Extract

เริ่มต้นด้วยการอ่านไฟล์ csv ด้วย pandas และหา Encoding ด้วย chardet ไม่งั้นจะขึ้นเป็นภาษาต่างดาว

Transform

แปลงข้อมูลด้วยการกำหนดคอลัมน์ของข้อมูลที่ต้องการ คือ LATITUDE, LONGITUDE, DATE,TIME, CONFIDENCE, TUMBOON, AUMPER, PROVINCE, NAME

ทำการกรองเอาเฉพาะคอลัมน์ NAME = เกษตกรรม และ มีค่า CONFIDENCE > 50

ทำการเปลี่ยนข้อมูลวันที่และเวลา เป็น yyyy-MM-dd hh:mm:ss เพื่อให้ใช้งานกับ mysql ได้

Load

นำข้อมูลที่แปลงแล้วเข้า sql เริ่มจากสร้างฐานข้อมูลด้วย script ให้ตรงกับข้อมูลที่กรองมาแล้ว

db\_cursor.execute("CREATE DATABASE Ex02")

db\_cursor.execute("USE Ex02")

db\_cursor.execute("""

    CREATE TABLE Hotspot\_Data (

        LATITUDE FLOAT,

        LONGITUDE FLOAT,

        DATE DATE,

        TIME TIME,

        CONFIDENCE INT,

        TUMBOON VARCHAR(255),

        AUMPER VARCHAR(255),

        PROVINCE VARCHAR(255),

        NAME VARCHAR(255),

        IMPORTED\_DATETIME TIMESTAMP

    )

""")

แล้วต่อมาทำการนำข้อมูลจาก Dataframe เข้า mysql ด้วยการ loop แต่ละ row และทำการจัดเก็บวันที่และเวลานำเข้าด้วย Timestamp

for i, row in filter.iterrows():

    db\_cursor.execute("""

        INSERT INTO Hotspot\_Data (LATITUDE, LONGITUDE, DATE, TIME, CONFIDENCE, TUMBOON, AUMPER, PROVINCE, NAME, IMPORTED\_DATETIME)

        VALUES (%s, %s, %s, %s, %s, %s, %s, %s, %s, CURRENT\_TIMESTAMP)

    """, (row['LATITUDE'], row['LONGITUDE'], row['DATE'], row['TIME'], row['CONFIDENCE'], row['TUMBOON'], row['AUMPER'], row['PROVINCE'], row['NAME']))

ได้ผลลัพธ์ดังนี้

