

```
1 !pip install dataprep
```

```
1 from dataprep.datasets import load_dataset  
2 from dataprep.eda import create_report, plot  
3 df = load_dataset("titanic")  
4 create_report(df)
```



# Overview

Dataset Statistics		Variable Types	
Number of Variables	15	Numerical	6
Number of Rows	891	Categorical	9
Missing Cells	869		
Missing Cells (%)	6.5%		
Duplicate Rows	107		
Duplicate Rows (%)	12.0%		
Total Size in Memory	398.1 KB		
Average Row Size in Memory	457.6 B		

# Variables

survived


numerical

Show Details

Distinct Count	2
Unique (%)	0.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%

Mean	0.3838
Minimum	0
Maximum	1
Zeros	549
Zeros (%)	61.6%
Negatives	0

survived



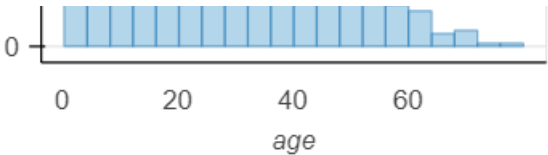
pclass	Frequency
1	220
2	180
3	480

sex	Count
male	570
female	320

A histogram showing the frequency distribution of the 'age' variable. The x-axis is labeled 'age' and the y-axis is labeled 'Frequency'. The distribution is unimodal and slightly right-skewed, with a peak frequency of approximately 95 at age 25.

Infinite (%)	0.0%
Memory Size	11.2 KB

negatives	0
Negatives (%)	0.0%

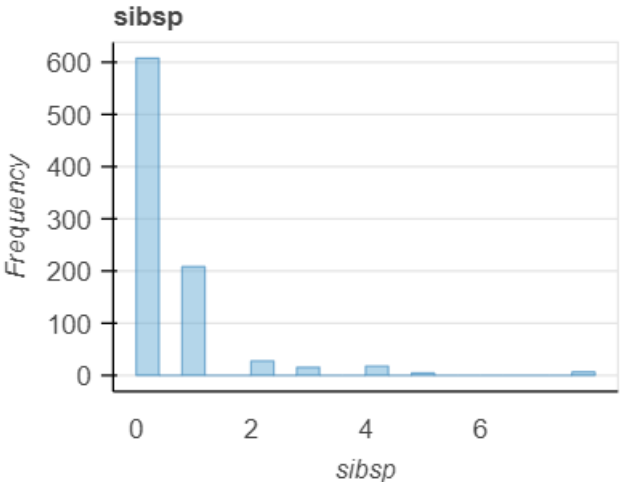


**sibsp**  
numerical

Show Details

Distinct Count	7
Unique (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	13.9 KB

Mean	0.523
Minimum	0
Maximum	8
Zeros	608
Zeros (%)	68.2%
Negatives	0
Negatives (%)	0.0%

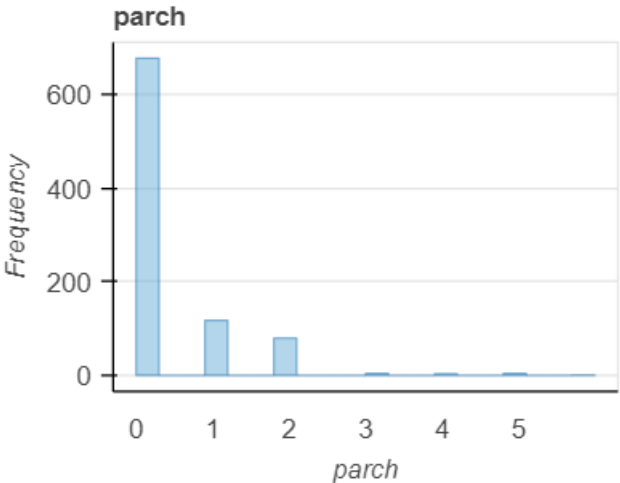


**parch**  
numerical

Show Details

Distinct Count	7
Unique (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	13.9 KB

Mean	0.3816
Minimum	0
Maximum	6
Zeros	678
Zeros (%)	76.1%
Negatives	0
Negatives (%)	0.0%

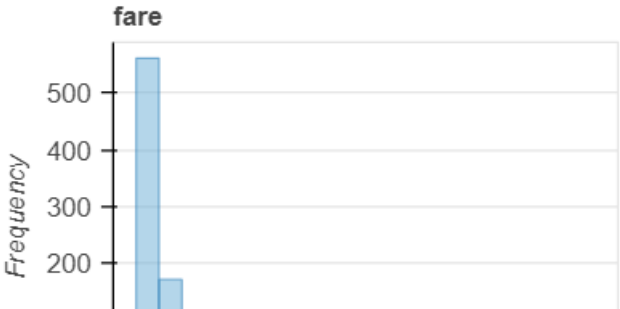


**fare**  
numerical

Show Details

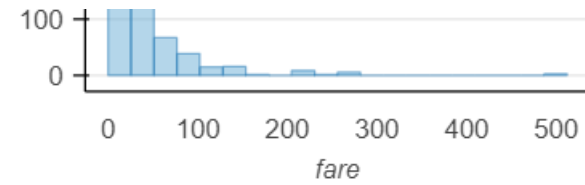
Distinct Count	248
Unique (%)	27.8%
Missing	0
Missing (%)	0.0%

