


Tourist behavior analysis using online user generated data

Yang Yang Ph.D.

Department of Tourism and Hospitality Management

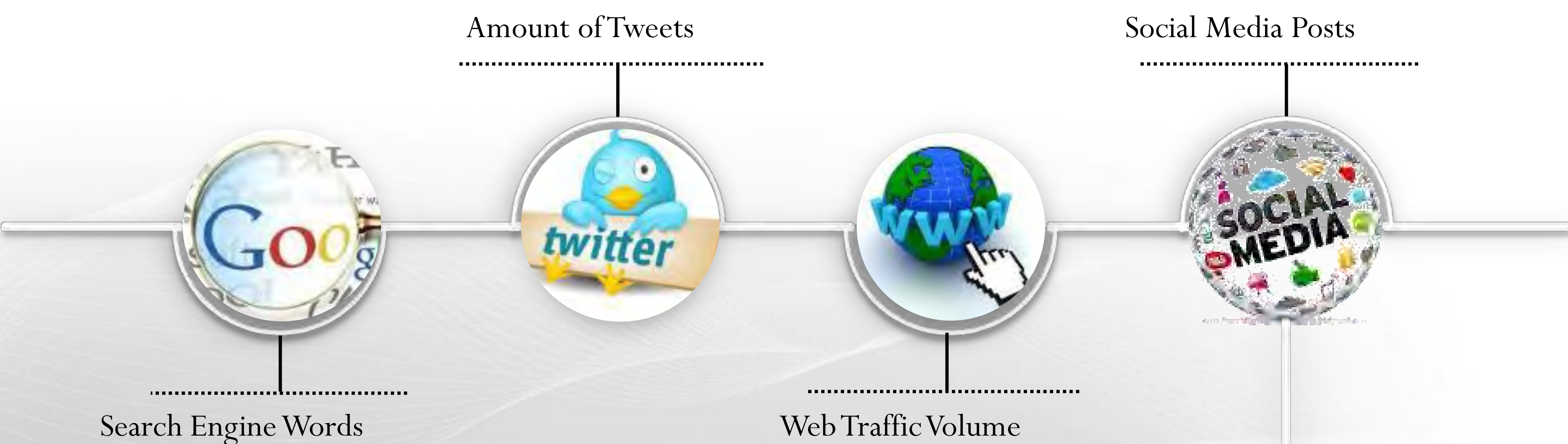
Temple University(yangy@temple.edu)

TEMPLE UNIVERSITY

School of Sport, Tourism and Hospitality Management

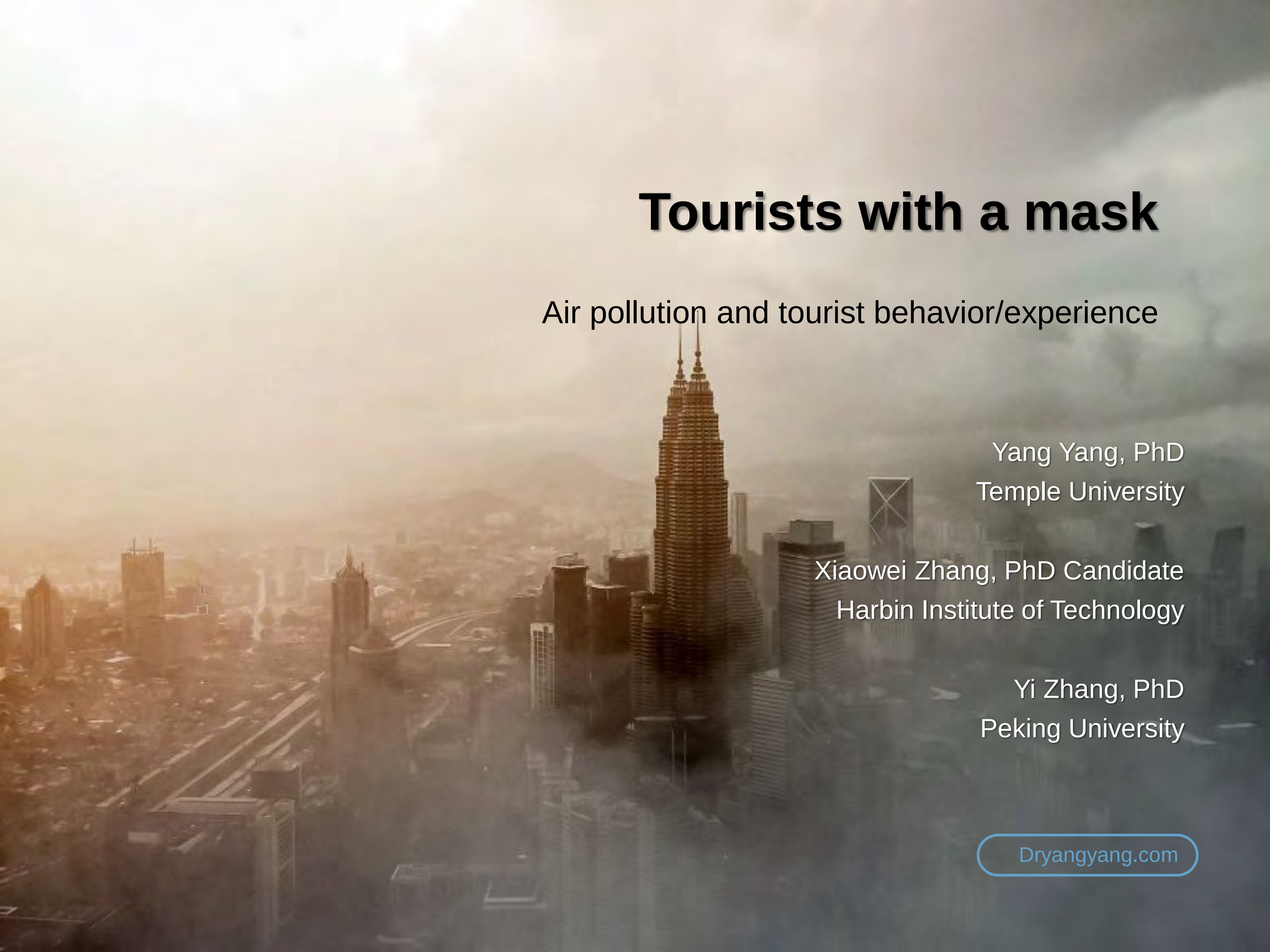
BIG DATA

BIG DATA



- New set of online data in this IT era!
- Provide comprehensive information on tourist behavior, experience, perception and satisfaction.

School of Tourism and Hospitality Management
TEMPLE UNIVERSITY



Tourists with a mask

Air pollution and tourist behavior/experience

Yang Yang, PhD
Temple University

Xiaowei Zhang, PhD Candidate
Harbin Institute of Technology

Yi Zhang, PhD
Peking University

Dryangyang.com

CONTENTS

Introduction

- Many tourist destinations are suffering from varying levels of air pollution.

- The situation is particularly worse in urban destinations.

POLLUTION

Foreign Tourists Skipping Delhi over Air Qu

By Niharika Lal | TNN · 12 December 2018 · TWC India

Tourists at Chandni Chowk market in Delhi (RAJESH MEHTA/ BCCL Delhi)

South China Morning Post

SIGN IN/UP

China

Air pollution takes toll on China's tourism

Shocking levels of air pollution have cast a pall over China's burgeoning tourism industry

AP

Published: 4:02pm, 13 Aug, 2013

1

1

Introduction

- Levels of particulate matter (PM) effectively indicate the level of severity of air pollution because PM affects the visibility of the atmosphere and induces a range of environmental and public health issues (Lanniello et al., 2011; Wang et al., 2001; Zhang et al., 2003).

- As a risk factor, air pollution heavily shapes tourists behavior and experiences. Due to air pollution, tourists may change their travel plan to avoid exposure in the air polluted areas.

- Tourists can be less satisfied with their experiences in the destination due to the quality the destination offers.

Introduction

- Smog is a frequently visible, sometimes literally tangible, generic term for air pollution deriving from multiple human activities including but not limited to the burning of fossil fuels and industrial processes (Watts, 2010).

- Beijing is the Chinese city with perhaps the most serious and widely publicized smog conditions.

- This study aims to leverage social media data (Weibo) to track the experience of tourists in Beijing toward Smog. Also, we compare the experience between local residents and tourists.

Introduction

Impact of Air Pollution

(Measured by PM 2.5 concentration level)

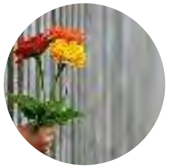
Smog Awareness

Whether the tourists are aware of the issue of air pollution, such as smog attacks

Behavioral Consequence

How tourists change their behavior in terms of location visited, travel scope, and duration.

Emotional Consequence



How tourists' emotion and sentiment change.

Health Consequence

The occurrence of health related issues such as illness and insomnia.

Major contributions

- **Micro**: Micro-level study of the impact of air pollution
- **Health**: Health consequence in tourism activities
- **Framework**: Comprehensive analytical framework of social media
- **D-Learning**: Deep learning tool as a supervised classifier

CONTENTS

Air pollution generally refers to the introduction of harmful particles and biological molecules into Earth's atmosphere.

Exposure to harmful particulate matter (PM) in extreme air pollution severely affects the respiratory system and results in symptoms like cough, chest tightness, headache, etc. (Gehring et al., 2013).

Prolonged exposure to air pollution affect the central nervous system (Fonken et al., 2011).

exposure to ambient air pollution has been found to increase the risk of low happiness (Zheng, Wang, Sun, Zhang, & Kahn, 2019), anxiety symptoms (Power et al., 2015), depressive symptoms (Zijlema et al., 2016), and suicide (Kim et al., 2010).

Behavioral responses to air pollution can be reflected in the social, economic, and political perspectives (Evans & Jacobs, 1981).

- Social responses usually refer to the impact of air pollution on an individual’s lifestyle.

- Economic responses refer to responses and activities that lead to economic transactions from the market place to mediate the effects of air pollution.

- Political responses refer to relevant policies, rules and individual opinions caused by air pollution and its control.

Air pollution is negatively associated with tourism performance indicators (Sajjad, Noreen, & Zaman, 2014)

Health and hedonic risks from air pollution are likely to make tourists feel anxious and change their intention to visit (Williams & Baláž, 2015).

Perceived pollution (measured by Google Trends) lowers inbound tourism to China (Xu & Reed, 2017, 2019)

The number of visitor arrivals to the Sun Moon Lake is negatively associated with the number of bard-air-quality days (Chen, Lin, & Hsu, 2017).

Air pollution has a significant direct negative effect on tourism, and the spillover effect is even larger than the direct effect (Deng, Li, & Ma, 2017).

Potential U.S. and Australian travelers expressed negative views about travel risks in China in general and about air quality in particular (Becken, et. al. 2016).

A positive relationship exists between smog concern on risk perception and trip satisfaction for international tourists to Beijing (Li, et. al., 2015).

Tourists' resilience to air pollution dependents on (1) health condition, (2) travel flexibility, and (3) impacts of haze pollution on experience (Zhang, et. al., 2015)

Residents’ and tourists’ opinions were divided on environmental issues in Hualien, Taiwan (Ku & Mak, 2017)

Due to the exposure to health hazards and impaired visibility, tourists tend to hold a lower level of satisfaction owing to air pollution (Peng & Xiao, 2018).

The pessimism stemming from air pollution tends to magnify tourists' suspicion of service providers during the travel and shape their interaction with these providers (Zhang, Hou, Li, & Huang, 2019).

Yan et al. (2019) found that the number of activities decreases as the concentration of air pollutants increases.

McKercher et al. (2015) found that tourists spend less time in an area with the presence of air pollution, and more time at shopping centers.

Advances of big data analytics

- Various types of big data available

GPS track of vehicles

Social media

Cell phone roaming

User Generate Content

Credit card transaction

Search engine query/ web traffic

Text analytics became particularly popular to explore text data, and popular methods include sentiment analysis, topic mining, and document classification (Zhang, Qiao, Yang, & Zhang, 2020).

Also, sophisticated artificial intelligence methods have been introduced to analyze big data in tourism (Zhang, Chen, et al., 2019).

CONTENTS

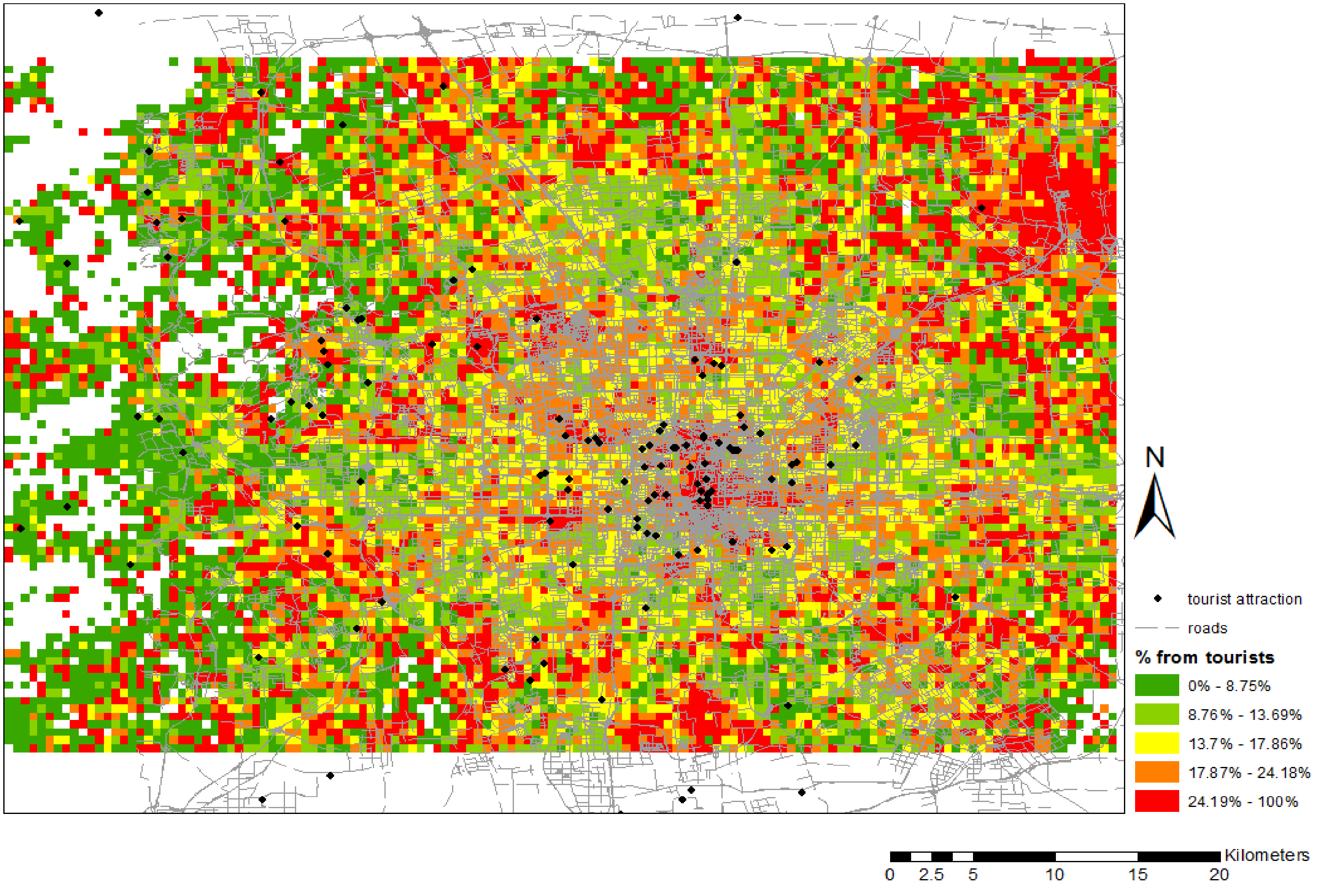
Data analysis

- We collected the social media data from weibo (Chinese counterpart of Twitter) using API. A total of more than 30 million Weibo posts were collected from Jan 1, 2013 to Dec 31, 2013 geocoded within Beijing core area.

