**นายภากรณ์ ธนประชานนท์ 62010694**

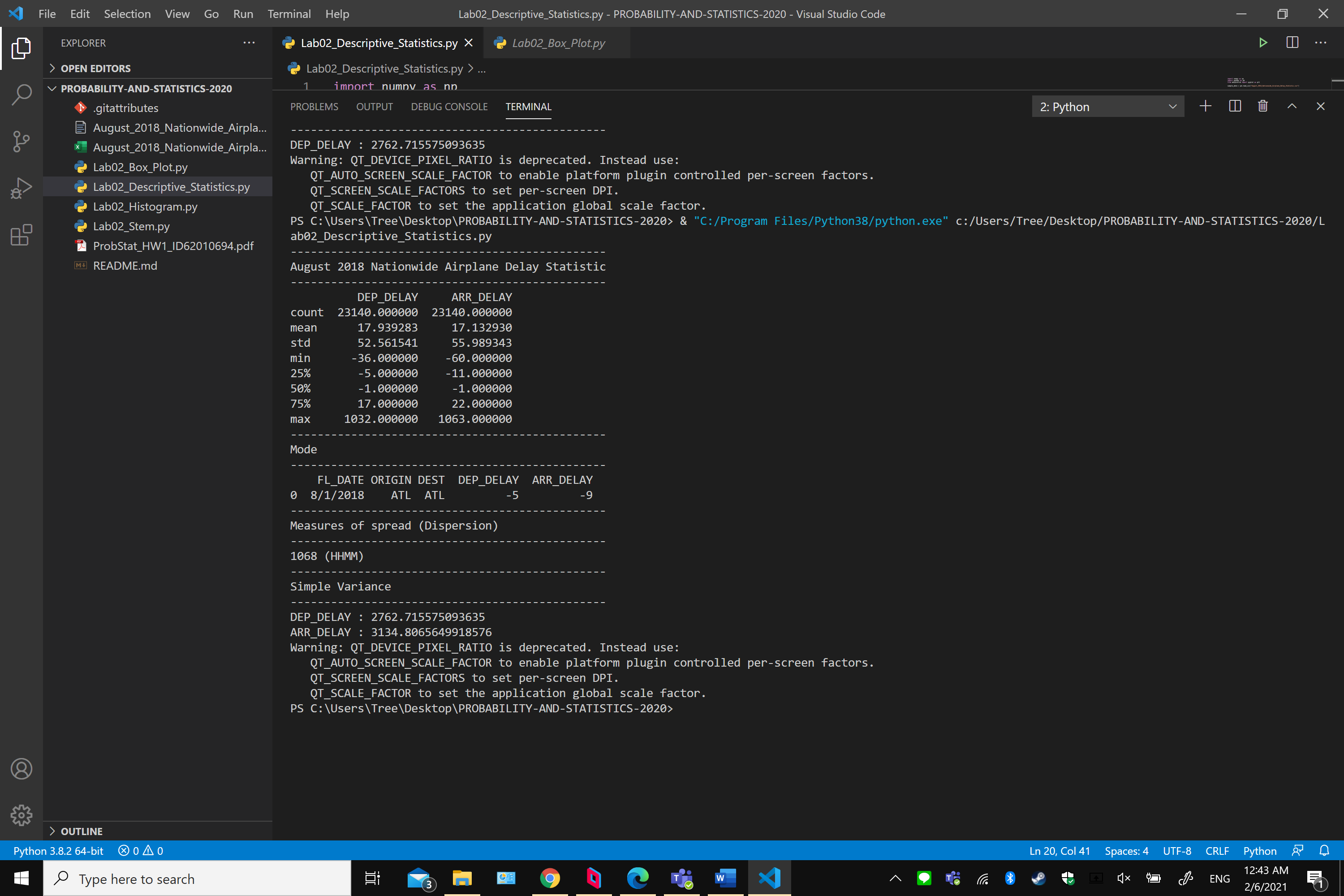
**Homework #2**

**สถิติความล่าช้าของเที่ยวบินภายในประเทศสหรัฐอเมริกาเดือนสิงหาคม**

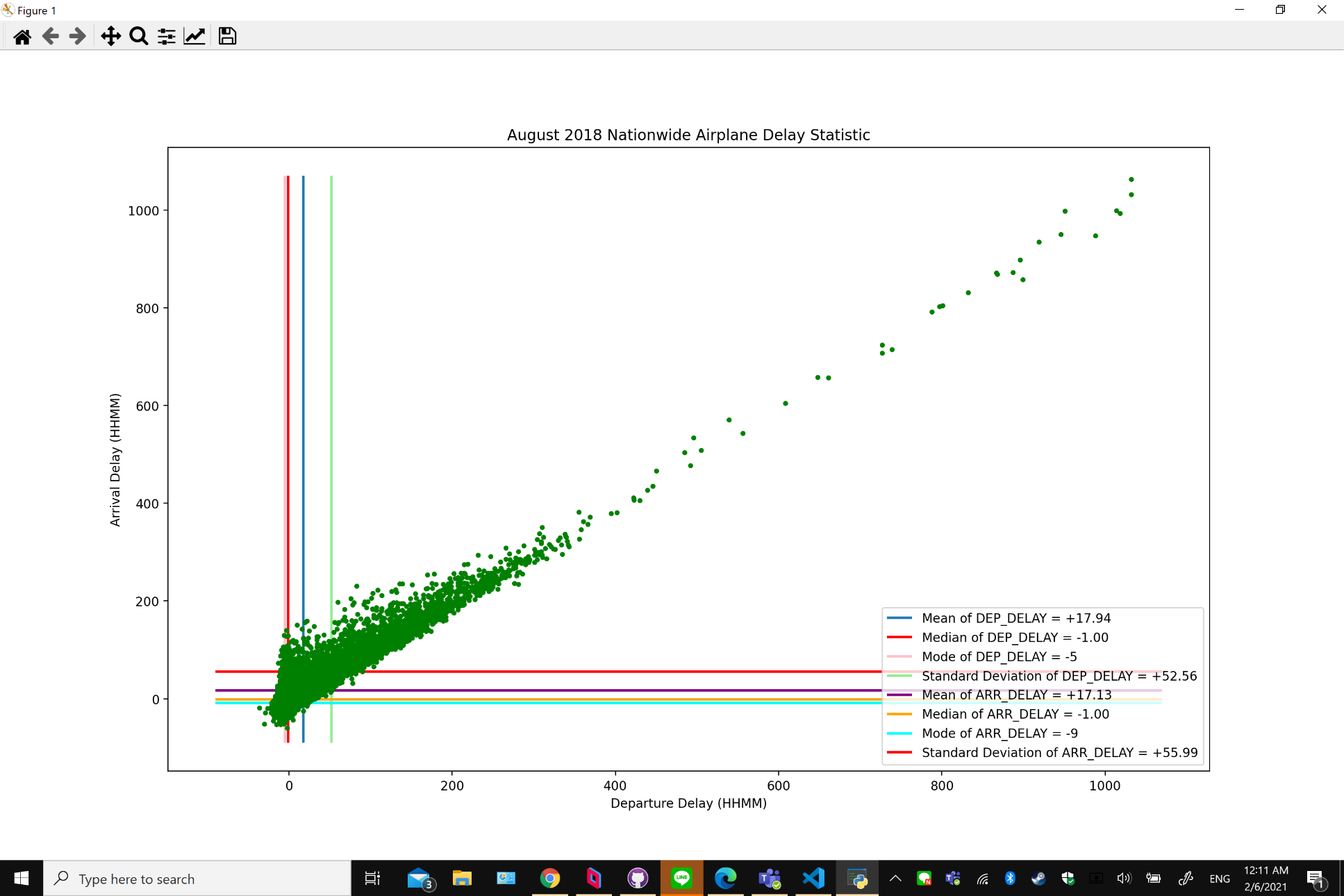
**2018 August 2018 Nationwide Airplane Delay Statistic**

