การพยากรณ์รายได้สำหรับอุตสาหกรรมการโรงแรมด้วยปัญญาประดิษฐ์

REVENUE FORECASTING FOR HOSPITALITY INDUSTRY USING ARTIFICIAL INTELLIGENCE

เจริญพร แสงจันทร์

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

2567

การพยากรณ์รายได้สำหรับอุตสาหกรรมการโรงแรมด้วยปัญญาประดิษฐ์

เจริญพร แสงจันทร์

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล
คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ปีการศึกษา 2567
ลิขสิทธิ์ของมหาวิทยาลัยศรีนครินทรวิโรฒ

REVENUE FORECASTING FOR HOSPITALITY INDUSTRY USING ARTIFICIAL INTELLIGENCE

CHAROENPORN SANGJAN

An Master's Project Submitted in Partial Fulfillment of the Requirements

for the Degree of MASTER OF SCIENCE

(Data Science)

Faculty of Science, Srinakharinwirot University

2024

Copyright of Srinakharinwirot University

สารนิพนธ์

เรื่อง

การพยากรณ์รายได้สำหรับอุตสาหกรรมการโรงแรมด้วยปัญญาประดิษฐ์

ของ

เจริญพร แสงจันทร์

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการข้อมูล

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....

(รองศาสตราจารย์ นายแพทย์ฉัตรชัย เอกปัญญาสกุล)

คณบดีบัณฑิตวิทยาลัย

.....

คณะกรรมการสอบปากเปล่าสารนิพนธ์

..... ที่ปรึกษาหลัก

(ผู้ช่วยศาสตราจารย์ ดร.รัตน์ชัยนันท์ ธรรมสุจริต)

..... ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.ศิฟาณี นุชิตประสิทธิ์ชัย)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.นุวีย์ วิวัฒนวัฒนา)

ชื่อเรื่อง	การพยากรณ์รายได้สำหรับอุตสาหกรรมการโรงแรมด้วยปัญญาประดิษฐ์
ผู้วิจัย	เจริญพร แสงจันทร์
ปริญญา	วิทยาศาสตรมหาบัณฑิต
ปีการศึกษา	2567
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร. รัตน์ชัยนันท์ ธรรมสุจริต

การพยากรณ์รายได้ของโรงแรมมีความสำคัญอย่างยิ่งต่อการวางแผนกลยุทธ์ด้านการบริหารรายได้ (Revenue Management) โดยเฉพาะอย่างยิ่งเมื่อต้องเผชิญกับความซับซ้อน และผันผวนของข้อมูล งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองอนุกรมเวลา ได้แก่ ARIMA, SARIMA, Prophet, ARIMAX และ SARIMAX ในการพยากรณ์จำนวนห้องพักที่ขายได้ (Room Sold) ซึ่งสัมพันธ์โดยตรงกับรายได้ของโรงแรม และเพื่อศึกษาอิทธิพลของตัวแปรสำคัญ อาทิ อัตราค่าห้องพักเฉลี่ย (ADR), ระยะเวลาการจองล่วงหน้า (Lead Time), จำนวนคืนที่เข้าพัก (Total Nights) และฤดูกาล ที่มีต่อประสิทธิภาพของแบบจำลอง Prophet ในการประเมินประสิทธิภาพของแบบจำลองจะใช้ตัวชี้วัดทางสถิติ ได้แก่ MAE, MAPE และ RMSE โดยใช้ชุดข้อมูล Hotel Booking Demand ซึ่งรวบรวมข้อมูลการจองโรงแรมในประเทศโปรตุเกส ระหว่างปี พ.ศ. 2561 ถึง2563 โดยผลการศึกษาแสดงให้เห็นว่า แบบจำลอง Prophet ที่ใช้ตัวแปรเสริม (Exogenous Variables) ให้ผลลัพธ์แม่นยำที่สุด โดยมีค่า MAPE = 2.71%, MAE = 0.2279 และ RMSE = 0.3078 รองลงมาคือ SARIMAX ที่มีค่า MAPE = 3.19% ซึ่งสามารถรวมปัจจัยฤดูกาล และพฤติกรรมลูกค้า ได้อย่างมีประสิทธิภาพ ในขณะที่ ARIMAX และ SARIMA ให้ผลลัพธ์ในระดับดี สำหรับแบบจำลอง Prophet ที่ไม่ใช้ตัวแปรเสริม และ ARIMA ซึ่งขาดโครงสร้างฤดูกาล และความสามารถในการรวมปัจจัยภายนอก พบว่ามีค่าความคลาดเคลื่อนสูงที่สุดในกลุ่ม ผลการวิเคราะห์นี้สะท้อนให้เห็นว่าแต่ละแบบจำลองมีความเหมาะสมแตกต่างกันตามลักษณะ และบริบทของข้อมูล โดย Prophet with Exogenous Variables เหมาะอย่างยิ่งสำหรับการวิเคราะห์พฤติกรรมเชิงลึก และ SARIMAX เหมาะกับการวิเคราะห์ที่ต้องการความเข้าใจเชิงฤดูกาลควบคู่กับปัจจัยภายนอก ผลการพยากรณ์ที่ได้สามารถนำไปประยุกต์ใช้ในการประมาณการรายได้ล่วงหน้า ติดตามเป้าหมาย และสนับสนุนการตัดสินใจเชิงกลยุทธ์ในธุรกิจโรงแรมได้อย่างมีประสิทธิภาพ

คำสำคัญ : การพยากรณ์รายได้ในอุตสาหกรรมโรงแรม, การวิเคราะห์อนุกรมเวลา, ปัญญาประดิษฐ์, Prophet, ARIMA, SARIMA, การจัดการรายได้

Title	REVENUE FORECASTING FOR HOSPITALITY INDUSTRY USING ARTIFICIAL INTELLIGENCE
Author	CHAROENPORN SANGJAN
Degree	MASTER OF SCIENCE
Academic Year	2024
Thesis Advisor	Assistant Professor Dr. Ratchainant Thammasudjarit

Forecasting hotel revenue plays a crucial role in strategic planning for revenue management, especially given the inherent complexity and volatility of market data. This study aims to compare the forecasting performance of five-time series models—ARIMA, SARIMA, Prophet, ARIMAX, and SARIMAX—for predicting the number of rooms sold, a key variable directly related to hotel revenue. Additionally, the study investigates the influence of key variables, including Average Daily Rate (ADR), Lead Time, Total Nights, and Seasonality, on the forecasting accuracy of the Prophet model. Model performance was evaluated using statistical metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE). The analysis utilized the Hotel Booking Demand dataset, which includes booking data from city and resort hotels in Portugal between 2018 and 2020. The results indicate that the Prophet model with exogenous variables provided the most accurate forecasts, achieving MAE = 0.2279, MAPE = 2.71%, and RMSE = 0.3078. SARIMAX ranked second with MAPE = 3.19%, demonstrating a strong capability to handle seasonality and consumer behavior patterns effectively. ARIMAX and SARIMA also yielded satisfactory results, whereas Prophet without Exogenous Variables and ARIMA—which lack structures to incorporate seasonality and external variables—produced the highest forecasting errors. These findings highlight that the appropriateness of each model depends on the nature and context of the data. Prophet with Exogenous Variables is particularly effective for analyzing complex consumer behavior, while SARIMAX is suitable for scenarios requiring detailed seasonal and external factor analysis. The forecasting results can be applied to revenue estimation, goal tracking, and strategic decision-making in the hotel industry with improved effectiveness.

Keyword : Hotel Revenue Forecasting Time Series Analysis Artificial Intelligence Prophet ARIMA SARIMA Revenue Management

กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ด้วยความอนุเคราะห์ และความช่วยเหลือจากหลายฝ่าย ข้าพเจ้าขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.รัตน์ชัยนันท์ ธรรมสุจริต อาจารย์ที่ปรึกษาสารนิพนธ์ ที่ได้มอบความช่วยเหลือ สนับสนุนทั้งในด้านวิชาการ งานวิจัย ความรู้ คำแนะนำ และเทคนิคต่างๆ ตลอดการจัดทำสารนิพนธ์ รวมถึงโอกาสอันมีค่าต่างๆ ที่มอบให้แก่ข้าพเจ้ามาโดยตลอด

ขอขอบพระคุณคณะกรรมการสอบสารนิพนธ์ทุกท่านที่ได้กรุณาให้คำแนะนำ และข้อเสนอแนะอันเป็นประโยชน์อย่างยิ่งสำหรับการปรับปรุงสารนิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น รวมทั้งขอขอบพระคุณบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ และคณาจารย์ทุกท่านในภาควิชาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่กรุณาให้ความรู้ คำแนะนำ และสนับสนุนการทำวิจัย

สุดท้ายนี้ ข้าพเจ้าขอขอบพระคุณบิดา มารดา และครอบครัว ซึ่งเป็นกำลังใจสำคัญและแรงสนับสนุนอันอบอุ่น คอยผลักดันให้ข้าพเจ้ามีพลังใจทุ่มเทให้กับการศึกษามาโดยตลอด ขอขอบคุณเพื่อน ๆ ทุกคนที่คอยให้กำลังใจ และให้คำปรึกษาเสมอมา ท้ายที่สุดข้าพเจ้าขอขอบคุณ ลี แทยง ผู้เป็นอีกหนึ่งกำลังใจสำคัญที่ช่วยเติมเต็มพลังให้ข้าพเจ้าดำเนินงานจนสารนิพนธ์ฉบับนี้สำเร็จลงได้

เจริญพร แสงจันทร์

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ญ

สารบัญรูปภาพ ..... ฎ

บทที่ 1 บทนำ 1

ความสำคัญ และความเป็นมาของงานวิจัย 1

วัตถุประสงค์ของการวิจัย 3

ขอบเขตของการวิจัย 3

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย 4

บทที่ 2 วรรณกรรม และงานวิจัยที่เกี่ยวข้อง 6

ข้อมูลเกี่ยวกับอุตสาหกรรมโรงแรม 6

ปัจจัยที่ส่งผลต่อรายได้ของโรงแรม 6

นิยามการจัดการรายได้ และวัตถุประสงค์ในการจัดการรายได้ 8

แบบจำลองการเรียนรู้ของเครื่อง 9

ARIMA (Autoregressive Integrated Moving Average) 9

ARIMAX (ARIMA with Exogenous Variables) 11

SARIMA (Seasonal Autoregressive Integrated Moving Average) 12

SARIMAX (Seasonal ARIMA with Exogenous Variables) 13

PROPHET 15

เกณฑ์ที่ใช้ในการวัดผลการพยากรณ์ 16
RMSE (Root Mean Squared Error) 16
MAE (Mean Absolute Error) 17
MAPE (Mean Absolute Percentage Error) 17
ผลงานวิจัยที่เกี่ยวข้อง 18
บทที่ 3 การดำเนินการวิจัย 28
การสร้างแบบจำลอง 28
การเก็บรวบรวมข้อมูล (Data Collection) 31
การสำรวจข้อมูล Exploratory Data Analysis (EDA) 33
การเตรียมข้อมูล (Data Preparation) 44
การคัดเลือกตัวแปร (Feature Selection) 45
การสร้าง และประเมินแบบจำลองพยากรณ์ 51
การตรวจสอบโครงสร้างของข้อมูลก่อนสร้างแบบจำลอง 53
แบบจำลอง Autoregressive Integrated Moving Average (ARIMA) 54
แบบจำลอง Seasonal Autoregressive Integrated Moving Average (SARIMA) 55
แบบจำลอง ARIMAX และแบบจำลอง SARIMAX 56
แบบจำลอง Prophet 58
บทที่ 4 ผลการดำเนินการวิจัย 60
ผลลัพธ์ของการสร้างแบบจำลอง Prophet 61
ผลของตัวแปรช่วย ต่อประสิทธิภาพของแบบจำลอง Prophet 62
ผลลัพธ์ของการสร้างแบบจำลอง ARIMA 64
ผลลัพธ์ของการสร้างแบบจำลอง SARIMA 66
ผลลัพธ์จากการสร้างแบบจำลอง ARIMAX 67
ผลลัพธ์จากการสร้างแบบจำลอง SARIMAX 69

ผลลัพธ์จากการเปรียบเทียบแบบจำลอง 70
บทที่ 5 สรุปผล อภิปรายผล และข้อเสนอแนะ 72
สรุปผลการวิจัย 72
อภิปรายผลการวิจัย 73
ข้อเสนอแนะ 74
บรรณานุกรม 76
ประวัติผู้เขียน 81

สารบัญตาราง

หน้า

ตาราง 1 โครงสร้างรายได้โรงแรมในประเทศไทย พ.ศ. 2559 2

ตาราง 2 ตารางแสดงตัวแปรในข้อมูล 32

ตาราง 3 ตารางแสดงรายละเอียดประเภทของข้อมูลแต่ละคอลัมน์ 35

ตาราง 4 ตารางแสดงอัตราการยกเลิกห้องพัก กับระยะเวลาการจองล่วงหน้า 42

ตาราง 5 ตารางแสดงจำนวนห้องพักที่มีผู้เข้าพักมากที่สุด 45

ตาราง 6 ตารางแสดงรายละเอียดข้อมูลแต่ละคอลัมน์ 46

ตาราง 7 ตารางแสดงค่าสถิติของจำนวนห้องที่ถูกขาย และรายได้ 47

ตาราง 8 ตารางแสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง 72

สารบัญรูปภาพ

หน้า

ภาพประกอบ 1 กราฟแสดงจำนวนการเดินทางระหว่างประเทศทั่วโลกตั้งแต่ปี 2005 ถึง 2023.....1

ภาพประกอบ 2 กรอบตัวแปรงานวิจัย.....4

ภาพประกอบ 3 ภาพแสดงกระบวนการสร้างแบบจำลองการพยากรณ์จำนวนห้องพักที่ขายได้.....29

ภาพประกอบ 4 ภาพแสดงค่าว่างของข้อมูล.....34

ภาพประกอบ 5 ภาพแสดงความสัมพันธ์ระหว่างตัวแปรสำคัญในการพยากรณ์รายได้ของโรงแรม 36

ภาพประกอบ 6 ภาพแสดงประเภทของโรงแรม และจำนวนแขกในแต่ละประเทศที่เข้าพัก 37

ภาพประกอบ 7 ภาพแสดงประเทศที่เข้าใช้บริการในโรงแรม.....38

ภาพประกอบ 8 ภาพแสดงแนวโน้มของอัตราค่าที่พักเฉลี่ยรายวัน (ADR) ในโรงแรมแต่ละประเภท 38

ภาพประกอบ 9 ภาพแสดงแนวโน้มของอัตราค่าที่พักเฉลี่ยรายวัน (ADR) ของโรงแรมในแต่ละเดือน 39

ภาพประกอบ 10 กราฟเปรียบเทียบค่า ADR ของ โรงแรมในเมือง และโรงแรมรีสอร์ท.....40

ภาพประกอบ 11 กราฟเปรียบเทียบค่าการกระจายตัวของค่าห้องพักเฉลี่ยต่อวัน.....40

ภาพประกอบ 12 กราฟแสดงประเทศที่เข้าใช้บริการโรงแรมสูงสุด 10 อันดับแรก 41

ภาพประกอบ 13 กราฟแสดงประเทศที่ดำเนินการยกเลิกเข้าพักสูงสุด 10 อันดับแรก.....41

ภาพประกอบ 14 กราฟแสดงอัตราการยกเลิกห้องพัก กับระยะเวลาการจองล่วงหน้า.....42

ภาพประกอบ 15 กราฟแสดงจำนวนผู้เข้าพัก และจำนวนคืนที่เข้าพักที่มีการยกเลิกการจอง.....43

ภาพประกอบ 16 กราฟแสดงอัตราการยกเลิกการจอง ตาม Market Segment 43

ภาพประกอบ 17 ภาพแสดงตัวอย่างจำนวนของ room sold และรายได้ของโรงแรม 46

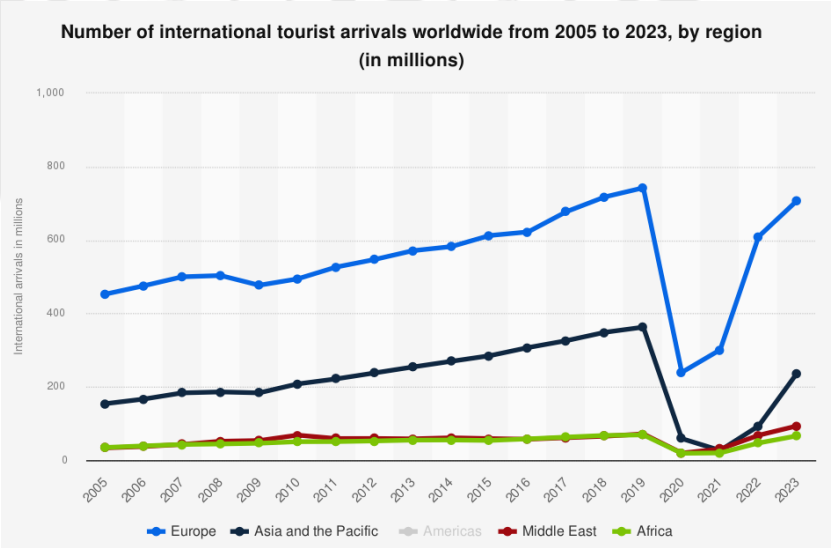
ภาพประกอบ 18 กราฟแสดงความสัมพันธ์ระหว่างตัวแปรสำคัญในการพยากรณ์รายได้ของโรงแรม.....48

ภาพประกอบ 19 แสดงการกระจายตัวของอัตราค่าห้องพักเฉลี่ยสำหรับห้องพักประเภท A 49
ภาพประกอบ 20 แสดงแนวโน้มรายได้ต่อเดือนสำหรับห้องพักประเภท A.....49
ภาพประกอบ 21 แสดงรายได้รวมจำแนกตามประเภทโรงแรม ได้แก่ City Hotel และ Resort Hotel 50
ภาพประกอบ 22 ภาพแสดงแนวโน้มของข้อมูลการจองห้องพักของห้องพักประเภท A.....50
ภาพประกอบ 23 กราฟแสดงแนวโน้มของอัตราค่าห้องพักเฉลี่ยรายวัน (ADR) และ Room Sold ...51
ภาพประกอบ 24 แสดงค่า Trend Seasonal และ Residual ของ Train Data.....53
ภาพประกอบ 25 ผลการวิเคราะห์ ACF และ PACF สำหรับ Train data 55
ภาพประกอบ 26 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง Prophet 62
ภาพประกอบ 27 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของ Prophet โดยมีตัวแปรช่วย 63
ภาพประกอบ 28 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง ARIMA.....65
ภาพประกอบ 29 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง SARIMA ..67
ภาพประกอบ 30 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง ARIMAX..68
ภาพประกอบ 31 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง SARIMAX70

บทที่ 1
บทนำ

ความสำคัญ และความเป็นมาของงานวิจัย

ในช่วงหลายปีที่ผ่านมาอุตสาหกรรมการท่องเที่ยว และการให้บริการ นับว่าเป็นหนึ่งอุตสาหกรรมที่มีการเติบโตอย่างต่อเนื่อง จากการขยายตัวของการท่องเที่ยวทั่วโลกที่เพิ่มมากขึ้น สังเกตได้จากจำนวนนักท่องเที่ยวที่เพิ่มขึ้นในแต่ละปี โดยอ้างอิงจากข้อมูลขององค์การการท่องเที่ยวโลก (World Tourism Organization: UN Tourism) ดังแสดงในภาพที่ 1พบว่าจำนวนนักท่องเที่ยวในแต่ละภูมิภาค โดยเฉพาะในยุโรปมีการเติบโตอย่างต่อเนื่องจนถึงปี 2019 ซึ่งเป็นปีที่มีการเดินทางสูงสุดก่อนที่จะลดลงอย่างมากในช่วงปี 2020 เนื่องจากผลกระทบจากการระบาดของโควิด-19 อย่างไรก็ตาม ในปี 2023 ผลการสำรวจชี้ให้เห็นถึงสัญญาณการฟื้นตัวอย่างชัดเจนในหลายภูมิภาคโดยเฉพาะในยุโรป และเอเชียแปซิฟิก ที่กลับมาใกล้เคียงกับระดับก่อนเกิดการระบาด รวมถึงมีแนวโน้มที่จะเพิ่มขึ้นอีก



ภาพประกอบ 1 กราฟแสดงจำนวนการเดินทางระหว่างประเทศทั่วโลกตั้งแต่ปี 2005 ถึง 2023 ที่มา Statista (1)

จากแนวโน้มการเติบโตของจำนวนนักท่องเที่ยวที่สูงขึ้น อุตสาหกรรมโรงแรมจึงมีบทบาทสำคัญในการรองรับนักท่องเที่ยวที่เพิ่มมากขึ้นไปด้วย อย่างไรก็ตามผู้ประกอบการในอุตสาหกรรมโรงแรมเผชิญกับความท้าทายในการบริหารจัดการรายได้ และการวางแผนทรัพยากรในระยะยาว เนื่องจากการแข่งขันทางธุรกิจที่สูงขึ้น รวมถึงการเปลี่ยนแปลงในพฤติกรรมของลูกค้า Contessi, Viverit (2) ด้วยเหตุนี้ความสามารถในการพยากรณ์ข้อมูลจึงมีความสำคัญต่อผู้ประกอบการ

ในอุตสาหกรรมโรงแรม เช่นการพยากรณ์จำนวนห้องพักที่ขายได้ หรือการพยากรณ์รายได้ ข้อมูลเหล่านี้จะช่วยให้ผู้ประกอบการสามารถประเมินรายได้ล่วงหน้าได้ รวมถึงสามารถวางแผนจัดการทรัพยากรต่างๆ ได้อย่างมีประสิทธิภาพ เช่น การจัดสรรพนักงาน การบริหารจัดการห้องพัก และการปรับกลยุทธ์ทางการตลาดให้เหมาะสม โดยข้อมูลที่แม่นยำสามารถนำไปใช้ในการวางแผน และจัดการเพื่อเพิ่มประสิทธิภาพในการดำเนินงาน Wu, Song (3)

การนำเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) มาประยุกต์ใช้เพื่อค้นหาความสัมพันธ์ ระหว่างตัวแปรที่มีผลกระทบต่อรายได้ของโรงแรม เช่น อัตราค่าห้องพักเฉลี่ย (Average Daily Rate: ADR) จำนวนห้องพักที่ถูกจอง และความต้องการพิเศษของลูกค้า ถือว่าเป็นสิ่งสำคัญ เนื่องจากตัวแปรเหล่านี้ถือเป็นรายได้หลักของโรงแรม จากการสำรวจของสำนักงานสถิติแห่งชาติในปี พ.ศ. 2559 รายงานว่ารายได้จากห้องพัก และบริการในห้องอาหารเป็นส่วนสำคัญที่สร้างรายได้ให้แก่โรงแรมในประเทศไทยกว่า 12,654 แห่ง รายละเอียดแสดงดังตารางที่ 1

ตาราง 1 โครงสร้างรายได้โรงแรมในประเทศไทย พ.ศ. 2559

ภาค	ประเภทรายได้				รวม	หน่วย
	ห้องพัก	อาหาร	ห้องประชุมสัมนา	อื่น ๆ		
กรุงเทพ ฯ	410,645.52	104,211.81	44,522.12	54,854.32	614,233.77	ล้านบาท
	66.85	16.97	7.25	8.93	100.00	ร้อยละ
ภาคกลาง	263,827.61	33,510.14	11,355.08	19,935.21	328,628.05	ล้านบาท
	80.28	10.20	3.46	6.07	100.00	ร้อยละ
ภาคเหนือ	77,627.05	19,274.06	6,758.35	8,135.38	111,794.84	ล้านบาท
	69.44	17.24	6.05	7.28	100.00	ร้อยละ
ภาคอีสาน	45,963.56	10,338.57	4,841.00	1,442.07	62,585.19	ล้านบาท
	73.44	16.52	7.74	2.30	100.00	ร้อยละ
ภาคใต้	484,820.05	105,409.11	10,180.24	9,462.15	609,871.56	ล้านบาท
	76.50	17.28	1.67	1.55	100.00	ร้อยละ
ทั่วประเทศ	1,282,883.80	272,743.70	77,656.79	93,829.13	1,727,113.41	ล้านบาท
	74.28	15.79	4.50	5.43	100.00	ร้อยบาท

หมายเหตุ.สรุปจากการสำรวจกิจการโรงแรมและเกสต์เฮาส์, สำนักงานสถิติแห่งชาติ (4), 2559, กรุงเทพฯ

จากข้อมูลข้างต้นพบว่ารายได้จากห้องพัก เป็นส่วนสำคัญที่สร้างรายได้ให้แก่โรงแรมโดยคิดเป็นร้อยละ 74.25 จากรายได้ทั้งหมด ผู้วิจัยจึงได้เลือกนำเทคนิคการวิเคราะห์อนุกรมเวลาสมัยใหม่ เช่น แบบจำลอง Prophet มาประยุกต์ใช้เพื่อพยากรณ์รายได้ของโรงแรมจากจำนวนห้องพักที่ขายได้

เนื่องจากเป็นเครื่องมือที่สามารถจัดการกับข้อมูลที่มีความซับซ้อน มีแนวโน้มเปลี่ยนแปลงในระยะยาว และรองรับลักษณะตามฤดูกาลได้อย่างมีประสิทธิภาพ การประยุกต์ใช้ Prophet ในการพยากรณ์รายได้ของโรงแรมในงานวิจัยนี้ จะพิจารณาใช้ตัวแปรสำคัญ เช่น อัตราค่าห้องพักเฉลี่ย (ADR) จำนวนคืนที่เข้าพัก (total nights) และการจองล่วงหน้า (lead time) มาประกอบการวิเคราะห์

เพื่อประเมินศักยภาพของแบบจำลองการพยากรณ์รายได้โรงแรม งานวิจัยนี้จึงเลือกนำเทคนิคการวิเคราะห์อนุกรมเวลาสมัยใหม่อย่างแบบจำลอง Prophet มาเปรียบเทียบกับแบบจำลองทางสถิติแบบดั้งเดิมที่ได้รับความนิยมในการพยากรณ์อนุกรมเวลา ได้แก่ ARIMA และ SARIMA นอกจากนี้เพื่อให้การเปรียบเทียบมีความครอบคลุม และสะท้อนศักยภาพของแบบจำลองได้อย่างแท้จริง จึงนำ ARIMAX และSARIMAX ซึ่งเป็นแบบจำลองที่สามารถนำตัวแปรเสริม ที่มีอิทธิพลต่อรายได้ของโรงแรมเข้ามาช่วยในการพยากรณ์มาใช้ในการศึกษาด้วย ผลการเปรียบเทียบนี้จะทำให้สามารถวิเคราะห์ข้อดี และข้อจำกัดของแต่ละแบบจำลองทั้งในกรณีที่ใช้ หรือไม่ใช้ตัวแปรเสริม โดยผลลัพธ์การประเมิน และการเปรียบเทียบนี้คาดว่าจะช่วยให้ได้ข้อมูลสำคัญต่อผู้ประกอบการสำหรับการวางแผนทรัพยากร การตัดสินใจเชิงกลยุทธ์ และการปรับตัวให้ทันกับการเปลี่ยนแปลงของตลาดในอุตสาหกรรมการโรงแรมได้อย่างมีประสิทธิภาพ

วัตถุประสงค์ของการวิจัย

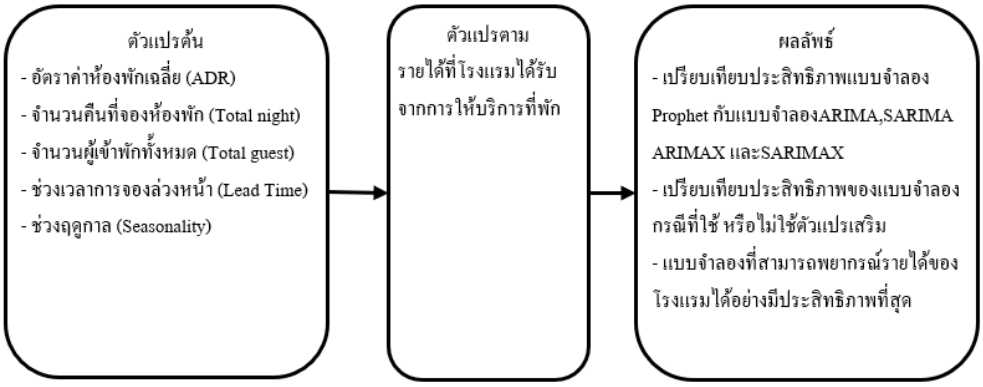
1. เพื่อทดสอบประสิทธิภาพของการใช้แบบจำลองพยากรณ์อนุกรมเวลาสมัยใหม่อย่าง Prophet เปรียบเทียบกับแบบจำลองทางสถิติที่นิยมใช้ เช่น ARIMA และ SARIMA รวมถึงเปรียบเทียบประสิทธิภาพกับแบบจำลองที่ใช้ตัวแปรเสริม (Exogenous Variables) อย่าง ARIMAX และ SARIMAX

2. เพื่อศึกษาผลกระทบของตัวแปรสำคัญ เช่น อัตราค่าห้องพักเฉลี่ย (ADR), ช่วงเวลาการจองล่วงหน้า (Lead Time), จำนวนคืนที่เข้าพัก (Total Nights) และฤดูกาล ต่อประสิทธิภาพของแบบจำลอง prophet ในการพยากรณ์รายได้ของโรงแรม

ขอบเขตของการวิจัย

การวิจัยนี้จะใช้ข้อมูลจากชุดข้อมูลความต้องการในการจองห้องพักในโรงแรม (Hotel Booking Demand Dataset) ซึ่งเผยแพร่โดย Antonio et al. (2019) (5) ข้อมูลดังกล่าวครอบคลุมการจองห้องพักของโรงแรมในประเทศโปรตุเกสสองประเภท ได้แก่ Resort Hotel และ City Hotel ประกอบด้วยข้อมูลในช่วงระหว่างปี 2018 ถึง 2020 ทั้งนี้ข้อมูลยังประกอบด้วยตัวแปรสำคัญที่ส่งผลต่อรายได้ของโรงแรม เช่น อัตราค่าห้องพักเฉลี่ย (ADR) จำนวนคืนที่เข้าพัก ช่วงเวลาการจองล่วงหน้า

(Lead Time) ประเภทของห้องพัก และส่วนลดต่าง ๆ ในการศึกษานี้จะมุ่งเน้นไปที่การพยากรณ์จำนวนห้องพักที่ขายได้ (Room Sold) แทนที่การพยากรณ์รายได้โดยตรง เนื่องจากค่า ADR มีความแปรปรวนสูงจากอิทธิพลของฤดูกาลท่องเที่ยว โปรโมชั่น และพฤติกรรมความต้องการของลูกค้า การพยากรณ์จำนวนห้องพักที่ขายได้ จึงเป็นวิธีการที่ช่วยควบคุมตัวแปรที่ผันผวนในเบื้องต้น เพื่อให้ได้ผลลัพธ์ที่มีเสถียรภาพ สามารถนำไปประยุกต์ใช้ในการบริหารจัดการทรัพยากรของโรงแรมได้อย่างมีประสิทธิภาพ ข้อมูลทั้งหมดจะถูกนำมาผ่านกระบวนการเตรียมข้อมูล เพื่อความเหมาะสมในการพยากรณ์ เช่น การจัดการกับค่าว่าง การเลือกใช้ตัวแปรที่มีความสำคัญ และการสร้างคุณลักษณะเพิ่มเติม เพื่อนำมาใช้ในการพัฒนาแบบจำลองการพยากรณ์รายได้ของโรงแรม โดยใช้เทคนิคการวิเคราะห์อนุกรมเวลา ได้แก่ Prophet นำมาเปรียบเทียบประสิทธิภาพกับแบบจำลองทางสถิติที่นิยมใช้ในการพยากรณ์ ได้แก่ ARIMA,SARIMA, ARIMAX และSARIMAX เพื่อประเมินความแม่นยำในการพยากรณ์รายได้ของโรงแรมแสดงดังภาพประกอบ 2



ภาพประกอบ 2 กรอบตัวแปรงานวิจัย

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. สามารถประเมินประสิทธิภาพของแบบจำลองพยากรณ์อนุกรมเวลาสมัยใหม่อย่าง Prophet โดยเปรียบเทียบกับแบบจำลองได้แก่ ARIMA, SARIMA, ARIMAX และ SARIMAXเพื่อระบุวิธีการพยากรณ์รายได้ที่มีความแม่นยำ และเหมาะสมที่สุดสำหรับอุตสาหกรรมโรงแรม

2. ให้ข้อมูลที่สำคัญเกี่ยวกับปัจจัยที่ส่งผลต่อรายได้ของโรงแรม ซึ่งจะเป็นประโยชน์ในการวางแผนกลยุทธ์ด้านการบริหารจัดการโรงแรม

3. ช่วยสนับสนุนการวางแผนกลยุทธ์ทางการตลาด เช่น การจัดโปรโมชั่น การกำหนดราคาห้องพักที่เหมาะสม และการส่งเสริมการขายเพื่อเพิ่มรายได้

4. เป็นแนวทางในการประยุกต์ใช้เทคโนโลยีปัญญาประดิษฐ์ในการพยากรณ์รายได้ของโรงแรมจากการพยากรณ์จำนวนห้องพักที่ขายได้ เพื่อสร้างความได้เปรียบในการแข่งขันและเพิ่มประสิทธิภาพในการดำเนินงาน

บทที่ 2
วรรณกรรม และงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยครั้งนี้ผู้วิจัยได้ศึกษาทฤษฎี และงานวิจัยที่เกี่ยวข้องกับการพัฒนาแบบจำลองพยากรณ์ข้อมูลอนุกรมเวลา โดยจะศึกษาแบบจำลองพยากรณ์ข้อมูลอนุกรมเวลาชื่อว่า Prophet ที่ออกแบบมาสำหรับการคาดการณ์ ข้อมูลที่มีแนวโน้ม และโครงสร้างตามฤดูกาล โดยจะนำเสนอตามหัวข้อดังต่อไปนี้

1. ข้อมูลเกี่ยวกับอุตสาหกรรมโรงแรม
2. วิธีการ และทฤษฎีที่เกี่ยวข้อง
3. งานวิจัยที่เกี่ยวข้อง

ข้อมูลเกี่ยวกับอุตสาหกรรมโรงแรม

โรงแรม หมายถึงโรงแรม และรีสอร์ท ตามความหมายที่พระราชบัญญัติโรงแรม (2547) (6)ได้ให้ความหมายไว้ว่า โรงแรมเป็นสถานที่พักชั่วคราวที่สร้างขึ้นเพื่อให้บริการที่พักแก่บุคคลทั่วไปโดยมีค่าตอบแทน ซึ่งมีการคิดค่าบริการเป็นรายวัน

กระทรวงท่องเที่ยว และกีฬา (2552) (7) ได้ให้ความหมายโรงแรมในเมือง และรีสอร์ทดังนี้

โรงแรม หมายถึง สถานที่พักที่สร้างขึ้นโดยมีวัตถุประสงค์ในการให้บริการที่พักชั่วคราวแก่ผู้เดินทาง หรือบุคคลทั่วไปโดยได้รับค่าตอบแทน ส่วนมากจะตั้งอยู่ในเขตพื้นที่เมือง เน้นสิ่งอำนวยความสะดวกด้านที่พัก เพื่อการใช้เวลาในเชิงธุรกิจ และการพักผ่อน

รีสอร์ท หมายถึง สถานที่พักที่มุ่งเน้นการพักผ่อนท่ามกลางธรรมชาติ มักตั้งอยู่ในพื้นที่ที่มีภูมิทัศน์สวยงาม โดยมีวัตถุประสงค์หลักในการให้บริการที่พักสำหรับผู้ที่ต้องการพักผ่อนชั่วคราว รีสอร์ทมักเน้นบริการที่พักที่ให้ความสะดวกสบายและเน้นการผ่อนคลาย

Budgets หมายถึง การประมาณการณ์รายได้ที่คาดว่าจะได้รับในช่วงระยะเวลาที่กำหนด

TA (Travel Agent) หมายถึง การจองผ่านตัวแทนการท่องเที่ยว

TO (Tour Operator) หมายถึง การจองผ่านผู้จัดทัวร์

ปัจจัยที่ส่งผลต่อรายได้ของโรงแรม

รายได้ของโรงแรมสามารถสะท้อนผ่านตัวชี้วัดต่างๆ ที่แสดงให้เห็นถึงประสิทธิภาพในการสร้างรายได้ของแต่ละโรงแรม โดย Ivanov (2014) (8)ได้ระบุถึงปัจจัยที่มีผลต่อการเปลี่ยนแปลงของ

รายได้ที่สำคัญหลายประการ ได้แก่ อัตราการเข้าพัก (Occupancy rate), อัตราค่าห้องพักเฉลี่ยรายวัน (Average Daily Rate: ADR), ระยะเวลาในการเข้าพัก (Room night) นอกจากนี้ยังมีปัจจัยอื่น ๆ จากการศึกษาของ Hesford และคณะ(2019) (9) ที่ระบุว่าจำนวนห้องพักที่ถูกจอง (Room sold) รายได้ต่อห้องพักที่ให้บริการ (Revenue Per Available Room: RevPAR) ช่องทางการจองห้องพักของลูกค้า (Market segment) ซึ่งแต่ละปัจจัยมีผลต่อการสร้างรายได้ในมุมมองที่แตกต่างกันออกไปดังนี้

ราคาห้องพักเฉลี่ยต่อคืน (Average Daily Rate: ADR)

เป็นค่าที่ใช้ในอุตสาหกรรมโรงแรม เพื่อช่วยวัดประสิทธิภาพโดยรวมของธุรกิจโรงแรม คำว่า *ADR* หมายถึงอัตราการให้บริการรายวันเฉลี่ยซึ่งถูกกำหนดโดยรายได้เฉลี่ยต่อห้องที่มีผู้เข้าพักในระยะเวลาที่กำหนด *Elphick (*10)

