```
'''
Kyle Tranfaglia
COSC311 - Lab03
Last updated 04/02/24
This program reads in two housing information data sets and does some
statistical analysis on certain attributes
'''
import stats
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt

# Read in data
Bejaia_Region = pd.read_csv('Bejaia_Region.csv')
SidiBel_Abbes_Region =pd.read_csv('Sidi-Bel_Abbes_Region.csv')

# Task 1: Mean values of four attributes of the "Bejaia Region Dataset"
for fire and no fire

attributes01 = ["Temperature", "RH", "Ws", "Rain"]  # Attribute list

# Get fire mean for each attribute
temp_fire_mean =
stats.mean(Bejaia_Region['Temperature'][Bejaia_Region['Classes'] == 'fire
'].values)
rh_fire_mean = stats.mean(Bejaia_Region['RH'][Bejaia_Region['Classes'] ==
'fire   '].values)
ws_fire_mean = stats.mean(Bejaia_Region['Ws'][Bejaia_Region['Classes'] ==
'fire   '].values)
rain_fire_mean = stats.mean(Bejaia_Region['Rain'][Bejaia_Region['Classes']
== 'fire   '].values)

# Get not fire mean for each attribute
temp_nofire_mean =
stats.mean(Bejaia_Region['Temperature'][Bejaia_Region['Classes'] == 'not
fire   '].values)
rh_nofire_mean = stats.mean(Bejaia_Region['RH'][Bejaia_Region['Classes']
== 'not fire   '].values)
ws_nofire_mean = stats.mean(Bejaia_Region['Ws'][Bejaia_Region['Classes']
== 'not fire   '].values)
```

```
rain_nofire_mean =
stats.mean(Bejaia_Region['Rain'][Bejaia_Region['Classes'] == 'not fire
'].values)

# List of fire means and no fire means
mean_fire = [temp_fire_mean, rh_fire_mean, ws_fire_mean, rain_fire_mean ]
mean_nofire = [temp_nofire_mean, rh_nofire_mean, ws_nofire_mean,
rain_nofire_mean ]

# Print both fire and not fire mean for each attribute
for i in range(len(attributes01)):
    print(attributes01[i] + "\nFire mean: " + str(round(mean_fire[i],
ndigits=6)) + "\nNo Fire mean: " + str(round(mean_nofire[i], ndigits=6)))

# Set up for plot
bar_width = 0.35
index = range(len(attributes01))

# Set up and plot bar chart with both fire and no fire means
plt.figure(figsize=(10, 6))
plt.bar(index, mean_fire, bar_width, label='Fire', color='red', alpha=0.7)
plt.bar([i + bar_width for i in index], mean_nofire, bar_width, label='Not
Fire', color='blue', alpha=0.7)
plt.xlabel('Attributes')
plt.ylabel('Mean Values')
plt.title('Mean Values of Attributes for Fire and Not Fire')
plt.xticks([i + bar_width / 2 for i in index], attributes01)
plt.legend()
plt.tight_layout()
plt.show()

# Alternative display: two individual bar plots
# plt.figure(figsize=(10, 6))
# plt.bar(range(len(attributes)), mean_fire, width=0.5)
# plt.title('Mean Values for Fire')
# plt.xlabel('Attributes')
# plt.ylabel('Mean Value')
# plt.xticks(ticks=range(len(attributes)), labels=attributes)
# plt.show()
```

```
# plt.figure(figsize=(10, 6))
# plt.bar(range(len(attributes)), mean_nofire, width=0.5)
# plt.title('Mean Values for No Fire')
# plt.xlabel('Attributes')
# plt.ylabel('Mean Value')
# plt.xticks(ticks=range(len(attributes)), labels=attributes)
# plt.show()

'''
Observations: Mean temperature during fire is higher than no fire, mean rh
during fire is lower than no fire, mean ws during fire is slightly lower
than no fire,
and mean rain during fire is significantly lower than no fire. Therefore,
we may infer that fire incidents might be associated with higher
temperatures,
occur more frequently during periods of lower relative humidity, have
minor correlation to wind peed yet may be associated with higher wind
speeds, and
occur more frequently during periods of lower rainfall.
'''

# Task 2: Using the "Sidi-Bel Abbes Region Dataset," calculate and show
the median values of four attributes

attributes02 = ['FFMC', 'DMC', 'DC', 'ISI']  # Attribute list

# Get median for each attribute
FFMC_median = stats.median(SidiBel_Abbes_Region['FFMC'].values)
DMC_median = stats.median(SidiBel_Abbes_Region['DMC'].values)
DC_fire_median = stats.median(SidiBel_Abbes_Region['DC'].values)
ISI_fire_median = stats.median(SidiBel_Abbes_Region['ISI'].values)

medians = [FFMC_median, DMC_median, DC_fire_median, ISI_fire_median]  #
List for all medians

print()  # Spacing for ouput

# Print median for each attribute
for i in range(len(attributes02)):
```

```
    print(attributes02[i] + "\nMedian: " + str(round(medians[i],
ndigits=6)))

# Set up and plot bar chart for medians
plt.figure(figsize=(10, 6))
plt.bar(attributes02, medians, color='skyblue')