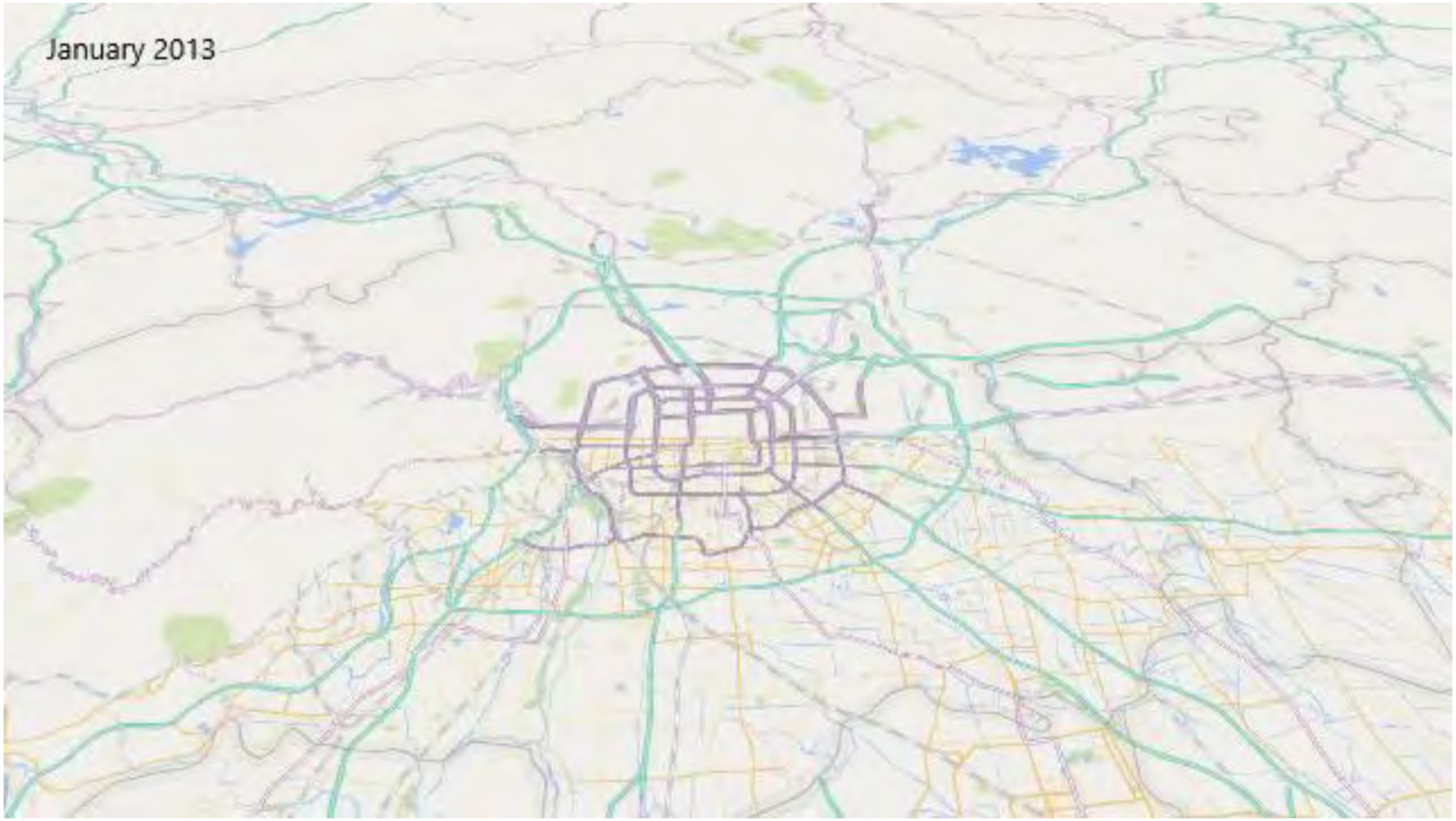
- The data cover the following information
  - Posting time,
  - Posting location (geographic coordinates)
  - Posting content
  - Self-reported residence of users
  - Check-in POI (and its type)
  - Posting device

- Separate residents and tourists
  - By self-reported residence
  - Not posting more than 9 out 12 months.

- We cleaned the data, ending up with more than 13 million.


N
tourist attraction
roads
% from tourists
0% - 8.75%
8.76% - 13.69%
13.7% - 17.86%
17.87% - 24.18%
24.19% - 100%
Kilometers
0
2.5
5
10
15
20

Data analysis



Residents
Tourists

Data Analysis

- Sentiment analysis

Opinion Mining within Natural Language Processing (NLP) that builds systems that try to identify and extract opinions

Chinese text segmentation: built to be the best Python Chinese word segmentation module.

Provides solutions from automatic Chinese words segmentation to psychological analysis.

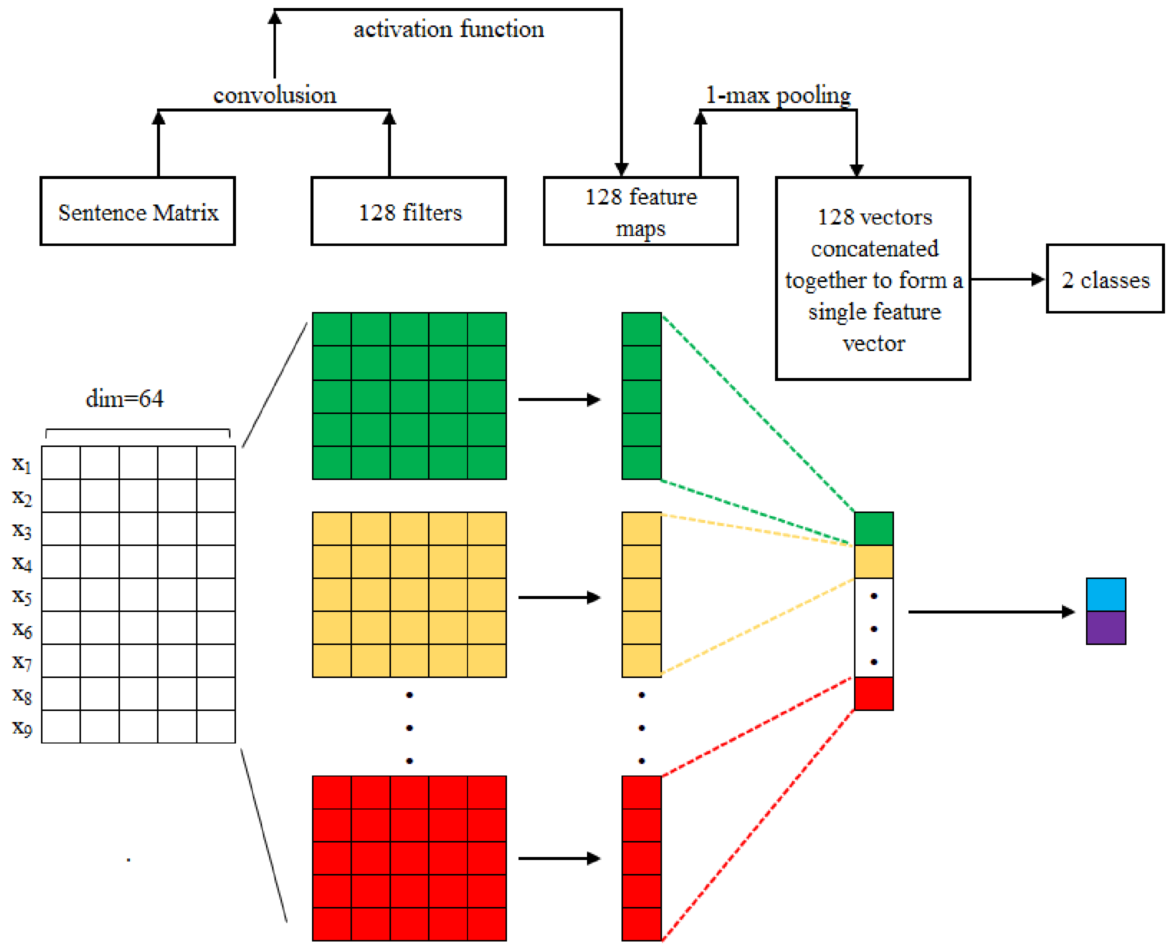
Data Analysis

- Deep learning classification

- manually check 14300 Weibo for health issues

- health-related keywords include coughing “咳嗽”， dizzy “头晕”， sickness “生病”， headache “头痛”, and difficulty breathing “呼吸困难”

- 11000 texts as the training dataset, 2200 texts as the testing dataset, and the other 1100 as the validation dataset



Data analysis-regression

- Dependent variables:

smog_tag

a dummy variable indicating if the smog ("雾霾") was mentioned

distance_TAM

geographic distance (in km) to Tian'anmen Square

radius

radius (in km) of a Weibo users' geotagged locations in a single day

sentiment

sentiment score based on the NLP dictionary

health_issue

a dummy variable indicating if any health-related issues

sleep_issue

a dummy variable indicating if any sleep-related keywords

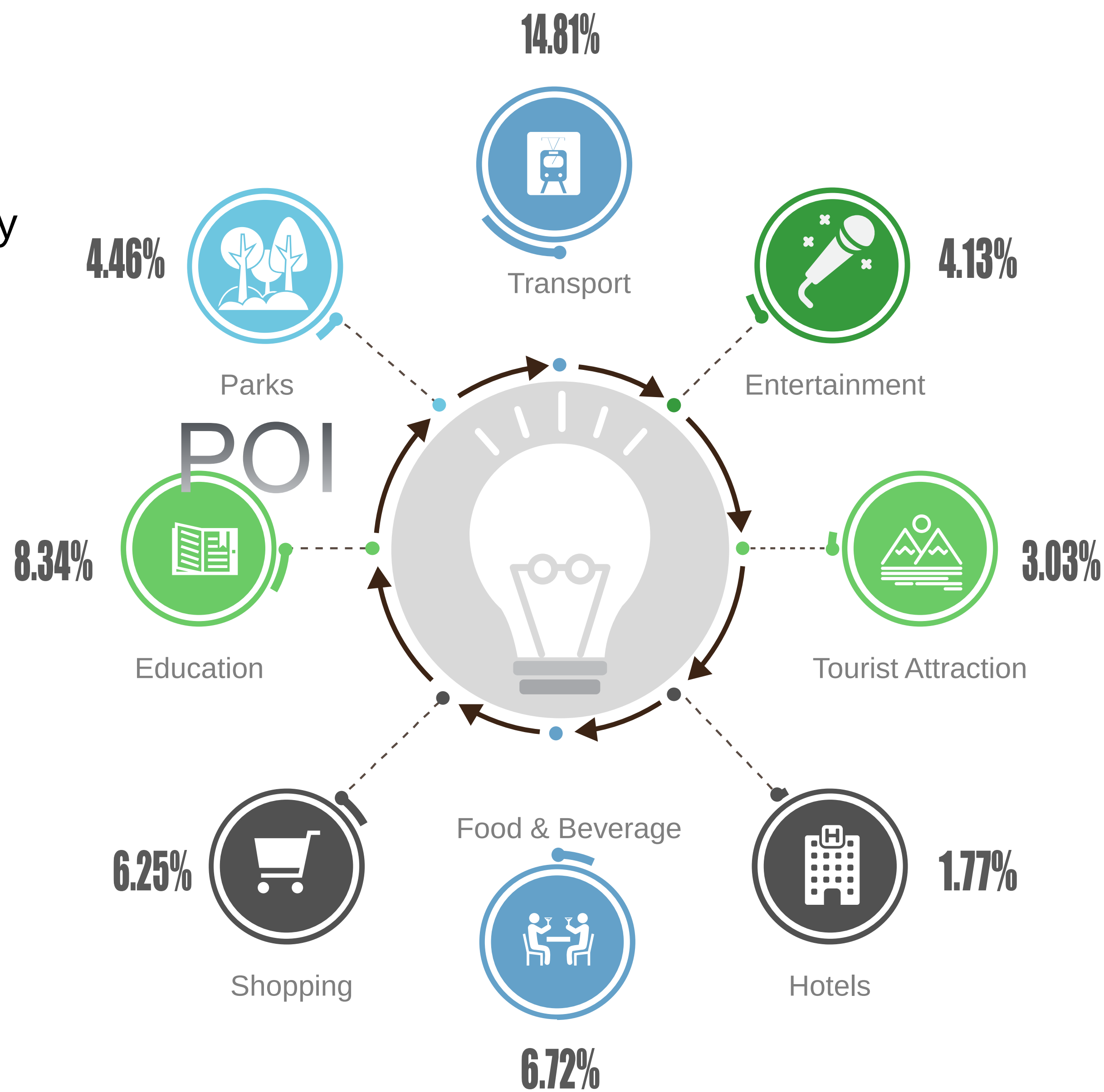
POI_type

a categorical variable indicating the type of points of interest (POIs).

Data Analysis

Categories of POI_type

- The original POI data consists of a large variety of type labels.
- We further merged them into eight types + others



Data Analysis

Major variable of interest

PM2_5: the concentration level of PM 2.5 (micrograms per cubic meter, µg/m3) every hour published by the U.S. embassy in Beijing.