Size	
Mean	32.2042
Minimum	0
Maximum	512.3292
Zeros	15
Zeros (%)	1.7%



Infinite	0
Infinite (%)	0.0%
Memory	13.9 KB

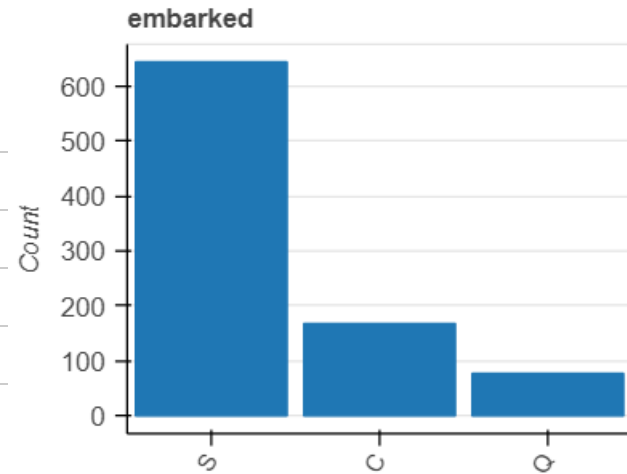
Negatives	0
Negatives (%)	0.0%



**embarked**  
categorical

Show Details

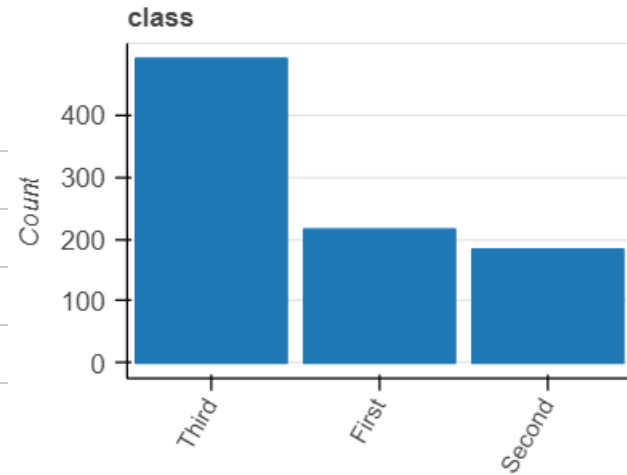
Distinct Count	3
Unique (%)	0.3%
Missing	2
Missing (%)	0.2%
Memory Size	57.3 KB



**class**  
categorical

Show Details

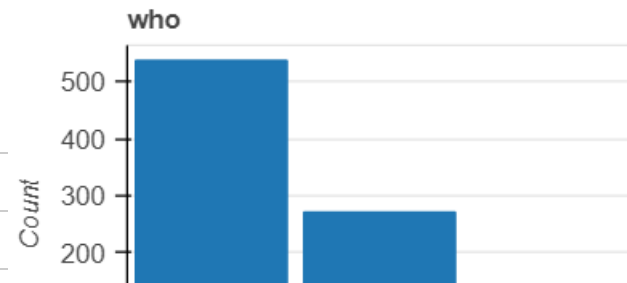
Distinct Count	3
Unique (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory Size	61.1 KB



**who**  
categorical

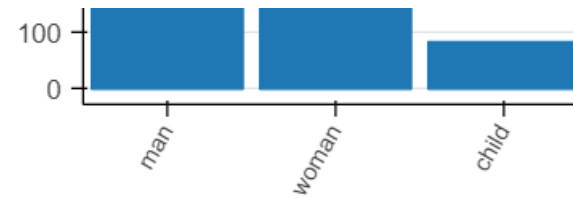
Show Details

Distinct Count	3
Unique (%)	0.3%
Missing	0

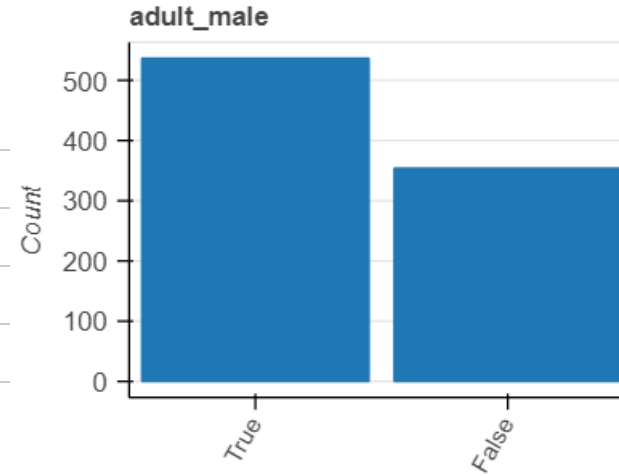


[Show Details](#)

<b>Missing (%)</b>	0.0%
<b>Memory Size</b>	59.9 KB

**adult\_male**  
categorical[Show Details](#)

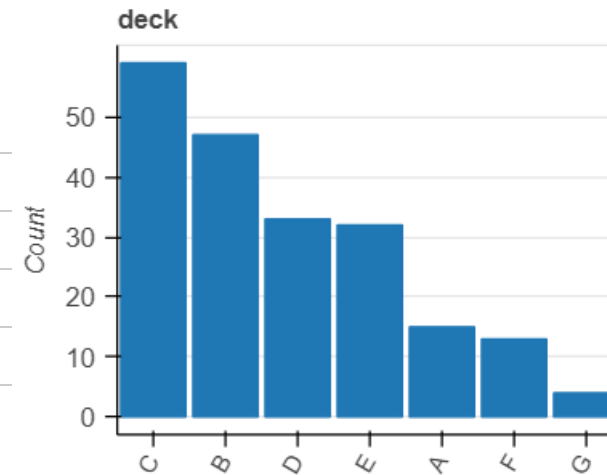
<b>Distinct Count</b>	2
<b>Unique (%)</b>	0.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory Size</b>	53.5 KB