plt.xlabel('Attributes')
plt.ylabel('Median Values')
plt.title('Median Values of Attributes')
plt.tight_layout()
plt.show()

# Task 3: Using the "Bejaia Region Dataset", calculate and show the
25-percent, 60-percent, and 75-percent quantiles of four attributes

attributes03 = ["Temperature", "RH", "Ws", "Rain"]  # Attribute list
quantiles = [0.25, 0.60, 0.75]  # Quantile list
temp_quantiles, rh_quantiles, ws_quantiles, rain_quantiles = [], [], [],
[]  # Attribute quantile lists

# Get Quantiles
for i in quantiles:

temp_quantiles.append(stats.quantile(Bejaia_Region['Temperature'].values,
i))
    rh_quantiles.append(stats.quantile(Bejaia_Region['RH'].values, i))
    ws_quantiles.append(stats.quantile(Bejaia_Region['Ws'].values, i))
    rain_quantiles.append(stats.quantile(Bejaia_Region['Rain'].values, i))

attribute_quantiles = [temp_quantiles, rh_quantiles, ws_quantiles,
rain_quantiles]  # List for all medians

print()  # Spacing for ouput

# Print median for each attribute
for i in range(len(attributes03)):
    print(attributes03[i] + "\nPercent Quantiles (0.25, 0.60, 0.75): ",
end="")
    for j in range(len(quantiles)):
        print(attribute_quantiles[i][j], "", end="")
```

```
    print()  # Move to the next line for the next attribute

# Positioning values for plotting
bar_width = 0.15
index = np.arange(len(attributes03))

plt.figure(figsize=(10, 6)) # Set figure size

# Set up bar plot of grouped bars for each attribute
for i in range(len(quantiles)):
    plt.bar(index + i * bar_width, [attribute_quantiles[j][i] for j in
range(len(attributes03))], bar_width, label=f'{quantiles[i]*100}%')

plt.xlabel('Attributes')
plt.ylabel('Quantile Values')
plt.title('Quantile Values of Attributes')
plt.xticks(index + bar_width * (len(quantiles) - 1) / 2, attributes03)
plt.legend(title='Quantiles')
plt.tight_layout()
plt.show()

# Task 4: Using the "Sidi-Bel Abbes Region Dataset", calculate and show
the standard deviation values of four attributes

attributes04 = ["Temperature", "Rain", "BUI", "FWI"]  # Attribute list

# Get standard deviation for each attribute
temp_std = stats.std(Bejaia_Region['Temperature'].values)
rain_std = stats.std(Bejaia_Region['Rain'].values)
bui_std = stats.std(Bejaia_Region['BUI'].values)
fwi_std = stats.std(Bejaia_Region['FWI'].values)

stds = [temp_std, rain_std, bui_std, fwi_std]  # List of all standard
deviations

print()  # Spacing for ouput

plt.figure(figsize=(10, 6))  # Set figure size
# Print standard deviation for each attribute
for i in range(len(attributes04)):
```

```
    print(attributes04[i] + "\nStandard Deviation: " + str(round(stds[i],
ndigits=6)))

# Set up and plot bar chart for standard deviations
plt.bar(attributes04, stds, color='violet')
plt.xlabel('Attributes')
plt.ylabel('STD Values')
plt.title('STD Values of Attributes')
plt.tight_layout()
plt.show()

# Task 5: Correlation between two attributes - Using the "Bejaia Region
Dataset", calculate and show the "correlation coefficient"
# between "RH" and the other attributes. Describe, if there is one, which
attribute has strongest positive and negative correlation with "RH"

attributes05 = ["Temperature", "Ws", "Rain", "FFMC", "DMC", "DC", "ISI",
"BUI", "FWI"]  # Attribute list

print()  # Spacing for ouput

# Print correlation coefficient between "RH" and each attribute
for i in attributes05:
    print(i + "\nCorrelation Coefficient: " +
str(round(stats.correlation(Bejaia_Region["RH"].values,
Bejaia_Region[i].values), ndigits=6)))

'''
Using standard application, strong positive range is (0.7 - 1.0), strong
negative range is (-1.0 - -0.7)
Given these ranges, there is no strong positive correlation nor a strong
negative correlation
Strongest postitive correlation: Rain 0.329
Strongest negative correlation: Temperature -0.660
'''

# Task 6: Assume you need to select some attributes or design some new
attributes to distinguish these two classes ("not fire" and "fire")
# as accurate as possible, which attributes you would like to select or
what new attributes you would like to design?
```

```
attributes06 = ["Temperature", "RH", "Ws", "Rain", "FFMC", "DMC", "DC",
"ISI", "BUI", "FWI"]  # Attribute list

print()  # Spacing for ouput

# Get a list of each attributes correlation coefficent to fire and not
fire class
for i in attributes06:
    print(i + "\nFire Correlation Coefficient: " +
