

Team members:

Sirisha Lanka - PES1201700294

Malavikka R - PES1201700794

Kevin Arulraj - PES1201700659

Assignment 3

Problem Statement:

Perform simple Extract Transform Load (ETL) functions with Python modules on two separate datasets and merge them into one.

Dataset:

Ship names #1, Ship names #2

Code:

```
import pandas as pd
a = pd.read_csv("DS1.csv")
b = pd.read_csv("DS2.csv")
d=a.rename(columns = {"name": "ship_name"})
e=b.rename(columns = {"ship": "ship_name"})
for i in d:
    if i not in e:
        for j in e:
            if j in i:
                f=e.rename(columns={j:i})
final = pd.concat([d,f],ignore_index=True)
final.to_csv('Merged.csv')
```

Results:

Merged.csv

	date	ship_name	nationality	tonnage (grt)	fate
0	07-Oct-41	Svend Foyn	United Kingdom	14795	Damaged
1	16-Feb-42	Monagas	Venezuela	2650	Sunk
2	16-Feb-42	San Nicholas	United Kingdom	2391	Sunk
3	16-Feb-42	Tia Juana	United Kingdom	2395	Sunk
4	22-Feb-42	J.N.Pew	United States	9033	Sunk
5	23-Feb-42	Sun	United States	9002	Damaged
6	23-Feb-42	Thalia	Panama	8329	Sunk
7	11-May-42	Cape of Good Hope	United Kingdom	4963	Sunk
8	24-May-42	Gonçalves Dias	Brazil	4996	Sunk
9	28-May-42	Alcoa Pilgrim	United States	6759	Sunk
10	03-Jun-42	M.F. Elliot	United States	6940	Sunk
11	09-Jun-42	Bruxelles	Belgium	5085	Sunk
12	09-Jun-42	Franklin K. Lane	United States	6589	Sunk
13	15-Jun-42	Cold Harbor	Panama	5010	Sunk
14	15-Jun-42	Scottsburg	United States	8010	Sunk
15	15-Jun-42	West Hardaway	United States	5702	Sunk
16	19-Aug-41	SS Aguila	United Kingdom	3255	Sunk
17	27-Nov-41	HMAS Parramatta	Royal Australian Navy	1060	Sunk
18	23-Dec-41	SS Shuntien	United Kingdom	3059	Sunk
19	26-Dec-41	SS Warszawa	Poland	2487	Sunk
20	10-Jun-42	MV Athene	Norway	4681	Sunk
21	10-Jun-42	SS Brambleleaf	United Kingdom	5917	Damaged