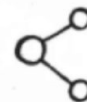# Unveiling the Links: TripAdvisor Ratings and Extrinsic Factors on Vacation Rentals in British Columbia, Canada

Group Three:
Anna Brabender, Audrey Nkrumah,
Fernando Lopez, and Saiyara Islam

Introduction : Our focus and purpose

Research has been conducted that show that there is indeed a significant relationship between internal factors such as number of amenities within a rental unit and the rating that rental unit gets from its renters. So we sought to find the following:

1. Does external factors/ locations define High-Rated Vacation Rentals?
2. The Relationship Between Ratings and Number of Reviews
3. Impact of External Factors on Ratings

Location of Interest :
British Columbia, Canada

External Locations of Interest:
- Tourist Sites
- Activity Centers
- Entertainment Spots
- Water Spots
- Spa Locations

Hypothesis

Relationship between the number of ratings and the average rating value

Hypothesis: If a vacation rental receives a high number of ratings, then the average rating of the rental is expected to be higher.

Null Hypothesis: There is no significant correlation between the number of ratings a vacation rental receives and the average rating value.

Correlation between the number of entertainment spots and a vacation rental rating

Hypothesis: If there are more entertainment spots around a vacation rental, then the vacation rental rating is expected to be higher.

Null Hypothesis: There is no significant correlation between the number of entertainment spots around a vacation rental and the rental's rating

Hypothesis

Is there a correlation between the number of spa spots and a vacation rental's rating?

Hypothesis: There is a positive correlation between the number of spa spots and a vacation rental's rating.

Null hypothesis: There is no correlation between the number of spa spots and a vacation rental's rating.

Is there a correlation between the number of activities (Clubs, community centers) and a vacation rental's rating?

Hypothesis: There is a positive correlation between clubs/community and vacation rental rating.

Null Hypothesis: There is no correlation/relationship between clubs/community centers and vacation rental ratings.

Is there a correlation between the number of tourism spots and a vacation rental's rating?

Hypothesis: There is a positive correlation between the number of tourism spots and a vacation rental's rating.

Null hypothesis: There is no correlation between the number of tourism spots and a vacation rental's rating.

Tools and Technologies

- The project was conducted using Python in a Jupyter Notebook environment.

- The primary libraries utilised included Pandas for data manipulation, Matplotlib and Seaborn for data visualisation, and requests for API interactions.

- The TripAdvisor and Geoapify APIs were used to gather relevant data.
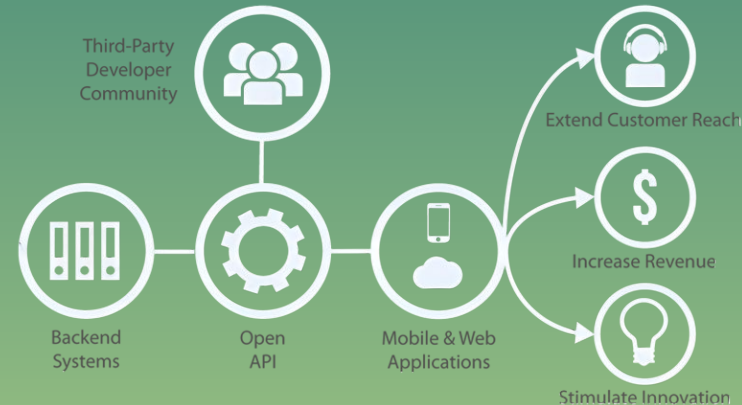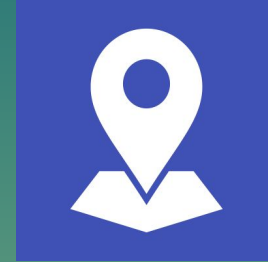
Data Collection

TripAdvisor API

- Obtained information about rental properties in BC
- Details included **property name, location, rating, Longitude** and **Latitude** coordinates.
- Allowed for the extraction of comprehensive details about each rental.

Geoapify API

- Gathered information about external locations near each rental.
- Categories of interest included **tourist sites, activity centres, entertainment spots, water spots,** and **spa locations**.
- Latitude and longitude coordinates were used to determine proximity.

Data Cleaning and Preprocessing

Handling Missing Data
- Missing or inconsistent data from the APIs were addressed through imputation or removal.
- Data types were standardised, and duplicate entries were handled.

