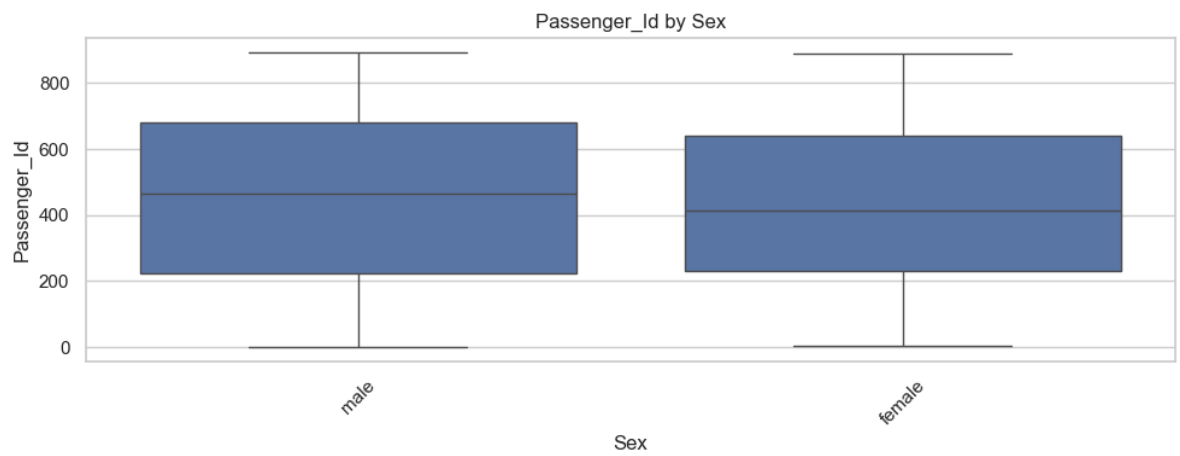
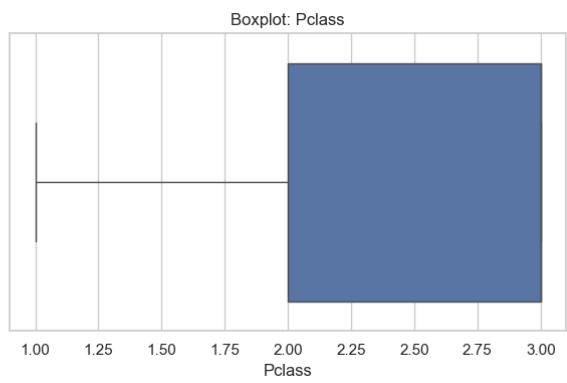
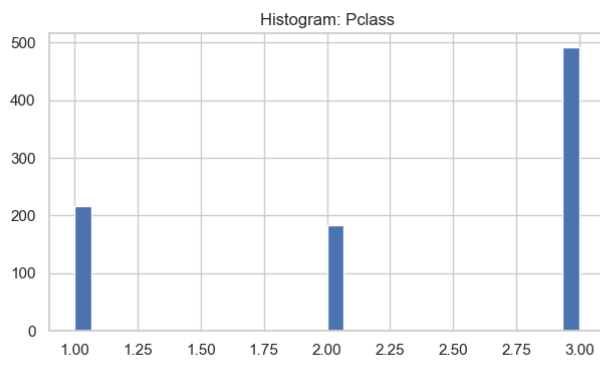
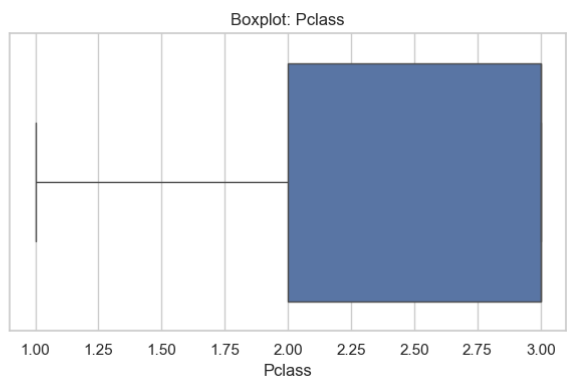
