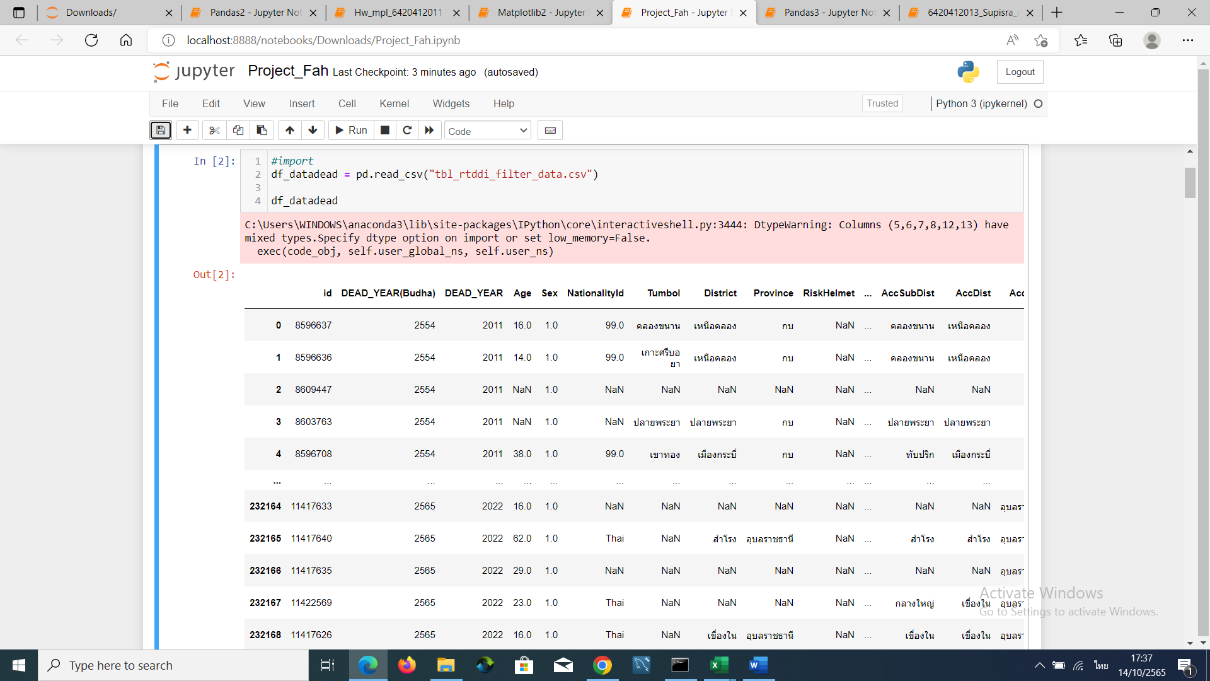
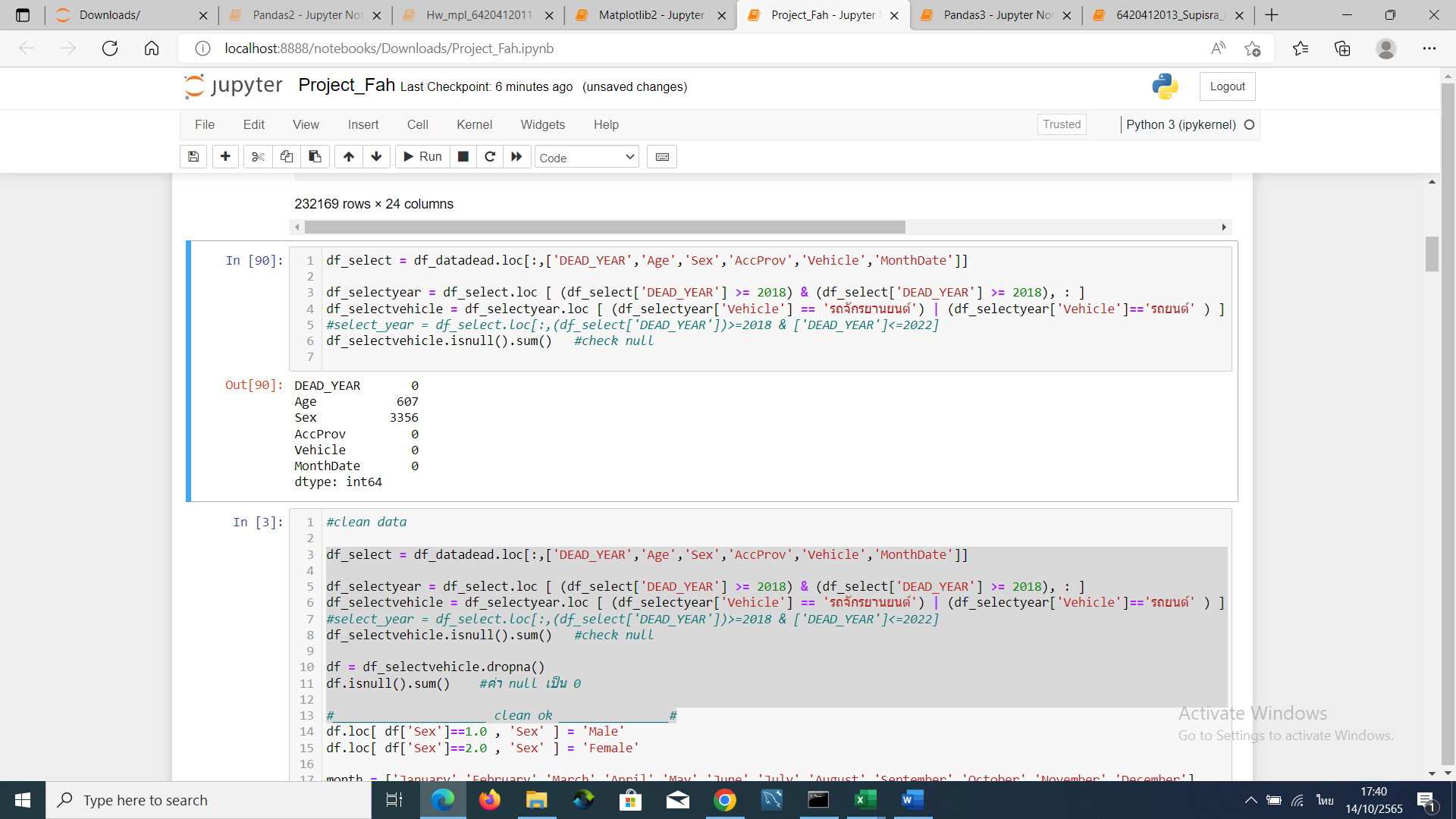
**Data Journey**

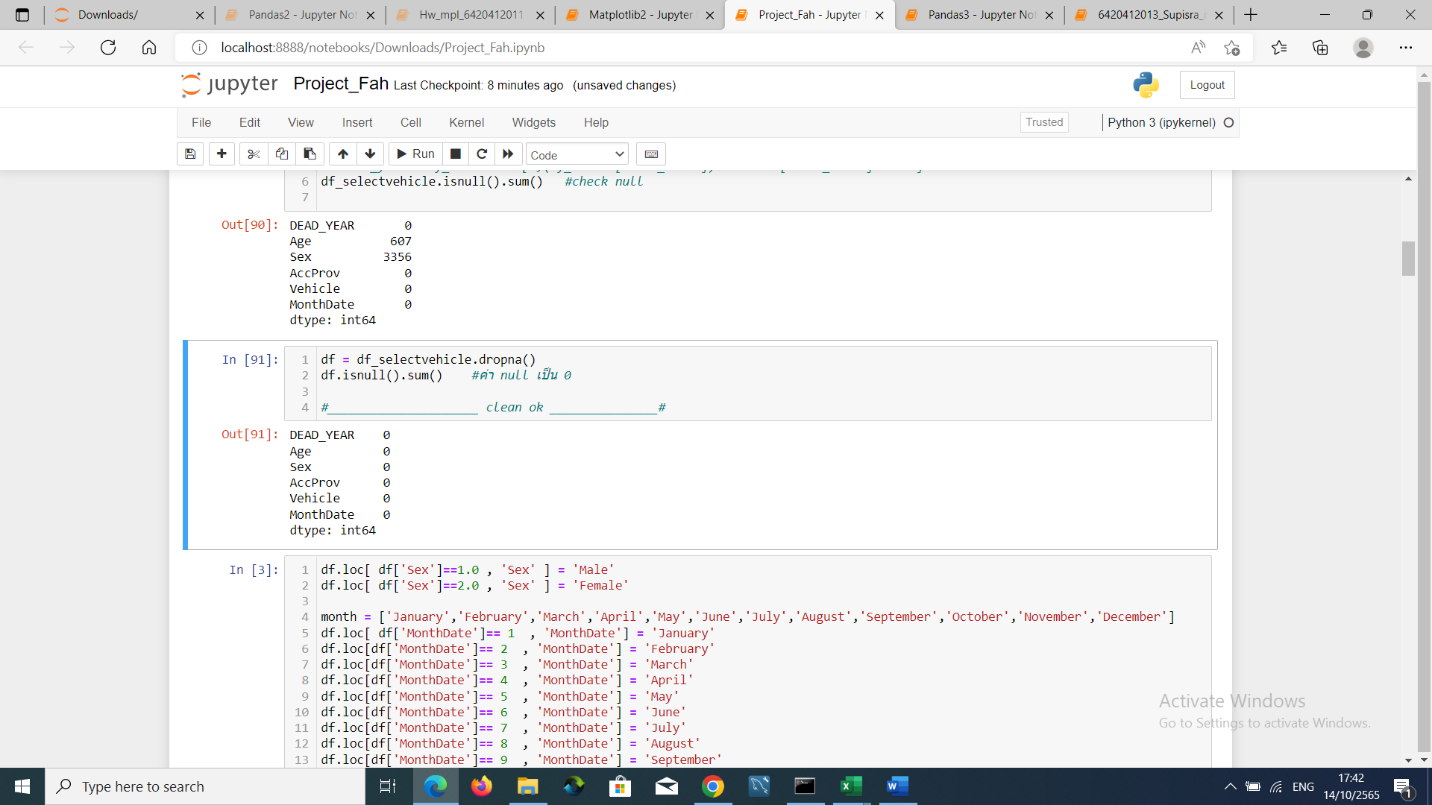
ข้อมูลชุดนี้เป็นข้อมูลผู้เสียชีวิตจากอุบัติเหตุบนท้องถนน โดยมีแหล่งข้อมูลจากสำนักงานสถิติแห่งชาติ จำนวนทั้งสิ้น 21 ตัวแปร จำนวนมากกว่า 2 แสนข้อมูล ซึ่งจะประกอบด้วยรายละเอียดของอุบัติเหตุและเสียชีวิตตั้งแต่ปีพ.ศ. 2554 ถึง 2565 (สิ้นสุดวันที่ 3 มิถุนายน) โดยตั้งสมมติฐานหรือคำถามที่ผู้จัดทำสงสัยคือ ระหว่างรถจักรยานยนต์กับรถยนต์ พาหนะใดก่อให้เกิดอุบัติหตุและเสียชีวิตมากกว่ากัน โดยผู้จัดทำคาดเดาว่าน่าจะเป็นรถจักรยานยนต์มากกว่า จึงทำการทดสอบ โดยมีขั้นตอนดังนี้ ในขั้นตอนที่ 1) นำข้อมูลมาดูภาพรวมเพื่อดูรายละเอียด พร้อมทั้งตัวแปรที่สามารถใช้งานได้ ในที่นี้สนใจตัวแปรคือ อายุ เพศ สถานที่ประสบอุบัติเหตุและเสียชีวิต ปีที่เสียชีวิต และยานพาหนะที่ประสบอุบัติเหตุ ซึ่งจากตัวแปรที่กล่าวมา มีค่า null อยู่ค่อนข้างเยอะ จึงทำการ clean data ดังภาพ