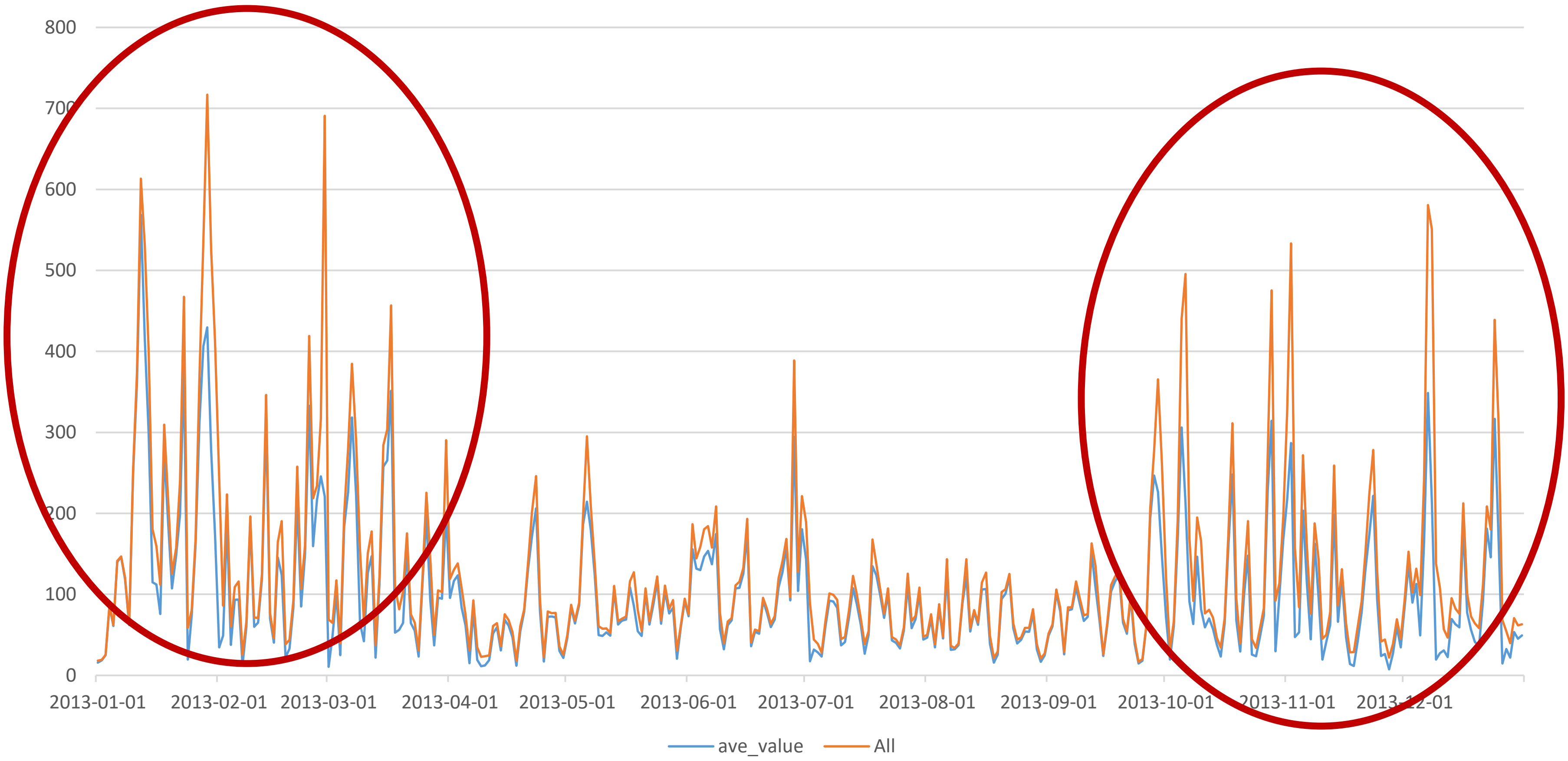
Control variables

sunny	windy	temp	weekend
holiday	tourist	Indensity	precipitation

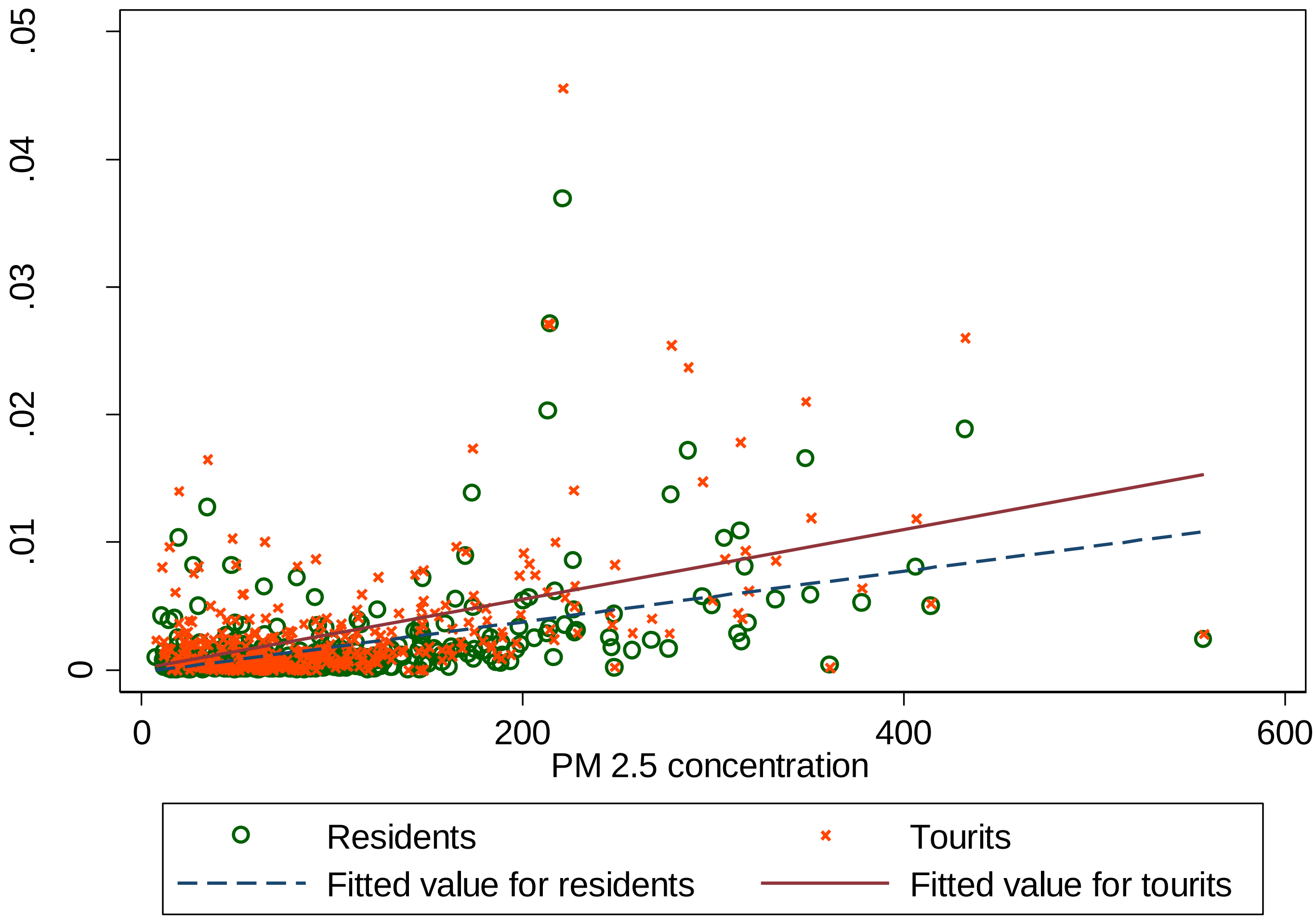
CONTENTS

Results

A clear similar pattern between daily PM 2.5 concentration and percentage of Weibo containing "Smog"

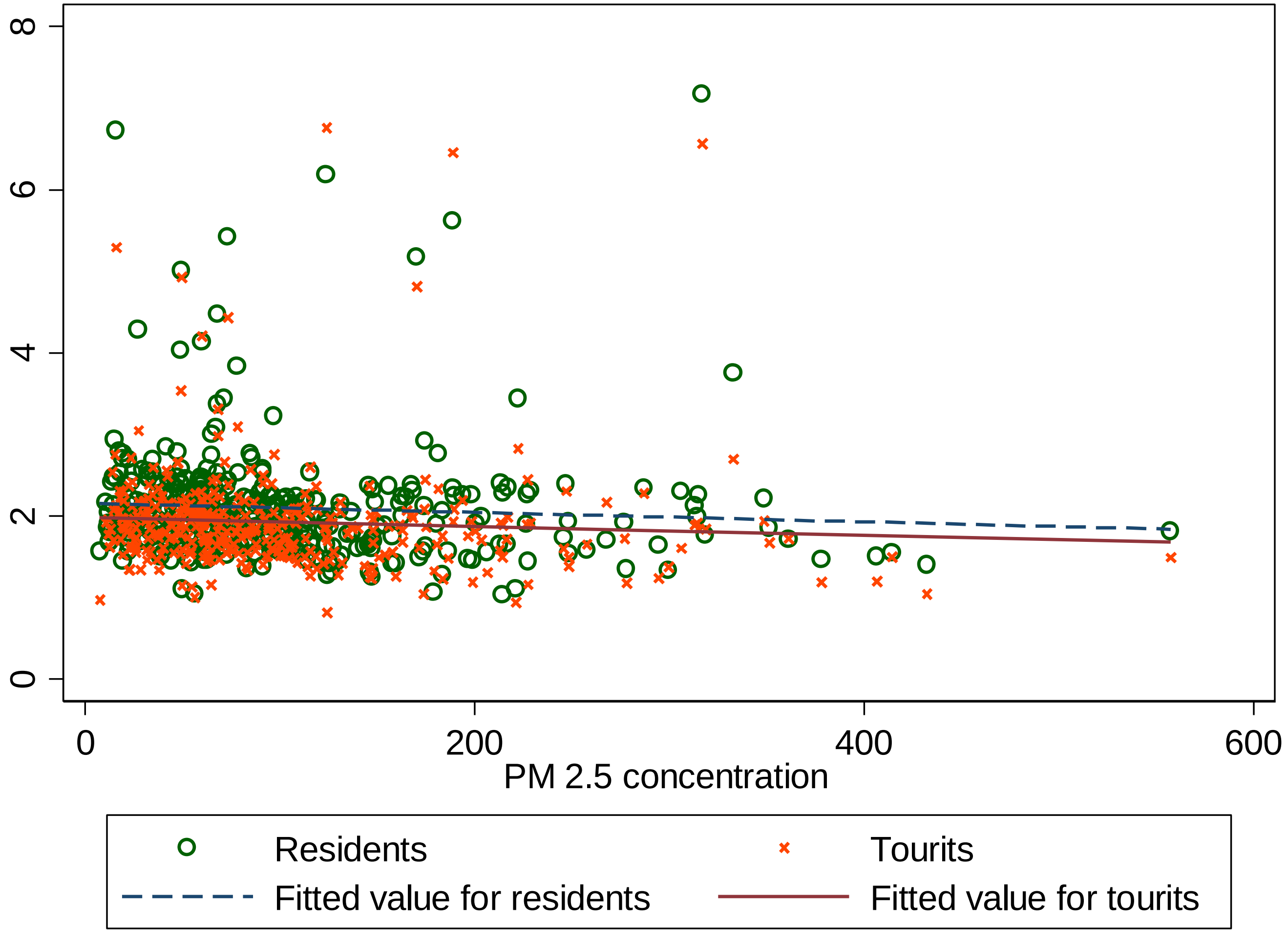


Results



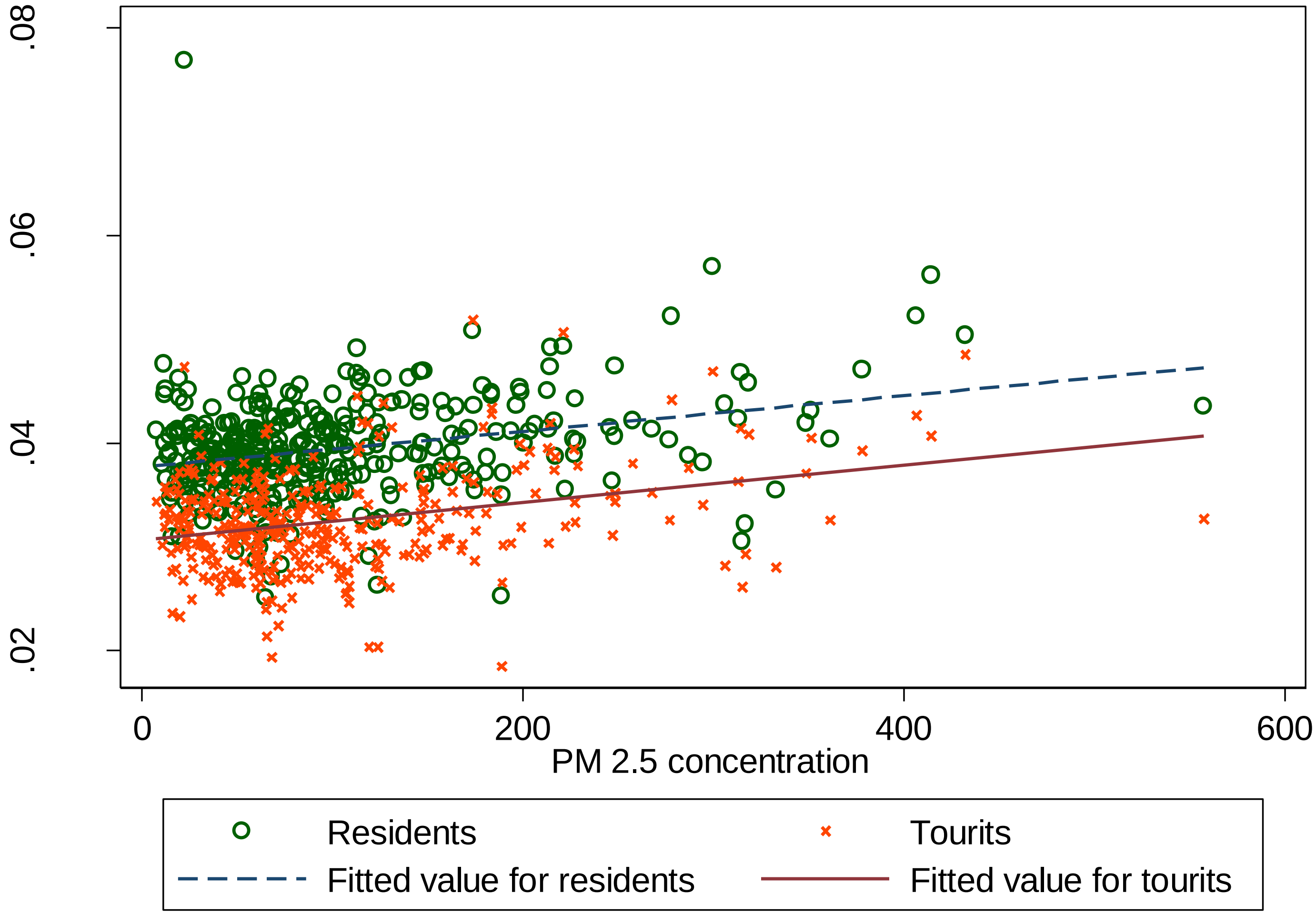
0
.01
.02
.03
.04
.05
0
200
400
600
PM 2.5 concentration
Residents
Tourits
Fitted value for residents
Fitted value for tourits

Results



0
2
4
6
8
0
200
400
600
PM 2.5 concentration
Residents
Tourits
Fitted value for residents
Fitted value for tourits

Results



.02
.04
.06
.08
0
200
400
600
PM 2.5 concentration
Residents
Tourits
Fitted value for residents
Fitted value for tourits

Results

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	smog_tag	distance_TAM	radius	sentiment	health_issue	sleep_issue
pm2_5*(tourist=0)	0.00398***	-0.000508***	0.000146***	-0.000605***	0.000262***	0.000244***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
pm2_5*(tourist=1)	0.00397***	-0.000103**	0.000213**	-0.000685***	0.000391***	0.000120
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
tourist	0.505***	-0.182***	0.923***	-0.187***	-0.193***	-0.200***
	(0.020)	(0.007)	(0.012)	(0.011)	(0.006)	(0.017)
sunny	0.151***	0.0300***	-0.0680***	-0.0436***	0.0259***	-0.0385***
	(0.015)	(0.004)	(0.006)	(0.007)	(0.004)	(0.010)
precipitation	0.132***	0.0336***	0.0499***	-0.0950***	-0.00410	0.00614
	(0.017)	(0.004)	(0.006)	(0.007)	(0.004)	(0.010)
windy	-0.439***	-0.0317***	0.0593***	0.00185	-0.000743	-0.0120
	(0.016)	(0.004)	(0.006)	(0.007)	(0.003)	(0.010)
temp	0.0713***	0.0240***	-0.00936***	-0.0161***	0.00148***	-0.00955***
	(0.002)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
temp2	-0.00480***	-0.000229***	0.000582***	0.000405***	-0.000159***	0.000243***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Results

- Tourists are more likely to mention smog issues than residents.
  - The smog issue becomes an integrated part of Beijing's city image, and tourists are more likely to mention it to present their integrated destination experience.

