

Worksheet: Social Media Likes vs Hashtags & Emojis

Code with Plot:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Generate random data for hashtags and emojis
hashtags = np.random.randint(0, 10, 50) # 50 values between 0 and 9
emojis = np.random.randint(0, 5, 50) # 50 values between 0 and 4

# Calculate likes based on hashtags and emojis
likes = 10 * hashtags + 15 * emojis + np.random.normal(0, 10, 50)

# Combine hashtags and emojis into a 2D array for regression
X2 = np.column_stack((hashtags, emojis))
y2 = likes

# Fit a linear regression model
model2 = LinearRegression().fit(X2, y2)

# Plot the data
plt.scatter(hashtags, y2, color='green', label='Hashtags')
plt.scatter(emojis, y2, color='purple', label='Emojis')
plt.xlabel('Count')
plt.ylabel('Likes')
plt.title('Likes vs Hashtags & Emojis')
plt.legend()
plt.show()
```

Explanation:

This code models how hashtags and emojis influence the number of likes on a social media post.

We generate random counts for hashtags and emojis using `np.random.randint`.

Likes are calculated based on a weighted formula, with added noise to reflect variability.

`np.column_stack` prepares the data for regression analysis.

We train a linear regression model to learn the relationship.

The plot shows how likes vary with the number of hashtags and emojis.

[Space for answers]

Questions:

1. How do hashtags and emojis affect likes?
2. Why is random noise added to the model?
3. What other features might influence likes?