Data Integration
- Data from TripAdvisor and Geoapify were integrated based on location coordinates.
- A consolidated dataset was created for further analysis.
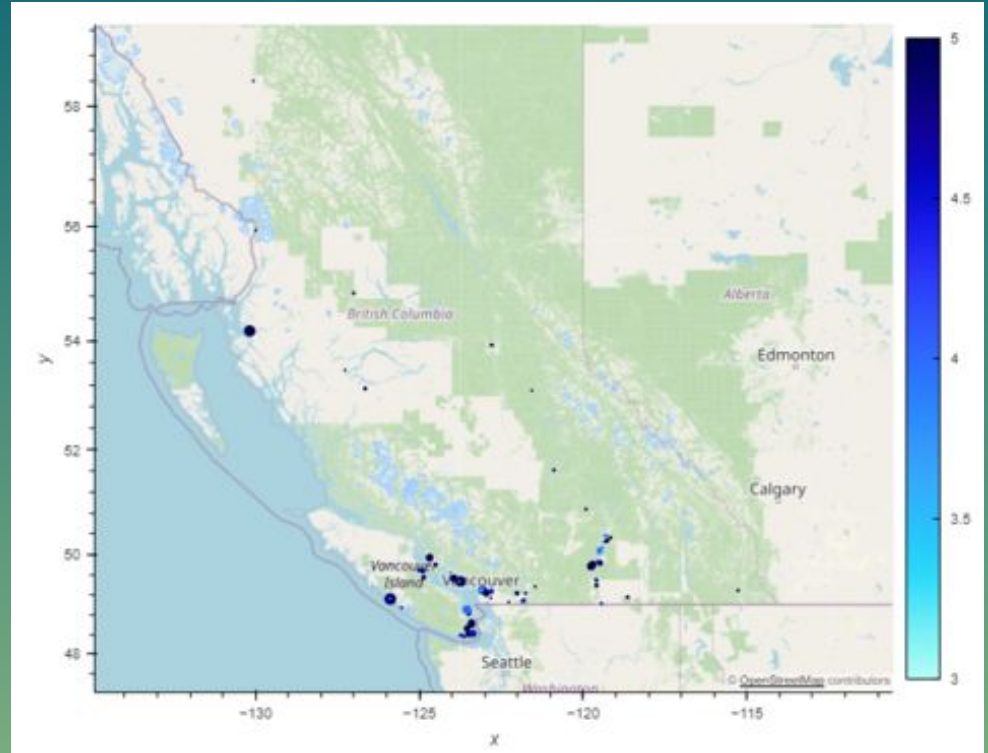
Map showing location of rentals, colour is based on rating, size is based on review count

Exploratory Data Analysis

Descriptive Statistics

- Initial data summaries and visualizations for rental data.

Correlation Analysis

- Scatter plots examined rental rating relationships with external locations.
- Weak correlation found, implying proximity to external locations doesn't strongly impact rental ratings.

Comparing the means of the top 20 highest rated rentals

Data examined: Looking at the mean of each of the external factors around the top 20 highest rated rentals
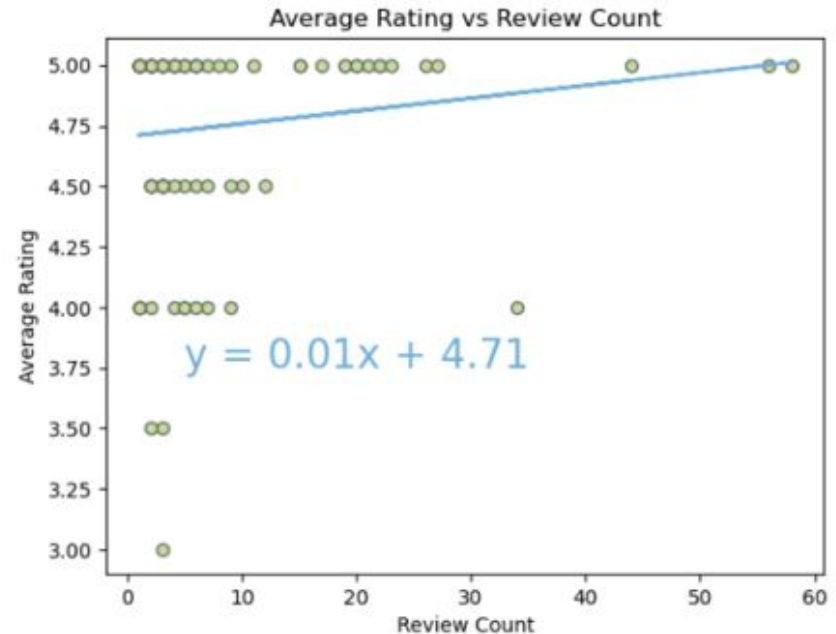