str(round(stats.correlation(Bejaia_Region['Classes'] == 'fire',
Bejaia_Region[i]), ndigits=6)))
    print("Not Fire Correlation Coefficient: " +
str(round(stats.correlation(Bejaia_Region['Classes'] == 'not fire',
Bejaia_Region[i]), ndigits=6)))

'''
Correlation Coefficient Results:
Temperature
Fire Correlation Coefficient: 0.101316
Not Fire Correlation Coefficient: -0.241286
RH
Fire Correlation Coefficient: 0.029412
Not Fire Correlation Coefficient: 0.174609
Ws
Fire Correlation Coefficient: 0.097343
Not Fire Correlation Coefficient: 0.091008
Rain
Fire Correlation Coefficient: -0.06466
Not Fire Correlation Coefficient: 0.678363
FFMC
Fire Correlation Coefficient: 0.143801
Not Fire Correlation Coefficient: -0.27639
DMC
Fire Correlation Coefficient: 0.380186
Not Fire Correlation Coefficient: -0.087554
DC
Fire Correlation Coefficient: 0.338891
Not Fire Correlation Coefficient: -0.11344
ISI
```

```
Fire Correlation Coefficient: 0.196957
Not Fire Correlation Coefficient: -0.143959
BUI
Fire Correlation Coefficient: 0.36975
Not Fire Correlation Coefficient: -0.096064
FWI
Fire Correlation Coefficient: 0.289918
Not Fire Correlation Coefficient: -0.110928

Top 3 Strongest correlations to fire: DMC (0.380186), BUI (0.36975), DC
(0.338891)
Top 3 Strongest correlations to not fire: Rain (0.678363), RH (0.174609),
WS (0.091008)

Recall observations from analysis if mean data:
Mean temperature during fire is higher than no fire, mean rh during fire
is lower than no fire, mean ws during fire is slightly lower than no fire,
and mean rain during fire is significantly lower than no fire. Therefore,
we may infer that fire incidents might be associated with higher
temperatures,
occur more frequently during periods of lower relative humidity, have
minor correlation to wind peed yet may be associated with higher wind
speeds, and
occur more frequently during periods of lower rainfall.

Thus, using the following data, the best attributes to select in order to
distinguish between "not fire" and "fire" events with the greatest
accuracy is DMC, BUI, DC,
and Rain. Each of these attributes have fairly high positive correlation
coeffients to a class relative to the other attributes. DMC, BUI, and DC
all have
correlation coefficients over 0.33, which although is not a strong
positive correlation, the combination of these attributes provides insight
into a the likeness
of a fire event. Rain and not fire has the strongest of the correlation
coefficents at 0.68 rounded which is nearly a strong positive correlation.
Therefore, this
relationship indicates that the Rain attribute may be useful in
determining a not fire event. In context of the attributes, DMC, BUI, DC,
and Rain are all useful in
```

```
distinguishing between "not fire" and "fire" events with an element of
accuracy based upon correlation coefficient representations of
relationships, but given previous
analyses with mean, we can also make further insinuations. For instance,
mean temperature during fire is higher than no fire, mean rh during fire
is lower than no fire,
mean ws during fire is slightly lower than no fire, and mean rain during
fire is significantly lower than no fire. With this, we can derive that
fire incidents might
be associated with higher temperatures, occur more frequently during
periods of lower relative humidity, and occur more frequently during
periods of lower rainfall.
The difference in mean for Ws is not notable, although, we may be able to
use the small variation to help identify "not fire" and "fire" events.
Despite the
usefulness of the identified attributes, there were not any very strong
positive or negative relationships nor outstanding differences in mean
besides Rain and
Temperature such that introducing a new attribute that fits this criteria
may be significant in distinguishing between "not fire" and "fire" events.
The new attribute
may be something such as geographic region or terrain type which may give
stronger insight into potential "not fire" and "fire" events. In
conclusion, the most ideal
way to distinguish between "not fire" and "fire" events with the greatest
accuracy is to introduce a new attribute with a larger correlation
coefficient and/or
variation in mean such as geographic region or terrain type, yet combining
the existing correlated attributes including DMC, BUI, DC, Temperature,
RH, and Rain
may provide a good accuracy.
'''
```