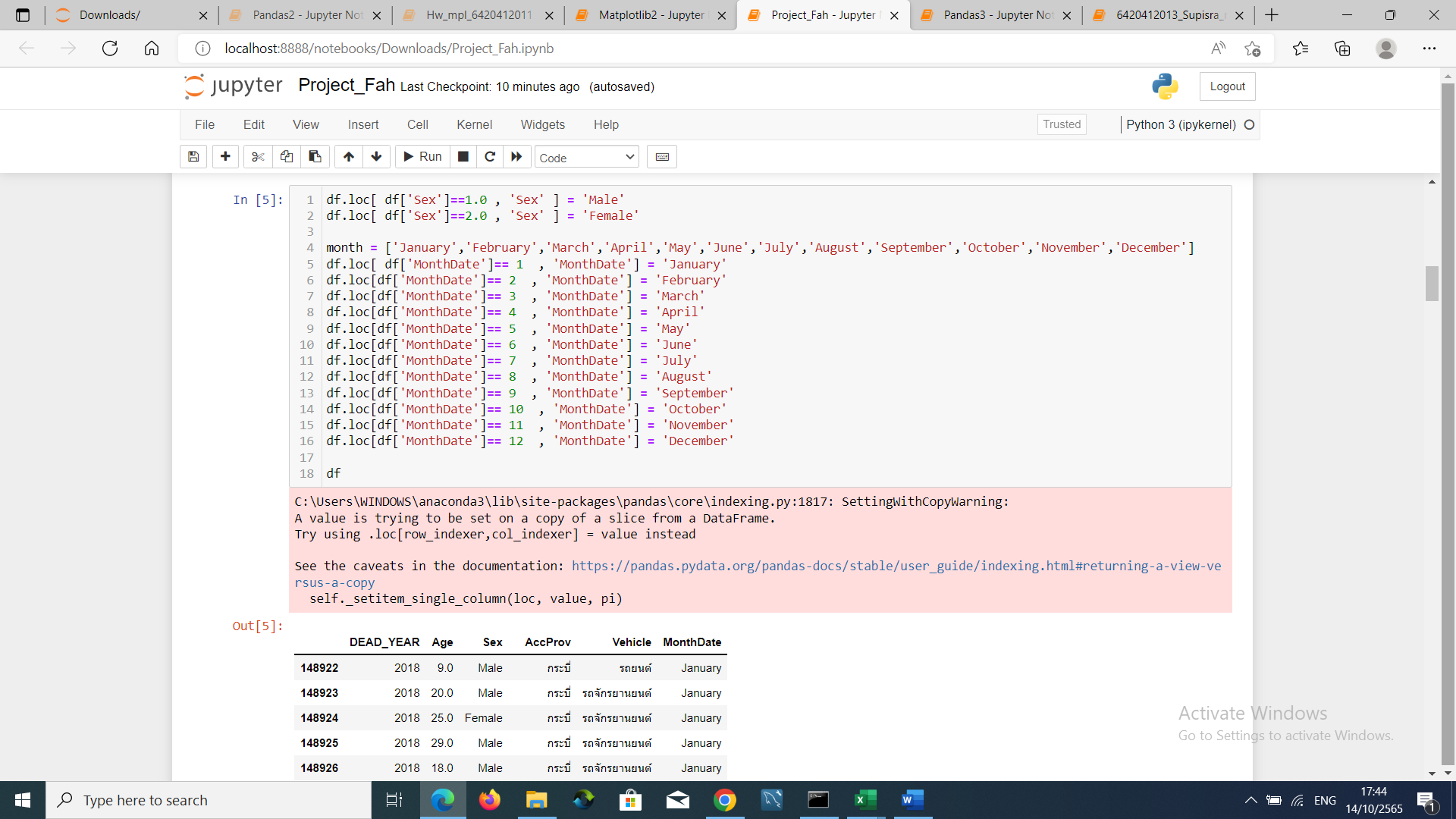


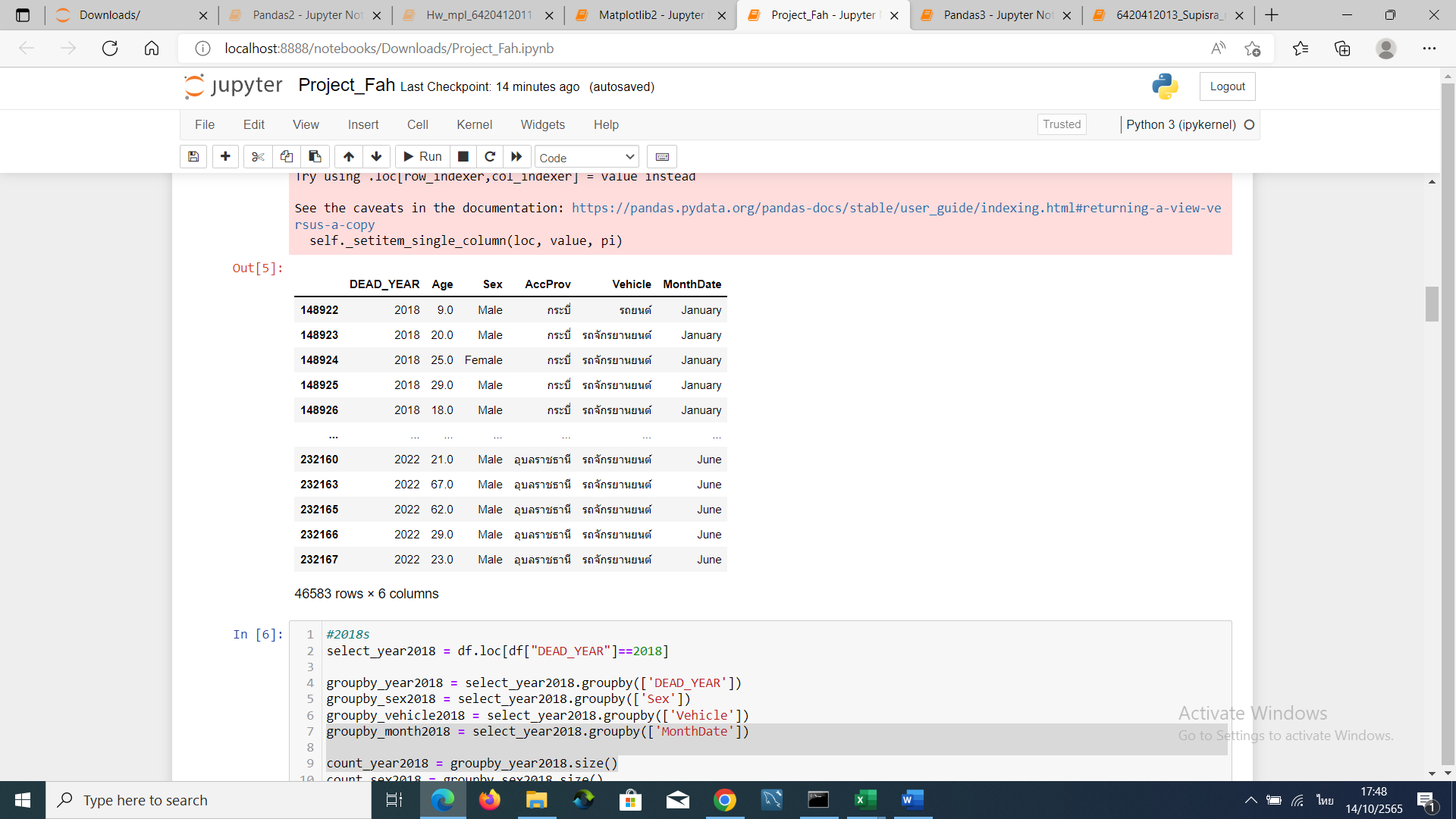
ทำการ import data เข้ามา



Check null เมื่อพบว่ามีค่า null ในตัวแปรที่จะนำมาวิเคราะห์ จึงทำการ

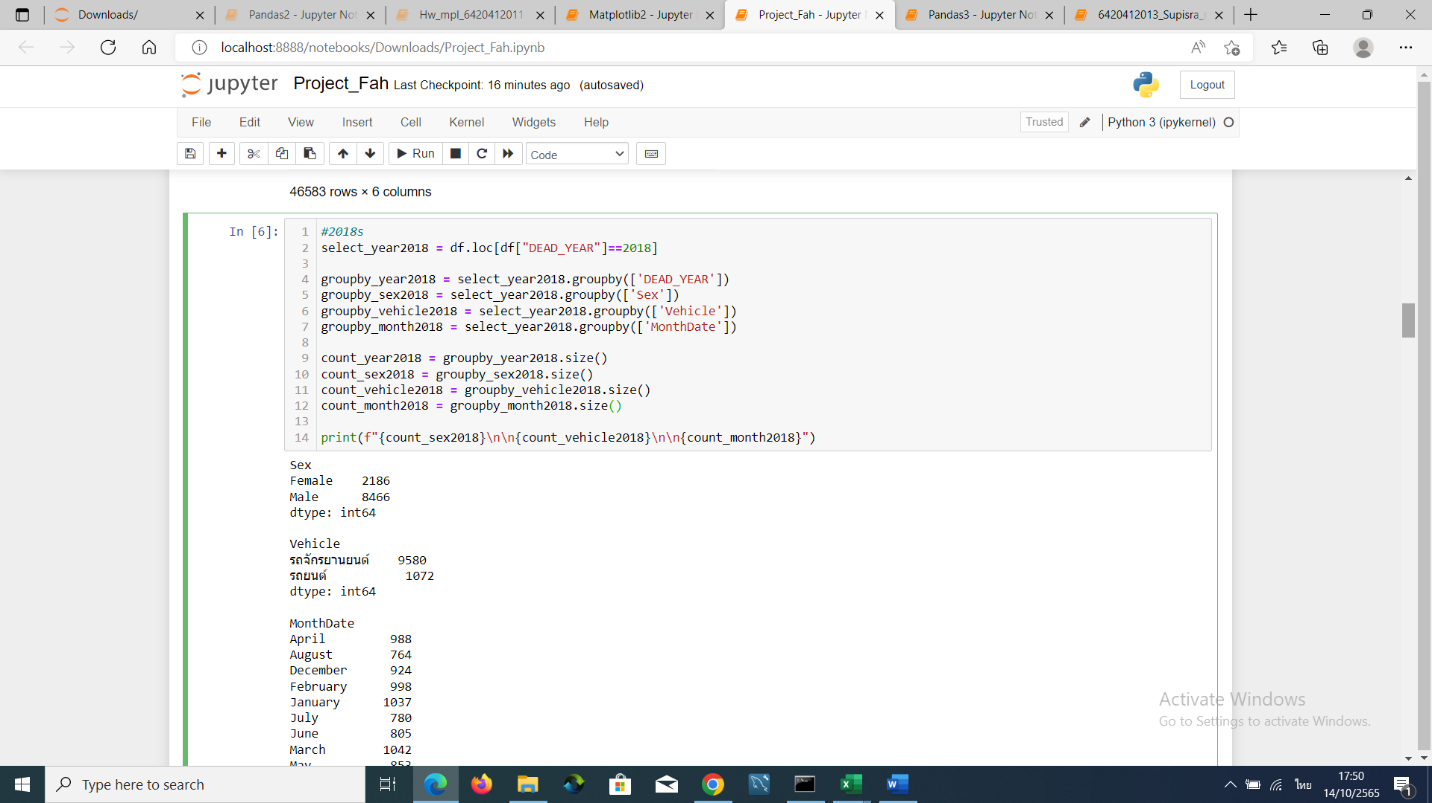


ขั้นนตอนที่ 2) เมื่อทำการ clean data ได้ตามที่ต้องการแล้ว จึงนำข้อชุดมูลที่เลือกมา แปลงค่าเป็นข้อมูลที่มีลักษณะอ่านง่ายยิ่งขึ้น โดยทำการแปลงเพศจาก ตัวเลข 1,2 เป็นเพศชาย(Male) และ เพศหญิง(Female) ตามลำดับ รวมถึงแปลงเดือนจากตัวเลขเป็นตัวหนังสือตามลำดับเดือนสากล ได้ดังนี้