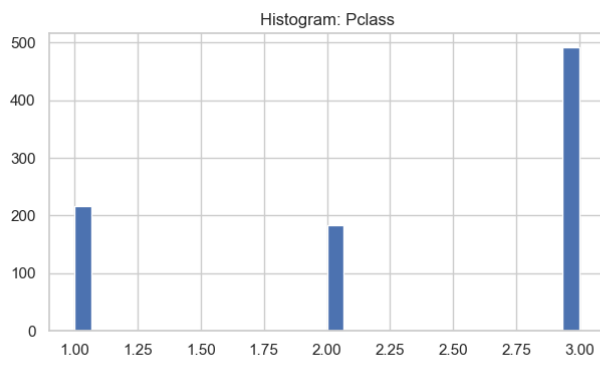
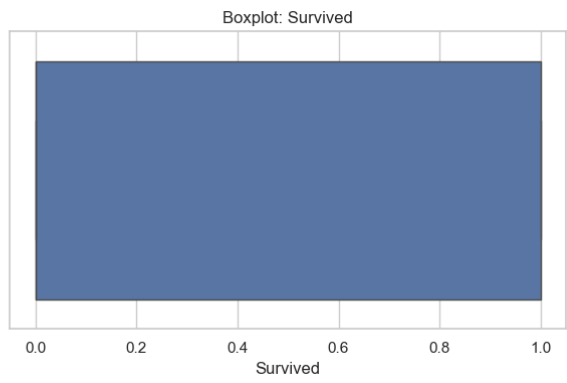
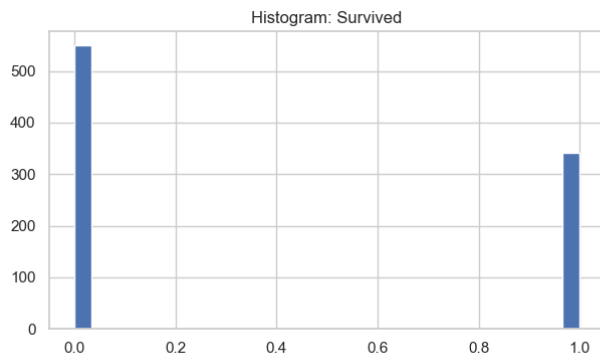