Task 1 and Task 6 responses (also in code as comments):

Observations: Mean temperature during fire is higher than no fire, mean rh during fire is lower than no fire, mean ws during fire is slightly lower than no fire,
and mean rain during fire is significantly lower than no fire. Therefore, we may infer that fire incidents might be associated with higher temperatures,

occur more frequently during periods of lower relative humidity, have minor correlation to wind peed yet may be associated with higher wind speeds, and
occur more frequently during periods of lower rainfall.

Correlation Coefficient Results:
Temperature
Fire Correlation Coefficient: 0.101316
Not Fire Correlation Coefficient: -0.241286
RH
Fire Correlation Coefficient: 0.029412
Not Fire Correlation Coefficient: 0.174609
Ws
Fire Correlation Coefficient: 0.097343
Not Fire Correlation Coefficient: 0.091008
Rain
Fire Correlation Coefficient: -0.06466
Not Fire Correlation Coefficient: 0.678363
FFMC
Fire Correlation Coefficient: 0.143801
Not Fire Correlation Coefficient: -0.27639
DMC
Fire Correlation Coefficient: 0.380186
Not Fire Correlation Coefficient: -0.087554
DC
Fire Correlation Coefficient: 0.338891
Not Fire Correlation Coefficient: -0.11344
ISI
Fire Correlation Coefficient: 0.196957
Not Fire Correlation Coefficient: -0.143959
BUI
Fire Correlation Coefficient: 0.36975
Not Fire Correlation Coefficient: -0.096064
FWI
Fire Correlation Coefficient: 0.289918
Not Fire Correlation Coefficient: -0.110928

Top 3 Strongest correlations to fire: DMC (0.380186), BUI (0.36975), DC (0.338891)
Top 3 Strongest correlations to not fire: Rain (0.678363), RH (0.174609), WS (0.091008)

Recall observations from analysis if mean data:

Mean temperature during fire is higher than no fire, mean rh during fire is lower than no fire, mean ws during fire is slightly lower than no fire,
and mean rain during fire is significantly lower than no fire. Therefore, we may infer that fire incidents might be associated with higher temperatures,
occur more frequently during periods of lower relative humidity, have minor correlation to wind peed yet may be associated with higher wind speeds, and
occur more frequently during periods of lower rainfall.

