In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
```

```
Matplotlib is building the font cache; this may take a moment.
```

In [2]:
```
titanic = pd.read_csv("train.csv")
```

In [3]:
```
titanic.head()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [4]:
```
titanic.info()
titanic.describe()
titanic['Sex'].value_counts()
titanic['Pclass'].value_counts()
titanic.isnull().sum()
```

```
<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[4]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [5]:
```
titanic.isnull().sum()
```

Out[5]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [6]:
```
titanic['Embarked'].fillna(titanic['Embarked'].mode()[0], inplace=True)
```

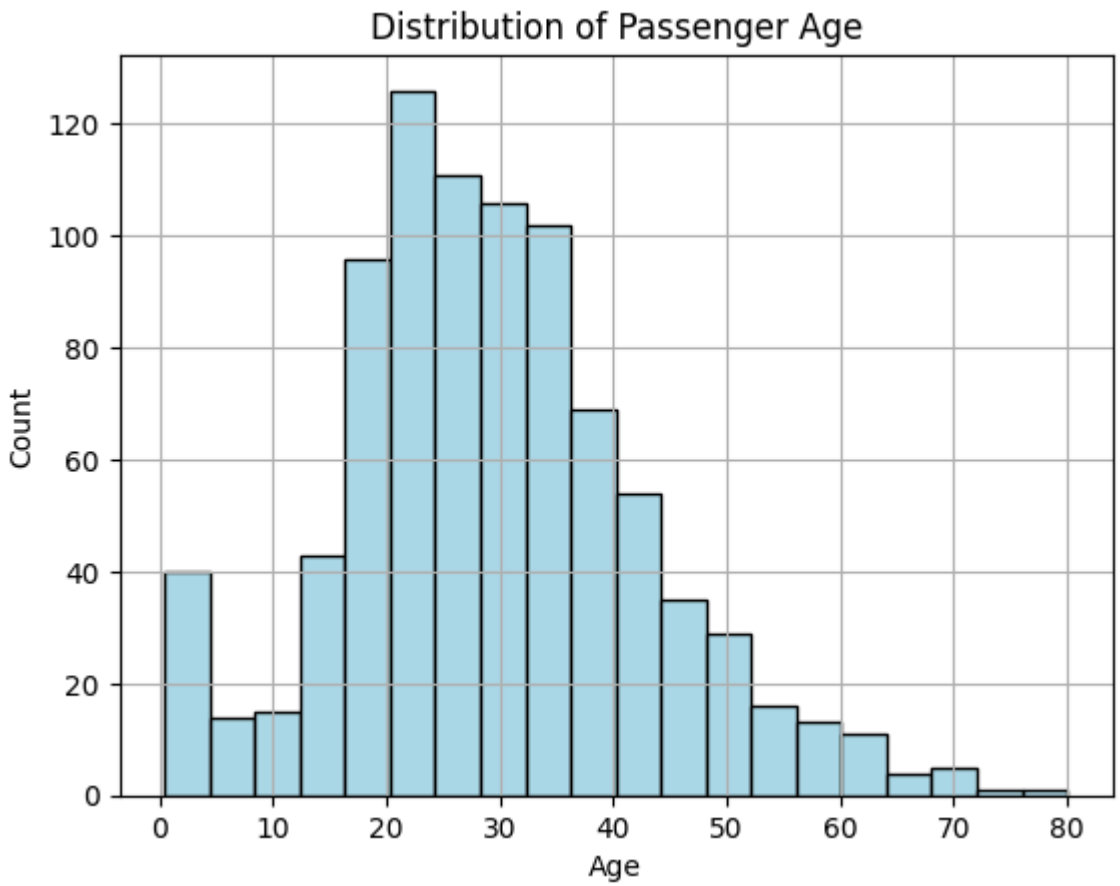
```
C:\Users\ASUS\AppData\Local\Temp\ipykernel_15716\565283480.py:1: FutureWarning: A value is trying to be set on a copy of a Data
Frame or Series through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we are set
ting values always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = d
f[col].method(value) instead, to perform the operation inplace on the original object.


