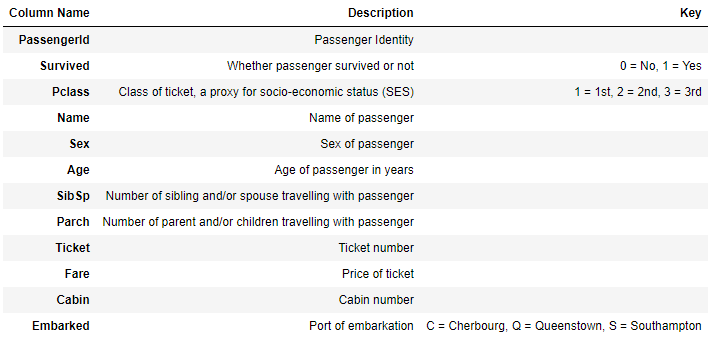
A**nalysis** of what sorts of people were most likely to **survive**. Predict whether passenger will **survive or not.**

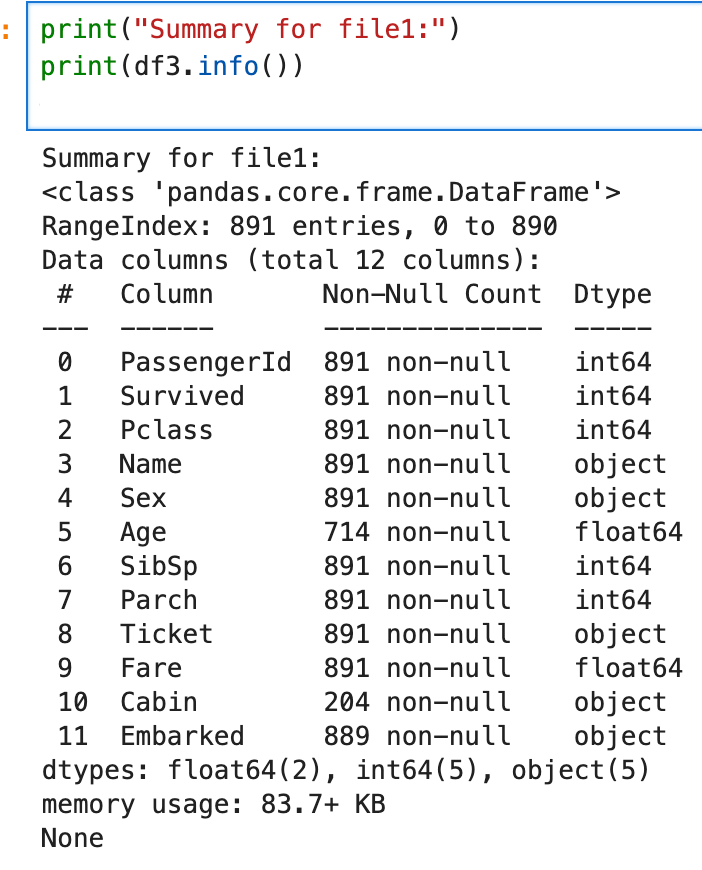
The training set includes passengers survival status (also know as the ground truth from the titanic tragedy) which along with other features like gender, class, fare and pclass (passenger class) is used to create the machine learning model.

The test set should be used to see how well the model performs on unseen data. The test set does not provide passengers survival status. We are going to use our model to predict passenger survival status.

