

# Assessment Math & Statistics module

---

PRESENTED BY  
**Haiyan Qin**





And by the way,  
I had to trick myself into  
believing I could be good at  
math and stats.

# Agenda

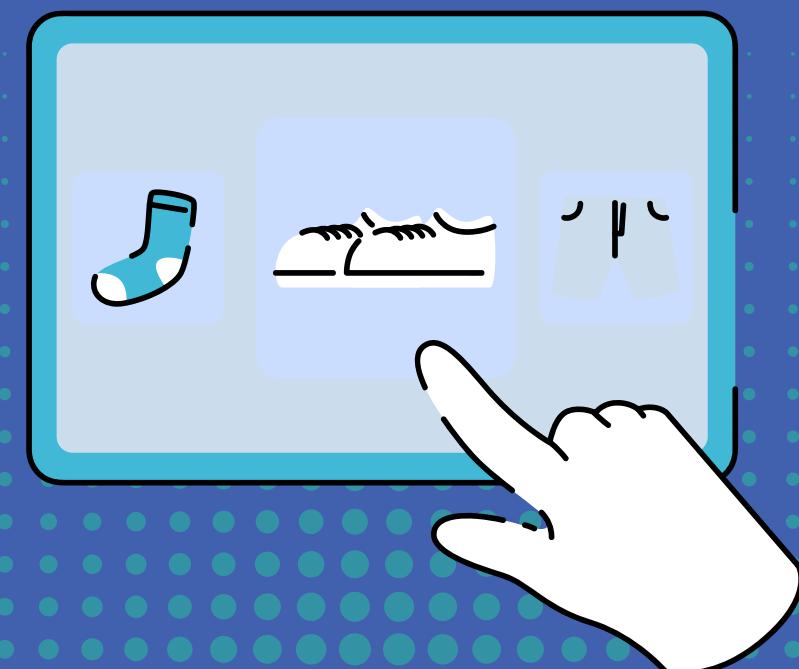
PRESENTED BY  
**Haiyan Qin**

- 1 INTRODUCTION
- 2 HYPOTHESIS
- 3 METRICS
- 4 EDA
- 5 STATISTICAL TEST & INFERENCE
- 6 STRATEGY
- 7 CONCLUSION

## CHAPTER 1

# Introduction

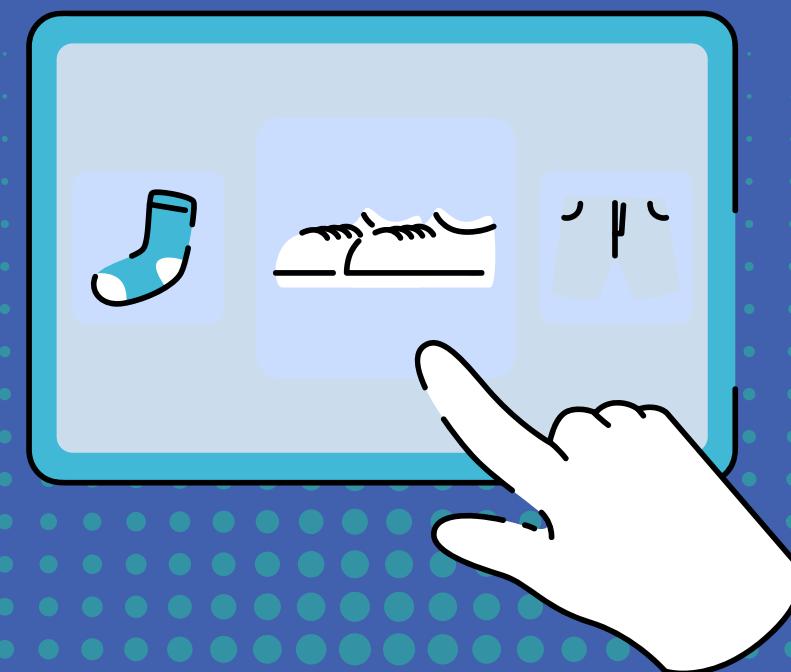
Our cooking equipment e-commerce website aims to enhance user engagement, boost sales, and improving the overall user experience by optimizing its product pages. As part of this pursuit, we have initiated an A/B test to assess the potential impact of switching the media rail layout from horizontal to vertical across all product pages.



