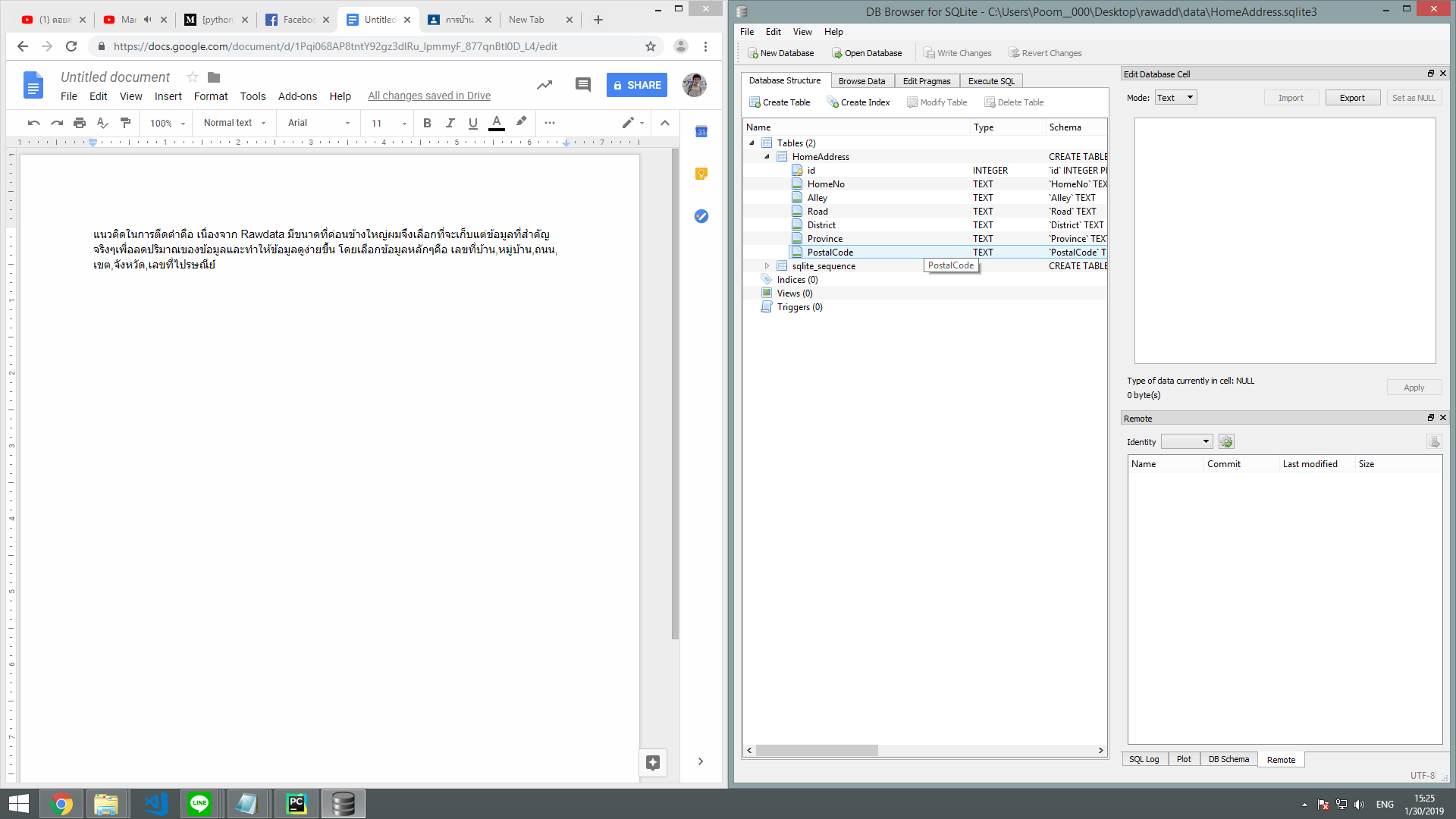
1.แนวคิดในการตีดคำคือ เนื่องจาก Rawdata มีขนาดที่ค่อนข้างใหญ่ผมจึงเลือกที่จะเก็บแต่ข้อมูลที่สำคัญจริงๆเพื่อลดปริมาณของข้อมูลและทำให้ข้อมูลดูง่ายขึ้น โดยเลือกข้อมูลหลักๆคือ เลขที่บ้าน,หมู่บ้าน,ถนน,

เขต,จังหวัด,เลขที่ไปรษณีย์



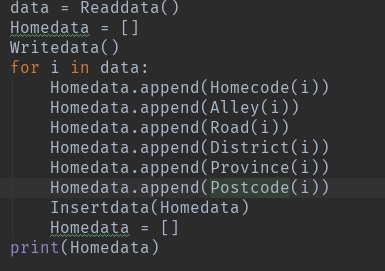
เนื่องจากว่าการ Procress ของข้อมูลมีเวลาการทำงานค่อนข้าวนานผมจึงตัดข้อมูลบางอย่างออกไปจากระบบ เช่น Data อาคาร