Mean of external factors for the 20 highest rated rentals
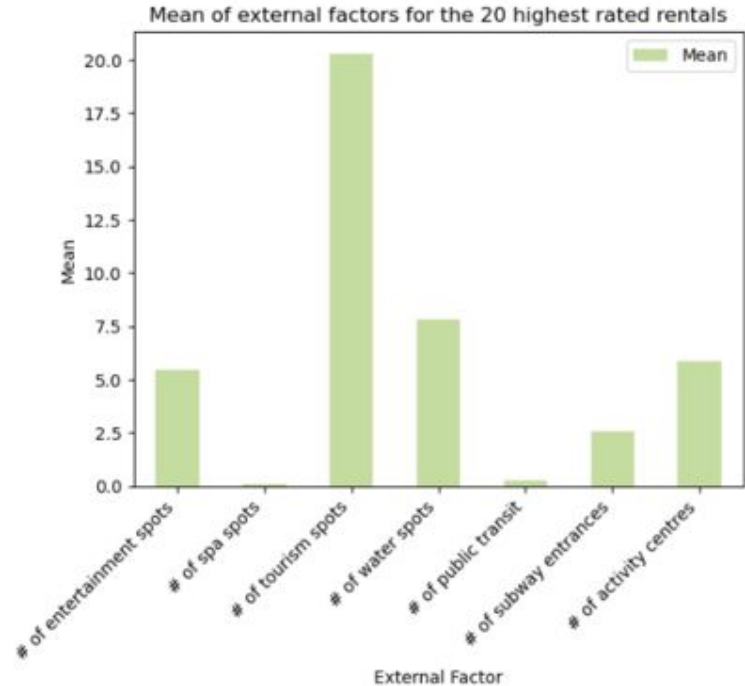
Average Rating Versus Review Count

R-value indicates there is no correlation between the average rating and average review count

P-value: 0.1932

the r value is: 0.0175714292483631128

Average Rating vs Review Count

$y = 0.01x + 4.71$

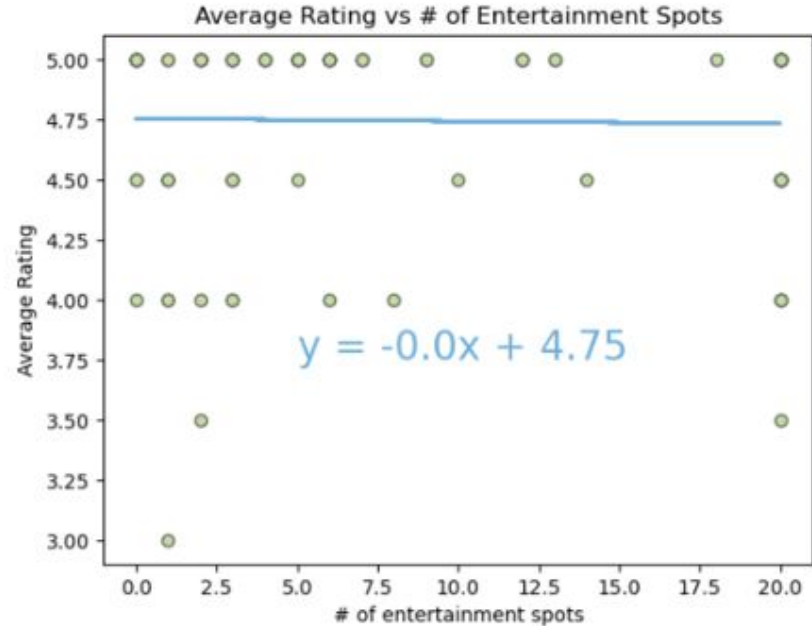# Number of Entertainment spots VS. Vacation Rentals Rating

R-value indicates there is no correlation between the average rating and # of entertainment spots

