Promising New Indicators for Tourism Statistics

By: Setia Pramana
Associate Professor
Politeknik Statistika STIS
BPS Statistics Indonesia

Delivered on:
UNESCAP Stat Cafe
Time: 27 August, 2021
BIG DATA SOURCES

**Exhaust data**
- Mobile phone data
- Financial transactions
- Online search and access logs
- Citizen card
- Postal data

**Sensing data**
- Satellite and UAV imagery
- Sensors in cities, transport and homes
- Sensors in nature, agriculture and water
- Wearable technology
- Biometric data
- Internet of Things (IoT)

**Digital Content**
- Social media data
- Web scraping
- Participatory sensing / crowdsourcing
- Health records
- Radio content

**What People Do**

**What People Say**
Online Booking site: Booking.com and Agoda.com

Review Sites: TripAdvisor

Google Trend

Google map
Online Booking Sites

Tourism statistics can be retrieved from various online accommodation booking service sites. These sites have quite diverse data such as the number of rooms available, rental prices, and other information.

Data Sources:

Data Acquisition Technique
Web scraping

Agoda: Collected Variables

01 Accommodation ID
02 Accommodation Name
03 Accommodation Type
04 Date of Collection
05 Location
06 Number of Rooms available
07 Total number of rooms
08 Room Prices
09 Star Rating

\[
ROR = \frac{\text{Total rooms} - \text{available rooms}}{\text{Total rooms}} \times 100
\]
TripAdvisor: Collected Variables

01
Name
Name of the reviewer

02
Title
Feedback to the accommodation from the reviewer

03
Date
Date of the reviewer stay in that accommodation

04
Review_text
Explanation about the accommodation from the reviewer (can be positive feedback or negative feedback/improvement feedback)
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>url</td>
<td>Link page of the review</td>
</tr>
<tr>
<td>place</td>
<td>Place that the reviewer visit</td>
</tr>
<tr>
<td>reviewer_name</td>
<td>Name of the reviewer</td>
</tr>
<tr>
<td>reviewer_country</td>
<td>Reviewer country of origin</td>
</tr>
<tr>
<td>room</td>
<td>Type of room that the reviewer used</td>
</tr>
<tr>
<td>long_stay</td>
<td>Duration of the stay from the reviewer</td>
</tr>
<tr>
<td>time_stay</td>
<td>Time of month when the reviewer visit</td>
</tr>
<tr>
<td>review_reaction</td>
<td>Reaction or comment feedback of the reviewer about the accommodation</td>
</tr>
<tr>
<td>review_score</td>
<td>Score given to the accommodation (1-10)</td>
</tr>
<tr>
<td>like_comment</td>
<td>Positive comment feedback from the reviewer about the accommodation</td>
</tr>
<tr>
<td>dislike_comment</td>
<td>Negative comment feedback/improvement feedback from the reviewer about the accommodation</td>
</tr>
</tbody>
</table>
1. Google Trends provides time-series data in the form of an index that shows the volume of queries entered by users in a region.

2. Users can set their search data to start from the last hour or another range with a maximum for Google Trend data ranging from 2004 to the present. However, this maximum range is only available in the form of monthly data.

3. It is a search ratio ranges from 100 to zero, where 100 indicates that the search term is relatively popular in the selected region.

4. According to Google, this normalization allows for smoother comparisons among search terms as search volume varies across different countries.

5. Google Trend Index of different keywords are then combined by using weight resulted from Principle Component Analysis (PCA).
## Keywords

<table>
<thead>
<tr>
<th>Bali</th>
<th>Maldives</th>
<th>Phuket</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tourist Bali</td>
<td>Tourist Maldives</td>
<td>Tourist Phuket</td>
</tr>
<tr>
<td>Hotel Bali</td>
<td>Hotel Maldives</td>
<td>Hotel Phuket</td>
</tr>
<tr>
<td>Holiday in Bali</td>
<td>Holiday in Maldives</td>
<td>Holiday in Phuket</td>
</tr>
<tr>
<td>Bali Tourism</td>
<td>Maldives Tourism</td>
<td>Phuket Tourism</td>
</tr>
<tr>
<td>Bali Tourist Place</td>
<td>Maldives Tourist Place</td>
<td>Phuket Tourist Place</td>
</tr>
<tr>
<td>Travel in Bali</td>
<td>Travel in Maldives</td>
<td>Travel in Phuket</td>
</tr>
<tr>
<td>Bali</td>
<td>Maldives</td>
<td>Phuket</td>
</tr>
</tbody>
</table>
The data collection process is carried out using the Python language with package `requests` and `urllib`.

