Thailand Domestic Tourism (2019–2023)

Exploratory Data Analysis and Preparatory Data Science Report

Cresht

June 18, 2025

Table of Contents

Acknowledgement

My expressive gratitude can not be described within a few pages, utmost to my admired lecturer who has laid the foundation of my preliminary research pattern in writing course. Through dedicated handouts, numerous trials and errors and valuable experience in terms of his research, i was managed to set on my first research report. Though containing unpleasant flaws, this consequently becomes the prototype for my further projects.

I would like also extend my sincere to experienced data scientists on Kaggle, Thaweewat R., for having provided explanatory datasets on Kaggle that are incredibly useful for ongoing Data Science practice. Additionally, this endeavor would not have been possible without data technique overview from Marília Prata.

Finally, i would be remiss in not mentioning my family, especially my sisters. Their indispensable encouragement has motivated me to keep on refinements and to pursue professional standards.

These analytic insights, compiling both programmatic and interpretative style, can be both informative and even controversial. Believing or not, loving or hating, my work could be rare sources for scientists - regardless of what fields they major - to adopt and learn from my shortcomings.

Abstract

This study examines Thailand's tourism trends from 2019 to 2023. A theme-based explanatory data analysis (EDA) is carried out to explore patterns in accommodations, tourism revenue, regional differences and unique province. Both well-known and lesser-known provinces are investigated – Chiang Mai, Chiang Rai, Bangkok, Phuket and Uthai Thani to provide a balanced overview. The findings reveal key factors such as geography, cuisine and cultural heritage positively affect to the occupancy rate. The study also analyzes the aftermath of COVID-19, strong signs of recovery during post-COVID and uncovers hidden potential of rural provinces. This research lays important groundwork for future modeling efforts and supports tourism development through modest interdisciplanary initiatives.

1. Introduction

The tourism industry of Thailand resumes becoming a key contributor to both national GDP and regional economy. Nevertheless, the COVID-19 pandemic caused unforeseen disruptions to international sightseeing patterns, exploring disparities in tourist infrastructure and indecisive policy responses. To uncover these abrupt changes, this study adopts a theme-based procedure that goes beyond the conventional framework, particularizing oriented themes: tourist accommodations, revenue classification, regional differences and hidden gem emergence. Provinces are selected not only based on prominence, but also for their scenic beauty and tourist appeal – such as Bangkok, Phuket, Chiang Mai, Chiang Rai. Uthai Thani is included into this analysis to perceive travel behaviour and local development. This diversity brings about a balanced perspective over fluctuating period and evolving aspirations.

Provinces are selected not only based on prominence, but also for their scenic beauty and tourist appeal – such as Bangkok, Phuket, Chiang Mai, Chiang Rai. Uthai Thani is included into this analysis to perceive travel behaviour and local development. This diversity brings about a balanced perspective over fluctuating period and evolving aspirations.

2. Methodology

This dataset uses a preprocessing pipeline to prepare for explanatory analysis and predictive modeling. The stripped sources, recreated by the author Theerawat R. and retrieved from Official Ministry of Tourism and Sports Statistics of Thailand, consists of monthly records across 77 provinces that grouped into five geographical regions on total revenue, visitors (foreign, Thai and all) and tourist accommodations (both numeric and percentage formats). Supporting the concise and descriptive data analysis, several libraries are ultilized:

- Numpy, Pandas for data manipulation
- Matplotlib (Seaborn), Plotly for visualisation
- Scikit-learn for data preprocessing

This data-preprocessing pipeline includes the following quinessential stages:

1. **Feature Engineering & Data Cleaning:** Categorical columns are gone through ordinal data cleaning process. The dataset is divided into three main components:

(a) Revenue class

(b) Tourist class

(c) Numeric tourist accommodation class

Each component is performed separately so as to avoid corruptions resulting from overlapping categorical structures —- subjective and objectively.

- Invalid / Missing values resolution: Initial inspection indicated that none of columns contained NaN values, so this phase was surpassed.
- Outlier detection and removal: Extreme and invalid data points are statistically excluded by thorough consideration over available procedures – Z-score & Interquartile Range (IQR) method.

  Ultimately, the IQR was chosen for its efficiency, but may also cause unrealistic values when comparing between contrasting successors.
- Duplicated Values Detection: Categorical variables were verified to contain only unique values; no duplication was found.

