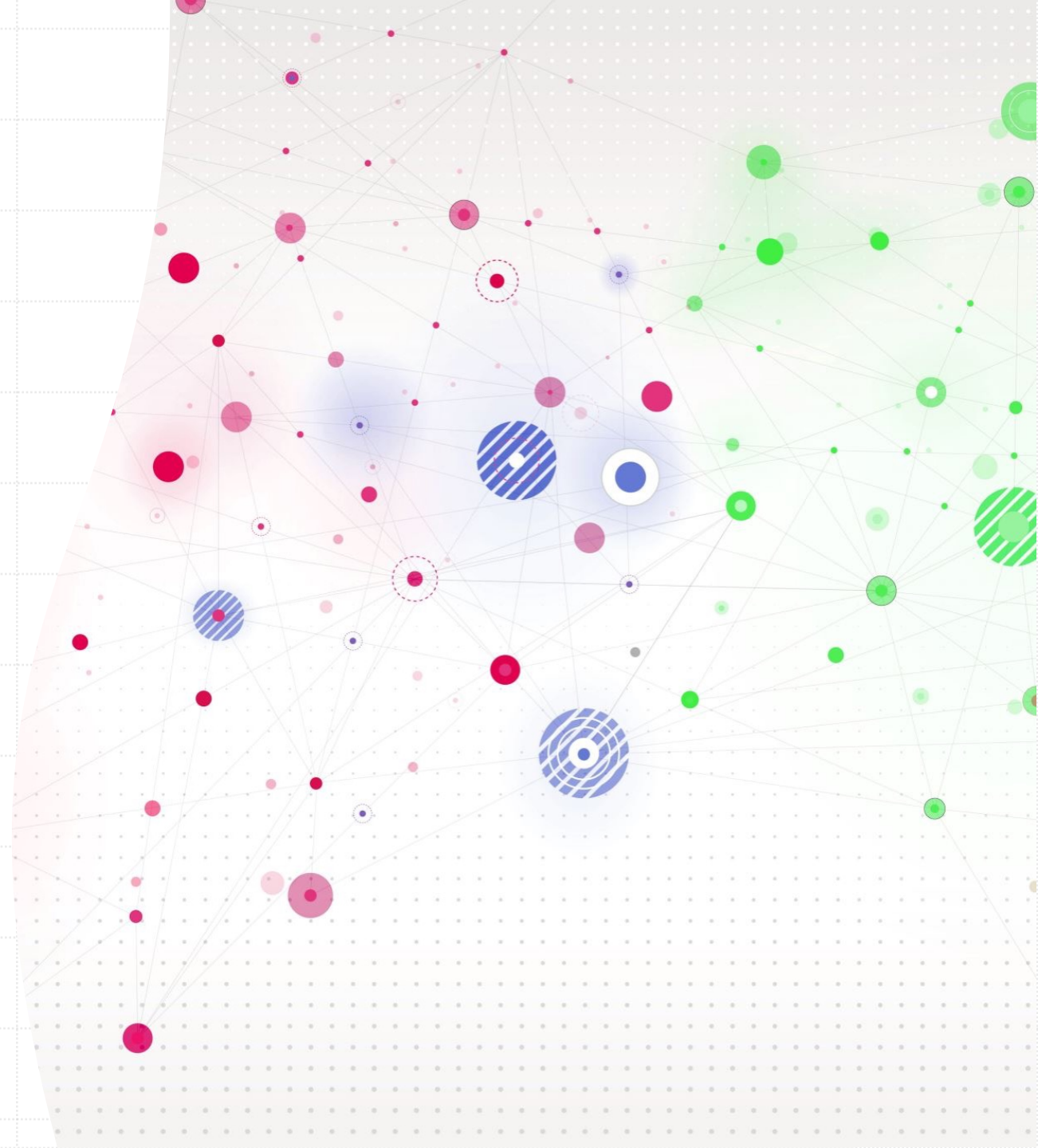


# Mastery Project #1

Prepared by Fereshteh Ranjbar

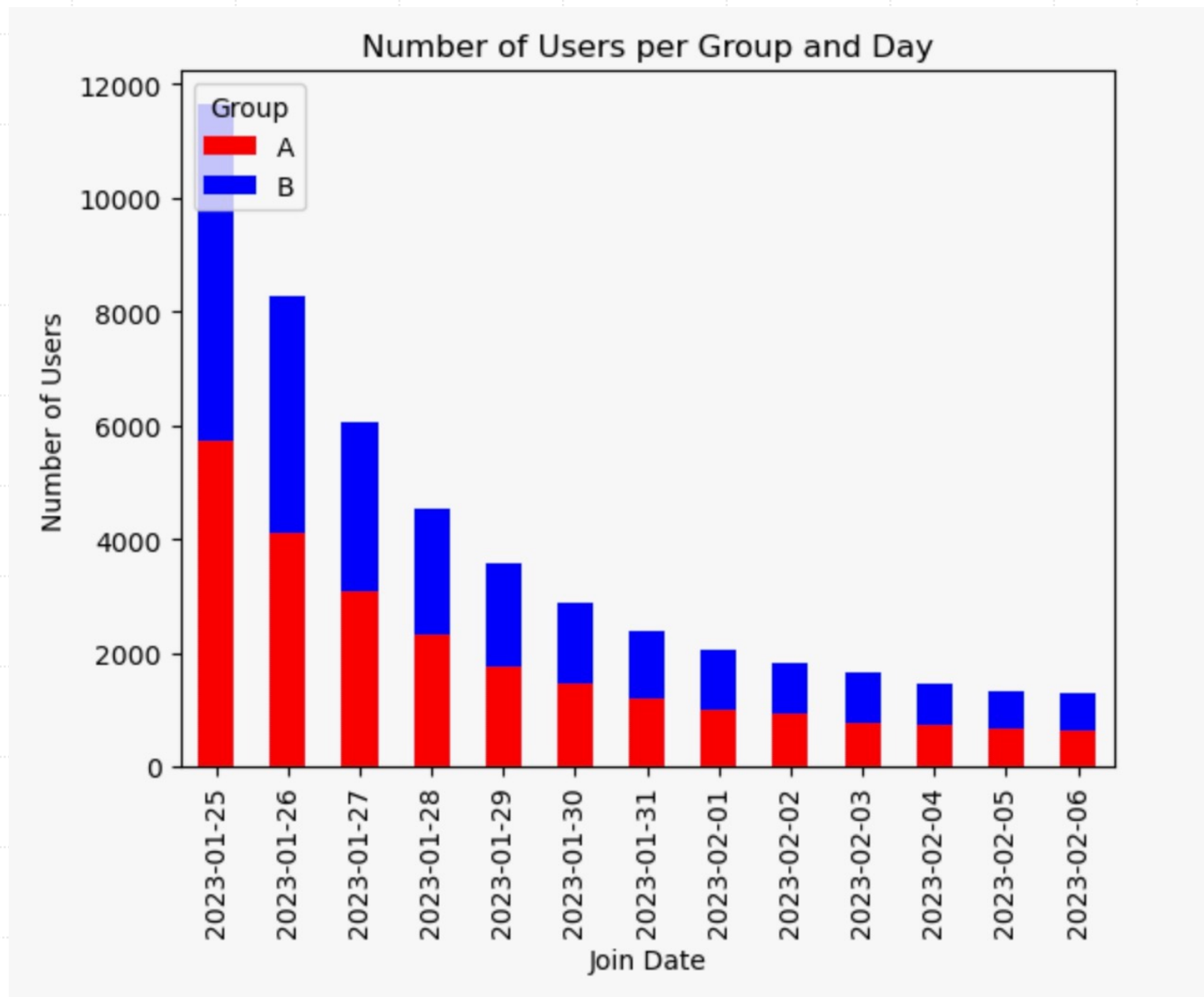


# Objective

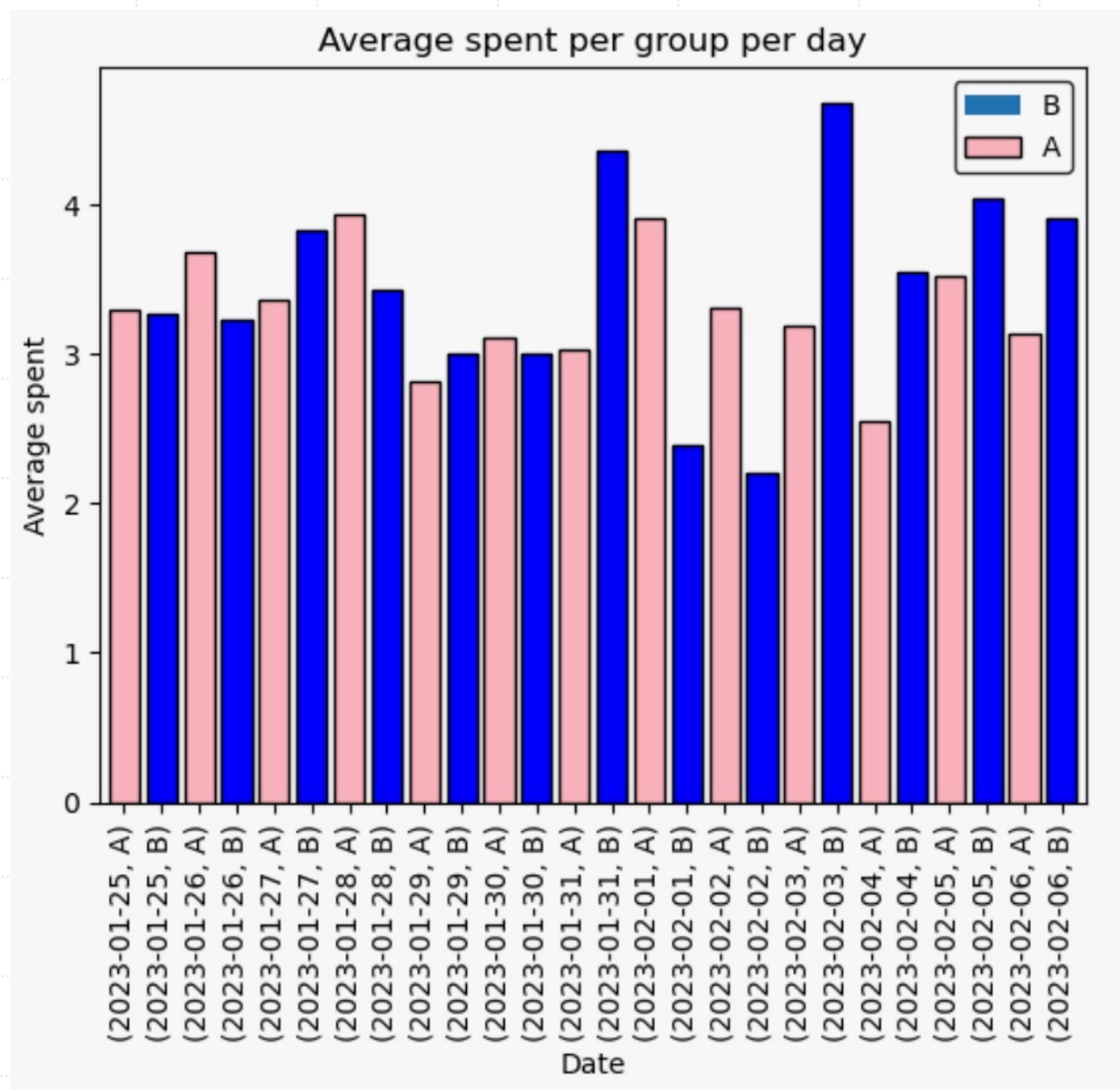
Based on the data provided, decide whether to launch the new version or not!



Number of  
visitors per day  
per group



# Average spent per day per group







# Conversion Rate

- In this project , I will use conversion rate in our hypothesis test
- *Conversion Rate* = 
$$\left( \frac{\text{Successful Conversions}}{\text{Total Visitors}} \right) * 100$$
- This is a good indicator of success rate.



# Conversion rate group A

```
1  with control as(
2      select count(distinct uid) as numcont
3      from groups as g
4      where g.group='A'
5  ),
6  numactA as(
7      select count(distinct activity.uid)
8      from activity
9      left join groups
10     on activity.uid=groups.uid
11     where groups.group = 'A'
12 )
13 SELECT CAST((SELECT * FROM numactA) AS FLOAT) /
14 (SELECT * FROM control) AS controlconvrate;
```

## Query Results

1 ROWS

**controlconvrate**

FLOAT8

3.9230990428459926

# 95% Confidence interval for group A

## Query Results

1 ROWS



**lower\_bound**

FLOAT8

**upper\_bound**

FLOAT8

3.679209516027415

4.1669885696645705

# Conversion rate group B (treatment)

```
1  with treat as(  
2      select count(distinct uid) as numcont  
3      from groups as g  
4      where g.group='B'  
5  ),  
6  numactB as(  
7      select count(distinct activity.uid)  
8      from activity  
9      left join groups  
10     on activity.uid=groups.uid  
11     where groups.group = 'B'  
12 )  
13 SELECT CAST((SELECT * FROM numactB) AS FLOAT)*100 /  
14 (SELECT * FROM treat) AS treatconvrate;
```

## Query Results

1 ROWS

**controlconvrate**

FLOAT8

4.630081300813008



# 95% Confidence interval for group B

## Query Results

1 ROWS



**lower\_bound**

FLOAT8

**upper\_bound**

FLOAT8

4.367485043175266

4.8926775584507505

# Hypothesis test

$$\hat{d} = \text{Conversion Rate}_B - \text{Conversion Rate}_A$$

Hypothesis:

$$H_0: \hat{d} = 0, \quad \hat{d} \sim N(0, SE_{pool})$$


$$H_a: \hat{d} \neq 0$$

If the following condition is true, then reject the null hypothesis. If not, then we cannot conclude anything.

$$\hat{d} > 1.96 * SE_{pool} \text{ or } \hat{d} < -1.96 * SE_{pool}$$

$$1.96 * SE_{pool} = 0.003528$$

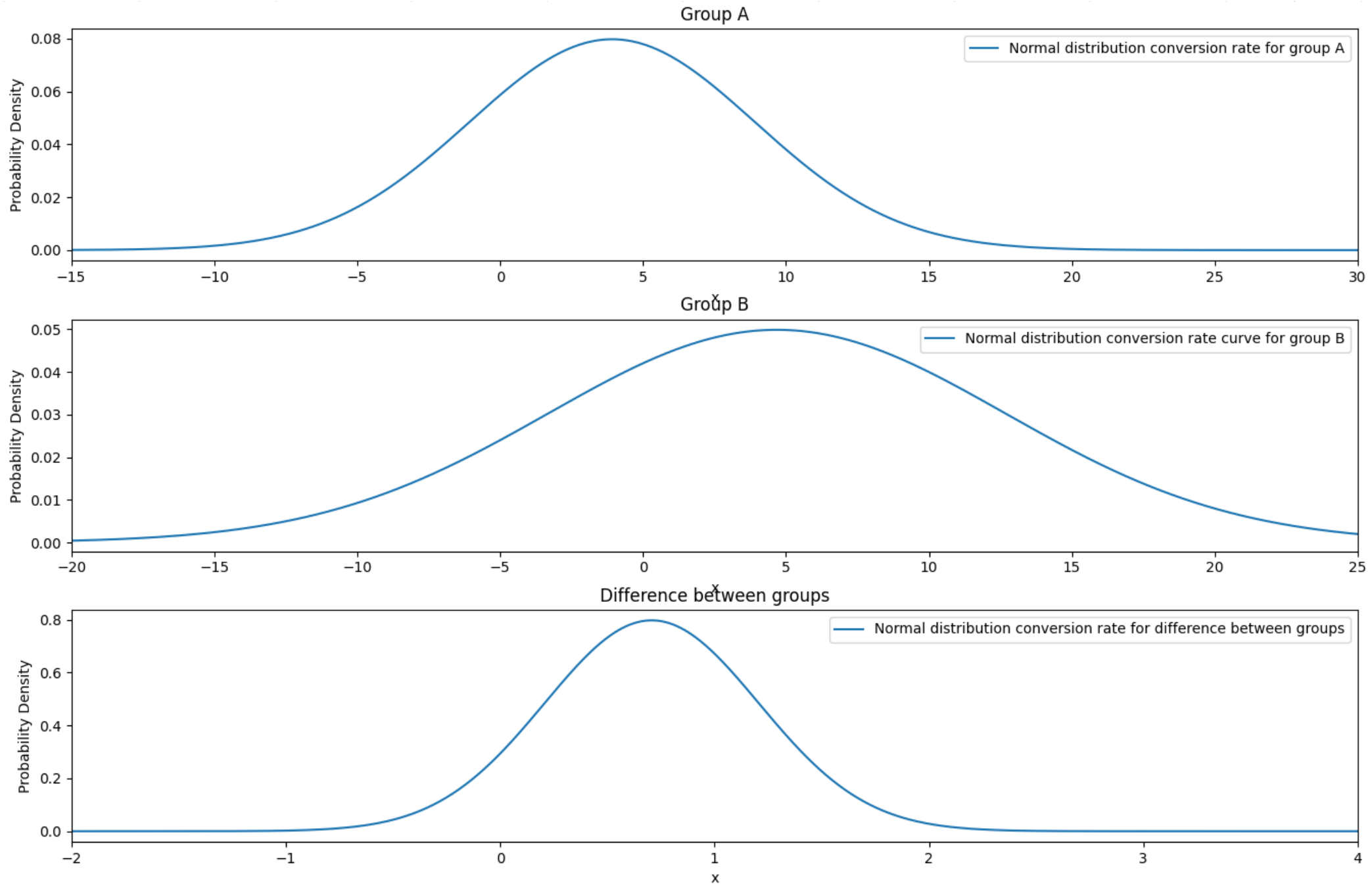
# Calculation


$$\text{Conversion rate}_B = 4.630$$

$$\text{Conversion rate}_A = 3.923$$

$$\hat{d} = 4.630 - 3.923 = 0.707$$

- We can see using normal distribution curves that the difference between conversion rates does not have an average of 0.





# Conclusion

- The calculated difference ( $\hat{d}$ ) between the conversion rates of the two groups, is found to be greater than the value of 0.003528, indicates a statistically significant distinction, reject the null hypothesis.
- Our Pandas visualization also shows evidence in favour of Group 'B'.
- So, launch the new version