In [ ]:
```
import pandas as pd
# !pip install tabula-py
from tabula import read_pdf

df = read_pdf("./international-schools-in-federal-loan-programs.pdf", pages="all")
# got the file from https://studentaid.gov/sites/default/files/international-schools-i
```

In [ ]:
```
len(df)
# combine all of these into one dataframe
df = pd.concat(df)
df
```

Out[ ]:

	OPEID	School Name	City	Country	Status
0	4240100	American University of Antigua College of Medi...	Coolidge	ANTIGUA	Eligible
1	4182900	Universidad Blas Pascal	Cordoba	ARGENTINA	Deferment Only
2	4131000	American University of Armenia	Yerevan	ARMENIA	Deferment Only
3	3136300	Australian Catholic University	North Sydney	AUSTRALIA	Deferment Only
4	1058800	Australian National University	Canberra	AUSTRALIA	Eligible
...	...	...	...	...	...
50	1015700	Cardiff University	Cardiff	WALES	Eligible
51	3478300	Royal Welsh College of Music and Drama	Cardiff	WALES	Eligible
52	858600	Swansea University	Swansea	WALES	Eligible
53	3547300	University of South Wales	Pontypridd	WALES	Eligible
54	1234100	University of Wales Trinity Saint David	Carmarthen	WALES	Eligible

758 rows × 5 columns

In [ ]:
```
# drop rows that are just deferment only
df = df[df["Status"] != "Deferment Only"]
# drop opeid and city
df = df.drop(columns=["OPEID", "City"])
```

In [ ]:
```
# i only want to study specific countries in the world, which are

countries_of_interest = ["england", "germany", "japan", "the netherlands", "switzerlar

df = df[df["Country"].str.lower().isin(countries_of_interest)]
```

In [ ]:
```
# upon checking the resulting dataframe, Ox and Cambridge have all of their individual

# if the row contains "Oxford" or "Cambridge", then just drop them
df = df[~df["School Name"].str.contains("University of Oxford -")]
df = df[~df["School Name"].str.contains("University of Cambridge -")]
```

```
# upon checking, there's some unis with certain words that do not have CS programs, so

some_words = ["Medicine", "Veterinary","Seminary", "Conservatoire", "Hotel", "Chiropra

for word in some_words:
    df = df[~df["School Name"].str.contains(word)]
```

In [ ]:
```
df["Country"].value_counts()
```

Out[ ]:
```
ENGLAND             117
CANADA               54
SCOTLAND             16
FRANCE                7
GERMANY               4
THE NETHERLANDS       4
SWITZERLAND           2
Name: Country, dtype: int64
```

In [ ]:
```
df
```

Out[ ]:

	School Name	Country	Status
49	Acadia University	CANADA	Eligible
51	Ambrose University	CANADA	Eligible
53	Bishop's University	CANADA	Eligible
54	Brandon University	CANADA	Eligible
56	Brock University	CANADA	Eligible
...	...	...	...
34	Graduate Institute of International and Develo...	SWITZERLAND	Eligible
43	Universiteit Maastricht	THE NETHERLANDS	Eligible
44	Universiteit Utrecht	THE NETHERLANDS	Eligible
45	Universiteit Van Amsterdam	THE NETHERLANDS	Eligible
46	Vrije Universiteit Amsterdam	THE NETHERLANDS	Eligible

204 rows × 3 columns

In [ ]:
```
df.to_csv("international-schools.csv", index=False)
```