2. **Data Preprocessing**: Moving onto preprocessing, the dataset although having near perfectly data entries, it contains descriptive columns that hinder modeling interpretation as machines do not understand verbal syntax as well as concerns about disparities in feature magnitudes.

   (a) Refined Date Formatting: The date column was converted into numeric columns – day, month, year for conveniece of feature aggregation.

   (b) Feature Compilation: Rearrange the cleaned sub-dataframes into a formal dataframe using `pd.concat()`

   (c) Preliminary Descriptive Feature Removal: Initial dataset contains two renditions of regional differences: Thai and Eng, since the explanatory analysis values accessibility and readibility, only the English version was retained.

   The dataset contains two categorical columns: `regional_eng` and `variable`. Nevertheless, it is important to remark: total revenue and tourists may be the combination of Thai and Foreign sources, which creates interpretational challenges. Furthermore, this assumption is also ambiguous as the author of the dataset didn't explain the variable inference such as foreign and domestic sources in revenue, thus each categorical sections are temporarily treated as different features to preserve flexibility and to avoid unforseen issues.

3. **Feature Encoding** Given earlier discussion about the unclear variables, a mixed-encoding approach is applied:

   (a) Label Encoding: `province_eng` — Containing 77 provinces which are appropriate for such high cardinality columns.

   (b) One-hot Encoding: `regional_eng` — Containing only 5 regions (low-cardinality column), applied due its robust interpretability and model compatibility.

4. **Feature Transformation**: In context of previous value distribution over total revenue, tourists and accomodations - which was previously done in initial inspection section, the KDE plot (Kernel Density Estimation) shows tendency of right-skewed data. To address this issue, several commonly-used transformers are regarded, specifically Box-Cox and Yeo Johnson methods. Yeo-Johnson transformation, while being proved for its non-strict positive data input, yieled less favourable results as Box-Cox one did despite strict input (allows for only positive input). Therefore, this dataset prefers the latter to the former.

5. **Feature Scaling**: This step was provided as an additional feature transformer since the distribution is already in balanced shape, though some of plots showing slightly less symmetric with unusual peaks, even after applying Standardization Method – aligning with Box – Cox Transformer.

   The final distribution was not perfectly consistent to some categorical variables.

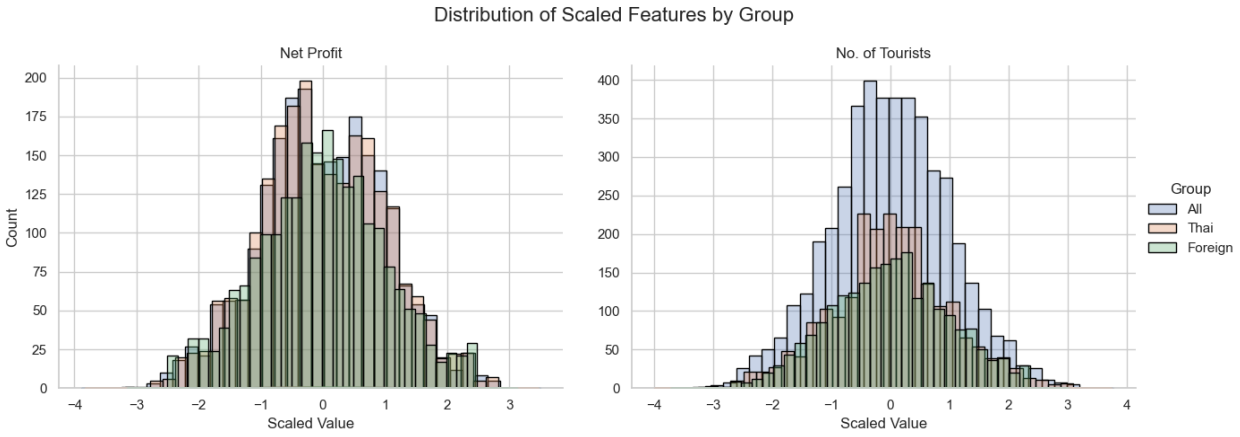
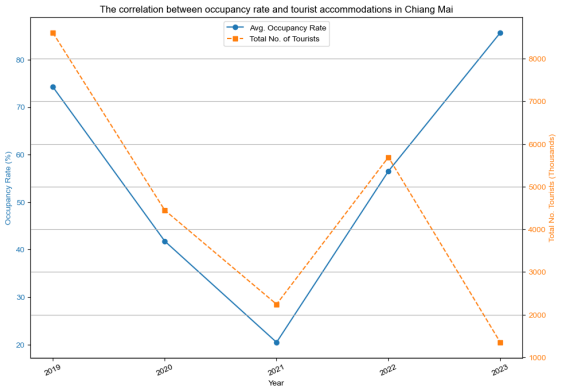