- The impact of air pollution on sentiment is similar between residents and tourists
  - A long-term hazard for residents, and a short-term 'shock' for tourists.

- Tourists are more vulnerable than residents for health-related issues
  - Tourists are less adaptive to an unfamiliar environment.

- Tourists do not suffer from sleep issues due to air pollution.
  - Many other factors shaping the sleep quality of tourists that may be more important than air pollution, such as travel fatigue, accommodation environment, jet lag, etc.

Results

Variables	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
	smog_tag	distance_TAM	radius	sentiment	health_issue	sleep_issue
pm2_5	0.00383***	-0.000290***	-0.000508***	-0.000727***	0.000482***	0.000281*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
pm2_5*lnday_of_tour	0.000187***	-0.000195***	-0.00000287	0.0000526	-0.0000970***	-0.000128*
	0.00383***	-0.000290***	-0.000508***	-0.000727***	0.000482***	0.000281*

- The longer a tourist stays in destination, the larger the impact of air pollution on smog awareness.
- The impact of air pollution on health and sleep issues becomes smaller as the tourist stays longer in Beijing.
  - Tourists started getting used to the polluted environment as they stay longer, and therefore, the impact of air pollution becomes less abrupt.

Results

Variables	Model 13	Model 14	Model 15	Model 16	Model 17	Model 18
	smog_tag	distance_TAM	radius	sentiment	health_issue	sleep_issue
pm2_5	0.00477***	-0.00151***	-0.0000557	0.000923	0.000404	0.000556
	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
pm2_5*lndistance	-0.000112	0.000147**	-0.0000677	-0.000236**	-0.00000774	-0.0000706
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

- Tourists from origins further away Beijing are more likely to visit places in the more remote area of Beijing in the presence of air pollution.
- Tourists traveling a longer distance to Beijing tend to leave a more negative sentiment in the presence of air pollution.
    - High opportunity cost of travel leads to a high expectation.

Results

Variables	Model 19	Model 20	Model 21	Model 22	Model 23	Model 24	Model 25	Model 26
	POI_type= transport	POI_type= parks	POI_type= entertainment	POI_type= education	POI_type= tourist attraction	POI_type= shopping	POI_type =hotels	POI_type= food and beverage
pm2_5 *(tourist=0)	-0.000125***	-0.000608***	0.000386***	-0.0000136	0.0000329	0.000437***	-0.0000661*	0.000367***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
pm2_5 *(tourist=1)	0.000227***	-0.000276***	0.000144***	-0.0000807**	-0.0000160	0.000227***	0.000531***	0.000342***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

- In the presence of air pollution, tourists are less likely to visit parks and more likely to visit entertainment, shopping, and food and beverage sites or stay in hotels to reduce outdoor exposure to pollutants.

- Different reaction patterns were found on Transport sites and Hotels between tourists and residents.

CONTENTS

Conclusion

- We conducted big data analytics on a large sample of Weibo posts to understand how air pollution influences tourist behavior and experience in an urban destination.

- We unveiled the impact of air pollution (measured by the concentration of PM 2.5) on tourists' awareness of smog, behavioral response to air pollution, emotional response to air pollution, and health consequence to air pollution.

- A higher level of air pollution leads to a higher level of smog awareness, more activities closer to the city center, a larger geographic scope of activities in a day, a more negative level of sentiment, and a higher probability of health issue reported.

Conclusion

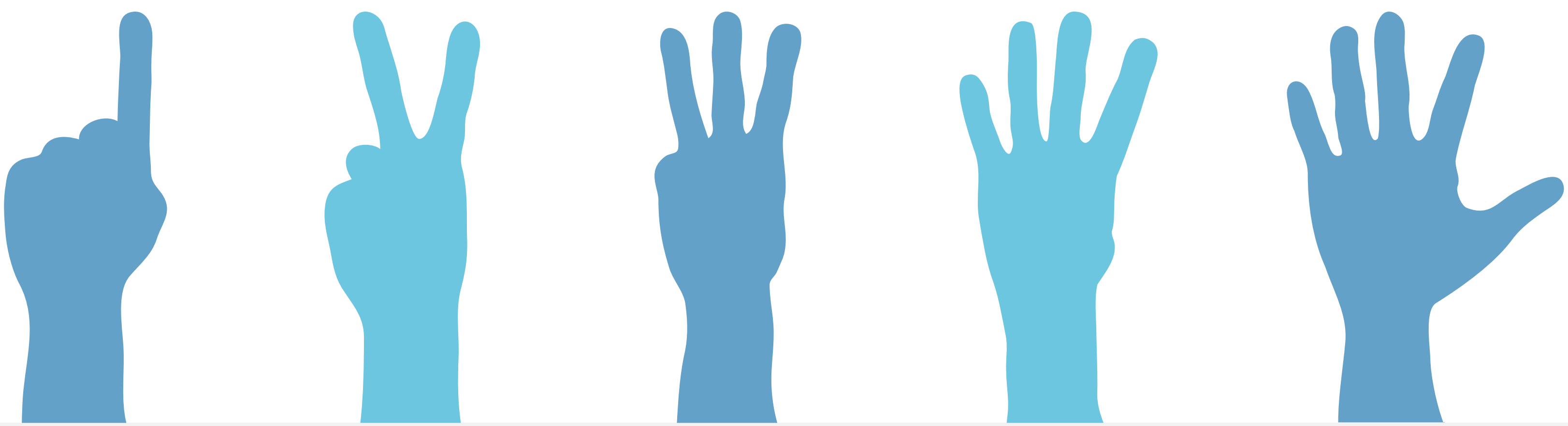
DMOs can better monitor tourist experience in a timely manner.

DMOs can better propose the crisis management plan in the presence of heavy air pollution.

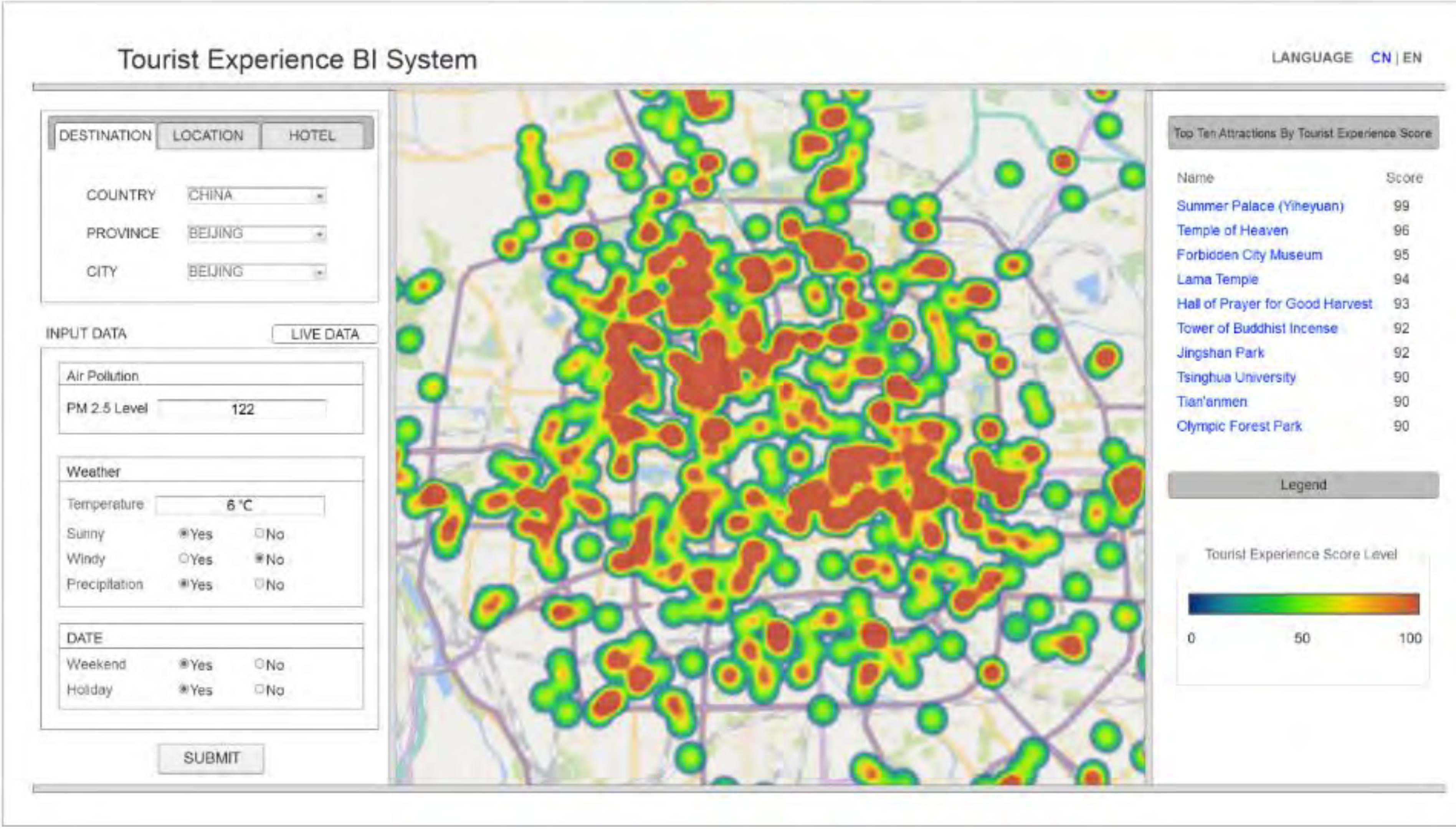
Destinations should provide necessary medicare information to vulnerable tourists in pollution attacks.

Attraction can better forecast the demand based on the level of air pollution

Use of geo-spatial tools analyzing the effects of global change on tourism.



Monitoring | **Crisis Mgmt.** | **Medicare** | **Forecasting** | **Route Design**


Tourist Experience BI System
LANGUAGE CN | EN
DESTINATION
LOCATION
HOTEL
COUNTRY CHINA
PROVINCE BEIJING
CITY BEIJING
INPUT DATA
LIVE DATA
Air Pollution
PM 2.5 Level 122
Weather
Temperature 6 °C
Sunny Yes No
Windy Yes No
Precipitation Yes No
DATE
Weekend Yes No
Holiday Yes No
SUBMIT
Top Ten Attractions By Tourist Experience Score
Name Score
Summer Palace (Yiheyuan) 99
Temple of Heaven 96
Forbidden City Museum 95
Lama Temple 94
Hall of Prayer for Good Harvest 93
Tower of Buddhist Incense 92
Jingshan Park 92
Tsinghua University 90
Tian'anmen 90
Olympic Forest Park 90
Legend
Tourist Experience Score Level
0
50
100

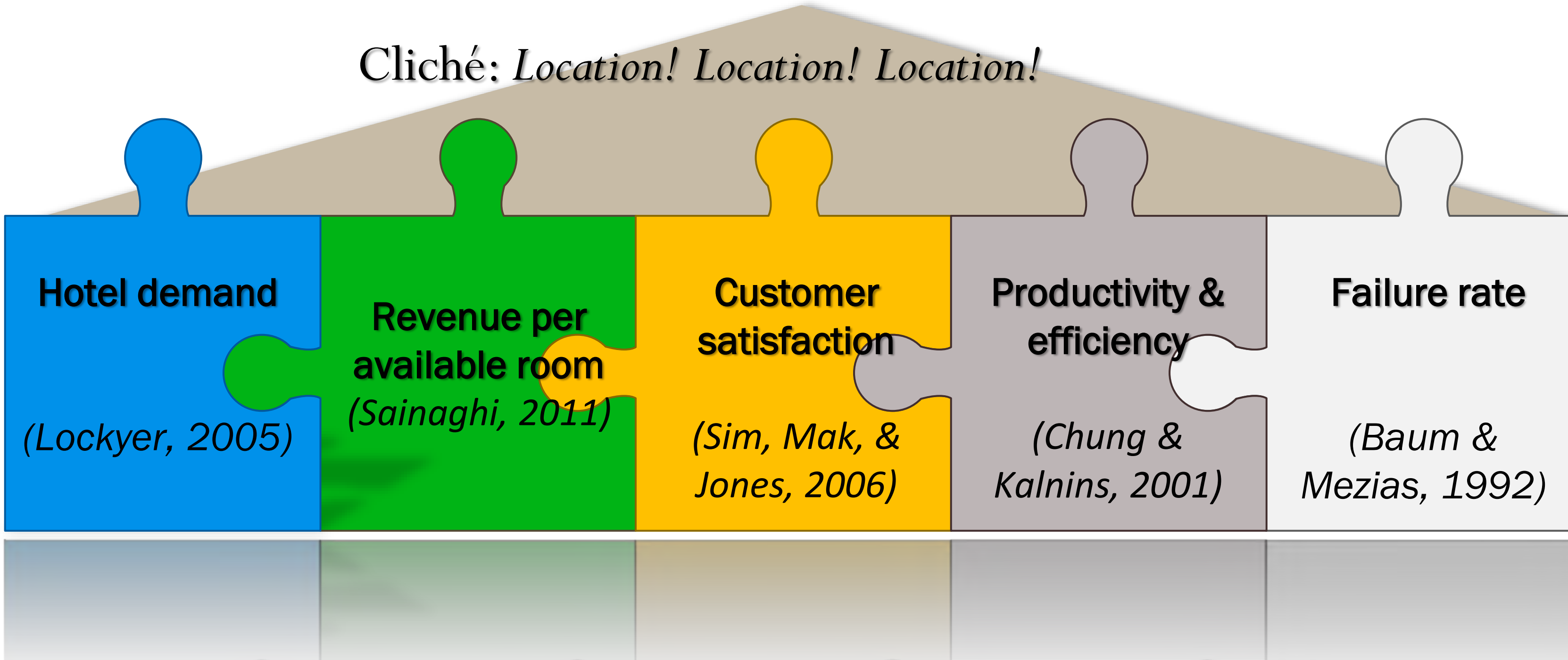
Part II

UNDERSTANDING GUEST SATISFACTION WITH URBAN HOTEL LOCATION: Evidence from big data

Outlines

- Introduction
- Literature
- Method and Data
- Results
- Conclusion and Implication

Background

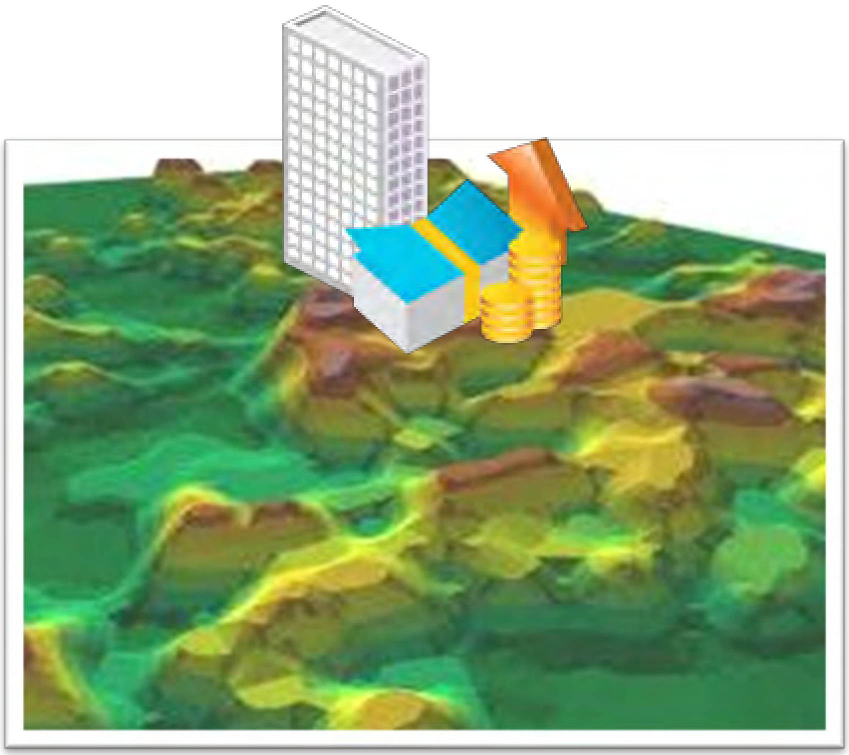


- Once located, difficult to re-locate (Urtasun and Gutierrez, 2006)!