สามารถคำนวณได้จากสมการ

$$\text{ราคาห้องพักเฉลี่ยต่อคืน}(ADR) = \frac{\text{รายได้ห้องพัก } (Room\ Revenue)}{\text{จำนวนห้องพักที่ขายได้ } (Number\ of\ Room\ Sold)} \quad (1)$$

รายได้ต่อห้องพักที่ให้บริการ (Revenue Per Available Room: RevPAR)

เป็นค่าที่ใช้ในอุตสาหกรรมการโรงแรม โดยจะให้ข้อมูลเกี่ยวกับจำนวนห้องที่ถูกขายในโรงแรม และรายได้ที่ได้รับจากการจอง Elphick (10) สามารถคำนวนได้จากสมการ

$$\text{รายได้ต่อห้องพักที่ให้บริการ } (RevPAR) = \frac{\text{รายได้รวมจากห้องทั้งหมด } (SUM\ Room\ Revenue)}{\text{จำนวนห้องที่ให้บริการ}(SUM\ Number\ of\ Room\ Sold)} \quad (2)$$

หรือสามารถใช้สมการ

$$\text{รายได้ต่อห้องพักที่ให้บริการ } (RevPAR) = \text{ ราคาห้องพักเฉลี่ยต่อคืน}(ADR) \times Occupancy\ Rate \quad (3)$$

อัตราการเข้าพัก (Occupancy Rate: OCC)

เป็นค่าที่ใช้ในอุตสาหกรรมการโรงแรม โดยแสดงถึงเปอร์เซ็นต์ของห้องพักทั้งหมดที่มีการขายในช่วงเวลาที่กำหนด โดยอัตราการเข้าพักที่สูง จะแสดงให้เห็นว่าโรงแรมมีประสิทธิภาพในการใช้ทรัพยากร และโรงแรมสามารถใช้ห้องพักได้อย่างมีประสิทธิภาพ สามารถคำนวนได้จากสมการ

$$\text{อัตราการเข้าพัก } (\text{OCC}) = \frac{\text{จำนวนห้องที่ขายได้ } (Room\ Sold)}{\text{จำนวนห้องพักที่โรงแรมมีทั้งหมด } (\text{Total Available Rooms})} \times 100 \quad (4)$$

Room Sold *เป็น*จำนวนห้องพักที่ถูกจอง หรือขายได้ โดยพิจารณาจากข้อมูลที่ไม่มีการยกเลิก

Room Night เป็นตัวชี้วัดสำคัญในการวิเคราะห์ประสิทธิภาพของโรงแรม เนื่องจากช่วยให้โรงแรมสามารถติดตาม และประเมินความต้องการของตลาด รวมถึงการจัดการรายได้สามารถคำนวนได้จากสมการ

$$\text{Room Nights} = \text{Room Sold} \times \text{Number of Nights per Room} \quad (5)$$

โดยสรุปแล้วปัจจัยต่างๆ ที่กล่าวมานี้ล้วนแต่เป็นปัจจัยที่ส่งผลสำคัญต่อรายได้ของโรงแรม โดยผู้บริหาร และนักจัดการรายได้ของโรงแรมจำเป็นต้องให้ความสำคัญกับการวิเคราะห์ และบริหารจัดการปัจจัยเหล่านี้ เพื่อให้เกิดประโยชน์สูงสุดต่อผลประกอบการทางการเงินขององค์กร

นิยามการจัดการรายได้ และวัตถุประสงค์ในการจัดการรายได้

การจัดการรายได้ (Revenue Management) หมายถึง กระบวนการบริหารจัดการที่นำเอาระบบสารสนเทศและกลยุทธ์ด้านราคามาประยุกต์ใช้ในการจัดสรร และจำหน่ายห้องพักของโรงแรมให้สอดคล้องกับกลุ่มเป้าหมายที่เหมาะสม ภายใต้ช่วงเวลาและราคาที่เหมาะสมที่สุด ทั้งนี้ เพื่อมุ่งสู่เป้าหมายในการเพิ่มประสิทธิภาพการสร้างรายได้สุทธิ (Net Revenues) และผลกำไรจากการดำเนินงานรวม (Gross Operating Profit: GOP) โดยการเสนอผลิตภัณฑ์ที่เหมาะสมแก่ลูกค้าที่ถูกต้อง ผ่านช่องทางการจัดจำหน่ายที่เหมาะสม ในเวลาที่เหมาะสม พร้อมการกำหนดราคาและการสื่อสารที่สอดคล้องกับความคาดหวังของลูกค้า และสภาวะตลาด Ivanov and Zhechev (11), (12) แนวคิดดังกล่าวได้ถูกปรับใช้ในธุรกิจโรงแรม และธุรกิจบริการอื่น ๆ โดย Denizci Guillet and Mohammed (13) ได้กล่าวไว้ว่าตั้งแต่ช่วงกลางทศวรรษ 1980 เป็นต้นมา มีการนิยามใหม่ของคำว่า yield จากเดิมที่เน้นเฉพาะรายได้ต่อหน่วยที่มีจำกัด (inventory) ให้ครอบคลุม และสะท้อนถึงมิติทางยุทธศาสตร์มากขึ้น โดยนิยามใหม่ของ yield จึงหมายถึง รายได้ต่อหน่วยสินค้าคงคลังที่พร้อมจำหน่าย (Revenue per Available Inventory) เพื่อให้สอดคล้องกับขอบเขตของการบริหารจัดการรายได้ในเชิงกลยุทธ์ที่กว้างขึ้น

วัตถุประสงค์หลักของการจัดการรายได้ คือ การเพิ่มศักยภาพในการสร้างรายได้สุทธิและผลกำไรจากการดำเนินงานรวมให้อยู่ในระดับสูงสุดภายใต้ทรัพยากรที่มีอยู่อย่างจำกัด โรงแรมจึงจำเป็นต้องบริหารความสมดุลระหว่างอัตราการเข้าพัก (Occupancy Rate) และระดับราคาห้องพัก (Average Daily Rate: ADR) ให้เหมาะสมกับสภาพการณ์ของตลาด เพื่อให้ได้รับผลตอบแทนจากการจำหน่ายห้องพักและบริการต่างๆ ของโรงแรมอย่างเต็มประสิทธิภาพ นอกจากนี้ โรงแรมยังจำเป็นต้องคำนึงถึงการสร้างความพึงพอใจและประสบการณ์ที่ดีให้แก่ลูกค้า ควบคู่ไปกับการสร้างรายได้และผลกำไรที่ยั่งยืนแก่ธุรกิจในระยะยาว

แบบจำลองการเรียนรู้ของเครื่อง

การพยากรณ์รายได้มีบทบาทสำคัญในการช่วยให้โรงแรมสามารถวางแผนกลยุทธ์การบริหารจัดการได้อย่างมีประสิทธิภาพ โดยใช้แบบจำลองการพยากรณ์ที่ทำนายจำนวนห้องที่ขายได้ (Room Sold) ซึ่งสามารถนำมาคำนวณรายได้ของโรงแรมล่วงหน้า การคาดการณ์รายได้จึงเป็นประโยชน์อย่างมากในสภาพแวดล้อมที่มีการแข่งขันสูงของอุตสาหกรรมการโรงแรม เพราะหากพบว่ารายได้ที่คาดการณ์ไว้มีแนวโน้มไม่ถึงเป้าหมายที่ตั้งไว้ ผู้ประกอบการสามารถปรับกลยุทธ์ทางการตลาด หรือโปรโมชั่นได้ทันท่วงที เพื่อเพิ่มโอกาสในการบรรลุเป้าหมายรายได้ที่ตั้งไว้ แบบจำลองที่ใช้ในการพยากรณ์มีดังนี้

ARIMA (Autoregressive Integrated Moving Average)

เป็นแบบจำลองพยากรณ์อนุกรมเวลาที่เหมาะสำหรับการพยากรณ์ข้อมูลที่ไม่มีฤดูกาล และมีลักษณะของข้อมูลที่คงที่ โดย ARIMA ถูกออกแบบมาเพื่อตรวจจับ และพยากรณ์รูปแบบในข้อมูลที่แสดงแนวโน้ม (Trend) หรือการเปลี่ยนแปลงในระดับที่ไม่สม่ำเสมอ มุ่งเน้นการเปลี่ยนข้อมูลให้คงที่ ผ่านกระบวนการ Differencing และวิเคราะห์ความสัมพันธ์เชิงเวลาเพื่อคาดการณ์ข้อมูลในอนาคต Stellwagen and Tashman (14), (15) จุดเด่นของ ARIMA คือความไม่ซับซ้อนใช้งานง่าย และมีประสิทธิภาพสูงกับข้อมูลที่ไม่มีฤดูกาล แต่มีข้อจำกัด คือไม่สามารถจัดการข้อมูลที่มีฤดูกาล หรือข้อมูลที่มีการเปลี่ยนแปลงอย่างรวดเร็ว และซับซ้อนได้ดีนักกระบวนการของ ARIMA สามารถแสดงด้วยสมการต่อไปนี้

$$\phi_p(L)(1-L^m)^d X_t = \theta_q(L)\epsilon_t \qquad (6)$$

โดยที่

X_t ค่าของข้อมูลในช่วงเวลา *t*

ϵ_t ข้อผิดพลาดแบบสุ่ม (*White Noise)*

$\phi_p(L)$ ตัวดำเนินการ *Autoregressive (AR)* ของลำดับ *p*

$\theta_q(L)$ ตัวดำเนินการ *Moving Average (MA)* ของลำดับ *q*

$(1-L^m)^d$ ตัวดำเนินการ *Differencing* เพื่อแปลงข้อมูลให้เป็น *Stationary*

L ตัวดำเนินการเลื่อนย้อนหลัง (Backward Shift Operator)

องค์ประกอบสำคัญของ ARIMA

Autoregressive (AR) ใช้ค่าก่อนหน้า (p) ของข้อมูลในการคาดการณ์ค่าในอนาคต

Differencing (I) ใช้ความแตกต่าง (d) เพื่อแปลงข้อมูลที่ไม่คงที่ให้เป็น Stationary

Moving Average (MA) ใช้ค่าข้อผิดพลาดในอดีต (q) เพื่อแก้ไขการพยากรณ์

การกำหนดพารามิเตอร์

p คือลำดับของตัวแปร Autoregressive (AR) ที่กำหนดโดยวิธี Partial Autocorrelation Function (PACF)

d คือจำนวนครั้งของการ Differencing เพื่อทำให้ข้อมูลเป็น Stationary

q คือลำดับของตัวแปร Moving Average (MA) ที่กำหนดโดยวิธี Autocorrelation Function (ACF)

การพยากรณ์ด้วย ARIMA

หลังจากพิจารณาค่าพารามิเตอร์ p,d,q และสร้างแบบจำลอง ARIMA กระบวนการพยากรณ์จะใช้ข้อมูลที่ปรับให้เป็น Stationary ผ่านการ Differencing ($Z_t = (1-L)^d X_t$) และจากนั้นจะใช้สมการ ARMA เพื่อคาดการณ์ค่า Z_t ค่าที่พยากรณ์ $\hat{X}_T + h$ แสดงดังสมการ

$$\hat{X}_T + h = \hat{Z}_T + h - \sum_{j=1}^{d} \binom{d}{j} (-1)^j \hat{X}_T + h - j \qquad (7)$$

$\hat{X}_T + h$ ค่าที่พยากรณ์ของข้อมูลเดิมที่เวลา T + h

$\hat{Z}_T + h$ ค่าที่พยากรณ์ของข้อมูลหลังจากทำ differencing ที่เวลา T + h

d ลำดับของการ differencing ใน ARIMA (ใน ARIMA (p, d, q))

$\sum_{j=1}^{d} \binom{d}{j}$ ผลรวมของพจน์ย้อนกลับจาก differencing เพื่อคืนค่ากลับไปดังเดิม

$\hat{X}_T + h - j$ ค่าพยากรณ์ที่จุดเวลาก่อนหน้า T + h – j

จากสมการ (7) หลังจากที่แบบจำลองได้ทำการพยากรณ์ข้อมูลในรูปของข้อมูลที่ถูกแปลงให้อยู่ในรูปแบบคงที่ (Stationary) ด้วยการทำ Differencingแล้ว ค่าที่ได้จากการพยากรณ์ในรูปแบบนี้ยังไม่ใช่ค่าจริงที่สามารถนำไปใช้งานได้โดยตรง ขั้นตอนที่สำคัญถัดมาจึงเป็นการทำ Inverse Differencing โดยการย้อนกลับนี้จะต้องอาศัยการคำนวณจากค่าที่พยากรณ์ไว้ในอดีตตามลำดับของการทำ Differencing เพื่อเปลี่ยนผลลัพธ์ที่ได้จากการพยากรณ์ของแบบจำลองให้กลับมาอยู่ในรูปแบบของข้อมูลเดิม ซึ่งมีหน่วยเป็นรายได้จริง หรือจำนวนห้องพักที่ขายได้ เพื่อให้สามารถนำผลลัพธ์ที่ได้ไปใช้ประโยชน์ในการวางแผนและตัดสินใจในเชิงกลยุทธ์ได้อย่างถูกต้อง

ในงานวิจัยนี้ แบบจำลอง ARIMA ถูกนำมาใช้ในการพยากรณ์จำนวนห้องพักที่ขายได้ ในช่วงเวลาที่ไม่มีลักษณะฤดูกาลอย่างชัดเจน โดยมุ่งเน้นการวิเคราะห์แนวโน้มในระยะยาวของรายได้ เพื่อสนับสนุนการวางแผนกลยุทธ์ด้านการตลาด และการบริหารรายได้ของผู้ประกอบการโรงแรมอย่างมีประสิทธิภาพ โดยเฉพาะในช่วงเวลาที่พฤติกรรมการเข้าพักของลูกค้าไม่มีรูปแบบซ้ำในลักษณะฤดูกาล

ARIMAX (ARIMA with Exogenous Variables)

ARIMAX เป็นแบบจำลองที่พัฒนาต่อยอดจากแบบจำลอง ARIMA โดยเพิ่มตัวแปรอิสระภายนอก (exogenous variables) เข้าไปในกระบวนการพยากรณ์ เพื่ออธิบายความผันแปรของอนุกรมเวลาที่อาจได้รับอิทธิพลจากปัจจัยภายนอก ซึ่งมีผลต่อค่าของอนุกรมเวลา ณ ช่วงเวลานั้น ๆ Banaś and Utnik-Banaś (16) ด้วยเหตุนี้ แบบจำลอง ARIMAX จึงมีความยืดหยุ่น และแม่นยำมากยิ่งขึ้นในการพยากรณ์ข้อมูลอนุกรมเวลาที่ได้รับอิทธิพลจากปัจจัยภายนอก เช่นอุตสาหกรรมการโรงแรม และการท่องเที่ยว จำนวนห้องพักที่ถูกจองในแต่ละเดือนอาจไม่ได้ขึ้นอยู่กับข้อมูลยอดจองในอดีตเพียงอย่างเดียว แต่ยังได้รับอิทธิพลจากปัจจัยภายนอก เช่น ช่วงวันหยุดยาว เทศกาลพิเศษ แบบจำลอง ARIMAX สามารถแสดงด้วยสมการดังต่อไปนี้ Chatfield (17)

$$\phi_p(L)(1-L)^d X_1 = \theta_q(L^m)\epsilon_t + \beta(L)Z_t \quad (8)$$

$\phi_p(L)$ ตัวดำเนินการ *Autoregressive (AR)* ของลำดับ *p*

$(1-L^m)^d$ ตัวดำเนินการ *Differencing* เพื่อแปลงข้อมูลให้เป็น *Stationary*

$\theta_q(L)$ ตัวดำเนินการ *Moving Average (MA)* ของลำดับ *q*

X_t ค่าของข้อมูลในช่วงเวลา *t*

ϵ_t ข้อผิดพลาดแบบสุ่ม (*White Noise)*

Z_t ตัวแปรอิสระภายนอก *(*Exogenous variables) ที่เวลา t

$\beta(L)$ ค่าสัมประสิทธิ์ของตัวแปรอิสระ ซึ่งอาจอยู่ในรูปแบบล่าช้า (lag) ได้

จากสมการที่(8) เนื่องจากแบบจำลอง ARIMAX พัฒนาต่อยอดมาจากแบบจำลอง ARIMA ดังนั้นจึงยังคงมีขั้นตอนที่จัดการกับอนุกรมเวลาที่มีแนวโน้ม (trend) ด้วยการทำ differencing เพื่อกำจัดแนวโน้มออกก่อน ซึ่งช่วยให้แบบจำลอง ARIMAX สามารถใช้กับข้อมูลที่มีแนวโน้มระยะยาวได้ดี และเมื่อเพิ่มตัวแปรภายนอกเข้าไป แบบจำลองจะสามารถจับความสัมพันธ์ระหว่างปัจจัยภายนอกกับตัวแปรหลักได้โดยตรง ทำให้เหมาะสมอย่างยิ่งสำหรับกรณีที่มีตัวแปรภายนอกซึ่งส่งผลต่อพฤติกรรมของข้อมูลอนุกรมเวลา อย่างไรก็ตาม หากข้อมูลมีฤดูกาลที่เด่นชัด เช่น มีรูปแบบซ้ำรายเดือนหรือรายไตรมาส การใช้แบบจำลอง ARIMAX เพียงอย่างเดียวอาจไม่เพียงพอ เนื่องจากแบบจำลองนี้ไม่มีองค์ประกอบที่จัดการฤดูกาลได้โดยตรง ดังนั้นอาจต้องรวมปัจจัยฤดูกาลเข้าไปเป็นตัวแปรภายนอกหรือทำการ seasonal differencing กับข้อมูล เพื่อทำให้ลักษณะของฤดูกาลหายไปก่อนสร้างแบบจำลอง ARIMAX

ในงานวิจัยนี้ แบบจำลอง ARIMAX ถูกนำมาใช้ในการพยากรณ์จำนวนห้องพักที่ขายได้ โดยเน้นการนำตัวแปรอิสระภายนอกที่เกี่ยวข้องกับรายได้ เช่น อัตราค่าห้องพักเฉลี่ย *(*ADR) จำนวนคืน

ที่เข้าพัก (total nights) หรือปัจจัยด้านฤดูกาล มาประกอบกับข้อมูลยอดขายในอดีต เพื่อเพิ่มความแม่นยำของการพยากรณ์ แบบจำลอง ARIMAX จึงเหมาะสำหรับสถานการณ์ที่รายได้ของโรงแรมได้รับอิทธิพลจากปัจจัยหลายด้าน ไม่ได้ขึ้นอยู่กับแนวโน้มในอดีตเพียงอย่างเดียว ในกรณีที่ฤดูกาลส่งผลอย่างมากในอนุกรมเวลา การใช้แบบจำลอง SARIMAX ที่จะกล่าวถึงต่อไป ซึ่งเป็นการเพิ่มองค์ประกอบฤดูกาลเข้าไปในแบบจำลองโดยตรงจะมีความเหมาะสมกว่า เนื่องจากแบบจำลอง SARIMAX ถูกออกแบบมาสำหรับข้อมูลที่มีทั้งแนวโน้ม ฤดูกาล และปัจจัยภายนอกโดยเฉพาะ

SARIMA (Seasonal Autoregressive Integrated Moving Average)

SARIMA เป็นหนึ่งในแบบจำลองการพยากรณ์เชิงสถิติที่ได้รับความนิยมสำหรับการวิเคราะห์ข้อมูลอนุกรมเวลา เป็นการพัฒนาเพิ่มเติมจากแบบจำลอง ARIMA โดยเพิ่มส่วนประกอบที่สามารถจัดการกับรูปแบบตามฤดูกาล (Seasonality) ในข้อมูลอนุกรมเวลาได้ Fatoumata&Christine (18) SARIMA เหมาะสำหรับข้อมูลที่แสดงลักษณะการเปลี่ยนแปลงตามฤดูกาล โดยสามารถวิเคราะห์และจับความสัมพันธ์ทั้งในระยะสั้นและตามฤดูกาลได้อย่างมีประสิทธิภาพแบบจำลอง SARIMA ทำงานโดยการรวมองค์ประกอบ 4 ส่วน ได้แก่

Autoregressive (AR) พิจารณาความสัมพันธ์ระหว่างค่าปัจจุบัน และค่าที่ผ่านมาในช่วงเวลาต่าง ๆ

Integrated (I) ใช้ความแตกต่างเพื่อเปลี่ยนข้อมูลที่ไม่คงที่ ให้เป็นข้อมูลคงที่

Moving Average (MA) พิจารณาความสัมพันธ์ระหว่างค่าปัจจุบัน และค่าความผิดพลาดจากช่วงเวลาก่อนหน้า

Seasonal Component เพิ่มพารามิเตอร์เชิงฤดูกาล เช่น ฤดูกาลรายเดือน หรือรายปี เพื่อจับลักษณะของข้อมูลที่เกิดซ้ำตามรอบเวลา การเพิ่มส่วนของฤดูกาลเข้าไปในแบบจำลอง ทำให้มีพารามิเตอร์ที่แยกย่อยออกเป็น Seasonal AR, Seasonal Differencing และ Seasonal MA เพื่อช่วยให้จับลักษณะเฉพาะของฤดูกาลที่อาจมีผลกระทบต่อข้อมูลได้ แสดงเป็นรูปแบบดังนี้

องค์ประกอบสำคัญของ SARIMA(p,d,q)(P,D,Q,m)

p,d,q ค่าพารามิเตอร์สำหรับส่วนที่ไม่ใช่ฤดูกาล

p: ลำดับของตัวแปรเชิงอัตโนมัติ (*Autoregressive - AR)*

d: ระดับของการทำ *Differencing* เพื่อทำให้ข้อมูลคงที่ (*Stationary)*

q: ลำดับของตัวแปรค่าเฉลี่ยเคลื่อนที่ (*Moving Average - MA)*

P,D,Q: ค่าพารามิเตอร์สำหรับส่วนที่เป็นฤดูกาล

P: ลำดับของตัวแปรเชิงอัตโนมัติในช่วงฤดูกาล

D: ระดับของการทำ *Differencing* ในช่วงฤดูกาล

Q: ลำดับของตัวแปรค่าเฉลี่ยเคลื่อนที่ในช่วงฤดูกาล

m: ความยาวของช่วงฤดูกาล เช่น m =12 หมายถึงข้อมูลรายเดือนที่มีฤดูกาลเป็นรายปี

สมการทั่วไปของ SARIMA สามารถเขียนได้ดังนี้

$$\phi_p(L)\phi_p(L^m)(1-L)^d(1-L^m)^D X_1 = \theta_q L\theta_Q(L^m)\epsilon_t \quad (9)$$

$\phi_p(L)$ ตัวดำเนินการ Autoregressive ที่ไม่ใช่ฤดูกาล

$\phi_p(L^m)$ ตัวดำเนินการ Autoregressive ในช่วงฤดูกาล

$\theta_q L$ ตัวดำเนินการ Moving Average ที่ไม่ใช่ฤดูกาล

$\theta_Q(L^m)$ ตัวดำเนินการ Moving Average ในช่วงฤดูกาล

$(1-L)^d$ ตัวดำเนินการ Differencing ที่ไม่ใช่ฤดูกาล

$(1-L^m)^D$ตัวดำเนินการ Differencing ในช่วงฤดูกาล

ϵ_t ข้อผิดพลาดแบบสุ่ม (White Noise)

จากสมการ *(9)* จะเห็นได้ว่าแบบจำลอง SARIMA ได้รับการพัฒนาต่อยอดจาก ARIMA โดยเพิ่มความสามารถในการจัดการกับลักษณะของข้อมูลที่มีฤดูกาลเข้ามาโดยเฉพาะ ทำให้สามารถวิเคราะห์ความสัมพันธ์ของข้อมูลได้ทั้งในระยะสั้น และในรอบฤดูกาลอย่างมีประสิทธิภาพ ด้วยการรวมองค์ประกอบของการทำ differencing ทั้งในส่วนที่ไม่ใช่ฤดูกาล และในระดับฤดูกาล รวมถึงการพิจารณาอิทธิพลจากค่าก่อนหน้า และความผิดพลาดในอดีตภายใต้กรอบของ Autoregressive และ Moving Average

ในงานวิจัยนี้ SARIMA ถูกนำมาใช้เพื่อพยากรณ์รายได้ของโรงแรมซึ่งมีลักษณะของข้อมูลที่แสดงรูปแบบตามฤดูกาลอย่างชัดเจน เช่น ช่วงเทศกาลท่องเที่ยวหรือช่วงนอกฤดูท่องเที่ยว การเลือกใช้แบบจำลองนี้มีวัตถุประสงค์เพื่อให้สามารถจับแนวโน้มของรายได้ในช่วงเวลาต่าง ๆ ได้อย่างแม่นยำ ช่วยให้ผู้ประกอบการสามารถวางแผนเชิงกลยุทธ์ได้อย่างเหมาะสม เช่น การกำหนดราคาห้องพัก การบริหารจัดการต้นทุน หรือการจัดสรรทรัพยากรในแต่ละช่วงเวลาให้มีประสิทธิภาพมากยิ่งขึ้น

SARIMAX (Seasonal ARIMA with Exogenous Variables)

แบบจำลอง SARIMAX เป็นการผสมผสานระหว่างแบบจำลอง ARIMA ที่มีส่วนประกอบฤดูกาล(SARIMA) กับการรวมตัวแปรภายนอกเข้าไว้ด้วยกัน จึงนับเป็นแบบจำลองที่ครอบคลุมกรณี

อนุกรมเวลาซับซ้อนที่มีทั้งแนวโน้ม ฤดูกาล และปัจจัยภายนอกที่ส่งผลต่ออนุกรมเวลาพร้อมกัน ในเชิงโครงสร้าง SARIMAX ประกอบด้วย 4 องค์ประกอบหลักเช่นเดียวกับ SARIMA เพื่อรองรับรูปแบบซ้ำตามฤดูกาลของข้อมูล ควบคู่ไปกับการนำข้อมูลปัจจัยภายนอกเข้ามาช่วยในการพยากรณ์ ทำให้แบบจำลองนี้มีศักยภาพในการพยากรณ์ที่แม่นยำขึ้น และตอบสนองต่อการเปลี่ยนแปลงจากปัจจัยภายนอกได้ดียิ่งขึ้น โดยการดำเนินการของแบบจำลองแสดงดังสมการดังต่อไปนี้ Chatfield (17)

$$\phi_p(L^m)\phi_p(L)(1-L)^d(1-L^m)^D X_1 = \theta_q(L^m)\epsilon_t + \beta(L)Z_t \quad (10)$$

$\phi_p(L)$ ตัวดำเนินการ Autoregressive ที่ไม่ใช่ฤดูกาล

$\phi_p(L^m)$ ตัวดำเนินการ Autoregressive ในช่วงฤดูกาล

$\theta_q L$ ตัวดำเนินการ Moving Average ที่ไม่ใช่ฤดูกาล

$\theta_Q(L^m)$ ตัวดำเนินการ Moving Average ในช่วงฤดูกาล

$(1-L)^d$ ตัวดำเนินการ Differencing ที่ไม่ใช่ฤดูกาล

$(1-L^m)^D$ ตัวดำเนินการ Differencing ในช่วงฤดูกาล

ϵ_t ข้อผิดพลาดแบบสุ่ม (White Noise)

Z_t ตัวแปรอิสระภายนอก (Exogenous variables) ที่เวลา t

$\beta(L)$ ค่าสัมประสิทธิ์ของตัวแปรอิสระ ซึ่งอาจอยู่ในรูปแบบล่าช้า (lag) ได้

จากสมการที่ (10) สามารถอธิบายความสัมพันธ์ได้ว่าแบบจำลอง SARIMAX มีทุกองค์ประกอบของแบบจำลอง ARIMA ที่ขยายเพิ่มเติมด้วยส่วนฤดูกาล และ ตัวแปรอิสระภายนอก จึงสามารถจัดการทั้งความสัมพันธ์ในตัวข้อมูล แนวโน้ม รูปแบบที่เกิดขึ้นซ้ำ ๆ ตามฤดูกาล และอิทธิพลจากปัจจัยภายนอกที่มีผลกระทบต่อข้อมูล โดยการรวมตัวแปรภายนอกเข้ามายังช่วยให้แบบจำลองสามารถสะท้อนการเปลี่ยนแปลงของข้อมูล ที่อาจเกิดขึ้นตามปัจจัยภายนอกได้ เช่น สภาวะเศรษฐกิจ หรือช่วงฤดูท่องเที่ยว ซึ่งส่งผลต่อรูปแบบตามฤดูกาลของข้อมูล

ในงานวิจัยนี้ SARIMAX ถูกนำมาใช้เพื่อพยากรณ์จำนวนห้องพักที่ขายได้ของโรงแรม โดยนอกจากจะคำนึงถึงแนวโน้ม และฤดูกาลในข้อมูลแล้ว ยังนำตัวแปรสำคัญที่อาจส่งผลต่อรายได้ เช่น อัตราค่าห้องพักเฉลี่ย (ADR) หรือระยะเวลาการจองล่วงหน้า (lead time) เข้ามาเป็นตัวแปรอิสระภายนอก เพื่อเพิ่มความแม่นยำในการพยากรณ์ให้สอดคล้องกับสถานการณ์ทางธุรกิจที่เปลี่ยนแปลงอยู่เสมอ ช่วยสนับสนุนการวางแผนกลยุทธ์ และการตัดสินใจที่เหมาะสมสำหรับผู้ประกอบการโรงแรม

PROPHET

เครื่องมือที่ถูกพัฒนาโดย Facebook เพื่อใช้ในการพยากรณ์ข้อมูลเชิงเวลา โดย Prophet ถูกออกแบบมาให้สามารถรับมือกับข้อมูลที่มีแนวโน้มตามฤดูกาล โดยสามารถแยกความสัมพันธ์ระหว่างฤดูกาลที่เกิดขึ้นทุกปี และความเปลี่ยนแปลงของรายได้ในแต่ละเดือน รวมถึงจัดการกับความไม่สม่ำเสมอของข้อมูลได้ดี Taylor&Benjamin (19)แบบจำลองนี้จึงสามารถจัดการกับข้อมูลที่มีการเติบโตเชิงเส้น และเชิงทวีคูณ รวมถึงสามารถปรับแต่งตามความซับซ้อนของข้อมูลได้ ทำให้ Prophet เป็นอีกหนึ่งเครื่องมือที่เหมาะสำหรับการพยากรณ์รายได้ของโรงแรม

จากสมการ (11) จะเห็นได้ว่าแบบจำลอง Prophet ถูกออกแบบมาเพื่อรองรับการวิเคราะห์ข้อมูลอนุกรมเวลาที่มีแนวโน้ม และรูปแบบตามฤดูกาลอย่างชัดเจน โดยพัฒนาในลักษณะของแบบจำลองเชิงบวก (Additive Model) ซึ่งประกอบด้วยองค์ประกอบหลัก 3 ส่วน ได้แก่ แนวโน้มของข้อมูล (Trend) รูปแบบฤดูกาล (Seasonality) และปัจจัยสุ่ม หรือความผิดพลาด (Residuals) โดยแบบจำลองมีความยืดหยุ่นในการจัดการกับข้อมูลที่มีลักษณะไม่เป็นเชิงเส้น มีการเปลี่ยนแปลงของแนวโน้ม หรือได้รับผลกระทบจากปัจจัยภายนอก แสดงดังสมการต่อไปนี้

$$y(t) = g(t) + s(t) + \epsilon t \qquad (11)$$

โดยที่

g(t) เป็นฟังก์ชันแนวโน้ม ซึ่งอาจเป็นเชิงเส้น หรือเชิงลอจิสติกก็ได้

s(t) เป็นฟังก์ชันของฤดูกาล ซึ่งอาจเป็นรายปี รายเดือน หรือรายวัน

h(t) เป็นฟังก์ชันของปัจจัยอื่น ๆ ที่คำนึงถึงปัจจัยที่อาจมีผลต่อรายได้

ϵt : เป็นเทอมของข้อผิดพลาด (error term) ที่เกิดขึ้นในแต่ละช่วงเวลา

นอกจากโครงสร้างพื้นฐานของแบบจำลอง Prophet ที่นำเสนอในรูปแบบ Additive Model แล้วแบบจำลองนี้ยังมีความโดดเด่นในด้านการวิเคราะห์แนวโน้ม และฤดูกาลที่ซับซ้อน โดยใช้แนวทางแบบ saturating growth model และ piecewise linear model สำหรับการจับแนวโน้มของข้อมูลที่มีการเติบโตอย่างรวดเร็ว หรือมีจุดเปลี่ยนของแนวโน้มในช่วงเวลาต่าง ๆ อย่างชัดเจน ส่วนในด้านของฤดูกาลแบบจำลองได้นำเทคนิค อนุกรมฟูเรียร์ (Fourier Series) มาใช้ในการสร้างฟังก์ชันที่สามารถอธิบายรูปแบบที่เกิดซ้ำตามรอบเวลาได้อย่างยืดหยุ่น และแม่นยำ นอกจากนี้ Prophet ยังสามารถจัดการกับปัจจัยแทรกซ้อนอื่น ๆ ที่อาจส่งผลกระทบต่อข้อมูล เช่น วันหยุดนักขัตฤกษ์ หรือเหตุการณ์เฉพาะเจาะจง โดยอนุญาตให้ผู้ใช้สามารถเพิ่มตัวแปรภายนอก (external Exogenous variables) เข้าไป

ในแบบจำลอง และตั้งค่าพารามิเตอร์ต่าง ๆ ให้เป็นอิสระต่อกัน เพื่อให้แบบจำลองมีความยืดหยุ่นสูงในการเรียนรู้ และจับรูปแบบที่ซับซ้อนของข้อมูลจริง

ในงานวิจัยนี้ Prophet จะถูกนำมาใช้ในการพยากรณ์ห้องพักที่ขายได้ โดยสามารถรองรับข้อมูลที่มีแนวโน้มชัดเจน รวมถึงข้อมูลที่ได้รับอิทธิพลจากปัจจัยภายนอก เช่น วันหยุดพิเศษช่วงเทศกาล หรือเหตุการณ์สำคัญที่มีผลต่อการเปลี่ยนแปลงรายได้ ด้วยความสามารถในการระบุจุดเปลี่ยนแปลงของแนวโน้ม (changepoints) และรองรับลักษณะข้อมูลที่มีฤดูกาลในรูปแบบต่าง ๆ ได้อย่างยืดหยุ่น จะช่วยสนับสนุนการตัดสินใจเชิงกลยุทธ์ในการบริหารรายได้ การกำหนดราคาห้องพัก และการวางแผนการตลาดของผู้ประกอบการโรงแรมให้มีประสิทธิภาพสูงสุด

เกณฑ์ที่ใช้ในการวัดผลการพยากรณ์

การวัดความแม่นยำของผลการพยากรณ์มีความสำคัญอย่างยิ่งต่อกระบวนการบริหารจัดการรายได้ของโรงแรม (Revenue Management) เนื่องจากแต่ละเกณฑ์มีจุดเด่นที่แตกต่างกัน และผลลัพธ์ของการพยากรณ์มีผลโดยตรงต่อการตัดสินใจด้านราคา และการจัดการห้องพัก Schwartz, Ma (20)โดยทั่วไปแล้วเกณฑ์ที่นิยมใช้ในการวัดผลความแม่นยำของการพยากรณ์ในอุตสาหกรรมโรงแรมประกอบด้วยค่าความคลาดเคลื่อนที่สำคัญหลายรูปแบบ โดยงานวิจัยนี้ได้เลือกใช้เกณฑ์วัดผลการพยากรณ์ดังนี้

RMSE (Root Mean Squared Error)

RMSE เป็นเกณฑ์ที่คำนวณจากค่าความผิดพลาดของการพยากรณ์ที่ถูกนำมายกกำลังสองก่อนจะถอดรากที่สองในขั้นตอนสุดท้าย วิธีนี้จะช่วยเน้นข้อผิดพลาดที่มีขนาดใหญ่มากขึ้น ทำให้เห็นได้ชัดเจนดังสมการต่อไปนี้

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - A_i)^2} \quad (12)$$

โดยที่

n จำนวนข้อมูลทั้งหมดที่ใช้ประเมิน

F_i ค่าที่แบบจำลองพยากรณ์ ในลำดับที่ i

A_i ค่าจริงที่เกิดขึ้นจริงในลำดับที่ i

จากสมการ (12) RMSE เป็นตัวชี้วัดที่ใช้วัดประสิทธิภาพของแบบจำลอง โดยวัดจากระยะห่างระหว่างค่าที่พยากรณ์กับค่าจริง ด้วยผลลัพธ์ที่ช่วยเน้นข้อผิดพลาดที่มีขนาดใหญ่จะช่วยให้ผู้บริหารตระหนักถึงความเสี่ยงของข้อผิดพลาดเหล่านั้น ข้อดีอีกประการหนึ่งคือ RMSE มีหน่วยชัดเจน

ตรงกับข้อมูลที่ใช้ เช่น จำนวนห้องพักที่จอง หรือรายได้เฉลี่ยต่อห้องพัก (ADR) ทำให้สะดวกในการตีความผล และเปรียบเทียบความแม่นยำระหว่างแบบจำลอง

MAE (Mean Absolute Error)

MAE เป็นเกณฑ์ที่คำนวณค่าความผิดพลาดโดยตรง โดยใช้ค่าคลาดเคลื่อนแบบสัมบูรณ์ โดยจะไม่พิจารณาว่าค่าที่พยากรณ์จะมากกว่า หรือน้อยกว่าค่าจริง อีกทั้งไม่มีการยกกำลังสอง จึงทำให้ข้อผิดพลาดแต่ละจุดมีน้ำหนักเท่ากันทั้งหมด กล่าวคือ MAE จะไม่เน้นขยายผลข้อผิดพลาดที่มีค่าสูงมากเกินไป แต่จะสะท้อนค่าความผิดพลาดโดยเฉลี่ยของแบบจำลองอย่างตรงไปตรงมา เหมาะสำหรับการประเมินความแม่นยำในภาพรวมที่เข้าใจง่าย โดยเฉพาะการบริหารรายได้แบบรายวัน หรือรายเดือนที่ไม่มีความผันผวนสูงมากดังสมการต่อไปนี้

