

รายงาน

Analysis data of Hotel & Resort

จัดทำโดย

น.ส. วิวารวรรณ	ลือศักดิ์ชกร	6110406210
น.ส. วีรินทร์	เนียมเนบ	6110406228
นาย กิตติรัช	จึงโสภณวิทวัส	6110450022
นาย สหรัฐ	ชุตภา	6110450995
นาย วรณ	สรรพกิจ	6110451398

เสนอ

อาจารย์เสฏฐวิทย์ เกิดผล

รายงานฉบับนี้เป็นส่วนหนึ่งของรายวิชา

Introduction to Data Science

สารบัญ

คำอธิบายข้อมูลตามคอลัมน์.....	2
สาเหตุที่มาของการเลือกข้อมูลชุดนี้.....	4
สิ่งที่น่าสนใจจากการวิเคราะห์ข้อมูลเบื้องต้น.....	4
คำถามองค์ความรู้เชิงลึก	4
สาเหตุที่น่าสนใจของคำถามองค์ความรู้เชิงลึก	4
กระบวนการได้มาซึ่งองค์ความรู้.....	5
ผลลัพธ์องค์ความรู้ที่ได้.....	6

แหล่งที่มาของข้อมูล

ข้อมูลเปรียบเทียบการเข้าพักและยกเลิกการจอง ระหว่าง โรงแรมในเมือง และ โรงแรมรีสอร์ท ในช่วง กรกฎาคม 2015 – สิงหาคม 2017 อ้างอิงจาก บทความ Hotel Booking Demand Datasets ในหนังสือ Data in Brief เล่มที่ 22, กุมภาพันธ์ 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)
3. **lead_time** คือ วันที่ผ่านไประหว่างวันที่มีการจองเข้าสู่ระบบ PMS และ วันที่มาถึง
4. **arrival_date_year** คือ ปีที่มาเข้าพักที่โรงแรม
5. **arrival_date_month** คือ เดือนที่มาเข้าพักที่โรงแรม
6. **arrival_date_week_number** คือ จำนวนสัปดาห์ในปีที่มาเข้าพักที่โรงแรม
7. **arrival_date_day_of_month** คือ วันที่ที่มาเข้าพักที่โรงแรม
8. **stays_in_weekend_nights** คือ จำนวนวันที่มีการจอง หรือ เข้าพัก ค้างคืนในโรงแรมช่วงวันสุดสัปดาห์(วันเสาร์ และ วันอาทิตย์)
9. **stays_in_week_nights** คือ จำนวนวันที่มีการจอง หรือ เข้าพัก ค้างคืนในโรงแรมช่วงวันธรรมดา(วันจันทร์ถึงวันศุกร์)
10. **adults** คือ จำนวนผู้ใหญ่ที่มาเข้าพัก
11. **children** คือ จำนวนเด็กที่มาเข้าพัก
12. **babies** คือ จำนวนทารกที่มาเข้าพัก
13. **meal** คือ ประเภทอาหารที่มีการจองไว้ แบ่งเป็น SC(ไม่มีอาหาร), BB(อาหารเช้า), HB(อาหารเช้าและอีกมื้อ), FB(อาหารเช้า, อาหารกลางวัน, อาหารเย็น)
14. **country** คือ ประเทศที่มา