## CHAPTER 1

# Introduction

This A/B test serves as a fundamental methodology for evaluating the performance and effectiveness of the proposed layout change in relation to key performance metrics essential for enhancing user engagement and driving conversions.



# Benefits

---

- Raise Profile Visibility
- Enhance User Engagement
- Optimize Space Utilization
- Improve Aesthetic Allure
- Ensure Multi-Device Compatibility

# Drawbacks

---

- User Adaptability Concerns
- Risk of Cluttering Pages
- Impact Page Performance & Conversions
- Accessibility and Inclusivity Barriers

CHAPTER 2

# Hypothesis





We predict that by switching the media rail layout to a vertical orientation on product pages will **increase user engagement** on pages, resulting in **more clicks on media**, more page views, a **longer average time spent on the pages**, and ultimately, a **higher likelihood of product sales**.

## **Null Hypothesis (H0)**

The repositioning of the media rail will not affect user engagement.

## **Alternative Hypothesis (H1)**

The repositioning of the media rail does affect user engagement.

# Why A/B test ?

Through an A/B test is the most suitable method for testing our hypothesis:

- Direct comparison
- Controllable
- Statistical
- Real-Time Insights
- Scalability

# Alternative Methods ?

- User Survey / Interview
- User Ability Test
- A/Bn Test / Multivariable Test
- Bayesien Test
- Bendit Test
- Heuristic Analysis

