### Create pace variable

import csv

import pandas as pd

def tpstmp(csvfile):

b=csv.reader(csvfile)

next(b)

tst=[]

tend=[]

pace=[]

for row in b:

if row[2]=='' or row[3]=='' or row[4]=='' or row[5]=='' or float(row[4])==0.0 or float(row[5])==0.0 or float(row[4])>1000.0 or (float(row[5])/(float(row[4])/3600.0))>70.0 or row[8]=='' or row[9]=='':

continue

else:

pace.append((float(row[4])/3600.0)/(float(row[5])))

tst.append((row[2]))

tend.append((row[3]))

return pace,tst,tend

def comar(csvfile):

b=csv.reader(csvfile)

next(b)

pcom=[]

dcom=[]

for row in b:

if row[2]=='' or row[3]=='' or row[4]=='' or row[5]=='' or float(row[4])==0.0 or float(row[5])==0.0 or float(row[4])>1000.0 or (float(row[5])/(float(row[4])/3600.0))>70.0 or row[8]=='' or row[9]=='':

continue

else:

pcom.append(int(row[8]))

dcom.append(int(row[9]))

return pcom,dcom

with open("C:/Users/Akshay/Desktop/Project/Taxi\_Trips.csv",encoding='utf-8',newline='') as csvfile:

pace,tst,tend=tpstmp(csvfile)

"""d = {'Start time stamp':tst, 'End time stamp': tend,'Pace':pace}

df = pd.DataFrame(d)"""

pcom,dcom=comar(csvfile)

### To convert data to datetime format and then sort as per Start time stamp

d = {'Start time stamp':tst, 'End time stamp': tend,'Pace':pace, 'Start Community Area':pcom, 'End Community Area': dcom}

df = pd.DataFrame(d)

df['Start time stamp'] = pd.to\_datetime(df['Start time stamp'])

df.index = df['Start time stamp']

del df['Start time stamp']

df=df.sort\_index()

### To extract first week of data

df\_week1=df[0:113856]