



9 บทเรียนง่าย ๆ ที่ไม่ควรพลาดเกี่ยวกับ การจัดการและการวิเคราะห์ข้อมูล กรณี COVID-19

รองศาสตราจารย์ ดร. ลัดดาวัลย์ แก้วกิติพงษ์
สาขาวิชาระบบสารสนเทศเพื่อการจัดการ

**“ หากข้อมูลที่น่าไปใช้งานต่อไม่ได้เป็นตัวแทนประชากร
ที่ดี ไม่ได้สะท้อนภาพความเป็นจริง ก็อาจจะทำให้เข้าใจ
สถานการณ์ผิดและตัดสินใจผิดได้ ”**

นาทนี่สิ่งที่เป็นกระแสหลักที่ทุกคนต่างให้ความสนใจคงหนีไม่พ้นเรื่องราวเกี่ยวกับ COVID-19 ในฐานะนักวิชาการด้านการบริหารจัดการข้อมูลสารสนเทศ นี่เป็นโอกาสที่ดีที่จะหยิบยกแง่มุมต่าง ๆ ที่เป็นเรื่องพื้นฐาน เป็นหลักการง่าย ๆ แต่สำคัญมากสำหรับนักวิเคราะห์ข้อมูลมาบอกเล่าสู่กันฟัง

1. ความสำคัญของการเลือกกลุ่มตัวอย่างและขนาดของตัวอย่าง - ท่านผู้อ่านหลายท่านเมื่อได้เห็นสถิติตัวเลขการติดเชื้อของประเทศไทยคงมีความสงสัยเกิดขึ้นกันบ้างว่า เอ...ที่ตัวเลขผู้ติดเชื้อของประเทศไทยไม่สูงมากนักเป็นเพราะเราไม่ได้ตรวจให้ทั่วถึงหรือเปล่านะ อาจจะมีหลายคนหรือเปล่าที่เป็นแต่ไม่ได้รับการตรวจหรือตรวจแต่เฉพาะในกรุงเทพหรือเปล่าจึงได้มีแต่ผู้ป่วยในกรุงเทพฯเยอะมาก หากท่านผู้อ่านท่านใดเคยดูคิดขึ้นมาแบบนี้ ยินดีด้วยค่ะ วิกฤต COVID-19 ได้ปลุกความเป็นนักวิเคราะห์ข้อมูลในตัวคุณแล้ว นักวิเคราะห์ข้อมูลที่ดีต้องรู้จักตั้งคำถามถึงวิธีการเก็บข้อมูล และการเลือกกลุ่มตัวอย่าง สำหรับนักสถิติและนักวิเคราะห์ข้อมูลแล้ว การตรวจสอบว่าข้อมูลนั้นครอบคลุม เป็นตัวแทนของประชากรที่ดีหรือไม่นั้นเป็นสิ่งสำคัญมาก หากข้อมูลที่ท่านนำไปใช้งานต่อไม่ได้เป็นตัวแทนประชากรที่ดี นั่นคือไม่ได้สะท้อนภาพความเป็นจริง ก็อาจจะทำให้ท่านเข้าใจสถานการณ์ผิดและตัดสินใจผิดได้

2. ความสำคัญของการทำ Data Visualization – ตัวอย่างรายงานอัปเดตสถานการณ์ผู้ติดเชื้อ COVID-19 จากเว็บไซต์ <https://covid19.workpointnews.com/> นั้นเป็นตัวอย่างของการทำ Visualization หรือการนำเสนอข้อมูลในรูปแบบรูปภาพที่ทำให้เข้าใจง่ายและเห็นข้อมูลที่ต้องการไฮไลท์ได้ในทันที ยกตัวอย่างเช่น การนำเสนอข้อมูลในลักษณะของแผนที่ Heatmap ที่แสดงให้เห็นว่าภูมิภาคใดจังหวัดใดมีผู้ติดเชื้อมากหรือน้อย โดยใช้เฉดสีจากแดงเข้ม แดงอ่อน เป็นชมพู และเทาเข้ามาช่วยนั้น ทำให้ผู้อ่านสามารถเห็นได้ภายในพริบตาเดียวกับที่มองแผนที่ว่าภูมิภาคหรือจังหวัดใดเริ่มเป็นโซนอันตรายหรือการแพร่ระบาดได้เริ่มกระจายไปถึงจุดใดมากน้อยขนาดไหนบ้างแล้ว ในยุค Big Data ที่มีข้อมูลจำนวนมากเกิดขึ้นในเวลาอันรวดเร็วความสามารถในการทำ Data Visualization โดยสามารถเลือกรูปแบบในการนำเสนอข้อมูลให้เข้าใจได้ง่ายที่สุดและสื่อความหมายได้มากที่สุดนั้นสำคัญมาก

