**Airfoils with Data Analytics**

**By Ruangyot Nanchiang**

**Simulation with JavaFoil**

ก่อนที่เราจะนำข้อมูลของ Airfoil แต่ละอันที่เราได้ทำการเลือก มาทำการ simulation ใน JavaFoil เสียก่อน เนื่องจากว่าสามารถประมาณค่าต่างๆได้รวดเร็วกว่า Anasys ซึ่งเงื่อนไขในการทำ Simulation ใน JavaFoil มีดังต่อไปนี้

1. Re = 4000000
2. Mach Number = 0.3
3. Cd and Angle of Attack at Cl=0.5
4. Thickness at 0.75c

จากนั้นก็นำข้อมูลที่ได้จากการทำ Simulation มาเก็บไว้ในโปรแกรม Excel และ save เป็น .csv เพื่อนำไป import และวิเคราะห์ต่อโดยใช้ Python

**Data Analytics with Machine Learning**

ในขั้นตอนนี้เราจะนำข้อมูลที่ได้มาจากการทำ Simulation จาก JavaFoil มาดำเนินการร่วมกับ Machine Learning เพื่อแบ่ง Airfoils ออกเป็นกลุ่มต่างๆ โดยจะวิเคราะห์จากตัวแปรที่ได้ทำการ Simulation มาก่อนหน้า ได้แก่ Cd, Cl/Cd และ t0.75c

Model ที่ได้เลือกใช้คือ K means clustering algorithm เนื่องจาก K means clustering เป็น algorithm ที่ใช้ในการแบ่งแยกกลุ่มข้อมูลจากกลุ่มใหญ่ๆให้เป็นกลุ่มย่อย โดยจะทำการหาความสัมพันธ์จากตัวแปรในข้อมูลทั้งหมด และแบ่งกลุ่มออกมา การจะบอกได้ว่าแบ่งเป็นกี่กลุ่มจึงจะเหมาะสม เราจะพิจารณาจาก silhouette score ยิ่งเข้าใกล้ 1 มากเท่าไหร่ นั่นหมายความว่าการแบ่งกลุ่มของข้อมูลยิ่งมีประสิทธิภาพ(โดยก่อนเริ่มทำการใช้ K means clustering เราต้องมีการ transform ขนาดของข้อมูลก่อนเสมอ เพื่อให้ค่าของข้อมูลอยู่ในช่วงที่มีการกระจายอย่างเหมาะสม)

