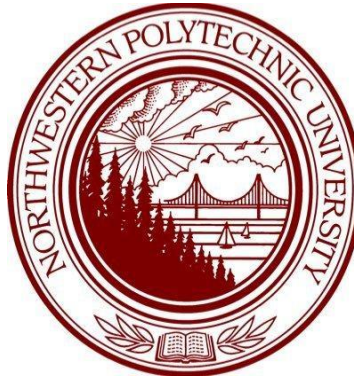


Summer Mania



Prepared For:

Dr. James Nysather

IT501 Business Analytics and Applications

Summer 2021

Prepared By:

Ananya Adyasha

Cindy Bulacan

Dennise Arcilla

Mai Thai

Nur Farabi

Prepared on:

August 18, 2021

QR CODE for Team Questionnaire:



INTRODUCTION

We are the Summer Mania group - Ananya Adyasha, Dennise Arcilla, Cindy Bulacan, Mai Thai, Nur Farabi. We are MBA students at Northwestern Polytechnic University researching how the summer season affects a person's activities and buying decisions as part of our IT 501 Business Analytics and Applications Part B course project.

For this study, Ananya focused on clothing, Dennise on food and beverages, Cindy for water bill/shower behavior, Mai on vacation, and Nur for sports. Our group developed a questionnaire using Google forms, generated graphs using Python programming (Pycharm IDE) and drafted conclusions based on data analysis. Hypothesis and study results are discussed more in detail in each section.

SECTION 1

SUMMER CLOTHING

Analysis by Ananya Adyasha

MOTIVATION

Due to severe temperatures, our choice of clothing and attire change as per the season. I wanted to understand the choice of clothing participants make and its correlation with gender. I was curious to check if the summer spending on clothes is related to a participant's household income.

METHOD

In order to collect responses, I created a survey via Google Forms. Aside from the demographic questions our team has, here are the 5 clothing-specific questions I created:

Q.1 How likely are you to shop for summer clothing every year?

Choice 1: Very unlikely

Choice 2: Unlikely

Choice 3: Neutral

Choice 4: Likely

Choice 5: Very likely

Q.2 How much do you spend on shopping for summer clothing? (Answer in dollars)

Q.3. How important is it to buy summer clothing that are in fashion or trend for that particular year?

Choice 1: Not at all important

Choice 2: Less important

Choice 3: Neutral

Choice 4: Bit important

Choice 5: Very important

Q.4 What are the types of clothing you prefer to buy in the summer?

Choice 1: Beach Wear

Choice 2: Shorts/Skirts

Choice 3: Dresses

Choice 4: Sandals/Flipflops

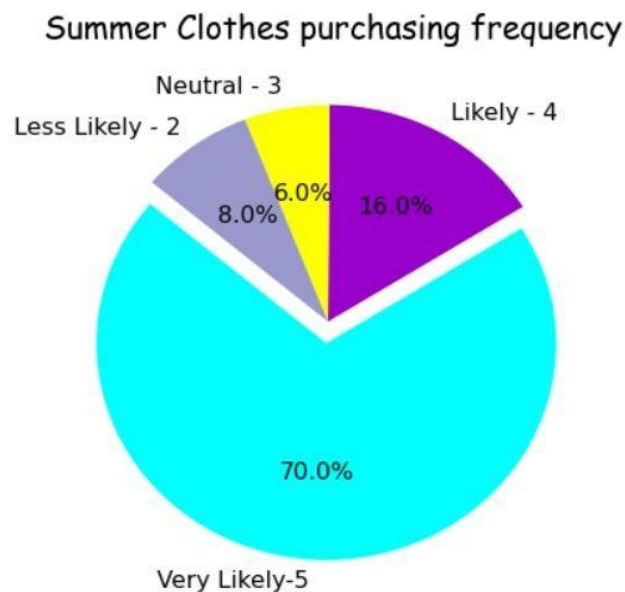
Choice 5: Others

Q.5 How many times in a week do you shop for summer clothing?

MECHANICS:

I used Python and related packages in order to generate graphs like Pie chart, Histogram, Bar plot, Scatter Plot and Boxplot. Using Python, it helped me analyze the data I gathered and provide a clear picture of my findings.

PIE CHART



```

# Import the necessary modules
import pandas as pd
import matplotlib.pyplot as plt

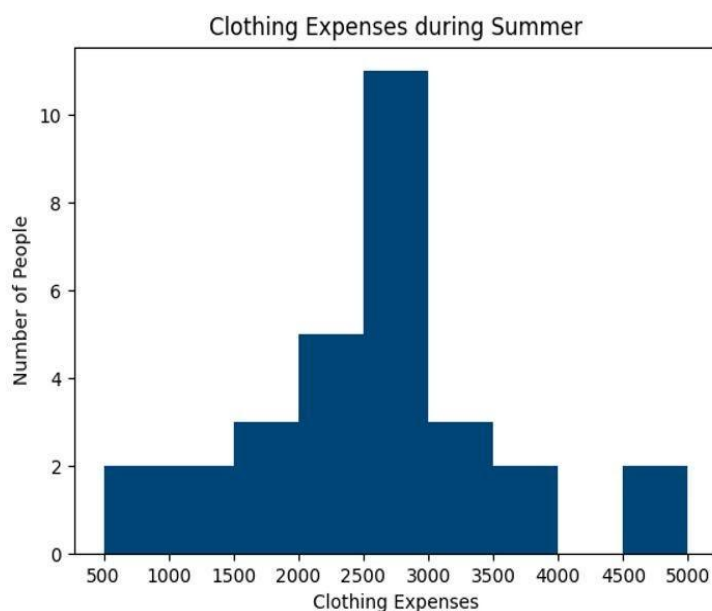
# Initialize the lists for X and Y
df = pd.read_csv('Summer_Shopping.csv')
print(df.head())
plt.rcParams['font.size'] = 11.0
Portal_data = df['Shopping']
PastOrder_data = df['Responses']
colors = ["#00FF00", "#9900CC", "#FF3300", "#8b4726"]
explode = (0.1, 0, 0, 0)

# Plot the data using pie() method
plt.pie(PastOrder_data, labels=Portal_data, explode= explode, colors=colors, autopct='%1.1f%%',
shadow=False, startangle=140)
plt.title(" Summer Clothes purchasing frequency", fontdict={'fontname':'Comic Sans MS' , 'fontsize': 15})

# Show the plot
plt.show()

```

HISTOGRAM:



```

# Import the necessary modules
import pandas as pd
import matplotlib.pyplot as plt

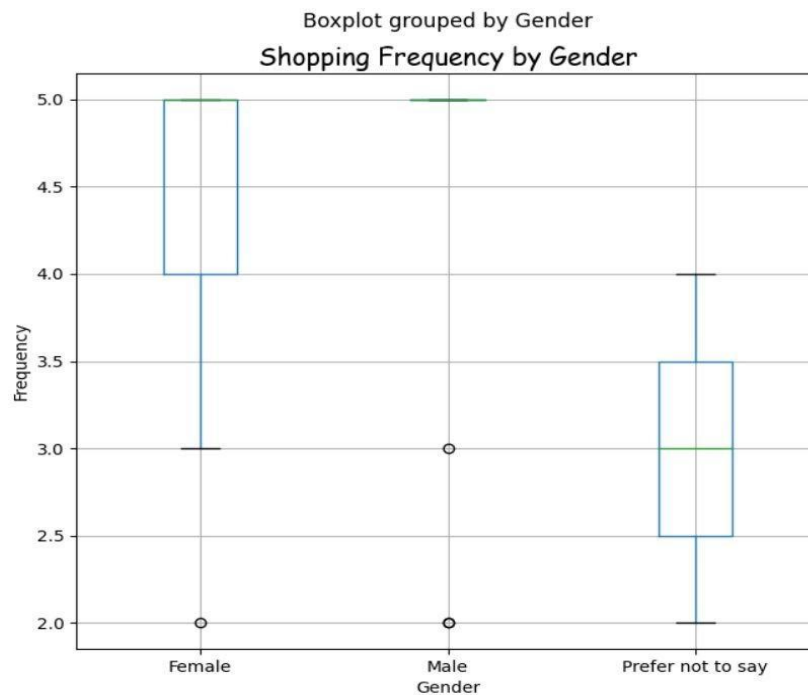
# Initialize the lists for X and Y
His= pd.read_csv('Histogram.csv')
print(His.head(3))
bins = [500,1000,1500,2000,2500,3000,3500,4000,4500,5000]

# Plot the data using hist() method
plt.hist(His.ClothingExpenses, bins=bins, color='#004578')
plt.xticks(bins)
plt.ylabel("Number of People")
plt.xlabel("Clothing Expenses")
plt.title('Clothing Expenses during Summer')

# Show the plot
plt.show()

```

BOX PLOT:



```

# Import the necessary modules
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use('default')

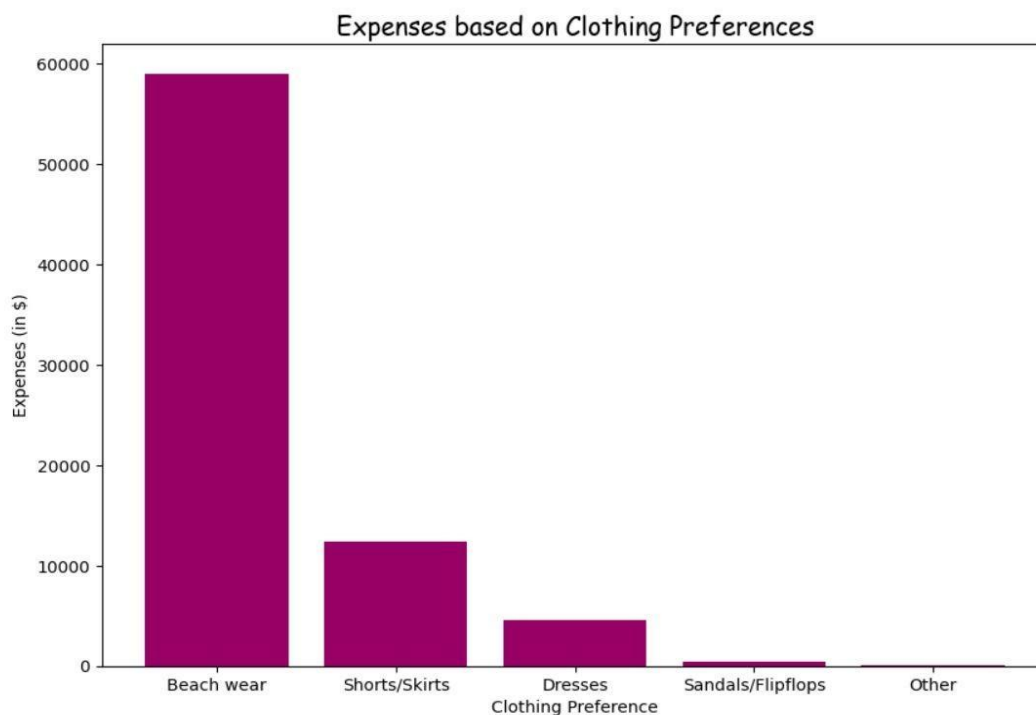
# Initialize the lists for X and Y
box = pd.read_csv('BoxPlot.csv', sep=',', na_values='.')
print(box.head(3))

# Plot the data using boxplot() method
boxes = box.boxplot(column=['summer_clothing'], by='Gender', figsize=(5,7), grid=True)
plt.title("Shopping Frequency by Gender", fontdict={'fontname':'Comic Sans MS', 'fontsize':
15})
plt.ylabel('Frequency')

# Show the plot
plt.show()

```

BAR PLOT:



```

# Import the necessary modules
import pandas as pd
import matplotlib.pyplot as plt

# Initialize the lists for X and Y
data = pd.read_csv('BarPlot.csv')

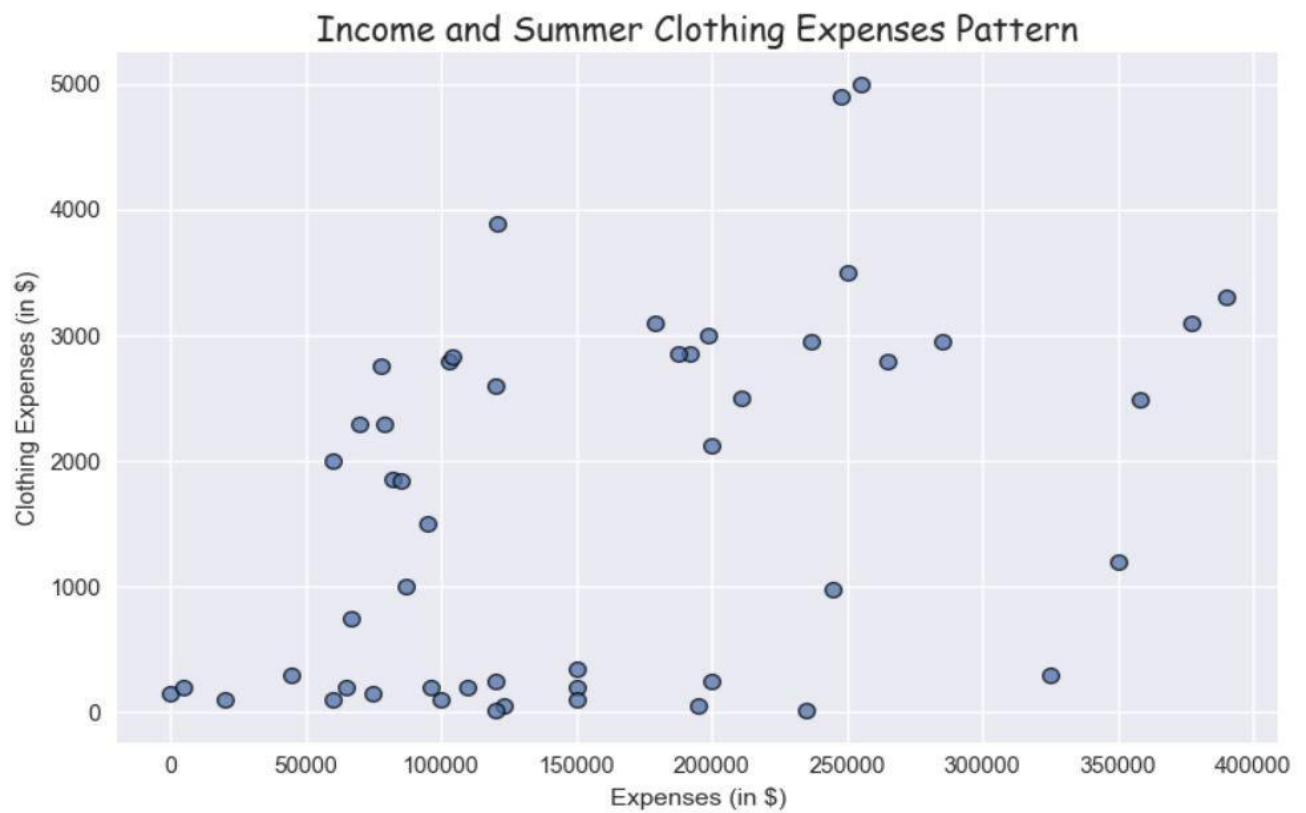
df = pd.DataFrame(data)

print(df.head())
X = list(df.iloc[:, 0])
Y = list(df.iloc[:, 1])

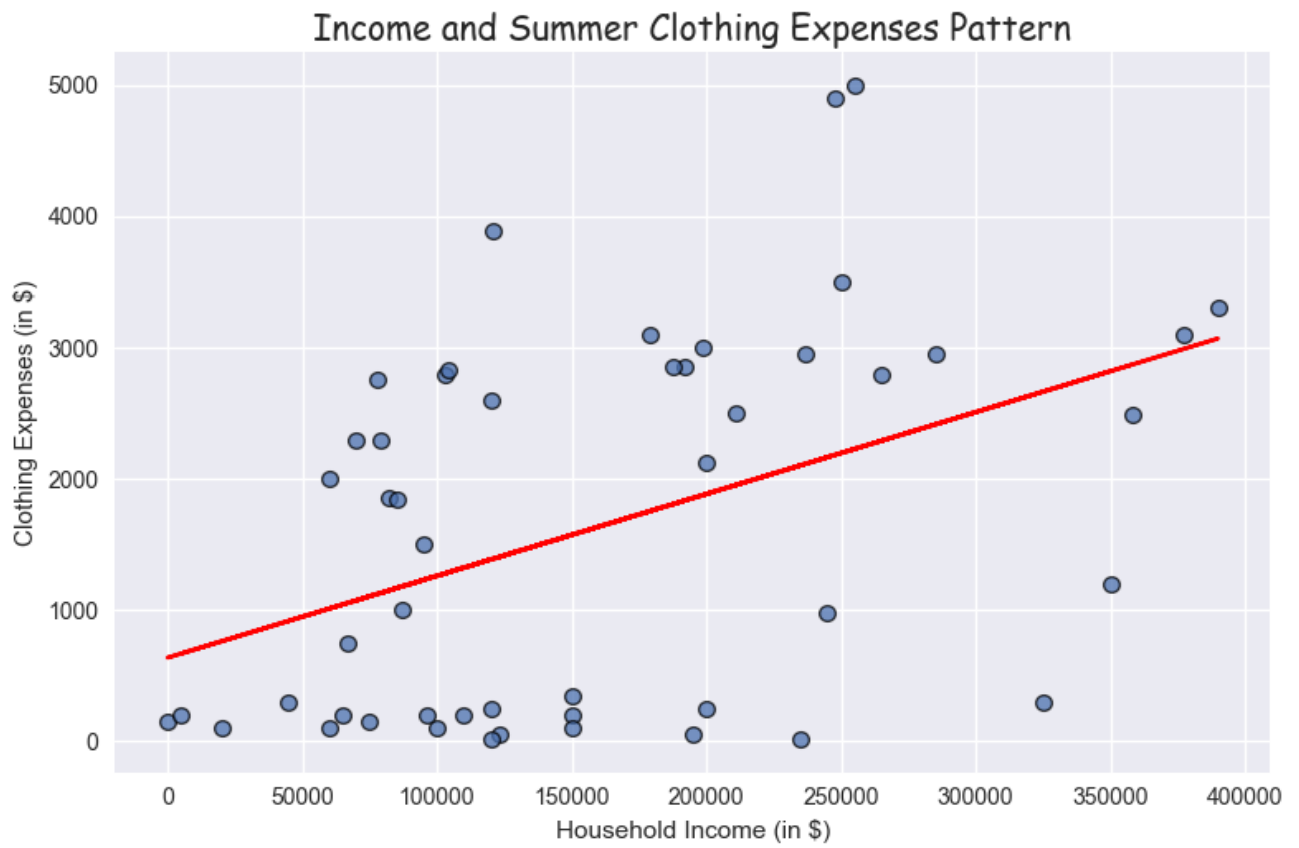
plt.figure(figsize=(10,7))

```

SCATTER PLOT:



SCATTER PLOT with REGRESSION LINE:



```
# Import the necessary modules
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use('seaborn')

# Initialize the lists for X and Y
sp = pd.read_csv('ScatterPlot.csv')
print(sp.head())
Inc= sp['Income']
Expenses= sp['clothing_spend']

# Plot the data using scatter() method
plt.scatter(Inc, Expenses, edgecolor= 'black', linewidth= 1, alpha= 0.75)

plt.title('Income and Summer Clothing Expenses Pattern', fontdict={'fontname':'Comic Sans MS', 'fontsize': 15})
plt.xlabel('Expenses (in $)')
plt.ylabel('Clothing Expenses (in $)')
plt.tight_layout()

# Show the plot
plt.show()
```

MESSAGE:

Based on 50 responses to our questionnaire, here are the primary findings:

**** Majority of our respondents are Female (68%) and 28% are male, 4% prefer not to say.**

**** 78% is aged between 18-34 years old**

**** 70% said they're likely to shop for summer clothes every year and almost 25% of participants spend around \$2500 - \$3000 in buying summer clothes.**

** 56% of respondents were female and have a shopping frequency of 5 times per week for summer clothes.

** Almost 85% of respondents prefer beach wear as a preference, however there isn't a correlation between household income and clothing expenses.

** Since, younger people are more likely to shop for clothes and beach wear during summer, the manufacturer and retail stores should keep that in mind while creating designs, so the clothes are trendier and attract more young customers.

=====

SECTION 2

WATER USAGE / SHOWER BEHAVIOR

Analysis by Cindy Bulacan

MOTIVATION

To know the most affected water usage during summer and how the summer season affects the shower behavior of people.

METHOD

To create and prepare a survey about the water usage and shower behavior of people during summer which will be given to the class. It comprises a total of 7 questions leading to the analysis of their shower behavior:

1. What is your household water bill for the month of June 2021?
2. Based on your experience, which water usage/consumption is the MOST affected during summer?
 - Shower
 - Toilet
 - Faucet
 - Clothes Washer
 - Others
3. When do you prefer to take your showers during summer?
 - Morning
 - Afternoon
 - Night
4. On average, how many times in a day do you take a shower during summer?
 - 1
 - 2
 - 3
 - 4
 - more than 4

- 5.) On average, how long do you take a shower during summer? (Answer in minutes)
- 6.) At what weather temperature do you start taking more showers in the summer than usual?
(Answers in Fahrenheit)
- 60 - below 67
 - 67 - below 75
 - 75 - below 82
 - 82 - 90
 - more than 90
- 7.) How likely does the high temperature in summer affect your shower behavior?
- Very Likely
 - Likely
 - Neutral
 - Not Likely
 - Very Unlikely
-

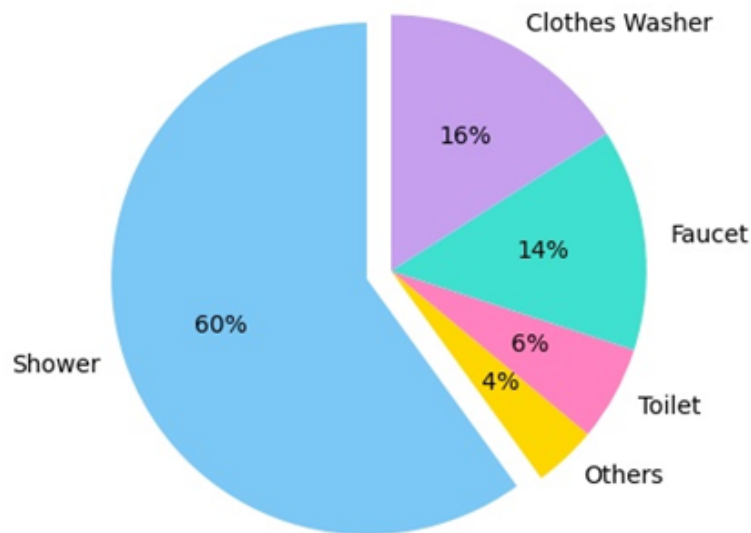
MECHANICS

Create a pie chart, bar plot, histogram, box plot, and scatter plot on Python using the survey results based on 50 respondents.

PIE CHART

The pie chart shows the most affected water usage/consumption during summer.

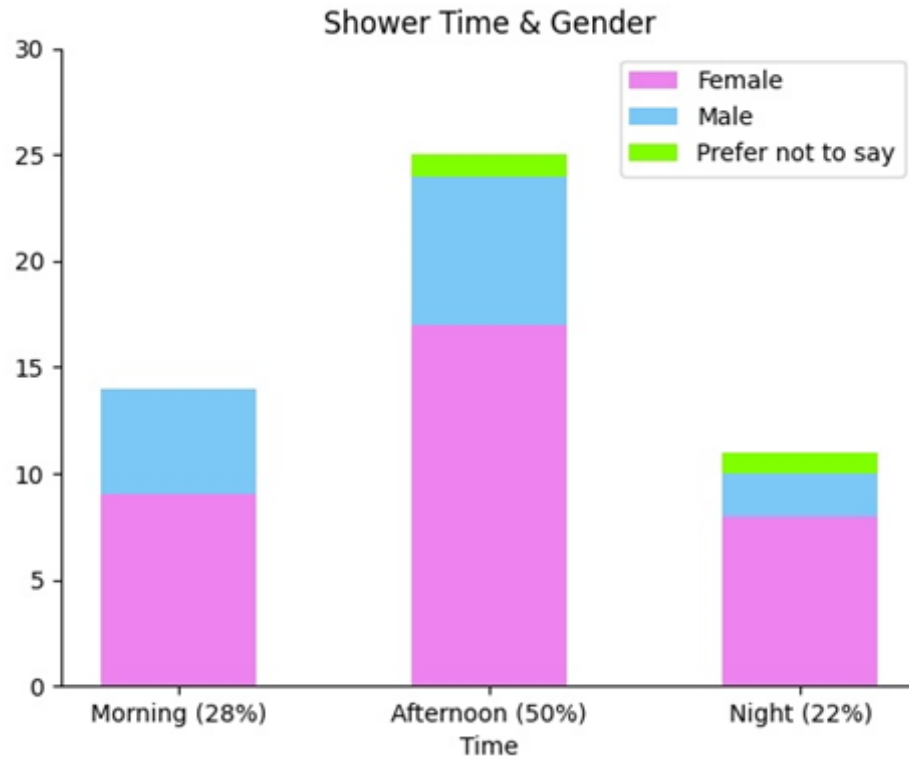
Most Affected Water Usage During Summer



```
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('piechart.csv')
water = data["Water Usage"]
response = data["Response"]
colors = ["#7BC8F6", "#FFD700", "#FF81C0", "#40E0D0", "#C79FEF"]
myexplode = [0.1, 0, 0, 0, 0]
plt.pie(response, labels=water, colors=colors, autopct='%2.0f%%', startangle=90, explode=myexplode)
plt.title("Most Affected Water Usage During Summer")
plt.show()
```

BAR PLOTS

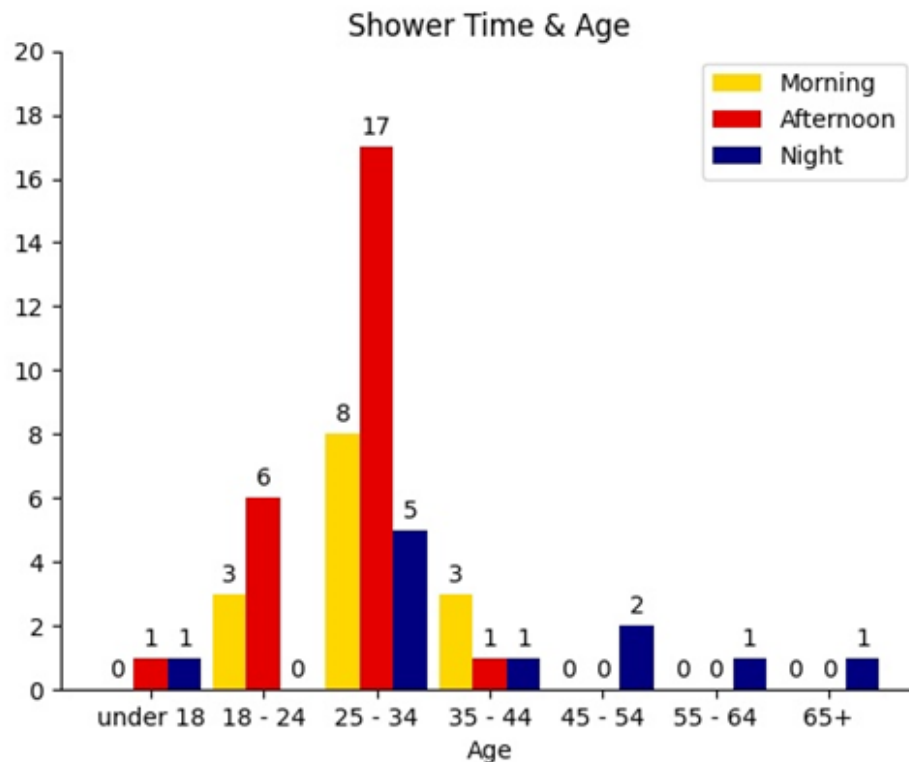
The first bar plot shows relevant data about the shower time preference by gender.



```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
data = pd.read_csv('barplot.csv')
print(data.head(3))
result = data[['Time', 'Resp']]
print(result)
x = ['Morning (28%)', 'Afternoon (50%)', 'Night (22%)']
y1 = pd.array([9, 17, 8])
y2 = pd.array([5, 7, 2])
y3 = pd.array([0, 1, 1])
plt.bar(x, y1, width=0.50, color='#EE82EE')
plt.bar(x, y2, width=0.50, bottom=y1, color='#7BC8F6')
plt.bar(x, y3, width=0.50, bottom=y1 + y2, color='#7FFF00')
plt.legend(["Female", "Male", "Prefer not to say"], loc='upper right')
plt.title("Shower Time & Gender")
```

```
plt.xlabel('Time')
plt.yticks(np.arange(0, 31, 5))
ax = plt.subplot()
ax.spines["top"].set_visible(False)
ax.spines["right"].set_visible(False)
plt.show()
```

The second bar plot shows analysis on the shower time preference by age group.



```
import matplotlib.pyplot as plt
import numpy as np
morning = [0, 3, 8, 3, 0, 0, 0]
afternoon = [1, 6, 17, 1, 0, 0, 0]
night = [1, 0, 5, 1, 2, 1, 1]
x = np.arange(7)
bar_width = 0.3
fig, ax = plt.subplots()
rects1 = ax.bar(x, morning, bar_width, color='#FFD700', label='Morning')
rects2 = ax.bar(x+bar_width, afternoon, bar_width, color='#E50000', label='Afternoon')
rects3 = ax.bar(x+2*bar_width, night, bar_width, color='#000080', label='Night')
ax.set_title('Shower Time & Age')
ax.set_xticks(x + bar_width)
ax.set_xticklabels(('under 18', '18 - 24', '25 - 34', '35 - 44', '45 - 54', '55 - 64', '65+'))
ax.legend(loc='upper right')
ax.bar_label(rects1, padding=3)
ax.bar_label(rects2, padding=3)
```



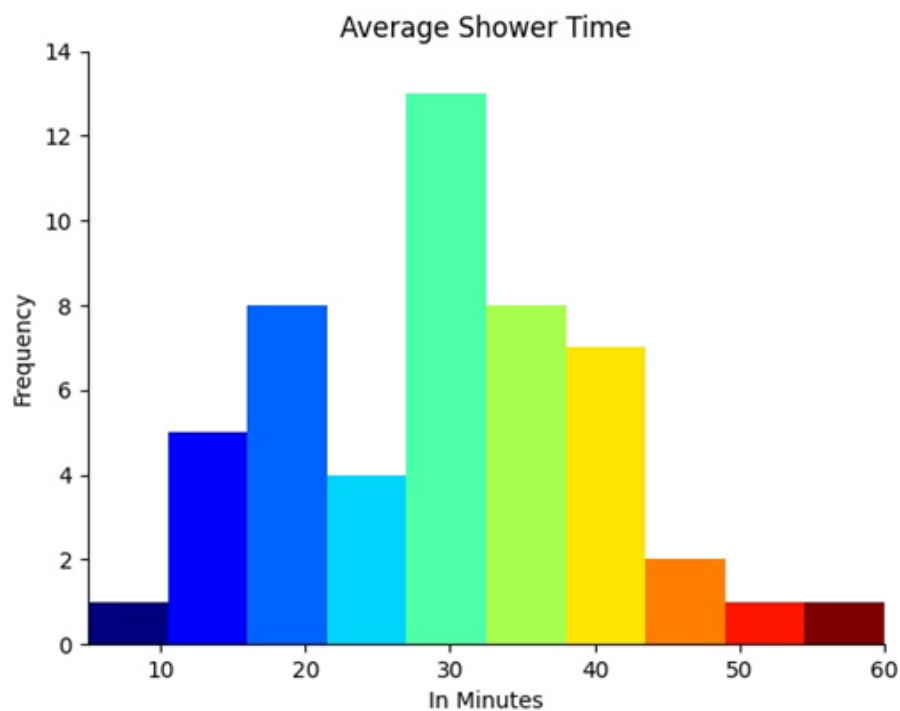
```

ax.bar_label(rects3, padding=3)
plt.yticks(np.arange(0, 21, 2))
plt.xlabel('Age')
ax = plt.subplot(111)
ax.spines["top"].set_visible(False)
ax.spines["right"].set_visible(False)
plt.show()

```

HISTOGRAM

The histogram shows the frequency distribution of the time spent in the shower in minutes.



```

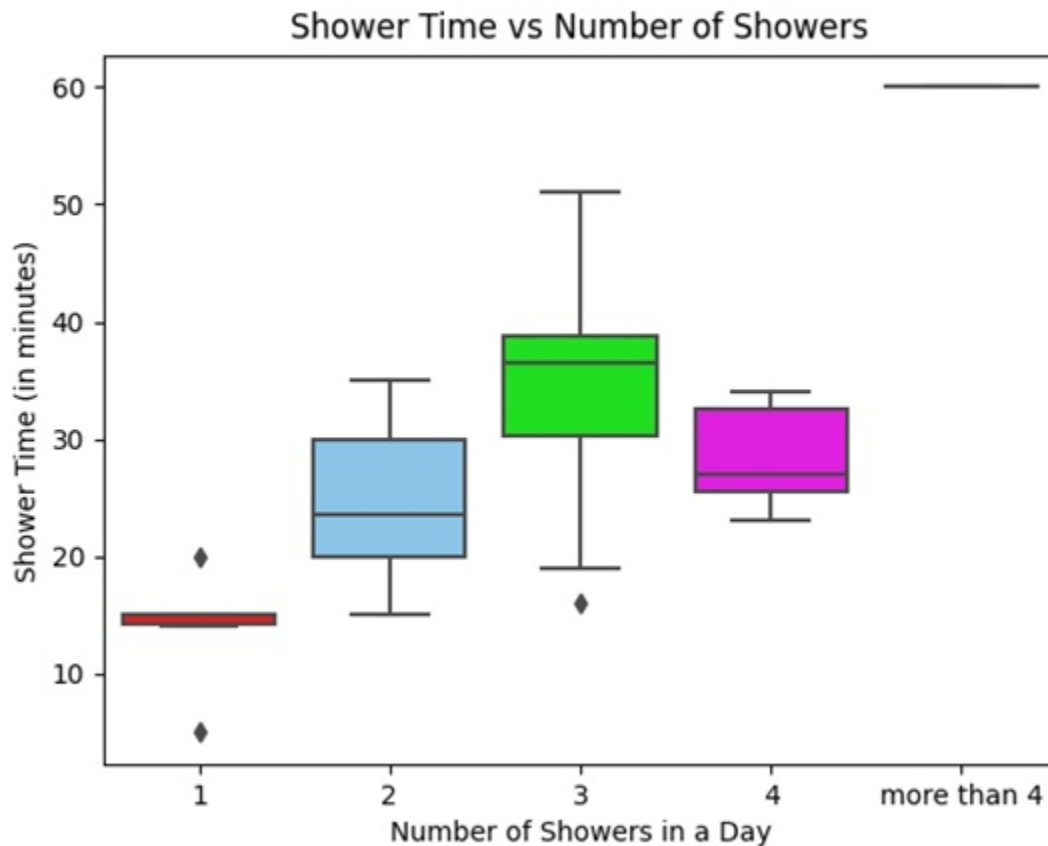
import pandas as pd
import matplotlib.pyplot as plt
import pylab as pl
hist = pd.read_csv('histogram.csv')
print(hist)
N, bins, patches = pl.hist(data=hist, x='SH_time', bins=10)
jet = pl.get_cmap('jet', len(patches))
for i in range(len(patches)):
    patches[i].set_facecolor(jet(i))
plt.xlabel('In Minutes')
plt.ylabel('Frequency')
plt.title("Average Shower Time")
plt.axis([5, 60, 0, 14])
ax = plt.subplot(111)
ax.spines["top"].set_visible(False)

```

```
ax.spines["right"].set_visible(False)
plt.show()
```

BOX PLOT

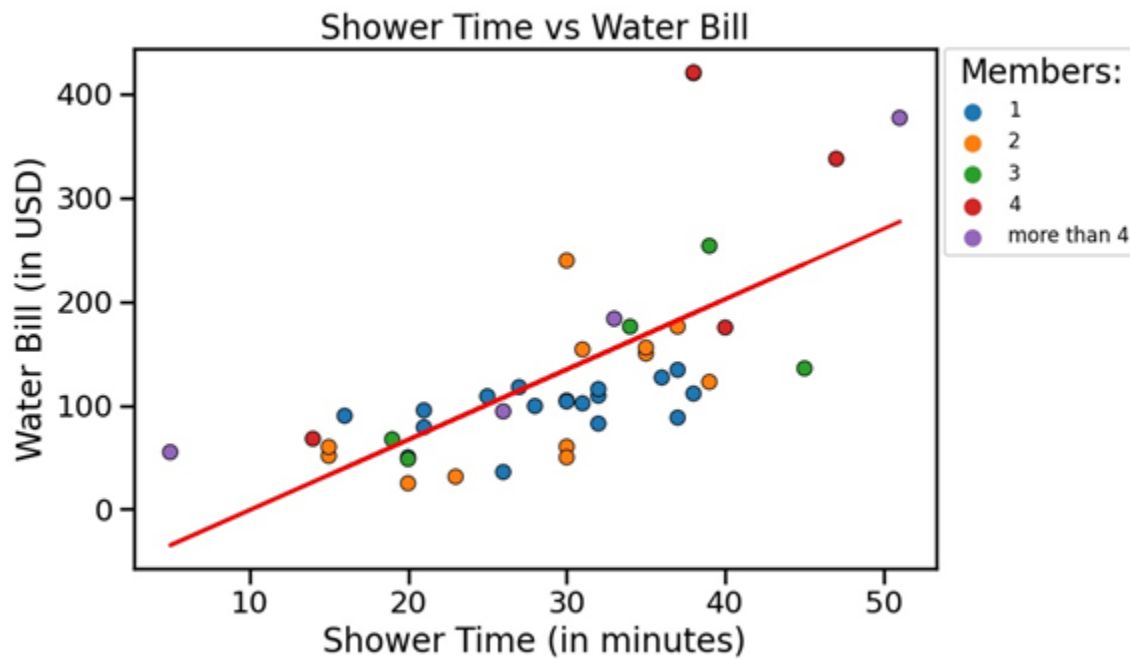
The box plot shows the distribution of the time spent in the shower with the number of showers in a day and its skewness and outliers.



```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
box = pd.read_csv('boxplot.csv')
my_pal = {"1": "red", "2": "#7BC8F6", "3": "lime", "4": "magenta", "more than 4": "yellow"}
sns.boxplot(x='shower_times', y='shower_minutes', data=box, palette=my_pal)
plt.xlabel('Number of Showers in a Day')
plt.ylabel('Shower Time (in minutes)')
plt.title('Shower Time vs Number of Showers')
plt.show()
```

SCATTER PLOTS

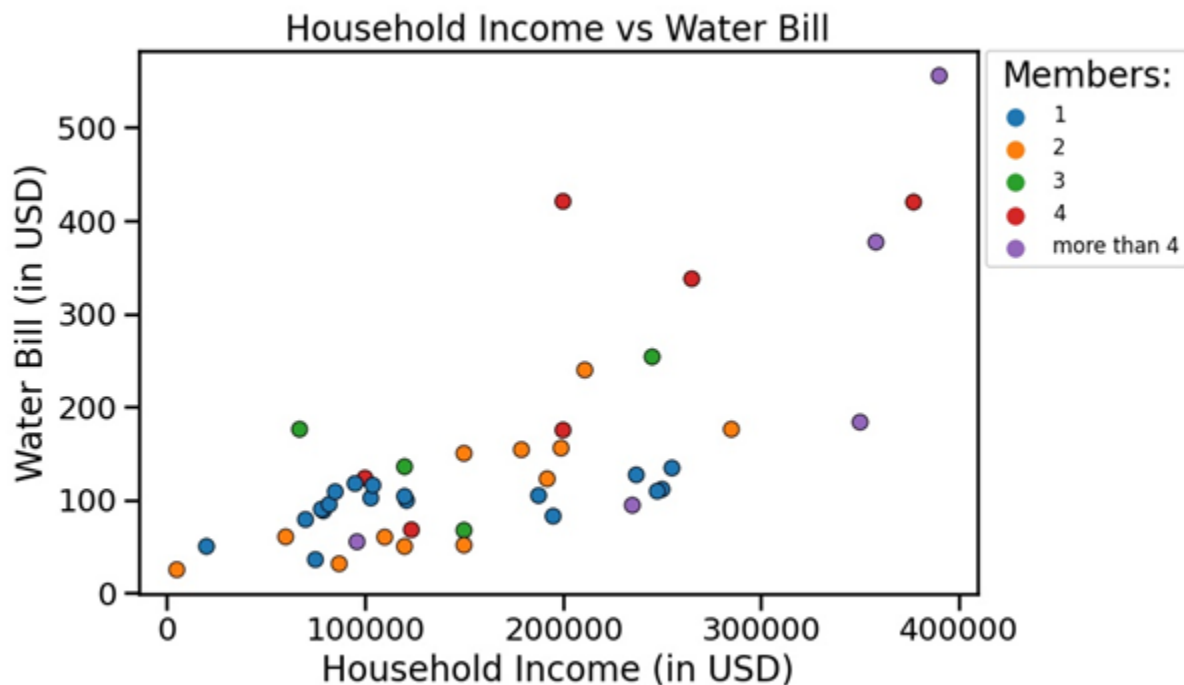
The first scatter plot shows the analysis of the relationship between the time spent in the shower and water bill of the respondents.



```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import numpy as np
data = pd.read_csv('scat1.csv')
print(data)
time = data['time']
water_bill = data['water_bill']
HH = data['HH']
m, b = np.polyfit(time, water_bill, 1)
sns.set_context("talk", font_scale=1.1)
plt.figure(figsize=(10,6))
sns.scatterplot(x="time", y="water_bill", hue="HH", data=data, edgecolor='k')
plt.plot(time, m*time+b, color='red')
plt.legend(bbox_to_anchor=(1.01, 1), borderaxespad=0, prop={'size': 12}, title='Members:')
```

```
plt.savefig("place_legend_outside_plot_Seaborn_scatterplot.png", format='png', dpi=100)
plt.xlabel("time")
plt.ylabel("water bill")
plt.title('Shower Time vs Water Bill')
plt.xlabel('Shower Time (in minutes)')
plt.ylabel('Water Bill (in USD)')
plt.tight_layout()
plt.show()
```

The second scatter plot shows the analysis of the relationship between the household income and water bill of the respondents.



```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
data = pd.read_csv('scat2.csv')
print(data)
income = data['income']
water_bill = data['water_bill']
HH = data['HH']
sns.set_context("talk", font_scale=1.1)
plt.figure(figsize=(10,6))
sns.scatterplot(x="income", y="water_bill", hue="HH", data=data, edgecolor='k')
plt.legend(bbox_to_anchor=(1.01, 1), borderaxespad=0, prop={'size': 12}, title='Members:')
plt.savefig("place_legend_outside_plot_Seaborn_scatterplot.png", format='png', dpi=100)
plt.xlabel("income")
plt.ylabel("water bill")
```

```
plt.title('Household Income vs Water Bill')
plt.xlabel('Household Income (in USD)')
plt.ylabel('Water Bill (in USD)')
plt.tight_layout()
plt.show()
```

MESSAGE

- Shower is the most affected water usage during summer with 60%, followed by clothes washer with 16%, then faucet with 14%, and toilet and “others” comprising the other 10%.
- 50% of the respondents prefer to take their showers in the afternoon during summer.
- Younger people prefer to take their showers in the morning and afternoon while older people prefer at night.
- The average time spent in one shower during summer is around 27 to 32 minutes.
- 52% of the respondents say that on average they take 3 showers in a day during summer.
- 88% of the respondents say that the high temperature during summer “very likely” affects their shower behavior.
- Time spent in the shower and the water bill has a positive linear relationship. The longer time spent in the shower, the higher the water bill.
- If people want to save water or lower their water bills, I would recommend them to reduce their time spent in the shower and preferably take showers during night since it is not so hot during that time and it will likely reduce their shower time.

=====

SECTION 3

FOOD / DRINKS

Analysis by Dennise Arcilla

MOTIVATION

During summer, most people like to eat or drink cold food or beverages. I am curious on how much people are willing to spend in a day to buy cold beverages. I would like to know what is the most popular food / drink for most people during summer. I also would like to confirm my hypothesis that there is an increase in consumption of cold beverages during summer and how much is the average consumption of cold beverage in terms of milliliters a person consumes in a day during the summer.

METHOD

In order to validate my assumption, I created a survey via Google Forms. Aside from the demographic questions our team has, here are the 6 food/beverage-specific questions I made:

1. How likely are you to eat/drink at least one cold food/drink a day during summer?

Choice 1: Very unlikely

Choice 2: Unlikely

Choice 3: Neutral

Choice 4: Likely

Choice 5: Very likely

2. How much do you spend on cold beverages in a day during summer? (Answer in dollars)

3. Does your cold beverage consumption increase or decrease during summer?

Choice 1: Increase

Choice 2: Decrease

4. On average, what is the change in the increase or decrease of your cold beverage during summer?

Choice 1: 1 less beverage than usual

Choice 2: 2 less beverage than usual

Choice 3: 1 more beverage than usual

Choice 4: 2 more beverage than usual

5. How many milliliters of cold beverage do you consume in a day during summer? 1 bottle of water is equal to 500mL. (Answer in mL)

6. Which of the following is your go-to food/drink to suppress the heat?

Choice 1: Ice cream / Frozen yogurt

Choice 2: Shakes /Fruit Juices

Choice 3: Chilled fruits

Choice 4: Iced cold water

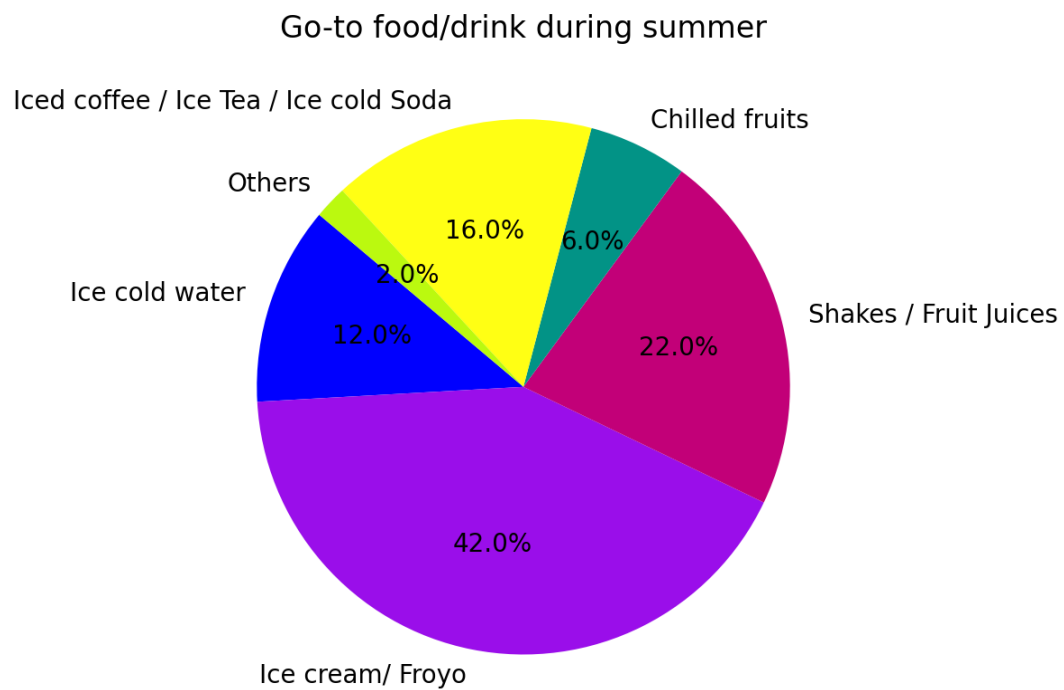
Choice 5: Iced coffee / Iced tea / Ice cold soda

Choice 6: Others

MECHANICS

I used Python in order to generate graphs like Pie chart, Histogram, Barplot, Scatterplot and Boxplot. Using this method, it was easier for me to analyze the data I gathered.

PIE CHART



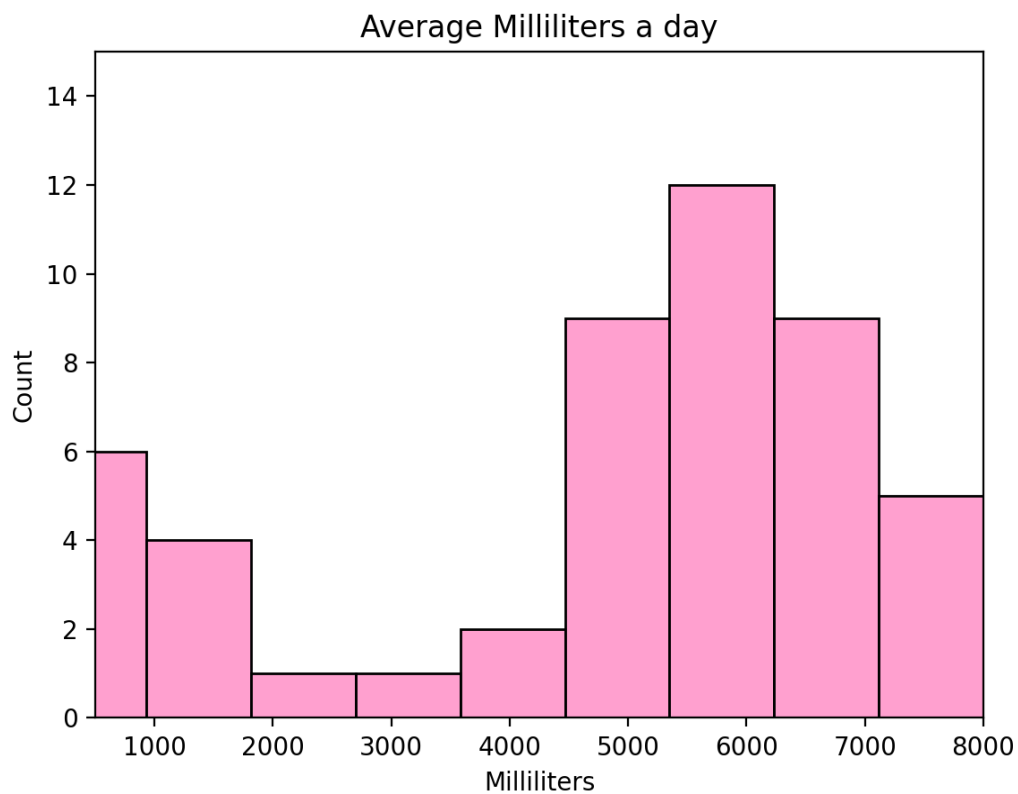
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('piechart.csv')

product_data = df["GoTo"]
Res_data = df["Count"]
colors = ["#0000FF", "#9A0EEA", "#C20078", "#029386", "#FFFF14", "#BBF90F"]
plt.title("Go-to food/drink during summer")
plt.pie(Res_data, labels=product_data, colors=colors,
autopct='%1.1f%%', shadow=False, startangle=140)

plt.show()
```


HISTOGRAM

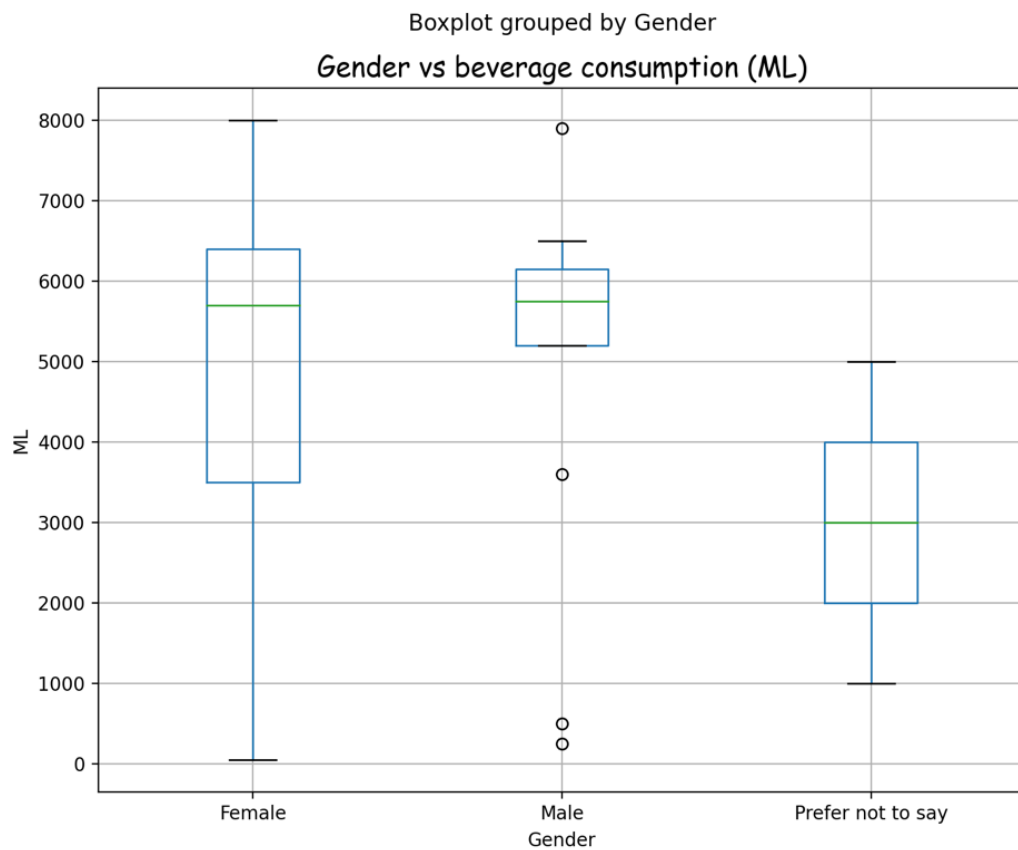


```
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns

values = pd.read_csv('histogram.csv')
print(values)
sns.histplot(data=values, x='ML', color="#FF81C0", bins=9)
plt.ylabel('Count')
plt.xlabel('Milliliters')
plt.title("Average Milliliters a day")
plt.axis([500, 8000, 0, 15])

plt.show()
```

BOX PLOT



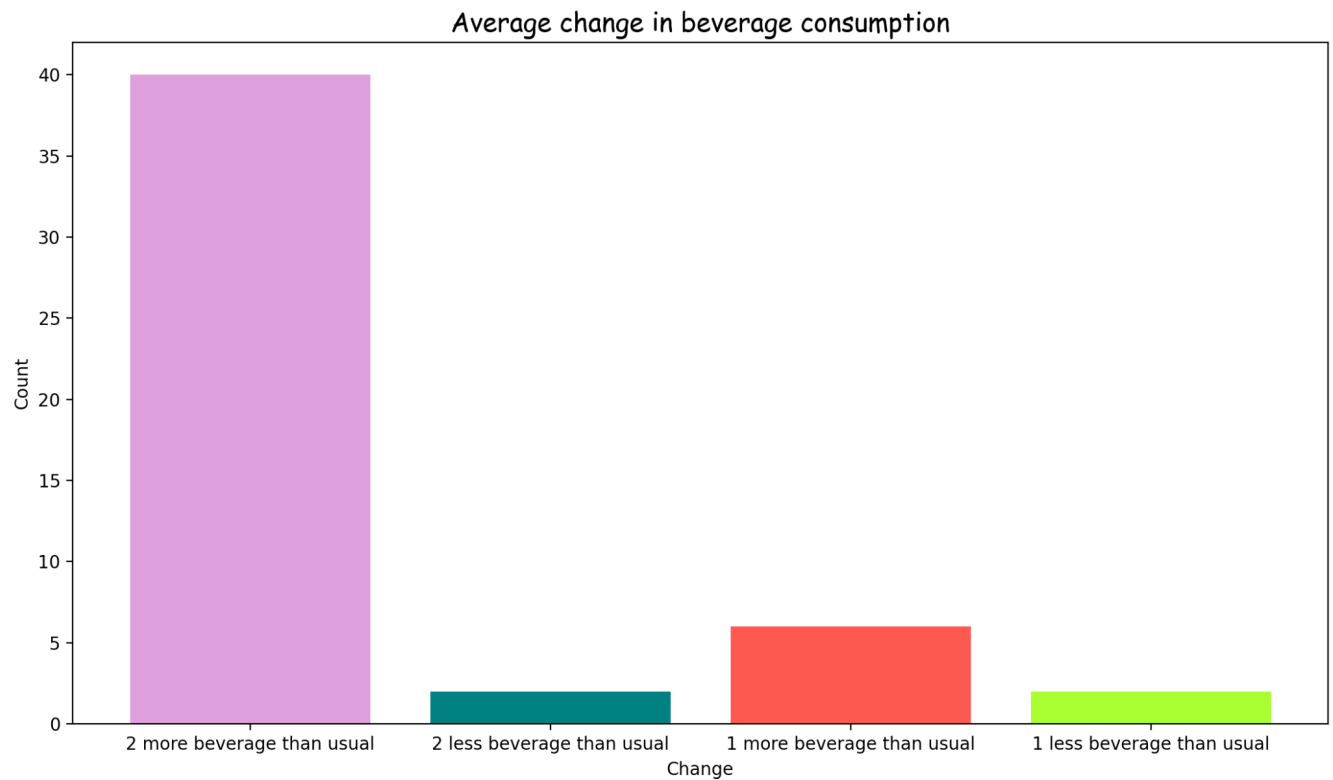
```
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use('default')

box = pd.read_csv('boxplot.csv', sep=',', na_values='.')
print(box.head(3))

boxes = box.boxplot(column=['ML'], by='Gender', figsize=(5,7), grid=True)
plt.title("Gender vs beverage consumption (ML)", fontdict={'fontname':'Comic Sans MS' , 'fontsize': 15})
plt.ylabel('ML')

plt.show()
```

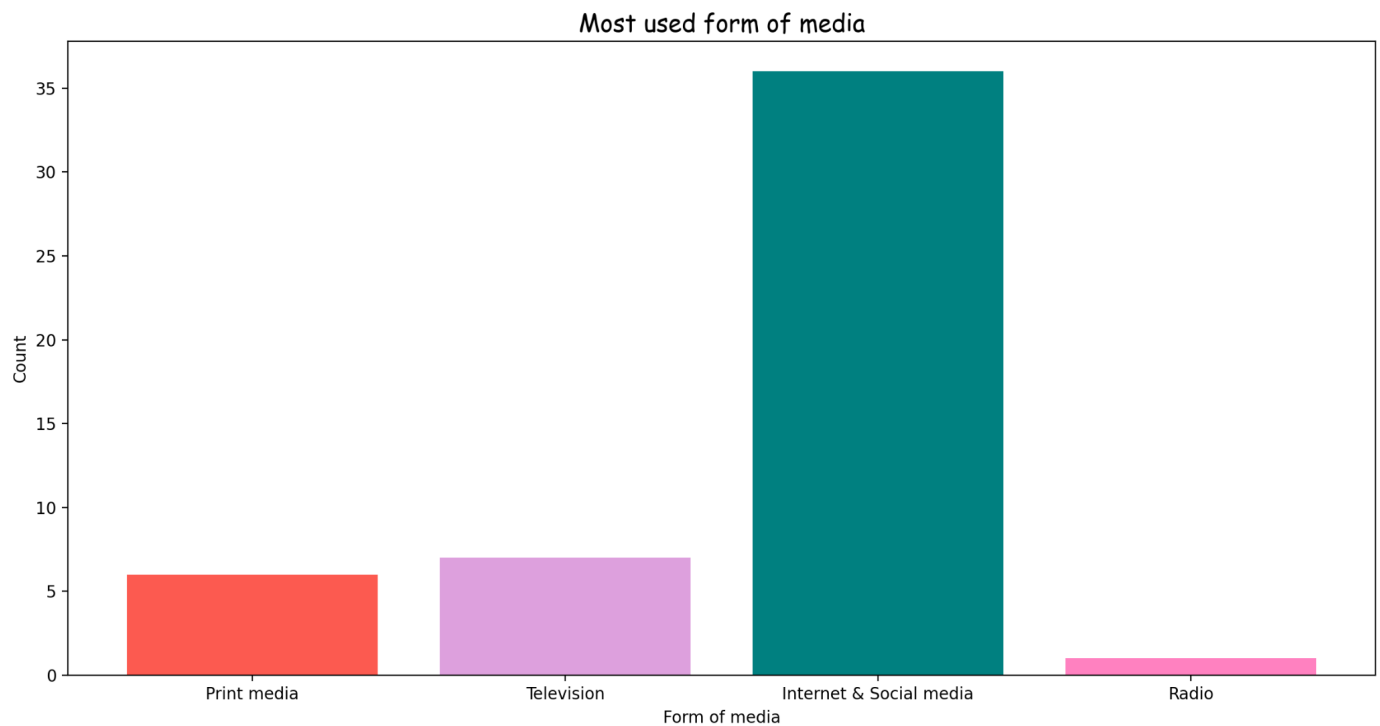
BAR PLOT



```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('barplot.csv')
df = pd.DataFrame(data)
print(df.head())
X = list(df.iloc[:, 0])
Y = list(df.iloc[:, 1])
plt.figure(figsize=(10,7))
colors = ['#DDA0DD', '#008080', '#FC5A50', '#AAFF32']
plt.bar(X, Y, color=colors)
plt.title('Average change in beverage consumption', fontdict={'fontname':'Comic Sans MS', 'fontsize': 15})
plt.xlabel("Change")
plt.ylabel("Count")
plt.show()
```

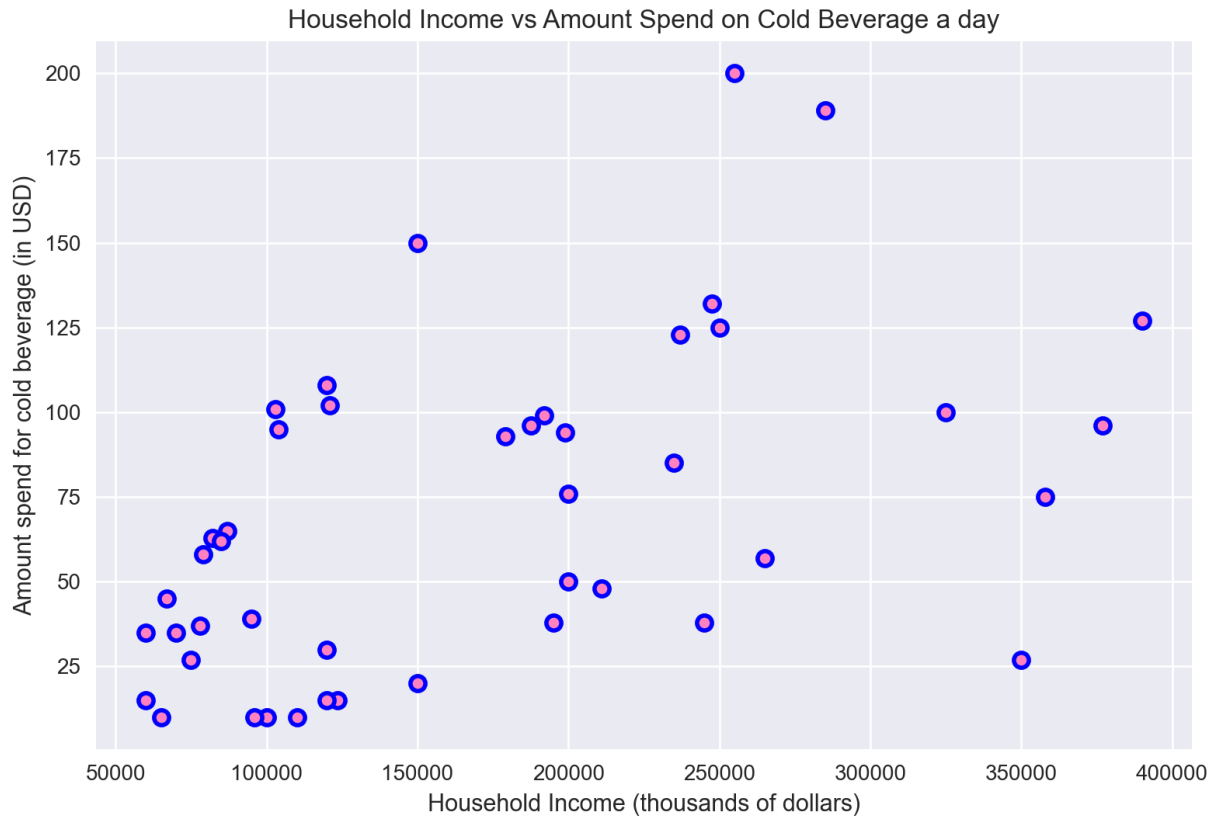
Demographic barplot which has a connection to my analysis stated in the message/conclusion part.



```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('media.csv')
df = pd.DataFrame(data)
print(df.head())
X = list(df.iloc[:, 0])
Y = list(df.iloc[:, 1])
plt.figure(figsize=(20, 7))
colors = ['#FC5A50', '#DDA0DD', '#008080', '#FF81C0']
plt.bar(X, Y, color=colors)
plt.title('Most used form of media', fontdict={'fontname': 'Comic Sans MS', 'fontsize': 15})
plt.xlabel("Form of media")
plt.ylabel("Count")
plt.show()
```

SCATTER PLOT



```
import pandas as pd

import matplotlib.pyplot as plt

plt.style.use('seaborn')

data = pd.read_csv('incspend.csv')

print(data)

Income = data['Income']

Spend = data['Spend']

plt.scatter(Income, Spend, s=48, facecolor='#FF81C0', edgecolor='#0000FF', linewidth=2)

plt.title('Household Income vs Amount Spend on Cold Beverage a day')

plt.xlabel('Household Income (thousands of dollars)')

plt.ylabel('Amount spend for cold beverage (in USD)')

plt.tight_layout()

plt.show()
```

MESSAGE

Based on 50 responses to our questionnaire, here are the primary findings:

- ** Majority of our respondents are Female (68%) and 28% are male, 4% prefer not to say.
- ** 78% are aged between 18-34 years old
- ** Most of them answered that the media they use is usually Internet or Social Media
- ** There is a 100% increase in beverage consumption during summer
- ** 80% said there is an increase of 2 more beverage than their usual consumption during summer
- ** 42% respondents said Ice cream or frozen yogurt are their go-to food during summer. Second is shakes and fruit juices at 22%
- ** During summer, most people prefer to eat ice cream or drink shakes. We can suggest ice cream / shakes businesses to add more flavors and advertise more through social media since most people use it.

=====

SECTION 4

SUMMER VACATION

Analysis by Mai Thai

MOTIVATION

The summer is often a good time for summer trips. Most people prefer to have a long summer trip. However, I have never known how money influences people's decision making on summer vacation. Therefore, I would like to know the importance of summer vacation and how people make decisions for their summer trips.

METHOD

To get the data for analyzing people's attitude with summer vacation, I created five questions about summer vacation for getting data.

1. During summer break (vacation), the most thing you want to do:

- Camping
- Staying at home
- Signing up for interesting summer program
- Traveling
- None of the above.

2. According to The Weather Channel, the temperatures are above average from July through September. Do you think it is a good time for traveling during summer vacation?

- Yes
- No

3. How far will you travel for summer vacation? (Answer in Miles)

4. How important is the cost to your summer vacation?

- Extremely important
- Very important
- Moderately important
- Slightly important
- Not at all important

5. How much will you spend for your summer break? (Answer in US Dollar)

6. How long is your vacation?

5 days

8 days

1 week

2 weeks

None of the above

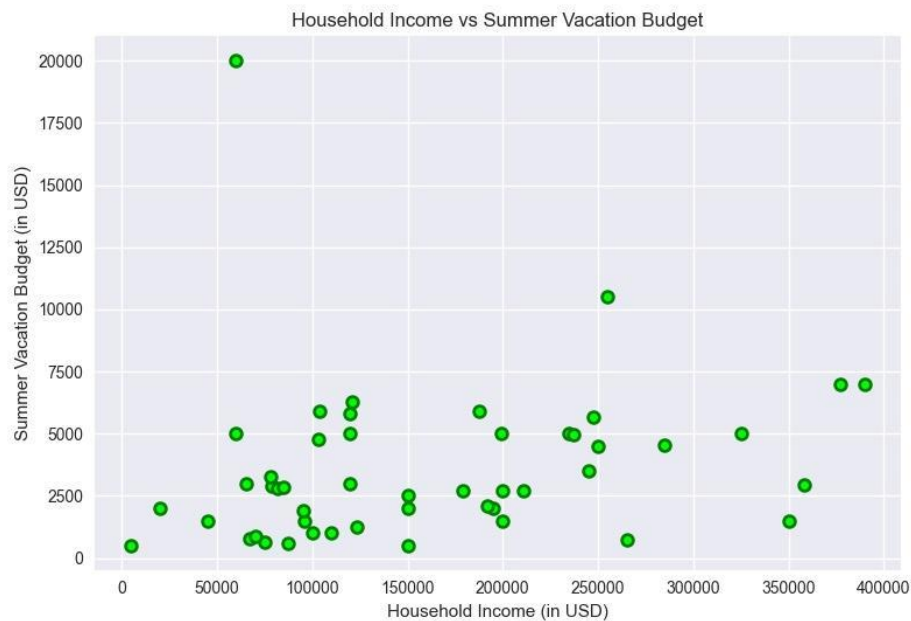
MECHANICS

Based on 50 responses, I used a scatter plot, pie chart, histogram, boxplot and bar chart for analyzing.

SCATTER PLOT

The scatter plot shows the relationship between household income and summer vacation budget.

BEFORE




```

import pandas as pd
import matplotlib.pyplot as plt

plt.style.use('seaborn')

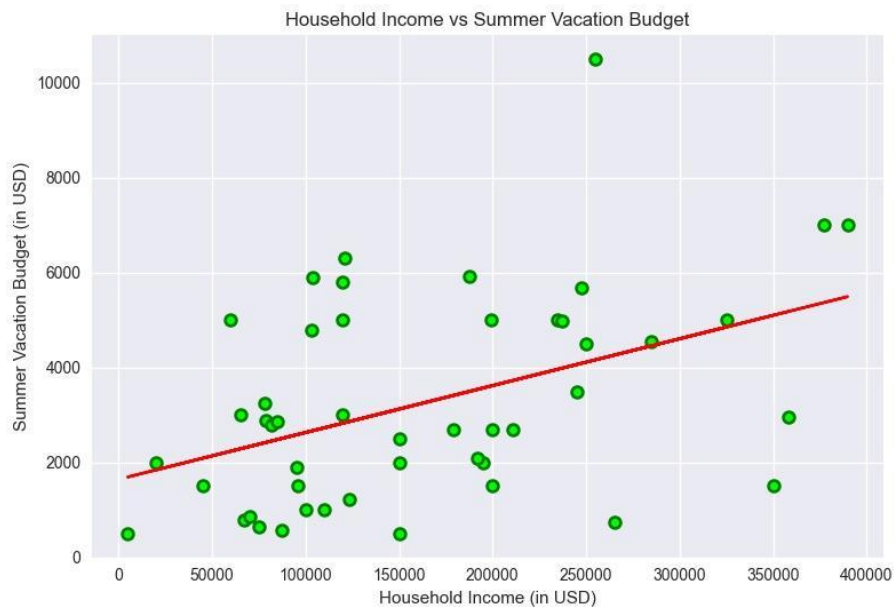
data = pd.read_csv('mai.csv')
print(data)
Income = data['Income']
Spend = data['Spend']

plt.scatter(Income, Spend, s=49, facecolor='lime', edgecolor='green', linewidth=2)
plt.title('Household Income vs Summer Vacation Budget')
plt.xlabel('Household Income (in USD)')
plt.ylabel('Summer Vacation Budget (in USD)')
plt.tight_layout()
plt.show()

```

The scatter plot above shows an outlier but it does not make sense because a household with an income of \$55,000 cannot spend \$200,000 for summer vacation. On the other hand, a household with an income around \$400,000 spends only \$7,000 for summer vacation. In this case, I assume a respondent had misunderstood the question, so he or she gave a wrong answer.

AFTER



```

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

plt.style.use('seaborn')

data = pd.read_csv('mai.csv')
print(data)
Income = data['Income']
Spend = data['Spend']
m, b = np.polyfit(Income, Spend, 1)

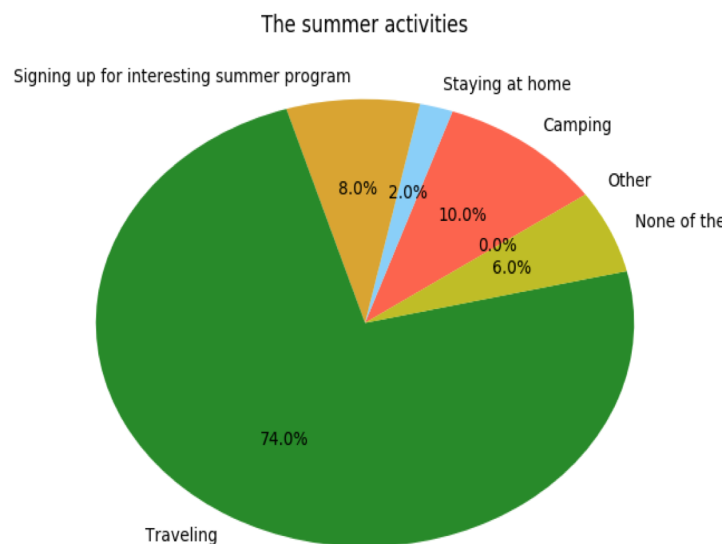
plt.scatter(Income, Spend, s=49, facecolor='lime', edgecolor='green', linewidth=2)
plt.plot(Income, m*Income+b, color='red')
plt.title('Household Income vs Summer Vacation Budget')
plt.xlabel('Household Income (in USD)')
plt.ylabel('Summer Vacation Budget (in USD)')
plt.tight_layout()
plt.show()

```

After deleting the outliers, I had another scatter plot as above. The scatter plot has a strong correlation positive linear between household income and summer vacation budget. It means that household income increases, then the summer vacation budget also increases.

PIE CHART

The chart shows the number of respondents for each summer activity.



```

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

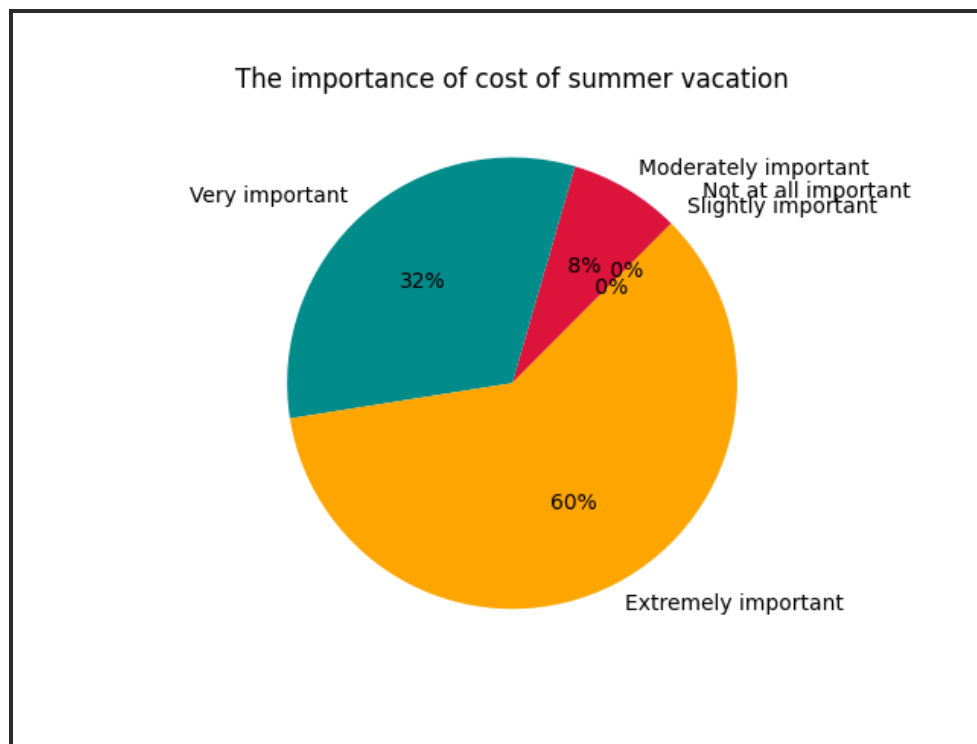
vac_type = pd.read_csv('/Users/maithai/Desktop/NPU/IT501/Project/Part B/Vacation_types.csv')
vac_type_col = vac_type['vacations']
vac_results = vac_type['vacation_results']
colors = ['tomato', 'lightskyblue', 'goldenrod', 'forestgreen', 'y', 'pink']

plt.title("The summer activities")
plt.pie(vac_results, labels = vac_type_col, colors = colors, autopct='%1.1f%%', startangle=35)

plt.show()

```

Another pie chart is about the importance of the cost of summer vacation. It means how respondents are interested in the cost for summer vacation.



```

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

impo = pd.read_csv("important.csv")

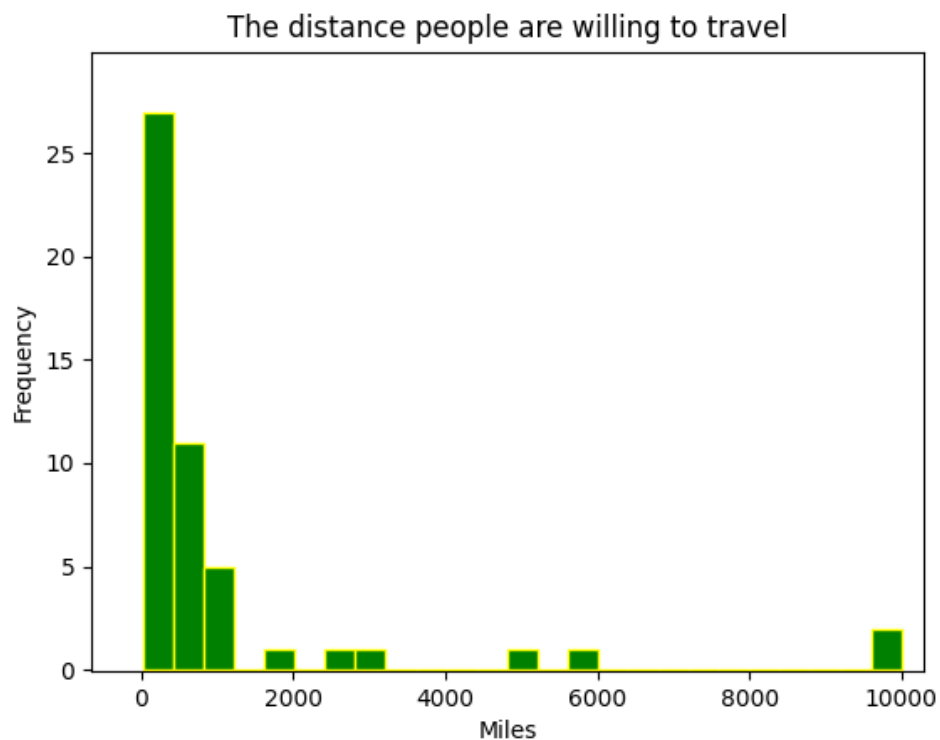
impo_cost = impo['likely']
result = impo["Results"]
color = ['magenta','yellowgreen','crimson','darkcyan','orange']

plt.pie(result,labels=impo_cost,colors=color,autopct='%2.0f%%',startangle=45)
plt.title('The importance of cost of summer vacation')
plt.show()

```

HISTOGRAM

The histogram below indicates the distance that respondents are willing to travel on summer vacation.



```

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

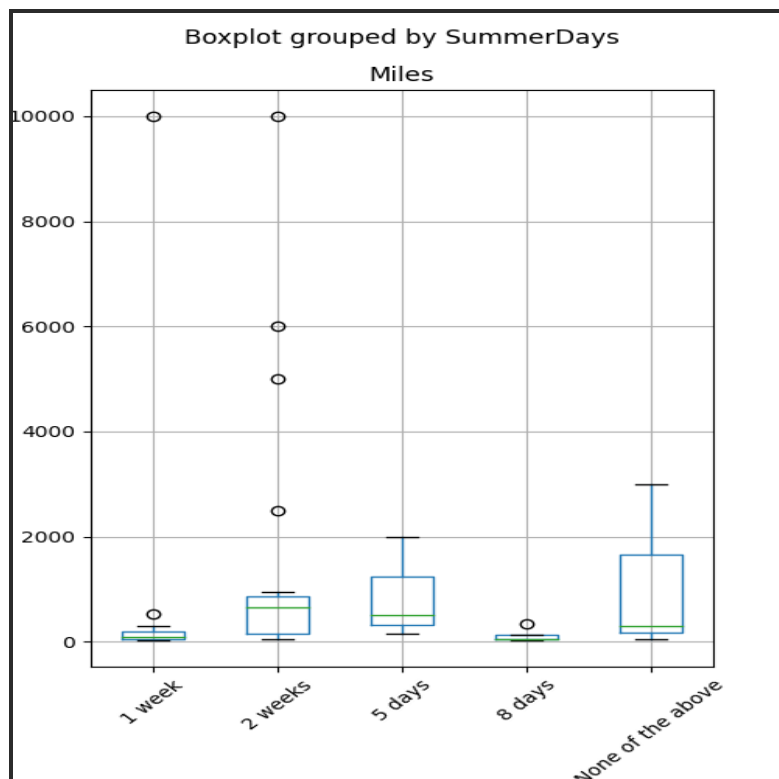
plt.hist(df['summer_miles'], bins=25, color='green', edgecolor="yellow")
plt.xlim([50,11000])
plt.ylim([0,30])
plt.title("The distance people are willing to travel")
plt.xlabel("Miles")
plt.ylabel("Frequency")

plt.show()

```

BOX PLOT

The boxplot below shows the distribution of the distance for traveling with the length of summer vacation



```

import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("Summer Mania.csv", encoding="ISO-8859-1")
miles_days = df.boxplot(column=['Miles'], by='SummerDays', figsize=(5,7),grid=True)

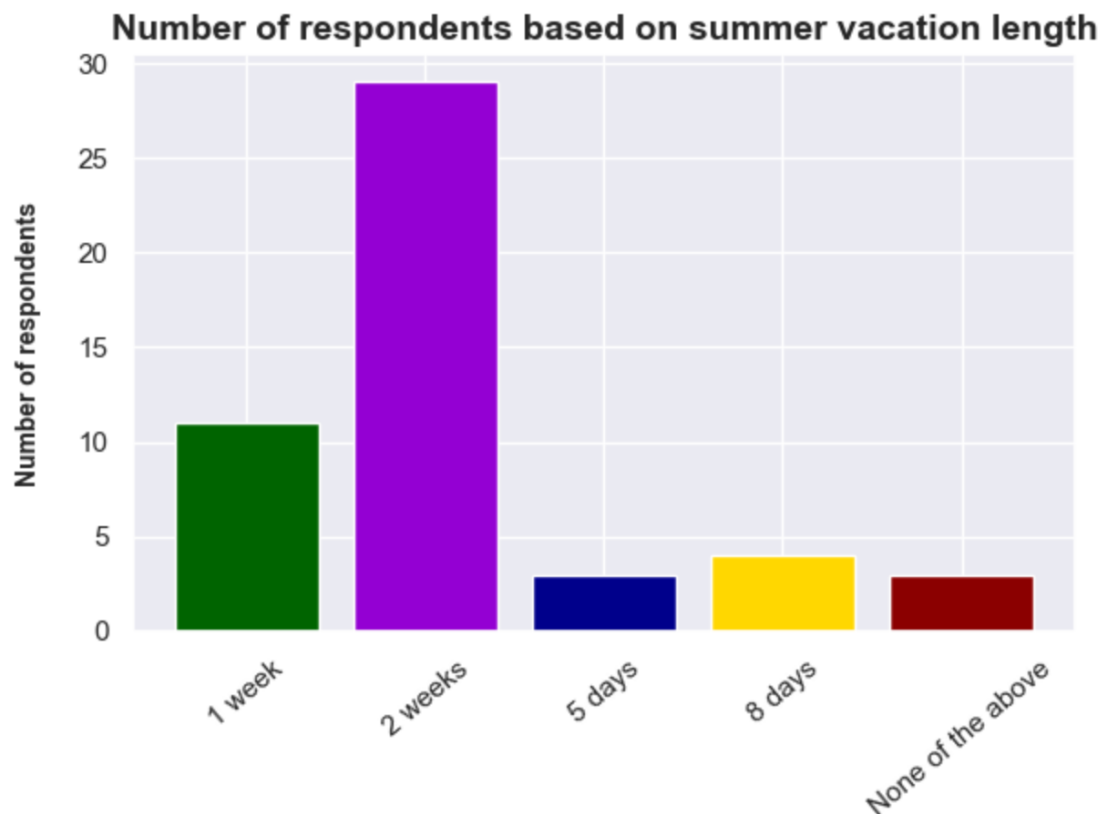
plt.ylabel('Miles')
plt.xticks(rotation=45)

plt.show()

```

BAR PLOT

The bar plot shows the number of respondents with the ideal vacation length.



```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

df1=(df.groupby('SummerDays').size())
colors = ['darkgreen','darkviolet','darkblue','gold','darkred']
df1.plot(kind='bar', figsize=(5,7), rot=45, color = colors)

plt.xlabel("The length of summer vacation",labelpad=14)
plt.ylabel("Number of respondents",labelpad=14)
plt.title("Number of respondents based on summer vacation length")

plt.show()
```

MESSAGE

- The household income and summer vacation budget have a strong correlation positive linear between. It means that household income increases, then the summer vacation budget also increases.
- 74% of people chose traveling as their most summer activity that they want to do in summer vacation. 8% for signing up for an interesting summer program, 2% for staying at home, 10% for camping, and 6% for none of the above. I assume 6% of people want to do other things that are not in the choices.
- 60% chose the cost as their most concern of summer activities. 32% of people think that the cost is somewhat important, and 8% think it is kind of important.
- 54% of people preferred a short distance trip within 400 miles, 22% within 700 miles, 8% within 1,200 miles, but still 6% are willing to travel within 10,000 miles.

- People who prefer 2 weeks for summer vacation are willing to travel further than the others. The distance trip for 1 week, 8 days, and none of the above is not high.
- 58% of people chose 2 weeks as the perfect length for summer vacation. 22% prefer 1 week for summer vacation, 6% for 5 days, 8% for 8 days, and 6% for no choice.
- If the cost is important to people's decision about summer vacation, then I suggest people choose vacations that fit their budget as well as their household income.
- However, there are still some families who want to have a perfect summer trip but their budget does not allow it, so airlines should offer discounts for long distance flights. With discounts, low-income households will have options to have a perfect summer trip. The application of discount programs not only benefits low-income households, but also helps airlines gain more customers in the summer and helps to improve the image and commercial value of the airline.

=====

SECTION 5

SPORTS

Analysis by Nur Farabi

INTRODUCTION

I am Nur A Farabi from the Summer Mania group. As our 2nd project of course IT501 we are doing an analysis business in summer about various products and services by collecting data from surveys and making suitable graphs using python to understand the market.

MOTIVATION

Some businesses depend on the season. Our goal in this project is to decide which products or services people use most in summer. I choose to collect data on Sports on summer and analyze that.

METHOD

Google forms Questionnaire:

The questions I asked about sports in summer are as below:

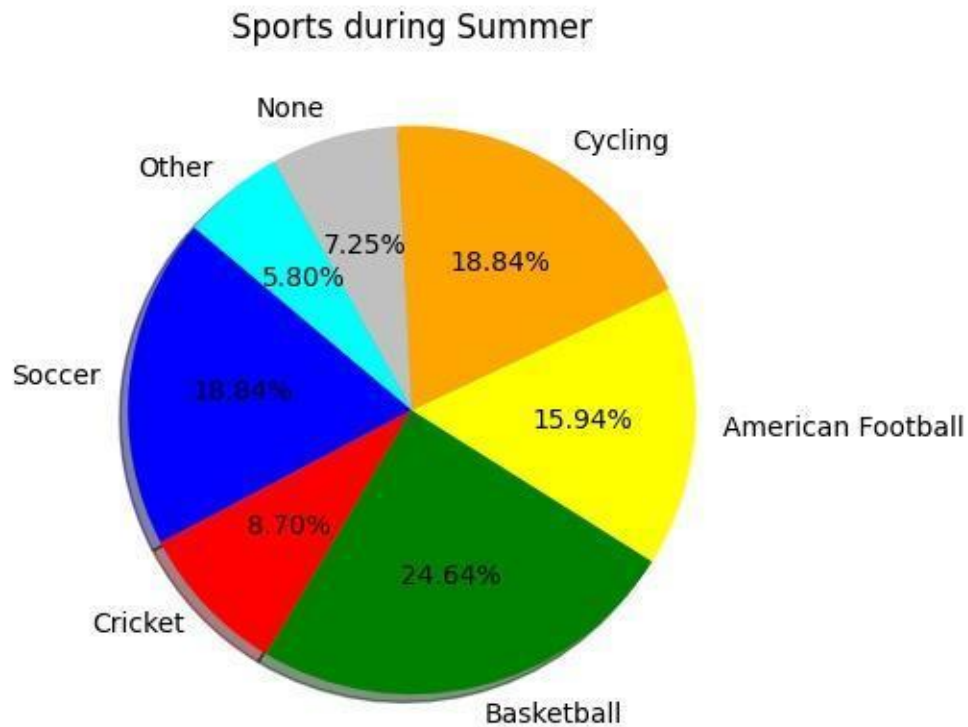
1. What is your weekly estimated time spent outside for sports during summer? (Answer in minutes)
2. How likely do you want to participate white water rafting in summer?
 - a. Extremely likely
 - b. Likely
 - c. Not sure
 - d. Don't like
 - e. Extremely don't like
3. How likely do you want to participate in kayaking in summer?
 - a. Extremely likely

- b. Likely
 - c. Not sure
 - d. Don't like
 - e. Extremely don't like
4. What sports do you enjoy during summer?
- a. Soccer
 - b. Cricket
 - c. Basketball
 - d. American Football
 - e. Cycling
 - f. None
 - g. Swimming
 - h. Ice Hockey
 - i. Tennes
 - j. Running, volleyball , swimming
5. How much do you spend in a week for sports during summer? (Answer in dollars)

MECHANICS

I am using Python programming and some libraries to analyze our data. Depending on our data I am using scatter plot, bar plot and pie chart to describe our data.

PIE CHART



```
import pandas as pd
import matplotlib.pyplot as plt

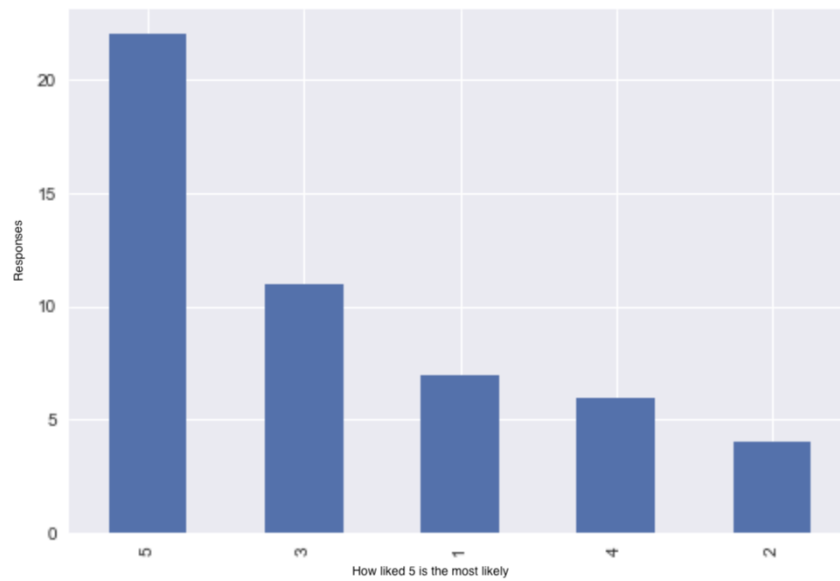
df = pd.read_csv('nurpiechart.csv')

product_data = df["Sports"]
Res_data = df["Response"]
colors = ["blue", "red", "green", "yellow", "orange", "silver", "aqua"]

plt.pie(Res_data, labels=product_data, colors=colors,
        autopct='%1.2f%%', shadow=True, startangle=140)
plt.title("Sports during Summer")
plt.show()
```

BAR PLOT

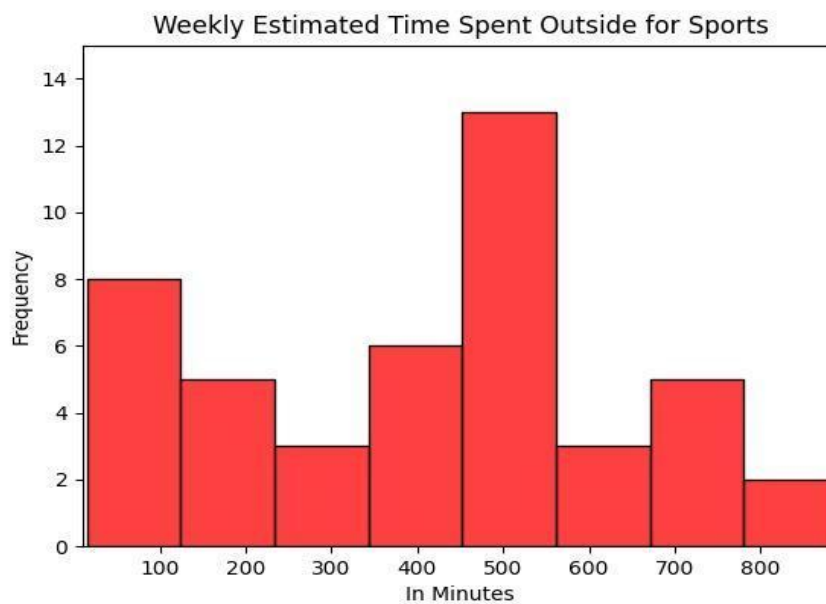
Water Activity



```
pd.value_counts(survey.White_water_rafting).plot.bar()
```

HISTOGRAM

Weekly Time Spend on summer



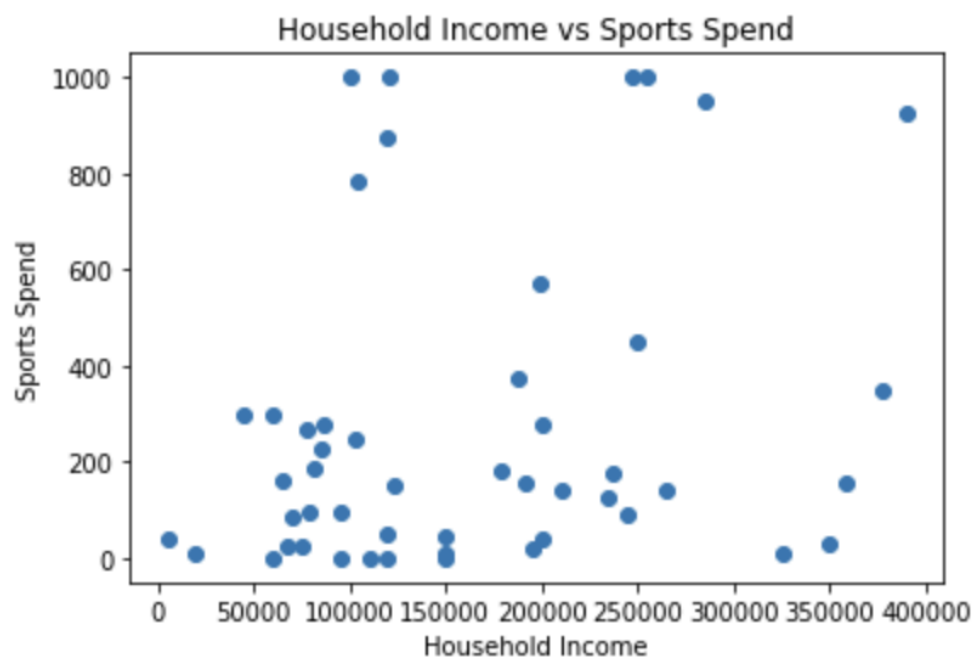
```

import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
values = pd.read_csv('nurhistogram.csv')
print(values)
sns.histplot(data=values, x='Minutes', color="red", bins=8)
plt.ylabel('Frequency')
plt.xlabel('In Minutes')
plt.title("Weekly Estimated Time Spent Outside for Sports")
plt.axis([10, 890, 0, 15])
plt.show()

```

SCATTER PLOT

Income vs spend



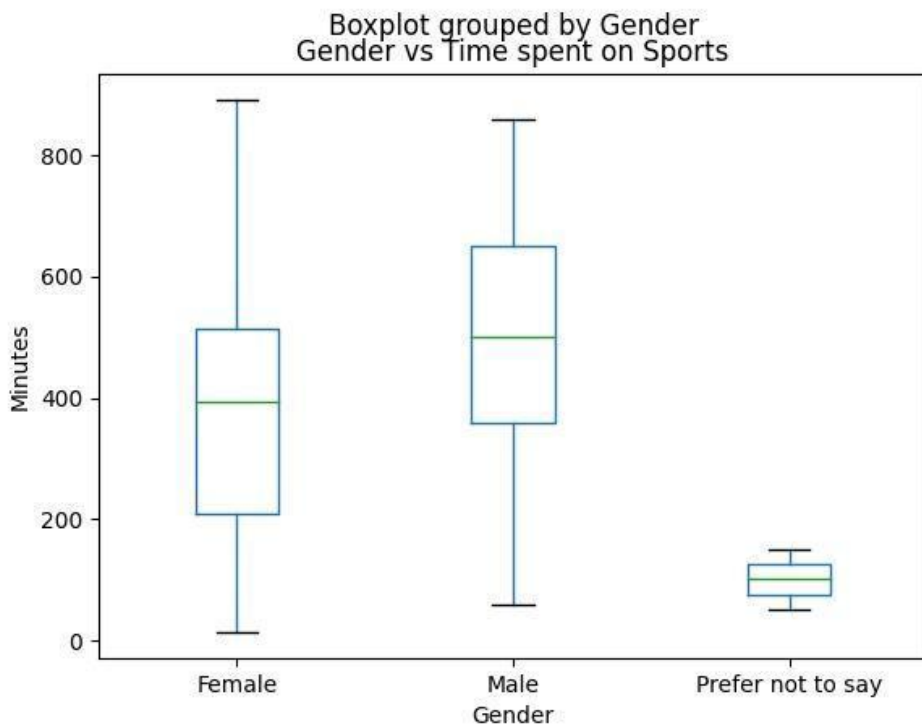
```

# X axis parameter shows sports spend:
yaxis =
np.array([1000,10,0,300,40,10,0,150,10,160,45,0,300,0,0,25,125,280,30,25,50,20,40,96,180,155,
450,1000,950,175,1000,376,999,
247,876,785,570,923,350,87,280,269,185,228,98,140,92,140,155])
# Y axis parameter shows the income:
xaxis =
np.array([100000,20000,96000,60000,5000,325000,60000,123456,150000,65000,150000,15000
0,
45000,110000,120000,75000,235000,87000,350000,67000,120000,195000,200000,79000,17900
0,192000,250000,255000,285000,237000,
247585,187654,121000,102957,120000,104000,199000,389958,377000,70000,199999,78000,82
000,85000,95000,211000,245000,265000,358000])
color_arr = np.random.rand(26)
plt.title('Household Income vs Sports Spend')
plt.ylabel('Sports Spend')
plt.xlabel('Household Income')
plt.scatter(xaxis , yaxis)

```

BOX PLOT

Activities outside depend on gender



```
import pandas as pd
import matplotlib.pyplot as plt
plt.style.use('default')

box = pd.read_csv('nurboxplot.csv', sep=',', na_values='.')
print(box.head(3))

boxes = box.boxplot(column=['Minutes'], by='Gender', grid=False)
plt.title("Gender vs Time spent on Sports")
plt.ylabel('Minutes')

plt.show()
```

MESSAGE

Based on 50 responses, we can see that:

- 60% are in the 25 - 34 aged group.
 - 68% are female.
 - People like to spend money on tools or services
 - Most people like to do some sort of outside activity.
 - Most of the people want to spend \$150 to \$300 for sports tools or services.
 - Most of the people like kinds of water Sports like Kayaking and white water rafting
 - Most of the people like basketball
 - Based on my survey and my group survey, I recommends on sports that any travel sports package will fit on \$150 to \$300.
 - Though both genders like outside sports and they both spend significant time on outside sports, Sports clothing should focus on women's clothing as they are the frequent buyers.
-

CONCLUSION

Our team conducted research about how the summer season affects one's activities and purchasing decisions.

The results of our research showed that our hypotheses were confirmed and the participant's preferences related to food/drinks, clothing, shower frequency, choice of vacation and summer sports change with sweltering heat. Overall, we saw that the time spent in the shower and the water bill has a positive linear relationship. The longer time spent in the shower, the higher the water bill. When it comes to vacation, the household income and vacation budget have a strong positive correlation. It means that household income increases, then the summer vacation budget also increases. People also spend moderate time on sports on a weekly basis, and spend at least \$150 to \$300 weekly for sports tools and service. There is also a significant increase in beverage consumption during summer, most of the respondents said the increase was at least 2 more beverages than their usual consumption. Female and younger people are also likely to shop for summer clothes every year and they spend at least \$2500-\$3000 to buy new clothing.

Overall, our team was able to make inferences based on people's activities and behavior based on the collected data and assessment of results and our initial hypothesis was correct.