www.bircu-journal.com/index.php/birci
Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Humanities and Social Sciences
ISSN 2615-3076 Online)
ISSN 2615-1715 (Print)

Online Travel Agent Service and Customer Spending Behavior: The influence of Age Generation, User's location, PayLater Status during COVID 19 Pandemic towards Spending

Candra Harja Kurnia[1], Hendra Achmadi[2]
[1,2]*Faculty of Economics and Business, Universitas Pelita Harapan, Indonesia*
Candraharja@gmail.com, Hendra.achmadi@uph.edu

Abstract

This necessity of continually improving the service to online customers is an important prerequisite for a better and successful business in all industries, including the Online Travel Agent industry. Mostly during Covid 19 where the travel industry hit the hardest, purchase and booking has emerged as the main objective for OTA to survive. Therefore, understanding the different customer groups in one of the biggest OTA in Indonesia is indispensably crucial. The purpose of this research is to examine whether User's gender, location (cities), and PayLater status have positive correlation on User's spending, and to eventually help OTA in providing approaches to different customers to increase their sales and bookings especially during the pandemic year.

Keywords
online travel agent service; customer; customer spending behavior; PayLater; covid 19

I. Introduction

Online travel shopping has been a rising trend for the past decade and is expected to continue to soar even much more rapidly in the upcoming years. While it is true that during the pandemic people don't really have travel plans and might not travel anywhere, travel agents are still alive selling products and services beyond what most people usually imagine. Apart from selling plane tickets, travel agents often provide varieties of products. These product offerings include: accommodations (both international and domestic), attraction tickets, offline/online events, apartment rentals, COVID-19 tests, car rentals, car transfer services, and also travel insurances. Most Online Travel Agents in Indonesia including traveloka and tiket.com, who has been established year ago, and has been growing steadily since then.

The outbreak of this virus has an impact of a nation and Globally (Ningrum *et al*, 2020). The presence of Covid-19 as a pandemic certainly has an economic, social and psychological impact on society (Saleh and Mujahiddin, 2020). Covid 19 pandemic caused all efforts not to be as maximal as expected (Sihombing and Nasib, 2020).

Despite the growing market and the advancement of technology in the industry, there hasn't been enough research on the characteristics of online travel agent's customers and loyal users. COVID-19 pandemic took a toll on the sales of most companies' sales, including those in Online Travel Agents. It is crucial for OTA to strategize effectively to understand who their customers are, and how to best retain and attract existing customers to keep on purchasing from them. With a variety of product offerings, Online Travel Agent indisputably has a wide range of customers. Each customer is unique in their own ways, and every unique person has their own demand and preferences; this makes it even more difficult for companies to give out the best strategies, but at the same time, they should not generalize their whole customers into one big group.

DOI: *https://doi.org/10.33258/birci.v5i1.4391*

Budapest International Research and Critics Institute-Journal (BIRCI-Journal)
Volume 5, No 1, February 2022, Page: 6653-6663
e-ISSN: 2615-3076 (Online), p-ISSN: 2615-1715 (Print)
www.bircu-journal.com/index.php/birci
email: birci.journal@gmail.com

The powerful action of clustering all customers into several niche groups has been known to be effective in knowing who are the customers with the best purchase intentions based on their lifetime spending in a company, but less is known about who are the groups of customers, whether there are several variables that impact and influence a certain group of customers to spend more, specifically in the Online Travel Agents. Customers can be grouped in so many different ways, and it is specifically much more important for online platforms to step up on their game to cluster them in more niche behaviors. Once clusters have been found, then OTA will be able to target their clusters specifically with the right marketing strategies, to make sure that they keep on coming back to the platform instead of their competitors. However, before any clustering is done by the company, we need to know first what variables are important in knowing how to segment the customers, and what are the effects they might have on customers' spending habits in the Online Travel Agents.

With that being said, this research aims to analyze the impact of these different variables: user's age, location, and whether they enroll in PayLater (credit scheme feature) on users' spending. For every cluster that is generated, we are going to analyze the spending pattern based on each user's lifetime spending inside the platform. We are planning to answer these following questions: Does age have any impact on customer's spending? Does the user's location (cities) have any impact on customer's spending? Does the PayLater status have any impact on customer's spending? If so, all OTA should really push for Paylater on certain cities across ages.

II. Review of Literature

2.1 Age Generation

According to the Glossary of Statistical Terms, age is defined as the interval of time between the day, month and year of birth and the day and year of occurrence of the event expressed in the largest completed unit of solar time such as years for adults and children and months, weeks, days, hours or minutes of life, as appropriate, for infants under one year of age (Gregorian calendar). On top of that, in the world we are living in, age is classified into different groups as generations. Each generation has different personalities and therefore might have a different impact on customer purchase. Through descriptive analysis on this research, we find that Millennial's households indeed have different consumption patterns. As stated by, here are the different generations:

Table 1. Age Group Table

Generation Groups	Range	Age in 2022
Gen Z	1997 - 2012	10 - 25
Millennials	1981 - 1996	26 - 41
Gen X	1965 - 1980	42 - 57
Baby Boomers	1946 - 1964	56 - 76

2.2 Location (City)

Cities are viewed as globally interconnected entities and thus as global city networks, especially with regard to the discussion of global cities and world cities. This is accompanied by the increasing urbanization process, which is also reflected in the network society. Due to their position as outstanding economic, political and cultural anchor centers, they have a special significance in the hierarchical system of cities. Urban areas are classified as: large metropolitan/megapolitan areas if they have a population of 1.5 million or more; metropolitan areas if their population is between 500 000 and 1.5 million; medium-size urban areas if their population is between 200 000 and 500 000; and, small urban areas if their population is between 50 000 and 200 000. Individuals in different cities might have different behavior towards

For the Cities variable, we are going to match the definition of different city groups, find out what cities are included under each group, and see which cities are available in the database. We will then use two cities per group as a sample that represent each group.

Table 2. City Classification

Group	**Size of the Population**
Large Metropolitan / Megapolitan	More than 1.5 Mio
Metropolitan	Between 500,000 to 1.5 Mio
Medium-sized Urban Area	Between 200,000 to 500,000

It seems that individual spending activity exhibits a statistically significant superlinear scaling with city size.

2.3 PayLater

The development of e-commerce has had an impact on payment methods which were originally only known as cash payments, paper-based, card based, and are now being introduced with new electronic-based payment systems such as m-banking, virtual accounts, and payment applications. Paylater is a type of online credit payment. Paylater allows its users to have a credit card without a card and without a complicated manufacturing process. The results of this study indicate that the ease of use of technology pay later has an effect of 6.4% on the impulsive buying behavior of users of ecommerce in Indonesia. From the results of data processing, it can be concluded that the ease of use of Paylater technology by users of e-commerce in Indonesia is very good and users who use Paylater tend to make impulse buying when shopping.

Table 3. Paylater Definition

PayLater Status	**Definition**
Approved	Credit is approved by the OTA
Rejected	Credit is rejected by the OTA
Inactive	User not yet register for credit limit

2.4 Data Mining

Data mining is a process of discovering knowledge from a data warehouse. This knowledge can be classified in different rules and patterns that can help users/organizations to analyze collective data and predict decision processes. Centralized database of any organization is known as Data warehouse, where all data is stored in a single huge database. Data mining is a method that is used by organizations to get useful information from raw data. Software's are implemented to look for needed patterns in huge amounts of data (data warehouse) that can help businesses to learn about their customers, predict behavior and improve marketing strategies.

2.5 Data Mining Processing Steps

According to (Achmadi, 2020), said that the stages of the data mining process are starting from data selection from data sources to target data which are often referred to as datasets that are used as the basis for data processing, then the process is continued with data processing or data cleansing, here the data preparation begins for further processing, for example whether the data has number type or factor or date, and then the data in the data cleansing is also done by removing special characters, then after that the transformation is carried out, namely transforming the data from the cleansing data into the target data, the process then is to do data mining or data model based on a method that is suitable for the data, and the last is the process of interpreting the knowledge obtained from processing the data. And the data mining process stages

2.6 Hypotheses

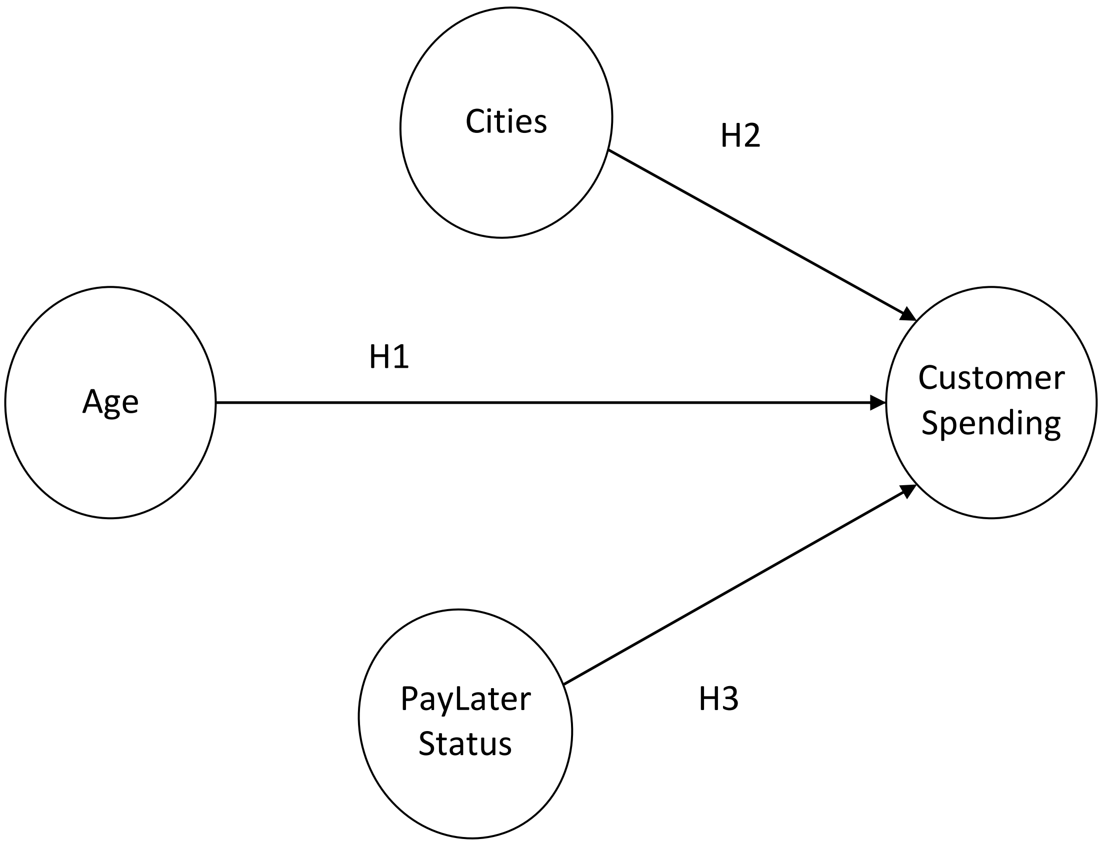

Figure 1. *Research Model*

The three relationships we are about to test are:

a. Relationship between **Age Generation** and **Customer Spending**
   H0: Age generation has no influence on customer spending in an Online Travel Agent
   Ha: Age generation has an influence on customer spending in an Online Travel Agent
b. Relationship between **Cities** and **Customer Spending**
   H0: User's locations in city has no influence on customer spending in an Online Travel Agent
   Ha: User's locations in city has an influence on customer spending in an Online Travel Agent
c. Relationship between **PayLater Status** and **Customer Spending**
   H0: User's PayLater Status has no influence on customer spending in an Online Travel Agent
   Ha: User's PayLater Status has an influence on customer spending in an Online Travel Agent

III. Research Method

3.1 Data Preparation

R Software is used to test out the hypotheses stated on the earlier section. Statistical methodology testing will be conducted on this research. Data points were collected from transacting users in a leading Online Travel Agents in Indonesia. Based on the literature review and research framework shown earlier, the independent variables for this research are Age Generation, Cities, and PayLater Status.These variables will be tested and measured to see whether they each have any influence on customer spending in an Online Travel Agent. Data points from the database were collected from the year 2021, the pandemic year. Database includes all transacting data per user ID, which is a unique identification number for each OTA's customer. 102,786 user data was derived, but not all of the users have complete information on our independent variables. So we had to cross match the data to another database that gives out the independent variable data column, and take out only the data points with complete information. There were 88,684 data points after the cross matching method, and we had to further cleanse the data to impute all the null and invalid data points. After the cleansing, we got 56,770 data points that we can use for the statistical analysis and testing. All the final data points have each of the user's information along with each of their age, locations in cities, and PayLater status.

Once data was ready, data points were divided into each independent variable for random sampling and testing. Once it was divided, we did a random sampling in each independent variable group (age, locations in cities, PayLater status), that gave us each 300 data points per group. Another cleansing that we did was removing the outliers from the data points per group. The method that we used to remove the outliers was the Interquartile range method. From this [https://online.stat.psu.edu/stat200/lesson/3/3.2], We can use the IQR method of identifying outliers to set up a "fence" outside of Q1 and Q3. Any values that fall outside of this fence are considered outliers. To build this fence we take 1.5 times the IQR and then subtract this value from Q1 and add this value to Q3. This gives us the minimum and maximum fence posts that we compare each observation to. Any observations that are more than 1.5 IQR below Q1 or more than 1.5 IQR above Q3 are considered outliers.

3.2 Hypothesis Testing

To test out the hypothesis, there were several steps and methods that we did. The first test that we did was the Shapiro-Wilk Method. The Shapiro-Wilk method is dedicated to checking normality with known mean value μ0, i.e. to testing the hypothesis H0: X ∼ N μ0, σ2, where X is the random variable of interest. In this case, our X is the IDR (Rp) amount of promotion usage of each user. The results can be shown as follow:

```
        Shapiro-Wilk normality test

data:  FinalDataMegapolitan$sum_gbv
W = 0.87923, p-value = 5.847e-14


        Shapiro-Wilk normality test

data:  FinalDataMetropolitan$sum_gbv
W = 0.8546, p-value = 1.321e-15


        Shapiro-Wilk normality test

data:  FinalDataMediumSized$sum_gbv
W = 0.84867, p-value = 1.578e-15


        Shapiro-Wilk normality test

data:  FinalDataGenZ$sum_gbv
W = 0.87824, p-value = 8.1e-14


        Shapiro-Wilk normality test

data:  FinalDataMillenials$sum_gbv
W = 0.87911, p-value = 1.58e-13
```

```
        Shapiro-Wilk normality test

data:  FinalDataGenX$sum_gbv
W = 0.85867, p-value = 6.005e-15


        Shapiro-Wilk normality test

data:  FinalDataBabyboomers$sum_gbv
W = 0.84026, p-value = 1.853e-10


        Shapiro-Wilk normality test

data:  FinalDataActivePaylater$sum_gbv
W = 0.88628, p-value = 3.228e-13


        Shapiro-Wilk normality test

data:  FinalDataInactivePaylater$sum_gbv
W = 0.86131, p-value = 6.394e-15


        Shapiro-Wilk normality test

data:  FinalDataRejectedPaylater$sum_gbv
W = 0.81987, p-value < 2.2e-16
```

Figure 2. *Shapiro Wilk Normality Test Result*

From the above image we can see that all of the data samples were not normally distributed (P value <0.05). Due to that, we took the next step to test it out again by using the Kruskal-Wallis test. The Kruskal-Wallis one-way analysis-of-variance-by-ranks test (or H test) is used to determine whether three or more independent groups are the same or different on some variable of interest when an ordinal level of data or an interval or ratio level of data is available. The results show the same result, all variables are shown to be statistically significant, with p-value of less than 0.05 for all of them.

Another test that we did was the Conover-Iman test, conducted to test for median difference. This test was conducted only on significant variables (in this case all three of them), to test the median difference between each sub-variable and see which sub-variable influences more on our dependent variable (customer spending).

```
data: x and group
Kruskal-Wallis chi-squared = 70.5585, df = 3, p-value = 0


                           Comparison of x by group
                                (No adjustment)
Col Mean-|
Row Mean |   BabyBoom       GenX       GenZ
---------+---------------------------------
    GenX |  -1.610161
         |     0.0539
         |
    GenZ |   5.174070   8.441763
         |    0.0000*    0.0000*
         |
Millenia |   2.135099   4.638597  -3.742764
         |    0.0165*    0.0000*    0.0001*

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure 3. *Kruskal-Wallis & Conover Iman Test Result (Spending Across Age Generation)*

```
data: x and group
Kruskal-Wallis chi-squared = 17.0885, df = 2, p-value = 0


                           Comparison of x by group
                                (No adjustment)
Col Mean-|
Row Mean |   Medium-S   Megapoli
---------+----------------------
Megapoli |  -3.207745
         |    0.0007*
         |
Metropol |  -3.929340  -0.709235
         |    0.0000*     0.2392

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure 4. *Kruskal-Wallis & Conover Iman Test Result (Spending Across City Type)*

```
Kruskal-Wallis chi-squared = 19.2689, df = 2, p-value = 0


                           Comparison of x by group
                                (No adjustment)
Col Mean-|
Row Mean |   approved   inactive
---------+----------------------
inactive |   1.612843
         |     0.0536
         |
rejected |   4.380774   2.779188
         |    0.0000*    0.0028*

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure 5. *Kruskal-Wallis & Conover Iman Test Result (Spending Across Paylater Status)*

IV. Results and Discussion

Table 4. Hypothesis Testing Result (Kruskal – Wallis)

Number	Figure	P-Value	Conclusion
1	Ha: Age generation has an influence on customer spending in an Online Travel Agent	0 (P-value < 0.05)	1. Reject Null Hypothesis 2. Age Generation has an influence on customer spending in an Online Travel Agent
2	Ha: User's locations in city has an influence on customer spending in an Online Travel Agent	0 (P-Value < 0.05)	1. Reject Null Hypothesis 2. User's locations in city has an influence on customer spending in an Online Travel Agent
3	Ha: User's PayLater Status has an influence on customer spending in an Online Travel Agent	0 (P-Value < 0.05)	1. Reject Null Hypothesis 2. User's PayLater Status has an influence on customer spending in an Online Travel Agent

The results show that all three variables (age generation, user's locations in city, and Paylater status) do influence customer's spending in an Online Travel Agent. From the Conover-Iman test, we get this table:

Table 5. Variables Testing Result (Conover - Iman)

Number	Figure	Conclusion
1	Figure 3	1. Baby Boomer & Generation X has no difference in spending 2. Millennial's spending is different with all the age group significantly 3. Generation Z's spending is different with all the age group significantly
2	Figure 4	1. Customer living in Megapolitan and Metropolitan area have no difference in customer spending 2. Customer living in Medium Sized City customer's spending is different significantly
3	Figure 5	1. Customer with Active and Inactive PayLater status have no difference in customer spending 2. Customer with Rejected PayLater status spending is different significantly

Table 6. Segment and Spending Median & Mean

	segment	gbv_median	gbv_mean
1	GenZ	1453400	1924375
2	Millenials	2040866	2630651
3	GenX	2945525	4372505
4	Babyboomers	2677125	3565338
5	Megapolitan	2423788	3323186
6	Metropolitan	2385652	3454751
7	Medium Sized	1621600	2451309
8	ActivePaylater	2460204	3185104
9	InactivePaylater	2082700	2781448
10	RejectedPaylater	1565200	2378363

Firstly, with age generation, the table above and the figures conclude that within Age Generations, Generation X & Baby Boomers spend the most. There is no significant difference in spending level between Baby Boomers and Generation X. The test result indicates that Millennials spend less than both Generation X and Baby boomers significantly. Lastly Generation Z spend the least and proven significant compared with the others.

Secondly, within the city levels, we see that there is no significant spending level between Metropolitan and Megapolitan area, however Medium Sized city is considerably lower compared to both Metropolitan and Megapolitan. Both mean and median data for Medium Sized city continue to support that (Row 5-7).

Lastly, within PayLater status, we can see those users with Rejected PayLater status tend to spend less, proven significant by the test. There is no significant difference between Active and Inactive, indicating some users can be a potential active PayLater users (Row 8-10).

V. Conclusion

It is quite interesting to see that the three variables (Age generations, User's location/ the type of city they live in, and PayLater status) all have an influence on User's Spending pattern. And as we can see from the results and discussions above, we see that older people (GenX and Baby Boomers), people from larger cities (Large Metropolitan and Metropolitan), and those with PayLater status (which means they can do credit schemes) tend to spend more on an Online Travel Agent. With the information from this research, Online Travel Agent can target these groups of customers more when applying any marketing strategy.

