Can this Data be normally distributed ?

Correct question: Has the data been sampled from a distribution that is close to the normal (Gaussian) bell curve?

Methods to test for normality:

- Graphical: Frequency distributions i.e. Histograms
  - Can easily determine the curve shape
  - Isn't accurate for a smaller sample
- Analytical: Shapiro-Wilk
  - gives a single valued result which can be compared against alpha
  - also not too accurate

Solution: Use a combination of graphical and analytical methods to determine result

Define:

- Null Hypothesis (P > 0.05) = Values **are** sampled from a population that follows a normal distribution
- Alternate Hypothesis (P <= 0.05) = Values **are not** sampled from a population that follows a normal distribution

In [ ]:
```python
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
from scipy.stats import shapiro
```

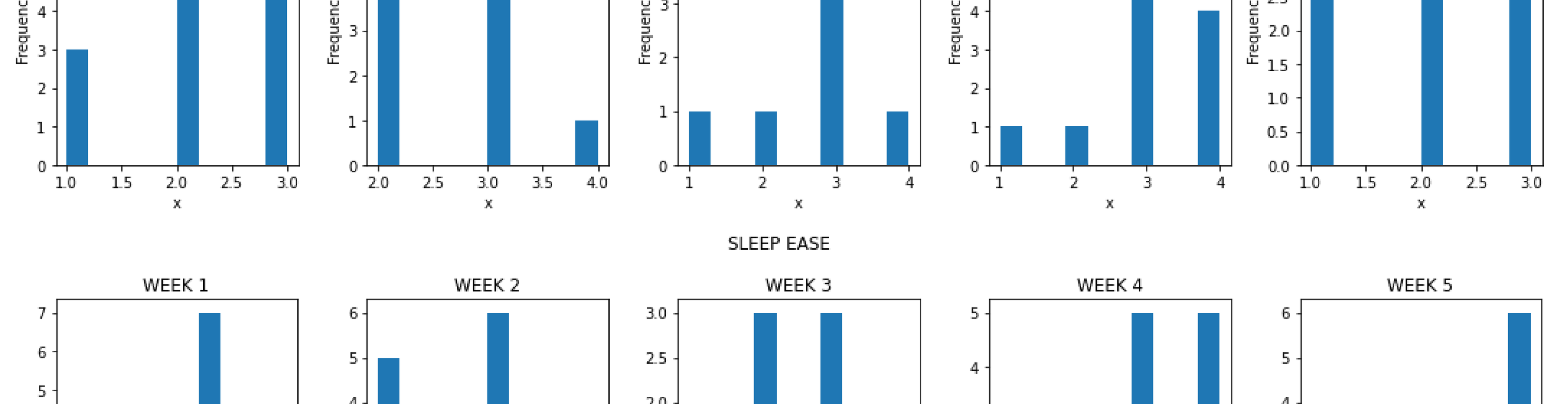
In [ ]:
```python
# State the file name and sheet name
file_name = 'quest_1.xlsx'
sheet_name = 'Sheet1'
```

