**REPORT**

**Pairplot**

**Trends and Relationships**

**1. Fare vs Survived**

**Trend:** People who paid **higher fares** tend to have a higher chance of **survival** (orange dots).

**Interpretation:** This could indicate that **passengers in higher classes** (who paid more) had **better access to lifeboats** or priority during evacuation.

**2. PassengerId vs Survived**

**Trend:** No clear pattern or correlation.

**Interpretation:** PassengerId is just a unique identifier; it doesn't hold analytical meaning regarding survival.

**3. Fare Distribution**

The **distribution is skewed to the right**, with a few passengers paying extremely high fares.

Many passengers paid **very low fares**, possibly representing third-class or economy.

**4. Survived (Hue)**

The hue distinguishes survivors (1) and non-survivors (0). In most plots, you can visually see more **blue dots (non-survivors)** among low-fare payers, and **orange dots (survivors)** become more prominent as fare increases.

**Heatmap**

| **Feature Pair** | **Correlation** | **Interpretation** |
| --- | --- | --- |
| **Survived & Pclass** | -0.34 | Passengers in **higher classes** (lower number) were **more likely to survive** |
| **Survived & Fare** | 0.26 | People who paid **higher fare** had **higher survival** chances |
| **Survived & Age** | -0.08 | Slight negative trend: **younger passengers** had slightly better survival |
| **Pclass & Fare** | -0.55 | Strong negative: **Lower class → lower fare** |
| **SibSp & Parch** | 0.41 | Families with **siblings** often had **parents/children** too onboard |
| **Age & Pclass** | -0.37 | **Older passengers** tended to travel in **higher class** |

**Histogram**

| **Feature** | **Observations** |
| --- | --- |
| **PassengerId** | **Uniform distribution – just a unique identifier; no analysis value.** |
| **Survived** | **0 (Not Survived) > 1 (Survived); more passengers died than survived.** |
| **Pclass** | **Class 3 had the most passengers, followed by Class 1 and 2. Indicates majority were in lower class.** |
| **Age** | **Distribution is right-skewed. Most passengers were young adults (20–40 years); fewer were children or seniors.** |
| **SibSp** | **Most passengers had 0 siblings/spouses onboard; some had 1 or 2. Large families were rare.** |
| **Parch** | **Most had 0 parents/children onboard. Very few traveled with family.** |

**Boxplots**

1. **PassengerId: Shows even distribution (as expected); not useful for outlier detection.**
2. **Survived: Binary, hence boxplot not meaningful.**
3. **Pclass: Discrete values; boxplot confirms class range is between 1–3.**
4. **Age: Clear presence of outliers (e.g., children and elderly), most passengers are between ~20 and ~40.**
5. **SibSp & Parch:**
   * **Strong presence of outliers (e.g., some passengers with 4–5+ siblings/spouses or parents/children).**
   * **Majority had small families or traveled alone.**

**SUMMARY OF FINDINGS**

**1. PassengerId**

* **Histogram Insight: Uniform distribution from start to end.**
* **Boxplot Insight: No meaningful statistical insight — this is just a unique identifier.**
* **Conclusion: Not useful for analysis or modeling.**

**2. Survived**

* **Histogram Insight: More people died (0) than survived (1).**
* **Boxplot Insight: Binary variable — boxplot isn’t useful here.**
* **Conclusion: Target variable for classification; imbalanced classes.**

**3. Pclass (Passenger Class)**

* **Histogram Insight: Most passengers were in 3rd class, followed by 1st and 2nd.**
* **Boxplot Insight: All values between 1 and 3; categorical but represented numerically.**
* **Conclusion: Strong class-based skew — likely important for survival prediction.**

**4. Age**

* **Histogram Insight: Right-skewed — most passengers were aged between 20–40 years.**
* **Boxplot Insight: Outliers present — especially very young and very old passengers.**
* **Conclusion: Likely important; consider imputing missing values and handling outliers.**

**5. SibSp (Siblings/Spouses Aboard)**

* **Histogram Insight: Most passengers had 0 or 1 sibling/spouse aboard.**
* **Boxplot Insight: Strong presence of outliers (values up to 5+).**
* **Conclusion: Family structure may influence survival; rare large families are outliers.**

**6. Parch (Parents/Children Aboard)**

* **Histogram Insight: Most passengers had no parents or children aboard.**
* **Boxplot Insight: Few extreme values — many passengers were alone.**
* **Conclusion: Like SibSp, this could influence survival; outliers present.**