```
import pandas as pd
import numpy as np

# Load your dataset into a DataFrame (replace 'your_dataset.csv' with your actual file)
df = pd.read_csv('/content/gender_submission.csv')

# Display basic information about the dataset
print("Original Dataset Info:")
print(df.info())

# Remove rows with missing values
df_cleaned = df.dropna()
```

```
Original Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 2 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  418 non-null    int64
 1   Survived     418 non-null    int64
dtypes: int64(2)
memory usage: 6.7 KB
None
```

```
import numpy as np
from scipy.stats import zscore

# Assuming you have a DataFrame named df_cleaned
# Replace this with actual data loading or cleaning process
# df_cleaned = ...

numeric_columns = ['PassengerId', 'Survived']  # Replace with actual column names
z_scores = np.abs(zscore(df_cleaned[numeric_columns]))
threshold = 3
outlier_mask = (z_scores < threshold).all(axis=1)
df_cleaned_no_outliers = df_cleaned[outlier_mask]
```

```
# Display information about the cleaned dataset
print("\nCleaned Dataset Info:")
print(df_cleaned_no_outliers.info())
```

```

Cleaned Dataset Info:
<class 'pandas.core.frame.DataFrame'>
Int64Index: 418 entries, 0 to 417
Data columns (total 2 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  418 non-null    int64
 1   Survived     418 non-null    int64
dtypes: int64(2)
memory usage: 9.8 KB
None
```