Introduction

- Guests' enjoyment of the experiences and utility embedded in hotel stays is heavily contingent upon hotel location.

- Hotel location as one of the top attributes affecting hotel selection and satisfaction (Lee et al. 2010; Shoval et al. 2011), but very few researchers have scrutinized guest satisfaction with hotel location and its determinants.



Introduction

- Scholars treated location components largely as latent constructs and used subjective self-rated items.
- Location convenience and accessibility can be more objectively evaluated by modern geo-spatial analytical tools
- We use online review data about hotel location to unveil factors determining guest satisfaction with urban hotel location

Introduction

- Potential contributions

1. Based on the gravity location allocation model, we propose a theoretical framework.

2. The location evaluations we use are based on "revealed preferences" instead of "stated preferences."

3. Reveal how to remedy location disadvantages by offering certain types of services.

4. Apply geo-spatial tools such as geographic information system (GIS) and remote sensing applications.

Introduction

- Advantage of online review data:

1. covering a more representative sample of hotel guests based on their actual hotel stays, as well as actual hotel location attributes on "revealed preferences."

2. "stated preferences" used in traditional surveys entangled with so-called laboratory effects associated with sample selection bias and plausible subjective manipulation (Yacouel and Fleischer 2012)

Literature

- Location is the only attribute of a hotel that is relatively fixed (Bull 1994).

- An ideal location is associated with larger accommodation demand, better firm performance and guest satisfaction, and a lower failure rate (Xiang and Krawczyk 2016).

- Scholars with different academic backgrounds such as geography, marketing and economics have utilized a number of theories such as analytic hierarchy process (Chou, Hsu, and Chen 2008), agglomeration theory (Kalnins and Chung 2004), and bid rent theory (Egan and Nield 2000).

Literature

- An exhaustive review of hotel guest satisfaction in the service management literature also suggests that convenient location is rated by travelers as one of the top attributes affecting hotel selection and satisfaction (Li et al. 2015; Magnini, Crotts, and Zehrer 2011; Saleh and Ryan 1992; Tsai, Yeung, and Yim 2011).

- Researchers have explored the constructs or the sub-attributes of hotel location satisfaction in rather limited ways.

Convenient location

Location convenience and accessibility

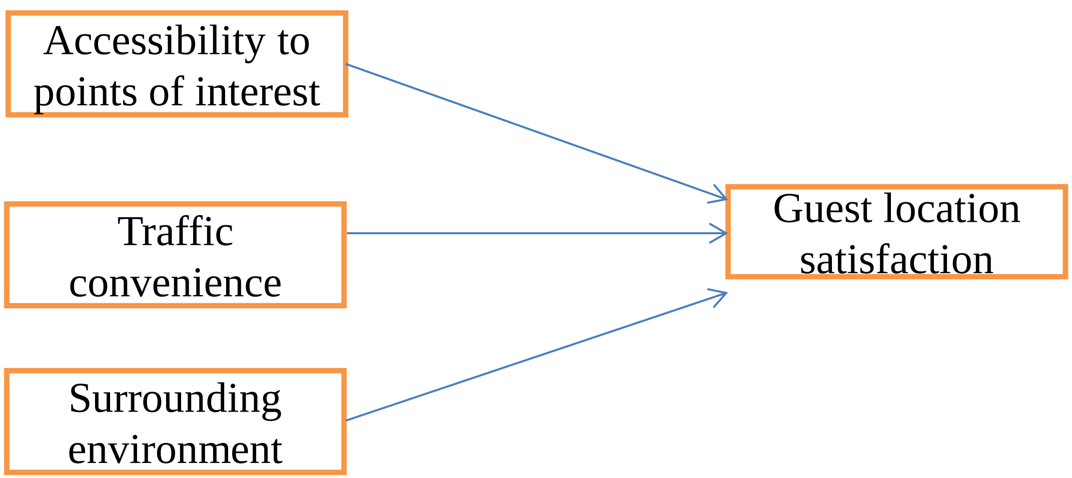
Proximity to public transit

Neighborhood condition

Proximity to nightlife

Literature

- Werczberger and Berechman (1988) proposed a gravity spatial allocation framework for understanding location satisfaction factors for residential houses.



Literature

- *<u>Accessibility to points of interest</u>*
  - Guests prefer a hotel location close to a destination's various points of interest (Weaver 1993), including (Dolnicar and Otter 2003; Lu and Stepchenkova 2012):
    - transportation portals (mainly airports)
    - central business district (CBD),
    - convention centers,
    - tourist attractions,
    - entertainment venues
  - It is an economically rational choice for hotel guests to minimize their transportation costs in terms of time and money.

Literature

- *Traffic convenience*
  - Traffic convenience measures how easy it is for guests leave and return to the hotel location using various types of local transportation.
  - It may reduce guests' travel costs (Canina, Enz, and Harrison 2005).
  - Examples include convenient access to both public transit (such as subways, freeways, local bus and metro service) and private transportation (such as taxi service) (Lee and Jang 2013; Zhou et al. 2014; Li, Ye, and Law 2013).

Literature

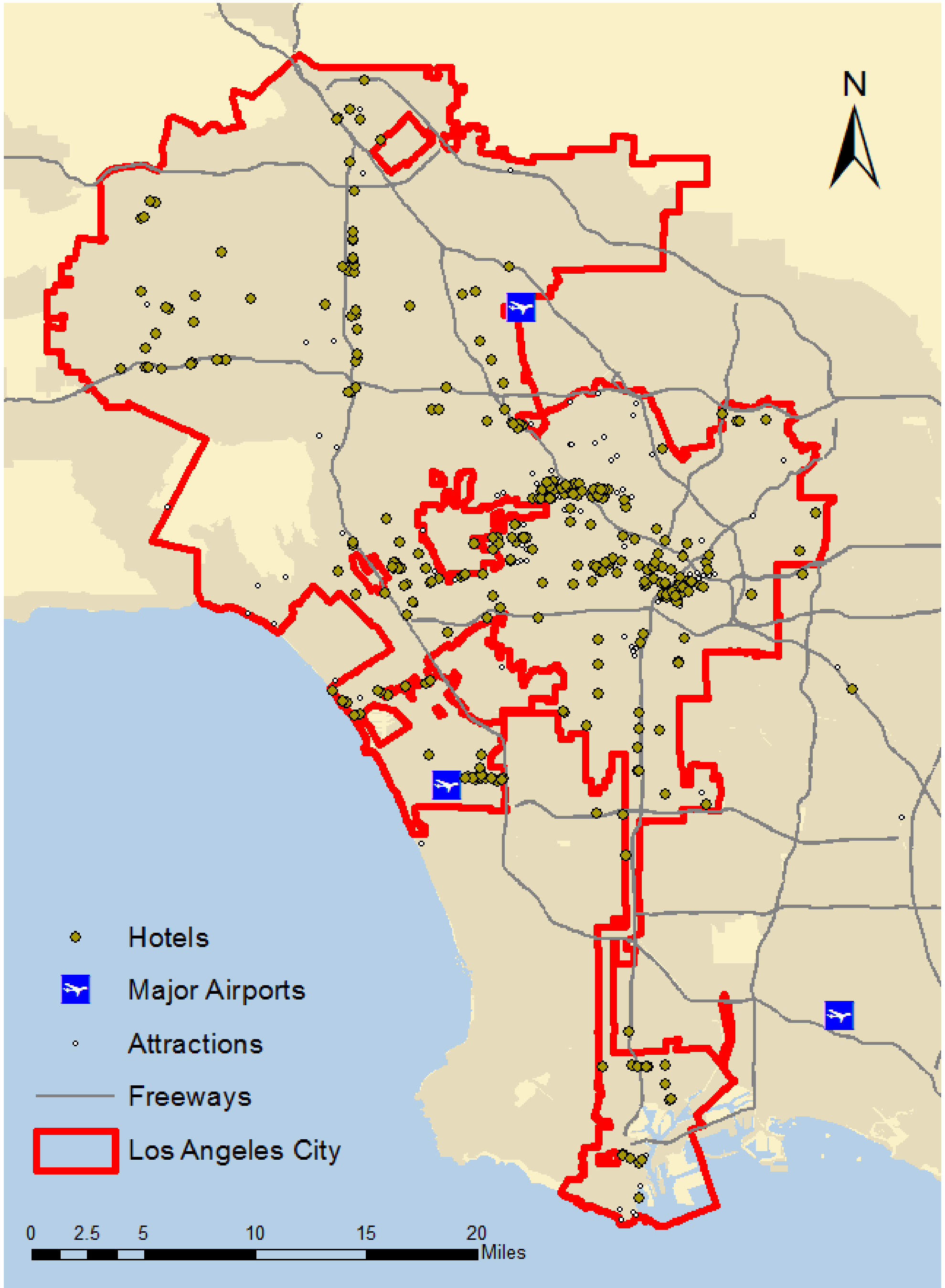
- *Surrounding environment*
  - The location of a property, along with its own local conditions, must be in accord with the surrounding environment (Medlik 1966).
  - A property's surrounding environment includes elements that are seemingly exogenous to the hotel property but endogenous to guests' evaluations of the hotel location (Bull 1994).
  - Examples include, but are not limited to: neighborhood environment (i.e., quietness, air quality and land use on adjoining blocks), public safety and security in the immediate vicinity, public areas and infrastructure (such as restaurants), and diversified culture (Bull 1994).

Model and data

- We chose the city of Los Angeles as our research area; all hotel properties within this area are included in our sample.

- The city of Los Angeles is the second largest city in the United States after New York, currently ranked sixth in the U.S. hotel market.

- The city offers a robust economy with a plethora of tourism activities, people of diverse ethnicities representing a wide range of cultures, various natural and man-made attractions, weather suitable for year-round travel.


N
Hotels
Major Airports
Attractions
Freeways
Los Angeles City
0 2.5 5 10 15 20
Miles

Model and data

- We retrieved hotel online review data from TripAdvisor, one of the most popular online third-party travel intermediaries providing hotel reviews of high credibility.

- Check-in date within one year.

- We collected data on **hotel attributes** (name, address, star rating, overall rating, location, price and rank), **reviewer's rating** (ratings of room quality, service, sleep, cleanliness, location and value), and **reviewer characteristics** (age, gender, number of cities visited, membership level and total TripCollective points).

United States › California (CA) › Los Angeles › Los Angeles Hotels

Four Seasons Hotel Los Angeles at Beverly Hills

1,107 Reviews | #3 of 348 Hotels in Los Angeles | Certificate of Excellence

Hotel website | Hotel packages | 00 1 310-273-2222 | 300 S Doheny Dr, Los Angeles, CA 90048

PriceFinder

travelerfood
California
Level 5 **Contributor**

88 reviews
25 hotel reviews
47 helpful votes

"really perfect"

Reviewed May 18, 2016

Really a wonderful place. William was so helpful and friendly. The place is stunning yet warm and friendly. We had massages and they were excellent. I highly recommend this hotel. Great sleep, great shower and bath, great climate control.

Room Tip: We had a suite which was kind of a weird suite. Recommend normal room

See more room tips

Stayed February 2016, traveled with family

Location
Service
Sleep Quality

Model and data

- We applied a mixed-effect ordered logit model to unveil factors determining the location satisfaction scores.
  - Location satisfaction score is ordinal with five levels, thus ordered logit model families can be used (Agresti 2010).

  - The mixed-effect model is able to incorporate the hierarchical structure of our data, which has reviewer-level information nested within hotel-level information.

$$\Pr(y_{ij} = m \mid \mathbf{x}_{ij}, \mathbf{z}_j, u_j) = \Pr(\tau_{m-1} \le y^* < \tau_m \mid \mathbf{x}_{ij}, \mathbf{z}_j, u_j)$$

$$= F(\tau_m - \mathbf{x}_{ij}\beta - \mathbf{z}_j\delta - u_j) - F(\tau_{m-1} - \mathbf{x}_{ij}\beta - \mathbf{z}_j\delta - u_j)$$

$$y^*_{ij} = \mathbf{x}_{ij}\beta + \mathbf{z}_j\delta + u_j + \varepsilon_{ij}$$

Research Methods

STR hotel census data

Leading data sources on hotel supply at property-level

TripAdvisor attraction data

“Things to do” data with >10 reviews of different types.

Transit score data

A 0-100 score summing the relative "usefulness" of nearby transit routes

Airport data

Airport enplanement data from Bureau of Transportation Statistics

Info Group

Estimated revenue data of local business units.

‘Greenness’ data

Normalized difference vegetation index (NDVI) from the Google Earth

Other Major Data Sources

Model and data

<u>*Accessibility variables*</u>

- *accessibility_attractions*: We weighted the tourist attraction accessibility index by popularity (i.e., number of reviews on the TripAdvisor website). We only kept attractions with a minimum of 10 reviews and eliminated all guided tour items.
- *accessibility_airports*: We weighted the airport accessibility index by passenger traffic. Only these five major airports are used to construct this variable.
- *university*: This is a dummy variable indicating whether the hotel property is located within 1 km of one of the leading research universities in Los Angeles: the University of Southern California, the University of California, Los Angeles, and the California Institute of Technology.

Model and data

Traffic convenience variables

- *metro*: The distance (in km) from to the closest metro station.
- *freeway*: indicates whether the hotel is located next to a freeway in the So uthern California Freeway system.

