DOI: https://doi.org/10.34123/jurnalasks.v17i1.797

Jurnal Aplikasi Statistika & Komputasi Statistik
Vol. 17 No. 1, June 2025
e- ISSN: 2615-1367, p-ISSN:2086-4132

Available online at https://jurnal.stis.ac.id/

The Implementation of Geospatial Analysis on Hotel Occupancy Rate

Muhammad Fachry Nazuli[1], Satria Bagus Panuntun[2], Addin Maulana[3], Takdir[4], Setia Pramana[5]

[1]BPS-Statistics Lhokseumawe City, Lhokseumawe, Indonesia, [2]Directorate of Statistical Analysis and Development, BPS-Statistics Indonesia, Jakarta, Indonesia, [3]Center for Industrial, Services and Trade Research, National Research and Innovation Agency, Jakarta, Indonesia, [4]University of Tsukuba, Japan, [5]Politeknik Statistika STIS, Jakarta, Indonesia

**Corresponding Author: E-mail address: [1]fachry.nazuli@bps.go.id, [2]satria.bagus@bps.go.id, [3]addi001@brin.go.id, [4]takdir-kde@kde.cs.tsukuba.ac.jp, [5]setia.pramana@stis.ac.id*

ARTICLE INFO

Article history:
Received 2 September, 2024
Revised 5 December, 2024
Accepted 10 December, 2024
Published 24 February, 2025

Keywords:
Big Data; Google Maps; Locations; Spatial Regression; Web-Scraping.

Abstract

Introduction/Main Objectives: One of the main attributes of hotel selection and customer satisfaction is its location. **Background Problems:** Strategic location leads to higher demand for accommodation. Accommodation demand is reflected in hotel occupancy levels, which indicate the percentage of reserved rooms at a specific period. **Novelty:** This study aims to investigate the effect of spatial location on hotel occupancy rates by analyzing data collected in online hotel reservation applications. A study related to the effects of location and hotel occupancy has never been conducted in Indonesia. **Research Methods:** We use data from hotels located in the province of Yogyakarta, which contains 245 hotels spread over three regencies/cities, namely Yogyakarta City, Sleman Regency, and Bantul Regency. We conducted a spatial regression analysis, namely the Spatial Error Model (SEM), with a spatial weight matrix using a radius of 3.2 km. **Finding/Results:** We found that spatial locations affect the occupancy rates of hotels based on the online hotel reservation application that we observed. These spatial locations include the distance from the hotel to the airport, the distance from the hotel to the bus stop, and the number of nearby restaurants, offices, and hotels.

1. Introduction

The geographic position is an important aspect of the economic performance of a hotel. Location significantly affects the probability of a Hotel's survival [1]. As much as 20% of hotel bankruptcy rates are caused by location [2]. Hotels make location their marketing base. For example, many luxury hotels utilize their coastal area as a marketing attraction [3].

Tourists, whether traveling for business or leisure, regard the hotel's location as a critical factor influencing hotel selection [4]. The geographical location of a hotel can be strongly related to higher room occupancy rates, revenue per available room, and profit from the hotel [5-6]. The hotel must carefully choose its location to avoid the difficulties of relocating.

Customer satisfaction is also closely related to the location of the hotel. The ideal location leads to greater demand for accommodation, better company performance, and higher customer satisfaction [7]. A convenient location is considered one of the main factors influencing hotel selection and the satisfaction of business and leisure tourists. They prefer a location close to the services and facilities available[8]. Satisfaction will be derived not solely from the internal amenities of a hotel but also from the surrounding facilities. Location and proximity to city facilities, such as the central business district, transportation network, sea, lakeside, or tourist attractions, determine the market value of a hotel [9]. Customers usually prefer hotels closer to these facilities, which can increase the hotel's market value. Proximity to the city center can drive increased guest satisfaction and hotel demand [7].

In general, three factors influence hotel selection, which also affects hotel occupancy rates, namely accessibility to the point of interest (POI), transportation convenience, and the surrounding environment [10]. Accessibility to points of interest (POI) includes transportation portals (airports, stations, bus stops, etc.), central business districts, tourist attractions, entertainment venues, and so on. The convenience of transportation is related to the convenience of hotel guests to depart and return to the hotel using public transportation. Environmental factors are closely related to air quality, public safety and security, public infrastructure (such as restaurants, parks, etc.), and culture.

Yogyakarta is a province that has a very strong tourist attraction. As a province with fairly high tourist attractions, it had a total of 108,599 foreign tourists in 2019 and 195,778 in 2020 [11]. This significant decrease is due to the COVID-19 pandemic, which has caused people to reduce their level of mobility. The research conducted by Pramana [12] found an impact of the pandemic on the mobility index, number of flights, occupancy rates, and other big data [12]. The COVID-19 pandemic has significantly affected the accommodation, retail, transportation, and manufacturing sectors, contributing to the current account deficit (CAD) [13]. According to the KBLI (Indonesian Standard Industrial Classification) 2020, the hotel industry categorized as *category I* a sector that provides accommodation, food, and drink [14].

With today's technological advances, hotel reservations are available online. Several applications provide services for online hotel reservations. Online hotel reservations are considered more efficient than making them in person. Customers can make reservations before their departure day. The application is connected to the hotel upon a reservation. Thus, the room occupancy rate can be directly calculated based on the number of rooms available and the number of rooms that have been reserved on a certain date. The data velocity and variation of hotel reservation applications are high, which can be qualified as big data.