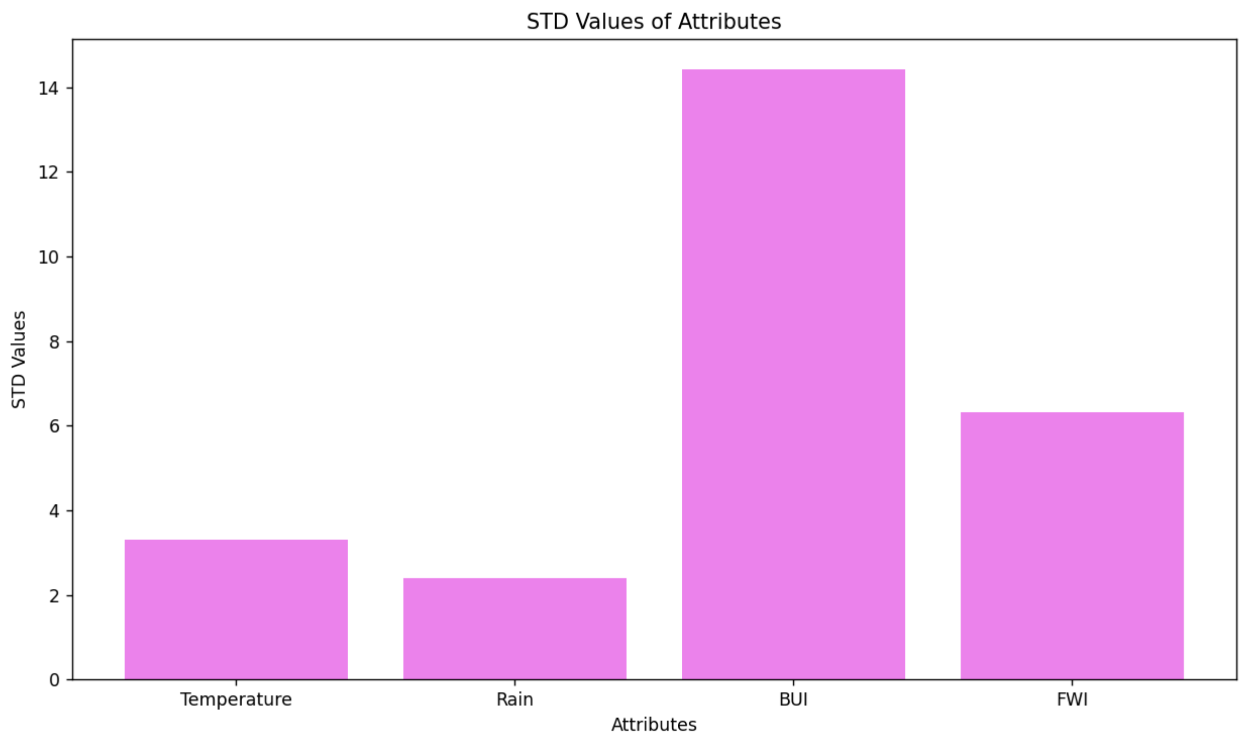
Thus, using the following data, the best attributes to select in order to distinguish between "not fire" and "fire" events with the greatest accuracy is DMC, BUI, DC,
and Rain. Each of these attributes have fairly high positive correlation coefficients to a class relative to the other attributes. DMC, BUI, and DC all have
correlation coefficients over 0.33, which although is not a strong positive correlation, the combination of these attributes provides insight into a the likeness
of a fire event. Rain and not fire has the strongest of the correlation coefficents at 0.68 rounded which is nearly a strong positive correlation. Therefore, this
relationship indicates that the Rain attribute may be useful in determining a not fire event. In context of the attributes, DMC, BUI, DC, and Rain are all useful in
distinguishing between "not fire" and "fire" events with an element of accuracy based upon correlation coefficent representations of relationships, but given previous
analyses with mean, we can also make further insinuations. For instance, mean temperature during fire is higher than no fire, mean rh during fire is lower than no fire,
mean ws during fire is slightly lower than no fire, and mean rain during fire is significantly lower than no fire. With this, we can derive that fire incidents might
be associated with higher temperatures, occur more frequently during periods of lower relative humidity, and occur more frequently during periods of lower rainfall.
The differennce in mean for Ws is not notable, although, we may be able to use the small variation to help identify "not fire" and "fire" events. Despite the
usefulness of the identified attributes, there were not any very strong positive or negative relationships nor outstanding differences in mean besides Rain and
Temperature such that introducing a new attribute that fits this criteria may be significant in distinguishing between "not fire" and "fire" events. The new attribute
may be something such as geographic region or terrain type which may give stronger insight into potential "not fire" and "fire" events. In conclusion, the most ideal
way to distinguish between "not fire" and "fire" events with the greatest accuracy is to introduce a new attribute with a larger correlation coefficient and/or
variation in mean such as geographic region or terrain type, yet combining the existing correlated attributes including DMC, BUI, DC, Temperature, RH, and Rain
may provide a good accuracy.

