1. What data do you have?

The data we have is a collection of 78 questions and 217 answers from a Google Form where only 5 questions pertain to our overarching question. For these questions, we have both numeric and qualitative data which can be standardized into levels 1-5 depending on the set of answer choices. The questions relating to our project are:

1. How many meals do you typically eat per day?
   - Type of Data: Numeric
   - Answer Choices:
     - 1 meal (1)
     - 2 meals (2)
     - 3 meals (3)
     - more than 3 meals (4)
2. On average, how many cups of water do you drink per day?
   - Type of Data: Numeric
   - Answer Choices:
     - Less than 1 liter (about 34 oz)(1)
     - 1- 2 liters (about 34-68 oz)(2)
     - 2 - 3 liters (about 68-102 oz)(3)
     - 3 - 4 liters (about 102-135 oz)(4)
     - More than 4 liters (about 135 oz)(5)
3. How often do you eat breakfast?
   - Type of Data: Qualitative (Categorical)
   - Answer Choices:
     - Never (1)
     - Rarely (2)
     - Occasionally (3)
     - Most days (4)
     - Every day (5)
4. On average, how would you describe your energy levels throughout the day?
   - Type of Data: Qualitative (Categorical)
   - Answer Choices:
     - Very low (1)
     - Low (2)
     - Neutral (3)
     - High (4)
     - Very high (5)
5. How often do you consume caffeinated beverages?
   - Type of Data: Qualitative (Categorical)
   - Answer Choices:
     - Never (1)
     - Occasionally (2)
     - Sometimes (3)

- Often (4)
- Always (5)

Data Standardization:

- For the qualitative data, responses can be standardized into numerical values (1-5) to facilitate statistical analysis. For example, if someone answers "Occasionally" for the question "How often do you eat breakfast?", it can be standardized to 2.
- The numeric responses (e.g., number of meals per day or cups of water) are already suitable for analysis but may need further cleaning, such as handling missing or inconsistent entries.

---

2. What would you like to know?

Does a student's liquid intake(water and caffeinated beverages) and frequency and type of meals affects their energy levels throughout the day?

---

3. Explore, present and analyze the data. (Generate statistics, perform visualizations)

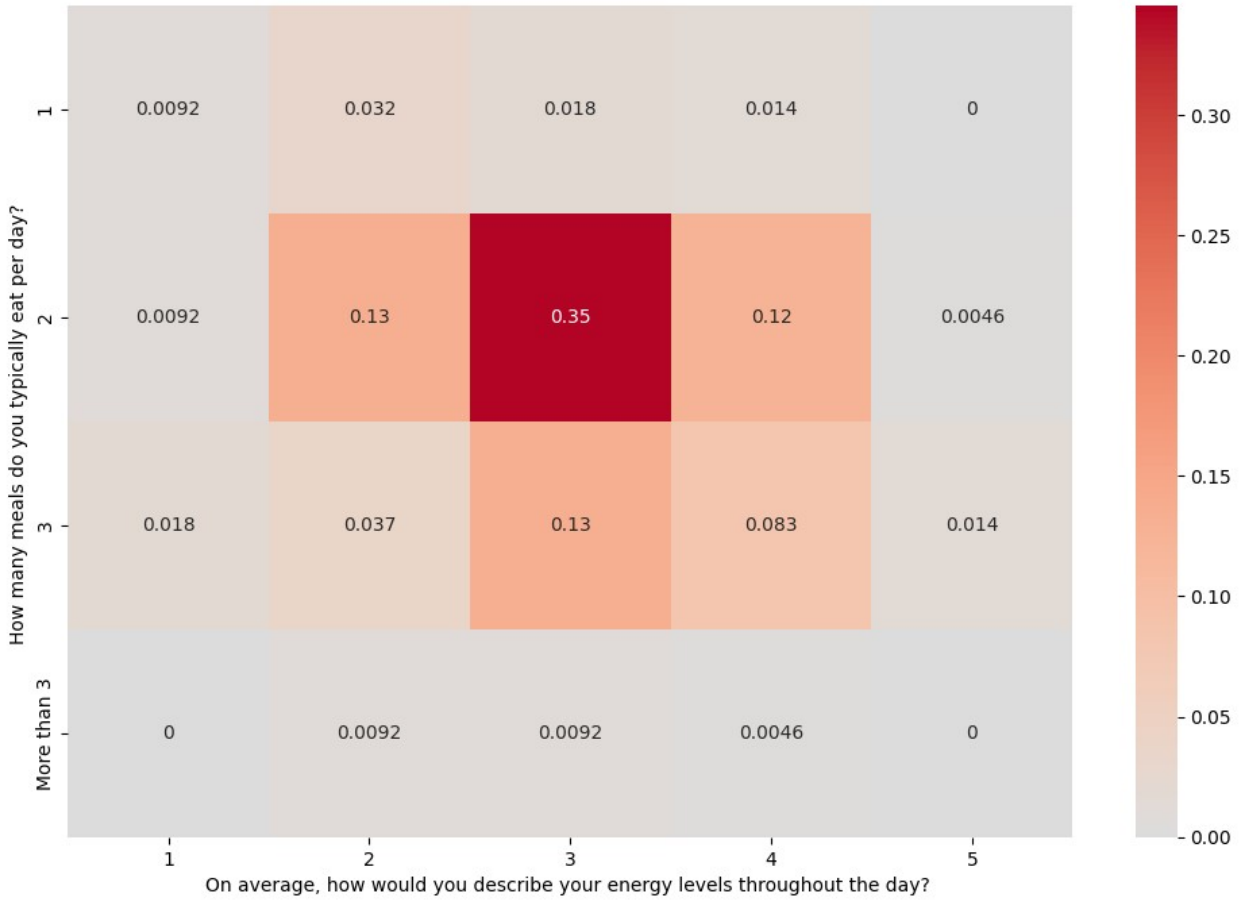
```
from google.colab import drive
drive.mount('/content/drive')
```

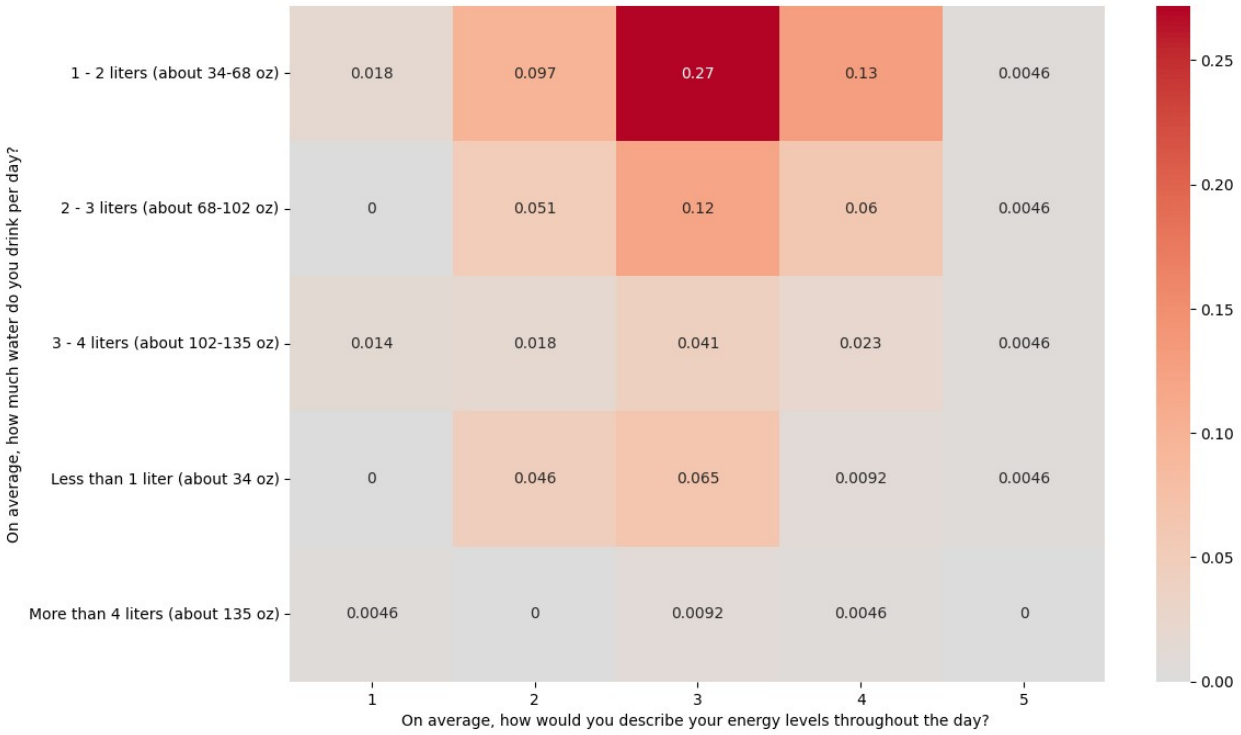
```
Drive already mounted at /content/drive; to attempt to forcibly
remount, call drive.mount("/content/drive", force_remount=True).
```

