

Automating Exploratory Data Analysis

Accelerating insights with Pandas Profiling

```
891,0,3,  
Braund, Mr. Owen Harris,  
male,22,1,0,  
A/5 21171,7.25,,S  
  
Braund, Mr. Owen Harris,  
male,22,1,0,  
A/5 21171,7.5,,S  
  
plt.hist(df['Age'])  
sns.boxplot(x='Pclass', y='Fare', data=df)  
  
import pandas as pd  
df = pd.read_csv('titanic.csv')  
df.isnull().sum()  
plt.show()
```



Stop writing code for every single plot

In traditional EDA, you manually write code for univariate and bivariate analysis, asking basic questions one by one. This is time-consuming.

The Manual Approach

```
df.describe()  
df.info()  
plt.scatter(x='age', y='income')  
plt.hist(df['age'])  
sns.heatmap(df.corr())  
df.isnull().sum()  
df.isnull().sum()  
plt.fundgs(df['age'], 'ncome')  
sns.mearanap.scatter(x='age', y='income'), 'age']))  
plt.scatter(x='inage', y='Income')  
plt.hist(df['usge'])  
plt.skitrepor(df.ratio)
```

High Effort • Repetitive • Slow Insights !

The Pandas Profiling Approach

```
df.profile_report()
```



HTML Report

One Line of Code • Instant Report • Holistic View ✓

"Behind the scenes, the library does a lot of the heavy lifting for you."

Generating a complete report in four lines of code

```
# 1. Installation  
!pip install pandas-profiling
```

→ One-time setup

```
# 2. Import and Run  
from pandas_profiling import ProfileReport
```

```
# Create the report object  
prof = ProfileReport(df)
```

→ Pass your existing DataFrame (e.g., Titanic dataset)

```
# Export to interactive HTML  
prof.to_file(output_file='output.html')
```

→ Generates a standalone HTML file you can share

Note: The process may take a moment depending on dataset size, but the result is a standalone webpage containing deep analytics.

The Overview: A bird's-eye view of data health

Before analysing specific columns, you instantly grasp the size, complexity, and cleanliness of your data.

Dataset statistics	
Number of variables:	12
Number of observations:	891
Missing cells:	891 (8.1%)
Duplicate rows:	0 (0.0%)
Total size in memory:	83.7 KiB

Variable types	
Numeric:	5
Categorical:	7

Calculated automatically across the entire dataset.

Automated Alerts and Warnings

The tool proactively flags data quality issues so you don't have to hunt for them.

High Cardinality

NAME has a high cardinality: 891 distinct values.

Warning: Too many unique categories.

Missing

CABIN has 687 / 77.1% missing values.

Flagged as a candidate to potentially drop.

Zeros

FARE has 15 / 1.7% zeros.

Potential errors or free tickets.

Uniform

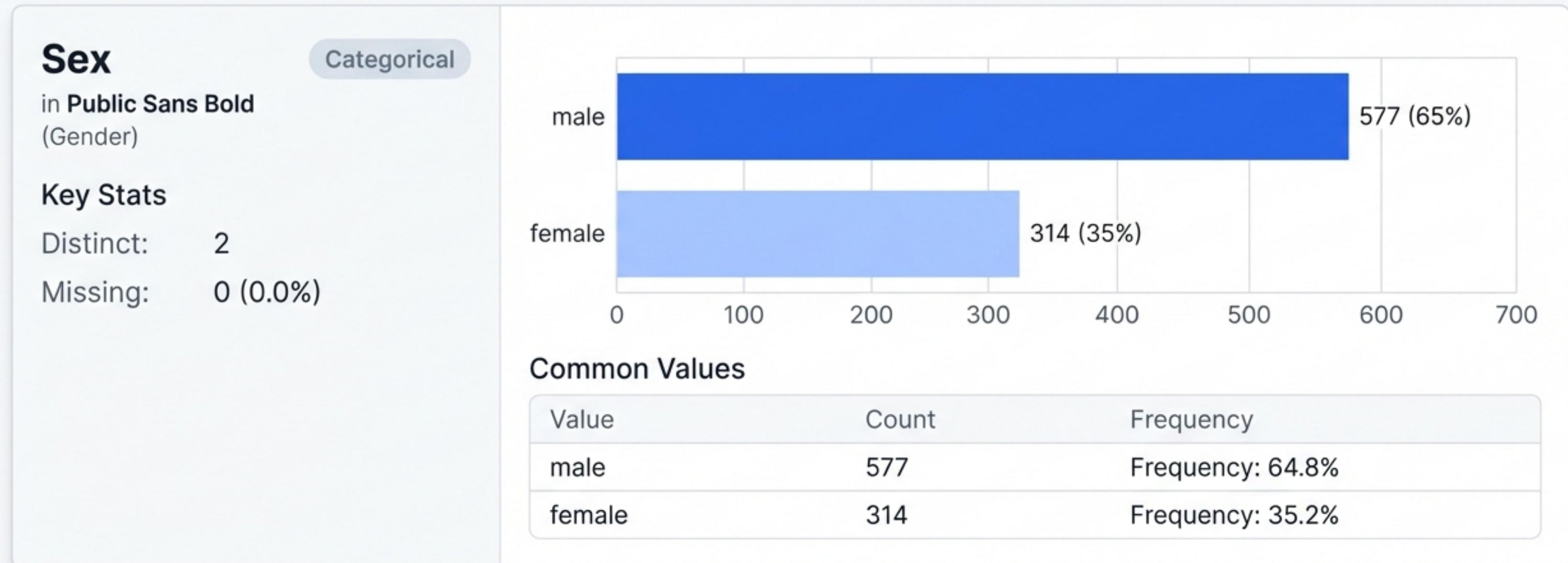
PASSENGERID is uniformly distributed.

Identifies unique identifiers effectively.

Univariate Analysis: Numerical Variables



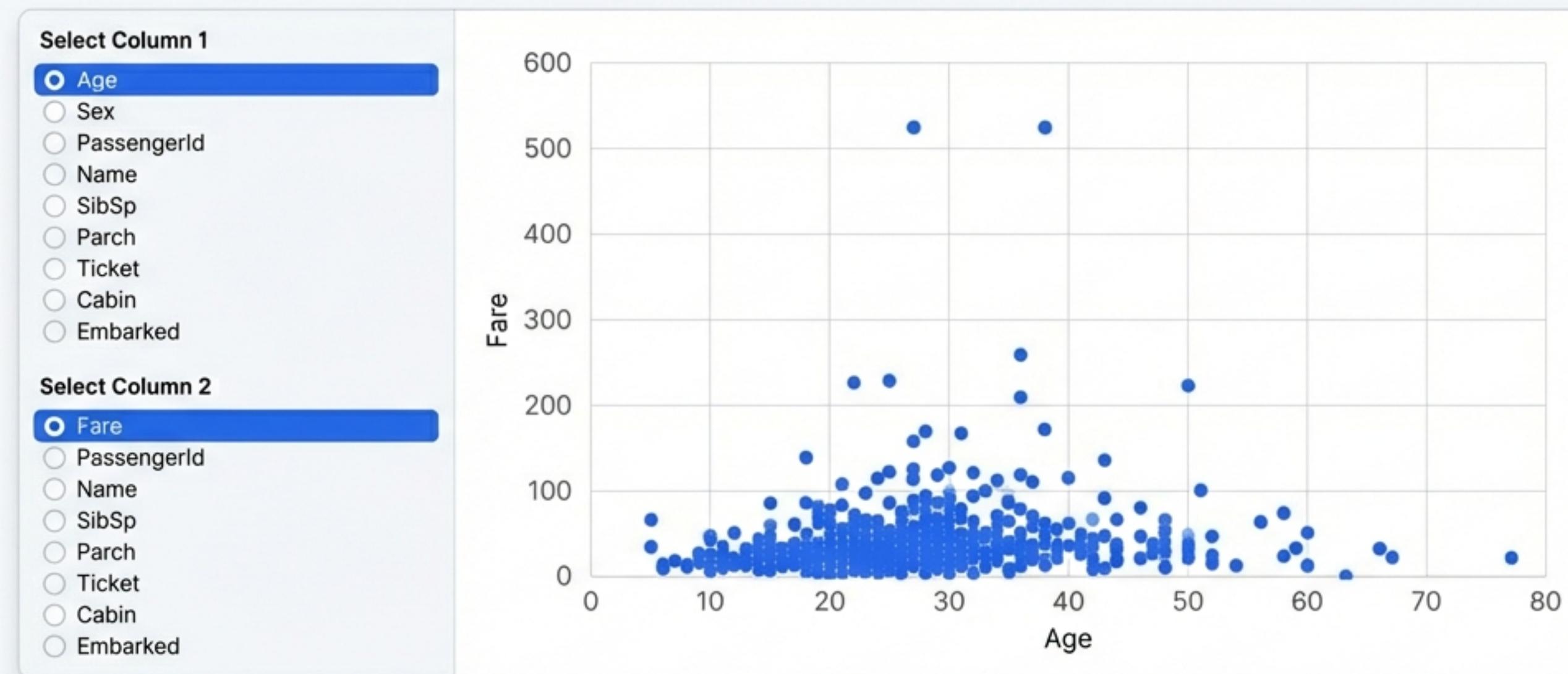
Univariate Analysis: Categorical Variables