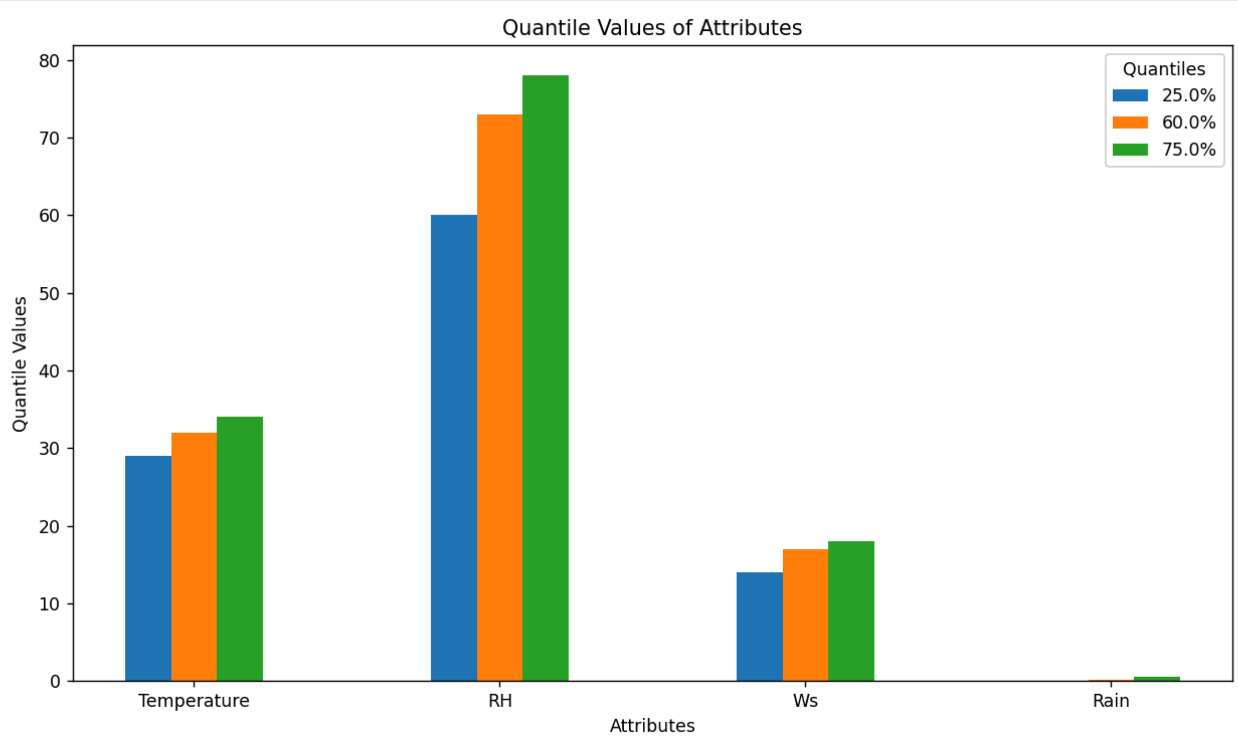
Output:

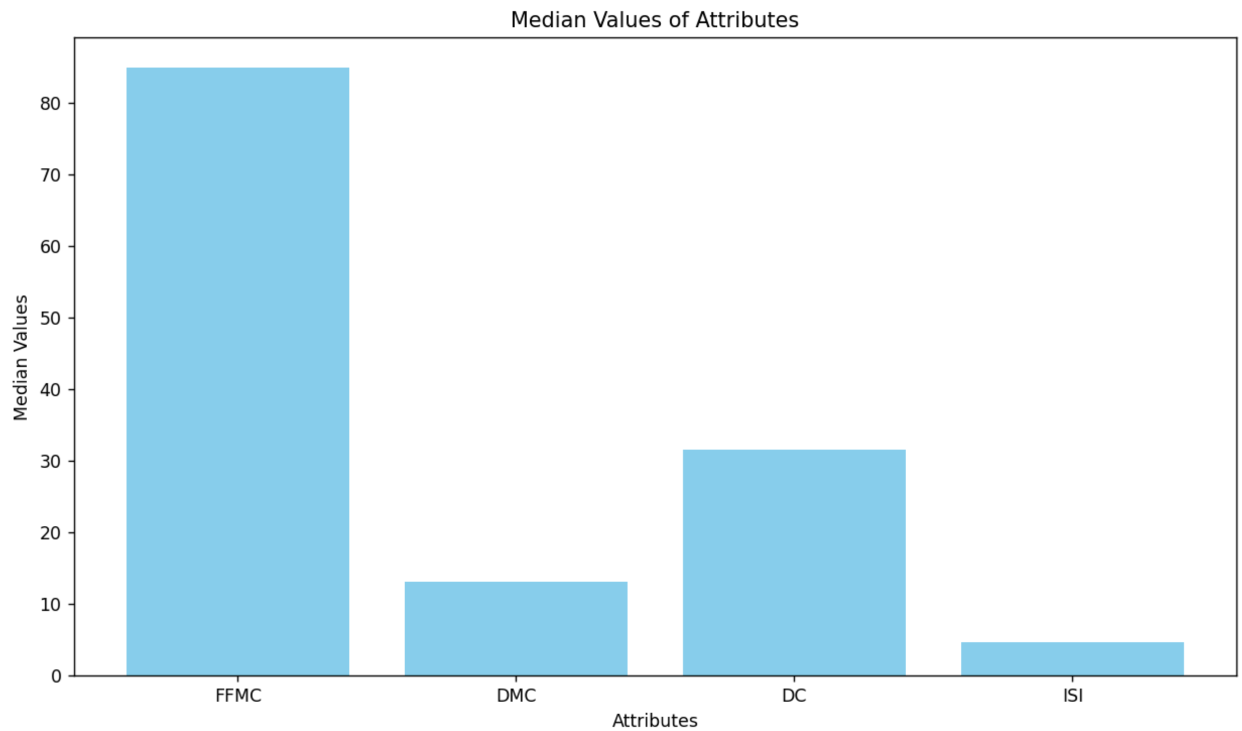
```
Temperature
Fire mean: 32.886792
No Fire mean: 29.816667
RH
Fire mean: 62.90566
No Fire mean: 71.6
Ws
Fire mean: 15.45283
No Fire mean: 16.266667
Rain
Fire mean: 0.015094
No Fire mean: 1.131667

FFMC
Median: 84.85
DMC
Median: 13.15
DC
Median: 31.5
ISI
Median: 4.6

Temperature
Percent Quantiles (0.25, 0.60, 0.75): 29 32 34
RH
Percent Quantiles (0.25, 0.60, 0.75): 60 73 78
Ws
Percent Quantiles (0.25, 0.60, 0.75): 14 17 18
Rain
Percent Quantiles (0.25, 0.60, 0.75): 0.0 0.1 0.5

Temperature
Standard Deviation: 3.306765
Rain
Standard Deviation: 2.399314
BUI
Standard Deviation: 14.414859
FWI
Standard Deviation: 6.317002
```

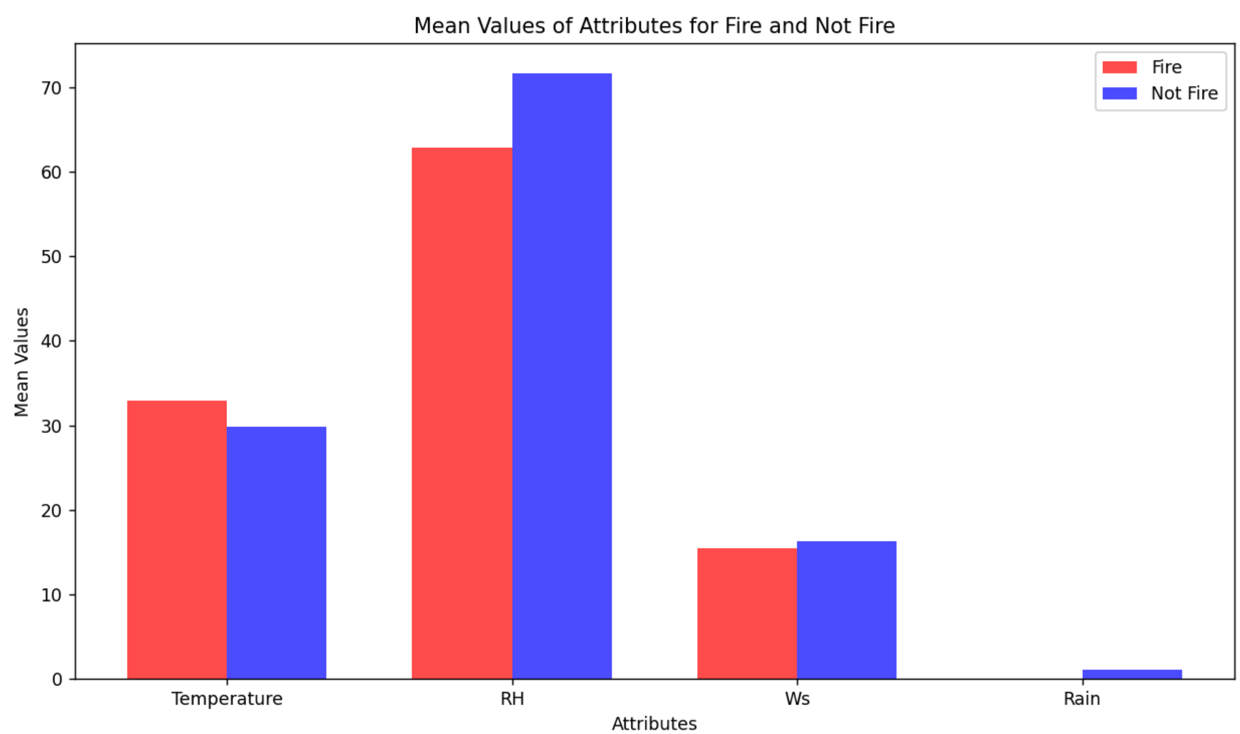
```
Temperature
Correlation Coefficient: -0.660151
Ws
Correlation Coefficient: 0.245774
Rain
Correlation Coefficient: 0.329163
FFMC
Correlation Coefficient: -0.653153
DMC
Correlation Coefficient: -0.34708
DC
Correlation Coefficient: -0.314271
ISI
Correlation Coefficient: -0.58641
BUI
Correlation Coefficient: -0.338233
FWI
Correlation Coefficient: -0.476067

Temperature
Fire Correlation Coefficient: 0.101316
Not Fire Correlation Coefficient: -0.241286
RH
Fire Correlation Coefficient: 0.029412
Not Fire Correlation Coefficient: 0.174609
Ws
Fire Correlation Coefficient: 0.097343
Not Fire Correlation Coefficient: 0.091008
Rain
Fire Correlation Coefficient: -0.06466
Not Fire Correlation Coefficient: 0.678363
FFMC
Fire Correlation Coefficient: 0.143801
Not Fire Correlation Coefficient: -0.27639
DMC
Fire Correlation Coefficient: 0.380186
Not Fire Correlation Coefficient: -0.087554
DC
Fire Correlation Coefficient: 0.338891
Not Fire Correlation Coefficient: -0.11344
ISI
Fire Correlation Coefficient: 0.196957
Not Fire Correlation Coefficient: -0.143959
```

```
BUI
Fire Correlation Coefficient: 0.36975
Not Fire Correlation Coefficient: -0.096064
FWI
Fire Correlation Coefficient: 0.289918
Not Fire Correlation Coefficient: -0.110928
```


STD Values of Attributes
STD Values
Attributes
Temperature
Rain
BUI
FWI
0
2
4
6
8
10
12
14


Quantile Values of Attributes
Quantile Values
Attributes
Quantiles
25.0%
60.0%
75.0%
Temperature
RH
Ws
Rain
0
10
20
30
40
50
60
70
80


Median Values of Attributes
Median Values
Attributes
FFMC
DMC
DC
ISI
0
10
20
30
40
50
60
70
80


Mean Values of Attributes for Fire and Not Fire
Fire
Not Fire
Mean Values
Attributes
Temperature
RH
Ws
Rain
0
10
20
30
40
50
60
70