COMPETITIVE SET SELECTION COMPARISON IN THE HOTEL INDUSTRY: CONTRASTING THE HOTELS' AND CUSTOMERS' PERSPECTIVES

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Ling Zhang

December 2018

© 2018 Ling Zhang

ABSTRACT

Despite the importance of competitive set, the lodging industry does not have a market convention on how to select the competitive sets. This study aims to assess hotels' competitive sets from customers' views. And the main objectives were threefold: 1) compare competitive set identified by hotels with those selected by customers and find out the attributes associating with the dissimilarity and similarity; 2) discover how customers identify the hotel competitors; 3) analyze hotels and customers competitive sets selections. By analyzing the data from Smith Travel Research (STR) and TripAdvisor, we find a low match between hotels' and customers' views of competitive sets selection and they have different opinions on attributes (price, size, class, distance, rank and score) in the selection process. Additionally, in forming consideration sets and choice sets, customers perceive those attributes differently. Finally, hotels' and customers' competitive sets selection are more complex than just using those six attributes.

BIOGRAPHICAL SKETCH

Ling Zhang was born in Huzhou, China in 1994. She graduated from Zhejiang Gongshang University with a Bachelor of Management degree in Business Administration. After graduating, she was admitted by School of Hotel Administration at Cornell University and started her master program.

Ling Zhang had internships in both hotels and hospitality consulting firms. During the internships, she gained hospitality insights and developed her interests in hospitality. She will be graduating in December 2018. With the help of her advisor, she developed the understanding of revenue management and learned valuable research methods. She will start her career in the hospitality industry in the near future.

ACKNOWLEDGMENTS

I must express my very profound gratitude to Professor Chris K. Anderson, my chair advisor. Without his instruction, I would never have been able to finish my master thesis. Professor Anderson is smart, knowledgeable and passionate about research. He guided me to form a clear research question and helped me obtain the valuable data from Smith Travel Research and TripAdvisor. The door to Prof. Anderson's office was always open whenever I ran into a trouble spot or had a question about the research.

And I would like to thank Professor Vincent W. Slaugh for encouragement and assistance. He is friendly and very willing to help.

Finally, I truly appreciate my parents for always supporting me and encouraging me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

1 INTRODUCTION

Smith Travel Research (STR) defines competitive sets (comp sets) as a group of hotels by which a property can compare itself to the groups' aggregate performance. Overall, the importance of defining the right competitive sets will never be overstated in the lodging industry. Many scholars argue that competitive sets identification is an inevitable step in valuation, performance evaluation and strategy formulation. (Morgan and Dev, 1994; Kim and Canina, 2011; Haynes, 2016).

However, there are no market guidelines of select the competitive sets in the hotel industry. Some previous researches have put effort into figuring out the process of managers' competitive sets selection (Clack and Montgomery, 1999; Mohammed et al., 2014) and into figuring out the attributes that managers use to form competitive sets (Yesawich, 1987; Baum and Mezias, 1992; Bull, 1994; Baum and Lant, 2003; Kim and Canina, 2011; Coleman, 2011; Li and Netessine, 2012; Mohammed et al., 2014; Lee, 2014). However, Kim and Canina (2011) argue that competitive sets selection remains a complex issue and its implications are not fully understood. As long as this problem is not solved, hotel managers can select the competitive sets that will give better performance for their own hotels (Webb and Schwartz, 2017).

Customer view is critical to set the competitive set, so many studies indicate that hotel managers should consider customers' perspective to select the right competitive sets (Coleman, 2011; Li et al., 2014; Haynes, 2015). Therefore, Li and Netessine (2012) use the price-matching method and network analysis to find out the difference between competitive sets form hotel side and choice sets constructed by customers. Mohammed et al. (2014) interview the hotel managers and surveys the hotel guests. Although their studies give some insights into match and mismatch of hotel managers and customers, their data and methods aren't without limitations. The price-matching method Li and Netessine (2012) utilized to find the competitive sets selected by hotel managers is not appropriate since some previous researchers (Mathews, 2000; Coleman, 2011) note the price alone is not a good classify. Additionally, the data collected by Mohammed et al. (2014) is from a single hotel in Hongkong. Consequently, there may exist some bias and limitations.

By using the data from Smith Travel Research (STR) and TripAdvisor, we aim to assess the accuracy of competitive sets choice by hotel managers. Moreover, we contrast and compare the attributes hotel managers may use to select the comp sets and consumers may use to form the choice set. Smith Travel Research (STR) is the leading benchmarking service provider in the United States, and increasingly around the world. STR gathers, processes and redistributes performance information of over 30,000 US hotels, and, if applicable, alternate comp set or sets each week (Smith and Zheng, 2011). TripAdvisor is the largest travel website in the world, with more than 315 million reviewers (active and inactive) and over 500 million reviews of hotels, restaurants, attractions and other travel-related businesses. With the development of the Internet, travelers increasing use travel reviews to arrange their trips (Buhalis & Law, 2008; Litvin, Goldsmith, & Pan, 2008). More than 80 percent of individuals search online, and most of them visit 26 websites and spend more than two hours in travel research (Trend, 2013). Therefore, our research will give more deeper insights.

The rest of the paper proceeds as follows. First, we research the literature review. Second, we discuss the data sample. Third, we analyze the results. Fourth, we present the conclusion and implication based on the results. Finally, we state the limitations and suggestions for further studies.

CHAPTER 2

2 LITERATURE REVIEW

2.1 Selection Process

2.1.1 Hotel Side

Some researchers (Chernatony et al., 1993; Baum and Lant, 2003) state that managers will use mental models to simplify the categorization process of forming a comp set. Specifically, they will choose some key attributes among all the attributes a firm exhibits and then compare attributes of firms to identify similarity. If the managers consider a greater similarity between their own firms and the potential competitors, these potential competitors are more likely to be selected into the comp set. Clack and Montgomery (1999) point out there is a five-step procedure that managers will use to select a comp set: (1) form representation of target firms; (2) retrieve "competitor" category representation from memory; (3) assess the similarity of target firms to category representation; (4) classify target firms; (5) store target firm classification in memory. A most recent study (Mohammed et al, 2014) argue that hotel managers will follow a three-step process to identify their competitors: (1) defining the corporate identity of the hotel; (2) scanning the market for potential competitors; (3) matching and choosing hotels with similar corporate identities.

2.1.2 Customer Side

Robert and Latin (1991) claim that customers use a multi-staged process to make decisions, and at each stage, they will reduce available alternatives. Jones and Chen (2010) point out that customers hotel selection follows a two-stage process: (1) form a consideration set; (2) form a smaller choice set. Noone and Robson (2014) also note a similar idea. They find that customers who book hotels online follows two stages: (1) browsing (examining search results); (2) deliberation (click through to selected properties to obtain more information). They also indicate that after deliberation, online customers are also likely to go back to browsing and then deliberation again.

Compare how managers choose their comp sets and how customers select their final hotel, we can clearly see there some similarities and differences. One similarity is that they both follow a multi-stage process to choose a comp set or a choice set. And during the process, they will use some key attributes to form the competitor sets. However, these two processes have a variety of differences and some of the difference is inevitable. Among all the differences, the major one is the attributes used during the processes.

2.2 Selection Attributes

2.2.1 Hotel Side

Determining the right attributes to select the right comp sets is not an easy and straightforward task (Kim and Canina, 2011). Many researchers note some key attributes to select the competitive sets, but none of them are exactly the same (Webb and Schwartz, 2017; Mohammed et al., 2014; Li and Netessine, 2012; Kim and Canina, 2011; Baum and Lant, 2003; Mathews, 2000; Yesawich, 1987). Among all the key attributes, price, location, product type and size are frequently used.

Table 1: Attributes used by hotels in comp set selection process

Author	Year	Attributes
Webb and Schwartz	2017	product class, price, and location, customers' behavior and reveled preferences
Mohammed et al.	2014	product type, price, location, service delivery quality and sales channels
Li and Netessine	2012	price, location, restaurant and room service in hotel, meeting space, complimentary breakfast, loyalty program, full-service amenities, and brand (STR)
Kim and Canina	2011	product type or ADR
Baum and Lant	2003	price, location proximity and size
Mathews	2000	price, proximity and segment
Yesawich	1987	location, price and substitute

Table 2: Comparisons of the mainly used attributes in the comp set selection process

	price	location	size	product type
useful	Yesawich, 1987; Baum and Mezias, 1992; Bull, 1994; Baum and Lant, 2003; Kim and Canina, 2011; Coleman, 2011; Li and Netessine, 2012; Mohammed et al., 2014; Lee, 2014	Yesawich, 1987; Baum and Mezias, 1992; Bull, 1994; Mathews, 2000; Baum and Lant, 2003; Mohammed et al., 2014; Lee, 2014; Ferrer et al., 2018	Baum and Mezias, 1992; Baum and Lant, 2003; Mohammed et al., 2014	Yesawich, 1987; Morgan and Dev, 1994
Limitation	Mathews, 2000; Coleman, 2011	Baum and Lant, 2003; Mohammed et al., 2014	Mathews, 2000	Kim and Canina, 2011
Definition	1. actual rate (Yesawich, 1987; Coleman, 2011; Li and Netessine, 2012) 2. published rate (Baum and Mezias, 1992; Mohammed et al, 2014) 3. ADR (Kim and Canina, 2011)	1. distance between two hotels (Baum and Lant, 2003) 2. market (Li and Netessine, 2012)	1. number of rooms (Mohammed et al., 2014) 2. size of the rooms (Mohammed et al., 2014)	scale (Kim and Canina, 2011)
Method	within +-15% (Yesawich, 1987)	1. 3-mile radius (Canina and Enz, 2006) 2. within a walking distance (Mohammed et al., 2014)	1. larger size (Clark and Montgomery, 1999; Chen andHambrick, 1995) 2. similary size (Mohammed, et al., 2014)	Similarity (Kim and Canina, 2011)
interaction with other variables	1. location (Yesawich, 1987; Lee, 2014) 2. substitute (Yesawich, 1987; 3. product type (Morgan and Dev, 1994; Lee, 2014)	1. product type (Mohammed et al., 2014) 2. customer type (Mohammed et al., 2014)	prodcut type (Baum and Lant, 2003)	price (Mohammed et al., 2014)

Price:

While most of the researchers (Yesawich, 1987; Baum and Mezias, 1992; Bull, 1994; Baum and Lant, 2003; Kim and Canina, 2011; Coleman, 2011; Li and Netessine, 2012; Mohammed et al., 2014; Lee, 2014) argue that price is an essential attribute to select the right competitive set, others (Mathews, 2000; Coleman, 2011) maintain that there is some limitation of using price. Mathews (2000) claim that using price can be problematic due to price discounting. Coleman (2011) also indicates that price alone is not a suitable factor to identify rivals. Although the price is a common conformity factor to select competitive sets, there are many different opinions on the definition of price. The market convention says that when choosing the competitive sets, we should use the published rate (Mohammed et al, 2014). However, Yesawich (1987) claims that actual rates are more accurate than published rates. In addition, Yesawich (1987) suggest that the competitor hotels should have the actual price within ± 15 percent actual rate of your own hotels. However, Mohammed et al. (2014) convey that managers are not able to set the margins to define comparable rates although they believe the price is very critical to select the competitive sets. Baum and Mezias (1992) argue that mid-price hotels are very vulnerable to be considered as competitors by luxury and economy hotels. However, they will not potentially compete with each other. And Mathews (2000) point out that price only cannot identify the right primary competitors. Except for actual rate and published rate, Kim and Canina (2011) conclude that ADR (Average daily rate) will be a useful index to separate the competitor hotels and noncompetitor hotels after using cluster analysis. Nowadays, customers will punish wrong pricing quickly and choose another low-price hotel because the increase in price transparency, indicating that price is an essential factor to identify competitive sets (Li and Netessine, 2012). Therefore, although a variety of researches have focused on room rate, there is still no accordance of how to use price to determine the comp sets.

<u>Location:</u>

Location is a plausible classifier in choosing competitive sets (Yesawich, 1987; Baum and Mezias, 1992; Bull, 1994; Mathews, 2000; Baum and Lant, 2003; Mohammed et al., 2014; Lee, 2014; Ferrer et al., 2018). However, Baum and Lant (2003) state that managers pay too much attention to location. Furthermore, how to define the spatial margin is not clear, the hotel managers in the interview point out the close proximity can be defined as "within a walking distance". But walking distance is another vague definition and different hotel managers and clients will have different understanding (Mohammed et al. 2014). In hotel industry, the convention is 3-mile radius (Canina and Enz, 2006). But Canina and Enz (2006) also claim that depending on the hotel segment and targeted guest, the radius will be varied. With development of internet, hotels' information becomes more transparent, and currently, hotels may want to consider choosing competitors that are located farther away but that offer attractive services and rates. Additionally, leisure travelers are more likely to perceive a wider radius when they think about "within a walk distance" than business travelers (Canina and Enz, 2006; Mohammed et al. 2014). Mohammed et al. (2014) also point out that hotel clients will consider a wider radius than managers.

<u>Size:</u>

While some former studies have showed that size is a vital classification variable for identifying competitors in hotel industry (Baum and Mezias, 1992; Baum and Lant, 2003; Mohammed et al., 2014), Mathews (2000) point out that size is not important to rivalry as thought. Moreover, defining how will hotels compete with rivals according the size are not easy (Mohammed et al, 2014). There are different opinions. Some writers argue that companies will compete with largest firms in their industry (Clark and Montgomery, 1999; Chen and Hambrick, 1995). Alternatively, others claim that similarly-sized companies are the most direct competitors (Porter,1979).

<u>Product type:</u>

Another commonly used characteristic is product type. The definition of product type varies among researchers. A common way is scale/class. The highest four quality scale, including luxury, upper upscale, upscale and upper midscale are more likely to select less quality properties in their competitive sets, while midscale and economy will select higher quality hotels (Smith and Zheng, 2011). Mohammed et al. (2014) maintain that higher quality hotels will compete a wider radius than lower quality hotels. The reason is the luxury hotels may have fewer proximate competitor hotels.

2.2.2 Customer Side

According to the previous research, hotel consumers pay attention to the following attributes: price, location, facilities, size, service and image. However, in this study, we mainly care about the attributes they use to select the choice set instead of the attributes customers used for the final hotel decision. Noone and Robson (2014) and UNWTO et al. (2015) point out that the attributes that customers used in the two stage of hotel selection is different. Therefore, we mainly focus on the attributes used for choice set selection.

Jones and Chen (2012) maintain that online customers will use product type, facilities, price to from the consideration set use reviews, picture, star-rating and price to form the choice set. and Noone and Robson (2014) indicate that online hotel consumers will use firm-supplied information (hotel name, images, price and location) and user ratings to select the choice set. And UNWTO et al. (2015) official hotel classifications are used as a filter mechanism to select the choice set.