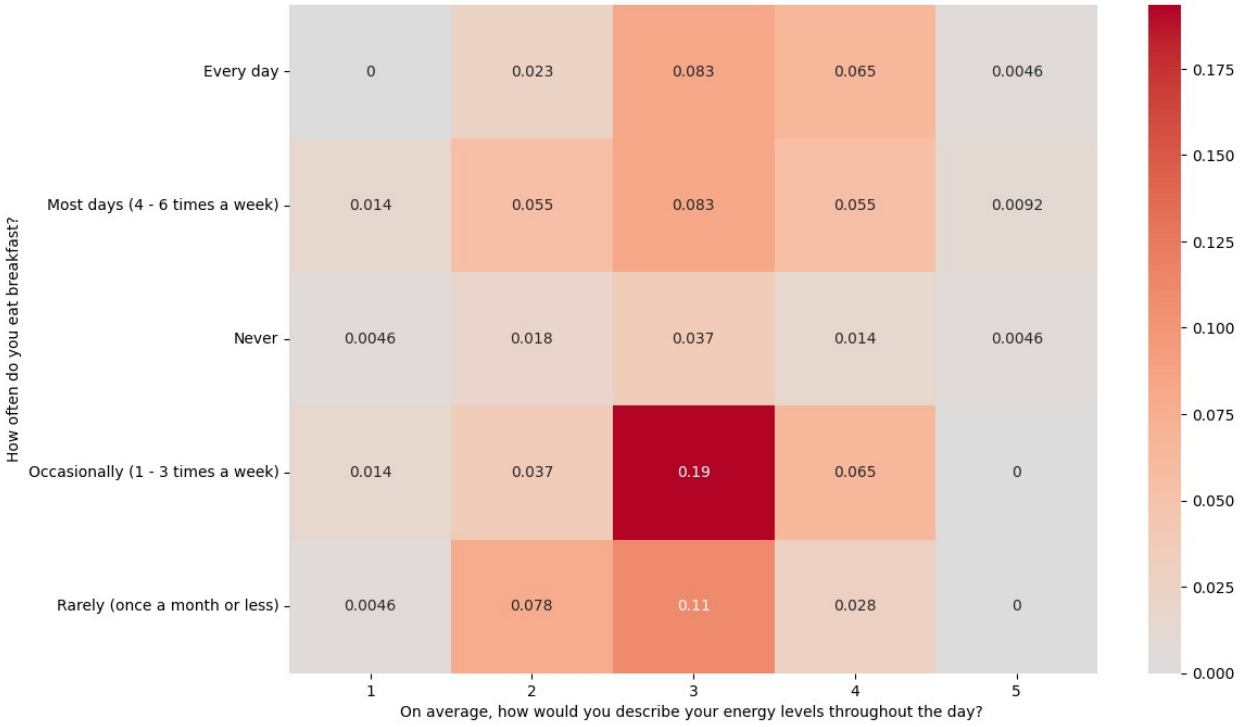
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from scipy.stats import chi2_contingency
from scipy.stats import stats
from statsmodels.stats.multicomp import pairwise_tukeyhsd
df = pd.read_csv('/content/drive/MyDrive/MINI_PROJECT/survey.csv')
```

```
#pivot table
group_columns = [
    'How many meals do you typically eat per day? ',
    'On average, how much water do you drink per day?',
    'How often do you eat breakfast?',
]

measurement_column = 'On average, how would you describe your energy
levels throughout the day? '

df[measurement_column] = pd.to_numeric(df[measurement_column], errors
= 'coerce')
```

```
for col in group_columns:
  counts = df.pivot_table(
  index = col, columns=measurement_column,
  values="CS105",
  aggfunc="count"
  )
  #Heat map
  counts = counts.fillna(0)
  joint = counts/counts.sum().sum()
  plt.figure(figsize=(12, 8))
  sns.heatmap(joint, center= 0, cmap = "coolwarm",annot= True)
```



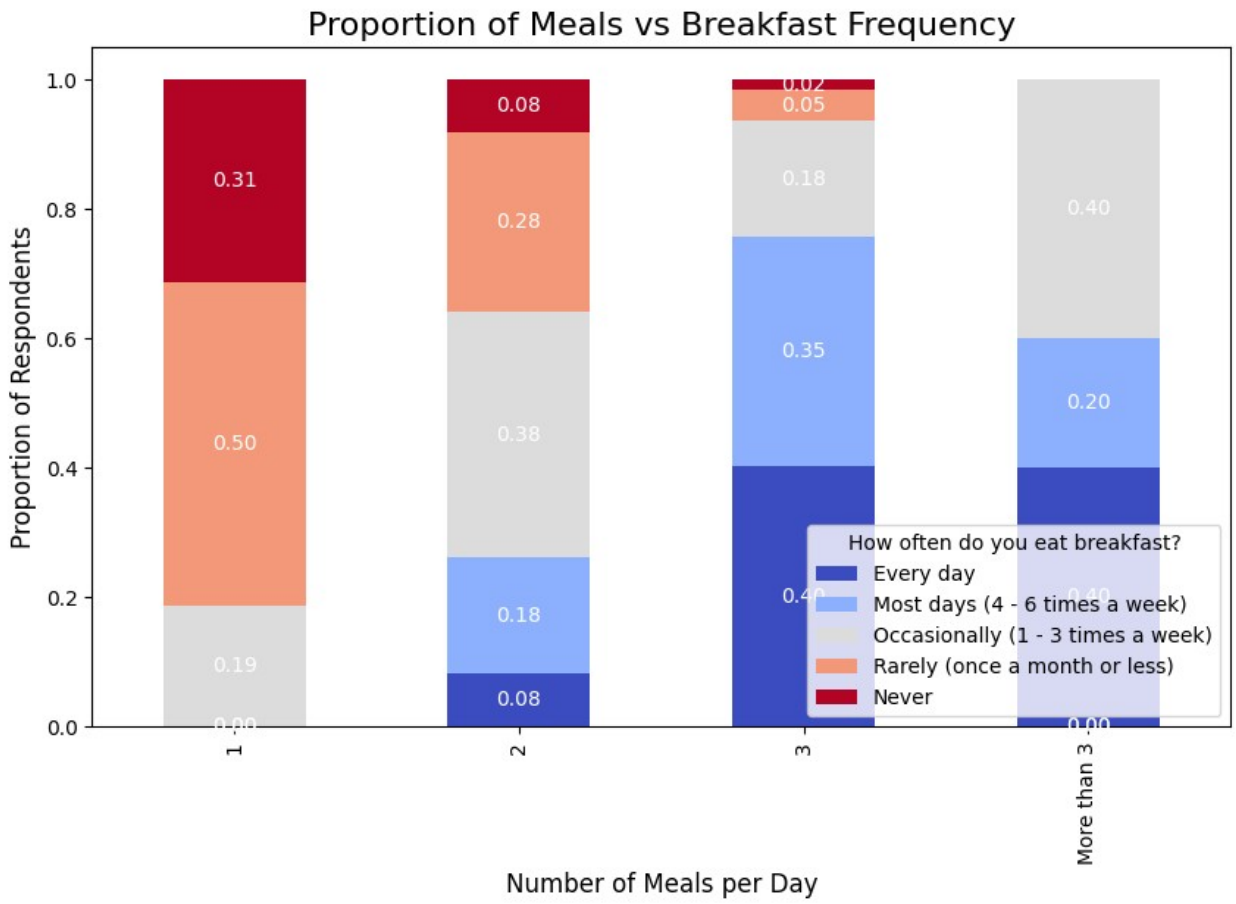




Looking at these heat maps, we can see that the highest concentration of respondants had an energy level of 3, but is hard to grasp the correlation just from the heatmap.