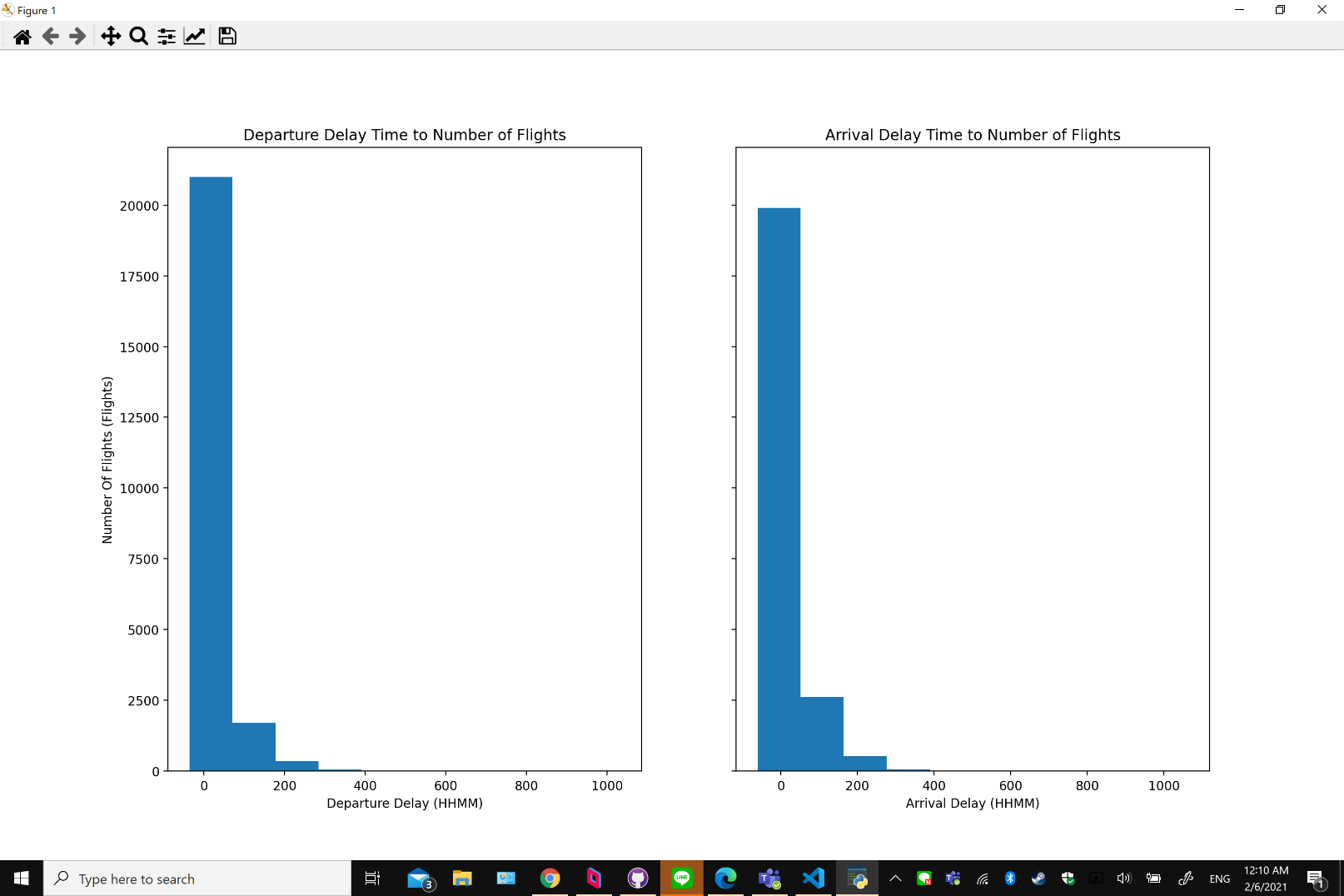
เลือกใช้ Column : DEP\_DELAY ( Departure Delay (HHMM) ) และ ARR\_DELAY ( Arrival Delay (HHMM) ) เพื่อหาความสัมพันธ์ว่า และเวลาล่าช้าขาออกมีผลต่อเวลาล่าช้าขาเข้าหรือไม่