    Radius examined:10km
    Limit: 20 entertainment spots
    P-value: 0.8455

the r value is: 0.0003977590056564513

Average Rating vs # of Entertainment Spots

$y = -0.0x + 4.75$
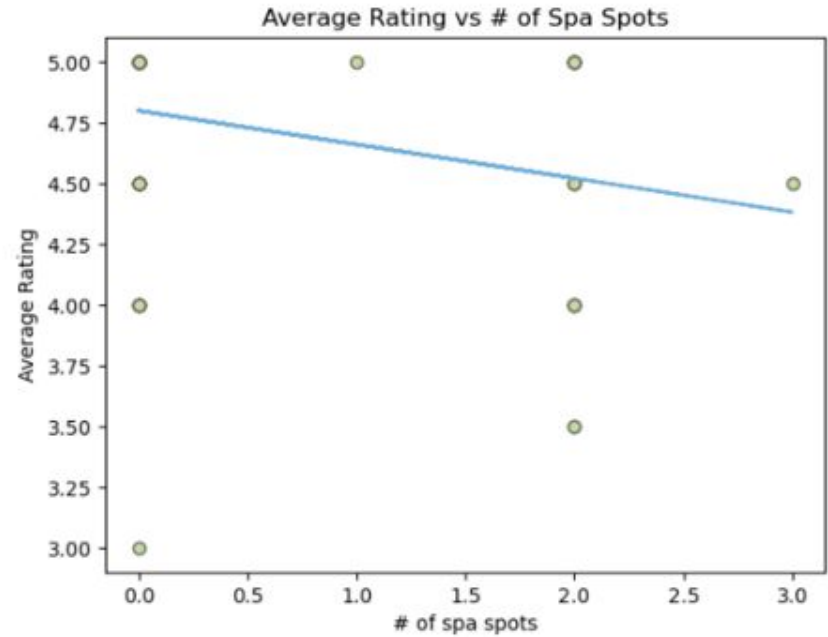
Number of spa spots VS. Vacation Rentals Rating

R-value indicates there is no correlation between the average rating and number of spa spots.

Radius examined:10km
Limit: 20 spa spots
P-value: 0.0087

the r value is: 0.06955647496124177

Average Rating vs # of Spa Spots
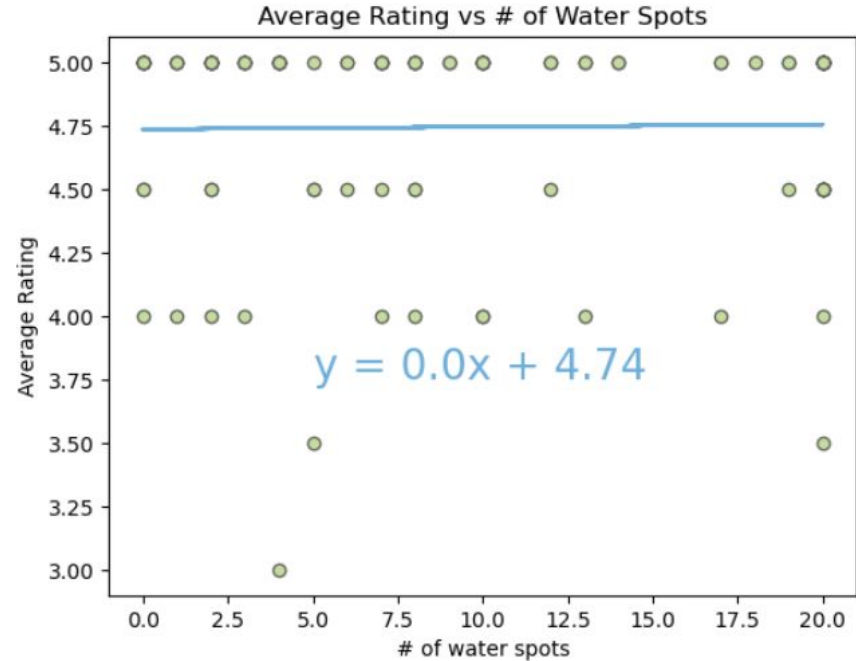
Number of water spots VS. Vacation Rentals Rating

R-value indicates there is no correlation between the average rating and number of water spots.