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - A_i| \quad (13)$$

โดยที่

n จำนวนข้อมูลทั้งหมดที่ใช้ประเมิน

F_i ค่าที่แบบจำลองพยากรณ์ ในลำดับที่ i

A_i ค่าจริงที่เกิดขึ้นจริงในลำดับที่ i

MAPE (Mean Absolute Percentage Error)

MAPE เป็นเกณฑ์ที่แสดงผลในรูปแบบของเปอร์เซ็นต์ และไม่มีหน่วยวัด (scale-free) จึงเหมาะสำหรับการเปรียบเทียบความแม่นยำระหว่างข้อมูลที่ต่างกัน เช่น โรงแรมที่มีขนาดต่างกันหรือช่วงเวลาที่แตกต่างกัน อย่างไรก็ตามข้อจำกัดสำคัญของ MAPE คือปัญหาความไม่สมมาตร เนื่องจากการพยากรณ์เกินจริง จะให้ค่าความผิดพลาดสูงกว่าการพยากรณ์ต่ำกว่าความจริง แม้ว่าจะมีขนาดข้อผิดพลาดเท่ากันก็ตามแสดงดังสมการต่อไปนี้

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{F_i - A_i}{A_i}\right| \times 100 \quad (14)$$

โดยที่

n จำนวนข้อมูลทั้งหมดที่ใช้ประเมิน

F_i ค่าที่แบบจำลองพยากรณ์ ในลำดับที่ i

A_i ค่าจริงที่เกิดขึ้นจริงในลำดับที่

ผลงานวิจัยที่เกี่ยวข้อง

ปี 2008 Song และ Li (21) ได้นำเสนอบทความวิจัยที่ทำการสำรวจแนวโน้มล่าสุดในการสร้างแบบจำลองและการพยากรณ์อุปสงค์การท่องเที่ยว (Tourism Demand Modelling and Forecasting) โดยรวบรวมงานวิจัยที่ตีพิมพ์ตั้งแต่ปี ค.ศ. 2000 จำนวนทั้งหมด 121 ฉบับ งานวิจัยพบว่าการพยากรณ์อุปสงค์การท่องเที่ยวมีการใช้เทคนิคที่หลากหลายมากขึ้น ซึ่งแบ่งออกเป็นสองกลุ่มหลัก ได้แก่ แบบจำลองอนุกรมเวลา (Time Series Models) เช่น ARIMA, SARIMA, Naïve และ Exponential Smoothing ซึ่ง ARIMA และ SARIMA เป็นแบบจำลองที่ได้รับความนิยมสูงสุด โดยเฉพาะ SARIMA ที่เหมาะสำหรับข้อมูลที่มีฤดูกาล และอีกกลุ่มคือแบบจำลองเชิงเศรษฐมิติ (Econometric Models) เช่น ECM (Error Correction Model), VAR (Vector Autoregressive), ADLM (Autoregressive Distributed Lag Model), TVP (Time Varying Parameter Model) และ AIDS (Almost Ideal Demand System) ซึ่งแบบจำลองกลุ่มนี้มีจุดเด่นในการวิเคราะห์ความสัมพันธ์ของข้อมูล และอธิบายปัจจัยต่าง ๆ ที่ส่งผลต่ออุปสงค์ได้ชัดเจนขึ้นนอกจากนี้ งานวิจัยได้สรุปว่าไม่มีแบบจำลองใดที่สามารถพยากรณ์ได้แม่นยำที่สุดในทุกสถานการณ์ แต่ความแม่นยำของแบบจำลองขึ้นอยู่กับลักษณะของข้อมูล เช่น ความถี่ของข้อมูลแบบรายปี รายไตรมาส รายเดือน ช่วงเวลาในการพยากรณ์ และรูปแบบของข้อมูลที่นำมาวิเคราะห์ เช่น รายได้ หรือค่าใช้จ่ายของนักท่องเที่ยว อย่างไรก็ตามงานวิจัยยังชี้ให้เห็นถึงข้อจำกัด เช่น งานวิจัยส่วนใหญ่ยังคงใช้ข้อมูลรายปีเป็นหลักซึ่งไม่เพียงพอต่อการพยากรณ์ระยะสั้น ขาดการใช้วิธีเชิงสถิติในการทดสอบความแตกต่างของประสิทธิภาพระหว่างแบบจำลอง และยังขาดความชัดเจนในการเลือกแบบจำลองในการเปรียบเทียบประสิทธิภาพอีกด้วย

ปี 2003 Weatherford และ Kimes (22)ได้นำเสนอบทความวิจัยที่เปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ที่หลากหลายสำหรับการจัดการรายได้ของโรงแรม (Hotel Revenue Management) โดยมีเป้าหมายหลักเพื่อค้นหาวิธีการพยากรณ์ที่มีความเหมาะสมและแม่นยำที่สุดในการคาดการณ์ความต้องการห้องพัก ซึ่งเป็นปัจจัยสำคัญในการบริหารจัดการรายได้ของธุรกิจโรงแรม

ในงานวิจัยนี้ได้มีการศึกษาและเปรียบเทียบวิธีการพยากรณ์ทั้งในกลุ่มวิธีการเชิงสถิติ เช่น แบบจำลอง ARIMA และ Exponential Smoothing และวิธีการพยากรณ์ที่ใช้การเรียนรู้ของเครื่อง (Machine Learning) รวมถึงแบบจำลองไฮบริด (Hybrid Models) ที่ผสมผสานเทคนิคต่าง ๆ เพื่อเพิ่มประสิทธิภาพในการพยากรณ์ โดยข้อมูลที่ใช้ประกอบด้วยข้อมูลการจองห้องพักของโรงแรมที่มีตัวแปรต่าง ๆ ที่ส่งผลต่อความต้องการ เช่น ช่วงเวลา การแข่งขันในตลาด และการเปลี่ยนแปลงราคา

ผลการศึกษาชี้ให้เห็นว่า การเลือกใช้วิธีการพยากรณ์ที่เหมาะสมนั้นขึ้นอยู่กับลักษณะของข้อมูล โดยสำหรับข้อมูลที่มีความผันผวนและความซับซ้อนสูงจะเหมาะสมกับการใช้แบบจำลองที่อาศัยการเรียนรู้ของเครื่อง หรือแบบจำลองไฮบริดมากกว่า ในขณะที่ข้อมูลที่มีความคงที่ หรือไม่ซับซ้อนมากนัก เช่น ข้อมูลที่ใช้ในการพยากรณ์ระยะสั้น ยังคงเหมาะสมกับแบบจำลองทางสถิติอย่าง ARIMA และ Exponential Smoothing มากกว่า งานวิจัยนี้จึงให้ข้อแนะนำที่สำคัญสำหรับการเลือกใช้แบบจำลองที่เหมาะสมในการพยากรณ์รายได้ของธุรกิจโรงแรม โดยคำนึงถึงลักษณะเฉพาะของข้อมูลที่มีอยู่เป็นหลัก

ปี 2019 Enache (23)ได้นำเสนอบทความวิจัยที่ทำการสำรวจวิธีการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ในการจัดการรายได้สำหรับอุตสาหกรรมการท่องเที่ยว โดยงานวิจัยเน้นการเพิ่มประสิทธิภาพของระบบจัดการรายได้ ผ่านการวิเคราะห์ข้อมูลเชิงลึกและการประมวลผลข้อมูลแบบอัตโนมัติแบบเรียลไทม์ เพื่อช่วยในการปรับราคาห้องพักและการบริหารจัดการความสามารถในการรองรับลูกค้าให้เหมาะสมกับอุปสงค์ในแต่ละช่วงเวลา

ในการจัดการรายได้ Machine Learning ถูกนำมาใช้เป็นเครื่องมือสำคัญในกระบวนการแบ่งกลุ่มลูกค้า ซึ่งช่วยให้ผู้บริหารโรงแรมสามารถระบุ และจำแนกกลุ่มลูกค้าที่มีลักษณะพฤติกรรมและระดับราคาที่ยอมรับได้แตกต่างกันออกจากกันได้อย่างแม่นยำ อีกทั้งยังมีการใช้ Machine Learning ในการวิเคราะห์พฤติกรรมของลูกค้า เพื่อปรับกลยุทธ์ด้านการตั้งราคาห้องพัก และการเสนอโปรโมชั่นให้เหมาะสม ตรงตามอุปสงค์ของแต่ละกลุ่มเป้าหมายมากขึ้น

นอกจากนี้ งานวิจัยยังกล่าวถึงการใช้เทคนิค Machine Learning เพื่อการพยากรณ์อุปสงค์ห้องพัก โดยใช้ข้อมูลจากพฤติกรรมการจองล่วงหน้า และข้อมูลประวัติการจองที่ผ่านมา ซึ่งช่วยให้ผู้บริหารโรงแรมสามารถคาดการณ์แนวโน้มของอุปสงค์ในช่วงเวลาต่าง ๆ ได้อย่างแม่นยำมากขึ้น งานวิจัยดังกล่าวชี้ให้เห็นว่าเทคนิค Machine Learning มีศักยภาพในการเพิ่มความแม่นยำ และประสิทธิภาพของการจัดการรายได้ ส่งผลให้โรงแรมสามารถเพิ่มรายได้สูงสุดและบริหารจัดการธุรกิจได้อย่างมีประสิทธิภาพยิ่งขึ้น

ปี 2016 Vagropoulos และคณะ(24)ได้นำเสนองานวิจัยที่เปรียบเทียบประสิทธิภาพของแบบจำลอง SARIMAX, SARIMA, modified SARIMA และ Artificial Neural Networks (ANN) ในการพยากรณ์การผลิตพลังงานแสงอาทิตย์แบบเชื่อมต่อกริดในระยะสั้น (Short-term PV Generation Forecasting) โดยงานวิจัยนี้เน้นการทดสอบความแม่นยำในการพยากรณ์ผ่านเกณฑ์การประเมินผล Normalized Root Mean Square Error (NRMSE) โดยข้อมูลที่นำมาใช้ในการทดสอบเป็นข้อมูลจริงจากโรงไฟฟ้าพลังงานแสงอาทิตย์ในประเทศกรีซผลการศึกษาพบว่าแบบจำลอง SARIMAX ซึ่งใช้ข้อมูล

ปัจจัยภายนอก อย่างเช่นข้อมูลการพยากรณ์ปริมาณรังสีดวงอาทิตย์ ที่ได้จากแบบจำลองการพยากรณ์สภาพอากาศเชิงตัวเลข (Numerical Weather Prediction: NWP) ช่วยให้ประสิทธิภาพในการพยากรณ์สูงขึ้นเมื่อเทียบกับ SARIMA แบบดั้งเดิมและ SARIMA ที่ผ่านการปรับปรุง โดยเฉพาะในช่วงฤดูหนาวและฤดูใบไม้ผลิที่สภาพอากาศเปลี่ยนแปลงบ่อย อย่างไรก็ตาม ในช่วงฤดูร้อนที่สภาพอากาศค่อนข้างนิ่ง แบบจำลอง SARIMA มีความแม่นยำมากกว่า SARIMAX เนื่องจากการรวมข้อมูลภายนอกกลับทำให้ประสิทธิภาพในการพยากรณ์ลดลง นอกจากนี้ แบบจำลอง ANN ยังมีประสิทธิภาพสูงที่สุดในการพยากรณ์แบบรายวัน

สำหรับการพยากรณ์ระหว่างวัน พบว่าแบบจำลอง SARIMA ซึ่งไม่ได้ใช้ข้อมูลภายนอกกลับมีประสิทธิภาพดีกว่าแบบจำลอง SARIMAX โดยเฉพาะเมื่อใช้ข้อมูลจริงในช่วงเช้าของวันในการปรับแบบจำลองเพิ่มเติมก่อนการพยากรณ์ ทำให้สามารถทำนายแนวโน้มพลังงานได้อย่างมีประสิทธิภาพแม้ในช่วงเวลาที่สภาพอากาศผันผวน

จากผลการศึกษาดังกล่าว ผู้วิจัยจึงได้สรุปว่า การนำข้อมูลปัจจัยภายนอกมาร่วมใช้ในแบบจำลองอนุกรมเวลานั้นมีประโยชน์ในการพยากรณ์ระยะสั้นแบบรายวัน แต่สำหรับการพยากรณ์แบบระหว่างวัน การใช้ข้อมูลภายนอกอาจไม่จำเป็น และการใช้ SARIMA แบบพื้นฐานจะให้ผลที่ดีกว่า ทั้งนี้ผลการวิจัยขึ้นอยู่กับลักษณะของข้อมูลที่นำมาใช้ในแต่ละช่วงเวลาเป็นสำคัญ

ปี 2021 Suwatanapongched ได้นำเสนอบทความวิจัยเกี่ยวกับการพยากรณ์ความต้องการห้องพักรายวันของโรงแรมโดยใช้ข้อมูลที่ผ่านกระบวนการแปลงข้อมูลก่อนการพยากรณ์ งานวิจัยนี้ใช้ข้อมูลจริงจากโรงแรมระดับ 4 ดาวในประเทศไทยระหว่างปี ค.ศ. 2016–2019 โดยการแปลงข้อมูลประกอบด้วยการทำให้ข้อมูลเรียบ (Data Smoothing) เพื่อลดความผันผวนของข้อมูล และการปรับค่าผิดปกติ (Outlier Adjustment) จากนั้นจึงนำข้อมูลที่ผ่านกระบวนการแปลงมาใช้ในแบบจำลองการพยากรณ์ที่หลากหลาย ทั้งแบบจำลองเชิงเส้น และไม่เป็นเชิงเส้น ได้แก่ Same Day Last Year, Holt-Winters, SARIMA และ TBATS เพื่อเปรียบเทียบผลลัพธ์การพยากรณ์

ผลการศึกษาพบว่า แบบจำลอง SARIMA ที่ใช้ข้อมูลหลังการแปลงมีประสิทธิภาพในการพยากรณ์สูงสุด โดยมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) เท่ากับ 6.18 นอกจากนี้ งานวิจัยยังชี้ให้เห็นว่าการนำข้อมูลผ่านกระบวนการแปลงก่อนการพยากรณ์สามารถเพิ่มความแม่นยำในการพยากรณ์ได้อย่างมีนัยสำคัญเมื่อเปรียบเทียบกับข้อมูลที่ไม่ได้ผ่านการแปลง ผลการศึกษานี้จึงเน้นย้ำถึงความสำคัญของการเตรียมข้อมูลให้เหมาะสมก่อนที่จะนำไปใช้ในแบบจำลองพยากรณ์ เพื่อให้ได้ผลลัพธ์ที่ถูกต้องและมีประสิทธิภาพสูงสุด(25)

ปี 2021 Sison, Li และ Han (26)ได้นำเสนอบทความวิจัยที่สำรวจการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) และการเรียนรู้เชิงลึก (Deep Learning) เพื่อการพยากรณ์ความต้องการเดินทาง (Travel Demand Forecasting) ในอุตสาหกรรมการท่องเที่ยว งานวิจัยนี้ได้เปรียบเทียบประสิทธิภาพของแบบจำลองแบบดั้งเดิม เช่น ARIMA และ Exponential Smoothing กับเทคนิคการพยากรณ์ที่ทันสมัยขึ้น เช่น Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN) และแบบจำลองไฮบริด (Hybrid Models) ที่ผสมผสานเทคนิคต่างๆ เข้าด้วยกันขั้นตอนสำคัญของงานวิจัยประกอบด้วยการคัดเลือกชุดข้อมูลที่เกี่ยวข้องกับอุตสาหกรรมการท่องเที่ยว และการเตรียมข้อมูลให้พร้อมด้วยกระบวนการปรับปรุงข้อมูล ก่อนจะนำไปใช้ในการพยากรณ์ ผลการศึกษาพบว่าแบบจำลองไฮบริด โดยเฉพาะอย่างยิ่งการผสมผสาน CNN กับการแบ่งกลุ่มข้อมูลโดยใช้ k-means clustering สามารถลดค่าความผิดพลาดในการพยากรณ์ได้อย่างมีประสิทธิภาพ เห็นได้จากค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error - RMSE) ที่ลดลงอย่างมีนัยสำคัญเมื่อเปรียบเทียบกับแบบจำลองดั้งเดิม งานวิจัยนี้จึงแสดงให้เห็นว่าเทคนิค Machine Learning และ Deep Learning แบบไฮบริดมีศักยภาพที่ดีในการปรับปรุงประสิทธิภาพการพยากรณ์ความต้องการเดินทางในภาคการท่องเที่ยวอย่างชัดเจน

ปี 2022 Sirisha และคณะ (27)ได้นำเสนองานวิจัยเปรียบเทียบประสิทธิภาพของแบบจำลองทางสถิติ ARIMA, SARIMA และแบบจำลองเชิงลึกอย่าง LSTM ในการพยากรณ์กำไรของธุรกิจด้วยข้อมูลอนุกรมเวลา โดยงานวิจัยนี้เน้นการทดสอบความแม่นยำในการพยากรณ์กำไรผ่านเกณฑ์วัดผลต่าง ๆ ได้แก่ RMSE และ MAPE โดยข้อมูลที่นำมาใช้เป็นข้อมูลยอดขายจำนวน 1 ล้านรายการครอบคลุมระยะเวลา 46 ปี ตั้งแต่ปี ค.ศ.1972 ถึง 2017

ในการวิเคราะห์ข้อมูลได้ดำเนินการเตรียมข้อมูลโดยการปรับชุดข้อมูลให้เป็นแบบ stationary สำหรับ ARIMA ในขณะที่ SARIMA และ LSTM ใช้ข้อมูลในลักษณะที่ยังไม่ stationary จากการเปรียบเทียบผลลัพธ์จากทั้งสามแบบจำลอง ผลการศึกษาพบว่าแบบจำลอง LSTM มีประสิทธิภาพสูงสุดในการพยากรณ์รายได้ โดยมีความแม่นยำประมาณ 97.01% ตามด้วย SARIMA ที่ 94.38% และ ARIMA ที่ 93.84% ตามลำดับ

นอกจากนี้ งานวิจัยยังได้ทำการพยากรณ์กำไรในช่วง 5 ปีถัดไป โดยพบว่าแบบจำลอง LSTM ให้ผลลัพธ์ที่แม่นยำ และมีแนวโน้มในการพยากรณ์ที่ชัดเจนกว่าแบบจำลองทางสถิติ อย่างไรก็ตาม งานวิจัยยังระบุว่าข้อจำกัดของ LSTM คือใช้เวลาในการประมวลผลนานและต้องการทรัพยากรในการคำนวณที่มากกว่าแบบจำลองทางสถิติ ดังนั้น สำหรับข้อมูลที่มีขนาดใหญ่แต่ไม่ซับซ้อนมาก การใช้แบบจำลอง SARIMA อาจเหมาะสมมากกว่าในทางปฏิบัติ

ผลการวิจัยครั้งนี้สะท้อนให้เห็นว่าแบบจำลองเชิงลึก เช่น LSTM มีศักยภาพที่ดีในการพยากรณ์ข้อมูลอนุกรมเวลา แต่ในเชิงปฏิบัติการเลือกใช้แบบจำลองยังต้องคำนึงถึงทรัพยากรที่มีอยู่และความซับซ้อนของข้อมูลด้วย

ปี 2021 Huang และ Zheng (28) ได้นำเสนอบทความวิจัยที่มุ่งพัฒนาแบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Networks - ANN) สำหรับการพยากรณ์โหลดพลังงานระยะสั้น (Short-Term Energy Load Forecasting) ในงานวิจัยนี้ผู้วิจัยได้เสนอแบบจำลองที่เรียกว่า DeepEnergy ซึ่งมีพื้นฐานจากเทคนิค Convolutional Neural Networks (CNN) และเปรียบเทียบประสิทธิภาพการพยากรณ์กับอัลกอริธึมอื่นๆ ได้แก่ Support Vector Machine (SVM), Random Forest, Decision Tree, Multilayer Perceptron (MLP) และ Long Short-Term Memory (LSTM)

กระบวนการหลักในการพัฒนาประกอบด้วยการทำ Normalization ของข้อมูล และการฝึกอบรมแบบ Batch Training เพื่อป้องกันปัญหาการเรียนรู้มากเกินไป และช่วยปรับปรุงความแม่นยำในการพยากรณ์ ผลการศึกษาพบว่าแบบจำลอง DeepEnergy ที่พัฒนาขึ้น มีประสิทธิภาพสูงกว่าแบบจำลองอื่นๆ อย่างชัดเจน โดยมีค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error) ต่ำที่สุดคือ 11.65% ซึ่งสอดคล้องกับข้อดีของ CNN ในการจัดการข้อมูลอนุกรมเวลาที่มีความซับซ้อนสูง และมีข้อมูลปริมาณมาก งานวิจัยนี้จึงเน้นให้เห็นว่า CNN สามารถนำมาประยุกต์ใช้ในการพยากรณ์ข้อมูลอนุกรมเวลาได้อย่างมีประสิทธิภาพ และแม่นยำในบริบทที่มีความซับซ้อนสูง เช่น การพยากรณ์โหลดพลังงานระยะสั้น(28)

ปี 2022 Kaya และคณะ (29)ได้นำเสนอบทความวิจัยเกี่ยวกับการพัฒนาแบบจำลองพยากรณ์ความต้องการห้องพักของโรงแรม โดยใช้เทคนิคการแบ่งกลุ่มโรงแรม ร่วมกับการฝังคุณสมบัติ (Feature Embedding) และแบบจำลอง Attention-Long Short-Term Memory (Attention-LSTM) เพื่อพยากรณ์ข้อมูลอนุกรมเวลาอย่างมีประสิทธิภาพ ข้อมูลที่นำมาใช้ในการศึกษาประกอบด้วยข้อมูลการจองห้องพักของโรงแรมจำนวนกว่า 1,400 แห่งในประเทศตุรกี ระหว่างปี 2012-2020 โดยมีเป้าหมายในการพยากรณ์ความต้องการห้องพักล่วงหน้าระยะเวลา 4 สัปดาห์

ในขั้นตอนการศึกษาได้มีการใช้เทคนิค K-Means Clustering เพื่อจัดกลุ่มโรงแรมที่มีลักษณะคล้ายกัน โดยพิจารณาคุณสมบัติที่มีผลต่อความต้องการของลูกค้าจำนวน 10 คุณสมบัติสำคัญ จากนั้นนำข้อมูลที่ได้ไปผ่านกระบวนการฝังคุณสมบัติเพื่อช่วยให้การวิเคราะห์มีประสิทธิภาพยิ่งขึ้น โดยผลการทดลองพบว่า แบบจำลอง Attention-LSTM ที่ถูกนำมาใช้ในการพยากรณ์สามารถรักษาข้อมูลสำคัญของช่วงเวลาต่าง ๆ และจัดการกับความซับซ้อนของข้อมูลอนุกรมเวลาได้ดีกว่าแบบจำลองทั่วไป

ส่งผลให้แบบจำลองนี้มีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error: MAE) เท่ากับ 4.61 และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเปอร์เซ็นต์ (Mean Absolute Percentage Error: MAPE) เท่ากับ 23.97 นอกจากนี้ การศึกษายังพบว่าการใช้ข้อมูลที่ผ่านการฝังคุณสมบัติในลักษณะ 10 มิติ ช่วยเพิ่มประสิทธิภาพ และความแม่นยำของแบบจำลอง Attention-LSTM ได้ดีกว่าการใช้ข้อมูลดิบที่ไม่ได้ผ่านกระบวนการปรับปรุง งานวิจัยนี้จึงแสดงให้เห็นว่าการประยุกต์ใช้เทคนิคการแบ่งกลุ่มและการฝังคุณสมบัติร่วมกับแบบจำลอง Attention-LSTM ช่วยเพิ่มความแม่นยำในการพยากรณ์ความต้องการห้องพักในโรงแรมได้อย่างมีนัยสำคัญ ซึ่งเป็นประโยชน์อย่างมากต่อการจัดการรายได้ในอุตสาหกรรมโรงแรมในอนาคต

ปี 2018 Lee (30)ได้นำเสนอบทความวิจัยเกี่ยวกับการสร้างแบบจำลอง และการพยากรณ์ความต้องการห้องพักของโรงแรม โดยเน้นการใช้ข้อมูลการจองล่วงหน้า ซึ่งถือเป็นข้อมูลสำคัญในการบริหารจัดการรายได้โรงแรม ในงานวิจัยนี้ได้มีการนำเสนอแบบจำลองที่อาศัยกระบวนการ Non-Homogeneous Poisson Process (NHPP) และ Poisson Mixture Models เพื่อจำลองการมาถึงของการจองห้องพักที่เปลี่ยนแปลงตามเวลา และผู้วิจัยยังได้แนะนำวิธีการปรับปรุงแบบจำลองโดยการอัปเดตข้อมูลการจองอย่างต่อเนื่อง (Dynamic Updating) เพื่อเพิ่มความแม่นยำในการพยากรณ์

การศึกษานี้ได้นำข้อมูลการจองจริงจากโรงแรมจำนวน 69 แห่งในประเทศสหรัฐอเมริกามาทำการวิเคราะห์ โดยผลลัพธ์การทดลองแสดงให้เห็นว่า แบบจำลอง Poisson Mixture Models ร่วมกับเทคนิค Dynamic Updating ที่นำเสนอมีประสิทธิภาพสูงกว่าวิธีการพยากรณ์แบบดั้งเดิม เช่น แบบจำลองรีเกรสชัน (Regression Models) อย่างมีนัยสำคัญ โดยสามารถลดข้อผิดพลาดในการพยากรณ์ได้อย่างชัดเจน ผลการศึกษาจึงยืนยันถึงประโยชน์ของการใช้ข้อมูลการจองล่วงหน้าร่วมกับวิธีการแบบไดนามิกในการจัดการรายได้ของโรงแรม เพื่อให้สามารถบริหารจัดการทรัพยากร และวางแผนเชิงกลยุทธ์ได้อย่างมีประสิทธิภาพ

ปี 2022 Alharbi และ Csala (31) ได้ศึกษาวิธีการพยากรณ์ความต้องการใช้ไฟฟ้าในระยะยาวในประเทศซาอุดิอาระเบีย โดยเลือกใช้แบบจำลองอนุกรมเวลาแบบ SARIMAX เข้ามาในกระบวนการพยากรณ์ โดยงานวิจัยนี้ใช้ชุดข้อมูลย้อนหลังยาวนานกว่า 40 ปี (1980–2020) ทั้งในด้านการใช้ไฟฟ้า ผลิตไฟฟ้า และโหลดไฟฟ้าสูงสุด เพื่อสร้างแบบจำลองที่รองรับการพยากรณ์ล่วงหน้าในระยะยาวถึง 30 ปีข้างหน้า ผลลัพธ์จากการศึกษาพบว่าความต้องการใช้ไฟฟ้าในซาอุดิอาระเบียมีความผันผวนสูง และได้รับผลกระทบจากทั้งปัจจัยเศรษฐกิจ สภาพอากาศ และโครงสร้างประชากร ซึ่งล้วนแต่เปลี่ยนแปลงในระยะยาวอย่างมีนัยสำคัญ ด้วยเหตุนี้แบบจำลอง SARIMAX จึงให้ผลลัพธ์การพยากรณ์ที่แม่นยำกว่า

แบบจำลอง ARIMA และ SARIMA ซึ่งใช้เฉพาะข้อมูลในอดีตของตัวแปรเป้าหมายอย่างมีนัยสำคัญ โดยมีค่า MAPE เท่ากับ 5.42% และ RMSE ที่ 4,298.65 การเพิ่มตัวแปรภายนอกไม่เพียงแต่ช่วยลดความคลาดเคลื่อนในการพยากรณ์ แต่ยังสามารถตามทันการเปลี่ยนแปลงที่เกิดจากปัจจัยภายนอกที่มีผลต่อความต้องการใช้ไฟฟ้าในแต่ละช่วงเวลาได้ดี ตัวอย่างเช่น แบบจำลอง SARIMAX สามารถรับมือกับฤดูกาลที่เกิดขึ้นประจำปี และความเปลี่ยนแปลงทางเศรษฐกิจหรือโครงสร้างประชากรที่มีผลระยะยาวได้ดีกว่า

ปี 2022 Wang และ คณะ (32) ได้ดำเนินการวิจัยเพื่อเปรียบเทียบศักยภาพของแบบจำลอง ARIMA, SARIMA และ Prophet สำหรับการพยากรณ์จำนวนผู้ติดเชื้อ COVID-19 ใน 3 ประเทศ ได้แก่ สหรัฐอเมริกา บราซิล และอินเดีย โดยวิเคราะห์ทั้งข้อมูลรายวัน (daily new cases) และข้อมูลสะสม (cumulative cases) เพื่อให้เห็นข้อเด่นข้อด้อยของแต่ละแบบจำลองในลักษณะข้อมูลที่แตกต่างกัน

จากผลการวิเคราะห์ในกลุ่มข้อมูลรายวันของสหรัฐอเมริกา พบว่าแบบจำลอง Prophet สามารถให้ค่า RMSE และ MAE ต่ำกว่า SARIMA อย่างชัดเจน โดยค่า RMSE เท่ากับ 67,842.84 ค่า MAE เท่ากับ 50,774.03 สะท้อนให้เห็นว่า Prophet มีความสามารถในการจับแนวโน้ม และฤดูกาลที่ซับซ้อนในข้อมูลที่ผันผวนได้อย่างมีประสิทธิภาพ อย่างไรก็ตามเมื่อเปรียบเทียบในข้อมูลสะสม (cumulative cases) โดยเฉพาะในบราซิล และอินเดีย พบว่าแบบจำลอง ARIMA กลับแสดงประสิทธิภาพที่เหนือกว่า Prophet อย่างมีนัยสำคัญ โดยให้ค่า RMSE และ MAE ต่ำกว่าเกือบทุกกรณี เช่นในประเทศบราซิล ARIMA มีค่าRMSE เท่ากับ 53,256.66, MAE เท่ากับ 38,376.99 ในขณะที่ Prophet มีค่าRMSE เท่ากับ 985,393.24, MAE เท่ากับ 784,833.60 ซึ่งแสดงให้เห็นถึงความเหมาะสมของ ARIMA สำหรับข้อมูลที่มีลักษณะเป็นแนวโน้มต่อเนื่อง ในงานวิจัยนี้ชี้ให้เห็นว่า การเลือกใช้แบบจำลองต้องคำนึงถึงลักษณะเฉพาะของข้อมูล และความถี่ของชุดข้อมูล เช่น ในกรณีของข้อมูลรายวัน Prophet มีข้อได้เปรียบในด้านความสามารถในการจัดการกับฤดูกาล และจุดเปลี่ยนแปลงของแนวโน้ม ในขณะที่ ARIMA เหมาะกับข้อมูลสะสมที่มีแนวโน้มระยะยาวต่อเนื่องมากกว่า

ปี 2023 Munim และคณะ (33)ได้นำเสนองานวิจัยที่มีวัตถุประสงค์เพื่อประเมินประสิทธิภาพของแบบจำลอง Prophet ในการพยากรณ์ปริมาณการขนส่งตู้คอนเทนเนอร์ (Container Throughput) ในท่าเรือหลักของเอเชีย ได้แก่ ท่าเรือเซี่ยงไฮ้ ประเทศจีน ท่าเรือปูซาน ประเทศเกาหลีใต้ และท่าเรือนาโกย่า ประเทศญี่ปุ่น โดยใช้ข้อมูลรายเดือนระหว่างปี ค.ศ. 2010–2019 เพื่อทำการพยากรณ์ระยะสั้นในช่วง 12 เดือนข้างหน้า (ปี ค.ศ. 2020) งานวิจัยได้นำแบบจำลอง Prophet มาเปรียบเทียบกับแบบจำลองการ

พยากรณ์อนุกรมเวลาอื่นๆ เช่น ARIMA, SARIMA และ Holt-Winter's Exponential Smoothing (HWES) รวมถึงแบบจำลองไฮบริด (Hybrid Models) ที่รวมวิธีการดังกล่าวเข้าด้วยกัน

จากการศึกษาผู้วิจัยพบว่าการพยากรณ์ข้อมูลภายในกลุ่มตัวอย่าง (In-sample Forecasting) แบบจำลอง Prophet มีประสิทธิภาพสูงสุดในการพยากรณ์ปริมาณตู้คอนเทนเนอร์ โดยมีค่า MAPE เฉลี่ยอยู่ในช่วง 1–4% ในทั้งสามท่าเรือที่ศึกษา ซึ่งแสดงให้เห็นถึงความสามารถที่ยอดเยี่ยมของ Prophet ในการจับแนวโน้ม และฤดูกาลในข้อมูล สำหรับการพยากรณ์ข้อมูลนอกกลุ่มตัวอย่าง (Out-of-sample Forecasting) แบบจำลองไฮบริดที่ผสมผสานระหว่าง ARIMA, SARIMA และ HWES ด้วยวิธีการเฉลี่ยแบบถ่วงน้ำหนัก แสดงผลลัพธ์ที่ดีที่สุด โดยมีค่าความผิดพลาดในการพยากรณ์ต่ำสุดเมื่อเปรียบเทียบกับแบบจำลองที่ใช้งานแบบเดี่ยว ซึ่งสะท้อนให้เห็นถึงข้อได้เปรียบของการพยากรณ์แบบผสมผสานที่สามารถดึงจุดเด่นของแต่ละแบบจำลองมารวมกัน ทั้งนี้งานวิจัยยังได้เสนอให้ใช้เกณฑ์ประเมินประสิทธิภาพการพยากรณ์ที่หลากหลายร่วมกัน เช่น MAPE, MAE, RMSE และ MSE เพื่อเพิ่มความแม่นยำ และความสอดคล้องของแบบจำลองที่นำไปใช้ในเชิงปฏิบัติได้จริงในอนาคต

ปี 2024 Pablo Negre และคณะ (34) ได้ศึกษาการพยากรณ์ความต้องการของสินค้าในตลาดรองเท้าแฟชั่นที่มีลักษณะความผันผวนตามฤดูกาลสูง โดยเปรียบเทียบประสิทธิภาพระหว่างแบบจำลอง Prophet และ SARIMA ซึ่งเป็นแบบจำลองทางสถิติที่นิยมใช้สำหรับข้อมูลที่มีรูปแบบตามฤดูกาลชัดเจน โดยงานวิจวิจัยนี้เน้นการพยากรณ์ในระยะยาว 12 เดือนล่วงหน้า และมีการประเมินผลลัพธ์โดยใช้ตัวชี้วัดสำคัญ (KPIs) หลายรายการ เช่น MAE, RMSE และ MAPE เพื่อวิเคราะห์และเปรียบเทียบประสิทธิภาพของทั้งสองแบบจำลองอย่างละเอียด

ผลการศึกษาพบว่า แบบจำลอง Prophet ซึ่งมีโครงสร้างไม่เป็นเชิงเส้น ให้ผลลัพธ์ในการพยากรณ์ยอดขายรองเท้ารวมรายปีแม่นยำถึง 98.8% ซึ่งสูงกว่าแบบจำลอง SARIMA ที่ให้ความแม่นยำที่ 93% อย่างไรก็ตาม เมื่อลงรายละเอียดถึงระดับรายเดือน พบว่า SARIMA กลับมีประสิทธิภาพดีกว่า Prophet อย่างชัดเจน โดยมีค่า MAE ต่ำกว่า (83.9 เทียบกับ 158.8) และใช้เวลาประมวลผลเร็วกว่าถึง 15 เท่า จึงแสดงให้เห็นว่า SARIMA เหมาะสำหรับการพยากรณ์ที่ต้องการรายละเอียดสูงรายเดือน ขณะที่ Prophet เหมาะกับการวางแผนระยะยาวประจำปีมากกว่า นอกจากนี้ งานวิจัยยังได้นำเสนอแนวคิดในการจัดโครงสร้างฐานข้อมูลแบบเชิงสัมพันธ์ (Relational Database) เพื่อรองรับการจัดเก็บผลลัพธ์จากการพยากรณ์อย่างเป็นระบบ ซึ่งถือเป็นแนวคิดใหม่ที่สามารถนำไปต่อยอดในการพัฒนาระบบการพยากรณ์ในอนาคตได้

ปี 2024 Mazaraki และคณะ (35)ได้ทำการศึกษาการคาดการณ์รายได้สำหรับโรงแรมในเครือนานาชาติ โดยมีวัตถุประสงค์เพื่อพัฒนาระบบการวิเคราะห์เชิงสมดุล (Balanced Scorecard - BSC) เพื่อประเมิน และคาดการณ์รายได้ของโรงแรมในเครือนานาชาติ ผู้วิจัยใช้ตัวชี้วัดประสิทธิภาพหลัก (Key Performance Indicators - KPIs) เพื่อวิเคราะห์ผลการดำเนินงานในด้านต่างๆ เช่น รายได้เฉลี่ยต่อห้อง (Average Daily Revenue per Room), อัตราการเข้าพัก (Occupancy Rate), ส่วนแบ่งตลาด (Market Share), ประสิทธิภาพแรงงาน (Staff Productivity), การสูญเสียทรัพยากร (Resource Intensity) และต้นทุนในการปรับเปลี่ยนสู่ระบบดิจิทัล (Digitalisation Costs)