  titanic['Embarked'].fillna(titanic['Embarked'].mode()[0], inplace=True)
```

In [7]:
```
titanic.drop(columns='Cabin', inplace=True)
```

In [8]:
```
import numpy as np
mean_age = titanic['Age'].mean()
std_age = titanic['Age'].std()
titanic['Age'] = titanic['Age'].apply(
    lambda x: np.random.randint(int(mean_age - std_age), int(mean_age + std_age)) if pd.isnull(x) else x
)
```

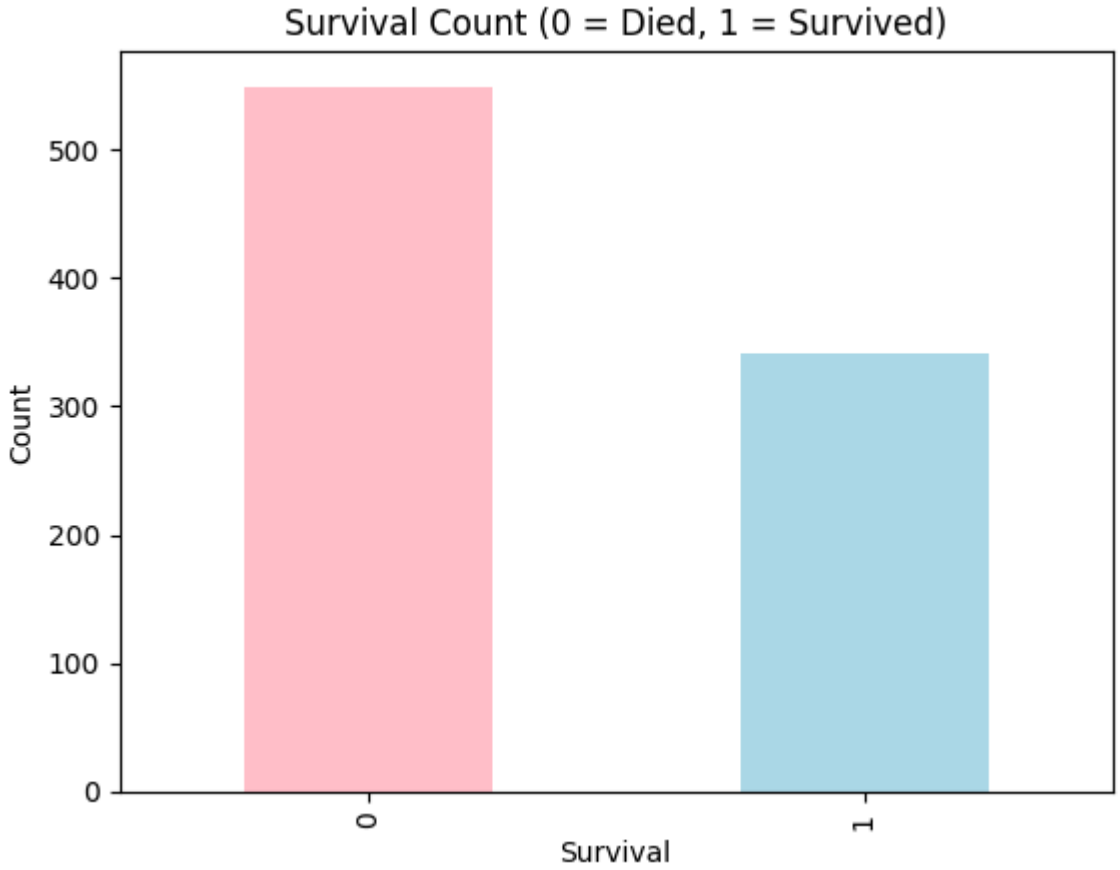
In [17]:
```
titanic['Age'].hist(bins=20, color='lightblue', edgecolor='black')
plt.title("Distribution of Passenger Age")
plt.xlabel("Age")
```

```
plt.ylabel("Count")
plt.show()
```



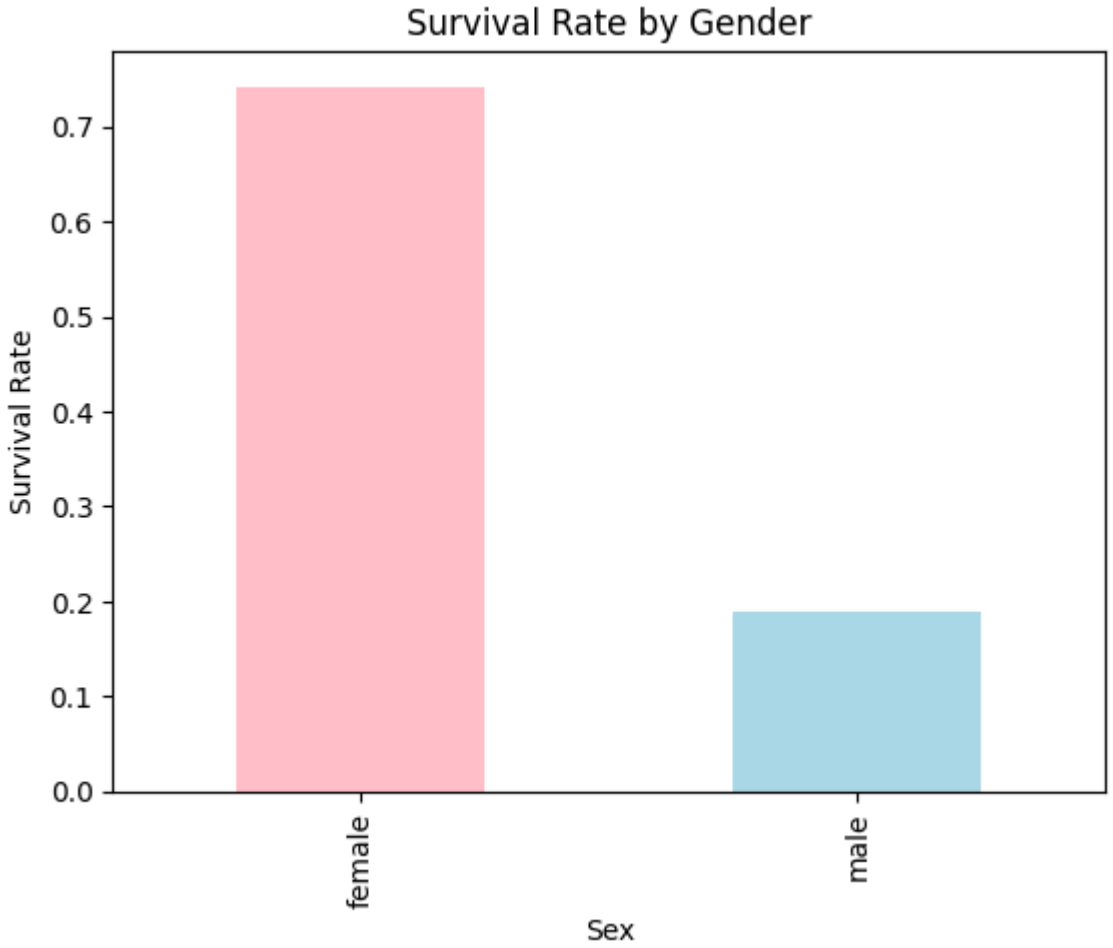
Most passengers are between 20 and 40 years old.