3. กราฟเส้นเหมาะกับการแสดงเทรนด์ – ต่อเนื่องจากความสำคัญของการเลือกรูปแบบการนำเสนอที่เหมาะสม กราฟเส้นที่ถ่านผู้อ่านอาจจะเคยเห็นผ่านตากันมาบ้างคือ กราฟเส้นการเพิ่มขึ้นของจำนวนผู้ติดเชื้อ COVID-19 โดยที่เส้นกราฟหนึ่งเส้นนั้นเป็นตัวแทนของข้อมูลจากหนึ่งประเทศ นี่เป็นการเลือกการแสดงผลได้เหมาะสมกับข้อมูลมากที่สุด เมื่อคุณต้องการจะมองเห็นเทรนด์ เช่น มีการเพิ่มขึ้นสูงมาก (กราฟเส้นตั้งชัน) มีการเพิ่มขึ้นแต่เริ่มเพิ่มขึ้นแบบไม่มากแล้ว (กราฟเส้นเริ่ม flat) กราฟเส้นนั้นช่วยไฮไลท์ข้อมูลเหล่านี้ได้จริงๆ นอกจากนี้กราฟเส้นยังช่วยให้ผู้อ่านเห็นแพทเทิร์นบางอย่างหรือความแตกต่างของชุดข้อมูลที่ซ่อนอยู่ได้ เช่น จากกราฟเส้นการเพิ่มขึ้นของจำนวนผู้ติดเชื้อทำให้เห็นความแตกต่างของการเพิ่มขึ้นของจำนวนผู้ติดเชื้อในกลุ่มประเทศที่ใส่หน้ากากอนามัย และกลุ่มประเทศที่ไม่ใส่หน้ากากอนามัยได้ชัดเจน

“ หน่วยงานข้อมูลมีความสำคัญมาก แต่ถูกมองข้ามไปในหลาย ๆ ครั้ง ”

4. รู้จักมองข้อมูลให้รอบด้าน – ต่อจากบทเรียนข้อก่อนหน้า กลุ่มประเทศที่ไม่ใส่หน้ากากอนามัยเหล่านั้นเขาไม่เห็นกราฟนี้หรือ เมื่อเห็นแล้วทำไมจึงยังไม่ใส่หน้ากากอีกล่ะ จากมุมมองของนักวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลต้องมองให้รอบด้าน อาจมีปัจจัยอื่นที่ส่งผลต่อการติดเชื้อมากกว่าแค่ใส่หรือไม่ใส่หน้ากากอนามัย เช่น ลักษณะอาหารการกิน ลักษณะนิสัย อายุ เพศ ซึ่งข้อมูลเหล่านี้ไม่ได้ปรากฏอยู่ในกราฟเส้นที่ถูกกล่าวถึงข้างต้นนี้เลย ดังนั้นผู้วิเคราะห์ข้อมูลจะด่วนสรุปไม่ได้ ยังต้องวิเคราะห์หาความสัมพันธ์หรืออิทธิพลของตัวแปรอื่น ๆ ให้รอบด้าน

5. ความสำคัญของหน่วยของข้อมูลกับการเปรียบเทียบข้อมูล - หน่วยของข้อมูลนั้นมีความสำคัญมาก แต่ถูกมองข้ามไปในหลาย ๆ ครั้ง เช่น การนับจำนวนการตรวจผู้ติดเชื้อ COVID-19 ในบางประเทศ นับโดยใช้หน่วยจำนวนคน ในบางประเทศนับโดยใช้หน่วยจำนวนครั้งการตรวจ ซึ่งทั้งสองหน่วยนี้ไม่สามารถนำมาเทียบกันโดยตรงได้ ทั้งนี้เนื่องจากโดยลักษณะของโรคแล้ว อาจไม่ได้ตรวจพบในครั้งแรก ผู้ป่วยหลายรายตรวจพบในครั้งที่ 2 หรือ 3 ดังนั้นจำนวนครั้งการตรวจจึงเป็นคนละหน่วยกับจำนวนคนที่ถูกตรวจ และไม่สามารถนำมาเทียบกันโดยตรงได้ว่าประเทศ ก. ทำการตรวจมากกว่าประเทศ ข.

“ นักวิเคราะห์ข้อมูลที่จำเป็นต้องตรวจสอบแหล่งที่มา และความเป็นปัจจุบันของข้อมูล ”

6. ความสำคัญของแหล่งที่มาและวันที่ของข้อมูล - จากตัวอย่างข้างต้น หากผู้เขียนไม่ได้ระบุว่า ผู้ติดเชื้อจำนวน 2 พันกว่ารายนั้นเป็นผู้ติดเชื้อนับถึงวันที่เท่าไร? และไม่บอกแหล่งที่มาใด ๆ เลย นั้นจะทำให้ความน่าเชื่อถือของข้อมูลนั้นด้อยลงไปมาก รวมถึงการจะนำไปใช้ต่อนั้นมีข้อจำกัด ผู้อ่านจะเกิดคำถามทันทีว่า ข้อมูลนี้เป็นข้อมูลตั้งแต่เมื่อไหร่ ยังสะท้อนภาพความเป็นจริงอยู่ไหม มีความเป็นปัจจุบันหรือไม่ และ Fake News ต่าง ๆ ที่เกิดขึ้นมากในช่วงนี้ส่วนหนึ่งก็เกิดจากการที่ผู้รับสารนั้นไม่ตรวจสอบแหล่งที่มาของข้อมูลและแชร์ต่อกันไปเรื่อย ๆ นั่นเอง นักวิเคราะห์ข้อมูลที่จำเป็นต้องรู้จักตรวจสอบแหล่งที่มาและความเป็นปัจจุบันของข้อมูลด้วย

7. ความแตกต่างระหว่าง Data Science กับ Business Analytics - สำหรับสายงานทางด้าน IT คงปฏิเสธไม่ได้ว่า Data Science และ Business Analytics เป็นหัวข้อและเป็นสายงานที่ถูกกล่าวถึงและได้รับความนิยมเป็นอย่างมาก สองคำนี้มักถูกใช้เรียกแทนกัน หรือเหมารวมว่าเป็นเรื่องของการวิเคราะห์เหมือนกัน อย่างไรก็ตาม หัวข้อ 2 หัวข้อนี้ต่างกันอยู่ค่อนข้างมากในหลักการ ผู้เขียนขอยกกรณีของ COVID-19 มาเปรียบเทียบให้เข้าใจได้ง่าย ๆ ดังนี้

Data Science กับกรณี COVID-19 อาจเป็นการนำข้อมูลลักษณะการเกิดอาการ ลักษณะ DNA พฤติกรรมการใช้ชีวิตอื่น ๆ การให้ยา และผลจากการให้ยาของกรณีทั้งหมดกว่า 3 แสนรายมาวิเคราะห์เพื่อให้เห็น Hidden Pattern หรือปัจจัยต่าง ๆ ที่จะส่งผลต่อการติดเชื้อ ความสามารถในการต่อต้านไวรัส หรือแบบจำลอง DNA ที่ทนทานต่อไวรัส เป็นต้น พุดง่าย ๆ คือความพยายามที่จะใช้คณิตศาสตร์ขั้นสูงในการหาแบบจำลองที่จะอธิบายสถานการณ์หรือสิ่งที่เราสนใจ ทั้งนี้ เนื่องจากการวิเคราะห์ข้อมูลทางสถิติทั่วไป อาจไม่สามารถช่วยให้เข้าใจสิ่งที่ซ่อนอยู่ในข้อมูลจำนวนมากมหาศาลที่ค่อนข้างจะไร้โครงสร้าง ในขณะที่ Business Analytics จะมุ่งความสนใจไปที่ข้อมูลและมุมมองที่เกี่ยวข้องกับธุรกิจ และข้อมูลที่ใช้สำหรับการวิเคราะห์นั้นค่อนข้างจะมีโครงสร้างมากกว่า เช่น ผู้ประกอบการ e-commerce ต่าง ๆ เมื่อนำข้อมูลการเข้าเยี่ยมชมเว็บไซต์และ Mobile App ของลูกค้ามาวิเคราะห์ อาจจะเริ่มมองเห็นแพทเทิร์นบางอย่างที่เปลี่ยนไปในพฤติกรรม เช่น การตัดสินใจซื้อที่ง่ายขึ้น Price Sensitive ลดลงในสินค้ากลุ่มใด ยอมจ่ายค่าขนส่งง่ายขึ้น ซึ่งการวิเคราะห์ข้อมูลเหล่านี้เพื่อไฮไลท์พฤติกรรมที่สำคัญของลูกค้าที่เปลี่ยนแปลงไปก็จะทำให้ผู้ประกอบการสามารถที่จะตัดสินใจและตอบสนองต่อพฤติกรรมและความต้องการที่เปลี่ยนไปของผู้บริโภคได้ดียิ่งขึ้น

