```
In [2]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import seaborn as sns
```

```
In [3]:  sns.set(style="whitegrid")
```

```
In [4]:  df = pd.read_csv("train.csv")

         df.head()
         df.tail()
         df.shape
```

Out[4]: (891, 12)
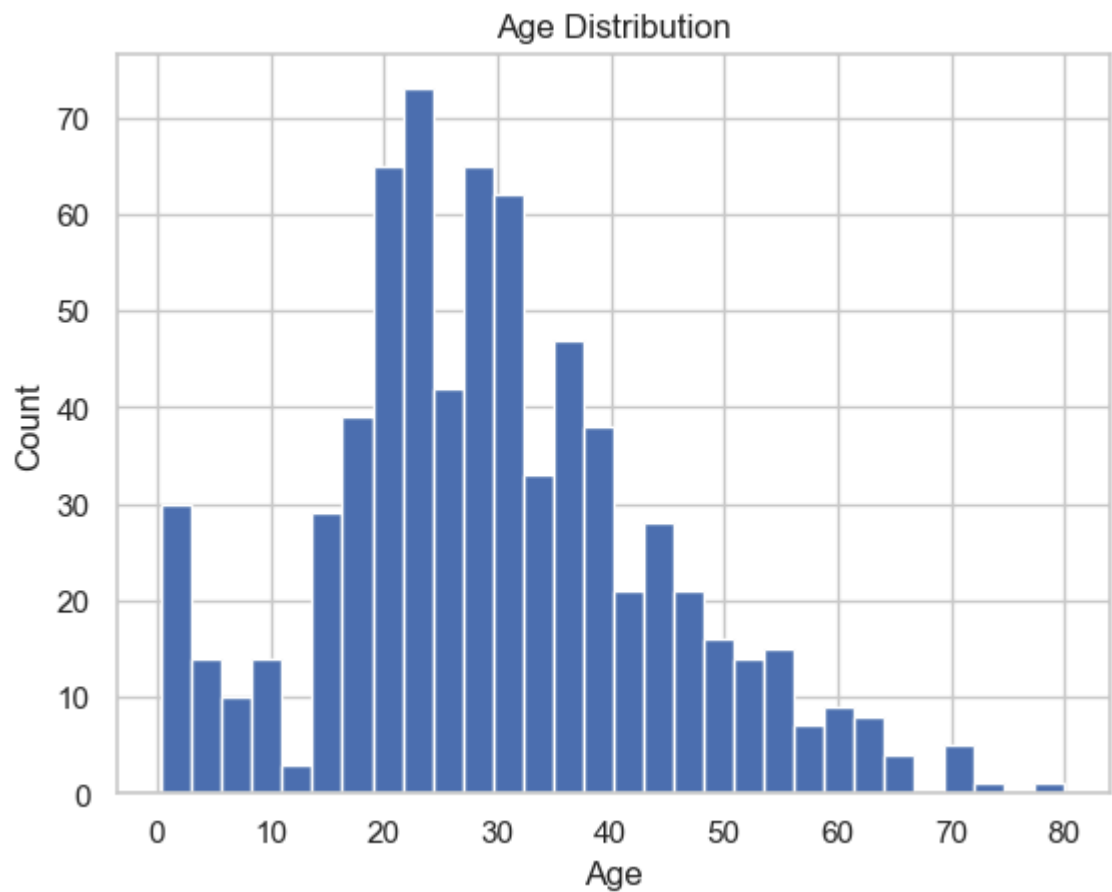
```
In [5]:  df.info()
         df.describe()
         df.isnull().sum()
         df.nunique()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   PassengerId  891 non-null     int64
 1   Survived     891 non-null     int64
 2   Pclass       891 non-null     int64
 3   Name         891 non-null     object
 4   Sex          891 non-null     object