**deck**

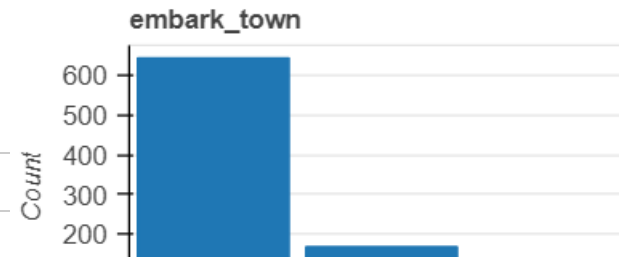
categorical

[Show Details](#)

<b>Distinct Count</b>	7
<b>Unique (%)</b>	3.4%
<b>Missing</b>	688
<b>Missing (%)</b>	77.2%
<b>Memory Size</b>	13.1 KB

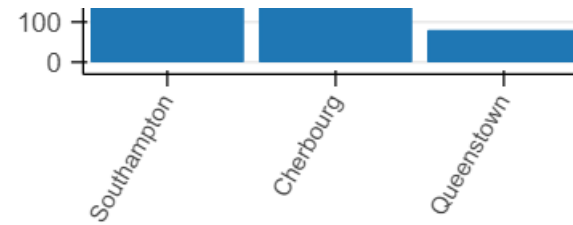
**embark\_town**  
categorical

<b>Distinct Count</b>	3
<b>Unique (%)</b>	0.3%
<b>Missing</b>	2



[Show Details](#)

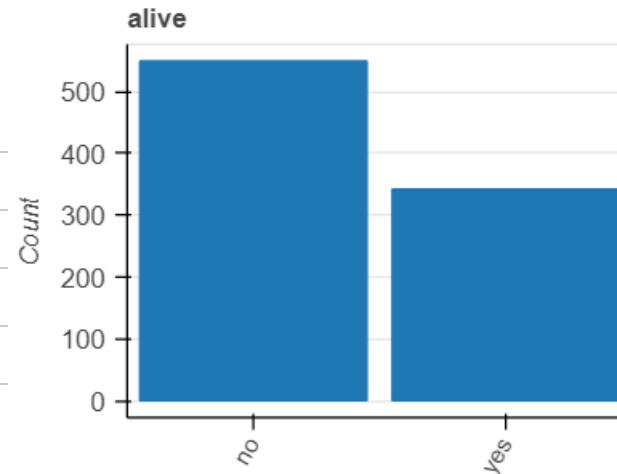
<b>Missing (%)</b>	0.2%
<b>Memory Size</b>	65.6 KB

**alive**

categorical

[Show Details](#)

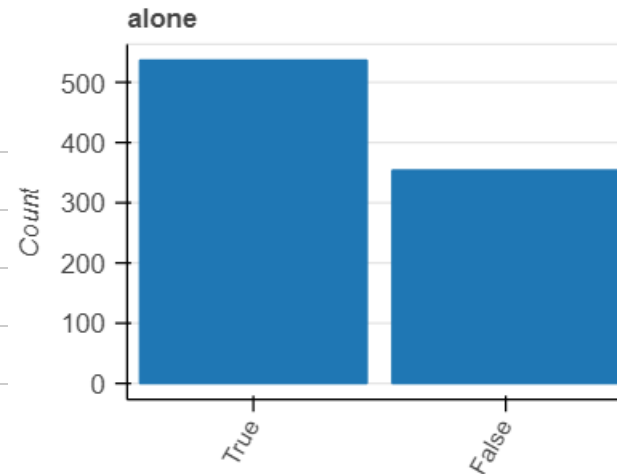
<b>Distinct Count</b>	2
<b>Unique (%)</b>	0.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory Size</b>	58.6 KB

