**Question 3**

1. 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')