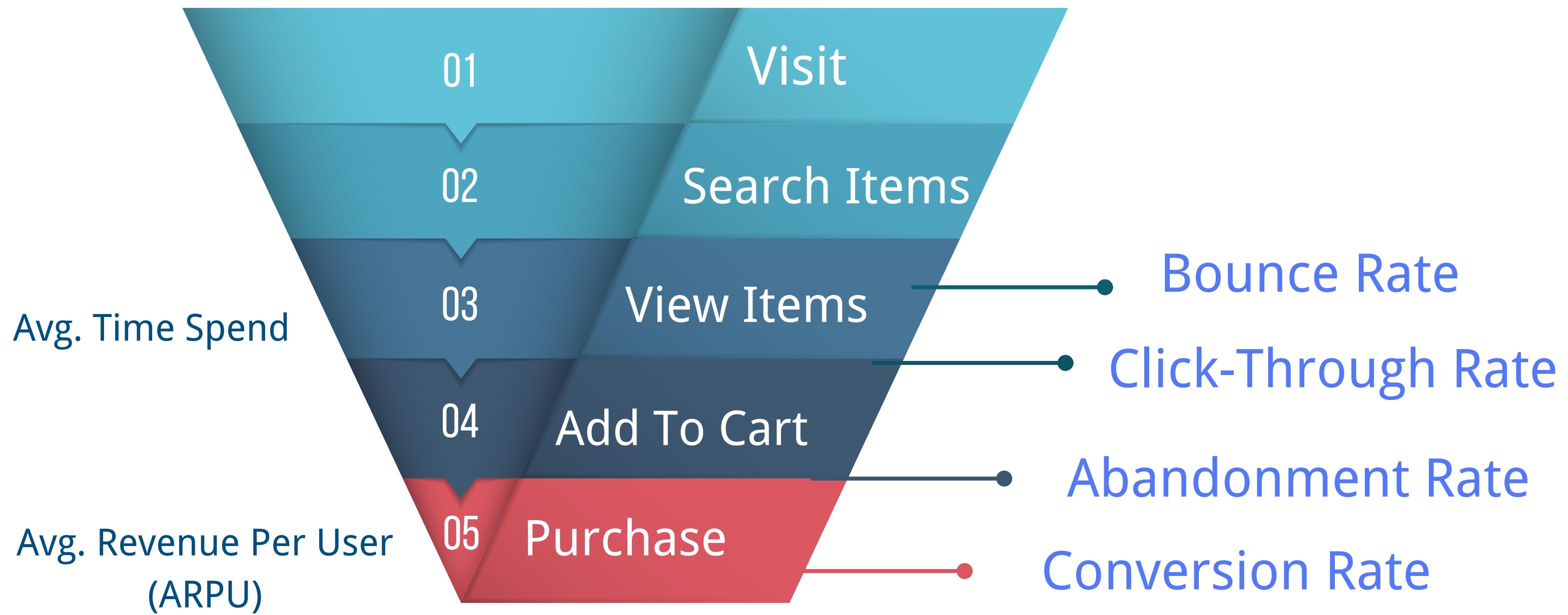
CHAPTER 3

# All Metrics



# Ideally User Flow

from Visit to Purchase



## ◆ PRIMARY METRIC

- Media Content Interaction
  - ['Interacted'] & ['Clicks on media']

## ◆ SECONDARY METRICS

- # of Page views or ['Churned']
- Conversion/GMV
- Avg. Time Spent
- Shopping cart abandonment

## ◆ MONITORING METRIC

- Avg. Revenue Per User

CHAPTER 4

# Exploratory Data Analysis



# From the given data

- **Variant:** Control group A and Treatment group B.
- **Number of page views:** How many times the product pages was viewed.  
(Apply for all products.)
- **GMV (in \$):** Gross Merchandise Value, likelihood to total sales.
- **Number of add to cart:** How many times products were added to the cart.
- **Clicks on media:** The number of clicks on media content.
- **Time on Page (sec):** The amount of time spent on the page, in seconds.
- **user\_id:** A unique identifier for each user.

**UNIT OF RANDOMIZATION**

# Metrics I wish to have from data:

.....

- **Timestamp:**
  - track user activity over time
  - identify trends
- **User segment multi-dimensional :**
  - targeted analysis
  - user behaviour across different segments
- **Returning user vs. new user:**
  - user retention and acquisition dynamics
- **Demographic information:**
  - user persona
- **Number of items bought:**
  - purchasing behaviour
  - inventory management
  - sales trends



# Converted Data

## (Group Summary)

The original information that has been refined and structured for easier use. This involves e.g. cleaning to remove errors, standardising formats for consistency, and aggregating data to a more manageable level for analysis.

Variant	Total Users	Total # of Clicks On Media	Total Purchase Amount \$	Total # of Page Views	Total # of Users Who Add Items to Cart	# of Users Who Complete Purchases	Total Time Spent on Page (sec)	# of Interacted Users	# of Churned Users
A	1000	1495	7689.99	4985	857	80	20543	780	46
B	1000	1324	9289.53	5317	896	97	20047	728	28

\*\*All presented data will be rounded to 2 decimal places.

# Key Performance Indicators

# MATTERS



## To Explore More...

- **Bounce Rate(Churn Rate)**

The percentage of users who view less than and equal 1 page.

- **Shopping Cart Abandonment Rate**

The percentage of users who indicated purchase intent by adding items to the cart but did not complete the purchase.

- **Conversion Rate(CR)**

The percentage of users who completed purchases.

- **Average Revenue Per User(ARPU)**

The average amount spent per user.



# KPIs



- Bounce Rate(Churn) =  $\frac{\text{\# of Users who view less than 2 pages}}{\text{Total \# of Users}} \times 100\%$
- Conversion Rate (CR) =  $\frac{\text{\# of Users Who Complete Purchases}}{\text{Total \# of Users}} \times 100\%$
- Average Revenue Per User =  $\frac{\text{Total Purchase Amount}}{\text{\# of Users Who Complete Purchases}} \times 100\%$   
(ARPU)
- Abandonment Rate =  $\frac{\text{\# of Users Who Abandoned Shopping Cart}}{\text{\# of Users Who Added Items to cart}} \times 100\%$

# KPIs

Variant	Bounce Rate %	Abandonment Rate %	Conversion Rate %	ARPU \$	Avg.Time Spent (sec)
A	4.6	90.67	8	96.13	20.54
B	2.8	89.17	9.7	95.77	20.05

\*\*All presented data will be rounded to 2 decimal places.

Overview	Variables	Interactions	Missing values	Sample
----------	-----------	--------------	----------------	--------

Overview	Alerts (7)	Reproduction
----------	------------	--------------

[Churned](#) is highly imbalanced (77.2%)

Imbalance

[Variant](#) is uniformly distributed

Uniform

[user\\_id](#) is uniformly distributed

Uniform

[user\\_id](#) has unique values

Unique

[GMV \(in \\$\)](#) has 1823 (91.1%) zeros

Zeros

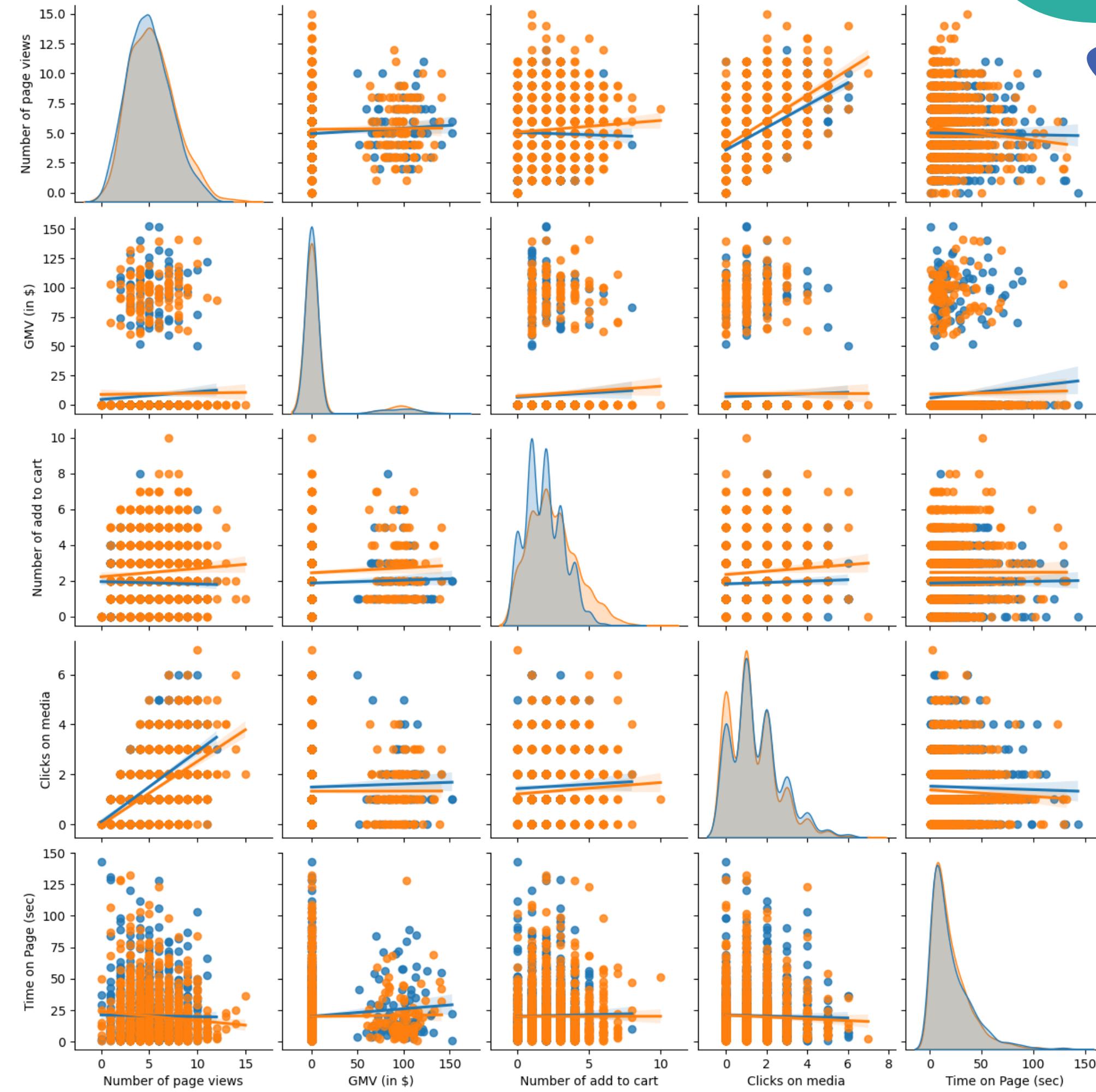
[Number of add to cart](#) has 247 (12.3%) zeros