```
df = pd.read_csv('/content/drive/MyDrive/MINI_PROJECT/survey.csv')
```

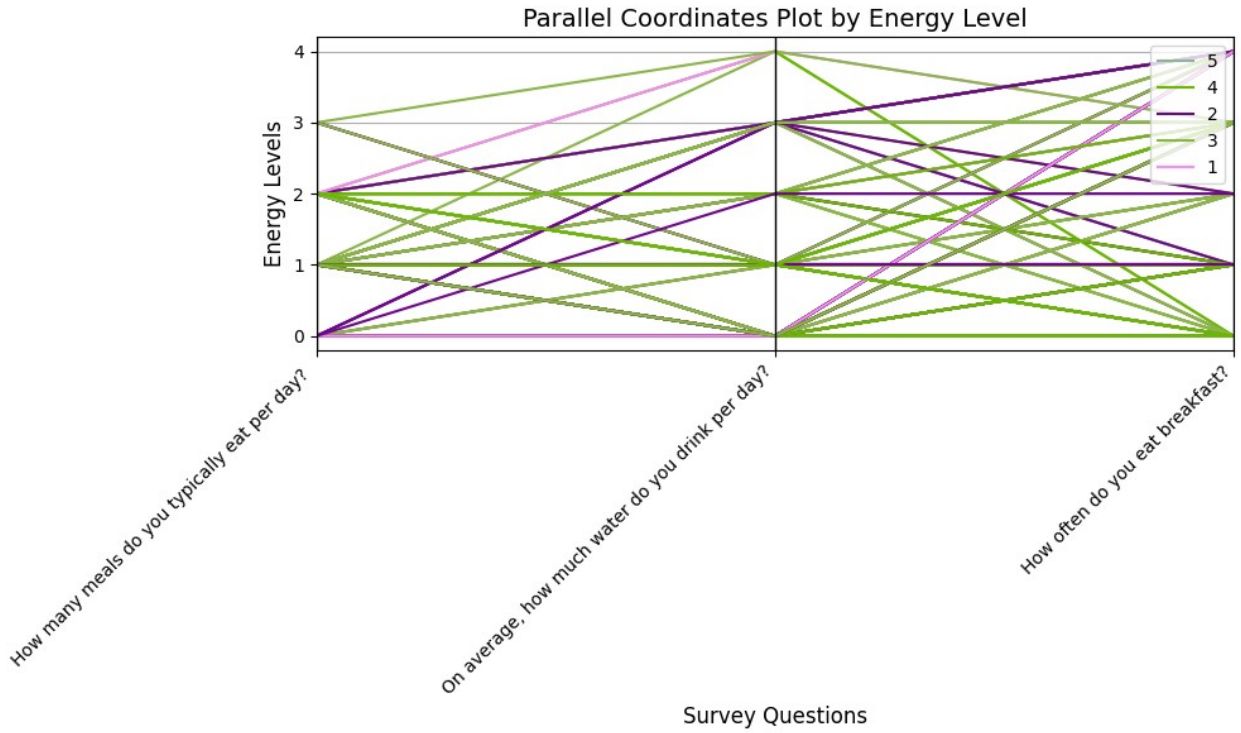
```
breakfast_order = ['Every day', 'Most days (4 - 6 times a week)',
'Occasionally (1 - 3 times a week)', 'Rarely (once a month or less)',
'Never']

counts = df.pivot_table(
    index="How many meals do you typically eat per day? ",
    columns="How often do you eat breakfast?",
    values="CS105",
    aggfunc="count"
)

counts = counts.fillna(0)
counts = counts[breakfast_order]
joint = counts.div(counts.sum(axis=1), axis=0)
# Create the stacked bar chart
ax = joint.plot(kind="bar", stacked=True, figsize=(10, 6),
cmap="coolwarm")

for p in ax.patches:
    height = p.get_height()
    width = p.get_width()
    x = p.get_x() + width / 2
    y = p.get_y() + height / 2

    ax.text(x, y, f'{height:.2f}', ha='center', va='center',
fontsize=10, color='white')

plt.title("Proportion of Meals vs Breakfast Frequency", fontsize=16)
plt.xlabel("Number of Meals per Day", fontsize=12)
plt.ylabel("Proportion of Respondents", fontsize=12)

plt.show()
```



This is a stacked graph comparing the number of meals per day and the frequency of people eating breakfast based on their respective meals per day. This graph helps us examine if there is a correlation between the number of meals per day and eating breakfast, and further allows us to distinguish whether we want to test factors other than energy in our hypothesis

```
df = pd.read_csv('/content/drive/MyDrive/MINI_PROJECT/survey.csv')


df['On average, how would you describe your energy levels throughout
the day? '] = (
    df['On average, how would you describe your energy levels
throughout the day? ']
    .astype('category')
    .cat.codes + 1
)

df['How many meals do you typically eat per day? '] = df['How many
meals do you typically eat per day? '].astype('category').cat.codes
df['On average, how much water do you drink per day?'] = df['On
average, how much water do you drink per
day?'].astype('category').cat.codes
df['How often do you eat breakfast?'] = df['How often do you eat
```

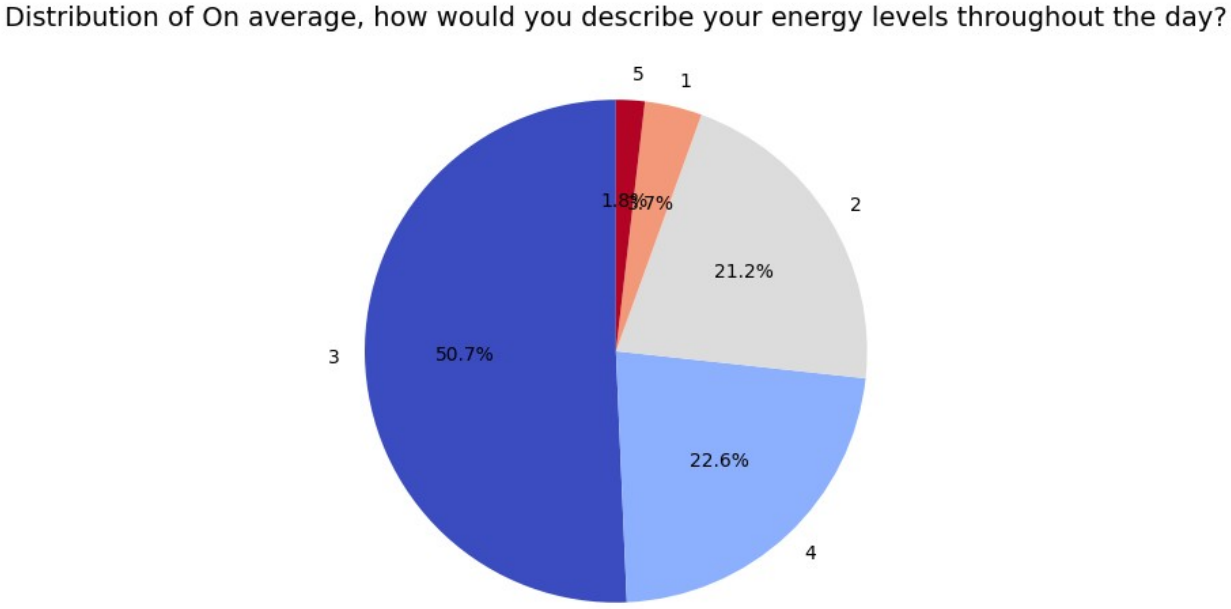
```
breakfast?'].astype('category').cat.codes

# Create parallel coordinates plot
plt.figure(figsize=(10, 6))
parallel_coordinates(df,
                     class_column='On average, how would you describe
your energy levels throughout the day? ',
                     cols=['How many meals do you typically eat per
day? ',
                           'On average, how much water do you drink
per day?',
                           'How often do you eat breakfast?'])