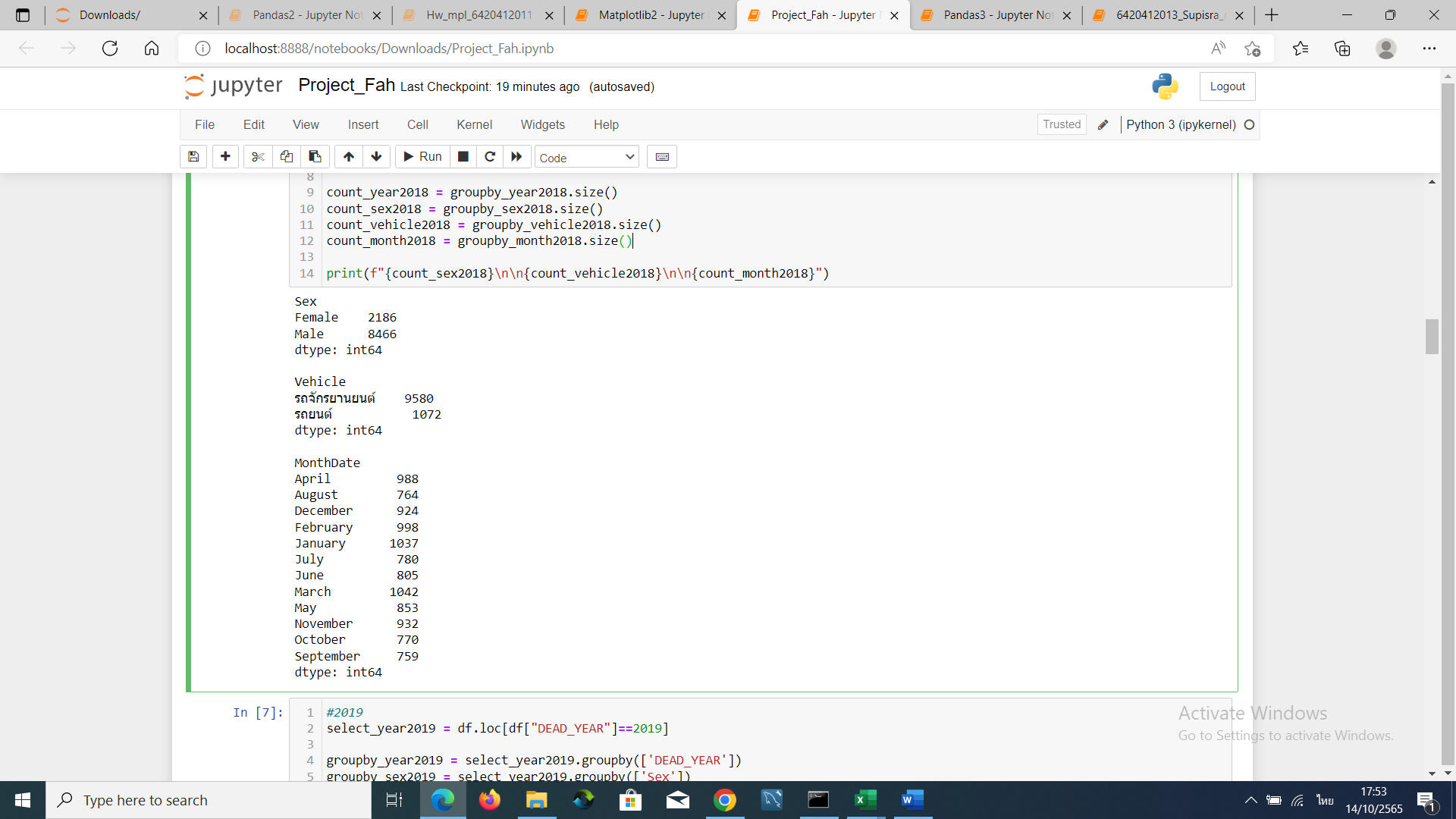
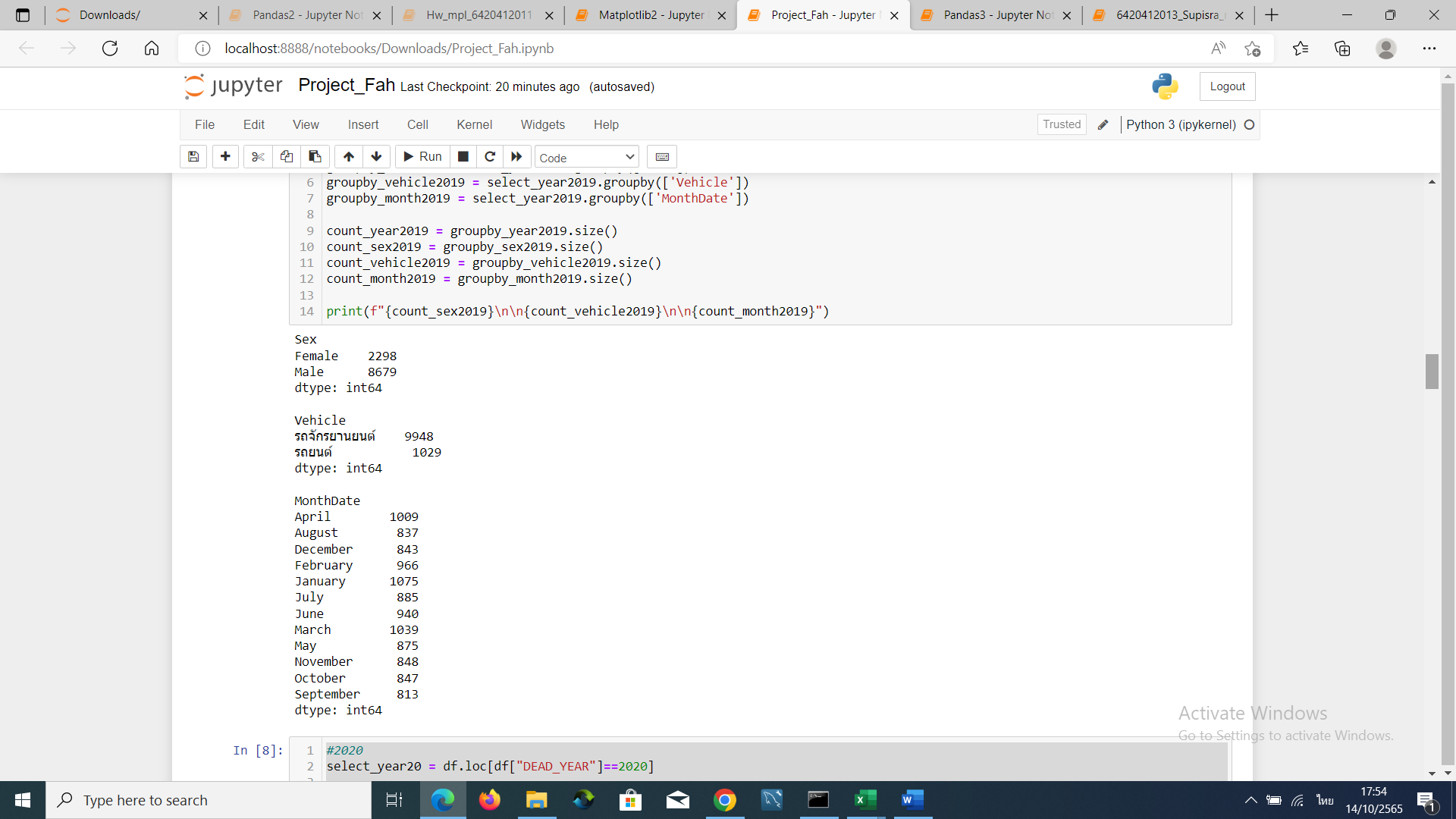
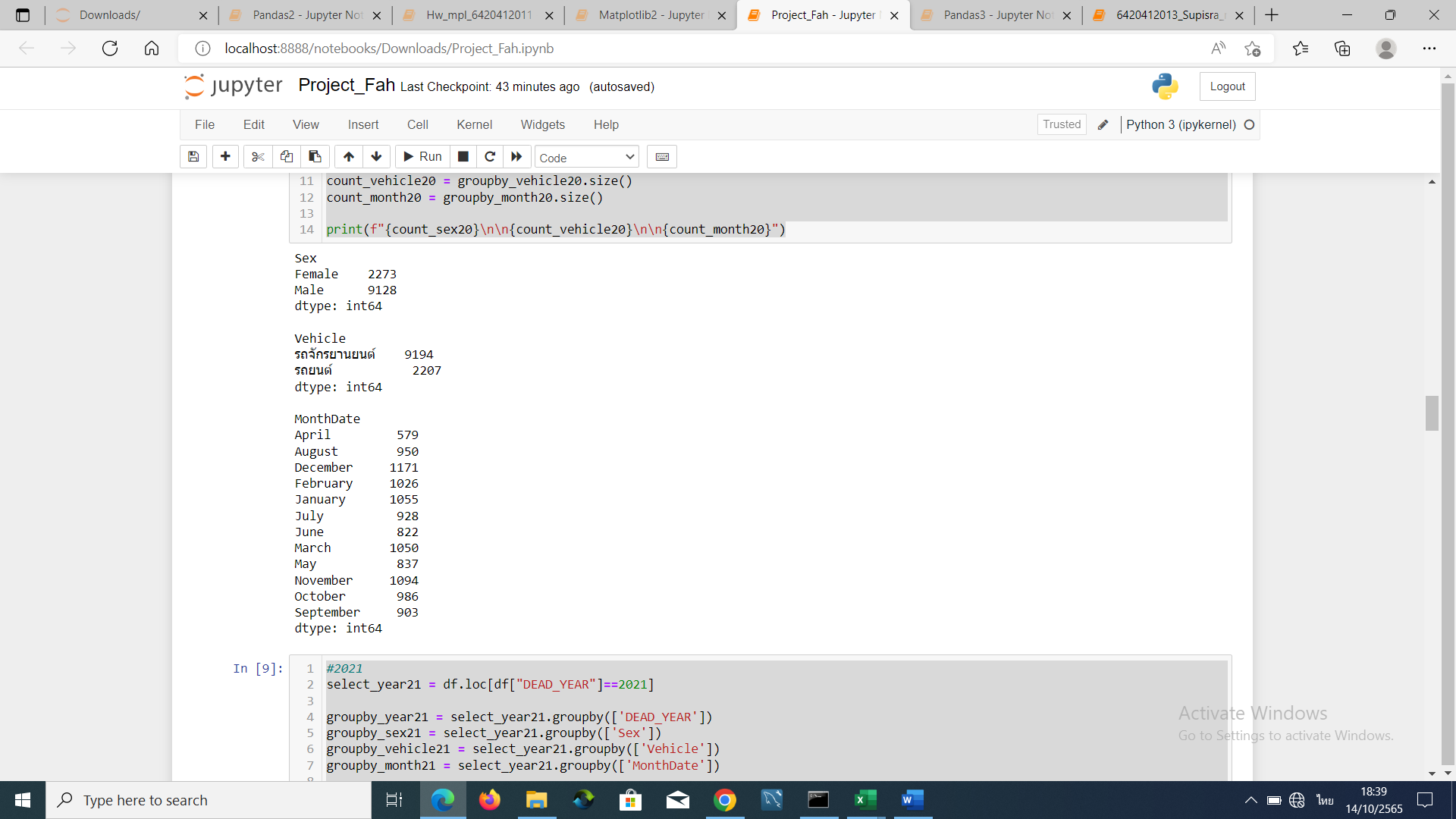


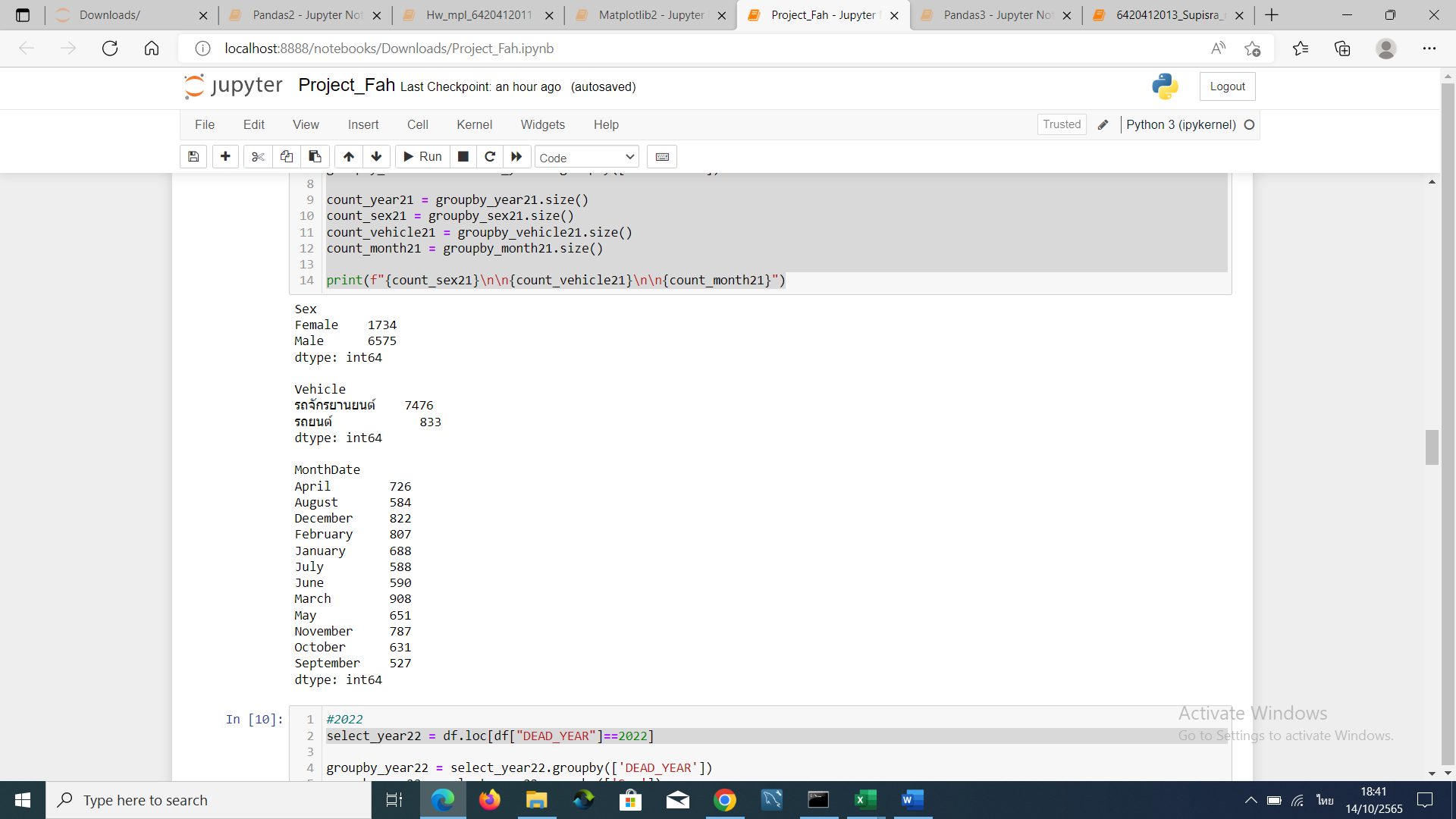
ตารางแสดงตัวแปรและแถวที่เราเลือกมา ได้ดังภาพข้างต้น

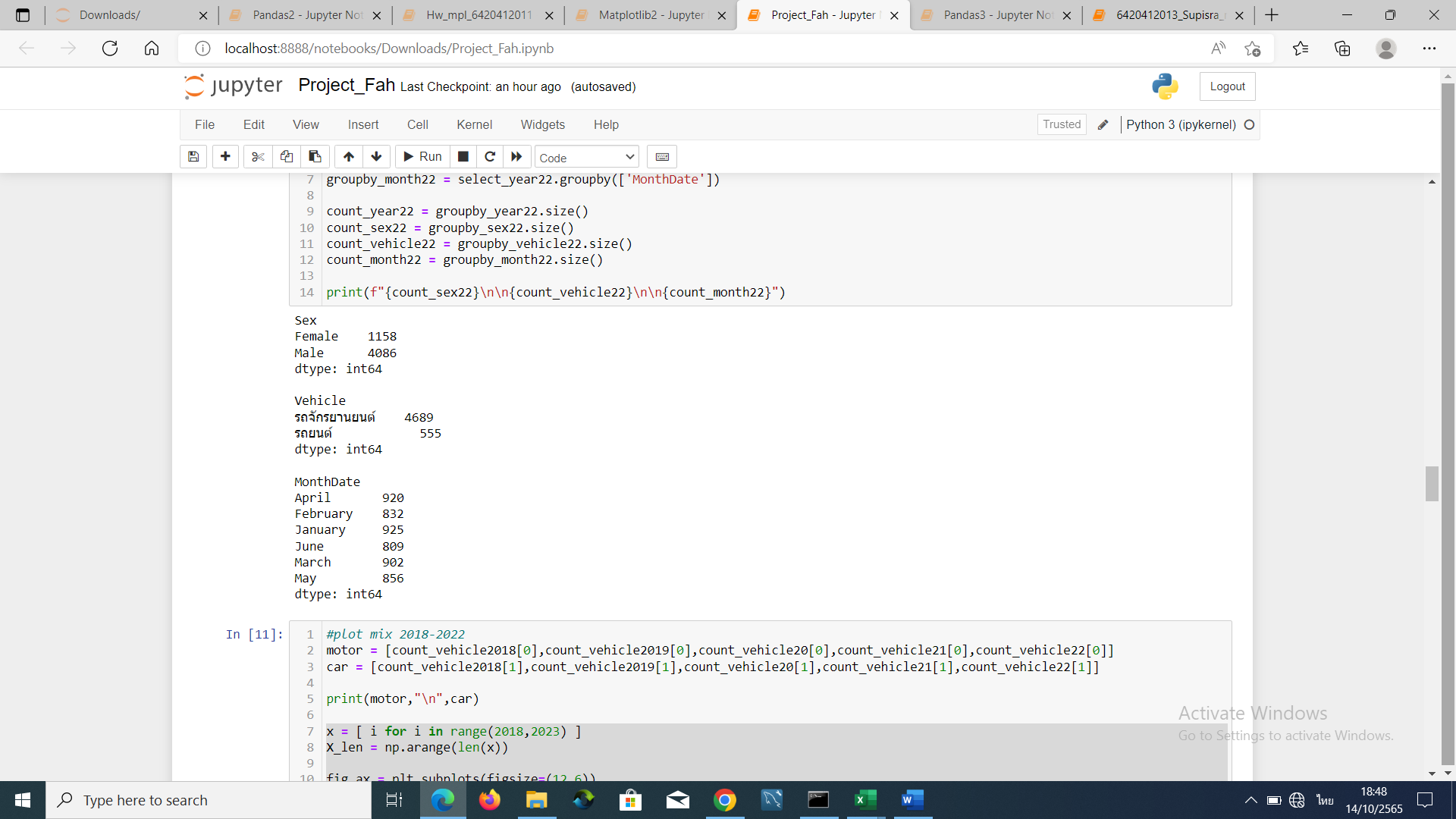
ขั้นตอนที่ 3) เป็นลำดับของการวิเคราะห์ข้อมูล โดยเริ่มจาก groupby ตัวแปรที่เราต้องการจะวิเคราะห์ ดังนี้



ซึ่งจะทำในลักษณะนี้จำนวน 5 ปีย้อนหลัง จะพบรายละเอียดจากการ groupby และ count ข้อมูลดังนี้

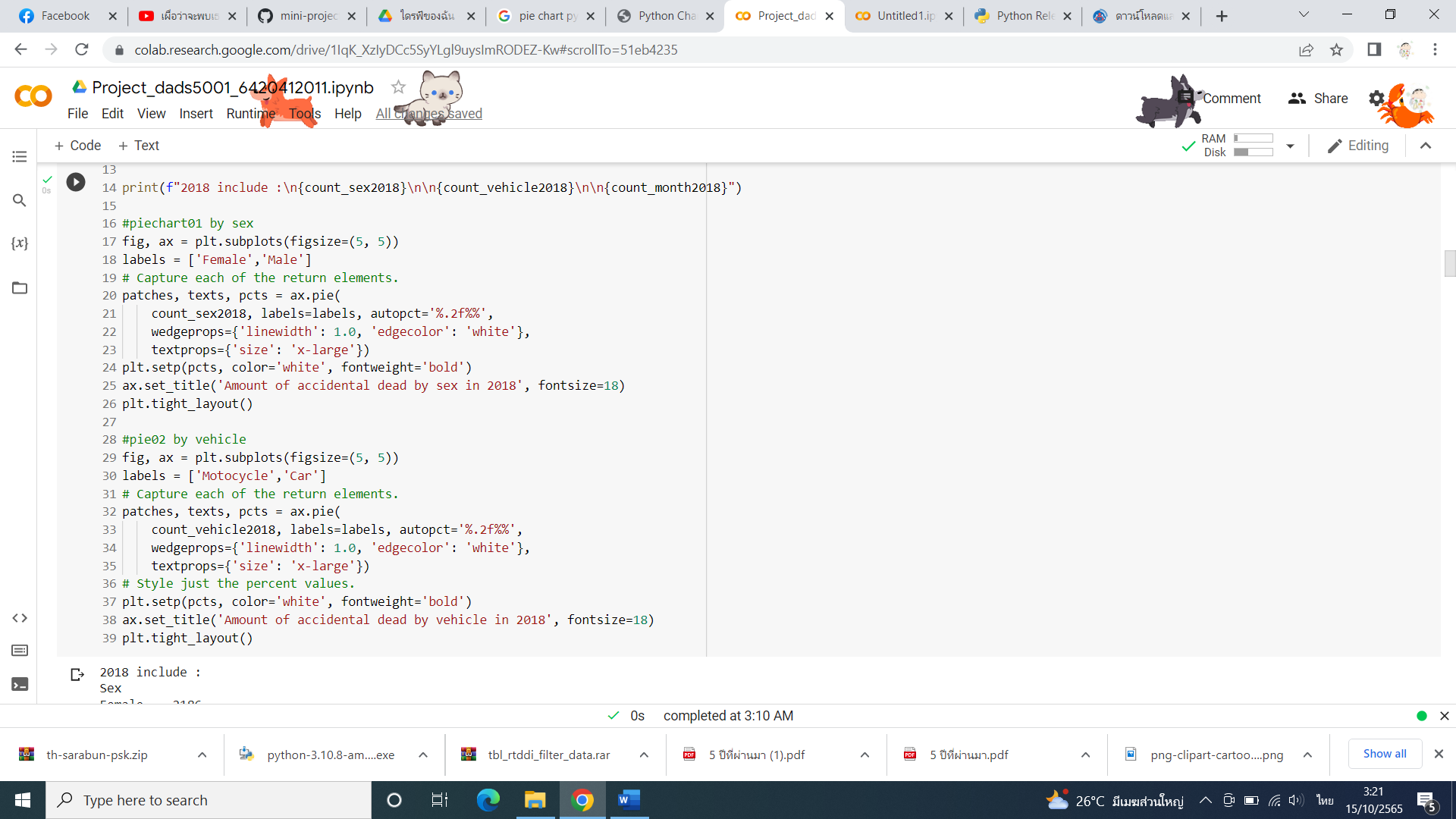
ในปี 2018 ในปี 2019 ในปี 2020



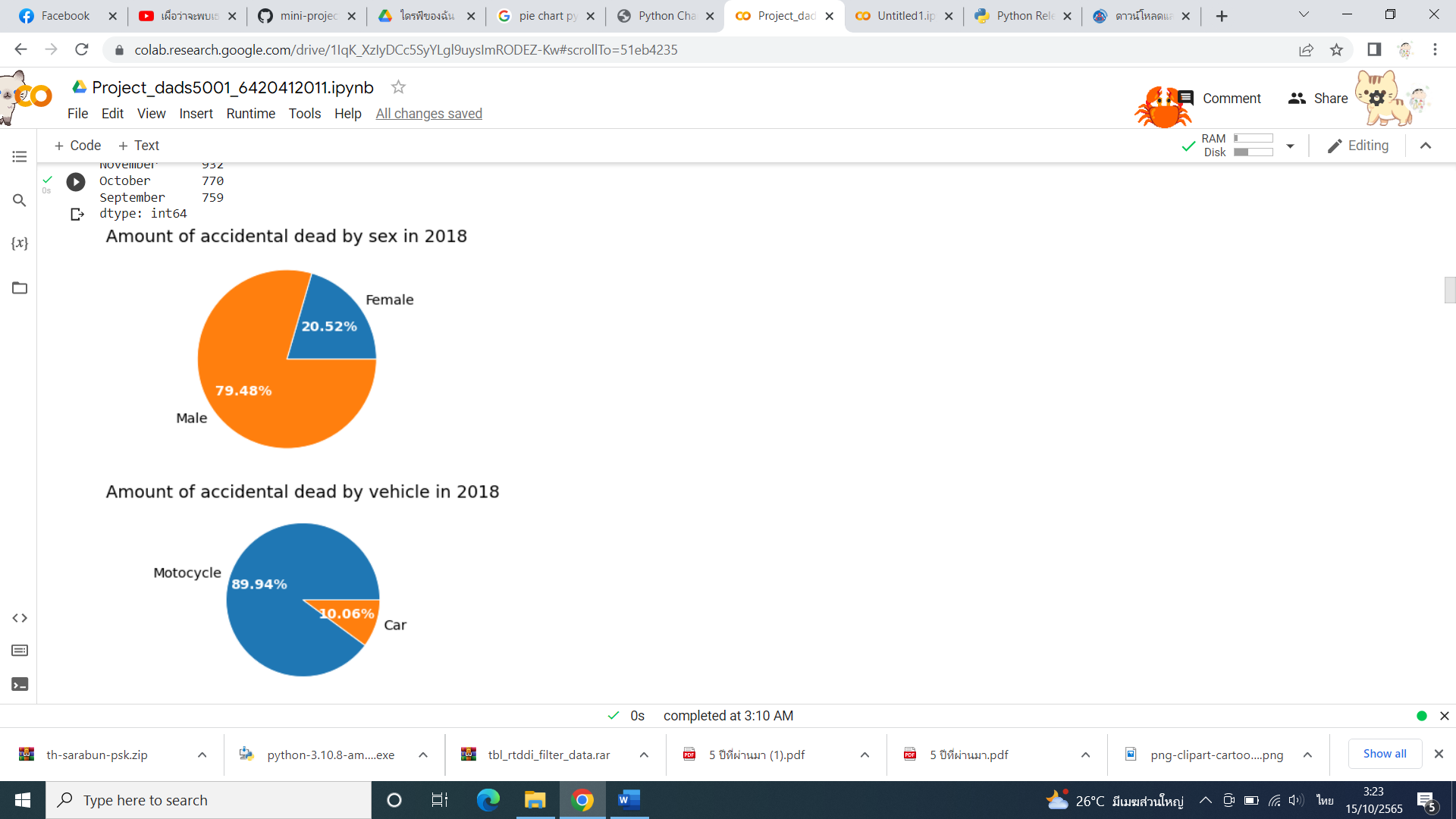


ในปี 2021 ในปี 2022

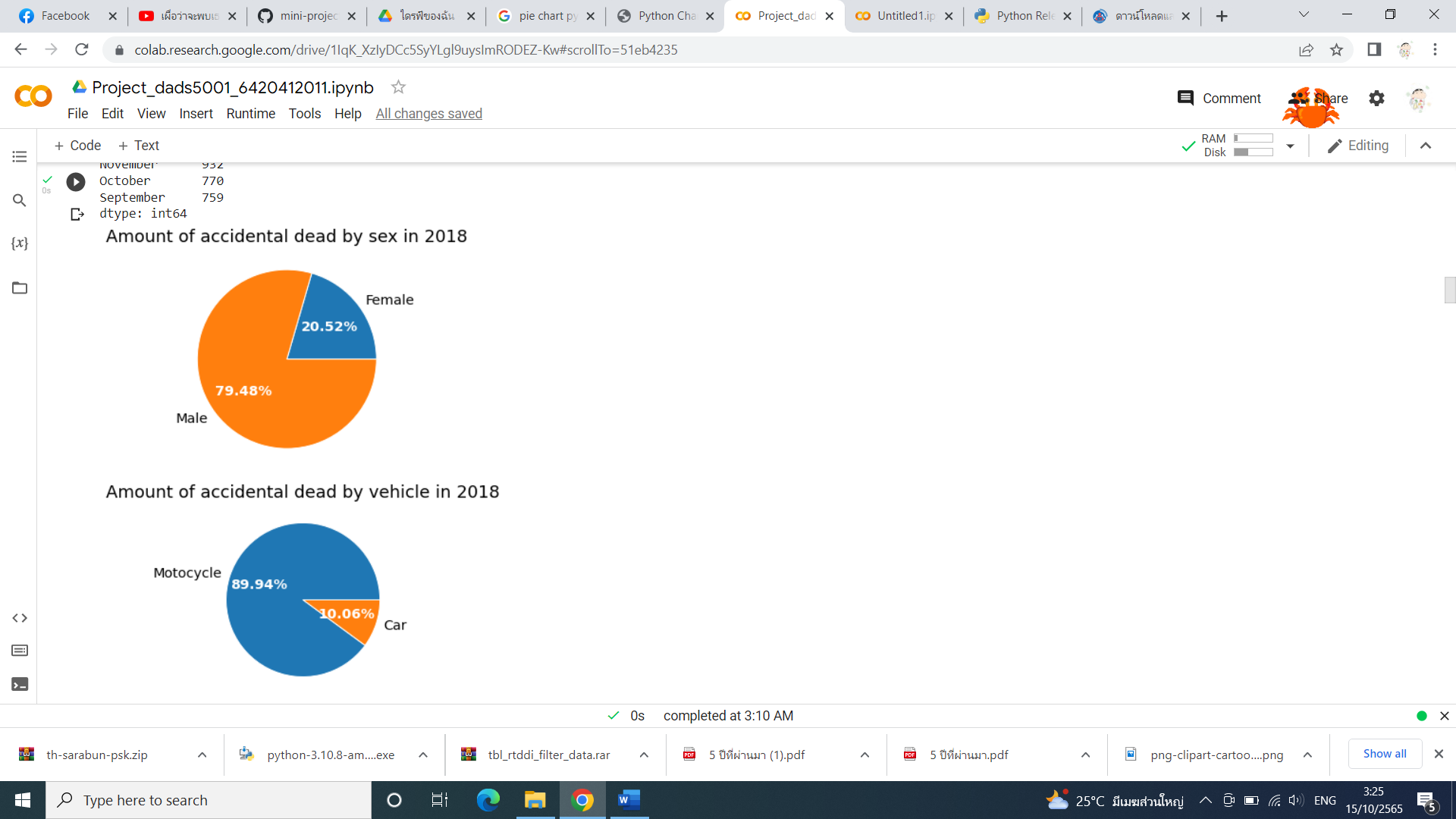
จัดทำกราฟวงกลมจำแนกตามเพศ และประเภทพาหนะ โดยทำซ้ำๆตั้งแต่ปี 2018 – 2022 ตามรายละเอียดที่กล่าวมาข้างต้น



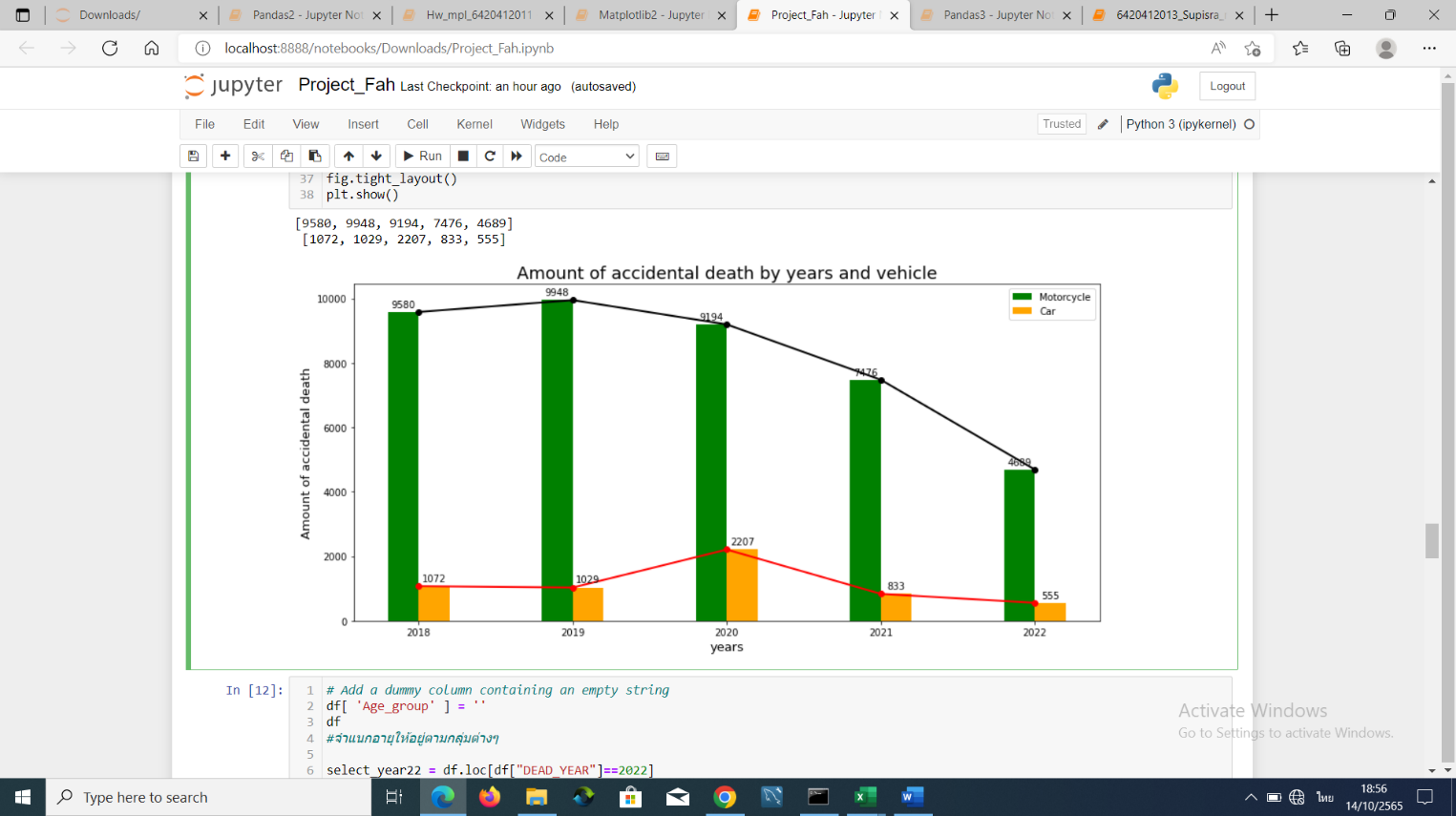
ตัวอย่างในปี 2018 พบว่า เพศชายเสียชีวิตคิดเป็นร้อยละ 79.48 และ เพศหญิงร้อยละ 20.52 (จำแนกตามเพศ)