**alone**

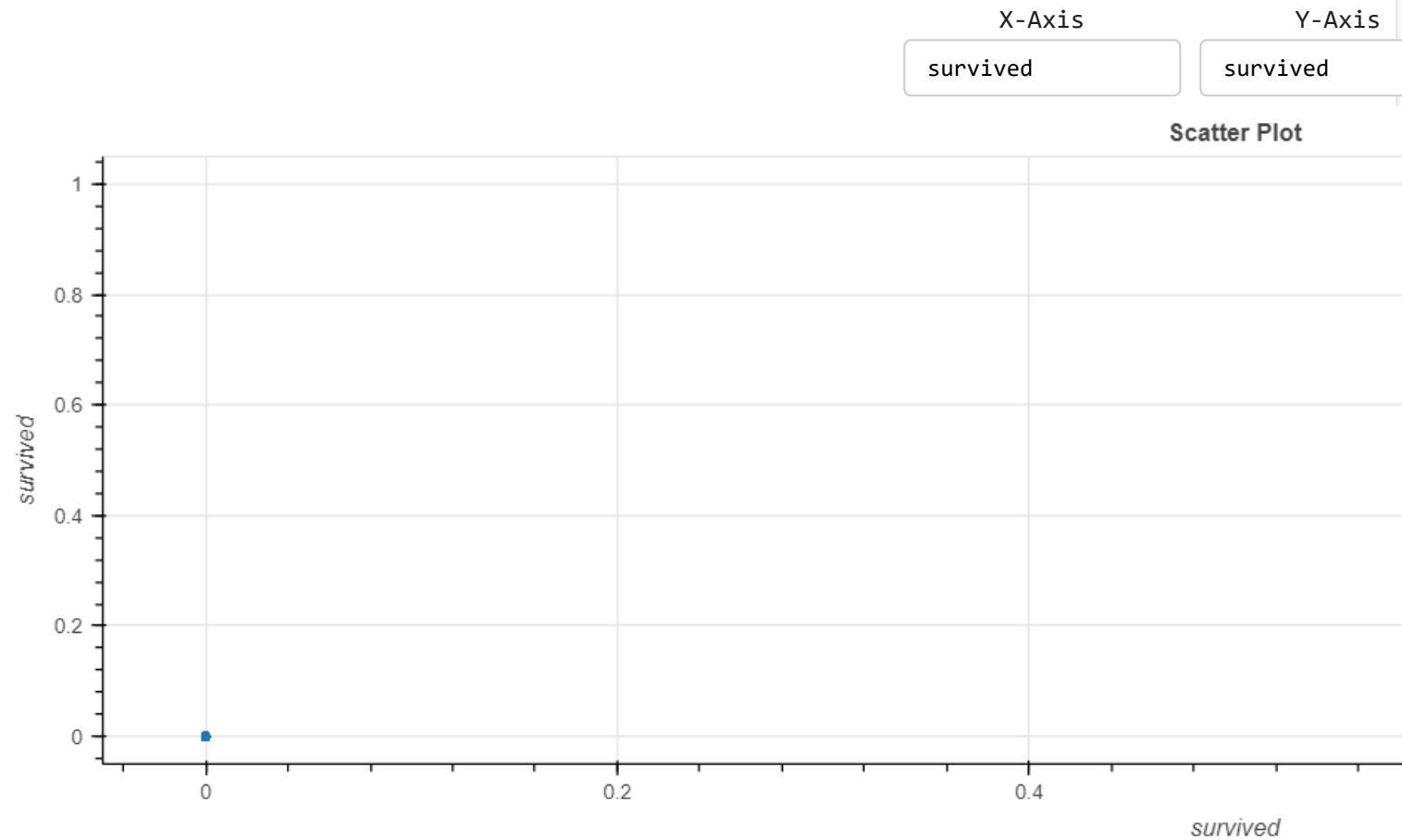
categorical

[Show Details](#)

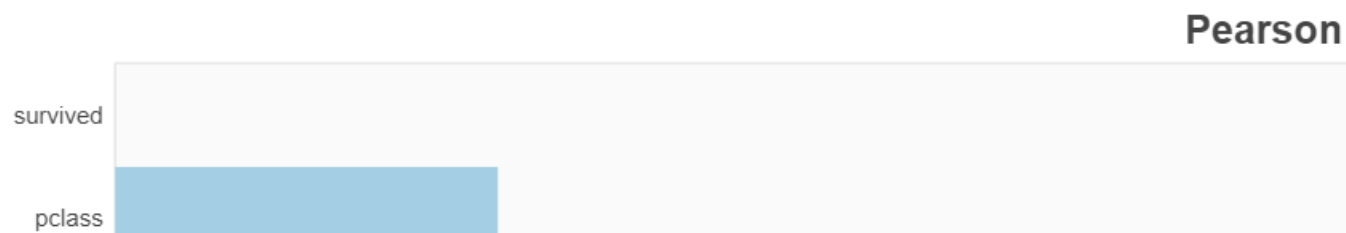
<b>Distinct Count</b>	2
<b>Unique (%)</b>	0.2%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory Size</b>	53.5 KB



