**รายงาน**

Analysis data of Hotel & Resort

**จัดทำโดย**

น.ส. วิวาวรรณ ลือศักดิ์กชกร 6110406210

น.ส. วีรินทร์ เนียมแนบ 6110406228

นาย กิตติธัช จึงโสภณวิทวัส 6110450022

นาย สหรัฐ ชุตาภา 6110450995

นาย วรุณ สรรพกิจ 6110451398

**เสนอ**

อาจารย์เสฏฐวิทย์ เกิดผล

**รายงานฉบับนี้เป็นส่วนหนึ่งของรายวิชา**

Introduction to Data Science

**สารบัญ**

[**คำอธิบายข้อมูลตามคอลัมน์** 2](#_Toc37350498)

[**สาเหตุที่มาของการเลือกข้อมูลชุดนี้** 3](#_Toc37350499)

[**สิ่งที่น่าสนใจจากการวิเคราะห์ข้อมูลเบื้องต้น** 4](#_Toc37350500)

[**คำถามองค์ความรู้เชิงลึก** 4](#_Toc37350501)

[**สาเหตุที่น่าสนใจของคำถามองค์ความรู้เชิงลึก** 4](#_Toc37350502)

[**กระบวนการได้มาซึ่งองค์ความรู้** 4](#_Toc37350503)

[**ผลลัพธ์องค์ความรู้ที่ได้** 5](#_Toc37350504)

**แหล่งที่มาของข้อมูล**

ข้อมูลเปรียบเทียบการเข้าพักและยกเลิกการจอง ระหว่าง โรงแรมในเมือง และ โรงแรมรีสอร์ท ในช่วง กรกฎาคม 2015 – สิงหาคม 2017 อ้างอิงจาก บทความ  [**Hotel Booking Demand Datasets**](https://www.sciencedirect.com/science/article/pii/S2352340918315191) ในหนังสือ Data in Brief เล่มที่ 22, กุมภาพันธุ์ 2019 เขียนโดย Nuno Antonio, Ana Almeida, Luis Nunes

<https://www.kaggle.com/jessemostipak/hotel-booking-demand>

# **คำอธิบายข้อมูลตามคอลัมน์**

1. **Hotel** คือ ประเภทของโรงแรมแบ่งเป็น Resort Hotel(โรงแรมรีสอร์ท) และ City Hotel(โรงแรมในเมือง)
2. **is\_canceled** คือ การแสดงข้อมูลว่าการจองถูกยกเลิกหรือไม่โดยที่แบ่งเป็น 1(ถูกยกเลิก) และ

0(ไม่ถูกยกเลิก)

1. **lead\_time** คือ วันที่ผ่านไประหว่างวันที่มีการจองเข้าสู่ระบบ PMS และ วันที่มาถึง
2. **arrival\_date\_year** คือ ปีที่มาเข้าพักที่โรงแรม
3. **arrival\_date\_month** คือ เดือนที่มาเข้าพักที่โรงแรม
4. **arrival\_date\_week\_number** คือ จำนวนสัปดาห์ในปีที่มาเข้าพักที่โรงแรม
5. **arrival\_date\_day\_of\_month** คือ วันที่ที่มาเข้าพักที่โรงแรม
6. **stays\_in\_weekend\_nights** คือ จำนวนวันที่มีการจอง หรือ เข้าพัก ค้างคืนในโรงแรมช่วงวันสุดสัปดาห์(วันเสาร์ และ วันอาทิตย์)
7. **stays\_in\_week\_nights** คือ จำนวนวันที่มีการจอง หรือ เข้าพัก ค้างคืนในโรงแรมช่วงวันธรรมดา(วันจันทร์ถึงวันศุกร์)
8. **adults** คือ จำนวนผู้ใหญ่ที่มาเข้าพัก
9. **children** คือ จำนวนเด็กที่มาเข้าพัก
10. **babies** คือ จำนวนทารกที่มาเข้าพัก
11. **meal** คือ ประเภทอาหารที่มีการจองไว้ แบ่งเป็น SC(ไม่มีอาหาร), BB(อาหารเช้า), HB(อาหารเช้าและอีกมื้อ), FB(อาหารเช้า, อาหารกลางวัน, อาหารเย็น)
12. country คือ ประเทศที่มา
13. **market\_segment** คือ การแบ่งส่วนตลาด โดยที่ TA หมายถึง Travel Agents และ TO หมายถึง Tour Operators
14. **distribution\_channel** คือ ช่องทางการซื้อ โดยที่ TA หมายถึง Travel Agents และ TO หมายถึง Tour Operators
15. **is\_repeated\_guest** คือ การแสดงข้อมูลว่าลูกค้าที่มาพักเคยมาพักหรือไม่โดยที่แบ่งเป็น 1 (เคยมาพัก) และ 0(ไม่เคยมาพัก)
16. **previous\_cancellations** คือ จำนวนครั้งที่ถูกยกเลิกการจองก่อนหน้าการจองรอบปัจจุบัน
17. **previous\_bookings\_not\_canceled** คือ จำนวนครั้งที่ไม่ถูกยกเลิกการจองก่อนหน้าการจองรอบปัจจุบัน
18. **reserved\_room\_type** คือ ประเภทห้องที่มีการจองไว้
19. **assigned\_room\_type** คือ ประเภทห้องที่เข้าพัก
20. **booking\_changes** คือ จำนวนรอบที่มีการแก้ไขการจองตั้งแต่วันที่มีการจองเข้าสู่ระบบ PMS จนถึงวันที่มาเข้าพัก หรือ วันที่ยกเลิก
21. **deposit\_type** คือ ประเภทของเงินค้ำประกันการจอง โดยแบ่งเป็น No Deposit(ไม่มีเงินค้ำประกัน),NonRefund(ขอเงินคืนไม่ได้),Refundable(ขอเงินคืนได้)
22. **agent** คือ รหัสของ  Travel Agents ที่ทำการจอง
23. **company** คือ รหัสของบริษัทที่ทำการจอง
24. **days\_in\_waiting\_list** คือ วันที่รอการยืนยันการจองจากทางโรงแรม
25. **customer\_type** คือ ประเภทการจอง แบ่งเป็น Contract, Group, Transient, Transient-party
26. **adr** คือ รายได้เฉลี่ยต่อวันโดยหามาจาก เงินที่ได้/จำนวนวันที่ค้างคืน
27. **required\_car\_parking\_spaces** คือ จำนวนที่จอดรถที่ลูกค้าต้องการ
28. **total\_of\_special\_requests** คือ จำนวนคำขอพิเศษจากลูกค้าเช่น เตียงคู่ หรือ ขอชั้นสูง
29. **reservation\_status** คือ สถานการณ์จองบ่งเป็น Canceled(ยกเลิกการจอง), Check-Out(เข้าพักและออกไปแล้ว), No-Show(ไม่ได้เข้าพักและไม่มีการติดต่อโรงแรม)
30. **reservation\_status\_date** คือ วันที่สถานการณ์การจองครั้งล่าสุดบันทึกไว้

# **สาเหตุที่มาของการเลือกข้อมูลชุดนี้**

มีการเก็บข้อมูลในระยะเวลานานเพียงพอที่จะสามารถวิเคราะห์ข้อมูลที่แม่นยำได้ และมีข้อมูลจำนวนมากทำให้สามารถวิเคราหะได้หลายเรื่อง เช่น ช่วงเวลาที่ควรไปเข้าพักที่สุด, โรงแรมที่สามารถยอมรับคำขอพิเศษได้ดีที่สุด

# **สิ่งที่น่าสนใจจากการวิเคราะห์ข้อมูลเบื้องต้น**

คอลัมน์ **adr** หากไม่มีการเข้าพักค้างคืนจะทำให้รายรับกลายเป็น 0

# **คำถามองค์ความรู้เชิงลึก**

1. จำนวนผู้เข้าพักแต่ละช่วงวัยส่งผลต่อการเลือกจองอาหารแต่ละประเภทมากแค่ไหน
2. โอกาสในการยกเลิกการจองโรงแรมของการจองล่วงหน้ามากกว่า 1 ปี การจองแบบไม่มีเด็กและทารก การจองแบบคืนเงินได้ ในแต่ละเดือนของแต่ละปีมีมากเท่าไร
3. ปัจจัยที่ส่งผลในการยกเลิกโรงแรม

# **สาเหตุที่น่าสนใจของคำถามองค์ความรู้เชิงลึก**

1. จำนวนผู้เข้าพักแต่ละช่วงวัยส่งผลต่อการเลือกจองอาหารแต่ละประเภทมากแค่ไหนเพื่อเป็นการวิเคราะห์จำนวนการจัดเตรียมอาหารสำหรับแขกที่เข้าพักโดยดูจากจำนวนและวัยของผู้เข้าพักในโรงแรม
2. โอกาสในการยกเลิกการจองโรงแรมของการจองล่วงหน้ามากกว่า 1 ปี การจองแบบไม่มีเด็กและทารก การจองแบบคืนเงินได้ ในแต่ละเดือนของแต่ละปีมีมากเท่าไร ดูความเสี่ยงในการโดนยกเลิกโรงแรมโดยเริ่มวิเคราะห์จากกลุ่มตัวอย่างข้างต้น
3. ปัจจัยที่ส่งผลในการยกเลิกโรงแรม เพื่อวิเคราะห์หาสาเหตุที่ส่งผลมากที่สุดสำหรับการโดนยกเลิกการจอง

# **กระบวนการได้มาซึ่งองค์ความรู้**

1. จำนวนผู้เข้าพักแต่ละช่วงวัยส่งผลต่อการเลือกจองอาหารแต่ละประเภทมากแค่ไหน
   1. จัดกลุ่มของผู้ใหญ่ เด็ก และ ทารกที่มาเข้าพัก โดย 0 หมายถึงไม่มีเข้าพัก 1 หมายถึงเข้าพัก

1-5 คน และ 2 หมายถึงเข้าพักมากกว่า 5 คน

* 1. เช็คจำนวนว่าแต่ละกลุ่มจองอาหารแบบไหนเป็นจำนวนเท่าไหร่
  2. เอาไปหารจำนวนทั้งหมด และ แปลงเป็นร้อยละ
  3. พล็อตกราฟเปรียบเทียบ

1. โอกาสในการยกเลิกการจองโรงแรมของการจองล่วงหน้ามากกว่า 1 ปี การจองแบบไม่มีเด็กและทารก การจองแบบคืนเงินได้ ในแต่ละเดือนของแต่ละปีมีมากเท่าไร
   1. เช็คว่าในแต่ละปีมีเดือนไหนบ้างที่มีการจอง
   2. หาจำนวนการยกเลิกการจองที่เข้าเงื่อนไขที่กำหนด
   3. เอาไปหารจำนวนที่เข้าพักทั้งหมดในเดือนนั้น และ แปลงเป็นร้อยละ
   4. พล็อตกราฟเปรียบเทียบ
2. ปัจจัยที่ส่งผลต่อการยกเลิกโรงแรม
   1. หาค่า nan
   2. แทนค่า nan
   3. เปลี่ยน agent, company, booking.previous\_cancellations, previous\_bookings\_not\_canceled, days\_in\_waiting\_list ให้เป็น 0,1

โดย 0 คือ ไม่มี และ 1 คือ มี

* 1. หาค่า column ที่เป็น continuousbooking และ qualitativebooking
  2. ตั้งค่าให้ค่าข้างต้นเป็น list
  3. หาค่าที่หลุดเกินค่าเกาะกลุ่ม และ แปลงค่าให้เป็นค่าสูงสุดที่เกาะกลุ่ม
  4. พล็อตกราฟดูช่องที่สีเข้มที่สุดใน column is\_canceled

# **ผลลัพธ์องค์ความรู้ที่ได้**

**คำถามองค์ความรู้เชิงลึก**

1.จำนวนผู้เข้าพักแต่ละช่วงวัยส่งผลต่อการเลือกจองอาหารแต่ละประเภทมากแค่ไหน

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

ส่งผลไม่มากนัก เนื่องจากผู้เข้าพักทุกช่วงวัยส่วนมากแล้วมักจะเลือก BB อยู่แล้ว ส่วนอาหารประเภทที่เหลือจะแตกต่างกันที่จำนวนนิดหน่อย

เมื่อร่วมกันเป็นภาพรวมแล้วจะได้กราฟนี้

A screenshot of a cell phone

Description automatically generated

2.โอกาสในการยกเลิกการจองล่วงหน้าของ การจองล่วงหน้ามากกว่า 1 ปี , การจองแบบไม่มีเด็กและทารก และ การจองแบบคืนเงินได้ ในแต่ละเดือนของปีมีมากเท่าไร

A screenshot of a cell phone

Description automatically generated

ปี 2015 เดือน 6 มีโอกาส 25%

เดือน 10 , 12 มีโอกาส 100%

A screenshot of a cell phone

Description automatically generated

ปี 2016 เดือน 2 , 6 มีโอกาส 0%

เดือน 3 , 4 , 7 , 11 , 12 มีโอกาส 100%

เดือน 5 มีโอกาส 80% เดือน 8 มีโอกาส 90%

เดือน 9 มีโอกาส 65 % เดือน 10 มีโอกาส 50%

A screenshot of a cell phone

Description automatically generated

ปี 2017 เดือน 1,2,3,4 มีโอกาส 100%

เดือน 5 กว่า 70% เดือน 6, 7, 8 กว่า 30%

3.ปัจจัยที่ส่งผลต่อการยกเลิกโรงแรม

A close up of a piece of paper

Description automatically generated

ปัจจัยที่ส่งผลต่อการยกเลิกโรงแรม คือ lead\_time , previous\_cancellations และ agent,days\_in\_waiting\_list