Figure 1: Distribution after Box-Cox Transformer and Standardization

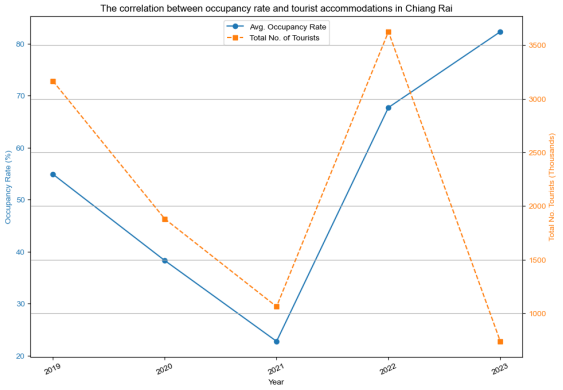
6. **Final Output Preparation:** These processed dataframes will be converted to different formats (commonly-used CSV or xlsx) for different purposes on predictive modeling tasks such as time series forecasting, regression modeling, etc.

3. Discussion

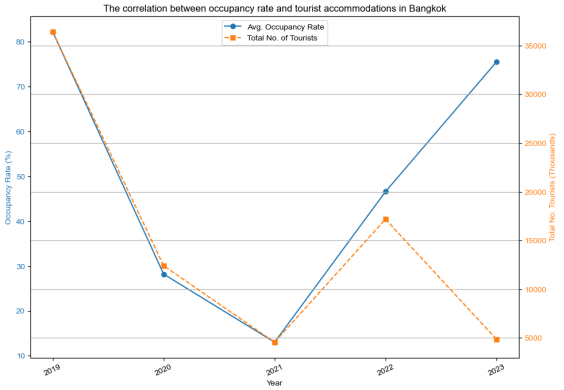
Task 1: Tourist Accommodations



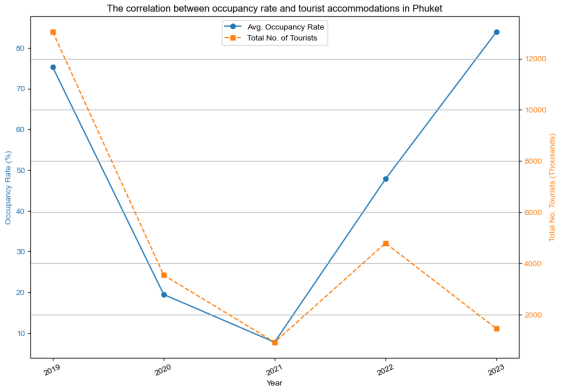
(a) Tourist Occupancy - Chiang Mai



(b) Tourist Occupancy - Chiang Rai



(c) Tourist Occupancy - Bangkok



(d) Tourist Occupancy - Phuket

Figure 2: Tourist Accommodation Distribution Among Selective Provinces

Observations:

The stable occupancy growth is witnessed in all provinces, with the exception of Chiang Rai, whose rate accounts for only approximately 60%. Despite a critical decline during 2020 – 2021, tourist occupancy began to show signs of recovery in 2022. Intriguingly, the number of tourist occupancies in Chiang Rai exceeded the pre-pandemic stage in 2019 (approximately 70% and 3,000 thousand accomodations).

Bangkok and Phuket have large proportion of accommodations – respectively more than 35,000 and 12,000 thousand. The concise explanation for such quantity can be attributed to the geography, particularly Phuket lies off west coast of mainland Thailand in the Andaman Sea and Bangkok situated in the Chao Phraya River delta in central coast. This aspect is

further specificed by 32% year-on-year increase in Phuket hotel room rates with ADR (Average Daily Rate) of 6,837 bath (Linab, 2024)[3] and 49% ADR increase in Bangkok luxury hotels, peaking at 15,000 baht per night[8]. Kittipong et al. (2019)[1] revealed that location such as beachfront and sea view significantly increased room prices, additionally high season perks help to raise the price up to 40% over normal seasons. Moreover, the proximity to tourist attractions and star rating had influenced significantly on price, particularly beachfront hotels charged 25 – 40% and over 50% during peak seasonal times from December to February (Weerawan et al., 2017)[10].