Previous studies have known that location influences or impacts a hotel's economic aspects. There are several critical aspects of the hotel economy, including the price per room [9], monthly income [15], profitability [16], hotel distributions [17-18], customer satisfaction [10], and others. However, no study related to the impact of location on occupancy rates has been conducted to the best of our knowledge.

This study aims to analyze the impact of location on the hotel/room occupancy rate using a big data approach. We use data from an online hotel reservation application as the main data source. We obtain locations for urban objects, such as restaurants, train stations, and airports from Google Maps.

2. Material and Methods

Before we look at how location influences hotel occupancy rates in the city of Yogyakarta, we can look at previous research related to hotels and location. Srimulyani [19] analyzed the hotel occupancy rate in West Nusa Tenggara Province. The purpose was to see the impact of COVID-19 on hotel occupancy rates in West Nusa Tenggara using a Big Data approach. Data was collected using the web-scraping method from online hotel reservation websites. They conducted a descriptive analysis of the collected data. They found that the early emergence of COVID-19 reduced the occupancy rate sharply, especially in high-star hotels, and the New Normal policy gradually increased the hotel occupancy rate.

Valentin and O'Neill [9] conducted a study that aims to investigate the significance of the hotel's location to the property market value of the hotel. The market value of a property is represented by the price per room. This research was conducted in Chicago using more than 600 hotels as the unit of observation. They use Ordinary Least Square Regression. The results show that proximity to the city center, namely the Loop, is the most influential factor in market value. One additional mile of the Loop will decrease the property value by 13% on average, for distances under 10 miles from the Loop and be relatively constant at distances above 10 miles from the Loop.

Yang [20] conducted a study that aims to look at the factors that contribute to the potential location of a hotel with a model that combines the characteristics of the hotel and its location. The potential hotel locations are categorized from 1 to 5 based on the ring road in Beijing City. The method used is the Ordered Logit Model. The results of this study indicate that the star level of the hotel, years of construction, service diversification, ownership, agglomeration effects, public service infrastructure, and accessibility (roads, subways, tourist sites) are important determinants in the selection of hotel locations.

Yang [10] conducted a study to determine the determinants of hotel guest satisfaction related to location based on reviews on the TripAdvisor website. The Ordinary Least Square Regression and the Ordered Logit Model were applied. They conclude that the accessibility of hotel properties to the location of tourist attractions, airports, universities, public transportation, green space, water coverage, and surrounding local businesses is vital to hotel reviews. Meanwhile, the variables of proximity to toll roads and crime rates have no significant effect on the model.

Fang [17] researched hotel location choices by developing a GWPR model in the Hong Kong region. The dependent variable used was the number of hotels located in certain sub-districts. They identified nine factors affecting the hotel's location. These include land area, green land, traffic land, commercial land, institutional land, station density, density of tourist attractions, population density, and the average income of residents in the area. The effect of each variable is different for each region. For example, transportation accessibility variables are not significant near the city center.

2.1. Data and variable

This study uses data from online hotel reservation applications. The observed hotels are located in Yogyakarta province. We analyze data from June 2019.

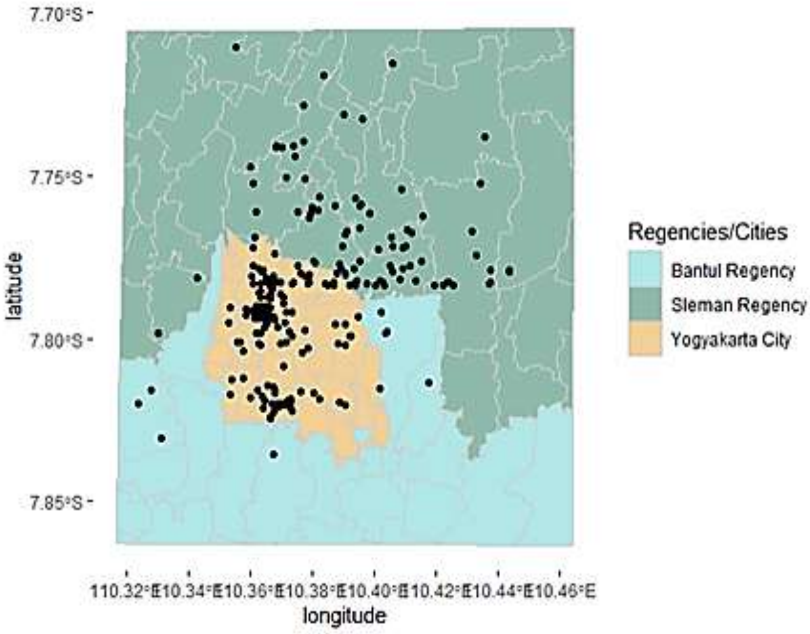

Figure 1. The distribution of hotels in the Special Region of Yogyakarta

In Figure 1, 245 hotels are detected from the application data spread over three regencies/cities in the Province of Yogyakarta, namely Yogyakarta City, Sleman Regency, and Bantul Regency. The dataset contains variables of the number of available, reserved, and occupied rooms. Thus, we can calculate the occupancy rate by using the following formula [15]:

$$occupancy_t = \frac{\Sigma(Total\ rooms - Available\ Rooms)}{\Sigma(Total\ rooms)} x100 \qquad (1)$$

The occupancy rate is the dependent variable in this study There are several variables that are considered to affect the hotel occupancy rate, including:

- hotel count: Number of hotels within a 1 km radius using Euclidean distance. It explains market competition or the agglomeration effect between hotels [17] due to the tendency of hotels to gather in one place.

- airport distance: This variable describes accessibility from the hotel to the airport or vice versa. The market value of a hotel will tend to increase along with its proximity to the airport, which impacts its occupancy rate [9].
- station distance: This variable describes accessibility from the hotel to the train station or vice versa. Access to train stations in urban areas determined hotel customer satisfaction [7].
- bus stop distance: This variable explains accessibility from the hotel to the nearest bus stop or subway [20].
- mall distance: This variable describes accessibility from the hotel to the nearest shopping center [15].
- university distance: This variable explains accessibility from the hotel to the nearest university [10].
- restaurant count: The sum of restaurants within a 1-kilometer radius using the Euclidean distance, excluding the in-house restaurant [10].
- attractions count: The number of attractions within a 1-kilometer radius of the hotel using Euclidean distance. Seven tourist attraction types include museums, theaters, amusement parks, state gardens, stadiums and arenas, beaches, historical sights, and shopping [10].
- office count: Number of offices located within 1 kilometer of the hotel using Euclidean distance. The number of offices has an impact on the type of business guests. The hotel will choose a location close to its potential market, such as shopping and business centers [17].

Of all the variables mentioned above, there are 2 data sources, including the online hotel reservation application to estimate room occupancy rates and Google Maps.

2.2. Analysis Method

a. Spatial Autocorrelation Analysis

Moran's test was applied to see the presence of spatial autocorrelation in hotel data [21]. The initial hypothesis for Moran's test shows the random distribution of the variables in certain areas. The alternative hypothesis is that the variable spreads with no random distribution and has a spatial autocorrelation. Morans' statistical value will be in the range of -1 and +1.

b. Spatial Analysis

Spatial analysis is applied if the Morans test rejects the initial hypothesis or the variables have spatial autocorrelation. The spatial analysis used is the Spatial Autoregressive (SAR) and Spatial Error Model (SEM) analysis. The results were then compared with the classical regression, namely the ordinary least square.

Before doing spatial modeling, it is necessary to test the Lagrange Multiplier. This test focuses on identifying the presence of spatial lag and spatial error in the model. If spatial lag is identified in the model, the Spatial Lag Model or Spatial Autoregressive Model (SAR) is used. Meanwhile, if the identified spatial error is in the model, the Spatial Error Model (SEM) is used. The following is a SAR model that explains the spatial lag (Wy) relationship [22]:

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2}$$

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2 I_n)$$

Where:

$\boldsymbol{y}$	: response variable vector
$\boldsymbol{x}$	: predictor variable matrix (n x p)
$\boldsymbol{\rho}$	: coefficient of spatial effect parameter of the predictor variable
$\boldsymbol{W}$	: spatial weighted matrix
$\boldsymbol{\beta}$	: coefficients of predictor variable parameters vector
$\boldsymbol{\epsilon}$	: error

Meanwhile, the spatial error model can be explained in the following equation [22]:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{3}$$

$$\boldsymbol{\epsilon} = \lambda \boldsymbol{W\epsilon} + \boldsymbol{u} \tag{4}$$

$$\boldsymbol{u} \sim N(0, \sigma^2 I_n)$$

Where:

$\boldsymbol{y}$: response variable vector
$\boldsymbol{x}$: predictor variable matrix (n x p)
$\boldsymbol{W}$: spatial weighted matrix
$\boldsymbol{\beta}$: coefficients of predictor variable parameters vector
$\boldsymbol{u}$: error vector that has autocorrelation
$\boldsymbol{\lambda}$: spatial effect parameter coefficient error
$\boldsymbol{\epsilon}$: error

The variable used is the occupancy rate as the dependent variable. As for the independent variables, a number of nine variables were used, as mentioned in the previous section. We performed a significant test on all independent variables.

3. Results and Discussion

In the process of achieving the research objectives, namely understanding the impact of location on hotel occupancy rates, we have undertaken several tasks, including literature review, development of a web-scrapper used to collect data, data preprocessing, the data analysis using descriptive and inferential analysis, model construction using spatial regression, and reviews the previous researches. The resulting spatial regression model shows that there are several location factors that have a significant effect on hotel occupancy rates. This research is useful for determining the location for the appropriate hotel development. The location factor is expected to provide an overview of the hotel's strategic location to increase the hotel's occupancy rate.

The use of big data in this study has several advantages, including the fact that data collection is easier to do than conventional methods. In addition, data collection through the scraping process does not incur a high cost. The data collection process is faster. However, using big data presents challenges, namely the need for better computing devices. Additionally, big data exhibits a significant data variance, requiring validation.

This study uses hotel data collected using a web-scraper that has been developed by Adhinugroho [23]. In this study, it was found that the occupancy rate data in the online hotel reservation application has the same pattern as the occupancy rate data issued by the Indonesian statistical BPS. This proves that this big data source can represent the official data effectively with additional location data.

In this study, spatial regression was used to assess the impact of location on hotel occupancy rates. We began with simple linear regression using ordinary least squares before moving to spatial regression (ols). In the modeling phase, it was found that the model did not meet the classical assumptions. Classical assumptions that are not met are normality, autocorrelation, and heteroscedasticity. Therefore, the ols model is not adequate for use with this data. Hence, we use a spatial regression model as an alternative considering that the data is spatial data. The complete results of this study are shown in the explanation below.

3.1. Data collection

We gathered data from various sources. The main data in this study is hotel reservation data on online hotel reservation applications. The hotel reservation application used is one of the largest hotel reservation applications in Indonesia, namely Agoda.com. Agoda data is hospitality data obtained using the web-scraping method. The scraper used was based on a previous study conducted by Adhinugroho in 2020 [23].

In addition to utilizing data from hotel reservation applications, this study utilizes urban object data acquired from Google Maps. A web-scraping process is carried out using Python programming to get the location of an object on Google Maps. By using the webdriver-based Selenium library, a robot is formed to retrieve urban objects with certain keywords.

The first step is to collect a list of the names of the villages in the Province of Yogyakarta. However, because the hotel distribution is not in the entire province of Yogyakarta, only all urban villages in Yogyakarta City and some villages in Sleman and Bantul districts are listed. Furthermore, the office of each sub-district was taken as a benchmark point for collecting object location data.

The next step is to do a keyword search for urban objects. There are several keywords used, namely “airport”, “bus stop”, “station”, “restaurant”, “office”, “company”, “mall”, “attractions”, and “university”. For some keywords such as “airport”, “bus stop”, “station”, etc., it only needs to be done at one benchmark point because it has issued object results in all provinces. Meanwhile, keywords such as “restaurant”, “office”, “company”, and “attractions” need to be repeated as many as the number of benchmark points. The results obtained in this step are a list of names of urban objects on Google Maps according to the number of benchmark points. Each page contains a maximum of 20 object names.

Then from the list of objects, we automate click events on the object names one by one. Then the HTML tag containing the name, address, category, number of reviews, rating, and page URL is taken. From the page URL, the longitude and latitude of the object are extracted using a string separation.

The disadvantage of using this method is the search results cannot be limited by a specific region. For example, if we want to retrieve restaurant data in the Gedongkiwo Sub-district, Google Maps cannot limit restaurants in that sub-district but it provides a list of all restaurants close to the benchmark point. Thus, it is necessary to carry out manual selections after retrieving all the data in the village. This process is necessary because the use of the Selenium library takes a long time for a single execution. An overview of Google Maps scraping can be seen in Figure 2.

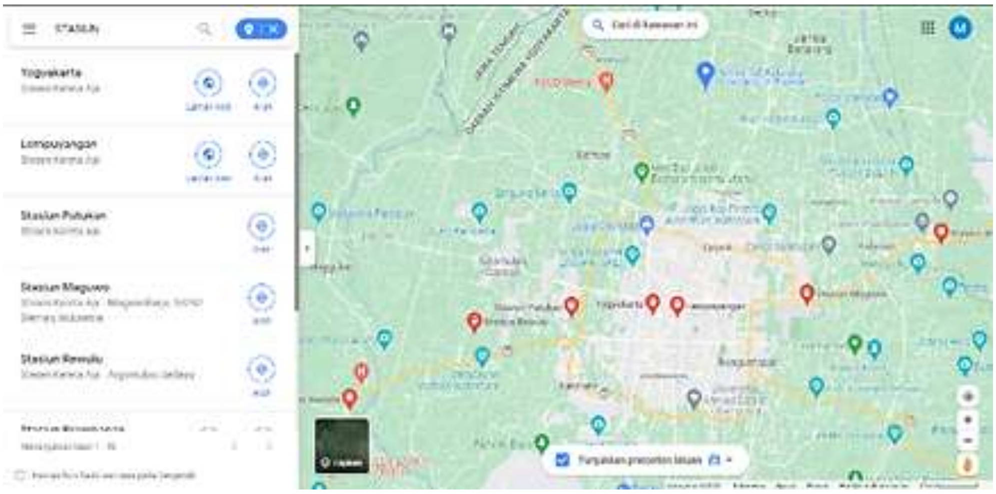

Figure 2. Google Maps scraping process

After the scraping process on Google Maps was carried out, urban objects were obtained from several categories, namely "airport", "bus stop", "station", "restaurant", “attractions”, etc. The number of objects obtained is different for each category. They can be seen in Table I below:

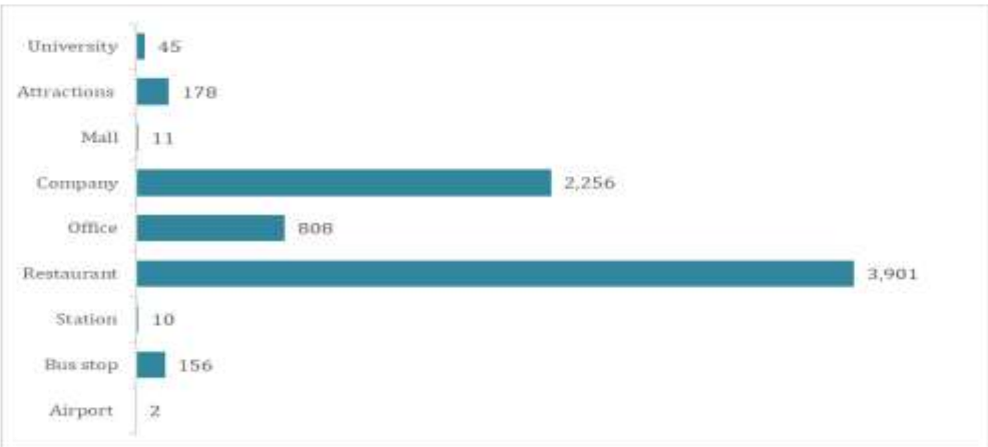

Figure 3. Number of Google Maps scraped objects

In Figure 3, the results from Google Maps data collection are mostly in the restaurant category, totaling 3,901 restaurants, followed by companies and offices, which, when combined, reach 3,064 objects. The lowest are airports and stations, which are 2 and 10, respectively. Figure 4 shows the distribution of each urban object from the scraping results of Google Maps.

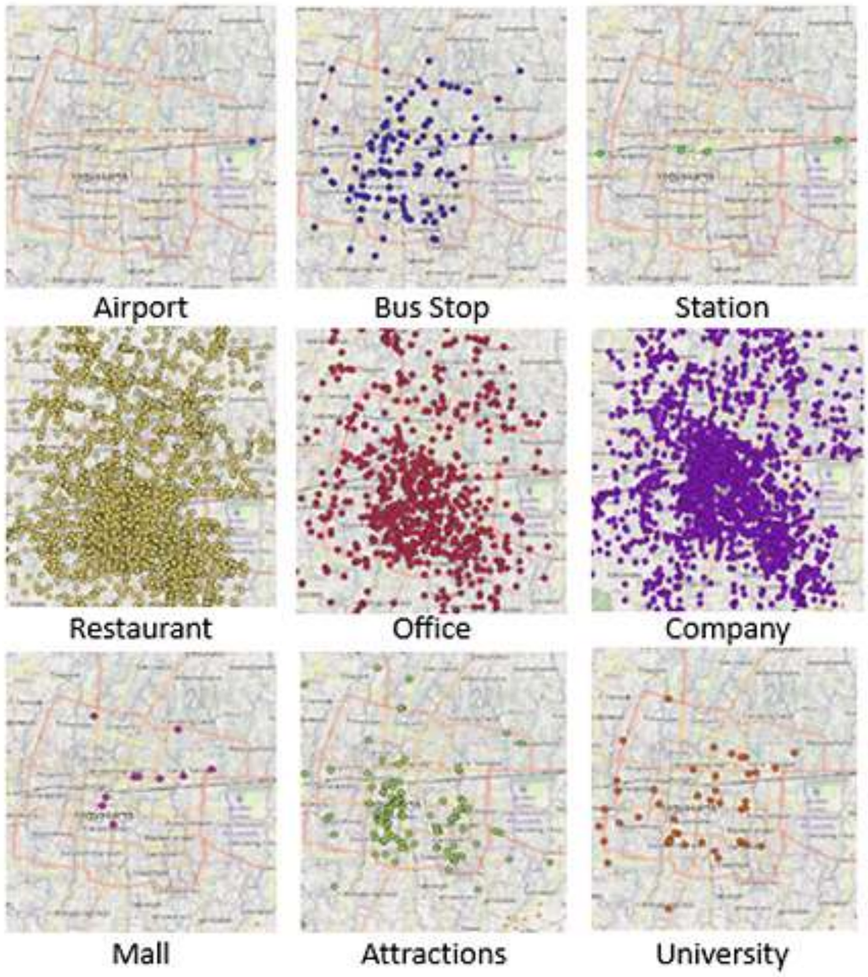

Figure 4. Distribution of Google Maps web-scraping data

To evaluate the web-scraper that has been built, the time needed to run it will be measured. Table 1 shows the time it takes to run the web-scraper for each stage. The average time required for Stage 1 and Stage 2 is relatively the same. The longest stage is in stage three, which is getting details of urban objects. This is because, after retrieving detailed object data, the web-scraper needs to return to the previous page, which takes twice as long as the previous stage. After getting urban objects from Google Maps, each hotel will measure the distance to the nearest object. Each existing hotel is paired with the nearest airport, station, bus stop, mall, and university using the Euclidean distance or the length of a straight line from two points. To calculate the distance traveled by two objects, measurements are made using the API from OpenStreetMap.

Table 1. Running time web-scraper

Stage	Running time	Number of data	Average time
Stage 1	1,251.726 seconds	206	6.08 seconds
Stage 2	6,418.425 seconds	1034	6.21 seconds
Stage 3	11,321.421 seconds	1000 (simulation)	11.36 seconds

The OpenStreetMap API is used because it is open source and offers free API access. However, it has a limit of 2,000 requests per day for each API. For objects that are categorized as restaurants, tourism objects, and offices and companies, the number of objects around the hotel is calculated using a radius of 1 km. For the effect of agglomeration or market competition between hotels, the same calculation method is used.

3.2. Data Preprocessing

Data preprocessing is conducted to prepare the data before it is used in the analysis stage. There are several tasks at the data preprocessing stage, such as checking for missing values, removing duplicates, eliminating inappropriate, and joining data. Data preprocessing is carried out on both hotel and urban object data. In the hotel data, several variables have the potential to be used in this study, such as the number of restaurants, the number of bars, and the price per room. However, these variables cannot be used because there are many missing values. Consequently, these variables are removed from our analysis. In hotel data the hotel ID, latitude, longitude, date, number of rooms reserved, and total number of rooms are used.

We initially filtered the hotel data according to the research period, which is June 2019. Additionally, only active hotels are taken. An active hotel on a day is characterized by having at least one hotel room reserved on that day. The hotels that were active during the research period were those that were active for at least 20 days in June 2019. In the Google Maps data, we began the preprocessing stage by eliminating duplicate data. Then, the data is filtered based on the description: *open or close*. The next step is to manually filter by name and category of urban objects. For example, an object named "shelter airport" with the category "airport", which does not have a matching category, is deleted.

3.3. Hotel Occupancy Rate Estimation

Based on the literature study, hotel occupancy is strongly influenced by the location of the hotel. Factors that affect hotel occupancy are distance to the airport, station, bus stop, mall, and university, as well as the number of offices, restaurants, tourist attractions, and hotels within a 1 km radius. Before formulating the model using regression, it is necessary to look at the relationship between the dependent and independent variables using the Pearson correlation value. The following graph shows the correlation of each independent variable with the dependent variable of hotel occupancy.

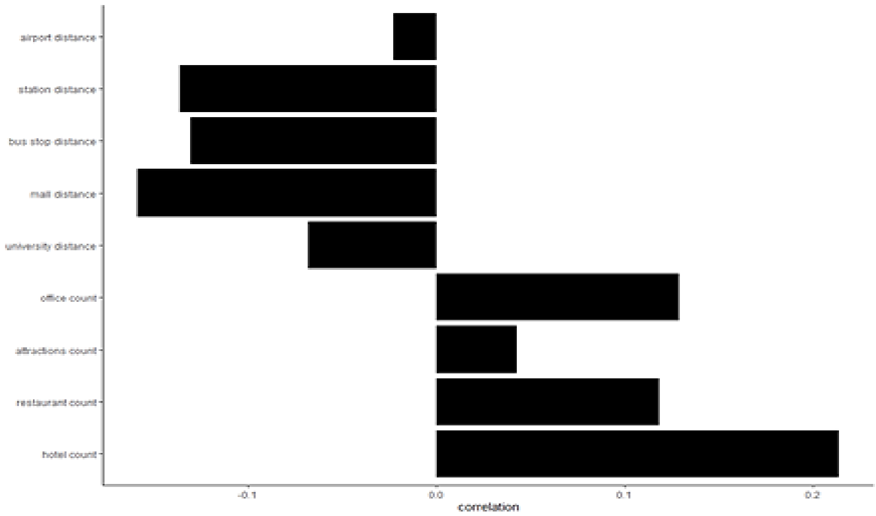

Figure 5. Variable correlation value

Figure 5 shows the pattern of relationships represented by the correlation value between the dependent variable and each independent variable. The variable "count" has a positive relationship with the occupancy rate. This indicates that the greater the number of urban objects within a 1 km radius of the hotel, the higher the hotel occupancy rate. While the variable "distance" has a negative relationship with the level of hotel occupancy. This indicates that the closer the hotel is to certain urban objects, the higher the occupancy rate of the hotel will be.

3.3.1. Spatial Autocorrelation Test

In regression modeling using spatial data, the most suitable autocorrelation test is the Global Moran's I test. This test is used to detect the spatial effect between neighboring observation units. The initial hypothesis or null hypothesis in this test is that the data is randomly distributed in a certain area, while the alternative hypothesis is that the data are not randomly distributed and have spatial autocorrelation. From the results of the Moran I test, the test statistic was 0.078014977 and the p-value was 0.008044. With a significance level of 5%, it can be concluded that the data are not randomly distributed and have spatial autocorrelation. Thus, spatial analysis can be an appropriate analytical step for the data in this study.

3.3.2. Spatial Regression

Before performing the spatial regression modeling, the Lagrange Multiplier was tested. The Lagrange Multiplier test aims to determine certain spatial effects on the model. The purpose of this test is to determine the spatial regression model that fits the data. If the Lagrange Multiplier test results in rejecting H0 in both tests, it can be concluded that spatial lag and spatial error exist in the model. The test will be continued with the Robust Lagrange Multiplier test. It is used because bias is generated by the Lagrange Multiplier test. The following table shows the results of testing the Lagrange Multiplier and Robust Lagrange Multiplier on spatial lag and spatial error.

Table 2. Result lagrange multiplier test

LM-test	statistic	p-value
Lagrange Multiplier (lag)	5.2009	0.02258**
Robust LM (lag)	2.5241	0.1121
Lagrange Multiplier (error)	9.4567	0.002104**
Robust LM (error)	6.78	0.009219**

Notes: **significant at 5%

Based on Table 2, it can be seen that the p-value in the spatial lag test has a value of 2.25%. Thus, at the 5% significance level, there is sufficient evidence of the occurrence of a spatial lag in the model. Meanwhile, when testing the spatial error in the model, the p-value is 0.21%. At the 5% significance level, there is sufficient evidence that there is an error link between regions or there is a spatial error. In the Lagrange multiplier test, significant results on spatial lag and spatial error are obtained. It will be continued with the Robust Lagrange Multiplier test. It can also be seen that the p-value of the spatial error is significant at a significance level of 5%, while the spatial lag is not significant. Therefore, the SEM model with the generalized methods of the moment estimation method is the right model to use. In this study, the autocorrelation test uses weights based on neighbors using the radius method. The distance or radius used is 3.2 km which is obtained from the maximum value of the distance between the hotel and the nearest other hotel. The following table shows the result of modeling using the SEM model with a spatial weighting matrix of the neighboring radius of 3.2 km.