Table 3: Attributes used by customers in the hotel competitor's selection process

Author	Year	Attributes
Lien et al.	2015	Brand image, perceived price, and perceived value
Mohammed et al.	2014	value for money, facilities, star rating and image, in-house benefits and online reviewer perceptions.
Liu, Law, et al., 2013	2013	Cleanliness, Location, Room, Service, Sleep Quality, Value
Bjorkelund, Burnett, & Norvag	2012	Shabby Bed, Clean Rats, Friendly Staff, Limited Parking, Good Room
Ariffin & Maghzi	2012	Personalization, Warm Welcome, Special Relationship, Straight from the Heart, Comfort
Sohrabi et al.	2012	Promenade and Comfort, Security and Protection, Network Services, Pleasure, Hotel Staff and Their Services, News and Recreational Information, Cleanliness and Room Comfort, Expenditure, Room Facilities, Parking
Kim and Canina	2011	usage, brands, preferences, and information
Jones and Chen	2011	non-smoking, swimming pool, high-speed internet, hot tub, fitness center, room service, price range, picture,reviews, star-ratings
Merlo &de Souza Joao, 2011	2011	Location, Size and Diversity, Characteristics of the Lobby, Characteristics of the Rooms, Parking
Albaladejo-Pina& Diaz-Delfa	2009	Type of Building, Location, Number of Bedrooms, Price per Room, Horses for Hire, Play Area, Meal Service, Swimming Pool, Sports Facilities, Mini-Farm, Bathroom, Type of Rent, 'Q' Quality Award, Booking
Hsieh, Lin, & Lin	2008	Problem-Solving Abilities by Service Personnel, Price Level, Sanitary Hot Spring Environment, Convenience of Traffic Route/Shuttle, Special Promotions, Convenience of Reservation Procedure, Food and Beverages Service
Lockyer	2005	Location, Price, Facilities, Cleanliness
Choi & Chu	2000	Staff Service Quality, Room Quality, General Amenities, Business Services, Value, Security, IDD Facilities
Ananth, DeMicco, Moreo, & Howey	1992	Cleanliness, Location, Room Rate, Security, Service Quality, Reputation of Hotel

2.3 Match and Mismatch

Mohammed et al. (2014) find out that eight out of 11 competitors (72.7%) are the hotels that both hotel managers and customers are considered as competitors. However, Li and Netessine (2012) argue that according to their research, the overall overlap rate is 49.5%.

Li and Netessine (2012) point out that independent hotels, hotels from different district, hotels with lower traveler reviews are more likely to be left out. Besides, they notice that hotels tend to compare with lower star ratings, lower price levels, and lower ranks hotels. In the case study, Mohammed conducted in 2014, hotel managers will choose five direct competitor hotels and six indirect competitors. However, hotel guests are only able to recall 2.33 hotels on average, and the standard deviation is 1.08. Furthermore, hotel managers utilize five attributes: price, product offering, location proximity, size and segment to choose the completive sets, and they will only choose three out of five to determine the competitive sets in practice. In comparison with managers, consumers mainly focus on products, process, value for money, service quality facilities, star rating, proximity, image, perceptions by online reviews and in-house benefits. In general, both hotel managers and consumers will use proximity, products and price to identify the competitors (Mohammed et al., 2014). In addition, Mohammed et al. (2014) argue that hotel guests will perceive a wider scope of distance than managers.

CHAPTER 3

3 DATA SAMPLE

We obtain two datasets to compare the difference between hotels' and hotel customers' perspectives of competitors sets. While we use Smith Travel Research's data to stand for hotels' perspective, we adopt the data from TripAdvisor to represent customers' opinions of competitive set. Furthermore, as our main topic is to contract the difference, it is required to join the two datasets together, so we use a shared ID to combine the STR and TripAdvisor data. That is, each hotel will have a unique ID and both the two datasets share this unique ID. Therefore, joining the tables together becomes much easier. After we merge the tables, contracting the different views of hotels and customers is more achievable. Before conducting further analyses, it is critical to understand the data and use descriptive analyses to summarize and reflect the main features of the data. The following two sessions will elaborate these two datasets clearer.

3.1 STR data

The STR data contains three types of dataset. The first dataset stores in a spreadsheet that contains the competitive set information for each subject hotel. The second dataset incorporates the features of each hotel (both subject hotels and competitor hotels) including: country, market, chain, owner company, management company, parent company, scale, operation, class, location, size code and open date. The third dataset is composed by the daily occupancy, ADR (average daily rate), local ADR, RevPAR (revenue per available room) and local RevPAR from 2010/01/01 to 2016/12/31 for each hotel.

We discuss the first dataset first. These data store in a pairwise format, each subject hotel will select a list of hotels as competitors and the following table discloses some information of the data.

Table 4: Summary of the first STR dataset

# records	9824
# unique subject hotels	1913
# unique competitor hotels	2150
Minimum rivals in a comp set	1
Maximum rivals in a comp set	23
Average rivals in a comp set	5.64
Average Name back ratio	0.35

The second dataset consists of 2349 unique hotels' information. According to the descriptive table of the second dataset, over sixty of hotels come from the United States; over fifty hotels are from Urban area; over fifty hotels have rooms between 75 and 299.

Table 5: Summary of the second STR dataset

Variable Name	**Category Levels / Range**
Country	6
Market	12
Chain	239
Owner company	567
Management company	452
Parent company	122
Operation	3
Scale	7
Class	6
Location	5
SizeCode	5
OpenDate	[1753, 2017]

Table 6: Summary for each variable[1]

Country	Market	Operation	Scale[2]	Class[3]	SizeCode[4]	Location
United States (1454; 61.9%);	Dallas, TX (338; 14.4%);	1: Chain Owned and/or Managed (858; 36.5%);	1: Luxury (139; 5.92%);	1: Luxury (243; 10.3%);	1: <75 (388; 16.5%);	*1: Urban (1292; 55%)*
Canada (78; 3.32%);	New York, NY (328; 14%);	2: Franchised (1008; 42.9%);	2: Upper Upscale (335; 14.3%);	2: Upper Upscale (451; 19.2%);	2: 75-149 (918; 39.1%);	*2: Suburban (666; 28.4%)*
United Kingdom (571; 24.3%);	Boston, MA (100; 4.26%);	3: Independent (483; 20.6%)	3: Upscale (483; 20.6%);	3: Upscale (600; 25.5%);	3: 150-299 (684; 29.1%);	*3: Airport (187; 7.96%)*
Finland (55; 2.34%);	Orlando, FL (299; 12.7%);		4: Upper Midscale (370; 15.8%);	4: Upper Midscale (442; 18.8%);	4: 300-500 (218; 9.28%);	*4: Resort (186; 7.92%)*
Australia (153; 6.51%);	Denver, CO (216; 9.2%);		5: Midscale (185; 7.88%);	5: Midscale (215; 9.15%);	5: >500 (141; 6%)	*5: Small Metro/Town (4; 0.17%)*
Poland (38; 1.62%)	San Francisco /San Mateo, CA (154; 6.56%);		6: Economy (355; 15.1%);	6: Economy (398; 16.9%)		
	Columbus, OH (19; 0.809%);		7: Independents (482; 20.5%)			
	London (571; 24.3%);					
	Edmonton (78; 3.32%);					
	Sydney (153; 6.51%);					
	Helsinki (55; 2.34%);					
	Warsaw (38; 1.62%)					

[1] The definitions of attributes are from STR and the number in the brackets are the number of hotels and percentage in that category.
[2] Scale Segments are a method by which branded hotels are grouped based on the actual average room rates. Independent hotels, regardless of their average room rates, are included as a separate Chain Scale category.
[3] Class is an industry categorization which includes chain-affiliated and independent hotels. The class for a chain-affiliated hotel is the same as its Chain Scale. An independent hotel is assigned a class based on its average daily rate (ADR), relative to that of the chain-affiliated hotels in its geographic proximity.
[4] SizeCode is defined by number of rooms a hotel possesses.

The third STR dataset consists of 2237 unique hotels and the table below gives the observation numbers, minimum, maximum, mean, standard deviation for each variable. There is no missing value in this dataset. The average hotels' ADR and RevPAR are higher than local. And the average occupancy during that period is 75.91%.

Table 7: Summary of the third STR dataset

Variable Name	Min	Max	Mean	SD
ADR	29.43	1356.85	162.03	123.16
Local ADR	27.65	972.85	149.12	106.30
RevPAR	2.71	1018.57	128.76	99.67
Local RevPAR	2.71	680.57	117.99	86.39
Occupancy	4.30	98.88	75.91	12.26

3.2 TripAdvisor data

TripAdvisor data was extracted from December 2017 and contains 30669 records with 1551 unique subject hotels (s_hotel) and 1847 unique competitor hotels (c_hotel) and 27 variables, including property name, address, country and city, etc. If the variable start with s, then it describes the subject property. And if the variable start with c, then it describes the competitor property.

According to the definition of attributes in TripAdvisor data, there are two types of data. One is measurement of customers' view of similar or competitive hotels. The second set of data focuses on the characteristics of the hotels. We will first focus on the attributes of the hotels.

3.2.1 Hotel Attributes

First and foremost, we analyze the geographic data to see whether they follow the similar pattern as the STR data, and we get the following table. The number of levels of country and city for subject hotel are the same as Smith Travel Research's data. However, the levels for competitor hotels' country and city are higher. Some of cities may stand for the same destination, such as San Francisco and South San Francisco, but the findings still imply that customers will consider a wider range of locations than hotels.

Table 8: Summary of geographic data

Variable Name	Levels
Subject hotel country	6
Subject hotel city	12
Competitor hotel country	9
Competitor hotel city	132

The following table shows that there is a little difference between STR data and TripAdvisor data. And the difference is due to the fact that we only get a subset of subject hotels from TripAdvisor compare to STR. We will handle this problem later.

Table 9: Summary of TripAdvisor's geographic attributes

Subject hotel country	Subject hotel city
United States (904; 58.3%);	Boston (64; 4.13%);
Canada (45; 2.9%);	Dallas (132; 8.51%);
United Kingdom (454; 29.3%)	Denver (95; 6.13%);
Finland (40; 2.58%);	Dublin (19; 1.23%);
Australia (85; 5.48%);	Edmonton (45; 2.9%)
Poland (23; 1.48%)	Helsinki (40; 2.58%)
	London (454; 29.3%)
	Helsinki (40; 2.58%)
	London (454; 29.3%)
	New York City (284; 18.3%)
	Orlando (201; 13%)
	San Francisco (109; 70.3%)
	Sydney (85; 54.8%)
	Warsaw (23; 14.8%)

Secondly, we summarize the average night rate in that month, average score given in reviews in that month, average rank in that month and number of reviews received in the past month we get the following summary table. According to the definition and analysis, the null value for score and reviews mean that the hotel didn't receive reviews during that month. The null value for rank is whether that hotel does not have a rank in TripAdvisor, or list in another two categories: B&B and Inns or Specialty Lodging. The null value for night rate is the hotel that does not list price on TripAdvisor. We will handle with these null values later.

Table 10: Summary of TripAdvisor's hotel attributes

Variables	Min	Max	mean	SD	median	#null
s_night_rate (USD)	42.68	946.14	205.98	116.32	180.12	700
c_night_rate (USD)	22.98	946.14	230.42	134.79	199.99	36
s_num_reviews	1	330	18.93	21.61	13	1977
c_num_reviews	1	318	30.31	34.91	21	710
s_rank_percentile	0.093	100	37.70	23.50	34.95	64
c_rank_percentile	0.093	100	27.49	22.17	22.35	71
s_avg_score	1	5	4.04	0.66	4.17	1977
c_avg_score	1	5	4.17	0.59	4.29	710

3.2.2 Customers competition identification measurement attributes

The second type of data in the TripAdvisor dataset are measures of customer activity across subject-competitor hotel pairs. Common sessions are defined as the number of times a customer views two (subject-competitor) hotels during a single visit to TripAdvisor. S_same_session_pvs measures the number of subject property pages viewed when the property is viewed in the same session as the competitor property while c_same_session_pvs means the number of competitor property pages viewed when viewed in the same session as the subject property. S_same_session_click measures the number of clicks the subject property received when viewed in the same session as the competitor while c_same_session_click measures the number of clicks the competitor property received when viewed in the same session as the subject property. S_uniques and c_uniques measure the number of unique viewed property. S_total_click and c_total_click means the total number of clicks received. Table 11 shows summary data for these five types or measurement.

Table 11: Summary of competition identification measurement attributes

Variables	Min	Max	Mean	Median	SD
common sessions	1	4536	31.4	10	107.45
s_same_session_pvs	0	55184	279	72	1114.96
c_same_session_pvs	0	55184	287.3	76	1127.17
s_same_session_clicks	0	1456	5.093	1	25.17
c_same_session_clicks	0	1456	5.474	1	25.36
s_uniques	68	550543	29883	18843	37726.46
c_uniques	140	550543	63320	38656	77776.64
s_total_clicks	0	16117	603.7	314	971.45
c_total_clicks	0	16117	1524	726	2290.31

CHAPTER 4

4 RESULTS

We develop our analyses in the following way. In the first part, we compare competitive set identified by hotels with those selected by customers and find the hotel attributes associated with differences across these two views by comparing data from STR and TripAdvisor. In the second part, we analyze how customers choose hotel competitors based solely on data from TripAdvisor. In the third part, we compare parts 1 and 2.

4.1 Contrast hotels' and customers' view of competitive sets

In this section we assess the similarity of a hotel and customer view of 'competing' properties. We define two concepts: Match and Mismatch. Match means both the hotels and customers consider the hotels as competitors while Mismatch means only hotels or customers consider these hotels as competitors. We extract the competitor hotels for each subject hotel that exists in both datasets, and we consider these competitor hotels as matched hotels. Then we find the competitor hotels for each subject hotel that only exists in the STR or TripAdvisor dataset (mismatched hotels). Then we utilize hotel attributes of interest to drive some insights and analyze the matched and mismatched groups. Namely, we utilize descriptive statistics and figures to assess whether these attributes were associated with match and mismatch. Next, we take advantage of multiple logistic regression model to detect whether these influenced attributes are associated with the mismatch rate.

