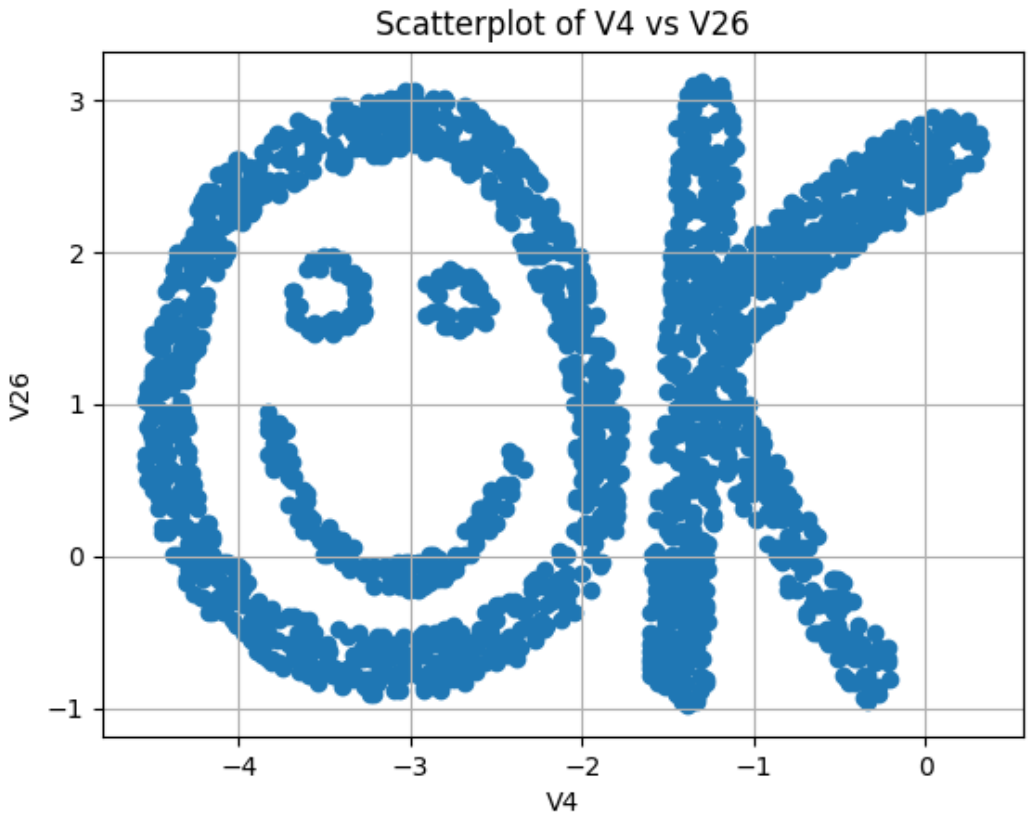
Problem1

```python
data = pd.read_csv('x.csv')

variances = data.var(numeric_only=True)

top2 = variances.nlargest(2).index.tolist()
var1, var2 = top2[0], top2[1]
print(f"Top 2 variables by variance: {var1}, {var2}")

plt.figure()
plt.scatter(data[var1], data[var2])
plt.xlabel(var1)
plt.ylabel(var2)
plt.title(f"Scatterplot of {var1} vs {var2}")
plt.grid(True)
plt.savefig("problem1.png")
plt.show()
```



Problem2

```python
A = np.array([[1.0, 2.0],
              [2.0, 1.618]])
w, v = np.linalg.eig(A)

order = np.argsort(-w)
w = w[order]
```

```python
v = v[:, order]

for i in range(2):
    v[:,i] = v[:,i] / np.linalg.norm(v[:,i])

print("Eigenvalues:", w)
print("Eigenvectors (columns):\n", v)
print("X^T X =\n", np.round(v.T @ v, 12))
recon = sum(w[i] * np.outer(v[:,i], v[:,i]) for i in range(2))
print("Reconstruction:\n", recon)
```

```
Eigenvalues: [ 3.33272948 -0.71472948]
Eigenvectors (columns):
 [[-0.65088847 -0.75917336]
 [-0.75917336  0.65088847]]
X^T X =
 [[1. 0.]
 [0. 1.]]
Reconstruction:
 [[1.    2.   ]
 [2.    1.618]]
```

Problem3

Target: For any random variable `X, Y` and any real scalar `a, b`:

$$E[aX + bY] = aE[X] + bE[Y] \tag{1}$$

$$\begin{aligned} E[aX + bY] &= \sum_{\omega \in \Omega} P(\omega)(aX(\omega) + bY(\omega)) \\ &= a\sum_{\omega \in \Omega} P(\omega)X(\omega) + b\sum_{\omega \in \Omega} P(w)Y(\omega) \\ &= aE[X] + bE[Y] \end{aligned} \tag{2}$$

Target: if `μ=E[X]`

$$Var[X] = E[(X-\mu)^2] = E[X^2] - E[X]^2 \tag{3}$$

$$\begin{aligned} E[(X-\mu)^2] &= E[X^2 - 2 \cdot \mu \cdot X + \mu^2] \\ &= E[X^2] - 2 \cdot \mu E[X] + \mu^2 \\ &= E[X^2] - 2 \cdot \mu \cdot \mu + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= E[X^2] - E[X]^2 \end{aligned} \tag{4}$$

Problem4

A

$$P(X|Y) = \frac{P(X \wedge Y)}{P(Y)} \tag{5}$$

$$P(Y|X) = \frac{P(X \wedge Y)}{P(X)} \tag{6}$$

$$\frac{P(X|Y)}{P(Y|X)} = \frac{P(X)}{P(Y)} \tag{7}$$

B

Define Boolean random variables:

- X = “person has pollen allergy” (true/false).
- Y = “test result is positive” (true/false).

Given

$$\begin{gathered} P(X) = 0.20 \quad P(\neg X) = 0.80 \\ P(Y \mid \neg X) = 0.23 \\ P(Y \mid X) = 1 - 0.15 = 0.85 \end{gathered} \tag{8}$$

$$P(Y) = P(Y \mid X)P(X) + P(Y \mid \neg X)P(\neg X) = 0.354 \tag{9}$$

$$P(X \mid Y) = \frac{P(X)P(Y \mid X)}{P(Y)} \approx 0.48 \tag{10}$$

Problem5

A

differentiate b

$$\begin{aligned} f'(b) &= \frac{1}{2}\sum_{i=1}^{n} 2x_i(bx_i - y_i) \\ &= \sum_{i=1}^{n} x_i(bx_i - y_i) \\ &= b\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i \end{aligned} \tag{11}$$

set `f'(b) = 0`

$$b = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{12}$$

$$f''(b) = \sum_{i=1}^{n} x_i^2 >= 0 \tag{13}$$

so `f'(b)` strict increase

so `f(b)` is min when `f'(b) = 0`

$$b = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{14}$$

B

$$f''(b) = \sum_{i=1}^{n} x_i^2 > 0 \tag{15}$$