Zeros

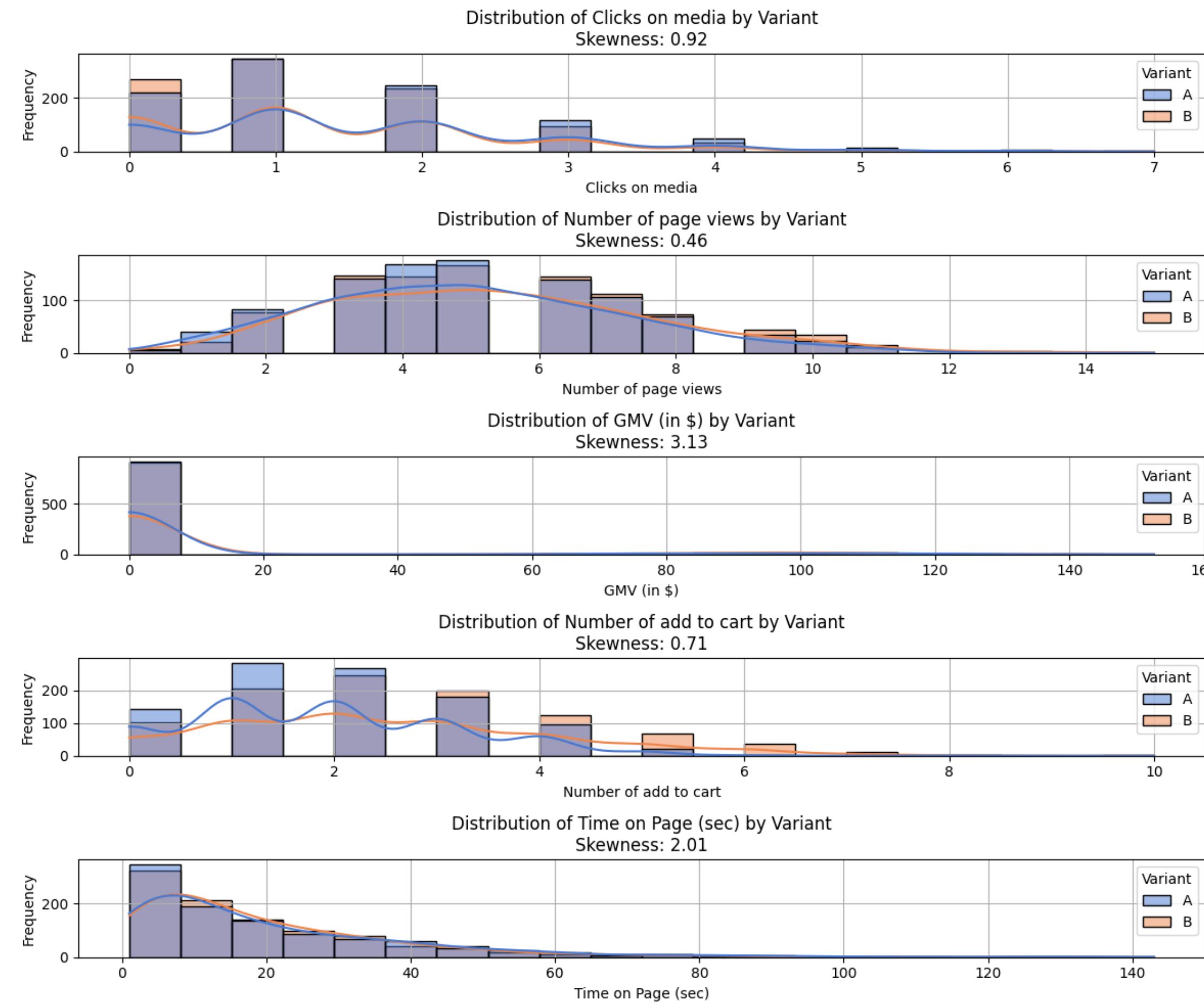
[Clicks on media](#) has 492 (24.6%) zeros

Zeros

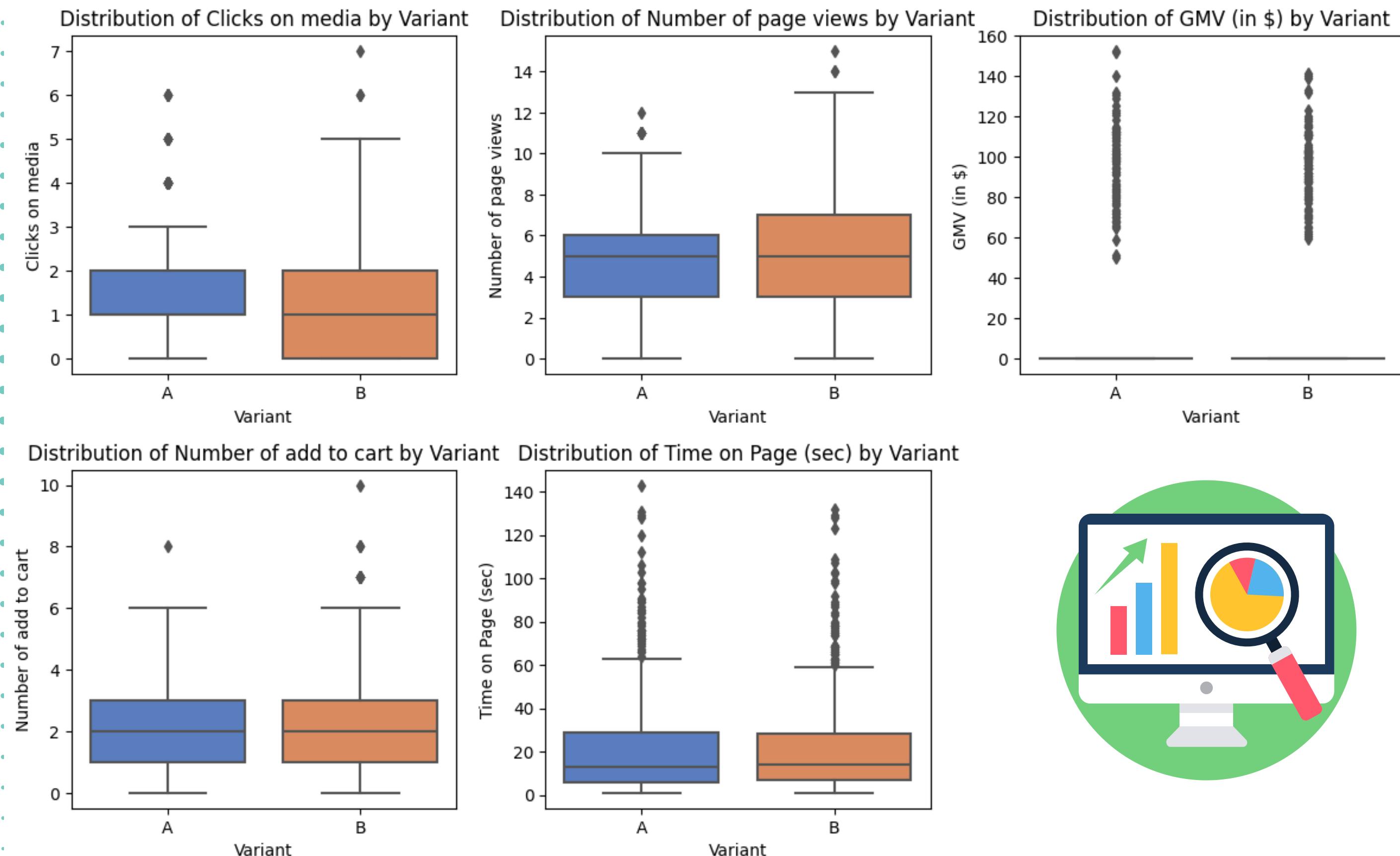
# Quickly identify patterns and relationships between variables