To begin with, we calculate the matched rate of hotel pairs. We use equation 4.1.1 to calculate the match rate – resulting in an average match rate of 42.3% across our sample. This number is lower than the matched rate found by Li and Netessine (2012) and Mohammed et al. (2014). We then separate the hotels by scale and operation type. We find that for the branded hotel, the match rate increase for higher scale, indicating that higher scale hotels' competitive sets views are more similar to customers'. Additionally, luxury, upper upscale and upscale hotels tended to have a larger size competitive sets than upper midscale, midscale and economy hotels. And this conclusion was the same as Smith and Zheng (2011). Furthermore, we find that chain managed hotels are more likely to select the same competitor hotels as customers than franchised. Therefore, chain managed hotels have a better competitive set management.

$$Matched\ rate = \frac{Number\ of\ matched\ hotels\ pairs}{Total\ hotel\ pairs\ in\ STR} \quad 4.1.1$$

Table 12: Match rate by subject hotel' scale

Scale	Matched rate	Avg. # comp
Luxury	54.0%	5.31
Upper Upscale	49.7%	5.41
Upscale	40.2%	5.54
Upper Midscale	40.0%	5.16
Midscale	37.2%	5.04
Economy	29.5%	4.86
Independents	45.8%	5.52

Table 13: Match rate by subject hotel' operation

Operation	Match Rate	Avg. # comp
Chain Owned and/or Managed	43.4%	5.17
Franchised	39.7%	5.32
Independent	45.8%	5.52

We further refine our measures of Match / Mismatch by focusses on subject hotels belonging to STR (STR mismatch rate) separately from those from TripAdvisor (TripAdvisor mismatch rate) as following:

$$STR\ Mismatch\ rate = 1 - \frac{Number\ of\ matched\ hotels\ pairs}{Total\ number\ of\ hotels\ pairs\ in\ STR} \quad 4.1.2$$

$$TripAdvisor\ Mismatch\ rate = 1 - \frac{Number\ of\ matched\ hotels\ pairs}{Total\ number\ of\ hotels\ pairs\ in\ TripAdvisor} \quad 4.1.3$$

On the basis of previous research, we conclude that the following independent variables were related to the competitive selection: ADR, rate, rank, score, size, class, scale, operation and location. These attributes, in our datasets, are stored in a pairwise format, which contained the information for both subject and competitor hotels. Therefore, it is necessary to find a way to measure the similarity and dissimilarity of these attributes.

According to the properties of different types of attributes (nominal, ordinal or quantitative), we adopt three different methods. For the numerical variables, such as ADR, rank and score, we compare the attributes of subject hotels and competitor hotels by obtaining new variables which are computed by the difference divided by the mean value of the two attributes value. A positive number indicates a higher value for the subject property. For example, the new variable sADR is calculated by the following formula.

$$sADR = \frac{subject\ hotel\ ADR - competitor\ hotel\ ADR}{(subject\ hotel\ ADR + competitor\ hotel\ ADR)/2} \qquad 4.1.4$$

For the ordinal categorical variables, such as size and class, we take the difference of the attributes between subject hotels and competitor hotels. For instance, we calculate the new variable size by the following formula. A positive number indicates a higher value for the subject property.

$$sSize = subject\ hotel\ SizeCode - competitor\ hotel\ SizeCode \qquad 4.1.5$$

Finally, we combine chain managed and franchised hotels as branded hotel and created four levels for operation type, including branded vs independent (BI), branded vs branded (BB), independent vs branded (IB) and independent vs independent (II). In each of these pairs, the first letter represents subject hotels and the second represents competitor hotels. The distance between the hotel pairs range from 3 meters to 16050967 meters. However, customers perceive distance differently for different cities. For example, customers will accept longer distance for Orlando than for New York City. Therefore, we standardize the distance by the subject hotels' city.

Based on our research purpose, we define a binary variable as the response variable, which contain two levels: match (the pair[5] exists on both TripAdvisor and STR datasets) and mismatch (the pair only exists in the STR or TripAdvisor dataset). Therefore, we examine the relationship between this binary variable with the newly defined variables as mentioned earlier.

Table 14: List of Variables of Interest

Data Sources	Variable types	Variable Names	Category Level/ Range	New Scaled Variable Names	Generate Method
STR	Categorical Variable	Class	6	sClass	4.1.2
		SizeCode	5	sSize	4.1.2
		Operation	3	sOperation	9 levels
	Numerical Variable	ADR	[29.43, 1356.85]	sADR	4.1.1
TripAdvisor	Numerical Variable	Rank Percentile	[0.093,100]	sRank	4.1.1
		Score	[1,5]	sScore	4.1.1
		Night rate (USD)	[22.98, 946.14]	sRate	4.1.1
		Distance (meters)	[3, 16050967]	sDistance	Scaled by market

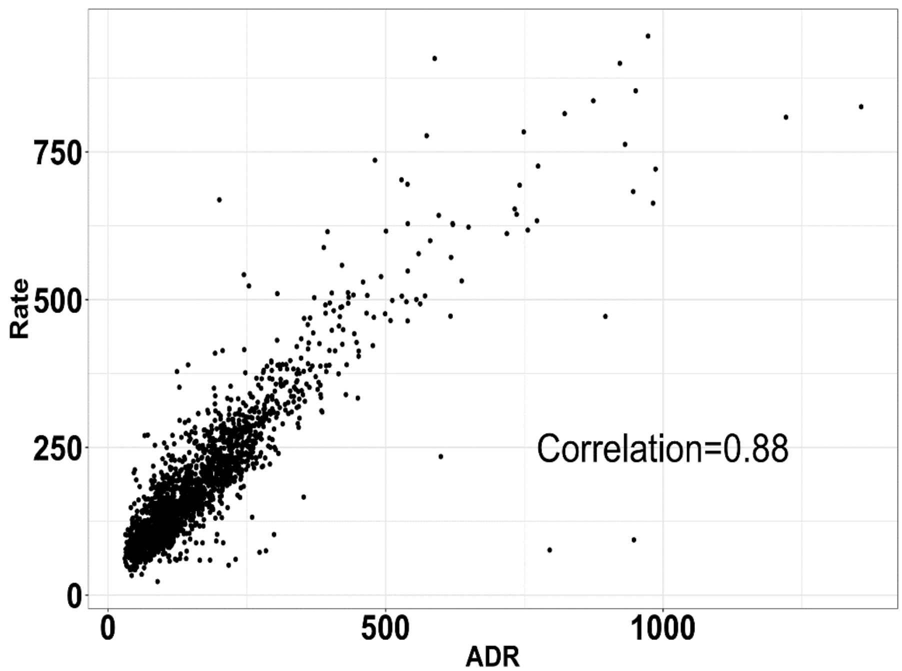
4.1.1 Price: ADR and rate

Literature has determined that hotels' room price is an essential attribute for both hotels and customers when selecting the rivals. However, the lodging industry has no convention for which type of price is the most suitable one. Therefore, we analyze both ADR from STR and rate from TripAdvisor website to find the more appropriate one. The Pearson correlation between ADR and rate is 0.88, indicating that they are similar. We determine that ADR performs better than rate because of the following reasons. To

[5] The pair means the record which contains the subject hotel and the competitor hotel that subject hotel choose.

begin with, ADR and rate both indicate hotels' actual night rate. However, the rate was extracted from the TripAdvisor website in December 2017. In contrast, ADR was extracted from 2010/01/01 to 2016/12/31 and contained average daily rate from all channels. Therefore, ADR includes more comprehensive information about hotels' room price than does rate. Hence, we adopt ADR to represent the hotels' room price and analyze the relationship between sADR and the response binary variable.

Figure 1: Correlation between TripAdvisor night rate and ADR



We examine whether matched and mismatched groups have the same ADR correlation. More specifically, we calculate the Pearson correlation of ADR between subject hotels and the competitor hotels they identified, and we utilize the Student's t-test to test whether ADR correlation are the same for both sets of groups. The p-value is $3.227e^{-13}$, indicating that there is a significant difference of ADR correlation between matched and mismatched groups. These results imply that price is perceived differently by hotels and customers. The estimated mean value of ADR correlation for the matched group is 0.70 and for the mismatched group is 0.67. It seems that hotels are more likely to select rivals with similar room price than customers. In order to verify this idea, we conduct further analyses.

We then plot the new defined sADR against the mismatch rate, which is calculated by the number of mismatched pairs over the total number of pairs. We find that when the difference between subject hotels and competitor hotels increase, the mismatch rate increases. Therefore, these results also imply that hotels and customers tend to select competitor hotels with similar ADR.

Figure 2: Relationship between sADR and STR mismatch rate

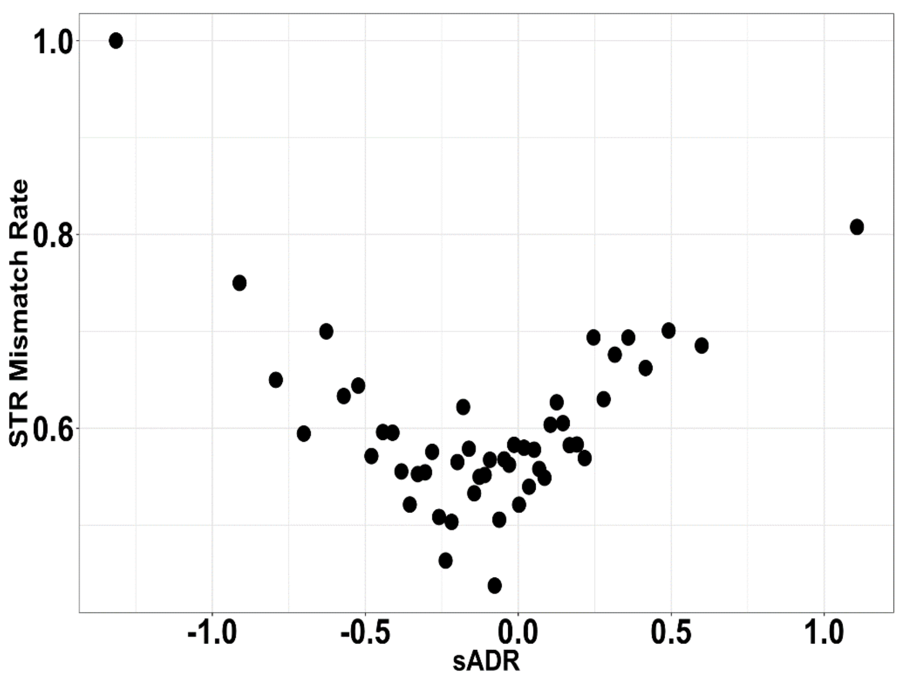
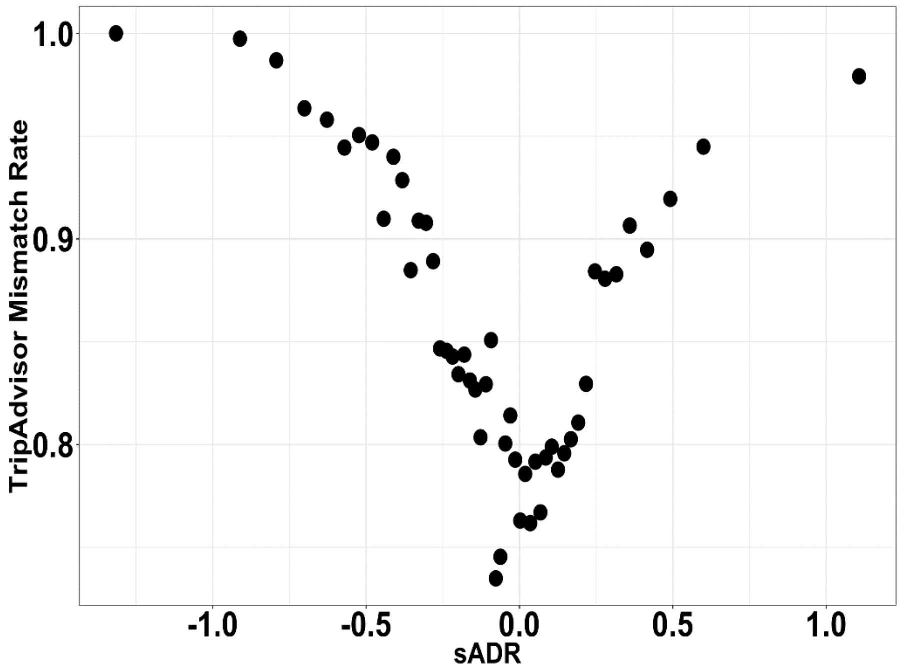


Figure 3: Relationship between sADR and TripAdvisor mismatch rate



4.1.2 Rank

Figure 4 and Figure 5 show that sRank has a negative linear relationship with STR mismatch rate but has a positive linear relationship with TripAdvisor mismatch rate, indicating that hotels tend to select lower TripAdvisor rank rivals while customers prefer to choose higher rank hotels as competitors.