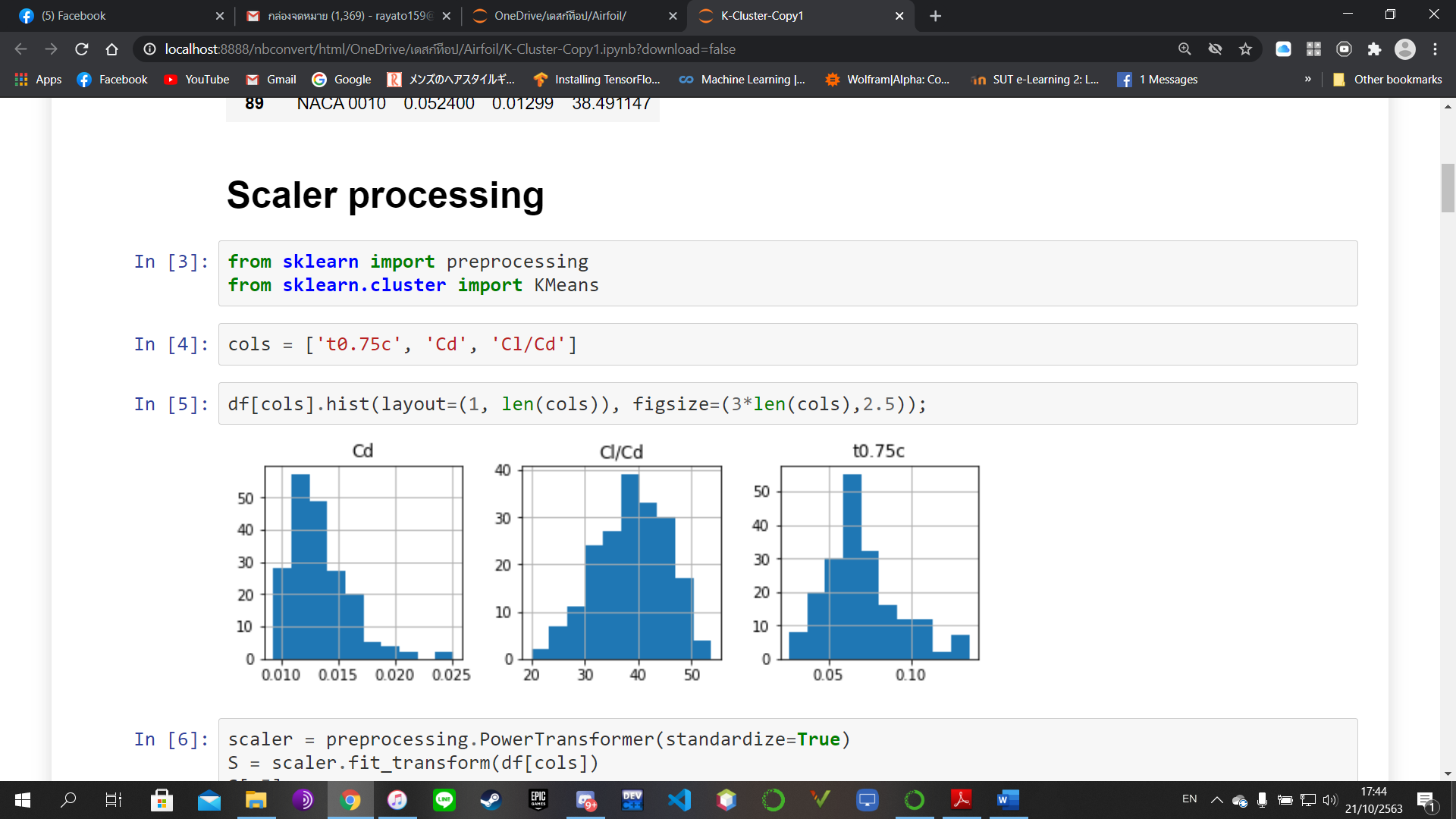


Figure 1 ก่อนทำการ transform ข้อมูล

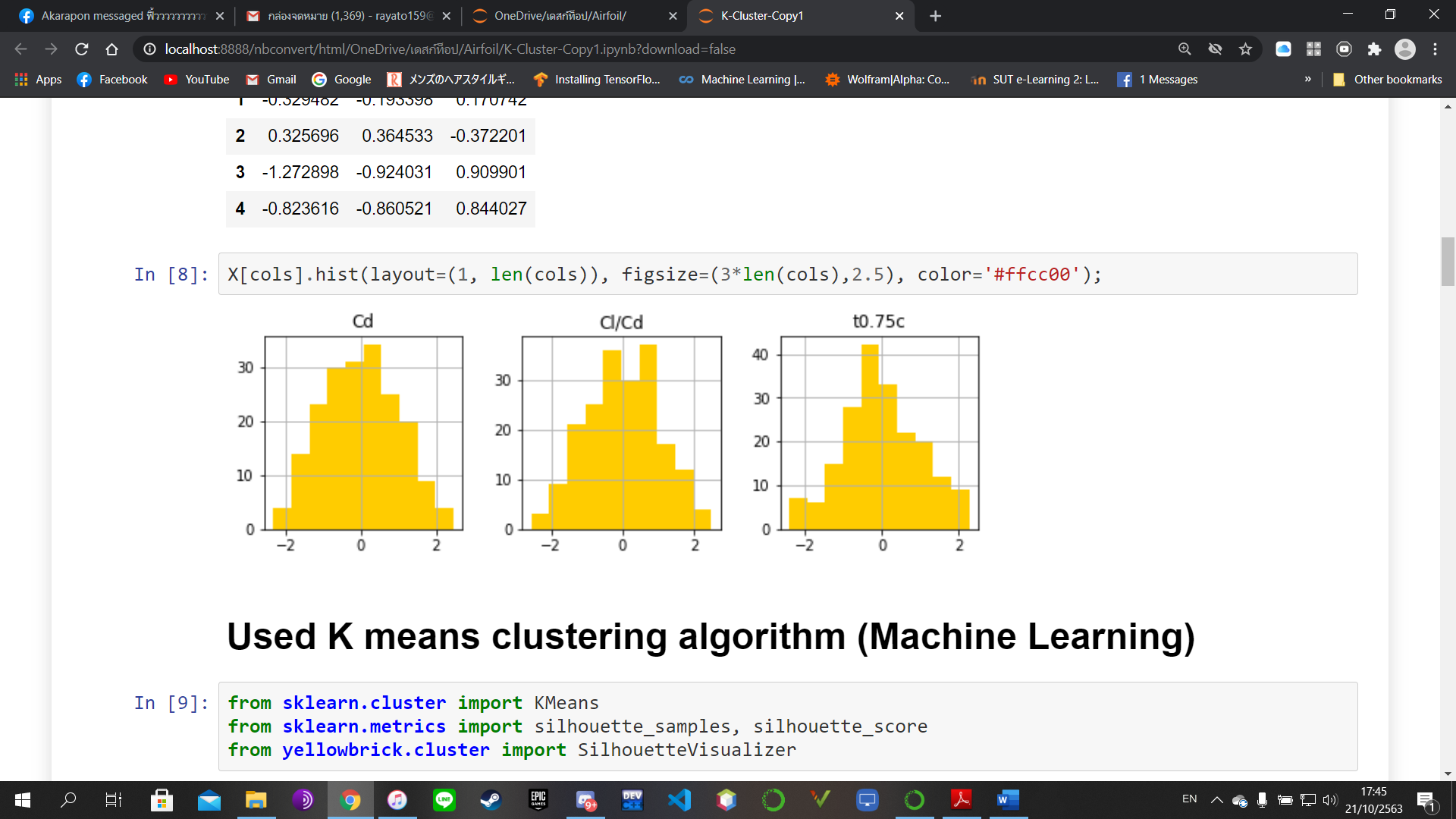
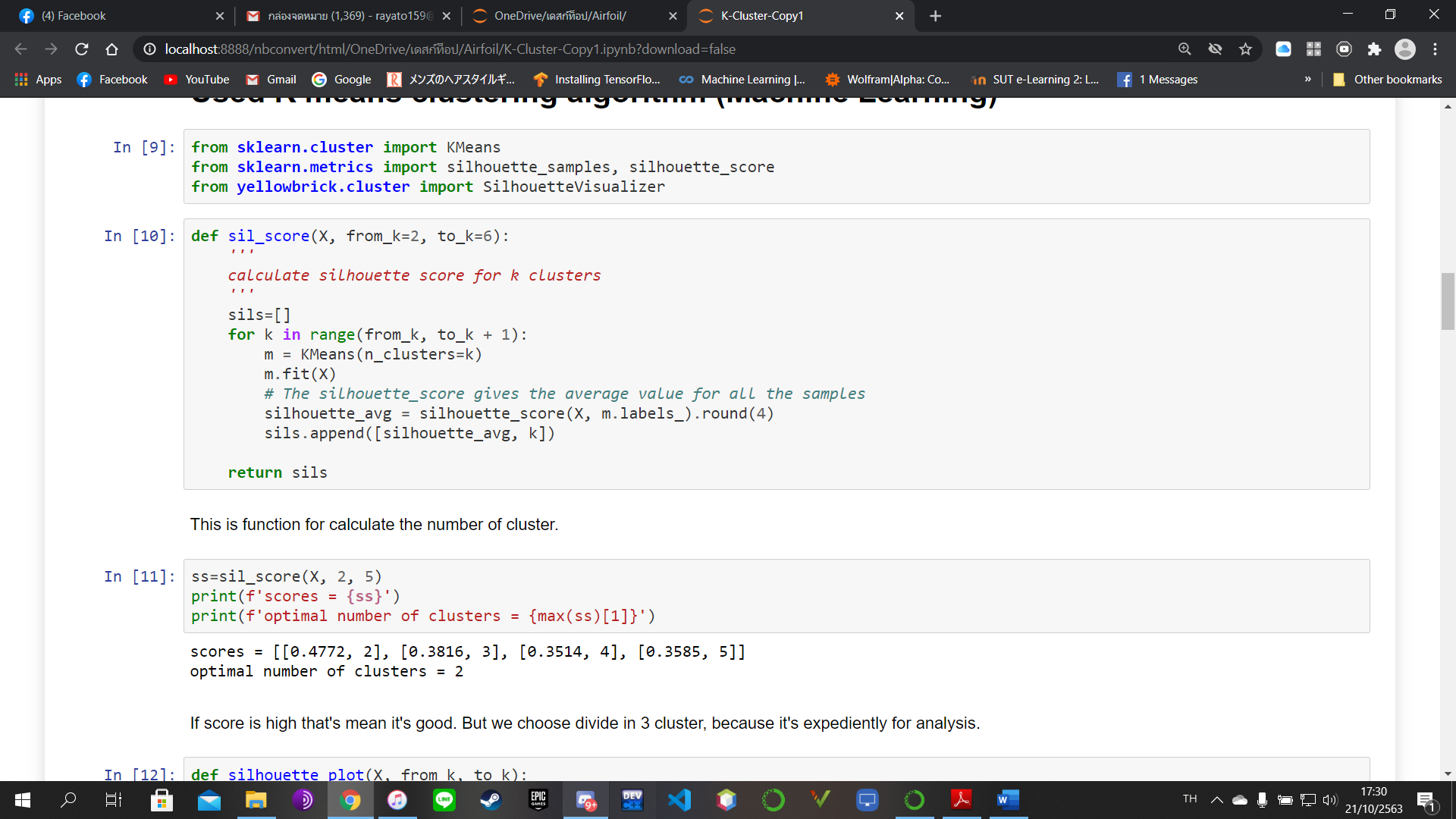


