```
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats
```

Null Hypothsis : mean(distance of golf balls hit off a regular 2-3/4" wooden tee) is equal to mean(distance of golf balls hit off a Stinger Competition tee)

Alternative hypothesis:mean(distance of golf balls hit off a regular 2-3/4" wooden tee) is not equal to mean(distance of golf balls hit off a Stinger Competition tee)bold text

```
tees = pd.read_csv('GolfTees.csv')
print(tees)
tees.shape
```

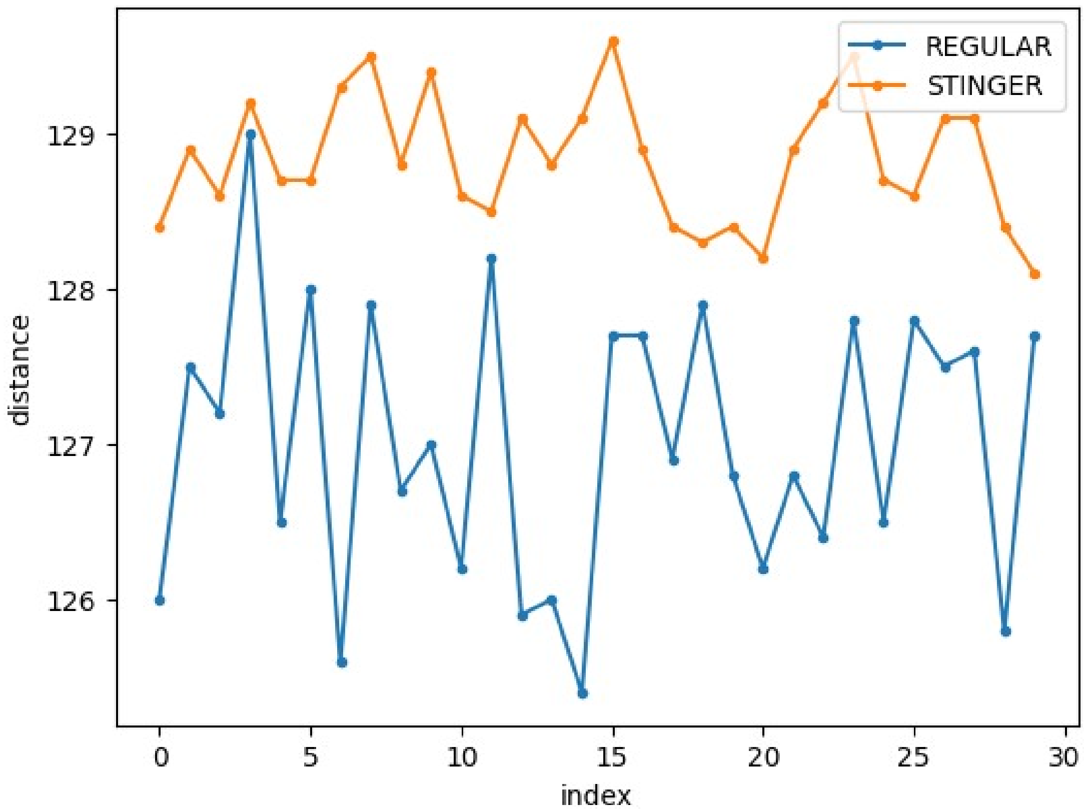
```
    REGULAR  STINGER
0     126.0    128.4
1     127.5    128.9
2     127.2    128.6
3     129.0    129.2
4     126.5    128.7
5     128.0    128.7
6     125.6    129.3
7     127.9    129.5
8     126.7    128.8
9     127.0    129.4
10    126.2    128.6
11    128.2    128.5
12    125.9    129.1
13    126.0    128.8
14    125.4    129.1
15    127.7    129.6
16    127.7    128.9
17    126.9    128.4
18    127.9    128.3
19    126.8    128.4
20    126.2    128.2
21    126.8    128.9
22    126.4    129.2
23    127.8    129.5
24    126.5    128.7
25    127.8    128.6
26    127.5    129.1
27    127.6    129.1
28    125.8    128.4
29    127.7    128.1
```

```
(30, 2)
```

Sample Size of the data is 30

```
tees.plot(style='.-')
plt.xlabel('index')
plt.ylabel('distance')
```