Figure 4: Relationship between sRank and STR mismatch rate

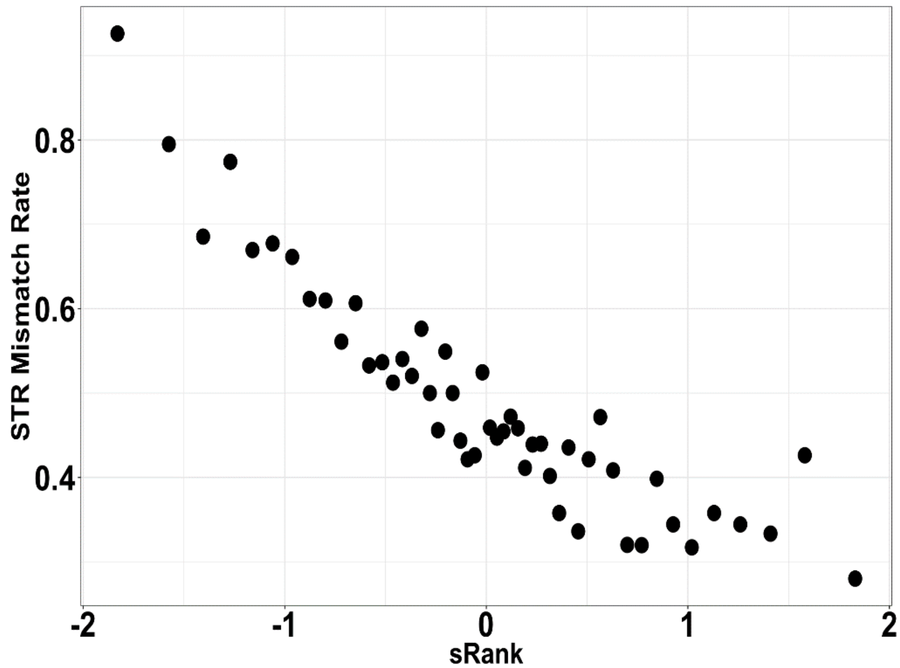
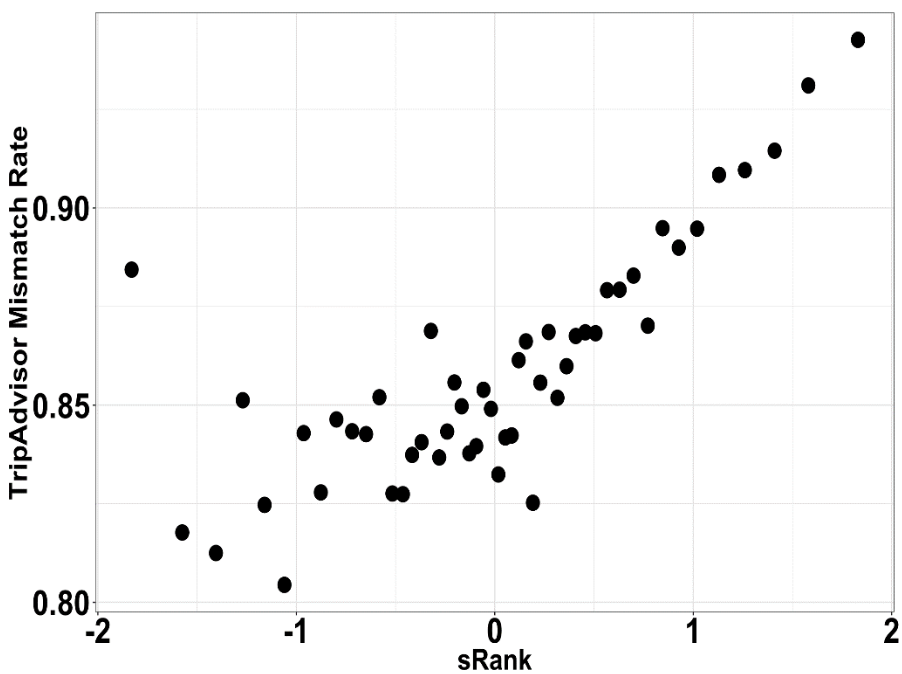


Figure 5: Relationship between sRank and TripAdvisor mismatch rate



4.1.3 Score

Figure 6 and Figure 7 show an opposite direction of sScore and mismatch rate, reflecting that hotels tend to select the lower TripAdvisor score rivals while customers prefer to select higher. Though the TripAdvisor score in our data is limited to a single month (December 2017), these scores are still meaningful.

Figure 6: Relationship between sScore and STR mismatch rate

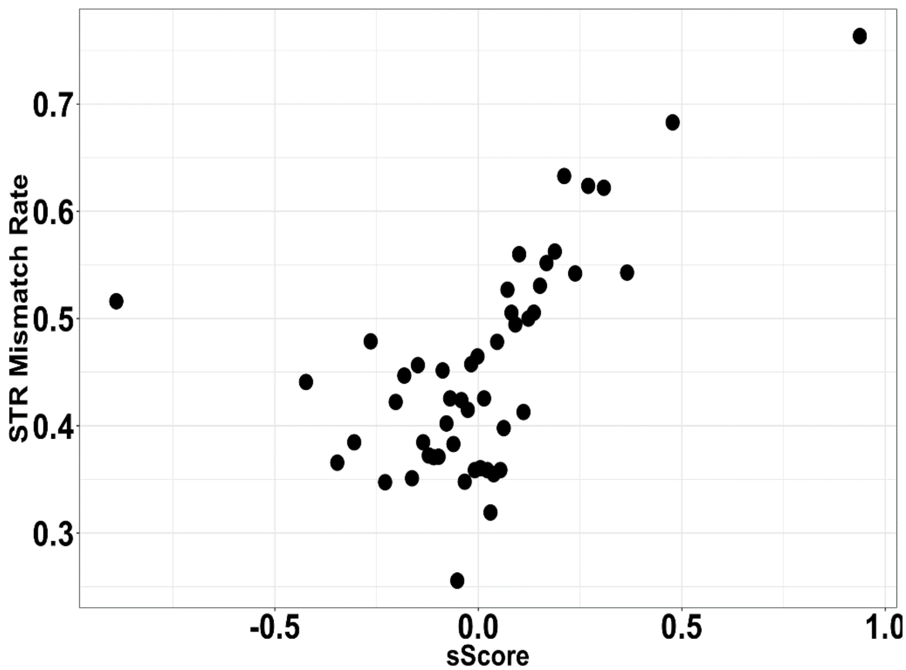
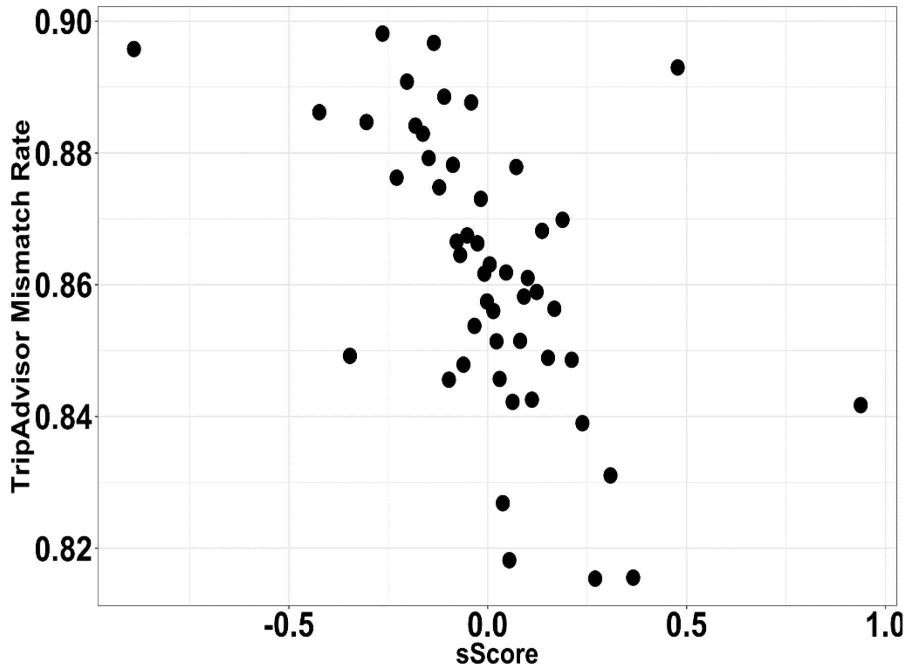


Figure 7: Relationship between sScore and TripAdvisor mismatch rate



4.1.4 Size

For the sSize, we found that the mismatch rate increase when sSize increase, indicating that hotels prefer to select smaller size hotels as competitors. These results contradict previous findings which claimed that companies tend to pick larger size competitors (Clark and Montgomery, 1999; Chen and Hambrick, 1995) or similar size hotels (Baum and Lant, 2003). However, customers tend to select different size hotels as competitors, indicating that compare to customers, hotel still tend to select similar size hotels.

Figure 8: Relationship between sSize and STR mismatch rate

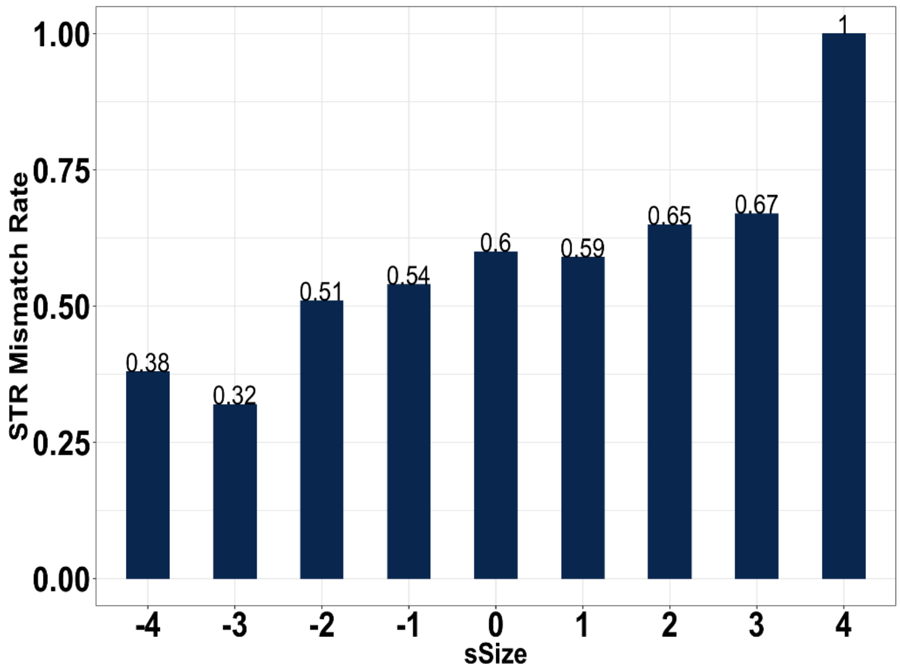
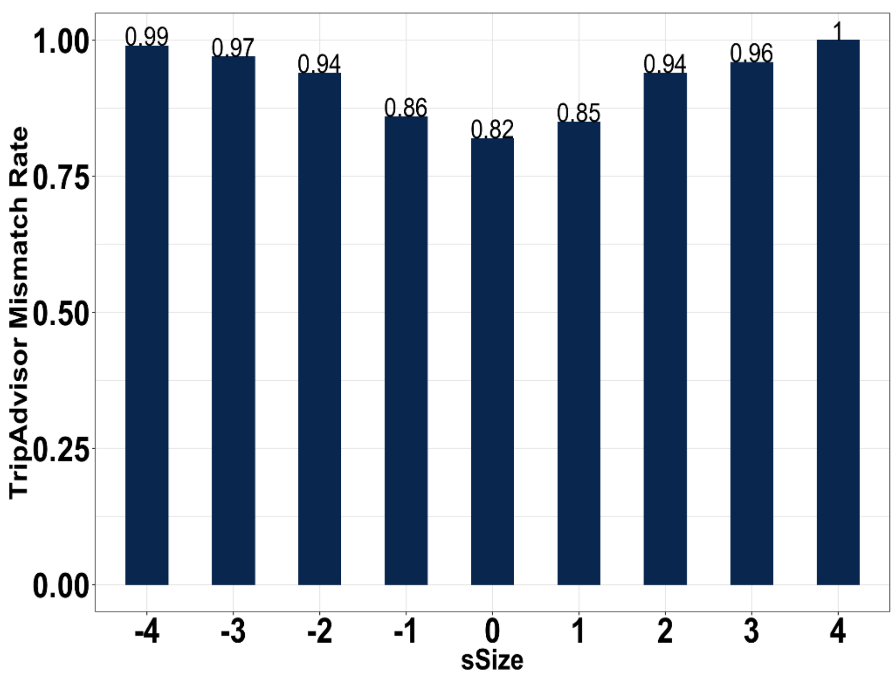


Figure 9: Relationship between sSize and TripAdvisor mismatch rate



4.1.5 Class and scale

According to STR's definitions of class and scale, these two measurements are very similar except that class includes the independent hotels while scale does not. Therefore, we only include sClass. Figure 10 indicates that when sClass decrease, the STR mismatch rate increase. Therefore, it seems that hotels are more likely to select the lower chain scale than are customers. Figure 11 exhibits that customers prefer to choose different class hotels as competitors.

Figure 10: Relationship between sClass and STR mismatch rate

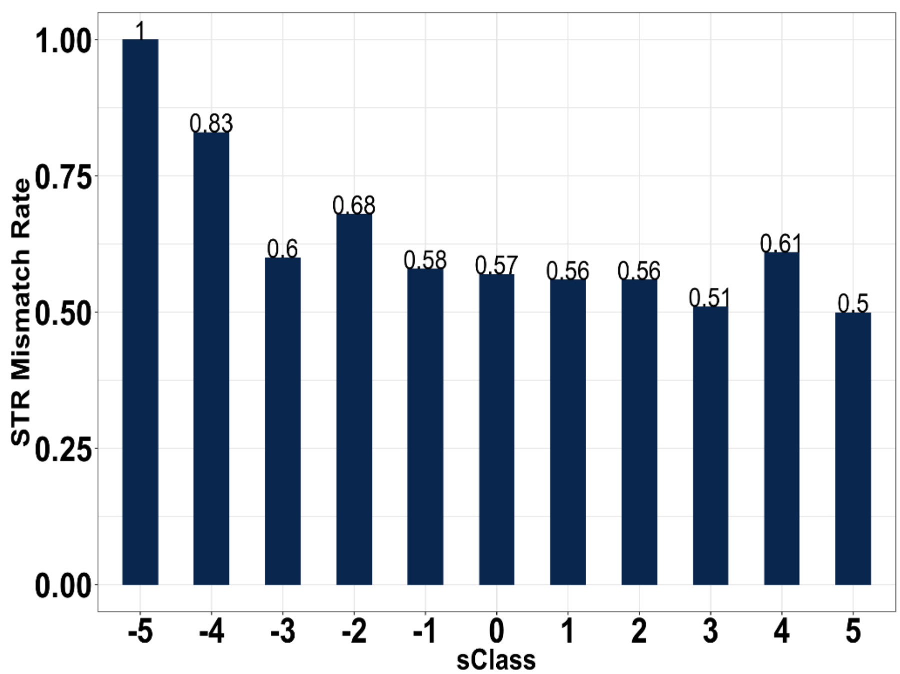
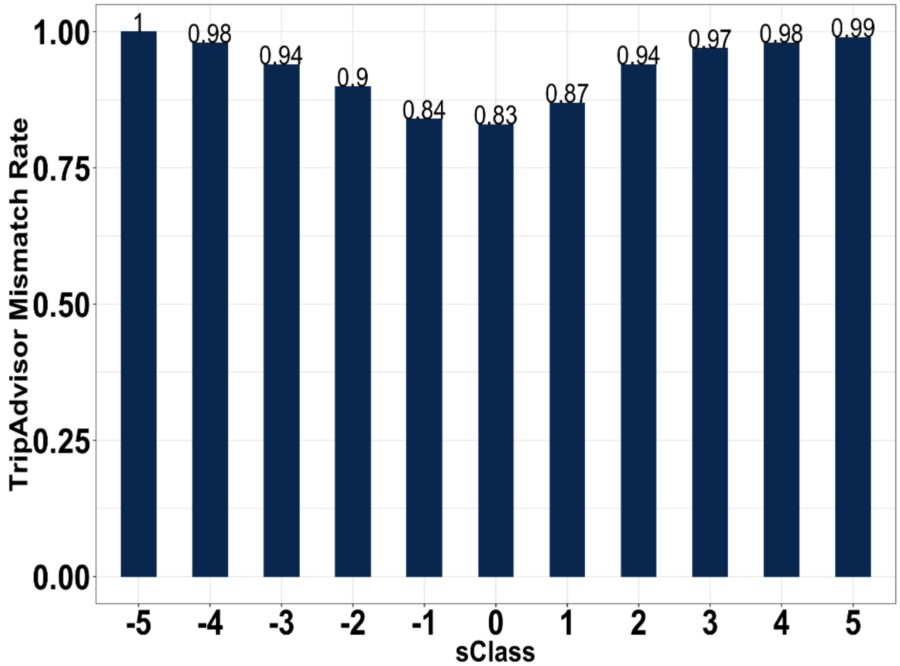


Figure 11: Relationship between sClass and TripAdvisor mismatch rate



4.1.6 Operation Type

We are also interested in whether hotels and customers view operation type differently. First, we test the mismatch rate of two groups—subject hotels and competitor hotels belong to the same operation type, and they belong to different operation type. As the left graph shown, the mismatch rates for the two groups are 0.58 and 0.57, indicating that there is no difference between the two groups. Therefore, we utilize the Students' t-test, the p-value is 0.22 which also proves our conclusion. Second, we involve further analyses and divide the data into four groups (BB, BI, IB, II) discussed above. Although the pattern is not clear, there appears some difference between the four groups. The p-value of the ANOVA test, which tested whether the means of the four groups are the same or not, is 6.265e-15. That number illustrates that these four groups are different.

