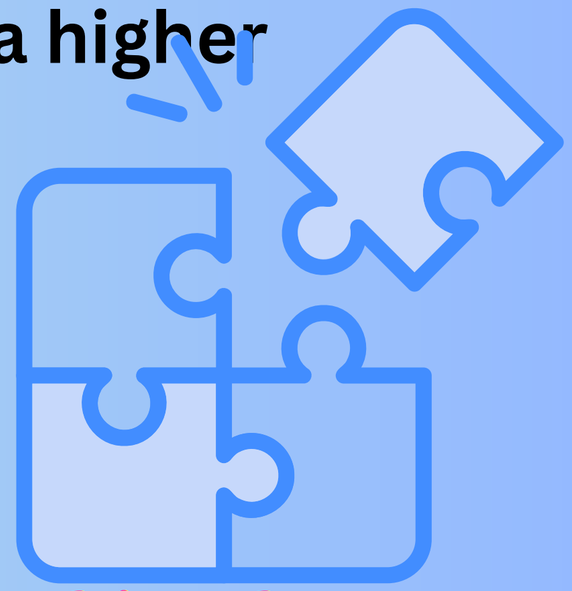
E-COMMERCE ANALYSIS

“NAVIGATING THE FUTURE OF ONLINE SHOPPING”

@Priyanshi Yadav

PROBLEM STATEMENT

An e-commerce company is evaluating two different website designs to see which one results in higher customer engagement. Design A is the current design, while Design B incorporates new features aimed at improving user experience. The company hypothesizes that Design B will lead to a higher average time spent on the website by users.



@Priyanshi Yadav

OBJECTIVE

To determine whether Design B results in a higher average time spent on the website compared to Design A.

DATASETS

current_design.csv: Contains data for user interactions with the current website design (Design A).

new_design.csv: Contains data for user interactions with the new website design (Design B).

@Priyanshi Yadav

Load the datasets

Import Necessary Libraries

```
[15]
import pandas as pd
import numpy as np
from scipy import stats as st
```

Define hypothesis

- Null Hypothesis (H0): Design B does not lead to a higher average time spent on the website compared to Design A (mean_DesignB <= mean_DesignA).
- Alternate Hypothesis (H1): Design B leads to a higher average time spent on the website compared to Design A (mean_DesignB > mean_DesignA).

1: Load the datasets

```
[16]
control = pd.read_csv("current_design.csv")
test = pd.read_csv("new_design.csv")
```

```
[17]
print(control.shape)
control.head()
```

```
(100, 2)
```

```
[18]
print(test.shape)
test.head()
```

```
(100, 2)
```

@Priyanshi Yadav

Calculate the mean and standard deviation

2: Calculate the mean and standard deviation of the time spent for both designs.

[19]
```
#control statistics

control_mean = control.time_spent_minutes.mean().round(2)
control_std = control.time_spent_minutes.std().round(2)
control_size = control.shape[0]
control_mean, control_std, control_size
```

```
(6.02, 0.62, 100)
```

[20]
```
#test statistics

test_mean = test.time_spent_minutes.mean().round(2)
test_std = test.time_spent_minutes.std().round(2)
test_size = test.shape[0]
test_mean, test_std, test_size
```

```
(8.06, 0.9, 100)
```

@Priyanshi Yadav

Test Using Rejection Region

3: Test using rejection region (i.e. critical z value)

[21]
```
a = (control_std**2/control_size)
b = (test_std**2/test_size)

Z_score = (test_mean - control_mean)/np.sqrt(a+b)
Z_score
```

```
18.66617228956536
```

[22]
```
# For a significance level of 5% (0.05) in a right-tailed test, the critical Z-value is approximately 1.645

alpha = 0.05 # significance level of 5%

critical_z_value = st.norm.ppf(1 - alpha)  # Right-tailed test at 5% significance level
critical_z_value
```

```
1.6448536269514722
```

[23]
```
Z_score > critical_z_value
```

```
True
```

@Priyanshi Yadav

Observation

As the calculated Z-score exceeds the critical value, it falls within the rejection region. Therefore, we reject the null hypothesis and conclude that Design B results in a higher average time spent on the website compared to Design A

@Priyanshi Yadav

Follow for more insights and real-world data analysis examples!"

LinkedIn

@Priyanshi Yadav