This is only natural considering the older generation might have already had the purchasing power to purchase more. Moreover people living in the Metropolitan and Megapolitan area might spend more due to larger income and lifestyle. Paylater (credit scheme) is proven to boost consumptive behavior, in which it can be utilized even more for the OTA.

However, these tests can't define the other types of business in other industries, since the data was obtained from an Online Travel Agent. Another limitation is that the data only has some sub-variables to be tested. For example, as seen in the Literature Review section, there are 4 types of urban cities, but in our data, we can only test 3 out of 4, since there was no data point found on this Online Travel Agent with the 1 sub-variable.

Despite the limitation, this research means that consumer spending is a very broad aspect that is influenced by tons of variables. Further research is needed to validate some theories and hypotheses that may arise.

References

Age range by generation. Beresford Research. (2022, January 4). Retrieved February 9, 2022, from https://www.beresfordresearch.com/age-range-by-generation/

Chan, Y., & Walmsley, R. P. (1997). Learning and understanding the kruskal-wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups. *Physical Therapy*, *77*(12), 1755–1761. https://doi.org/10.1093/ptj/77.12.1755

Comprehensive R Archive Network (CRAN). (n.d.). *Package conover.test*. CRAN. Retrieved February 9, 2022, from https://cran.r-project.org/web/packages/conover.test/index.html

Directorate, O. E. C. D. S. (n.d.). The OECD Glossary of Statistical terms. Retrieved February 9, 2022, from https://stats.oecd.org/glossary

Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan RAPIDMINER Dengan K-means Cluster Pada Daerah Terjangkit DEMAM Berdarah dengue (DBD) Berdasarkan provinsi. *Computer Engineering, Science and System Journal*, *3*(2), 173. https://doi.org/10.24114/cess.v3i2.9661

Mughal, M. J. (2018). Data Mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*, *9*(6). https://doi.org/10.14569/ijacsa.2018.090630

Ningrum, P. A., et al. (2020). The Potential of Poverty in the City of Palangka Raya: Study SMIs Affected Pandemic Covid 19. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)* Volume 3, No 3, Page: 1626-1634

Population by region - urban population by city size - OECD data. theOECD. (n.d.). Retrieved February 9, 2022, from https://data.oecd.org/popregion/urban-population-by-city-size.htm

Saleh, A., Mujahiddin. (2020). Challenges and Opportunities for Community Empowerment Practices in Indonesia during the Covid-19 Pandemic through Strengthening the Role of Higher Education. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*. Volume 3, No 2, Page: 1105-1113.

Sari, R. (2021). Pengaruh Penggunaan Paylater TERHADAP perilaku impulse buying pengguna e-commerce di Indonesia. *Jurnal Riset Bisnis Dan Investasi*, *7*(1), 44–57. https://doi.org/10.35313/jrbi.v7i1.2058

Shapiro–wilk test with known mean - researchgate. (n.d.). Retrieved February 9, 2022, from https://www.researchgate.net/publication/298706800_Shapiro-Wilk_test_with_known_mean

Sihombing, E. H., Nasib. (2020). The Decision of Choosing Course in the Era of Covid 19 through the Telemarketing Program, Personal Selling and College Image. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)* Volume 3, No. 4, Page: 2843-2850.

Sobolevsky, S., Sitko, I., Tachet des Combes, R., Hawelka, B., Murillo Arias, J., & Ratti, C. (2016). Cities through the prism of people's spending behavior. *PLOS ONE*, *11*(2). https://doi.org/10.1371/journal.pone.0146291

The Millennial's consumption: How are they any different ... (n.d.). Retrieved February 9, 2022, from https://www.researchgate.net/publication/328928441_The_Millennial's_Consumption_How_Are_They_Any_Different_From_Previous_Generation

Wagner, M., & Growe, A. (2021). Research on small and medium-sized towns: Framing a new field of inquiry. *World*, *2*(1), 105–126. https://doi.org/10.3390/world2010008

Xiang, Z., & Gretzel, U. (2010). Role of social media in Online Travel Information Search. *Tourism Management*, *31*(2), 179–188. https://doi.org/10.1016/j.tourman.2009.02.016