In [1]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv("https://raw.githubusercontent.com/datasciencedojo/datasets/
print(df.head())
```

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
```

In [2]:
```python
df_scores = df[['PassengerId', 'Name', 'Age', 'Fare', 'SibSp', 'Parch']]
df_scores = df_scores.drop_duplicates()
print(df_scores.head())
```

```
   PassengerId                                               Name   Age  \
0            1                            Braund, Mr. Owen Harris  22.0
1            2  Cumings, Mrs. John Bradley (Florence Briggs Th...  38.0
2            3                             Heikkinen, Miss. Laina  26.0
3            4       Futrelle, Mrs. Jacques Heath (Lily May Peel)  35.0
4            5                           Allen, Mr. William Henry  35.0

      Fare  SibSp  Parch
0   7.2500      1      0
1  71.2833      1      0
2   7.9250      0      0
3  53.1000      1      0
4   8.0500      0      0
```

In [3]:
```python
long = pd.melt(
    df_scores,
    id_vars=['PassengerId', 'Name'],
    value_vars=['Age', 'Fare', 'SibSp', 'Parch'],
    var_name='metric',
    value_name='value'
)
print(long.head())
```

```
   PassengerId                                               Name metric  \
0            1                            Braund, Mr. Owen Harris    Age
1            2  Cumings, Mrs. John Bradley (Florence Briggs Th...    Age
2            3                             Heikkinen, Miss. Laina    Age
3            4       Futrelle, Mrs. Jacques Heath (Lily May Peel)    Age
4            5                           Allen, Mr. William Henry    Age
```

```
       value
0     22.0
1     38.0
2     26.0
3     35.0
4     35.0
```

In [4]:
```python
wide = pd.pivot(
    long,
    index=['PassengerId', 'Name'],
    columns='metric',
    values='value'
)
print(wide.head())
```

```
metric                                                       Age      Fare
\
PassengerId Name
1           Braund, Mr. Owen Harris                          22.0    7.2500
2           Cumings, Mrs. John Bradley (Florence Briggs Tha... 38.0  71.2833
3           Heikkinen, Miss. Laina                           26.0    7.9250
4           Futrelle, Mrs. Jacques Heath (Lily May Peel)     35.0   53.1000
5           Allen, Mr. William Henry                         35.0    8.0500

metric                                                       Parch  SibSp
PassengerId Name
1           Braund, Mr. Owen Harris                           0.0    1.0
2           Cumings, Mrs. John Bradley (Florence Briggs Tha... 0.0    1.0
3           Heikkinen, Miss. Laina                            0.0    0.0
4           Futrelle, Mrs. Jacques Heath (Lily May Peel)      0.0    1.0
5           Allen, Mr. William Henry                          0.0    0.0
```

In [5]:
```python
print(df[['Age', 'Fare', 'SibSp', 'Parch']].aggregate(['sum', 'min', 'max']))
```

```
           Age         Fare   SibSp  Parch
sum   21205.17   28693.9493     466    340
min       0.42       0.0000       0      0
max      80.00     512.3292       8      6
```

In [6]:
```python
df.aggregate({
    'Age': ['min', 'mean', 'max'],
    'Fare': ['sum', 'mean', 'max'],
    'SibSp': ['sum', 'max'],
    'Parch': ['sum']
})
```

Out[6]:

	Age	Fare	SibSp	Parch
min	0.420000	NaN	NaN	NaN
mean	29.699118	32.204208	NaN	NaN
max	80.000000	512.329200	8.0	NaN
sum	NaN	28693.949300	466.0	340.0

In [7]:
```python
result = long.groupby('metric')['value'].agg(['sum', 'mean', 'max', 'min'])
print(result)
```

```
              sum        mean        max     min
metric
Age     21205.1700   29.699118    80.0000   0.42
```

```
Fare     28693.9493   32.204208   512.3292   0.00
Parch      340.0000    0.381594     6.0000   0.00
SibSp      466.0000    0.523008     8.0000   0.00
```

In [8]:
```python
count_result = long.groupby('metric')['value'].agg(['count'])
print(count_result)
```

```
        count
metric
Age       714
Fare      891
Parch     891
SibSp     891
```