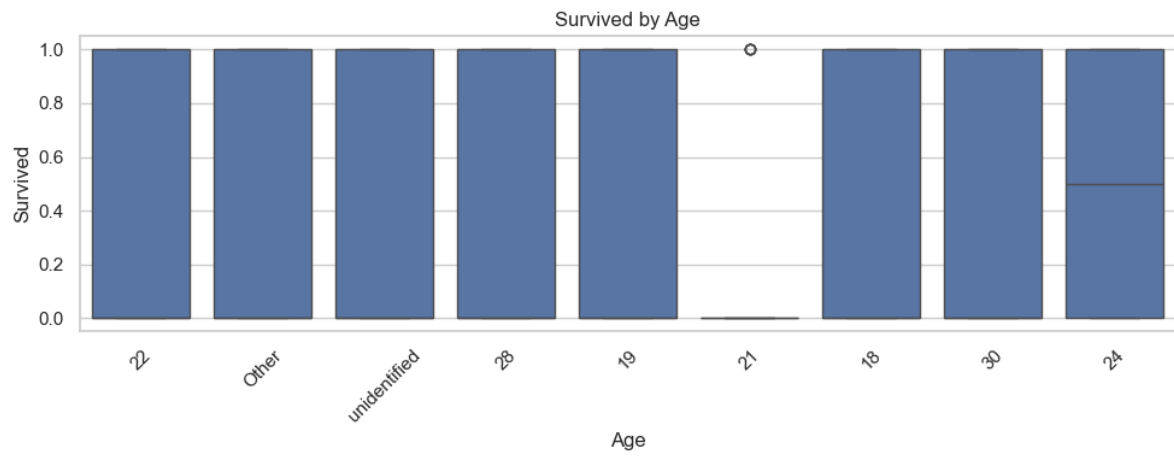
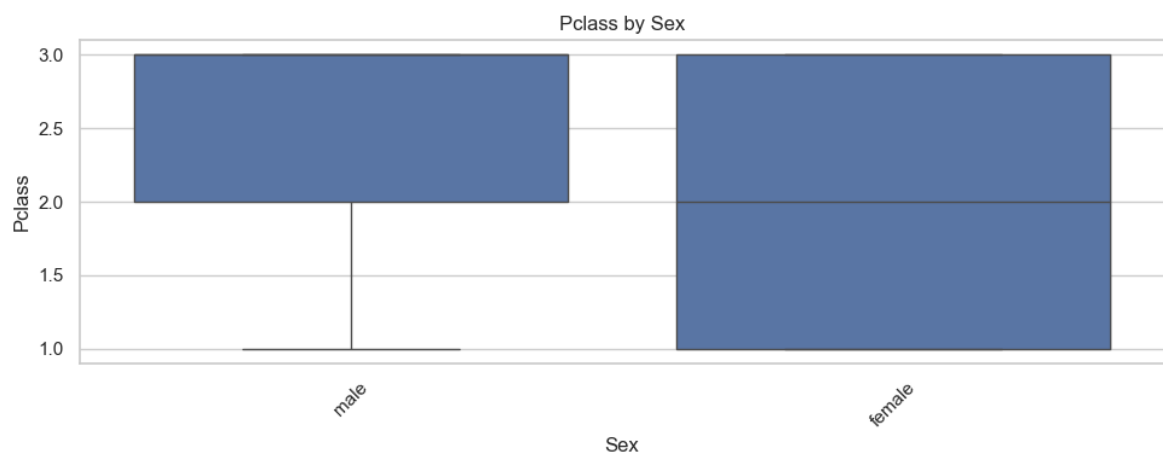
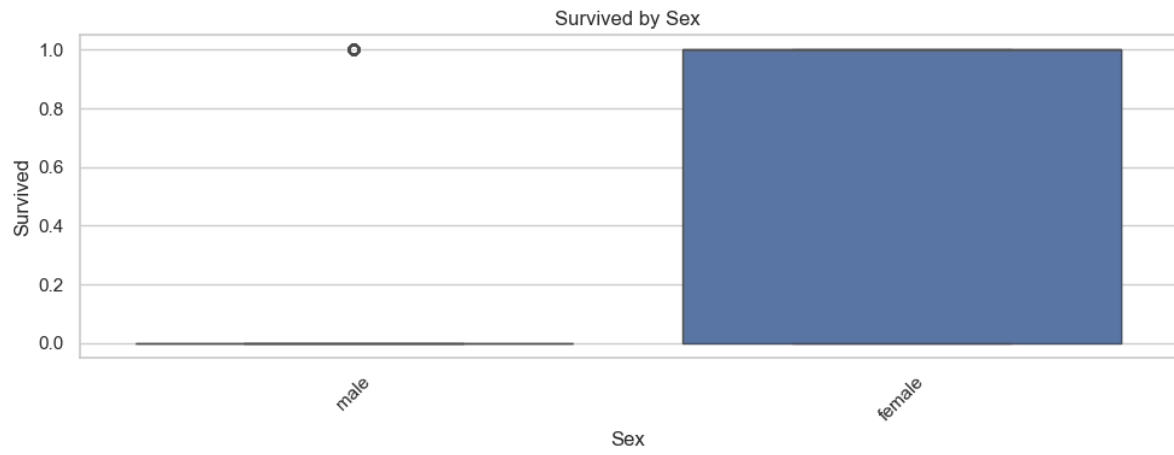
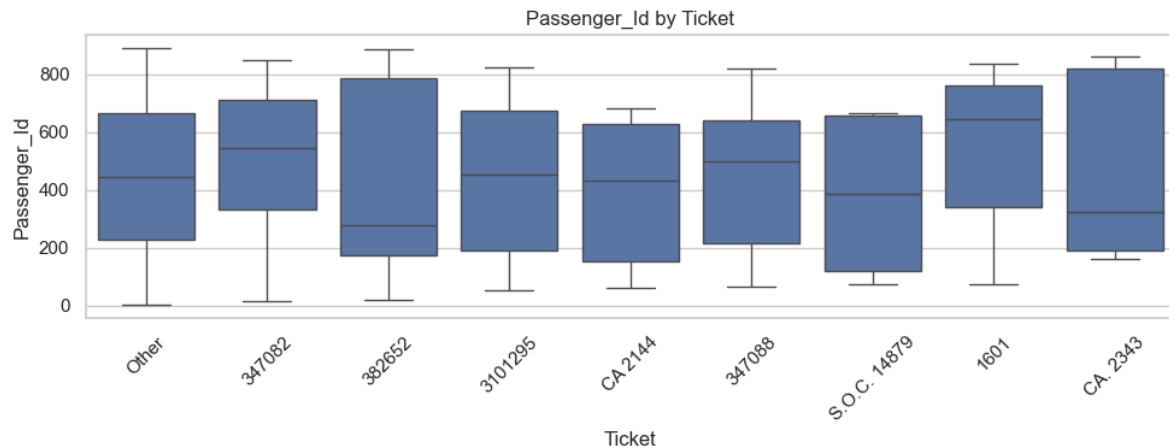