In short, tourism popularity and satisfaction have remarkably contributed to the price accommodation growth, resulting from seasonal demand planning and digital reputation management (e.g: rating systems on Agoda, Tripadvisor, etc.).

Task 2.1: Revenue Classification - Domestic, Foreign and Total

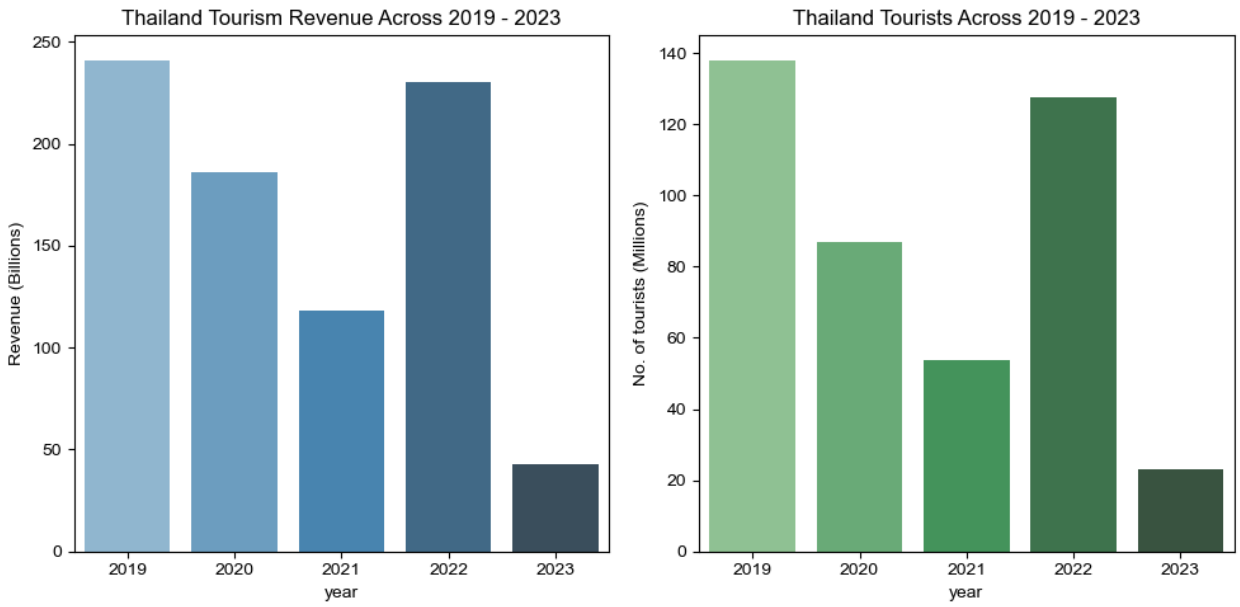


Figure 3: Thailand's Total Revenue And Vistors 2019 - 2023

Observations:

Thailand's average revenue sources from tourism witnessed a perk increase at the beginning of 2019 – estimatedly 250 billion baht and 140 million visitors to the country. Similar to the steep decline of occupancy rate, the domino effects also impacted the tourist income, as only half of 2019 total revenue and tourists was recorded. However, the tourist majority began to revive in 2022, leading to the near-preceding stage in both revenue and travelers.

Tourism continues to be a key driver for overall GDP growth of Thailand across 2019 – 2023,

as explicitly stated in 2022 with a 7.6% share in global GDP and a year-on-year measurement has been seen for an increase of 22% change in travel & tourism GDP (Travel & Tourism Council, 2023)[9]. Many factors contributed to strong performance in revenue, comprising heritage attractions such as The Grand Palace - Bangkok, cultural cuisine (Khao soi, Pad thai, Green / Red Curry, Mango Sticky Rice) and government initiatives in holiday strategy. According to Rakklin, P. et al (2025)[5], the Thai government established such as like visa exemptions for short-term visitors, digitalization emphasis on enhancing online booking systems with multilingual platform and campaigns promoting eco-tourism and wellness tourism.

These aspects indicate that Thailand's tourism is not purely market-driven, but instead reflects a meticulous collaboration between policymakers and local stakeholders that values technological advancement and sustainable development in long-term.