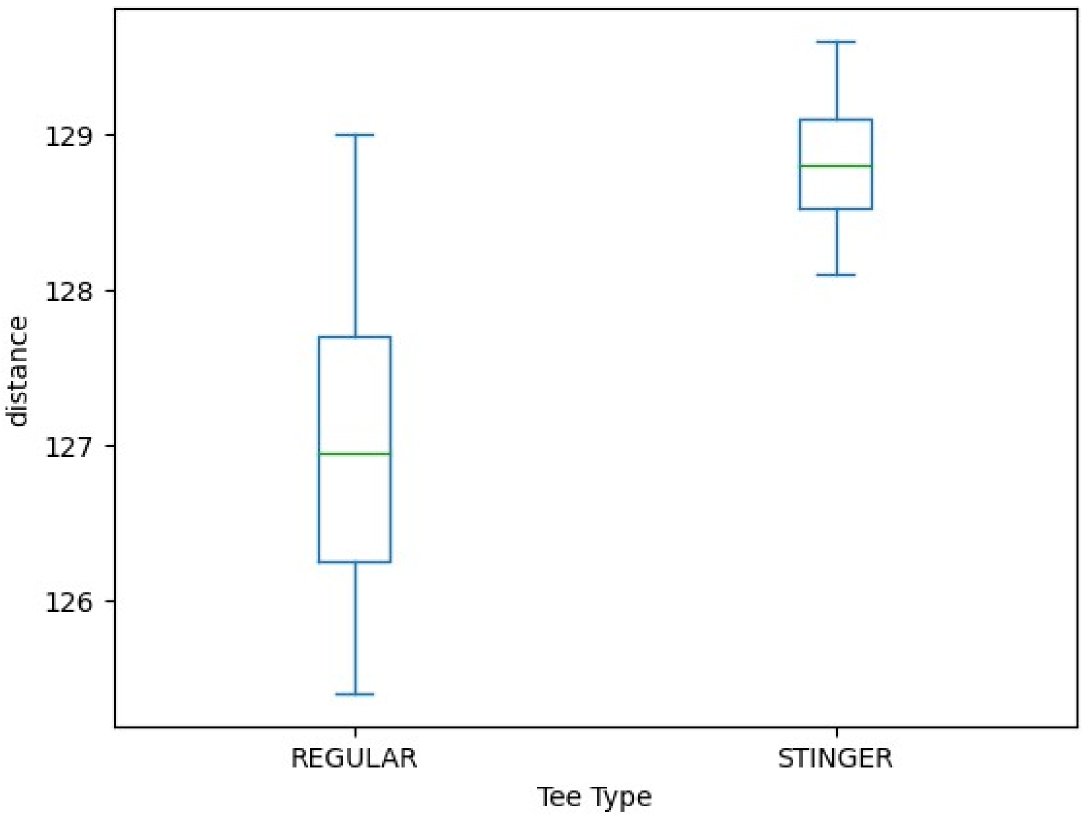
```
Text(0, 0.5, 'distance')
```



OUtliner

```
tees.plot.box()
plt.xlabel('Tee Type')
plt.ylabel('distance')
```

```
Text(0, 0.5, 'distance')
```



A box and whisker plot or diagram (otherwise known as a boxplot), is a graph summarising a set of data. The shape of the boxplot shows how the data is distributed and it also shows any outliers. The box plot of regular and stinger shows that the regular tees covers less distance when compared to the stinger tees based on the box plot And I personnally feel that the null hypothesis is not true that is I feel that the variation in tees would be contributing to the distance of the golf ball covered.Ex., let's say like we are using 10 golf ball with different tees.Based on the box plot it is clearly seen that the distance covered by the 10 golf balls differ with the usage of differnt tees.

```
Mean = tees.mean()
Standard_deviation =  tees.std()
Minimum = tees.min()
Maximum = tees.max()
print("Mean:",Mean)
print('\n')
print("Standard_deviation:",Standard_deviation)
print('\n')
print("Minimum:",Minimum)
print('\n')
print("Maximum:",Maximum)
```

```
Mean: REGULAR    127.006667
STINGER    128.833333
dtype: float64
```

```
Standard_deviation: REGULAR    0.894401
STINGER    0.410495
dtype: float64


Minimum: REGULAR    125.4
STINGER    128.1
dtype: float64


Maximum: REGULAR    129.0
STINGER    129.6
dtype: float64
```

