

```
import pandas as pd
```

```
df = pd.read_csv('/content/Large_Patient_Data.csv')
```

```
# Find duplicate rows in the DataFrame
duplicates = df[df.duplicated()]
```

```
# Display duplicate rows
print(duplicates)
```



Empty DataFrame

Columns: [Patient ID, Age, Diagnosis, Length of Stay, Hospital Department]
Index: []

```
df.isnull().sum()
```



0

Patient ID	0
Age	0
Diagnosis	0
Length of Stay	0
Hospital Department	0

dtype: int64

```
df.describe()
```



	Age	Length of Stay
count	10000.00000	10000.000000
mean	50.17870	15.313500
std	29.32195	8.706159
min	0.00000	1.000000
25%	24.00000	8.000000
50%	50.00000	15.000000
75%	76.00000	23.000000
max	100.00000	30.000000



```
# Check for missing values in each column
missing_values = df.isnull().sum()
```

```
# Display the result
print(missing_values)
```



```
Patient ID      0
Age             0
Diagnosis       0
Length of Stay  0
Hospital Department  0
dtype: int64
```

```
# Perform basic statistical analysis
statistics = {
    "Metric": ["Mean", "Median", "Standard Deviation"],
    "Age": [
        df["Age"].mean(),
        df["Age"].median(),
        df["Age"].std(),
    ],
    "Length of Stay": [
        df["Length of Stay"].mean(),
        df["Length of Stay"].median(),
        df["Length of Stay"].std(),
    ],
}

# Convert statistics to a DataFrame for better readability
stats_df = pd.DataFrame(statistics)

# Display the statistical analysis results
stats_df
```



	Metric	Age	Length of Stay
0	Mean	50.17870	15.313500
1	Median	50.00000	15.000000
2	Standard Deviation	29.32195	8.706159



Next
steps:

[Generate code with stats_df](#)[View recommended plots](#)[New interactive sheet](#)

