



Data Science คือการเอาข้อมูลที่มีเยอะแยะมากมายมหาศาลที่กระจัดกระจายมารวมกันแล้วจัดการให้มีค่า สมมติว่าข้อมูลเปล่าๆ มานั่นก็เหมือนน้ำมันดิบ ถ้าไม่ผ่านการกลั่นอะไรเลยก็ใช้ไม่ได้ แต่พอเรามี Data Science ก็เหมือนการกลั่นออกมาเป็นน้ำมันที่สามารถเอาไปใช้ในชีวิตประจำวันได้ ขั้นตอนของ Data Science เริ่มตั้งแต่เก็บข้อมูล → คลีนข้อมูล → วิเคราะห์หาความสัมพันธ์ → ทำโมเดลทำนายอนาคต → เอาผลลัพธ์ไปทำกราฟหรือดาชบอร์ดเพื่อให้คนอื่นเข้าใจได้ง่ายขึ้น

นอกจาก Data Science ยังมี Data Engineer ที่คอยสร้างระบบให้ข้อมูลไหลเข้ามาได้แบบเป็นระบบระเบียบ Data Analyst ที่ช่วยดูแลเบื้องต้น แล้วก็ออกแบบโมเดลให้เอาไปใช้จริง จังมีสกิลทั้งเขียนโปรแกรมเป็น (Python, R, SQL) เข้าใจสถิติ หลักการคณิตศาสตร์ ทำ Machine Learning ได้ และต้องรู้เรื่องธุรกิจหรือปัญหาที่จะเอาข้อมูลไปแก้

Bias คือความเอนเอียงของข้อมูล เช่น เก็บข้อมูลไม่ครอบคลุม เลือกใช้เฉพาะบางอย่าง ซึ่งถ้าไม่ระวังก็จะกลายเป็นโมเดลที่เอาไปใช้แล้วไม่ดี ผลไม่ตรงกับข้อมูลในความเป็นจริง เลยต้องมีการตรวจสอบและแก้ไขให้ดี