the text are tokenized, the words will be split into positive and negative regarding the list of words in the dictionary or corpus.

2.6 Proposed Motivation to Select Airbnb

Based on numerous previous studies, we realized that the 4 main selling point of Airbnb is: Price, Host, Amenities, and Location.

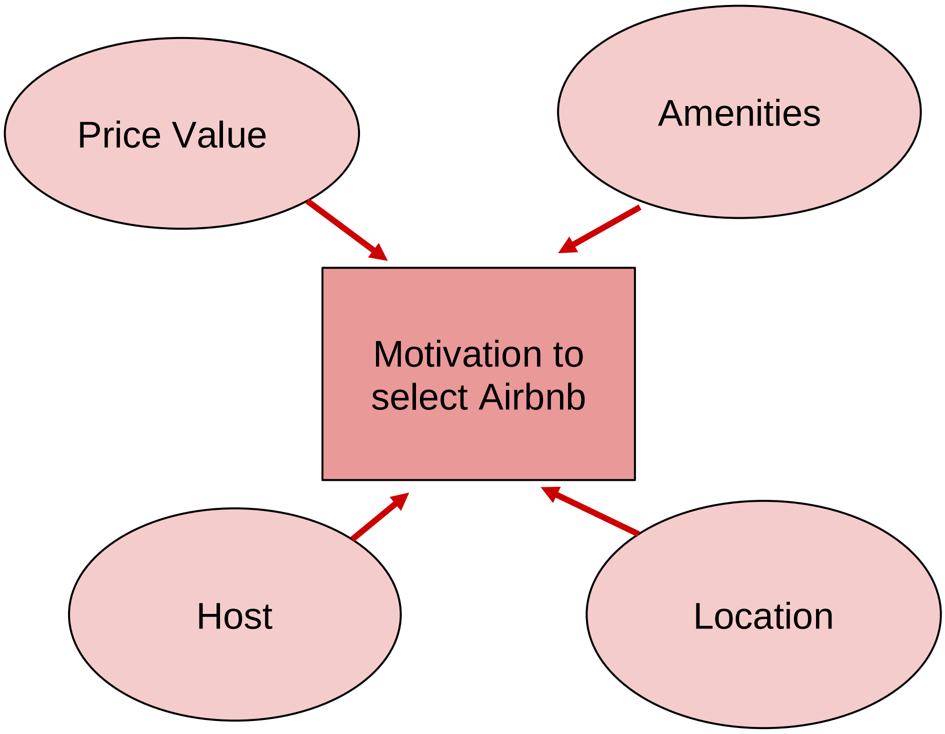

Figure 2.1: Proposed Motivation to Select Airbnb

Motivation	Definition	Literature
Price	- Price matches perceived value with the experience value **(Gunasekaran & Anandkumar, 2012)**. - Cheapness of homestay drives tourists to choose it **(Agyeiwaah, 2014)**. - Price drives customers to choose Airbnb (**Guttentag, Potwarka, & Havitz, 2017**; **Thapa, Anicar & Mailini, 2018**; **Calinao, Alcantara & Bermejo, 2019**, **Tran & Filimonau, 2019**)	Gunasekaran & Anandkumar (2012), Agyeiwaah (2014), Guttentag, Potwarka, & Havitz (2018), Thapa, Anicar, & Mailini (2018), Calinao, Alcantara, & Bermejo (2019), Tran & Filimonau (2019)
Host	- Cordial host **(Gunasekaran & Anandkumar, 2012)**. - Community co-prosperity builds trust, collaboration, interactions, and alliances between the homestay and its customers **(Toh, Tan & Yeo, 2016)**. - Most significant factors. Tourists want	Gunasekaran & Anandkumar (2012), Toh, Tan, & Yeo (2016), Guttentag, Potwarka, & Havitz (2018), Thapa, Anicar, & Mailini (2018), Calinao, Alcantara, & Bermejo (2019)

	to interact with their hosts and receive useful local tips from them **(Guttentag, Potwarka, & Havitz, 2018)**. - Warm welcome to guest **(Thapa et al, 2018)**. - Host being proactive and efficient in providing services and handling uncertain circumstances **(Thapa et al, 2018)**. - People select Airbnb to gain local experience that they could share with their families or friends **(Calinao, Alcantara & Bermejo, 2019)**.	
Amenities	- Accommodation facility is critical to achieving customers' satisfaction **(Toh, Tan & Yeo, 2016)**. - Accommodation offered by Airbnb is complete with amenities like TV, Wi-Fi, cooking utensils **(Calinao, Alcantara & Bermejo, 2019)**.	Toh, Tan, & Yeo (2016), Calinao, Alcantara & Bermejo (2019)
Location	-Tourists were interested in the safety atmosphere of the home environment **(Agyeiwaah, 2014)**. - Tourists think the location offered by Airbnb is safe **(Calinao, Alcantara & Bermejo, 2019)**. - Convenient location, household amenities are important for purchase intention **(Tran & Filimonau, 2019)**.	Agyeiwaah (2014), Calinao, Alcantara, & Bermejo (2019), Tran & Filimonau, (2019)

Table 2.1: Summary of Articles

2.6.1 Price

Prices always a vital topic in hospitality industries such as Airbnb (Zhang et al., 2017). A company or business able to gain financial success if they developed good pricing strategy. Spending on lodging is one of the major budgets for tourists. Airbnb has a competitive advantage on traditional hotels on the accommodation price (Chattopadhyay & Mitra, 2020). The average room price offered by Airbnb is usually lesser than room price that hotels offered in majority cities. For instance, the price of renting a whole apartment on Airbnb is 21% cheaper than staying in a hotel room and 49% less expensive to rent out a private room (Priceconomics, 2013).

Pricing is widely acknowledged to be one of the most critical factors to determining the long-term success of the accommodation industry (Wang & Nicolau, 2017). Various studies have proved that pricing plays an important role in affecting customers decision-making (Gunasekaran & Anandkumar, 2012; Agyeiwaah, 2014; Guttentag, Potwarka, & Havitz, 2018; Thapa, Anicar, & Mailini, 2018; Calinao, Alcantara, & Bermejo, 2019; Tran & Filimonau, 2019). Most tourists are attracted to select Airbnb because of the price. The prices offered by the Airbnb hosts mostly match the perceived value and are reasonably priced (Gunasekaran & Anandkumar, 2012; Calinao, Alcantara, & Bermejo, 2019). Hence, this has proved that price is playing an important role in the lodging industry.

Respondents in Agyeiwaah (2014) stated the cheap price will drive them to select accommodation. Guttentag, Potwarka, & Havitz (2018) also proved that compared to other travellers, youth travellers are more attracted to the cheap accommodation price offered by Airbnb hosts. They have conducted a survey and found that money savers are usually attracted by the price of Airbnb. In their study, money savers tend to be youth travellers. Youth travel is one of the fastest growing markets and a major contributor to the growth of world tourism. In 2010, the World Tourism Organization (UNWTO) found that there are 940 million international tourists travelling in the word and 20% of them are young people. They also estimated that there will be almost 300 million of international youth trips in 2020, which has 59% growth from 10 years ago (UNWTO, 2014).

2.6.2 Host

As a well-known home-sharing platform, Airbnb allows travellers to live together with hosts who rent out their house. Several researchers have found that the host of Airbnb is one of the significant factors to attract customers (Gunasekaran & Anandkumar, 2012; Guttentag, Potwarka, & Havitz, 2018; Thapa, Anicar, & Mailini, 2018; Barbosa, 2019). Hosts influenced tourists’ travel experience. Airbnb guests have cordial relationships with their hosts (Gunasekaran & Anandkumar, 2012).

Thapa, Anicar & Mailini (2018) have performed sentiment analysis on travellers to gain insight of what they expected from their hosts. As expected, a warm welcome from the hosts and will make them joyful. Moreover, hosts should be efficient in providing services and handling visitors’ difficulties. Airbnb hosts always show their friendliness by giving some local tips and travel advice. Taking care of visitors in terms of giving proper suggestions and personal touch are attentive to build trust while trust is very important antecedents of tourists' visit, revisit and recommendation (Thapa, Anicar, & Mailini, 2018).

Besides that, staying with hosts also offers a local experience. Many tourists are driven by this factor. Airbnb is well known with the peer-to-peer business model which allows interaction between hosts and guests. The hosts-guest relationship and social interaction is key to the hospitality experience. Consumers have more access to local experience in the context of Airbnb with P2P connection with local hosts (Moon et al. 2019). Guttentag and Potwarka (2017) argues that Airbnb not only enhances the interaction between hosts and guests but also allows the visitors to connect to the local community. Airbnb is promoting their "living with locals" so tourists not only experience the local life but also gain the opportunity to interact with locals (Tran & Filimonau, 2019). The experience of staying in Airbnb is unpredictable and exciting (Guttentag, Potwarka, & Havitz, 2018). Calinao, Alcantara, & Bermejo (2019) believed that most of the tourists selecting Airbnb are to share the local experience with their families and friends.

2.6.3 Amenities

The household amenities are cited as a key factor motivating travellers to choose Airbnb (Toh, Tan, & Yeo, 2016; Calinao, Alcantara, & Bermejo, 2019; Barbosa, 2019). Calinao, Alcantara & Bermejo (2019) stated in their study, accommodation offered by Airbnb is complete with amenities like TV, Wi-Fi, cooking utensils, refrigerator and so on. Additionally, customers the travellers choose Airbnb for the facilities and amenities offered by the host which matched their desired rentals (Liang et al., 2017).

