Summary of Findings

The overall survival rate for passengers in this dataset was low (approximately 38.4%). The most critical factors influencing survival were **gender, socio-economic status (passenger class/fare), and family size**.

- **Gender Bias:** Females were far more likely to survive than males.
- **Class/Wealth Advantage:** Passengers in 1st class, who paid the highest fares, had the best survival rates, while those in 3rd class had the worst.
- **Age and Family Structure:** Infants and young children were prioritized for survival. Traveling with a small family unit (2 to 4 members) improved survival chances compared to traveling alone or with a very large family.

Identified Relationships and Trends

The analysis reveals several strong relationships and trends, which would be essential for building a predictive model:

1. **Inverse Relationship between P class and Survival:** There is a clear negative correlation between passenger class number (1 being highest, 3 being lowest) and the probability of survival. As the class number increases (i.e., moving from 1st to 3rd class), the survival rate drops significantly.
2. **Strong Relationship between Sex and Survival:** Gender is the most apparent differentiator for survival probability, with females having a highly disproportionate survival rate over males. This is likely due to the implementation of the "women and children first" protocol.
3. **Non-Linear Trend with Family Size:** Survival probability is not monotonic with family size.
   - **Trend 1: Lower survival for isolation.** Traveling alone is detrimental to survival.
   - **Trend 2: Optimal survival for small groups.** Small family units (2-4 members) show the highest survival propensity.

o **Trend 3: Lower survival for large groups.** Very large families (5 or more members) have a low chance of survival.

4. **Relationship between `Fare` and Survival:** Higher fares (indicating wealth and typically 1st class status) are strongly associated with higher survival rates.

Visuals and Feature Distributions

Distribution of Passenger Ages (Histogram with KDE)

- The age distribution is bimodal (or highly peaked) with two clear modes: a small peak at the infant/very young child ages (around 0-5) and a large, dominant peak in the young adult ages (mid-20s).

Boxplot of Passenger Fares

- The boxplot confirms the extreme positive skew of the fare data. The interquartile range (the box) is very small, indicating most fares are low, but there are a large number of extreme outliers representing very high ticket prices (up to over $500).

Scatter Plot: Fare vs. Age, Colored by Survival

- Most passengers paid low fares, regardless of age.
- **Survivors (orange points)** are visibly more frequent among passengers who paid the highest fares (above $200).
- Survival appears mixed across all ages, though there is a cluster of survivors among very young children (Age 0-10).

Pair plot (Diagonal KDEs)

- **Survival by Pclass:** The distribution for non-survivors is heavily concentrated in 3rd class. The distribution for survivors is most prominent in 1st class.

  Survival by Fare: The non-survivor distribution is sharply peaked near $0. The survivor distribution is flatter and shifted toward higher fare values.