In [16]:
```
titanic['Survived'].value_counts().plot(kind='bar', color=['pink', 'lightblue'])
plt.title("Survival Count (0 = Died, 1 = Survived)")
plt.xlabel("Survival")
plt.ylabel("Count")
plt.show()
```



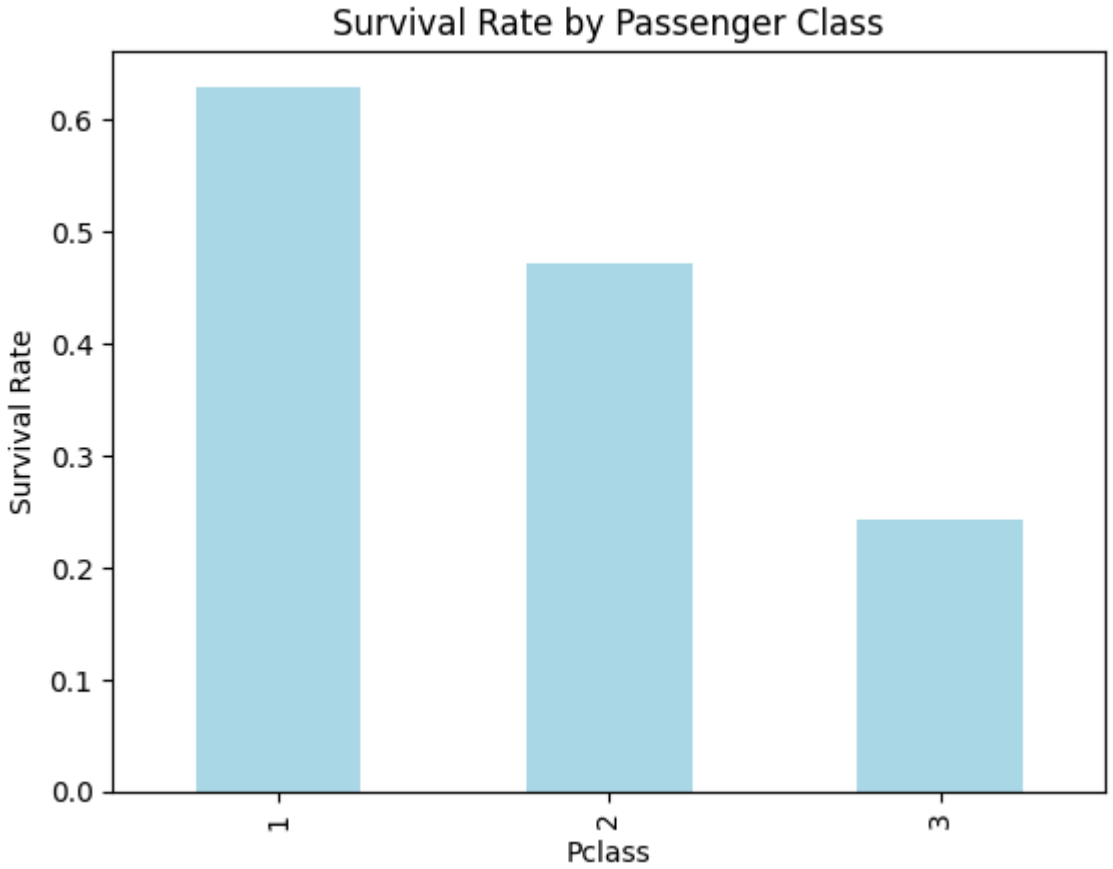
Survival rate was less than 40% overall.

