

```
import pandas as pd
import os

# Ensure the file path is correct
file_path = '/content/drive/MyDrive/titanic.csv'
```

```
# Load the CSV file into a DataFrame
df = pd.read_csv(file_path)
```

```
filtered_df = df[df['Age'] > 30]
```

```
missing_values = df.isnull().sum()
print(missing_values)
```

```
↔ Passengerid    0
   Age           0
   Fare          0
   Sex           0
   sibsp         0
   zero          0
   zero.1        0
   zero.2        0
   zero.3        0
   zero.4        0
   zero.5        0
   zero.6        0
   Parch         0
   zero.7        0
   zero.8        0
   zero.9        0
   zero.10       0
   zero.11       0
   zero.12       0
   zero.13       0
   zero.14       0
   Pclass        0
   zero.15       0
   zero.16       0
   Embarked      2
   zero.17       0
   zero.18       0
   Survived      0
dtype: int64
```

Start coding or [generate](#) with AI.

```
# b. Fill missing values in the 'Age' column with the mean
df['Age'].fillna(df['Age'].mean(), inplace=True)
```

```
df.dropna(subset=['Fare'], inplace=True)
```

```
# Step 4: Calculate summary statistics
```

```
# a. Descriptive statistics for the entire DataFrame
summary_stats = df.describe()
print("Summary statistics for the DataFrame:\n", summary_stats)
```



Summary statistics for the DataFrame:

	Passengerid	Age	Fare	Sex	sibsp \
count	1309.000000	1309.000000	1309.000000	1309.000000	1309.000000
mean	655.000000	29.503186	33.281086	0.355997	0.498854
std	378.020061	12.905241	51.741500	0.478997	1.041658
min	1.000000	0.170000	0.000000	0.000000	0.000000
25%	328.000000	22.000000	7.895800	0.000000	0.000000
50%	655.000000	28.000000	14.454200	0.000000	0.000000
75%	982.000000	35.000000	31.275000	1.000000	1.000000
max	1309.000000	80.000000	512.329200	1.000000	8.000000

	zero	zero.1	zero.2	zero.3	zero.4	...	zero.12	zero.13	zero.14 \
count	1309.0	1309.0	1309.0	1309.0	1309.0	...	1309.0	1309.0	1309.0
mean	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
std	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
min	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
25%	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
50%	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
75%	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0
max	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0

	Pclass	zero.15	zero.16	Embarked	zero.17	zero.18 \
count	1309.000000	1309.0	1309.0	1307.000000	1309.0	1309.0
mean	2.294882	0.0	0.0	1.492731	0.0	0.0
std	0.837836	0.0	0.0	0.814626	0.0	0.0
min	1.000000	0.0	0.0	0.000000	0.0	0.0
25%	2.000000	0.0	0.0	1.000000	0.0	0.0
50%	3.000000	0.0	0.0	2.000000	0.0	0.0
75%	3.000000	0.0	0.0	2.000000	0.0	0.0
max	3.000000	0.0	0.0	2.000000	0.0	0.0

	Survived
count	1309.000000
mean	0.261268
std	0.439494
min	0.000000
25%	0.000000
50%	0.000000
75%	1.000000
max	1.000000

[8 rows x 28 columns]

```
# b. Specific statistics for the 'Age' column
mean_age = df['Age'].mean()
median_age = df['Age'].median()
std_age = df['Age'].std()
```

```
print(f"Mean Age: {mean_age}")
print(f"Median Age: {median_age}") # Removed extra indent here
print(f"Standard Deviation of Age: {std_age}")
```

```
↔ Mean Age: 29.50318563789152
   Median Age: 28.0
   Standard Deviation of Age: 12.905240585464622
```