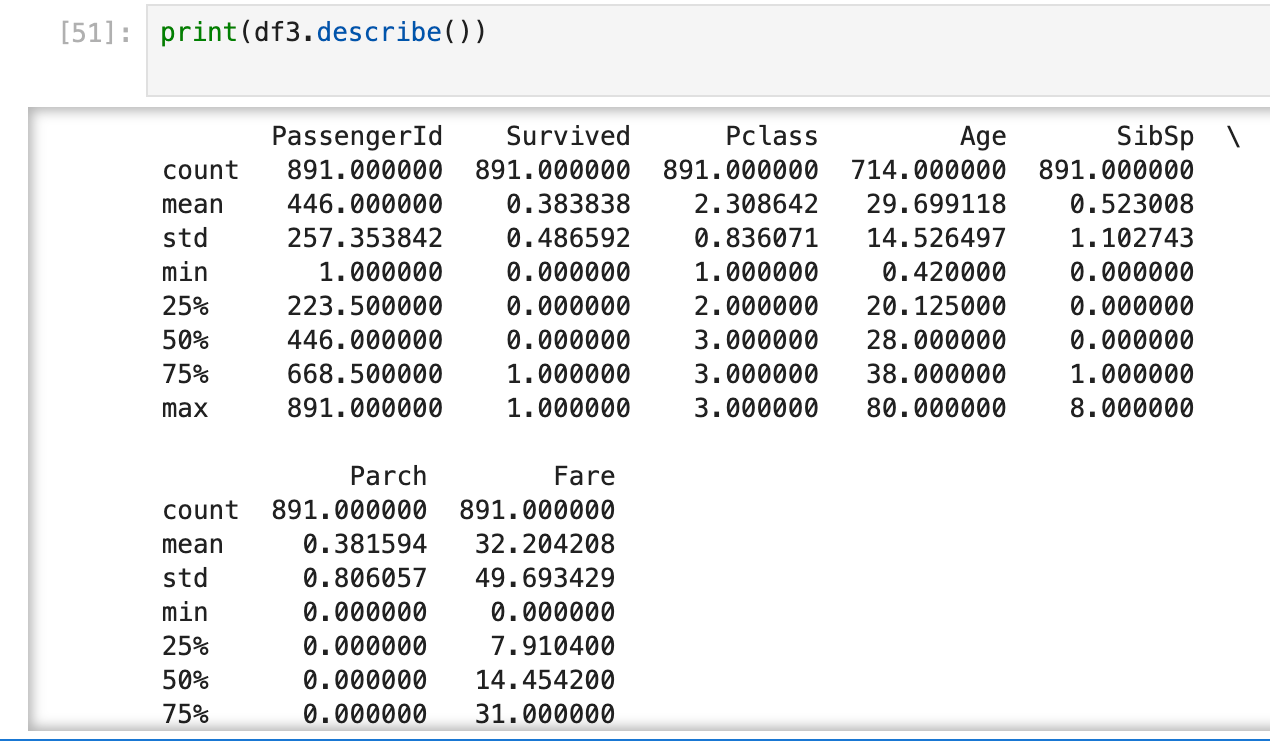
This is clearly a **Classification problem**. In predictive analytics, when the **target** is a categorical variable, we are in a category of tasks known as **classification tasks.**



The training-set has 891 rows and 11 features + the **target variable (survived).** 2 of the features are floats, 5 are integers and 5 are objects.



The training-set has 891 rows and 11 features + the ****target variable (Survived).**** 2 of the features are floats, 5 are integers and 5 are objects.



### **Conclusions from .describe() method**

****.describe()**** gives an understanding of the central tendencies of the numeric data.

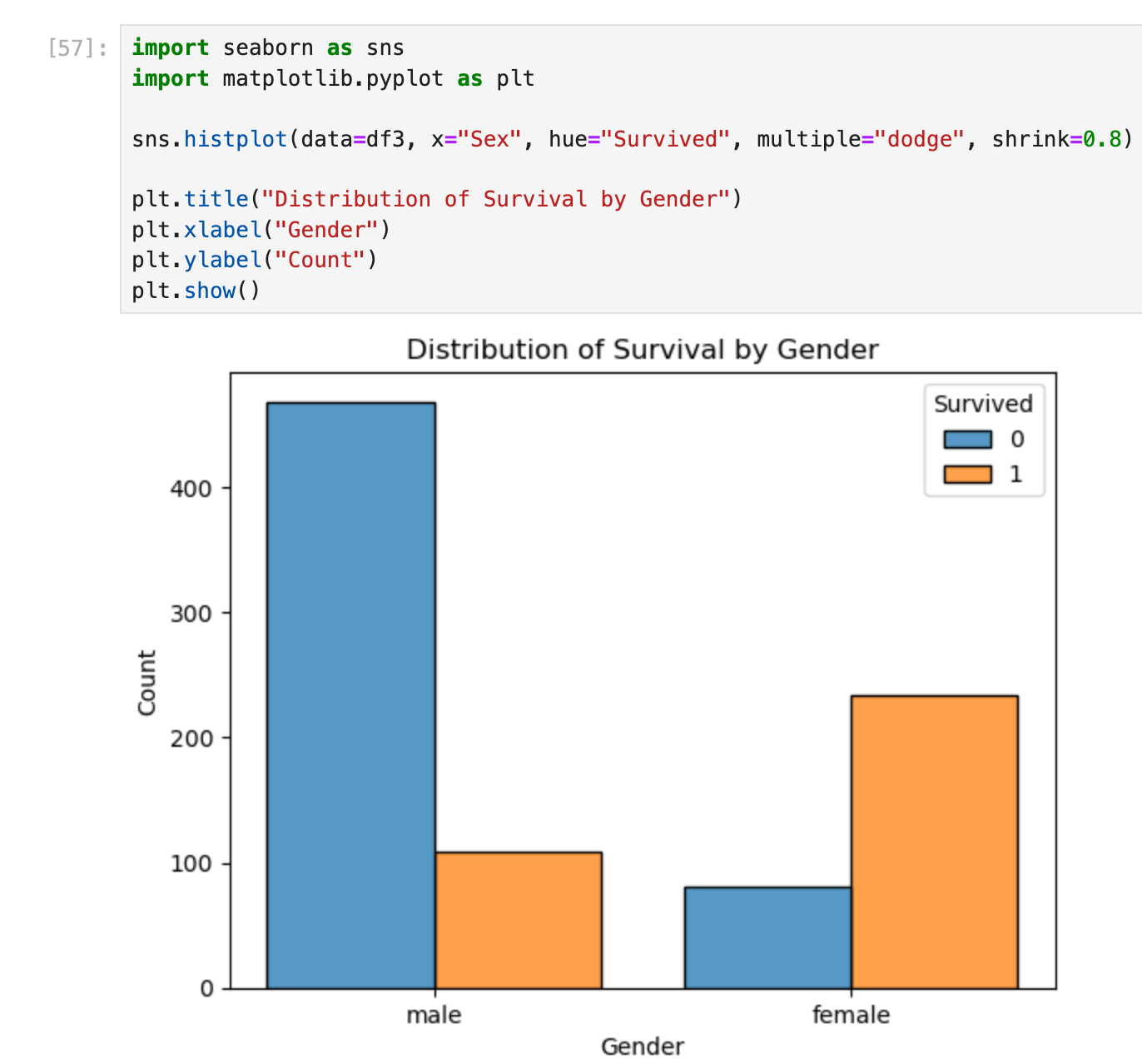
* Above we can see that ****38% out of the training-set survived the Titanic.****
* We can also see that the passenger age range from ****0.4 to 80 years old.****
* We can already detect some features that contain ****missing values****, like the ‘Age’ feature (714 out of 891 total).
* There's an ****outlier**** for the 'Fare' price because of the differences between the 75th percentile, standard deviation, and the max value (512). We might want to drop that value.