Surrounding environment variables

- *local_business:* The annual sales volume of all local businesses within 1 km of a hotel (Antipova 2015).
- *NDVI:* The average normalized difference vegetation index (NDVI) value of areas within 2 km of each hotel in 2015 based on remote sensing data that measures the greenness of a neighborhood.
- *water_coverage:* The percentage of the area within 2 km of each hotel covered by water from the 2011.
- *crime_rate*: The per-capita violent crime rates for Los Angeles Police Department- and sheriff-patrolled neighborhoods.

Model and data

Moderating variables

- *shuttle_bus*: Whether or not a hotel offers shuttle bus service to the airport.
- *free_parking*: Whether or not a hotel offers free parking.

Reviewer characteristics variables

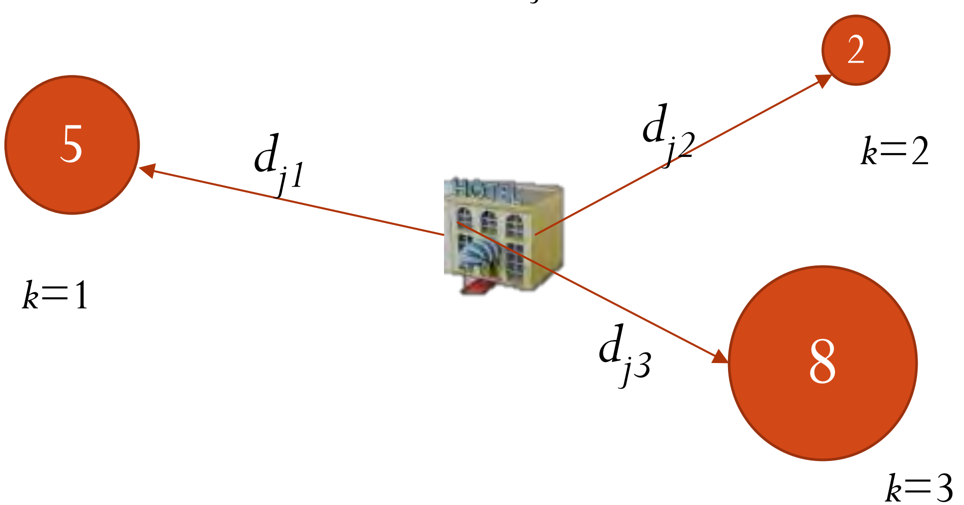
- *lnpoints*: This is the number of TripCollective points (in log) for each reviewer on the TripAdvisor website.
- *traveler_type*: This is a reviewer's traveler type for the hotel stay .
- *month*: This is the month of a reviewer's stay in the hotel reviewed.
- *traveler_age*: This is the age group of the reviewer: 13–17, 18–24, 25–34, 35–49, 50–64, and 65 and above.
- *traveler_gender*: This is the gender of the reviewer

Model and data

- For accessibly indexes, we use a gravity-type accessibility measure specified as:

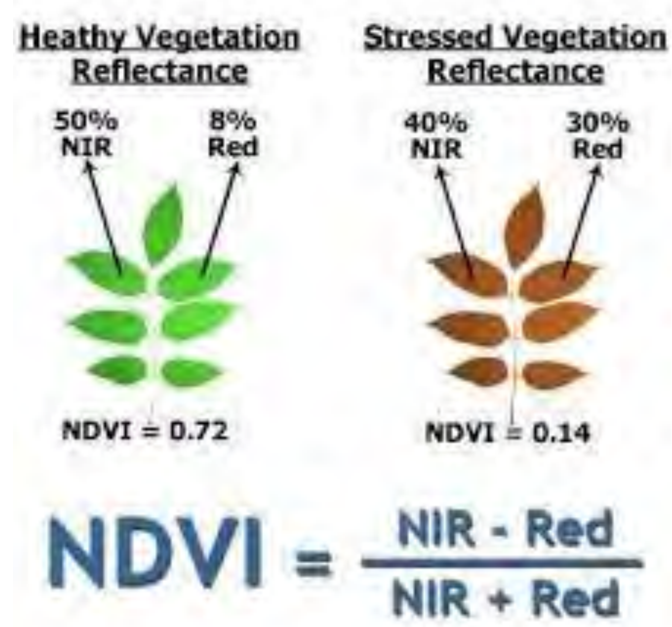
$$accessibility_j = \sum_{k=1}^{K} \frac{w_k}{d_{jk}^2}$$

- where the accessibility of hotel j with respect to a number of K points of interest is specified to be positively associated with the weight of point k, w_k, and negatively associated with the distance between hotel i and point k, d_{jk}



Model and data

- Using the Google Earth Engine platform, calculated the average NDVI values from LandSat 8 with 30 m spatial resolution.
- NDVI is a continuous measure of variation in green vegetation from 0 (bare ground) to 1(completely green).



Variable	Mean	Std. Dev.	Min	Max	Obs
accessibility_attractions	2.159	9.209	0.008	100	220
accessibility_airports	3.309	11.487	0.127	100	220
university	0.014	0.116	0	1	220
metro	2.606	3.269	0.033	14.270	220
freeway	0.227	0.420	0	1	220
local_business	4.195	5.810	0.063	30.107	220
NDVI	0.165	0.066	0.023	0.368	220
water_coverage	0.023	0.104	0	0.694	220
crime_rate	29.395	16.895	1.700	94.800	220
shuttle_bus	0.095	0.295	0	1	220
free_parking	0.436	0.497	0	1	220

Variable	Mean	Std. Dev.	Min	Max	Obs
lnpoints	6.599	1.899	0	12.813	8185
traveler_gender	0.538	0.499	0	1	2398
Variable	**Category**	**Freq.**	**Percentage**	**Cumulated %**	**Obs**
traveler_type=1	Couple travelers	2196	26.83	26.83	8185
traveler_type=2	Business travelers	2059	25.16	51.99	8185
traveler_type=3	Solo travelers	670	8.19	60.17	8185
traveler_type=4	Family travelers	2383	29.11	89.29	8185
traveler_type=5	Travelers with friends	877	10.71	100.00	8185
traveler_age=1	Age 18–24	54	2.60	2.60	2074
traveler_age=2	Age 25–34	398	19.19	21.79	2074
traveler_age=3	Age 35–49	826	39.83	61.62	2074
traveler_age=4	Age 50–64	662	31.92	93.54	2074
traveler_age=5	Age 65 and above	134	6.46	100.00	2074

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
	All	All	All	All	All
accessibility_attractions	0.0206***		0.0204***	0.0205***	0.0206***
	(0.007)		(0.007)	(0.007)	(0.007)
accessibility_airports	0.0250***	0.0225***	0.0517***	0.0252***	0.0250***
	(0.007)	(0.008)	(0.010)	(0.007)	(0.007)
accessibility_airports shuttle_bus*			-0.0267*** (0.008)		
accessibility_airports free_parking*				-0.0156* (0.008)	
*accessibility_attractions*free_parking*					0.00313 (0.220)
university	1.433***	1.009**	1.432***	1.428***	1.433***
	(0.271)	(0.513)	(0.271)	(0.272)	(0.271)
metro	-0.0446**	-0.0737**	-0.0427**	-0.0452**	-0.0446**
	(0.021)	(0.035)	(0.021)	(0.021)	(0.021)
freeway	0.0106	-0.0508	0.0191	0.00513	0.0105
	(0.166)	(0.239)	(0.167)	(0.166)	(0.166)
local_business	0.0662***	0.0422**	0.0665***	0.0653***	0.0662***
	(0.013)	(0.017)	(0.013)	(0.013)	(0.013)
NDVI	9.629***	9.076***	9.711***	9.532***	9.625***
	(1.752)	(2.352)	(1.758)	(1.764)	(1.814)
water_coverage	5.299***	3.737***	5.247***	5.258***	5.298***
	(0.895)	(0.961)	(0.885)	(0.899)	(0.896)
crime_rate	0.00461	0.0175	0.00519	0.00425	0.00458

Results

- *lnpoints* is estimated to be negative, suggesting that more experienced reviewers are more demanding about hotel location.
- Among different traveler types, business travelers (*traveler_type* = 2) are more demanding about hotel location compared to baseline couple travelers (*traveler_type* = 1)
- Younger guests are more demanding about hotel locations and tend to post lower scores for hotel location satisfaction.

- By offering airport shuttle service, hotels with poor airport accessibility can significantly improve the location satisfaction scores posted by guests

Variable	Model 6	Model 7	Model 8	Model 9	Model 10
	Couple travelers	Business travelers	Solo travelers	Family travelers	Travelers with friends
accessibility_attraction	0.0259**	0.0200**	0.0328**	0.0245	0.0186
	(0.010)	(0.009)	(0.015)	(0.020)	(0.015)
accessibility_airports	0.0269***	0.0288**	0.0577***	0.0532***	0.0460***
	(0.010)	(0.011)	(0.014)	(0.011)	(0.014)
accessibility_airports shuttle_bus*	-0.00139	-0.0168**	-0.0420***	-0.0365***	-0.0274***
	(0.007)	(0.008)	(0.009)	(0.007)	(0.010)
university	0.0110	1.887***	0.176	1.809***	1.609
	(0.296)	(0.333)	(1.004)	(0.250)	(1.188)
metro	-0.00997	-0.0474**	-0.0358	-0.0333	-0.0158
	(0.034)	(0.022)	(0.033)	(0.021)	(0.043)
freeway	-0.255	0.393**	-0.370	-0.119	0.131
	(0.196)	(0.173)	(0.263)	(0.191)	(0.288)
local_business	0.0547***	0.0408***	0.0484**	0.0530***	0.0719***
	(0.015)	(0.014)	(0.023)	(0.018)	(0.024)
NDVI	7.456***	3.509*	6.467**	8.996***	8.957***
	(1.784)	(2.043)	(2.921)	(2.258)	(3.085)
water_coverage	4.099***	3.136***	4.608***	3.941***	3.894***
	(0.819)	(0.915)	(1.629)	(1.159)	(1.096)
crime_rate	0.00293	-0.00709	0.00196	0.00527	0.00697
	(0.008)	(0.011)	(0.012)	(0.010)	(0.012)

Results

- Airport accessibility is most important for solo travelers, and the moderating effect of shuttle bus service is most substantial
  - solo travelers have the inherent disadvantage of high per capita transportation costs for car rental and taxi service to airports.

- Travelers with friends tend to engage in more social and entertainment activities (So and Lehto 2007), and a stronger local business presence provides more opportunities for these activities.

- NDVI is highest for family travelers because the tranquility and relaxation of these locations can be particularly family friendly (So and Lehto 2007)

Attr'n type	All	Couple travelers	Business travelers	Solo travelers	Family travelers	Travelers with friends
Museums and theaters						
Amusement and theme parks						
State parks and gardens						
Stadiums and arenas						
Beaches						
Historical attractions						
Shopping attractions						

Results

- Compared to other travelers, solo travelers place a high level of importance on accessibility to attractions less associated with social activities, such as state parks and gardens, beaches, and historical attractions.

- Travelers with friends highly value accessibility to stadiums and arenas; sports and entertainment events hosted by these attractions provide ideal socialization opportunities for travelers of this type.

- Only business travelers with higher travel budgets prefer hotels located close to shopping attractions.

Conclusions

- Overall, our findings empirically support the importance of accessibility to points of interest, traffic convenience and surrounding environment to guest satisfaction with hotel location.

- Offering free parking and airport shuttle service could mitigate the hotel location problem related to a hotel’s inferior airport accessibility.

- Different types of travelers seem to have heterogeneous location preferences toward different tourist attractions.

Implications

- Hotels with poor airport accessibility can mitigate their location disadvantages by providing airport shuttle service
- Hotels are advised to work together with their local communities to improve the natural environment of their neighborhoods and foster a better business environment to encourage local commerce
- Hotel operators need to emphasize different location advantages when communicating to different target market segments
- Factors unveiled can be incorporated into an evaluation tool for selecting new hotel locations and used in competitive assessments

Part III

Sleepless nights in hotels? ***Understanding tourist' sleep quality from online reviews***

Outlines

- Introduction
- Literature
- Method and Data
- Results
- Conclusion and Implication

Introduction

- Sleep is an important physiological state for humans; although nobody really knows why, human beings simply cannot live without it (Jones, 2015).

- Sleep is a significant part of travel and hotel stays. Great sleep has even become a form of hedonic consumption for travelers (Valtonen and Moisander, 2012).

- In most tourism and hospitality studies, scholars have briefly acknowledged the importance of sleep to travelers without engaging in further analysis.

Introduction

- Objective:
  - investigate how unique travel- and hotel-related factors (facilitators and barriers) contribute to subjective sleep quality for travelers who stay overnight in hotels.

- Method
  - we collected a large corpus of TripAdvisor review data for hotels in Los Angeles to explore and unveil how different factors help influence the level of sleep quality rated by reviewers.

Introduction

- Potential contributions

1. we develop a conceptual framework to classify determinants of hotel sleep quality
2. we analyze online review data from TripAdvisor, which contain a rich and representative sample of information generated directly by hotel guests
3. we employ a big data approach, using sentiment analysis, geographic information system (GIS) techniques, remote sensing applications and econometric models to quantify and analyze qualitative online review data.

Literature

- Sleep is a necessary biological mechanism by which human beings recuperate (Carskadon and Rechtschaffen, 2000; Roth, 2003).

- Sleep has different dimensions, such as duration, continuity (i.e., amount and distribution of sleep and wakefulness), and architecture (i.e., stages of sleep) (Mezick, et al., 2008)

- Sleep is fragile, and sleep quality is affected by myriad factors (Roth, 2003).

Literature

- We propose a framework/typology to classify potential sleep factors among hotel guests into two meaningful domains, traveler characteristics and hotel characteristics

Literature

- *Demographic factors*
- Demographic differences such as age, gender and race/ethnicity demonstrate significant effects on sleep quality (Basner, Spaeth, and Dinges, 2014).

- Sharma and Panda (2014) found that middle-aged working people suffer from poor sleep quality more than their younger and elderly counterparts.

- Most research has confirmed that on average and across the whole age range, females sleep significantly better than males (Basner, et al., 2014).

Literature

- *Biopsychosocial factors*
- including but not limited to physical and psychological health, family structure, social support, personal relationships and sleep habits (Knutson, 2013).

- Poor physical health can disturb sleep though physiologic arousal, which interferes with good sleep hygiene (Krueger and Friedman, 2009)

- Psychosocial disorders (e.g., depression, stress, loneliness) may also impair sleep by increasing physiological, psychological and pathological arousal (Knutson, 2013),

Literature

- *Tripographic factors*
- Long haul travel and/or uncomfortable modes of transportation could result in travel fatigue, which in turn, could negatively affect sleep quality (Silva, Pascoal, Silva, and Paiva, 2016).
- Sleep quality for those who travel with family members and kids may be affected by others staying in the same room, whereas guests traveling by themselves may not be physically bothered by other people when they go to sleep
- Experienced travelers have accumulated adequate experience addressing sleep problems

Literature

- *Hotel location*
- Low neighborhood quality (such as noise, lack of green space, unpleasantness, etc.) can lead to poor sleep quality due to disruptive sleeping conditions and poor housing environments (Hale and Do, 2007; Hale, et al., 2013).
- Green spaces and parks in urban areas as noise reducers (Gidlöf-Gunnarsson and Öhrström, 2007; Wells, Evans, and Cheek, 2016), which may help improve the sleep of those who are nearby

Literature

- *Hotel facilities*
- Low quality buildings are likely to have been built using cheap construction materials that do not insulate the indoor environment from noise.

