

3.2) การหาข้อมูลเชิงลึกจาก Interactive Dashboard

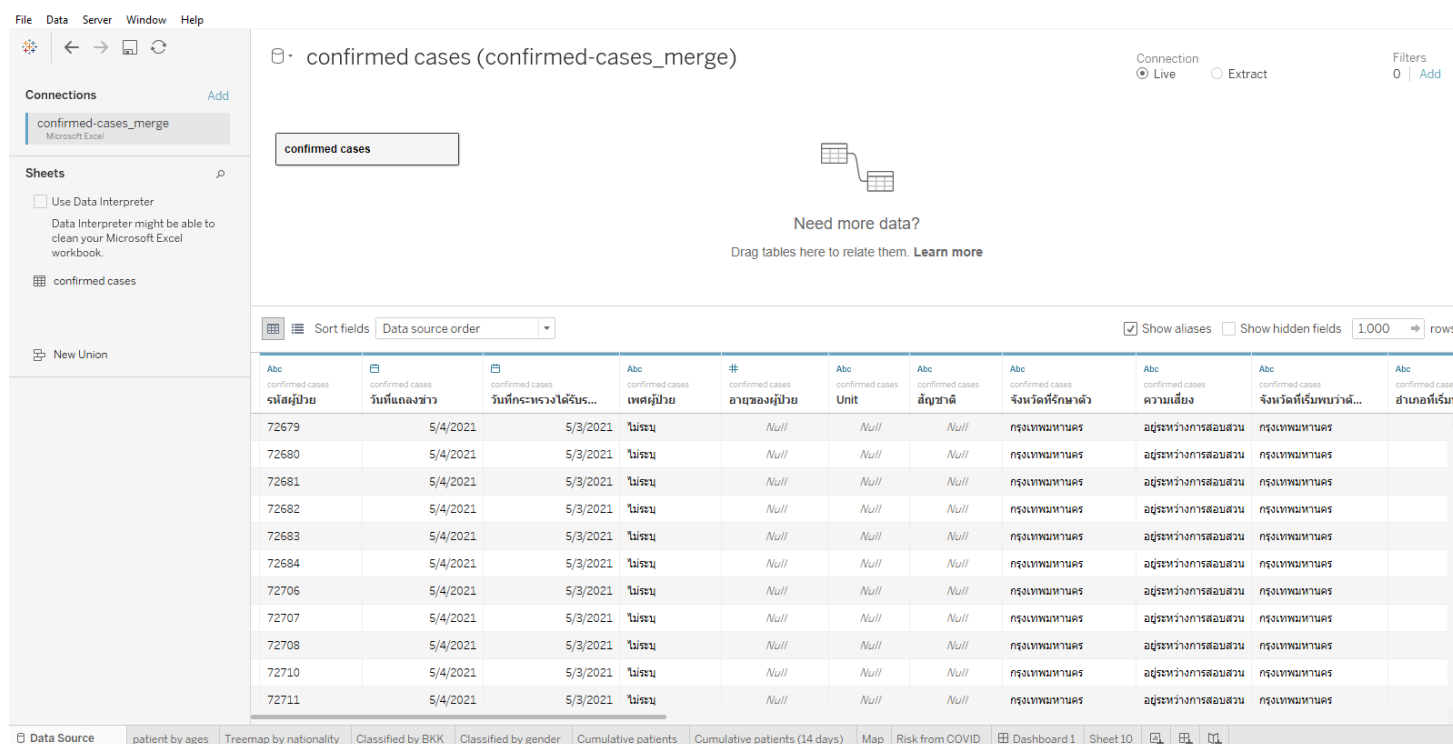
3.2.1) ที่มา ความสำคัญ และวัตถุประสงค์ของการทำงาน

หนึ่งในวิธีการนำเสนอข้อมูลจากการทำ Data Visualization ให้ผู้อื่นเข้าใจคือ การทำ Dashboard ซึ่งเป็นการนำแต่ละ worksheet ที่สร้างเอาไว้ในโปรแกรมมาจัดวางให้อยู่ในหน้าเดียว โดยหนึ่งในเทคนิคการสร้าง Dashboard ที่สามารถทำให้ผู้ใช้ทำการหาข้อมูลเชิงลึก (Data insight) เพิ่มเติม คือการสร้าง Interactive Dashboard โดยผู้ใช้งานสามารถทำการเจาะลึกรายละเอียดของข้อมูลลงไปได้ตามการกระทำ (Action) ที่ได้ออกแบบไว้ในโปรแกรม ซึ่ง Tableau Desktop เป็นหนึ่งในโปรแกรมที่สามารถทำได้ ไม่ว่าจะเป็นการทำการกรอง (Filter), การไฮไลต์ส่วนสำคัญ (Highlight), การไปลิงค์ URL หรือ Sheet ที่ผู้ใช้ได้ออกแบบไว้ เป็นต้น

ทั้งนี้ในการสร้าง Interactive Dashboard ดังกล่าวจะใช้ข้อมูลของผู้ป่วยโควิดจากกรมควบคุมโรค และใช้เทคนิค Filter สำหรับให้ผู้ใช้งาน Dashboard นั้นสามารถเห็นข้อมูลเฉพาะบริเวณที่กดเข้าไปใน sheet ที่ได้ตั้งไว้ใน Dashboard โดยข้อมูลที่ทำกรกดจะประยุกต์ใช้กับ sheet ต่างๆ ที่ผู้สร้าง Dashboard ได้กำหนดไว้

3.2.2) วิธีการทำงาน

ขั้นตอนแรกของการทำ Data Visualization คือ การนำเข้าชุดข้อมูลเข้าโปรแกรม Tableau โดยให้ทำการเชื่อมต่อข้อมูลกับไฟล์ excel “confirmed-cases_merge” ใน sheet ที่ชื่อว่า “confirmed cases” ซึ่งเมื่อเชื่อมต่อสำเร็จแล้ว จะสังเกตเห็น sheet ของ “confirmed cases” ปรากฏในบริเวณของ Canvas ของ Tableau รวมไปถึงภาพรวมของตารางต่างๆใน sheet นั้น ในบริเวณด้านล่างของ Canvas ดังแสดงในรูปที่ 3.3