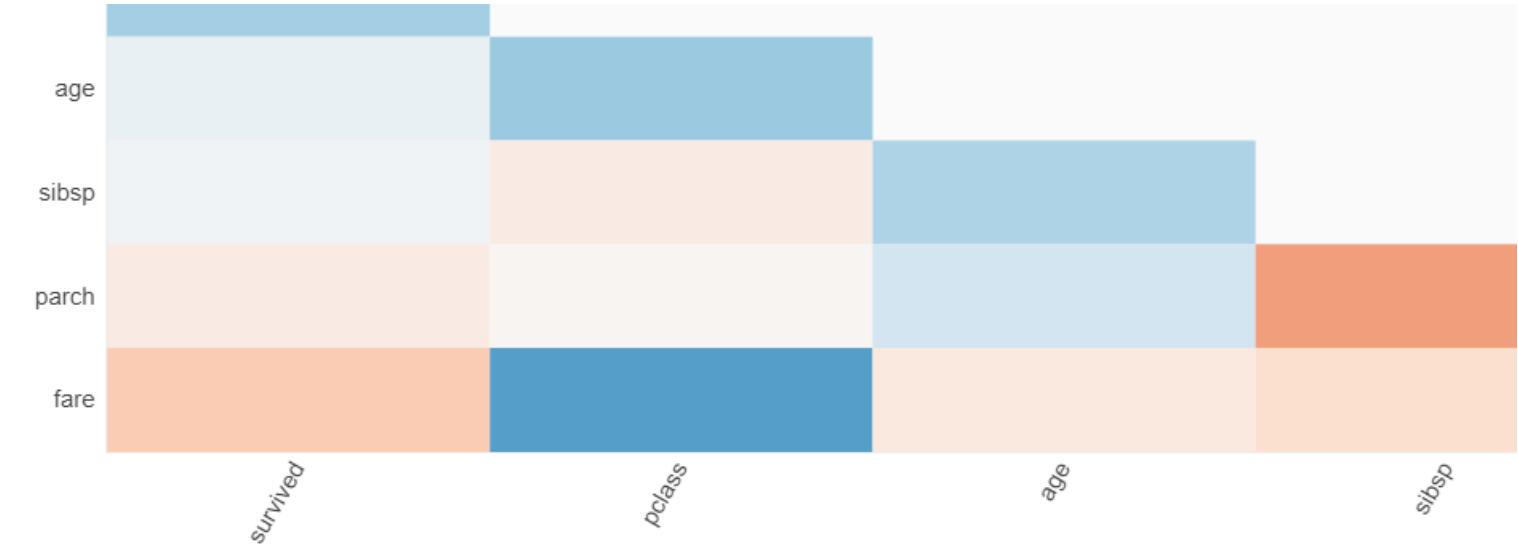
## Interactions



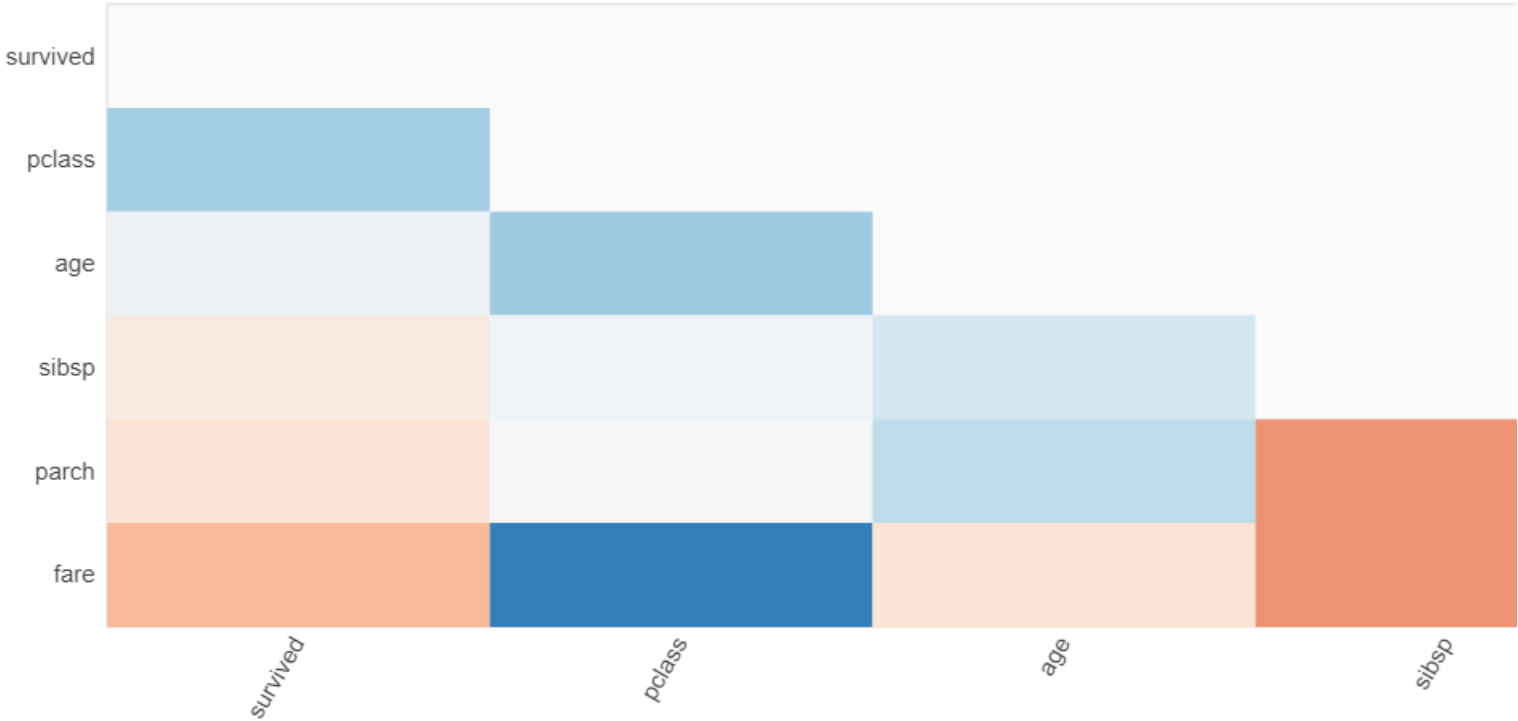
## Correlations



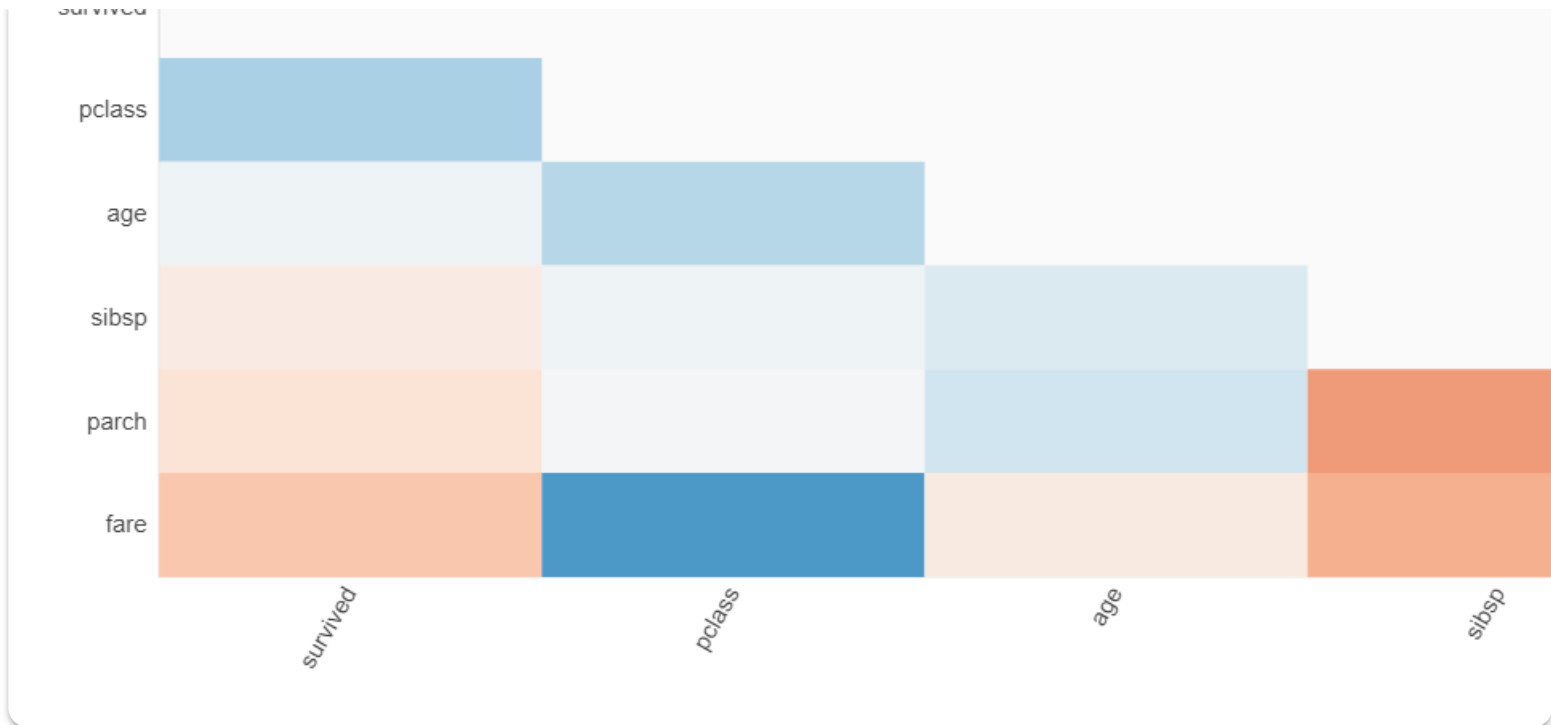