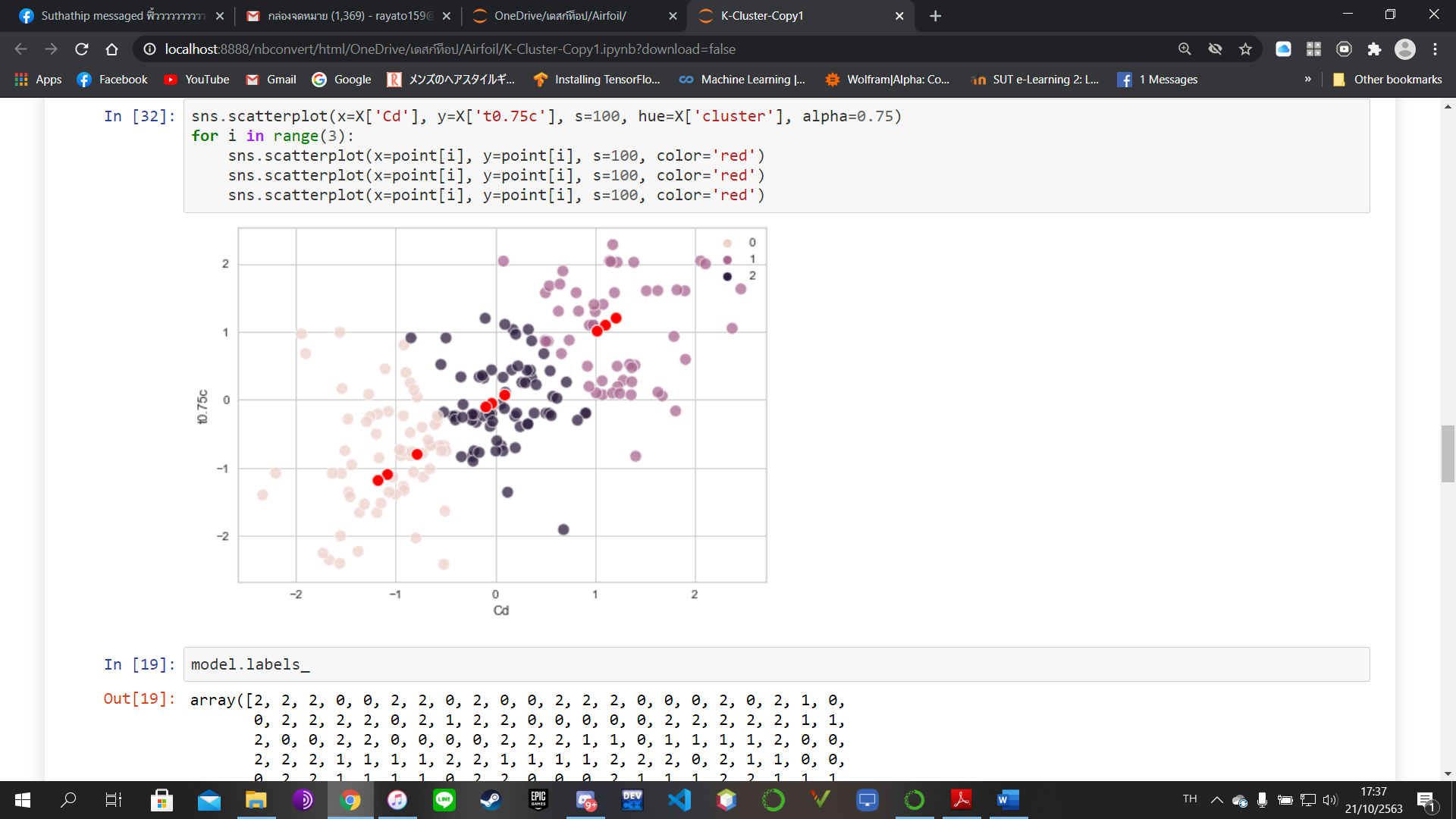
Figure 2 หลังทำการ transform ข้อมูล

จากข้อมูล Airfoils ทั้งหมดที่มี เราได้ทำการ coding ให้ Machine Learning ได้ทำการแบ่งข้อมูลออกมา โดยมี silhouette score ของการแบ่งดังนี้



จะสังเกตได้ว่า ถ้าหากเราแบ่งข้อมูลออกเป็น 2 กลุ่ม จะมี silhouette score สูงถึง 0.4772 แต่เนื่องจากว่า เราต้องการแบ่งออกเป็น 3 กลุ่ม เพื่อง่ายต่อการเปรียบเทียบระหว่าง t0.75c กับ Cl/Cd ซึ่งการแบ่งเป็น 3 กลุ่ม จะได้ silhouette score = 0.3816 รองจากการแบ่งข้อมูลออกเป็น 2 กลุ่ม

หลังจากที่เราได้ทำการแบ่งกลุ่มของข้อมูลทั้งหมดแล้ว Machine Learning จะทำการคำนวณหาจุด Mean ของตัวแปรในแต่ละจุดให้และแยกประเภทของข้อมูลให้ดังภาพ