15. **market_segment** คือ การแบ่งส่วนตลาด โดยที่ TA หมายถึง Travel Agents และ TO หมายถึง Tour Operators
16. **distribution_channel** คือ ช่องทางการซื้อ โดยที่ TA หมายถึง Travel Agents และ TO หมายถึง Tour Operators
17. **is_repeated_guest** คือ การแสดงข้อมูลว่าลูกค้าที่มาพักเคยมาพักหรือไม่ โดยที่แบ่งเป็น 1 (เคยมาพัก) และ 0 (ไม่เคยมาพัก)
18. **previous_cancellations** คือ จำนวนครั้งที่ถูกยกเลิกการจองก่อนหน้าการจองรอบปัจจุบัน
19. **previous_bookings_not_canceled** คือ จำนวนครั้งที่ไม่ถูกยกเลิกการจองก่อนหน้าการจองรอบปัจจุบัน
20. **reserved_room_type** คือ ประเภทห้องที่มีการจองไว้
21. **assigned_room_type** คือ ประเภทห้องที่เข้าพัก
22. **booking_changes** คือ จำนวนรอบที่มีการแก้ไขการจองตั้งแต่วันที่มีการจองเข้าสู่ระบบ PMS จนถึงวันที่มาเข้าพัก หรือ วันที่ยกเลิก
23. **deposit_type** คือ ประเภทของเงินค้ำประกันการจอง โดยแบ่งเป็น No Deposit(ไม่มีเงินค้ำประกัน), NonRefund(ขอเงินคืนไม่ได้), Refundable(ขอเงินคืนได้)
24. **agent** คือ รหัสของ Travel Agents ที่ทำการจอง
25. **company** คือ รหัสของบริษัทที่ทำการจอง
26. **days_in_waiting_list** คือ วันที่รอการยืนยันการจองจากทางโรงแรม
27. **customer_type** คือ ประเภทการจอง แบ่งเป็น Contract, Group, Transient, Transient-party
28. **adr** คือ รายได้เฉลี่ยต่อวันโดยมาจาก เงินที่ได้/จำนวนวันที่ค้างคืน
29. **required_car_parking_spaces** คือ จำนวนที่จอดรถที่ลูกค้าต้องการ
30. **total_of_special_requests** คือ จำนวนคำขอพิเศษจากลูกค้าเช่น เตียงคู่ หรือ ขอชั้นสูง
31. **reservation_status** คือ สถานการณ์การจองแบ่งเป็น Canceled(ยกเลิกการจอง), Check-Out(เข้าพักและออกไปแล้ว), No-Show(ไม่ได้เข้าพักและไม่มีการติดต่อโรงแรม)
32. **reservation_status_date** คือ วันที่สถานการณ์การจองครั้งล่าสุดบันทึกไว้

สาเหตุที่มาของการเลือกข้อมูลชุดนี้

มีการเก็บข้อมูลในระยะเวลาอันยาวนานเพียงพอที่จะสามารถวิเคราะห์ข้อมูลที่แม่นยำได้ และมีข้อมูลจำนวนมากทำให้สามารถวิเคราะห์ได้หลายเรื่อง เช่น ช่วงเวลาที่ควรไปเข้าพักที่สุด, โรงแรมที่สามารถยอมรับค่าขอพิเศษได้ดีที่สุด

สิ่งที่น่าสนใจจากการวิเคราะห์ข้อมูลเบื้องต้น

คอลัมน์ **adr** หากไม่มีการเข้าพักค้างคืนจะทำให้รายรับกลายเป็น 0

คำถามองค์ความรู้เชิงลึก

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

สาเหตุที่น่าสนใจของคำถามองค์ความรู้เชิงลึก

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

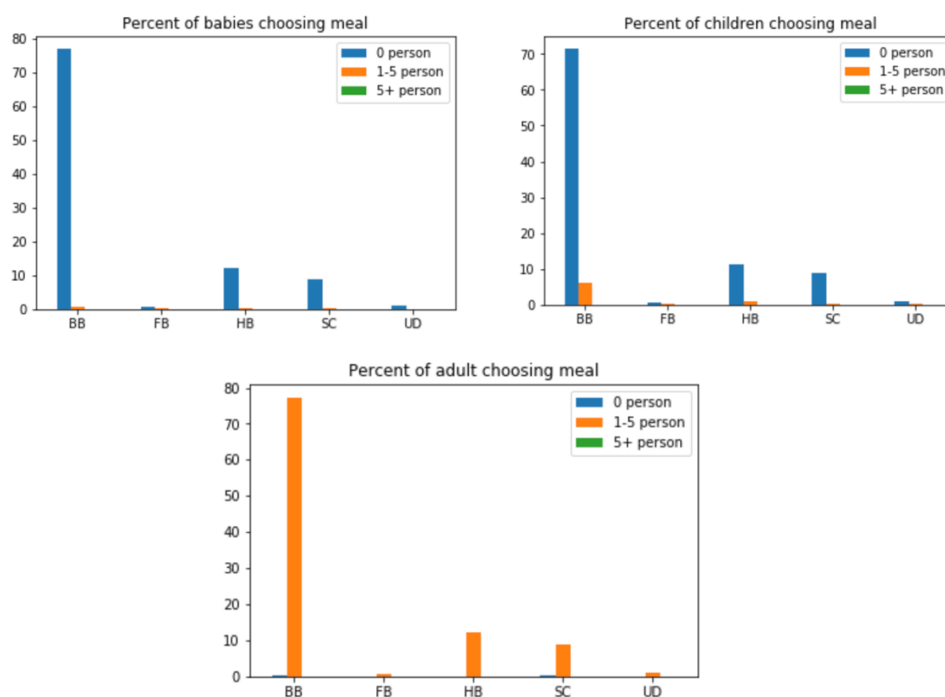
กระบวนการได้มาซึ่งองค์ความรู้

1. จำนวนผู้เข้าพักแต่ละช่วงวัยส่งผลต่อการเลือกจองอาหารแต่ละประเภทมากแค่ไหน
 - 1.1 จัดกลุ่มของผู้ใหญ่ เด็ก และ ทารกที่มาเข้าพัก โดย 0 หมายถึงไม่มีเข้าพัก 1 หมายถึงเข้าพัก 1-5 คน และ 2 หมายถึงเข้าพักมากกว่า 5 คน
 - 1.2 เช็คว่าจำนวนของแต่ละกลุ่มจองอาหารแบบไหนเป็นจำนวนเท่าไร
 - 1.3 เอาไปหารจำนวนทั้งหมด และ แปลงเป็นร้อยละ
 - 1.4 พล็อตกราฟเปรียบเทียบ
2. โอกาสในการยกเลิกการจองโรงแรมของการจองล่วงหน้ามากกว่า 1 ปี การจองแบบไม่มีเด็ก และทารก การจองแบบคืนเงินได้ ในแต่ละเดือนของแต่ละปีมีมากเท่าไร
 - 2.1 เช็คว่าในแต่ละปีมีเดือนไหนบ้างที่มีการจอง
 - 2.2 หาจำนวนการยกเลิกการจองที่เข้าเงื่อนไขที่กำหนด
 - 2.3 เอาไปหารจำนวนที่เข้าพักทั้งหมดในเดือนนั้น และ แปลงเป็นร้อยละ
 - 2.4 พล็อตกราฟเปรียบเทียบ
3. ปัจจัยที่ส่งผลต่อการยกเลิกโรงแรม
 - 3.1 หาค่า nan
 - 3.2 แทนค่า nan
 - 3.3 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
 - 3.4 หาค่า column ที่เป็น continuousbooking และ qualitativebooking
 - 3.5 ตั้งค่าให้ค่าข้างต้นเป็น list
 - 3.6 หาค่าที่หลุดเกินค่าเกาะกลุ่ม และ แปลงค่าให้เป็นค่าสูงสุดที่เกาะกลุ่ม
 - 3.7 พล็อตกราฟดูช่องที่สีเข้มที่สุดใน column is_canceled

ผลลัพธ์องค์ความรู้ที่ได้

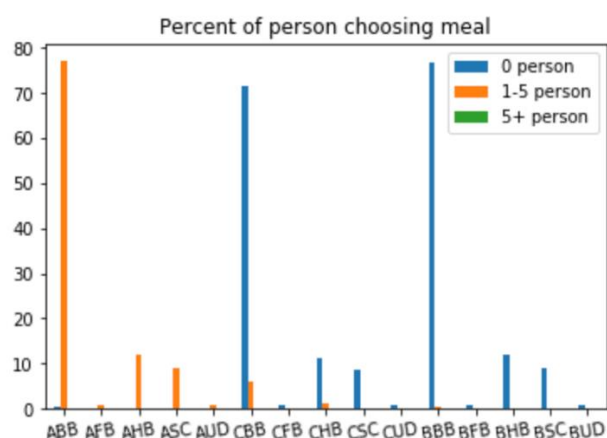
คำถามองค์ความรู้เชิงลึก

1.จำนวนผู้เข้าพักแต่ละช่วงวัยส่งผลต่อการเลือกของอาหารแต่ละประเภทมากแค่ไหน

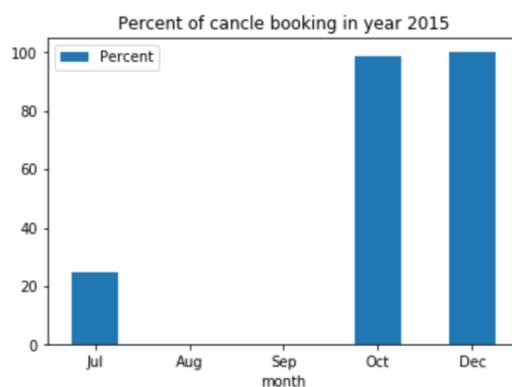


ส่งผลไม่มากนัก เนื่องจากผู้เข้าพักทุกช่วงวัยส่วนมากแล้วมักจะเลือก BB อยู่แล้ว ส่วนอาหารประเภทที่เหลือจะแตกต่างกันที่จำนวนนิดหน่อย

เมื่อรวมกันเป็นภาพรวมแล้วจะได้กราฟนี้

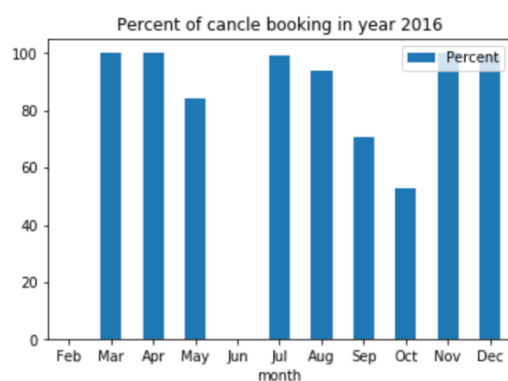


2.โอกาสในการยกเลิกการจองล่วงหน้าของ การจองล่วงหน้ามากกว่า 1 ปี , การจองแบบไม่มีเด็กและทารก และ การจองแบบคืนเงินได้ ในแต่ละเดือนของปีมีมากเท่าไร



ปี 2015 เดือน 6 มีโอกาส 25%

เดือน 10 , 12 มีโอกาส 100%

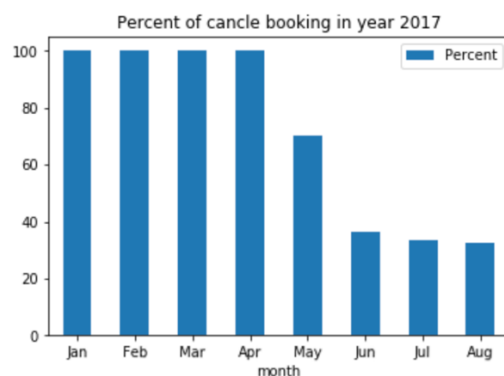


ปี 2016 เดือน 2 , 6 มีโอกาส 0%

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

เดือน 5 มีโอกาส 80% เดือน 8 มีโอกาส 90%

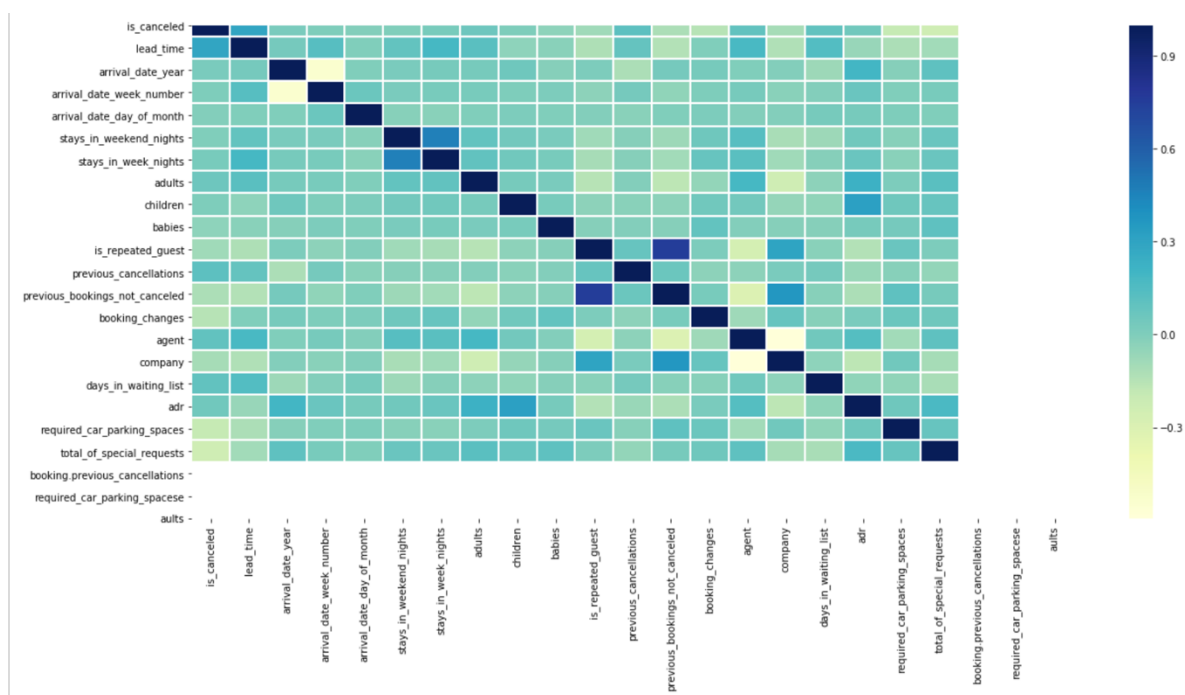
เดือน 9 มีโอกาส 65 % เดือน 10 มีโอกาส 50%



ปี 2017 เดือน 1,2,3,4 มีโอกาส 100%

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

3.ปัจจัยที่ส่งผลต่อการยกเลิกโรงแรม



ปัจจัยที่ส่งผลต่อการยกเลิกโรงแรม คือ lead_time , previous_cancellations และ agent,days_in_waiting_list