Based on the right-side plot below, we find that although the pattern is not clear, customers and hotels view operation type differently when choosing the comp set.

Figure 12: Relationship between sOperation and STR mismatch rate

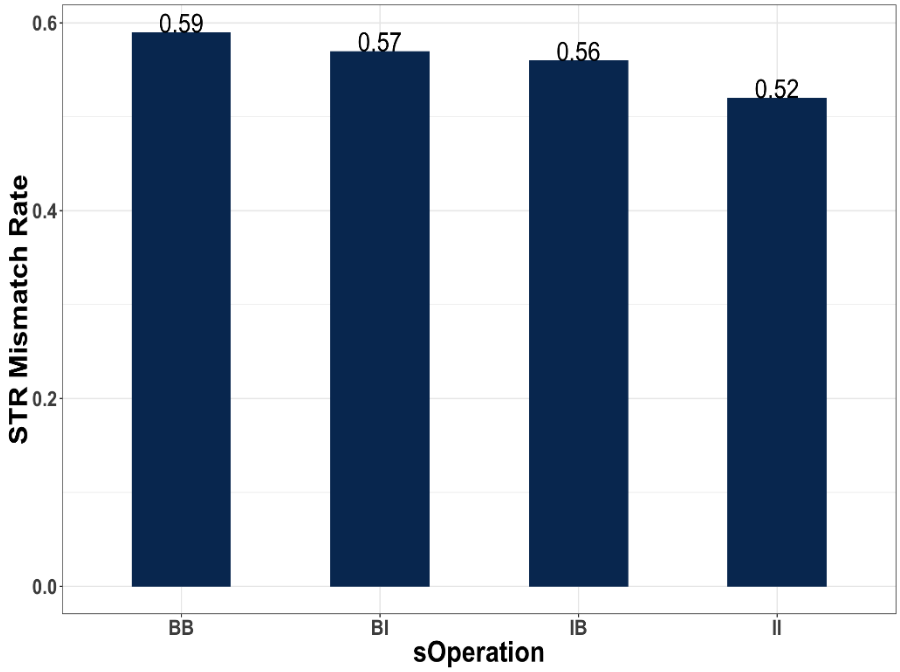
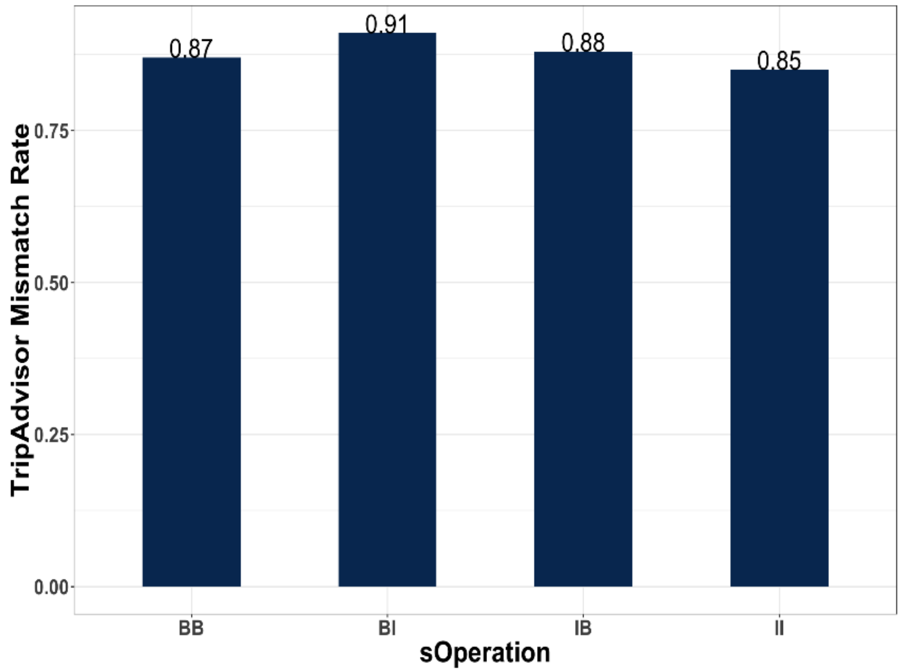


Figure 13: Relationship between sOperation and TripAdvisor mismatch rate



4.1.7 Interaction of the attributes

The graphs below have shown some evidence of the interaction between rank, ADR and score. Every chart includes three colors of points. The black points incorporate all the data points, while the gray ones and light gray ones only contain a subset of the data points. Specifically, gray points mean that the ADR of competitor hotels are lower than that of subject hotels. On the contrary, the light gray ones mean that the ADR of competitor hotels are higher than that of subject hotels.

Figure 14: Interaction between sScore and sADR on STR mismatch rate

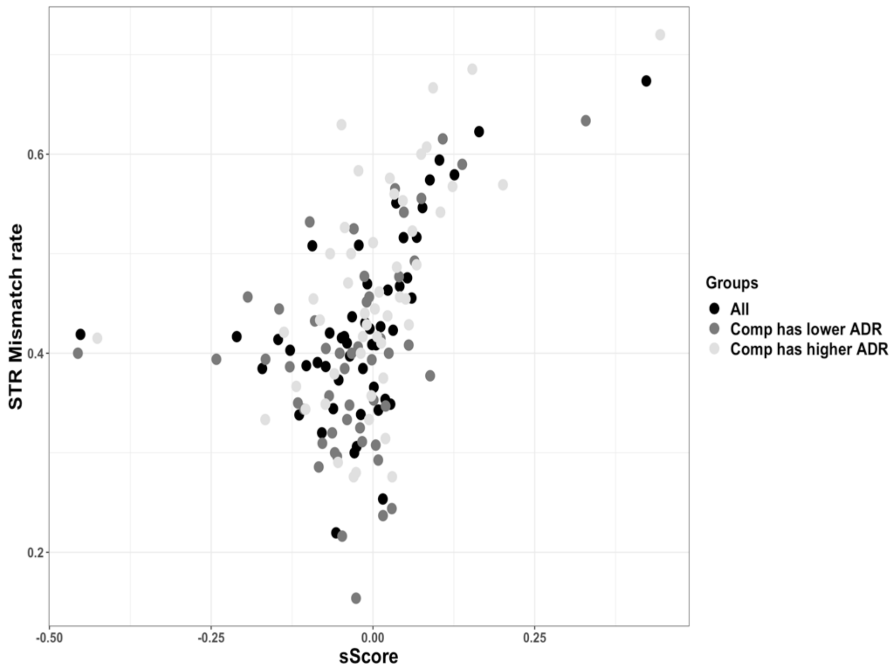


Figure 15: Interaction between sScore and sADR on TripAdvisor mismatch rate

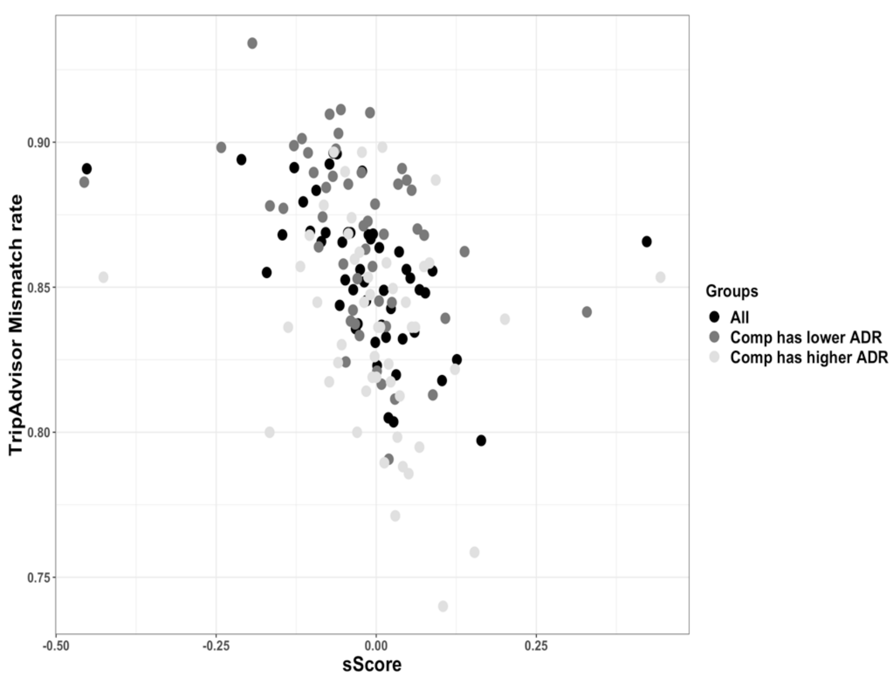


Figure 16: Interaction between sRank and sADR on STR mismatch rate

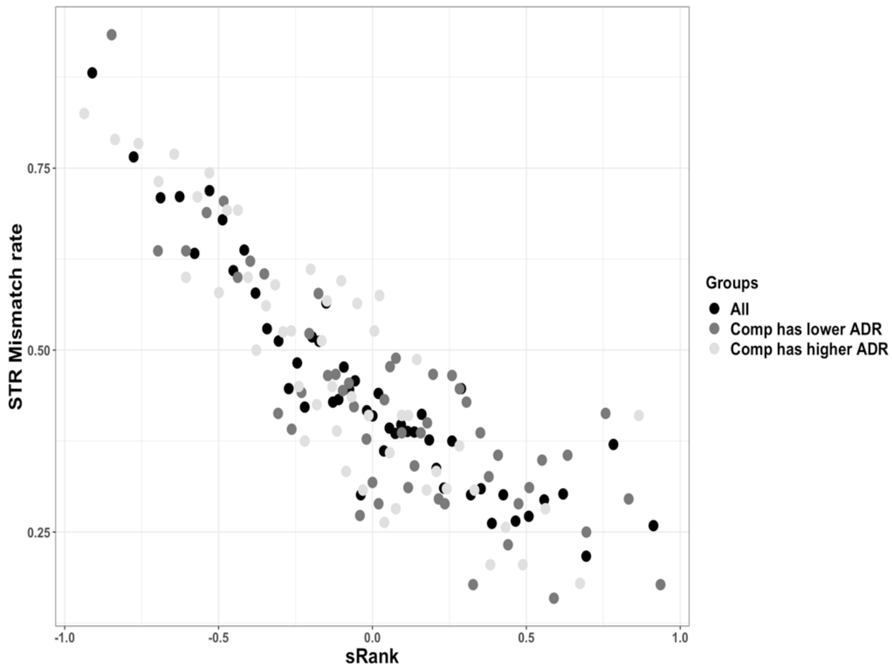
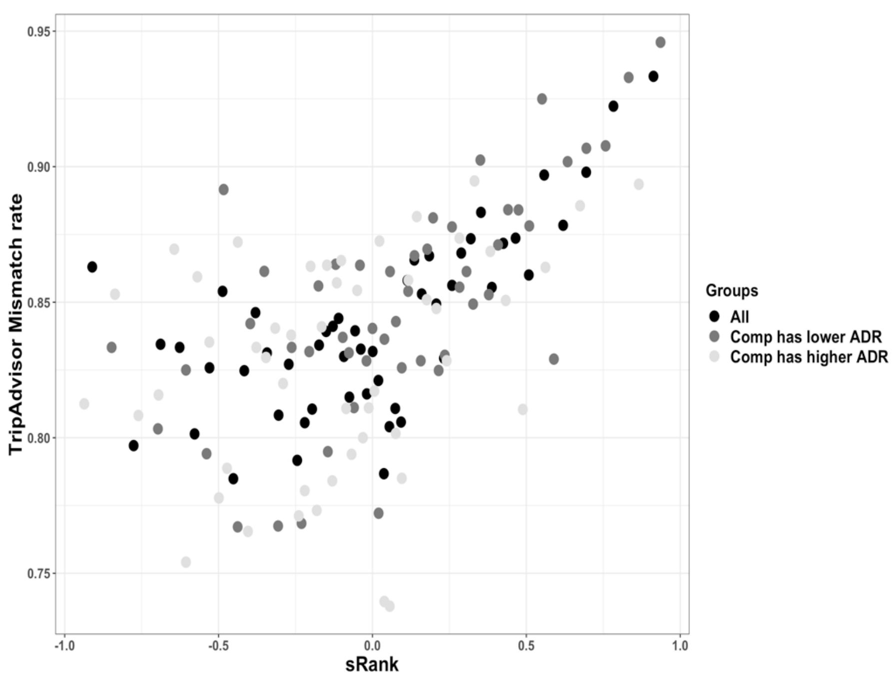


Figure 17: Interaction between sRank and sADR on TripAdvisor mismatch rate



4.1.8 Multiple logistic regressions

Based on the attributes analyses above, the mismatch rate is influenced by ADR, class, size, rank, score. The multiple logistic regression results in Table 15 and Table 16 verify these relationships. In the two models, all the numerical predictors are standardized and interaction between sScore and sADR is included.

For sScore and sADR, a positive number indicates a higher value for the subject property. Figure 18 shows the net effects of the interaction between TripAdvisor score and ADR. The figure indicates the strong interaction between score and ADR especially for competitor hotels of lower ADR.