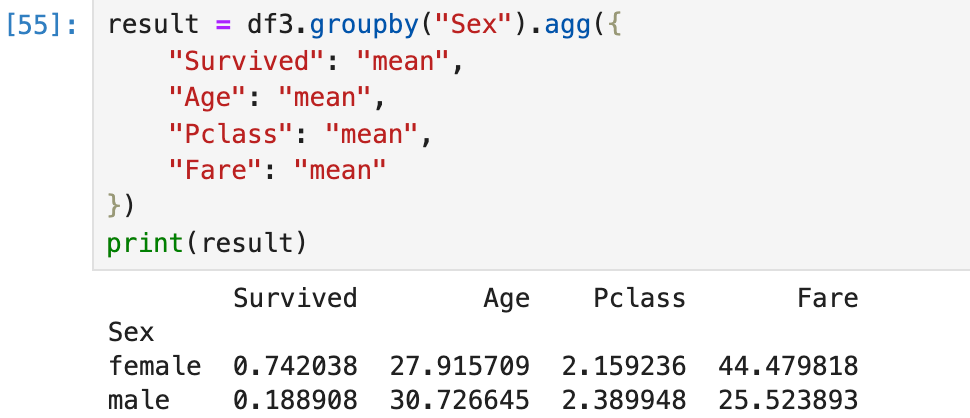


Conclusion

Passenger Class 1 has surrived more than class 3 and class 2

Maximum passenger travelling from class 3 has not survived.





