**1.1 Introduction**

This project aims to analyse whether showing advertisements leads to more user conversions using A/B testing. The dataset is split into two groups: one that saw ads ("ad") and one that did not ("no ads"). The primary objective is to determine if the ad exposure significantly affected conversion behaviour.

**1.2 Proposed Solution : - A/B Testing**

**NOTE : Prerequisites & Learning Goals**

To understand or replicate this project, you should know:

* Basics of statistics (mean, variance, distribution)
* Hypothesis testing & p-value interpretation
* Concepts of A/B testing
* Python libraries: pandas, matplotlib, seaborn, scipy
* Data visualization and interpretation

**What is A/B Testing?**

A/B Testing is a type of randomized experiment used to compare two versions of something to determine which performs better. In the context of this project:

* Group A (Control): Users who did not see any ads.
* Group B (Treatment): Users who were shown ads.

The goal is to see if Group B converts more often than Group A, and if the difference is statistically significant.

**1.2.1. A/B Testing Process has its own process: -**

1. Define the Hypothesis
   * Null Hypothesis (‘H0’): Ads do not impact conversion rates.
   * Alternative Hypothesis (‘H1’): Ads increase conversion rates.
2. Split Users Randomly into A and B groups.
3. Expose Only Group B to ads.
4. Measure the Outcome: Did users convert or not?
5. Apply Statistical Tests to evaluate if the observed difference is real or just due to chance.

**1.2.2. Statistical Concepts Explained**

**1. Hypothesis Testing:**

* It's a method of making decisions using data.
* The null hypothesis assumes no difference.
* We try to reject it using a test.

**2. P-Value:**

* A p-value measures the probability that the observed difference happened by random chance.
* A low p-value (typically < 0.05) means we reject the null hypothesis.

**3. T-Test:**

* A t-test compares the means of two groups.
* Assumes normal distribution and equal variance.
* If these assumptions are not met, use a non-parametric test like Mann-Whitney U.

**4. Mann-Whitney U Test:**

* Used when data is not normally distributed.
* Compares the rank of values between two groups.

Ads A/B Experiment Summary Report

**Dataset Overview-**

**What’s in your data?**

You have a table (df) that looks like this:

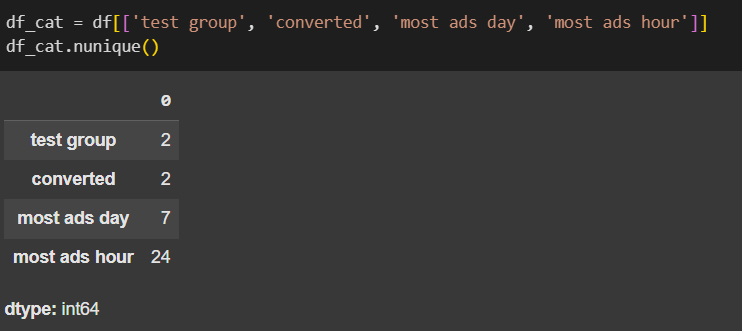
| **user id** | **test group** | **converted** | **total ads** | **most ads day** | **most ads hour** |
| --- | --- | --- | --- | --- | --- |
| 1069124 | ad | False | 130 | Monday | 20 |
| 1119715 | ad | False | 93 | Tuesday | 22 |

**Summary**

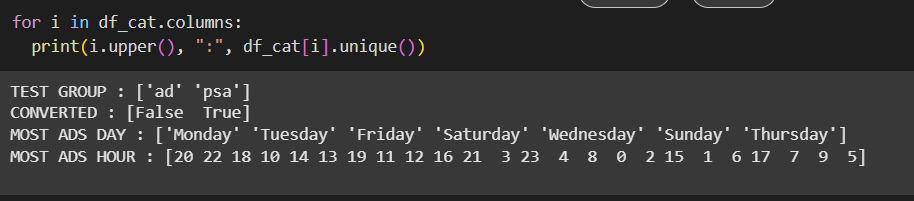
1. Two groups: "ad" (users shown ads) and "no\_ad" (users not shown ads).
2. Key variables: converted (True/False), total ads, most ads day/hour.

Let me explain each column:

* user id: Just a name/number for a person who saw ads.
* test group: Did the person see ads or not? (ad or no\_ad)
* converted: Did the person do what the company wanted (like sign up, buy)? (True or False)
* total ads: How many ads they saw.
* most ads day: The day they saw the most ads.
* most ads hour: The hour when they saw the most ads.

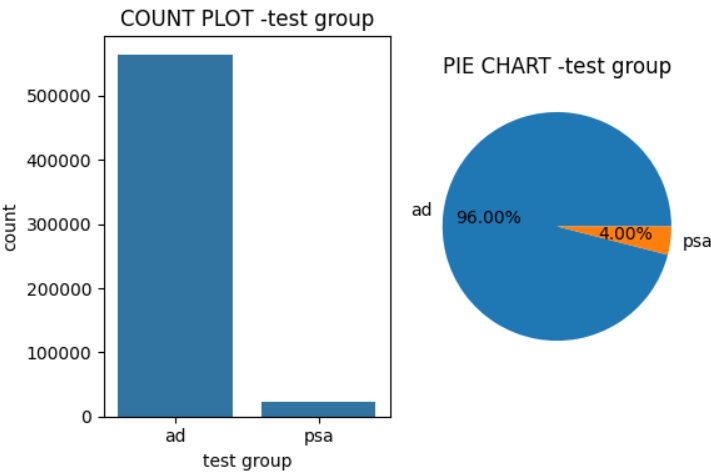
**Next Step** : Find no of unique values in each category

Find what are the type of values :



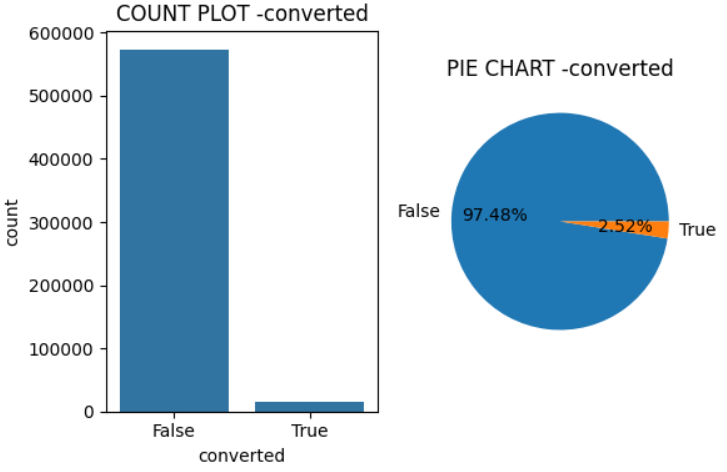
**Next Step :** Going through each plot (graph) in notebook

It shows how many people were in:



* "ad" group – saw ads
* "no ad" group – didn’t see ads

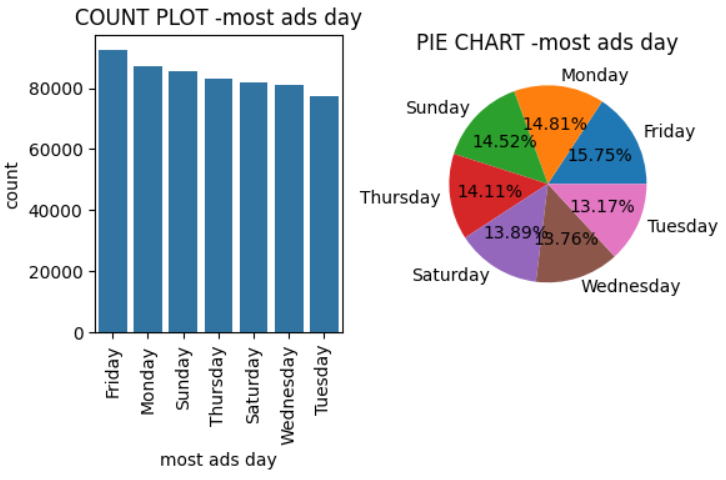
Pie Chart gives you the portion/percentage of each group like slicing a pizza.  
- To check if we have a fair number of users in both groups. (Balanced data is important.)



It shows how many people:

* converted (True) – did the thing we wanted (maybe clicked or signed up)
* didn’t convert (False)

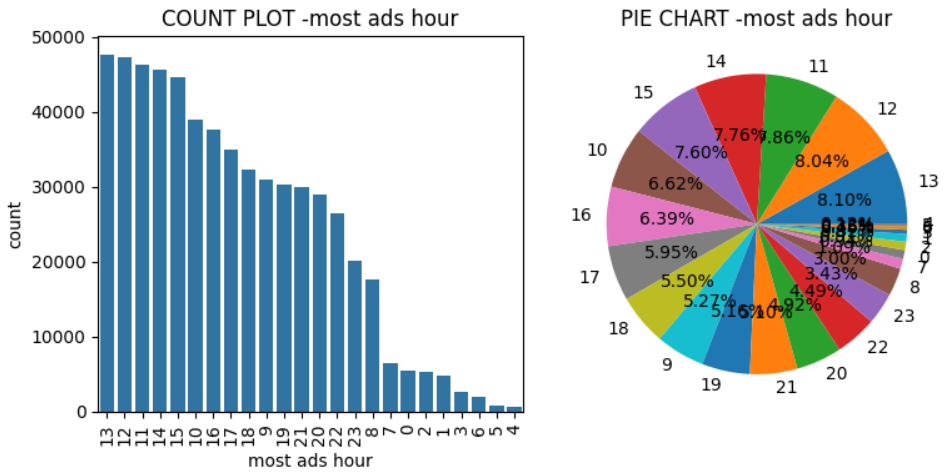
Pie Chart tells us the success vs failure rate.  
- To see if the campaign worked overall or not.



3. Most Ads Day — Count Plot & Pie Chart

It shows which day people saw the most ads — Monday, Tuesday, etc.

Pie chart tells how often each day shows up.  
To know which days ads were shown the most — maybe to plan better next time.

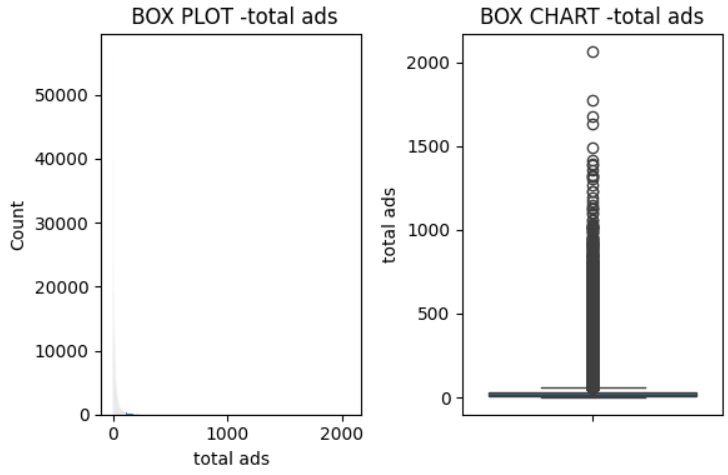


Most Ads Hour — Count Plot & Pie Chart

It shows the hour of the day when users saw the most ads (e.g., 14 = 2 PM).

Pie chart shows popular hours.

To find best times to show ads — smart timing gets more attention.

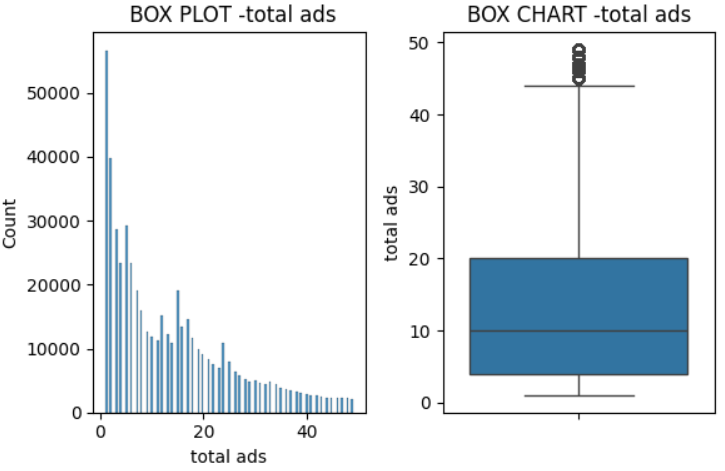


Total Ads — Histogram & Boxplot

What’s shown?

* Histogram tells how many people saw how many ads.
* Boxplot shows average, minimum, maximum, and if there are outliers.

Why?  
To understand ad exposure — are some people seeing too many ads?



6. Total Ads < 50 — Zoomed In

Same as above, but only for people who saw fewer than 50 ads (to ignore extreme cases).

Zooming in helps see the pattern better if too many extreme values are messing up the graph.

1. Test Group Count / Pie Plot

Conclusion:  
You likely had a fairly balanced number of users in both the "ad" and "no\_ad" groups. That’s good — it means your test setup was fair.

2. Converted Count / Pie Plot

Conclusion:  
You probably saw more users did not convert than those who did. This is common in most campaigns — the majority don’t take action.

3. Most Ads Day

Conclusion:  
One or two days (like Monday or Tuesday) probably showed up more. This tells you that on those days, ads were shown more often — maybe due to scheduling or traffic spikes.

4. Most Ads Hour

Conclusion:  
Certain hours (like 20 = 8 PM or 22 = 10 PM) were peak times for showing ads. Useful insight if you want to target ads better in the future.

5. Total Ads Histogram & Boxplot

Conclusion:

* Some users saw very few ads, some saw a lot.
* Boxplot likely showed outliers — people who saw way more ads than others.
* The average number of ads might be moderate, but the distribution is not symmetric.

6. Zoomed Plot: Total Ads < 50

Conclusion:  
Most users likely saw under 20–30 ads. This zoom helps you see the real pattern without being distracted by extreme values.

7. Test Group vs Converted Plot

Conclusion:  
This is the most important:

* If both "ad" and "no\_ad" groups had similar conversion rates, then:

Ads didn’t really make a difference.

* If "ad" group had higher conversion, then:

Ads helped.

* But based on the test you did (Mann-Whitney), the result was **not significant**.

**Final Overall Conclusion (from all plots):**

* People in both groups (ad vs. no\_ad) behaved **similarly**, and the **number of ads they saw didn’t really change** whether they converted or not.

**2. What Do the Statistical Tests Say?**

* You ran the **Mann-Whitney U test**, which is used when data isn’t normal.
* It gave a **p-value of 1.0** → which means:

“No statistically significant difference between the ad and no-ad groups.”

**Statistical message**:

“Nope, no real difference. Showing ads didn’t help conversions.”

**Final Answer: Yes, They Say the Same Thing!**

Both the **visual plots** and the **math tests** are telling you the same story:

**Showing ads didn’t increase conversions in this data.**  
People behaved pretty much the same whether they saw ads or not.

This is great — it means your analysis is consistent, and your conclusion is strong.