

R.Sathish  
192324204

DATE:09/02/2026

EXP:16

```
16. Frequency Distribution of Words (Customer Reviews)

[1] ✓ 0s
from collections import Counter

reviews = [
    "This product is very good",
    "Good quality and good performance",
    "Product quality is average",
    "Very good and useful product"
]

words = " ".join(reviews).lower().split()
frequency = Counter(words)
print(frequency)

Counter({'good': 4, 'product': 3, 'is': 2, 'very': 2, 'quality': 2, 'and': 2, 'this': 1, 'performance': 1, 'average': 1, 'useful': 1})
```

EXP:17

```
17. Word Frequency from CSV + Bar Plot

[3e] ✓ 0s
import pandas as pd
from collections import Counter
import matplotlib.pyplot as plt

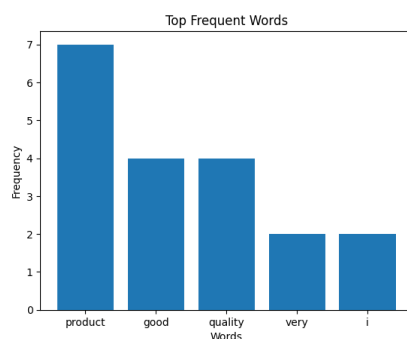
df = pd.read_csv("data.csv")
stop_words = {"the", "is", "and", "but"}

words = []
for text in df["feedback"]:
    for w in text.lower().split():
        if w not in stop_words:
            words.append(w)

freq = Counter(words)
top_words = freq.most_common(5)

plt.bar([w for w, _ in top_words], [c for _, c in top_words])
plt.xlabel("Words")
plt.ylabel("Frequency")
plt.title("Top Frequent Words")
plt.show()
```

OP:



## EXP:18

```
18. Age vs %Fat Analysis

import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm

age = [23,23,27,27,27,39,41,47,49,50,52,54,54,56,57,58,58,60,61]
fat = [9.5,26.5,7.8,17.8,31.4,25.9,27.4,27.2,31.2,
      34.6,42.5,28.8,33.4,30.2,34.1,32.9,41.2,35.7]

df = pd.DataFrame({"Age": age, "Fat": fat})

print(df.describe())

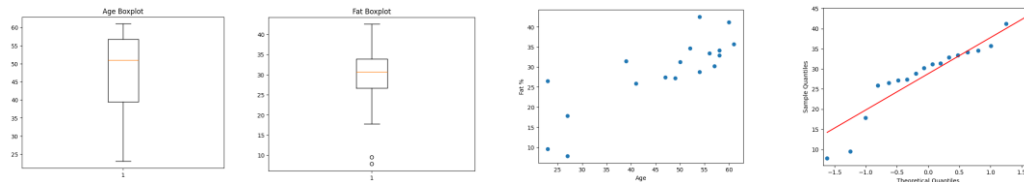
plt.boxplot(age)
plt.title("Age Boxplot")
plt.show()

plt.boxplot(fat)
plt.title("Fat Boxplot")
plt.show()

plt.scatter(age, fat)
plt.xlabel("Age")
plt.ylabel("Fat %")
plt.show()

sm.qqplot(df[["Fat"]], line='s')
plt.show()
```

OP:



## EXP:19

```
19. Confidence Interval (Drug vs Placebo)

import numpy as np
from scipy import stats
np.random.seed(1)
drug = np.random.normal(10, 2, 50)
placebo = np.random.normal(5, 2, 50)

ci_drug = stats.t.interval(0.95, len(drug)-1,
                           np.mean(drug), stats.sem(drug))
ci_placebo = stats.t.interval(0.95, len(placebo)-1,
                              np.mean(placebo), stats.sem(placebo))

print("Drug CI:", ci_drug)
print("Placebo CI:", ci_placebo)

Drug CI: (np.float64(9.392266700207967), np.float64(10.505673907761432))
Placebo CI: (np.float64(4.844162698510661), np.float64(5.74255951012553))
```

## EXP:20

```
20. A/B Testing (Conversion Rate)

import numpy as np
from scipy import stats
A = np.random.normal(0.12, 0.02, 100)
B = np.random.normal(0.14, 0.02, 100)
t_stat, p_val = stats.ttest_ind(A, B)
print("t-statistic:", t_stat)
print("p-value:", p_val)

t-statistic: -6.221644677990623
p-value: 2.873413301359196e-09
```