These amenities are the selling point of Airbnb. For example, most Airbnb will provide cooking utensils while most hotels will not offer. For budget travellers, selecting Airbnb can help them to save money as they can cook by themself rather than spend money eating outside. Guttentag, Potwarka, & Havitz (2018) conducted a research and found that most pragmatic novelty seekers are drawn to the novelty of Airbnb and the household benefits offered by the accommodation.

Based on Calinao, Alcantara, & Bermejo (2019), millennial leisure travellers attracted by amenities such as garden, gym, swimming pool and a lot of Airbnb accommodations have a unique preposition related to this. For travellers who are having a road trip in other countries, parking is necessary. Many tourists are satisfied by the fixed parking lots offered in Airbnb so they no need to waste their time finding parking space in a strange country (Calinao, 2019). Guttentag and Potwarka (2017) believed that most of the home seekers are attracted to Airbnb by household amenities. This is logical that home seekers tend to be on relatively long trips and in relatively large travel parties, as such trip characteristics make the amenities that a home offers particularly attractive. Most Airbnb are fully equipped and provided a home environment (Calinao, Alcantara, & Bermejo, 2019).

2.6.4 Location

An accommodation location, as well as the way it relates to various activities in the area are undoubtedly important factors in choosing accommodation (Agyeiwaah, 2014; Calinao, Alcantara, & Bermejo, 2019; Tran & Filimonau, 2019). Amenities, interior design, price of accommodation can be changed easily according to trend, but location is difficult to change. Hence, location is very crucial in establishing accommodation future success. Consumer demand for accommodation location always affected by various factors such as transportation accessibility, parking convenience, closeness to city or attractions, supermarket and so on.

Most respondents in the Calinao, Alcantara, & Bermejo (2019) study agreed with the location of accommodation that Airbnb offered are usually near to convenience stores, clinics, and transport links so they are accessible to their destination and tourist attractions. Tourists hope the accommodation area they stay in has smooth traffic so they can reach their destination on time according to their planning.

Agyeiwaah (2014) believed that tourists were interested in the safe atmosphere of the home environment provided by Airbnb. Seldom customers will choose area with known criminal to stay. Most tourists are attracted to Airbnb because they think the location offered is safe. Respondents in Calinao, Alcantara, & Bermejo (2019) study stated that Airbnb mostly offered condominium units where the regulations and policies are strictly implied. This will guarantee guests a nice and safe stay at their rented place. Moreover, Airbnb provided the information of the accommodation location on the official website to build trust as trust positively affects customer choice of accommodation. (Calinao, Alcantara, & Bermejo, 2019). Besides that, the uniqueness of a homestay location attracts customers to visit (Thapa, Anicar, & Mailini, 2018). For example, one of the most popular listings of Airbnb - treehouse in Atlanta.

3 Methodology

3.1 Text Mining

This project was carried out through text mining process. Text mining process in this project includes various stages: Business Understanding, Data Collection, Data Pre-processing, Visualization and Evaluation.

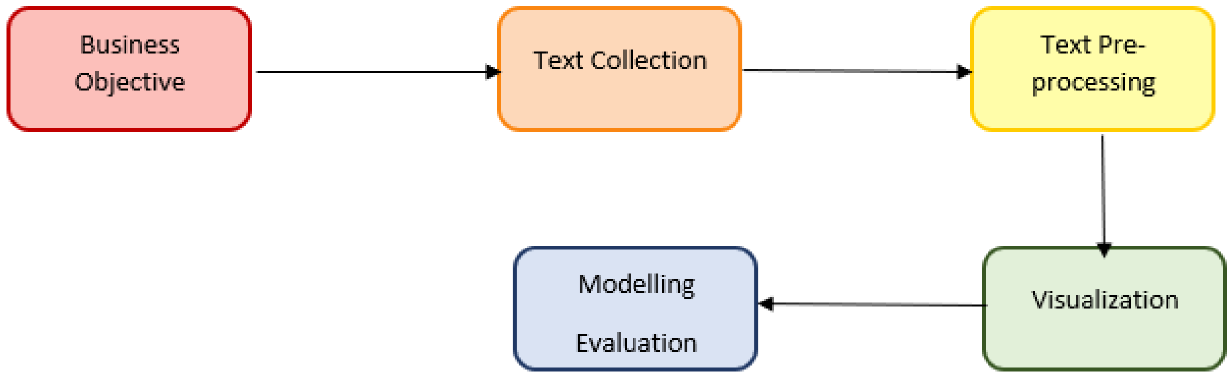

Figure 3.1: Text Mining Process

3.2 Business Understanding

The reason of this project was carried out is to investigate the motivation of western and Asian to select Airbnb as their accommodation while traveling or working in other countries. Identifying what customers care about the most so that Airbnb can make improvement to attract new customers and retain old customers.

3.3 Data Collection

In total, there are 5507 data were collected. The dataset consists of data form year 2019 To 2020. **Table 3.1** shows the variables included in the dataset. The dataset was named all.csv. The objective of this project is to find out the motivation of western and Asian in choosing Airbnb. Hence, the dataset was then split into wester.csv and Asian.csv according to the continent and location variables inside the dataset for further analysis. There are 1395 rows consist in Asian data set while western dataset contains 4112 rows in total. The all.csv dataset also split inside the R studio into 70% training data and 30% testing data for the modelling. The table below shows the variables inside the dataset:

Variables	Description
Date	Date of text written
Screen Name	Name users used online
Full Name	Real name of users
Text	Text about Airbnb posted
App	App where the text scrapped from
Location (Continent)	Continent where the text posted
Location	Location where the text posted

Table 3.1: Dataset Variables

3.3.1 Data Set Splitting

Since one of the objectives of this project is to see whether price, host, amenities and location have any effect on Airbnb's customers accommodation choice. The data collected were distributed into the four different aspects through Excel. The filter function in Excel was used to find the text which contains price, host, amenities/facilities and location. There are 316 values in price dataset, 522 values in host dataset, 256 values in amenities dataset and 349 values in location dataset.

The dataset also split into two datasets Asian.csv and Western.csv at first according to the countries in location variables. This is because the objective of this project is to investigate is there any difference in motivation of Asians and Westerners to choose Airbnb as their accommodation choice. The 15 countries in Asian.csv includes Japan, Indonesia, India, Thailand, Malaysia, China and more while the 6 countries in Western.csv includes United States of America (USA), United Kingdom (UK), Canada, France, Germany and Mexico. In total there are 21 countries form Asian and Western included in this study. There are 1395 rows inside the Asian dataset and 4112 rows in Western dataset.

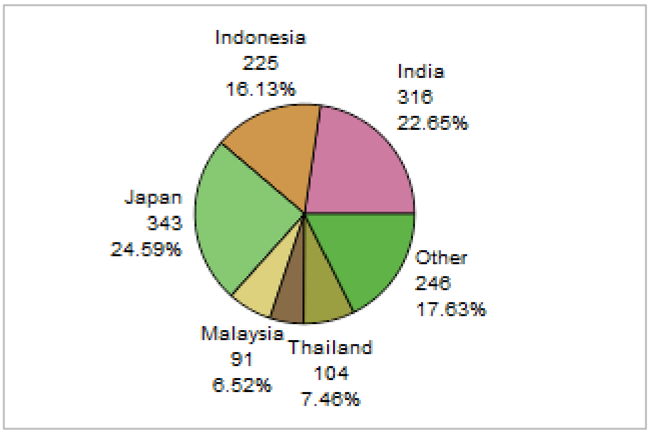

Figure 3.2: Pie Chart on Asian Countries

According to **Figure 3.2**, we can know that the tweets are gathered form which Asian countries. Most of the tweets from Asian are posted from Japan which have approximately 24.59% from overall Asian countries. The second most is India with 22.66% then followed by Indonesia, Thailand, and Malaysia with 16.13%, 7.46%, 6.52%.

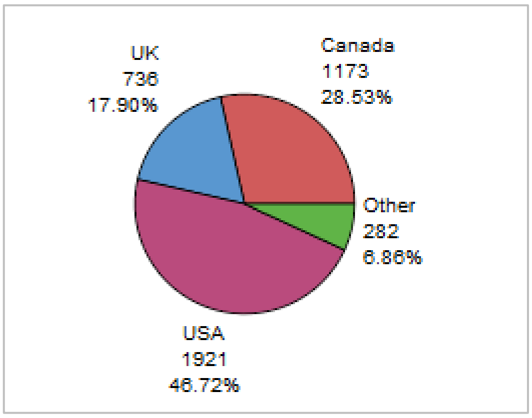

Figure 3.3: Pie Chart of Western Countries

Based on **Figure 3.3**, there are 46.72% of texts were posted in USA then followed by Canada which have 28.53% of texts were posted in this country. UK placed number three with 17.90%. The other countries like Mexico, French, and others only have 6.86% text posted.

3.4 Data Pre-processing

This process starts by scraping information and saving it into an Excel file. So, the scrapped data are all saved in an Excel file as a dataset. Then, text mining techniques are applied to clean the irrelevant words from the text collected. After a corpus based on the text created, text mining techniques are reapplied to obtain a frequency table of terms and a word cloud. The Excel file is later transformed into a .csv (Comma-separated value) so that one can analyse the unstructured texts in the R statistics tool. The data pre-processing is done to remove unnecessary content from tweet text and find out the root out from the words.