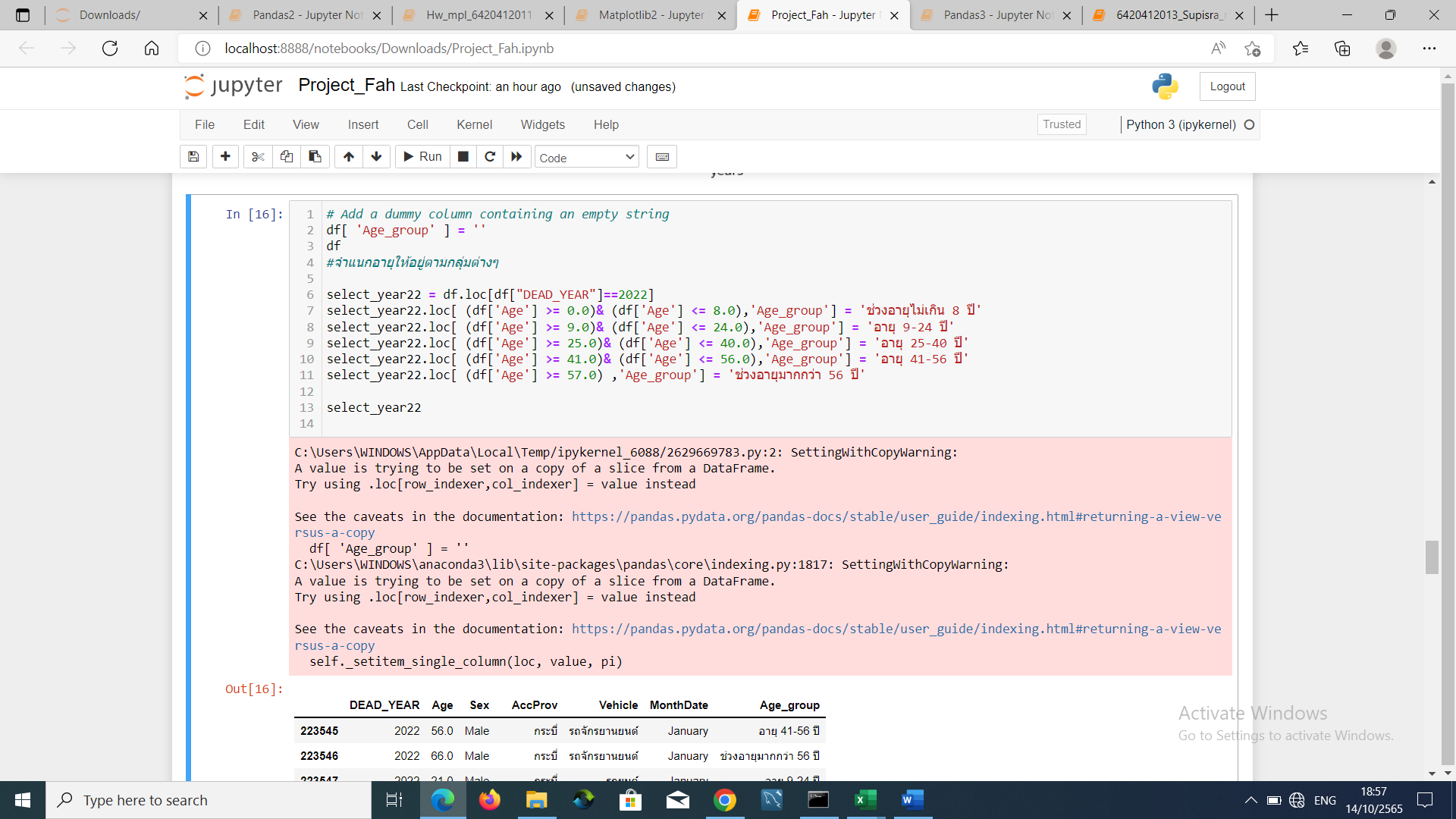
ตัวอย่างในปี 2018 พบว่า รถจักรยานยนต์ก่อให้เกิดอุบัติเหตุและเสียชีวิตคิดเป็นร้อยละ 89.94 และ รถยนต์ร้อยละ 10.06 (จำแนกตามพาหนะ)

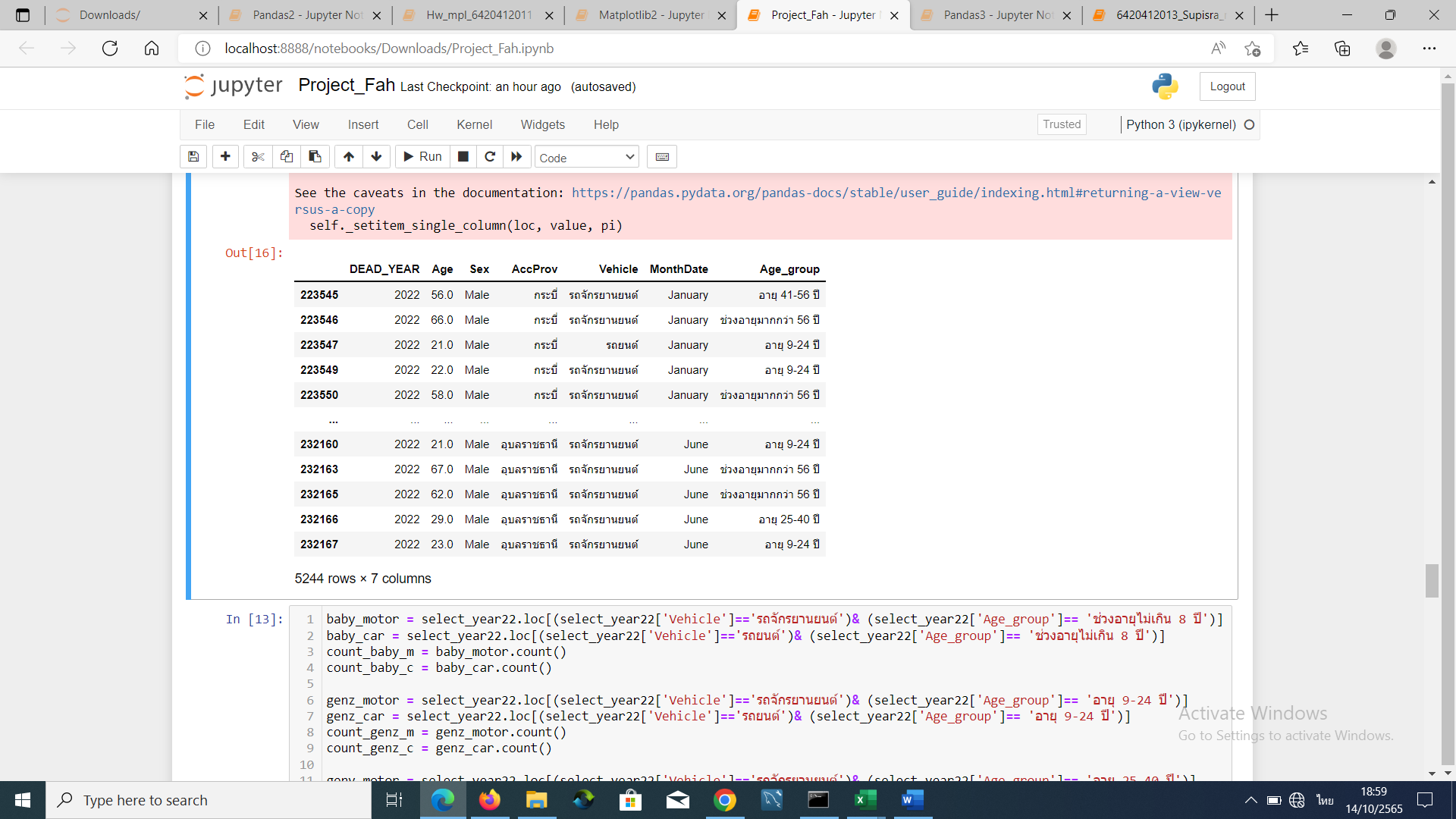


ขั้นตอนที่ 4) เป็นการทำ Visualization ที่จะนำสิ่งที่เราวิเคราะห์ได้ มา show ให้เห็นภาพมากยิ่งขึ้น โดยทางผู้จัดทำได้ทำ นำข้อมูลผู้ประสบอุบัติเหตุและเสียชีวิตรายปีและจำแนกตามพาหนะ(รถยนต์และรถจักรยานยนต์) มานำเสนอ จะได้ลักษณะดังนี้

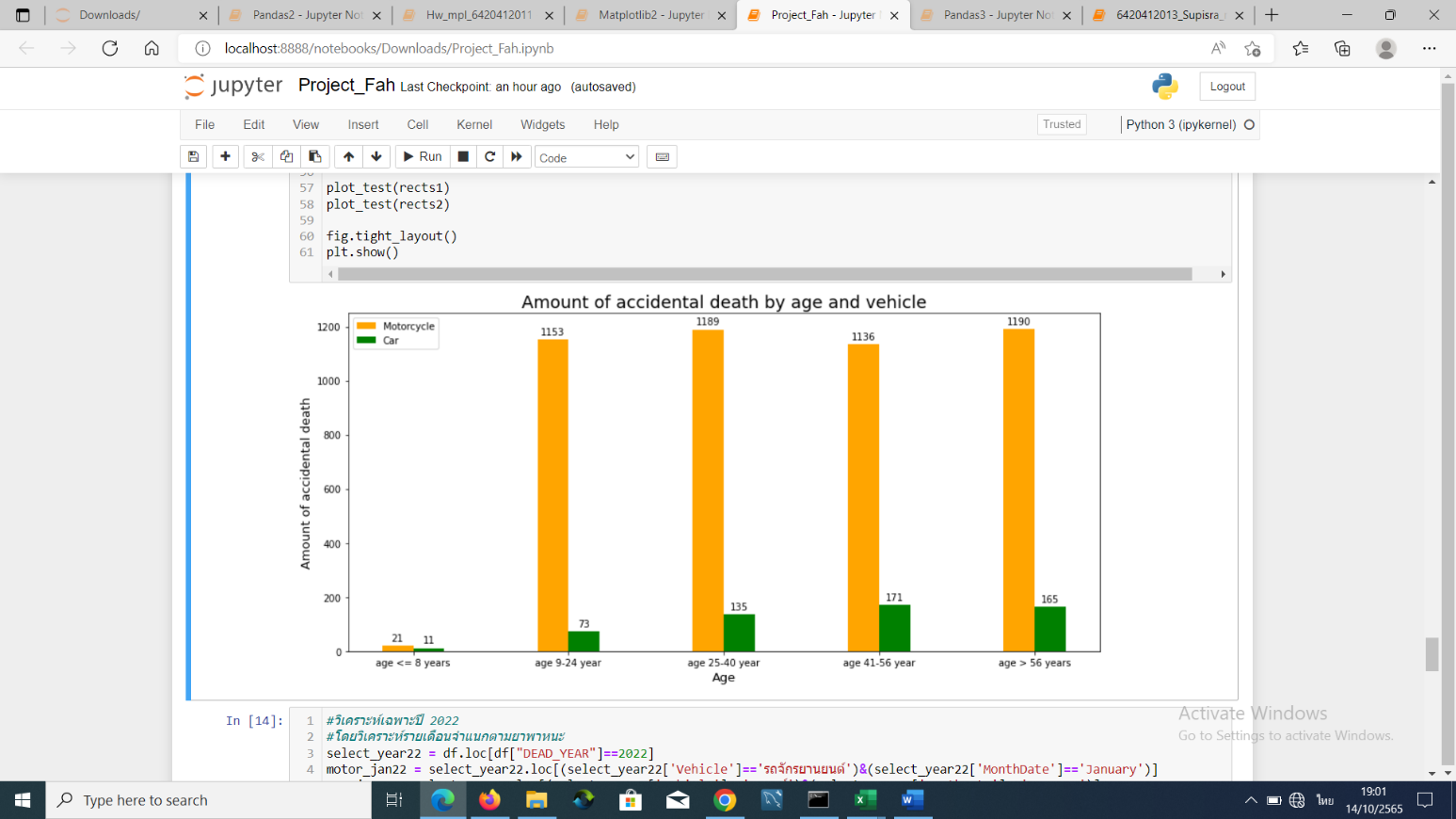


ในระหว่างการจัดได้พบว่า ช่วงอายุของผู้เสียชีวิตมาในลักษณะข้อมูลดิบ จึงจัดทำ คอลัมน์ดัมมี่ขึ้นมาชื่อว่า Age\_group ดังภาพ





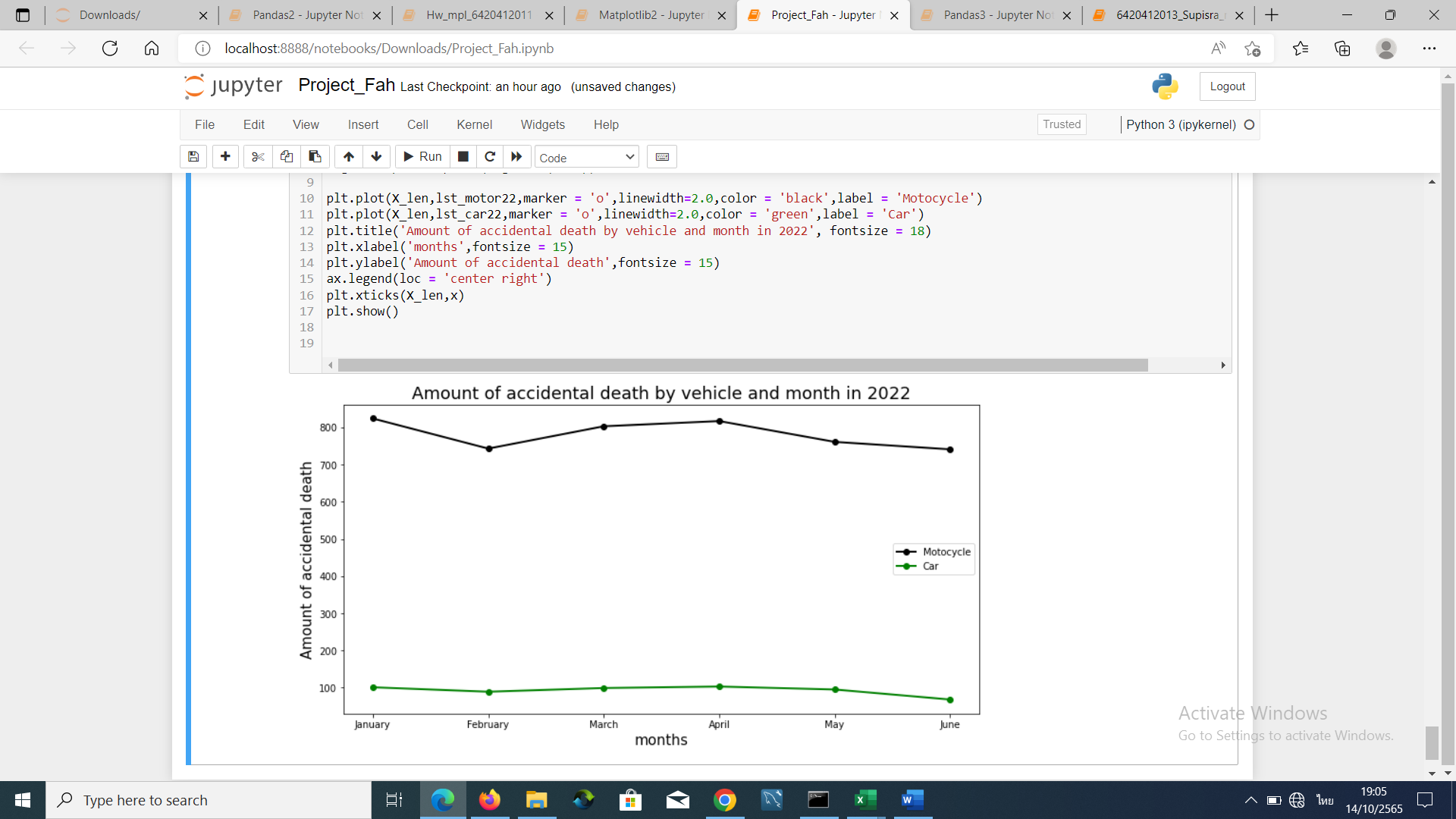
เมื่อนำตัวแปรที่พึ่งสร้างมาใหม่มา ทำการสร้าง bar chart จะได้



และสร้างทำการเลือกข้อมูลที่มีเงื่อนไข โดยพิจารณาเฉพาะปีล่าสุด คือ 2022



เมื่อนำมา plot เป็น line ดังนี้



สามารถสรุปได้ว่า สมมติฐานที่ตั้งไว้ว่า รถจักรยานยนต์กับรถยนต์ พาหนะประเภทใดก่อให้เกิดอุบัติเหตุและเสียชีวิตได้มากที่สุด คำตอบจากการวิเคราะห์ข้อมูลคือ รถจักรยานยนต์ ซึ่งสอดคล้องกับคำตอบที่ผู้จัดทำตั้งไว้

ปัญหาที่พบในการทำงานครั้งนี้ เนื่องจากในวันที่ 14 ตุลาคม 2565 เวลาประมาณ 01.55 น. ผู้จัดทำกำลังจะทำ pie chart ที่เป็นส่วนเสริมในแต่ละปี เครื่องโน้ตบุ้คเกิดปัญหาขึ้น เมื่อนำไปซ่อมพบว่า ฮาร์ดดิสเสีย และได้ส่งซ่อม ทำให้ในเช้าวันเดียวกันต้องจัดทำงานขึ้นใหม่หมด \*\*อาจจะมีการแก้ไขงาน เพิ่มในส่วนของกราฟลงไป เพราะในส่วนนี้ทำเพื่อนำเสนอ เป็นการแก้ไขปัญหาเฉพาะหน้าก่อน\*\* รวมถึงพบปัญหาว่าข้อมูลชุดนี้ ในตัวแปรบางมีข้อมูลไม่ครบบ้าง ตอบไม่ตรงกับคำถามบ้าง จึงต้องทำการ drop ข้อมูลทิ้งลงไปบางส่วน และเลือกข้อมูลบางส่วนมาทำการวิเคราะห์ รวมถึงปัญหาในการทำ groupby ที่ผู้จัดทำลืมทำการ .size() เพื่อนับจำนวนข้อมูลที่สนใจ