

## IntroToDataStructures \_ Workshop Challenge Week 2

READING THE DATA FROM SOURCE FILE: data/sets/titanic\_data.csv

From Challenge 6 - Calculate the MEAN MODE MEDIAN

```

In [1]: from statsCalculation import StatsCalculation

# Create instance and get summary table
csv_path = 'data/sets/titanic_data.csv'

stats = StatsCalculation(csv_path=csv_path)

summary_df = stats.summary_table()

# Display with styling
styled_summary = summary_df.style.set_caption("📊 Titanic Summary Statistics")\
    .format(precision=2)\
    .set_table_styles([
        {'selector': 'thead th', 'props': [('background-color', '#dbe9f4'), ('color', 'black')]},
        {'selector': 'tbody td', 'props': [('border', '1px solid #ccc')]}
    ])\
    .set_properties(**{'font-family': 'Arial', 'font-size': '14px'})

styled_summary

```

Out[1]: 📊 Titanic Summary Statistics

	Mean	Median	Mode
<b>Survived</b>	0.39	0.00	None
<b>Pclass</b>	2.31	3.00	None
<b>Age</b>	29.47	28.00	None
<b>Siblings/Spouses Aboard</b>	0.53	0.00	None
<b>Parents/Children Aboard</b>	0.38	0.00	None
<b>Fare</b>	32.31	14.45	None

From Challenge 6 - Calculate the variance

```

In [2]: #variance

summary_df = stats.varianceMethod()
# Style for notebook display
varianceData = summary_df.style.set_caption("📊 Titanic Spread & Distribution Statistics")\
    .set_table_styles([
        {'selector': 'thead th', 'props': [('background-color', '#f0f8ff'), ('color', 'black')]},
        {'selector': 'tbody td', 'props': [('border', '1px solid #ccc')]}
    ])\
    .set_properties(**{'font-family': 'Arial', 'font-size': '14px'})

varianceData

```

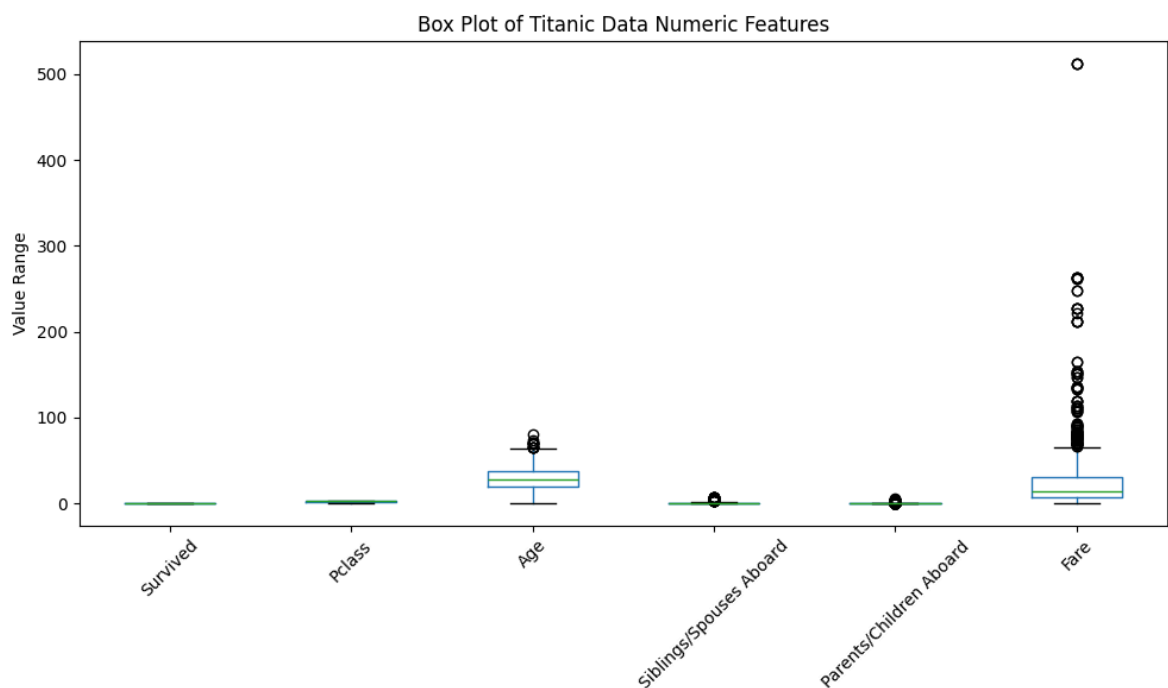
Out[2]:

Titanic Spread &amp; Distribution Summary

	Variance	Std Dev	Min	Q1	Q2 (Median)	Q3
<b>Survived</b>	0.240000	0.490000	0.000000	0.000000	0.000000	1.000000
<b>Pclass</b>	0.700000	0.840000	1.000000	2.000000	3.000000	3.000000
<b>Age</b>	199.430000	14.120000	0.420000	20.250000	28.000000	38.000000
<b>Siblings/Spouses Aboard</b>	1.220000	1.100000	0.000000	0.000000	0.000000	1.000000
<b>Parents/Children Aboard</b>	0.650000	0.810000	0.000000	0.000000	0.000000	0.000000
<b>Fare</b>	2478.250000	49.780000	0.000000	7.920000	14.450000	31.140000