In the location attribute, there are same countries with different name. For example, Japan has different values in the attributes such as 東京, Tokyo and so on. Hence, these values are replaced with Japan to make the dataset clean and easy to read. Besides that, a new column named "resident" is added in the dataset. The variable resident only contains two value which are Asian and Western. This variable is added to split the dataset inside R studio more conveniently.

3.4.1 Data Cleaning

While both datasets were running inside the R studio, *distinct()* function was used to remove duplications. *sapply()* function is then applied to the datasets to check the number of missing value. There is no missing value inside the datasets but there are various unknown columns appeared inside the dataset which might be caused by while splitting the datasets. Hence, those unknown attributes x, x.1, and x.2 are removed.

3.4.2 Text cleaning

Before emotion computing, stop words in Tweet text need to be removed. Stop words are words which are filtered out prior to, or after, processing of natural language data. Stop words are common words that carry less important meaning than keywords. These stop words are some of the most common, short function words such as is, where, when, what and so on which are not contributing to this study. Stop words removal is a process of removing these stop words. Hence, to find out emotion from a text, all unnecessary content must be removed and put the useful words into an array.

The following is an algorithm used to remove stop words:

1. **Install *library(tm)***

   Install text mining packages within R studio. This package was installed to convert text into corpus and be able to perform text mining procedure. This package has several functions, which allow conversion of unstructured into structured data by reducing dimensionality of data and keeping relevant information.

2. **Build corpus**

   Corpus is the R object that holds the tweet text. The corpus is created by the corpus function from the *library(tm)*.

3. **Switch to lowercase**

   The text inside the corpus should be converted into lower case to prevent case sensitive stop words. For example, the word “host” and “Host” will be considered the same word.

4. **Remove punctuation**

   Punctuation marks hidden inside the tweet text are removed to prevent failure in the results. The punctuation marks include full stop, comma, brackets and so on used in writing to separate sentences.

5. **Remove numbers**

   Delete the numbers included in tweet text to prevent the effect of numbers on the results.

6. **Delete spaces**

   Remove extra white spaces inside the tweet text.

7. **Remove stop words**

   The unknown words inside tweet text are removed. There are a lot of emojis used by users. Those emojis counted as unknown. Besides that, only tweets in English were retained. The stop words removed include ‘airbnb’.

8. **Stemming**

   Stemming is a process of removing part of word. A word has one root-base form but having different variations. For instance, “book” is a root-base word and “booking”, “booked” are the different forms of single word. They are different words but have the same meaning.

The removal of stop words is based on the elimination of preposition, interjections, numbers and other irrelevant to the analysis. This is to prevent those keywords that have no contribution to this study appearing as the most frequently used word list (Feinerer, 2018). Thus, ensure a correct analysis of the core words used in the text is obtained.

3.5 Visualization

Various analyses were employed to answer the research questions guiding this study. All analyses were conducted through Excel, R studio and SAS Enterprise Guide software. Excel is used to separate the dataset manually into western dataset and Asian dataset for further analysis. R studio is used to run word cloud, sentiment analysis, emotion classification, modelling, and evaluation.

SAS Enterprise Guide shows the basic analysis – pie chart. Various techniques were used to visualize the text for better understanding. The techniques used include word cloud, sentiment analysis, emotion classification, and bar plot.

3.5.1 Word Cloud

Word clouds also known as tag cloud or text clouds. Word clouds are usually used to display the most frequent words appearing in the texts. The word size is related with the frequency of words. When the word appears bigger and bolder, the more frequent it is mentioned in a text. Similar frequency words will be in similar size and colour. Word cloud representations can support readers to distil large sets of text to more easily identify patterns. This provides quick look for formative purposes (DePaolo & Wilkinson, 2014).

Word cloud is used in this study to observe what is the positive and negative words mentioned the most in the Tweet text. This is to have a quick look on why customers satisfied and dissatisfied. The more the words appear in the Tweet text, the more they care about that things.

3.5.2 Sentiment Analysis

Sentiment analysis also called opinion mining implies extract specific opinions, emotions and sentiments from the Tweet text. It is widely used to track attitudes and feelings on the web especially for measuring performance. Sentiment analysis is the automated process of analysing text data and classifying opinion, these system extract attributes of the expression such as polarity, subject, and opinion holder (Liu, 2012).

Sentiment analysis will be implied in this project to identify the positive and negative words in the Tweet text. As mentioned before, many people believed that products with positive feedbacks ore comments have good quality while negative reviews associated with low-quality products (Lee, Park & Han, 2008). There are about 98% of people plan make their travel choice depending on online reviews and 84% of them will be affected (Gretzel & Yoo, 2008).

Sentiment score will be calculated to make comparison within Asians and Western. The sentiment score is to figure out the attitude of Asians and Westerners towards the 4 aspects. The sentiment analysis can be act as summary of Airbnb performance evaluation by the customers.

3.5.3 Emotion Classification

There are 8 emotions will be shown in the emotion classification analysis. The emotions include anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Anticipation, joy, surprise and trust will be considered as positive emotion while anger, disgust, fear and sadness are considered as negative emotions (Thapa, Anicar, & Mailini, 2018). Emotion processing is currently a hot and active area in the field of Natural Language Processing (NLP). Textual emotion detection or classification is one task that many scholars and researchers concentrate on

The emotion classification is useful to see what Asians and Westerners think about the 4 motivations: price, host, amenities and location. The emotion classification is given the role as a summary of Airbnb service quality evaluation by their customers.

3.6 Modelling

In reference to data science, modelling means formulating every step and gathering the techniques required to achieve the solution. Data modelling is a process of creating a data model for an information system by applying formal data modelling techniques. The machine learning techniques used to in this project are Naïve Bayes, Support Vector Machine, Decision Tree and Random Forest.

3.6.1 Naive Bayes

Naïve Bayes focuses on Bayes' Theorem which is known as a classification algorithm. Naïve Bayes analyse the relationship between each attribute and each class. Naïve Bayes model is easy to build, and it is very effective for large data sets. Naïve Bayes is believed to surpass highly complex classification techniques as well as usability (Ray, 2017). The Bayes' Theorem is shown below:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

P(c|x) = c is target variable, x is the predictor
P(c) = The prior probability of the target variable
P(x) = The prior probability of the predictor
P(x|c) = The likelihood of the prior probability of the predictor given target

3.6.2 Decision Tree

Decision Tree uses a tree structure to specify sequences of decisions and consequences. Decision Tree is a supervised learning technique that builds classification or regression models in the form of tree structure. Decision Tree's comprehensibility able to uncover small or large data structure and predict the value (Swamy & Hunumanthappa, 2012). It breaks down dataset into smaller and smaller subsets while at the same time an associated Decision Tree is incrementally developed. Decision Tree algorithm explain the relationship with attributes and the comparative significance of attributes. Furthermore, it can handle both numerical and categorical data. As stated by Jain (2017), the core algorithm for building Decision Tree called ID3 by J. R. Quinlan which employs a top-down, greedy search through space of possible branches with no backtracking. ID3 uses Entropy and Information Gain to construct a Decision Tree.

A Decision Tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogenous). ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogenous the entropy is zero and if the sample is an equally divided it has entropy of one (Jain, 2017). To build Decision Tree, we need to calculate two types of entropy using frequency tables as follows:

- Entropy using the frequency table of one attribute

$$E(S) = \sum_{I=1}^{C} -P_i \log^2 P_i$$

- Entropy using the frequency table of two attribute

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$

The information gain is based on the decrease in entropy after a dataset is split on an attribute:

$$Gain(T, X) = Entropy(T) - Entropy(T, X)$$

3.6.3 Random Forest

The random forest is a supervised learning method that operates by constructing multiple Decision Tree during training phase. he decision of most of the tree is chosen by Random Forest as the final decision. It merges the decisions of multiple decision trees in order to find an answer, which represents the average of all these decision trees. It uses labelled data to "learn" how to classify unlabelled data. Random Forest algorithm is used to solve both regression and classification problems, making in diverse model that is widely used by engineers. With Random Forest, there is no overfitting. The use of multiple trees is to reduce the risk of overfitting. Besides that, it is fast to train with test data (Schott, 2019).

When performing Random Forest based on classification data, Gini index. The formula used to decide how nodes on a decision tree branch:

$$Gini = 1 - \sum_{i=1}^{c} (pi)^2$$

Pi = relative frequency of the class observed in dataset
c = number of classes

This formula uses the class and probability to determine the Gini of each branch of node, determining which of the branches is more likely to occur. Entropy also can be used to determine how nodes branch in a decision tree.