Task 2.2: COVID-19 & Aftermath

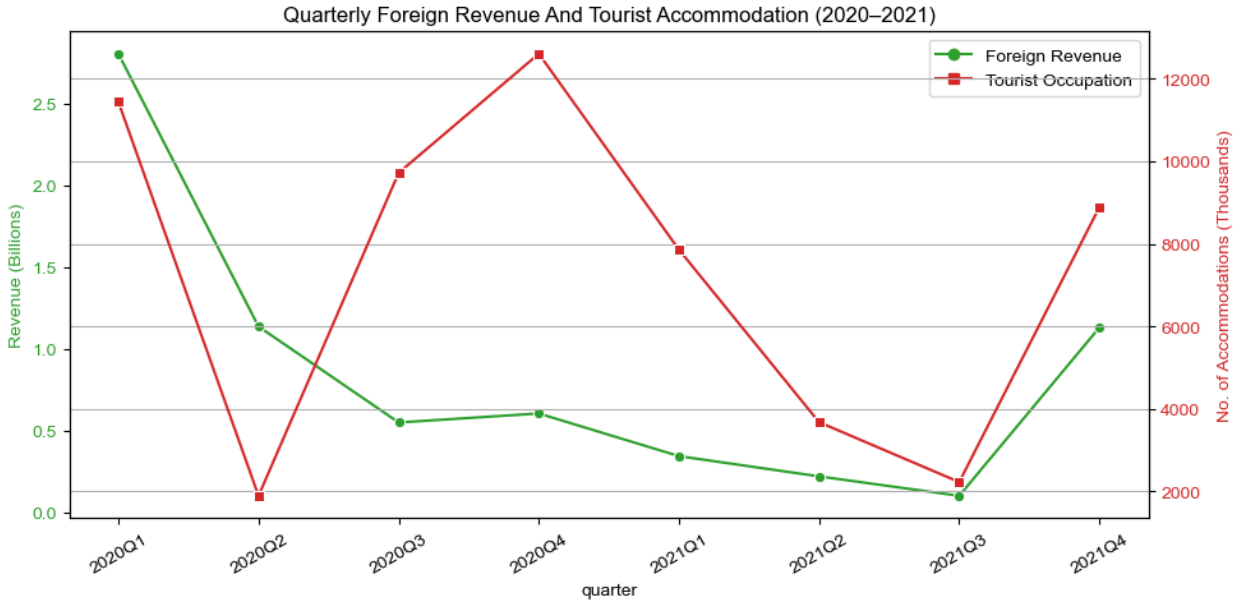


Figure 4: Thailand's Foreign Revenue And Accommodation 2020 - 2021

Observations:

Besides previously strong tourism revenue, Thailand tourism, as well as global tourism in general, experienced the aftermath of COVD-19 outbreak that began in early stages of 2020, causing widespread damage to a key contributor to the national GDP.

In the summer of 2020, the number of occupied tourist accommodations dropped acutely from around 12 million to approximately 2 million. Due to concerns about subtle but severe

symptoms posing threat on public health, economics closures were initiated as safest procedure, which in turn resulted in significant job losses in the hospitality and entertainment zones - 12% of skilled workers wages and especially 17% among unskilled larborers (Manop, U., 2020)[4]. During Q4 of 2020, a slight recovery was observed in number of tourist occupation returned to over 12 million to the original state, whereas the tourism revenue remained weak, recorded in slight increase. This short progress was interrupted by the emergence of Delta variant, causing devastating effects to the revenue that declined massively in Q2 – Q3 of 2021.

It was not until late year that signs of economics recovery appeared in accordance to accelerated vaccination programs, leading to positive recovery in both income and occupancy rates.

