```python
import numpy as np
import pandas as pd
```

```python
df = pd.read_csv("IRIS.csv")
df
```

```
sepal_length

sepal_width

petal_length

petal_width

species

0

5.1

3.5

1.4

0.2

Iris-setosa

1

4.9

3.0

1.4

0.2

Iris-setosa

2

4.7

3.2

1.3

0.2

Iris-setosa

3

4.6
```

3.1

1.5

0.2

Iris-setosa

4

5.0

3.6

1.4

0.2

Iris-setosa

...

...

...

...

...

...

145

6.7

3.0

5.2

2.3

Iris-virginica

146

6.3

2.5

5.0

1.9

Iris-virginica

147

6.5

3.0

5.2

2.0

Iris-virginica

148

6.2

3.4

5.4

2.3

Iris-virginica

149

5.9

3.0

5.1

1.8

Iris-virginica

150 rows × 5 columns

```
sub_df =
df[["sepal_length","sepal_width","petal_length","petal_width","species
"]]
sub_df.head()
```

sepal_length

sepal_width

petal_length

petal_width

species

0

5.1

3.5

1.4

0.2

Iris-setosa

1

4.9

3.0

1.4

0.2

Iris-setosa

2

4.7

3.2

1.3

0.2

Iris-setosa

3

4.6

3.1

1.5

0.2

Iris-setosa

4

5.0

3.6

1.4

0.2

Iris-setosa

```
#Q1. Central Tendency Without directly calculating the mean, determine
which sepal measurement (sepal_length or sepal_width) has a higher
average by observing the dataset.

#Explanation = sepal_length has higher average because sepal_length
has higher mean than the sepal_width.
```

```
sub_df.describe()
```

sepal_length

sepal_width

petal_length

petal_width

count

150.000000

150.000000

150.000000

150.000000

mean

5.843333

3.054000

3.758667

1.198667

std

0.828066

0.433594

1.764420

0.763161

min

4.300000

2.000000

1.000000

0.100000

25%

5.100000

2.800000

1.600000

0.300000

50%

5.800000

3.000000

4.350000

1.300000

75%

6.400000

3.300000

5.100000

1.800000

max

7.900000

4.400000

6.900000

2.500000

```
#Q2. Dispersion Compare the variability of petal measurements by
checking the difference between the ranges of petal_length and
petal_width. Which feature shows more spread?

#Explanation = range of petal_length has more variability

#range of petal_length
r1 = sub_df['petal_length'].max() - sub_df['petal_length'].min()
r1
```

```
5.9
```

```
#range of petal_width
r2 = sub_df['petal_width'].max() - sub_df['petal_width'].min()
r2
```

```
2.4
```

```
#Q3. Quartiles Identify which feature, sepal_length or sepal_width,
has a higher interquartile range (IQR), indicating greater variability
in the middle 50% of the data.

#Explanation = IQR 1 has the greater variability. so, sepal_length has
more variability.
```

```
#IQR of sepal_length
IQR1 = sub_df['sepal_length'].quantile(0.75) -
sub_df['sepal_length'].quantile(0.25)
IQR1
```

```
1.3000000000000007
```

```
#IQR of sepal_width
IQR2= sub_df['sepal_width'].quantile(0.75) -
sub_df['sepal_width'].quantile(0.25)
IQR2
```

```
0.5
```

```
#Q4. Standard Deviation Insight Based on the computed standard
deviations of sepal_length and sepal_width, determine which attribute
exhibits more variability and discuss what this indicates about the
consistency of the two features.

#Explanation = sepal_length has more standard deviation which means
sepal_length data set has more variability but less consistency and
sepal_width has less standard deviation which means sepal_width has
less consistency.
```

```
#Standard Deviation of sepal_length
sd1 = sub_df["sepal_length"].std()
sd1
```

```
0.8280661279778629
```

```
#Standard Deviation of sepal_width
sd2 = sub_df["sepal_width"].std()
sd2
```

```
0.4335943113621737
```

```
#Q5. Shape of Distribution Without plotting, assess whether the
petal_length distribution is skewed. Use a descriptive statistic that
measures skewness and infer the distribution shape.
```

```
#Explanation = The data of petal_length is left skewed because mean of
petal length is less than the median
```

```
#Mean
mean= sub_df["petal_length"].mean()
mean
```

```
3.758666666666666
```

```
#Median
median= sub_df["petal_length"].median()
median
```

```
4.35
```

```
#Skew
skew = sub_df["petal_length"].skew()
skew
```

```
-0.27446425247378287
```

```
#Q6. Symmetry For which attribute, sepal_length or sepal_width, is the
skewness closer to zero? What does this imply about the symmetry of
its distribution?
```

```
#Explanation = Both sepal_length and sepal_width are right skewed data
and both attributes has not skewness which is closer to zero
```

```
#Mean of sepal_length
mean_l=sub_df["sepal_length"].mean()
mean_l
```

```
5.843333333333334
```

```
#Median of sepal_length
median_l=sub_df["sepal_length"].median()
median_l
```

```
5.8
```

```
#Skew
skew_l = sub_df["sepal_length"].skew()
skew_l
```

```
0.3149109566369728
```

```
#Mean of sepal_width
mean_w=sub_df["sepal_width"].mean()
mean_w
```

```
3.0540000000000003
```

```
#Median of sepal_width
median_w=sub_df["sepal_width"].median()
median_w
```

```
3.0
```

```
#Skew
skew_w=sub_df["sepal_width"].skew()
skew_w
```

```
0.3340526621720866
```

```
#Q7. Coefficient of Variation Calculate the coefficient of variation
(CV) for sepal_length and petal_length. Which feature has greater
relative variability, and what does this tell you about the dataset?

#Explanation = A higher CV indicates greater disparity or variation in
the data set
```

```
mean = sub_df["sepal_length"].mean()
mean
```

```
5.843333333333334
```

```
sd = sub_df["sepal_length"].std()
sd
```

```
0.8280661279778629
```

```
cv_s = sd/mean*100
```

```
mean = sub_df["petal_length"].mean()
mean
```

```
3.758666666666666
```

```
sd = sub_df["petal_length"].std()
sd
```

```
1.7644204199522617
```

```
cv_p = sd/mean*100
```

```
#Q8. Outlier Detection using IQR Identify the outliers in the
sepal_length feature using the IQR method (define outliers as values
that fall below Q1 - 1.5 * IQR or above Q3 + 1.5 * IQR). How many
outliers are present?

#Explanation = there is no outliers
```

```
#Outliers of sepal_length
lower_bound = sub_df['sepal_length'].quantile(0.25) - 1.5 * IQR1
lower_bound
```

```
3.1499999999999986
```

```
upper_bound = sub_df['sepal_length'].quantile(0.75) + 1.5 * IQR1
upper_bound
```

```
8.350000000000001
```

```
outliers = sub_df[(sub_df['sepal_length'] < lower_bound) |
(sub_df['sepal_length'] > upper_bound)]
print("Outliers Rows:")
print(outliers)
```

```
Outliers Rows:
Empty DataFrame
Columns: [sepal_length, sepal_width, petal_length, petal_width,
species]
Index: []
```