From Challenge 7 - Display the data with a box-whisker plot

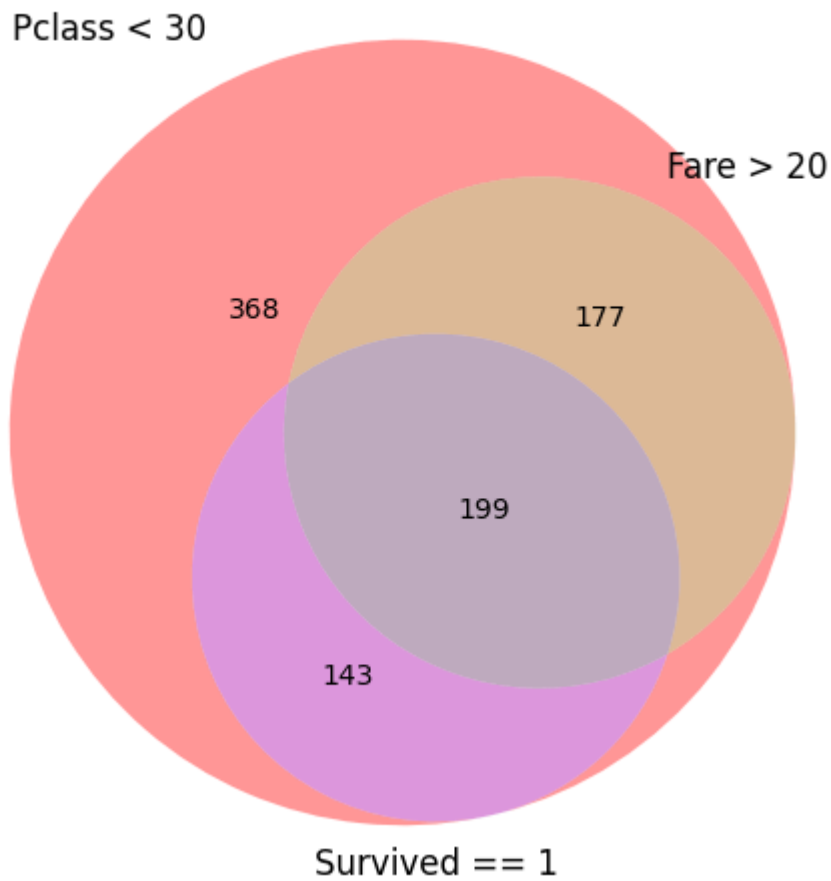
In [3]: stats.boxPlotData()



From Challenge 7 - Display the data with a Venn Diagram

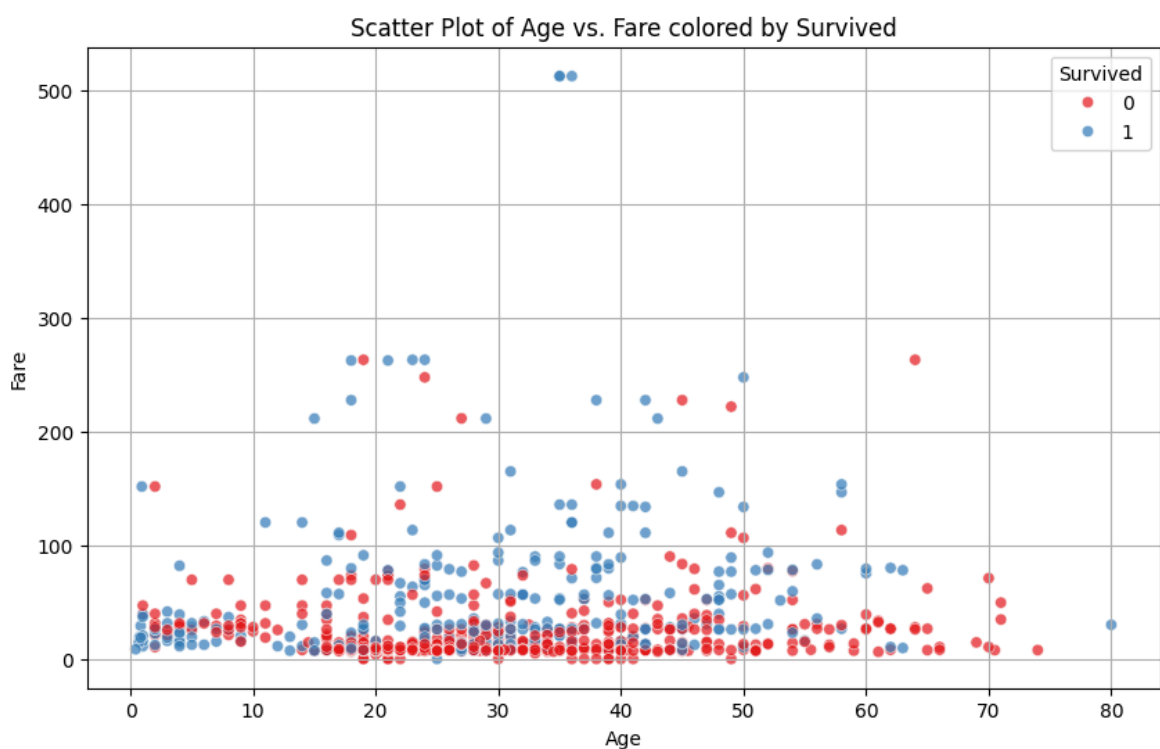
In [4]: stats.vennDiagramData(col1='Pclass', col2='Fare', col3='Survived', threshold1=30