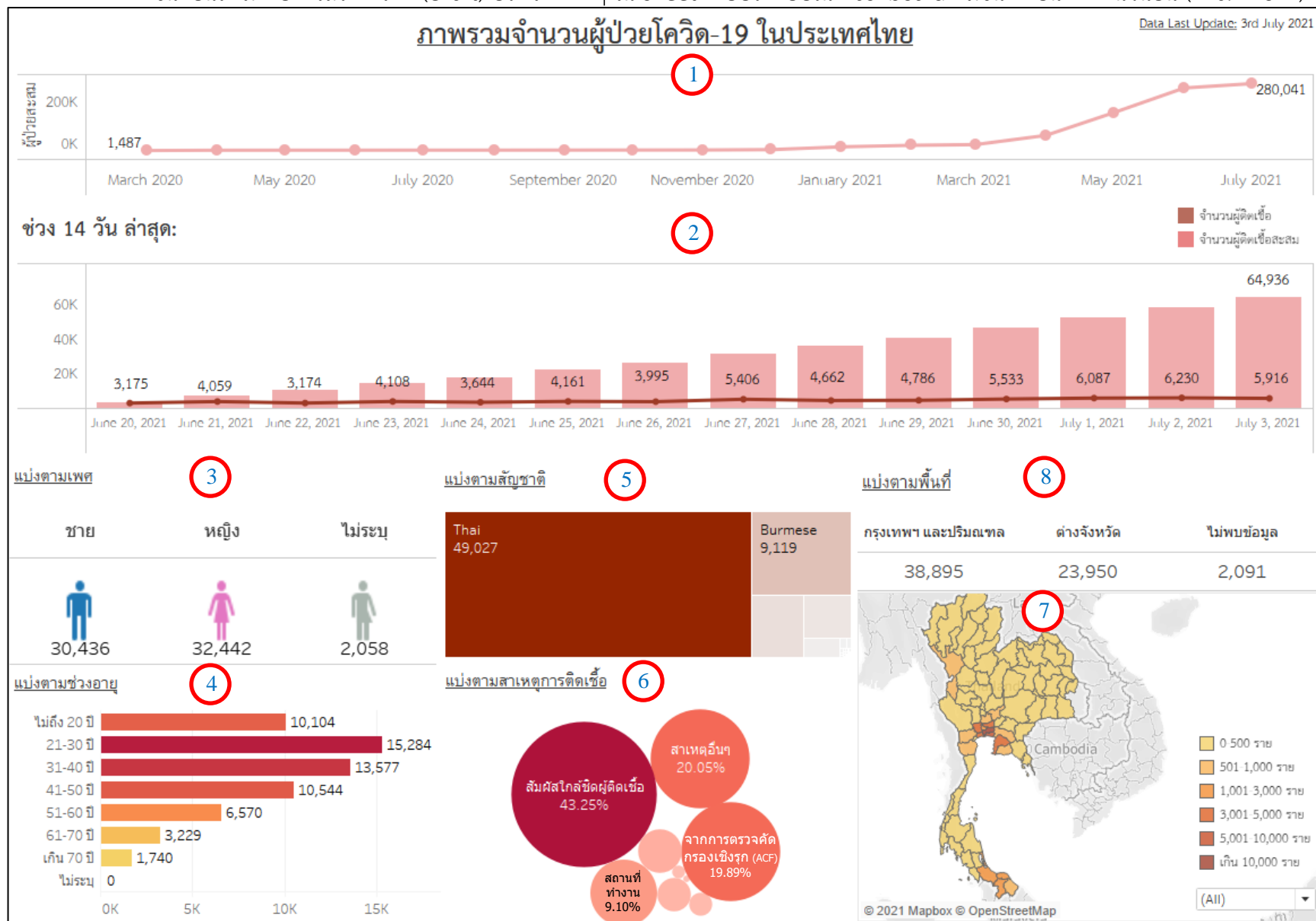


Abc confirmed cases รหัสผู้ป่วย	confirmed cases วันที่แสดงข่าว	confirmed cases วันที่กระทรวงได้รับ...	Abc confirmed cases เพศผู้ป่วย	# confirmed cases อายุของผู้ป่วย	Abc confirmed cases Unit	Abc confirmed cases สัญชาติ	Abc confirmed cases จังหวัดที่รักษาตัว	Abc confirmed cases ความเสี่ยง	Abc confirmed cases จังหวัดที่เริ่มพบว่...	Abc confirmed cases อำเภอที่เริ่มพ...
72679	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72680	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72681	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72682	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72683	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72684	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72706	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72707	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72708	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72710	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	
72711	5/4/2021	5/3/2021	ไม่ระบุ	Null	Null	Null	กรุงเทพมหานคร	อยู่ระหว่างการสอบสวน	กรุงเทพมหานคร	

รูปที่ 3.3: หน้าต่างของโปรแกรม Tableau ใน Tab ของ Data Source

หลังทำการเชื่อมต่อชุดข้อมูลใน sheet “confirmed cases” ของไฟล์ “confirmed-cases_merge.xlsx” แล้ว

ขั้นตอนถัดมาคือการสร้างกราฟ (chart) ประเภทต่างๆ ใน sheet เพื่อประกอบใน Dashboard ที่มีขนาดเป็น A4 แนวนอน (1169 x 827) ดังแสดงในรูปที่ 3.4



รูปที่ 3.4: Dashboard ที่แสดงถึงข้อมูลผู้ป่วยโควิดในประเทศไทย

โดยกราฟต่างๆ มีรายละเอียดการสร้างที่แตกต่างกันต่อไปดังนี้

1 สร้างกราฟเส้น (Line Chart) แสดงจำนวนผู้ป่วยสะสมต่อเดือนที่ผ่านมา โดยให้แกน y เป็นสิ่งที่สนใจ ซึ่งเป็นผู้ป่วยโควิดสะสม และแกน x เป็นเดือนที่ผ่านมา เพื่อแสดงถึงแนวโน้มการเพิ่มขึ้นของผู้ป่วยในแต่ละเดือนว่าแตกต่างกันอย่างมีนัยยะสำคัญมากหรือน้อยเพียงใด โดยกราฟเส้นดังกล่าวจะกำหนดให้มีการปรับสูตรคำนวณสำหรับข้อมูล “confirmed cases (Count)” จากการนับยอด (Count) เป็นยอดสะสม (Running Total) รวมถึงทำการปรับข้อมูล “วันที่กระทรวงได้รับรายงาน” จาก Discrete field [Default] เป็น Continuous field และทำการแตกลำดับชั้น (Hierarchy) จาก YEAR [Default] เป็น MONTH