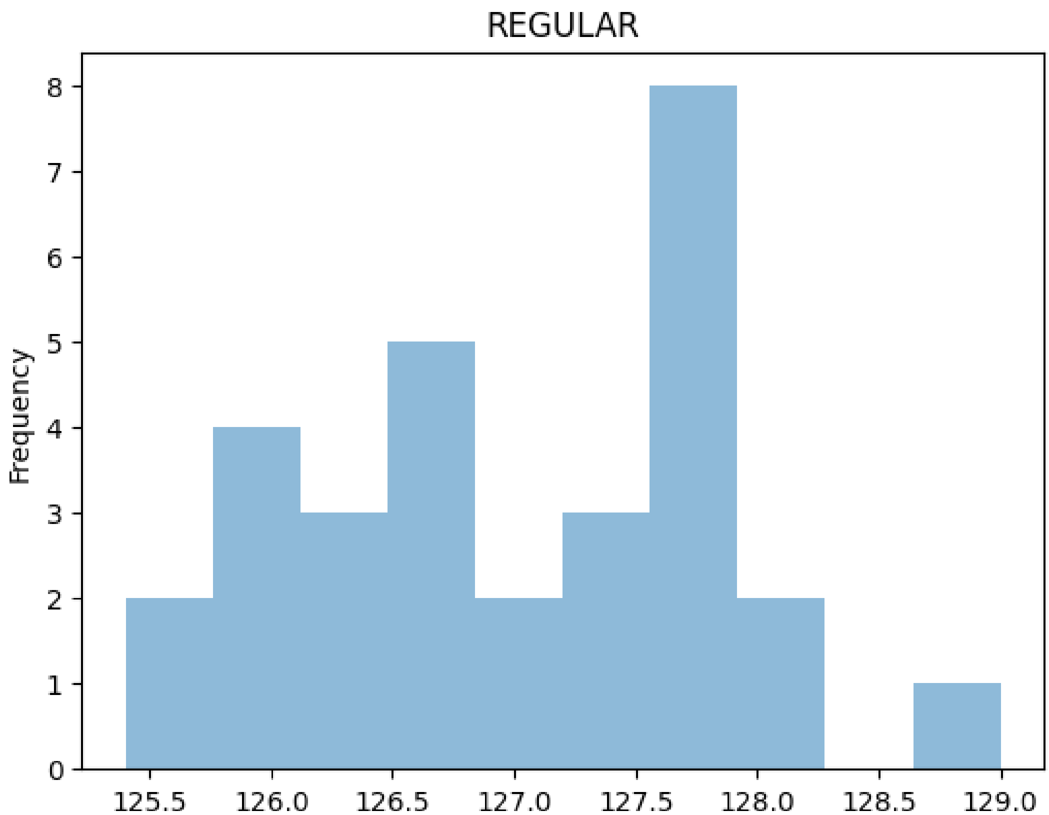
Yes,the values we calculated are consistent with the characteristics observed in the box plots.The box plots for both the 'REGULAR' and 'STINGER' show a symmetrical distribution, as indicated by the approximately centered 'REGULAR' and 'STINGER' median lines within the box. This symmetry suggests that the mean and median are same, aligning with our calculated output.

Moreover, the length of the Interquartile Range (IQR) serves as a measure of variability. In this case, the 'REGULAR' box plot has a longer IQR compared to 'STINGER,' indicating higher variability in the 'REGULAR' tee type. This aligns with our standard deviation calculations, where a higher IQR corresponds to a higher standard deviation and vice versa.

And aslo both the minimum and maximum values are approximately the same from the output that we got for both tee types as we can see from the box plot. This reinforces the consistency of our values with the box plots, confirming that there are no unusual data points or extremes in the dataset

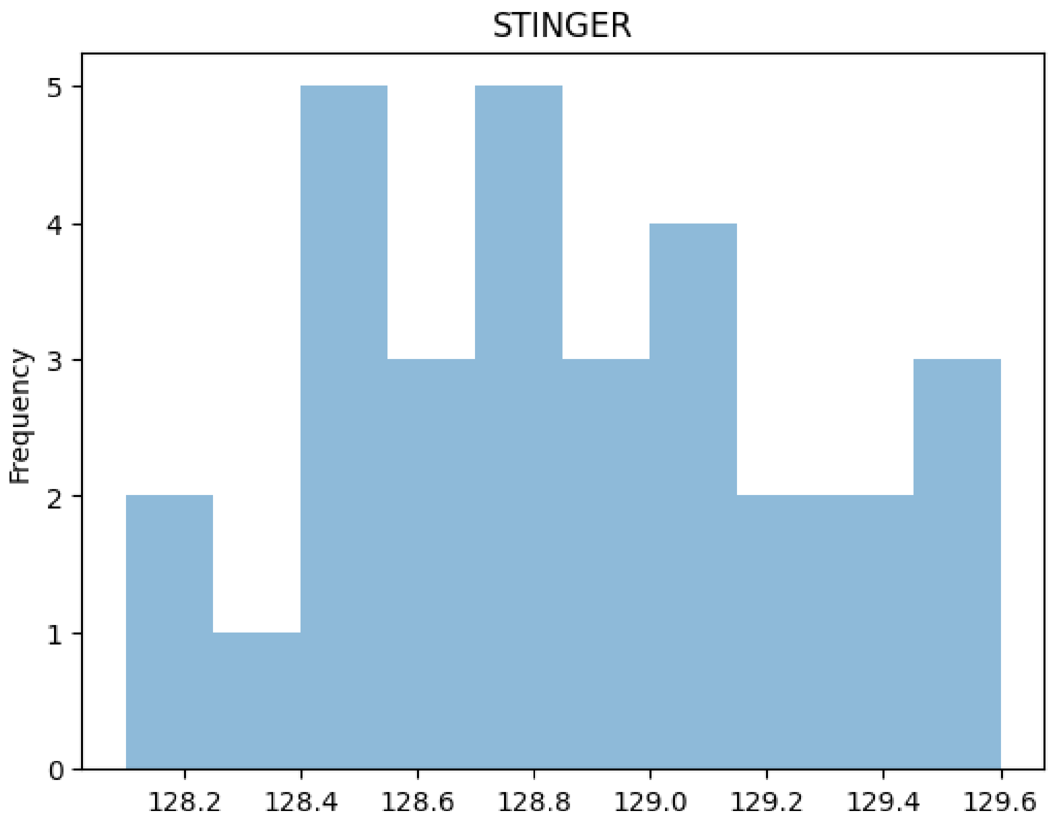
```
tees['REGULAR'].plot.hist(alpha=0.5, bins=10)
plt.title('REGULAR')
```

```
Text(0.5, 1.0, 'REGULAR')
```




```
tees['STINGER'].plot.hist(alpha=0.5, bins=10)
plt.title('STINGER')

Text(0.5, 1.0, 'STINGER')
```