In [9]:
```python
scaling = df.copy()

min_fare = np.min(scaling['Fare'])
max_fare = np.max(scaling['Fare'])

scaling['norm_fare'] = (scaling['Fare'] - min_fare) / (max_fare - min_fare)
print(scaling[['Fare', 'norm_fare']].head())
```
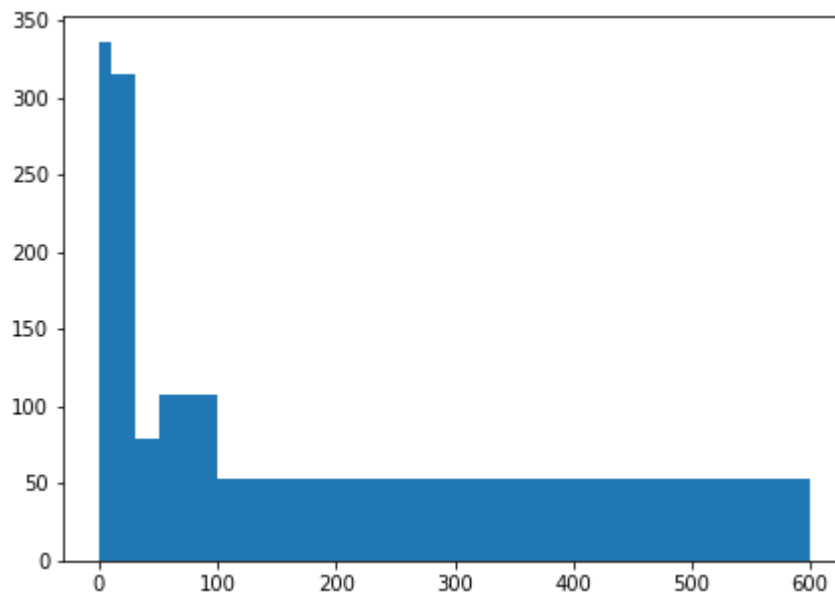
```
      Fare   norm_fare
0   7.2500    0.014151
1  71.2833    0.139136
2   7.9250    0.015469
3  53.1000    0.103644
4   8.0500    0.015713
```

In [10]:
```python
mean_fare = np.mean(scaling['Fare'])
std_fare = np.std(scaling['Fare'])

scaling['standardized_fare'] = (scaling['Fare'] - mean_fare) / std_fare
print(scaling[['Fare', 'standardized_fare']].head())
```
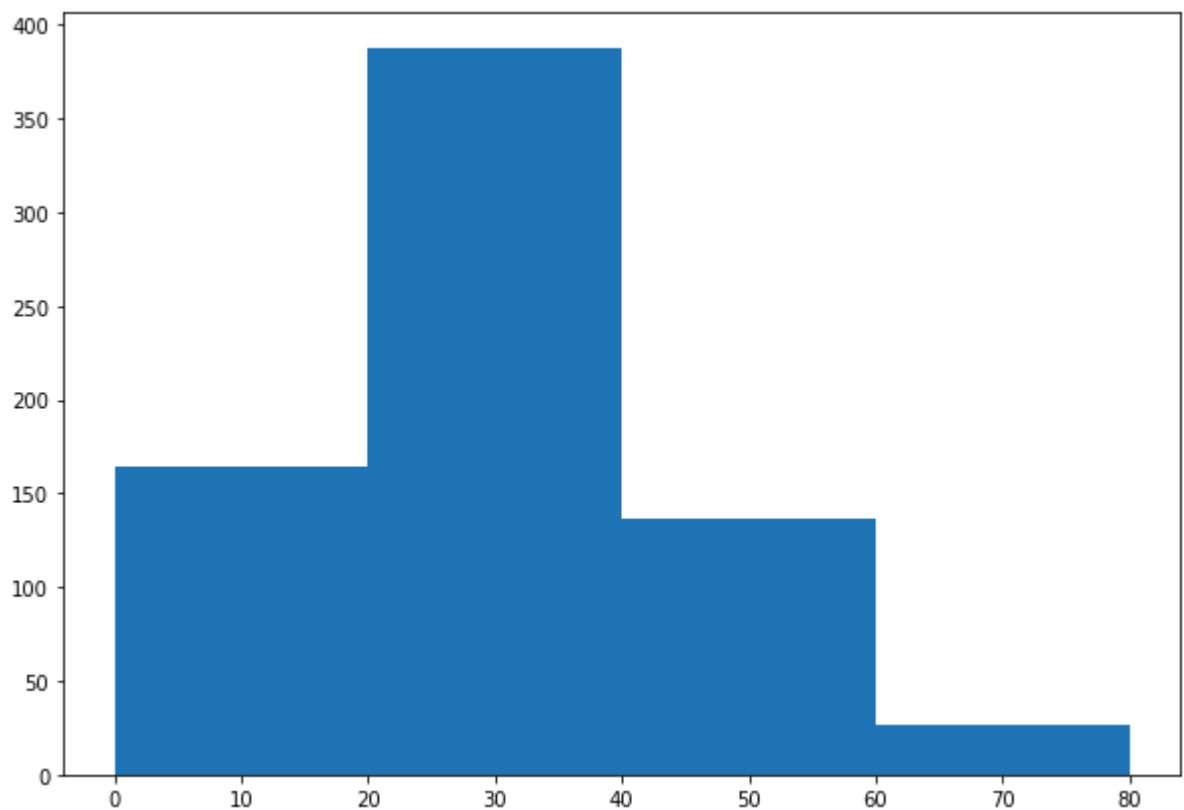
```
      Fare   standardized_fare
0   7.2500           -0.502445
1  71.2833            0.786845
2   7.9250           -0.488854
3  53.1000            0.420730
4   8.0500           -0.486337
```

In [11]:
```python
fig, ax = plt.subplots(figsize=(7,5))
ax.hist(df['Fare'], bins=[0, 10, 30, 50, 100, 600])
plt.show()
```

In [12]:
```python
a = np.array(df['Age'].dropna())

fig, ax = plt.subplots(figsize=(10,7))
ax.hist(a, bins=[0, 20, 40, 60, 80])
plt.show()
```


In [ ]: