Question 3

A. Given a CSV file on this link as the data source, please create a python script to reformat the data to JSON files with output like this:

Code:

```
import pandas as pd
import datetime
import re

# Path of the input CSV file
datapath = r"driver_registration.csv"

# Path of the output JSON file
datatarget = r"driver_registration.json"

# Assign column names to the data and read the CSV file
colnames = ['id',
'date_created',
'date_last_modified',
'active_date',
'name',
'phone',
'resign_date',
'resign_reason',
'status',
'tipe',
'area',
"CONCAT('operator_',id)",
'modified_by',
'vehicle_type',
'helmet_qty',
'jacket_qty',
'vehicle_brand',
'vehicle_year',
'bike_type',
'first_ride_bonus_awarded',
'is_doc_completed']
data = pd.read_csv(datapath, names=colnames, header=None)

# Function reformat_date():
```

```
def reformat_date(date_column):
    new_date = []

    for x in date_column:
        date_split = re.split('[- :]', x)
        date_prep = []

        if len(date_split) == 6:
            for i in date_split:
                date_prep.append(int(i))
            date_reformat = datetime.datetime(date_prep[0], date_prep[1], date_prep[2], date_prep[3],
date_prep[4], date_prep[5])
            date_reformatted = date_reformat.strftime('%Y-%m-%dT%H:%M:%S.000Z')
            new_date.append(date_reformatted)
        elif len(date_split) == 3:
            for i in date_split:
                date_prep.append(int(i))
            date_reformat = datetime.datetime(date_prep[0], date_prep[1], date_prep[2])
            date_reformatted = date_reformat.strftime('%Y-%m-%dT%H:%M:%S.000Z')
            new_date.append(date_reformatted)

    date_column = new_date
    return date_column

# Take all date columns and reformat
data['date_created'] = reformat_date(data['date_created'])
data['date_last_modified'] = reformat_date(data['date_last_modified'])
data['active_date'] = reformat_date(data['active_date'])

# Take all phone numbers and prepend a '+' string
data['phone'] = "+" + data['phone'].astype(str)

# Take all first_ride_bonus_awarded values and prepend a '+' string
data['first_ride_bonus_awarded'] = data['first_ride_bonus_awarded'].replace({'\\0' : '\u0000'})

# Format data into a JSON file
data.to_json(datatarget, indent = 2, orient = 'records')
```