2.ขั้นตอนการทำงาน

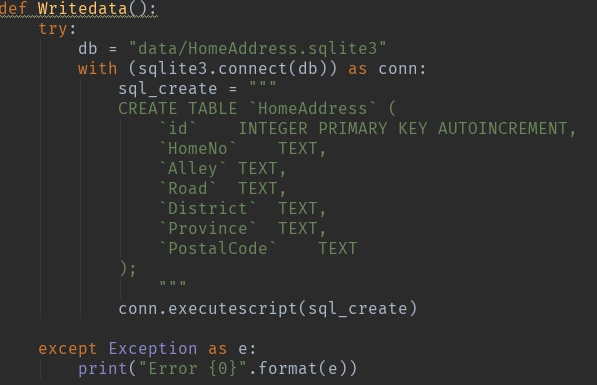
2.1 เริ่มจาก อ่านไฟล์จาก rawaddress ก่อนและนำมาเก็บใน string



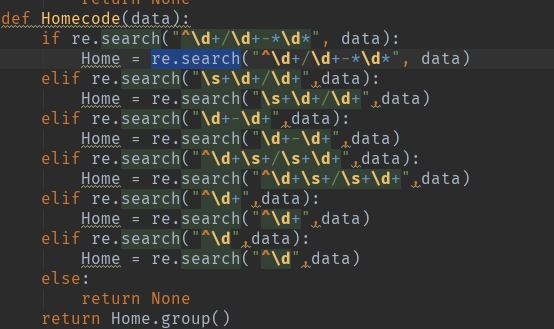
2.2 และทำงานตามระดับดังนี้



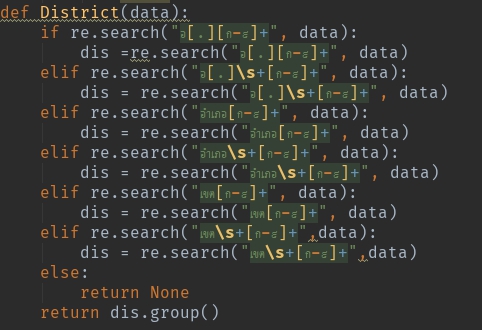
2.3 ก่อนหน้านั้นทำการสร้าง ฐานข้อมูล เตรียมเอาไว้ก่อนดังนี้



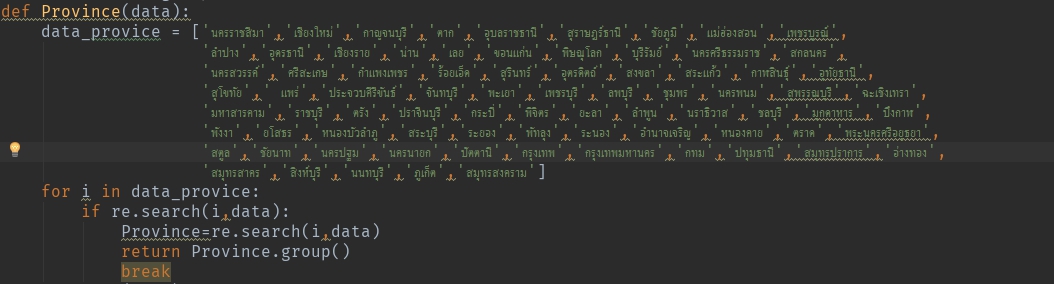
2.4 หลังจากนั้นเราทำการตัดคำต่างๆโดยใช้ re.search เข้ามาช่วยเพื่อให้ตัดคำได้แม่นยำมากขึ้นโดยมีการทำงานคร่าวๆดังนี้



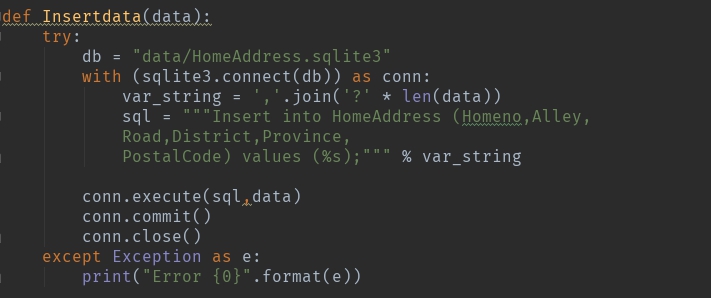
โดยการทำงานต้ดคำของ Homecode นั้นจะทำการหาโดยการหาตัวเลขข้างหน้า ข้างหลังและตรงกลางตามรูป

