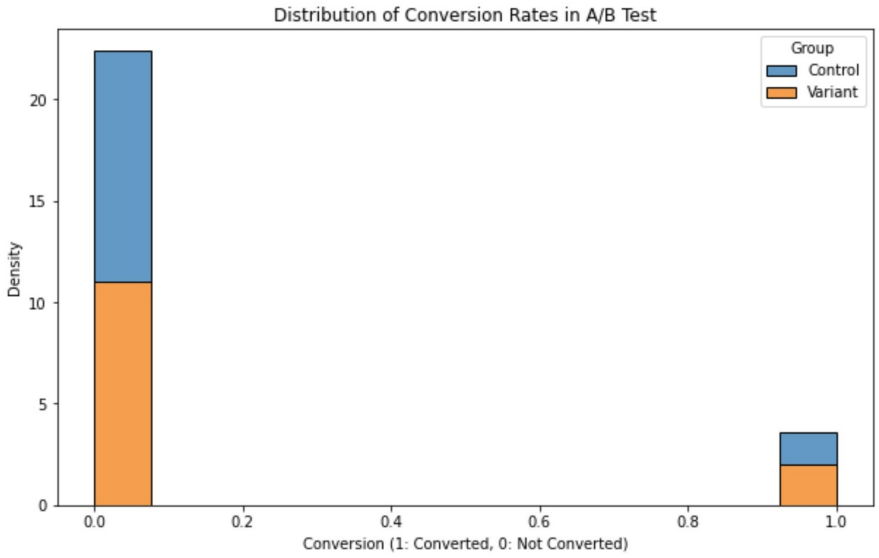
Project 4: A/B Testing for Website Optimization

```
import pandas as pd
import numpy as np
from scipy.stats import norm
import matplotlib.pyplot as plt
import seaborn as sns

# Set seed for reproducibility
np.random.seed(42)

# Number of users in each group
num_control_group = 1000
num_variant_group = 1200

# Conversion rates
conversion_rate_control = 0.12  # 12% conversion rate for control group
conversion_rate_variant = 0.15  # 15% conversion rate for variant group

# Generating synthetic data for A/B testing
control_group = np.random.choice([0, 1], size=num_control_group, p=[1 - conversion_rate_cont
variant_group = np.random.choice([0, 1], size=num_variant_group, p=[1 - conversion_rate_vari

# Create a DataFrame
ab_test_data = pd.DataFrame({
    'UserID': range(1, num_control_group + num_variant_group + 1),
    'Group': ['Control'] * num_control_group + ['Variant'] * num_variant_group,
    'Converted': np.concatenate([control_group, variant_group])
})

# Visualize the distribution of conversion rates in both groups
plt.figure(figsize=(10, 6))
sns.histplot(data=ab_test_data, x='Converted', hue='Group', multiple="stack", stat="density'
plt.title('Distribution of Conversion Rates in A/B Test')
plt.xlabel('Conversion (1: Converted, 0: Not Converted)')
plt.ylabel('Density')
plt.show()

# Save the data to a CSV file
ab_test_data.to_csv('ab_test_data.csv', index=False)
```



Here's an interpretation based on the synthetic data generated:

1. Conversion Rates:

Control Group (12%): The proportion of users who converted (e.g., made a purchase, signed up) in the control group is 12%. Variant Group (15%): The variant group, which experienced the changes or optimizations, has a higher conversion rate of 15%.

2. Visualizing the Distribution:

The histogram visually represents the distribution of conversion rates in both the control and variant groups. You can observe the shift in the distribution of the variant group to the right, indicating a higher concentration of converted users.

3. Statistical Significance:

Conduct statistical tests, such as a chi-squared test or a t-test, to determine if the observed differences in conversion rates are statistically significant. If the p-value is below a predefined significance level (e.g., 0.05), you may conclude that the observed differences are not due to random chance.

4. Impact on Conversion:

The higher conversion rate in the variant group suggests that the implemented changes had a positive impact on user behavior. The increase from 12% to 15% represents a relative improvement of 25%.

5. Practical Significance:

Consider not only statistical significance but also practical significance. Even if the results are statistically significant, assess whether the observed improvement is practically meaningful for the business.

6. Business Recommendations:

Based on the analysis, it is recommended to implement the changes from the variant group on the live website to potentially improve overall conversion rates.

7. Future Steps:

Monitor the long-term impact of the changes on user behavior. Consider additional metrics such as user engagement, revenue per user, or any other relevant KPIs.