Table 15: Model with interaction for STR mismatch

Parameters	Estimate	Std. Error	t-Value	P value	Significance
(Intercept)	-0.295	0.038	-7.675	1.66E-14	***
sADR	0.009	0.038	0.225	0.822161	
Squared sADR	0.077	0.021	3.724	0.000196	***
sRank	-0.554	0.039	-14.276	<2.00E-16	***
sScore	0.096	0.037	2.620	0.008785	**
sClass	-0.036	0.036	-0.993	0.320478	
sSize	0.207	0.034	6.101	1.06E-09	***
sADR X sScore	0.101	0.035	2.887	0.003895	**

Figure 18: Interaction between sScore and sADR on STR mismatch rate

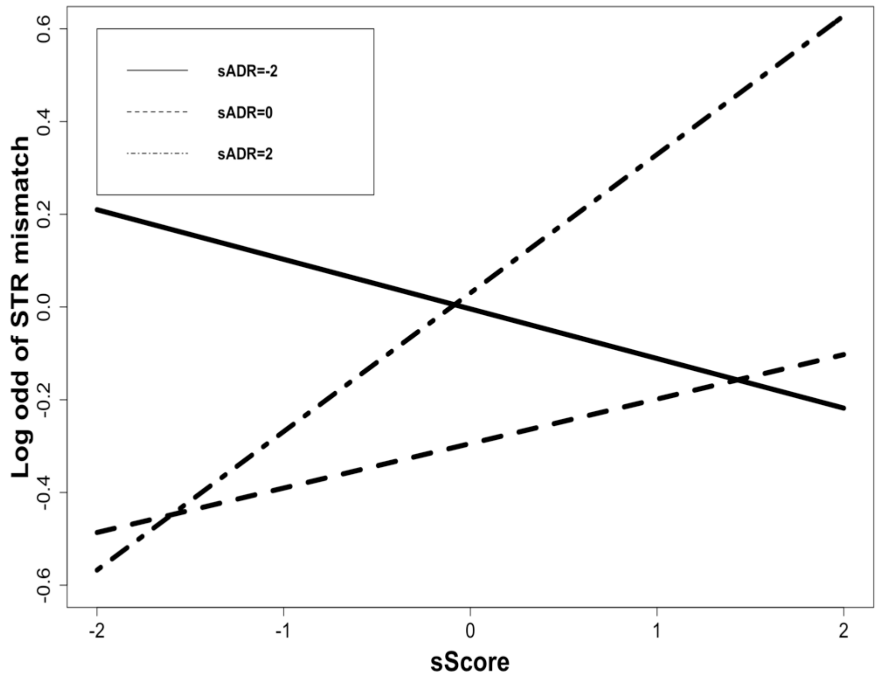


Table 16 shows results for the similar model as in Table 15 but for TripAdvisor mismatch versus STR (Table 15) with Figure 19 showing the interaction of score and ADR – it is interesting to note that the interaction observed with STR is not present with TripAdvisor. Table 17 and Figure 20 show the net effect of sADR in both models as sADR is represented twice – once as sADR and once as sADR squared. The results indicate that sADR has more influence on TripAdvisor mismatch rate than STR mismatch rate.

Table 16: Model for TripAdvisor mismatch rate

	Estimate	Std. Error	t-Value	P value	Significance
(Intercept)	1.729	0.044	39.587	<2.00E-16	***
sADR	-0.278	0.041	-6.740	1.59E-11	***
Squared sADR	0.624	0.040	15.739	<2.00E-16	***
sRank	0.174	0.030	5.807	6.36E-09	***
sScore	0.005	0.028	0.194	0.846	
Same sClass	-0.243	0.049	-4.922	8.58E-07	***
Same sSize	-0.548	0.050	-10.929	<2.00E-16	***
sADR X sScore	-0.027	0.038	-0.722	0.47	

Figure 19: Interaction between sScore and sADR on TripAdvisor mismatch rate

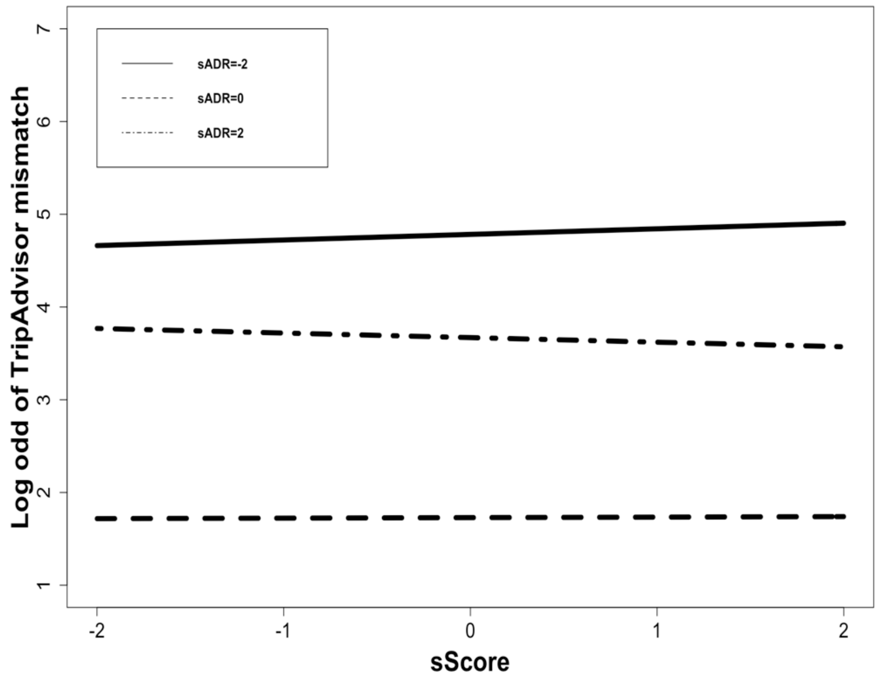
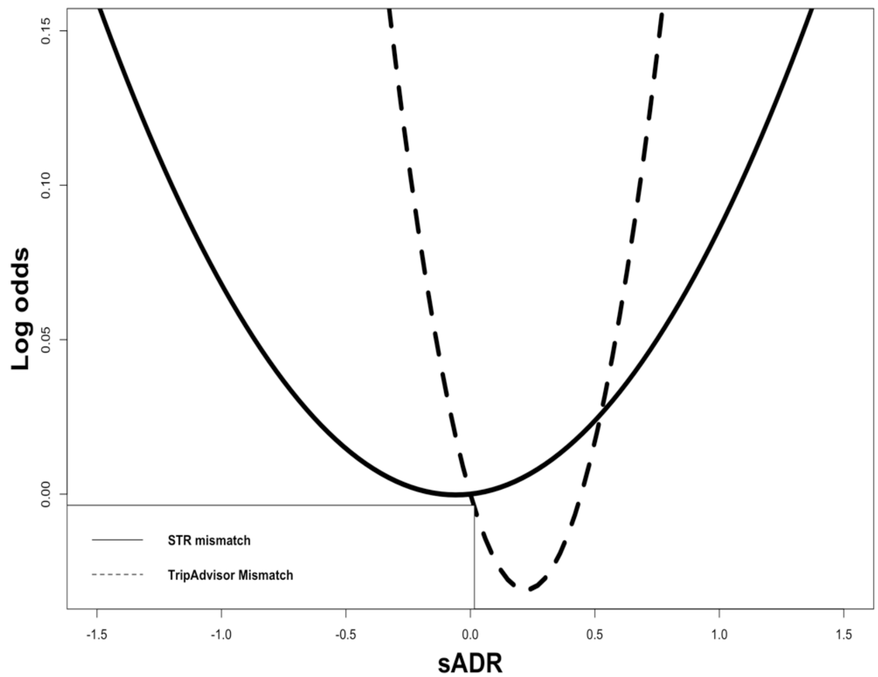


Table 17: Net effect of sADR on both models

	Net sADR effect
On Log odds of STR mismatch	0.086+0.164*sADR
On Log odds of TripAdvisor mismatch	0.346+1.248*sADR

Figure 20: Net effect of sADR on both models



4.2 Who are my competitors?

In this part, we aim to figure out how hotel customers are forming competitive sets. In part 1, we find that all the five variables (sRate, sScore, sRank, sClass and sSize) are related to mismatch rate, indicating that hotels and customers have different views of those attributes when selecting the hotels. Therefore, in this part, we analyze the same attributes and try to find the relationship between customers' choice and those attributes. The following tables show the variables of interest:

Table 18: Variables of Interest

Variable Types	Variable Names	Mean	Standard Deviation	Max	Min
Numerical	sDistance	0	0.99	-1.16	53.6
	sRate	-0.08	0.35	-1.56	1.26
	sScore	-0.03	0.20	-1.33	1.31
	sRank	0.34	0.82	-1.96	1.99
Ordinal categorical	sClass	11	-	-	-
	sSize	9	-	-	-

We use two type of dependent variables – common session and scaled page views (absolute value of the subject hotel's page views minus competitor hotels' page views) – to represent customers' views of competitive sets. Literature reviews state that consumers follow a multi-stage process to choose hotel competitors. Namely, customers first use some attributes to form consideration sets and then use other attributes to determine choice sets from these consideration sets. The common session is defined as two hotels shown in the same session while page views is defined as the number of times subject hotels or competitor hotels being clicked when they are in the same session. Therefore, we use common session to define the consideration set and use scaled page views to define the choice set.

4.2.1 Poisson regression results

The outcome common session is count data, so we utilize Poisson regression offset by count[6] model to conduct further analyses. The numerical attributes – sRank, sScore and sRate – are standardized. Accordingly, our model for common session is as follows:

[6] Count is the sum of all common sessions for each subject hotel, and we offset by count because the number of common sessions varies a lot for different subject hotels.

$$\ln(\text{Common sessions}) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \ln(count) \quad 4.2.1$$

The other outcome scaled page views is calculated by the absolute value of the subject hotel's page views minus competitor hotels page views. We use the model 4.2.2 – Poisson regression offset by the average value of subject hotels' page views and comp hotels' page views – to further analyze.

$$ln(|subject\ hotel\ page\ views - comp\ hotel\ page\ views|) = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \ln\left((subject\ hotel\ page\ views + comp\ hotel\ page\ views)/2\right) \quad 4.2.2$$

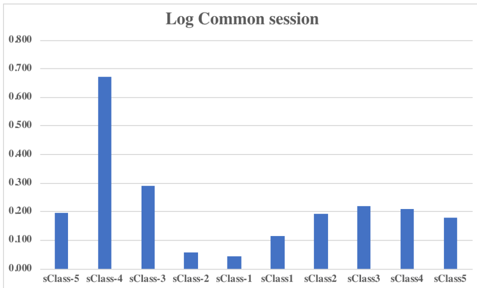
Based on the results in Table 19, sSize, sClass, sDistance, sRank, sScore and sRate all influence customers' selection of both the consideration set and the choice set. If comp hotels' price increase, the likelihoods of being chosen into the consideration set and choice set both first increase and then decrease, indicating that customers have a price threshold in mind during the hotel selection process. While customers are more likely to select lower score hotels as competitors in determining the consideration set, they prefer to choose higher score competitor hotels in forming the choice set.

Customers are more likely to consider two hotels as competitors when common sessions are larger and when scaled page views are smaller. Therefore, in order to compare two concepts, we change the direction of scaled page views and create a new concept called negative scaled page views to do further analyses. Figure 21 illustrates that customers may prefer higher class competitors in forming the consideration set, while customers tend to select similar class hotels as competitors in forming the choice set. Customers may prefer similar size competitors in choosing the consideration set while customers are more likely to select smaller size hotels as competitors in forming the choice set. Therefore, for the interaction model for common session (Table 20), sSize is treated as a continues variable and sClass is treated as a dummy variable (subject and competitor hotels have same class or not). Similarly, for interaction model for scaled page views (Table 21), sClass is treated as a continuous variable while sSize is treated as a dummy variable (Same size or not).

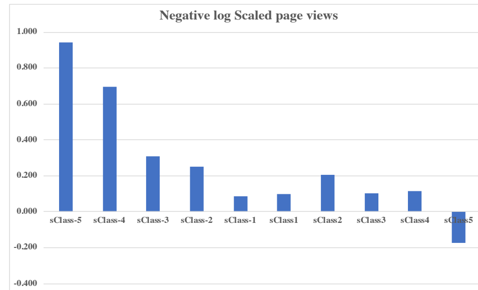
Table 19: Multiple Poisson regressions results

Parameters	Consideration Set			Choice Set		
	Estimate	P value	Significance	Estimate	P value	Significance
(Intercept)	-2.731	<2.00E-16	***	-1.459	<2.00E-16	***
sDistance	-0.161	<2.00E-16	***	0.076	<2.00E-16	***
sScore	0.062	4.64E-08	***	0.026	0.00143	**
sRate	0.046	0.000121	***	-0.031	0.000169	***
Squared sRate	-0.072	<2.00E-16	***	0.056	<2.00E-16	***
sRank	-0.067	6.07E-11	***	0.117	<2.00E-16	***
sClass-5	-0.942	0.502026		0.195	0.860073	
sClass-4	-0.702	0.000633	***	0.679	1.66E-08	***
sClass-3	-0.328	4.30E-05	***	0.310	2.44E-09	***
sClass-2	-0.254	1.03E-08	***	0.058	0.071071	.
sClass-1	-0.088	0.000256	***	0.044	0.011579	*
sClass1	-0.099	2.33E-05	***	0.114	2.25E-12	***
sClass2	-0.207	2.34E-07	***	0.195	2.01E-13	***
sClass3	-0.110	0.142437		0.232	2.71E-06	***
sClass4	-0.115	0.447071		0.208	0.039888	*
sClass5	0.173	0.441288		0.171	0.270704	
sSize-4	-0.227	0.004623	**	0.310	2.48E-09	***
sSize-3	-0.119	0.003469	**	0.277	<2.00E-16	***
sSize-2	-0.154	6.83E-08	***	0.170	<2.00E-16	***
sSize-1	-0.155	6.66E-11	***	0.167	<2.00E-16	***
sSize1	-0.124	1.81E-06	***	0.058	0.002365	**
sSize2	-0.066	0.060828	.	0.067	0.006606	**
sSize3	-0.373	5.22E-08	***	0.043	0.381533	
Size4	-0.441	0.141974		-0.007	0.973528	

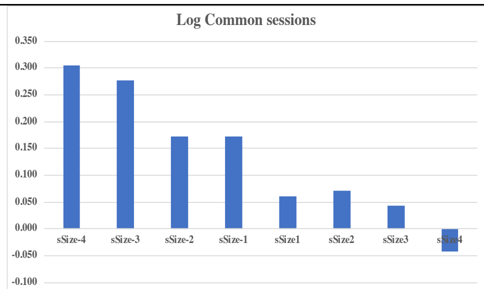
Figure 21: Relationship between sClass or sSize and log common session or negative log scaled page views



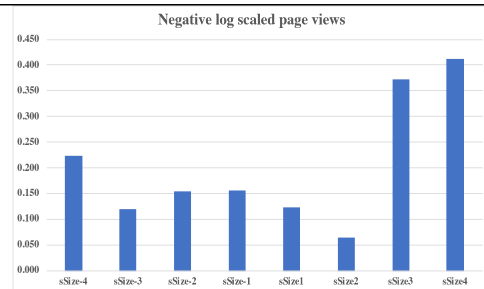
log common session VS sClass



negative log scaled page views VS sClass



log common session VS sSize



negative log scaled page views VS sSize

4.2.2 Predicted Interactions

Table 19 shows that customers perceive score differently but rate similarly in determining the consideration sets and choice sets. To further understand customers' competitor hotels selection, we analyze the interaction between score and rate for both common session and scaled page views. Table 21 displays result for the similar model (same model except Same sClass treats class as a dummy variable and sSize is a treated as a continuous variable) as in Table 20 but for scaled page views versus common session. The interaction between score and rate exhibits for selecting the choice set but not for selecting the consideration set. Figure 23 reveals that this interaction is stronger for competitor hotels with lower rates.