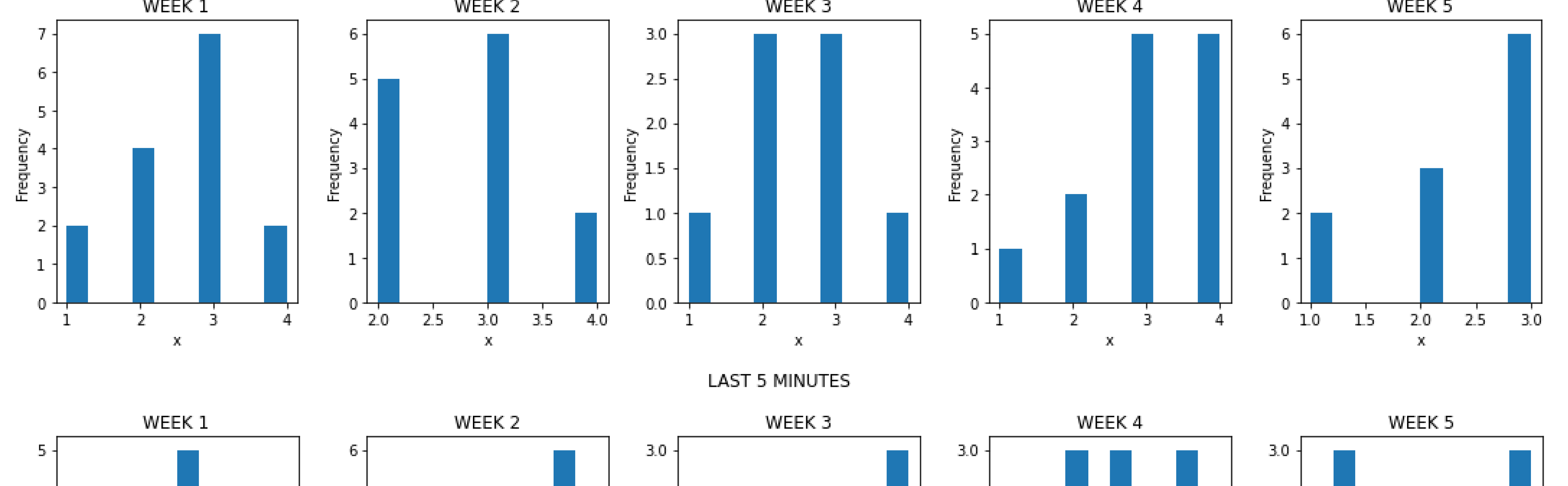
In [ ]:
```python
def display_cols(file_name, sheet_name):
    '''
    returns the list of all the columns in a dataframe along with the original dataframe
    '''
    df = pd.read_excel(file_name, sheet_name)

    # # specifically for quest_2 data
    # df.rename({'I’ve been feeling optimistic about the future':'OPTIMISM',
    # 'I’ve been feeling useful':'USEFUL',
    # '\n1 – very dim\xa0 \xa0 4 – very bright\n':'BRIGHT or DIM',
    # '1 – very warm/yellowish\xa0 \xa0 \xa0 4 – very cold /bluish':'YELLOW or BLUE',
    # 'dealing with problems ':'DEALING WITH PROBLEMS',
    # ' thinking clearly': 'THINKING CLEARLY',
    # '1 – not at all glaring\xa0 \xa0 \xa0 \xa04 – very glaring\n':'GLARING',
    # ' feeling close to other people': 'FEELING CLOSE',
    # 'LIGHTING OVERALL IMPRESSION':'OVERALL IMPRESSION'
    # }, inplace=True, axis=1)

    return df.columns, pd.DataFrame(df)

# printing the dataframe and the columns in it
print(f"Columns: {list(display_cols(file_name, sheet_name)[0])} \nDataframe: \n{display_cols(file_name, sheet_name)[1]}")
```