1. Collecting Hotel, tourist places, restaurant coordinates in the province using the Geocoding API

The data was collected using the web scraping package `selenium` method which was applied to the Python and Java programming languages.

1. Hotel, tourism attractions, restaurant URL Collection
2. Detailed location Data Collection
<table>
<thead>
<tr>
<th>ATTRIBUTE</th>
<th>TYPE</th>
<th>EXPLANATION</th>
<th>SAMPLE DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>String</td>
<td>place name</td>
<td>Angkringan Rumah Cerita</td>
</tr>
<tr>
<td>Full_Address</td>
<td>String</td>
<td>place address</td>
<td>Jl. Tukad Citarum No.25, Renon, Kec. Denpasar Sel., Kota Denpasar, Bali 80234</td>
</tr>
<tr>
<td>Website</td>
<td>String</td>
<td>place website (may NULL)</td>
<td>business.site</td>
</tr>
<tr>
<td>Plus_Code</td>
<td>String</td>
<td>place plus code</td>
<td>869M+FR Renon, Denpasar City, Bali</td>
</tr>
<tr>
<td>Rating_all</td>
<td>float</td>
<td>rating in overall</td>
<td>4.3</td>
</tr>
<tr>
<td>Reviews_counts</td>
<td>integer</td>
<td>total number of reviews</td>
<td>67</td>
</tr>
<tr>
<td>Review</td>
<td>String</td>
<td>review text (may NULL)</td>
<td>(Diterjemahkan oleh Google) Tempat kecil (Asli) Small places</td>
</tr>
<tr>
<td>Rating</td>
<td>float</td>
<td>rating given each review</td>
<td>3</td>
</tr>
<tr>
<td>time</td>
<td>String</td>
<td>time the reviews was given</td>
<td>setahun yang lalu</td>
</tr>
</tbody>
</table>
## ACCOMMODATIONS LIST

### Bali

<table>
<thead>
<tr>
<th>Type</th>
<th>Accommodation Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entire villa</td>
<td>1998</td>
</tr>
<tr>
<td>Guesthouse/bed and breakfast</td>
<td>1767</td>
</tr>
<tr>
<td>Hotel</td>
<td>1467</td>
</tr>
<tr>
<td>Resort villa</td>
<td>1186</td>
</tr>
<tr>
<td>Resort</td>
<td>1050</td>
</tr>
<tr>
<td>Villa</td>
<td>502</td>
</tr>
<tr>
<td>Homestay</td>
<td>407</td>
</tr>
<tr>
<td>Entire apartment</td>
<td>274</td>
</tr>
<tr>
<td>Entire bungalow</td>
<td>238</td>
</tr>
<tr>
<td>Hostel</td>
<td>190</td>
</tr>
</tbody>
</table>

### Maldives

- Maalhosmadulu Atholu
- North Ari Atholu
- Male Atholu
- South Ari Atholu
- Addu Atholu

---

*Map showing the distribution of accommodations across different areas.*
NUMBER OF REVIEWERS BY COUNTRY

Reviewer Country (by Year)

Year

2018  2019  2020

Australia

Australia

Indonesia
RESULTS: GOOGLE MAPS

Number of Reviews of Tourist Places

The Number of Tourism Place Reviews in Bali Over Time

The Number of Tourism Place Reviews in Phuket Over Time
• Various Data Sources can be used as alternative for Tourism Statistics
• The variables, information, and data availability are different across the platforms and sources.
• Need data cleaning, aggregation and integration to combine them into a system of official statistics
• These alternatives data sources can be used as complement of the existing tourism statistics
Thank You

Setia Pramana
Assoc. Professor Politeknik Statistika STIS,
Coordinator for Statistical Model Development
BPS-Statistics Indonesia
setia.pramana@stis.ac.id