4 Results

4.1 Asians Motivations

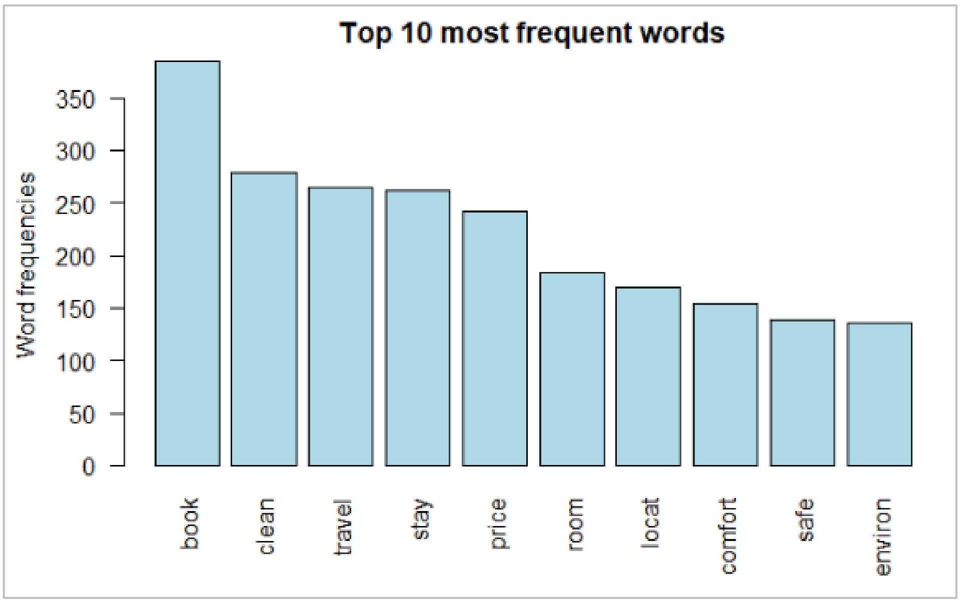

Figure 4.1: Top 10 Most Frequent Word in Asians

Word frequency technique was applied on the Tweet text. **Figure 4.1** shows the top 10 words appeared the most in the Asian's comments about Airbnb. We can obviously see that the most frequent keywords in the collected Asians' Tweet text is "book" which has 385. "clean" was the second most mentioned word used by Asian's Airbnb users followed by "travel", "stay", "price", “room”, “location”, “comfortable”, “safe”, and “environment”. We can see that Asians are more likely to select their accommodation form Airbnb because of the price, location and quality such as cleanliness, comfortability, and environment.

KEYWORDS	WORDS	ASSOCIATION
Book	Easy	0.53
	Convenient	0.43
	Airbnbfamilytravel	0.43
	Service	0.43
	Cheaper	0.41
	Website	0.38
Clean	Bedroom	0.86
	Enough	0.84
	Family vacation	0.40
Travel	Airbnbfamilytravel	0.50
	Budgetrip	0.44
	Trip	0.40
	Service	0.33
	Convenient	0.30
Stay	Facilities	0.47
	Airbnbfamilytravel	0.38
	Cozy	0.37
	Easy	0.33
	Affordable	0.32
	Stay facilities	0.32
	Cheaper	0.31
Price	Affordable	0.53
	Facilities	0.39
	Great	0.37
	Cheaper	0.37
	Accommodation	0.35
	Stay facilities	0.33
Room	Tidy	0.58
	Cheaper	0.32
	Privacy	0.32
	Accommodation	0.30
Location	Privacy	0.40
	Great	0.37
	Tidy	0.32
	Accommodation	0.30
Comfortable	Stay	0.54
	Price	0.49
	Airbnbfamilytravel	0.36
	Book	0.35
	Room	0.32
Safe	Location	0.58

	Tidy	0.55
	Room	0.45
	Accommodation	0.42
	Price	0.41
	Stay	0.39
	Great	0.35
	Privacy	0.31
	atmosphere	0.31

Table 4.1: Asians Word Association

Table 4.1 shows the words associated with the high frequency words which are highly occurred in the Tweet text - “book”, “clean”, “travel”, “stay”, “price”, “room”, and “location”.

As per **Table 4.1**, “book” is associated with “easy” and “convenient” which the correlation value between the two words is 0.53. Asians think that the booking service provided by Airbnb is convenient and easy to use. Besides that, “book” also associated with “airbnbfamilytravel”, “service”, “cheaper”, and “website”. Since, the “book” is connected with “easy”, Asians might think the booking service is ease to use. According to Calinao, Alcantara & Bermejo (2019), the Airbnb website made their booking process easier and increase their satisfaction towards Airbnb. The ease of booking can influence the customers’ satisfaction.

As the second most mentioned words, “clean” is related with “bedroom”, “enough” and “family vacation”. “bedroom” and “enough” are highly related with “clean” compared with “family vacation” which the correlation values are 0.86 and 0.84. From the word association we can assume that Asians think the cleanliness of bedroom is one of the criteria that drives them to choose Airbnb.

In the Tweet text, “travel” has relationship with “airbnbfamilytravel”, “budgetrip”, “trip”, “service”, and “convenient”. We can see that “airbnbfamilytravel” is related with man keywords. This proved that most Asians travel with their family. We can know that most Asians travellers are family travellers and budget travellers.

The word “stay” relates to “facilities”, “airbnbfamilytravel”, “cozy”, “easy”, “affordable”, “stay facilities”, and “cheaper”. Asians mentioned “facilities” together when they mentioned “stay”. Hence, we can assume that facilities in stay is quite important. Besides that, they also hope the stays they selected are cozy and affordable.

There are connections between "price" with "affordable", "facilities", "great", "cheaper", "accommodation", and "stay facilities". Most Asians think that the accommodation price offered by Airbnb is affordable, great and cheaper.

Table 4.1 shows that "room" has correlation with "tidy", "cheaper", "privacy" and "accommodation". For Asians, they often mentioned the room offered by Airbnb is tidy. Furthermore, privacy also seems significant to them.

The word "location" is connecting with "privacy", "great", "tidy", and "accommodation". As we stated before, privacy is important to Asians, they hope their accommodation privacy is guaranteed.

Then, we talked about "comfortable". It is obviously that "comfortable" has relationship with "stay", "price", "airbnbfamilytravel", "book" and "room". Since they travel together with family members, Asians hope that they can stay in a comfortable toom.

Lastly is the "safe". "safe" relates to "location", "tidy", "room", "accommodation", "price", "stay", "great", "privacy", and "atmosphere". Asians really conveyed about what they want is a safe accommodation. Besides that, privacy also important.

4.2 Westerners Motivations

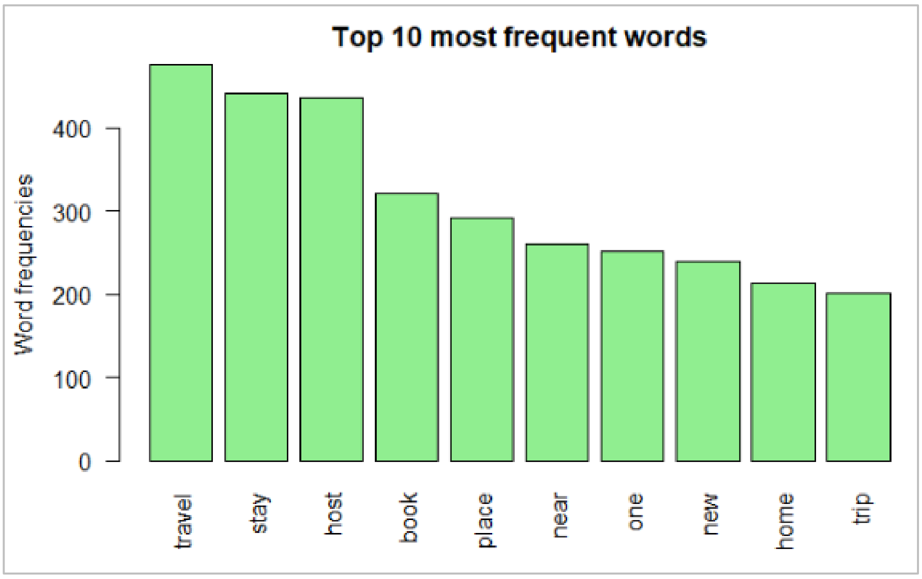

Figure 4.2: Top 10 Most Frequent Word in Westerners

According to **Figure 4.2**, the “travel” was the most frequent occurred word within Tweet text form Westerners which occurred 489 times. “stay” is the second most frequently used words in Westerners’ Tweet text, followed by “host”, “book”, “place”, “near”, “new”, “one”, “trip”, and “home”. For Westerners, host and near accommodation location are more important to them.

KEYWORDS	WORDS	ASSOCIATION
Travel	Solo	0.57
	Trip	0.38
Stay	Airport	0.47
	Mall	0.38
	Subway	0.38
	Shop	0.35
Host	Kind	0.47
	Responsible	0.36
	hosttip	0.36
	Local environment	0.36
	Get airbnb book	0.32
	helpful	0.30
Book	Location	0.52
Place	Book	0.53
	Convenient	0.51
	safe	0.30
Near	Mall	0.79
	Subway	0.79
	Shop	0.71
	Airport	0.45

Table 4.2: Westerners Word Association

Table 4.2 is the word association table in Westerners. As we can see that “travel” is connecting with “solo” and “trip”. Westerners are more enjoy travelling alone. Most of them are solo travellers.

The word “stay” also one of the top 10 most frequent words appeared in Asians’ Tweet text but different from Asians, “stay” in Westerners is associated with “airport”, “mall”, “subway”, “shop”, and “shop”. Westerners usually select their Airbnb accommodation because of the convenience of location. They wished that the location of their stay near to airport, mall, subway station or shops.

Based on **Table 4.2**, “host” is related to “kind”, “responsible”, “hosttip”, “help”, “localenvironment” and “getairbnbnbook”. Host seems quite important to drive Westerners to choose Airbnb as their accommodation. Most of them think Airbnb's hosts are kind and responsible. Besides that, they choose to live with host to get travel tips or advices from host. Living with host also make them having local environment.

There is a linkage between "book" and "location". There are 0.49 correlation value between the two words. From the associated word, we can know that they will book their lodging depend on the location.

