Planet Data Collection

Using the Open Exoplanet Catalogue database:
https://github.com/OpenExoplanetCatalogue/open_exoplanet_catalogue/

Data License

Follow instructions to get the xml file

```
import xml.etree.ElementTree as ET, urllib.request, gzip, io
url =
"https://github.com/OpenExoplanetCatalogue/oec_gzip/raw/master/systems
.xml.gz"
oec =
ET.parse(gzip.GzipFile(fileobj=io.BytesIO(urllib.request.urlopen(url).
read())))
```

Parse into Pandas DataFrame

Information on what each field means can be found here.

```
import pandas as pd

def parse(base):
    db = oec.findall(f".//{base}")

    exclude = ['star', 'videolink', 'binary'] if base in ['system',
```

```
'binary'] else ['planet']

    columns = set([attribute.tag for attribute in db[0] if
attribute.tag not in exclude])
    results = pd.DataFrame(columns=columns)

    for entry in db:
        data = {col : entry.findtext(col) for col in columns}
        if base in ['system', 'binary']:
            data['binaries'] = len(entry.findall('.//binary'))
            data['stars'] = len(entry.findall('.//star'))
        if base in ['system', 'star', 'binary']:
            data['planets'] = len(entry.findall('.//planet'))
        results = results.append(data, ignore_index=True)

    return results
```

Parse planet data

```
planets = parse('planet')
planets.head()
```

```
  lastupdate                                          description
discoveryyear  \
0   15/09/20  11 Com b is a brown dwarf-mass companion to th...
2008
1   15/09/20  11 Ursae Minoris is a star located in the cons...
2009
2   15/09/20  14 Andromedae is an evolved star in the conste...
2008
3   15/09/21  The star 14 Herculis is only 59 light years aw...
2002
4   15/09/21  14 Her c is the second companion in the system...
2006

  eccentricity periastrontime discoverymethod   mass semimajoraxis
period  \
0        0.231      2452899.6              RV   19.4          1.29
326.03
1         0.08     2452861.04              RV  11.20          1.54
516.22
2            0      2452861.4              RV    4.8          0.83
185.84
3        0.359           None              RV  4.975         2.864
1766
4        0.184           None              RV  7.679         9.037
9886

       name                list periastron
0  11 Com b  Confirmed planets       94.8
```

```
1  11 UMi b  Confirmed planets     117.63
2  14 And b  Confirmed planets          0
3  14 Her b  Confirmed planets     22.230
4  14 Her c      Controversial    189.076
```

Parse system data

```
systems = parse('system')
systems.head()
```

```
   rightascension     declination distance   constellation
name  \
0   12 20 43.0255  +17 47 34.3392     88.9  Coma Berenices    11 Com

1  15 17 05.88899  +71 49 26.0466    122.1      Ursa Minor    11 UMi

2  23 31 17.41346  +39 14 10.3092     79.2       Andromeda    14 And

3   16 10 24.3152  +43 49 03.4987     18.1        Hercules    14 Her

4  19 41 48.95343  +50 31 30.2153   21.146          Cygnus  16 Cygni


   binaries  planets  stars
0       0.0      1.0    1.0
1       0.0      1.0    1.0
2       0.0      1.0    1.0
3       0.0      2.0    1.0
4       2.0      1.0    3.0
```

Parse binary data

```
binaries = parse('binary')
binaries.head()
```

```
  separation          name positionangle  binaries  planets  stars
0      39.56      16 Cygni        133.30       1.0      1.0    3.0
1        3.4   16 Cygni AC           209       0.0      0.0    2.0
2      12.37   2M0441+2301         237.3       1.0      1.0    3.0
3     0.2323     2M 044145         79.61       0.0      0.0    2.0
4       None  2M 1938+4603          None       0.0      1.0    2.0
```

Parse star data

```
stars = parse('star')
stars.head()
```

```
    magJ   magV   magK metallicity  mass   magH temperature   magB
radius  \
0  2.943   4.74  2.282       -0.35   2.7  2.484        4742   5.74
```

```
19
1  2.876  5.024  1.939        0.04  1.80  2.091        4340  6.415
24.08
2  3.019   5.22  2.331       -0.24   2.2  2.608        4813   6.24
11
3  5.158   6.67  4.714        0.43   1.0  4.803      5311.0   7.57
0.708
4   5.09   5.95   4.43       0.096  1.11   4.72        5825   6.59
1.243
```

```
  spectraltype        name  planets
0       G8 III      11 Com      1.0
1        K4III      11 UMi      1.0
2        K0III      14 And      1.0
3         K0 V      14 Her      2.0
4          G2V  16 Cygni A      0.0
```

Save to CSVs

```
planets.to_csv('data/planets.csv', index=False)
binaries.to_csv('data/binaries.csv', index=False)
stars.to_csv('data/stars.csv', index=False)
systems.to_csv('data/systems.csv', index=False)
```