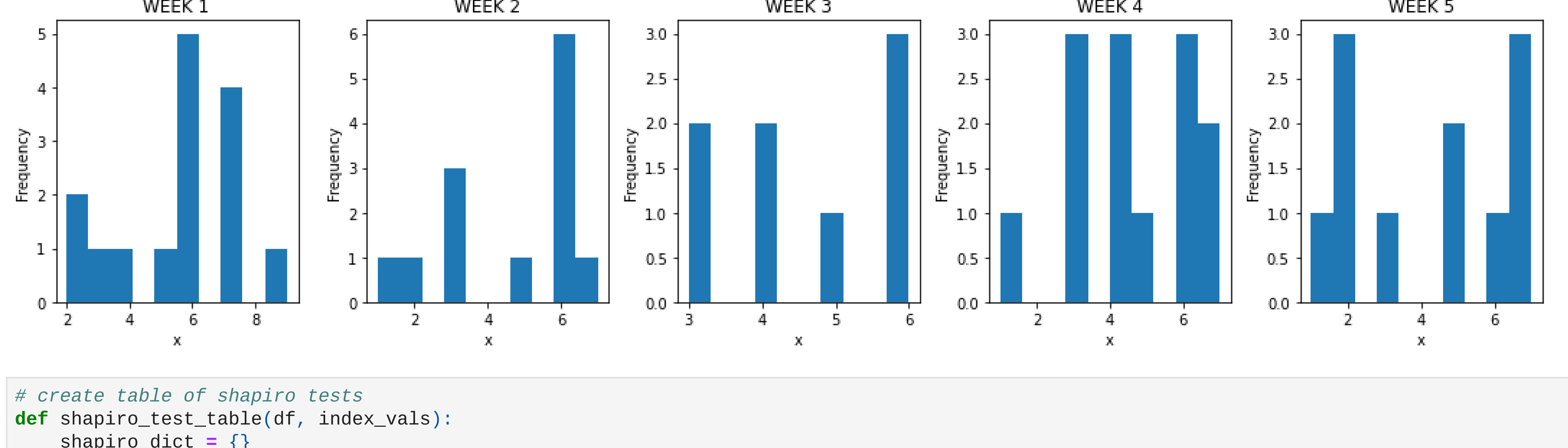
```
Columns: ['Start time', 'WEEK NO', 'What is your I.D.?', 'SLEEP QUALITY', 'SLEEP EASE', 'SLEEP TIME', 'WAKE TIME', 'NO. OF SLEEP HOURS', 'LAST 5 MI
NUTES', 'Unnamed: 9', 'Unnamed: 10', 'Unnamed: 11', 'Unnamed: 12', 'Unnamed: 13', 'Unnamed: 14', 'Unnamed: 15', 'Unnamed: 16', 'Unnamed: 17', 'Unna
med: 18']
Dataframe:
              Start time WEEK NO What is your I.D.?  SLEEP QUALITY  SLEEP EASE  \
0  2022-06-07 14:39:33  WEEK 1                S20              3           3
1  2022-06-07 18:14:57  WEEK 1                S22              1           1
2  2022-06-07 22:30:44  WEEK 1                S25              2           3
3  2022-06-08 12:09:54  WEEK 1                S04              1           2
4  2022-06-08 15:21:08  WEEK 1                S06              2           2
5  2022-06-08 16:30:00  WEEK 1                S14              3           3
6  2022-06-08 20:07:22  WEEK 1                S08              2           3
7  2022-06-08 23:39:42  WEEK 1                S23              3           4
8  2022-06-09 00:00:00  WEEK 1                S17              3           3
9  2022-06-09 14:36:42  WEEK 1                S21              2           3
10 2022-06-09 18:15:49  WEEK 1                S12              1           1
11 2022-06-09 21:08:13  WEEK 1                S11              2           2
12 2022-06-09 23:04:52  WEEK 1                S10              2           2
13 2022-06-10 15:17:00  WEEK 1                S27              3           4
14 2022-06-11 12:24:57  WEEK 1                S31              2           3
15 2022-06-14 17:03:03  WEEK 2                S22              2           2
16 2022-06-14 20:09:07  WEEK 2                S25              3           2
17 2022-06-15 14:03:21  WEEK 2                S04              2           3
18 2022-06-15 17:28:05  WEEK 2                S06              3           3
19 2022-06-15 19:30:00  WEEK 2                S14              3           3
20 2022-06-15 20:01:08  WEEK 2                S08              3           2
21 2022-06-16 00:00:00  WEEK 2                S17              4           4
22 2022-06-16 19:09:58  WEEK 2                S12              3           3
23 2022-06-16 20:47:24  WEEK 2                S05              2           2
24 2022-06-17 11:12:17  WEEK 2                S10              2           2
25 2022-06-17 15:15:15  WEEK 2                S31              2           4
26 2022-06-17 17:05:15  WEEK 2                S21              2           3
27 2022-06-17 21:50:54  WEEK 2                S20              3           3
28 2022-06-22 14:38:02  WEEK 3                S04              2           2
29 2022-06-22 14:39:20  WEEK 3                S21              3           3
30 2022-06-22 19:30:00  WEEK 3                S14              3           3
31 2022-06-22 20:08:22  WEEK 3                S08              1           1
32 2022-06-23 00:00:00  WEEK 3                S17              4           4
33 2022-06-23 08:07:22  WEEK 3                S05              3           2
34 2022-06-24 11:32:00  WEEK 3                S27              3           2
35 2022-06-21 14:44:00  WEEK 3                S20              3           3
36 2022-06-28 14:29:40  WEEK 4                S20              3           3
37 2022-06-29 14:31:04  WEEK 4                S04              1           1
38 2022-06-29 14:53:20  WEEK 4                S21              3           3
39 2022-06-29 16:19:48  WEEK 4                S05              3           4
40 2022-06-29 20:00:00  WEEK 4                S14              3           3
41 2022-06-29 20:47:34  WEEK 4                S08              3           3
42 2022-06-29 22:51:51  WEEK 4                S06              3           2
43 2022-06-30 00:00:00  WEEK 4                S17              4           4
44 2022-06-30 18:45:48  WEEK 4                S12              4           4
45 2022-06-30 22:59:29  WEEK 4                S05              4           4
46 2022-07-01 11:15:00  WEEK 4                S27              3           4
47 2022-07-01 07:31:42  WEEK 4                S10              2           2
48 2022-07-01 15:07:59  WEEK 4                S31              4           3
49 2022-07-05 14:33:03  WEEK 5                S20              1           1
50 2022-07-06 13:11:31  WEEK 5                S04              2           3
51 2022-07-06 17:00:00  WEEK 5                S14              3           3
52 2022-07-06 17:14:20  WEEK 5                S06              3           2
53 2022-07-06 20:08:20  WEEK 5                S08              3           3
54 2022-07-07 11:41:45  WEEK 5                S05              1           2
55 2022-07-08 07:37:34  WEEK 5                S10              2           2
56 2022-07-08 09:15:06  WEEK 5                S01              1           1
57 2022-07-08 17:01:04  WEEK 5                S31              2           3
58 2022-07-08 17:06:03  WEEK 5                S21              2           3
59 2022-07-09 22:27:27  WEEK 5                S12              3           3

          SLEEP TIME        WAKE TIME NO. OF SLEEP HOURS  LAST 5 MINUTES  \
0   10:30 – 12:00 am   6:00 - 7:30 am        5 - 7 hours               5
1     After 12:00 am   7:30 - 9:00 am        5 - 7 hours               9
2   10:30 – 12:00 am   7:30 - 9:00 am        7 - 9 hours               7
3     After 12:00 am   4:30 - 6:00 am        4 - 5 hours               6
4     After 12:00 am   4:30 - 6:00 am        4 - 5 hours               2
5    9:00 – 10:30 pm   4:30 - 6:00 am        5 - 7 hours               4
6     After 12:00 am   After 10:30 am        7 - 9 hours               6
7     After 12:00 am   7:30 - 9:00 am        7 - 9 hours               6
8    9:00 – 10:30 pm   6:00 - 7:30 am        5 - 7 hours               3
9   10:30 – 12:00 am   4:30 - 6:00 am        4 - 5 hours               6
10    After 12:00 am   7:30 - 9:00 am        4 - 5 hours               7
11    After 12:00 am   After 10:30 am        7 - 9 hours               7
12  10:30 – 12:00 am   7:30 - 9:00 am        5 - 7 hours               7
13  10:30 – 12:00 am   After 10:30 am  More than 9 hours               6
14    After 12:00 am   Before 4:30 am  Less than 4 hours               2
15    After 12:00 am   7:30 - 9:00 am        5 - 7 hours               5
16    After 12:00 am  9:00 – 10:30 am        7 - 9 hours               7
17    After 12:00 am   4:30 - 6:00 am        4 - 5 hours               6
18    After 12:00 am   4:30 - 6:00 am        4 - 5 hours               6
19   9:00 – 10:30 pm   4:30 - 6:00 am        7 - 9 hours               3
20    After 12:00 am   After 10:30 am        7 - 9 hours               2
21   9:00 – 10:30 pm   6:00 - 7:30 am        5 - 7 hours               3
22    After 12:00 am   6:00 - 7:30 am        5 - 7 hours               1
23    After 12:00 am   4:30 - 6:00 am        4 - 5 hours               6
24  10:30 – 12:00 am   6:00 - 7:30 am        4 - 5 hours               6
25  10:30 – 12:00 am   4:30 - 6:00 am        4 - 5 hours               3
26  10:30 – 12:00 am   4:30 - 6:00 am        5 - 7 hours               6
27  10:30 – 12:00 am   6:00 - 7:30 am        5 - 7 hours               6
28    After 12:00 am   7:30 - 9:00 am        7 - 9 hours               6
29  10:30 – 12:00 am   4:30 - 6:00 am        5 - 7 hours               6
30   9:00 – 10:30 pm   4:30 - 6:00 am        7 - 9 hours               4
31    After 12:00 am   After 10:30 am        7 - 9 hours               3
32   9:00 – 10:30 pm   6:00 - 7:30 am        5 - 7 hours               3
33  10:30 – 12:00 am   6:00 - 7:30 am        5 - 7 hours               5
34    After 12:00 am   6:00 - 7:30 am        5 - 7 hours               4
35  10:30 – 12:00 am   6:00 - 7:30 am        5 - 7 hours               6
36   9:00 – 10:30 pm   6:00 - 7:30 am        7 - 9 hours               6
37    After 12:00 am   4:30 - 6:00 am        4 - 5 hours               7
38  10:30 – 12:00 am   4:30 - 6:00 am        5 - 7 hours               4
39  10:30 – 12:00 am   4:30 - 6:00 am        4 - 5 hours               4
40   9:00 – 10:30 pm   4:30 - 6:00 am        7 - 9 hours               4
41    After 12:00 am   After 10:30 am  More than 9 hours               3
42    After 12:00 am   4:30 - 6:00 am        4 - 5 hours               3
43   9:00 – 10:30 pm   6:00 - 7:30 am        5 - 7 hours               3
44    After 12:00 am   6:00 - 7:30 am        5 - 7 hours               1
45   9:00 – 10:30 pm  9:00 – 10:30 am  More than 9 hours               6
46  10:30 – 12:00 am   6:00 - 7:30 am        5 - 7 hours               5
47  10:30 – 12:00 am   6:00 - 7:30 am        5 - 7 hours               7
48  10:30 – 12:00 am   4:30 - 6:00 am        5 - 7 hours               6
49   9:00 – 10:30 pm   4:30 - 6:00 am  Less than 4 hours               7
50    After 12:00 am   7:30 - 9:00 am        7 - 9 hours               5
51   9:00 – 10:30 pm   4:30 - 6:00 am        7 - 9 hours               3
52    After 12:00 am   4:30 - 6:00 am        4 - 5 hours               2
53    After 12:00 am   After 10:30 am        7 - 9 hours               2
54  10:30 – 12:00 am   4:30 - 6:00 am        5 - 7 hours               5
55    After 12:00 am   7:30 - 9:00 am        5 - 7 hours               6
56    After 12:00 am   7:30 - 9:00 am        5 - 7 hours               7
57  10:30 – 12:00 am   4:30 - 6:00 am  Less than 4 hours               1
58  10:30 – 12:00 am   4:30 - 6:00 am        5 - 7 hours               7
59    After 12:00 am   After 10:30 am        7 - 9 hours               2

    Unnamed: 9  Unnamed: 10  Unnamed: 11  Unnamed: 12  Unnamed: 13  \
0          NaN          NaN          NaN          NaN          NaN
1          NaN          NaN          NaN          NaN          NaN
2          NaN          NaN          NaN          NaN          NaN
3          NaN          NaN          NaN          NaN          NaN
4          NaN          NaN          NaN          NaN          NaN
5          NaN          NaN          NaN          NaN          NaN
6          NaN          NaN          NaN          NaN          NaN
7          NaN          NaN          NaN          NaN          NaN
8          NaN          NaN          NaN          NaN          NaN
9          NaN          NaN          NaN          NaN          NaN
10         NaN          NaN          NaN          NaN          NaN
11         NaN          NaN          NaN          NaN          NaN
12         NaN          NaN          NaN          NaN          NaN
13         NaN          NaN          NaN          NaN          NaN
14         NaN          NaN          NaN          NaN          NaN
15         NaN          NaN          NaN          NaN          NaN
16         NaN          NaN          NaN          NaN          NaN
17         NaN          NaN          NaN          NaN          NaN
18         NaN          NaN          NaN          NaN          NaN
19         NaN          NaN          NaN          NaN          NaN
20         NaN          NaN          NaN          NaN          NaN
21         NaN          NaN          NaN          NaN          NaN
22         NaN          NaN          NaN          NaN          NaN
23         NaN          NaN          NaN          NaN          NaN
24         NaN          NaN          NaN          NaN          NaN
25         NaN          NaN          NaN          NaN          NaN
26         NaN          NaN          NaN          NaN          NaN
27         NaN          NaN          NaN          NaN          NaN
28         NaN          NaN          NaN          NaN          NaN
29         NaN          NaN          NaN          NaN          NaN
30         NaN          NaN          NaN          NaN          NaN
31         NaN          NaN          NaN          NaN          NaN
32         NaN          NaN          NaN          NaN          NaN
33         NaN          NaN          NaN          NaN          NaN
34         NaN          NaN          NaN          NaN          NaN
35         NaN          NaN          NaN          NaN          NaN
36         NaN          NaN          NaN          NaN          NaN
37         NaN          NaN          NaN          NaN          NaN
38         NaN          NaN          NaN          NaN          NaN
39         NaN          NaN          NaN          NaN          NaN
40         NaN          NaN          NaN          NaN          NaN
41         NaN          NaN          NaN          NaN          NaN
42         NaN          NaN          NaN          NaN          NaN
43         NaN          NaN          NaN          NaN          NaN
44         NaN          NaN          NaN          NaN          NaN
45         NaN          NaN          NaN          NaN          NaN
46         NaN          NaN          NaN          NaN          NaN
47         NaN          NaN          NaN          NaN          NaN
48         NaN          NaN          NaN          NaN          NaN
49         NaN          NaN          NaN          NaN          NaN
50         NaN          NaN          NaN          NaN          NaN
51         NaN          NaN          NaN          NaN          NaN
52         NaN          NaN          NaN          NaN          NaN
53         NaN          NaN          NaN          NaN          NaN
54         NaN          NaN          NaN          NaN          NaN
55         NaN          NaN          NaN          NaN          NaN
56         NaN          NaN          NaN          NaN          NaN
57         NaN          NaN          NaN          NaN          NaN
58         NaN          NaN          NaN          NaN          NaN
59         NaN          NaN          NaN          NaN          NaN

    Unnamed: 14  Unnamed: 15  Unnamed: 16  Unnamed: 17  Unnamed: 18
0           NaN          NaN          NaN          NaN          NaN
1           1.0          2.0          3.0          4.0          5.0
2           NaN          NaN          NaN          NaN          NaN
3           NaN          NaN          NaN          NaN          NaN
4           NaN          NaN          NaN          NaN          NaN
5           3.0          2.0          2.0          3.0          1.0
6           1.0          3.0          3.0          1.0          2.0
7           2.0          2.0          3.0          3.0          3.0
8           1.0          3.0          1.0          3.0          3.0
9           2.0          3.0          4.0          3.0          3.0
10          3.0          3.0          3.0          3.0          1.0
11          2.0          4.0          3.0          3.0          2.0
12          3.0          3.0          3.0          4.0          1.0
13          3.0          2.0          NaN          4.0          2.0
14          2.0          2.0          NaN          4.0          2.0
15          1.0          2.0          NaN          3.0          3.0
16          2.0          2.0          NaN          2.0          NaN
17          2.0          3.0          NaN          4.0          NaN
18          3.0          NaN          NaN          NaN          NaN
19          2.0          NaN          NaN          NaN          NaN
20          NaN          NaN          NaN          NaN          NaN
21          NaN          NaN          NaN          NaN          NaN
22          3.0          2.0          2.0          3.0          1.0
23          1.0          3.0          3.0          1.0          2.0
24          2.0          2.0          3.0          3.0          3.0
25          1.0          3.0          1.0          3.0          3.0
26          2.0          3.0          4.0          3.0          3.0
27          3.0          3.0          3.0          3.0          1.0
28          2.0          4.0          3.0          3.0          2.0
29          3.0          3.0          3.0          4.0          1.0
30          3.0          2.0          NaN          4.0          2.0
31          2.0          2.0          NaN          4.0          2.0
32          1.0          2.0          NaN          3.0          3.0
33          2.0          2.0          NaN          2.0          0.0
34          2.0          3.0          NaN          4.0          0.0
35          3.0          NaN          NaN          NaN          0.0
36          2.0          NaN          NaN          NaN          0.0
37          NaN          NaN          NaN          NaN          NaN
38          NaN          NaN          NaN          NaN          NaN
39          NaN          NaN          NaN          NaN          NaN
40          NaN          NaN          NaN          NaN          NaN
41          NaN          NaN          NaN          NaN          NaN
42          NaN          NaN          NaN          NaN          NaN
43          NaN          NaN          NaN          NaN          NaN
44          NaN          NaN          NaN          NaN          NaN
45          NaN          NaN          NaN          NaN          NaN
46          NaN          NaN          NaN          NaN          NaN
47          NaN          NaN          NaN          NaN          NaN
48          NaN          NaN          NaN          NaN          NaN
49          NaN          NaN          NaN          NaN          NaN
50          NaN          NaN          NaN          NaN          NaN
51          NaN          NaN          NaN          NaN          NaN
52          NaN          NaN          NaN          NaN          NaN
53          NaN          NaN          NaN          NaN          NaN
54          NaN          NaN          NaN          NaN          NaN
55          NaN          NaN          NaN          NaN          NaN
56          NaN          NaN          NaN          NaN          NaN
57          NaN          NaN          NaN          NaN          NaN
58          NaN          NaN          NaN          NaN          NaN
59          NaN          NaN          NaN          NaN          NaN
```