Radius examined: 10km
Limit: 20 water spots
P-value: 0.8765

the r value is: 0.00025284312259425744

Average Rating vs # of Water Spots

y = 0.0x + 4.74

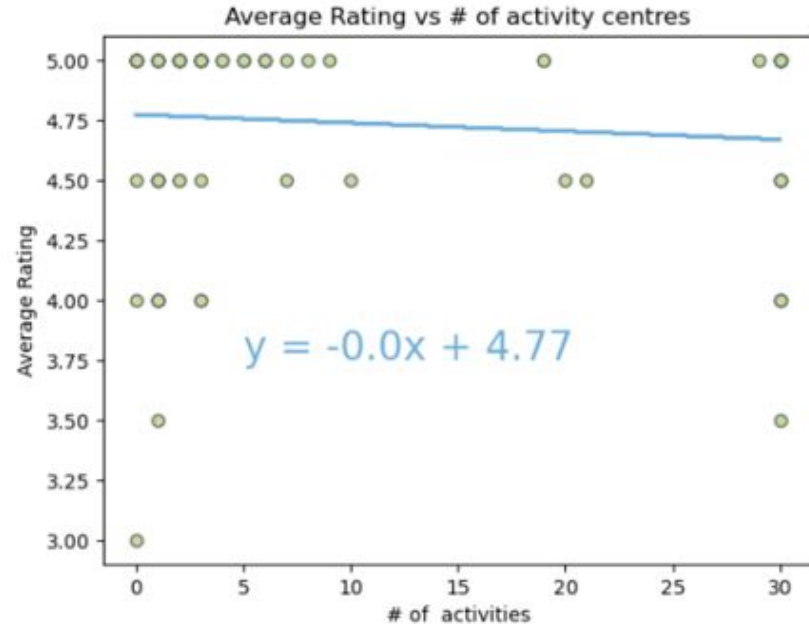# Number of Activities (Clubs, community centres) VS. Vacation Rentals Rating

R-value indicates there is no correlation between the average rating and the number of activity centres

Radius examined:10km
Limit: 40 activity centres
P-value: 0.3743

the r value is: 0.00823075585072367

Average Rating vs # of activity centres

$y = -0.0x + 4.77$

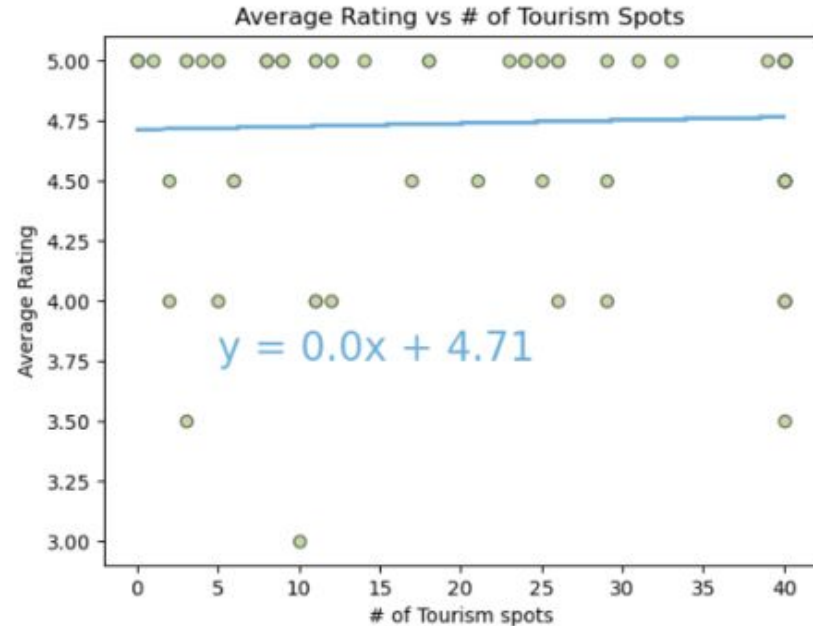# Number of Tourism spots VS. Vacation Rentals Rating

R-value indicates there is no correlation between the average rating and the number of activity centres

Radius examined: 10km
Limit: 40 tourism spots
P-value: 0.6502

the r value is: 0.002150466370403442

Average Rating vs # of Tourism Spots

y = 0.0x + 4.71

Findings

- **Data Selection:** Local transportation, entrenatiments, activities, natural and man-made landmarks are spots that benefit from reviews as tourists rely on them to judge experiential goods (Chu et al., 2022).

