import csv

from faker import Faker

import datetime

import random

def datagenerate(records, headers):

fake = Faker('en\_US')

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

with open("cars\_newest17.csv", 'wt') as csvFile:

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

writer.writeheader()

less=0

medium = 0

high = 0

for i in range(records):

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

year = date.year

miles = random.randint(1000,5000)

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 = 0.05

elif condition is "fair":

condition\_prob = 0.05

elif condition is "good":

condition\_prob = 0.20

else:

condition\_prob = 0.70

if paint is "blue":

paint\_prob = 0.2

elif paint is "red":

paint\_prob = 0.2

elif paint is "black":

paint\_prob = 0.3

else:

paint\_prob = 0.3

if make is "Honda":

make\_prob = 0.05

elif make is "Hyundai":

make\_prob = 0.05

elif make is "Audi":

make\_prob = 0.40

elif make is "Mercedes-Benz":

make\_prob = 0.30

elif make is "Minicooper":

make\_prob = 0.10

else:

make\_prob = 0.10

if fuel is "diesel":

fuel\_prob = 0.4

else:

fuel\_prob = 0.6

if pay is "cash":

pay\_prob = 0.2

else:

pay\_prob = 0.4

if tube is "tubed":

tube\_prob = 0.4

else:

tube\_prob = 0.6

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

year\_prob = 0.10

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

year\_prob = 0.2

else:

year\_prob = 0.7

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

miles\_prob = 0.70

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

miles\_prob = 0.25

else:

miles\_prob = 0.05

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

if(int(weight) in range(0, 3)):

label = "<=20K SGD"

less = less +1

elif(int(weight) in range(3, 6)):

label = ">20KSGD and <=30K SGD"

medium = medium +1

else:

label = "50K SGD"

high = high +1

if ((not(label == "<=20K SGD" and less>=1000)) and (not(label == ">20KSGD and <=30K SGD" and medium>=1000)) and (not(label == "50K SGD" and high>=1000)) ):

writer.writerow({

"VIN" : identity,

"miles\_traveled": miles,

"vehicle\_condition" : condition,

"colour" : paint,

"tyre\_type": tube,

"make\_year" : year,

"manufacturer" : make,

"fuel\_type" : fuel,

"market\_price" : pay,

"payment\_mode" : pay,

"market\_price" : label

})

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

records = 60000

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

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

datagenerate(records, headers)

print("CSV generation complete!")