#Assignment 2 : Introduction to Inferential Statistics

```
import pandas as pd
import numpy as np
from scipy import stats

debugging_df = pd.read_csv('debugging.csv')

debugging_df.head()
```

```
debugging_df.info()
debugging_df.isnull().sum()
debugging_df['Time Taken to fix the bug'].describe()
```

```
<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
```

```
# Assuming the data follows a normal distribution, get mean and standard deviation
mean_time = debugging_df['Time Taken to fix the bug'].mean()
std_time = debugging_df['Time Taken to fix the bug'].std()

# Probability that X < 3 hours
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 X > 2 hours
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}")
# Find the 50th percentile (median) of the debugging time
percentile_50 = np.percentile(debugging_df['Time Taken to fix the bug'], 50)
print(f"The 50th percentile (median) of the debugging time: {percentile_50}")
```

```
Probability that debugging requires less than 3 hours: 0.4956422029421937
Probability that debugging requires more than 2 hours: 0.8112874434344626
The 50th percentile (median) of the debugging time: 3.005
```