จากการศึกษาผู้วิจัยพบว่าแบบจำลองการคาดการณ์ที่เหมาะสมที่สุดคือแบบจำลองความสัมพันธ์แบบยกกำลัง (Power Law Model) ซึ่งมีค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination) ที่ 0.89 แสดงให้เห็นถึงความสัมพันธ์ที่ชัดเจนของตัวแปรที่นำมาวิเคราะห์ โดยตัวแปรที่มีผลต่อกำไรของโรงแรมอย่างมีนัยสำคัญ ได้แก่ ต้นทุนเฉลี่ยต่อห้อง, อัตราการเข้าพัก, ส่วนแบ่งตลาด และต้นทุนด้านดิจิทัล ทั้งนี้ งานวิจัยเสนอว่าการใช้ระบบ Balanced Scorecard ที่มีการประเมินผลแบบหลายมิติทั้งตัวชี้วัดเชิงปริมาณ และคุณภาพ จะช่วยให้โรงแรมสามารถบริหารจัดการรายได้ ได้อย่างมีประสิทธิภาพ และสามารถคาดการณ์รายได้ในอนาคตได้แม่นยำมากขึ้น ส่งผลให้ผู้บริหารสามารถวางแผนกลยุทธ์ได้อย่างมีประสิทธิผลในสภาพแวดล้อมทางธุรกิจที่ผันผวน และมีการแข่งขันสูง

ปี 2024 Mane และคณะ (36) ได้นำเสนองานวิจัยเกี่ยวกับการพยากรณ์รายได้โดยใช้เทคนิคการเรียนรู้ของเครื่องแบบลำดับเวลา (Sequential Machine Learning) โดยเปรียบเทียบประสิทธิภาพของสองแบบจำลอง คือ Long Short-Term Memory (LSTM) และ Prophet บนชุดข้อมูลรายวันของบริษัทค้าปลีกในประเทศสหรัฐอเมริกา ระหว่างปี 2018 ถึง 2020 จากการศึกษาผู้วิจัยพบว่าแบบจำลอง LSTM มีประสิทธิภาพในการพยากรณ์ที่ดีกว่าแบบจำลอง Prophet โดยมีค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์เปอร์เซ็นต์ (Mean Absolute Percentage Error) เท่ากับ 12.493% และค่าความคลาดเคลื่อนรากที่สองของค่าความคลาดเคลื่อนเฉลี่ยกำลังสอง (Root Mean Squared Error) ที่ 839.40 ขณะที่แบบจำลอง Prophet มีค่า MAPE ที่ 14.41% และ RMSE ที่ 978.15 อย่างไรก็ตาม แบบจำลอง Prophet สามารถอธิบายแนวโน้ม และฤดูกาลของข้อมูลได้ละเอียดกว่า ทั้งแนวโน้มรายสัปดาห์ รายปี และผลกระทบของวันหยุดพิเศษ งานวิจัยนี้จึงแสดงให้เห็นถึงความสำคัญของการเลือกใช้แบบจำลองให้เหมาะสมกับลักษณะข้อมูล โดยแนะนำว่า LSTM เหมาะสำหรับข้อมูลที่เน้นความสัมพันธ์เชิงเวลาที่ชัดเจน ในขณะที่ Prophet มีจุดเด่นในการจัดการข้อมูลที่มีผลกระทบจากปัจจัยภายนอก เช่น วันหยุดต่าง ๆ ได้ดีกว่า

ปี 2024 คุณ Kwarteng และ Andreevich(37) ได้ศึกษาศักยภาพของแบบจำลองอนุกรมเวลาอย่างARIMA, SARIMA และ Prophet ในการพยากรณ์ความต้องการใช้ยาต้านเบาหวานในแต่ละเดือนในประเทศออสเตรเลีย โดยใช้ข้อมูลจริงจาก Medicare Australia ซึ่งมีลักษณะของข้อมูลที่ทั้งแนวโน้มเพิ่มขึ้นอย่างต่อเนื่อง และมีรูปแบบฤดูกาลที่ชัดเจน งานวิจัยนี้ให้ความสำคัญกับการประเมินความแม่นยำของแบบจำลองผ่านค่า MAE เป็นหลัก รวมทั้งพิจารณา RMSE และ MAPE เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลอง ผลการวิเคราะห์พบว่าแบบจำลอง Prophet ซึ่งเป็นแบบจำลองที่พัฒนาโดย Facebook มีค่า MAE ต่ำที่สุดที่ 0.74 ขณะที่ SARIMA มี MAE เท่ากับ 2.18 และ ARIMA มีค่า 3.02 ตามลำดับ นอกจากนี้ Prophet ยังแสดงศักยภาพเหนือกว่าในแง่ของการจัดการกับแนวโน้มที่ไม่เป็นเชิงเส้น (non-linear trend) และรูปแบบฤดูกาลที่อาจมีการเปลี่ยนแปลง หรือซับซ้อนมากกว่าที่แบบจำลองเชิงสถิติดั้งเดิมอย่าง ARIMA และ SARIMA สามารถรองรับได้

จากการศึกษางานวิจัยนี้ พบว่าแบบจำลอง Prophet มีความเหมาะสมกับชุดข้อมูลที่มีความซับซ้อนทั้งในแง่ของแนวโน้ม และฤดูกาล สามารถนำไปใช้ในงานพยากรณ์จริงในบริบทที่ข้อมูลมีความเปลี่ยนแปลงหลากหลาย ตัวอย่างเช่น การพยากรณ์อุปสงค์ยาในภาคสาธารณสุข ซึ่งอาจได้รับอิทธิพลจากปัจจัยภายนอก หรือพฤติกรรมของผู้บริโภคที่เปลี่ยนแปลงตลอดเวลา

ปี 2025 คุณ Lee Geun-Cheol (38) ได้ดำเนินการวิจัยเพื่อพัฒนาแบบจำลองการพยากรณ์จำนวนนักท่องเที่ยวขาเข้ารายเดือนในประเทศสิงค์โปร์ เพื่อสนับสนุนการวางแผนเชิงกลยุทธ์ และการจัดสรรทรัพยากรอย่างมีประสิทธิภาพ โดยเฉพาะในบริบทหลังวิกฤต COVID-19 ที่พฤติกรรม และปริมาณนักท่องเที่ยวที่เปลี่ยนแปลงอย่างรวดเร็ว จากการทดลองทดสอบเปรียบเทียบแบบจำลอง SARIMAX กับแบบจำลองแบบดั้งเดิมอย่าง SARIMA, Holt–Winters, Prophet และแบบจำลองเชิงลึกเช่น LSTM และ RNN พบว่า SARIMAX สามารถให้ค่า MAPE ต่ำสุดที่ 7.32% ในการพยากรณ์จำนวนนักท่องเที่ยวระยะเวลา 24 เดือน (ปี 2023–2024) ซึ่งดีกว่าแบบจำลองอื่น ๆ อย่างชัดเจน จากการศึกษางานวิจัยยังพบว่าแบบจำลอง Prophet ซึ่งเป็นแบบจำลองสมัยใหม่ที่ได้รับความนิยม กลับให้ผลลัพธ์ไม่เป็นอย่างที่คาด โดยเฉพาะในช่วงที่มีวิกฤตการณ์อย่าง COVID-19 เนื่องจาก Prophet ให้ความสำคัญกับข้อมูลระยะสั้นมากเกินไป จึงไม่สามารถจับแนวโน้มฟื้นตัวของนักท่องเที่ยวหลังวิกฤตได้อย่างเหมาะสม

บทที่ 3
การดำเนินการวิจัย

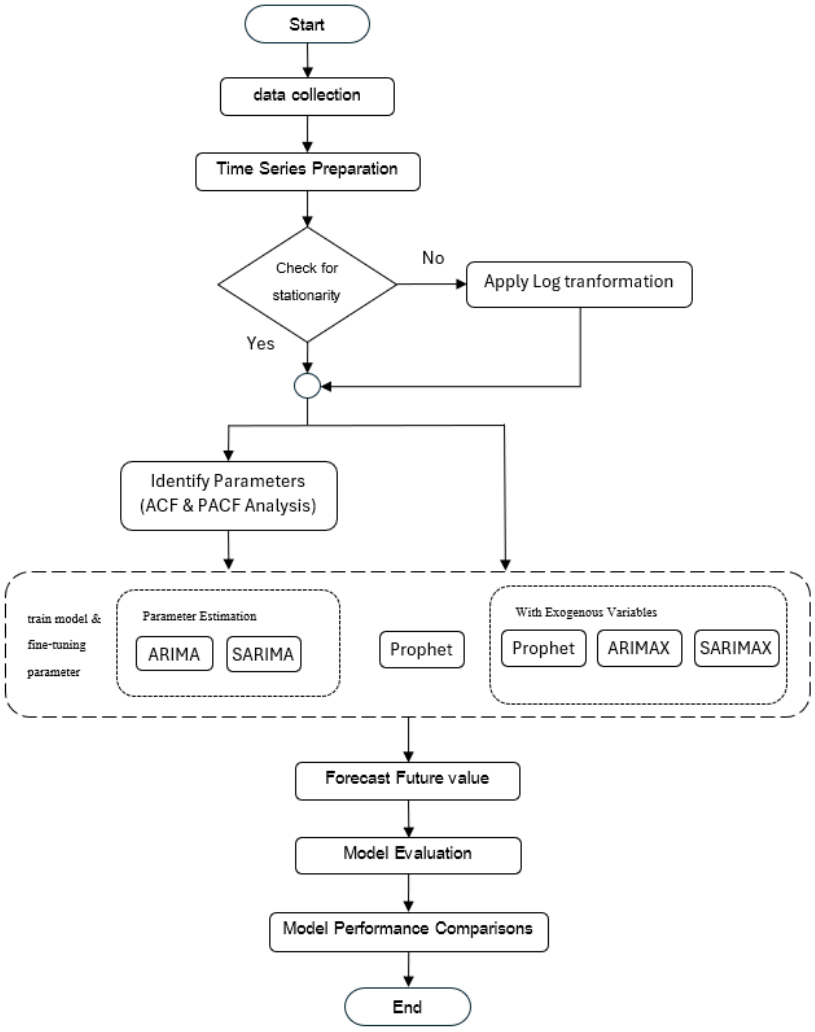
ในงานวิจัยนี้ผู้วิจัยได้ดำเนินการตามขั้นตอนอย่างเป็นระบบ โดยแบ่งออกเป็น 5 ขั้นตอนหลัก ดังนี้

1. การสร้างแบบจำลอง
2. การรวบรวมชุดข้อมูล (Data Collection)
3. การตรวจสอบข้อมูลเชิงลึก (Exploratory Data Analysis: EDA)
4. การเตรียมข้อมูล (Data Preparation)
5. การสร้าง และประเมินแบบจำลองพยากรณ์

การสร้างแบบจำลอง

การพยากรณ์รายได้ของโรงแรมในงานวิจัยนี้ เริ่มต้นจากการกำหนดประเด็นปัญหา และวัตถุประสงค์ของการศึกษา โดยมุ่งเน้นไปที่การพยากรณ์จำนวนห้องพักที่ขายได้ (Room Sold) ซึ่งเป็นตัวแปรสำคัญที่สะท้อนศักยภาพในการสร้างรายได้หลักของโรงแรมโดยตรง ทั้งนี้งานวิจัยเลือกที่จะไม่คาดการณ์รายได้โดยตรง จากการนำค่าพยากรณ์ของจำนวนห้องพักไปคูณกับอัตราค่าห้องพักเฉลี่ยรายวัน (ADR) เนื่องจากต้องการควบคุมความแปรปรวนของตัวแปร ADR ซึ่งมีความผันผวนสูงจากปัจจัยภายนอกต่างๆ เช่น ช่วงเทศกาล, โปรโมชั่น และฤดูกาลท่องเที่ยว เพื่อเน้นความถูกต้องแม่นยำและประสิทธิภาพของแบบจำลองในการพยากรณ์จำนวนห้องพักที่ถูกจอง โดยแบบจำลองหลักที่นำมาใช้ในการศึกษานี้คือ Prophet ซึ่งเป็นแบบจำลองที่มีประสิทธิภาพสูงในการจัดการข้อมูลอนุกรมเวลาที่มีลักษณะเป็นแนวโน้ม และมีรูปแบบตามฤดูกาลชัดเจน มาเปรียบเทียบกับแบบจำลองทางสถิติแบบดั้งเดิมที่เป็นที่นิยมใช้อย่าง แบบจำลอง ARIMA และ SARIMA ตลอดจนแบบจำลองที่สามารถรองรับปัจจัยเสริมจากภายนอกอย่างแบบจำลอง ARIMAX และ SRIMAX เพื่อให้ทราบถึงจุดเด่นข้อจำกัด และประสิทธิภาพของแบบจำลอง โดยผลลัพธ์ที่ได้จากการพยากรณ์ จะถูกนำไปสนับสนุนการตัดสินใจเชิงกลยุทธ์ในการบริหารจัดการรายได้ของโรงแรมในอนาคตต่อไป

ภาพประกอบแสดงกระบวนการสร้างแบบจำลอง



ภาพประกอบ 3 ภาพแสดงกระบวนการสร้างแบบจำลองการพยากรณ์จำนวนห้องพักที่ขายได้

จากภาพประกอบที่ 3 การพัฒนาแบบจำลองเพื่อใช้ในการพยากรณ์รายได้ ถูกออกแบบอย่างเป็นระบบ โดยมีวัตถุประสงค์เพื่อเปรียบเทียบศักยภาพของแบบจำลองอนุกรมเวลาหลายประเภทในการพยากรณ์จำนวนห้องพักที่ขายได้ ซึ่งถือเป็นตัวแปรสำคัญในการคาดการณ์รายได้ของโรงแรมและการบริหารรายได้ของโรงแรม กระบวนการทั้งหมดสามารถสรุปเป็นขั้นตอนสำคัญได้ดังนี้

1. การเก็บรวบรวมข้อมูล (Data Collection) ในงานวิจัยนี้ใช้ชุดข้อมูล Hotel Booking Demand Dataset ซึ่งครอบคลุมตัวแปรสำคัญที่ส่งผลต่อรายได้ของโรงแรม เช่นจำนวนคืนที่พัก ช่วงเวลาที่ทำการจอง จำนวนแขกที่เข้าพัก และรายละเอียดอื่น ๆ ที่อาจสามารถสนับสนุนการสร้างแบบจำลองที่มีความแม่นยำยิ่งขึ้น

2. การเตรียมข้อมูลอนุกรมเวลา (Time Series Preparation) ข้อมูลจะถูกนำมาผ่านกระบวนการเตรียมความพร้อม ได้แก่ การตรวจสอบ และเติมค่าที่ขาดหาย การลบข้อมูลผิดปกติ และการปรับโครงสร้างข้อมูลให้อยู่ในรูปแบบที่เหมาะสมต่อการวิเคราะห์อนุกรมเวลา

3. การตรวจสอบความคงที่ของข้อมูล (Check for Stationarity) ในขั้นตอนนี้จะใช้วิธีทางสถิติ Augmented Dickey-Fuller (ADF) Test เพื่อประเมินว่าข้อมูลมีค่าเฉลี่ย และความแปรปรวนคงที่หรือไม่ ซึ่งเป็นเงื่อนไขสำคัญสำหรับการใช้แบบจำลองเชิงสถิติ หากข้อมูลยังไม่ stationary จะดำเนินการปรับเปลี่ยนข้อมูลต่อไป

4. การแปลงข้อมูล (Log Transformation) หากผลการทดสอบพบว่าข้อมูลไม่ stationary จะมีการปรับเปลี่ยนข้อมูลด้วยฟังก์ชันลอการิทึม (Log Transformation) เพื่อช่วยลดความผันผวน และสร้างเสถียรภาพให้กับข้อมูล ก่อนนำไปวิเคราะห์ในขั้นตอนถัดไป

5. การระบุค่าพารามิเตอร์ (*Identify Parameters: ACF & PACF Analysis*) ใช้เทคนิคการวิเคราะห์ความสัมพันธ์ภายในข้อมูล ได้แก่ Autocorrelation Function (ACF) และ Partial Autocorrelation Function (PACF) เพื่อกำหนดค่าพารามิเตอร์ที่เหมาะสมของแต่ละแบบจำลอง เช่น ARIMA กำหนดค่าของ p ลำดับของตัวแปรเชิงอัตโนมัติ (Autoregressive), ค่า d ระดับของการ Differencing ที่ไม่ใช่เชิงฤดูกาล และค่า q ลำดับของตัวแปรค่าเฉลี่ยเคลื่อนที่ (Moving Average) ในส่วนของแบบจำลอง SARIMA ซึ่งมีพารามิเตอร์ในส่วนของฤดูกาล P, D, Q และ m ซึ่งความยาวของรอบฤดูกาลเพิ่มเข้ามา

6. การสร้าง และปรับแต่งแบบจำลอง (Model Building & Parameter Estimation) ในขั้นตอนนี้ดำเนินการฝึกฝน (Train) และปรับแต่ง (Tune) แบบจำลองแต่ละประเภทให้เหมาะสมกับข้อมูลของโรงแรม โดยแบบจำลองARIMA และ SARIMA จะเป็นการประเมินพารามิเตอร์จากข้อมูลในอดีต สำหรับแบบจำลอง Prophet จะเป็นการแยกวิเคราะห์องค์ประกอบแนวโน้ม (Trend) ฤดูกาล (Seasonality) และ Changepoint ในส่วนของแบบจำลองARIMAX, SARIMAX และ Prophet with Exogenous Variables จะมีการนำตัวแปรภายนอกที่มีส่วนช่วยให้ประสิทธิภาพของแบบจำลองดีขึ้น เช่น ฤดูกาลท่องเที่ยว หรืออัตราค่าห้องพักเฉลี่ย เพื่อเสริมความแม่นยำของแบบจำลอง

7. การพยากรณ์ค่าในอนาคต (Forecast Future Value) นำแบบจำลองที่ผ่านการฝึกฝนมาทำนายจำนวนห้องพักที่คาดว่าจะขายได้ในช่วงเวลาข้างหน้า ซึ่งข้อมูลที่พยากรณ์ได้นี้สามารถนำไปแปลงกลับจาก log scale เพื่อให้สอดคล้องกับข้อมูลในเชิงธุรกิจจากนั้นจึงนำไปคูณกับค่า ADR ซึ่งเป็นอัตราค่าห้องพักเฉลี่ยในช่วงเวลานั้น เพื่อคำนวณรายได้ที่คาดว่าจะได้รับ และสามารถนำข้อมูลดังกล่าว

ไปสะสมเป็นยอดรายได้รวม (Cumulative Revenue) เพื่อวิเคราะห์แนวโน้ม และประเมินศักยภาพในการบรรลุเป้าหมายรายได้

8. การประเมินประสิทธิภาพของแบบจำลอง (Model Evaluation) ในส่วนนี้จะประเมินความแม่นยำของแบบจำลองโดยใช้ตัวชี้วัด อย่าง MAE, RMSE และ MAPE เพื่อเปรียบเทียบผลลัพธ์ที่ได้จากแบบจำลองแต่ละประเภทอย่างเป็นรูปธรรม

9. การเปรียบเทียบประสิทธิภาพแบบจำลอง (Model Performance Comparisons) นำผลการประเมินของแบบจำลองแต่ละประเภทมาเปรียบเทียบในเชิงประจักษ์ ทั้งในด้านความแม่นยำ การตอบสนองต่อฤดูกาล และความสามารถในการประมวลผลตัวแปรภายนอก เพื่อเลือกแนวทางที่เหมาะสมที่สุดสำหรับบริบทของธุรกิจโรงแรม

ขั้นตอนทั้งหมดนี้ถูกออกแบบให้เป็นระบบ และครอบคลุมทั้งกระบวนการสร้าง เตรียมข้อมูล สร้างแบบจำลอง ทดสอบ และประเมินผล เพื่อให้ได้แนวทางการพยากรณ์ที่เหมาะสมกับลักษณะข้อมูล และบริบทของธุรกิจโรงแรม ซึ่งสามารถนำไปประยุกต์ใช้ในการวางแผนกลยุทธ์ และการบริหารจัดการรายได้ในโลกความเป็นจริงได้อย่างมีประสิทธิภาพ

การเก็บรวบรวมข้อมูล (Data Collection)

ในการวิจัยครั้งนี้ ข้อมูลที่ใช้ในการพยากรณ์รายได้ของโรงแรม นำมาจากชุดข้อมูลสาธารณะบนเว็บไซต์ Kaggle ชื่อชุดข้อมูล " Hotel Booking Demand Dataset " ซึ่งเผยแพร่โดย Antonio et al. (2019) (5) โดยข้อมูลดังกล่าวครอบคลุมการจองห้องพักของโรงแรมสองประเภท ได้แก่ Resort Hotel และ City Hotel โดยเป็นข้อมูลในช่วงระหว่างเดือนกรกฎาคม ค.ศ. 2018 ถึงเดือนสิงหาคม ค.ศ. 2020 ในชุดข้อมูลนี้ประกอบไปด้วย 31 คอลัมน์ และ 141,947 แถว ซึ่งครอบคลุมตัวแปรสำคัญที่ส่งผลต่อรายได้ของโรงแรม ตัวอย่างเช่น ประเภทของโรงแรม จำนวนคืนที่พัก ช่วงเวลาที่ทำการจอง จำนวนแขกที่เข้าพัก อัตราค่าห้องพักเฉลี่ยรายวัน และรายละเอียดอื่น ๆ ข้อมูลนี้จะถูกนำมาประยุกต์ใช้ในการสร้างแบบจำลองพยากรณ์จำนวนห้องพักที่ขายได้ เพื่อนำไปคาดการณ์รายได้ของโรงแรม ซึ่งเป็นวัตถุประสงค์หลักของการวิจัยในครั้งนี้ รายละเอียดของข้อมูลจะแสดงดังในตารางที่ 2

ตาราง 2 ตารางแสดงตัวแปรในข้อมูล

Variable	Description
hotel	ประเภทของโรงแรม เช่น Resort Hotel, City Hotel
is_canceled	สถานะการยกเลิกการจอง (0 = ไม่ยกเลิก, 1 = ยกเลิก)
lead_time	จำนวนวันที่มีการจองล่วงหน้าก่อนวันที่เข้าพัก
arrival_date_year	ปีที่ลูกค้าเข้าพัก
arrival_date_month	เดือนที่ลูกค้าเข้าพัก
arrival_date_week_number	สัปดาห์ที่ลูกค้าเข้าพัก
arrival_date_day_of_month	วันที่ลูกค้าเข้าพัก
stays_in_weekend_nights	จำนวนคืนที่เข้าพักในวันสุดสัปดาห์
stays_in_week_nights	จำนวนคืนที่เข้าพักในวันธรรมดา
adults	จำนวนผู้ใหญ่ที่เข้าพัก
children	จำนวนเด็กที่เข้าพัก
babies	จำนวนทารกที่เข้าพัก
meal	แผนการอาหาร เช่น BB (Bed & Breakfast), HB (Half Board), FB (Full Board), SC (Special Cost)
country	ประเทศของลูกค้า
is_repeated_guest	ลูกค้าเคยเข้าพักที่โรงแรมนี้มาก่อนหรือไม่ (0 = ไม่เคย, 1 = เคย)
previous_cancellations	จำนวนการจองที่เคยยกเลิกก่อนหน้านี้
previous_bookings_not_canceled	จำนวนการจองที่ไม่เคยยกเลิกก่อนหน้านี้
reserved_room_type	ประเภทห้องที่ลูกค้าจอง
assigned_room_type	ประเภทห้องที่โรงแรมจัดสรรให้
booking_changes	จำนวนการเปลี่ยนแปลงที่ลูกค้าทำกับการจอง
days_in_waiting_list	จำนวนวันที่ลูกค้ารอในรายชื่อสำรอง
market segment	ช่องทางการตลาดที่ลูกค้าจองห้องพัก เช่น Corporate ลูกค้าองค์กร หรือบริษัท Online TA ลูกค้าที่ซื้อผ่านตัวแทนท่องเที่ยวออนไลน์, Online TA/TO ลูกค้าที่ซื้อผ่านตัวแทนท่องเที่ยวแบบออฟไลน์, Direct ลูกค้าที่ซื้อโดยตรง ไม่ผ่านตัวแทน, Groups กลุ่มลูกค้าที่ซื้อเป็นกลุ่ม เช่น กลุ่มทัวร์, Aviation ลูกค้าที่เกี่ยวข้องกับสายการบิน, Complementary ลูกค้าที่ได้รับส่วนลดเพิ่มเติม หรือเป็นโปรโมชั่นพิเศษ Undefined ไม่มีการกำหนดกลุ่มลูกค้า หรืออาจเป็นข้อมูลที่ขาดหายไป,

ตารางที่ 2 ตารางแสดงตัวแปรในข้อมูล (ต่อ)

Variable	Description
agent	รหัสตัวแทนการจอง
company	รหัสบริษัทที่เกี่ยวข้องกับการจอง (ถ้ามี)
deposit type	ประเภทการฝากเงิน เช่น
	No Deposit ไม่มีการวางมัดจำล่วงหน้าในการจองห้องพัก,
	Refundable Deposit ต้องชำระเงินเต็มจำนวน ไม่สามารถขอคืนเงินได้
	หากทำการยกเลิกการจอง
	Refundable ต้องวางเงินมัดจำในการจองห้องพัก แต่สามารถขอเงินคืน
	ได้หากทำการยกเลิกภายในเวลาที่กำหนด
customer_type	ประเภทของลูกค้า เช่น Transient ลูกค้าที่จองห้องพักสำหรับเข้าพักระยะสั้น
	Transient-Party ลูกค้าที่จองห้องพักเป็นกลุ่มขนาดเล็ก
	Contract ลูกค้าที่จองห้องพักภายใต้สัญญา หรือข้อตกลงระยะยาวกับโรงแรม
	Group ลูกค้าที่จองห้องพักเป็นกลุ่มใหญ่
Adr	อัตราค่าห้องพักเฉลี่ยต่อวัน (average daily rate)
required_car_parking_spaces	จำนวนที่จอดรถที่ลูกค้าต้องการ
total_of_special_requests	จำนวนคำขอพิเศษที่ลูกค้าร้องขอ
reservation_status	สถานะการจอง เช่น Canceled สถานะการจองถูกยกเลิก,
	Check-Out เช็กเอาท์, No-Show ลูกค้าที่ไม่ได้เข้าพักตามที่ได้จองไว้ และไม่ได้แจ้งยกเลิก
reservation_status_date	วันที่เกิดการเปลี่ยนแปลงสถานะการจอง

การสำรวจข้อมูล Exploratory Data Analysis (EDA)

การวิเคราะห์ข้อมูลเชิงสำรวจ (EDA) เป็นขั้นตอนสำคัญในการวิจัยนี้ มีวัตถุประสงค์เพื่อให้เข้าใจโครงสร้าง คุณลักษณะ และรูปแบบที่ซ่อนอยู่ในชุดข้อมูล ตลอดจนการตรวจสอบคุณภาพของข้อมูลก่อนที่จะเข้าสู่กระบวนการสร้างแบบจำลองพยากรณ์ โดยจะทำการสำรวจข้อมูล และวิเคราะห์ข้อมูลของโรงแรมตั้งแต่เดือนกรกฎาคม ค.ศ. 2018 ถึงเดือนสิงหาคม ค.ศ. 2020 ข้อมูลนี้ประกอบไปด้วย 31 คอลัมน์ และ 141,947 แถว ซึ่งเป็นข้อมูลเกี่ยวข้องการจองห้องพักของโรงแรม

ก่อนเริ่มกระบวนการ EDA จะต้องทำการเตรียมข้อมูล โดยทำการรวมข้อมูลในปี 2018 2019 และ 2020 พร้อมทั้งนำข้อมูลเสริมจากตาราง meal_cost และ market_segment มารวมให้เป็นข้อมูลเดียวกัน เพื่อเพิ่มมิติของข้อมูลในการวิเคราะห์ หลังจากนั้นจึงทำการตรวจสอบค่าว่างที่อยู่ในข้อมูล

เพื่อให้สามารถกำหนดวิธีการจัดการ เช่น การแทนค่าด้วยค่าเฉลี่ย การลบแถวที่ข้อมูลไม่สมบูรณ์ หรือการใช้วิธีการ Imputation ตามความเหมาะสม เพื่อป้องกันผลกระทบต่อการวิเคราะห์และการสร้างแบบจำลองในขั้นตอนถัดไป แสดงตัวอย่างข้อมูลดังภาพประกอบ

children	8	previous_bookings_not_canceled	0
babies	0	reserved_room_type	0
meal	0	assigned_room_type	0
country	625	booking_changes	0
market_segment	0	deposit_type	0
distribution_channel	0	agent	19555
is_repeated_guest	0	company	133822

ภาพประกอบ 4 ภาพแสดงค่าว่างของข้อมูล

จากภาพประกอบ 4 พบว่าในคอลัมน์ children มีค่าว่างทั้งหมด 8 ค่า country มีค่าว่าง 625 agent มีค่าว่าง 19,555 และcompany มีค่าว่าง 133,822 จึงต้องทำการกรองข้อมูล และทำการสร้างตัวแปรใหม่เพื่อใช้ในการพยากรณ์ จากการวิเคราะห์ข้อมูลเพิ่มเติมพบรายละเอียดของข้อมูลมีดังนี้

1. ชุดข้อมูลประกอบด้วย 141,123 แถว และ 31 คอลัมน์ มีประเภทโรงแรมทั้งหมด 2 ประเภท คือ โรงแรมรีสอร์ท (Resort Hotel) และโรงแรมในเมือง (City Hotel)

2. ข้อมูลช่วงเวลาการเข้าพักของลูกค้ามีตั้งแต่ปี 2018 ถึง 2020 มีบางแถวในคอลัมน์ children และ country ที่มีค่า missing จึงทำการลบแถวที่ไม่มีข้อมูลสำคัญออก

3. ในส่วนของคอลัมน์ agent ที่มีค่าว่าง 19,555 และ company ที่มีค่าว่าง 133,822 เนื่องจากว่าเป็นข้อมูลที่ไม่ส่งผลต่อประสิทธิภาพของแบบจำลอง และมีค่าว่างค่อนข้างมาก จึงทำการลบคอลัมน์ทั้งสองออกจากชุดข้อมูล

4. พบว่าในคอลัมน์ adr (อัตราค่าห้องพักต่อวัน) มีบางแถวที่มีค่าเป็น 0 ซึ่งอาจเกิดจากการยกเลิกการจอง หรือข้อผิดพลาดในการบันทึกข้อมูล จึงทำการกรองแถวที่มีค่า adr เป็น 0 ออกไป

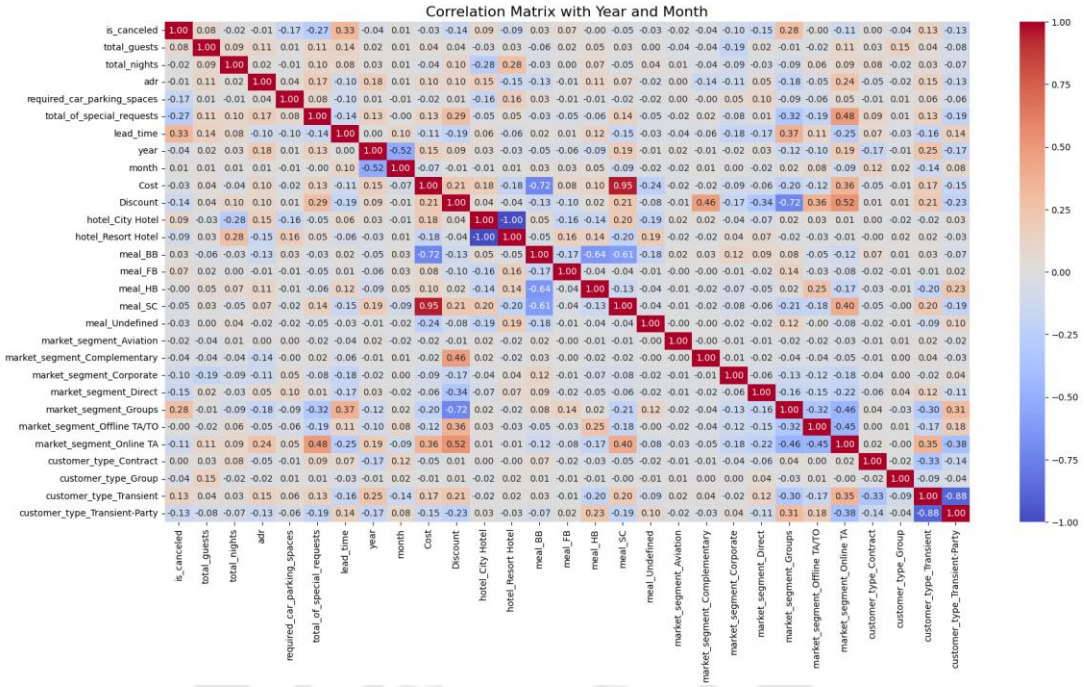
ตาราง 3 ตารางแสดงรายละเอียดประเภทของข้อมูลแต่ละคอลัมน์

No.	Variable	Description
1	hotel	object
2	is_canceled	int64
3	lead_time	int64
4	arrival_date_year	int64
5	arrival_date_month	object
6	arrival_date_week_number	int64
7	arrival_date_day_of_month	int64
8	stays_in_weekend_nights	int64
9	stays_in_week_nights	int64
10	adults	int64
11	children	float64
12	babies	int64
13	meal	object
14	country	object
15	market_segment	object
16	is_repeated_guest	int64
17	previous_cancellations	int64
18	previous_bookings_not_canceled	int64
19	reserved_room_type	object
20	assigned_room_type	object
21	booking_changes	int64
22	deposit_type	object
23	days_in_waiting_list	int64
24	customer_type	object
25	adr	float64
26	required_car_parking_spaces	int64
27	total_of_special_requests	int64
28	reservation_status	object
29	reservation_status_date	datetime64[ns]
30	Cost	float64
31	Discount	float64

เพื่อให้เข้าใจลักษณะพื้นฐานของข้อมูล ตลอดจนรูปแบบการกระจายตัวของตัวแปรสำคัญ และความสัมพันธ์ระหว่างตัวแปรต่าง ๆ อย่างชัดเจน การวิเคราะห์ข้อมูลเชิงสำรวจ (Exploratory Data Analysis) จึงเป็นขั้นตอนที่ขาดไม่ได้ในงานวิจัยนี้ โดยจะใช้สถิติเชิงพรรณา และการนำเสนอข้อมูลผ่านกราฟ และแผนภูมิต่าง ๆ เช่น Heatmap, Histogram, Boxplot และกราฟเส้น เพื่อช่วยในการค้นหาแนวโน้มเบื้องต้น รูปแบบการเปลี่ยนแปลงตามเวลา (Trend) ฤดูกาล (Seasonality) รวมถึงตรวจสอบความสัมพันธ์ หรือความเชื่อมโยงระหว่างตัวแปรที่อาจมีผลต่อผลลัพธ์การพยากรณ์

การดำเนินการ EDA นี้ นอกจากจะช่วยให้มองเห็นภาพรวม และจุดเด่นของข้อมูลได้อย่างครอบคลุมแล้ว ยังเป็นแนวทางสำคัญในการคัดเลือกคุณลักษณะ (Feature Selection) และการปรับปรุงข้อมูลก่อนเข้าสู่กระบวนการสร้างแบบจำลองพยากรณ์ต่อไป

ทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรสำคัญในการพยากรณ์รายได้ของโรงแรม โดยใช้ Heat Map ช่วยให้เราเข้าใจพฤติกรรมของผู้เข้าพัก และปัจจัยที่ส่งผลต่อการจองเข้าใช้บริการโรงแรม รายละเอียดข้อมูล แสดงตัวอย่างดังภาพประกอบที่ 5



ภาพประกอบ 5 ภาพแสดงความสัมพันธ์ระหว่างตัวแปรสำคัญในการพยากรณ์รายได้ของโรงแรม

จากผลลัพธ์ของความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ที่แสดงใน Heatmap มีข้อสังเกตที่สำคัญดังนี้

1. is_canceled กับ lead_time มีความสัมพันธ์ทางบวกที่ค่อนข้างสูง (+0.33) ซึ่งแสดงว่าการจองที่มี Lead Time หรือช่วงเวลาล่วงหน้าเป็นเวลานานมีผลอัตราการยกเลิกการจองห้องพัก จากผลลัพธ์ยังพบว่าการจองระยะยาวมักเป็นของ Group Booking โดยมีค่าความสัมพันธ์ (+0.37) มีโอกาสถูกยกเลิกสูง

2. adr กับ market_segment_Online TA มีความสัมพันธ์ทางบวก (+0.24) แสดงว่าการจองผ่านช่องทาง Online Travel Agents (OTA) มีส่วนเกี่ยวข้องกับการที่โรงแรมมีค่า ADR สูงขึ้น

3. Discount กับ special request มีความสัมพันธ์ทางบวก (+0.29) หมายความว่าการจองที่มีการจองที่มีการระบุคำขอพิเศษ มักจะมีการใช้ส่วนลด

4. special request market_segment_Online TA มีความสัมพันธ์ทางบวก (+0.48) ลูกค้าที่จองผ่าน Online Travel Agents (OTA) มักมีความคาดหวังและทำ Special Requests มากกว่า เนื่องจากOTA อาจมีฟีเจอร์ที่ทำให้ลูกค้ากรอก Special Requests ได้สะดวก