"place" is bonding with "book", "convenient" and "safe". As stated previously, the location convenience seems like one of the important criteria for Westerners to choose Airbnb. They will consider the convenience of the place when they are booking their accommodation. Furthermore, they want their accommodation placed in a safe location.

Similar to the "stay", "near" has relationship with, "mall", "subway", "shop" and "airport". Westerners had mentioned many times in the Tweet text that they hope the place they stay will be near to mall, subway stations, shops or airport.

4.3 Price, Host, Amenities and Location Sentiment Analysis

PRICE

negative <dbl>	positive <dbl>	sentiment <dbl>
136	1611	1475

HOST

negative <dbl>	positive <dbl>	sentiment <dbl>
222	1092	870

AMENITIES

negative <dbl>	positive <dbl>	sentiment <dbl>
75	1516	1441

LOCATION

negative <dbl>	positive <dbl>	sentiment <dbl>
106	1327	1221

Figure 4.3: Number of Sentiment Words

As per **Figure 4.3**, the positive words in all the four aspects price, host, amenities and location are more than negative words. In price, there are 136 negative words and 1611 positive words. Host has 222 negative words and1092 positive words. There are 75 negative words and 1516 positive words in amenities. The location got 106 negative words and 1327 positive words.

Figure 4.4: Sentiment Comparison Word Cloud

From **Figure 4.3** we already knew that people have more positive attitudes than negative attitudes toward the four aspects. **Figure 4.4** shows the positive and negative words in the 4 aspects – price, host, amenities and location. In the price sentiment word cloud, we can see that people have positive attitude to affordable, reasonable, great and nice Airbnb price while they showed negative attitude when the price value is expensive, unacceptable and overpriced. The host word cloud shows that helpful, friendly and nice host made them have positive attitude while they have negative attitude if the hosts are terrible or scamming them. The amenities sentiment word cloud proves that Airbnb customers like clean, comfortable and cozy environment and dislike boring, terrible environment. In the location sentiment word cloud, people choose like Airbnb when the location is clean, safe, convenient and great but they do not like Airbnb when the location is bad, weird, and with problem.

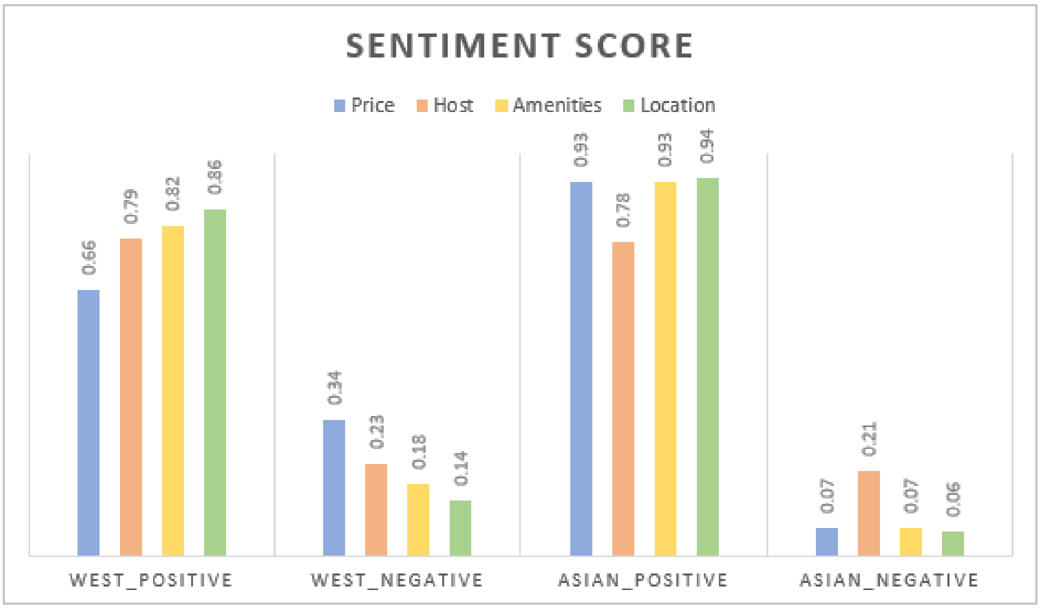

Figure 4.5: Sentiment Score

Figure 4.5 shows the sentiment score of Asians and Westerners on the 4 motivations – price, host, amenities, and location. The aspect location received the highest positive score in both Asians and Westeners. As stated before, Asians care about the location privacy while Westerners care about the location convenience. Price value has the lowest positive score in Westerners while host has the lowest positive score in Asians. Not only price, amenities also has high positive score in Asians. The two motivations that gain high positive score in both Asians and Westerners are location and amenities.

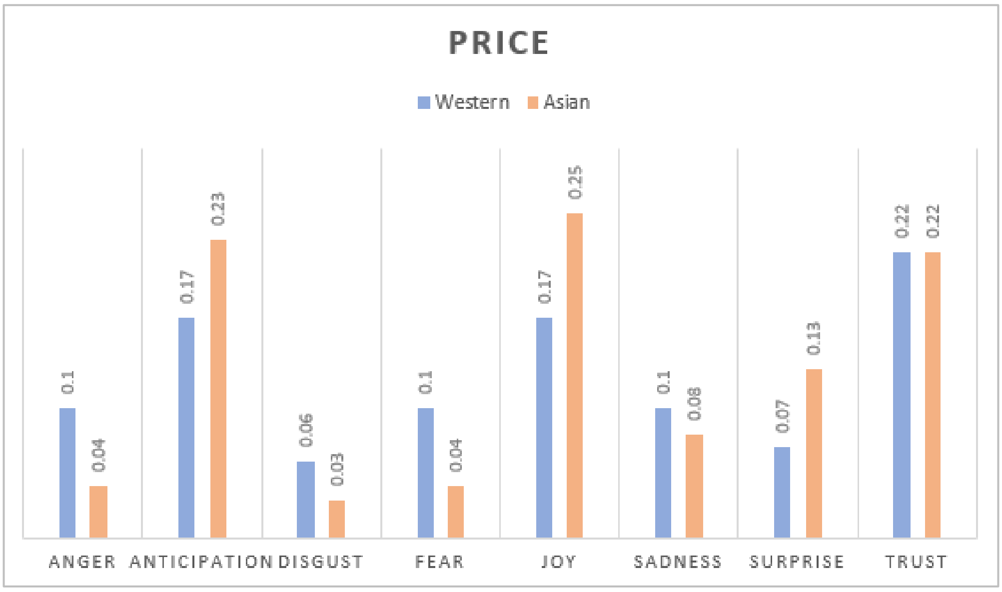

Figure 4.6: Price Emotion Score

Figure 4.6 is the emotion score for Airbnb's price. There are 8 emotions in total include anger, anticipation, disgust, joy, sadness, surprise and trust. The anticipation, joy, surprise and trust are considered as positive emotions while the rest are considered as negative emotions (Thapa, Anicar, & Mailini, 2018). Overall, Asians have better emotion score than Westerners. The anticipation, joy, surprise emotion score in Asians are all better than Westerners and they have same trust score. They also think the price offered is what they expected. Furthermore, Asians also has lower anger, disgust, fear, sadness emotion score than Westerners. The high trust score means they think price is trustable.

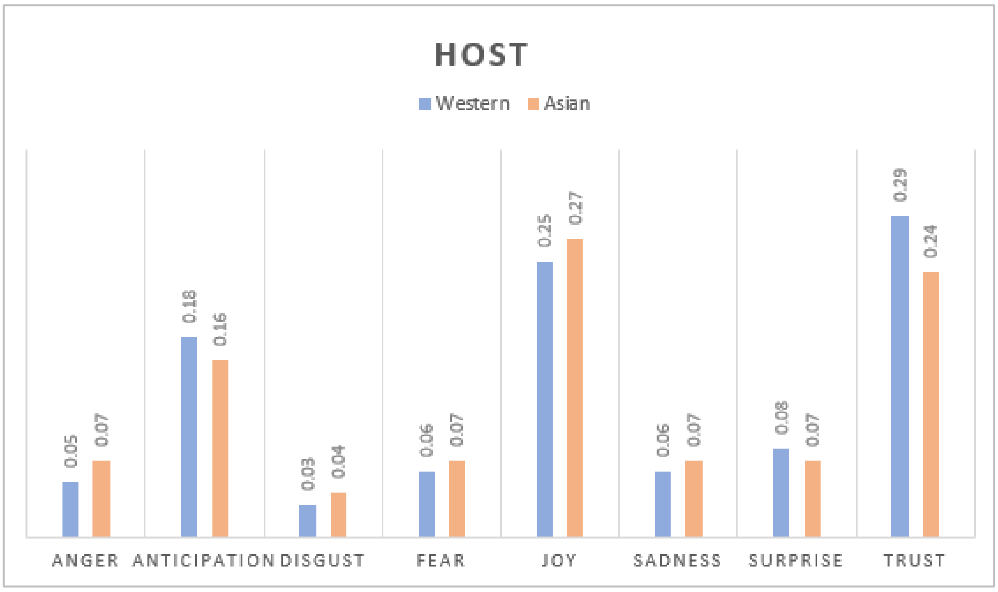

Figure 4.7: Host Emotion Score

Figure 4.7 shows the emotion score of Airbnb's hosts within Westerners and Asians. Overall, both Asians and Westerners have high good emotion score towards the motivation – host. The good emotions – anticipation, joy, surprise and trust all have higher score compared to bad emotion – anger, disgust, fear, and sadness. Both Asians and Westerners think that Airbnb hosts are trustable. Trust is very important factors for customer to visit and make recommendation (Thapa, Anicar, & Mailini, 2018). Furthermore, the joy score also high in both Asians and Westerners. As we stated in **Figure 4.4**, they mentioned in their Tweet text that Airbnb's hosts are nice, helpful. This probably make them feel joy.