In [11]:
```
titanic.groupby('Sex')['Survived'].mean().plot(kind='bar', color=['pink', 'lightblue'])
plt.title("Survival Rate by Gender")
plt.ylabel("Survival Rate")
plt.show()
```



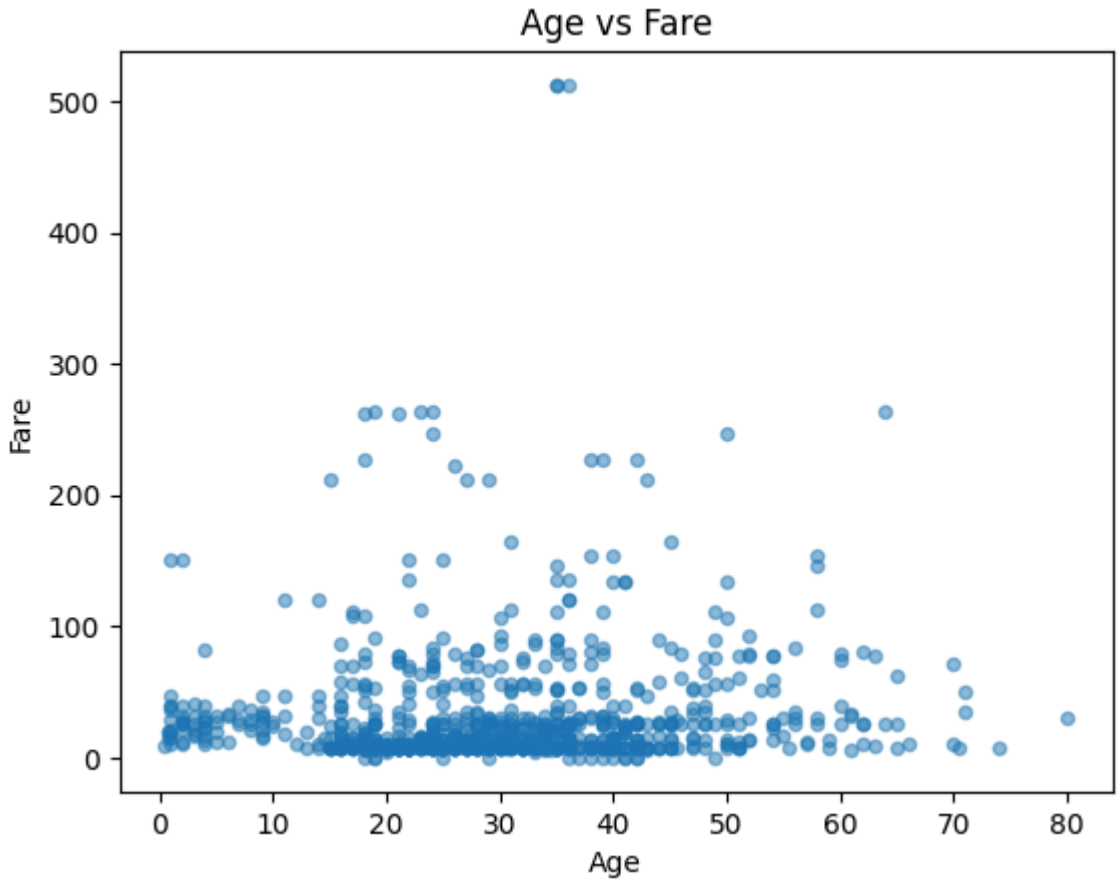
Females had a much higher survival rate than males.

In [15]:
```
titanic.groupby('Pclass')['Survived'].mean().plot(kind='bar', color='lightblue')
plt.title("Survival Rate by Passenger Class")
plt.ylabel("Survival Rate")
plt.show()
```



3rd class passengers had the lowest chance of survival.

In [13]:
```
titanic.plot.scatter(x='Age', y='Fare', alpha=0.5)
plt.title("Age vs Fare")
plt.show()
```



There’s no strong link between age and fare overall.

Final Summary

The analysis of the Titanic dataset reveals that most passengers were between 20 and 40 years old. The overall survival rate was below 40%, showing that the majority of passengers did not survive. Gender had a major impact on survival, as females had a much higher chance of surviving than males. Class also played a crucial role — passengers in 3rd class had the lowest survival rate compared to those in higher classes. Interestingly, there was no strong relationship found between a passenger’s age and the fare they paid.

In [ ]: