

# Frequency Analysis Report

---

Titanic Dataset

---

Aryan Sawant

August 1, 2025

## Introduction

This report presents a frequency and correlation analysis of the Titanic passenger dataset. Key steps include: dataset setup, frequency and contingency tables, probability calculations, and correlation visualizations.

## 1 Setup and Data Loading

Python Code:

```
1 import seaborn as sns
2 import pandas as pd
3 import matplotlib.pyplot as plt
4
5 df = sns.load_dataset('titanic')
6 df.shape
```

**Explanation:** We import required libraries and load the Titanic dataset as a DataFrame.

## 2 Frequency Analysis of Passenger Class

Python Code:

```
1 # Absolute Frequency
2 abs_freq = df['class'].value_counts()
3 # Relative Frequency (%)
4 rel_freq = (abs_freq / df['class'].notna().sum()) * 100
5 # Cumulative Frequency
6 cum_freq = abs_freq.cumsum()
7 # Combine into a DataFrame
8 freq_table = pd.DataFrame({
9     'Absolute Frequency': abs_freq,
10    'Relative Frequency (%)': rel_freq.round(2),
```

```

11     'Cumulative Frequency': cum_freq
12 })
13 # Added a total row
14 total_row = pd.DataFrame({
15     'Absolute Frequency': [freq_table['Absolute Frequency'].sum()],
16     'Relative Frequency (%)': [freq_table['Relative
17         Frequency (%)'].sum()],
18     'Cumulative Frequency': [',']
19 }, index=[ 'Total'])
20 freq_table = pd.concat([freq_table, total_row])
21 freq_table

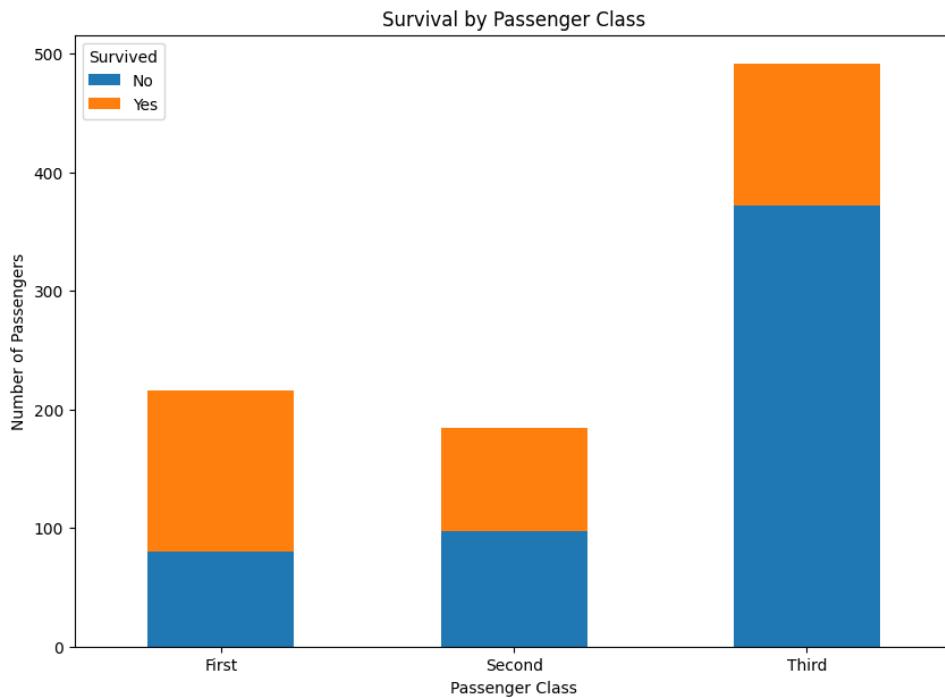
```

**Result Table:**

Class	Absolute Freq.	Relative Freq. (%)	Cumulative Freq.
Third	491	55.11	491
First	216	24.24	707
Second	184	20.65	891
Total	891	100.00	

Table 1: Passenger Class Frequency

### 3 Survival by Passenger Class

**Visualization:****Python Code:**

```

1 plt.figure(figsize=(10,6))
2 sns.countplot(data=df, x='class', hue='survived')
3 plt.title('Survival by Passenger Class')
4 plt.xlabel('Passenger Class')
5 plt.ylabel('Count')
6 plt.legend(title='Survived', labels=['No', 'Yes'])
7 plt.show()

```

## 4 Contingency Table and Probabilities

Python Code:

```

1 contingency_table = pd.crosstab(df['sex'], df['survived'],
2                                 margins=True)
3 print(contingency_table)

```

Contingency Table:

	survived	0	1	All
sex				
female	81	233	314	
male	468	109	577	
All	549	342	891	

Probability Calculations:

```

1 total_passengers = contingency_table.loc['All', 'All']
2
3 # Joint Probability: female and survived
4 prob_female_and_survived = contingency_table.loc['female', 1] /
5     total_passengers
6
7 # Marginal Probabilities
8 prob_female = contingency_table.loc['female', 'All'] /
9     total_passengers
prob_survived = contingency_table.loc['All', 1] /
10    total_passengers
11
12 # Conditional Probabilities
prob_survived_given_female = contingency_table.loc['female',
13         1] / contingency_table.loc['female', 'All']
prob_female_given_survived = contingency_table.loc['female',
14         1] / contingency_table.loc['All', 1]

```

Results:

- $P(\text{Sex} = \text{female}, \text{Survived} = 1)$ : 0.2615
- $P(\text{Sex} = \text{female})$ : 0.3524
- $P(\text{Survived} = 1)$ : 0.3838

- $P(\text{Survived} = 1 \mid \text{Sex} = \text{female})$ : 0.7420
- $P(\text{Sex} = \text{female} \mid \text{Survived} = 1)$ : 0.6813

## 5 Correlation Analysis

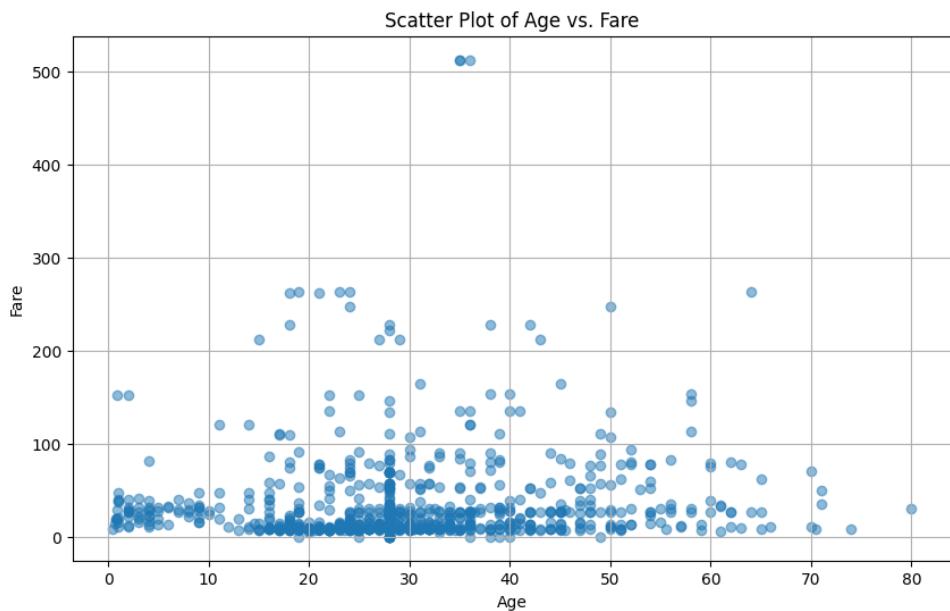
Python Code:

```

1 df_corr = df[['age', 'fare']].copy()
2 age_median = df_corr['age'].median()
3 df_corr['age'].fillna(age_median, inplace=True)
4
5 pearson_corr = df_corr.corr(method='pearson')
6 age_fare_corr = pearson_corr.loc['age', 'fare']
7 print(f"Pearson correlation (age, fare): {age_fare_corr:.4f}")
  
```

**Result:** Pearson correlation between Age and Fare: 0.0967 (very weak positive linear relationship).

Scatter Plot:

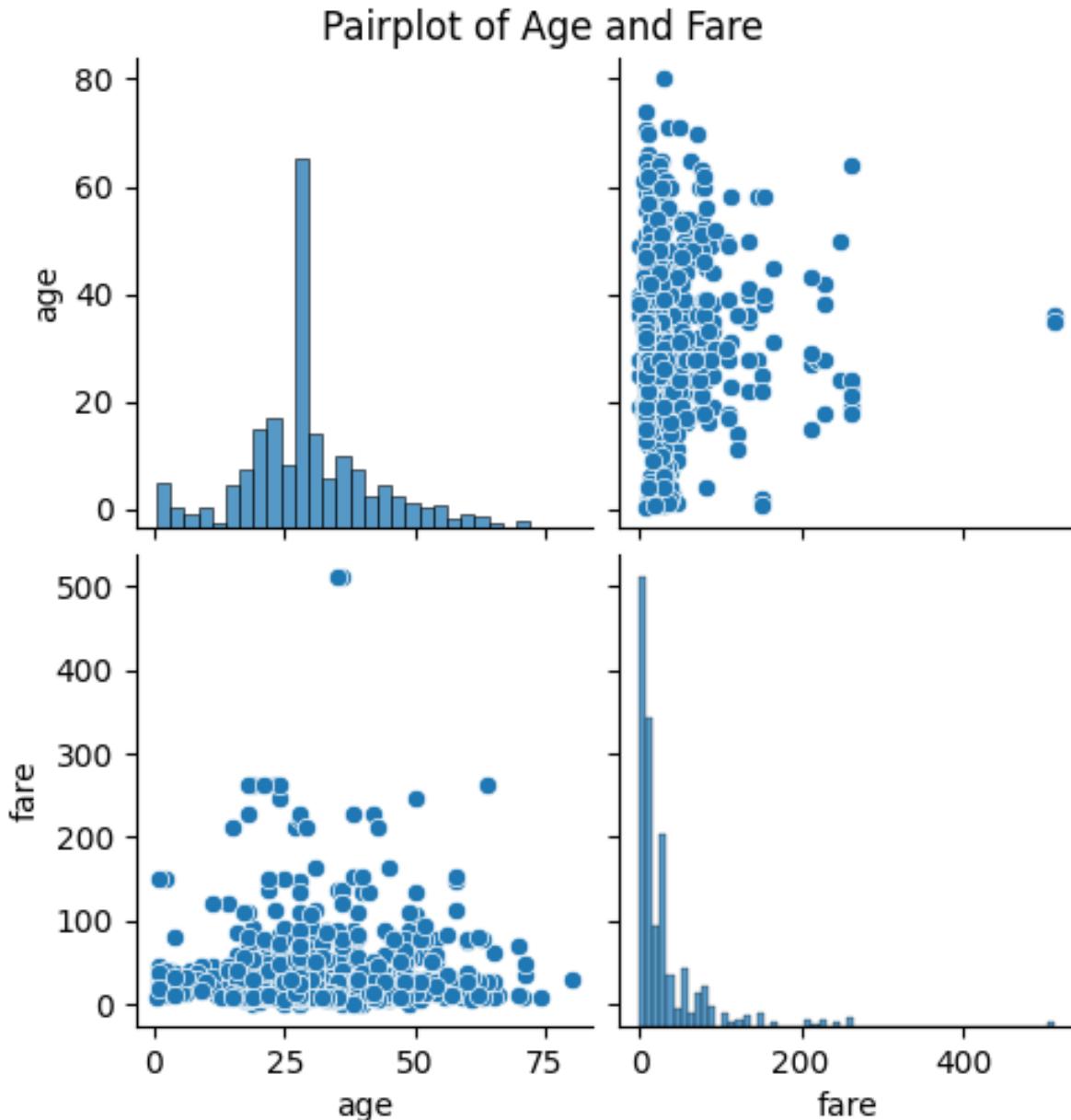


Python Code:

```

1 plt.figure(figsize=(10, 6))
2 sns.scatterplot(data=df_corr, x='age', y='fare')
3 plt.title('Fare vs Age')
4 plt.xlabel('Age')
5 plt.ylabel('Fare')
6 plt.show()
  
```

## Pairplot of Numerical Variables



```
1 numerical_vars = ['survived', 'pclass', 'age', 'sibsp', ''
                  'parch', 'fare']
2 sns.pairplot(df[numerical_vars].dropna())
3 plt.show()
```

## Conclusion

This report has provided frequency, contingency, probability, and correlation analyses of the Titanic dataset, with enhanced readability and visual formatting for clarity.