

```
In [3]: 1 from dataprep.datasets import load_dataset
        2 from dataprep.eda import create_report
        3 df = load_dataset("titanic")
        4 create_report(df)
```

NumExpr defaulting to 8 threads.

Out[3]:

DataPrep Report

Overview

Variables

Interactions

Correlations

Missing Values

# Overview

Dataset Statistics	
Number of Variables	12
Number of Rows	891
Missing Cells	866
Missing Cells (%)	8.1%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	315.7 KB
Average Row Size in Memory	362.8 B
Variable Types	Numerical: 3 Categorical: 9

Dataset Insights	
PassengerId is uniformly distributed	Uniform
Age has 177 (19.87%) missing values	Missing
Cabin has 687 (77.1%) missing values	Missing
Fare is skewed	Skewed
Name has a high cardinality: 891 distinct values	High Cardinality
Ticket has a high cardinality: 681 distinct values	High Cardinality
Cabin has a high cardinality: 147 distinct values	High Cardinality
Survived has constant length 1	Constant Length
Pclass has constant length 1	Constant Length
SibSp has constant length 1	Constant Length

1

2

# Variables

Sort by Feature order ☐ Reverse order

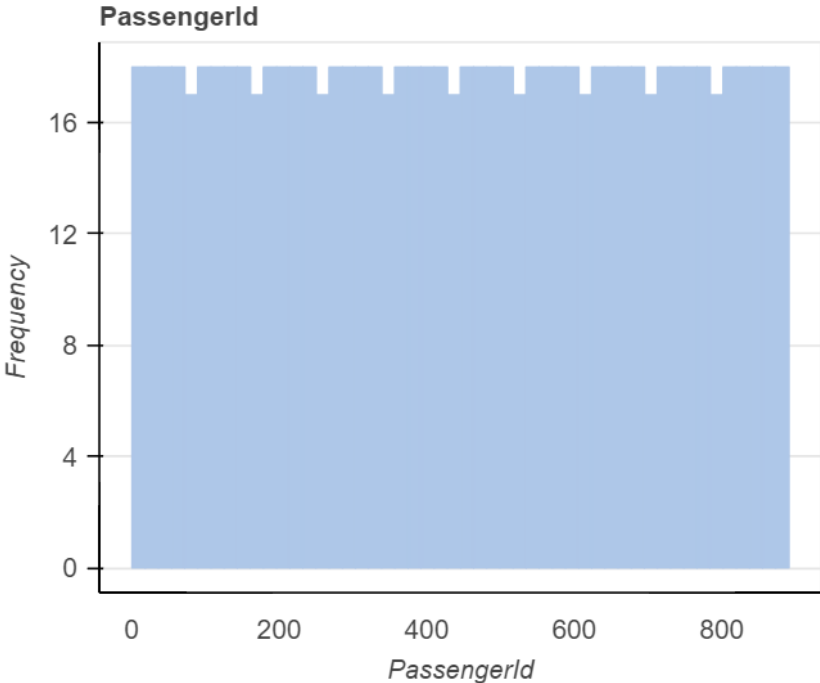
PassengerId

numerical

Show Details

Approximate Distinct Count	891
Approximate Unique (%)	100.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	14256

Mean	446
Minimum	1
Maximum	891
Zeros	0
Zeros (%)	0.0%
Negatives	0
Negatives (%)	0.0%

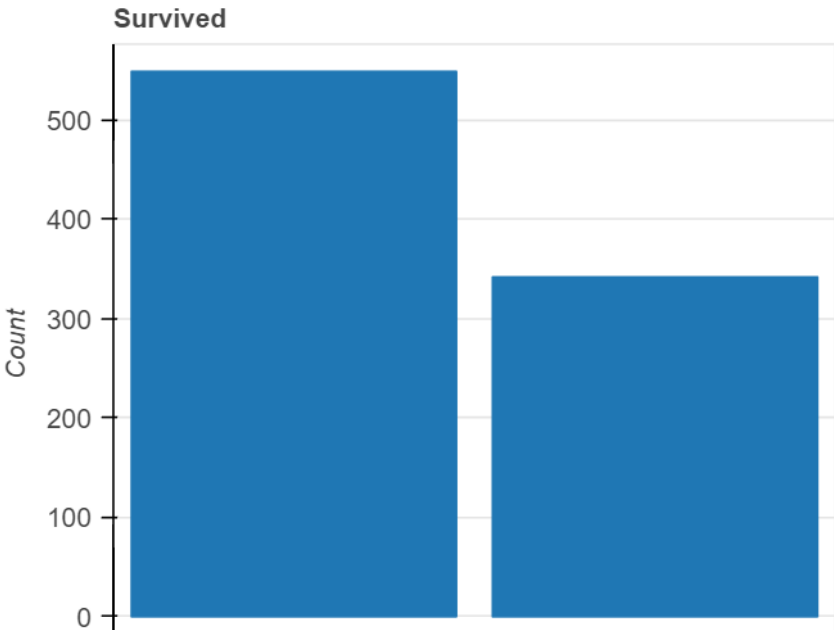


Survived

categorical

Show Details

Approximate Distinct Count	2
Approximate Unique (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory Size	58806