16)The distribution of the numbers, as visually assessed from the provided plots, suggests that the 'REGULAR' tee data appears to have a left-skewed shape, while the 'STINGER' tee data seems to resemble a normal distribution.

```
sw_r = stats.shapiro(tees['REGULAR'])
sw_s = stats.shapiro(tees['STINGER'])
print(sw_r)
print(sw_s)

ShapiroResult(statistic=0.964442253112793, pvalue=0.40010493993759155)
ShapiroResult(statistic=0.9688685536384583, pvalue=0.5086992383003235)

nt_r = stats.normaltest(tees['REGULAR'])
nt_s = stats.normaltest(tees['STINGER'])
print(nt_r)
print(nt_s)

NormaltestResult(statistic=1.061684097167462,
pvalue=0.5881095443126476)
NormaltestResult(statistic=2.024846979700686,
pvalue=0.36333736729193306)
```

In a hypothesis test, the evidence against a null hypothesis is measured by the p-value. The null hypothesis in the Shapiro-Wilk and normal test is that the distribution of the data is normal. When the p-value is low, usually less than 0.05, it means that the null hypothesis is significantly supported by the data. On the other hand, a high p-value indicates insufficient evidence to reject the null hypothesis.

Result of the Shapiro-Wilk test: The 'REGULAR' tee's p-value is 0.4001. The 'STINGER' tee's p-value is 0.5087. The p-values are both over 0.05.Where we can not reject the null hypothesis for any of the two types of Tee.

Normaltest Outcomes: The 'REGULAR' tee's p-value is 0.5881. The 'STINGER' tee's p-value is 0.3633. Both of the normal test's p-values are more than 0.05.In this too we are unable to reject the null hypothesis.

20)Preferred Test: When comparing the means of two groups, particularly when the Standard Deviation are not equal, I advise utilizing the nonpooled t-test.

Tests to Avoid: If the data are independent, we should avoid using tests that comes under the dependent. Tests that assume equal Standard Deviation when they differ should also be avoided.

Potentially Better Options: The nonpooled t-test fits well given the possibility of unequal Standard Deviation.

```
r = tees['REGULAR']
s = tees['STINGER']
tt = stats.ttest_ind(r, s, equal_var=True)
print(tt)
```

```
TtestResult(statistic=-10.16667963779591, pvalue=1.676195302841857e-
14, df=58.0)
```

22)Since the p-value (1.676195302841857e-14) is substantially smaller than the significance level (α = 0.05), we reject the null hypothesis. This implies that there's enough evidence to make the conclusion that there's a considerable difference in mean distances between the 'REGULAR' and 'STINGER' tees:

The t-test results indicate that, with a 95% confidence level, there is a significant difference between the mean distances that golf balls hit off standard wooden tees and Stinger Competition tees. The negative test value (-10.16667963779591) indicates that one tee type's distance is much less than the other.

23)Significant risks arise when errors occur in the decision-making process and a company believes the Stinger Competition Golf Tee will perform well when in reality it does not. This can affect resource allocation, as companies may invest effort and resources to promote or produce products based on inaccurate information. Such a scenario can affect not only the financial aspects but also the overall strategy and image of the company. Therefore, to avoid such mistakes, it is important to carefully consider the potential implications and have an informed decision-making process.