Problem statement

IT industry records the amount of time a software engineer needs to fix a bug in the initial phase of software development in 'debugging.csv'.

Let

X = Time needed to fix bugs

X is a continuous random variable. Let's see the distribution of X and answer the below questions.

1. Find the probability that a randomly selected software debugging requires less than three hours

2. Find the probability that a randomly selected software debugging requires more than two hours

3. Find the 50th percentile of the software debugging tire

```
import pandas as pd
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt

df = pd.read_csv('/Users/vishal/Desktop/CSV files/Debugging.csv')

df.head()
```

	Bug ID	Time Taken to fix the bug
0	12986	2.42
1	12987	2.03
2	12988	2.74
3	12989	3.21
4	12990	3.40

```
df.tail()
```

	Bug ID	Time Taken to fix the bug
2093	15079	4.17
2094	15080	1.05
2095	15081	2.50
2096	15082	2.85
2097	15083	2.64

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2098 entries, 0 to 2097
Data columns (total 2 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   Bug ID                     2098 non-null   int64
 1   Time Taken to fix the bug  2098 non-null   float64
dtypes: float64(1), int64(1)
memory usage: 32.9 KB
```

```
df.describe()
```

	Bug ID	Time Taken to fix the bug
count	2098.000000	2098.000000
mean	14034.500000	3.012531
std	605.784753	1.147148
min	12986.000000	1.010000
25%	13510.250000	2.010000
50%	14034.500000	3.005000
75%	14558.750000	4.030000
max	15083.000000	5.000000

Check the Distribution of 'Time Taken to Fix the Bug'

```python
# Plot a histogram to visualize the distribution
plt.hist(df['Time Taken to fix the bug'], bins=30, edgecolor='k',
alpha=0.7)
plt.title('Distribution of Time Taken to Fix Bugs')
plt.xlabel('Time (hours)')
plt.ylabel('Frequency')
plt.show()
```

Distribution of Time Taken to Fix Bugs
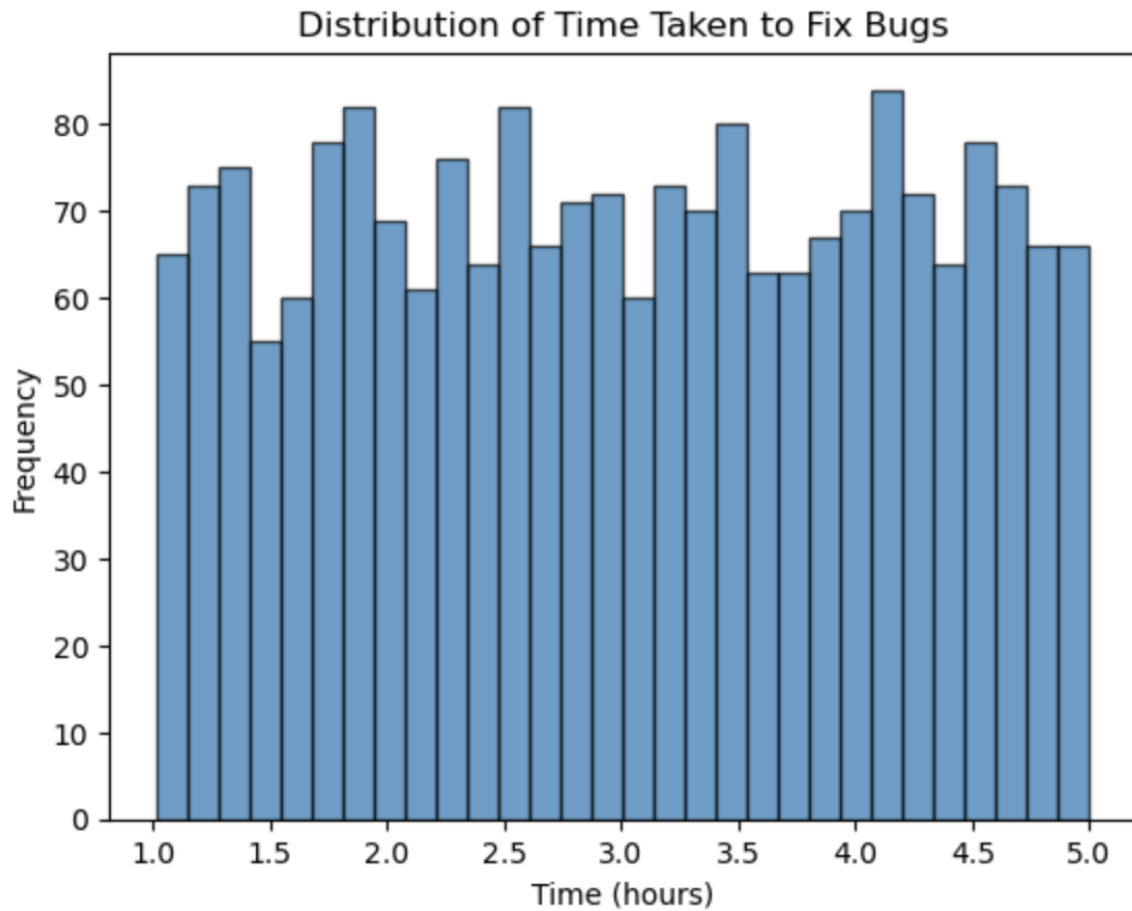
Calculate the Mean and Standard Deviation

```
mean_time = df['Time Taken to fix the bug'].mean()
std_time = df['Time Taken to fix the bug'].std()
print(f"Mean Time to Fix: {mean_time}")
print(f"Standard Deviation of Time to Fix: {std_time}")
```

```
Mean Time to Fix: 3.012530981887512
Standard Deviation of Time to Fix: 1.1471482047102495
```

Find the probability that Time Taken to Fix is less than 3 hours

```python
prob_less_than_3 = stats.norm.cdf(3, loc=mean_time, scale=std_time)
print(f"Probability that debugging requires less than 3 hours:
{prob_less_than_3}")
```

    Probability that debugging requires less than 3 hours: 0.4956422029421937

Find the Probability That Debugging Takes More Than 2 Hours

```python
prob_more_than_2 = 1 - stats.norm.cdf(2, loc=mean_time,
scale=std_time)
print(f"Probability that debugging requires more than 2 hours:
{prob_more_than_2}")
```

    Probability that debugging requires more than 2 hours: 0.8112874434344626

Find the 50th Percentile (Median)

```python
percentile_50 = stats.norm.ppf(0.5, loc=mean_time, scale=std_time)
print(f"50th percentile (median) of time to fix bugs:
{percentile_50}")
```

    50th percentile (median) of time to fix bugs: 3.012530981887512