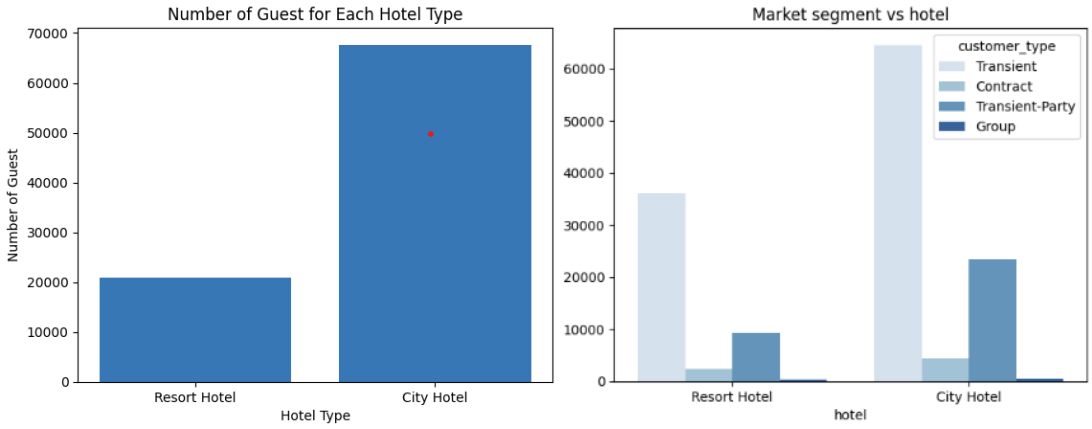
5. hotel_Resort Hotel กับ total_nights มีความสัมพันธ์ทางบวก (+0.28) จากข้อมูลพบว่าโรงแรมรีสอร์ทมักถูกจองในระยะเวลาที่นานกว่าลูกค้าเข้าพักหลายคืนเพื่อการพักผ่อน ต่างจากโรงแรมในเมืองที่ลูกค้ามักจะเข้าพักระยะสั้น ซึ่งมีความสัมพันธ์ทางลบ (-0.28)

6. Lead Time กับ market_segment_Groups มีความสัมพันธ์ทางบวก (+0.37) กลุ่มลูกค้าประเภท Groups มักจองล่วงหน้านาน เพราะการจองแบบกลุ่มต้องมีการวางแผนมากกว่าการจองแบบรายบุคคล

7. market_segment_Online TA กับ Discount มีความสัมพันธ์ทางบวก (+0.52) ลูกค้าที่จองผ่าน Online Travel Agents มักได้รับส่วนลดพิเศษ เพราะเป็นช่องทางที่มักมีโปรโมชั่นพิเศษทั้งจากเว็บไซต์ และจากโรงแรม เพื่อดึงดูดลูกค้า

8. market_segment_Offline TA/TO กับ Discount มีความสัมพันธ์ทางบวก (+0.36) เอเจนซี่ท่องเที่ยวออฟไลน์ (TA/TO) อาจได้รับส่วนลดจากโรงแรมเพื่อดึงดูดลูกค้า

ทำการวิเคราะห์หาจำนวนนักท่องเที่ยวที่เข้าใช้บริการโรงแรมทั้งสองประเภท

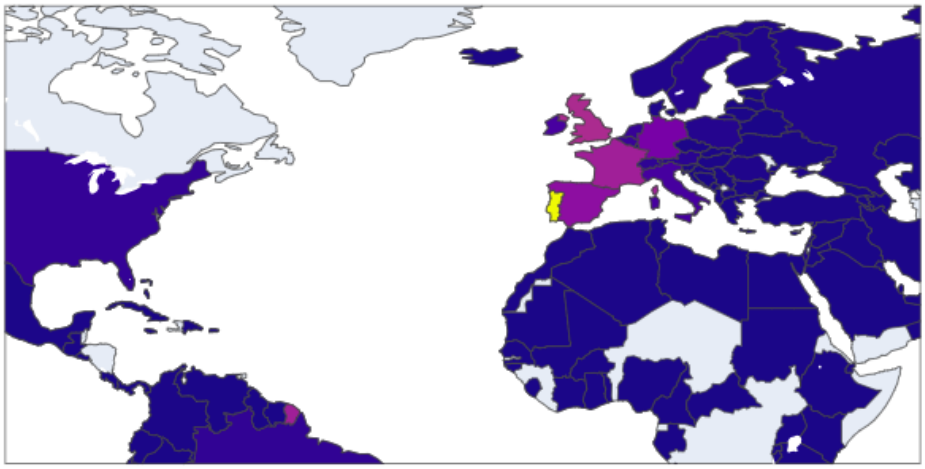


(a) (b)

ภาพประกอบ 6 ภาพแสดงประเภทของโรงแรม และจำนวนแขกในแต่ละประเทศที่เข้าพัก

จากภาพประกอบที่ 6 ภาพประกอบ (a) แสดงให้เห็นจำนวนผู้เข้าใช้บริการโรงแรมในเมืองมีปริมาณมากกว่าโรงแรมประเภทรีสอร์ท นอกจากนี้ ภาพประกอบ (b) ยังแสดงให้เห้นว่าลูกค้าที่จองห้องพักสำหรับการเข้าพักระยะสั้น นั้นมีปริมาณมากที่สุด รองลงมาคือลูกค้าที่จองเข้าพักระยะสั้นแบบเป็นกลุ่ม ตามมาด้วยลูกค้าที่มาจากการทำสัญญากับโรงแรม และสุดท้ายคือกลุ่มขนาดใหญ่

ทำการวิเคราะห์หาว่าประเทศใดเข้ารับบริการพักที่โรงแรมเยอะที่สุด



ภาพประกอบ 7 ภาพแสดงประเทศที่เข้าใช้บริการในโรงแรม

จากข้อมูลที่แสดงในภาพประกอบ 7 พบว่าประเทศที่เข้าพักมากที่สุดเป็นอันดับ 1 คือประเทศโปรตุเกส โดยจำนวนที่เข้าพักมี 14231 ห้อง ในส่วนอันดับ 2 คือประเทศฝรั่งเศสที่มีจำนวนที่เข้าพักทั้งหมด 5364 อันดับ 3 คือประเทศอังกฤษ ที่มีจำนวนที่เข้าพักทั้งหมด 5220 ห้อง อันดับ 4 คือประเทศเยอรมันนี ที่มีจำนวนที่เข้าพักทั้งหมด 4407 ห้อง และอันดับที่ 5 คือประเทศสเปน ที่มีจำนวนที่เข้าพักทั้งหมด 4323 ห้องซึ่งเป็นไปตามที่คาดไว้เนื่องจากข้อมูลมาจากโรงแรมที่ตั้งในประเทศโปรตุเกส ดังนั้นผู้ใช้บริการจากประเทศอื่น ๆ จึงเป็นประเทศจากทวีปยุโรปในฝั่งตะวันตกที่ใกล้กับประเทศโปรตุเกส

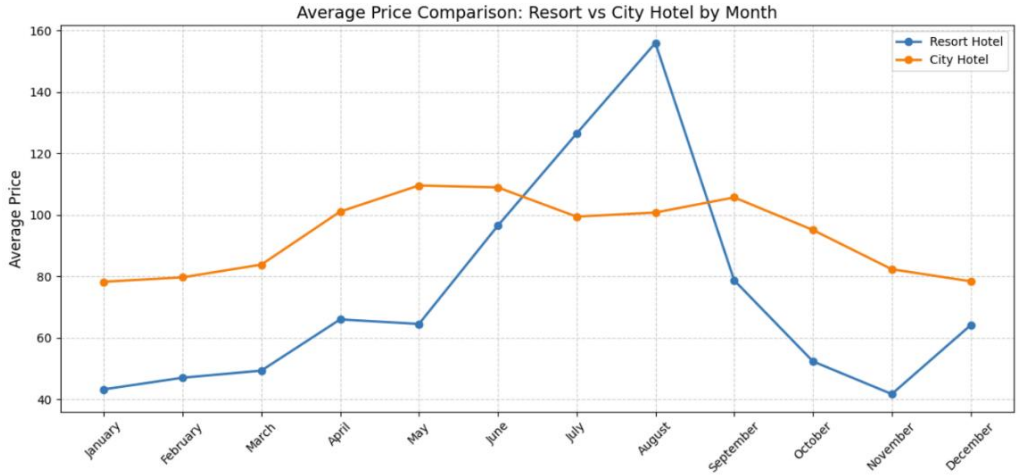
ทำการวิเคราะห์หาแนวโน้มของของอัตราค่าที่พักต่อหนึ่งคืน (ADR) ว่ามีการเปลี่ยนแปลงไปอย่างไรในแต่ละวัน

ภาพประกอบ 8 ภาพแสดงแนวโน้มของอัตราค่าที่พักเฉลี่ยรายวัน (ADR) ในโรงแรมแต่ละประเภท

จากภาพประกอบที่ 8 แสดงแนวโน้มอัตราค่าห้องพักเฉลี่ยรายวัน (ADR) ตามช่วงเวลา โดยจำแนกตามประเภทโรงแรม พบว่าโรงแรมรีสอร์ท มีความผันผวนของราคาสูงกว่าโรงแรมในเมือง โดยเฉพาะในช่วงเดือนกรกฎาคม กับสิงหาคม ซึ่งเป็นช่วงฤดูท่องเที่ยวของโปรตุเกส ค่า ADR ของ

โรงแรมรีสอร์ทเพิ่มขึ้นสะท้อนถึงความต้องการที่สูงขึ้นของนักท่องเที่ยว ขณะที่ช่วงโลว์ซีซั่นในเดือนมกราคม ถึงมีนาคม ค่า ADR ของโรงแรมรีสอร์ทลดลงอย่างมาก อาจเป็นเพราะจำนวนนักท่องเที่ยวที่ลดลง ทำให้โรงแรมจึงต้องปรับราคาลงเพื่อจูงใจลูกค้า ส่วนโรงแรมในเมืองมีแนวโน้มของราคาที่เสถียรกว่า เพราะพึ่งพาลูกค้ากลุ่มนักธุรกิจ หรือนักเดินทางทั่วไปที่ไม่ได้ขึ้นอยู่กับฤดูกาลมากนัก

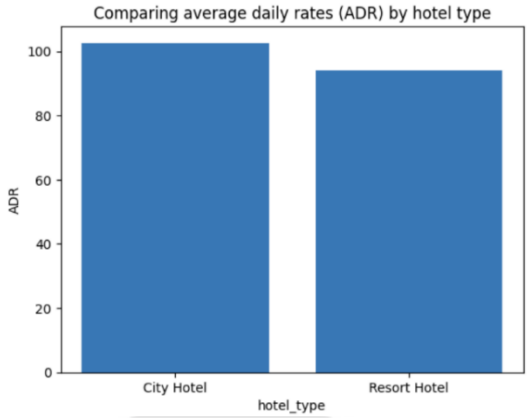
เพื่อให้เข้าใจถึงความแตกต่างด้านราคาค่าห้องพัก ระหว่างโรงแรมทั้งสองประเภทในแต่ละช่วงเวลา จึงทำการพล็อตกราฟเปรียบเทียบแนวโน้มของอัตราค่าที่พักเฉลี่ยรายวัน (ADR) ในแต่ละเดือนของโรงแรมรีสอร์ท และโรงแรมในเมือง ดังที่ปรากฏในภาพประกอบ 9 โดยกราฟนี้ช่วยให้เห็นภาพรวมและความแตกต่างของระดับราคาห้องพักตลอดทั้งปีได้ชัดเจนมากขึ้น



ภาพประกอบ 9 ภาพแสดงแนวโน้มของอัตราค่าที่พักเฉลี่ยรายวัน (ADR) ของโรงแรมในแต่ละเดือน

จากภาพประกอบที่ 9 แสดงให้เห็นแนวโน้มของอัตราค่าห้องพักเฉลี่ยรายวันในแต่ละเดือนของรีสอร์ท และโรงแรมในเมือง กราฟแสดงให้เห็นว่าในช่วงเดือนสิงหาคม ซึ่งมักเป็นช่วงฤดูกาลท่องเที่ยวของประเทศโปรตุเกส โรงแรมรีสอร์ทมีอัตราค่าห้องพักเฉลี่ยสูงกว่าของโรงแรมในเมืองอย่างมีนัยสำคัญ ในขณะที่ตลอดช่วงเวลาอื่น ๆ โรงแรมในเมืองจะมีอัตราค่าห้องพักเฉลี่ยที่สูงกว่า รวมทั้งลักษณะเส้นกราฟค่อนข้างคงที่ เมื่อเทียบกับรีสอร์ท แนวโน้มดังกล่าวสะท้อนให้เห็นถึงฤดูกาลท่องเที่ยวและความต้องการที่พักที่เปลี่ยนแปลงไปตามช่วงเวลา โดยเฉพาะในช่วงที่อัตราค่าห้องพักเฉลี่ยรายวันสูงสุด ของรีสอร์ทช่วงกลางปี ซึ่งสอดคล้องกับพฤติกรรมนักท่องเที่ยวที่นิยมเดินทางไปยังแหล่งท่องเที่ยวธรรมชาติในฤดูหยุดยาว หรือช่วงเทศกาล

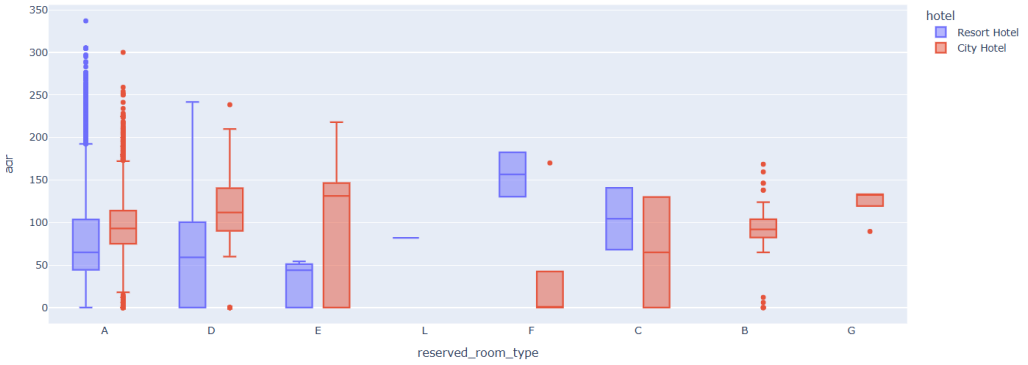
ทำการวิเคราะห์หาความสัมพันธ์ระหว่างประเภทของโรงแรม และอัตราค่าห้องพักเฉลี่ยต่อวัน



ภาพประกอบ 10 กราฟเปรียบเทียบค่า ADR ของ โรงแรมในเมือง และโรงแรมรีสอร์ท

จากภาพประกอบ 10 กราฟแสดงให้เห็นว่า โรงแรมในเมืองมีอัตราค่าห้องพักเฉลี่ยสูงกว่าโรงแรมรีสอร์ท

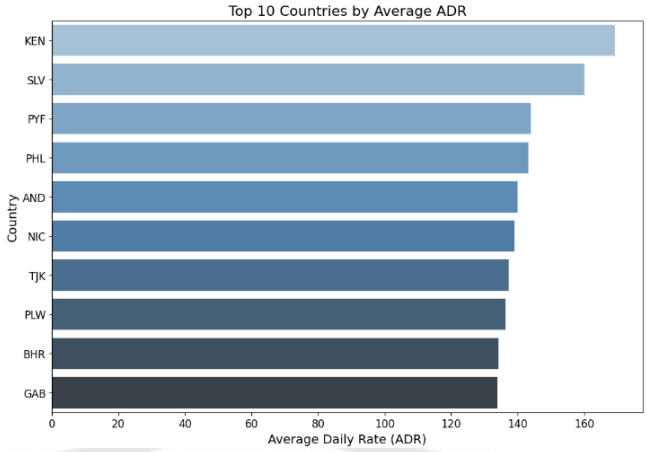
ทำการเปรียบเทียบการกระจายตัวของค่าห้องพักเฉลี่ยต่อวัน สำหรับแต่ละประเภทของห้องพักที่จอง และดูความแตกต่างของค่าห้องพักเฉลี่ยต่อวัน ระหว่างโรงแรมทั้งสองประเภท



ภาพประกอบ 11 กราฟเปรียบเทียบค่าการกระจายตัวของค่าห้องพักเฉลี่ยต่อวัน

จากภาพประกอบ 11 กราฟ Box Plot แสดงให้เห็นว่าห้องพักประเภท A ในโรงแรมในเมือง มีค่าอัตราค่าห้องพักเฉลี่ย(ADR) สูงกว่าห้องพักประเภท A ของโรงแรมรีสอร์ท อย่างไรก็ตามค่า ADR ของโรงแรมรีสอร์ทมีการกระจายตัวของค่า ADR ที่กว้างกว่า รวมถึงมี outliers จำนวนมากทั้งด้านล่าง และด้านบน แสดงให้เห็นว่าราคาของห้องพักประเภท A ในโรงแรมรีสอร์ท อาจมีการเปลี่ยนแปลงตามฤดูกาล หรือโปรโมชั่นโดยขึ้นอยู่กับปัจจัยอื่นๆร่วมด้วย เช่น ขนาดของห้องพัก วิว หรือสิ่งอำนวยความสะดวกเพิ่มเติม

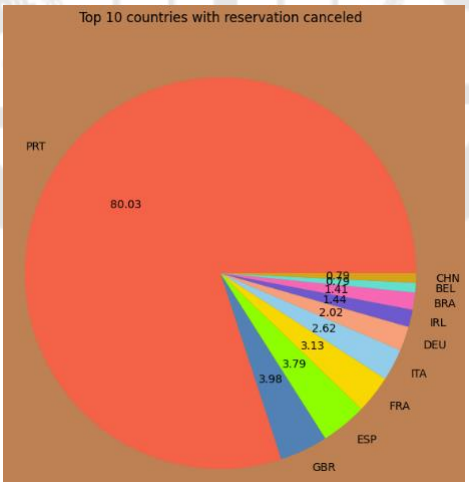
ทำการวิเคราะห์หาประเทศที่เข้าใช้บริการโรงแรมสูงสุด 10 อันดับแรก



ภาพประกอบ 12 กราฟแสดงประเทศที่เข้าใช้บริการโรงแรมสูงสุด 10 อันดับแรก

จากภาพประกอบที่ 12 แสดงให้เห็นรายได้เฉลี่ยของห้องพักต่อวัน (ADR) จากประเทศที่เข้าใช้บริการโรงแรมสูงสุด 10 อันดับแรก พบว่าประเทศโปรตุเกสมีการยกเลิกการจองมากที่สุดถึงร้อยละ 80.03

ทำการวิเคราะห์หาประเทศที่ดำเนินการยกเลิกเข้าพักสูงสุด 10 อันดับแรก



ภาพประกอบ 13 กราฟแสดงประเทศที่ดำเนินการยกเลิกเข้าพักสูงสุด 10 อันดับแรก

จากภาพประกอบที่ 13 แสดงให้เห็นว่าประเทศโปรตุเกสดำเนินการยกเลิกเข้าพักสูงสุดจาก 10 อันดับแรก เมื่อนำไปวิเคราะห์ร่วมกับภาพประกอบที่ 12 พบว่าประเทศที่เข้าใช้บริการห้องพักมากเป็นอันดับต้นๆ เมื่อหาค่าเฉลี่ยของ ADR ที่สูงที่สุดในแต่ละประเทศกลับมีค่าเฉลี่ยที่ต่ำกว่าเนื่องจากมีอัตราการเข้าพักมาก ขณะเดียวกันก็มีอัตราการยกเลิกห้องพักมากเช่นกัน

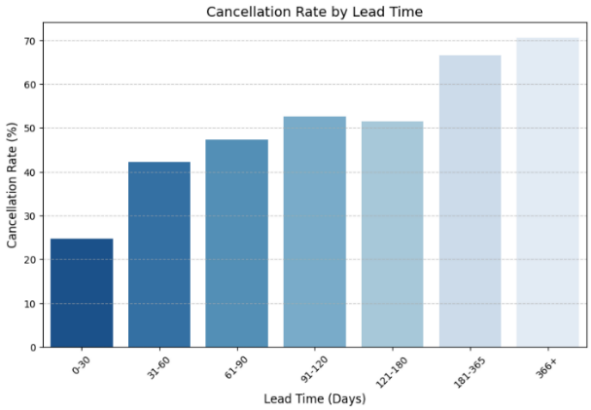
วิเคราะห์ข้อมูลการยกเลิกการจอง โดยพิจารณาจากระยะเวลาการจองห้องพักล่วงหน้า แสดงดังตารางที่ 4

ตาราง 4 ตารางแสดงอัตราการยกเลิกห้องพัก กับระยะเวลาการจองล่วงหน้า

Lead time (day)	total_bookings	canceled_bookings	Cancellation_rate (%)
0-30	21,607	5,325	24.645
31-60	12,057	5,090	42.216
61-90	9,041	4,283	47.373
91-120	7,886	4,147	52.587
121-180	11,425	5,877	41.440
181-365	19,640	13,089	66.645
366+	2,941	2,077	70.622

จากตารางที่ 4 แสดงให้เห็นว่าอัตราการยกเลิกห้องพักเพิ่มขึ้นตามระยะเวลาการจองล่วงหน้า (Lead Time) โดยการจองที่ใกล้วันเข้าพัก 0 ถึง30วัน มีอัตราการยกเลิกต่ำสุดที่ 24.64% ในขณะที่การจองล่วงหน้าเกิน 6 เดือน (366+ วัน) มีอัตราการยกเลิกสูงสุดที่ 70.62% ซึ่งสะท้อนว่าลูกค้าที่จองล่วงหน้านานมักมีแนวโน้มยกเลิกสูงขึ้น

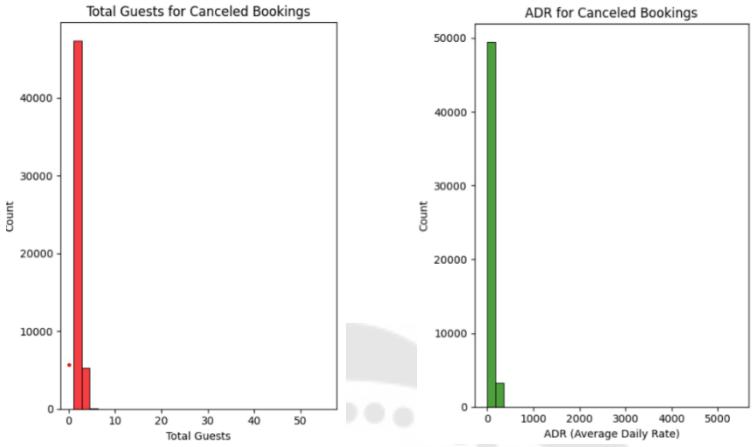
วิเคราะห์ข้อมูลการยกเลิกการจอง โดยพิจารณาจากระยะเวลาการจองห้องพักล่วงหน้า



ภาพประกอบ 14 กราฟแสดงอัตราการยกเลิกห้องพัก กับระยะเวลาการจองล่วงหน้า

จากภาพประกอบ 14 กราฟแสดงให้เห็นว่าอัตราการยกเลิกห้องพักมีแนวโน้มเพิ่มขึ้นตามระยะเวลาการจองล่วงหน้า (Lead Time) โดยการจองที่ใกล้วันเข้าพัก (0-30 วัน) มีอัตราการยกเลิกต่ำที่สุด (24.64%) ในขณะที่การจองล่วงหน้ามากกว่า 6 เดือนมีอัตราการยกเลิกสูงสุด (>70%)

วิเคราะห์ข้อมูลที่เกี่ยวข้องกับการยกเลิกการจอง โดยพิจารณาจากจำนวนผู้เข้าพัก และจำนวนคืนที่เข้าพัก

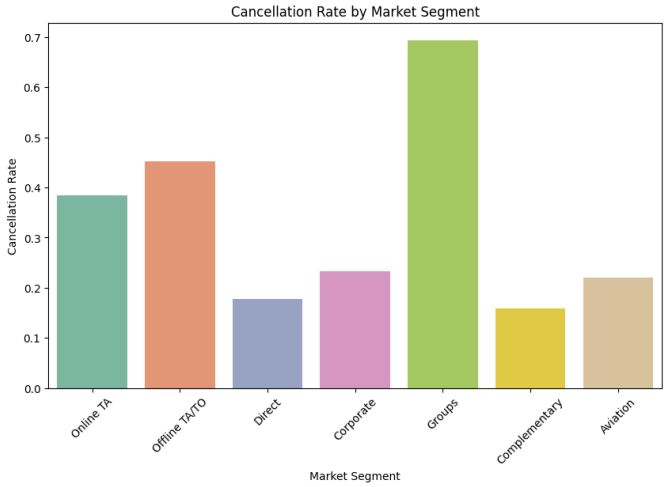


(b) (b)

ภาพประกอบ 15 กราฟแสดงจำนวนผู้เข้าพัก และจำนวนคืนที่เข้าพักที่มีการยกเลิกการจอง

จากภาพประกอบ 15 กราฟ (a) แสดงให้เห็นว่าการจองที่มีจำนวนแขกที่ประมาณ 1-2 คน เป็นกลุ่มที่ทำการยกเลิกการจองห้องพักซึ่งอาจจะสืบเนื่องกับราคาตามประเภทห้องพัก และจำนวนผู้เข้าพักน้อยจึงมีความยืดหยุ่นสูงในการวางแผน และยกเลิกห้องพัก ส่วนผลลัพธ์ที่แสดงในกราฟ (b) พบว่าการจองห้องที่มีราคาห้องพักเฉลี่ยต่ำ จะมีแนวโน้มถูกยกเลิกการจองมากกว่า โดยเฉพาะราคาห้องพักเฉลี่ยที่อยู่ในช่วง 0-200 ซึ่งอาจเป็นเพราะลูกค้าที่จองห้องพักราคาถูกมักเป็นกลุ่มที่มีความอ่อนไหวต่อราคา เมื่อพบห้องพักราคาที่ถูกกว่า จึงเกิดการยกเลิกการจองเดิม

วิเคราะห์อัตราการยกเลิกการจอง ตาม Market Segment



ภาพประกอบ 16 กราฟแสดงอัตราการยกเลิกการจอง ตาม Market Segment

จากภาพประกอบ 16 กราฟแสดงอัตราการยกเลิกการจอง ตาม Market Segment พบว่า กลุ่ม Groups มีอัตราการยกเลิกสูงสุด เนื่องจากเป็นการจองแบบหมู่คณะที่มีโอกาสเปลี่ยนแปลงสูง รองลงมา คือ Offline TA/TO และ Online TA ซึ่งสะท้อนว่าการจองผ่านตัวแทน และแพลตฟอร์มออนไลน์มักถูกยกเลิกบ่อย เพราะลูกค้าสามารถจองหลายโรงแรม และยกเลิกได้ง่าย ในทางกลับกัน Direct Booking และ Complementary มีอัตราการยกเลิกต่ำสุด เพราะลูกค้าจองโดยตรงมักมีความตั้งใจเข้าพักสูง แสดงให้เห็นว่าลูกค้าที่จองผ่านโรงแรมโดยตรง และกลุ่มลูกค้าพิเศษที่ใช้โปรโมชั่นมีแนวโน้มเข้าพักมากกว่ากลุ่มอื่น ๆ

การเตรียมข้อมูล (Data Preparation)

เพื่อให้สามารถใช้ข้อมูลอย่างมีประสิทธิภาพในการสร้างแบบจำลองพยากรณ์รายได้ของโรงแรม ข้อมูลจึงถูกจัดเตรียมให้อยู่ในรูปแบบอนุกรมเวลารายวัน (Daily Time Series) โดยมีขั้นตอนดังนี้

1. แปลงคอลัมน์ arrival_date ซึ่งระบุวันที่เข้าพักของผู้จองห้องพัก โดยแปลงให้อยู่ในรูปแบบวันที่ (datetime) เพื่อให้สามารถจัดการข้อมูลเชิงเวลาได้อย่างมีประสิทธิภาพ

2. ตั้งค่า Index โดยกำหนดให้คอลัมน์ arrival_date เป็น Index ของชุดข้อมูล เพื่อความสะดวกในการจัดกลุ่มข้อมูลตามช่วงเวลา และการประมวลผลแบบเชิงเวลา

3. สร้าง Time Series รายวัน โดยข้อมูลจำนวนห้องพักที่ขายได้ (room_sold) ถูกจัดกลุ่มตามช่วงเวลาสุดท้ายของแต่ละวันโดยใช้คำสั่ง resample เพื่อสร้างข้อมูล Time Series รายวัน เพื่อให้เหมาะกับการใช้ในการพยากรณ์

4. การตรวจสอบโครงสร้างข้อมูล (Decomposition) โดยทำการแยกองค์ประกอบของข้อมูลอนุกรมเวลา ได้แก่ แนวโน้ม (Trend), ความเป็นฤดูกาล (Seasonality) และความผิดปกติ (Residual) เพื่อเข้าใจพฤติกรรมของข้อมูลในเชิงลึกและประกอบการเลือกแบบจำลอง

5. การตรวจสอบความอยู่กับที่ของข้อมูล (Stationarity) ในขั้นตอนนี้จะเป็นการทดสอบความคงที่ของข้อมูลโดยใช้วิธี Augmented Dickey-Fuller (ADF) Test และวิเคราะห์กราฟ ACF/PACF หากข้อมูลยังไม่คงที่ (non-stationary) จะทำการแปลงข้อมูลด้วยเทคนิค log transformation เพื่อให้ข้อมูลมีลักษณะเหมาะสมต่อการสร้างแบบจำลอง ARIMA และSARIMA

การคัดเลือกตัวแปร (Feature Selection)

ในการพัฒนาแบบจำลองการพยากรณ์ที่มีประสิทธิภาพ และเหมาะสมกับปัญหาวิจัย กระบวนการคัดเลือกตัวแปร เพื่อให้ได้ตัวแปรที่มีอิทธิพลต่อการพยากรณ์จำนวนห้องพักที่ขายได้มากที่สุด การเลือกตัวแปรที่เหมาะสมไม่เพียงช่วยลดความซับซ้อนของแบบจำลอง แต่ยังส่งผลโดยตรงต่อความแม่นยำของผลลัพธ์ และการตีความเชิงธุรกิจ

ชุดข้อมูลที่ใช้ในงานวิจัยนี้ประกอบด้วยตัวแปรที่เกี่ยวข้องกับการจองห้องพัก เช่น อัตราค่าห้องพักเฉลี่ย (ADR), Lead Time, จำนวนคืนที่เข้าพัก (Total Nights), ประเภทห้องพัก, สถานะการจอง, ประเภทลูกค้า, และปัจจัยอื่น ๆ ที่เกี่ยวข้อง เพื่อให้สามารถนำข้อมูลไปใช้วิเคราะห์และพยากรณ์รายได้ได้อย่างเหมาะสม ผู้วิจัยจึงดำเนินการคัดเลือกประเภทห้องพักที่มีจำนวนการเข้าพักสูงสุด โดยอาศัยสมมติฐานตามหลักการจัดการโรงแรมว่า โรงแรมส่วนใหญ่มักมีห้องประเภทมาตรฐาน (Standard Room) มากที่สุด จากนั้นจึงสรุปจำนวนผู้เข้าพักในแต่ละประเภทห้องเพื่อกำหนดตัวแปรเป้าหมายในการวิเคราะห์ ดังแสดงในตารางที่ 5

ตาราง 5 ตารางแสดงจำนวนห้องพักที่มีผู้เข้าพักมากที่สุด

No.	Assigned room type	Count
1	A	88,570
2	D	29,230
3	E	9,081
4	F	4,370
5	G	2,935
6	C	2,824
7	B	2,662
8	H	840
9	I	397
10	K	204
11	L	2

จากตารางแสดงจำนวนห้องพักที่มีผู้เข้าพักมากที่สุด ทำให้สามารถระบุได้ว่าห้องประเภท A เป็นห้อง standard room ที่มักจะจำกัดจำนวนผู้เข้าพักอยู่ที่ 1-2 คน ดังนั้นจึงทำการคำนวณ room_sold โดยแยกข้อมูลตามประเภทห้อง A เพื่อความเรียบง่ายและความสม่ำเสมอในการวิเคราะห์ เนื่องจากมีจำนวนการจองสูงสุดในข้อมูล นอกจากนี้ เมื่อทราบจำนวน room sold ก็จะสามารถหารายได้ของโรงแรมได้ โดยนำไปคูณกับอัตราค่าห้องพักเฉลี่ยรายวัน (ADR) ดังแสดงในภาพประกอบที่17

ข้อมูลเฉพาะ assigned_room_type ประเภท A พร้อม room_sold และ revenue:

	assigned_room_type	total_guests	room_sold	adr	revenue
0	A	2.0	1.0	82.00	82.00
10	A	2.0	1.0	136.33	136.33
11	A	3.0	2.0	110.30	220.60
12	A	2.0	1.0	82.00	82.00
14	A	2.0	1.0	73.80	73.80

ภาพประกอบ 17 ภาพแสดงตัวอย่างจำนวนของ room sold และรายได้ของโรงแรม

เพื่อเพิ่มประสิทธิภาพ และความแม่นยำในการพยากรณ์ข้อมูลอนุกรมเวลา การสร้างคุณลักษณะเพิ่มเติม (Feature Engineering) จากข้อมูลที่ได้รวบรวมมา โดยมุ่งเน้นไปที่การเลือกข้อมูลสำคัญเชิงเวลา และข้อมูลเชิงธุรกิจที่เกี่ยวข้องกับการจองห้องพักของโรงแรม การสร้างคุณลักษณะเพิ่มเติมนี้มีวัตถุประสงค์เพื่อให้แบบจำลองสามารถเข้าใจโครงสร้าง และบริบทของข้อมูลได้ลึกซึ้งยิ่งขึ้น ซึ่งประกอบด้วยขั้นตอนสำคัญ ดังนี้

1. สร้างตัวแปรเชิงเวลา (Temporal Features) จากข้อมูลวันที่ที่ ระบุในแต่ละรายการ ทำการแยกข้อมูลออกเป็นตัวแปรย่อย ได้แก่ วัน (Day), เดือน (Month) และปี (Year)

2. สร้างตัวแปรด้านปริมาณ (Quantity-related Features) เพื่อสะท้อนถึงปริมาณการใช้บริการของลูกค้าในแต่ละครั้ง โดยทำการสร้างตัวแปร total_nights ซึ่งเป็นผลรวมของจำนวนคืนที่ลูกค้าเข้าพัก และ total_guest ซึ่งแสดงถึงจำนวนผู้เข้าพัก

3. สร้างตัวแปรเป้าหมาย (Target Features) สำหรับการวิเคราะห์ และพยากรณ์ในงานวิจัยนี้ โดยสร้างตัวแปรสำคัญ ได้แก่ room_sold ซึ่งหมายถึงจำนวนห้องพักที่ถูกขายในแต่ละช่วงเวลา และ revenue หรือรายได้ที่ได้จากการคูณจำนวนห้องพักที่ขายได้กับอัตราค่าห้องพักเฉลี่ย (ADR) ในแต่ละวัน โดยรายละเอียดของข้อมูลตัวแปรทั้งหมด แสดงในตารางที่ 6

ตาราง 6 ตารางแสดงรายละเอียดข้อมูลแต่ละคอลัมน์

No.	Variable	Description
1	hotel	object
2	is_canceled	int64
3	lead_time	int64
4	arrival_date_year	int64
5	arrival_date_month	object
6	arrival_date_week_number	int64
7	arrival_date_day_of_month	int64
8	stays_in_weekend_nights	int64
9	stays_in_week_nights	int64
10	adults	int64

ตารางที่ 6 ตารางแสดงรายละเอียดข้อมูลแต่ละคอลัมน์ (ต่อ)

No.	Variable	Description
11	children	float64
12	babies	int64
13	meal	object
14	country	object
15	market_segment	object
16	is_repeated_guest	int64
17	previous_cancellations	int64
18	previous_bookings_not_canceled	int64
19	reserved_room_type	object
20	assigned_room_type	Object
21	booking_changes	int64
22	deposit_type	object
23	days_in_waiting_list	int64
24	customer_type	object
25	adr	float64
39	revenue	float64

เพื่อให้เห็นข้อมูลทางสถิติ ของตัวแปร room sold และ revenue จึงทำการตรวจสอบข้อมูลเชิงสถิติของคอลัมน์ใหม่ ดังแสดงตารางที่ 7

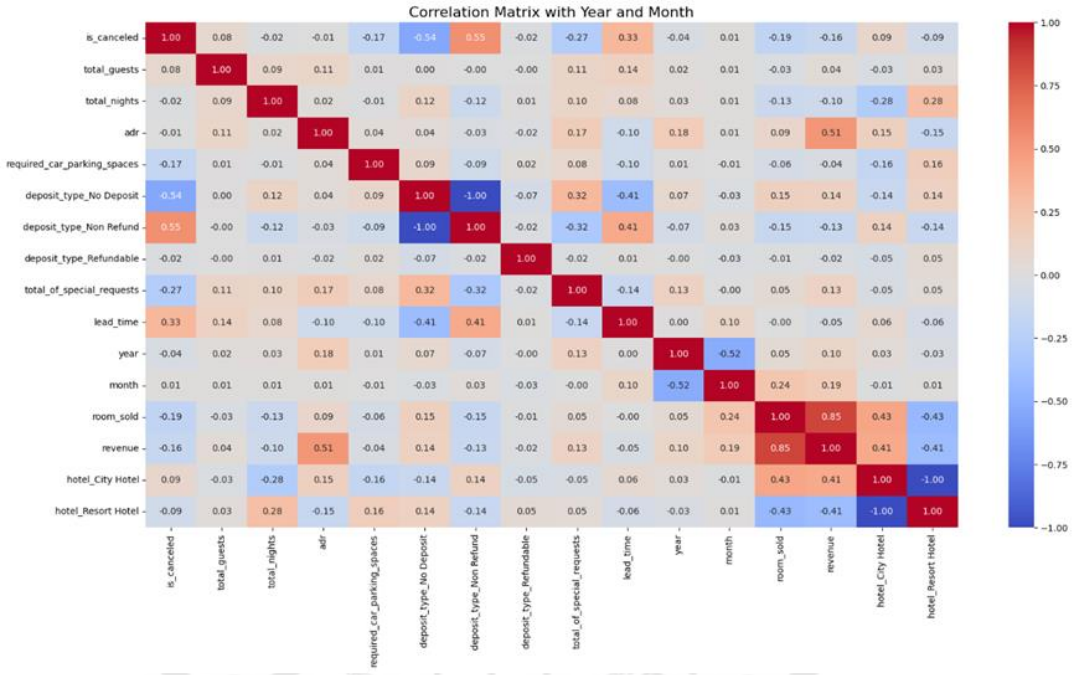
ตาราง 7 ตารางแสดงค่าสถิติของจำนวนห้องที่ถูกขาย และรายได้