2 สร้าง Dual Axis ระหว่างกราฟเส้น (Line chart) และกราฟแท่ง (Bar chart) สำหรับแสดงจำนวนผู้ติดเชื้อรายวันและจำนวนผู้ติดเชื้อสะสมตามลำดับ โดยมีตัวกรองเป็นวันที่กระทรวงได้รับรายงานเป็น 14 วันล่าสุด ซึ่งตรงกับช่วงเวลาที่เชื้อโควิด-19 จะแสดงอาการ ทั้งนี้อาจมีการปรับสีให้อยู่ในโทนของสีแดงที่เหมาะสม เพื่อแสดงให้เห็นถึงความรู้สึกอันตราย ซึ่งหลังจากสร้าง Dual Axis chart แล้วให้ทำการปรับ Synchronized Axis ให้มี scale เดียวกัน เพื่อป้องกันการตีความที่ผิดพลาดตาม Best Practice ของหลักการ Visual Analytics

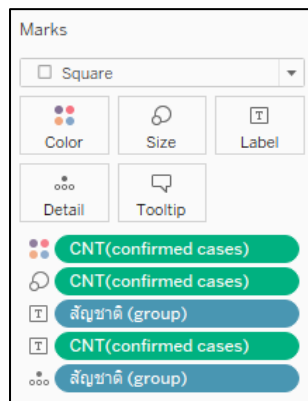
3 สร้าง crosstab ซึ่งแบ่งผู้ป่วยในช่วง 14 วันล่าสุดตามเพศ คือ ชาย หญิง และไม่ระบุ โดยให้แกนตั้งเป็นเพศ และแกนนอนเป็นจำนวนผู้ป่วยที่แบ่งตามเพศ จากนั้นทำการ Hide Axis แล้วกำหนดให้แต่ละแถวที่แสดงเพศมีสัญลักษณ์และสีแยกออกจากกันอย่างชัดเจน

4 สร้าง Horizontal bar chart โดยให้แกนตั้งเป็นช่วงอายุของผู้ป่วย และแกนนอนเป็นจำนวนผู้ป่วยตามแต่ละช่วงอายุ โดยมีการโค้ดตั้งเงื่อนไขของช่วงอายุใน Calculated field ที่ชื่อ “Age_category” ดังแสดงในรูปที่ 3.5 และมีการปรับโทนสีของแท่งกราฟเป็น Gradient Color ของ Red-Gold ตามจำนวนผู้ป่วยของแต่ละช่วงอายุ



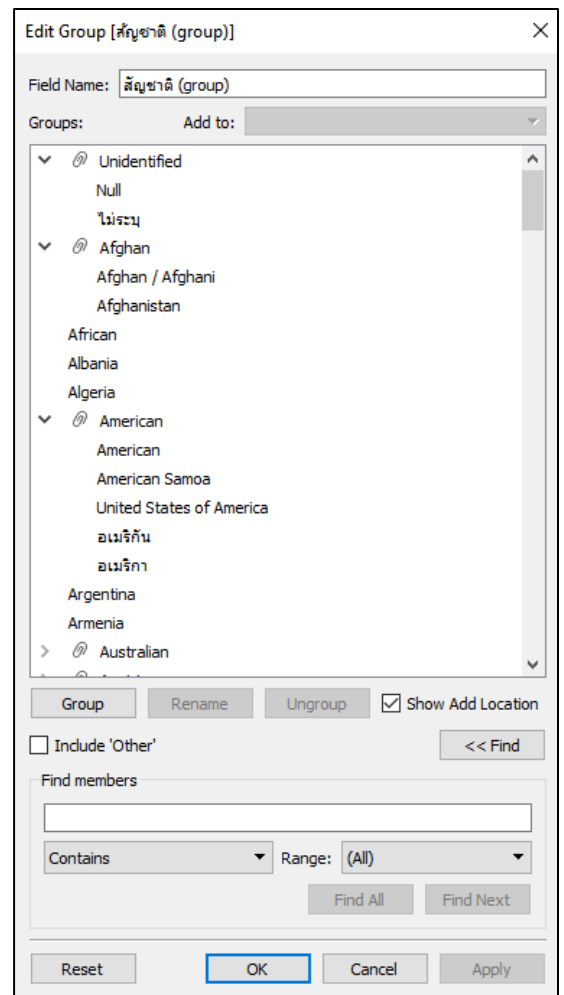
รูปที่ 3.5: โค้ดตั้งใน calculated field ที่ชื่อ “Age_category”

5 สร้าง Tree Map โดยมีขนาดและความเข้มของสีแปรตามสัดส่วนของผู้ป่วยของแต่ละสัญชาติในช่วง 14 วันล่าสุด ทั้งนี้สัญชาติผู้ป่วยจำเป็นต้องมีการจัดกลุ่มข้อมูล เช่น กลุ่ม Unidentified สำหรับข้อมูล “สัญชาติ” ที่เว้นว่าง (null) หรือ “ไม่ระบุ”, กลุ่ม “Afghan” สำหรับข้อมูล “สัญชาติ” ที่เป็น “Afghan / Afghani” หรือ “Afghanistan” เป็นต้น ดังแสดงในรูปที่ 3.6 ทั้งนี้กำหนด Mark card เป็น “Square” ดังแสดงในรูปที่ 3.7 และภายในสีเหลี่ยมมีการระบุชื่อสัญชาติที่มีการจัดกลุ่มแล้ว รวมถึงจำนวนผู้ป่วยตามสัญชาติกลุ่มนั้นๆ

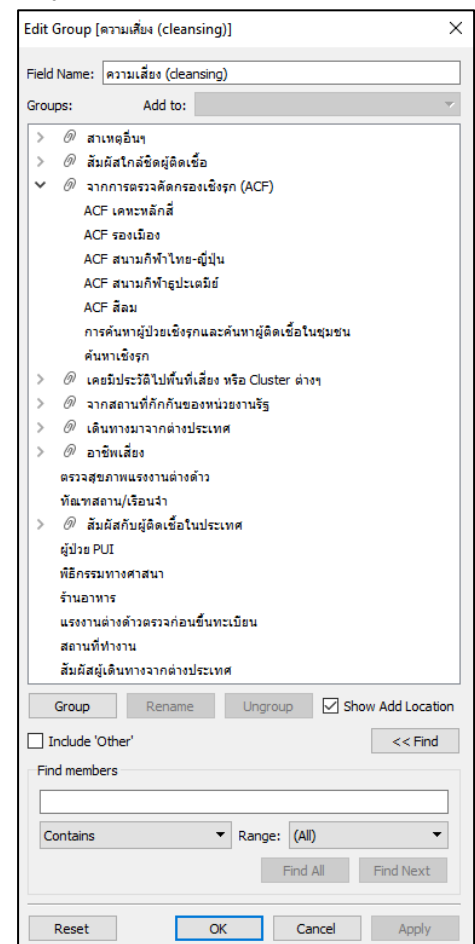


รูปที่ 3.7: Mark Card ของ Tree map

6 สร้าง Bubble Chart โดยมีขนาดของวงกลมและความเข้มของสีแดงแปรตามสัดส่วนของผู้ป่วยของแต่ละความเสี่ยงในช่วง 14 วันล่าสุด ทั้งนี้ความเสี่ยงต่อการติดเชื้อของผู้ป่วยจำเป็นต้องมีการจัดกลุ่มข้อมูล เช่น ในกลุ่มความเสี่ยง “จากการตรวจคัดกรองเชิงรุก (ACF)” ประกอบไปด้วยผู้ป่วยที่มีข้อมูล “ความเสี่ยง” เป็น “ACF เคหะหลักสี่”, “ACF รองเมือง”, “ACF สนามกีฬาไทย-ญี่ปุ่น”, “ACF สนามกีฬาจุฬารัตน์”, “การค้นหาผู้ป่วยเชิงรุกและค้นหาผู้ติดเชื้อในชุมชน” และ “ค้นหาเชิงรุก” ดังแสดงในรูปที่ 3.8 อย่างไรก็ตามในส่วน Mark card นั้นคล้ายกับการสร้าง Tree map แต่ให้ปรับการแสดงผลเป็น “Circle” รวมถึงในส่วน confirmed cases ให้มีการปรับสูตรการคำนวณ Quick Table Calculation ให้เป็น “Percent of Total” เพื่อแสดงเปอร์เซ็นต์ของสาเหตุการติดเชื้อแต่ละชนิด เทียบกับสาเหตุการติดเชื้อทั้งหมด

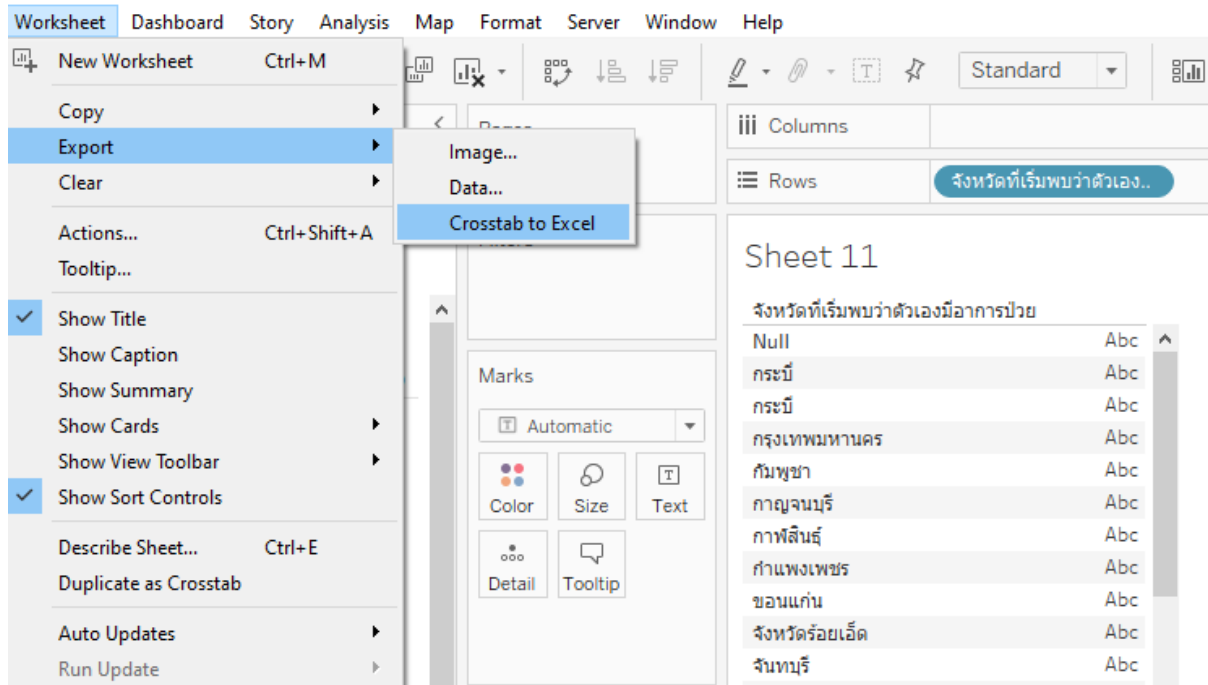


รูปที่ 3.6: การจัดกลุ่ม “สัญชาติ (group)”



รูปที่ 3.8: การจัดกลุ่ม “ความเสี่ยง (cleansing)”

7 สร้าง Area Map โดยแบ่งแต่ละจังหวัดให้แยกตามสีของจำนวนผู้ติดเชื้อสะสมในช่วง 14 วันล่าสุด โดยข้อมูล “จังหวัดที่เริ่มพบว่าตัวเองมีอาการ” ต้องได้รับการแก้ไขหรือทำความสะอาดก่อนเนื่องจากอาจมีการสะกดผิด หรือข้อมูลอาจรอกมาผิด เช่น การกรอกอำเภอแทนจังหวัด เป็นต้น ซึ่งในโปรแกรมให้ทำการเปิด Sheet ใหม่ แล้วทำการสร้าง Crosstab โดยมีข้อมูล “จังหวัดที่เริ่มพบว่าตัวเองมีอาการป่วย” เป็นแถวแนวนิ่ง และทำการส่งออก crosstab ดังกล่าวให้อยู่ในไฟล์ excel ดังแสดงในรูปที่ 3.9



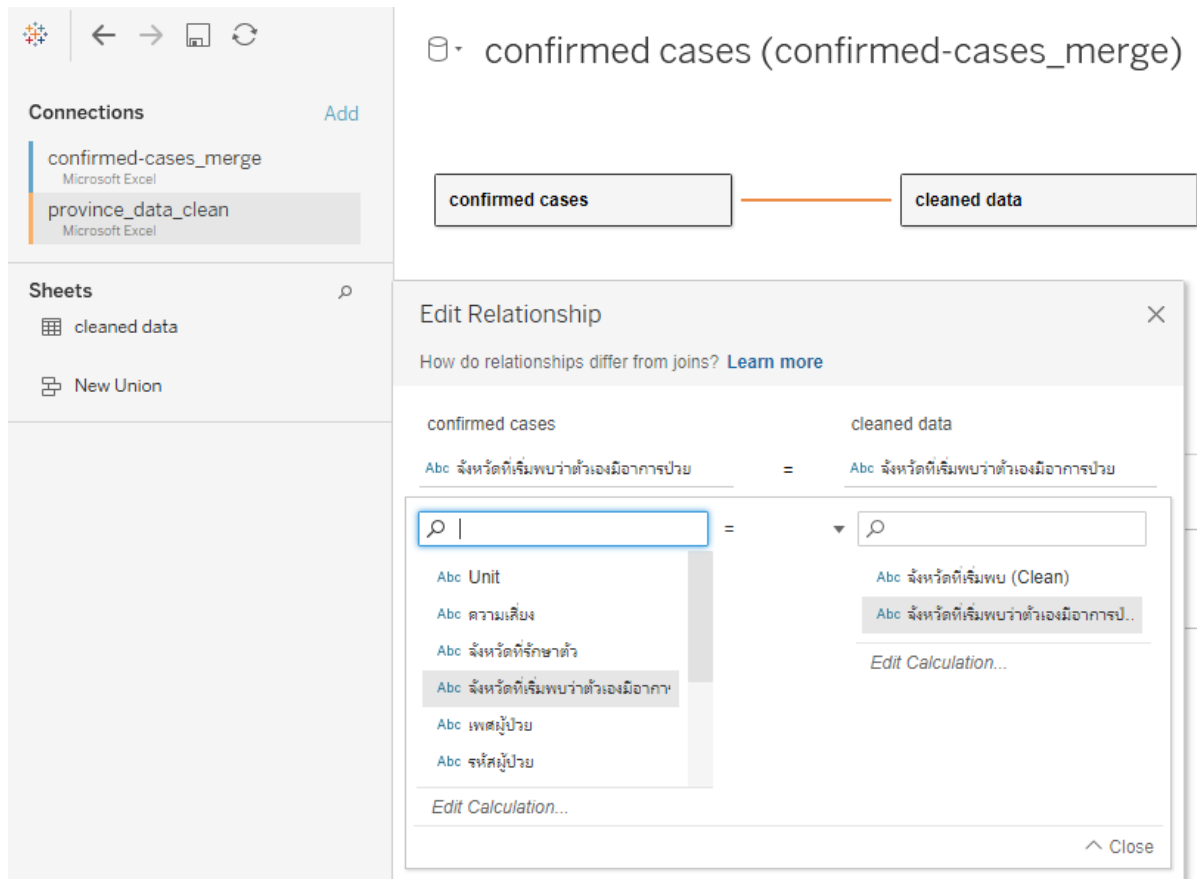
รูปที่ 3.9: การส่งออก crosstab ของข้อมูล

เมื่อทำการแปลงข้อมูลให้อยู่ในไฟล์ excel แล้วให้ทำการเปิดไฟล์และสร้าง column ใหม่ที่ชื่อว่า “จังหวัดที่เริ่มพบ (clean)” และเปลี่ยนชื่อ sheet เป็น “cleaned data” จากนั้นให้ทำการแก้ไขชื่อจังหวัดให้อยู่ในรูปของชื่อจังหวัดที่มีการสะกดที่ถูกต้องโดยไม่มีคำว่า “จังหวัด” นำหน้า รวมถึงชื่อประเทศอื่นให้เปลี่ยนเป็น “ต่างประเทศ” และในกรณีที่ เป็นช่องว่าง (Null) ให้เปลี่ยนเป็น “ไม่พบข้อมูล” ดังแสดงในรูปที่ 3.10

	A	B
1	จังหวัดที่เริ่มพบว่าตัวเองมีอาการป่วย	จังหวัดที่เริ่มพบ (clean)
2		ไม่พบข้อมูล
3	กระบี่	กระบี่
4	กระบี่	กระบี่
5	กรุงเทพมหานคร	กรุงเทพมหานคร
6	กัมพูชา	ต่างประเทศ
7	กาญจนบุรี	กาญจนบุรี
8	กาฬสินธุ์	กาฬสินธุ์
9	กำแพงเพชร	กำแพงเพชร
10	ขอนแก่น	ขอนแก่น
11	จังหวัดร้อยเอ็ด	ร้อยเอ็ด
12	จันทบุรี	จันทบุรี

รูปที่ 3.10: การแก้ไขข้อมูล “จังหวัดที่เริ่มพบว่าตัวเองมีอาการป่วย” ใน Excel

หลังจากเสร็จสิ้นการแก้ไขข้อมูล “จังหวัดที่เริ่มพบว่าตัวเองมีอาการป่วย” นั้นให้ทำการเชื่อมต่อข้อมูล โดยเปิดไปที่ Tab “Data Source” และทำการเชื่อมข้อมูลในไฟล์ excel ใน sheet ที่ได้แก้ไขก่อนหน้านี้ จากนั้นทำการเชื่อมความสัมพันธ์ด้วย Key ที่ตรงกัน ดังที่แสดงไว้ในรูปที่ 3.11 เพื่อให้สามารถเรียกใช้ข้อมูลที่ผ่านการแก้ไขในโปรแกรม Tableau



รูปที่ 3.11: การเชื่อมต่อข้อมูล “จังหวัดที่เริ่มพบว่าตัวเองมีอาการป่วย” ซึ่งได้รับการแก้ไขแล้ว