Quickly understand class balance without running `value_counts()` manually.

Interactions: Visualising relationships

The Interactions tab performs automatic bivariate analysis, plotting two columns against each other to reveal patterns.

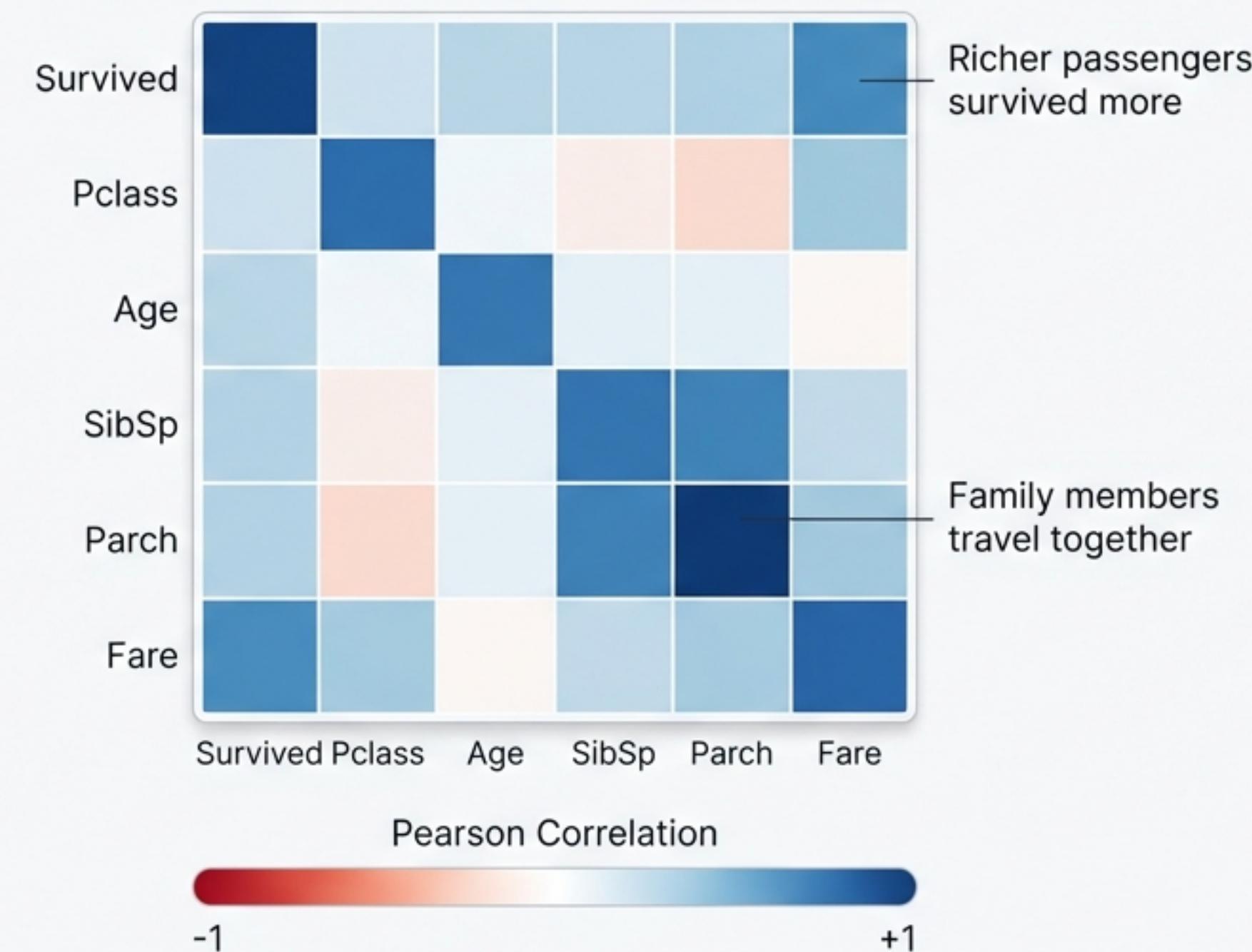


Discovering how variables influence each other.
Example: Age vs. Fare.

This functions similarly to a manual scatter plot but is generated for every permutation of variables automatically.

Correlations: Identifying dependencies

Instantly see which input columns influence your target variable, or which features are redundant.



Missing Values Analysis

Visualising gaps to decide between imputation or dropping.



Sample Data Inspection

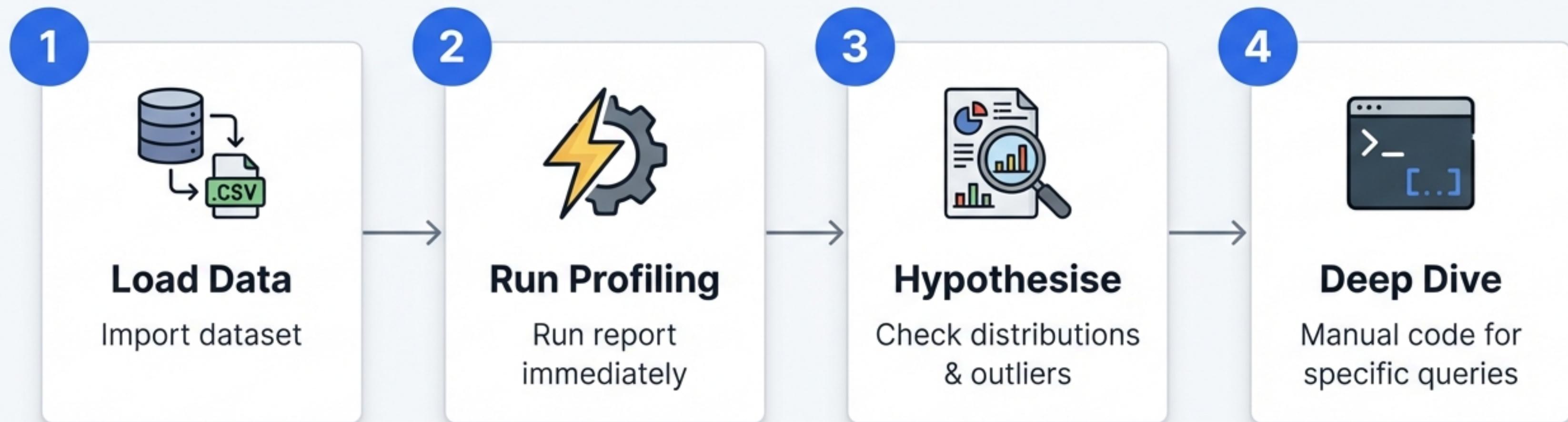
The final sanity check: viewing the raw data.

PassengerId	Survived	Pclass	Name	Sex	Age	...
1	0	3	Braund, Mr. Owen Harris	male	22	...
2	1	1	Cumings, Mrs. John Bradley	female	38	...
3	1	3	Heikkinen, Miss. Laina	female	26	...
4	1	1	Futrelle, Mrs. Jacques Heath	female	35	...
5	0	3	Allen, Mr. William Henry	male	35	...

The report concludes by displaying the First 10 Rows (Head) and Last 10 Rows (Tail). This ensures the data was loaded correctly and matches the statistics observed.

The Recommended Workflow

How to integrate Pandas Profiling into your daily routine.



“Pick any dataset, run this tool, and practice reading the report.
It is the tool that will teach you self-exploration.”