- Hotel star rating, a measure of the quality of a hotel’s rooms and service, appears to be significantly associated with reviews related to sleep quality (Park and Nicolau, 2017).

Literature

- *Hotel sleeping environment*
- Environmental factors including lighting, noise, bedding, room temperature, air quality and so forth.

- The thermal environment, including the mattress and bedding (e.g., sheets, duvet and pillow), play a vital role in sleep quality (Lin and Deng, 2008).

- Mattress supports the human body and distributes weight in a way that allows the muscles and intervertebral disks to recover (LeBlanc, Evans, Schneider, Wendt Iii, and Hedrick, 1994), the bedding can help maintain a comfortable thermal environment (Lin and Deng, 2008).

Outlines

- Introduction
- Literature
- **Method and Data**
- Results
- Conclusion and Implication

Model and data

- We chose the city of Los Angeles as our research area; all hotel properties within this area are included in our sample.

- Most TripAdvisor reviewers are randomly selected to rate their sleep quality and provide other sub-ratings after providing an overall rating of their hotel experience.

- We collected data by crawling TripAdvisor using an automated program that obtains the content and rating of each review and reviewer characteristics. The program is able to crawl the hotel characteristics.

Jakarta, Indonesia
Level 6 Contributor
903 reviews
307 hotel reviews
720 helpful votes

"We hope it is better next time"

Reviewed 1 week ago

We looked forward to another delightful stay at Courtyard LAX after a great overnight in 2013. But the first shuttle van passed us up at Bradley Terminal and the air vent in our room 628 blew directly on to the bed. We were too tired to pack up and change rooms, so we wound up sleeping with the window open despite the noise of aircraft landing and taking off. Check-out in the morning was fast, and the shuttle bus took us right to the airport without waiting for other passengers.

Stayed February 2017, traveled as a couple

Sleep Quality
Cleanliness
Service

Level 6 Contributor

Report

- TripAdvisor member since 2006
- 65+ man from Jakarta, Indonesia

Review distribution (900)

Excellent (448)
Very good (326)
Average (106)
Poor (21)
Terrible (2)

1,705 Contributions
720 Helpful votes
287 Cities visited

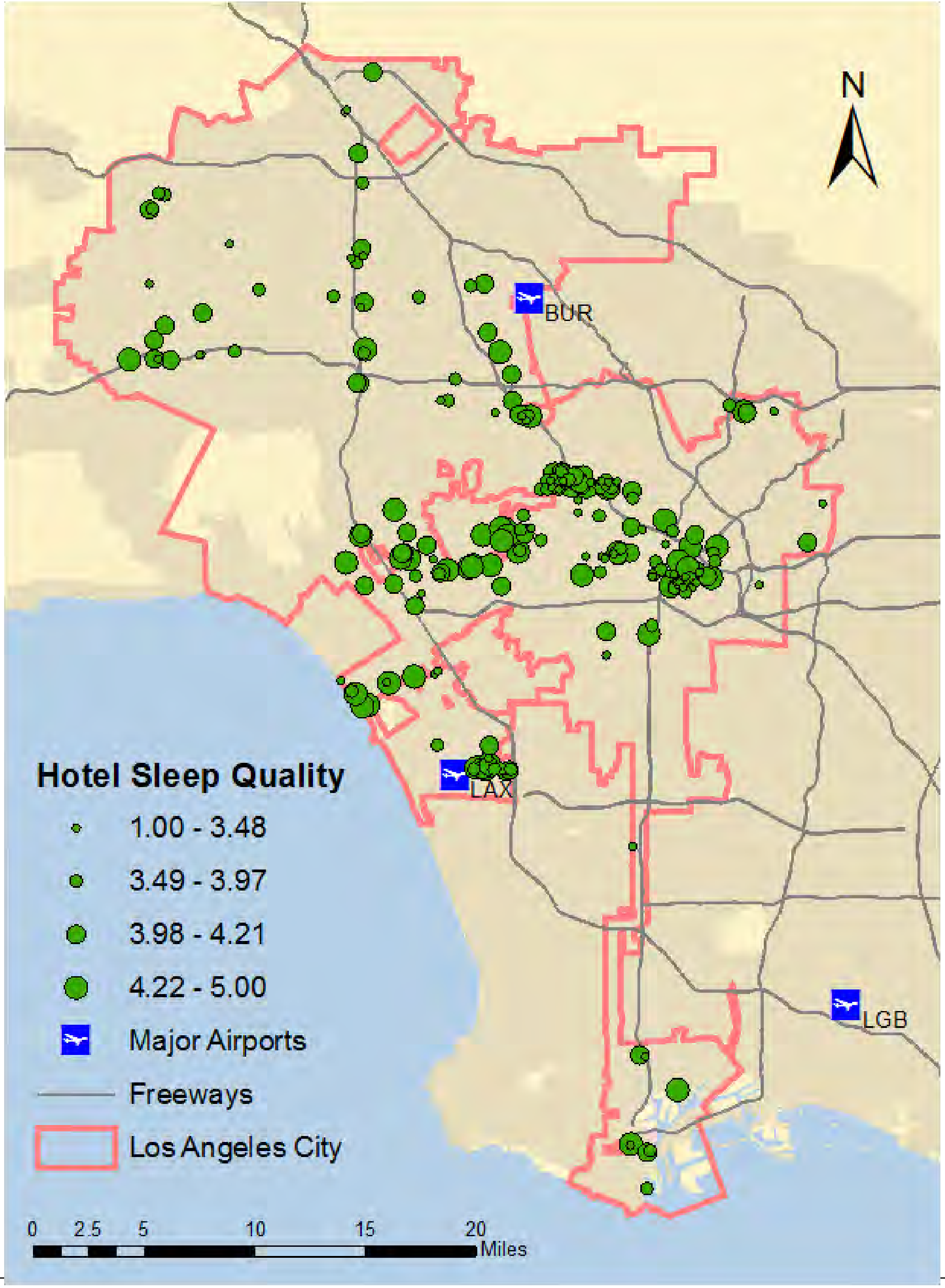
Message
Full profile

Jakarta, Indonesia
Level 6 Contributo
903 reviews
307 hotel reviews
720 helpful votes

X after a
p at
tly on to
we
f aircraft
d the
er

Stayed February 2017, traveled as a couple

Sleep Quality
Cleanliness
Service


N
BUR
LAX
LGB
Hotel Sleep Quality
1.00 - 3.48
3.49 - 3.97
3.98 - 4.21
4.22 - 5.00
Major Airports
Freeways
Los Angeles City
0 2.5 5 10 15 20
Miles

Model and data

Sentiment analysis of sleeping environment

- Apply three-step text mining process to convert unstructured text data into structured data (Zhou, Wang, and Li, 2017).
  - Identify sleep-relate text using a list of keywords associated with sleeping environment factors based on results reported by Pallesen, et al. (2016)
  - Use the Natural Language Toolkit to break the text of reviews into words and/or other meaningful elements (Bird, Klein, and Loper, 2009).
  - Perform a sentiment analysis based on the keywords defined in the first step to get sentiment scores (Philander and YunYing, 2016).

Variable	Definition
sleep_quality	Self-reported evaluation of sleep quality from 1 to 5
Demographic characteristics	
traveler_age	Traveler age: 1 = 18–24; 2 = 25–34; 3 = 35–49; 4 = 50–64; 5 = 65 and above
traveler_gender	Traveler gender: 1 = male; 2 = female.
Tripographic characteristics	
traveler_type	Traveler type: 1 = couple travelers; 2 = business travelers; 3 = solo travelers; 4 = family travelers; 5 = travelers with friends
ln*cities*	Log of number of different cities visited as shown in reviewer profile
ln*distance*	Log of geographical distance between the home city and Los Angeles (in km
longitude_dif	Difference in longitude between the home city and Los Angeles
Hotel facilities	
floors	Number of floors of the hotel's major building with guest accommodations
star	Star rating for the hotel on TripAdvisor
Hotel location	
restaurant	Number of restaurants within 1 km of the hotel
freeway	An indicator of a freeway within 200 m of the hotel
airport	An indicator of an airport within 2 km of the hotel
canopy_cover	Tree canopy cover percentage within 500 m of hotel
Control variable	
expertise	Reviewer's expertise class on TripAdvisor from 0 to 6.

Model and data

- We geo-coded each hotel's address using GIS software and performed spatial analysis to generate relevant location-specific variables

- We measured the restaurant density around each hotel using the restaurant business data from ESRI's 2015 Business Analyst database.

- We measured the greenness and serenity of each hotel's neighborhood using the National Land Cover Database (NLCD) 2011 to calculate tree canopy cover (Sander, Polasky, and Haight, 2010).

Model and data

Variable	Obs	Mean	Std. Dev.	Min	Max
lncities	9,555	4.100	1.239	0.693	8.342
lndistance	6,566	7.596	2.255	0	9.723
longitude_dif	6,566	53.625	49.041	0	173.682
floors	229	5.760	5.902	1	35
star	229	2.731	0.798	1	5
restaurant	229	99.362	96.775	1	474
highway	229	0.236	0.425	0	1
airport	229	0.022	0.021	0	1
canopy_cover	229	4.636	13.592	0	100
expertise	9,555	4.306	1.462	0	6

Variable	Category	Freq.	Percentage	Cumulative %	Obs
sleep_quality = 1		273	2.86	2.86	9,555
sleep_quality = 2		417	4.36	7.22	9,555
sleep_quality = 3		1,490	15.59	22.82	9,555
sleep_quality = 4		3,164	33.11	55.93	9,555
sleep_quality = 5		4,211	44.07	100	9,555
traveler_type = 1	Couple travelers	3,210	33.59	33.59	9,555
traveler_type = 2	Business travelers	2,227	23.31	56.9	9,555
traveler_type = 3	Solo travelers	807	8.45	65.35	9,555
traveler_type = 4	Family travelers	2,329	24.37	89.72	9,555
traveler_type = 5	Travelers with friends	982	10.28	100	9,555
traveler_age = 1	Age 18–24	275	2.88	2.88	9,555
traveler_age = 2	Age 25–34	1,957	20.48	23.36	9,555
traveler_age = 3	Age 35–49	3,581	37.48	60.84	9,555
traveler_age = 4	Age 50–64	3,002	31.42	92.26	9,555
traveler_age = 5	Age 65 and above	740	7.74	100	9,555
traveler_gender = 1	Male	5,081	53.18	53.18	9,555
traveler_gender = 2	Female	4,474	46.82	100	9,555

Results

- We searched the review content using the defined list of sleep-related keywords associated with hotel sleeping environment (Pallesen, et al., 2016).

- We also manually excluded irrelevant reviews like those highlighting pool temperature and lighting problems when taking showers.

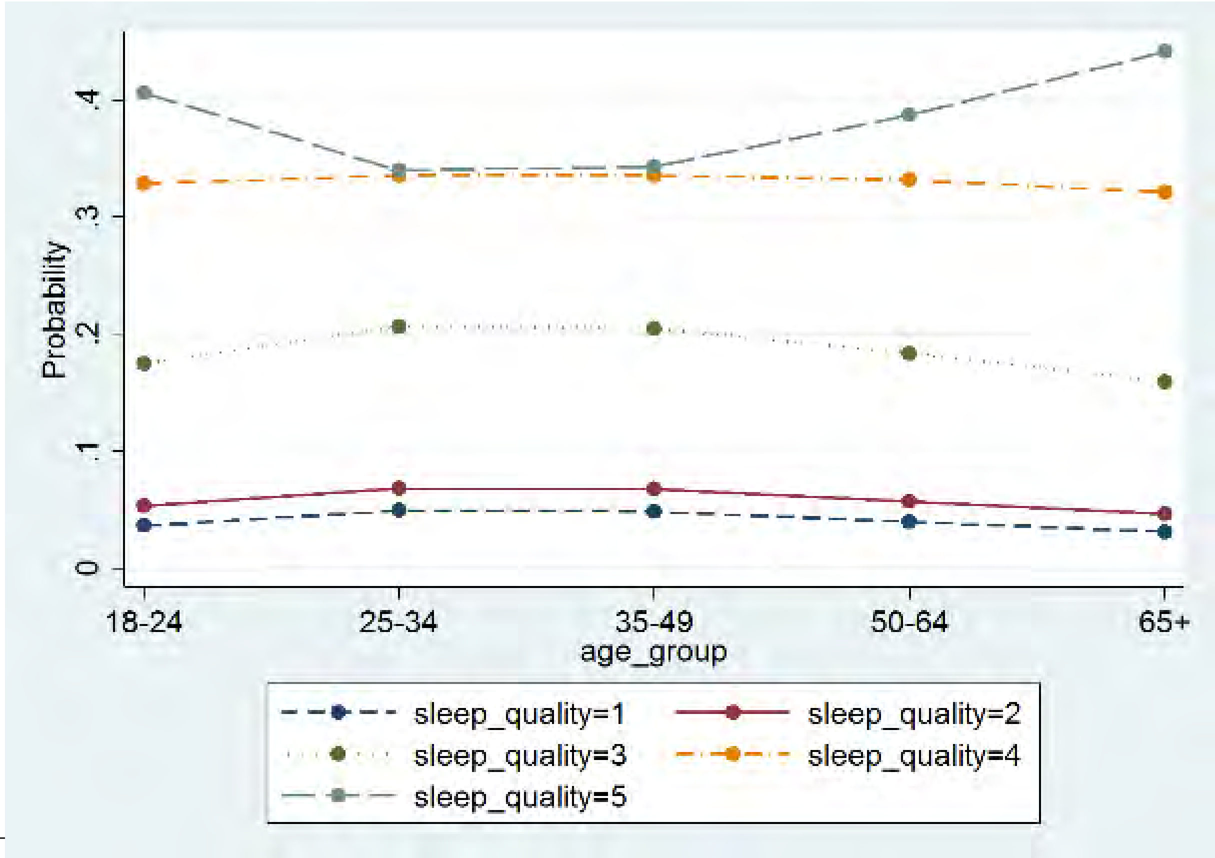
- The results demonstrate the applicability and accuracy of using sentiment analysis to generate an evaluation measure of each aspect related to sleeping environment in a hotel setting.

Factor	Observations	Mean of score	Std. Dev. of score	Example
Pillow	373	0.306	0.270	"Beds and pillows are glorious :)" (score = 0.500) "The one thing I hate is tons of pillows on the bed that are all down" (score = -0.176)
Mattress	127	0.205	0.306	"Very comfortable beds with mattress cushion covers" (score = 0.560) "The comfort level of the mattress was so poor that my sleep was terrible" (score = -0.548)
Bedding	170	0.264	0.296	"Very comfortable bedding" (score = 0.523) "The carpet is dingy shag and the bedding is outdated" (score = -0.400)
Temperature	251	0.162	0.268	"Room temperature - I admittedly like my room on the cold side but it seems as though the insulation of our room was weak and by mid-afternoon it got very warm" (score = 0.041) "The temperature was impossible to control" (score = -0.483)
Noise	1337	0.146	0.233	"I expected some street noise but was surprised how good the sound proofing was" (score = 0.431) "The noise was going on all night with doors being slammed and loud talking in the corridors" (score = -0.236)
Lighting	225	0.227	0.285	"It was nice to have a dimmer so we could have the light on at night without being too intrusive" (score = 0.313) "The noise is pretty bad and also the lights from the Holiday Inn sign creates a green glow into the room" (score = -0.217)
Ventilation	39	0.095	0.242	"On the upside, the ventilation in the room was terrific" (score = 0.407) "Poor ventilation which kept the bathroom hot and steamy after showering" (score = -0.250)

	Model 1	Model 2	Model 3	Model 4	Model 5
traveler_age = 1	0.327***	0.402***	0.402***	0.339***	0.342***
	(0.117)	(0.138)	(0.138)	(0.117)	(0.116)
traveler_age = 3	0.0176	0.0657	0.0668	0.0134	0.0131
	(0.053)	(0.067)	(0.067)	(0.053)	(0.053)
traveler_age = 4	0.235***	0.242***	0.244***	0.234***	0.232***
	(0.062)	(0.068)	(0.068)	(0.062)	(0.062)
traveler_age = 5	0.494***	0.536***	0.539***	0.501***	0.497***
	(0.091)	(0.114)	(0.113)	(0.091)	(0.091)
traveler_gender = 2	0.219***	0.210***	0.211***	0.223***	0.223***
	(0.037)	(0.042)	(0.043)	(0.037)	(0.037)
traveler_type = 2	-0.297***	-0.282***	-0.280***	-0.313***	-0.313***
	(0.056)	(0.065)	(0.064)	(0.055)	(0.055)
traveler_type = 3	0.309***	0.356***	0.357***	0.317***	0.315***
	(0.087)	(0.090)	(0.090)	(0.087)	(0.087)
traveler_type = 4	-0.128**	-0.0989	-0.0984	-0.122*	-0.124**
	(0.062)	(0.066)	(0.066)	(0.063)	(0.063)
traveler_type = 5	0.189***	0.166**	0.167**	0.187***	0.190***
	(0.070)	(0.079)	(0.079)	(0.070)	(0.070)
lncities	0.00663	-0.0111	-0.0110	0.00667	0.00666
	(0.020)	(0.025)	(0.025)	(0.020)	(0.020)
expertise	-0.0539***	-0.0544***	-0.0546***	-0.0557***	-0.0559***
	(0.016)	(0.020)	(0.020)	(0.016)	(0.016)
lndistance		0.00145			
		(0.011)			
longitude_dif			0.000164		
			(0.000)		

Results

- Guests in the 18–24 and 50 and older age groups reported better sleep quality compared to the reference group, with the 65 and older age group reporting the highest quality.



Results

- The odds for female guests reporting a higher sleep quality level are 24.5% (exp(0.219) -1) times higher than the odds for male guests.

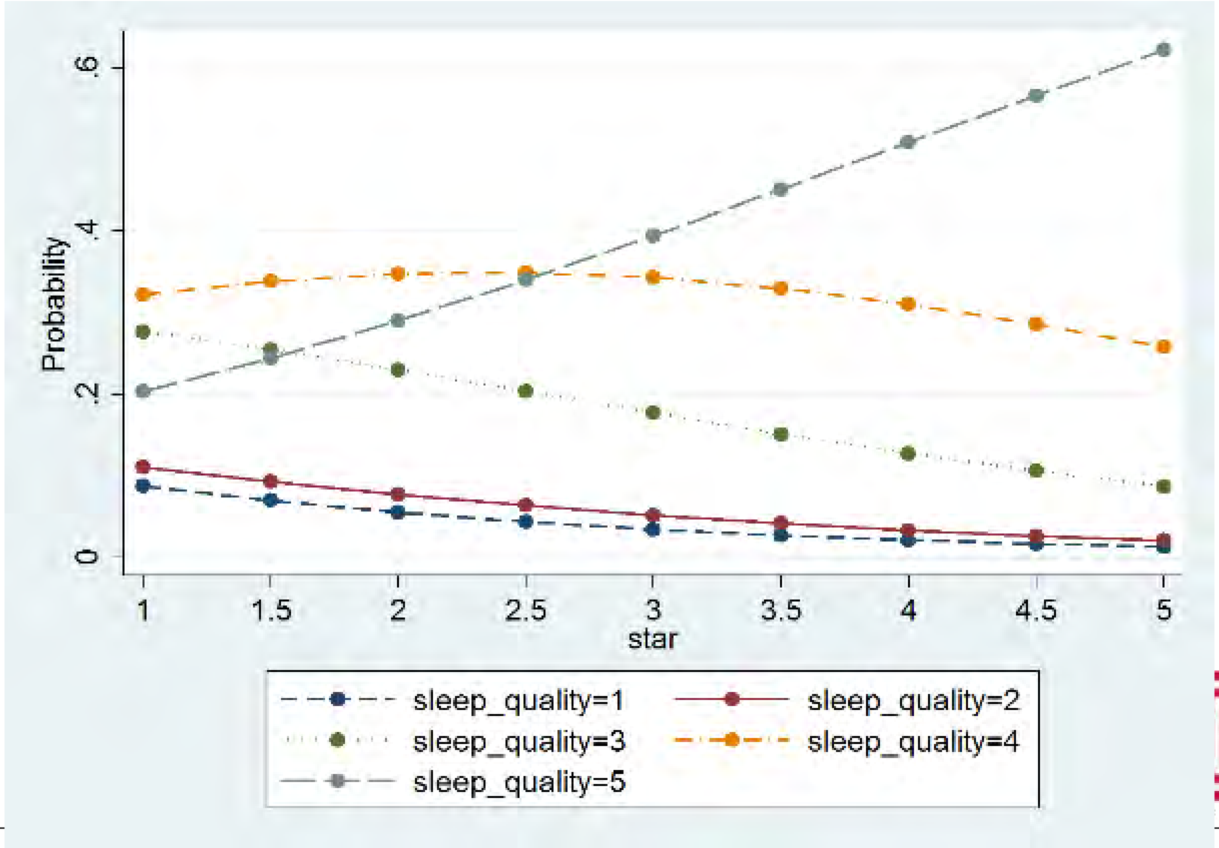
- Business travelers (traveler_type = 2) and family travelers (traveler_type = 4) have 25.7% (exp(-0.297)-1) and 12.0% (exp(-0.128)-1) lower odds of reporting higher-level sleep quality, respectively; whereas solo travelers (traveler_type = 3) and travelers with friends (traveler_type = 5) have 36.2% (exp(0.309)-1) and 20.8% (exp(0.189)-1) higher odds of doing so, respectively.

- The influence of travel distance/time zone on sleep quality is insignificant.

	Model 1	Model 2	Model 3	Model 4	Model 5
floors				0.0196*	0.0255**
				(0.011)	(0.011)
star				0.511***	0.512***
				(0.104)	(0.101)
restaurant					-0.00117**
					(0.001)
freeway					-0.0587
					(0.126)
airport					-0.0533
					(0.374)
canopy_cover					0.0000147
					(0.003)
Monthly effect	Controlled	Controlled	Controlled	Controlled	Controlled
Yearly effect	Controlled	Controlled	Controlled	Controlled	Controlled
var(μ)	0.646***	0.582***	0.582***	0.366***	0.361***
	(0.115)	(0.116)	(0.116)	(0.066)	(0.064)
N (observations)	9555	6566	6566	9555	9555
N (hotels)	229	222	222	229	229
AIC	23167.2	16009.2	16009.1	23093.1	23097.4
BIC	23375.0	16212.9	16212.8	23315.3	23348.1
ll	-11554.6	-7974.6	-7974.5	-11515.6	-11513.7

Results

- One more floor in a hotel building leads to 2% (exp(0.0196) -1) higher odds of reporting a higher sleep quality level
- One more star leads to 66.7% (exp(0.511) -1) higher odds of doing so



Results

- Only restaurant is estimated to be statistically significant, suggesting that a neighborhood with a denser restaurant population could be associated with lower sleep quality ratings.

- We do not find any empirical evidence to support the role of freeway proximity, airport proximity and tree canopy cover in determining hotel guests' sleep quality in general.

	Model 6	Model 7	Model 8	Model 9
I_pillow	0.161	-0.550***	0.137	-0.558***
	(0.134)	(0.192)	(0.133)	(0.191)
I_mattress	-0.991***	-1.604***	-0.993***	-1.590***
	(0.234)	(0.267)	(0.237)	(0.273)
I_bedding	0.00114	-0.233	-0.0592	-0.272
	(0.192)	(0.247)	(0.198)	(0.255)
I_temperature	-0.744***	-0.989***	-0.738***	-0.986***
	(0.128)	(0.126)	(0.128)	(0.132)
I_noise	-0.750***	-0.898***	-0.764***	-0.903***
	(0.071)	(0.081)	(0.073)	(0.082)
I_lighting	-0.211*	-0.476***	-0.229*	-0.495***
	(0.128)	(0.153)	(0.134)	(0.160)
I_vent	-0.769***	-0.725**	-0.795***	-0.749**
	(0.259)	(0.285)	(0.266)	(0.297)
S_pillow		2.473***		2.431***
		(0.461)		(0.468)
S_mattress		2.974***		2.895***
		(0.731)		(0.757)
S_bedding		0.800		0.678
		(0.557)		(0.567)
S_temperature		2.131***		2.075***
		(0.483)		(0.486)
S_noise		1.213***		1.154***
		(0.222)		(0.222)
S_lighting		1.284***		1.304***
		(0.415)		(0.436)

Results

- We use the prefix "I_" to indicate that a sleeping environment factor was mentioned, and the prefix "S_" to represent the sentiment score for this factor after sentiment analysis.

- Estimates for other "S_" prefixed variables—S_mattress, S_temperature, S_noise and S_lighting—are positive and significant, suggesting that more positive evaluations of mattresses, temperature, noise and lighting are associated with a higher levels of reported sleep quality.

- S_mattress has the largest coefficient, which again highlights the important role mattresses play in determining guests' sleep quality.

Conclusions

- We unveiled a U-shaped relationship between age and sleep quality ratings, and found that females are more likely to report better sleep quality. Solo travelers reporting the highest level of sleep quality in general, and business travelers reporting the lowest.
- Our results do not provide any evidence to support travel distance and jet lag as factors that negatively affect sleep quality
- Restaurant density around the hotel, the number of floors in a hotel building and hotel star ratings are associated with travelers' sleep quality at the hotel level
- Our findings reveal the importance of mattresses, pillows, room temperature, noise and lighting, with mattresses having the largest impact.

Implications

- Our study thus complements and enriches the current sleep research on the priority and order of these environmental factors on sleep quality for people in general, and travelers in particular.

- We crafted novel ways to incorporate and convert unstructured data from online reviews into structured variables that can be used in econometric models.

- Our study significantly improves practical understandings and awareness of how various trip- and hotel-related factors affect travelers' sleep.

Other examples

Does time dull the pain?

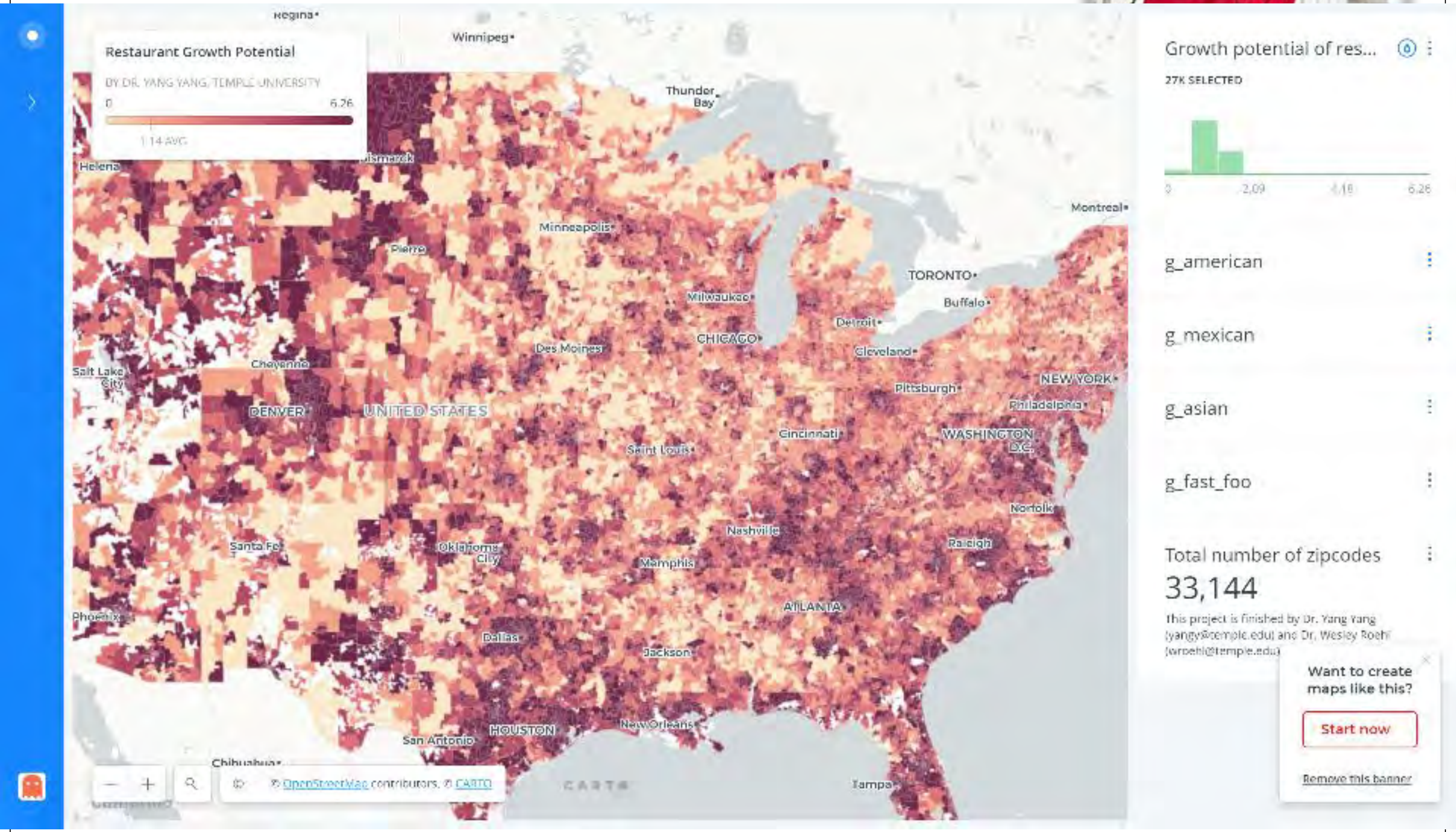



Dec 14, 2019
5th World Research Summit for Tourism and Hospitality
Orlando, FL

From zero to hero: A journey toward an experienced online reviewer

Yang Yang
Xiaowei Zhang
Department of Tourism and Hospitality Management
Temple University(yangy@temple.edu)

TEMPLE UNIVERSITY
School of Sport, Tourism and Hospitality Management



https://sites.temple.edu/yangyang/map/maps-projecting-u-s-restaurantscape/

Findings

www.dryangyang.com

yangy@temple.edu