Table 22 and Figure 24 show the net effect of sRate in both models since sRate is presented twice – once as sRate and once as sRate squared. The results imply that customers may perceive rate similarly during the two process – selecting consideration set and choice set.

Table 20: Interaction model for common session

Parameters	Estimate	Std. Error	t-Value	P Value	Significance
(Intercept)	-2.853	0.017	-168.996	<2.00E-16	***
sDistance	-0.159	0.014	-11.064	<2.00E-16	***
sRate	0.031	0.011	2.907	0.00365	**
Squared sRate	-0.079	0.008	-10.472	<2.00E-16	***
sScore	0.053	0.012	4.608	4.09E-06	***
sClass	-0.021	0.003	-6.671	2.60E-11	***
Same sSize	0.140	0.018	7.679	1.67E-14	***
sRank	-0.065	0.010	-6.460	1.07E-10	***
sRate X sScore	0.017	0.012	1.402	0.16087	

Figure 22: Interaction between sScore and sRate on Log common session

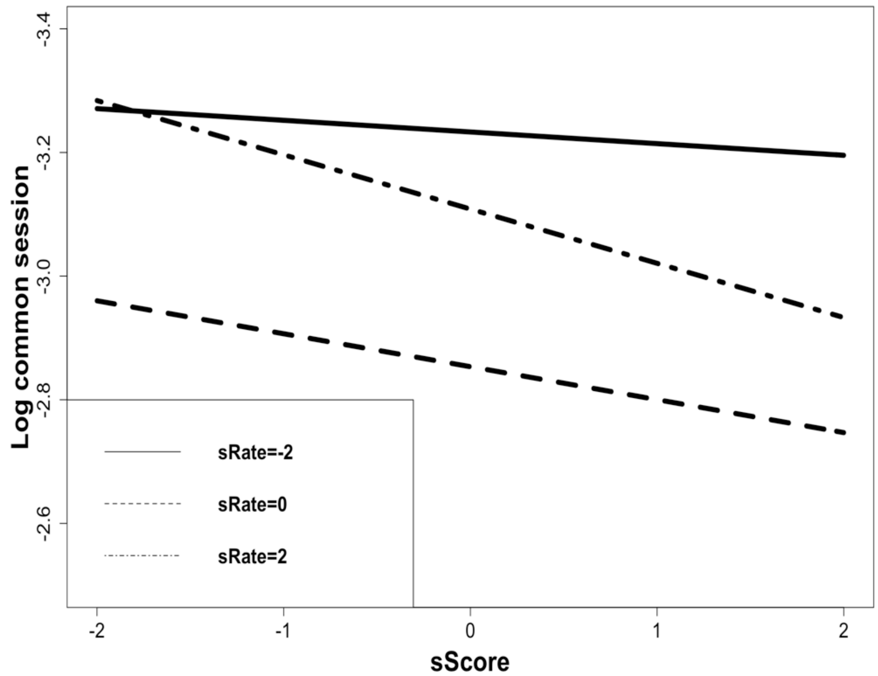


Table 21: Interaction model for scaled page views

	Estimate	Std. Error	t Value	P Value	Significance
(Intercept)	-1.320	0.014	-94.017	<2.00E-16	***
sDistance	0.079	0.005	14.371	<2.00E-16	***
sRate	-0.050	0.007	-6.945	3.88E-12	***
Squared sRate	0.056	0.005	12.195	<2.00E-16	***
sRank	0.105	0.007	14.314	<2.00E-16	***
Same sClass	-0.107	0.013	-8.490	<2.00E-16	***
sSize	0.015	0.003	5.369	7.99E-08	***
sScore	0.035	0.008	4.311	1.63E-05	***
sRate X sScore	0.018	0.008	2.196	0.0281	*

Figure 23: Interaction between sScore and sRate on Negative log scaled page views

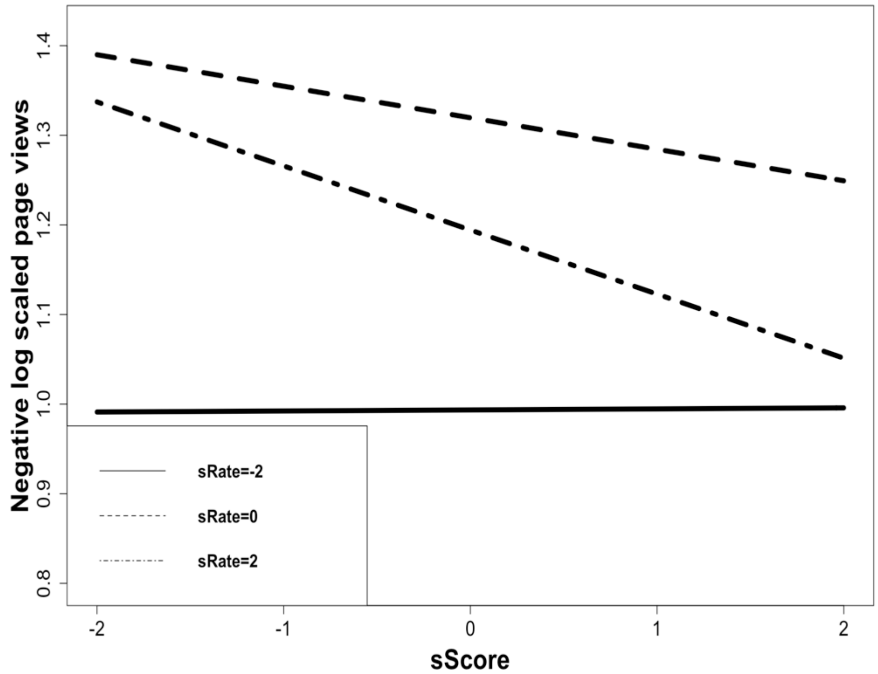
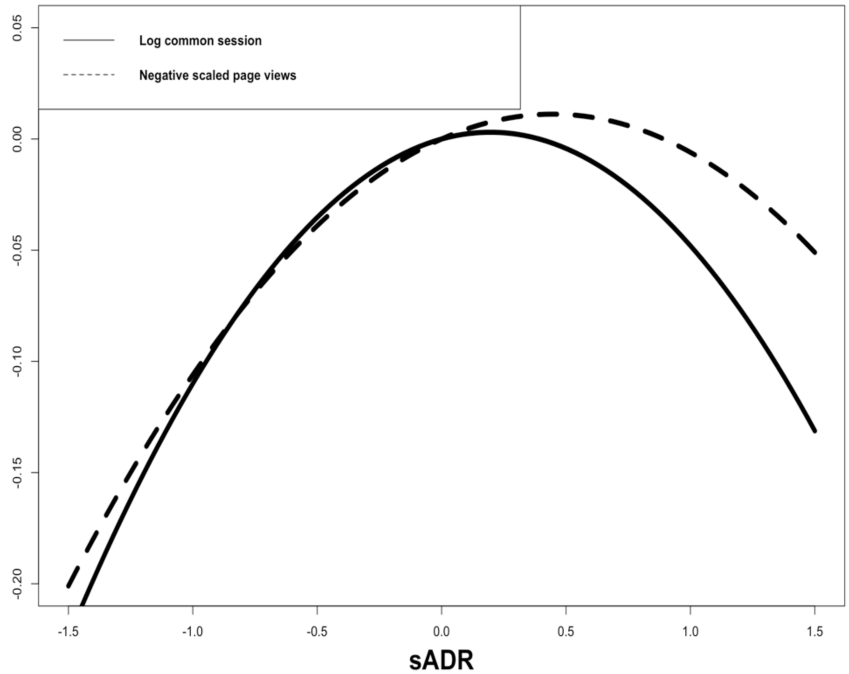


Table 22: Net effect of sRate on both models

	Net sRate effect
On Log common session	-0.048-0.159*sRate
On Negative log scaled page views	-0.006-0.113*sRate

Figure 24: Net effect of sRate on both models



4.3 Assess hotels' current comp set

In part 1, we find that hotels and customers views of competitive sets are different. In part 2, we observe how customers choose competitor hotels. In this part, we examine hotels' current competitive sets based on customers' views. To demonstrate it, we use TripAdvisor match rate (1- TripAdvisor mismatch rate) to measure the similarity between hotels and customers' views of competitive sets. The TripAdvisor mismatch rate is predicted using the final model (Table 23) in part 1. The higher the match rate, the more similar of hotels and customers' views of competitive sets. Furthermore, we apply the common session and negative scaled page views – they are calculated based on the final models in part 1 (Table 24 and Table 25) – to represent customers' view of competitive sets. Customers are more likely to consider two hotels as competitors when the common session or negative scaled page views is large. The correlations between the predicted log common session and predicted negative log scaled page views are 0.63 and 0.67, indicating that hotels and customers are using the attributes (price, size, class, geographic distance, TripAdvisor rank and TripAdvisor score) to select the competitor hotels. However, these results also imply that they will also use other attributes to make competitor selection decisions.

Table 23 Final model for TripAdvisor mismatch rate

Parameters	Estimate	Std. Error	t-Value	P Value	Significance
(Intercept)	1.728	0.044	39.597	<2.00E-16	***
sADR	-0.278	0.041	-6.753	1.45E-11	***
Squared sADR	0.618	0.039	15.990	<2.00E-16	***
sRank	0.173	0.027	6.320	2.61E-10	***
Same Class	-0.241	0.049	-4.894	9.86E-07	***
Same Size	-0.547	0.050	-10.927	<2.00E-16	***

Table 24: Final model for log common session

Parameters	Estimate	Std. Error	t-Value	P Value	Significance
(Intercept)	-2.854	0.017	-169.065	<2.00E-16	***
sDistance	-0.160	0.014	-11.100	<2.00E-16	***
sRate	0.032	0.011	2.986	0.00283	**
Squared sRate	-0.076	0.007	-10.603	<2.00E-16	***
sScore	0.057	0.011	5.039	4.71E-07	***
sClass	-0.020	0.003	-6.609	3.94E-11	***
Same sSize	0.140	0.018	7.654	2.02E-14	***
sRank	-0.065	0.010	-6.440	1.22E-10	***

Table 25: Final model for log scaled page views

	Estimate	Std. Error	t Value	P Value	Significance
(Intercept)	-1.320	0.014	-94.017	<2.00E-16	***
sDistance	0.079	0.005	14.371	<2.00E-16	***
sRate	-0.050	0.007	-6.945	3.88E-12	***
Squared sRate	0.056	0.005	12.195	<2.00E-16	***
sRank	0.105	0.007	14.314	<2.00E-16	***
Same sClass	-0.107	0.013	-8.490	<2.00E-16	***
sSize	0.015	0.003	5.369	7.99E-08	***
sScore	0.035	0.008	4.311	1.63E-05	***
sRate X sScore	0.018	0.008	2.196	0.0281	*

Figure 25: Relationship between predicated log CS and TripAdvisor matched rate

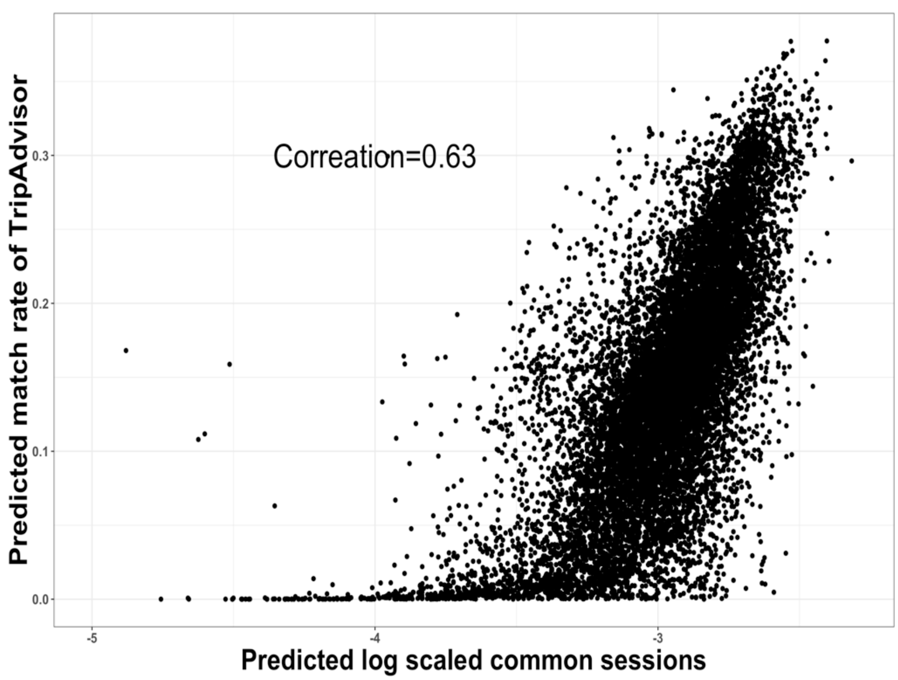
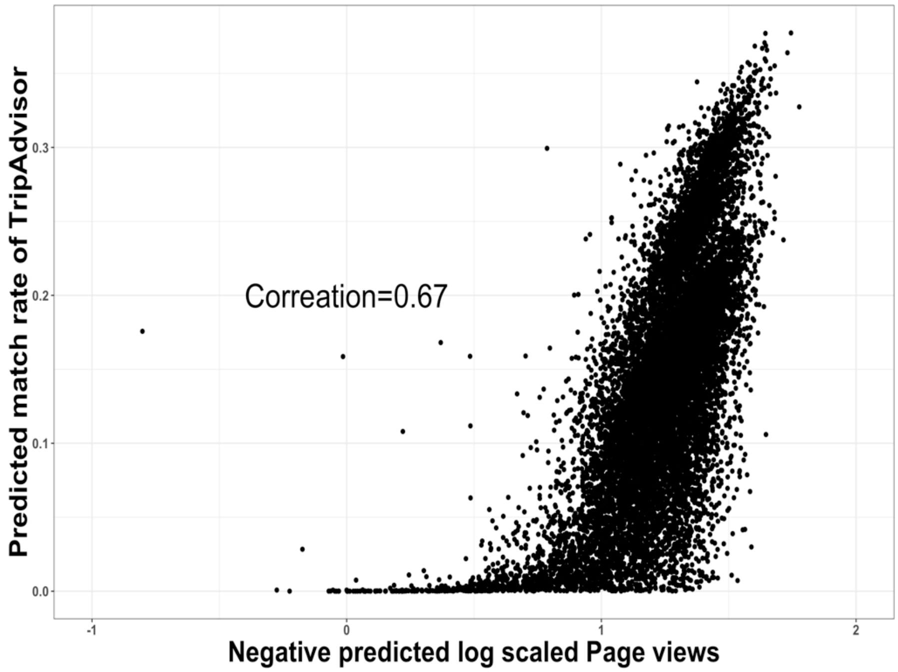


Figure 26: Relationship between negative predicted log PVS and TripAdvisor matched rate



CHAPTER 5

5 CONCLUSIONS

The purpose of this study is to use customers' views to assess hotels' competitive sets. Based on this purpose, our threefold goals are: 1) compare competitive set identified by hotels with those selected by customers and find the attributes associating with the dissimilarity and similarity; 2) find out how customers identify the hotel competitors; 3) analyze hotels and customers' competitive sets selection.