Conclusion / Interpretation

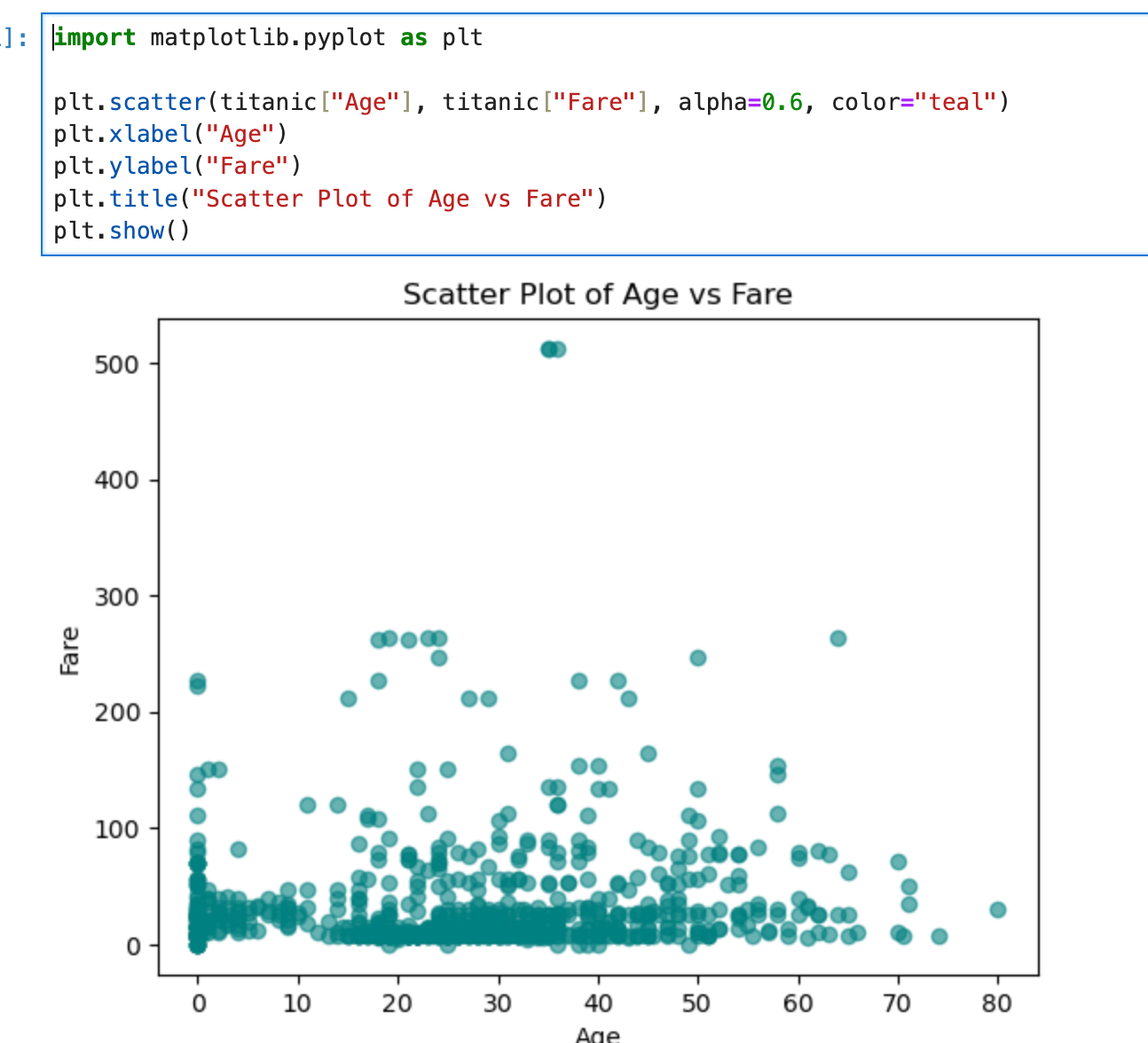
74% of the women survied while only 18% of the male survived

women had much higher survival rate.

Women were on average in slightly higher passenger classes.

Women paid higher fares --> more first/second class.

Age difference between sexes was small.



**Majority cluster**:

Many passengers fall into **younger/middle ages (20–40 years)** with **low fares (<50)**

This is expected since most were **3rd class passengers**

**High fare group**:

A small number of passengers paid **very high fares (>200)**

These are almost all **1st class passengers**, often with better survival chances

Some very young ages (below 10) appear at different fare ranges — family travel cases

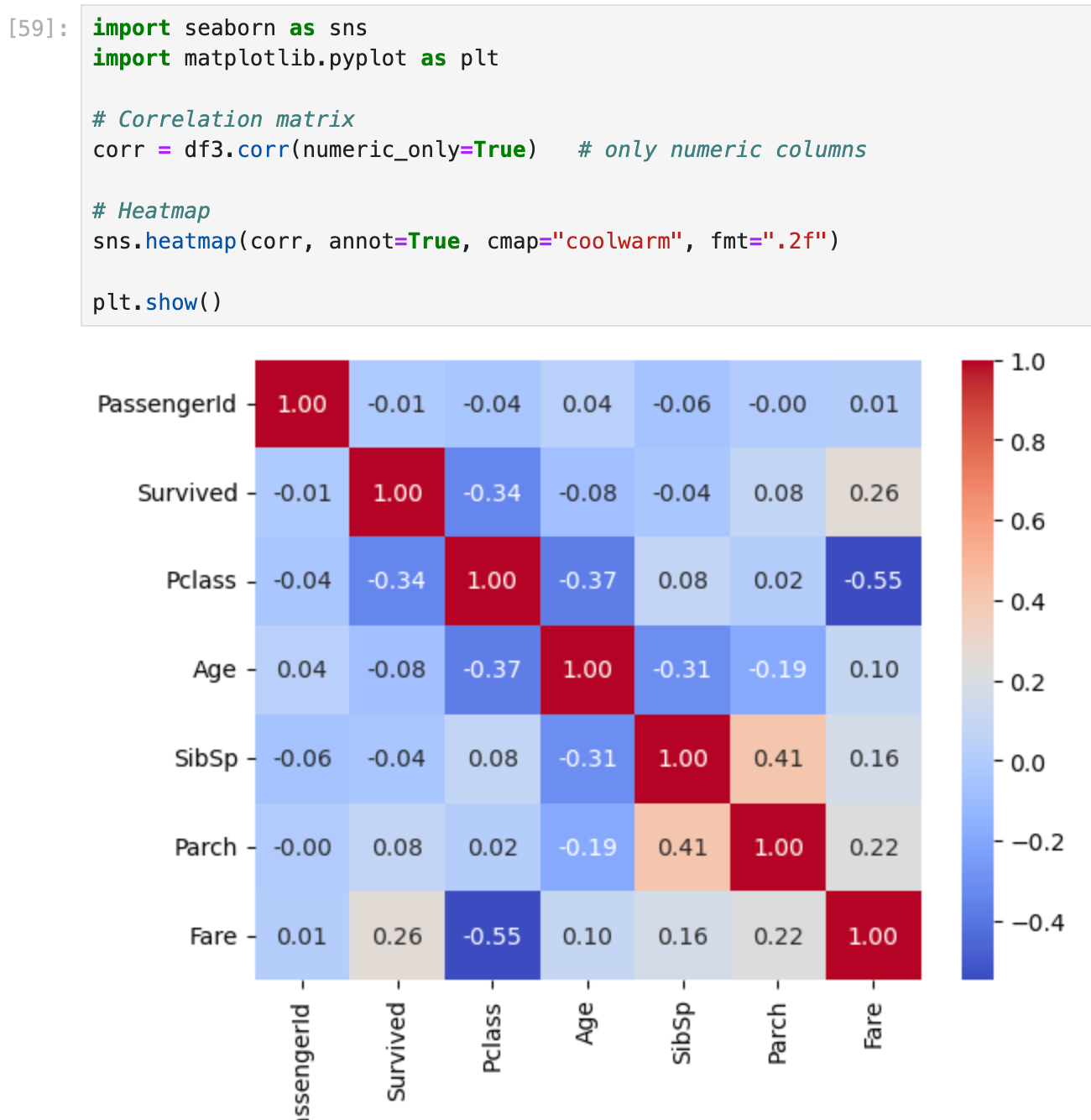
**Sparse older ages**:

Fewer elderly passengers, mostly with lower fares

**Age and Fare are not strongly correlated** (young and old both appear across fare ranges).

The **fare distribution is highly skewed**: most paid very little, a few paid a lot.

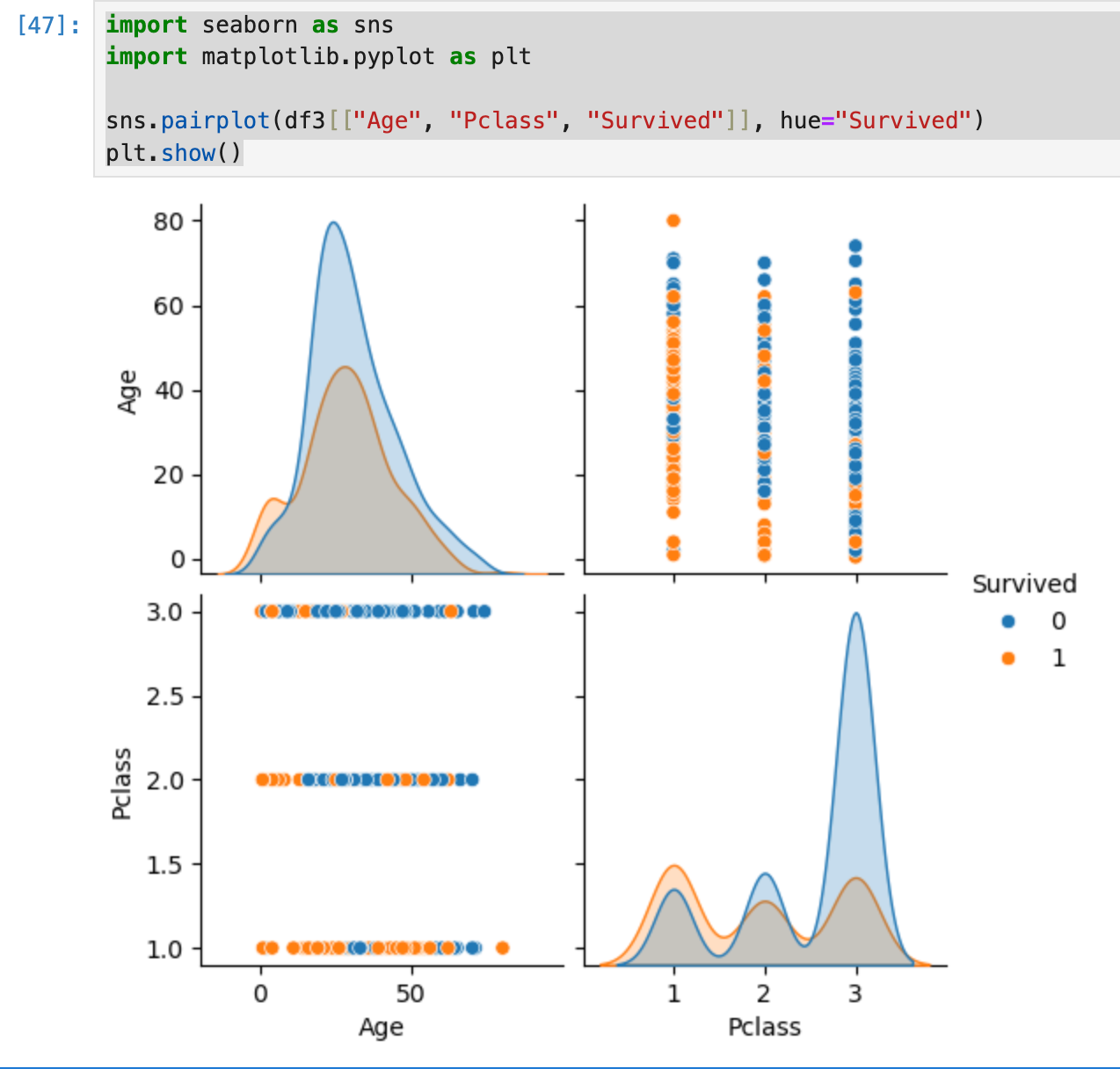
This hints at the importance of **class (Pclass)** in survival analysis, since fare is strongly tied to passenger class.