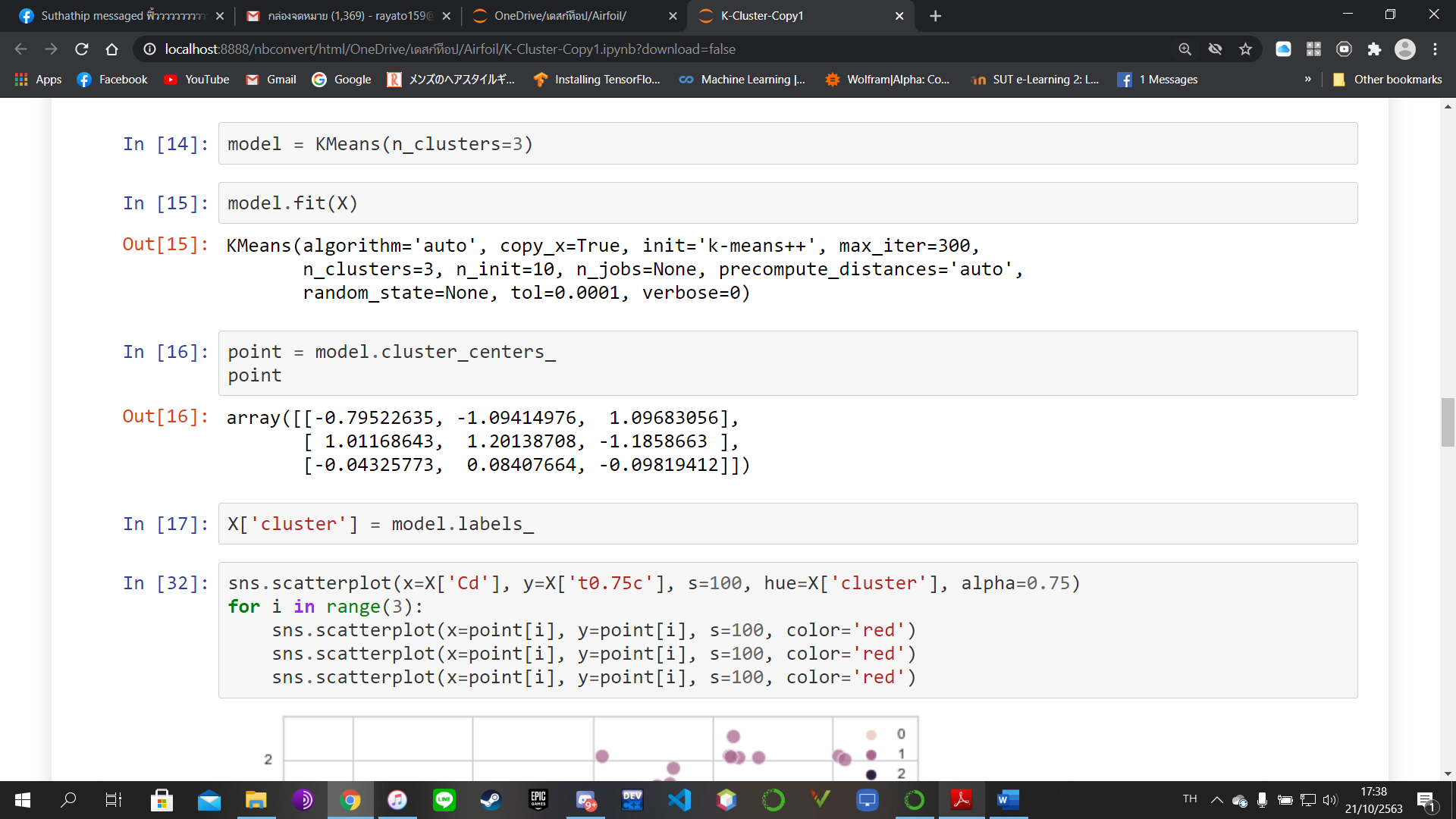


Figure 3 พิกัดของการแบ่งกลุ่มข้อมูล

เมื่อทำการแบ่งข้อมูลเรียบร้อยแล้ว ก็ทำการ save เป็นไฟล์ใหม่ เพื่อไปทำการเลือกต่อว่า Airfoils อันไหนจะมีประสิทธิภาพดีที่สุดและหนาที่สุด

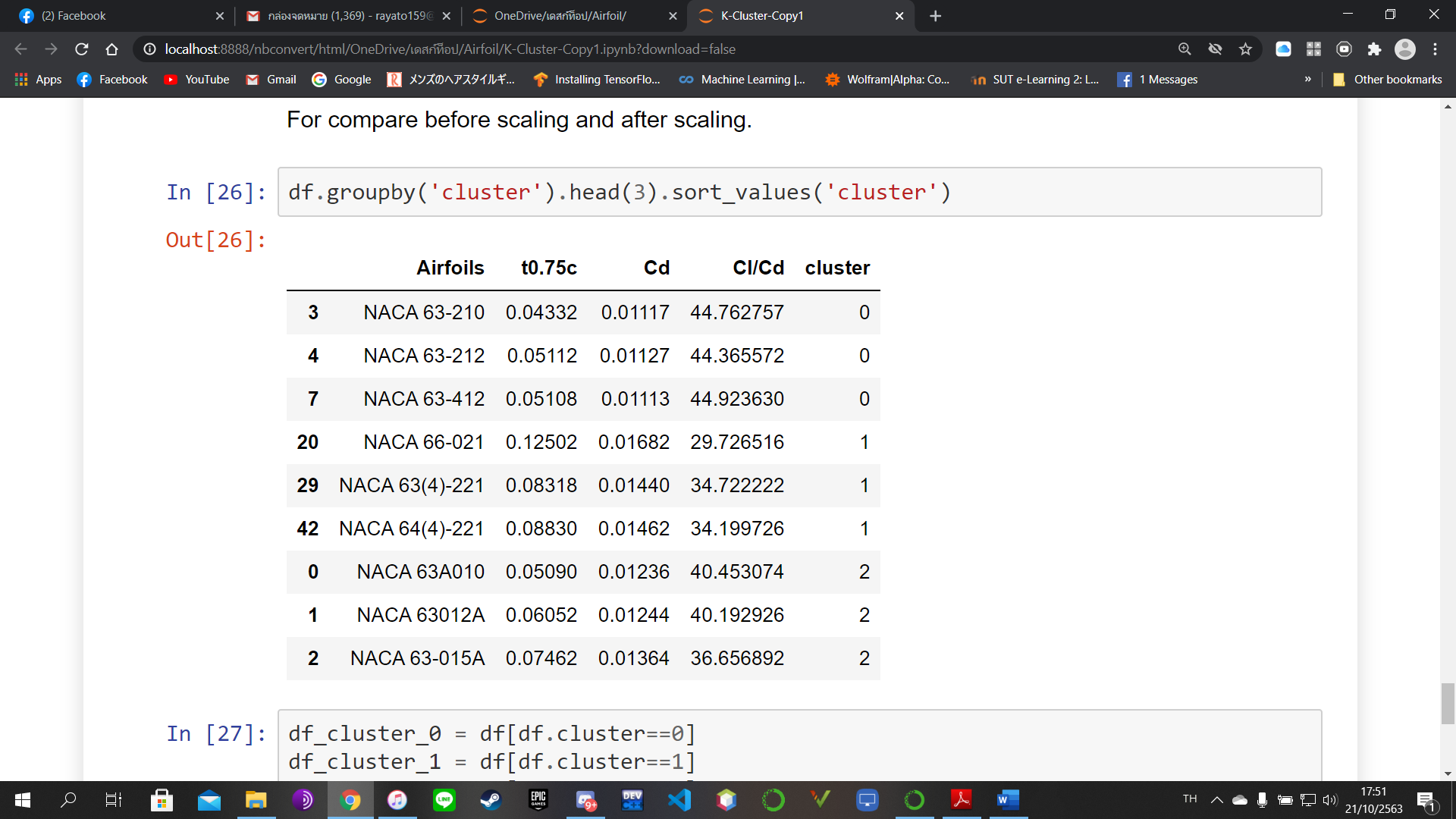
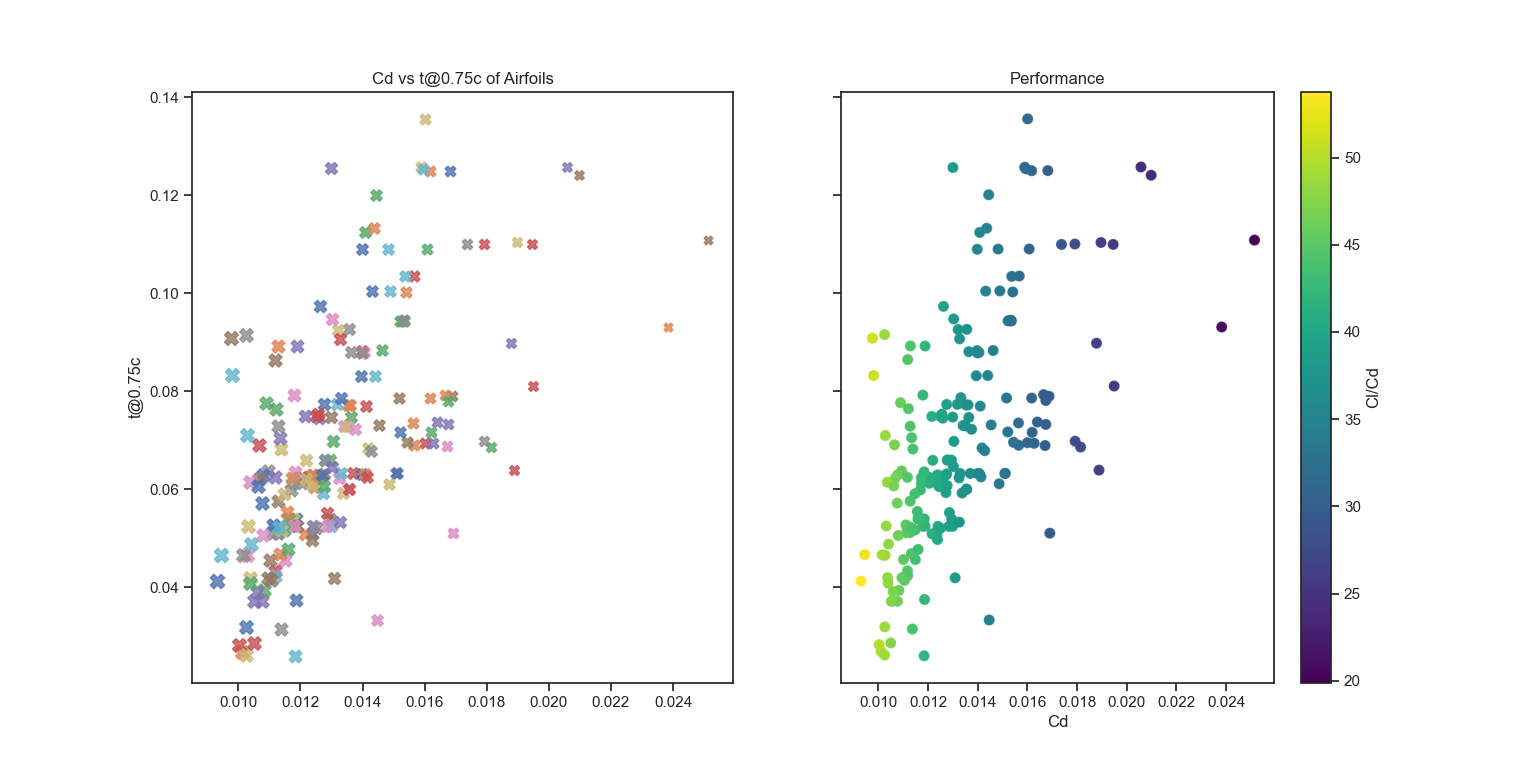