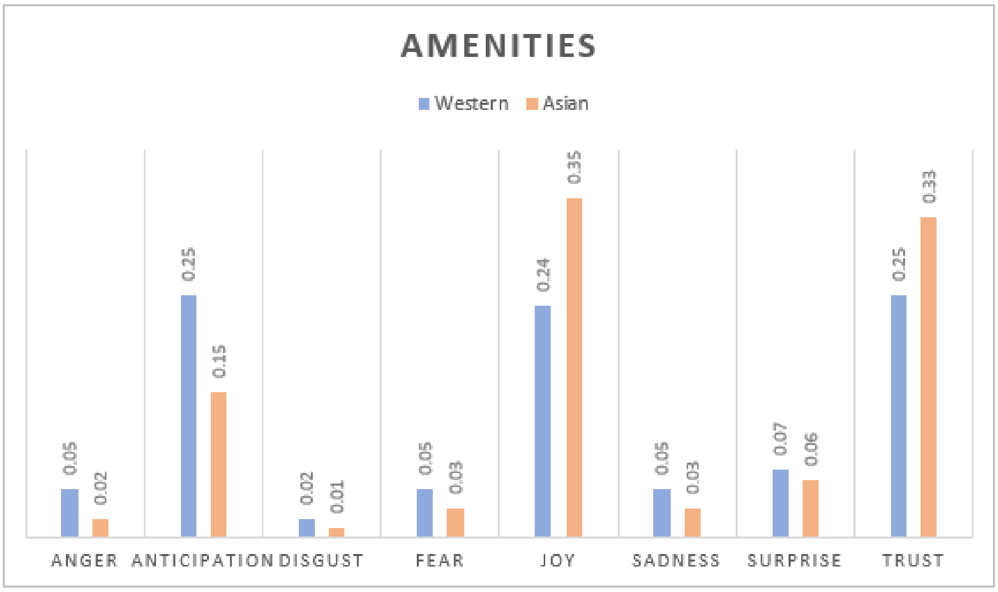

Figure 4.8: Amenities Emotion Score

The emotion score in **Figure 4.8** shows that Asians have high score in anticipation, joy, surprise and trust score. Although Westerners have higher score in anticipation and surprise compared to Asians, but the they also have higher anger, disgust, sadness score. According to Thapa, Anicar, & Mailini (2018), trust is one of the most important antecedents of tourist' visit, revisit or recommendations. As we can see that the trust score in Asians for amenities is quite high. As we can see that, the joy and trust score in Asian is very high. They feel happy towards the amenities offered by Airbnb. The anticipation score in Westerners is higher than Asians which means the amenities offered are what they expected.

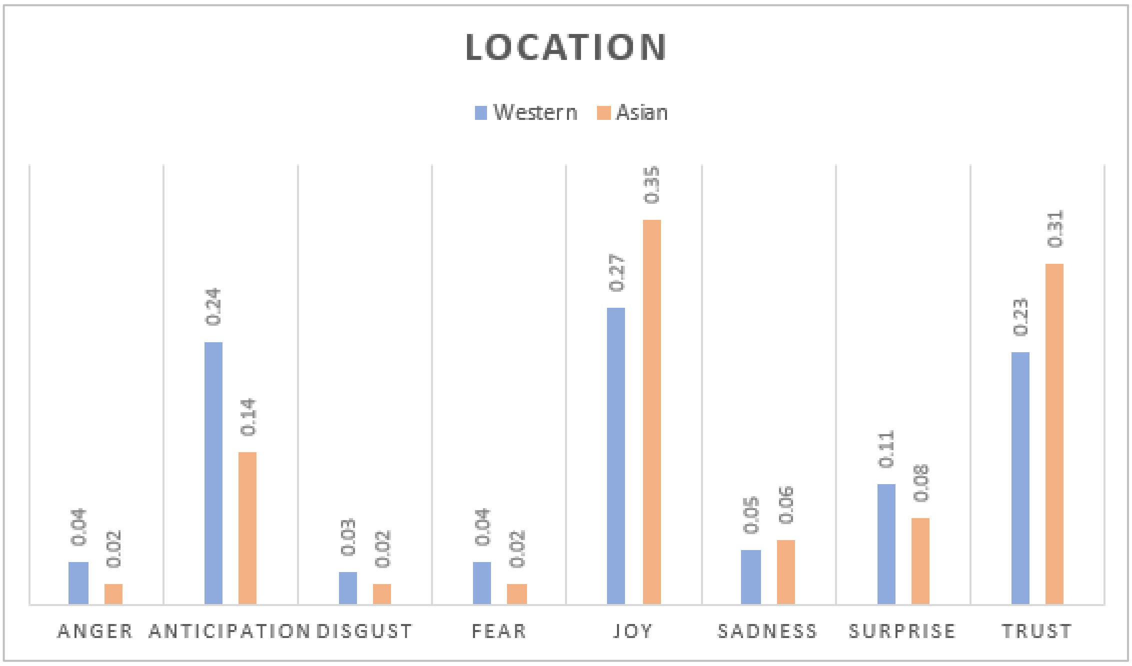

Figure 4.9: Location Emotion Score

Figure 4.9 shows the emotion of Airbnb users towards different attributes – price, host, amenities, and location. Both Asians and Westerners have good emotions towards the location of Airbnb. They have few anger, disgust, fear, and sadness score. As shown in the sentiment score in **Figure 4.5**, location performs the best in both Asians and Westerners. They trust about and feel happy with the location offered by Airbnb. Overall, they are very satisfied with this motivation. The emotions in the form of anger, disgust, fear and sadness show when customers are dissatisfied. The score for negative emotions in **Figure 4.9** is all very low. Very few of them dissatisfied with the location provided. The anticipation score in Westerners always higher than Asians which means the Airbnb always meet their expectations.

4.4 Modelling

Modelling	Sentiment	Accuracy	F Measure	Precision	Recall
Naïve Bayes	*Negative*	0.6237	0.4281	0.3683	0.5104
	Neutral		0.6265	0.5839	0.6758
	Positive		0.7022	0.8144	0.6172
Decision Tree	*Negative*	0.6316	0.1748	0.0969	0.8929
	Neutral		0.6481	0.8590	0.5204
	Positive		0.7056	0.6306	0.8010
Random Forest	*Negative*	0.7405	0.4310	0.3140	0.6864
	Neutral		0.7165	0.7623	0.6759
	Positive		0.8310	0.8637	0.8005

Table 4.3 Comparison of Model

Different kinds of modelling algorithms were used to make comparison for the sentiment classification. The accuracy, F Measure, Precision and Recall were collected through **Table 5.3**. The values give an overview of the performance of the model. The overall dataset was splitted into 70% of training and 30% of testing dataset.

Overall, the Random Forest performs the best. Random Forest is the best model to classify sentiment analysis. Overall, all the algorithms have more than 50% accuracy. Random Forest has the highest accuracy among all with 74% accuracy but the other two only have less than 65% accuracy. Besides that, the precision and recall rate in Random Forest has the highest value among all the algorithms. This indicates that most of the positive samples are correct, and only about 20% or below of the samples are labeled wrong.

5 Discussion

5.1 Asians Motivations

According to the word association table in **Table 4.1**, we can know that the Asians are more likely to travel with their families or having a budget trip. Agoda (2018) has conducted a survey with 10,784 respondents from different countries across Asian, UK, USA and Australis and found Asian families go on twice the amount of family holidays compared with Western travellers. Family trip is one of the travel trends in Asia countries.

Since most Asians are family travellers and budget travellers, they are more sensitive to the accommodation price value. Due to the budget trip, they need to find cheaper and affordable stay during their vacation. Hence, compare with Westerners they will more sensitive with their accommodation price. The word “price” has appeared as top 10 most frequent words in **Figure 4.1**. When we take a look into the **Table 4.1** Airbnb’s price is associated with “affordable”, “great”, and “cheaper”. Furthermore, the sentiment word cloud in **Figure 4.4** also shows that most people will show positive attitude when the price is great, affordable, suitable and nice. The sentiment score and emotion score of Asians also higher than Westerners. The high score of anticipation in **Figure 4.6** also proved that price offered always meet their expectation. Agyeiwaah (2014) also proved that the affordability and cheapness of price drive customers to choose Airbnb. Calinao, Alcantara, & Bermejo (2019) also conducted a study in Filipinos, and stated customers’ satisfaction relates to the price. Airbnb customers will consider whether the accommodation price offered suits their budget or not.

The aspect host is less important to Asians. Although the host score is quite high in Asians, but they mentioned about Airbnb’s host lesser compared to price, amenities and location. This can only prove that the hosts they met are kind, helpful and nice but not an influential motivation for them to select Airbnb. Although the trust score in **Figure 4.7** is quite high and they trust Airbnb’s hosts but they still not preferred to live with someone they are not familiar with.