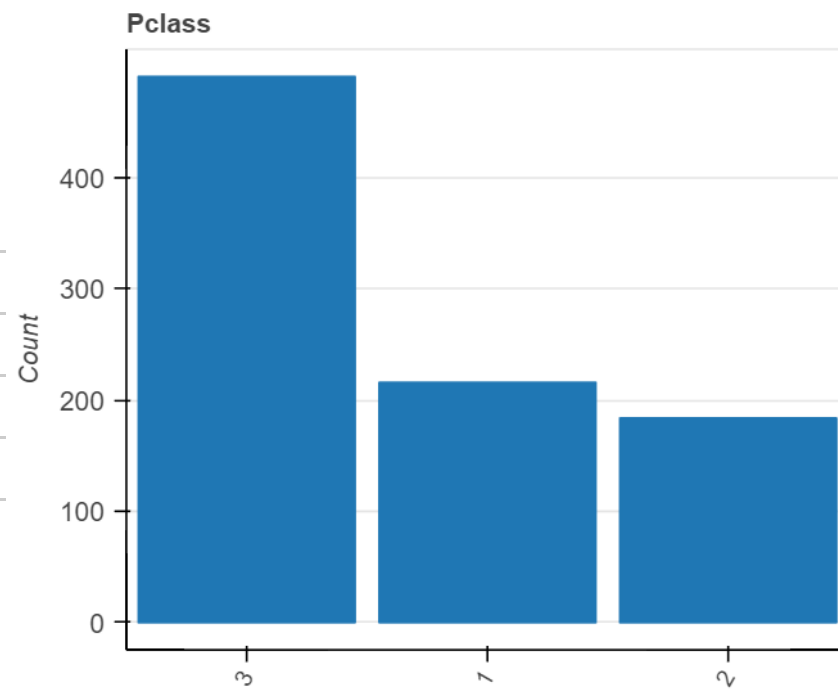


### Pclass

categorical

Show Details

Approximate Distinct Count	3
Approximate Unique (%)	0.3%
Missing	0
Missing (%)	0.0%
Memory Size	58806

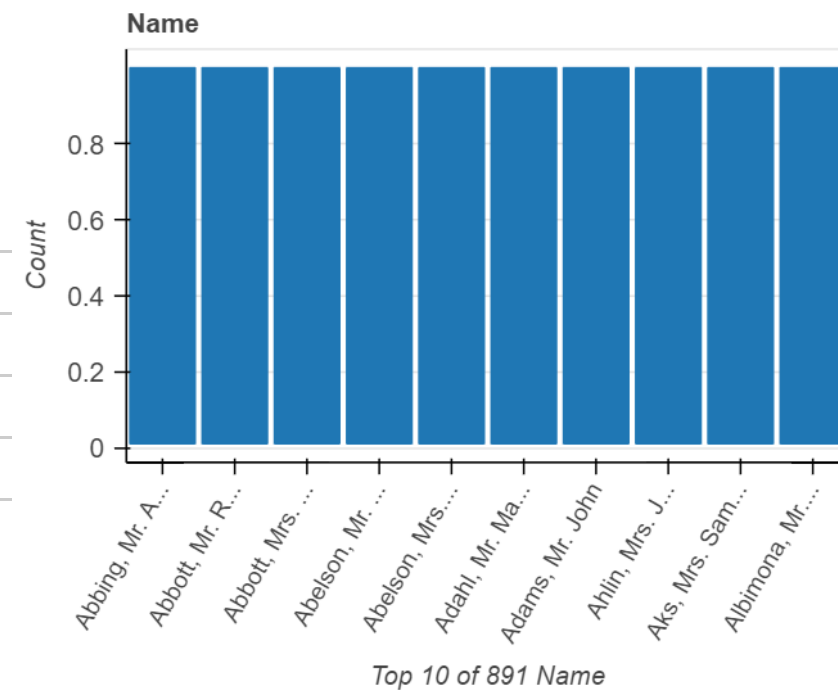


### Name

categorical

Show Details

Approximate Distinct Count	891
Approximate Unique (%)	100.0%
Missing	0
Missing (%)	0.0%
Memory Size	81941