Summary Statics	Room_sold	Revenue
count	793.00	793.0
mean	5,028.56	466,647.5
std	4,837.81	458,949.3
min	73.00	3,394.6
25%	1,979.00	151,836.4
50%	3,814.00	352,199
75%	6,098.00	584,606
max	43,452.00	3,693,974

จากตารางที่ 7 แสดงค่าสถิติของจำนวนห้องที่ถูกขาย (Room Sold) และ รายได้ (Revenue) พบว่าจำนวนห้องพักที่ถูกขายมีค่าเฉลี่ยอยู่ที่ประมาณ 5,028 ห้อง เมื่อพิจารณาค่าต่ำสุด และค่าสูงสุด จะพบว่ามีความแตกต่างกันอย่างมาก กล่าวคือมีช่วงที่ขายห้องพักได้น้อยเพียง 73 ห้อง และในบางช่วงสามารถขายได้สูงถึง 43,452 ห้อง ซึ่งแสดงถึงลักษณะของอุตสาหกรรมที่ได้รับผลกระทบจากฤดูกาลและปัจจัยภายนอก ในด้านรายได้ พบว่ารายได้เฉลี่ยของโรงแรมอยู่ที่ประมาณ 466,647.5 หน่วย โดยมีค่าต่ำสุดอยู่ที่ 3,394.6 หน่วย และค่าสูงสุดสูงถึง 3,693,974 หน่วย ซึ่งสอดคล้องกับแนวโน้มของจำนวน

ห้องพักที่ถูกขาย ในช่วงเวลาที่มีความต้องการสูง เช่นช่วงเทศกาล โรงแรมจะมีรายได้สูงกว่าปกติ ขณะที่ในช่วงโลว์ซีซั่น รายได้ก็จะลดต่ำลงตามไปด้วย

ทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรสำคัญในการพยากรณ์รายได้ของโรงแรม โดยใช้ Heat Map เพื่อหาปัจจัยที่ส่งผลต่อรายได้โรงแรม รายละเอียดข้อมูลแสดงตัวอย่างดังแสดงในภาพประกอบที่ 18



ภาพประกอบ 18 กราฟแสดงความสัมพันธ์ระหว่างตัวแปรสำคัญในการพยากรณ์รายได้ของโรงแรม

จากผลลัพธ์ของความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ที่แสดงใน Heatmap มีข้อสังเกตที่สำคัญดังนี้

1. room_sold กับ revenue มีความสัมพันธ์เชิงบวกสูง (+0.85) แสดงว่าจำนวนห้องพักที่ถูกขายส่งผลโดยตรงต่อรายได้ ซึ่งสมเหตุสมผล เนื่องจากห้องพักที่ขายได้มากขึ้นจะสร้างรายได้แก่โรงแรมเพิ่มขึ้นโดยตรง

2. revenue กับ ADR มีความสัมพันธ์ทางบวก (+0.51) ซึ่งบ่งชี้ว่าการเพิ่มขึ้นของค่า ADR ส่งผลต่อรายได้โดยตรง เนื่องจากราคาห้องพักที่สูงขึ้น รายได้ของโรงแรมก็เพิ่มขึ้น

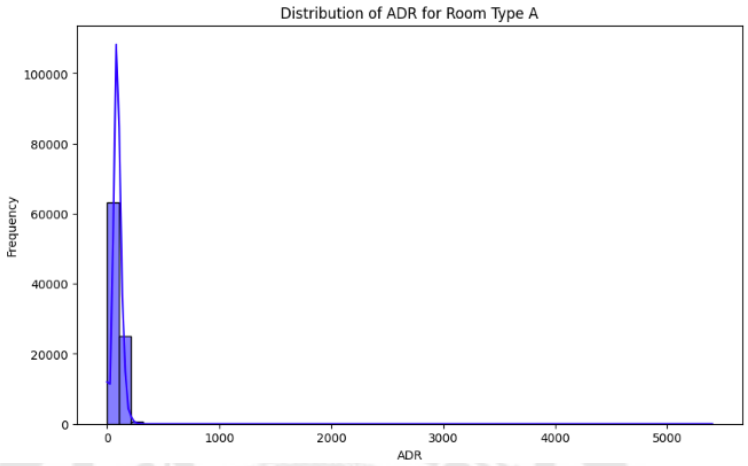
3. is_canceled กับ total_of_special_requests มีความสัมพันธ์เชิงลบ (-0.27) แสดงว่าการจองที่มีคำขอพิเศษมาก มีแนวโน้มยกเลิกน้อย ซึ่งอาจสะท้อนถึงความตั้งใจในการเข้าพัก

4. deposit_type_No Deposit กับ is_canceled มีค่าความสัมพันธ์ติดลบ (-0.54) หมายความว่า ถ้าจองแบบ No Deposit จะมีโอกาสยกเลิกน้อยลง

5. hotel_City Hotel กับ room_sold มีความสัมพันธ์ทางบวก (+0.43) ซึ่งแสดงว่าโรงแรมในเมือง (City Hotel) มีค่า room sold สูงกว่าเมื่อเทียบกับรีสอร์ท ซึ่งมีความสัมพันธ์ทางลบกับ

room sold (-0.43) แต่ประเภทของโรงแรมอาจไม่ใช่ตัวแปรหลักที่ส่งผลต่อราคาดูได้จากค่าความสัมพันธ์กับ ADR ที่มีความสัมพันธ์ทางบวกไม่สูงมาก

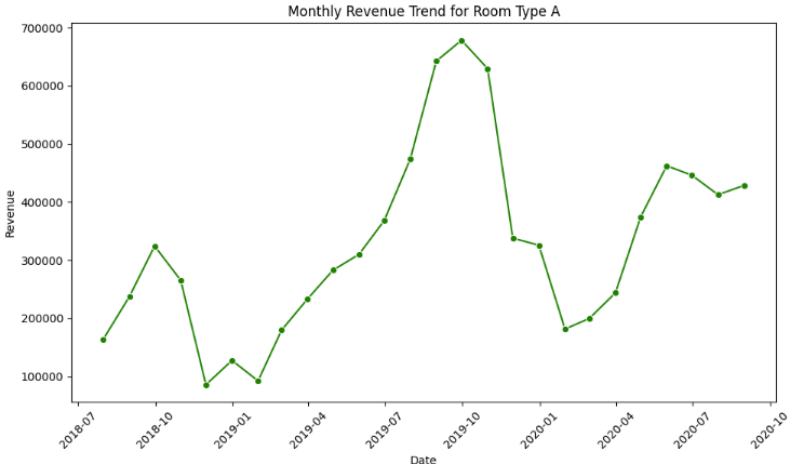
ทำการวิเคราะห์อัตราค่าห้องพักเฉลี่ยของห้องพักประเภท A แสดงดังภาพประกอบที่ 19



ภาพประกอบ 19 แสดงการกระจายตัวของอัตราค่าห้องพักเฉลี่ยสำหรับห้องพักประเภท A

จากภาพประกอบ 19 แสดงให้เห็นว่าค่า ADR สำหรับห้องพักประเภท A มีการกระจายตัวที่ส่วนใหญ่กระจุกตัวอยู่ในช่วง 0 ถึง 500 โดยมีความหนาแน่นสูงสุดในช่วง 100 ถึง 200 ซึ่งแสดงให้เห็นว่าห้องประเภทนี้มักกำหนดราคาห้องพักเฉลี่ยในช่วงราคาดังกล่าว อย่างไรก็ตาม การกระจายตัวของข้อมูลแสดงลักษณะ Skewed Right ซึ่งบ่งชี้ว่ามีบางการจองที่มีค่า ADR สูงเกิน 500 ซึ่งอาจถือเป็น Outliers ที่เกิดจากสถานการณ์พิเศษ เช่น การจองในช่วงเทศกาล หรือราคาห้องพักภายใต้เงื่อนไขเฉพาะ

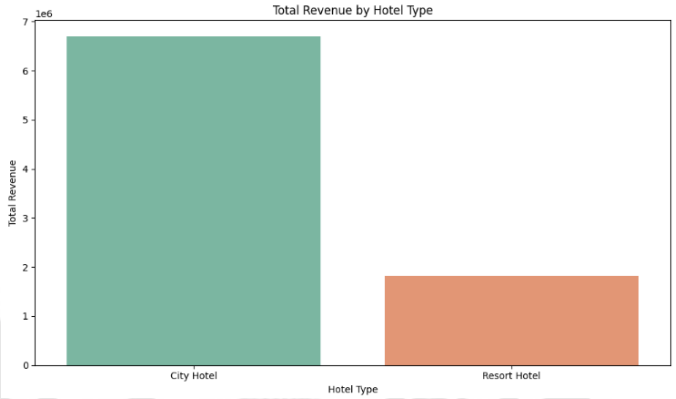
วิเคราะห์แนวโน้มรายได้ต่อเดือนสำหรับห้องพักประเภท A แสดงดังภาพประกอบที่ 20



ภาพประกอบ 20 แสดงแนวโน้มรายได้ต่อเดือนสำหรับห้องพักประเภท A

จากภาพประกอบ 20 แสดงให้เห็นลักษณะการเปลี่ยนแปลงของข้อมูลในลักษณะ Seasonal Trend โดยรายได้สูงสุดเกิดขึ้นในเดือนตุลาคม 2019 และต่ำสุดในช่วงต้นปี โดยการเพิ่มขึ้นของรายได้ในปี 2019 อาจเกิดจากปัจจัยเชิงบวก เช่น ความต้องการที่พักที่เพิ่มขึ้นในฤดูกาลท่องเที่ยว

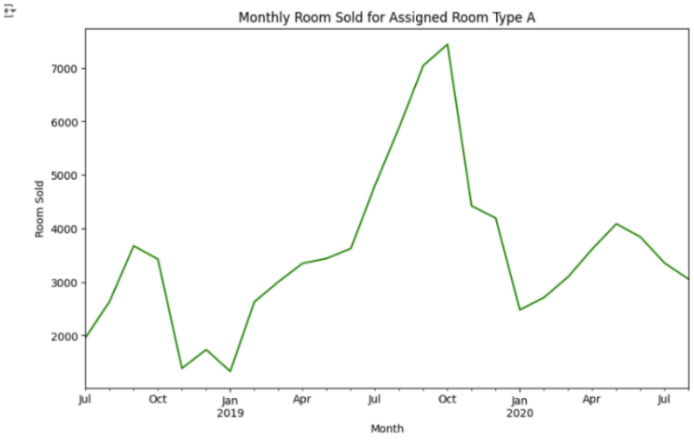
วิเคราะห์รายได้ของโรงแรมโดยจำแนกตามประเภทของโรงแรม แสดงดังภาพประกอบ 21



ภาพประกอบ 21 แสดงรายได้รวมจำแนกตามประเภทโรงแรม ได้แก่ City Hotel และ Resort Hotel

จากภาพประกอบ 21 พบว่าโรงแรม City Hotel มีรายได้รวมสูงกว่าประเภท Resort Hotel อย่างชัดเจน สะท้อนถึงความนิยม และความต้องการที่พักในพื้นที่เมืองสูงกว่า ในขณะที่โรงแรม Resort Hotel มีรายได้รวมต่ำกว่า ซึ่งอาจเกิดจากลักษณะเฉพาะของโรงแรมที่มุ่งเน้นการให้บริการในพื้นที่พักผ่อน ซึ่งมีความต้องการตามฤดูกาล ความแตกต่างนี้แสดงถึงลักษณะตลาด และความต้องการที่แตกต่างกันระหว่างสองประเภทโรงแรม

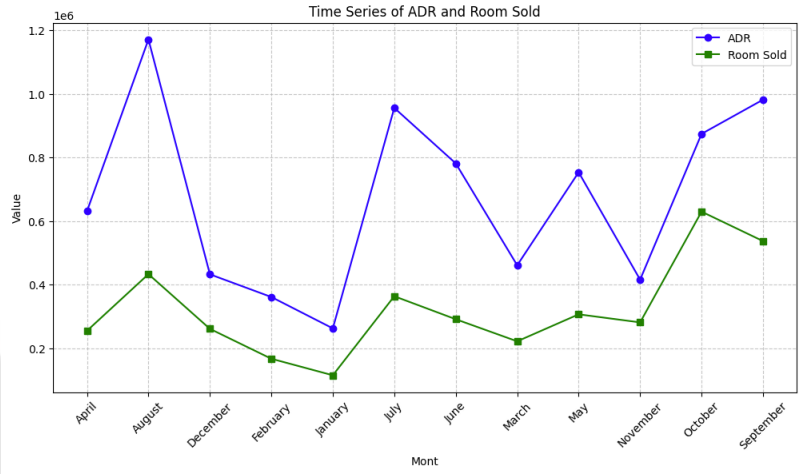
ทำการตรวจสอบแนวโน้มข้อมูลจากข้อมูล Time Series รายเดือนที่สร้างขึ้น เพื่อวิเคราะห์แนวโน้ม และความผันผวนของจำนวนห้องพักที่ขายได้ในช่วงเดือนกรกฎาคม 2018 ถึงเดือนตุลาคม 2020 แสดงดังภาพประกอบที่ 22



ภาพประกอบ 22 ภาพแสดงแนวโน้มของข้อมูลการจองห้องพักของห้องพักประเภท A

จากภาพประกอบ 22 กราฟแสดงจำนวนห้องพักที่ขายได้รายเดือนของห้องพักประเภท A ระหว่างปี 2018 ถึง2020 ชี้ให้เห็นถึงลักษณะตามฤดูกาลอย่างชัดเจน โดยจำนวนห้องพักที่ขายได้เพิ่มขึ้นในช่วงฤดูท่องเที่ยว คือเดือน กรกฎาคม ถึงตุลาคม และลดลงในช่วงโลว์ซีซั่น คือเดือนมกราคม ถึงเมษายน ที่อาจเกิดจากปัจจัยภายนอก เช่น สภาพเศรษฐกิจ หรือการลดลงของการเดินทาง

วิเคราะห์หาความสัมพันธ์ระหว่างอัตราค่าห้องพักเฉลี่ยรายวัน (ADR) และ Room Sold แสดงดังภาพประกอบที่ 23



ภาพประกอบ 23 กราฟแสดงแนวโน้มของอัตราค่าห้องพักเฉลี่ยรายวัน (ADR) และ Room Sold

ภาพประกอบ 23 แสดงแนวโน้มของอัตราค่าห้องพักเฉลี่ยรายวัน (ADR) และ Room Sold มีแนวโน้มไปในทิศทางเดียวกันในแต่ละเดือน พบว่า ADR และจำนวนห้องที่ขายได้มีความสัมพันธ์กับฤดูกาล โดยมีจุดสูงสุดใน เดือนกรกฎาคม ถึงสิงหาคม และตุลาคม ซึ่งเป็นช่วงฤดูท่องเที่ยว ของโปรตุเกส ช่วงนี้ทั้ง ADR และจำนวนห้องที่ขายได้เพิ่มขึ้นอย่างชัดเจน สะท้อนถึงความต้องการเข้าพักที่สูง ในทางกลับกัน ช่วง มกราคม ถึงกุมภาพันธ์ ซึ่งเป็นช่วงโลว์ซีซั่น ค่า ADR และจำนวนห้องที่ขายได้ลดลงอย่างมาก แสดงให้เห็นว่าช่วงนี้โรงแรมอาจมีการลดราคาหรือมีจำนวนนักท่องเที่ยวลดลง โดยรวมแล้ว ADR และ Room Sold ช่วงที่มีนักท่องเที่ยวมาก ADR มักสูงขึ้น และช่วงที่ความต้องการที่พักลดลง ADR ก็มักลดลงด้วย การพยากรณ์จึงควรคำนึงถึงฤดูกาล ซึ่งเป็นปัจจัยสำคัญในการคาดการณ์แนวโน้มราคา และจำนวนห้องที่ขายในอนาคต

การสร้าง และประเมินแบบจำลองพยากรณ์

ในงานวิจัยนี้ ได้สร้างแบบจำลองเพื่อคาดการณ์รายได้ของโรงแรม จากการพยากรณ์จำนวนห้องพักที่ขายได้ ซึ่งเป็นตัวแปรหลักที่สะท้อนรายได้ของโรงแรมโดยตรง โดยใช้ภาษา Python และ

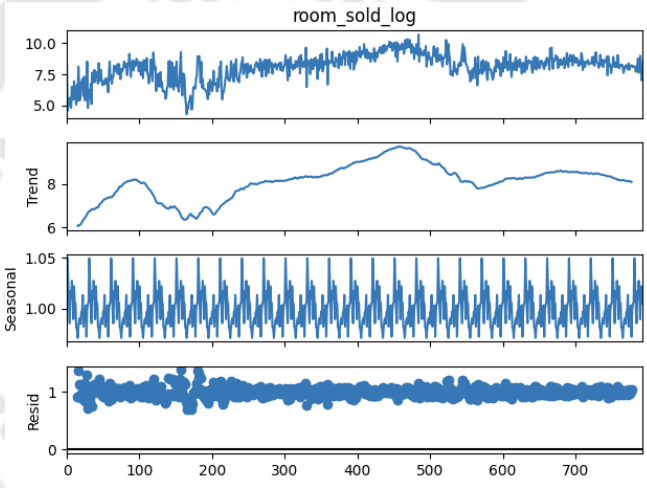
ดำเนินการปรับค่าพารามิเตอร์ให้เหมาะสมเพื่อเพิ่มประสิทธิภาพของแบบจำลอง โดยวิธีการที่ใช้ประกอบด้วยแบบจำลองที่ได้รับความนิยมในงานวิเคราะห์อนุกรมเวลา ได้แก่ ARIMA , SARIMA, ARIMAX, SARIMAX และ Prophet การนำแบบจำลอง ARIMAX และ SARIMAX เข้ามาเปรียบเทียบโดยตรงกับ Prophet ซึ่งเป็นแบบจำลองที่สามารถรองรับตัวแปรเสริม (Exogenous Variables) ได้ เพื่อศึกษาผลกระทบของตัวแปรสำคัญต่อความแม่นยำในการพยากรณ์ยอดขายห้องพักโดยการเตรียมข้อมูลสำหรับสร้างแบบจำลองนั้น ข้อมูลธุรกรรมการจองห้องพักในแต่ละวันจะถูกแปลงให้อยู่ในรูปแบบอนุกรมเวลารายวัน (Daily Time Series) และแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ข้อมูลฝึกสอน (Training Data) ซึ่งประกอบด้วยข้อมูล 80% แรก และข้อมูลทดสอบ (Testing Data) อีก 20% เพื่อประเมินประสิทธิภาพการพยากรณ์ของแบบจำลองในสถานการณ์ที่ใกล้เคียงกับการใช้งานจริง

โดยเครื่องมือที่ใช้วัดประสิทธิภาพของแบบจำลอง คือ Mean Absolute Percentage Error (MAPE) , Mean Absolute Error (MAE) และ Root Mean Squared Error (RMSE) ซึ่งเป็นตัวชี้วัดที่นิยมใช้ในงานวิเคราะห์อนุกรมเวลา เพื่อตรวจสอบความแม่นยำในการพยากรณ์ของจำลองแต่ละประเภท ผลลัพธ์ที่ได้จากการทดสอบ จะถูกนำมาเปรียบเทียบ เพื่อระบุแบบจำลองที่มีความเหมาะสม และประสิทธิภาพดีที่สุดสำหรับการพยากรณ์จำนวนห้องพักที่ขายได้

สำหรับข้อมูลที่ใช้ในการพยากรณ์ คือตัวแปร Daily data ซึ่งแสดงจำนวนห้องที่ขายได้ในหนึ่งวัน โดยข้อมูลนี้ถูกใช้เป็นข้อมูลนำเข้า ของแบบจำลองในการพยากรณ์จำนวนห้องพักที่ขายได้ในอนาคต การประเมินประสิทธิภาพของแบบจำลองพิจารณาจากความสามารถในการพยากรณ์จำนวนห้องพักที่ขายได้ในช่วงเวลาทดสอบ จะถูกนำมาเปรียบเทียบระหว่างแบบจำลองต่าง ๆ โดยผลลัพธ์ที่ได้จากแบบจำลองแต่ละประเภทจะอยู่ในรูปแบบ log scale เพื่อควบคุมความผันผวน และทำให้แบบจำลองเรียนรู้ข้อมูลได้อย่างมีประสิทธิภาพ และเพื่อระบุแบบจำลองที่มีความเหมาะสมที่สุดสำหรับการนำไปพยากรณ์รายได้ของโรงแรม โดยหากนำไปใช้เป็นข้อมูลสนับสนุนการวางแผนรายได้ จะต้องแปลงค่าที่พยากรณ์ได้จาก log scale กลับเป็นค่าจริง (inverse transformation) เพื่อให้สอดคล้องกับจำนวนห้องพักจริงที่คาดว่าจะขายได้ แล้วจึงนำจำนวนห้องพักที่คาดการณ์ได้ไปคูณกับค่า ADR (Average Daily Rate) ของแต่ละวันเพื่อคำนวณรายได้ที่คาดว่าจะได้รับในแต่ละวัน ซึ่งสอดคล้องกับวิธีการที่ใช้ในธุรกิจโรงแรมทั่วไป นอกจากนี้เมื่อได้ค่ารายได้ในแต่ละวันแล้ว ยังสามารถนำไปคำนวณยอดรายได้รวม (Cumulative Revenue) เพื่อประเมินความสามารถในการบรรลุเป้าหมายรายได้ตามที่ตั้งไว้ และการวางแผนกลยุทธ์ในอนาคต

การตรวจสอบโครงสร้างของข้อมูลก่อนสร้างแบบจำลอง

ก่อนดำเนินการสร้างแบบจำลองพยากรณ์ ควรทำการตรวจสอบโครงสร้างของข้อมูล โดยเฉพาะการวิเคราะห์แนวโน้ม (Trend) ลักษณะตามฤดูกาล (Seasonality) และค่าอื่น ๆ (Residual) เพื่อทำความเข้าใจรูปแบบการเปลี่ยนแปลงของตัวแปร room_sold อย่างลึกซึ้ง ซึ่งมีความสำคัญอย่างยิ่งในการเลือกใช้แบบจำลองที่เหมาะสม หากข้อมูลมีแนวโน้มระยะยาว หรือมีฤดูกาลชัดเจน อาจจำเป็นต้องใช้แบบจำลองที่สามารถรองรับลักษณะเหล่านี้ เช่น SARIMA หรือ Prophet ในงานวิจัยนี้ได้เลือกใช้เทคนิค Time Series Decomposition เพื่อให้เห็นโครงสร้างภายในของข้อมูลอย่างชัดเจน แสดงดังภาพประกอบ 24 ที่แสดงผลการแยกองค์ประกอบของข้อมูล room_sold จากชุดข้อมูลฝึกสอน (Train Data) ออกเป็น 3 ส่วนหลัก ได้แก่ Trend, Seasonality, และ Residual



ภาพประกอบ 24 แสดงค่า Trend Seasonal และ Residual ของ Train Data

แนวโน้ม (Trend) แสดงให้เห็นว่าจำนวนห้องพักที่ขายได้มีการเพิ่มขึ้นอย่างต่อเนื่องในช่วงปี 2019 ก่อนจะลดลงเล็กน้อยในต้นปี 2020 ซึ่งสะท้อนถึงแนวโน้มของข้อมูล ในฤดูกาลที่มีความต้องการสูง ส่วนฤดูกาล (Seasonality) แสดงรูปแบบที่เกิดซ้ำเป็นประจำทุกปี เช่น ยอดขายที่พุ่งสูงในช่วงเดือนสิงหาคม ซึ่งสัมพันธ์กับช่วงเทศกาลท่องเที่ยว ขณะที่ค่าคงเหลือ (Residual) แสดงความเบี่ยงเบนของข้อมูลที่ไม่สามารถอธิบายได้ด้วยแนวโน้ม หรือฤดูกาล โดยภาพรวมมีความผันผวนไม่สูงมากนัก แสดงให้เห็นว่าโครงสร้างหลักของข้อมูลสามารถอธิบายได้ดีผ่านสององค์ประกอบแรก

การตรวจสอบองค์ประกอบเหล่านี้ช่วยให้ทราบว่าข้อมูลมีทั้งแนวโน้มระยะยาว และฤดูกาลที่ชัดเจน จึงสมควรเลือกใช้แบบจำลองที่สามารถรองรับโครงสร้างลักษณะดังกล่าว เช่น SARIMA ซึ่งออกแบบมาเพื่อจัดการกับฤดูกาลโดยตรง หรือ Prophet ซึ่งมีความยืดหยุ่นสูงในการประมวลผล

แนวโน้มไม่เป็นเชิงเส้นและฤดูกาลหลายระดับ ซึ่งการตัดสินใจในขั้นตอนนี้มีส่วนสำคัญในการส่งผลต่อความแม่นยำของผลการพยากรณ์ในขั้นตอนถัดไป

แบบจำลอง Autoregressive Integrated Moving Average (ARIMA)

ในงานวิจัยนี้ ได้ใช้แบบจำลอง Autoregressive Integrated Moving Average (ARIMA) เพื่อพยากรณ์จำนวนห้องพักที่ขายได้ต่อวัน (Daily room sold) สำหรับห้องพักประเภท A โดยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่

1. ข้อมูลฝึกสอน (Training Data) ใช้ข้อมูล 80% แรกของชุดข้อมูล ตั้งแต่เดือนกรกฎาคม 2018 ถึงเดือนตุลาคม 2019

2. ข้อมูลทดสอบ (Testing Data) ใช้ข้อมูล 20% สุดท้ายของชุดข้อมูล ตั้งแต่เดือนพฤศจิกายน 2019 ถึงเดือนตุลาคม 2020

การทำงานของ ARIMA เหมาะสมสำหรับข้อมูลอนุกรมเวลาที่มีความคงที่ (Stationary) ซึ่งในกรณีที่ข้อมูลไม่มีความคงที่ จะต้องดำเนินการปรับข้อมูล เพื่อเปลี่ยนข้อมูลให้มีความคงที่ก่อนนำไปสร้างแบบจำลอง โดยมีขั้นตอนดังนี้

3. จากการทำ Time Series Decomposition พบว่าข้อมูลมีความแปรปรวน จึงควรใช้ Log Transformation เพื่อช่วยลดผลกระทบของค่าผิดปกติที่สูงมาก

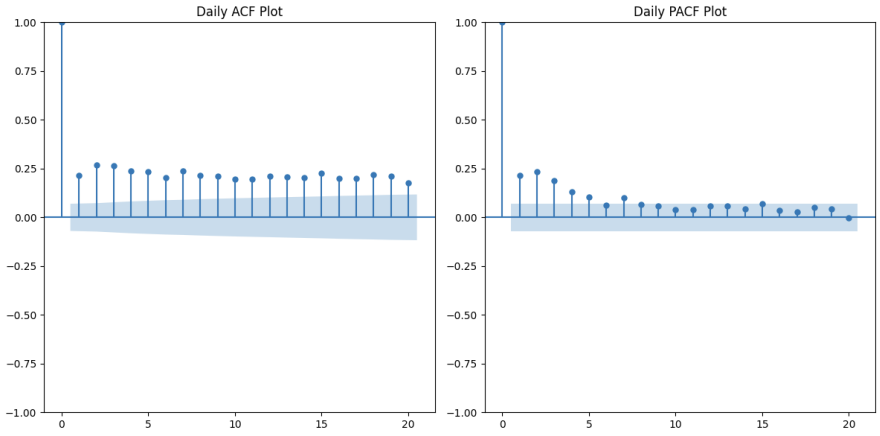
4. ตรวจสอบความคงที่ของข้อมูลโดยใช้กราฟ ADF Test (Augmented Dickey-Fuller Test) เพื่อตรวจสอบว่าข้อมูลมีความคงที่หรือไม่ โดยเทียบกับค่าวิกฤต = 0.05 หากค่า P-value < 0.05 แสดงว่าข้อมูลมีความคงที่ (Stationary) โดยผลลัพธ์ที่ได้คือ ADF Statistic มีค่า -3.634662156787051 ถือว่าค่า ADF ต่ำมาก ซึ่งหมายความว่าข้อมูลมีแนวโน้มเป็น Stationary p-value มีค่า 0.005131909459014963 ถือว่าค่า P-value < 0.05 หมายความว่าข้อมูลเป็น Stationary

5. พบว่าข้อมูลเป็น Stationary แล้ว สามารถ Train-Test Split ได้เลย แบ่งเป็น

Train Data: 633 records

Test Data: 159 records

ใช้กราฟ ACF (Autocorrelation Function) และ PACF (Partial Autocorrelation Function) เพื่อกำหนดพารามิเตอร์ของแบบจำลอง โดยผลลัพธ์แสดงดังภาพประกอบที่ 25



ภาพประกอบ 25 ผลการวิเคราะห์ ACF และ PACF สำหรับ Train data

จากการวิเคราะห์พบว่าพารามิเตอร์ที่เหมาะสมสำหรับแบบจำลองคือ ARIMA(1, 1, 1)

1. สร้างแบบจำลอง ARIMA ด้วยพารามิเตอร์ที่ได้จากการวิเคราะห์ และทำการฝึกแบบจำลองโดยใช้ชุดข้อมูลฝึกสอน (train_data)

2. พยากรณ์ข้อมูล โดยใช้แบบจำลอง ARIMA ที่ผ่านการฝึกฝนเพื่อพยากรณ์จำนวนห้องพักที่ขายได้ในช่วงเวลาทดสอบ โดยเปรียบเทียบค่าจริง และค่าที่พยากรณ์

3. ทำการประเมินประสิทธิภาพของแบบจำลอง โดยใช้ตัวชี้วัดในการประเมินความแม่นยำของแบบจำลองคือ MAE, MAPE, และ RMSE

แบบจำลอง Seasonal Autoregressive Integrated Moving Average (SARIMA)

ในงานวิจัยนี้ ได้ใช้แบบจำลอง Seasonal Autoregressive Integrated Moving Average (SARIMA) เพื่อพยากรณ์จำนวนห้องพักที่ขายได้ต่อวัน (Daily room sold) สำหรับห้องพักประเภท A โดยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่

1. ข้อมูลฝึกสอน (Training Data) ใช้ข้อมูล 80% แรกของชุดข้อมูล ตั้งแต่เดือนกรกฎาคม 2018 ถึงเดือนตุลาคม 2019

2. ข้อมูลทดสอบ (Testing Data) ใช้ข้อมูล 20% สุดท้ายของชุดข้อมูล ตั้งแต่เดือนพฤศจิกายน 2019 ถึงเดือนตุลาคม 2020

แบบจำลอง SARIMA เหมาะสำหรับการพยากรณ์ข้อมูลอนุกรมเวลาที่มีลักษณะตามฤดูกาล (Seasonality) และจำเป็นต้องมีความคงที่ (Stationary) เช่นเดียวกับแบบจำลอง ARIMA ดังนั้น

ในกรณีที่ข้อมูลยังไม่มีความคงที่ จะต้องดำเนินการปรับข้อมูล เพื่อกำจัดลักษณะตามฤดูกาล เพื่อให้ข้อมูลมีความคงที่ก่อนที่จะนำไปสร้างแบบจำลอง SARIMA โดยมีขั้นตอนดังนี้

1. เนื่องจากข้อมูลมีความแปรปรวน ใช้ Log Transformation เพื่อช่วยลดผลกระทบของค่าผิดปกติที่สูงมาก

2. เช่นเดียวกับแบบจำลอง ARIMA ใช้ ACF (Autocorrelation Function) และ PACF (Partial Autocorrelation Function) เพื่อกำหนดพารามิเตอร์ของแบบจำลอง โดยผลลัพธ์แสดงดังภาพประกอบที่ 25 จากผลการวิเคราะห์ ACF และ PACF กำหนดค่าพารามิเตอร์ของแบบจำลอง โดยแบ่งเป็น (p,d,q) พารามิเตอร์ของ ARIMA และ(P,D,Q,m) พารามิเตอร์สำหรับ SARIMA ผลจากการวิเคราะห์พบว่าค่าพารามิเตอร์ที่เหมาะสมคือ SARIMA(2,1,2)(1,0,2)(7)

3. ตรวจสอบความคงที่ของข้อมูล (Stationarity) โดยใช้ ADF Test (Augmented Dickey-Fuller Test) เพื่อตรวจสอบว่าข้อมูลมีความคงที่หรือไม่ โดยพิจารณาค่า p-value เมื่อเทียบค่าวิกฤต = 0.05 หากค่า P-value < 0.05 แสดงว่าข้อมูลมีความคงที่ โดยผลลัพธ์ที่ได้คือ ADF Statistic มีค่า -3.634662156787051 ถือว่าค่า ADF ต่ำมาก ซึ่งหมายความว่าข้อมูลมีแนวโน้มเป็น Stationary p-value มีค่า 0.005131909459014963 ถือว่าค่า P-value < 0.05 หมายความว่าข้อมูลเป็น Stationary

4. พบว่าข้อมูลเป็น Stationary แล้ว สามารถ Train-Test Split ได้เลย แบ่งเป็น

Train Data: 633 records

Test Data: 159 records

5. สร้างแบบจำลอง SARIMA โดยใช้ค่าพารามิเตอร์ที่กำหนดจากการวิเคราะห์ นำมาสร้างแบบจำลอง SARIMA และทำการฝึกฝนแบบจำลองด้วยข้อมูลฝึกสอน (Training Data)

6. ใช้แบบจำลอง SARIMA ที่ผ่านการฝึกฝน ทำการพยากรณ์จำนวนห้องพักที่ขายได้ในช่วงเวลาทดสอบ (Testing Data) แล้วนำมาเปรียบเทียบค่าจริง (Test Data) กับค่าที่พยากรณ์ (Forecast)

7. ทำการประเมินประสิทธิภาพของแบบจำลอง โดยใช้ตัวชี้วัดในการประเมินความแม่นยำของแบบจำลองคือ MAE, MAPE, และ RMSE

แบบจำลอง ARIMAX และแบบจำลอง SARIMAX

ในงานวิจัยนี้ ได้ใช้แบบจำลอง *ARIMAX (ARIMA with Exogenous Variables)* และแบบจำลอง*SARIMAX (SARIMA with Exogenous Variables)* ซึ่งเป็นแบบจำลองที่พัฒนาต่อยอดจาก *ARIMA* และ*SARIMA* โดยเพิ่มตัวแปรอิสระภายนอก (Exogenous Variables) เข้ามาในกระบวนการ

พยากรณ์ เช่น อัตราค่าห้องพักเฉลี่ย (ADR), Lead Time, จำนวนคืนที่เข้าพัก และปัจจัยฤดูกาล เพื่อช่วยเพิ่มประสิทธิภาพในการพยากรณ์ของแบบจำลอง โดยจะทำการพยากรณ์จำนวนห้องพักที่ขายได้ต่อวัน(Daily room sold) สำหรับห้องพักประเภท A แบ่งข้อมูลออกเป็น 2 ชุด ได้แก่

1. ข้อมูลฝึกสอน (Training Data) ใช้ข้อมูล 80% แรกของชุดข้อมูล ตั้งแต่เดือนกรกฎาคม 2018 ถึงเดือนตุลาคม 2019

2. ข้อมูลทดสอบ (Testing Data) ใช้ข้อมูล 20% สุดท้ายของชุดข้อมูล ตั้งแต่เดือนพฤศจิกายน 2019 ถึงเดือนตุลาคม 2020

ขั้นตอนการสร้างแบบจำลอง ARIMAX และ SARIMAX มีรายละเอียดดังนี้

1. เนื่องจากข้อมูลเป็นชุดเดียวกันที่ใช้ในการพยากรณ์ ข้อมูลจึงผ่านการทำ *log transformation* เพื่อช่วยลดผลกระทบของค่าผิดปกติที่สูงมาก

2. เตรียมข้อมูล และทำการเลือกตัวแปรอิสระ (*Exogenous variables)* ที่ใช้ร่วมในการพยากรณ์ เช่น *ADR, lead time, high season, total nights, total guests*

3. ตรวจสอบความคงที่ของข้อมูล (*Stationarity)* เช่นเดียวกับแบบจำลอง *ARIMA* และ *SARIMA*โดยใช้ *ADF Test (Augmented Dickey-Fuller Test)* เพื่อตรวจสอบว่าข้อมูลมีความคงที่หรือไม่ โดยพิจารณาค่า *p-value* เมื่อเทียบค่าวิกฤต = 0.05 หากค่า *P-value* < 0.05 แสดงว่าข้อมูลมีความคงที่ โดยผลลัพธ์ที่ได้คือ *ADF Statistic* มีค่า *-3.634662156787051* ถือว่าค่า *ADF* ต่ำมาก ซึ่งหมายความว่าข้อมูลมีแนวโน้มเป็น *Stationary p-value* มีค่า *0.005131909459014963* ถือว่าค่า *P-value* < 0.05 หมายความว่าข้อมูลเป็น *Stationary*

4. พบว่าข้อมูลเป็น Stationary แล้ว สามารถ Train-Test Split ได้เลย แบ่งเป็น

Train Data: 633 records

Test Data: 159 records

5. สร้างแบบจำลอง *ARIMAX* หรือ *SARIMAX* โดยใช้ข้อมูลฝึกสอน *(Traing Data)* พร้อมกับชุดของ *Exogenous Variables* ที่เลือกไว้ แล้วทำการฝึกแบบจำลองด้วยชุดข้อมูล *Training*

6. ใช้แบบจำลอง *ARIMAX* หรือ*SARIMAX* ที่ผ่านการฝึกฝน ทำการพยากรณ์จำนวนห้องพักที่ขายได้ในช่วงเวลาทดสอบ (*Testing Data)* แล้วนำมาเปรียบเทียบค่าจริง (*Test Data)* กับค่าที่พยากรณ์ (*Forecast)*

7. ทำการประเมินประสิทธิภาพของแบบจำลอง โดยใช้ตัวชี้วัดในการประเมินความแม่นยำของแบบจำลองคือ *MAE, MAPE,* และ *RMSE*

แบบจำลอง Prophet

ในงานวิจัยนี้ ได้ใช้แบบจำลอง Prophet เพื่อพยากรณ์จำนวนห้องพักที่ขายได้ต่อวัน (Daily room sold) สำหรับห้องพักประเภท A โดยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่

1. ข้อมูลฝึกสอน (Training Data) ใช้ข้อมูล 80% แรกของชุดข้อมูล ตั้งแต่เดือนกรกฎาคม 2018 ถึงเดือนกุมภาพันธ์ 2020

2. ข้อมูลทดสอบ (Testing Data) ใช้ข้อมูล 20% สุดท้ายของชุดข้อมูล ตั้งแต่เดือนมีนาคม 2020 ถึงเดือนสิงหาคม 2020

ขั้นตอนการสร้างแบบจำลอง Prophet มีรายละเอียดดังนี้

1. เนื่องจากข้อมูลเป็นชุดเดียวกันที่ใช้ในการพยากรณ์ ข้อมูลจึงผ่านการทำ *log transformation* เพื่อช่วยลดผลกระทบของค่าผิดปกติที่สูงมาก

2. เตรียมข้อมูลสำหรับ *Prophet* โดยทำการแปลงข้อมูล *Datetime Index* ให้เป็นคอลัมน์ และตั้งชื่อคอลัมน์เป็น *ds (*วันที่) และ *y (*ค่าที่ต้องการพยากรณ์) เพื่อให้ตรงตามรูปแบบที่ *Prophet* ต้องการ

3. เลือกตัวแปรอิสระ (*Exogenous variables)* ที่ใช้ร่วมในการพยากรณ์ เช่น *ADR, lead time, high season, total nights, total guests*

4. แบ่งข้อมูลออกเป็น ชุดข้อมูลฝึกสอน (*Training Data)* และ ชุดข้อมูลทดสอบ (*Testing Data)* โดยใช้ช่วงเวลา 80% สำหรับการฝึก และ 20% สำหรับทดสอบ

5. สร้างแบบจำลอง *Prophet* โดยตั้งค่าดังนี้

5.1 *seasonality_mode='Multiplicative'* เนื่องจากข้อมูลแสดงลักษณะของแนวโน้มแบบเพิ่มขึ้นอย่างสม่ำเสมอ

5.2 เปิดใช้งาน *yearly_seasonality=True* และ*weekly_seasonality=True* เพื่อรองรับรูปแบบตามฤดูกาล

5.3 กำหนดค่า *changepoint_prior_scale* เพื่อลดการ *overfitting*

5.4 เพิ่มตัวแปรช่วยพยากรณ์ (*Exogenous variables)* ได้แก่ *ADR, lead time, high season, total nights, total guests* ทำการพยากรณ์ข้อมูลกำหนดช่วงเวลาที่ต้องการพยากรณ์ โดยกำหนดไว้ที่ *6* เดือนข้างหน้า และใช้แบบจำลอง *Prophet* ที่ผ่านการฝึกฝนทำการพยากรณ์จำนวนห้องพักที่ขายได้ในช่วงเวลาทดสอบ (*Testing Data)*

5.5 ทำการค้นหาจุดเปลี่ยนแปลงสำคัญ (*Changepoints)* ในข้อมูลอนุกรมเวลา ซึ่งแสดงถึงช่วงเวลาที่แนวโน้มของข้อมูลเปลี่ยนแปลงจุดเปลี่ยนที่ค้นพบจะแสดงด้วยเส้นสีแดงเพื่อช่วยในการวิเคราะห์แนวโน้ม

5.6 เปรียบเทียบค่าที่พยากรณ์ โดยทำการเปรียบเทียบค่าจริง (*Actual)* กับค่าที่พยากรณ์ (*Forecast)* โดยใช้ข้อมูลทดสอบ (*Testing Data)*

5.7 ทำการประเมินผลการพยากรณ์ด้วย *MAE*, *MAPE*, และ *RMSE* เพื่อวัดความแม่นยำของแบบจำลอง

บทที่ 4
ผลการดำเนินการวิจัย

ในการวิจัยนี้ ผู้วิจัยได้ดำเนินการวิเคราะห์ และพัฒนาแบบจำลองสำหรับการพยากรณ์จำนวนห้องพักที่ขายได้ (Room Sold) ซึ่งเป็นตัวแปรที่มีความสัมพันธ์โดยตรงกับรายได้ของโรงแรม และถือเป็นองค์ประกอบสำคัญในการบริหารรายได้ (Revenue Management) โดยมุ่งเน้นการศึกษาความแม่นยำในการพยากรณ์ของแบบจำลองแต่ละประเภท เพื่อสนับสนุนการวางแผนด้านรายได้ และการจัดสรรทรัพยากรของโรงแรมได้อย่างมีประสิทธิภาพ ทั้งนี้รายได้ของโรงแรม จะสามารถคำนวณได้จากการนำจำนวนห้องพักที่ขายได้จากการพยากรณ์ไปคูณกับอัตราค่าห้องพักเฉลี่ย (ADR) ในช่วงเวลานั้น ๆ เพื่อควบคุมตัวแปรด้านราคาที่มีความผันผวนสูง และเพื่อประเมินประสิทธิภาพของแบบจำลองในการพยากรณ์รายได้ของโรงแรมจากจำนวนห้องพักที่ขายได้

การสร้างแบบจำลอง ทำการเลือกแบบจำลองวิเคราะห์อนุกรมเวลาที่ได้รับความนิยม ได้แก่ Prophet, ARIMA, ARIMAX, SARIMA, SARIMAX เพื่อศึกษาความแม่นยำในการพยากรณ์ของแบบจำลองแต่ละประเภท และเพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์ของแบบจำลองที่ใช้ตัวแปรเสริม (Exogenous Variables) โดยสามารถแบ่งการดำเนินการออกเป็นหัวข้อดังนี้

1. ผลลัพธ์ของการสร้างแบบจำลอง Prophet
2. ผลลัพธ์ของการสร้างแบบจำลอง ARIMA
3. ผลลัพธ์ของการสร้างแบบจำลอง SRIMA
4. ผลลัพธ์ของการสร้างแบบจำลอง ARIMAX
5. ผลลัพธ์ของการสร้างแบบจำลอง SARIMAX
6. ผลลัพธ์จากการเปรียบเทียบของแบบจำลอง Prophet กับ ARIMA, ARIMAX, SARIMA, SARIMAX
7. ผลลัพธ์จากการเปรียบเทียบของแบบจำลอง Prophet กรณีไม่ใช้ และใช้ตัวแปรเสริม

ผลลัพธ์ของการสร้างแบบจำลอง Prophet

Prophet เป็นแบบจำลองการวิเคราะห์อนุกรมเวลาที่ทันสมัย มีจุดเด่นในการรองรับข้อมูลที่มีแนวโน้ม (Trend) และฤดูกาล (Seasonality) อย่างชัดเจน รวมถึงสามารถวิเคราะห์จุดเปลี่ยนแปลงแนวโน้ม (changepoints) ที่เกิดขึ้นได้

การตั้งค่าพารามิเตอร์ของ Prophet ผู้วิจัยได้กำหนดค่าพารามิเตอร์ของแบบจำลอง Prophet ดังต่อไปนี้ โดยเลือก seasonality_mode คือ multiplicative เพื่อให้แบบจำลองสามารถจับฤดูกาลที่สัมพันธ์กับระดับของข้อมูล, yearly_seasonality คือ True เพื่อให้แบบจำลองจับลักษณะของฤดูกาลที่เกิดขึ้นในแต่ละปี, weekly_seasonality คือ True เพื่อให้แบบจำลองจับลักษณะของฤดูกาลที่เกิดขึ้นในแต่ละอาทิตย์ และ changepoint_prior_scale คือ 0.02 เพื่อลดการ overfitting และเพิ่มความสามารถของแบบจำลองในการจับจุดเปลี่ยนแนวโน้ม

ตัวแปรช่วยในการพยากรณ์ (Exogenous variables) ในการเพิ่มประสิทธิภาพของแบบจำลองในการพยากรณ์จำนวนห้องที่ขายได้ (room_sold) ผู้วิจัยได้เลือกใช้ตัวแปรเสริมที่สำคัญ ได้แก่ Lead time หรือระยะเวลาการจองก่อนเข้าพัก, High Season หรือตัวแปรที่ระบุว่าช่วงเวลานั้นอยู่ในฤดูกาลท่องเที่ยวหรือไม่, *Total night* หรือจำนวนคืนที่เข้าพักทั้งหมด, Total guest หรือจำนวนแขกที่เข้าพักทั้งหมด

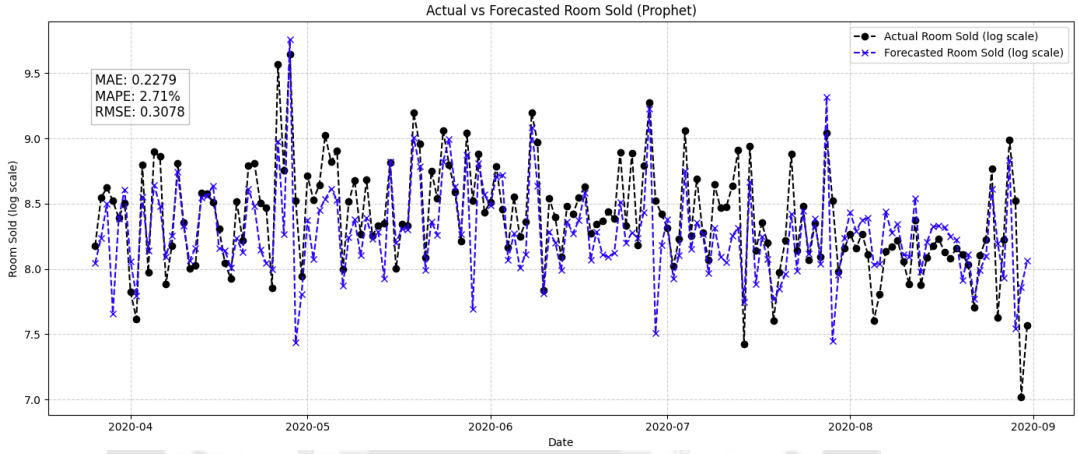
หลังจากที่ได้ดำเนินการฝึกแบบจำลอง Prophet ด้วยข้อมูลชุดฝึกสอน (Training Data) และทดสอบด้วยข้อมูลชุดทดสอบ (Testing Data) พบว่า Prophet มีศักยภาพในการจำลองแนวโน้มเชิงเวลา และฤดูกาลของข้อมูลจำนวนห้องพักที่ขายได้ในแต่ละเดือนได้ดี โดยเฉพาะอย่างยิ่งเมื่อมีการนำตัวแปรเสริม เช่น Lead Time, High Season, Total night และ Total guest เข้ามาใช้ร่วมในการพยากรณ์ ประสิทธิภาพของแบบจำลองในการพยากรณ์จำนวนห้องพัก มีรายละเอียดดังนี้

Mean Absolute Error (MAE) เท่ากับ 0.2279 แสดงถึงค่าคลาดเคลื่อนเฉลี่ยต่อเดือนเมื่อเทียบกับค่าจริง ซึ่งถือว่าอยู่ในระดับต่ำ

Mean Absolute Percentage Error (MAPE) เท่ากับ 2.71% แสดงถึงค่าความคลาดเคลื่อนเฉลี่ยประมาณร้อยละ 2.71 ของค่าจริงในแต่ละเดือน ถือเป็นค่าความผิดพลาดที่ค่อนข้างต่ำ

Root Mean Square Error (RMSE) เท่ากับ 0.3078 แสดงถึงระดับของข้อผิดพลาดโดยรวมของการพยากรณ์ โดยค่านี้บ่งชี้ว่า แบบจำลอง Prophet สามารถรักษาความแม่นยำในการติดตามแนวโน้ม และฤดูกาลของข้อมูลได้อย่างต่อเนื่อง

เมื่อเปรียบเทียบผลการพยากรณ์กับข้อมูลจริงในช่วงทดสอบ(Testing Data) ดังแสดงในภาพประกอบที่ 26 พบว่า Prophet สามารถจับแนวโน้มจำนวนห้องพักที่ขายได้ในแต่ละวันได้ดีในส่วนใหญ่ โดยเฉพาะในช่วงเดือนเมษายน ถึงกรกฎาคม ปี พ.ศ. 2020 ซึ่งค่าที่พยากรณ์มีทิศทางสอดคล้องกับค่าจริง อย่างไรก็ตามในบางช่วงเวลา โดยเฉพาะช่วงต้นเดือนพฤษภาคม และปลายเดือนสิงหาคม พบว่าข้อมูลจริงมีการเปลี่ยนแปลงที่ค่อนข้างผันผวน และฉับพลัน ในขณะที่ค่าที่พยากรณ์ได้จากแบบจำลองมีการตอบสนองช้ากว่าเล็กน้อย ส่งผลให้เกิดความคลาดเคลื่อนที่ชัดเจนในช่วงดังกล่าว



ภาพประกอบ 26 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง Prophet

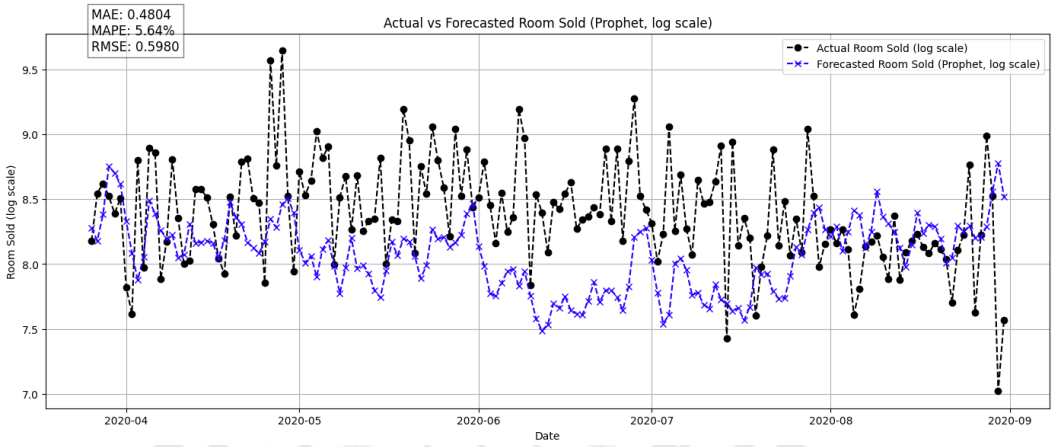
จากภาพประกอบ 26 ผลลัพธ์ข้างต้นสะท้อนว่า แม้แบบจำลอง Prophet จะมีความสามารถในการพยากรณ์ข้อมูลที่มีลักษณะของแนวโน้ม และฤดูกาลได้อย่างมีประสิทธิภาพ อย่างไรก็ตามในช่วงที่ข้อมูลมีการเปลี่ยนแปลงอย่างฉับพลัน หรือมีปัจจัยที่ไม่สามารถอธิบายได้จากตัวแปรอิสระที่ใช้ในการฝึกแบบจำลอง อาจส่งผลให้ความแม่นยำของค่าพยากรณ์ลดลง จึงนับว่าเป็นข้อจำกัดในการติดตามข้อมูลที่มีความเปลี่ยนแปลงอย่างรวดเร็ว หรือช่วงที่ข้อมูลมีความผันผวนสูง

ผลของตัวแปรช่วย ต่อประสิทธิภาพของแบบจำลอง Prophet

เพื่อประเมินผลของตัวแปรช่วย (Exogenous variables) ที่มีต่อประสิทธิภาพของแบบจำลอง Prophet ผู้วิจัยได้ทำการเปรียบเทียบความแม่นยำของแบบจำลองระหว่างกรณีที่ไม่มีการใช้ตัวแปรเสริม และกรณีที่เพิ่มตัวแปรเสริมที่เกี่ยวข้องกับพฤติกรรมการจองห้องพัก ได้แก่ Lead Time, High Season, Total Night และ Total Guest ซึ่งเป็นตัวแปรที่มีนัยสำคัญทางธุรกิจในการอธิบายพฤติกรรมความต้องการของลูกค้า นำมาใช้ในการฝึกแบบจำลอง

จากผลการวิเคราะห์เชิงเปรียบเทียบพบว่า การไม่ใช้ตัวแปรช่วยส่งผลให้ค่าความคลาดเคลื่อนของแบบจำลองเพิ่มขึ้นอย่างชัดเจน โดยแบบจำลอง Prophet ที่ไม่ใช้ Exogenous variables มีค่า Mean

Absolute Error (MAE) เท่ากับ อยู่ที่ 0.4804, Root Mean Square Error (RMSE) เท่ากับ 0.5980 และ Mean Absolute Percentage Error (MAPE) สูงถึง 5.64% จากผลลัพธ์ชี้ให้เห็นว่าแบบจำลองไม่สามารถติดตามแนวโน้มของข้อมูลได้อย่างแม่นยำในกรณีที่ไม่มีข้อมูลบริบทเพิ่มเติมมาสนับสนุน แสดงดังภาพประกอบที่ 27



ภาพประกอบ 27 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของ Prophet โดยมีตัวแปรช่วย

จากภาพประกอบที่ 27 กราฟแสดงผลการเปรียบเทียบจำนวนห้องพักที่ขายได้จริงกับค่าที่พยากรณ์โดยใช้แบบจำลอง Prophet ในกรณีที่ไม่ได้เพิ่มตัวแปรช่วย (Exogenous variables) จะเห็นได้ชัดว่าแบบจำลองไม่สามารถติดตามแนวโน้มของข้อมูลจริงได้อย่างมีประสิทธิภาพ แสดงให้เห็นลักษณะของค่าพยากรณ์ที่แปรปรวนเกินจริง หรือผิดเพี้ยนจากแนวโน้มจริงของข้อมูล โดยเฉพาะช่วงกลางเดือนพฤษภาคม และต้นเดือนกรกฎาคม 2020 ที่ข้อมูลจริงมีแนวโน้มลดลง แต่แบบจำลองกลับพยากรณ์สูงกว่าความเป็นจริงอย่างมีนัยสำคัญซึ่งไม่สอดคล้องกับรูปแบบของข้อมูลจริงที่มีการเปลี่ยนแปลงอย่างต่อเนื่องในลักษณะโค้งระหว่างเดือนเมษายน ถึงกันยายน 2020

ในทางตรงกันข้าม แบบจำลอง Prophet ที่มีการเพิ่มตัวแปรช่วย มีค่า MAE ลดลงเหลือเพียง 0.2279 หน่วย, RMSE ลดลงเหลือ 0.3078 และ MAPE ลดลงเหลือ 2.71% แสดงให้เห็นถึงการปรับปรุงประสิทธิภาพของแบบจำลองอย่างมีนัยสำคัญ ดังแสดงในภาพประกอบที่ 26

จากการเปรียบเทียบผลลัพธ์ระหว่างแบบจำลอง Prophet ที่ไม่มีตัวแปรเสริมกับแบบจำลองที่มีการเพิ่มตัวแปรเสริม พบว่า ตัวแปรช่วย (Exogenous variables) มีบทบาทสำคัญในการยกระดับประสิทธิภาพของแบบจำลอง โดยเฉพาะตัวแปรอย่าง Lead Time ซึ่งช่วยให้แบบจำลองสามารถเรียนรู้แนวโน้มการจองที่กำลังจะเกิดขึ้นในอนาคตได้ล่วงหน้า ส่งผลให้การตอบสนองของแบบจำลองแม่นยำมากยิ่งขึ้นในช่วงที่มีการเปลี่ยนแปลงของความต้องการของลูกค้า ขณะเดียวกันตัวแปร High Season มีบทบาทในการสะท้อนลักษณะตามฤดูกาลของอุตสาหกรรมโรงแรม ซึ่งเป็นปัจจัยที่ส่งผลกระทบอย่างมี

นัยสำคัญต่อความต้องการห้องพัก ส่วนตัวแปร Total Night และ Total Guest ช่วยอธิบายขนาดของการเข้าพักและพฤติกรรมของผู้บริโภค ซึ่งมีผลโดยตรงต่อจำนวนห้องที่ถูกจองจริงในแต่ละวัน

ผลการวิเคราะห์ชี้ให้เห็นว่า การเพิ่มตัวแปรเสริมที่มีความสัมพันธ์กับพฤติกรรมการจอง และปัจจัยแวดล้อมทำให้ Prophet สามารถเรียนรู้โครงสร้างของข้อมูลได้ลึกซึ้งขึ้น ติดตามแนวโน้มได้อย่างแม่นยำ และสามารถตอบสนองต่อสถานการณ์ที่มีความผันผวน เช่น ช่วงเทศกาลท่องเที่ยว หรือเหตุการณ์เฉพาะที่ส่งผลต่อการจอง

ผลลัพธ์ของการสร้างแบบจำลอง ARIMA

ARIMA เป็นแบบจำลองอนุกรมเวลาทางสถิติที่อาศัยหลักการของ Autoregressive (AR) Moving Average (MA) และ Differencing (I) เพื่อปรับข้อมูลให้คงที่ (Stationary) ก่อนทำการพยากรณ์

การกำหนดค่าพารามิเตอร์ของ ARIMA ในการวิจัยนี้ ได้ทำการวิเคราะห์ และกำหนดค่าพารามิเตอร์โดยอาศัยฟังก์ชัน Autocorrelation Function (ACF) และ Partial Autocorrelation Function (PACF) ผลจากการวิเคราะห์ และทำการปรับแต่งค่าพารามิเตอร์ ได้ค่าพารามิเตอร์ที่เหมาะสม ดังนี้

p = 1 (ค่าของตัวแปรอัตโนมัติ)

d = 1 (จำนวนครั้งที่ทำ Differencing เพื่อให้ข้อมูลคงที่)

q = 1 (ค่าของ Moving Average)

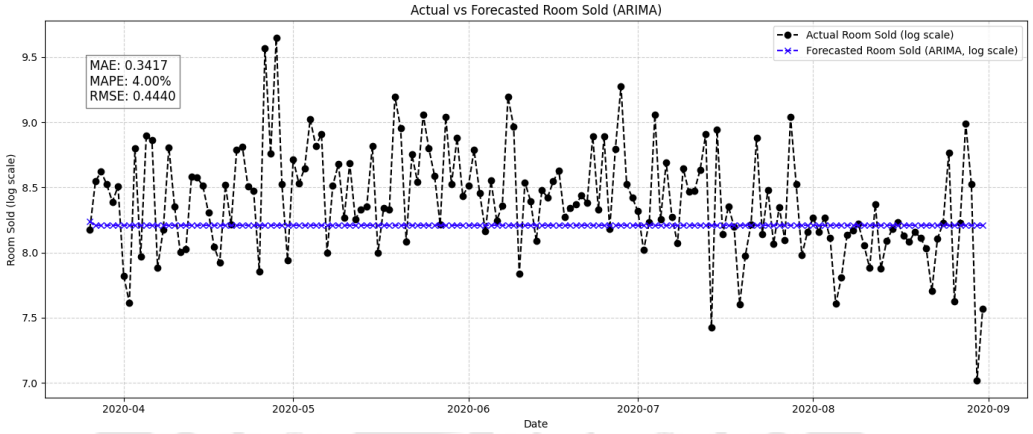
หลังจากการฝึกแบบจำลอง ARIMA ด้วยข้อมูลชุดฝึกสอน (Training Data) และการทดสอบด้วยชุดข้อมูลทดสอบ (Testing Data) พบว่า ARIMA มีศักยภาพในการจำลองแนวโน้มของข้อมูลจำนวนห้องพักที่ขายได้ในแต่ละวันได้ในระดับปานกลาง โดยผลการประเมินค่าความคลาดเคลื่อนจากการพยากรณ์จำนวนห้องพัก (Room Sold) มีรายละเอียดดังนี้

Mean Absolute Error (MAE) เท่ากับ 0.3417 สะท้อนถึงค่าความคลาดเคลื่อนเฉลี่ยต่อวันเมื่อเทียบกับค่าจริง ค่าดังกล่าวอยู่ในระดับที่ยอมรับได้ในบริบทของข้อมูลนี้

Mean Absolute Percentage Error (MAPE) เท่ากับ 4.00% แสดงให้เห็นว่าโดยเฉลี่ยค่าที่พยากรณ์มีความคลาดเคลื่อนประมาณร้อยละ 4 เมื่อเทียบกับค่าจริง ถือเป็นค่าความผิดพลาดที่ยังยอมรับได้

Root Mean Square Error (RMSE) เท่ากับ 0.444 บ่งชี้ถึงระดับข้อผิดพลาดรวมของการพยากรณ์ โดยค่าที่ได้แสดงให้เห็นถึงความผันผวนของข้อผิดพลาดในแต่ละวันซึ่งยังอยู่ในเกณฑ์ที่ยอมรับได้

เมื่อนำผลการพยากรณ์จากแบบจำลอง ARIMA มาเปรียบเทียบกับข้อมูลจริงในช่วงทดสอบ (Testing Data) ดังแสดงในภาพประกอบที่ 28 จะพบว่าค่าที่พยากรณ์ได้จากแบบจำลอง ARIMA มีลักษณะค่อนข้างราบเรียบและเปลี่ยนแปลงช้ากว่าค่าจริงอย่างเห็นได้ชัด โดยเฉพาะในช่วงที่ข้อมูลจริงเกิดความผันผวนหรือเปลี่ยนแปลงฉับพลัน เช่น ในเดือนพฤษภาคมและกรกฎาคม พ.ศ. 2020 ที่จำนวนห้องพักที่ขายได้มีการแกว่งตัวอย่างรุนแรง แต่ค่าพยากรณ์จาก ARIMA ยังคงอยู่ใกล้ระดับค่าเฉลี่ยของข้อมูลโดยไม่ตอบสนองต่อการเปลี่ยนแปลงเหล่านั้น



ภาพประกอบ 28 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง ARIMA

จากภาพประกอบที่ 28 แสดงให้เห็นถึงข้อจำกัดของแบบจำลอง ARIMA ในการพยากรณ์ข้อมูลที่มีความผันผวนสูง และเปลี่ยนแปลงอย่างฉับพลัน เนื่องจากโครงสร้างของ ARIMA อาศัยเพียงความสัมพันธ์ของข้อมูลในอดีตเป็นหลัก โดยไม่สามารถใช้ปัจจัยภายนอก หรือองค์ประกอบทางฤดูกาลเข้ามาอธิบายความเปลี่ยนแปลงในข้อมูลได้โดยตรง ส่งผลให้ค่าพยากรณ์ที่ได้ มีความแปรปรวนต่ำกว่าค่าจริง อย่างเห็นได้ชัด และไม่สามารถติดตามแนวโน้ม หรือขนาดของการเปลี่ยนแปลงในแต่ละวันได้อย่างแม่นยำ ด้วยเหตุนี้ทำให้เห็นข้อจำกัดของ ARIMA โดยเฉพาะอย่างยิ่งสำหรับข้อมูลในภาคธุรกิจโรงแรมที่มักได้รับอิทธิพลจากปัจจัยภายนอกหลากหลาย เช่น ฤดูกาล การจัดโปรโมชั่น หรือเหตุการณ์พิเศษ ซึ่งส่งผลต่อความต้องการห้องพัก และยอดขายอย่างมีนัยสำคัญ กล่าวคือ แม้ ARIMA จะสามารถติดตามแนวโน้มหลักของข้อมูลในอดีตได้ในระดับหนึ่ง แต่ยังขาดความสามารถในการตอบสนองต่อความผันผวน หรือโครงสร้างข้อมูลที่ซับซ้อน ทำให้ประสิทธิภาพในการพยากรณ์ในบริบทที่มีการเปลี่ยนแปลงรวดเร็ว หรือได้รับผลกระทบจากปัจจัยแฝงหลายด้าน ยังอยู่ในระดับจำกัด

ผลลัพธ์ของการสร้างแบบจำลอง SARIMA

SARIMA เป็นแบบจำลองที่พัฒนาต่อเนื่องมาจากแบบจำลอง ARIMA โดยมีการเพิ่มองค์ประกอบของฤดูกาล (Seasonality) เข้าไปในโครงสร้างของแบบจำลอง ส่งผลให้เหมาะสมกับข้อมูลที่มีลักษณะแนวโน้มและความผันผวนตามฤดูกาลอย่างชัดเจน ซึ่งเหมาะสมกับข้อมูลธุรกิจโรงแรมที่มีฤดูกาลท่องเที่ยวและช่วงเทศกาลที่ส่งผลต่อยอดขายห้องพักอย่างชัดเจน

ในการศึกษาวิจัยครั้งนี้ ได้มีการกำหนดค่าพารามิเตอร์สำหรับแบบจำลอง *SARIMA* ซึ่งแบ่งออกเป็น *2* ส่วนคือ องค์ประกอบที่ไม่ใช่ฤดูกาล (*Non-seasonal Component)* และองค์ประกอบฤดูกาล (*Seasonal Component)* ดังรายละเอียดต่อไปนี้

*p = 2, d = 1, q = 2 (*ค่าของ *ARIMA)*

*P = 1, D = 0, Q = 2, S = 7 (*ค่าของ *Seasonal Component)* โดยค่าความถี่ของฤดูกาล (*S=7)* จะสอดคล้องกับการสังเกตช่วงเปลี่ยนแปลงที่เกิดขึ้นแบบเป็นอาทิตย์ในข้อมูลโรงแรม

หลังจากที่ได้ดำเนินการฝึกแบบจำลอง *SARIMA* ด้วยข้อมูลชุดฝึกสอน (*Training Data)* และทดสอบด้วยข้อมูลชุดทดสอบ (*Testing Data)* พบว่า *SARIMA* มีความสามารถในการจับแนวโน้มเชิงเวลา และรูปแบบตามฤดูกาลของข้อมูลจำนวนห้องพักที่ขายได้ (*Room Sold)* ได้อย่างแม่นยำในระดับหนึ่ง โดยเฉพาะในช่วงที่ข้อมูลมีลักษณะต่อเนื่อง และแสดงพฤติกรรมตามฤดูกาลอย่างชัดเจน โดยการประเมินค่าความคลาดเคลื่อนในการพยากรณ์ของจำนวนห้องพักมีรายละเอียดดังนี้

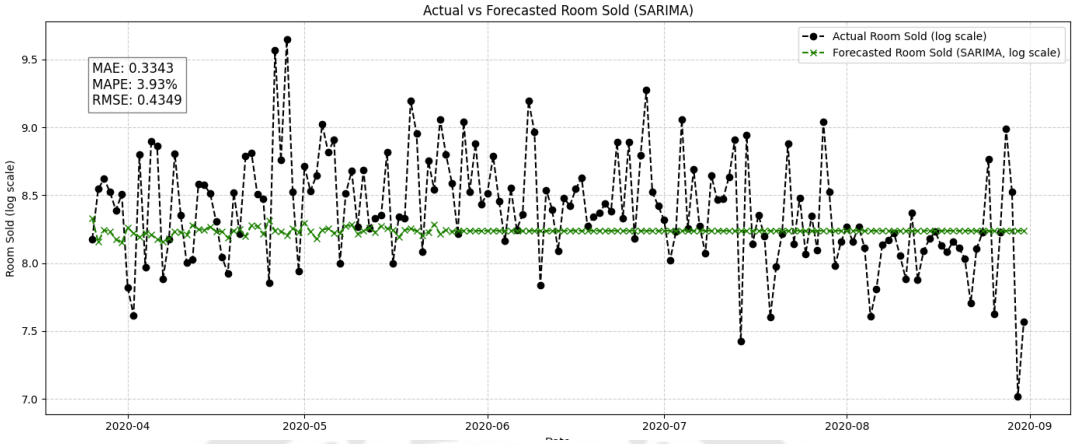
Mean Absolute Error (MAE) เท่ากับ *0.33* สะท้อนถึงค่าความคลาดเคลื่อนเฉลี่ยในแต่ละวันระหว่างค่าที่พยากรณ์ได้กับค่าจริง ค่าดังกล่าวอยู่ในระดับที่ยอมรับได้ในบริบทของข้อมูลนี้

Mean Absolute Percentage Error (MAPE) เท่ากับ *3.92%*แสดงให้เห็นว่าค่าการพยากรณ์ของ *SARIMA* มีความคลาดเคลื่อนเฉลี่ยเพียงประมาณร้อยละ *3.92* เมื่อเทียบกับค่าจริงในแต่ละวัน ถือเป็นค่าความผิดพลาดที่ยอมรับได้ในบริบทของข้อมูลนี้

Root Mean Square Error (RMSE) เท่ากับ *0.43* บ่งชี้ถึงระดับของข้อผิดพลาดโดยรวม โดยค่าที่ได้แสดงให้เห็นถึงความผันผวนของข้อผิดพลาดในแต่ละวันซึ่งยังอยู่ในเกณฑ์ที่ยอมรับได้

เมื่อเปรียบเทียบค่าที่พยากรณ์ได้จากแบบจำลอง *SARIMA* กับข้อมูลจริงในช่วงทดสอบ (*Testing Data)* ดังแสดงในภาพประกอบที่ *29* จะพบว่าค่าที่พยากรณ์ด้วย *SARIMA* มีแนวโน้มค่อนข้างราบเรียบ และมีความแปรปรวนน้อยกว่าค่าจริง อย่างเห็นได้ชัด โดยเฉพาะในช่วงเดือนมีนาคม ถึงกรกฎาคม *2020* แบบจำลองสามารถติดตามระดับกลางของแนวโน้มข้อมูลได้ในระดับ

หนึ่ง แต่ยังขาดความยืดหยุ่นในการตอบสนองต่อความผันผวนรายวัน หรือการเปลี่ยนแปลงที่เกิดขึ้นอย่างรวดเร็ว ของชุดข้อมูลทดสอบที่มีแนวโน้มแตกต่างไปจาก รูปแบบฤดูกาลเดิม



ภาพประกอบ 29 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง SARIMA

จากภาพประกอบที่ 29 แสดงให้เห็นว่าค่าที่พยากรณ์จาก SARIMA มักจะอยู่ในช่วงค่าเฉลี่ยของข้อมูลมากกว่าการจับแนวโน้มที่มีการขึ้นลงจริงในแต่ละวัน ทำให้แบบจำลองแสดงผลลัพธ์ที่มีค่าความคลาดเคลื่อนโดยรวมในระดับที่รับได้ แต่ยังมีข้อจำกัดในการจับลักษณะความผันผวน ที่เกิดขึ้นจริงในธุรกิจโรงแรม ด้วยผลลัพธ์นี้จึงทำให้เห็นข้อจำกัดสำคัญของแบบจำลอง ในการรับมือกับข้อมูลที่มีรูปแบบซับซ้อน และการเปลี่ยนแปลงอย่างเฉียบพลัน แม้ว่าจะมีองค์ประกอบฤดูกาลเข้ามาช่วยอธิบายข้อมูล แต่แบบจำลองก็ยังให้ผลลัพธ์ที่โน้มเอียงไปทางค่าเฉลี่ยมากกว่าการตอบสนองต่อความแปรปรวนในรายละเอียด ดังนั้น หากต้องการเพิ่มความแม่นยำในการพยากรณ์ในบริบทที่มีปัจจัยแฝงหรือฤดูกาลเปลี่ยนแปลงอย่างรวดเร็ว อาจจำเป็นต้องพิจารณาการใช้แบบจำลองที่สามารถรวมตัวแปรเสริมหรือปรับโครงสร้างให้ยืดหยุ่นมากยิ่งขึ้น เช่น SARIMAX หรือ Prophet เพื่อเพิ่มศักยภาพในการจับสัญญาณที่มีความซับซ้อนในข้อมูลจริง

ผลลัพธ์จากการสร้างแบบจำลอง ARIMAX

แบบจำลอง ARIMAX (ARIMA with Exogenous Variables) เป็นแบบจำลองอนุกรมเวลาที่พัฒนาต่อยอดจาก ARIMA โดยเพิ่มความสามารถในการนำตัวแปรเสริม (Exogenous Variables) ที่มีอิทธิพลต่อจำนวนห้องพักที่ขายได้ (Room Sold) เข้ามาอยู่ในโครงสร้างของแบบจำลอง ซึ่งช่วยให้สามารถอธิบายความเปลี่ยนแปลงของข้อมูลได้อย่างลึกซึ้งยิ่งขึ้น โดยเฉพาะในกรณีที่ข้อมูลได้รับผลกระทบจากปัจจัยภายนอก เช่น ฤดูกาล การจัดโปรโมชั่น หรือเหตุการณ์เฉพาะ

ในการวิจัยนี้ ได้กำหนดพารามิเตอร์ของ ARIMAX โดยอาศัยผลจากฟังก์ชัน Autocorrelation Function (ACF) และ Partial Autocorrelation Function (PACF) เช่นเดียวกับการกำหนดค่าของ ARIMA ร่วมกับการพิจารณาตัวแปรเสริมที่มีความสัมพันธ์เชิงสถิติอย่างมีนัยสำคัญ ได้แก่ Lead Time, High Season, Total Night และ Total Guest โดยค่าพารามิเตอร์ที่เหมาะสมสำหรับแบบจำลอง ARIMAX มีดังนี้

p = 2 (ค่าของตัวแปรอัตโนมัติ)

d = 1 (จำนวนครั้งที่ทำ Differencing เพื่อให้ข้อมูลคงที่)

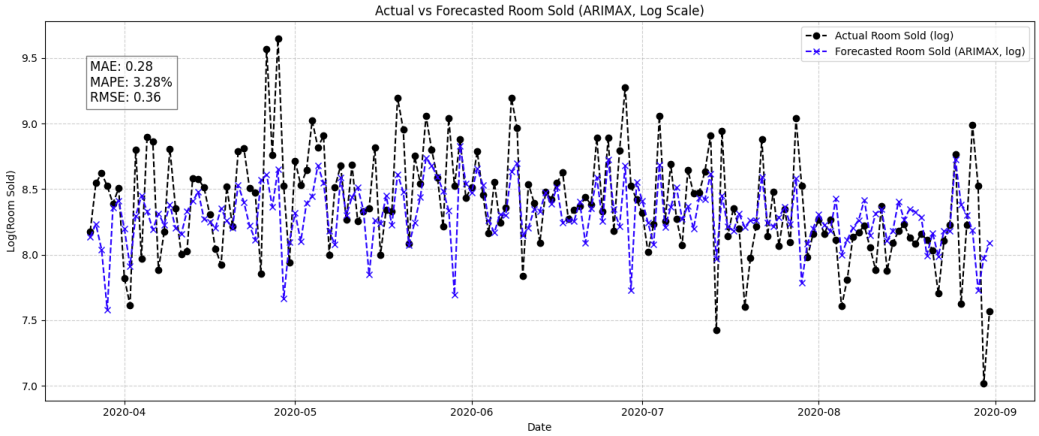
q = 2 (ค่าของ Moving Average)

เมื่อฝึกแบบจำลองด้วยชุดข้อมูลฝึกสอน (Training Data) และประเมินด้วยชุดข้อมูลทดสอบ (Testing Data) พบว่าแบบจำลอง ARIMAX แสดงผลลัพธ์ที่มีความแม่นยำสูงกว่าแบบจำลอง ARIMA โดยไม่ใช้ตัวแปรเสริม ดังแสดงในค่าประเมินผลการพยากรณ์ดังนี้

Mean Absolute Error (MAE) *เท่ากับ*0.28 สะท้อนค่าความคลาดเคลื่อนเฉลี่ยต่อวันที่อยู่ในระดับต่ำ แสดงว่าแบบจำลองสามารถติดตามค่าจริงได้อย่างใกล้เคียง

Mean Absolute Percentage Error (MAPE) เท่ากับ 3.28% แสดงให้เห็นว่าโดยเฉลี่ยแบบจำลองมีค่าความคลาดเคลื่อนเพียงร้อยละ 3.28 เมื่อเทียบกับค่าจริง ซึ่งถือว่าอยู่ในระดับที่ดี

จากภาพประกอบที่ 30 แสดงผลการพยากรณ์จำนวนห้องพักที่ขายได้ (Room Sold) โดยใช้แบบจำลอง ARIMAX เมื่อเปรียบเทียบกับข้อมูลจริงในช่วงชุดข้อมูลทดสอบ (Testing Data) จะเห็นได้ว่าแบบจำลองสามารถติดตามแนวโน้มของข้อมูลได้อย่างใกล้เคียงกับค่าจริงในแต่ละวัน โดยเฉพาะในช่วงที่ข้อมูลมีความผันผวน หรือมีการเปลี่ยนแปลงเฉียบพลัน เช่น ในช่วงเดือนเมษายน ถึงกรกฎาคม ปี พ.ศ. 2020 พบว่าแบบจำลองสามารถปรับตัวตามพฤติกรรมของข้อมูลได้ดี



ภาพประกอบ 30 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง ARIMAX

จากภาพประกอบที่ 30 ค่าที่พยากรณ์จาก ARIMAX มีลักษณะการเคลื่อนไหวที่ใกล้เคียงกับข้อมูลจริง ในหลายช่วงเวลา ซึ่งแสดงให้เห็นถึงศักยภาพของแบบจำลองในการเรียนรู้ความสัมพันธ์ของแนวโน้มจากอดีต และตอบสนองต่อความแปรปรวนที่เกิดขึ้นแบบรายวันได้อย่างเหมาะ ผลลัพธ์ที่ได้จากการพยากรณ์ พบว่าข้อได้เปรียบของแบบจำลอง ARIMAX ในการนำตัวแปรเสริมอย่าง Lead Time, High Season, Total Night และ Total Guest มาใช้ร่วมในการพยากรณ์ ช่วยให้แบบจำลองสามารถอธิบายความแปรปรวนของข้อมูลเป้าหมายได้แม่นยำยิ่งขึ้น เมื่อเทียบกับแบบจำลอง ARIMA ที่อาศัยเฉพาะข้อมูลในอดีต

ผลลัพธ์จากการสร้างแบบจำลอง SARIMAX

SARIMAX (Seasonal ARIMA with Exogenous Variables) เป็นแบบจำลองอนุกรมเวลาที่พัฒนาต่อยอดจาก SARIMA โดยเพิ่มความสามารถในการนำตัวแปรภายนอก (Exogenous Variables) เข้ามาใช้ร่วมในการพยากรณ์ ซึ่งช่วยให้แบบจำลองสามารถวิเคราะห์ข้อมูลที่มีทั้งฤดูกาล (Seasonality) และผลกระทบจากปัจจัยแวดล้อมได้ในเวลาเดียวกัน เหมาะอย่างยิ่งสำหรับข้อมูลในบริบทของธุรกิจโรงแรม ที่ได้รับอิทธิพลจากทั้งฤดูกาลการท่องเที่ยว และพฤติกรรมความต้องการของลูกค้า

ในการวิจัยนี้ ได้กำหนดค่าพารามิเตอร์ของแบบจำลอง SARIMAX โดยอาศัยผลจากกราฟ ACF, PACF และการทดสอบความเป็น Stationary ของข้อมูล เช่นเดียวกับแบบจำลอง SARIMA ควบคู่กับการคัดเลือกตัวแปรเสริมที่มีความสัมพันธ์กับ Room Sold อย่างมีนัยสำคัญ ได้แก่ Lead Time, High Season, Total Night และ Total Guest พารามิเตอร์ที่เหมาะสมที่ได้จากการปรับแต่งคือ

p = 2, d = 1, q = 2 (ส่วนของ ARIMA)

P = 1, D = 1, Q = 1, S = 7 (องค์ประกอบฤดูกาล โดย S รอบฤดูกาลของข้อมูลรายสัปดาห์)

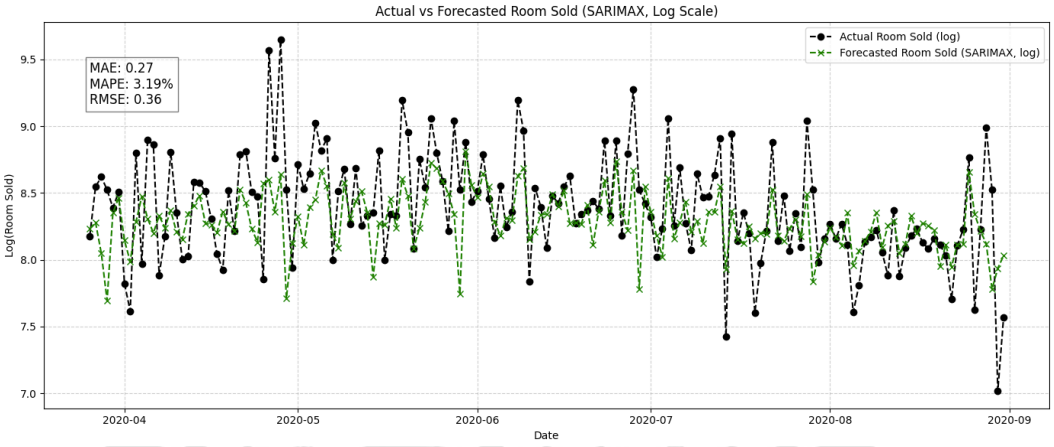
หลังจากฝึกแบบจำลองด้วยชุดข้อมูลฝึกสอน (Training Data) และทดสอบกับชุดข้อมูลทดสอบ (Testing Data) พบว่าแบบจำลอง SARIMAX ให้ผลลัพธ์ที่น่าพึงพอใจ โดยสามารถอธิบายพฤติกรรมข้อมูลได้ทั้งในมิติของฤดูกาล และผลกระทบจากปัจจัยภายนอก โดยมีค่าประเมินประสิทธิภาพของแบบจำลองดังนี้

Mean Absolute Error (MAE) เท่ากับ 0.27 แสดงถึงค่าความคลาดเคลื่อนเฉลี่ยรายวันอยู่ในระดับต่ำ

Mean Absolute Percentage Error (MAPE) เท่ากับ 3.19% ซึ่งถือว่าอยู่ในเกณฑ์ความแม่นยำสูง

Root Mean Square Error (RMSE) เท่ากับ 0.36 แสดงถึงความคลาดเคลื่อนโดยรวมที่อยู่ในระดับต่ำเช่นกัน

จากภาพประกอบที่ 31 ซึ่งเปรียบเทียบค่าพยากรณ์ของ SARIMAX กับข้อมูลจริงในช่วงทดสอบ (Testing Data) จะเห็นว่าแบบจำลองสามารถติดตามแนวโน้มของจำนวนห้องพักที่ขายได้ได้อย่างใกล้เคียงกับค่าจริง เช่น ในช่วงเดือนพฤษภาคมถึงกรกฎาคม แบบจำลองสามารถตอบสนองต่อแนวโน้มที่เปลี่ยนแปลงได้ดีกว่าแบบจำลองที่ไม่รวมตัวแปรเสริม



ภาพประกอบ 31 กราฟแสดงการเปรียบเทียบค่าที่พยากรณ์ กับค่าจริงของแบบจำลอง SARIMAX

จากภาพประกอบที่ 31 ซึ่งเปรียบเทียบค่าพยากรณ์ของ SARIMAX กับข้อมูลจริงในช่วงทดสอบ จะเห็นว่าแบบจำลองสามารถติดตามแนวโน้มของจำนวนห้องพักที่ขายได้ ได้อย่างใกล้เคียงกับค่าจริง โดยเฉพาะในช่วงที่ข้อมูลมีความผันผวนสูง หรือเปลี่ยนแปลงอย่างเฉียบพลัน เช่นในช่วงเดือนพฤษภาคมถึงกรกฎาคม แบบจำลองสามารถตอบสนองต่อแนวโน้มที่เปลี่ยนแปลงได้ดี แสดงให้เห็นถึงความสามารถในการพยากรณ์ที่เหนือกว่า SARIMA โดยเฉพาะในบริบทของข้อมูลที่มีทั้งฤดูกาล และผลกระทบจากปัจจัยแวดล้อม การผสมผสานทั้งองค์ประกอบฤดูกาล และตัวแปรเสริมเข้าด้วยกัน ทำให้แบบจำลองสามารถอธิบายพฤติกรรมที่ซับซ้อนของข้อมูลโรงแรมได้อย่างมีประสิทธิภาพยิ่งขึ้น

ผลลัพธ์จากการเปรียบเทียบแบบจำลอง

จากการศึกษาเปรียบเทียบแบบจำลองอนุกรมเวลา ได้แก่ Prophet, ARIMA, SARIMA, ARIMAX และ SARIMAX โดยมีเป้าหมายเพื่อประเมินความแม่นยำในการพยากรณ์จำนวนห้องพักที่ขายได้ (Room Sold) เพื่อนำไปคาดการณ์รายได้ของโรงแรมผ่านการวัดค่าความคลาดเคลื่อนจากการพยากรณ์ด้วยตัวชี้วัดทางสถิติ ได้แก่ MAE, MAPE และ RMSE จากการเปรียบเทียบผลลัพธ์พบว่าแบบจำลอง ARIMA ซึ่งไม่รองรับฤดูกาล หรือปัจจัยภายนอก ให้ผลลัพธ์ในระดับปานกลาง โดยมีค่า

MAE เท่ากับ 0.3417, MAPE เท่ากับ 4.00% และ RMSE เท่ากับ 0.444 แม้จะสามารถติดตามแนวโน้มหลักของข้อมูลได้บ้าง แต่ยังขาดความสามารถในการรับมือกับความผันผวน และฤดูกาลที่มีผลต่อพฤติกรรมผู้บริโภคในอุตสาหกรรมโรงแรมอย่างชัดเจน ถัดมาคือ แบบจำลอง SARIMA ซึ่งพัฒนาต่อยอดจาก ARIMA ทำให้สามารถจำลองรูปแบบซ้ำตามฤดูกาลได้ดี โดยมีค่า MAE เท่ากับ 0.33, MAPE เท่ากับ 3.92% และ RMSE เท่ากับ 0.43 อย่างไรก็ตาม แบบจำลองนี้ยังไม่สามารถตอบสนองต่อการเปลี่ยนแปลงเฉียบพลันที่ไม่ได้เกิดซ้ำในรอบฤดูกาล เช่นเหตุการณ์พิเศษที่ไม่เกิดประจำปี ส่วนแบบจำลอง Prophet ที่ไม่ได้ใช้ตัวแปรเสริม ให้ผลการพยากรณ์ที่ด้อยลงอย่างชัดเจน โดยมีค่า MAE เท่ากับ 0.4804, MAPE เท่ากับ 5.64% และ RMSE เท่ากับ 0.5980 แสดงให้เห็นว่า แม้แบบจำลองจะถูกออกแบบมาเพื่อรองรับแนวโน้ม และฤดูกาลได้ในระดับหนึ่ง แต่เมื่อไม่มีข้อมูลเสริม แบบจำลองกลับไม่สามารถตอบสนองต่อบริบทพฤติกรรมลูกค้าที่เปลี่ยนแปลงได้อย่างแม่นยำ อย่างไรก็ตาม เมื่อมีการใช้ตัวแปรเสริมร่วมกับแบบจำลอง Prophet คือ Lead Time, High Season, Total Night และ Total Guest ผลลัพธ์กลับเปลี่ยนไปอย่างมีนัยสำคัญ โดยให้ค่า MAE ลดลงเหลือ 0.2279, MAPE ต่ำที่สุดเพียง 2.71% และ RMSE เหลือเพียง 0.3078 ซึ่งเป็นค่าต่ำที่สุดในทุกเกณฑ์การวัด แสดงให้เห็นถึงบทบาทสำคัญของตัวแปรเสริม ในการช่วยให้แบบจำลอง Prophet เรียนรู้แนวโน้มที่ซับซ้อนได้อย่างแม่นยำ และยืดหยุ่นมากยิ่งขึ้น นอกจากนี้พบว่าแบบจำลอง ARIMAX ซึ่งมีการเพิ่มตัวแปรเสริมเข้าไปในโครงสร้างของ ARIMA โดยยังไม่มีองค์ประกอบฤดูกาล ค่า MAE ที่ได้คือ 0.2765, MAPE เท่ากับ 3.28% และ RMSE เท่ากับ 0.363 ถือว่าอยู่ในระดับที่ดี สะท้อนว่าตัวแปรเสริม ช่วยให้แบบจำลองสามารถอธิบายความแปรปรวนที่เกิดจากพฤติกรรมผู้บริโภคได้ดีกว่า ARIMA แบบดั้งเดิมอย่างชัดเจน และสุดท้ายคือแบบจำลอง SARIMAX ซึ่งผสานความสามารถในการรองรับฤดูกาลของ SARIMA เข้ากับตัวแปรเสริม ทำให้สามารถจัดการกับข้อมูลที่ซับซ้อนทั้งเชิงฤดูกาล และพฤติกรรมลูกค้าได้อย่างครบถ้วน โดยมีค่า MAE เท่ากับ 0.2701, MAPE เท่ากับ 3.19% และ RMSE เท่ากับ 0.3579 ซึ่งอยู่ในระดับที่ดี ทั้งในแง่ความแม่นยำ และเสถียรภาพ โครงสร้างของแบบจำลองยังเปิดโอกาสให้ผู้ใช้งานควบคุมพารามิเตอร์ต่าง ๆ ได้อย่างละเอียด เหมาะสำหรับการใช้งานเชิงวิเคราะห์ที่ต้องการความแม่นยำสูง

จากผลการเปรียบเทียบทั้งหมดสามารถสรุปได้ว่า แบบจำลอง Prophet ที่ใช้ตัวแปรเสริมเหมาะสมที่สุดสำหรับบริบทของการพยากรณ์รายได้โรงแรม เนื่องจากสามารถนำตัวแปรที่เกี่ยวข้องกับพฤติกรรมของลูกค้า และฤดูกาลมาใช้ในการเรียนรู้ ทำให้แบบจำลองสามารถปรับตัวตามข้อมูลจริงได้อย่างยืดหยุ่น และแม่นยำ ในขณะที่แบบจำลองอย่าง SARIMAX ก็มีจุดแข็งในด้านโครงสร้างการควบคุมที่ละเอียด เหมาะสำหรับการวิเคราะห์เชิงลึก และกรณีที่ต้องการความสามารถในการตีความเชิงสถิติอย่างเข้มงวด

บทที่ 5
สรุปผล อภิปรายผล และข้อเสนอแนะ

สรุปผลการวิจัย

ในบริบทของอุตสาหกรรมการโรงแรมซึ่งมีการแข่งขันสูง และได้รับผลกระทบจากปัจจัยแวดล้อมที่เปลี่ยนแปลงอยู่เสมอ การพยากรณ์รายได้จึงเป็นเครื่องมือสำคัญในการวางแผนกลยุทธ์ทางธุรกิจ โดยเฉพาะในด้านการบริหารรายได้ (Revenue Management) งานวิจัยนี้มุ่งเน้นการวิเคราะห์และเปรียบเทียบประสิทธิภาพของแบบจำลองอนุกรม 5 แบบ ได้แก่ Prophet, ARIMA, ARIMAX, SARIMA และ SARIMAX เพื่อทำนายจำนวนห้องพักที่ขายได้ (Room Sold) ซึ่งสัมพันธ์โดยตรงกับรายได้ของโรงแรม และสามารถนำไปประเมินรายได้โดยคูณกับอัตราค่าห้องพักเฉลี่ยต่อวัน (ADR)

ผลการเปรียบเทียบประสิทธิภาพของแบบจำลองทั้งห้าเทคนิคสามารถวัดได้จากค่าความคลาดเคลื่อน ได้แก่ Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) และ Root Mean Square Error (RMSE) โดยสามารถสรุปผลการทดลองได้ดังตารางที่ 8

ตาราง 8 ตารางแสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง

Model	MAE	MAPE	RMSE
Prophet	0.4804	5.64	0.5980
Prophet with Exogenous variables	0.2279	2.71%	0.3078
ARIMA (1,1,1)	0.3417	4.00%	0.444
ARIMAX (2,1,2)	0.2765	3.28%	0.363
SARIMA (2,1,1)(1,0,2,7)	0.33	3.92%	0.43
SARIMAX (3,1,2)(1,0,2,7)	0.2701	3.19%	0.3579

จากตารางที่ 8 แสดงผลการเปรียบเทียบค่าความคลาดเคลื่อนของแบบจำลอง พบว่าแบบจำลอง Prophet ที่ใช้ตัวแปรเสริม ให้ผลลัพธ์ดีที่สุดในทุกตัวชี้วัด ทั้ง MAE, MAPE และ RMSE ซึ่งสะท้อนถึงศักยภาพของแบบจำลองในการตอบสนองต่อแนวโน้ม และพฤติกรรมของลูกค้าได้อย่างแม่นยำ และยืดหยุ่น ขณะที่ SARIMAX แม้จะมีค่าความคลาดเคลื่อนสูงกว่าเล็กน้อย แต่โดดเด่นในด้านเสถียรภาพ และความสามารถในการควบคุมพารามิเตอร์เชิงฤดูกาลอย่างละเอียด เหมาะสำหรับการใช้งานที่ต้องการวิเคราะห์เชิงลึกในบริบทที่ข้อมูลมีทั้งฤดูกาล และพฤติกรรมแปรผัน ในส่วนของแบบจำลองอย่าง ARIMAX, SARIMA, Prophet without Exogenous variables และ ARIMA ให้ผลลัพธ์

ในระดับดีถึงปานกลางตามลำดับ โดยมีข้อดีเฉพาะด้านที่เหมาะสมกับลักษณะข้อมูลแตกต่างกัน เช่น ARIMA เหมาะกับข้อมูลที่ไม่มีฤดูกาล ในขณะที่ SARIMA เหมาะกับข้อมูลที่มีรูปแบบซ้ำตามฤดูกาลแต่ไม่ซับซ้อน ผลลัพธ์จากการศึกษานี้ พบว่าการเลือกใช้แบบจำลองควรคำนึงถึงทั้งลักษณะข้อมูล ความซับซ้อนของปัจจัยแวดล้อม และวัตถุประสงค์ของการพยากรณ์ เช่น หากต้องการความแม่นยำเชิงพฤติกรรมของลูกค้า Prophet with Exogenous variables จะเป็นทางเลือกที่เหมาะสมที่สุด แต่หากต้องการควบคุมโครงสร้าง และวิเคราะห์เชิงเทคนิคอย่างละเอียด SARIMAX ก็ถือเป็นอีกทางเลือกที่มีประสิทธิภาพสูงเช่นกัน

อภิปรายผลการวิจัย

จากการดำเนินการวิจัยโดยใช้ข้อมูล Hotel Booking Demand เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองอนุกรมเวลาสำหรับการพยากรณ์จำนวนห้องพักที่ขายได้ ซึ่งเป็นตัวแปรสำคัญที่สะท้อนรายได้ของโรงแรม พบว่าแบบจำลองแต่ละประเภทมีข้อดี และข้อจำกัดที่แตกต่างกันตามโครงสร้างของแบบจำลอง และความสามารถในการรองรับลักษณะเฉพาะของข้อมูล ไม่ว่าจะเป็นแนวโน้มระยะยาว ฤดูกาล หรือปัจจัยภายนอกที่มีผลต่อพฤติกรรมของลูกค้า

ผลการศึกษาแสดงให้เห็นว่าแบบจำลอง Prophet ที่มีการเพิ่มตัวแปรช่วย ให้ผลลัพธ์ที่แม่นยำที่สุด โดยมีค่า MAPE ต่ำที่สุดที่ 2.71% และค่าความคลาดเคลื่อน MAE และ RMSE อยู่ในระดับต่ำที่สุดเมื่อเทียบกับแบบจำลองอื่น การเพิ่มตัวแปรภายนอกที่สะท้อนพฤติกรรมการจอง เช่น Lead Time, High Season, Total Night และ Total Guest มีบทบาทสำคัญในการช่วยให้แบบจำลอง Prophet สามารถเรียนรู้แนวโน้ม และตอบสนองต่อสถานการณ์ที่ซับซ้อนในธุรกิจโรงแรมได้อย่างมีประสิทธิภาพ แบบจำลองนี้ยังมีความยืดหยุ่นสูง ไม่ต้องกำหนดพารามิเตอร์เชิงเทคนิคที่ซับซ้อน และสามารถวิเคราะห์ข้อมูลที่มีโครงสร้างไม่เป็นเชิงเส้นได้ดี ในทางตรงกันข้ามเมื่อไม่มีการใช้ตัวแปรเสริมแบบจำลอง Prophet without Exogenous variables กลับให้ผลลัพธ์ที่ด้อยลงอย่างชัดเจน โดยมีค่า MAPE เพิ่มขึ้นเป็น 5.64% แสดงให้เห็นถึงข้อจำกัดของแบบจำลองเมื่อขาดข้อมูลบริบทที่ช่วยอธิบายพฤติกรรมของลูกค้าอย่างลึกซึ้ง ผลลัพธ์นี้ชี้ว่า Prophet มีความสามารถในการรองรับความซับซ้อนได้ดีขึ้นมากเมื่อมี Exogenous variables สนับสนุน สำหรับแบบจำลอง ARIMA ซึ่งเป็นแบบจำลองเชิงสถิติพื้นฐานที่ไม่มีโครงสร้างรองรับฤดูกาล หรือปัจจัยแวดล้อม พบว่ามีค่า MAPE อยู่ที่ 4.00% แม้จะสามารถติดตามแนวโน้มหลักของข้อมูลได้บ้าง แต่กลับมีข้อจำกัดในการตอบสนองต่อความผันผวนเฉียบพลัน หรือพฤติกรรมผู้บริโภคที่แปรเปลี่ยนตามฤดูกาล ซึ่งพบได้บ่อยในธุรกิจโรงแรม ในขณะที่ SARIMA ซึ่งเป็นการขยายโครงสร้างจาก ARIMA โดยเพิ่มองค์ประกอบฤดูกาลเข้าไป สามารถติดตาม

รูปแบบที่เกิดซ้ำเป็นรอบได้ดีกว่า โดยมีค่า MAPE ลดลงมาอยู่ที่ 3.92% อย่างไรก็ตาม แบบจำลองนี้ยังคงมีข้อจำกัดในด้านการตอบสนองต่อสถานการณ์ที่ไม่ได้เป็นไปตามรูปแบบฤดูกาล เช่น ช่วงโปรโมชั่น หรือเหตุการณ์เฉพาะหน้า ที่มีผลต่อพฤติกรรมการจองของลูกค้า ในส่วนของการเปรียบเทียบแบบจำลองที่มีการใช้ตัวแปรเสริมเพิ่มเติม ได้แก่ ARIMAX และ SARIMAX พบว่าแบบจำลองทั้งสองสามารถเพิ่มประสิทธิภาพของแบบจำลองได้อย่างชัดเจน โดย ARIMAX มีค่า MAPE อยู่ที่ 3.28% ซึ่งดีกว่า ARIMA อย่างมีนัยสำคัญ ขณะที่ SARIMAX ซึ่งรวมทั้งโครงสร้างฤดูกาล และตัวแปรเสริมเข้าด้วยกัน ให้ค่า MAPE เท่ากับ 3.19% ซึ่งแม้จะยังสูงกว่า Prophet with Exogenous variables เล็กน้อย แต่แสดงให้เห็นถึงความเสถียรในการวิเคราะห์ข้อมูลที่มีความซับซ้อนเชิงฤดูกาลร่วมกับพฤติกรรมลูกค้า โดยเฉพาะในบริบทที่ผู้ใช้งานต้องการควบคุมพารามิเตอร์แบบจำลองในเชิงเทคนิคอย่างละเอียด

ผลการวิจัยนี้ชี้ให้เห็นว่าแบบจำลอง Prophet เมื่อใช้ร่วมกับตัวแปรเสริมที่สะท้อนพฤติกรรมของผู้บริโภค มีประสิทธิภาพสูงสุดในการพยากรณ์รายได้ของโรงแรม ในขณะที่ SARIMAX ให้ผลลัพธ์ที่แม่นยำ และมีเสถียรภาพสูง เหมาะสำหรับการใช้งานในบริบทที่ข้อมูลมีลักษณะเชิงฤดูกาลร่วมกับปัจจัยแวดล้อม ส่วนแบบจำลองอื่น ๆ มีประสิทธิภาพในระดับปานกลาง เหมาะกับสถานการณ์ที่มีข้อมูลเชิงเดียว หรือไม่ซับซ้อนมากนัก ทั้งนี้การเลือกใช้แบบจำลองจึงควรพิจารณาจากลักษณะของข้อมูล และจุดประสงค์ของการใช้งานเป็นสำคัญ เพื่อให้ผลลัพธ์ที่ได้สามารถนำไปใช้วางแผนกลยุทธ์ทางธุรกิจได้อย่างมีประสิทธิภาพ

ข้อเสนอแนะ

จากผลการวิจัยพบว่าแบบจำลองแต่ละประเภทมีจุดเด่น และข้อจำกัดที่แตกต่างกัน ซึ่งควรพิจารณาให้เหมาะสมกับลักษณะข้อมูล และวัตถุประสงค์ของการใช้งานในทางปฏิบัติ หากข้อมูลมีลักษณะเป็นแนวโน้มเชิงเวลาโดยไม่มีฤดูกาลที่ชัดเจน แบบจำลอง ARIMA สามารถนำมาใช้งานได้สะดวกเนื่องจากโครงสร้างเรียบง่าย และไม่ต้องการข้อมูลเสริมมากนัก อย่างไรก็ตาม แบบจำลองนี้มีข้อจำกัดในการรับมือกับพฤติกรรมที่เปลี่ยนแปลงเฉียบพลัน หรือซับซ้อน ในกรณีที่ข้อมูลแสดงรูปแบบตามฤดูกาลอย่างชัดเจน แบบจำลอง SARIMA จะสามารถจับรูปแบบซ้ำเหล่านี้ได้ดี โดยไม่ต้องใช้ปัจจัยภายนอกเพิ่มเติม แต่หากต้องการวิเคราะห์เชิงลึกที่รวมทั้งฤดูกาล และพฤติกรรมผู้บริโภค แนะนำให้ใช้แบบจำลอง SARIMAX ซึ่งสามารถควบคุมองค์ประกอบตามฤดูกาล และนำตัวแปรเสริม เช่น Lead Time, High Season หรือ Total Guest เข้ามาช่วยอธิบายข้อมูลได้อย่างมีประสิทธิภาพ สำหรับองค์กรที่เน้นความยืดหยุ่นในการเพิ่มตัวแปรอธิบาย และไม่ต้องการการตั้งค่าทางเทคนิคมากนัก แบบจำลอง Prophet เป็นอีกทางเลือกที่น่าสนใจ โดยเฉพาะเมื่อใช้ร่วมกับตัวแปรช่วยในการพยากรณ์ ซึ่งช่วยให้

แบบจำลองสามารถเข้าใจแนวโน้มที่ไม่เป็นเชิงเส้น และพฤติกรรมผู้บริโภคที่แปรผันตามเวลาได้ดียิ่งขึ้น ทั้งนี้ การไม่ใช้ตัวแปรเสริม จะลดความสามารถของ Prophet ลงอย่างชัดเจน จึงควรให้ความสำคัญกับการคัดเลือกตัวแปรเสริมที่เหมาะสม ในกรณีที่ข้อมูลมีปริมาณไม่มาก เช่น โรงแรมเปิดใหม่ หรือยังไม่มีข้อมูลย้อนหลังเพียงพอ การใช้แบบจำลองอนุกรมเวลาอาจไม่เหมาะสมในช่วงเริ่มต้น แนะนำให้ใช้เทคนิคเชิงสถิติอย่าง Pickup Forecasting ซึ่งอาศัยข้อมูลการจองสะสมในช่วงใกล้วันเข้าพักเพื่อคาดการณ์ระยะสั้น โดยไม่จำเป็นต้องใช้ข้อมูลในอดีตยาวนาน วิธีนี้สามารถให้แนวโน้มเบื้องต้นที่ช่วยในการวางแผนทรัพยากรและการบริหารห้องพักได้อย่างมีประสิทธิภาพ

ส่วนแนวทางการนำแบบจำลองไปใช้ในทางปฏิบัติ กรณีที่ได้ผลลัพธ์จากแบบจำลองพยากรณ์ที่ผ่านการประมวลผลในรูปแบบ log scale ควรทำการแปลงค่าที่พยากรณ์ กลับคืนเป็นค่าจริง (inverse transform) เพื่อให้ได้จำนวนห้องพักที่คาดว่าจะขายได้ในแต่ละวัน จากนั้นจึงนำไปคูณกับค่าเฉลี่ยรายวันของ ADR (Average Daily Rate) ที่แท้จริงในช่วงเวลานั้น ๆ เพื่อประมาณรายได้ประจำวัน (Daily Revenue) อย่างถูกต้อง และสะท้อนสถานการณ์จริงของโรงแรม นอกจากนี้เมื่อได้รายได้รายวันแล้ว สามารถนำไปคำนวณเป็นยอดรายได้สะสม (Cumulative Revenue) เพื่อเปรียบเทียบกับเป้าหมายที่ตั้งไว้ เช่น งบประมาณประจำเดือน หรือเป้าหมายเชิงกลยุทธ์ วิธีนี้จะช่วยให้ผู้บริหารสามารถติดตามการดำเนินงานรายได้ในแต่ละช่วงเวลา และปรับกลยุทธ์ได้อย่างทันท่วงที

บรรณานุกรม

1. Number of international tourist arrivals worldwide from 2005 to 2023,by region [Internet]. 2024. Available from: https://www.statista.com/statistics/186743/international-tourist-arrivals-worldwide-by-region-since-2010/.

2. Contessi D, Viverit L, Pereira LN, Heo CY. Decoding the future: Proposing an interpretable machine learning model for hotel occupancy forecasting using principal component analysis. International Journal of Hospitality Management. 2024;121:103802.

3. Wu DC, Song H, Shen S. New developments in tourism and hotel demand modeling and forecasting. International Journal of Contemporary Hospitality Management. 2017;29(1):507-29.

4. รายงานฉบับสมบูรณ์ สำรวจการประกอบกิจการโรงแรมและเกสต์เฮาส์ พ.ศ. 2559 [Internet]. 2559 [cited 5 มิถุนายน 2567]. Available from: https://data.go.th/th/dataset/0101_17_0023.

5. Antonio N, de Almeida A, Nunes L. Hotel booking demand datasets. Data in Brief. 2019;22:41-9.

6. พระราชบัญญัติโรงแรม, (2547).

7. กระทรวงการท่องเที่ยวและกีฬา. มาตรฐานที่พักเพื่อการท่องเที่ยว(โรงแรม/รีสิร์ท/เซอร์วิสอพาร์ทเมนท์/เกสต์เฮาส์)2552.

8. Ivanov S. Hotel Revenue Management: From Theory to Practice2014.

9. Hesford JW, Turner MJ, Mangin N, Thomas CR, Hoffmann K. Competitor Monitoring and Revenue Performance: Evidence from the Hospitality Industry. In: Burney LL, Malina MA, editors. Advances in Management Accounting. Advances in Management Accounting. 31: Emerald Publishing Limited; 2019. p. 1-22.

10. Elphick D. Hotel ADR: คู่มือของคุณเกี่ยวกับอัตรารายวันเฉลี่ย (Average Daily Rate) 2024 [updated 2/06/2024. Available from: https://www.siteminder.com/th/r/hotel-adr/.

11. Ivanov S, Zhechev V. Hotel Revenue Management – A Critical Literature Review. Tourism. 2011;60.

12. Klein R, Koch S, Steinhardt C, Strauss AK. A review of revenue management:

Recent generalizations and advances in industry applications. European Journal of Operational Research. 2020;284(2):397-412.

13. Denizci Guillet B, Mohammed I. Revenue management research in hospitality and tourism. International Journal of Contemporary Hospitality Management. 2015;27(4):526-60.

14. Stellwagen E, Tashman L. ARIMA: The Models of Box and Jenkins. Foresight: Int J Appl Forecast. 2013:28-33.

15. Box GE, and Jenkins, G. M. Time series analysis. Forecasting and control. 5 ed. United States of America1976.

16. BanaŚ J, Utnik-BanaŚ K. Evaluating a seasonal autoregressive moving average model with an exogenous variable for short-term timber price forecasting. Forest Policy and Economics. 2021;131:102564.

17. Chatfield C. The Analysis of Time Series

An Introduction. Edition S, editor. New York2003 29 July 2003.

18. Fatoumata&Christine. Time Series Analysis and Modeling to Forecast: a Survey. 2021.

19. Taylor&Benjamin. Forecasting at Scale. 2018.

20. Schwartz Z, Ma J, Webb T. The MSapeMER: a symmetric, scale-free and intuitive forecasting error measure for hospitality revenue management. International Journal of Contemporary Hospitality Management. 2024;36(6):2035-48.

21. Song H, Li G. Tourism demand modelling and forecasting - A review of Recent research. Tourism Management. 2008;29:203-20.

22. Weatherford LR, Kimes SE. A comparison of forecasting methods for hotel revenue management. International Journal of Forecasting. 2003;19(3):401-15.

23. Enache MC. Machine Learning in Tourism Revenue Management. 2019.

24. Vagropoulos SI, Chouliaras GI, Kardakos EG, Simoglou CK, Bakirtzis AG, editors. Comparison of SARIMAX, SARIMA, modified SARIMA and ANN-based models for short-term PV generation forecasting. 2016 IEEE International Energy Conference (ENERGYCON); 2016 4-8 April 2016.

25. Suwatanapongched NPP. Forecasting hotel daily room demand with transformed data using time series methods. 2021.

26. Sison N, Li L, Han M, editors. Survey of Machine Learning and Deep Learning Techniques for Travel Demand Forecasting. 2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI); 2021 18-21 Oct. 2021.

27. Sirisha U, Belavagi M, Attigeri G. Profit Prediction Using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison. IEEE Access. 2022;PP:1-.

28. Huang L, Zheng W. Novel deep learning approach for forecasting daily hotel demand with agglomeration effect. International Journal of Hospitality Management. 2021;98:103038.

29. Kaya K, Yılmaz Y, Yaslan Y, Öğüdücü ŞG, Çıngı F. Demand forecasting model using hotel clustering findings for hospitality industry. Information Processing & Management. 2022;59(1).

30. Lee M. Modeling and forecasting hotel room demand based on advance booking information. Tourism Management. 2018;66:62-71.

31. Alharbi FR, Csala D. A Seasonal Autoregressive Integrated Moving Average with Exogenous Factors (SARIMAX) Forecasting Model-Based Time Series Approach. Inventions. 2022;7(4):94.

32. Wang Y, Yan Z, Wang D, Yang M, Li Z, Gong X, et al. Prediction and analysis of COVID-19 daily new cases and cumulative cases: times series forecasting and machine learning models. BMC Infectious Diseases. 2022;22.

33. Munim ZH, Fiskin CS, Nepal B, Chowdhury MMH. Forecasting container throughput of major Asian ports using the Prophet and hybrid time series models. The Asian Journal of Shipping and Logistics. 2023;39(2):67-77.

34. Negre P, Alonso RS, Prieto J, García Ó, de-la-Fuente-Valentín L. Prediction of footwear demand using Prophet and SARIMA. Expert Systems with Applications. 2024;255:124512.

35. Mazaraki A, Kulyk M, Zubko T. REVENUE FORECASTING SCENARIOS FOR INTERNATIONAL HOTEL CHAINS. Baltic Journal of Economic Studies. 2024;10:207-14.

36. Mane V, Patil R, Patwardhan S, Pethkar P, editors. Revenue Prediction using Sequential Machine Learning. 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC); 2024 3-5 Oct. 2024.

37. Andreevich SBKPA. Comparative Analysis of ARIMA, SARIMA and Prophet Model in Forecasting. http://www.sciencepg.com/journal/rd: Science Publishing Group; 2024 18 October 2024. 110-20 p.

38. Lee G-C. A Data-Driven Approach to Tourism Demand Forecasting: Integrating Web Search Data into a SARIMAX Model. Data [Internet]. 2025; 10(5).

ประวัติผู้เขียน