Using the data from Smith Travel Research and TripAdvisor, we find that the STR match rate was 42.3%, indicating that hotels' and customers' choices of competitive sets are different. Hotels may need to pay more attention to customers' opinions since many researchers have claimed the importance of customers in choosing the right competitive set (Coleman, 2011; Li et al., 2014; Haynes, 2015). Chain managed hotels have higher match rates than franchised hotels, reflecting that chain managed hotels may pay more attention to customers' point of views when selecting the competitive sets. And when the chain scale decrease - from luxury to economy - the match rate decrease. This result indicates that higher chain scale hotels' competitive sets views are more similar to customers' than lower chain scale hotels. In addition, we find that compared to customers, hotels tend to select the competitors with smaller size, lower rank, lower score, lower class. On the contrary, customers are more likely to choose different size, different class, higher rank and higher score hotels as competitors. Therefore, hotels and customers perceive attributes differently when selecting the competitor hotels.

Furthermore, customers will use size, class, geographic distance, price, rank and score to determine the consideration set and then use the same six attributes to identify the choice set. It is interesting to note that they perceive these attributes differently. That is, they will choose lower score, similar class and smaller size rivals into the consideration set and higher score, higher class and similar size rivals into the choice set. The interaction between rate and score is observed for the choice set selection but not for consideration selection, implying that the selection for the choice set is more complicated than for the consideration set.

Finally, both hotels and customers will use size, class, geographic distance, price, rank and score to select the competitor hotels. However, their selection processes are more complex and may involve more attributes to select the competitive sets.

CHAPTER 6

6 LIMITATIONS AND FURTURE RESEARCH

Although we use substantial data from the leading benchmarking service provider and the largest travel website in the world (Smith Travel Research and TripAdvisor) to conduct our research, we still face two limitations. To begin with, our data does not provide with comprehensive customer types for the customer side analysis. Namely, TripAdvisor may only capture the transient business and leisure customers and we lack the information of managed business customers' point of view. And we might lose some insights of the customers who will use other channels rather than TripAdvisor. This limitation, however, is mitigated as TripAdvisor is the largest travel website in the world. Secondly, our data may not take seasonality into consideration for the customer side analysis, since the data we collected from TripAdvisor was on December 2017. Seasonality may have effect on the competitive set choice of customers.

Accordingly, further studies may want to include the managed business travelers as well and provide a more general customers type to stand for the customer side's completive set selection. Moreover, researchers can incorporate whole year' data the avoid the seasonality problems. With the hindsight, hotels may come up with a better competitive set which pay close attention to the customers' preference.

REFENRENCES

Anderson, Chris K., and Benjamin Lawrence. 2014. “The Influence of Online Reputation and Product Heterogeneity on Service Firm Financial Performance.” *Service Science* 6 (4): 217–28.

Baruca, Petra Zabukovec, and Žana Čivre. n.d. “How Do Guests Choose a Hotel?,” 10.

Baum, Joel A. C., and Stephen J. Mezias. 1992. “Localized Competition and Organizational Failure in the Manhattan Hotel Industry, 1898-1990.” *Administrative Science Quarterly* 37 (4): 580.

Baum, Joel A.C, and Theresa K Lant. 2003. “HITS AND MISSES: MANAGERS’ (MIS)CATEGORIZATION OF COMPETITORS IN THE MANHATTAN HOTEL INDUSTRY.” In *Advances in Strategic Management*, 20:119–56. Bingley: Emerald (MCB UP).

Berezina, Katerina, Anil Bilgihan, Cihan Cobanoglu, and Fevzi Okumus. 2016. “Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews.” *Journal of Hospitality Marketing & Management* 25 (1): 1–24.

Cai, Liping A., and J. S. Perry Hobson. 2004. “Making Hotel Brands Work in a Competitive Environment.” *Journal of Vacation Marketing* 10 (3): 197–208.

Campa-Planas, Fernando, Antonio Alvarez-Ferrer, and Juan Pablo Gonzales-Bustos. 2018. “Identification of the Key Factors for Success in the Hotel Sector.” *Intangible Capital* 14 (1): 74.

Canina, Linda. n.d. “Revenue Management in U.S. Hotels: 2001 - 2005,” 18.

Chatterjee, Patrali, and Yawei Wang. 2012. “Online Comparison Shopping Behavior of Travel Consumers.” *Journal of Quality Assurance in Hospitality & Tourism* 13 (1): 1–23.

Chernatony, Leslie De, Kevin Daniels, and Gerry Johnson. 1994. “Competitive Positioning Strategies Mirroring Sellers’ and Buyers’ Perceptions?” *Journal of Strategic Marketing* 2 (3): 229–48.

Chiang, Chun-Fang, and SooCheong Shawn Jang. 2007. “The Effects of Perceived Price and Brand Image on Value and Purchase Intention: Leisure Travelers’ Attitudes Toward Online Hotel Booking.” *Journal of Hospitality & Leisure Marketing* 15 (3): 49–69.

Chu, Raymond K.S., and Tat Choi. 2000. “An Importance-Performance Analysis of Hotel Selection Factors in the Hong Kong Hotel Industry: A Comparison of Business and Leisure Travellers.” *Tourism Management* 21 (4): 363–77.

Clark, Bruce H., and David B. Montgomery. 1999. “Managerial Identification of Competitors.” *Journal of Marketing* 63 (3): 67.

Crnojevac, Igor H, and Saša Karl. 2010. “ETourism: A Comparison of Online and Offline Bookings and the Importance of Hotel Attributes” 34 (1): 14.

Gallego, Guillermo, Lin Li, and Richard Ratliff. 2009. “Choice-Based EMSR Methods for Single-Leg Revenue Management with Demand Dependencies.” *Journal of Revenue and Pricing Management* 8 (2–3): 207–40.

Hatten, Kenneth J., and Mary Louise Hatten. 1987. “Strategic Groups, Asymmetrical Mobility Barriers and Contestability.” *Strategic Management Journal* 8 (4): 329–42.

Haynes, Natalie. 2016. “The Evolution of Competitor Data Collection in the Hotel Industry and Its Application to Revenue Management and Pricing.” *Journal of Revenue and Pricing Management* 15 (3–4): 258–63.

Hu, Ya-Han, Kuanchin Chen, and Pei-Ju Lee. 2017. “The Effect of User-Controllable Filters on the Prediction of Online Hotel Reviews.” *Information & Management* 54 (6): 728–44.

Jain, Pryas. 2013. "ANALYZING CUSTOMERS' PREFERENCES IN SELECTING HOTEL SEGMENT: AN EMPIRICAL CASE STUDY," no. 4: 6.

Jones, Peter, and Meng-Mei Chen. 2011. "Factors Determining Hotel Selection: Online Behaviour by Leisure Travellers." *Tourism and Hospitality Research* 11 (1): 83–95.

Kim, Jin-young, and Linda Canina. 2011. "Competitive Sets for Lodging Properties." *Cornell Hospitality Quarterly* 52 (1): 20–34.

Lee, Seul Ki. 2015. "Quality Differentiation and Conditional Spatial Price Competition among Hotels." *Tourism Management* 46 (February): 114–22.

Lee, Seul Ki, and SooCheong (Shawn) Jang. 2013. "Asymmetry of Price Competition in the Lodging Market." *Journal of Travel Research* 52 (1): 56–67.

Li, Gang, Rob Law, Huy Quan Vu, Jia Rong, and Xinyuan (Roy) Zhao. 2015. "Identifying Emerging Hotel Preferences Using Emerging Pattern Mining Technique." *Tourism Management* 46 (February): 311–21.

Li, Jun, and Serguei Netessine. 2012. "Who Are My Competitors? - Let the Customer Decide." *SSRN Electronic Journal*.

Lien, Che-Hui, Miin-Jye Wen, Li-Ching Huang, and Kuo-Lung Wu. 2015. "Online Hotel Booking: The Effects of Brand Image, Price, Trust and Value on Purchase Intentions." *Asia Pacific Management Review* 20 (4): 210–18.

Liu, James N.K., and Elaine Yulan Zhang. 2014. "An Investigation of Factors Affecting Customer Selection of Online Hotel Booking Channels." *International Journal of Hospitality Management* 39 (May): 71–83.

Livengood, R Scott, and Rhonda K Reger. n.d. "That's Our Turf! Identity Domains and Competitive Dynamics," 20.

Mathews, Vinitia E. 2000. “Competition in the International Hotel Industry.” *International Journal of Contemporary Hospitality Management* 12 (2): 114–18.

McCarthy, Laura, and Debra Stock. n.d. “How Travelers Use Online and Social Media Channels to Make Hotel-Choice Decisions,” 24.

Mccleary, Ken W., Pamela A. Weaver, and Joe C. Hutchinson. 1993. “Hotel Selection Factors as They Relate to Business Travel Situations.” *Journal of Travel Research* 32 (2): 42–48.

Mohammed, Ibrahim, Basak Denizci Guillet, and Rob Law. 2014. “Competitor Set Identification in the Hotel Industry: A Case Study of a Full-Service Hotel in Hong Kong.” *International Journal of Hospitality Management* 39 (May): 29–40.

Morgan, Michael S., and Chekitan S. Dev. 1994. “Defining Competitive Sets of Hotel Brands Through Analysis of Customer Brand Switching.” *Journal of Hospitality & Leisure Marketing* 2 (2): 57–91.

Öğüt, Hulisi, and Bedri Kamil Onur Taş. 2012. “The Influence of Internet Customer Reviews on the Online Sales and Prices in Hotel Industry.” *The Service Industries Journal* 32 (2): 197–214.

Pan, Bing, Lixuan Zhang, and Rob Law. 2013. “The Complex Matter of Online Hotel Choice.” *Cornell Hospitality Quarterly* 54 (1): 74–83.

Perrucci, Alexa Angelica. n.d. “Consumer Perception of Hotel Competitive Sets,” 68.

Peteraf, Margaret A., and Mark E. Bergen. 2003. “Scanning Dynamic Competitive Landscapes: A Market-Based and Resource-Based Framework.” *Strategic Management Journal* 24 (10): 1027–41.

Porac, Joseph F, and Howard Thomas. n.d. “Taxonomic Mental Models in Competitor Definition,” 18.

Rhee, Hosung Timothy, and Sung-Byung Yang. 2015. "How Does Hotel Attribute Importance Vary among Different Travelers? An Exploratory Case Study Based on a Conjoint Analysis." *Electronic Markets* 25 (3): 211–26.

Robson, Stephani K A. n.d. "Using Eye Tracking to Obtain a Deeper Understanding of What Drives Online Hotel Choice," 20.

Smith, Randell A., and Lucy Zheng. 2011. "A Look at Comp Sets: A Historical Perspective That Shapes Today's Way of Thinking." *Cornell Hospitality Quarterly* 52 (4): 371–73.

Sohrabi, Babak, Iman Raeesi Vanani, Kaveh Tahmasebipur, and Safar Fazli. 2012. "An Exploratory Analysis of Hotel Selection Factors: A Comprehensive Survey of Tehran Hotels." *International Journal of Hospitality Management* 31 (1): 96–106.

Stringam, Betsy Bender, John Gerdes, and Dawn M. Vanleeuwen. 2010. "Assessing the Importance and Relationships of Ratings on User-Generated Traveler Reviews." *Journal of Quality Assurance in Hospitality & Tourism* 11 (2): 73–92.

Teng, Shasha, Kok Wei Khong, Alain Yee-Loong Chong, and Binshan Lin. 2017. "Examining the Impacts of Electronic Word-of-Mouth Message on Consumers' Attitude." *Journal of Computer Information Systems* 57 (3): 238–51.

Tsai, Henry, Sylvester Yeung, and Phoebe H. L. Yim. 2011. "Hotel Selection Criteria Used by Mainland Chinese and Foreign Individual Travelers to Hong Kong." *International Journal of Hospitality & Tourism Administration* 12 (3): 252–67.

Tuominen, Pasi. n.d. "The Influence of TripAdvisor Consumer-Generated Travel Reviews on Hotel Performance," 11.

United Nations World Tourism Organization (UNWTO), Anita Blomberg-Nygard, and Chris K. Anderson. 2016. "United Nations World Tourism Organization Study on Online

Guest Reviews and Hotel Classification Systems: An Integrated Approach.” *Service Science* 8 (2): 139–51.

Victorino, Liana, Rohit Verma, Gerhard Plaschka, and Chekitan Dev. 2005. “Service Innovation and Customer Choices in the Hospitality Industry.” Edited by Allard C.R. van Riel. *Managing Service Quality: An International Journal* 15 (6): 555–76.

Vinod, B. 2013. “Leveraging BIG DATA for Competitive Advantage in Travel.” *Journal of Revenue and Pricing Management* 12 (1): 96–100.

Webb, Timothy, and Zvi Schwartz. 2017. “Revenue Management Analysis with Competitive Sets: Vulnerability and a Challenge to Strategic Co-Opetition among Hotels.” *Tourism Economics* 23 (6): 1206–19.

Zhang, Dan, and William L. Cooper. 2005. “Revenue Management for Parallel Flights with Customer-Choice Behavior.” *Operations Research* 53 (3): 415–31.