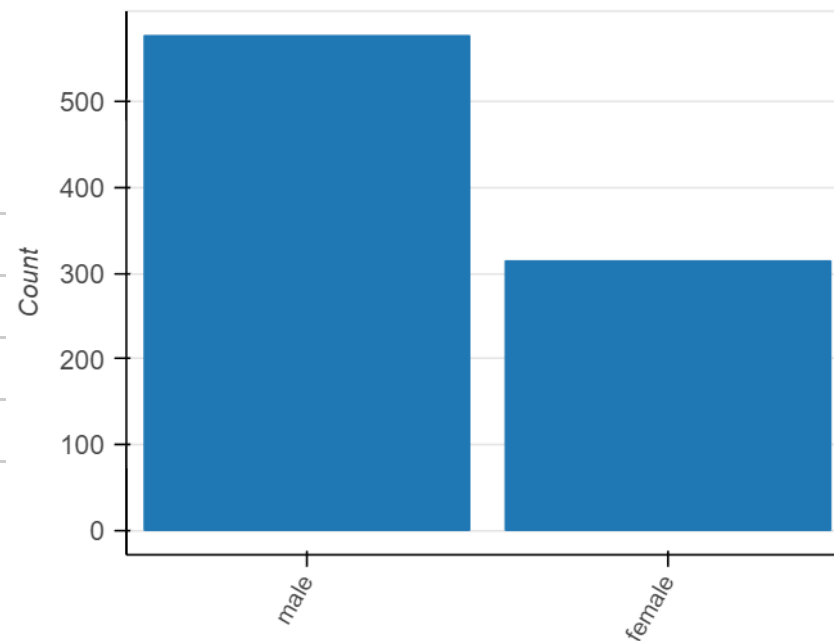
### Sex

## Sex

categorical

Show Details

Approximate Distinct Count	2
Approximate Unique (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory Size	62107



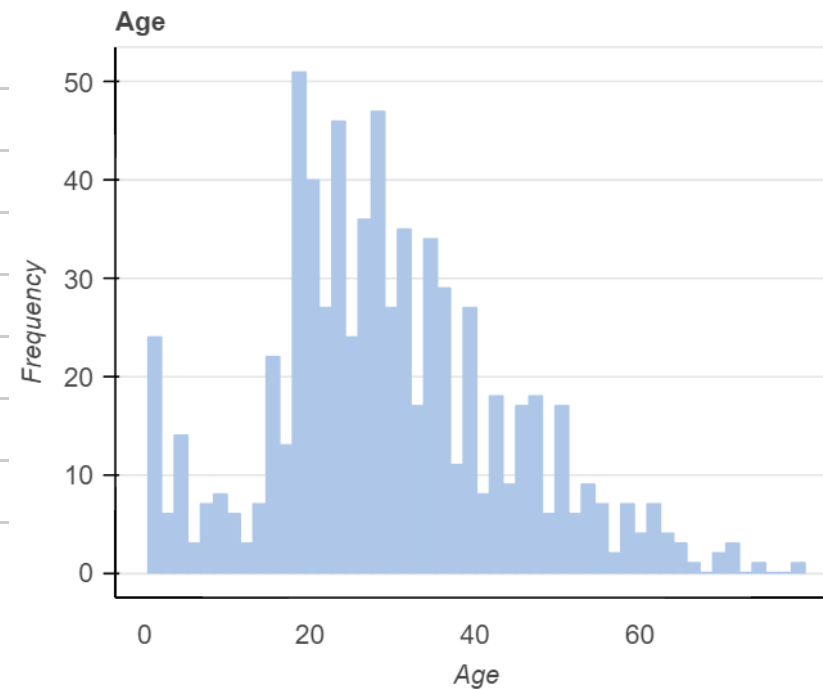
## Age

numerical

Show Details

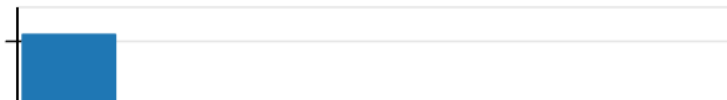
Approximate Distinct Count	88
Approximate Unique (%)	12.3%
Missing	177
Missing (%)	19.9%
Infinite	0
Infinite (%)	0.0%
Memory Size	11424

Mean	29.6991
Minimum	0.42
Maximum	80
Zeros	0
Zeros (%)	0.0%
Negatives	0
Negatives (%)	0.0%