This includes the insight and information about the titanic csv file









Findings

1. Correlation Heatmap (top image)

- **Strongest relationships:**
 - **Pclass ↔ Fare: -0.55** (higher class = much more expensive tickets)
 - **Pclass ↔ Survived: -0.34** (1st class survived more)
 - **SibSp ↔ Parch: $+0.41$** (people with siblings/spouses often had parents/children aboard)
 - **Fare ↔ Survived: $+0.26$** (more expensive tickets → higher survival chance)

- **Almost no correlation between PassengerId and anything (as expected — it's just an index).**

2. Distribution plots (strip + violin plots)

- **PassengerId: uniformly distributed 1–891**
- **Survived: binary (0 = died, 1 = survived), ~38% survived**
- **Pclass: 1st > 3rd > 2nd class in count**
- **SibSp & Parch: heavily skewed — most people traveled alone or with 1 family member**
- **Fare: very right-skewed (many cheap tickets, a few extremely expensive ones)**

3. Individual histograms & boxplots

- **PassengerId: perfectly uniform**
- **Survived: ~549 died, ~342 survived (38.4% survival rate)**
- **Pclass: 1st class \approx 216, 2nd \approx 184, 3rd \approx 491 passengers**

4. Key bivariate insights by Sex

- **Almost identical number of male and female passengers (~577 male, ~314 female)**
- **Survival rate: women $\approx 74\%$, men $\approx 19\%$ → huge gender effect**
- **Pclass by Sex: women slightly over-represented in 1st class**

5. By Age groups

- **Most passengers 20–40 years old**
- **Survival rate roughly similar across adult age groups, but noticeably lower for 70+ and very young children in some bins**

6. By Ticket (first few characters)

- **Distribution fairly even across major ticket prefixes, no single ticket type dominates**

Key Takeaways from the entire EDA

- **The three strongest predictors of survival appear to be:**
 - 1. Sex (female >> male)**
 - 2. Pclass (1st > 2nd > 3rd)**
 - 3. Fare (correlated with class but adds some extra signal)**

- **Family variables (SibSp, Parch) have moderate positive correlation with each other but weaker direct link to survival.**
- **Age has some effect but is much weaker than sex and class in this dataset.**