plt.xticks(rotation=45, ha='right', fontsize=10)
plt.xlabel('Survey Questions', fontsize=12)
plt.ylabel('Energy Levels', fontsize=12)
plt.title('Parallel Coordinates Plot by Energy Level', fontsize=14)
plt.tight_layout()
plt.show()
```



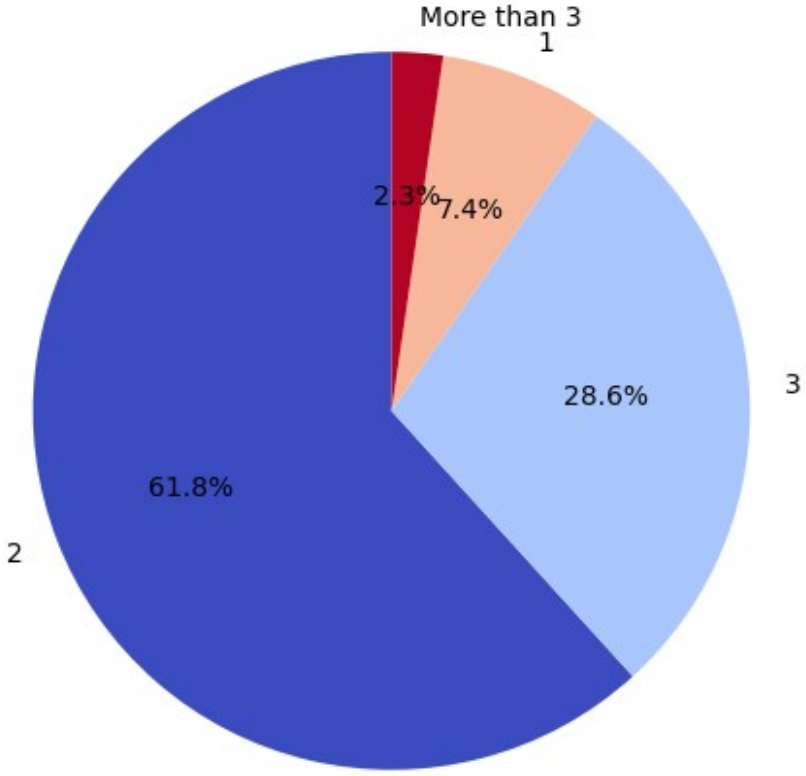
This parallel graph attempts to show the trend of energy levels compared to our other survey questions. As we can see from the graph, there are very limited correlations between all three questions, so we should separate many of the survey questions into separate graphs and tests.

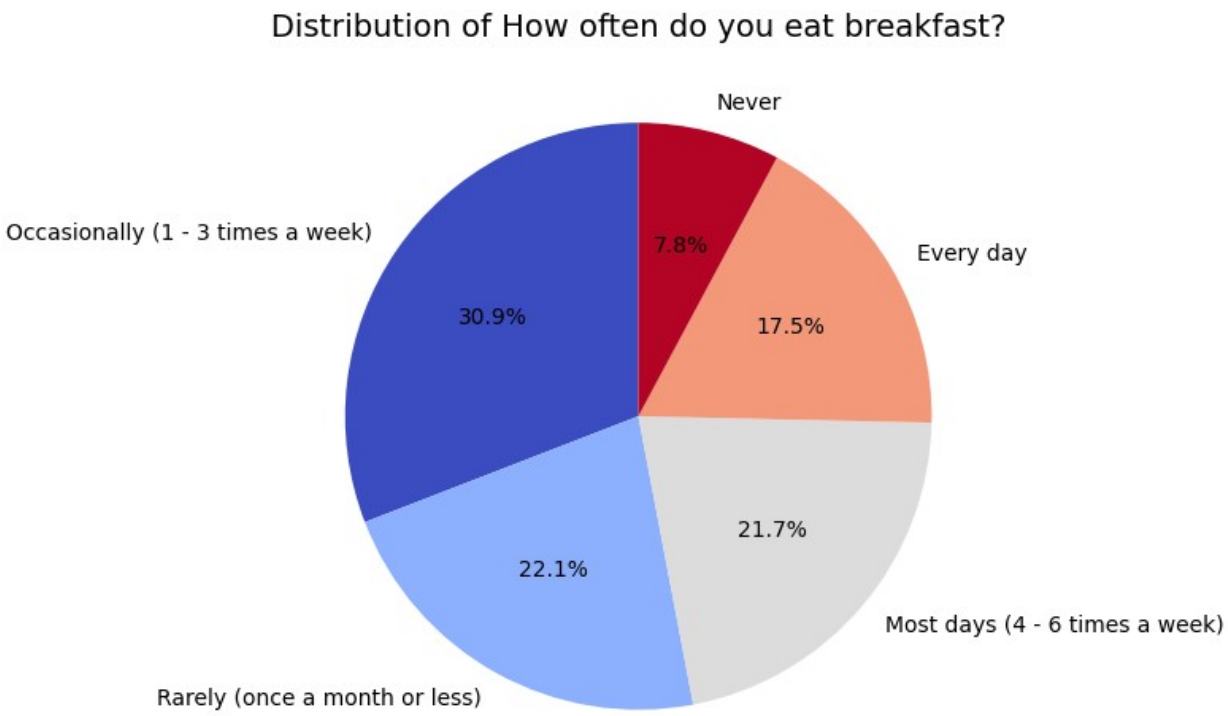
```
df = pd.read_csv('/content/drive/MyDrive/MINI_PROJECT/survey.csv')

# Select the relevant columns
```

```
df = df.iloc[:, 26:31]

# List of columns to create pie charts for
columns = [
    'On average, how would you describe your energy levels throughout
the day? ',
    'How many meals do you typically eat per day? ',
    'How often do you eat breakfast?'
]

for column in columns:
    counts = df[column].value_counts()
    plt.figure(figsize=(6, 6))
    counts.plot.pie(autopct='%1.1f%%', startangle=90, cmap='coolwarm',
legend=False)
    plt.title(f'Distribution of {column.strip()}', fontsize=14)
    plt.ylabel('')

    plt.show()
```



Distribution of How many meals do you typically eat per day?



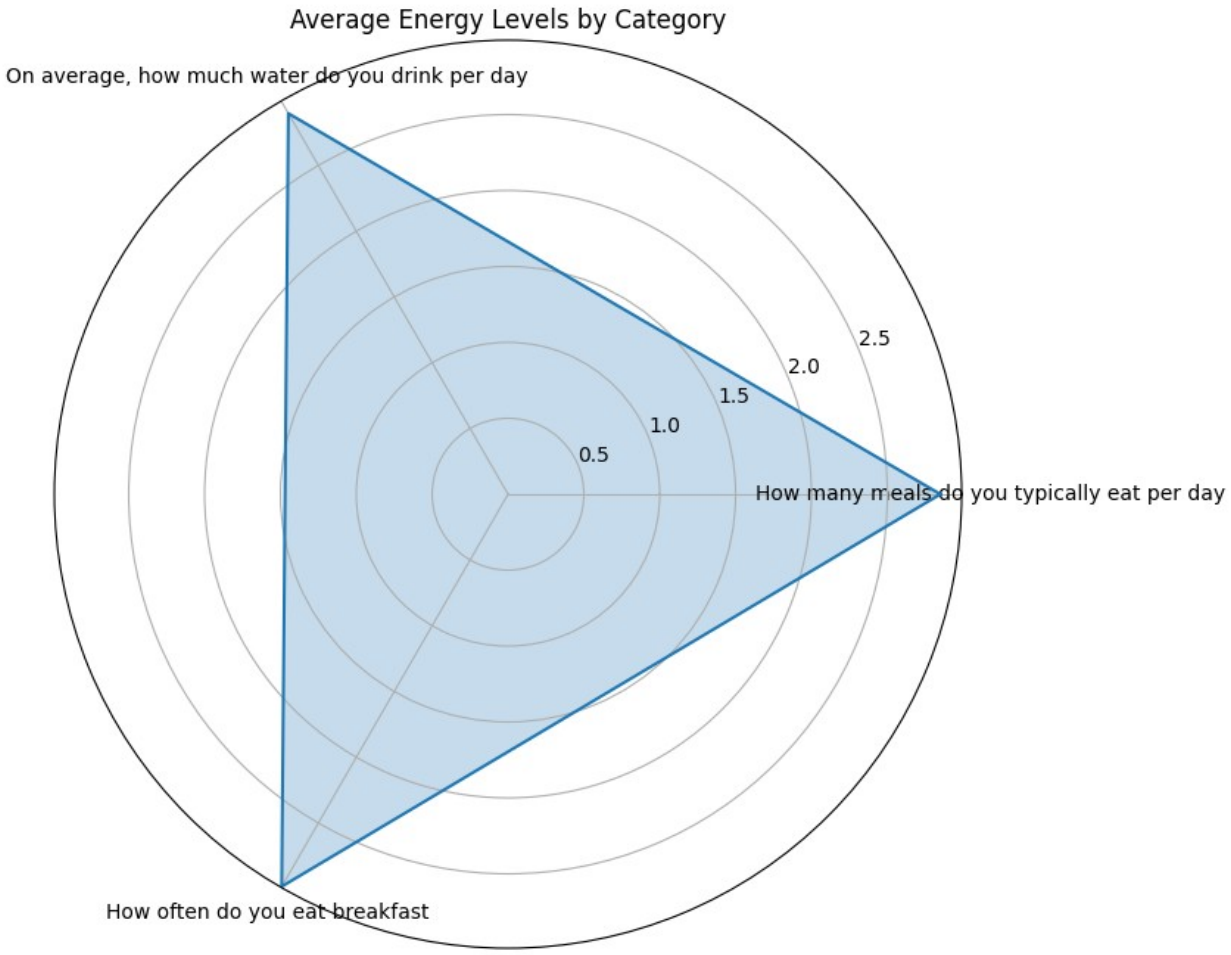


Here we can see the most common repsonses for each of our surveys and we can further split these into graphs comparing these polls with energy levels.

```
means = {}

group_columns = [
    'How many meals do you typically eat per day? ',
    'On average, how much water do you drink per day?',
    'How often do you eat breakfast?',
]

measurement_column = 'On average, how would you describe your energy
levels throughout the day? '

df[measurement_column] = pd.to_numeric(df[measurement_column], errors
= 'coerce')

for col in group_columns:
  means[col] = df.groupby(col)[measurement_column].mean()

angles = np.linspace(0, 2 * np.pi, len(group_columns), endpoint =
False)

values = [means[col].mean() for col in group_columns]
values += values[:1]
angles = np.concatenate((angles, [angles[0]]))
```

```
fig = plt.figure(figsize=(8, 8))
ax = fig.add_subplot(111, projection='polar')
ax.plot(angles, values)
ax.fill(angles, values, alpha=0.25)
ax.set_xticks(angles[:-1])
labels = [col.split('?')[0].strip() for col in group_columns]
ax.set_xticklabels(labels)
plt.title('Average Energy Levels by Category')
print("\nAngles and corresponding categories:")
for angle, label, value in zip(np.degrees(angles[:-1]), labels,
values[:-1]):
    print(f"Angle: {angle:>7.2f}°, Category: {label:<40}, Value:
{value:.2f}")
plt.show()
```

```
Angles and corresponding categories:
Angle:    0.00°, Category: How many meals do you typically eat per
day, Value: 2.85
Angle:  120.00°, Category: On average, how much water do you drink per
day, Value: 2.90
Angle:  240.00°, Category: How often do you eat breakfast          ,
Value: 2.98
```



Analysis:

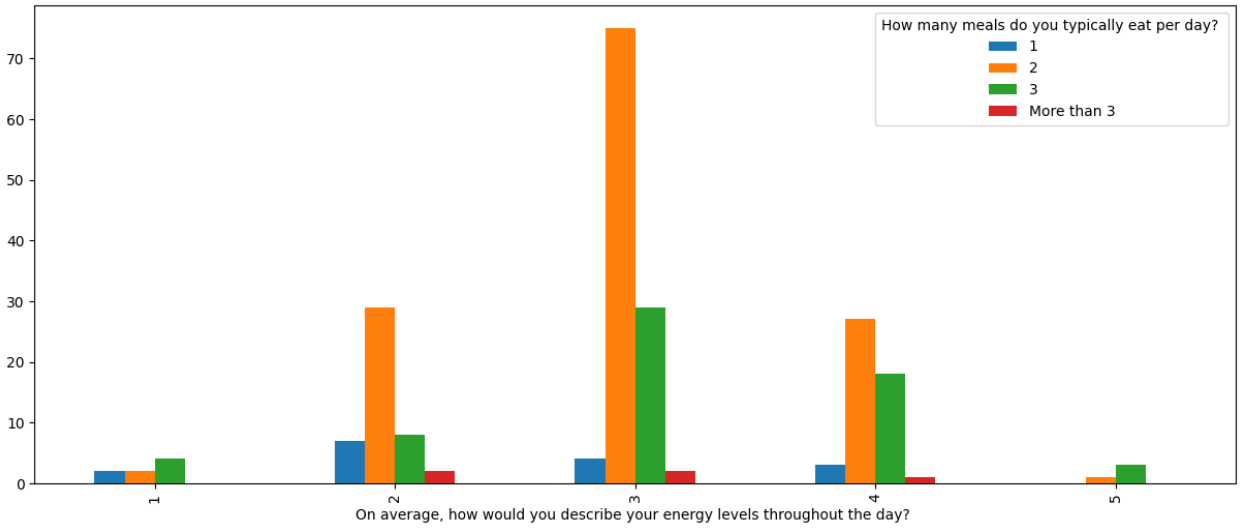
- The Triangle shape seems to be relatively balanced, which indicates that all the factors had similar influence on energy levels.
- The size of the pentagon is relatively large, meaning that the factors had substancial relation with energy levels

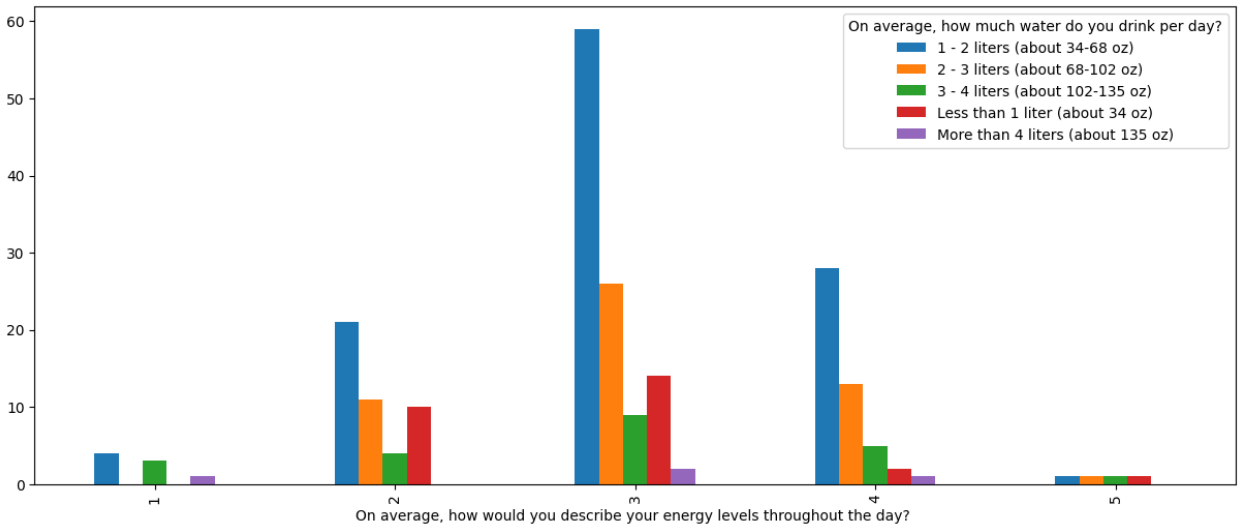
However, we can note that the spider graph is not helpful when it comes to our research, as the mean energy level throughout all three questions were very similar

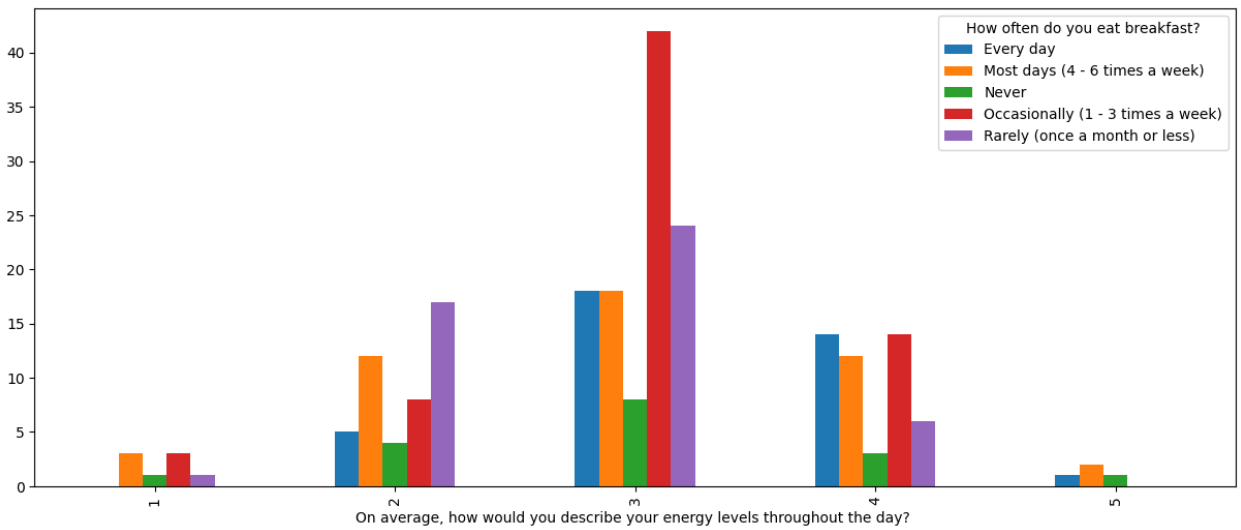
```
df = pd.read_csv('/content/drive/MyDrive/MINI_PROJECT/survey.csv')

df.set_index('CS105', inplace=True)

df = df.sort_values(by='CS105', ascending=False)

df_filtered = df[['How many meals do you typically eat per day? ',
                  'On average, how much water do you drink per day?',
                  'How often do you eat breakfast?',
                  'On average, how would you describe your energy
levels throughout the day? ']]
```

```python
columns_to_plot = ['How many meals do you typically eat per day? ',
                   'On average, how much water do you drink per day?',
                   'How often do you eat breakfast?']

for col in columns_to_plot:
    df_filtered.groupby('On average, how would you describe your
energy levels throughout the day? ').\
        apply(lambda x:
x[col].value_counts()).unstack().plot(kind='bar', figsize=(15, 6))
```

```
<ipython-input-44-479729086e2a>:18: DeprecationWarning:
DataFrameGroupBy.apply operated on the grouping columns. This behavior
is deprecated, and in a future version of pandas the grouping columns
will be excluded from the operation. Either pass
`include_groups=False` to exclude the groupings or explicitly select
the grouping columns after groupby to silence this warning.
  apply(lambda x: x[col].value_counts()).unstack().plot(kind='bar',
figsize=(15, 6))
<ipython-input-44-479729086e2a>:18: DeprecationWarning:
DataFrameGroupBy.apply operated on the grouping columns. This behavior
is deprecated, and in a future version of pandas the grouping columns
will be excluded from the operation. Either pass
`include_groups=False` to exclude the groupings or explicitly select
the grouping columns after groupby to silence this warning.
  apply(lambda x: x[col].value_counts()).unstack().plot(kind='bar',
figsize=(15, 6))
<ipython-input-44-479729086e2a>:18: DeprecationWarning:
DataFrameGroupBy.apply operated on the grouping columns. This behavior
is deprecated, and in a future version of pandas the grouping columns
will be excluded from the operation. Either pass
`include_groups=False` to exclude the groupings or explicitly select
the grouping columns after groupby to silence this warning.
  apply(lambda x: x[col].value_counts()).unstack().plot(kind='bar',
figsize=(15, 6))
```


How many meals do you typically eat per day?
1
2
3
More than 3
0
10
20
30
40
50
60
70
1
2
3
4
5
On average, how would you describe your energy levels throughout the day?


On average, how much water do you drink per day?
1 - 2 liters (about 34-68 oz)
2 - 3 liters (about 68-102 oz)
3 - 4 liters (about 102-135 oz)
Less than 1 liter (about 34 oz)
More than 4 liters (about 135 oz)
0
10
20
30
40
50
60
1
2
3
4
5
On average, how would you describe your energy levels throughout the day?


How often do you eat breakfast?
Every day
Most days (4 - 6 times a week)
Never
Occasionally (1 - 3 times a week)
Rarely (once a month or less)
0
5
10
15
20
25
30
35
40
1
2
3
4
5
On average, how would you describe your energy levels throughout the day?

Here we seperate our different questions and energy values to show all the freqencies of our differnt surverys. Using this we induced that there could be strong correlations between our tests and the energy levels.

Can you state any hypotheses or make predictions? Which tests can you apply to verify your hypothesis?

Hypotheses:

1. The number of meals eaten per day affects your energy levels throughout the day.
   - **Prediction**: As the number of meals eaten per day increases, energy levels throughout the day may improve due to a more consistent energy supply from food intake.
2. The amount of water consumed per day affects your energy levels throughout the day.
   - **Prediction**: Higher water intake is associated with higher energy levels, as dehydration can lead to fatigue and lower energy levels.
3. How often you eat breakfast affects your energy levels throughout the day.
   - **Prediction**: People who eat breakfast more regularly will report higher energy levels throughout the day compared to those who skip breakfast or eat it less frequently.

Tests to Verify Hypotheses:

To verify these hypotheses, we applied the following statistical tests:

1. **ANOVA (Analysis of Variance)**:
   - It compares the variance within each group to the variance between the groups to assess whether any observed differences are likely due to chance or if they reflect a true effect.
2. **Tukey's HSD Test**:
   - **Post-hoc analysis**: If the ANOVA shows significant differences in the groups, Tukey's HSD (Honestly Significant Difference) test can be used to determine which specific groups differ from
3. **Chi Square Analysis**

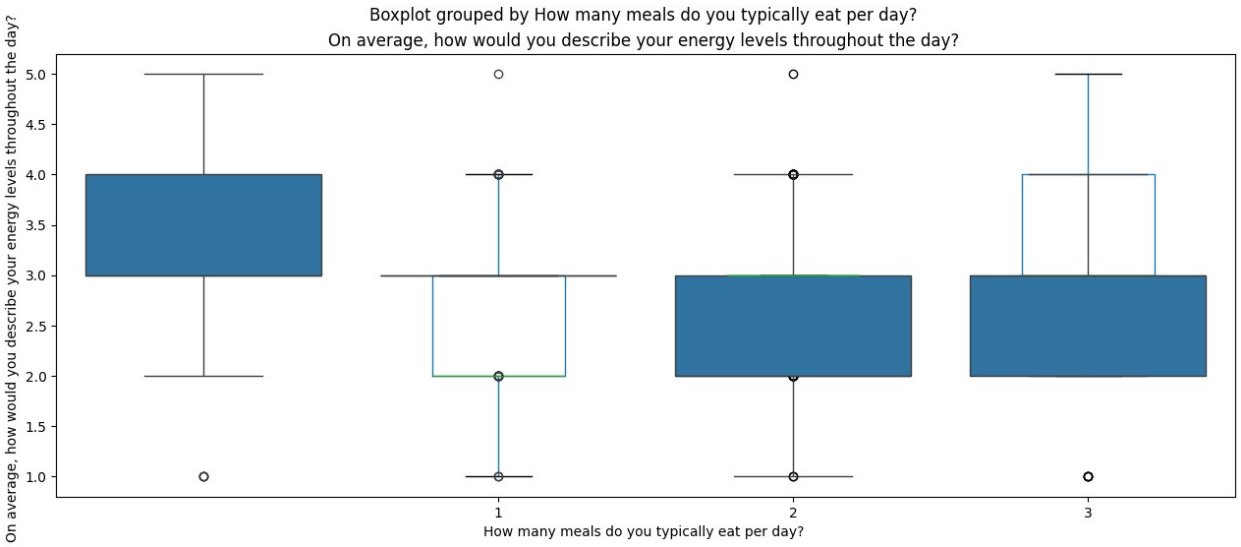
- It compares the observed frequencies with the expected frequencies, and if the calculated Chi-Square statistic exceeds the critical value from the Chi-Square distribution, the null hypothesis is rejected, indicating a significant relationship.

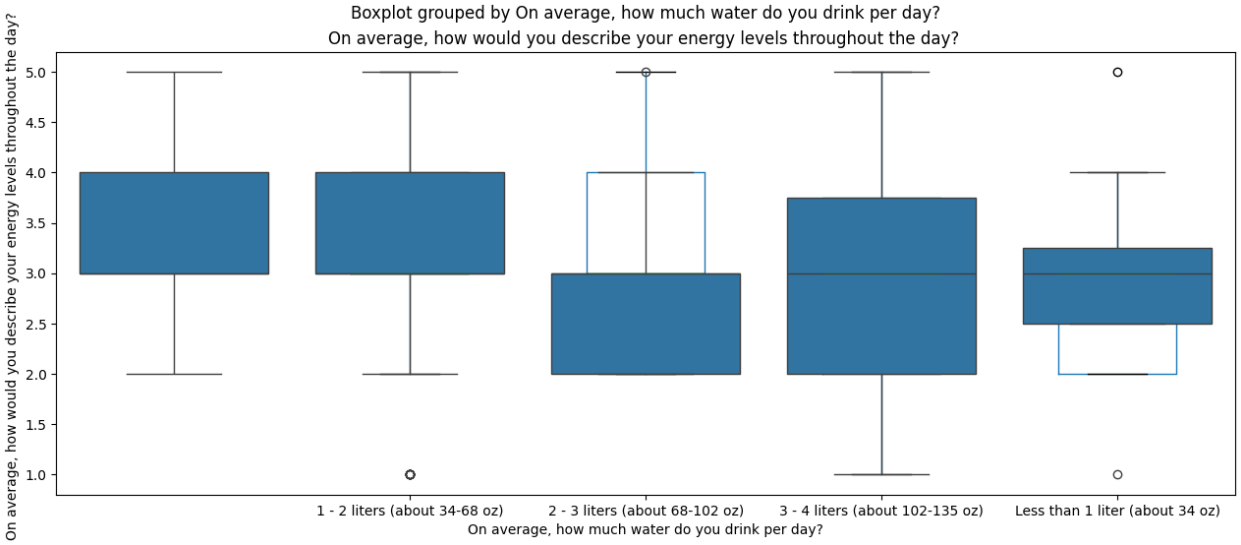
```
"""
Below is ANOVA analysis + Tukey test, followed by box plots on all the
categories in group_columns

"""
#import f_oneway from scipy.stats


group_columns = [
    'How many meals do you typically eat per day? ',
    'On average, how much water do you drink per day?',
    'How often do you eat breakfast?',
```

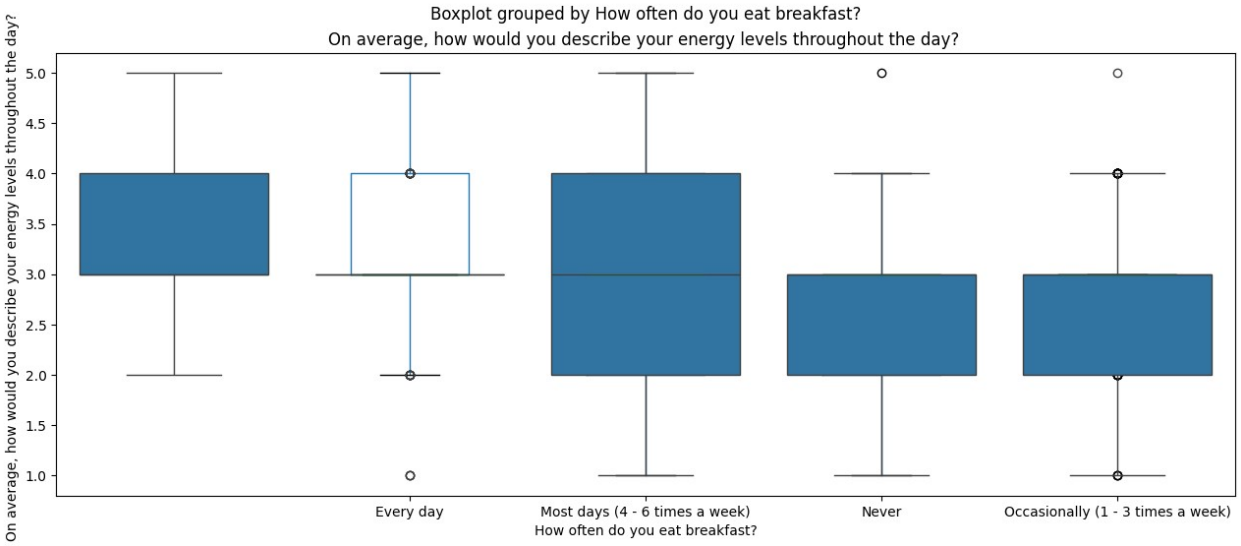
```
]

measurement_column = 'On average, how would you describe your energy
levels throughout the day? '

df[measurement_column] = pd.to_numeric(df[measurement_column], errors
= 'coerce')

def anova(df, group_col, measure_col):
  groups = [group[measure_col].values for name, group in
df.groupby(group_col)]

  f_statistic, p_value = stats.f_oneway(*groups)

  print(f"\nANOVA Results for {group_col}:")
  print(f"F-statistic: {f_statistic:.4f}")
  print(f"p-value: {p_value:.4f}")

  #Tukey test here, if significant (P value < 0.05)
  if p_value < 0.05:
    print("\nTukey's HSD Tests:")
    tukey = pairwise_tukeyhsd(
        endog = df[measure_col],
        groups = df[group_col],
        alpha = 0.05
    )
    print(tukey)

    #boxplots
  df.boxplot(column=measure_col, by=group_col, grid=False,
figsize=(15, 6))

    # Create a boxplot
  sns.boxplot(x=group_col, y=measure_col, data=df)

    # # Title and axis adjustments
    # plt.title(f"Energy Levels by {group_col}")
    # plt.suptitle('')  # To remove the default title generated by the
'by' argument in boxplot
    # plt.xticks(rotation=45)
    # plt.tight_layout()

    #plt.show()

  return f_statistic, p_value

#ANOVA for each group column
results = {}
for group_col in group_columns:
  print(f"ANOVA test for: {group_col}")
```

```
  f_stat, p_val = anova(df, group_col, measurement_column)
  results[group_col] = {'f_statistic': f_stat, 'p_value': p_val}


ANOVA test for: How many meals do you typically eat per day?

ANOVA Results for How many meals do you typically eat per day? :
F-statistic: 2.7010
p-value: 0.0466

Tukey's HSD Tests:

<ipython-input-35-d1be978ff5e6>:21: DeprecationWarning: Please import
`f_oneway` from the `scipy.stats` namespace; the `scipy.stats.stats`
namespace is deprecated and will be removed in SciPy 2.0.0.
  f_statistic, p_value = stats.f_oneway(*groups)

  Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
group1    group2   meandiff p-adj   lower  upper  reject
-----------------------------------------------------
     1           2   0.4701 0.1235 -0.0804 1.0207  False
     1           3    0.629 0.0291  0.0453 1.2127   True
     1 More than 3      0.3 0.8856 -0.7665 1.3665  False
     2           3   0.1589 0.5724 -0.1608 0.4786  False
     2 More than 3  -0.1701 0.9667 -1.1183  0.778  False
     3 More than 3   -0.329  0.815 -1.2968 0.6387  False
-----------------------------------------------------
ANOVA test for: On average, how much water do you drink per day?

ANOVA Results for On average, how much water do you drink per day?:
F-statistic: 0.8287
p-value: 0.5082

<ipython-input-35-d1be978ff5e6>:21: DeprecationWarning: Please import
`f_oneway` from the `scipy.stats` namespace; the `scipy.stats.stats`
namespace is deprecated and will be removed in SciPy 2.0.0.
  f_statistic, p_value = stats.f_oneway(*groups)
<ipython-input-35-d1be978ff5e6>:21: DeprecationWarning: Please import
`f_oneway` from the `scipy.stats` namespace; the `scipy.stats.stats`
namespace is deprecated and will be removed in SciPy 2.0.0.
  f_statistic, p_value = stats.f_oneway(*groups)

ANOVA test for: How often do you eat breakfast?

ANOVA Results for How often do you eat breakfast?:
F-statistic: 2.6213
p-value: 0.0359

Tukey's HSD Tests:
                           Multiple Comparison of Means - Tukey HSD,
```

```
FWER=0.05
=============================================================================
====================================
                group1                                group2
meandiff p-adj   lower   upper  reject
-----------------------------------------------------------------------------
-------------------------------------
                           Every day    Most days (4 - 6 times a week)
-0.332  0.321  -0.813   0.149  False
                           Every day                                Never  -
0.3483 0.5704 -0.9917  0.2951  False
                           Every day Occasionally (1 - 3 times a week)  -
0.2895 0.3887 -0.7373  0.1583  False
                           Every day     Rarely (once a month or less)  -
0.5603 0.0128 -1.0391 -0.0815   True
   Most days (4 - 6 times a week)                                Never  -
0.0163    1.0 -0.6403  0.6078  False
   Most days (4 - 6 times a week) Occasionally (1 - 3 times a week)
0.0426 0.9987  -0.377  0.4621  False
   Most days (4 - 6 times a week)     Rarely (once a month or less)  -
0.2283  0.636 -0.6807  0.2242  False
                             Never Occasionally (1 - 3 times a week)
0.0588 0.9988   -0.54  0.6576  False
                             Never     Rarely (once a month or less)
-0.212  0.882 -0.8343  0.4103  False
Occasionally (1 - 3 times a week)     Rarely (once a month or less)  -
0.2708 0.3838 -0.6878  0.1461  False
-----------------------------------------------------------------------------
-------------------------------------
```







How many meals do you typically eat per day?

We ran a one way anova test with the assumptions that the data follows a normal distribution and all the data collected from one person is independent of another. We are following a significant value of α=0.05

Ho: $\mu 1 = \mu 2 = \mu 3 = \mu 4$

Ha: at least one μi is different from the rest

Our F-statistic is 2.7010 and our p-value is 0.0466 so when we compare it to our critical F-value(2.414) and our significance level of 0.05, we reject the null hypothesis and conclude that at least one μi is different from the rest which lead us to run a post-hoc test which is a Tukey's test.

A Tukey's test concludes that two or more means are significantly different and lessens the chances of a type I error compared to other post-hoc tests. We ran a Tukey's test and found that we can rank the groups of how many meals people eat a day compared to how much energy

they have and we found that 3>1=2= more than 3 which suggests that 3 meals a day gives students more energy compared to eating 1,2,4, or more meals a day.

On average, how much water do you drink per day?

We ran a one way anova test with the assumptions that the data follows a normal distribution and all the data collected from one person is independent of another. We are following a significant value of α=0.05

Ho: μ1 = μ2 = μ3=μ4=μ5

Ha: at least one μi is different from the rest

Our F-statistic is 0.8287and our p-value is 0.5082 so when we compare it to our critical F-value(2.414) and our significance level of 0.05, we fail to reject the null hypothesis and conclude that μ1 = μ2 = μ3 = μ4 = μ5 which means that there is no significant difference between the amount of water a student drinks and how much energy they have throughout the day.

How often do you eat breakfast?

We ran a one way anova test with the assumptions that the data follows a normal distribution and all the data collected from one person is independent of another. We are following a significant value of α=0.05

Ho: μ1 = μ2 = μ3=μ4=μ5

Ha: at least one μi is different from the rest

Our F-statistic is 2.6213 and our p-value is 0.0359 so when we compare it to our critical F-value(2.414) and our significance level of 0.05, we reject the null hypothesis and conclude that at least one μi is different from the rest which lead us to run a post-hoc test which is a Tukey's test.

A Tukey's test concludes that two or more means are significantly different and lessens the chances of a type I error compared to other post-hoc tests. We ran a Tukey's test and found that we can rank the groups of how many people eat breakfast and how often they eat can affect their energy levels throughout the day 5>1=3=4>2. We found that the people who ate breakfast everyday had more energy compared to those who ate breakfast never, occasionally, and most days which still had more energy compared to those who rarely ate breakfast.

```
from scipy.stats import chi2_contingency
df = pd.read_csv('/content/drive/MyDrive/MINI_PROJECT/survey.csv')


topic_columns = [
    "How many meals do you typically eat per day? ",
    "On average, how much water do you drink per day?",
    "How often do you eat breakfast?",
    "On average, how would you describe your energy levels throughout
the day? "]

survey_data = df[topic_columns]
```

```python
def chi_square_test(column, energy_level):
    data = pd.crosstab(survey_data[column], survey_data[energy_level])
    chi, p, dfree, expected = chi2_contingency(data)

    print(f"Calculated Chi-square Value: {chi}")
    print(f"Calculated p-value: {p}")
    print(f"Degrees of Freedom: {dfree}")
    print("Expected Frequencies Table:")
    print(expected, "\n")

print("Meals per Day vs. Energy Levels")
chi_square_test("How many meals do you typically eat per day? ", "On
average, how would you describe your energy levels throughout the day?
")

print("\nWater Intake vs. Energy Levels")
chi_square_test("On average, how much water do you drink per day?",
"On average, how would you describe your energy levels throughout the
day? ")

print("\nBreakfast Frequency vs. Energy Levels")
chi_square_test("How often do you eat breakfast?", "On average, how
would you describe your energy levels throughout the day? ")
```

```
Meals per Day vs. Energy Levels
Calculated Chi-square Value: 22.30097577471199
Calculated p-value: 0.03428181991310866
Degrees of Freedom: 12
Expected Frequencies Table:
[[ 0.58986175  3.39170507  8.11059908  3.61290323  0.29493088]
 [ 4.94009217 28.40552995 67.92626728 30.25806452  2.47004608]
 [ 2.28571429 13.14285714 31.42857143 14.          1.14285714]
 [ 0.1843318   1.05990783  2.53456221  1.12903226  0.0921659 ]]


Water Intake vs. Energy Levels
Calculated Chi-square Value: 23.914472321527857
Calculated p-value: 0.09138922860020784
Degrees of Freedom: 16
Expected Frequencies Table:
[[ 4.16589862 23.95391705 57.28110599 25.51612903  2.08294931]
 [ 1.88018433 10.81105991 25.85253456 11.51612903  0.94009217]
 [ 0.81105991  4.66359447 11.15207373  4.96774194  0.40552995]
 [ 0.99539171  5.7235023  13.68663594  6.09677419  0.49769585]
 [ 0.14746544  0.84792627  2.02764977  0.90322581  0.07373272]]


Breakfast Frequency vs. Energy Levels
Calculated Chi-square Value: 26.628989738602925
```

```
Calculated p-value: 0.0457956333502961
Degrees of Freedom: 16
Expected Frequencies Table:
[[ 1.40092166  8.05529954 19.26267281  8.58064516  0.70046083]
 [ 1.73271889  9.96313364 23.82488479 10.61290323  0.86635945]
 [ 0.62672811  3.60368664  8.61751152  3.83870968  0.31336406]
 [ 2.47004608 14.20276498 33.96313364 15.12903226  1.23502304]
 [ 1.76958525 10.17511521 24.33179724 10.83870968  0.88479263]]
```

##**Chi-Square Conclusions**: Based on the results from each Chi-Square Value, here are our conclusions using a significance level of 0.05:

Meals per Day vs Energy Levels:

Null Hypothesis: There is no correlation between the number of meals eaten and the energy levels in a day.

With a calculated chi square value of ~22.301, and a calculated p-value of ~0.0343 < 0.05, we reject the null hypothesis. There exists a correlation between the number of meals eaten per day and the energy levels of that day.

Water Intake vs Energy Levels:

Null Hypothesis: There is no correlation between the amount of water intake and energy levels in a day.

With a calculated chi square value of ~23.914, and a calculated p-value of ~0.0914 > 0.05, we fail to reject the null hypothesis. There does not exist a correlation between the amount of water intake in a day and the energy levels of that day.

Breakfast Frequency vs Energy Levels

Null Hypothesis: There is no correlation between how often breakfast is eaten and energy levels in a day.

With a calculated chi square value of ~26.629, and a calculated p-value of ~0.0458 < 0.05, we reject the null hypothesis. There exists a correlation between how often breakfast is eaten and the energy levels in a day.

#**Final Thoughts**: After reviewing our results, we can see that most of our comparisons with energy levels show a link between energy levels and diet, aside from water intake which had a p-value greater than our selected signficance level (0.05).

Our hypotheses are mostly supported also through the chi square test, however, water intake failed to show any signs of correlation with energy levels. We concluded that this could be due to food intake playing a larger role in energy levels overall, overshadowing water intake as a whole. Whether or not you have a high water intake, it will not raise energy levels as high if food intake is low. It is evident that eating breakfast often and more meals eaten a day raise energy levels, which prove our two hypotheses true.

Our study indicates that diet plays a large role in our day-to-day lifestyle, and taking care of our health with proper diet plans is crucial to allow us to perform at the highest in terms of physicality. Even though higher water intake is generally known to be beneficial, it is more important to eat enough and healthy for more significantly higher energy levels.

####Project Contributions Aaron:

Created ANOVA, Tukey's HSD Tests, and Box Plots

Created Spider plots + analysis.

Tabito:

Worked on number 1, 2, heatmap, hypothesis, overall debugging throughout the entire project

Karen:

Worked on 1, 2, and analysis for ANOVA and Tukey

Nathan: Worked on 8 different graphs to help visualize the data and create hypotheses from them. Created 3 grouped bar graphs, 3 pie charts, a parallel graph, and a stacked bar graph. Wrote explanations for all and derived hypotheses.

Steven:

Created Chi-Square Statistic and Analysis

Added Final Thoughts