In [ ]:
```python
def get_data(df, col_list):
    '''
    returns the required cols from the dataframe along with the data
    '''
    df = pd.DataFrame(df)
    required_df = df.loc[:,col_list].set_index('WEEK NO')
    return pd.DataFrame(required_df), list(required_df.index.unique())

# stating required columns
col_list = ['WEEK NO', 'SLEEP QUALITY', 'SLEEP EASE', 'LAST 5 MINUTES']
# col_list = ['WEEK NO', 'SLEEP QUALITY', 'SLEEP EASE', 'LAST 5 MINUTES', 'OPTIMISM',
# 'SLEEP QUALITY WEEK', 'USEFUL', 'BRIGHT or DIM', 'RELAXED', 'YELLOW or BLUE', 'DEALING WITH PROBLEMS', 'THINKING CLEARLY',
# 'GLARING', 'FEELING CLOSE', 'OVERALL IMPRESSION']

req_df = get_data(display_cols(file_name, sheet_name)[1], col_list)[0]
week_nos = get_data(display_cols(file_name, sheet_name)[1], col_list)[1]
```

In [ ]:
```python
# separate histogram plots
def sep_hist(df, index_vals):
    for col in df:
        fig, ax = plt.subplots(figsize=(15,4), nrows=1, ncols=len(index_vals))
        fig.suptitle(col)
        i = 0
        for week in index_vals:
            ax[i].set_title(week)
            ax[i].set_xlabel('x')
            ax[i].set_ylabel('Frequency')
            ax[i].hist(np.array(df.loc[df.index == week,col]))
            i += 1
            plt.tight_layout() # adjust size of each plot
        plt.savefig(file_name.split(".")[0]+"_"+col+".png")

sep_hist(req_df, week_nos)
```







In [ ]:
```python
# create table of shapiro tests
def shapiro_test_table(df, index_vals):
    shapiro_dict = {}
    with pd.ExcelWriter(file_name.split(".")[0]+"_shapiro_test.xlsx", engine='xlsxwriter') as writer:
        for col in df:
            for week in index_vals:
                shapiro_dict[week] = shapiro(np.array(df.loc[df.index == week,col])).pvalue
            shapiro_df = pd.DataFrame(shapiro_dict, index=['pValue'])
            shapiro_df.T.to_excel(writer, sheet_name=col)

shapiro_test_table(req_df, week_nos)
```

In [ ]: