A comparative of Beach Pool Villa Prices Across Three OTAs

Agoda,Traveloka and Trip.com

Presented By

Thanapon Nitchaphatchanakul
66102010168

Tammarat Sukkha
6102010170

Phimchanok Thongpool
66102010181

Phattharamongkonchan Puttiworrapong
66102010182

Present To

ASST.PROF.DR.Ratchainan Thammasudjarit

This project is part of the study in DE282 Statistics for Data Engineering

Course at Srinakharinwirot University

Academic Year2024 , Semester 1

Introduction

This report is designed to present data and analysis on the prices of Pool Villa accommodation from popular accommodation booking websites such as Agoda, Traveloka, and Trips.com. The objective is to help users compare prices and make informed decisions about the most suitable accommodation. The data collected is the result of searches during a specific time period and is further analyzed statistically to show price trends and related factors.

Background and Significance

The research and data collection in this report were conducted using popular online accommodation booking platforms in Thailand and abroad, such as Agoda, Traveloka, and Trips.com. The data covers the accommodation name, province, search date, and accommodation price on the specified check-in date to obtain an overview of the market price during that period.

Objectives

-To compare accommodation prices from various platforms

-To calculate the mean and standard deviation of accommodation prices

-To perform statistical analysis, such as price difference analysis via ANOVA

-To provide information to interested parties in a format that is easy to read and understand

Data Collection Method

Data collection for this report was conducted using a clear and precise process to ensure quality and useful data for analysis. The data collection steps included:

Selection of data sources

Select popular accommodation booking websites:

-Agoda

-Traveloka

-Trips.com

Determination of variables and data required

Data collected were as follows:

-Property name (Pool Villa Name)

-Province of accommodation

-Search Date

-Check-in Date

-Prices from each platform (Agoda, Travelloka, Trips.com)

-Search process

Search Process:

-Search for accommodations on each website on the specified date and time.

-Specify the dates of your stay to get similar prices for comparison.

Data Recording:

-Record accommodation price data in Excel format with relevant details

-Use separate data recording of data sources from each platform to prevent confusion.

Verification:

-Check the consistency of the accommodation prices from each website.

-If there are any discrepancies or incomplete information, search again or update the incomplete information..

Data organization:

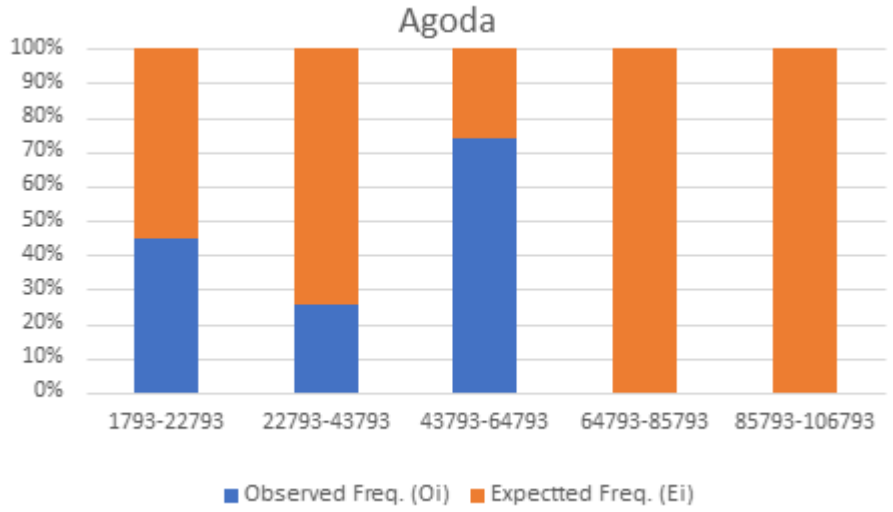
-Consolidate data from all platforms into a single file.

-Separate data into sheets for easy further analysis, such as calculating means and standard deviations.

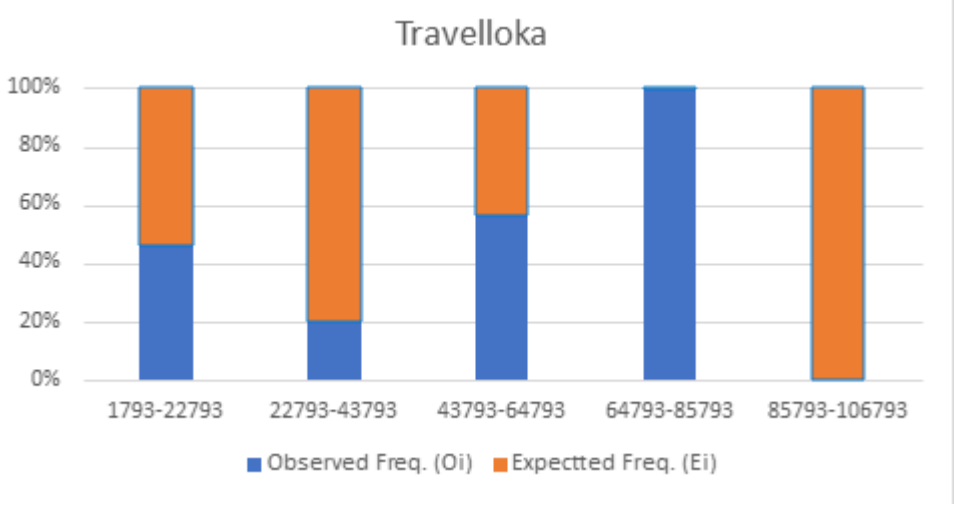
Pool Villa Name	Province	Search Date	Chickin Date	Agoda	Travelloka	Trips.com
Sun Diego Resort Pool Villa	ชลบุรี	26/10/2024	28/10/2024	24498	23147	16839
hyatt regency	ประจวบคีรีขันธ์	26/10/2024	28/10/2024	23000	22288	23000
V Villas Hua Hin	ประจวบคีรีขันธ์	26/10/2024	28/10/2024	22800	22800	22800
The Gems Mining Pool Villas Pattaya	ชลบุรี	26/10/2024	28/11/2024	13354	15503	19923
Banyan Tree Krabi	กระบี่	26/10/2025	11/11/2024	44036	44732	32076
sala Khao yai	นครราชสีมา	26/10/2024	17/11/2024	64363	65552	65552
Sri Panwa Phuket Luxury Pool Villa Hotel	ภูเก็ต	26/10/2024	17/11/2024	38274	40577	37537
เดอะ ไพรเวต พูล วิลลา แอท ศิริไลซ์ฮิลล์ เขาใหญ่	นครราชสีมา	27/10/2024	16/11/2024	17217	12615	18235
Villa Saifon	กระบี่	27/10/2024	19/11/2024	14667	15490	13373
อันดาร่า รีสอร์ท แอนด์ วิลล่า	ภูเก็ต	27/10/2024	20/11/2024	14614	14812	17718
Villa La Flora Kanchanaburi	กาญจนบุรี	27/10/2024	20/11/2024	9216	8770	9027
ลา มิเนียร่า พูล วิลล่า พัทยา	ชลบุรี	27/10/2024	21/11/2024	12364	12698	12744
Rice Villa ChiangMai	เชียงใหม่	27/10/2024	21/11/2024	16120	16797	15113
6BR Luxury Tropical Pool Villa PH125	เพชรบุรี	27/10/2024	25/11/2024	9614	9980	9262
Panwaburi Beachfront Resort	ภูเก็ต	4/11/2024	2/12/2024	10686	9088	9649
วิลล่า พาราเมาท์ 2	นครนายก	4/11/2024	2/12/2024	9159	10375	8983
ศิวานา วิลลา หัวหิน	ประจวบคีรีขันธ์	4/11/2024	6/12/2024	6937	8767	5729
เคซี แกรนด์ รีสอร์ท เกาะช้าง	ตราด	4/11/2024	6/12/2024	9811	9073	9730
เดอะลิตเติ้ล ซอร์ เขาหลัก บาย คะตะธานี	พังงา	4/11/2024	6/12/2024	18318	17991	17842
ทริเฮาส์ วิลล่า - สำหรับผู้ใหญ่เท่านั้น	พังงา	4/11/2024	6/12/2024	21356	17619	18716
Anasiri Poolvillas Rayong	ระยอง	4/12/2024	12/12/2024	10310	11595	10233
แทงโก้ ลั๊กซ์ บีช วิลล่า สมุย	สุราฎธานี	4/11/2024	13/12/2024	9575	9660	10231
chivani Pattaya	ชลบุรี	4/11/2024	15/12/2024	4098	5590	4378
เอ็กซ์-พี ขนอม ฮาร์เบอร์เบย์ รีสอร์ต	นครศรีธรรมราช	4/11/2024	15/12/2024	4261	4265	4279
อมตะปุระ ชีวิว วิลล่า 26	กระบี่	5/11/2024	18/12/2024	15612	15504	15948
Baan KanTiang See Villas	กระบี่	5/11/2024	18/12/2024	5476	5522	5603
เทรเวลลอดจ์ นิมมาน	เชียงใหม่	5/11/2024	18/12/2024	3121	3472	2924
Ana Anan Resort & Villas Pattaya	ชลบุรี	5/11/2024	18/12/2024	3630	3589	3139
Good Moments Pool Villa	นครราชสีมา	5/11/2024	18/12/2024	8678	10430	8683
Chiangkhan River Mountain Resort	เลย	5/11/2024	18/12/2024	3419	3413	3145

Data Exploration

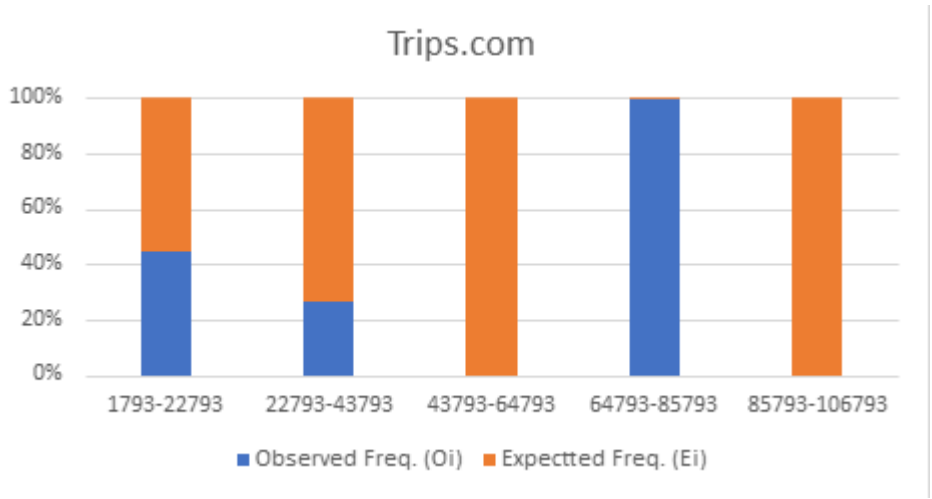
In this analysis, we will explore the relationship between accommodation prices and their frequency across various Online Travel Agencies (OTAs). This will be achieved by creating histograms in Excel, with the x-axis representing accommodation prices and the y-axis representing the frequency of accommodations. The frequency of accommodations will be categorized into five distinct bins for each OTA. Additionally, we will perform descriptive statistical analysis, including calculating the mean and standard deviation, using Excel.



Histogram displaying Agoda.com price



Histogram displaying Traveloka.com price



Histogram displaying Trip.com price

	Agoda	Traveloka	Trips.com
Mean	15619.46667	15723.8	15080.36667
St.Dev	13288.31558	13436.19746	12637.27552
Min	2924	3413	2924
Max	65552	65552	65552

The calculated results for each Online Travel Agency (OTA) are as follows:

- The mean price for Agoda is 15,619.47, with a standard deviation of 13,288.32.

- The mean price for Traveloka is 15,723.80, with a standard deviation of 13,436.20.

- The mean price for Trip.com is 15,080.37, with a standard deviation of 12,637.28.

Normality Test

This report aims to determine whether the price distributions of pool villas listed on Agoda, Traveloka, and Trip.com exhibit characteristics of a normal distribution. To evaluate this, a Chi-squared test for normality was applied to the price data. The prices were divided into five intervals (bins), and the observed frequencies were compared to the expected frequencies derived from a normal distribution.

The hypotheses tested are as follows:

- **H_0 (null hypothesis):** The price distribution aligns with a normal distribution.
- **H_a (alternative hypothesis):** The price distribution deviates from a normal distribution.

Given the use of five bins, the degrees of freedom (df) are calculated as $n-1=4n - 1 = 4n-1=4$. At a confidence level of 95% ($\alpha=0.05\alpha = 0.05\alpha=0.05$), the critical value for the Chi-squared statistic is approximately 9.49.

