

# Poking around the `titanic` dataset

## First in Python

```
import requests
from pathlib import Path

def download(url, output_file):
    if not Path(output_file).exists():
        r = requests.get(url)
        with open(output_file, "wb") as f:
            f.write(r.content)
    print("Download finished")

download("https://web.stanford.edu/class/archive/cs/cs109/cs109.1166/stuff/titanic.csv",
"examples/titanic.csv")
```

Download finished

```
import pandas as pd
df = pd.read_csv("examples/titanic.csv")
df["Name"] = df["Name"].str.split(" ").str.slice(0, 3).str.join(" ")
df = df.drop(df.filter(like="Aboard", axis=1).columns, axis=1)

print(df.head(5))
```

	Survived	Pclass	Name	Sex	Age	Fare
0	0	3	Mr. Owen Harris	male	22.0	7.2500
1	1	1	Mrs. John Bradley	female	38.0	71.2833
2	1	3	Miss. Laina Heikkinen	female	26.0	7.9250
3	1	1	Mrs. Jacques Heath	female	35.0	53.1000
4	0	3	Mr. William Henry	male	35.0	8.0500

Here's some `inline` text