Table 3 shows that there are five significant variables out of 9 independent variables at a significance level of 10%. At the 5% significance level, there are four significant variables. The significant variables are the distance to the airport, the distance to the bus stop, the number of offices, the number of restaurants, and the number of hotels in the vicinity. Therefore, the SEM model obtained is:

$$\begin{aligned}\widehat{Occupancy_i} = {} & 0.80393 + 0.005258\,airport\ distance_i^{*} + 0.008216\,station\ distance_i \\ & - 0.03822\,bus\ stop\ distance_i^{**} - 0.00701\,mall\ distance_i \\ & + 0.004436\,university\ distance_i - 0.00257\,\mathrm{restaurant\ count}_i^{**} + 0.000476\,offi \\ & + 0.000075\,attractions\ count_i + 0.006901\,hotel\ count_i^{**} + u_i^{*}\end{aligned}$$

$$\boldsymbol{u_i^{*} = -1.9734\Sigma_{j=1,i\neq j}^{n} w_{ij} u_j} \quad (5)$$

Notes: * significant at 10%; **significant at 5 %

Table 3. SEM model result

variable	coefficient	p-value
(intercept)	0.80393	< 0.0001**
airport distance	0.005258	0.087688*
station distance	0.008216	0.171019
bus stop distance	-0.03822	0.003868**
mall distance	-0.00701	0.435993
university distance	0.004436	0.564319
office count	0.000476	0.011438**
attractions count	-7.50E-05	0.946433
restaurant count	-0.00257	0.003601**
hotel count	0.006901	1.8E-05**
λ	-1.9734	4.52e-06**

Notes: *significant at 10%; **significant at 5%

Table 3 also shows that the coefficient has a p-value of less than 0.05. This indicates that spatial error has a significant effect on the model formed. The R-Square value obtained for the SEM model is 17.5%. This value increases when compared to the value in linear regression, which is only 6.7%. When compared with the AIC value, the SEM model has a lower AIC value of -136.14 while the linear regression model has an AIC value of -117.1. Therefore, the SEM model can be considered superior to the multiple linear regression model.

The distance to the airport significantly affects the hotel occupancy rate in the model. This is in line with the research by Valentin and O'Neill, 2018 [9] which explains that accessibility or distance to the airport affects the hotel economy. The agglomeration effect described by the number of hotels within a 1 km radius also affects hotel occupancy. Chung and Kalnins [24] confirmed that the number of hotels in the vicinity will affect high demand to increase the occupancy rate of a hotel. The variable number of restaurants around the hotel significantly affects the hotel's occupancy rate with a negative coefficient value. It is in line with the findings by Yang [20] that restaurants will affect non-star hotels that do not have restaurant facilities, while star hotels tend not to be located in areas with many small-business scale restaurants. It is in line with Fang [17] that a hotel's economy is significantly affected by the number of business units (described by the number of offices). The more business units around the hotel, the more the hotel accommodation will increase. The distance to the bus stop has a negative and significant relationship. It illustrates that the closer the hotel is to the bus stop, the higher the occupancy rate. However, this is not in line with Yang's research [20], which states that accessibility to bus stops will only affect hotels located on the city outskirts.

4. Conclusions

Web-Scraper Google Maps has been developed to generate data on urban objects, ranging from airports, stations, bus stops, restaurants, offices, companies, malls, and universities. The results of this web-scraping are used at the analysis stage as the dependent variable. Distances to airports, stations, bus stops, malls, and universities are used as the distance from the hotel to the nearest object. At the same time, the remaining objects are used to determine the number of urban objects around the hotel.

From the results of the experiment, the model constructed in this study is the Spatial Error Model (SEM). The model concludes that spatial location has an impact on the occupancy level of hotels in the online hotel reservation application in the Province of Yogyakarta. The impact of spatial location that affects the occupancy rate is the distance from the hotel to the airport, the distance from the hotel to the bus stop, the number of nearest restaurants, the number of nearest offices, and the number of closest hotels.

Ethics approval

Not required.

Acknowledgments

Not required.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

Derived data supporting the findings of this study are available from the corresponding author on request.

Credit Authorship

Muhammad Fachry Nazuli: Conceptualization, Methodology, Data Collection, Data Analysis, and Manuscript Writing. **Satria Bagus Panuntun**: Data Collection, Software Development. **Addin Maulana**: Manuscript Review and Editing. **Takdir**: Manuscript Review and Editing. **Setia Pramana**: Conceptualization, Methodology, Manuscript Review, Research Advisor.

References

[1] R. Lado-Sestayo, M. Vivel-Búa, and L. Otero-González, "Survival in the lodging sector: An analysis at the firm and location levels", *International Journal of Hospitality Management*, 59, 19 – 30, 2016. https://doi.org/10.1016/j.ijhm.2016.08.005

[2] R. Lado-Sestayo, and M. Vivel-Búa, "Diagnosis of bankruptcy hospitality and tourist destination; [Diagnosis de quiebra hotelera y destino turístico]". *Lurralde: Investigacion y Espacio*, 41, 149 – 174, 2018. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85026880026&partnerID=40&md5=ccc920fcfb7dc05ab6e36304ec94ed0f

[3] T. Kajla, S. Raj, S. Sharma, M. Joshi, and A. Kaur, "Key preferences of tourists during COVID-19 pandemic in luxury hotels: Evidence from qualitative data", *Tourism and Hospitality Research*, 22(4), 473 – 487, 2022. https://doi.org/10.1177/14673584211066742