Figure 4 ตัวอย่างของข้อมูลที่ผ่านการแบ่งโดย K means clustering แล้ว

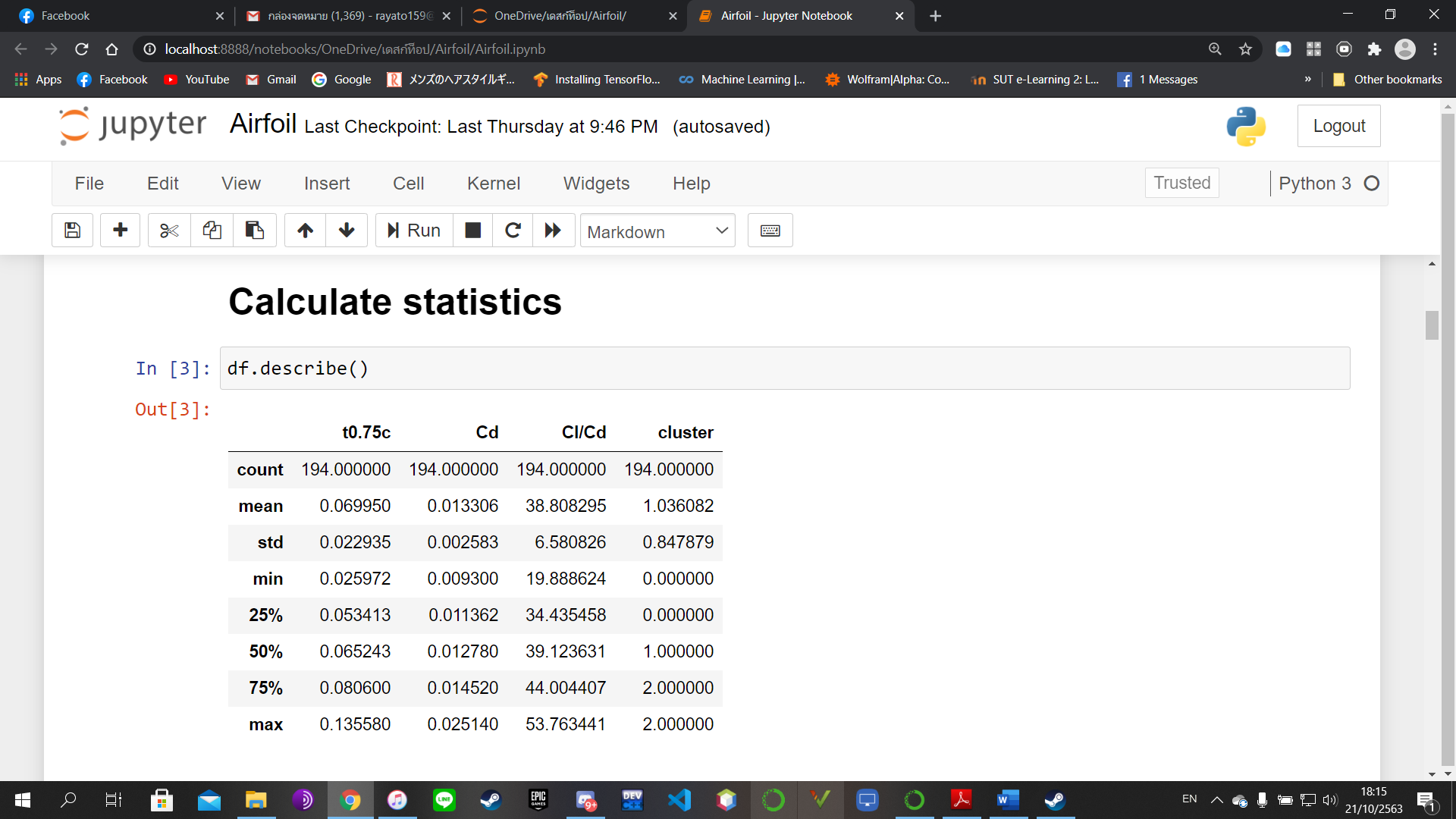
**Discussion Part**

ในขั้นตอนนี้เราจะทำการตัดสินใจเลือกว่า Airfoils อันไหนที่มีประสิทธิภาพสูงที่สุด และมีความหนาที่ 0.75c มากที่สุด จากกลุ่มข้อมูลของ Airfoils ที่แบ่งมาแล้วจากการใช้ Machine Learning

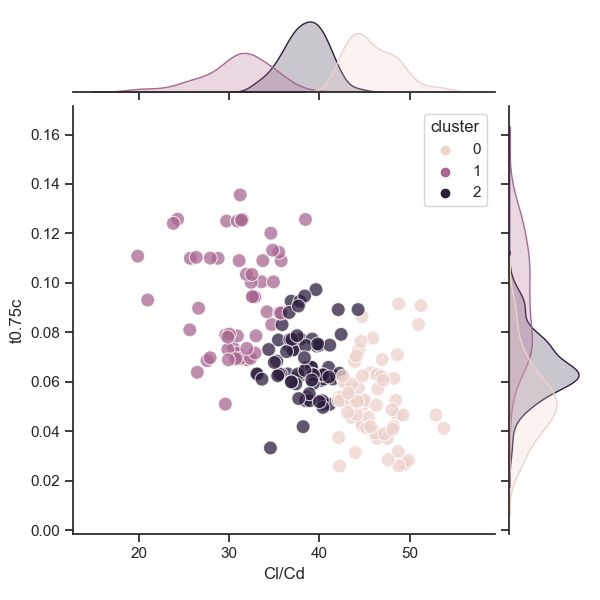
ก่อนอื่น เราต้องทำการ visualizer ข้อมูลโดยรวมก่อน เพื่อประกอบการตัดสินใจ และตรวจสอบความถูกต้องของข้อมูลเทียบจากหลักความเป็นจริงที่ควรจะเป็นได้



จากข้อมูลดังกล่าว ค่อนข้างมีความถูกต้อง เพราะว่าเป็นไปตามหลักความเป็นจริงคือ ถ้า Airfoils ยิ่งหนา Cd ก็จะยิ่งเยอะ และ Cl/Cd ก็จะน้อยลง



จากนั้น เราจะนำข้อมูลที่ได้ทำการแบ่งแล้วมาเลือก โดยกลุ่มที่เลือกจะเป็นกลุ่มที่ 0 เพราะเป็นกลุ่มที่มี Cl/Cd มากที่สุด



หลังจากที่ตัดสินใจได้แล้วว่าจะเลือกกลุ่มข้อมูลที่ 0 เราก็ต้องมาตั้งสมมุติฐานอีกทีว่ามันมีความหนาไหม ถ้าเทียบกับ ความหนาเฉลี่ยของข้อมูลกลุ่มที่ 1 ซึ่งเป็นกลุ่มที่มีความหนามากที่สุด โดยเราจะคัดเลือก Airfoils ในกลุ่มที่ 0 มา 5 ข้อมูล โดย sort จาก 0.75c และ Cl/Cd และนำไปผ่านกระบวนการ Hypothesis Test เพื่อให้แน่ใจว่า Airfoils กลุ่มนี้มีความหนาหรือไม่