## SibSp

600

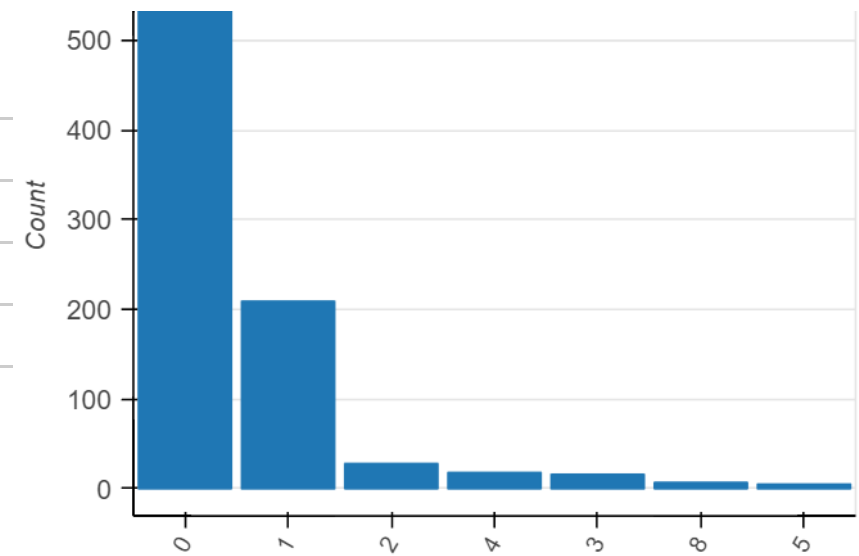


### SibSp

categorical

Show Details

Approximate Distinct Count	7
Approximate Unique (%)	0.8%
Missing	0
Missing (%)	0.0%
Memory Size	58806

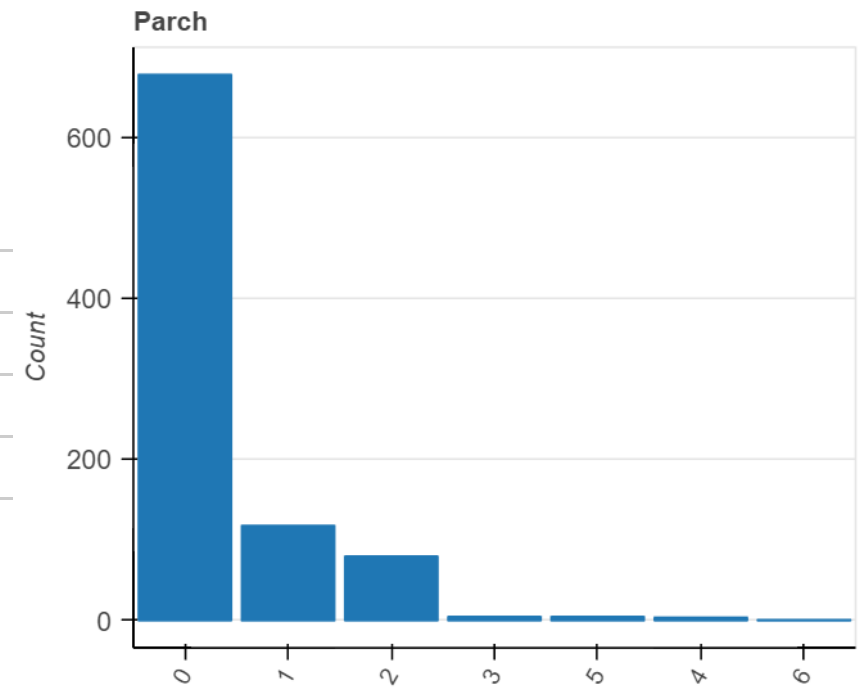


### Parch

categorical

Show Details

Approximate Distinct Count	7
Approximate Unique (%)	0.8%
Missing	0
Missing (%)	0.0%
Memory Size	58806



