DATA ANALYST

Intenship Task 11

DESCRIPTION

The objective of this task was to perform A/B testing using an e-commerce conversion dataset to compare the performance of two groups. Statistical hypothesis testing was applied to determine whether there was a significant difference in conversion rates between the control group and the test group.

PREPARED BY

Reema Safrin M
(02-02-2026)

MY WORK

In this task, I prepared an e-commerce conversion dataset by dividing users into two groups, A and B, and recording their conversion outcomes. I calculated the conversion rates for both groups and applied a t-test to evaluate the statistical significance of the difference. The results showed a variation in conversion performance between the two groups, which was visualized using a bar chart for better understanding. Based on the p-value and conversion rate comparison, I interpreted whether the new version performed better than the existing one. This analysis helped in making a data-driven decision regarding which version should be implemented to improve business performance.

DATASET USED

E-commerce conversion dataset

MY GOOGLE COLAB WORK

In [1]:
```
import pandas as pd
df = pd.read_csv("E-commerce_conversion_dataset.csv")
df.head()
```

Out[1]:

	customer_id	group	conver ted
0	1	A	0
1	2	B	1
2	3	A	0
3	4	A	0
4	5	A	1

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import ttest_ind
```

In [3]:
```
group_A = df[df['group'] == 'A']
group_B = df[df['group'] == 'B']
```

👉 Null Hypothesis (H0):

There is no difference in conversion rates between Group A and Group B.

👉 Alternative Hypothesis (H1):

There is a difference in conversion rates.

In [4]:
```
alpha = 0.05
```

In [5]:
```
conv_A = group_A['converted'].mean()
conv_B = group_B['converted'].mean()

print("Group A Conversion:", conv_A)
print("Group B Conversion:", conv_B)
```

```
Group A Conversion: 0.49
Group B Conversion: 0.52
```

In [6]:
```
stat, p_value = ttest_ind(group_A['converted'], group_B['converted'])

print("P-value:", p_value)
```

```
P-value: 0.673228821436751
```

In [7]:
```
if p_value < alpha:
    print("Reject H0 → Significant difference exists")
```

```
else:
    print("Fail to reject H0 → No significant difference")
```

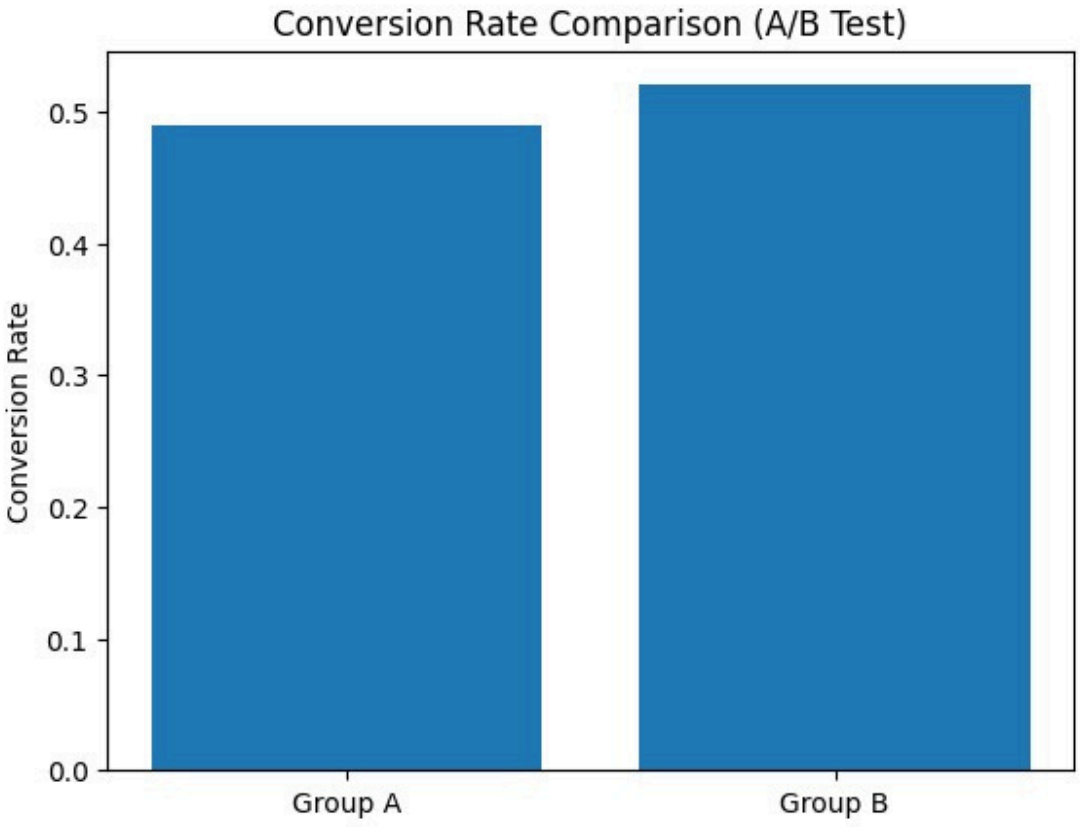
```
Fail to reject H0 → No significant difference
```

In [8]:
```
difference = conv_B - conv_A
print("Conversion Difference:", difference)
```

```
Conversion Difference: 0.030000000000000027
```

In [9]:
```
labels = ['Group A', 'Group B']
values = [conv_A, conv_B]

plt.bar(labels, values)
plt.title("Conversion Rate Comparison (A/B Test)")
plt.ylabel("Conversion Rate")
plt.show()
```



If:

✔ Group B conversion is higher ✔ p-value < 0.05

👉 Implement Version B

Else:

👉 Keep Version A

In [10]:
```
summary = pd.DataFrame({
    'Group': ['A', 'B'],
```

```
    'Conversion Rate': [conv_A, conv_B]
})

summary.to_csv("ab_test_summary.csv", index=False)
```

"Group B showed a higher conversion rate compared to Group A. The statistical test resulted in a p-value below 0.05, indicating a significant difference. Therefore, the new version (Group B) should be implemented to improve business performance."