## Venn Diagram: Overlap of Pclass, Fare, and Survived



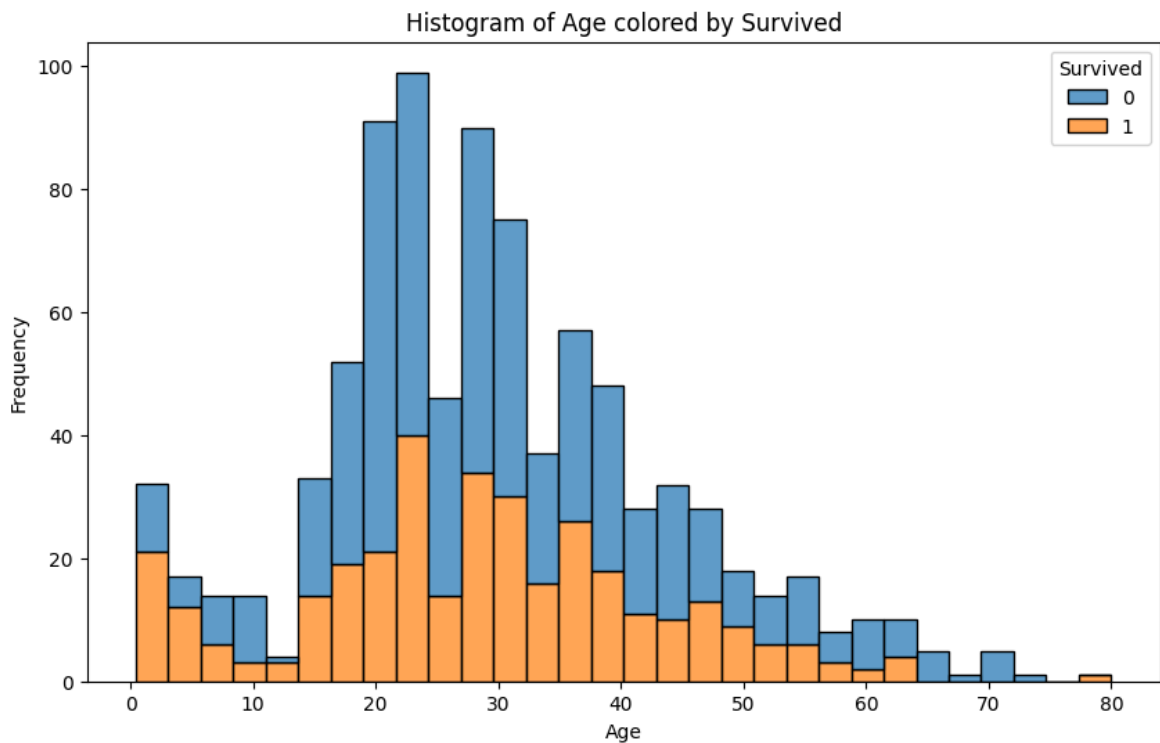
From Challenge 7 - Display the data with a scatter plot

```
In [5]: stats.scatterPlotDataSeaborn(x_col='Age', y_col='Fare', hue_col='Survived')
```



From Challenge 7 - Display the data with a Histogram

```
In [6]: stats.histogramDataSeaborn(column='Age', bins=30, hue_col='Survived')
```



SUMMARY:-

- count: number of non-null entries
- mean, std, min, max
- 25%, 50% (median), 75% percentiles
- For categorical columns: unique values, top value, and frequency

```
In [7]: stats.getSummary().style.set_caption("📊 Titanic Summary Statistics")
```

Out[7]: 📊 Titanic Summary Statistics

	count	unique	top	freq	mean	std	min	max
<b>Survived</b>	887.000000	nan	nan	nan	0.385569	0.487004	0.000000	0.000000
<b>Pclass</b>	887.000000	nan	nan	nan	2.305524	0.836662	1.000000	2.000000
<b>Name</b>	887	887	Mr. Owen Harris Braund	1	nan	nan	nan	nan
<b>Sex</b>	887	2	male	573	nan	nan	nan	nan
<b>Age</b>	887.000000	nan	nan	nan	29.471443	14.121908	0.420000	20.000000
<b>Siblings/Spouses Aboard</b>	887.000000	nan	nan	nan	0.525366	1.104669	0.000000	0.000000
<b>Parents/Children Aboard</b>	887.000000	nan	nan	nan	0.383315	0.807466	0.000000	0.000000
<b>Fare</b>	887.000000	nan	nan	nan	32.305420	49.782040	0.000000	7.000000