Agoda

Bin	Observed Freq. (Oi)	Expected Prop.	Expected Freq. (Ei)	(Oi-Ei)^2/Ei
1793-22793	24	0.705345507	29.62451131	1.067870019
22793-43793	4	0.277658781	11.66166879	5.033685115
43793-64793	2	0.016888128	0.709301392	2.348653078
64793-85793	0	0.000107519	0.004515813	0.004515813
85793-106793	0	6.42913E-08	2.70024E-06	2.70024E-06
Total	30	1	42	8.454726726

deg freedom	4
p-value	0.85712346

The test results revealed that Agoda had a Chi-squared value of **8.45**, which was lower than the critical value of 9.49. Therefore, we failed to reject the null hypothesis, indicating that the price distribution of Agoda **follows a normal distribution**. With **4 degrees of freedom**, the calculated p-value was **0.8571**, which is greater than the significance level (α=0.05\alpha = 0.05α=0.05). This further supports the conclusion that the price distribution for Agoda follows a normal distribution.

Traveloka

Bin	Observed Freq. (Oi)	Expected Prop.	Expected Freq. (Ei)	(Oi-Ei)^2/Ei
1793-22793	25	0.700601404	29.42525898	0.665513839
22793-43793	3	0.281048028	11.80401719	6.566469489
43793-64793	1	0.018220472	0.765259829	0.072005541
64793-85793	1	0.000130003	0.005460135	181.1511207
85793-106793	0	9.19235E-08	3.86079E-06	3.86079E-06
Total	30	1	42	188.4551134

deg freedom	4
p-value	1

For Traveloka, the Chi-squared value was **188.46**, which was significantly higher than the critical value of 9.49. Therefore, we rejected the null hypothesis, indicating that the price distribution of Traveloka **does not follow a normal distribution**.With **4 degrees of freedom**, the p-value was **1**. Despite the high p-value, the calculated Chi-squared value far exceeded

the critical value, confirming that the price distribution for Traveloka deviates significantly from a normal distribution.

Trips.com

Bin	Observed Freq. (Oi)	Expected Prop.	Expected Freq. (Ei)	(Oi-Ei)^2/Ei
1793-22793	25	0.729171179	30.62518952	1.033226492
22793-43793	4	0.259287342	10.89006837	4.359296977
43793-64793	0	0.011499674	0.4829863	0.4829863
64793-85793	1	4.1794E-05	0.00175535	567.6886918
85793-106793	0	1.09953E-08	4.61804E-07	4.61804E-07
Total	30	1	42	573.564202
deg freedom	4			

p-value	1

For Trip.com, the Chi-squared value was **573.56**, which was substantially higher than the critical value of 9.49. Therefore, we rejected the null hypothesis, indicating that the price distribution of Trip.com **does not follow a normal distribution**.

With **4 degrees of freedom**, the p-value was **1**. Despite this high p-value, the observed Chi-squared value far exceeded the critical threshold, confirming that the price distribution for Trip.com does not align with a normal distribution.

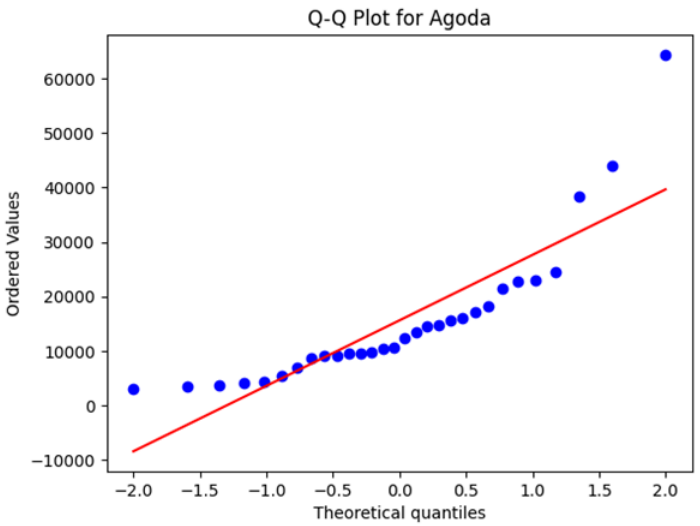
While the Chi-squared test provides initial insights, it has notable limitations. The test's accuracy heavily depends on how the bins are defined, and the method may lack precision with smaller sample sizes. To address these limitations, additional methods such as a **Q-Q plot** and the **Shapiro-Wilk test** were used to complement this analysis and assess the normality of the data more effectively.

Q-Q Plot

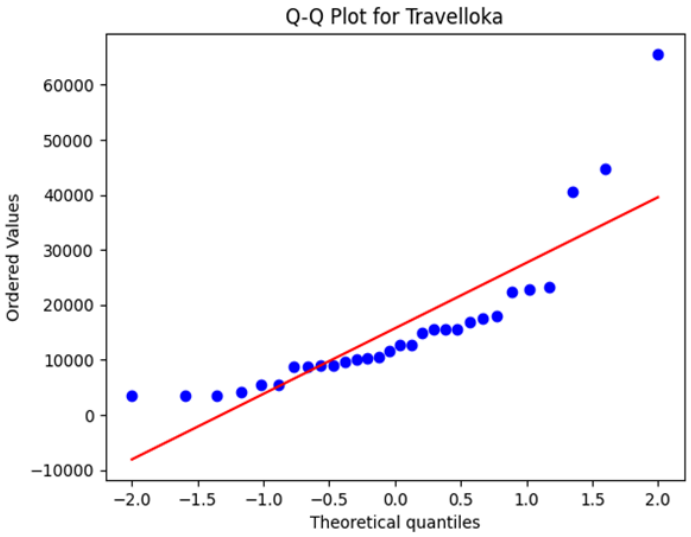
A Q-Q plot is a visual method used to evaluate whether a dataset's distribution aligns with a specific theoretical distribution, such as the normal distribution. It compares the quantiles of the observed data to those of the expected distribution. If the data conforms to the theoretical distribution, the points in the Q-Q plot will form a straight diagonal line. However, deviations from this line suggest discrepancies between the observed and expected distributions, highlighting features such as skewness or heavy tails.

Based on the Q-Q plots for prices from Agoda, Traveloka, and Trips.com, there are clear departures from the theoretical normal distribution, particularly at the extremes. These deviations indicate that the price data for all three platforms do not fit a normal distribution. Such observations align with the variability seen in the descriptive statistics, such as wide ranges and relatively high standard deviations, which point to non-normal characteristics in the datasets.

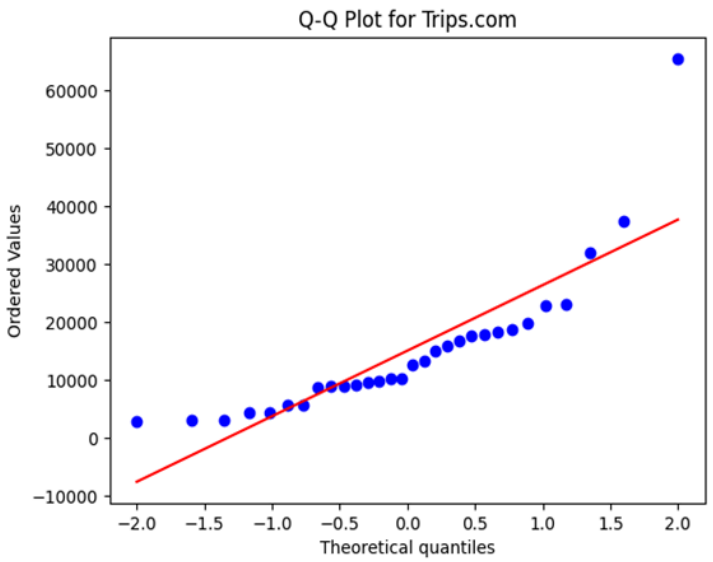
Agoda



Q-Q PLOT for Agoda.com



Q-Q PLOT for Traveloka.com



Q-Q PLOT for Trip.com

The Q-Q plots for prices from Agoda, Traveloka, and Trips.com indicate significant deviations from the theoretical normal distribution, suggesting that the price data do not follow a normal distribution. The points deviate from the 45-degree line, particularly in the upper tails, highlighting the presence of extreme values or heavy-tailed behavior. This pattern suggests variability in pricing, which is further supported by the descriptive statistics showing large standard deviations and wide ranges in the datasets.

Analysis of Price Distribution Across Agoda, Traveloka, and Trip.com

Q-Q Plot Insights

The Q-Q plots for prices from Agoda, Traveloka, and Trip.com reveal significant deviations from the theoretical normal distribution. The points deviate markedly from the 45-degree reference line, particularly in the tails. This suggests non-normality, likely due to heavy tails or the presence of outliers.

Descriptive Statistics

The summary statistics for the prices are as follows:

- **Mean Prices:** Agoda (15,619.47), Traveloka (15,723.80), Trip.com (15,080.37)
- **Standard Deviations:** Agoda (13,288.32), Traveloka (13,436.20), Trip.com (12,637.28)
- **Range:** Prices range from a minimum of 2,924 to a maximum of 65,552 across all platforms.