- **Descriptive Insights:** The dataset contained a diverse range of rental properties in British Columbia spread across urban, suburban and rural. Top 20 rentals had a high mean tourist attraction count. Tourism spots had the highest mean, followed by water spots and activity centres. Public transit and spa spots had extremely low mean values.

- **Correlation analysis:** Findings suggested no correlation between rental ratings and proximity to external locations. These findings are similarly to (Hernández et al., 2018) which found that geographic proximity between tourist attractions and vacation rentals do not necessarily overlap.

Challenges and Limitations

Data Quality & Quantity:

We limited the calculation of review count to 5 reviews as that was the limit in the free option for the API

We also limited the data collected on the rental properties to a total population of 97, regardless of ratings but as fairly distributed across the province

Causation vs. Correlation:

The analysis focused on correlation, and causation cannot be conclusively inferred from the findings.

Review Bias:

Only a select group of individuals review specific locations and the amounts reviews they recieve.

Relevance:

Is the tripadvisor database still a relevant source to represent tourist and consumer behaviour? It's possible external factors don't influence the rating as much as internal factors such as ambience and environment (Sangkaew & Zhu, 2020).

Future Directions

Possible alternate services as better indicators of tourism experience: local markets, glamping, sports events, and natural landmarks (Sangkaew & Zhu, 2020).

Incorporating sentiment analysis to explore what elements influence tourists' emotions. Looking for words related to crowdedness, and hygiene (Chu et al., 2022).

Examining shifts in rating based on the frequency of reviews across a period time (Tayeen et al., 2021).

Conclusions

- Successfully extracted data from TripAdvisor and Geoapify APIs.
- Conducted an in-depth analysis of rental ratings in British Columbia, considering their proximity to external locations.
- Identified weak correlations, suggesting we fail to reject our null hypothesis as a rule of thumb. Therefore, we concluded that:
- External factors does not define High-Rated Vacation Rentals
- There is no significant Relationships Between Ratings and Number of Reviews
- There are no significant impact of External Factors on Ratings
- Emphasizes the need for additional analysis to establish causation and to uncover any other influential factors affecting rental ratings.
- Lays the groundwork for future research endeavors.

References:

Boicheva, A. (2022). *The Best Free Presentation Icons For Your Project*. Graphic Mama Blog. GraphicMama.com. Retrieved October 7, 2023, from https://graphicmama.com/blog/free-presentation-icons/.

Çakırtaş, M., & Ozdemir, M. K. (Eds.). (2021). Big Data and Social Media Analytics. *Lecture Notes in Social Networks*. doi:10.1007/978-3-030-67044-3

Chu, M., Chen, Y., Yang, L., & Wang, J. (2022). Language interpretation in travel guidance platform: Text mining and sentiment analysis of TripAdvisor reviews. *Frontiers in psychology*, 13, 1029945. https://doi.org/10.3389/fpsyg.2022.1029945

Filieri, R., Acikgoz, F., Ndou, V., & Dwivedi, Y. (2020). Is TripAdvisor still relevant? the influence of review credibility, review usefulness, and ease of use on consumers' continuance intention. International Journal of Contemporary Hospitality Management. doi:10.1108/IJCHM-05-2020-0402

Hernández, J. M., Kirilenko, A. P., & Stepchenkova, S. (2018). Network approach to tourist segmentation via user generated content. *Annals of Tourism Research*, 73, 35–47. doi:10.1016/j.annals.2018.09.002

GIS Software. (n.d.). *Geoapify Pricing*. G2.com. Retrieved October 7, 2023, from https://www.g2.com/products/geoapify/pricing.

Sangkaew, N., & Zhu, H. (2020). Understanding Tourists' Experiences at Local Markets in Phuket: An Analysis of TripAdvisor Reviews. *Journal of Quality Assurance in Hospitality & Tourism*, 1–26. doi:10.1080/1528008x.2020.1848747

The New Tripadvisor Logo 2023 PNG. (2023). eDigital. Retrieved October 7, 2023, from https://www.edigitalagency.com.au/logos/new-tripadvisor-logo-png/.

Van der Zee, E., & Bertocchi, D. (2018). Finding patterns in urban tourist behaviour: a social network analysis approach based on TripAdvisor reviews. *Information Technology & Tourism*. doi:10.1007/s40558-018-0128-5