8. Data Governance (ธรรมาภิบาลข้อมูล) - เมื่อมีการเก็บข้อมูลจำนวนมาก โดยเฉพาะเป็นข้อมูลที่มีความอ่อนไหวและมีหลายฝ่ายต้องการจะนำข้อมูลไปใช้ การจัดทำ Data Governance คือ การกำหนดว่าข้อมูลใดบ้างเปิดเผยได้ และใครจะสามารถนำข้อมูลอะไรไปใช้ได้บ้าง เนื่องจากข้อมูลบางอย่างเป็นข้อมูลระดับส่วนบุคคลที่มีความอ่อนไหวมากและไม่ควรให้บุคคลที่ไม่เกี่ยวข้องเข้าถึงได้ ตัวอย่างเช่น ถึงแม้กรมควบคุมโรคติดต่อจะมีข้อมูลว่าผู้ติดเชื้อแต่ละรายชื่อจริง นามสกุล ที่อยู่ปัจจุบันอยู่ที่ใด แต่ข้อมูลเหล่านี้ไม่ควรถูกเข้าถึงได้ จึงต้องมีการกำหนดให้ชัดเจนว่าผู้ที่สนใจจะนำข้อมูลไปใช้วิเคราะห์จะสามารถเข้าถึงข้อมูลใดได้บ้าง ผู้เกี่ยวข้อง เช่น หน่วยงานภาครัฐอื่น ๆ โรงพยาบาลต่าง ๆ จะเข้าถึงข้อมูลใดได้บ้าง (ซึ่งอาจจะละเอียดกว่าข้อมูลที่เปิดเผยสำหรับนักวิเคราะห์หรือบุคคลทั่วไป)

หมายเหตุ: การทำ Data Governance นั้นครอบคลุมประเด็นอื่น ๆ อีกมากนอกจากเรื่องของการอนุญาตตามสิทธิ์ แต่ในที่นี้ยกเพียงตัวอย่างสั้น ๆ 1 ตัวอย่าง ที่จะเห็นภาพได้ง่ายและชัดเจนที่สุด

9. ข่าวดีสำหรับผู้ที่ต้องการพัฒนาทักษะไปสู่การเป็นนักวิเคราะห์ข้อมูล – หากท่านอยากฝึกฝนและพัฒนาทักษะการเป็นนักวิเคราะห์ข้อมูล ในปัจจุบัน ข้อมูลที่เปิดให้ใช้ฟรีมีอยู่มากเต็มไปหมด ตัวอย่างเช่น ข้อมูลที่กรมควบคุมโรค ได้เก็บรวบรวมเกี่ยวกับจำนวนผู้ติดเชื้อ ก็มีเปิดให้นักวิเคราะห์และนักพัฒนาแอปพลิเคชันต่าง ๆ สามารถไปดึงมาใช้ได้ฟรีผ่าน API ที่กรมควบคุมโรคพัฒนาขึ้น ซึ่งสามารถเข้าถึงได้จากลิงค์นี้ <http://covid19.ddc.moph.go.th/th/api> นอกจากข้อมูลแล้ว เครื่องมือดี ๆ สำหรับการวิเคราะห์และ Visualize นั้นก็มีอยู่มากเช่นกัน ไม่ว่าจะเป็น Microsoft Excel ที่ใช้กันอยู่ทั่วไปในองค์กรธุรกิจ หรือ Tool ที่ Microsoft ตั้งใจพัฒนาเพิ่มเติมขึ้นมาเพื่อช่วงชิงตลาดด้านการวิเคราะห์ข้อมูลอย่าง Microsoft Power Bi ซึ่งมีฟรีเวอร์ชันให้ใช้งานได้ด้วย หรือ Tool ค่ายทางเลือกที่กำลังเป็นที่นิยมและใช้กันอย่างแพร่หลายอย่าง Tableau และ Google Data Studio

สำหรับผู้ที่สนใจพัฒนาทักษะในการจัดการและวิเคราะห์ข้อมูล รวมถึงเพิ่มพูนมุมมองด้านธุรกิจ หลักสูตรปริญญาโท Management Information Systems คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์ เปิดรับสมัครถึงวันที่ 31 พฤษภาคม 2563 สนใจดูรายละเอียดการรับสมัครเพิ่มเติมได้ที่ www.mis.tbs.tu.ac.th