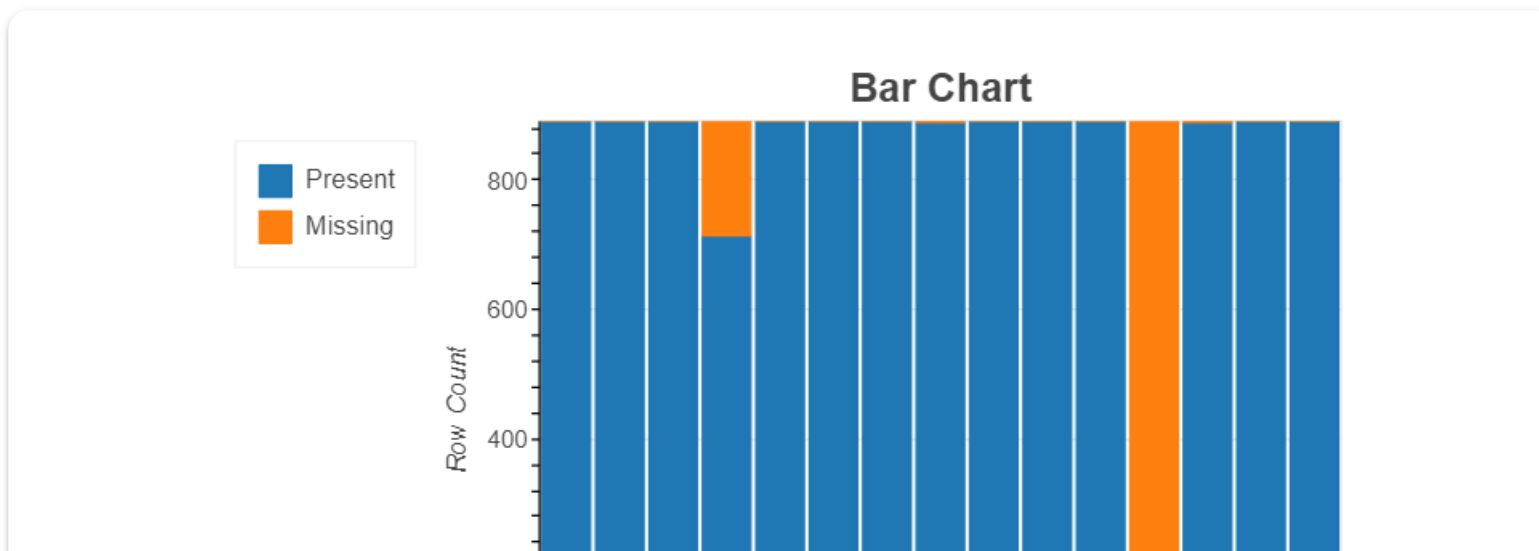
Spearman

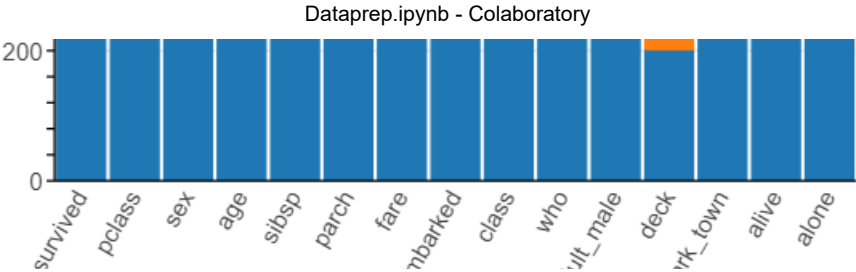


KendallTau



## Missing Values

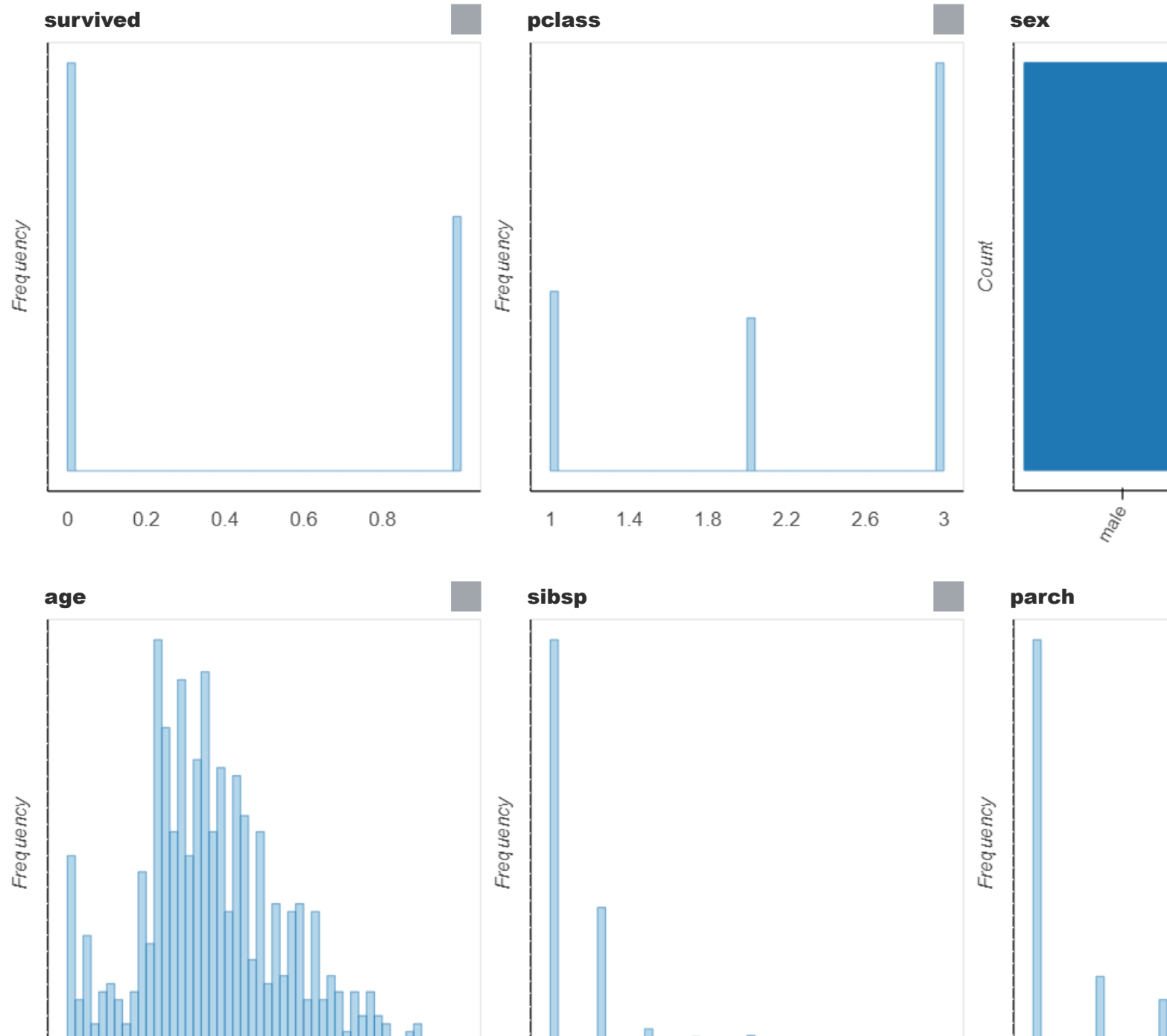


