import pandas as pd

import numpy as np

#Read files:

train = pd.read\_csv("C:\\Users\\unic\\Downloads\\train.csv")

test = pd.read\_csv("C:\\Users\\unic\\Downloads\\test.csv")

train['source']='train'

test['source']='test'

data = pd.concat([train, test],ignore\_index=True)

print(" train.shape, test.shape, data.shape")

data.apply(lambda x: sum(x.isnull()))

data.describe()

data.apply(lambda x: len(x.unique()))

#Filter categorical variables

categorical\_columns = [x for x in data.dtypes.index if data.dtypes[x]=='object']

#Exclude ID cols and source:

categorical\_columns = [x for x in categorical\_columns if x not in ['Name','Ticket','Fare']]

#Print frequency of categories

for col in categorical\_columns:

print ('\nFrequency of Categories for varible %s'%col)

print(" data[col].value\_counts()")

#Determine the average weight per item:

item\_avg\_weight = data.pivot\_table(values='source', index='Ticket')

#Get a boolean variable specifying missing Item\_Weight values

miss\_bool = data['source'].isnull()

#Impute data and check #missing values before and after imputation to confirm

print ('Orignal #missing: %d'% sum(miss\_bool))

data.loc[miss\_bool,'Name'] = data.loc[miss\_bool,'Ticket'].apply(lambda x: item\_avg\_weight[x])

print ('Final #missing: %d'% sum(data['Fare'].isnull()))