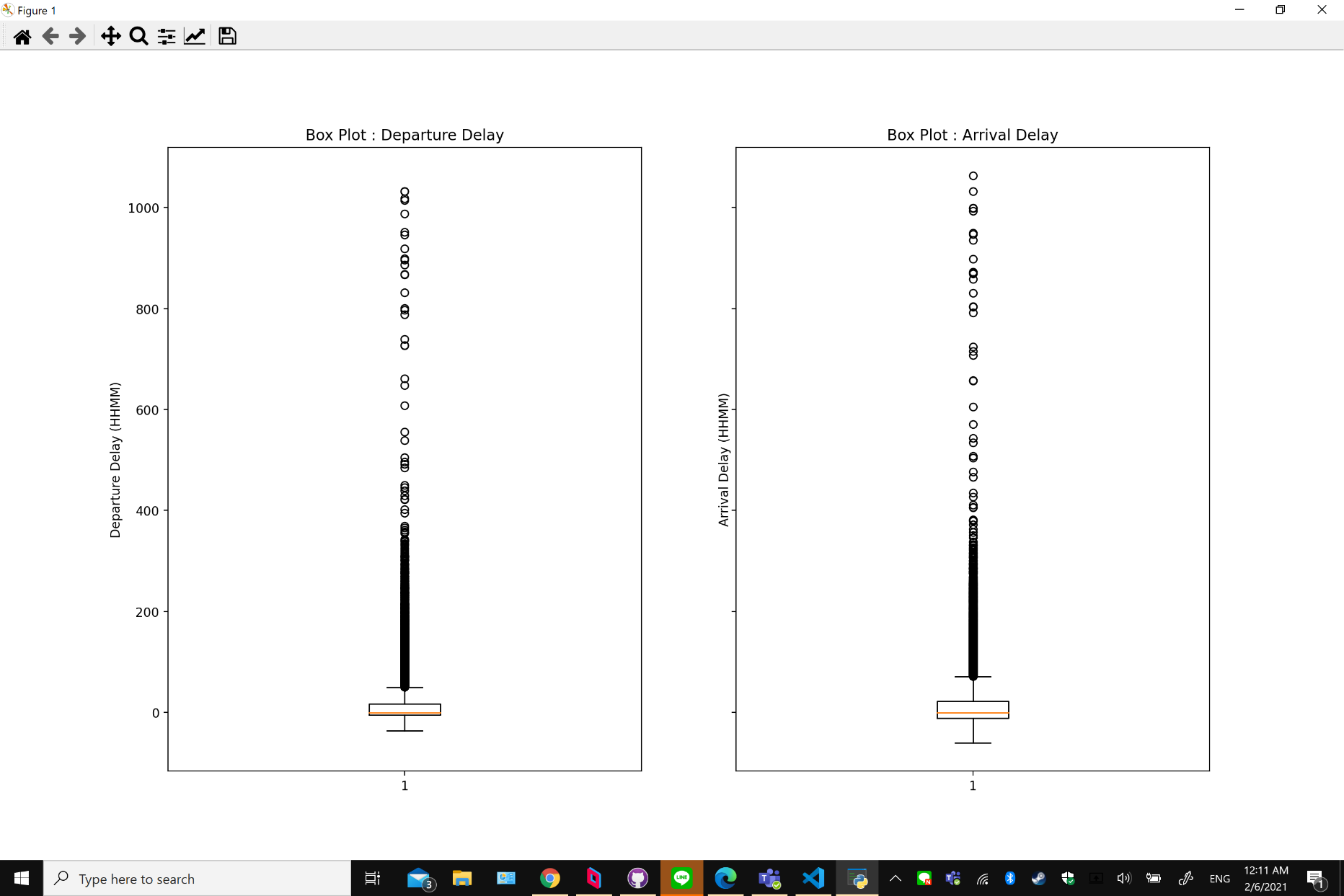
ข้อมูลสถิติต่างๆที่ได้จากชุดข้อมูลนี้

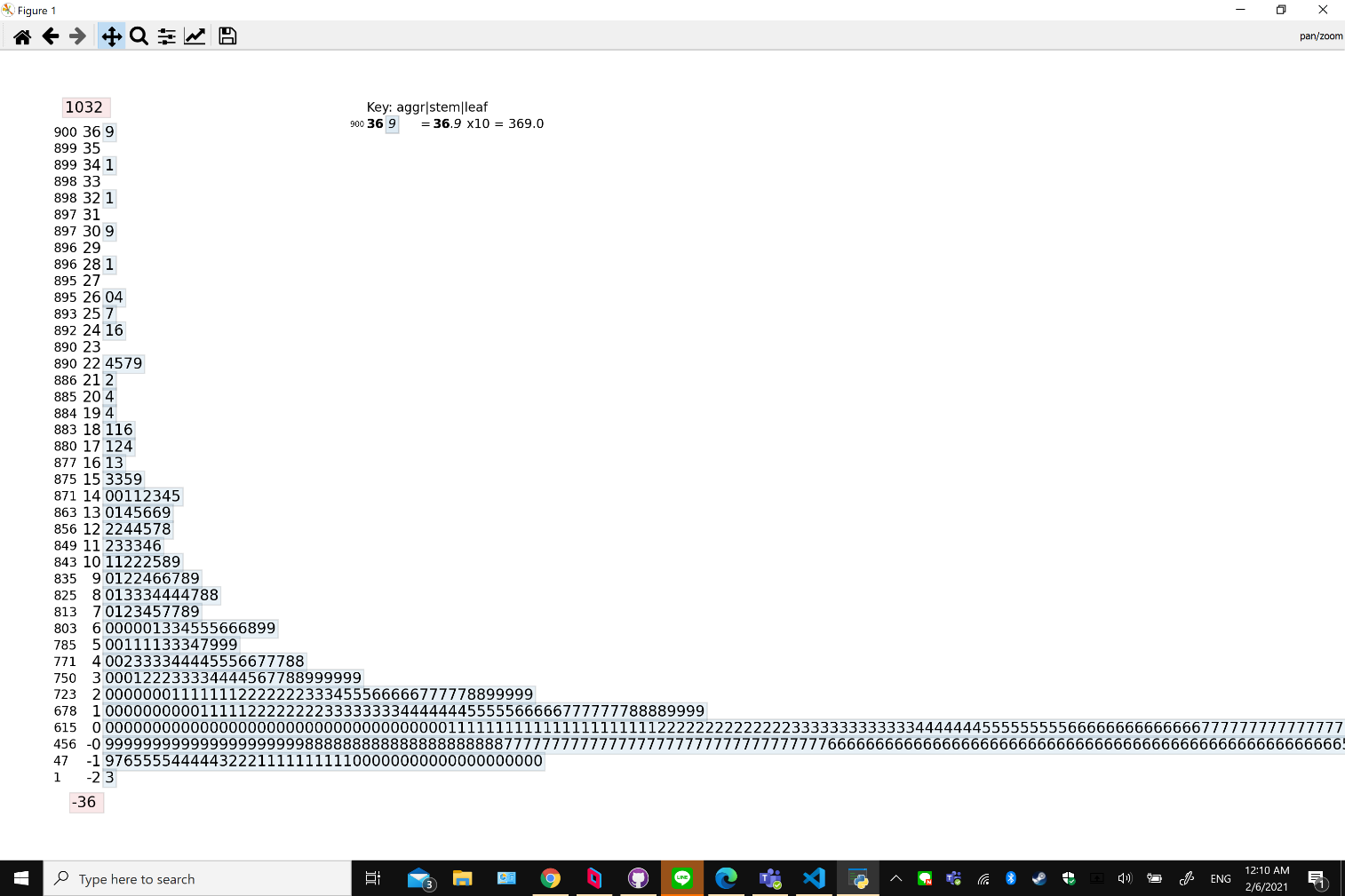
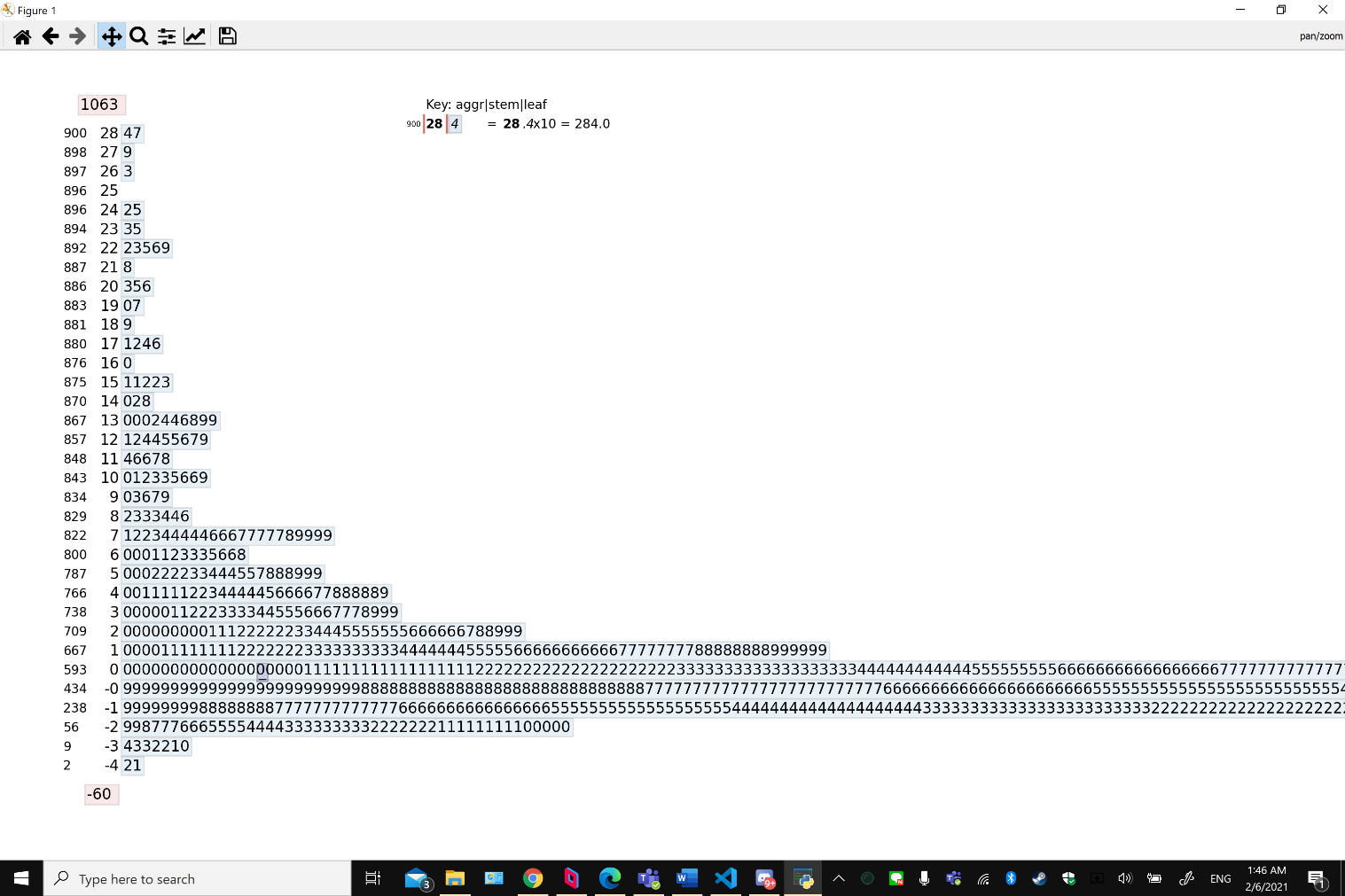


โดยาค่าสถิติพื้นฐานอื่นๆสามารถดูได้จากกราฟ Scatter Polt ด้านล่าง

**Scatter Plot**

**Histogram**

**Box Plot**

******Stem and Leave:** Departure Delay (HHMM)

**Stem and Leave:** Arrival Delay (HHMM)

**ให้ตัวแปรต้นเป็น** : DEP\_DELAY ( Departure Delay (HHMM) ) เวลาล่าช้าขาออก

**ตัวแปรตามเป็น** : ARR\_DELAY ( Arrival Delay (HHMM) ) เวลาล่าช้าขาเข้า

เหตุผล เพราะต้องการที่จะศึกษาประสิทธิภาพของการเดินทาง ด้วยเครื่องบินพาณิชย์เที่ยวบินภายในของประเทศสหรัฐอเมริกาว่าเวลาล่าช้าขาออกมีผลต่อเวลาล่าช้าขาเข้าหรือไม่ หรือว่าสองอย่างนี้ไม่มีความเกี่ยวข้องโดยกันสิ้นเชิง มีความเป็นไปได้หรือไม่ที่เครื่องบินออกช้าจะถึงจุดหมายตรงเวลา หรือเร็วกว่าเวลากำหนดเดิม