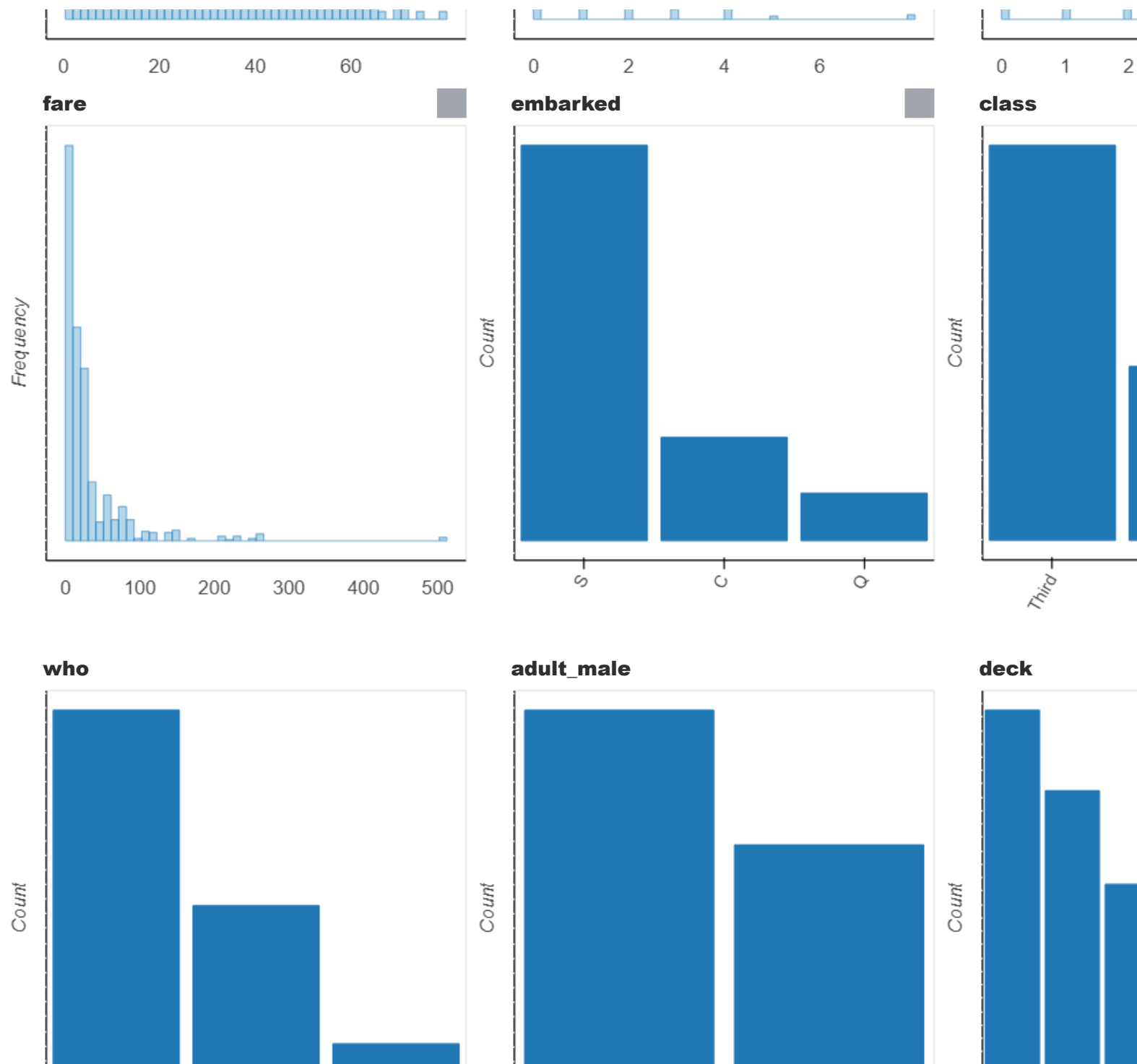


1

1 plot(df)

## Show Stats and Insights





```
1 from dataprep.eda.missing import plot_missing