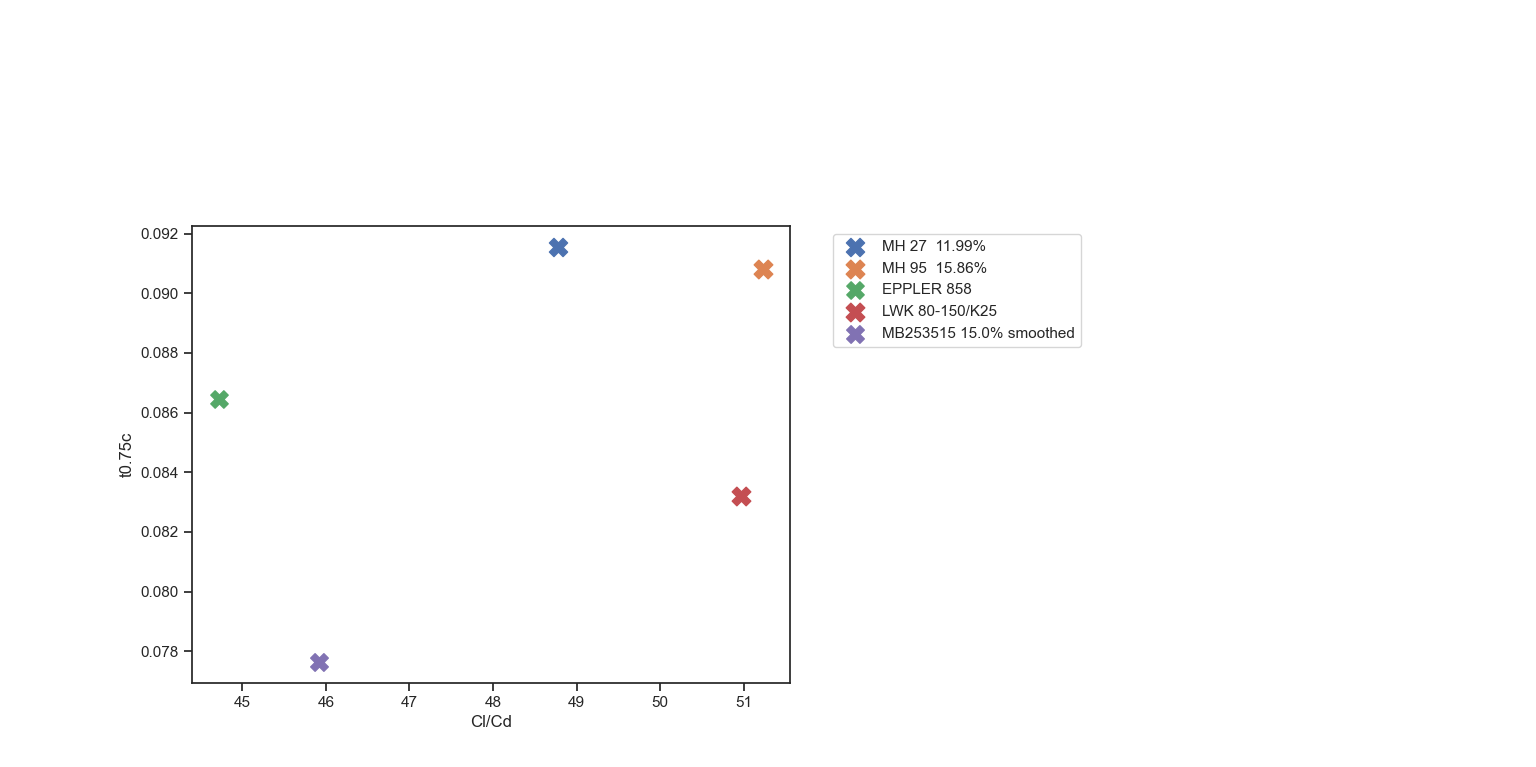


Figure 5 Airfoils กลุ่มที่มีคัดแล้ว

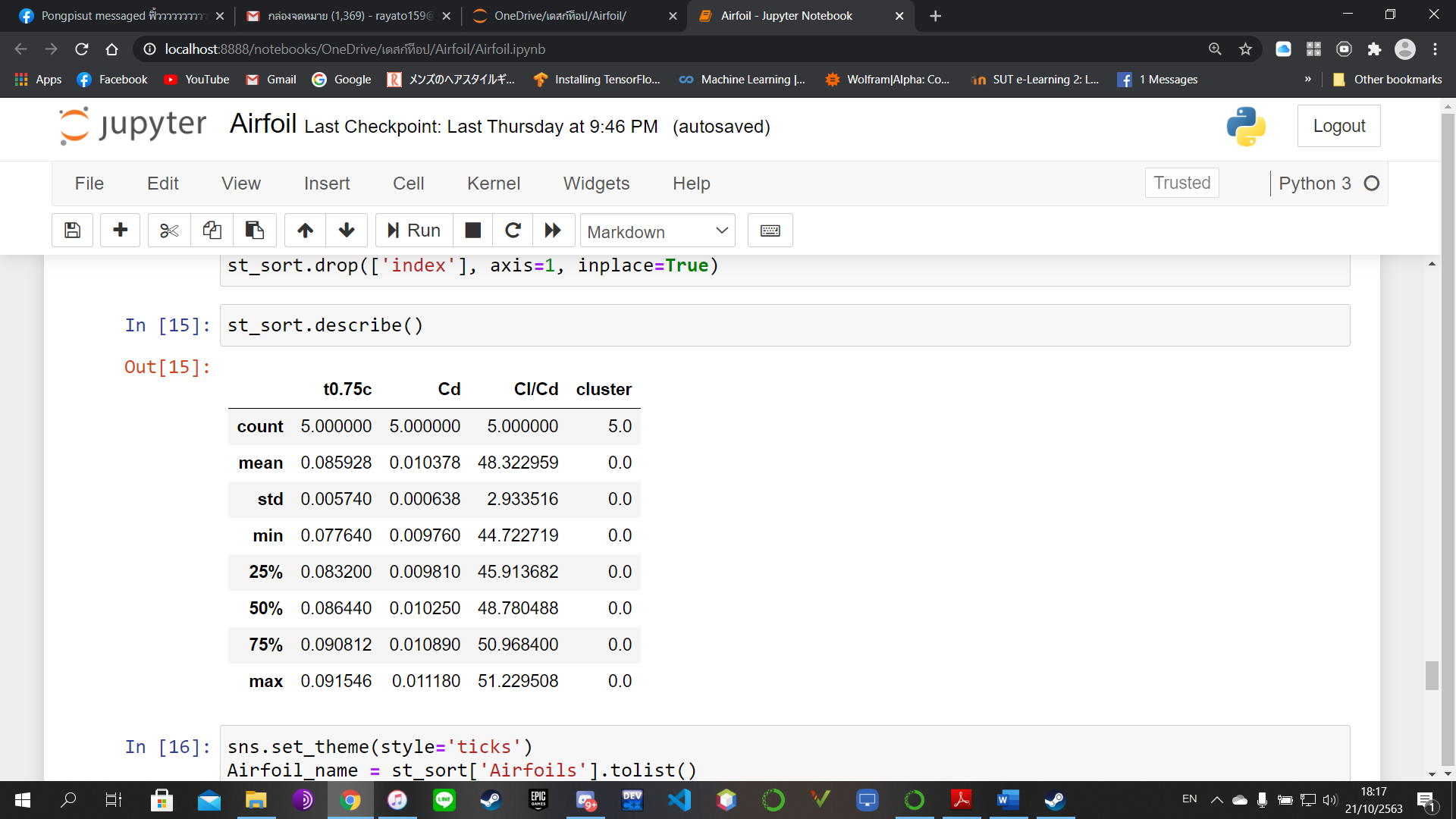
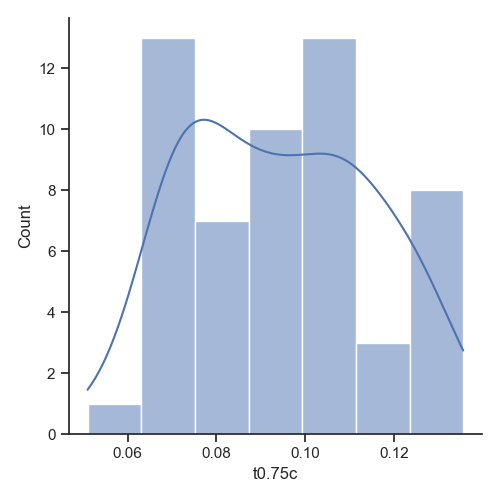


Figure 6 Statistics ของกลุ่ม Airfoil ที่ผ่านการ sort แล้ว

เนื่องจากข้อมูลกลุ่มที่ 1 มีการกระจายความหนาในช่วง mean ประมาณ 40% เราจึงจะกำหนดช่วงความเชื่อมั่นในการทำ Hypothesis test 0.4 ดังนั้น จะได้ค่า Alpha = 0.6 แบ่งออกเป็น Two way test จะได้ว่า Alpha = 0.3

\*\*\*หมายเหตุ ที่ต้องการ Test เป็น Two Ways เพราะว่า ต้องการทราบว่า Airfoils กลุ่มนี้มีความหนาเทียบเคียงกับ Airfoils กลุ่มที่มีความหนามากที่สุดได้หรือไม่