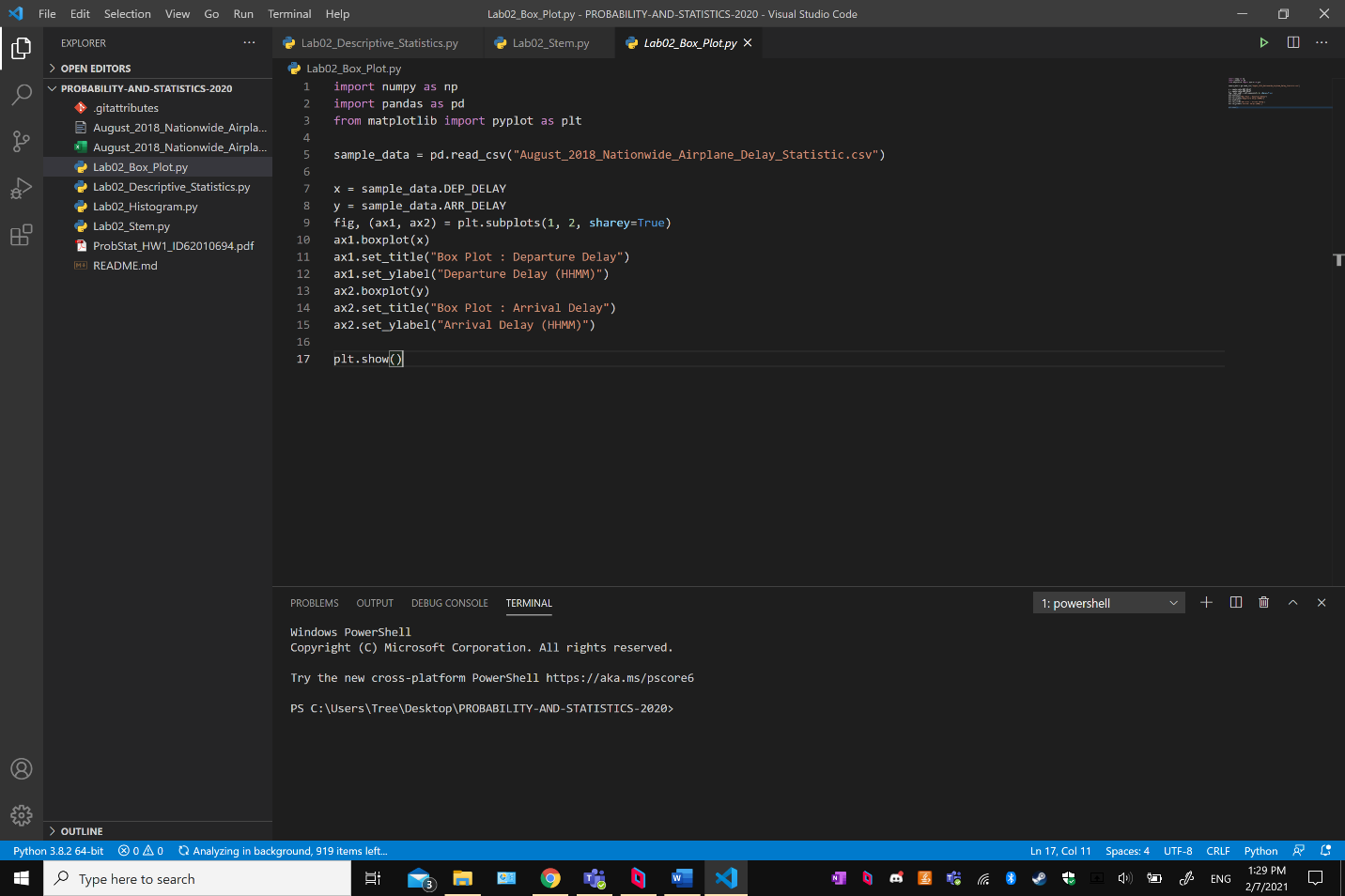
**Outlier:** เนื่องจากชุดข้อมูลนี้มี Outlier จำนวนมากจึงจะขอยกตัวอย่างค่าที่เป็น Outlier สูงสุดแทน

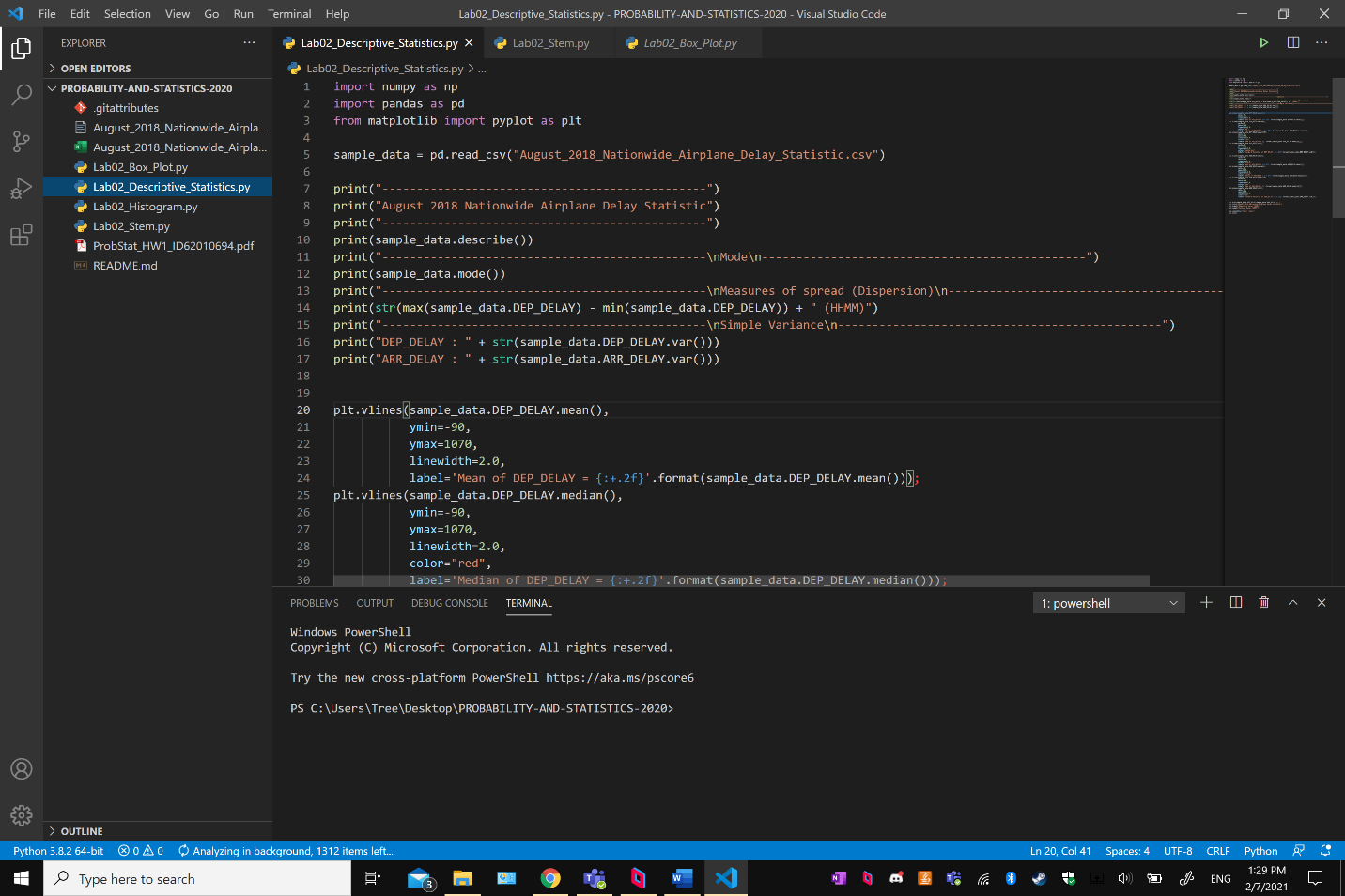
ตามแนวแกน X ( ค่าเวลาล่าช้าขาออก ) : 1032 (ล่าช้าไป 10 ชั่วโมง 32 นาที)