ตัวอย่าง ที่ 2 การหา เขต 

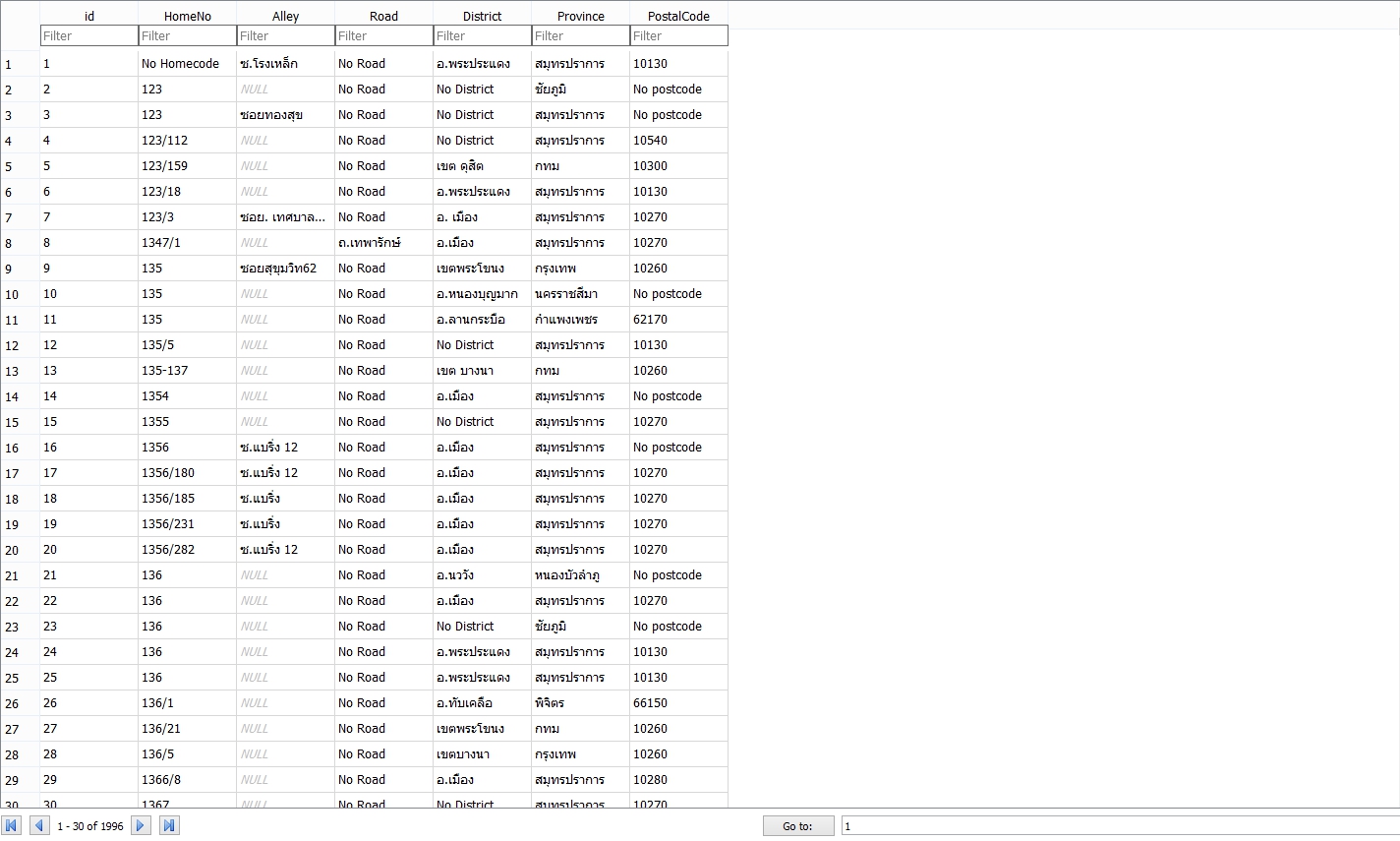
จากภาพจะเห็นได้ว่าจะมีการเช็คค่าทั้งตัวย่อ อ. อำเภอ หรือ เขต เพื่อที่จะรวมกลุ่มของตัวที่เกี่ยวข้องกับเขตทั้งหมดมาไว้ในตัวแปรเดียว

ตัวอย่างที่ 3 การหาจังหวัด เราจะทำการเก็บข้อมูลของ 77 จังหวัดไว้ และนำมา search ในข้อมูลทั้งหมดถ้าเจอแล้วให้หยุดการทำงานเพื่อที่จะลดระยะเวลาในการหา

หลังจากนั้นเราจะทำการ เก็บข้อมูลลงฐานข้อมูล



เก็บข้อมูลทั้งหมดมาไว้ใน table ที่เราได้สร้างไว้ก่อหน้านี้โดยข้อมูลที่ได้จะมีดังนี้



จะเห็นได้ว่าข้อมูลที่ data นั้นหาไม่พบจะมีค่าเป็น null เพื่อง่ายต่อการใช้งานต่อไป

แหล่งอ้างอิงการทำงานของโปรแกรมครับ

<https://medium.com/@minatonamikazerak/python-ลองเล่น-regular-expression-aeffdb07c794>

นาย วีรภัทร บุณยสิงห์ 58121098-6