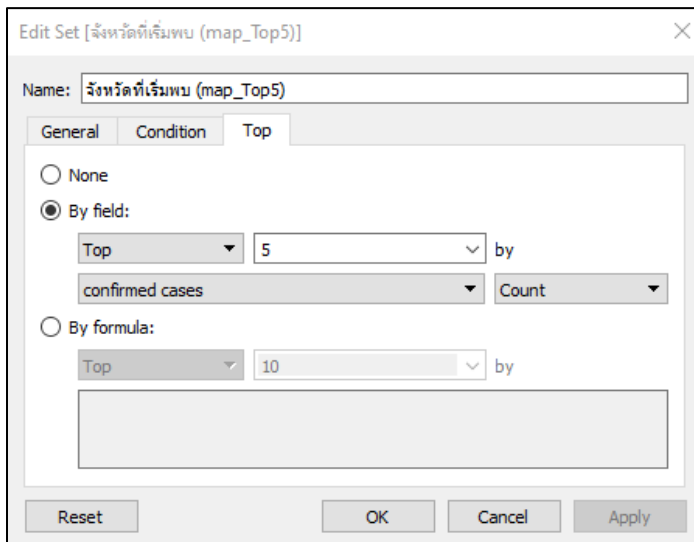
จากนั้นทำการสร้าง Calculated field ที่จัดจำแนกข้อมูล “confirmed cases” ตามจำนวนผู้ป่วยของแต่ละจังหวัด โดยมีการเขียนโค้ดดังที่แสดงในรูปที่ 3.12 และทำการสร้าง Area Map โดยใช้ข้อมูลที่ได้แก้ไข “จังหวัดที่เริ่มพบ (Clean)” ซึ่งมาจากการเชื่อมต่อข้อมูลในรูปที่ 3.11 โดยใช้เฉดสีแสดง ในการไล่ลำดับความรุนแรงของปริมาณผู้ติดเชื้อแต่ละจังหวัดที่ได้สร้างไว้ใน Calculated field



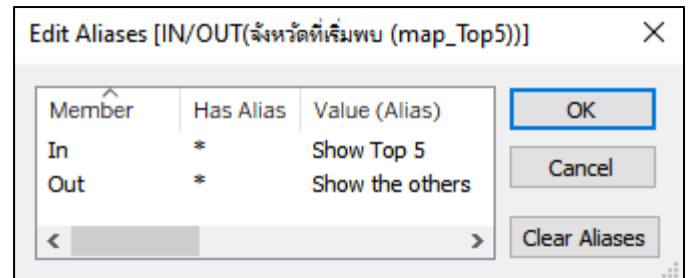
รูปที่ 3.12:

โค้ดใน Calculated field ที่ชื่อ “Count_category” สำหรับจัดประเภทความรุนแรงของปริมาณผู้ติดเชื้อแต่

นอกจากนั้นทำการสร้างเซตสำหรับ 5 จังหวัดที่พบผู้ป่วยสูงสุด ดังแสดงในรูปที่ 3.13 รวมถึงทำการแก้ไข Alias ของสมาชิกในเซต ดังแสดงในรูปที่ 3.14



รูปที่ 3.13: การสร้างเซต “จังหวัดที่เริ่มพบ (map_Top5)”

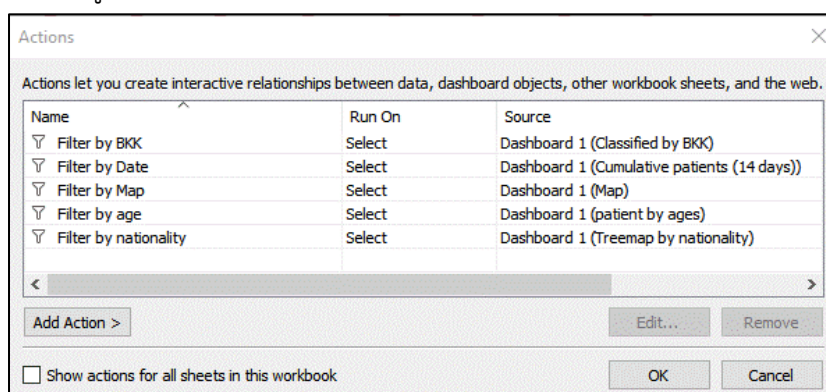


รูปที่ 3.14: การแก้ไข Alias ของสมาชิกในเซต
“จังหวัดที่เริ่มพบ (map_Top5)”

หมายเหตุ: สาเหตุที่ไม่ใช้ “Create Relationship” แต่ใช้ “Create Relationship” ในส่วนของการแก้ไขหรือทำความสะอาดข้อมูล เนื่องจากข้อมูล “จังหวัดที่เริ่มพบอาการป่วย” นั้นต้องนำไปใช้จัดกลุ่มจังหวัด “กรุงเทพและปริมณฑล” “จังหวัดที่เริ่มพบ (map_Top5)” หรือกลุ่มอื่นๆ ที่ใช้ข้อมูลจังหวัดที่ได้รับการแก้ไขแล้ว ซึ่งหากมีการอัปเดตข้อมูลเข้า Dashboard นี้ จะต้องทำการแก้ไขข้อมูล 2 รอบ ซึ่งเสียเวลามากกว่าการแก้ไขข้อมูลใน Excel แค่ครั้งเดียว ซึ่งก่อให้เกิดการสูญเสีย (Loss) ในการทำงาน และอาจจัดเป็นการทำงานที่ไม่มีประสิทธิภาพ

8 สร้าง Crosstab โดยมีแกนแนวนั่งเป็นกลุ่มจังหวัด “กรุงเทพและปริมณฑล”, “ต่างจังหวัด” และ “ไม่พบข้อมูล” ที่ใช้ข้อมูลจาก “จังหวัดที่เริ่มพบ (clean)” โดยกลุ่มจังหวัด “กรุงเทพและปริมณฑล” ประกอบไปด้วย “กรุงเทพมหานคร” “นครปฐม” “นนทบุรี” “ปทุมธานี” “สมุทรปราการ” และ “สมุทรสาคร” ในขณะที่กลุ่ม “ต่างจังหวัด” ประกอบไปด้วยจังหวัดอื่นๆ ภายนอกประเทศไทยที่ไม่ได้อยู่ใน “กรุงเทพและปริมณฑล”

หลังจากการนำ sheet ต่างๆ มาประกอบเป็น Dashboard ให้ทำการสร้าง Action โดยในที่นี้จะทำการสร้าง Filter ข้อมูลเมื่อเกิดการคลิกไปที่จุดใดจุดหนึ่งของ sheet ที่กำหนด ซึ่งคือ sheet หมายเลข 2, 4, 5, 7, 8 โดยกำหนดให้ประยุกต์ใช้กับการแสดงข้อมูลที่กดเข้าไปกับทุก sheet ยกเว้น sheet หมายเลข 1 ดังแสดงในรูปที่ 3.15



รูปที่ 3.15: Action ที่ถูกสร้างขึ้นใน Interactive Dashboard ซึ่งใช้ข้อมูลของผู้ป่วยโควิด-19 ในประเทศไทย

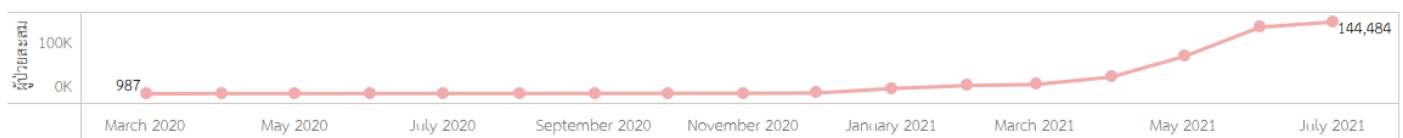
3.2.3) ผลลัพธ์ที่ได้ และสิ่งที่ได้เรียนรู้จากการทำงาน

เมื่อทำการคลิกตาม Action ต่างที่ได้สร้างไว้ใน Interactive Dashboard จะค้นพบ Insight ที่ซ่อนอยู่ภายในข้อมูล เช่น

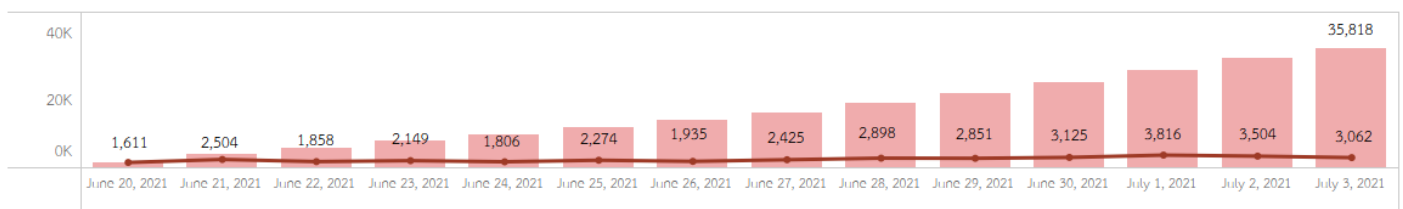
- ในบริเวณ Area Map เมื่อทำการเลือก “Top 5” สำหรับ 5 จังหวัดแรกที่พบผู้ป่วยสูงสุดในประเทศ พบว่าเป็นจังหวัดในกลุ่ม “กรุงเทพฯ และปริมณฑล” ทั้งหมด และสาเหตุการติดเชื้อส่วนใหญ่มาจากการสัมผัสใกล้ชิดผู้ติดเชื้อ รวมไปถึงกลุ่มคนที่ติดเชื้อมากที่สุด คือช่วงอายุ 21-30 ปี และหากเทียบยอดการติดเชื้อจะพบว่า ยอดผู้ติดเชื้อในช่วง 14 วันล่าสุด กินสัดส่วนมากที่สุด เมื่อเทียบกับข้อมูลทั้งหมด ดังนั้น 5 จังหวัดดังกล่าวจึงควรต้องมีการเฝ้าระวังเป็นพิเศษ

ภาพรวมจำนวนผู้ป่วยโควิด-19 ในประเทศไทย

Data Last Update: 3rd July 2021



ช่วง 14 วัน ล่าสุด:



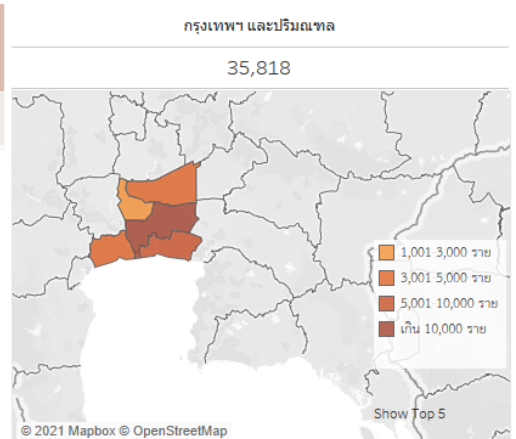
แบ่งตามเพศ



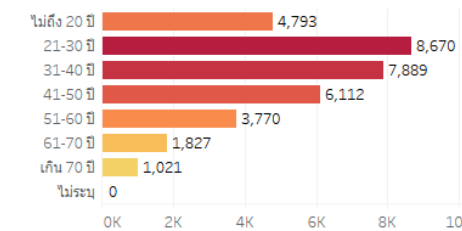
แบ่งตามสัญชาติ



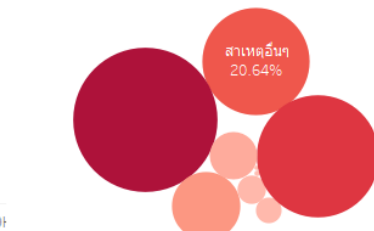
แบ่งตามพื้นที่



แบ่งตามช่วงอายุ



แบ่งตามสาเหตุการติดเชื้อ



รูปที่ 3.15: Interactive Dashboard ที่เปลี่ยนไปเมื่อเลือก Set Filter “Show Top 5” ที่บริเวณ sheet ของ Area Map