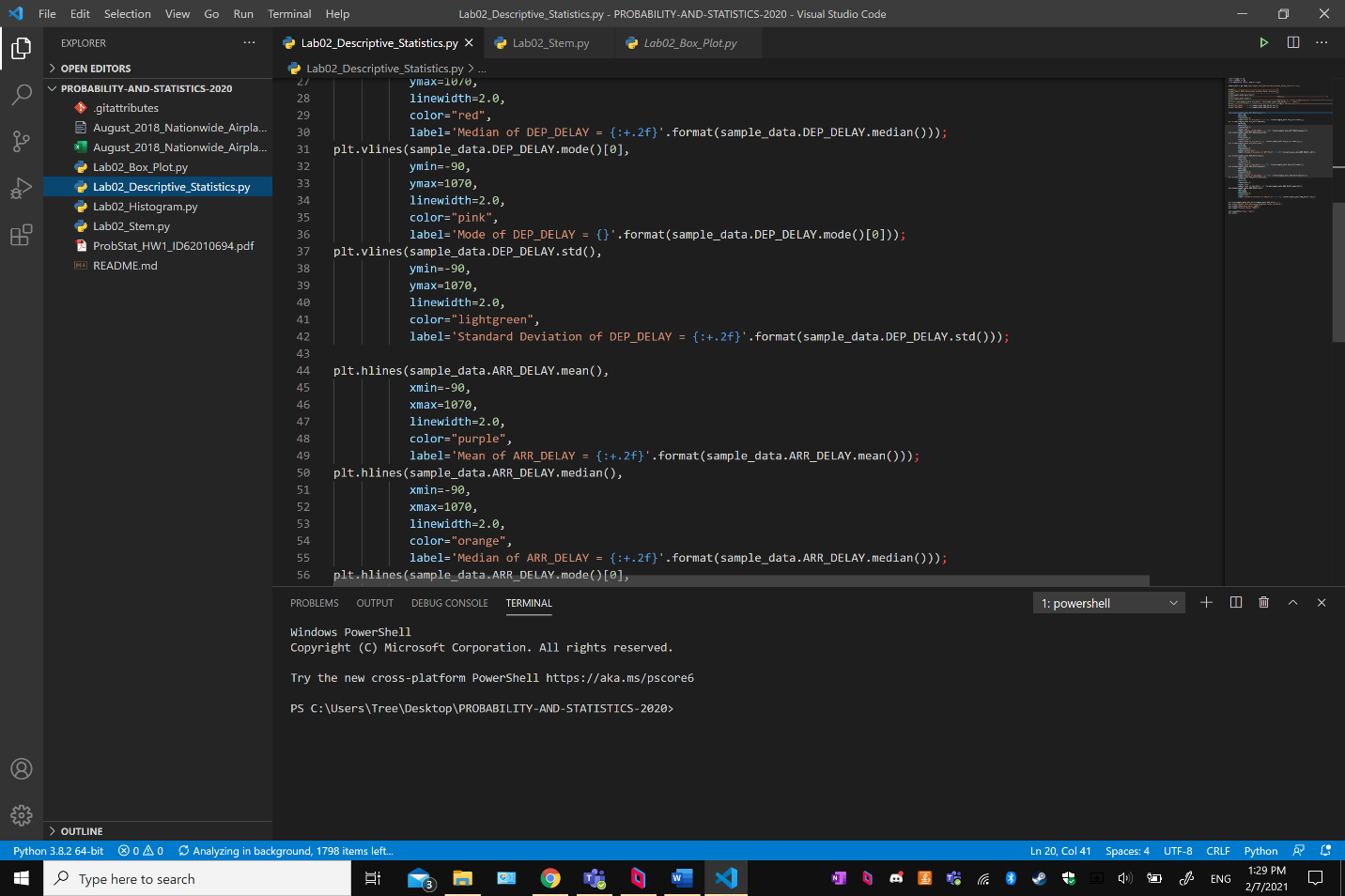
ตามแนวแกน Y ( ค่าเวลาล่าช้าขาเข้า ) : 1063 (ล่าช้าไป 11 ชั่วโมง 3 นาที) \*คาดว่าเป็นการผิดพลาดของการเก็บข้อมูล