### Ticket

Approximate Distinct Count	681
----------------------------	-----

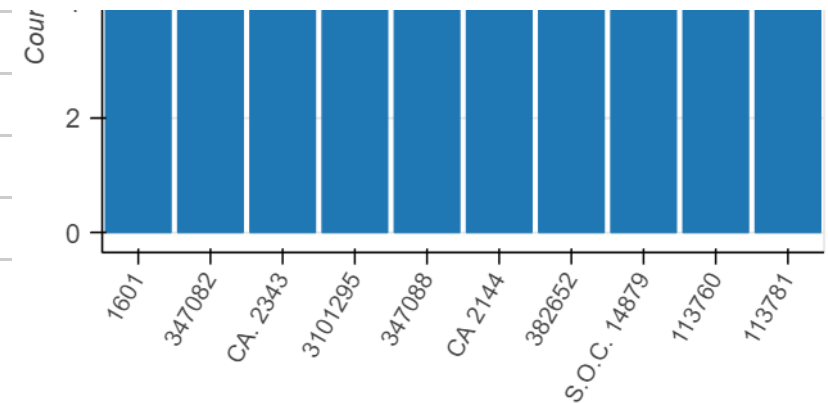


### Ticket

categorical

Show Details

Approximate Unique (%)	76.4%
Missing	0
Missing (%)	0.0%
Memory Size	63930



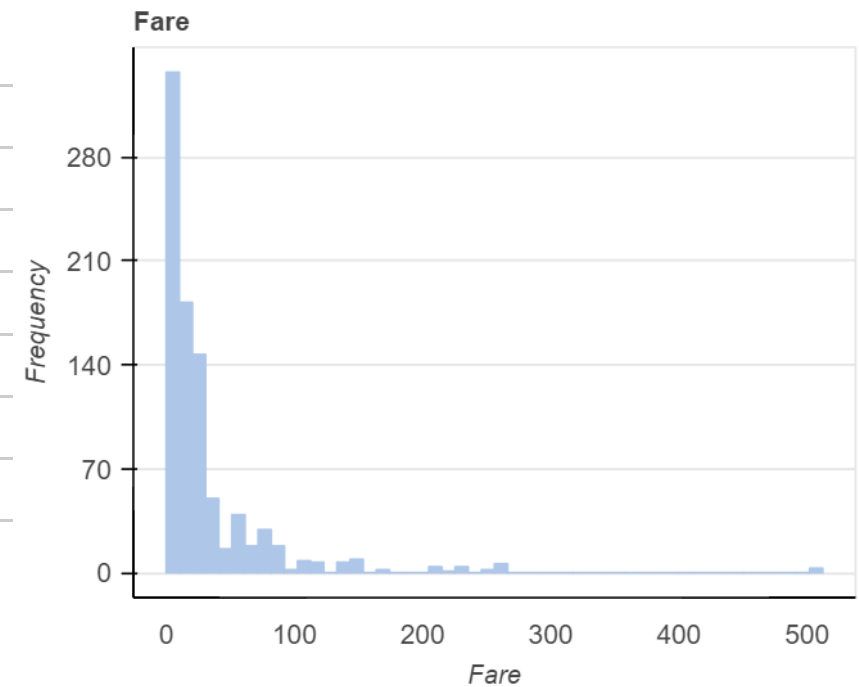
Top 10 of 681 Ticket

### Fare

numerical

Show Details

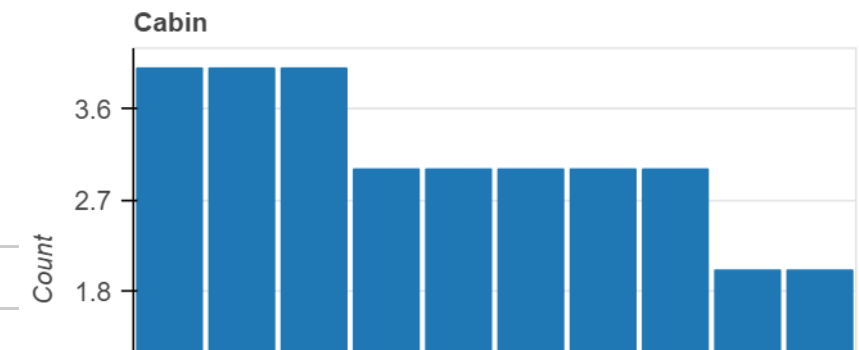
Approximate Distinct Count	248	Mean	32.2042
Approximate Unique (%)	27.8%	Minimum	0
Missing	0	Maximum	512.3292
Missing (%)	0.0%	Zeros	15
Infinite	0	Zeros (%)	1.7%
Infinite (%)	0.0%	Negatives	0
Memory Size	14256	Negatives (%)	0.0%



### Cabin

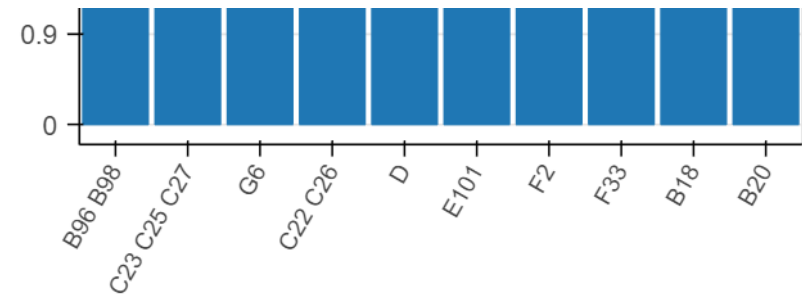
categorical

Approximate Distinct Count	147
Approximate Unique (%)	72.1%
Missing	687



Show Details

Missing (%)	77.1%
Memory Size	13992

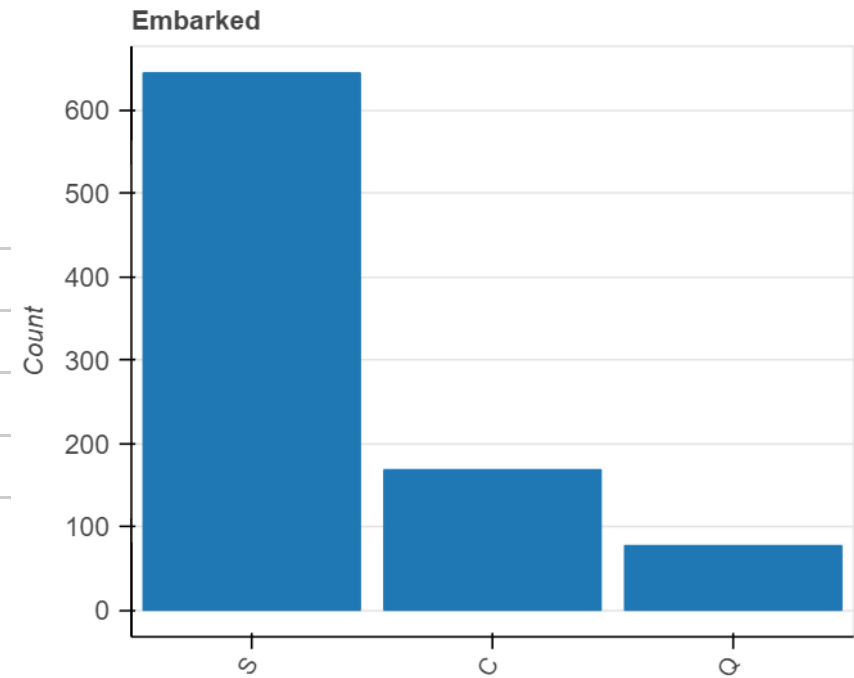


Top 10 of 147 Cabin

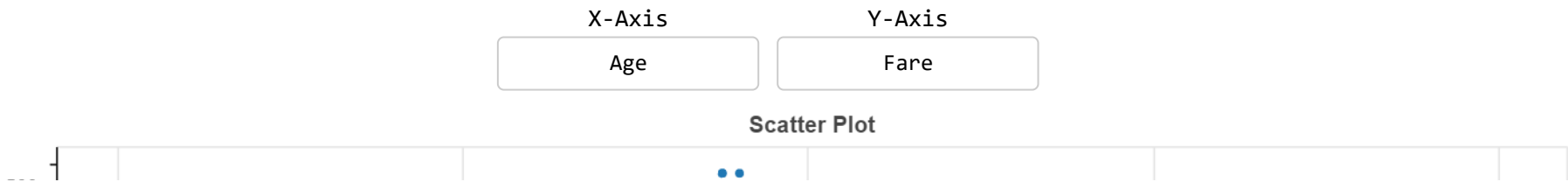
Embarked  
categorical

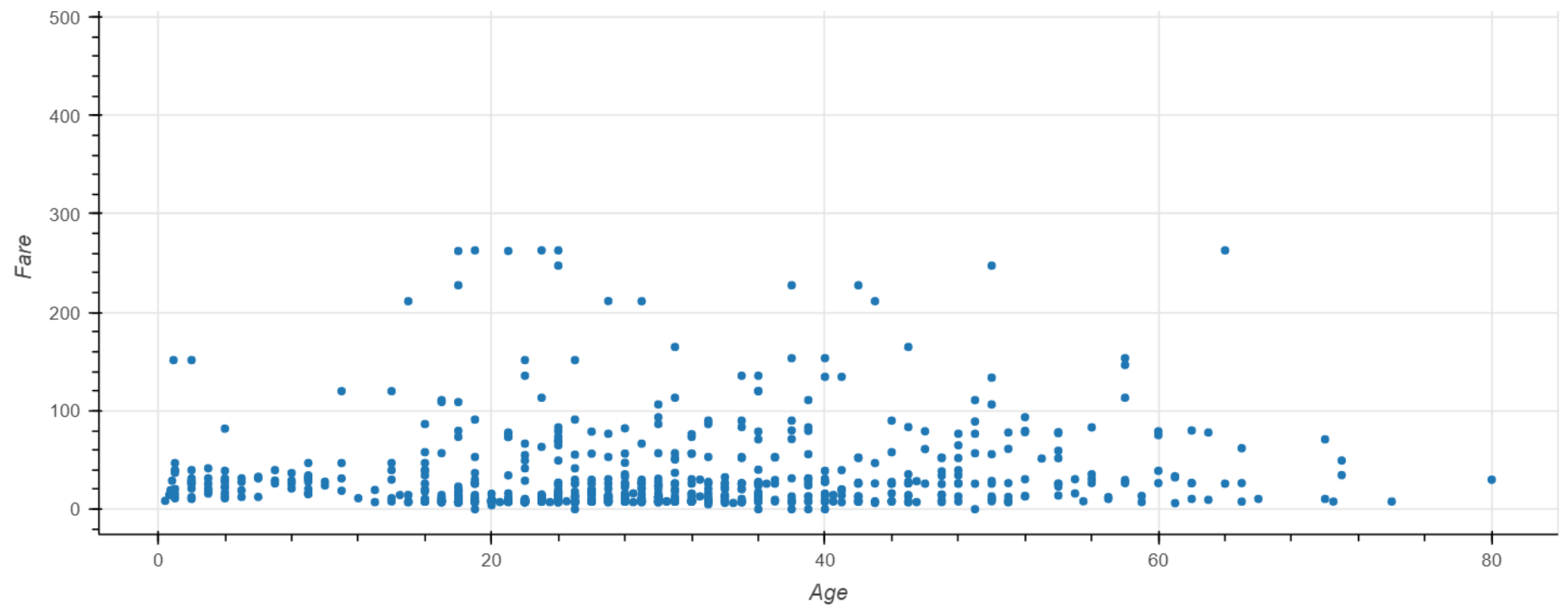
Show Details

Approximate Distinct Count	3
Approximate Unique (%)	0.3%
Missing	2
Missing (%)	0.2%
Memory Size	58674

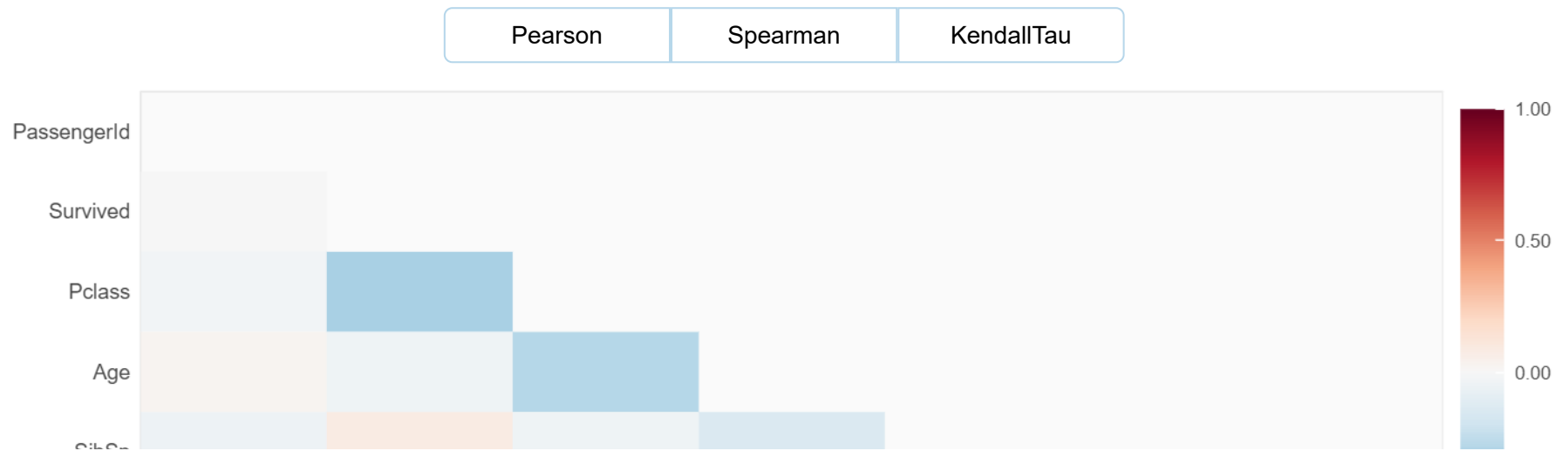


Interactions

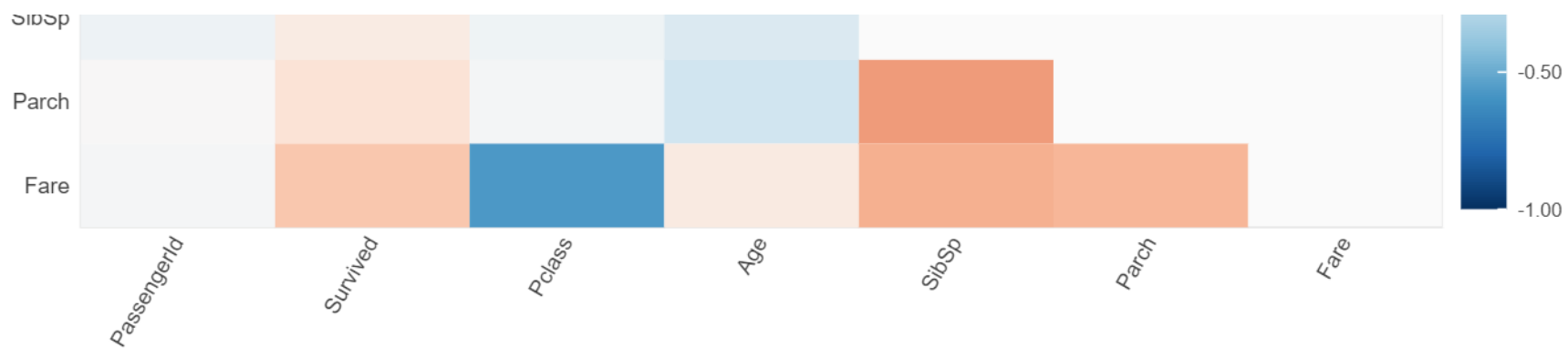




## Correlations







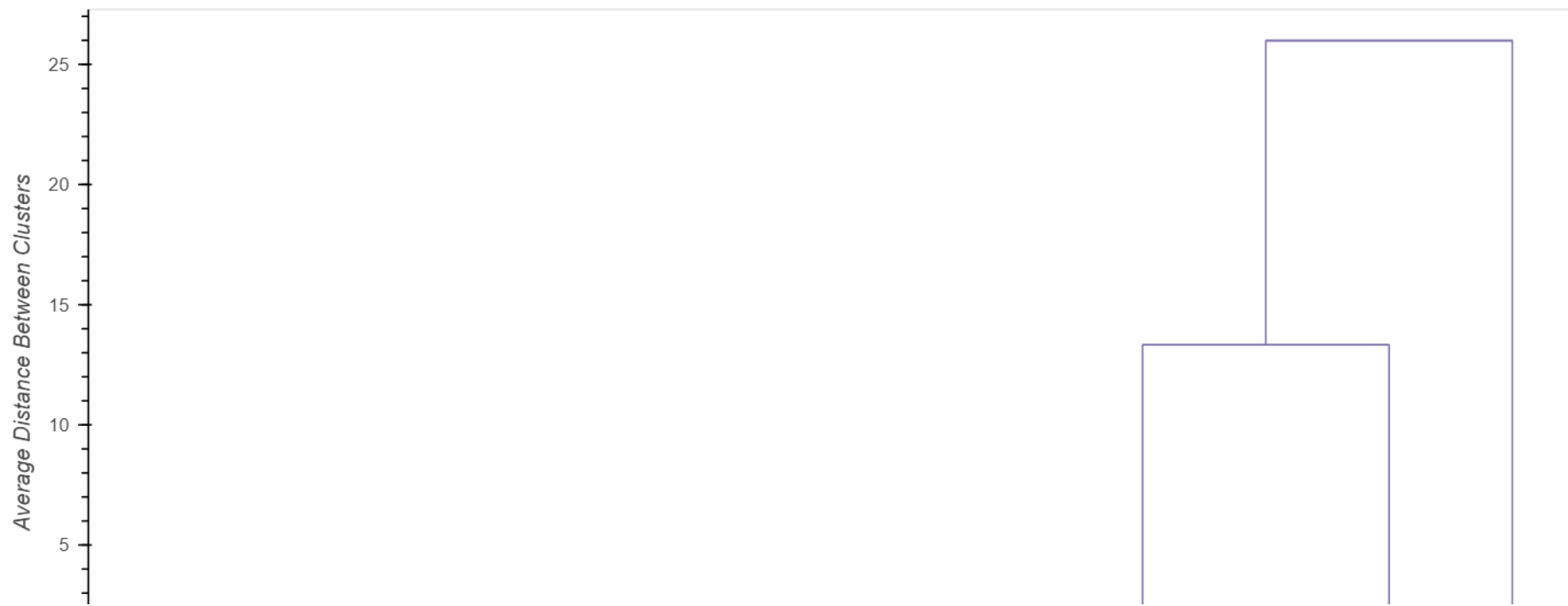
## Missing Values

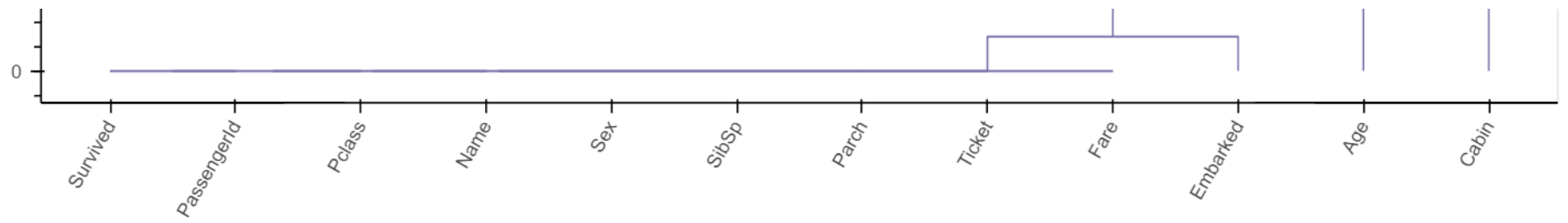
Bar Chart

Spectrum

Heat Map

Dendrogram





Report generated with [DataPrep \(https://dataprep.ai/\)](https://dataprep.ai/).

In [ ]: 1

In [ ]: 1