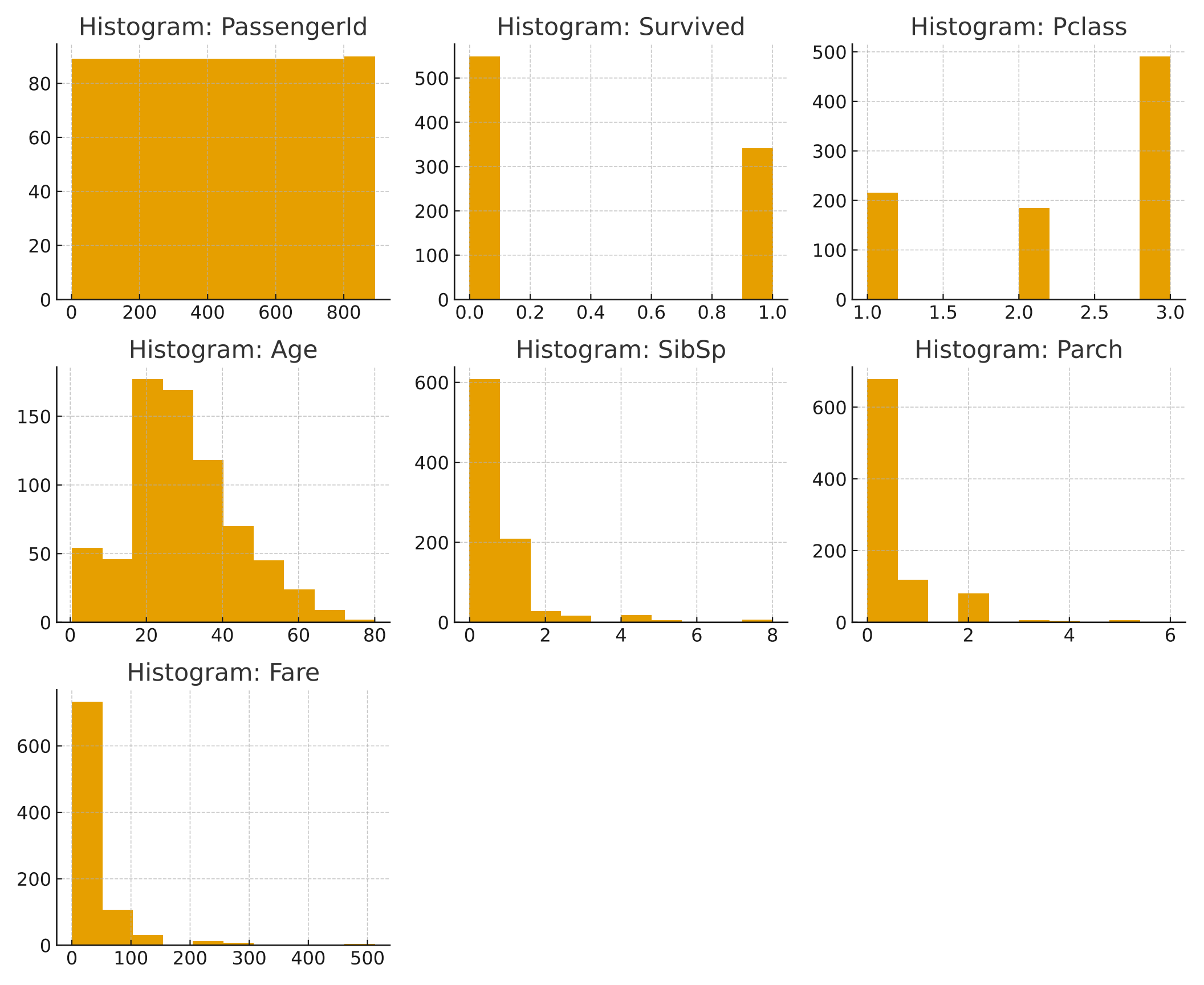
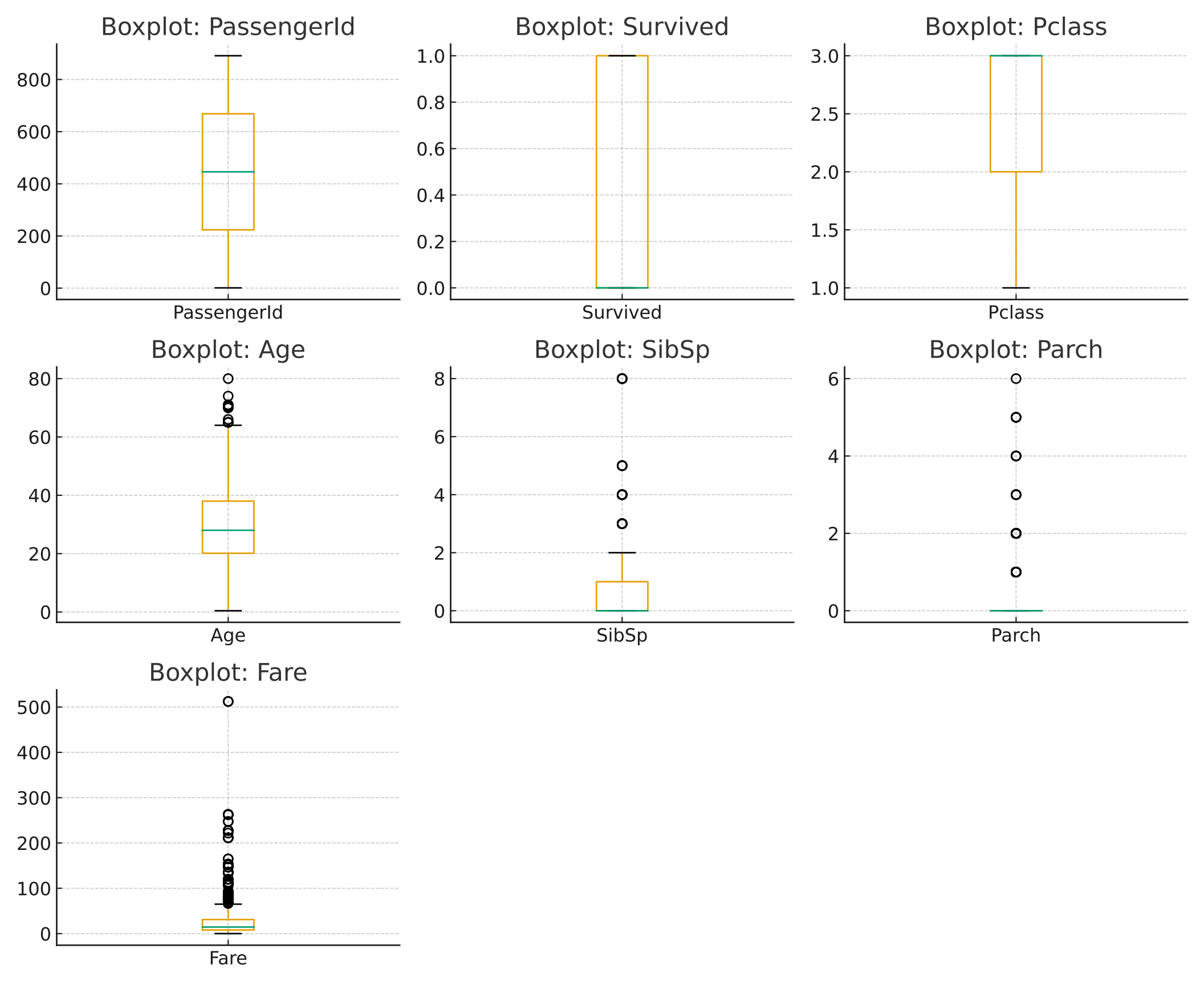
Exploratory Data Analysis Report

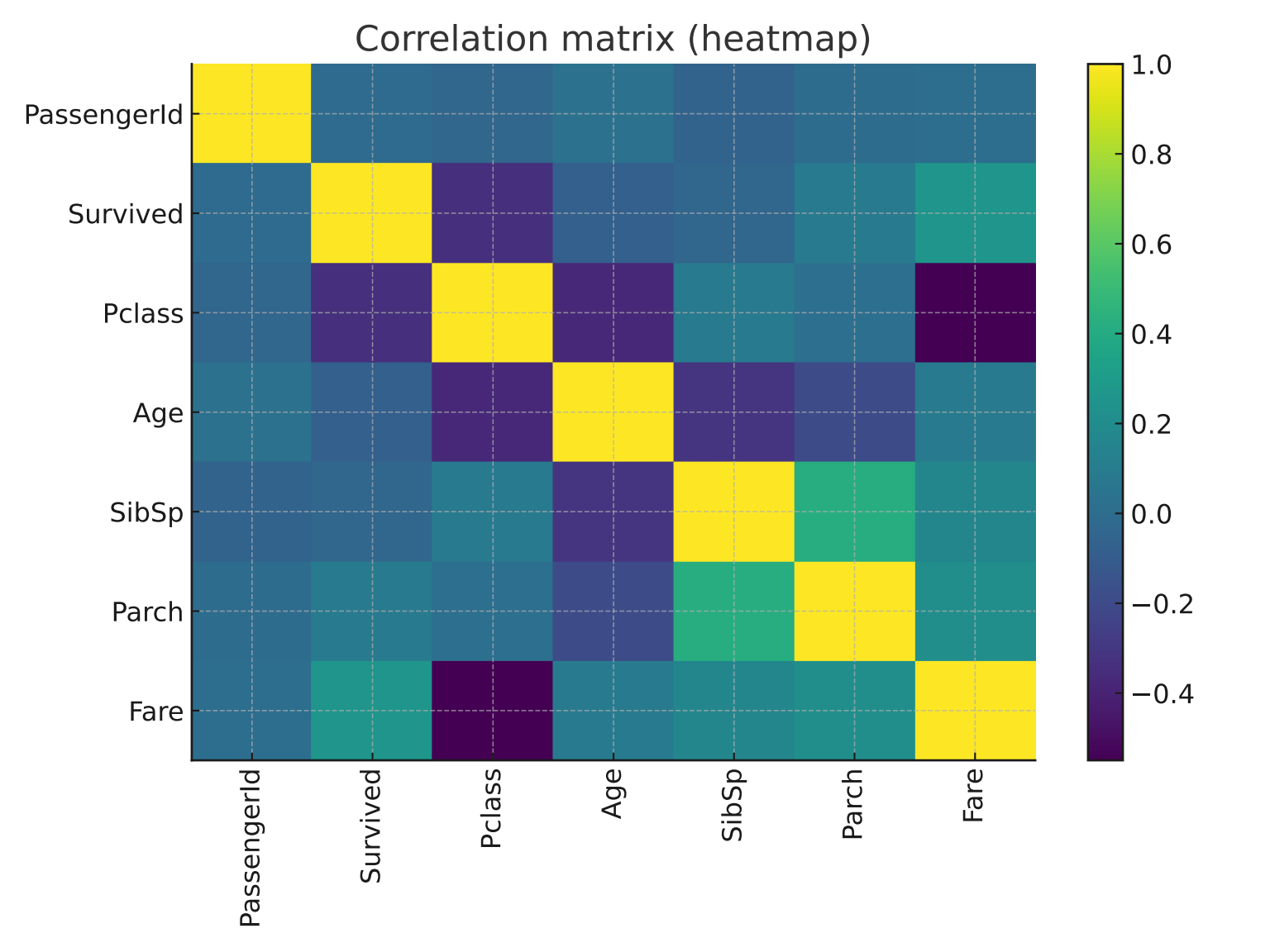
train.csv

Rows: 891 Columns: 12


Histogram: PassengerId
Histogram: Survived
Histogram: Pclass
Histogram: Age
Histogram: SibSp
Histogram: Parch
Histogram: Fare


Boxplot: PassengerId
PassengerId
Boxplot: Survived
Survived
Boxplot: Pclass
Pclass
Boxplot: Age
Age
Boxplot: SibSp
SibSp
Boxplot: Parch
Parch
Boxplot: Fare
Fare

Scatter matrix (pairwise relationships)


Correlation matrix (heatmap)
PassengerId
Survived
Pclass
Age
SibSp
Parch
Fare
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Top 10 rows sample

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr.	male	22.0	1	0	A/5 21171	7.25	nan	S
2	1	1	Cumings, M	female	38.0	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen,	female	26.0	0	0	STON/O2. 3	7.925	nan	S
4	1	1	Futrelle, Mr	female	35.0	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. W	male	35.0	0	0	373450	8.05	nan	S
6	0	3	Moran, Mr. J	male	nan	0	0	330877	8.4583	nan	Q
7	0	1	McCarthy, M	male	54.0	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Ma	male	2.0	3	1	349909	21.075	nan	S
9	1	3	Johnson, Mr	female	27.0	0	2	347742	11.1333	nan	S
10	1	2	Nasser, Mrs	female	14.0	1	0	237736	30.0708	nan	C

Dataset loaded from: /mnt/data/train.csv
Rows: 891, Columns: 12
Top observations:
Columns with missing values (count => %):
 - Cabin: 687 => 77.1%
 - Age: 177 => 19.87%
 - Embarked: 2 => 0.22%
Top absolute correlations (column pairs):
 - Fare vs Pclass: 0.55
 - SibSp vs Parch: 0.41
 - Pclass vs Age: 0.37
 - Survived vs Pclass: 0.34
 - SibSp vs Age: 0.31
 - Fare vs Survived: 0.26
 - Fare vs Parch: 0.22
 - Age vs Parch: 0.19
 - SibSp vs Fare: 0.16
 - Fare vs Age: 0.10