2 plot_missing(df)
```

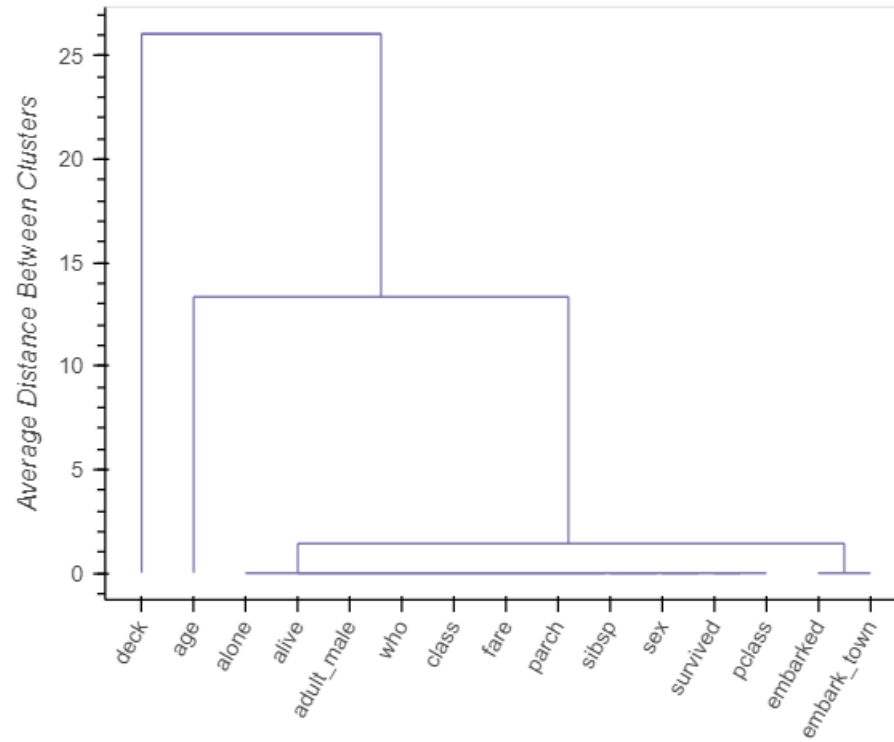
Stats

Bar Chart

Spectrum

Heatmap

Dendrogram



1

---

✓ 1s completed at 5:57 PM

