A|B Testing for emails

Connecting to sqlite and importing the dataset

In [1]:
```
import os
import sqlite3
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
db_path = r'C:\Users\dimit\SQLite_databases\Email_campaign.db'
conn = sqlite3.connect(db_path)
cursor = conn.cursor()

cursor.execute("SELECT name FROM sqlite_master WHERE type='table';")
tables = cursor.fetchall()
print("Tables in the database:", tables)
```

```
Tables in the database: [('Emails',), ('sqlite_sequence',), ('email_ab_test_data',)]
```

In [2]:
```
cursor.execute("SELECT * FROM email_ab_test_data;")
rows = cursor.fetchall()

# Print the results
for row in rows:
    print(row)
```

```
('A', 10200, 20, 9162, 9, 0.0009823182711198428)
('B', 9900, 9, 8802, 17, 0.0019313792319927288)
```

In [3]:
```
query = "SELECT * FROM email_ab_test_data"
df = pd.read_sql_query(query, conn)
df
```

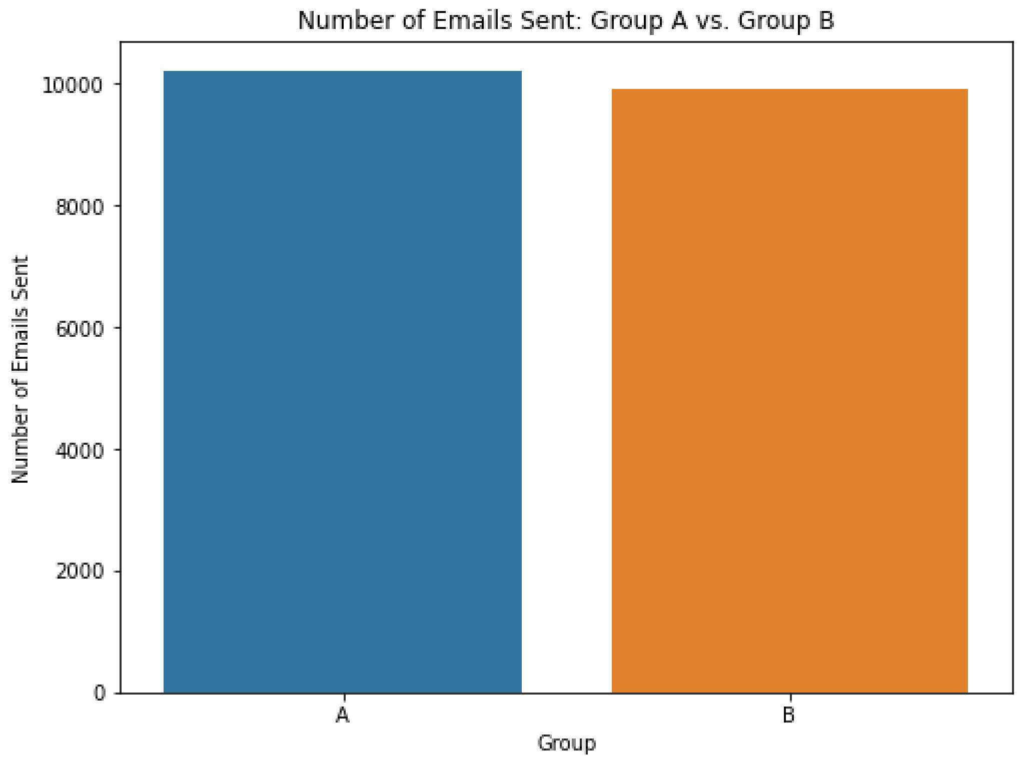
Out[3]:

	Dataset	Sent	Bounced	Opened	Sales	ConversionRate
0	A	10200	20	9162	9	0.000982
1	B	9900	9	8802	17	0.001931

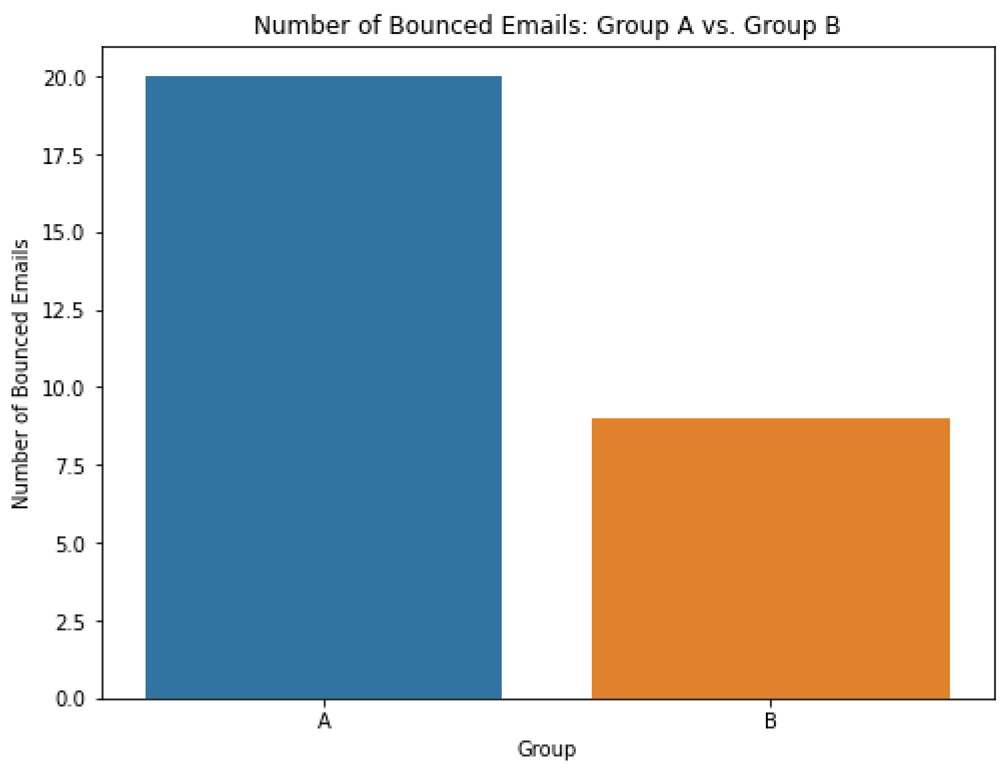
In [4]:
```
conn.close()
```

Visualizations

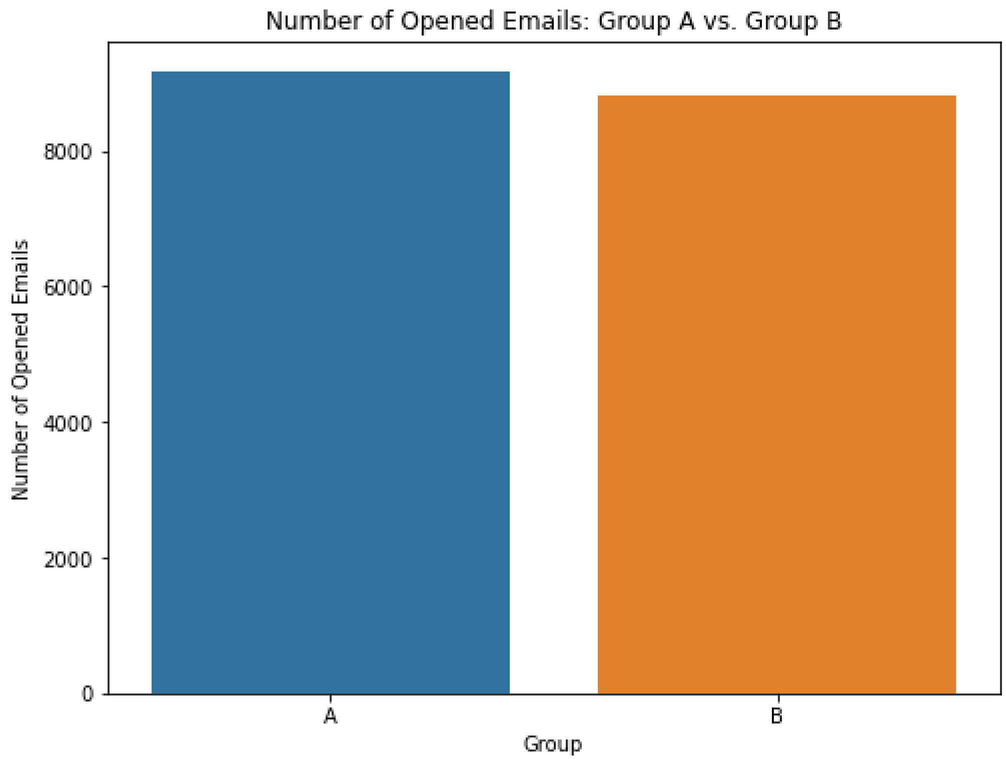
In [5]:
```
plt.figure(figsize=(8, 6))
sns.barplot(x='Dataset', y='Sent', data=df)
plt.title('Number of Emails Sent: Group A vs. Group B')
plt.xlabel('Group')
plt.ylabel('Number of Emails Sent')
plt.show()
```



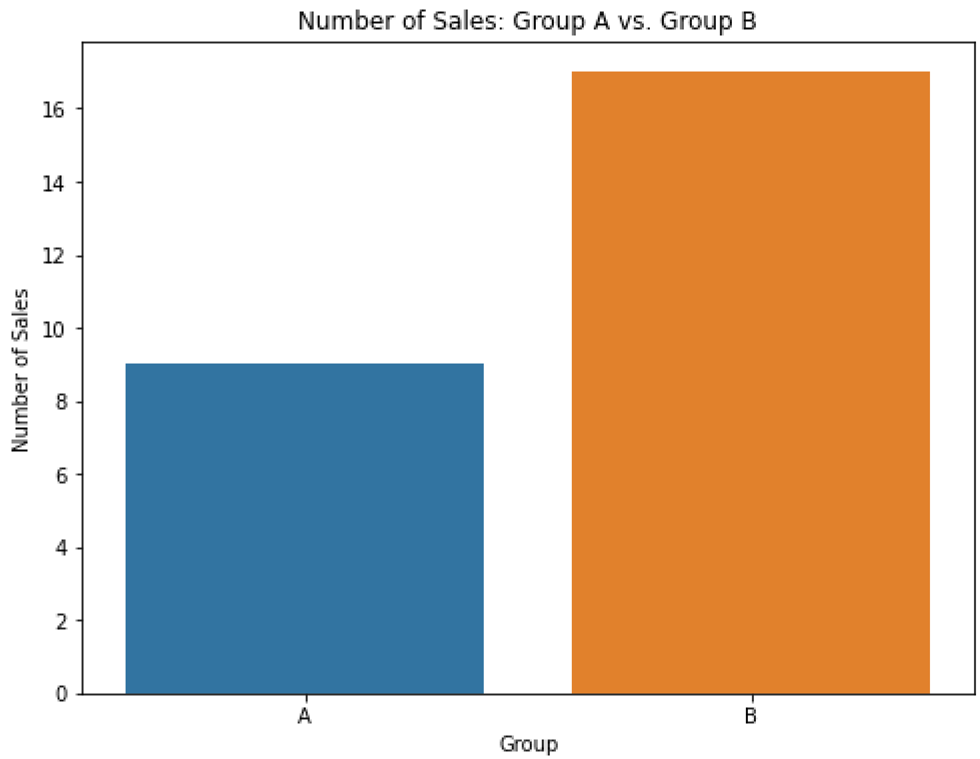
In [6]:
```
plt.figure(figsize=(8, 6))
sns.barplot(x='Dataset', y='Bounced', data=df)
plt.title('Number of Bounced Emails: Group A vs. Group B')
plt.xlabel('Group')
plt.ylabel('Number of Bounced Emails')
plt.show()
```



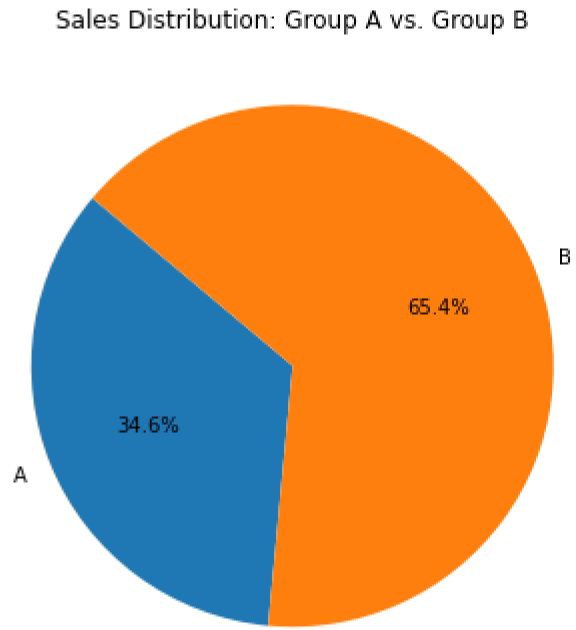
In [7]:
```
plt.figure(figsize=(8, 6))
sns.barplot(x='Dataset', y='Opened', data=df)
plt.title('Number of Opened Emails: Group A vs. Group B')
plt.xlabel('Group')
plt.ylabel('Number of Opened Emails')
plt.show()
```



In [8]:
```
plt.figure(figsize=(8, 6))
sns.barplot(x='Dataset', y='Sales', data=df)
plt.title('Number of Sales: Group A vs. Group B')
plt.xlabel('Group')
plt.ylabel('Number of Sales')
plt.show()
```



In [9]:
```
plt.figure(figsize=(8, 6))
plt.pie(df['Sales'], labels=df['Dataset'], autopct='%1.1f%%', startangle=140)
plt.title('Sales Distribution: Group A vs. Group B')
plt.show()
```



Chi-square for "Sent" vs. "Opened"

In [13]:
```
from scipy.stats import chi2_contingency
```

In [11]:
```
opened_A = df.loc[df['Dataset'] == 'A', 'Opened'].values[0]
sent_A = df.loc[df['Dataset'] == 'A', 'Sent'].values[0]
opened_B = df.loc[df['Dataset'] == 'B', 'Opened'].values[0]
sent_B = df.loc[df['Dataset'] == 'B', 'Sent'].values[0]

contingency_table = [
    [opened_A, sent_A - opened_A],  # Group A: Opened, Not Opened
    [opened_B, sent_B - opened_B]   # Group B: Opened, Not Opened
]

print(contingency_table)
```

```
[[9162, 1038], [8802, 1098]]
```

In [15]:
```
chi2, p, dof, expected = chi2_contingency(contingency_table)
```

In [16]:
```
print("Chi-Square Test: Sent vs. Opened")
print(f"Chi-square statistic: {chi2:.4f}")
print(f"P-value: {p:.4f}")
print(f"Degrees of freedom: {dof}")
print("Expected frequencies:")
print(expected)
```

```
Chi-Square Test: Sent vs. Opened
Chi-square statistic: 4.3274
P-value: 0.0375
Degrees of freedom: 1
Expected frequencies:
[[9116.05970149 1083.94029851]
 [8847.94029851 1052.05970149]]
```

P-value: 0.0375, Since 0.0375 is less than the typical significance level of 0.05, we reject the null hypothesis.

Interpretation: There is a statistically significant difference in open rates between Group A and Group B. Group A and B are significantly different regarding the opened emails rate.

Chi-square for "Opened" vs. "Sales"

In [18]:
```
sales_A = df.loc[df['Dataset'] == 'A', 'Sales'].values[0]
sales_B = df.loc[df['Dataset'] == 'B', 'Sales'].values[0]

contingency_table = [
    [sales_A, opened_A - sales_A],  # Group A: Sales, Not Sales (Opened)
    [sales_B, opened_B - sales_B]   # Group B: Sales, Not Sales (Opened)
]
```

In [19]:
```
contingency_table
```

Out[19]:
```
[[9, 9153], [17, 8785]]
```

In [20]:
```
chi2, p, dof, expected = chi2_contingency(contingency_table)
```

In [21]:
```
print("Chi-Square Test: Opened vs. Sales")
print(f"Chi-square statistic: {chi2:.4f}")
print(f"P-value: {p:.4f}")
print(f"Degrees of freedom: {dof}")
print("Expected frequencies:")
print(expected)
```

```
Chi-Square Test: Opened vs. Sales
Chi-square statistic: 2.1796
P-value: 0.1398
Degrees of freedom: 1
Expected frequencies:
[[  13.26052104 9148.73947896]
 [  12.73947896 8789.26052104]]
```

P-value: 0.1398, Since 0.1398 is greater than the typical significance level of 0.05, we fail to reject the null hypothesis.

Interpretation: There is no statistically significant difference in sales within the opened emails between Group A and Group B.

Chi-square for "Sent" vs. "Sales"

In [22]:
```
contingency_table = [
    [sales_A, sent_A - sales_A],  # Group A: Sales, Not Sales (Sent)
    [sales_B, sent_B - sales_B]   # Group B: Sales, Not Sales (Sent)
]
```

In [23]:
```
contingency_table
```

Out[23]:
```
[[9, 10191], [17, 9883]]
```

In [24]:
```
chi2, p, dof, expected = chi2_contingency(contingency_table)
```

In [25]:
```
print("Chi-Square Test: Sent vs. Sales")
print(f"Chi-square statistic: {chi2:.4f}")
print(f"P-value: {p:.4f}")
print(f"Degrees of freedom: {dof}")
print("Expected frequencies:")
print(expected)
```

```
Chi-Square Test: Sent vs. Sales
Chi-square statistic: 2.1026
P-value: 0.1471
Degrees of freedom: 1
Expected frequencies:
[[   13.19402985 10186.80597015]
 [   12.80597015  9887.19402985]]
```

P-value: 0.1471, Since 0.1471 is greater than the typical significance level of 0.05, we fail to reject the null hypothesis. Interpretation: There is no statistically significant difference in sales per sent email between Group A and Group B.