The CLT in Action - Sampling Means Loop

Question:

1. Set the seed to 104.
2. Take a sample of size 20 with replacement from the `num_users` column of `amir_deals` and calculate the mean.
3. Repeat this process 100 times using a for loop, storing the means in a list called `sample_means`.
4. Print the list of sample means.

Explanation of the Question:

This task demonstrates the central limit theorem by repeatedly sampling the data and calculating the means. The goal is to observe how the sample means form a distribution that approximates a normal distribution as more samples are taken.

Answer:

```
import pandas as pd
import numpy as np

# Set seed to 104
np.random.seed(104)

# Assuming amir_deals is a DataFrame with a column named 'num_users'
amir_deals = pd.DataFrame({
    'num_users': [10, 15, 12, 20, 25, 30, 22, 18, 14, 19, 28, 17]
})

# Sample 20 num_users with replacement from amir_deals and take mean
samp_20 = amir_deals['num_users'].sample(20, replace=True)
np.mean(samp_20)

sample_means = []
# Loop 100 times
for i in range(100):
  # Take sample of 20 num_users
  samp_20 = amir_deals['num_users'].sample(20, replace=True)
  # Calculate mean of samp_20
  samp_20_mean = np.mean(samp_20)
  # Append samp_20_mean to sample_means
  sample_means.append(samp_20_mean)
```

print(sample_means)

Explanation of the Answer:

The `np.random.seed(104)` ensures reproducibility. A single sample mean is calculated first using a sample of size 20 drawn with replacement. In a loop, 100 such samples are drawn, and their means are calculated and stored in the `sample_means` list. Finally, the list of sample means is printed.