The high variability, as indicated by the standard deviations and wide range, underscores the non-normality observed in the Q-Q plots.

ANOVA

To compare the mean prices across Agoda, Traveloka, and Trip.com, a one-way ANOVA was performed:

- **Between Groups Variation:** The sum of squares (SS) between groups is 7,155,207.09, with 2 degrees of freedom (df). The mean square (MS) is 3,577,603.54.
- **Within Groups Variation:** The sum of squares within groups is 14,987,532,499, with 87 degrees of freedom. The mean square is 172,270,488.50.
- **F-statistic:** The calculated F-value is 0.0208, which is far below the critical value (F-crit = 3.1013).
- **P-value:** The p-value of 0.9794 indicates no statistically significant difference in mean prices across the three platforms at any reasonable significance level.

Chi-Squared Goodness-of-Fit Test

The Chi-squared tests for Agoda, Traveloka, and Trip.com indicate that the observed frequencies deviate from expected frequencies under a normal distribution, particularly in the extreme bins. For Trip.com, in particular, the Chi-squared statistic is excessively high (573.56), showing substantial non-normality.

The results from the Q-Q plots, descriptive statistics, Chi-squared goodness-of-fit tests, and ANOVA collectively highlight the following:

1. **Non-normality:** The price data do not follow a normal distribution, as evidenced by Q-Q plots and Chi-squared tests.
2. **High Variability:** Large standard deviations and a wide range of values indicate variability influenced by extreme prices or outliers.
3. **No Significant Differences in Means:** Despite the variability, ANOVA shows no significant differences in the average prices across Agoda, Traveloka, and Trip.com (p-value = 0.9794).

These findings suggest that while prices across platforms have similar averages, the distributions exhibit non-normal characteristics with high variability, making parametric

statistical assumptions less appropriate for further analysis. Non-parametric or robust statistical methods may better capture the underlying patterns in the data.

Reference:

This analysis is based on the methods and concepts presented in the YouTube video

https://www.youtube.com/watch?v=jC8PN29kTaU

Anova

Based on the Analysis of Variance (ANOVA) results, we aim to determine whether there is a statistically significant difference in the average prices of beach pool villas across the three selected OTAs: Agoda, Traveloka, and Trip.com. Since the prices are collected from different platforms, it is crucial to assess whether the platform (OTA) has an influence on the prices or if the observed differences are due to random variation.

Null Hypothesis (H_0): The OTA platform has no effect on the prices of beach pool villas. The mean prices of beach pool villas are the same across Agoda, Traveloka, and Trip.com. μAgoda Price = μTraveloka Price = μTrip.com

Price p-value ≥ 0.05

Alternative Hypothesis (H_a): The OTA platform has an effect on the prices of beach pool villas.

The mean prices of beach pool villas differ across at least one OTA.
µAgoda Price ≠ µTraveloka Price ≠ µTrip.com Price

p-value < 0.05

Anova: Single Factor

SUMMARY

Groups	*Coun*	*Sum*	*Average*	*Variance*
Column 1	30	468584	15619.46667	176579330.9
Column 2	30	471714	15723.8	180531402.1
Column 3	30	452411	15080.36667	159700732.4

ANOVA

Source of Variati	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Grou	###	2	3577603.544	0.020767362	0.979451649	3.101295757
Within Groups	###	87	172270488.5			
Total	###	89				

The ANOVA Table

Based on the Analysis of Variance (ANOVA) conducted to determine whether there is a significant difference in the average prices of beach pool villas across the three OTA platforms (Agoda, Traveloka, and Trip.com), the results showed that the p-value is 0.9795, which is greater than the significance level of 0.05. Additionally, the F-value is 0.0208, which is lower than the critical F-value of 3.1013. These results suggest that there is no statistically significant difference in the average prices among the three OTAs.

At the 95% confidence level, we fail to reject the null hypothesis, which states that the OTA platforms do not influence the average prices of beach pool villas. This implies that the observed price differences are likely due to random variation rather than being attributable to the specific platform used.

However, it's important to note that the assumption of normality for the data may not have been met, as indicated by the Q-Q plot and Shapiro-Wilk test results. ANOVA assumes that

the data within each group is normally distributed, and when the data deviates from this assumption, the results of the ANOVA test can become unreliable. Therefore, given this violation of the normality assumption, the ANOVA results may not be fully valid. In such cases, alternative methods should be considered to ensure more accurate conclusions.