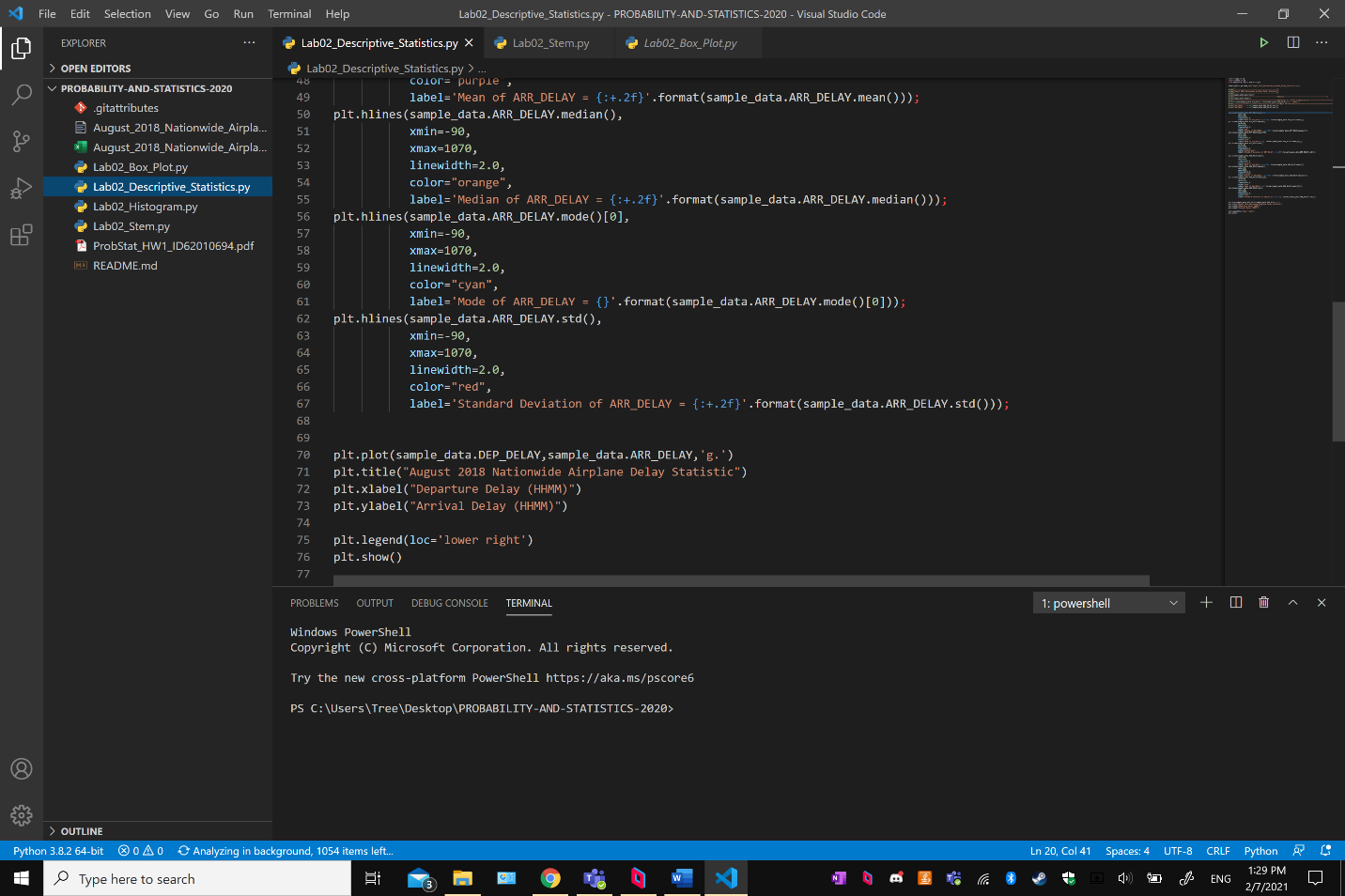
**บทวิเคราะห์ข้อมูลจากกราฟ** : จากชุดข้อมูลที่ได้นำมาจะสามารถเห็นได้ว่า ค่าเวลาล่าช้าขาออก ( Departure Delay (HHMM) ) ส่งผลโดยตรงกับ ค่าเวลาล่าช้าขาเข้า ( Arrival Delay (HHMM) ) อย่างชัดเจนในรูปแบบแปรผันตรง หากเที่ยวบินออกช้า เวลาถึงจะช้าตามไปด้วย มีความเป็นไปได้ที่เครื่องบินออกช้าจะถึงจุดหมายตรงเวลา หรือเร็วกว่าเวลากำหนดเดิม แต่จะไม่มีทางเร็วไปมากกว่าสองชั่วโมงอย่างแน่นอน มีเที่ยวบินที่ล่าช้าถึง 10 ชั่วโมงอยู่จริง และจากการหาค่า Mode เราสามารถระบุสนามบินทั้งต้นทางและปลายทางที่มีประวัติเครื่องบินล่าช้าสุดในเดือนสิงหาคม 2018 ได้ นั่นคือท่าอากาศยานนานาชาติฮาร์ทสฟิลด์–แจ็คสัน แอตแลนตา (Hartsfield-Jackson Atlanta International Airport) ซึ่งมี IATA Airport Code คือ ALT และจากทั้งชุดข้อมูลนี้ ค่าเวลาล่าช้าขาออก ( Departure Delay (HHMM) ) และ ค่าเวลาล่าช้าขาเข้า ( Arrival Delay (HHMM) ) ที่พบได้มากที่สุดคือ -5 (ออกไวไป 5 นาที) และ -9 (ถึงไวไป 9 นาที) ตามลำดับ สุดท้ายจากชุดข้อมูลนี้เราสามารถบอกได้ว่า เวลาล่าช้าของเครื่องบินนั้นจะอยู่ในช่วงถึงช้า 2 ชั่วโมงและออกช้า 2 ชั่วโมง เพราะข้อมูลจะเกาะกลุ่มกันอยู่ในช่วงซ้ายล่างของ Scatter Plot เที่ยวบินที่ล้าช้ามากกว่านี้มีอยู่จริง แต่ถือว่าเป็นจำนวนน้อยมากหากเทียบกับจำนวนข้อมูลที่เกาะกลุ่มนี้

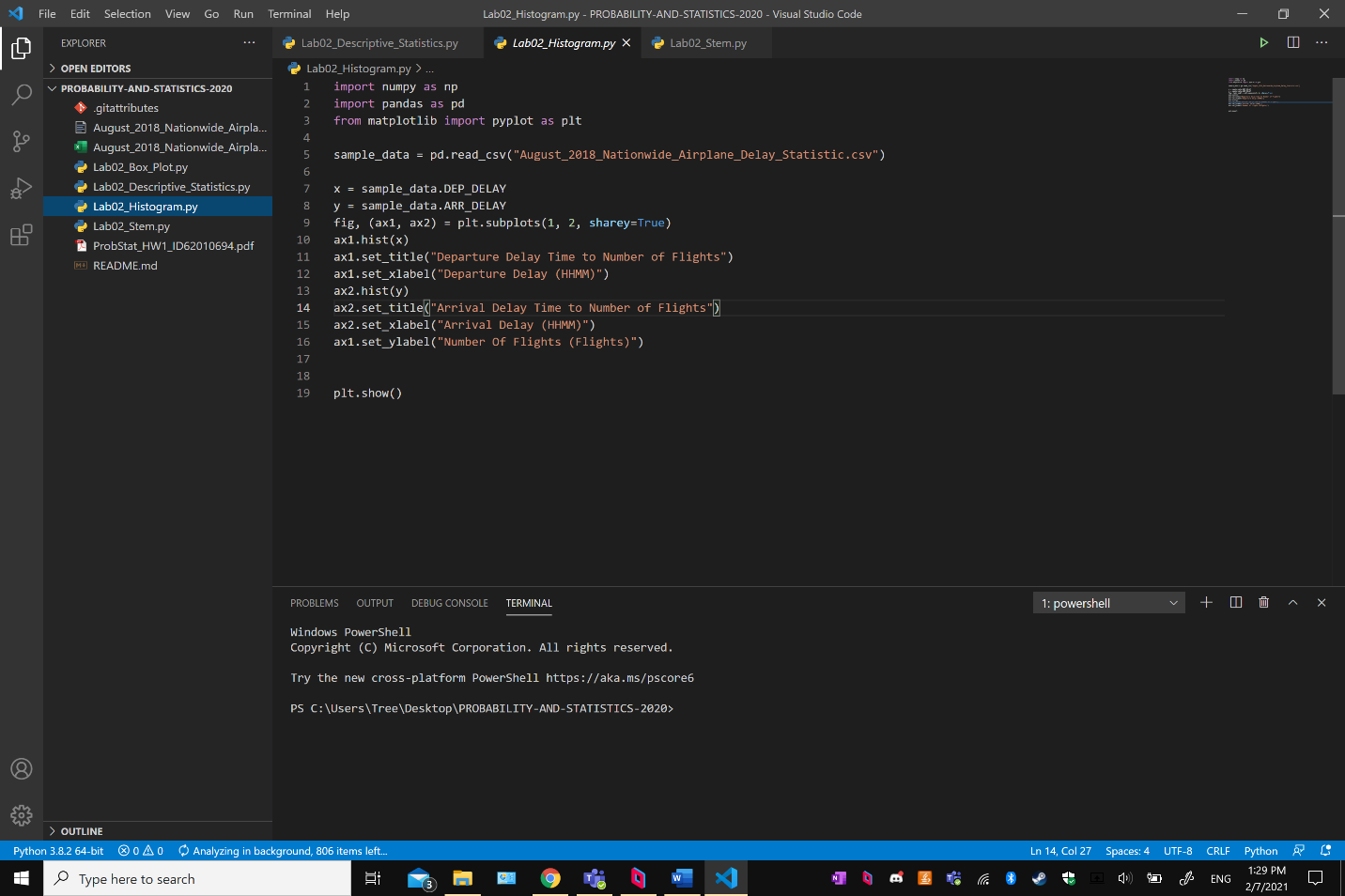
**เราจึงวิเคราะห์ได้ว่า เวลาล่าช้าทั้งขาออกจะส่งผลกับเวลาล่าช้าขาเข้าในเกือบจะทุกกรณีในรูปแบบแปรผันตรงอย่างเห็นได้ชัด หรือกล่าวอีกนัยหนึ่งคือ หากเครื่องออกช้า จะมีความเป็นไปได้สูงอย่างมากที่จะถึงที่หมายปลายทางล่าช้าอย่างแน่นอน**

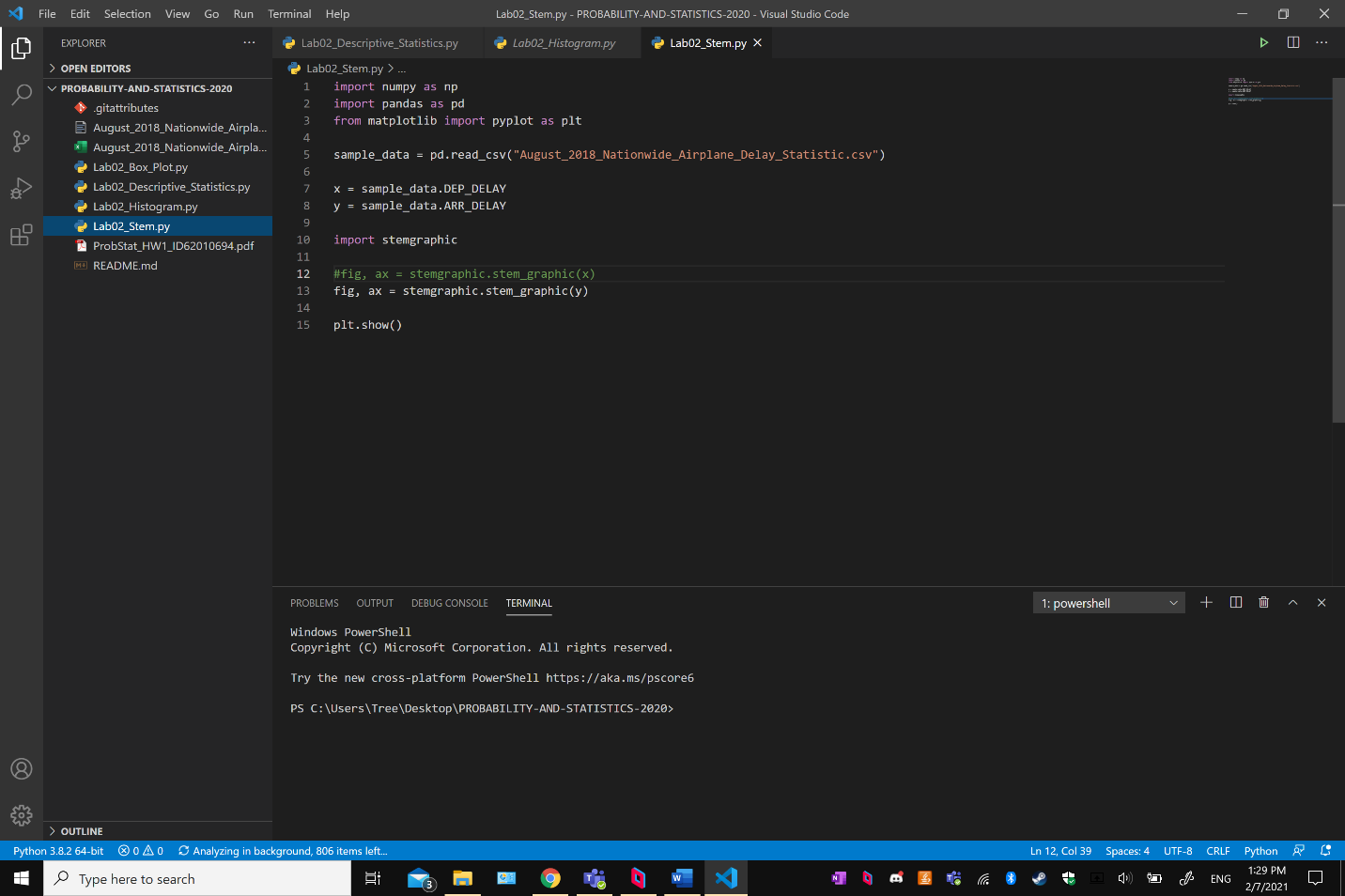
****

****

****

****

****

****