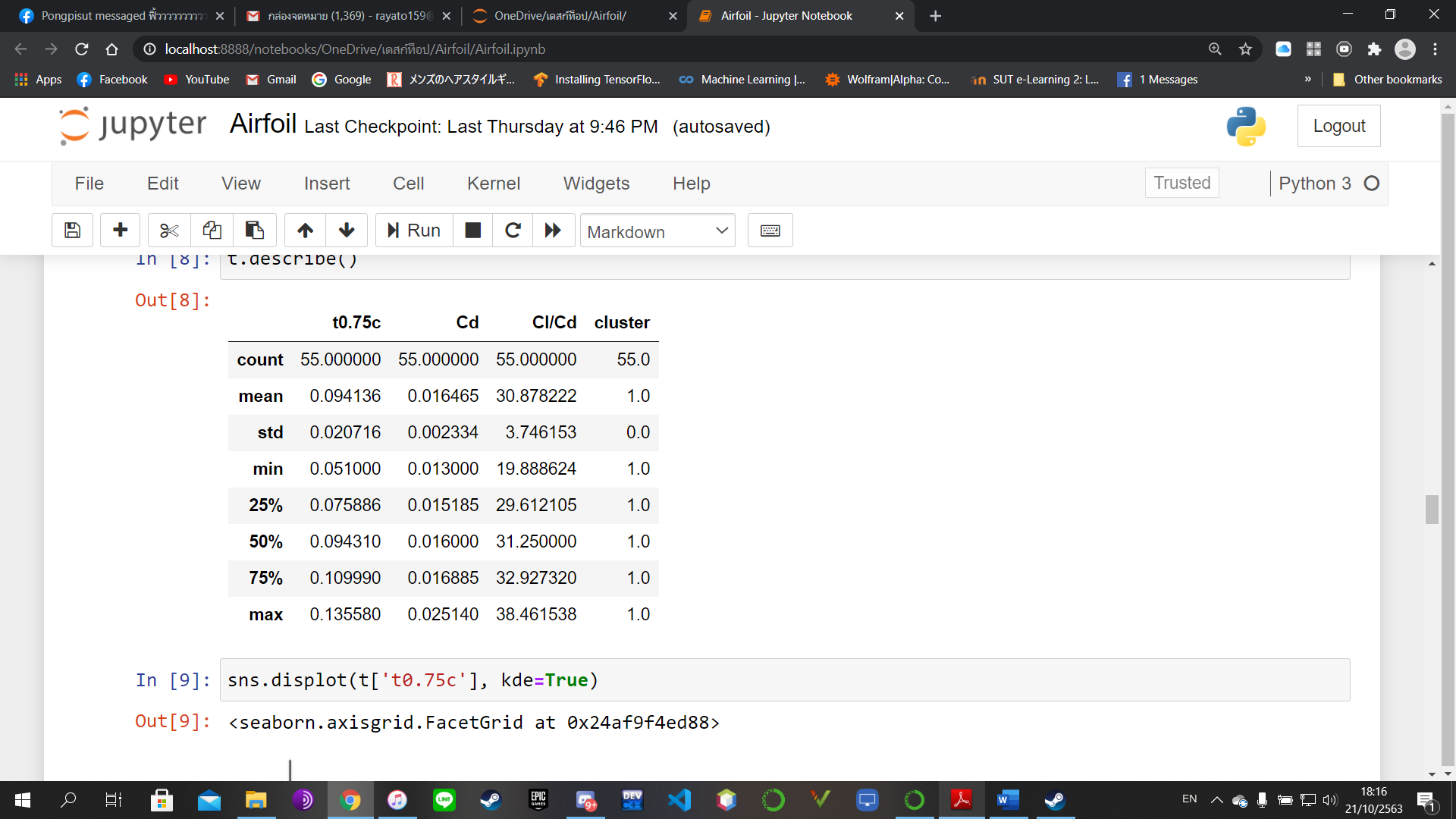
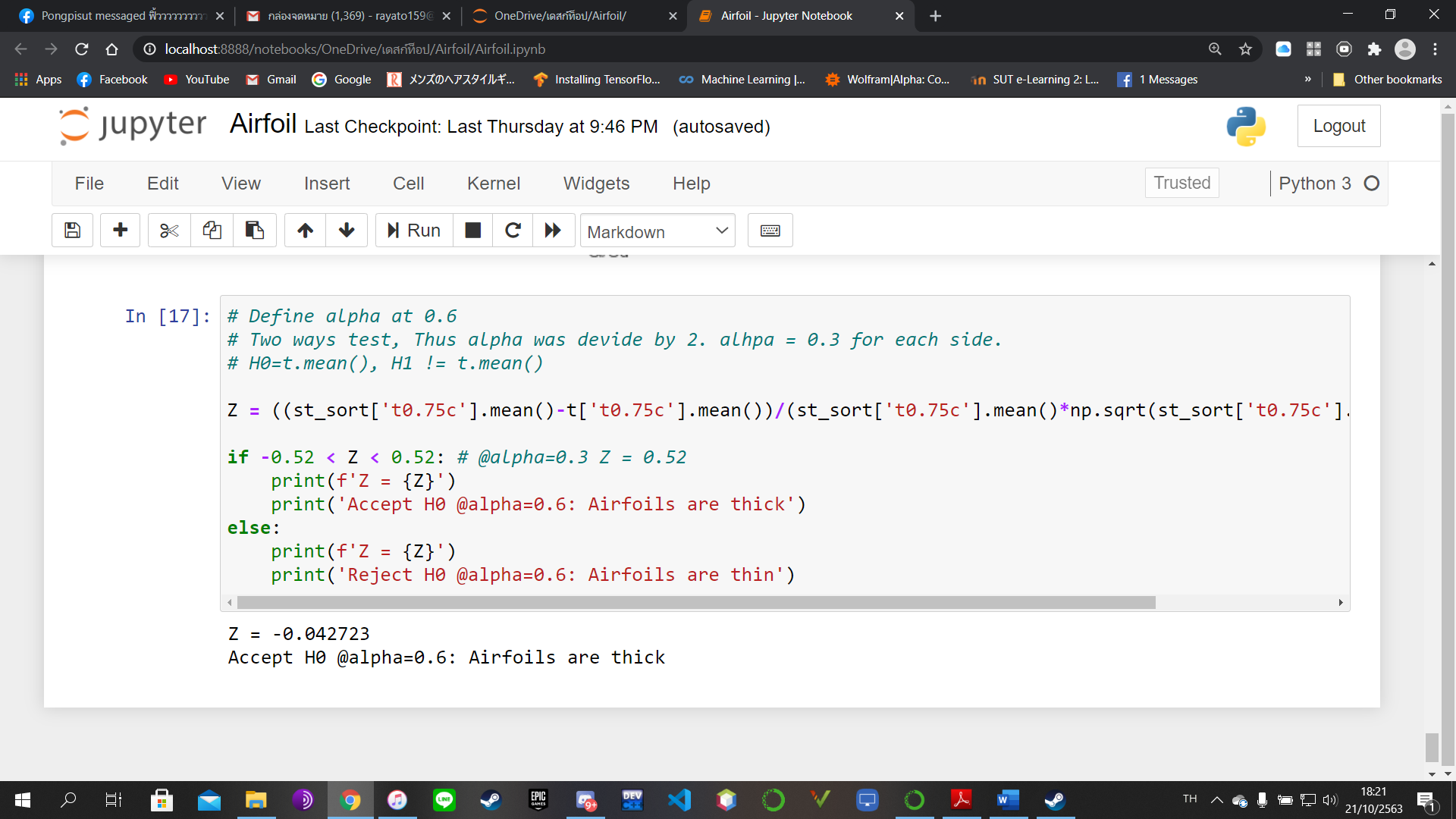


Figure 7 Statistics ของ Airfoils กลุ่มที่ 1



หลังจากที่ได้ทำการ Hypothesis Test แล้ว เราสามารถบอกได้ว่า Airfoils กลุ่มที่เลือกมานี้มีความหนาเทียบเคียงได้กับ กลุ่ม Airfoils ที่มีความหนามากที่สุด