Task 3: Regional Differences

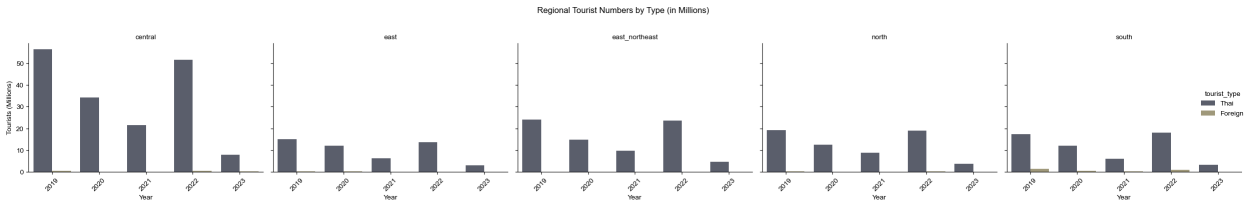


Figure 5: Regional Annual Visitors

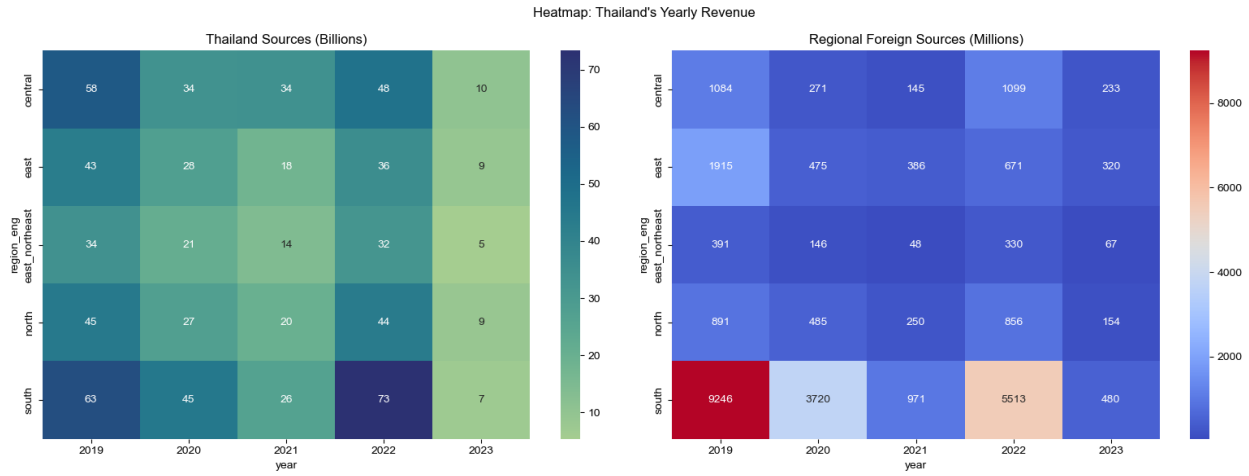


Figure 6: Regional Difference Revenue Heatmap

Observations:

An apparent tendency is indicated in tourism data: southern and central provinces generate higher revenue and attract more visitors than other places, while East-Northeast region lag

behind prior ones - for instance, this region generated only total of 34 billion baht in Thai revenue in 2019 and attracted around 20 million visitors.

Regarding foreign tourism revenue, southern areas surpassed with a total of 9.246 billion baht in 2019, following by 5.513 billion baht in 2022. The eastern region ranked second with 1.915 billion baht in 2019. Although the central region produced relatively low foreign tourism revenue, the domestic one held second place with the total of 58 billion baht in 2019.

These overlapping patterns in tourism revenue and visitor distributions can be ascribed to regional differences – such as viration in culture, common spoken languages and geographical distance between provinces. Reinforcing this perspective, Zihan et al. (2024)[2] stated that short-haul tourists prefer destinations with cultural similarity, often with desirable proximity though the minority may seek cultural contrast. In contrast, long-haul tourists are more likely to choose culturally distant destinations, where the novelty and difference enhance the appearance. The previous findings are also mirrored in Thailand, where regions such as East-Northeast, North and East show insignificant net profit. For instance, less well-known destinations such as Amphoe Chiang Khan of Loei Province, Tha Sadet Market (Nong Khai Province) contrast with astonding destinations like Phuket or metropolitan centers like Bangkok capital.

To some extent, the visualisation used in this analysis might not fully reflect the actual disparities due to the removal of significant outliers representing extreme values in tourism revenue and occupancy. This preprocessing step was taken so that machine could easily process well-distributed dataset, alongisde avoid potential distortion.