[4] M. Li, L. Fang, X. Huang, and C. Goh, "A spatial–temporal analysis of hotels in urban tourism destination", *International Journal of Hospitality Management*, vol. 45 34-43, 2015. https://doi.org/10.1016/j.ijhm.2014.11.005

[5] H. Luo, and Y. Yang, "Intra-metropolitan location choices for star-rated and non-rated budget hotels: The role of agglomeration economies", *International Journal of Hospitality Management*, 72-83, 2017, https://doi.org/10.1016/j.ijhm.2016.09.007

[6] R. J. G. Calinao, M. V Amores, E. M. A. Rellores, and F. T. G Tabla, "The Determinants of Hotel Room Price In Philippines: A Structural Equation Modeling Analysis", *Journal of Applied Structural Equation Modeling*, 6(1), 2022. https://doi.org/10.47263/JASEM.6(1)02

[7] Z. Xiang and M. Krawczyk, "What Does Hotel Location Mean for the Online Consumer? Text Analytics Using Online Reviews", *Information and Communication Technologies in Tourism*, 383-395, 2016. http://dx.doi.org/10.1007/978-3-319-28231-2_28

[8] G. Li, R. Law, H. Q. Vu, J. Rong, and X. R. Zhao, "Identifying emerging hotel preferences using Emerging Pattern Mining technique", *Tourism Management*, 46, 311-321, 2015. http://dx.doi.org/10.1016/j.tourman.2014.06.015

[9] M. Valentin, and J. W. O'Neill, "The Value of Location for Urban Hotels", *Cornell Hospitality Quarterl*, 5-24, 2018, https://doi.org/10.1177/1938965518777725
[10] Y. Yang, Z. Mao, and J. Tang, "Understanding Guest Satisfaction with Urban Hotel Location", *Journal of Travel Research*, 1-17, 2017. https://doi.org/10.1177%2F0047287517691153
[11] BPS Province D.I. Yogyakarta, Daerah Istimewa Yogyakarta Province in Figures 2021, BPS:2021 [Online]. Available: https://yogyakarta.bps.go.id/en/publication/2021/02/26/3a501d00eaa097f65efc96f9/provinsi-di-yogyakarta-dalam-angka-2021.html
[12] S. Pramana, D. Y. Paramartha, G. Y. Ermawan, N. F. Deli, W. Srimulyani, "Impact of COVID-19 pandemic on tourism in Indonesia", *Current Issues in Tourism*, 2022. https://doi.org/10.1080/13683500.2021.1968803
[13] Susilawati, R. Falefi, and A. Purwoko, "Impact of COVID-19's Pandemic on the Economy of Indonesia", *Budapest International Research and Critics Institute-Journal (BIRCI-Journal),* 1147-1156, 2020. 10.33258/birci.v3i2.954
[14] BPS, *Klasifikasi Baku Lapangan Usaha (KBLI) [Standard Classification of Business Fields 2020]*, (2020), BPS:2020 [Online]. Available: https://ppid.bps.go.id/upload/doc/KBLI_2020_1659511143.pdf
[15] Y. Yang, and Z. Mao, Z. "Location advantages of lodging properties: A comparison between hotels and Airbnb units in an urban environment", *Annals of Tourism Research*, 81, 2020. https://doi.org/10.1016/j.annals.2020.102861
[16] R. Lado-Sestayo, M. Vivel-Búa, and L. Otero-González, "Connection between hotel location and profitability drivers: an analysis of location-specific effects", *Current Issues in Tourism*, 1–18, 2018. https://doi.org/10.1080/13683500.2018.1538203
[17] L. Fang, H. Li, and M. Li, "Does hotel location tell a true story? Evidence from Geographically Weighted Regression Analysis of Hotels in Hong Kong", *Tourism Management*, 72, 78-91, 2019, https://doi.org/10.1016/j.tourman.2018.11.010
[18] L. Fang, Y. Xie, S. Yao, and T. Liu, "Agglomeration and/or differentiation at regional scale? Geographic spatial thinking of hotel distribution – a case study of Guangdong, China", *Current Issues in Tourism*, 1–17, 2020. https://doi.org/10.1080/13683500.2020.1792852
[19] W. Srimulyani, M. Faris, N. F. Deli, and S. Pramana, "Profile of Occupancy Rate Hotel in NTB During Pandemic Covid-19 With Big Data Approach", *Seminar Nasional Official Statistics*, 273-280, 2020, https://doi.org/10.34123/semnasoffstat.v2020i1.503.
[20] Y. Yang, K. K. F. Wong, T. Wang, "How do Hotels Choose Their Location? Evidence from Hotels in Beijing", *International Journal of Hospitality Management*, 675-685, 2012. https://doi.org/10.1016/j.ijhm.2011.09.003
[21] F. Yalcin, M. Mert, "Determination of hedonic hotel room prices with spatial effect in Antalya". *Economía, Sociedad y Territorio*, 18, 697-734, , 2018. Doi: 10.22136/est20181228.
[22] J. P. LeSage. *The Theory and Practice of Spatial Econometrics*. Toledo: University of Toledo, 1999.
[23] Y. Adhinugroho, A. P. Putra, M. Luqman, and E. Ermawan, "Development of online travel Web scraping for tourism statistics in Indonesia", *Information Research*, 25(4), 2020. https://doi.org/10.47989/irpaper885
[24] L. Woo and S. G. Mun, "Types of agglomeration effects and location choices of international hotels in an emerging market", *Tourism Management*, 77, 2019. https://doi.org/10.1016/j.tourman.2019.104034