Amenities also another motivation that Asians care about. Although the amenities or facilities not the top 10 most frequent words in Tweet text but it is highly association with most of the frequent words. As shown in **Table 4.1**, facilities are associated with price and stay. The stay facilities are quite important to them. We already knew that most Asians are family travellers. Furthermore, Asians also have high positive score and good emotion score towards amenities. This is supported by study from Tran and Filimonau (2019). They have conducted a study among Vietnamese and found that household amenities have positive effect on Airbnb’s purchase selection. A study by Toh, Tan, & Yeo (2016) has conveyed their opinion that stay’s facilities is significantly related with customer loyalty. We can see in **Figure 4.4**, Asians are in good mood when the facilities make them feel comfortable, clean and cozy.

Location is also appeared as the highly mentioned words in **Figure 4.1**. As stated in **Table 4.1**, location is having a relationship with privacy, great, tidy and accommodation. Privacy also associated with room. Privacy seems important o Asians. Hence, we can know that why they seldom mentioned host in their Tweet text. Rather living together to host family, they prefer staying without host. In **Figure 4.9**, They feel trust that Airbnb can provide location can guaranteed their privacy and safety. Unlike Westerners, Asians not really emphasis on the convenience of accommodation location. 63% of family travellers from Statista (2015) has revealed that their favourite wat to travel is with car. Hence, they are not care about the location convenience as Westerners because they have transportation.

Despite price, amenities and location, Asians also pay attention to the room quality. In **Figure 4.1**, we can see that the words clean, comfortable and safe are stated. Toh, Tan, and Yeo (2016) founded that the vital consideration for customers to revisit an accommodation is include whether the rooms are well-maintained and clean; whether the accommodation is in a safe and secure environment.

5.2 Westerners Motivations

Different form Asians, **Table 4.2** has clearly showed that most Westerners are solo travellers which means they preferred travelling alone rather than travel with others. According to Condorferries (2020), Agoda ‘s survey has found that 72% of American women have taken at least ne trip by themselves. Besides that, one person out of every six people in United States has taken a vacation alone.

Westerners did not mention much about price value of Airbnb. The top 10 most frequent words listed in **Figure 4.2** also not associated with any words that describe price such as affordable cheap or expensive. Hence, we can assume that price is not an important criterion for Westerners. Besides that, the positive score of prices is slightly low compared to the other aspects. Although they trust about the price offered in Airbnb, but price is not a vital motivation the drives them to choose Airbnb.

Compared with Asians, Westerners cared about Airbnb host. Host has appeared as the second highest mentioned words in their Tweet text. The positive score of host 0.79 is also considered high. The high positive score means that most of them think Airbnb host are helpful, friendly and kind. Besides that, the high trust scores in **Figure 4.7** proved that they think it is trustable to live with hosts. The word association in **Table 4.2** also shows that host in the Tweet text is related with king, responsible and helpful. Besides that, the host also associated with host tips and local environment. As we stated before, most of them are solo travellers. The Solo Traveler Readers Survey has found that 49.04% of solo travellers thinks that local guides help them to understand the history and the culture of the destination. Participants in Guttentag, Potwarka, & Havitz (2018) want to interact with their hosts and receive useful local tips from them. Gunasekaran & Anandkumar (2012) agreed that local culture and interaction with local provide an authentic experience which attract people to choose Airbnb. In **Table 4.2**, the word association shows that Westerners are attracted to the local environment.

Westerners are not very concerned about amenities. The high positive score and good emotion score only can represent that they are satisfied by the facilities offered by Airbnb but not the necessary things to drive them select their accommodation. They didn't mention much about Airbnb's facilities or amenities in their Tweet text.

Location is the motivation which is important in both Asians and Westerners. Although location is highly mentioned in Asians and Westerners Tweet text, but the requirements for location are different within them. Asians more favourable on privacy of location while Westerners more focus on the convenience and safety of the location. In **Table 4.2**, the word near is associated with airport, subway stations or mall which means they want their location near to these places and make them more convenient. Condorferries (2020) has stated that solo travellers are spending about 20% more on travel insurance than other travellers. Hence, we can know that they very care about their safety during vacation. The customers' satisfaction is affected by the location offered is safe or not (Calinao, Alcantara, & Bermejo, 2019). In Toh, Tan and Yeo (2016) study, one of the vital consideration of customers before choosing lodging is the safety of accommodation. Besides that, convenience of accommodation also important for them. Tran & Filimonau (2019) stated that Airbnb's customer agreed that it is important for location to be convenient.

6 Conclusion and Future Work

6.1 Conclusion

Nowadays, the advance of sharing economy has made Airbnb become a significant impact on the tourism accommodation industry. Hence, it is important to discover the motivations from the consumer side for using this innovative service. Understanding customers is a vital way to attract new customers and retain the potential customers. There is a big culture difference between Asians and Westerners. So, this study is to find out is there any difference of motivations between Asians and Westerners to select their travel accommodation. This study has been trying to find some answer to those questions by revealing the motivations that lay behind the feedbacks, reviews or comments that posted online and social media from the customers.

The study reveals the most relevant motivations for Asians are price, amenities, location, and room environment. Since most Asians are budget travellers so it is important for the price of accommodation match their budget. Amenities not only important to Asians, they also seem like quite satisfied with the amenities offered by Airbnb as the positive score is high. Satisfaction is related with customers' intention to choose Airbnb again (Calinao, Alcantara, & Bermejo, 2019). Compared to the convenience, Asians more care about the location that can guaranteed their privacy and safety. Hence, host is not considered as one of the motivations that drive them to select Airbnb because they care about their privacy. Hence, most of them will not willingly to live together with strangers. The room environment includes the cleanliness, and safely of the room they rented. Besides that, cozy and comfortable environment also highly mentioned in their Tweet text.

Different from Asians, Westerners are solo travellers and only attracted by host and location of Airbnb. They preferred to interact with local host to gain local experience in a sharing house rather than staying alone. Host makes customers have a feeling being welcomed at someone's house (Thapa, Anicar, & Mailini, 2018). This provides a home feeling for Airbnb customers when they travel alone in a strange country. The convenience and safety of location is another factor that make them want to choose Airbnb. Since they travel alone, it is important for them to choose accommodation which placed in a safe place. Besides that, accommodation near to subway stations, mall, or airport can reduce their burden.

6.2 Implications and Future Studies

This study offers a theoretical foundation in future hospitality studies on the sharing economy. The new proposed text analytics method could be a useful tool to enable scholar to better understand customer experience evaluation from text. The present work also contributes to the literature by mining customers Tweet text. This study also provides practical advice to Airbnb investors and hosts regarding methods to make improvement according what customers really value and care about.

The competition in tourism and hospitality industry is quite fierce understanding the customers is important to stand out from the competitors. This study really investigates the motivations lay on customers' feedbacks and comments that were posted through social media. Through this study, Airbnb host can learn what the Asians customers and Western customers preferred. This help them designing better strategies and enhance their performance. Furthermore, the sentiment analysis and emotional classification can act as the service performance evaluation by Airbnb customers which is helpful for host. The performance evaluation is helpful for those host to take measure, make improvement and stand out of the competitors. The study also provides a planning for those who intended to rent out their house through Airbnb.

6.3 Limitation

The limitation in this study is the Tweet text are more scrapped from the South-East Asia countries include India, Indonesia, Malaysia, Thailand and so on. As we know that people from Asian countries like China, Taiwan, Hong Kong seldom write their comments in English. The Chinese comments are not included in the project. Moreover, they are more constantly convey their feedback in their countries microblogging site such as Weibo. The language problems also might limit those customers who can't speak English to choose accommodation in English-speaking countries.

References

Agyeiwaah, E. (2014). International tourists' motivations to choose homestay: Do their socio-demographics have any influence? *Tourism and Hospitality Research.* 13(16). doi: 10.1177/1467358413517895.

Airbnb. (2019, February 25). The Emerging Markets: Powering Airbnb's Global Growth Report. Retrieved from https://news.airbnb.com/the-emerging-markets-powering-airbnbs-global-growth-report/

Airbnb. (2020). About us. Retrieved from https://news.airbnb.com/about-us/

Alaei, A., R., Becken, S. & Stantic, B. (2017). Sentiment analysis in tourism: Capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191. doi: 10.1177/0047287517747753.

Antoniadis, K., Zafiropoulos, K. & Varna, V. (2015). Communites of followers in tourism Twitter accounts of European countries. *European Journal of Tourism, Hospitality and Recreation*, 6(1), 11-26.

Aslam, S. (2020, February 10). Twitter by the numbers: Stats, demographics & fun facts. Retrieved from https://www.omnicoreagency.com/twitter-statistics/

Aslam, S. (2020, October 28). Twitter by the numbers: Stats, demographics & fun facts. Retrieved from https://www.omnicoreagency.com/twitter-statistics/#:~:text=Twitter%20Demographics,(users%20outside%20the%20U.S.)

Calinao, R. J., Alcantara, A. M., Bermejo, L. I. L. (2019). Customer satisfaction and level of motivation among Airbnb customers in Makati City, Philippines. *Journal of International Tourism and Hospitality Management.* 4(2).

Chattopadhyay, M., & Mitra, S., K. (2020). What Airbnb host listings influence peer-to-peer tourist accommodation price? *Journal of Hospitality & Tourism Research (Washington, D.C.)*, 44(4), 109634802091021-623. doi: 10.1177/1096348020910211.

Condorferries. (2020). Solo travel statistics 2020. Retrieved from https://www.condorferries.co.uk/solo-travel-statistics#:~:text=Solo%20travel%20%231%20trend%20for,already%20taken%20a%20holiday%20alone.&text=25%25%20of%20millennials%20in%20the,in%20the%20next%2012%20months.

Cui, Y. H., Niu, J. S. & Guo, R. (2015). Effects of negative online reviews on consumers' attitudes and behavioural intentions toward online products. *Economic Management Journal*. 4(4), 75-83.

Ćurlin, T., Jaković, B. & c, I. (2019). Twitter usage in tourism: Literature review. *Business System Research*, 10(1), 102-119, doi: 10.2478/bsrj-2019-0008.

Dang, S., & Ahmad, P. H. (2015). A review of text mining techniques associated with various application areas. *International Journal of Science and Research (IJSR)*, 4(2), 2461-2466.

Depaolo, C., & Wilkinson, K. (2014). Get your head into clouds: Using word clouds for analyzing qualitative assessment data. *TechTrends*, 58(3), 38-44. doi: 10.1007/s11528-014-0750-9.

Elrhoul, M. (2014, November 20). Three new insights for travel brands on Twitter. Retrieved from https://blog.twitter.com/en_us/a/2014/three-new-insights-for-travel-brands-on-twitter.html

Ert, E., Fleischer, A. & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*. 55, 62-73, doi: 10.1016/j.tourman.2016.01.013.

Fang, B., Ye, Q., Kucukusta, D., and Law, R. (2016), Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52, 498–506. doi: 10.1016/j.tourman.2015.07.018.

Feinerer, I. (2018, December 12). Introduction to the tm package Text Mining in R. Retrieved from https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82-89. doi: 10.1145/2436256.2436274.

Gretzel, U. & Yoo, K. H. (2008). Use and impact of online travel reviews. *Laboratory for Intelligent Systems in Tourism*, 35-46. doi: 10.1007/978-3-211-77280-54.

Güçlü, B., Roche, D. & Marimon, F. (2019). City characteristics that attract Airbnb travellers: Evidence from Europe. *International Journal for Quality Research*, 14(1), 271-290, doi: 10.24874/IJQR14.01-17

Gunasekaran, N., & Anandkumar, V. (2012). Factors of Influence in Choosing Alternative Accommodation: A Study with Reference to Pondicherry, A Coastal Heritage Town. *Procedia Social and Behavioral Sciences*, 62, 1127-1132.doi: 10.1016/j.sbspro.2012.09.193.

Guttentag, D. (2015). Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, 18(12), 1192-1217. doi: 10.1018/13683500.2013.827159.

Guttentag, D., Smith, S., Potwarka, L. & Havitz, M. (2018). Why Tourists Choose Airbnb: A Motivation-Based Segmentation Study. *Journal of Travel Research*, 57(3), 342-359, doi: 10.1177/0047287517696980.

Jain, R. (2017, March 21). Decision tree. It begins here. Retrieved from https://medium.com/@rishabhjain_22692/decision-trees-it-begins-here-93ff54ef134.

Lee, J., Park, D. H., Han, T. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*. 7(3), 341-352. doi: 10.1016/j.elerap.2007.05.004.

Leung, D. et al. (2013). Social media in tourism and hospitality: A literature review. *Journal of Travel & Tourism Marketing*, 30(1-2), 3-22. doi: 10.1080/10548408.2013.750919.

Liu, B. (2012). Sentiment analysis and opinion mining. Retrieved from https://doi.org/10.2200/S00416ED1V01Y201204HLT016.

Luo, Y. (2018). What Airbnb's reviews can tell us? An advanced latent aspect rating analysis approach. *Graduate Theses and Dissertations.* doi: 10.31274/etd-180810-6033.

McCarthy, L., Stock, D. & Verma, R. (2010). How travelers use online and social media channels to make hotel-choice decisions. *Cornell Hospitality Report,* 10(18), 4-18.

Moon, H., G. et al. (2019). Peer-to-peer interactions: Perspectives of Airbnb guests and hosts. *International Journal of Hospitality Management*, 77(77), 405-414. doi: 10.1016/j.ijhm.2018.08.004.

Parente, R. C., Geleitla, J. G. & Rong, K. (2018). The sharing economy globalization phenomenon: A research agenda. *Journal of International Management*, 24(1), 52-64, doi: 10.1016/k.intman.2017.10.001.

Philander, K., & Zhong, Y., Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Mangement*, 55, 16-24. doi: 10.1016/j.ijhm.2016.02.001.

Priceconomics (2013, June 17). Airbnb vs hotels: A price comparison. Retrieved from https://priceonomics.com/hotels/

Quratulain, R., Sajjad, H., & Sayeed, G. (2016). Lexicon-based sentiment analysis of teacher's evaluation. *Applied Computational Intelligence and Soft Computing*, 2016, 1-12. doi: 10.1155/2016/2385429.

Radovanovic, M., & Ivanovic, M. (2008). Text mining: Approaches and applications. *Novid Sad Journal of Mathematics*¸ 38(3), 227-234.

Ráthonyi, G. (2013). Influence of social media on tourism - especially among students of the University of Debrecen. *Applied Studies in Agribusiness and Commerce,* 7(1), 105-112. doi:10.19041/Apstract/2013/1/18.

Ray, S. (2017, September 11). 6 easy steps to learn Naïve Bayes algorithm with codes in Python and R. Retrieved from https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

Raymond, A. F. & Charlie, Z. (2019, October 7). Analysis and machine learning modelling of New York City Airbnb data. Retrieved from https://nycdatascience.com/blog/student-works/analysis-and-machine-learning-modeling-of-new-york-city-airbnb-data/

Schott, M. (2019, April 25). Random forest algorithm for machine learning. Retrieved from https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb#:~:text=When%20using%20the%20Random%20Forest,better%20decision%20for%20your%20forest.

Schuller, B., Mousa, A., E., & Vryniotis, V. (2015). Sentiment analysis and opinion mining: On optimal parameters and performances: Sentiment analysis and opinion mining. *Wiley Interdisciplinary reviews. Data Mining and Knowledge Discovery*, 5(5), 255-263. doi: 10.1002/widm.1159.

Sinoara, R., A., Antunes, J., & Rezende, S., O. (2017). Text mining and semantics: A systematic mapping study. *Journal of the Brazilian Computer Society*, 23(1), 1-20. doi: 10.1186/s13173-017-0058-7.

Solotraveler. (2020, November 10). Solo travel statistics and data: 2020. Retrieved from https://solotravelerworld.com/about/solo-travel-statistics-data/

Statista. (2015, September 28). Preferred modes of transportation when taking a family vacation in the U.S. 2015. Retrieved from https://www.statista.com/statistics/470196/us-family-vacation-transport-preferences/

Suchánek,P., Richter, J., & Králová, M. (2015). Customer satisfaction, product quality and performance of companies. *Review of Economic Perspectives*, 14(4), 319-344. doi: 10.1515/revecp-2015-0003.

Swamy, M., N., & Hanumanthappa, M. (2012). Predicting academic success from student enrolment data using decision tree technique. *International Journal of Applied Information Systems*, 4(3). doi: 10.5120/ijais12-450654.

Tam, A., S. (2019, May 28). Airbnb's growth strategy: How they attract and retain 250 million users. Retrieved from https://www.growthmanifesto.com/airbnb-growth-strategy

Thapa, B., Anicar, D., M., & Malini, D., H. (2018). Unraveling Tourists' Preferred Homestay Attributes from Online Reviews: A Sentiment Analysis Approach. *International Journal of Pure and Applied Mathematics*, 119(15), 1567-1585. doi: 10.13140/RG.2.2.20453.65768.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2), 406-418. doi: 10.1002/asi.21462.

Tighe, P., J. et al. (2015). The painful tweet: Text, sentiment and community structure analyses of tweets pertaining to pain. *Journal of Medical Internet Research*, 17(4), e84-e84. doi: 10.2196/jmir.3769.

Toh, S. T. J., Tan, C. L., & Yeo, S. F. (2016). What are the attractiveness aspects that influence customer loyalty to homestay? A study in Taiwan. *Jurnal Pengurusan.* 48, pp. 201-208. doi: 10.17576/pengurusan-2016-48-16.

Tran, T. H. & Filimonau, V. (2019). The (de)motivation factors in choosing Airbnb amongst Vietnamese. *Journal of Hospitality and Tourism Management.* 42, 130-140, doi: 10.1016/j.jhtm.2019.10.011.

U.S. Travel Association. (2015, October 29). Travel facts and statistics. Retrieved from http://www.ustravel.org/news/press-kit/travel-facts-and-statistics

UNWTO. (2014). The power of youth travel. Retrieved from https://www.wysetc.org/wp-content/uploads/2014/12/wysetc-unwto-report-english_the-power-of-youth.pdf

Varma et al. (2016). Airbnb: Exciting innovation or passing fad? *Tourism Management Perspective 20*, 228-237. doi: 10.1016/j.tmp.2016.09.002.

Wang, D., & Nicolau, J., L. (2017). Price determinants of sharing economy based accommodation rental: A study listings from 33 cities n Airbnb.com. *International Journal of Hospitality Management*, 62, 120-131. doi: 10.1016/j.ijhm.2016.12.007.

Zhang, L., Yan, Q., & Zhang, L. (2018). A computational framework for understanding antecedents of guests' perceived trust towards hosts on Airbnb. *Decision Support System.* 115, 106-116. doi: 10.1016/j.dss.2018.10.002.

Zhang, Z. et al. (2017). Key factors affecting the price of Airbnb listings: A geographically weighted approach. *Sustainability (Basel, Switzerland)*, 9(9), 1635. doi: 10.3390/su9091635.