 5   Age          714 non-null     float64
 6   SibSp        891 non-null     int64
 7   Parch        891 non-null     int64
 8   Ticket       891 non-null     object
 9   Fare         891 non-null     float64
 10  Cabin        204 non-null     object
 11  Embarked     889 non-null     object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
Out[5]:  PassengerId    891
         Survived         2
         Pclass           3
         Name           891
         Sex              2
         Age             88
         SibSp            7
         Parch            7
         Ticket         681
         Fare           248
         Cabin          147
         Embarked         3
         dtype: int64
```
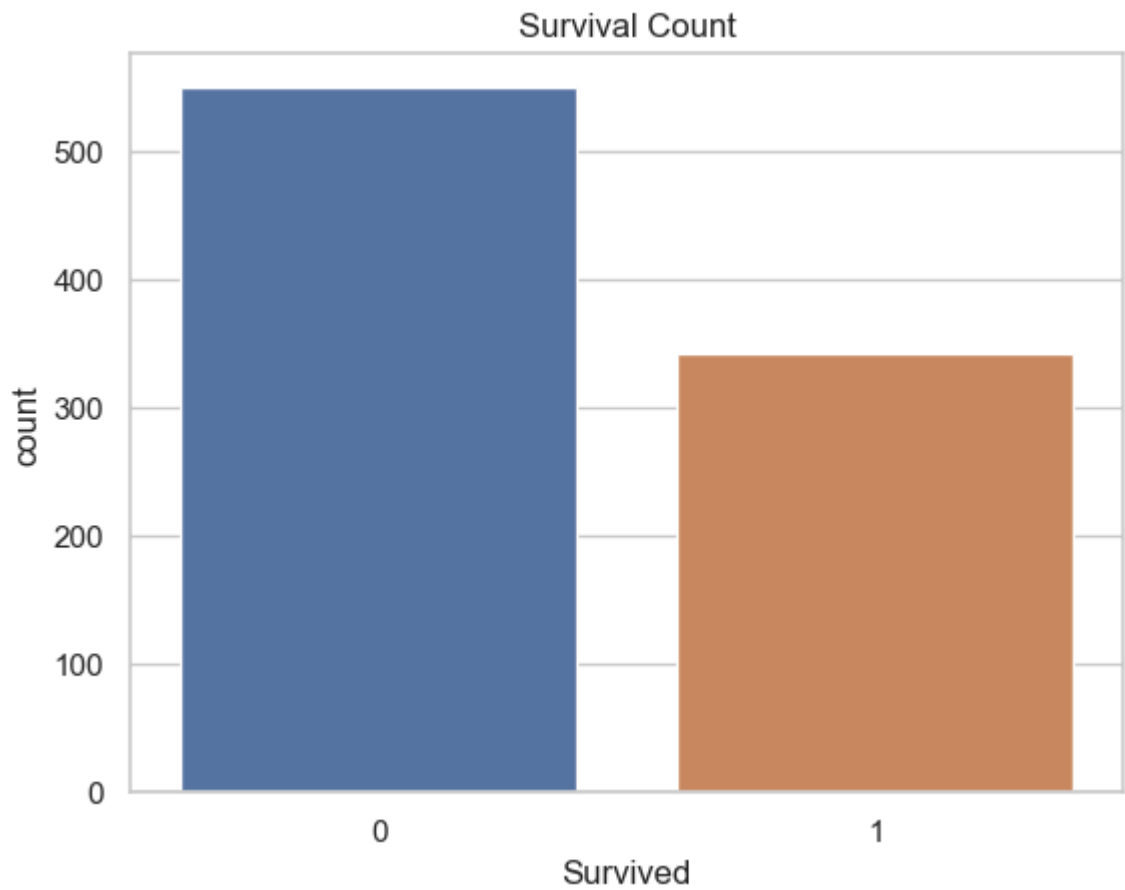
```
In [6]:  plt.hist(df['Age'].dropna(), bins=30)
         plt.xlabel("Age")
         plt.ylabel("Count")
```

```
plt.title("Age Distribution")
plt.show()
```
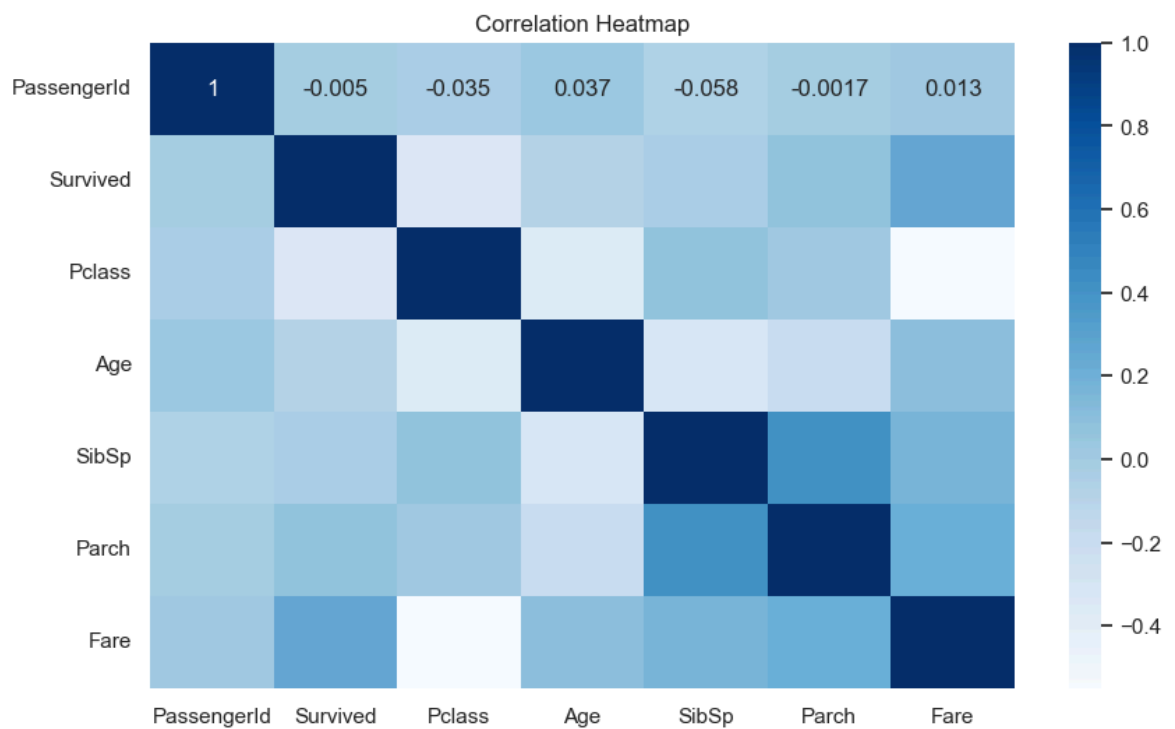

Age Distribution

```
In [7]:   sns.countplot(x="Survived", data=df)
          plt.title("Survival Count")
          plt.show()
```

Survival Count


```
In [8]:  num_df = df.select_dtypes(include=['int64', 'float64'])

         plt.figure(figsize=(10,6))
         sns.heatmap(num_df.corr(), annot=True, cmap="Blues")
         plt.title("Correlation Heatmap")
         plt.show()