### **Correlation Matrix and Heatmap**

We notice from the heatmap above that:

* Parents and sibling like to travel together (Parch red squares)
* Age has a high negative correlation with number of siblings



#### a) ****Age vs Survived****

younger passengers (children) tend to have **higher survival**

Older passengers are more concentrated in **non-survived**

#### b) ****Pclass vs Survived****

**1st class passengers** have a higher proportion of survivors

**3rd class passengers** show a larger number of non-survivors

**2nd class** is in between

#### c) ****Age vs Pclass****

Scatter shows distribution of ages within each class

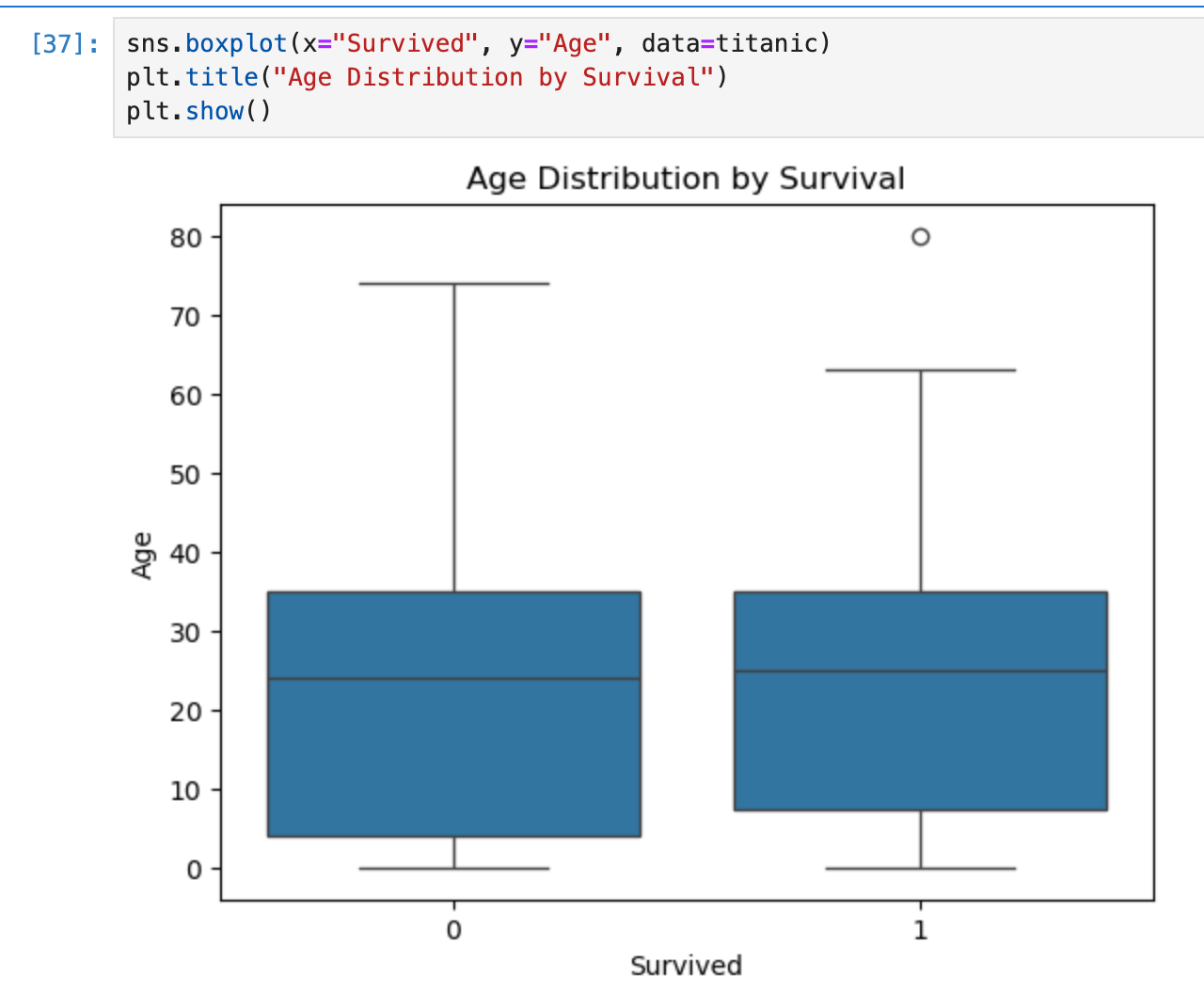
Younger passengers in 1st class likely survived

Middle-aged adults in 3rd class likely did not survive

### 3️Overall Insights

### ****Survival was influenced by both age and passenger class****

### Children and passengers in ****higher classes had better survival rates****

Adults in **lower classes had the lowest** 

**Median age:**

Survivors (1) often have **lower median age** than non-survivors (0)

Indicates **younger passengers had higher survival rates**

Non-survivors may have a wider age range (more adults and older passengers)

Survivors often cluster in younger ages but can also include some adults.

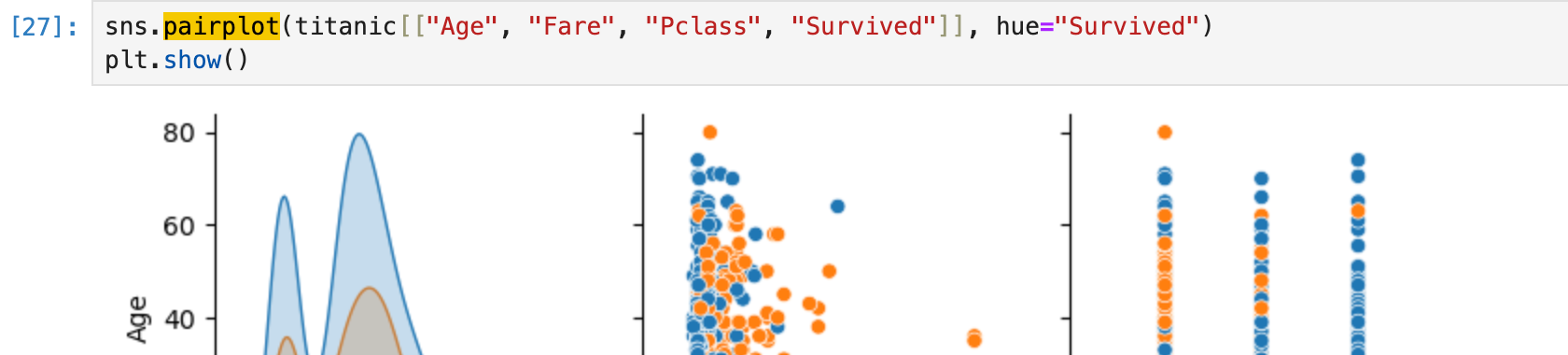
Very young children or very old passengers may appear as dots outside whiskers

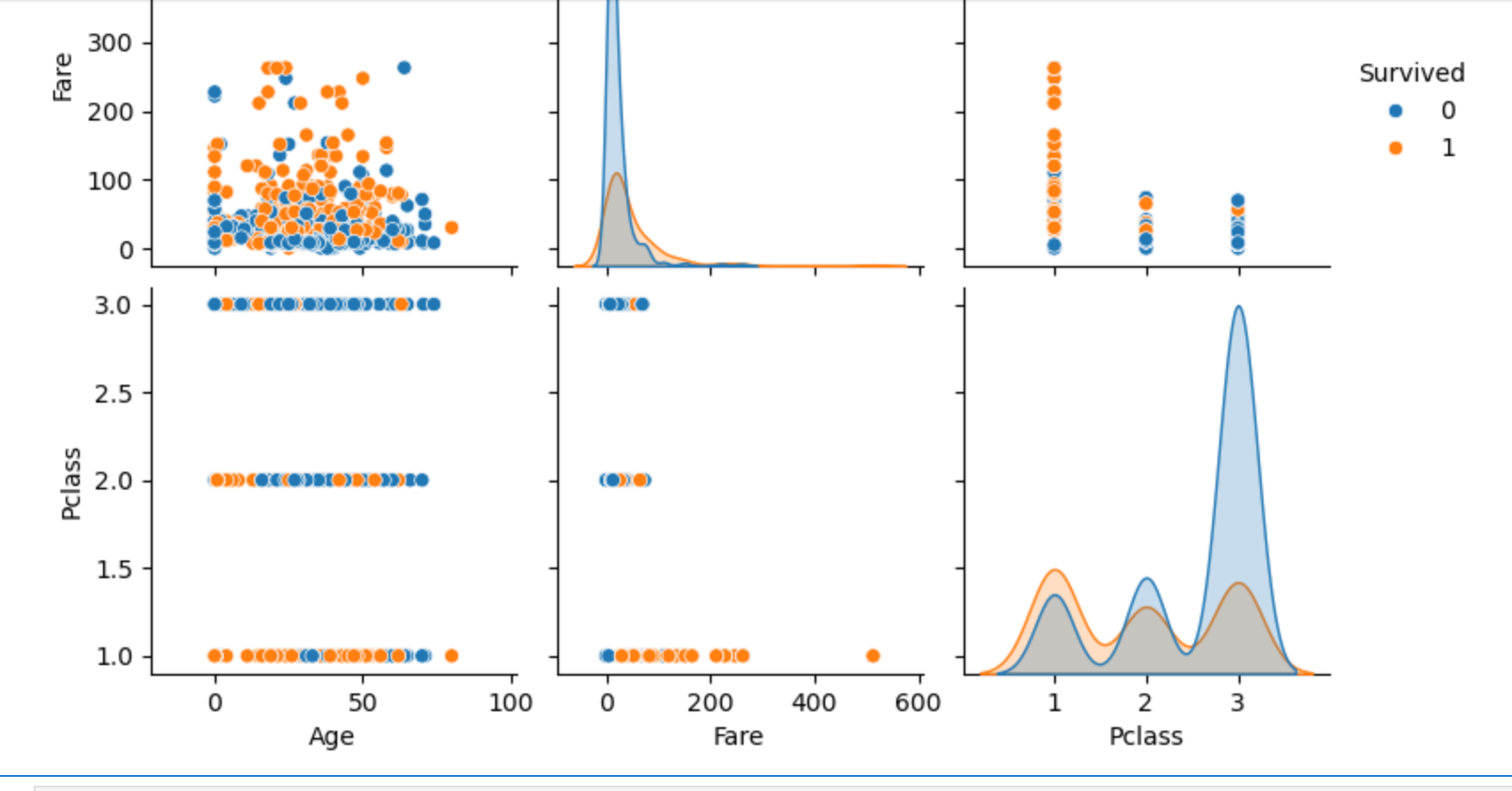
Can indicate special survival cases (e.g., children in 3rd class)

**Age played an important role in survival:**

Younger passengers were more likely to survive.

Older passengers had a higher mortality.





**Age vs Survived**

Survivors are spread across all ages.

Slightly higher survival for **children** compared to older passengers (supporting the "women and children first" rule).

**Fare vs Survived**

Clear pattern: **higher fares → more survivors**.

Many who paid low fares (3rd class) had lower survival rates.

**Pclass vs Survived**

Strong negative relationship:

**Pclass 1 (wealthier passengers)** → higher survival.

**Pclass 3 (lower class)** → low survival.

Confirms the socioeconomic survival bias on Titanic

**Age vs Fare (color by Survived)**

Among high-fare passengers, survival is much more likely.

Younger children from higher classes had better chances of survival.

Age distribution: Most passengers were **20–40 years old**, but survival spread is uneven.

Fare distribution: Skewed right, with a small number of very high fares (many of them survived).

Pclass distribution: Majority were in **3rd class**, but survival was lowest there.

## ****Summary of Findings****

**Survival Rate Patterns**

Survival was **not random** — it strongly depended on passenger demographics and ticket class.

**Women and children** had better chances compared to men.

**Impact of Pclass & Fare**

**Pclass (1st, 2nd, 3rd) was the strongest predictor** of survival.

1st Class → higher survival.

3rd Class → lowest survival

**Fare** (ticket price) aligns with class → passengers who paid more had much higher survival chances.

**Age Influence**

Children (lower age group) had relatively better survival than middle-aged adults.

Elderly passengers had lower survival probability.

**Family Size (SibSp + Parch = Relatives)**

Being **completely alone** lowered survival chances.

Having **too many relatives** also reduced survival (harder to escape together).

Moderate family size (1–2 relatives) had slightly better survival odds.

**Gender**

**Females survived much more than males** → aligns with the “Women and Children First” policy.

**Relationships from Visuals**

**Boxplot (Age vs Survived)** → survivors were distributed across ages, but survival was higher for children.

**Scatter (Age vs Fare)** → higher fare linked with survival, regardless of age.

**Pairplot (Age, Fare, Pclass, Survived)** → clear clusters show that survival strongly correlates with **high fare + low Pclass**.