- ในบริเวณของ Tree map เมื่อทำการคลิกที่สัญชาติ “Burmese” จะพบว่านอกจากจังหวัดในประเทศไทยที่ติดกับชายแดนประเทศเมียนมาจะมีผู้ติดเชื้อแล้ว แต่ในบริเวณจังหวัด “กรุงเทพฯ และปริมณฑล” ทั้งหมดกลับพบด้วยเช่นกัน โดยเฉพาะในจังหวัด “กรุงเทพมหานคร” ที่มีสี่ของยอดผู้ติดเชื้อเข้มข้นที่สุด ซึ่งอาจตีความได้ว่าในจังหวัดดังกล่าวมีการจ้างแรงงานพม่าค่อนข้างเยอะ และความเป็นอยู่ หรือกิจวัตรประจำวันอาจเสี่ยงทำให้ติดเชื้อ เช่น ค่ายก่อสร้างที่ระหว่างวันมีอุณหภูมิที่สูงมากจนคนงานกลุ่มนี้ทนไม่ไหวจึงได้ทำการถอดหน้ากากอนามัย ซึ่งอาจทำให้ติดเชื้อ ดังนั้นหน่วยงานรัฐควรเฝ้าระวังและออกมาตรการการ

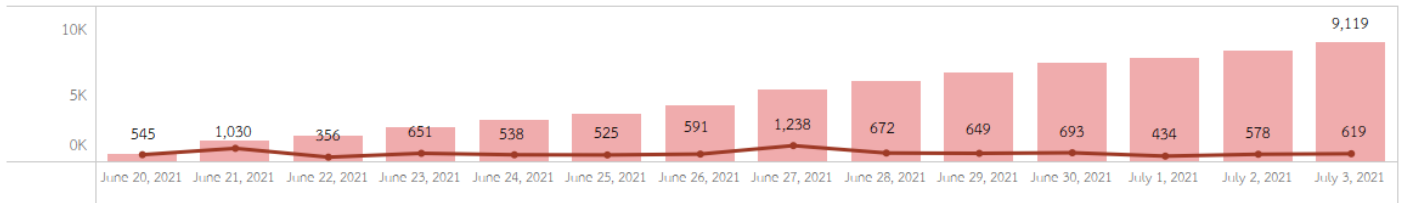
รับมือเพื่อหลีกเลี่ยงการเกิดพฤติกรรมที่เสี่ยงให้ติดเชื้อของกลุ่มคนเหล่านี้ โดยมีจุดประสงค์สำหรับป้องกันไม่ให้เกิดการติดเชื้อหรือคลัสเตอร์การระบาดที่รุนแรงไปกว่านี้ เป็นต้น

ภาพรวมจำนวนผู้ป่วยโควิด-19 ในประเทศไทย

Data Last Update: 3rd July 2021



ช่วง 14 วัน ล่าสุด:



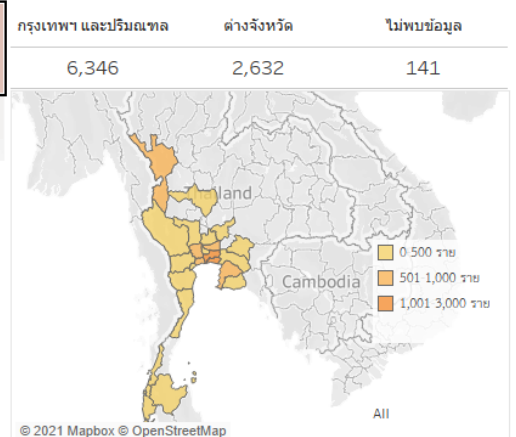
แบ่งตามเพศ



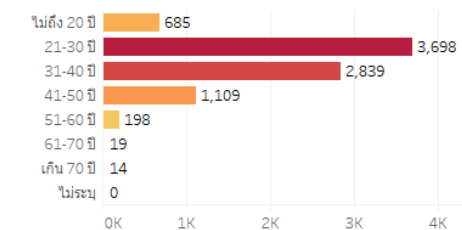
แบ่งตามสัญชาติ



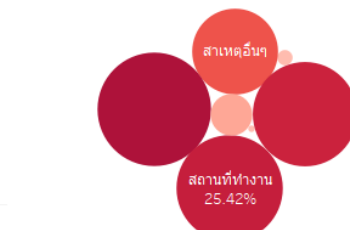
แบ่งตามพื้นที่



แบ่งตามช่วงอายุ



แบ่งตามสาเหตุการติดเชื้อ



รูปที่ 3.16: Interactive Dashboard ที่เปลี่ยนไปเมื่อเลือกสัญชาติ “Burmese” ที่บริเวณ sheet ของ Tree Map

นอกจากการหาข้อมูลที่แอบแฝง จาก Interactive Dashboard แล้ว หลักสำคัญที่สุดก่อนที่จะเริ่มการวิเคราะห์ข้อมูลก็คือการทำความสะอาดข้อมูล เช่น การจัดการข้อมูลที่หายไป ข้อมูลที่สะกดผิด ฯลฯ จากคำพูดในวิชาการควบคุมคุณภาพ (Quality Control, QC) ที่ได้กล่าวไว้ว่า “Garbage in, garbage out” ซึ่งหมายความว่าหากวัตถุดิบที่ใช้ในการกระบวนการผลิตมีคุณภาพที่ไม่ดีหรือแย่มากๆ สิ่งที่จะออกมาซึ่งเป็นสินค้าหรือผลิตภัณฑ์ ที่แม้จะมีกระบวนการผลิตที่ดีเยี่ยม ก็จะออกมามีคุณภาพที่แย่มากๆ ดังนั้นในการวิเคราะห์ข้อมูล ซึ่งครอบคลุมไปถึงการทำ Data Visualization ด้วย หากข้อมูลที่เราไปใช้วิเคราะห์ไม่ผ่านการเตรียมการหรือทำความสะอาด (Cleansing) ที่ดี ผลลัพธ์จากการวิเคราะห์ข้อมูลเหล่านั้นก็จะเกิดการตีความที่ผิดพลาดขึ้นได้ ซึ่งในทางธุรกิจแล้ว หากฝ่ายอื่นๆ ในองค์กรซึ่งนำผลลัพธ์ที่ผิดพลาดไปใช้ เช่น ฝ่ายจัดทำกลยุทธ์ ก็อาจทำให้เกิดการวางแผนที่ผิดพลาดจากผลการวิเคราะห์ข้อมูลที่ผิด ซึ่งอาจนำไปสู่การสูญเสีย และสูญเสียเปล่าภายในองค์กรได้