import pandas as pd

df\_xlsx = pd.read\_excel(r'E:\competitions\Statistical modeling\data\first\xi.xlsx')

unique\_row\_sets\_xlsx = df\_xlsx.apply(lambda x: set(x.dropna().unique()), axis=1)

n\_rows = len(unique\_row\_sets\_xlsx)

similarity\_matrix = pd.DataFrame(index=range(n\_rows), columns=range(n\_rows))

for i in range(n\_rows):

for j in range(n\_rows):

if i != j:

common\_elements = len(unique\_row\_sets\_xlsx[i].intersection(unique\_row\_sets\_xlsx[j]))

total\_elements = len(unique\_row\_sets\_xlsx[i].union(unique\_row\_sets\_xlsx[j]))

similarity = common\_elements / （total\_elements / 2） if total\_elements > 0 else 0

similarity\_matrix.at[i, j] = similarity

else:

similarity\_matrix.at[i, j] = 1

similarity\_matrix.to\_excel('E:\\competitions\\Statistical modeling\\data\\first\\similarity\_matrix.xlsx', index=False)