```
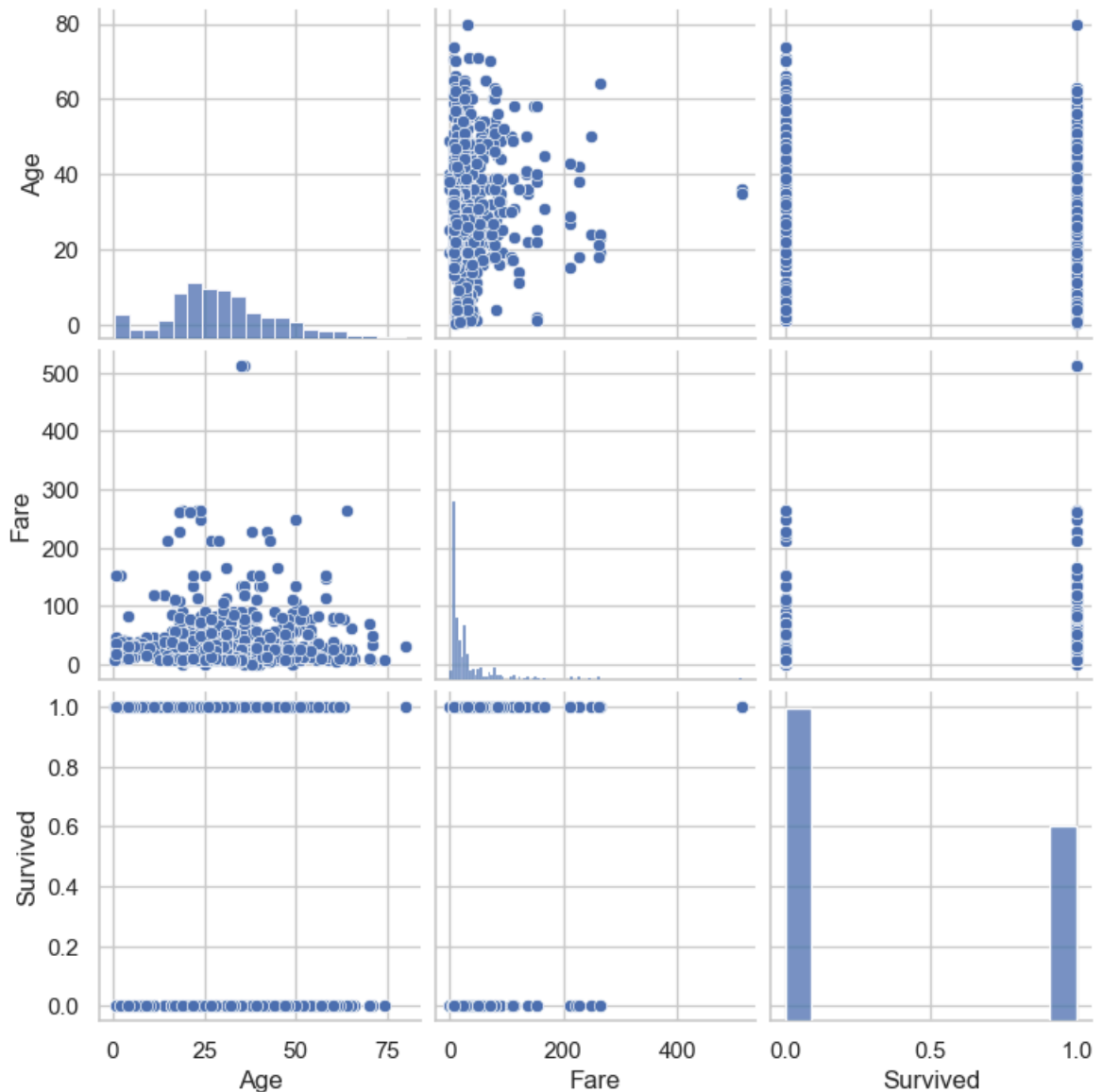


```
In [9]:  sns.pairplot(df[['Age','Fare','Survived']])
         plt.show()
```

```
C:\Users\arunk\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\arunk\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\arunk\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarnin
g: use_inf_as_na option is deprecated and will be removed in a future version. Co
nvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```
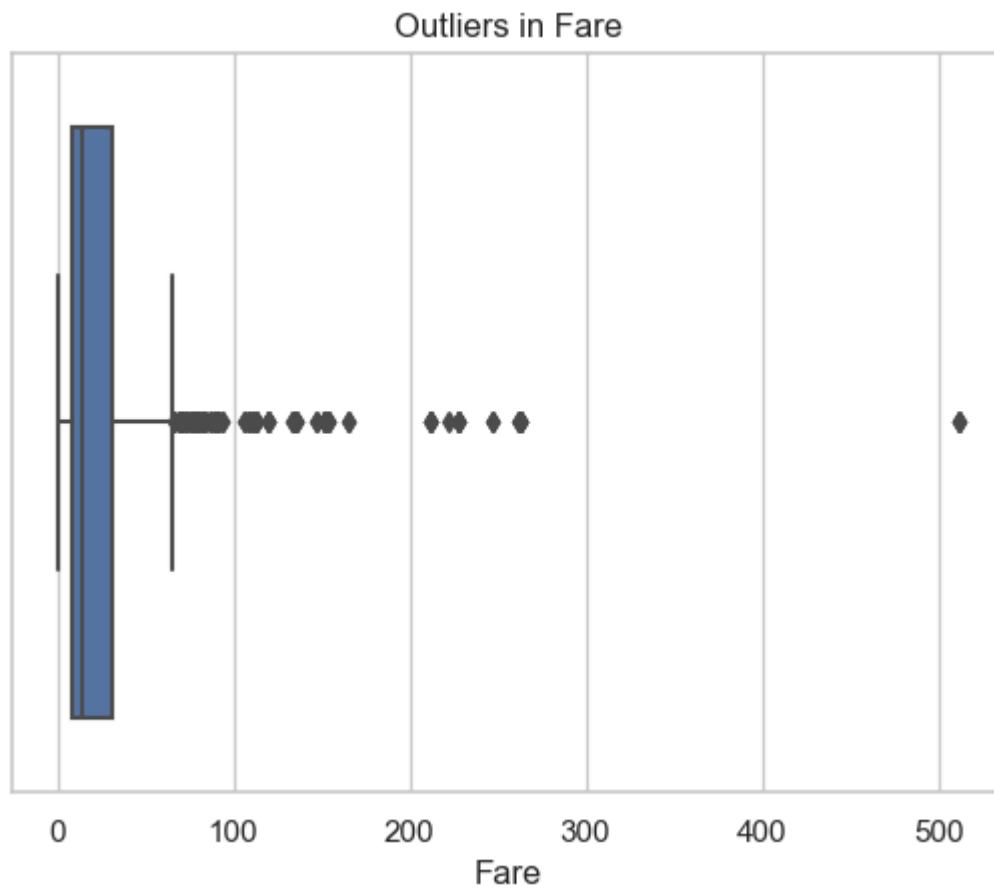


```
In [10]:  sns.boxplot(x=df["Fare"])
          plt.title("Outliers in Fare")
          plt.show()
```

Outliers in Fare

Fare

```
In [ ]:
```