Task 4: The Hidden Gem

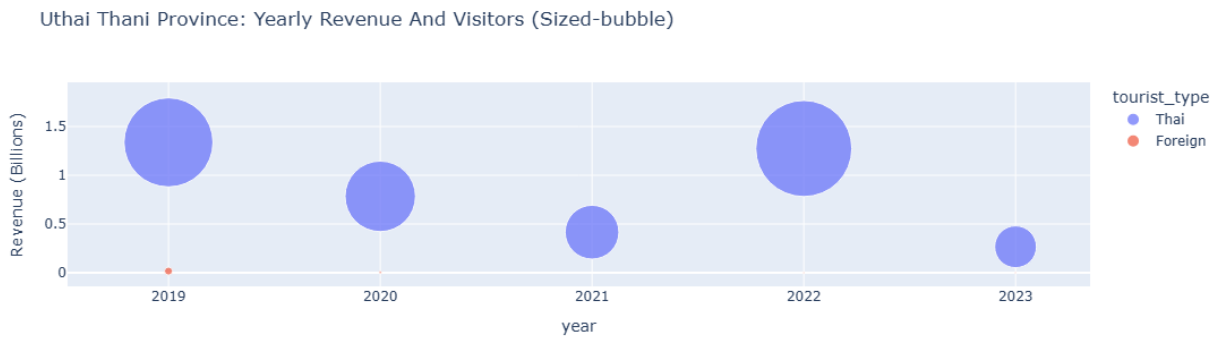


Figure 7: Visual Uthai Thani's Tourism Development

Table 1: **Uthai Thani's Tourism Development Table**

Year	Tourist Type	Revenue (Billions)	Visitors (Millions)
2019	Thai	1.3378	0.7661
	Foreign	0.0167	0.0063
2020	Thai	0.7838	0.4800
	Foreign	0.002	0.007
2021	Thai	0.4161	0.2799
	Foreign	0	0
2022	Thai	1.2747	0.8914
	Foreign	0.00057	0.0003
2023	Thai	0.2663	0.1691
	Foreign	0.00039	0.0002

Observations:

Uthai Thani is not well-known among foreign visitors, primarily due to its geographical constraints. Nevertheless, the province proceeds to attract substantial number of domestic tourists, deriving from the abundance of pristine nature and beautiful temples.

In terms of domestic tourism revenue and visitors, Uthai Thani exhibits the divergence comparing to other destinations as attracting only approximately 766,000 visitors and generated 1.3378 billion baht. Apparently, this streamline sharply declined during COVID-era, reducing from 479,957 visitors to 279,911 visitors while the currency format was dropped to million-baht range (around 783.8 million in 2020 and 416.1 million in 2021). Despite the unpromising phase, the province proceeded to make recovery with revenue rising up to 1.2747 billion baht and attracted more than 891 thousand visitors, drawn by Thai government efforts to boost domestic travel by providing subsidies for hotel stays and flight for travelers, for instance Rao Tiew Duay Gun (We Travel Together) with a budget of $640 million invested and 0.0234 trillion in Thai baht, especially high-end ones from foreign markets (Steve, S., 2021)[6]. International tourism in Uthai Thani remains minimal. In 2019, the province attracted modest revenue of 16.67 million baht and purely 6261 visitors, whereas these figures dropped to zero in 2021, ultimately making deliberate growth of 277 travelers and generating only 570 thousand baht. These insights suggest that while the province holds stable domestic revenue, faces challenges in stimulating foreign tourist curiosity due to its

limited infrastructure and geographic remoteness, particularly buses depart from the Mo Chit Bus Terminal to Uthai Thani with only a few departures per day or regular ones from Nakhon Sawan[Thailandee].

Regardless of constraints in attracting foreign tourists, Uthai Thani still retains tourism potential that could be facilitated, uncover the latent appeal so as to enhance local economic gains. Notable sites such as Wat Tha Sung (also known as Crystal Temple) – providing sophisticated golden architecture and the Golden Catsle, offers photogenic overview. Nature beauty - Hua Pa Tat, known for its collapsed ceiling located in a lush forest with rare and ancient tree species, broadening opportunities for long-haul tourism, previously mentioned by Zihan et al. (2024)[2]. Additionally, Ko Thepho River Island for those wishing to enjoy the landscape of rice fields and local life, has been increasingly popular among Thai people who recommend going on bike tours for best experiences. Strategically, Uthai Thani could emerge these hidden features combining with social media marketing and engaging digital storytelling, which not only to diversify revenue streamlines but also to distribute advantages more impartially among rural communities.

4. Limitations

While the analysis provides indepth perspectives with references, limitations are spotted which hinder the interpretability and extensibility of notebook.

1. **Hypothetical Statistics:** The data interpretation for each task could be more profound if providing statistical tests on specific categories as shown in detail:

    - *Task 1:* Implementation of Pearson - Spearman correlation, ANOVA/t-test on acommodation popularity.
    - *Task 2:* Additional hypothesis test of ANOVA/t-test on comparing revenue/guests across pre- and post-COVID.
    - *Task 3:* Use of cluster analysis to identify province features (climate / seasonality, historical buildings), ANOVA on checking the myth that tourist type contributes to high / low revenue.
    - *Task 4:* Provision of external aspects, taking resemblance on a blend mixture between Explanatory Data Analysis (EDA) and society experience.
    - *Additional Requirements:* Predicate the difference between hot-pick tourist attractions and the majority, this hypothesis partially proves that pandemic closures

critically affect to the national revenue, which sets up challenges addressing the pure dependency on particular fields.

2. **Categorical Variable Distribution:** The distribution of datset contains unsymmetric points, slightly skewed curves or unusual shapes such as M-shape / Mountain shape, whereas only `no_tourist_all` and `no_tourist_thai` meet up perfect Gaussian distribution. These unusual details might result from the loss of significant outliers represented by revenue / tourists of popular destinations: Bangkok, Phuket, etc.

3. **Missing Uthai Thani Savestate Plot:** Owing to the unsucessful implementation of kaleido libraries for exporting image, this issue remains unsolvable.

4. **Technical Heaviness:** The analysis is considered to be programmatic-based with vision to Data Science, potentially resulting in lack of accessibility and less informative insights.

5. Conclusion and Future Work

This detailed storytelling helps to broaden beyond the numeric values in dataset. This analysis reveals that geographical characteristics significantly influence on total profit, accommodation price. A remarkable economic tourism recovery in post-pandemic is observed alongsides abrupt decreases in Q2-2020 and Q3-2021. Additionally, the potential advantages of less well-known province Uthai Thani is explored.

For further implementation, predictive factors will be expanded and thereby lay important groundwork for Machine Learning algorithms. Minor improvements will also be made, involving the integration of up-to-date dataset as of 2025 where feasible.

References

References

[1] Kittipong, T. and Wichit, L. (2019). An alternative hotel pricing technology: Hedonic price model for pricing beach resort revenue in thailand. *International Journal of Building, Urban, Interior and Landscape Technology (BUILT)*.

[2] Li, Z., Zhang, H., Xu, Y., Fu, T., Wang, H., and Tang, G. (2024). How does cultural distance affect tourism destination choice? evidence from the correlation between regional culture and tourism flow. *Applied Geography*, 166:103261.

[3] Linab (2024). Phuket tourism boom drives hotel prices to record levels. NANALOCAL.

[4] Manop, U. (2020). Thailand economic focus: Tourism industry plummets amid covid-19 pandemic. United Nations, Thailand.

[5] Rakklin, P. and Laphet, J. (2025). Public sector approaches to hospitality and tourism development in thailand using information databases. *Journal of Information Systems Engineering and Management*, 10(42s).

[6] Steve, S., Jan, S., and Voramon, S. (2021). Reimagining travel: Thailand tourism after the covid-19 pandemic. *McKinsey and Company*.

[Thailandee] Thailandee. All you need to know to visit uthai thani. Retrieved from Thailandee.com.

[8] Thai.news (2025). Koh samui tourism boom: The white lotus' impact on luxury hotel prices in 2025. March 3.

[9] Travel and Tourism Council (2023). Travel and tourism economic impact 2023, thailand.

[10] Weerawan, P. and Kobkul, C. (2017). Determinants of hotel and resort room rates in the west beach of thailand: Hedonic pricing approach. *RMUTT Global Business and Economics Review*, 12(2).

Appendix

Appendix A

Analysis of Thailand Tourism Economics Post-pandemic

To support the clarity of economic impact of global travel & tourism, sources are retrieved from World Travel & Tourism Council (See References).

Category	**2019**	**2022**	**2023 Forecast**
Travel & Tourism GDP (percentage share of global GDP)	10.4%	7.6%	9.2%
Change in Travel & Tourism GDP (Year on Year)	–	+22% (+$1.4 trillion GDP gain)	+23.3% (+$1.8 trillion GDP gain)
Jobs supported by Travel & Tourism	334m (1 in 10 of global employment)	295m (1 in 11 of global employment)	320m (1 in 10 of global employment)
2014–2019 Jobs	1/5 of all global net new jobs were created by Travel & Tourism	–	–
Change in Travel & Tourism Jobs (Year on Year)	–	+22m New jobs	+24m New jobs

Table 2: Economic Impact of Global Travel & Tourism (WTTC)