import csv

from faker import Faker

import datetime

import random

import math

def datagenerate(records, headers):

fake = Faker('en\_US')

fake1 = Faker('en\_GB') # To generate phone numbers

with open("cars\_newest62.csv", 'wt') as csvFile:

writer = csv.DictWriter(csvFile, fieldnames=headers)

writer.writeheader()

less=0

medium = 0

high = 0

for i in range(records):

date\_today = fake.date\_between(start\_date="-30y",end\_date="today")

year\_now = date\_today.year

miles = random.randint(13000,390000)

condition = random.choice(['very bad', 'fair', 'good', 'excellent'])

paint = random.choice(['red','blue','black','white'])

tube = random.choice(['tubed','tubeless'])

make = random.choice(['Honda','Hyundai','Audi','Mercedes-Benz','MiniCooper','Toyota'])

fuel = random.choice(['petrol','diesel'])

pay = random.choice(['cash','card'])

identity = fake.isbn10(separator = '-')

if condition is "very bad":

condition\_prob = 4

elif condition is "fair":

condition\_prob = 6

elif condition is "good":

condition\_prob = 15

else:

condition\_prob = 75

if paint is "blue":

paint\_prob = 20

elif paint is "red":

paint\_prob = 20

elif paint is "black":

paint\_prob = 30

else:

paint\_prob = 30

if make is "Honda":

make\_prob = 5

elif make is "Hyundai":

make\_prob = 5

elif make is "Audi":

make\_prob = 40

elif make is "Mercedes-Benz":

make\_prob = 30

elif make is "Minicooper":

make\_prob = 10

else:

make\_prob = 10

if fuel is "diesel":

fuel\_prob = 40

else:

fuel\_prob = 60

if tube is "tubed":

tube\_prob = 40

else:

tube\_prob = 60

if int(year\_now) in range(1989,2001):

year\_prob = 10

elif int(year\_now) in range(2001,2011):

year\_prob = 20

else:

year\_prob = 70

if int(miles) in range(13000,130000):

miles\_prob = 60

elif int(miles) in range(130000,195000):

miles\_prob = 35

else:

miles\_prob = 5

#weight = (5\*(miles\_prob + condition\_prob + year\_prob)) + (3\*(fuel\_prob + make\_prob)) + ((paint\_prob +tube\_prob)/10) #Simple Fn

#weight = (5\*(miles\_prob + condition\_prob + year\_prob)) + (3\*(fuel\_prob + make\_prob)) + ((paint\_prob +tube\_prob)/10)

#weight = ((miles\_prob + condition\_prob + year\_prob)\*\*3) + ((fuel\_prob + make\_prob)\*\*2) + ((paint\_prob +tube\_prob)) #Moderate Fn

#weight = math.ceil(weight/1000)

weight = ((miles\_prob + condition\_prob + year\_prob)\*\*4) + (((fuel\_prob + make\_prob)\*\*3)) + (paint\_prob +tube\_prob)\*\*2 #Complex Fn

weight = weight/10000

if(int(weight) in range(0, 10000)):

label = "Less than 20K SGD"

less = less +1

elif(int(weight) in range(10000, 33000)):

label = "Between 20K SGD and 50K SGD"

medium = medium +1

else:

label = "Greater than 50K SGD"

high = high +1

if ((not(label == "Less than 20K SGD" and less>=5000)) and (not(label == "Between 20K SGD and 50K SGD" and medium>=5000)) and (not(label == "Greater than 50K SGD" and high>=5000)) ):

writer.writerow({

"VIN" : identity,

"miles\_traveled": miles,

"vehicle\_condition" : condition,

"colour" : paint,

"tyre\_type": tube,

"make\_year" : year\_now,

"manufacturer" : make,

"fuel\_type" : fuel,

"payment\_mode" : pay,

"market\_price" : label

})

if \_\_name\_\_ == '\_\_main\_\_':

records = 95000

headers = ["VIN", "miles\_traveled", "vehicle\_condition", "colour", "tyre\_type", "make\_year",

"manufacturer", "fuel\_type","payment\_mode","market\_price"]

datagenerate(records, headers)

print("CSV generation complete!")