# Are columns distribution differs between A and B?

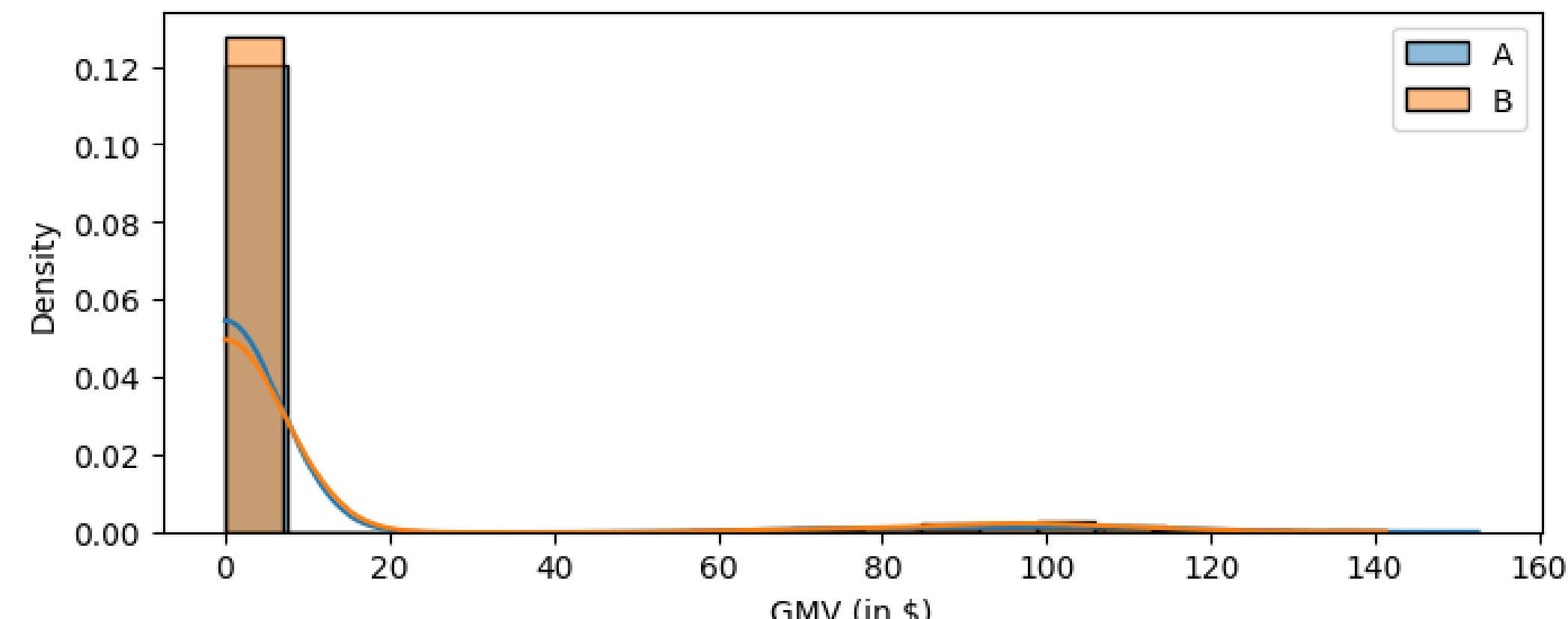


# Are columns distribution differs between A and B?

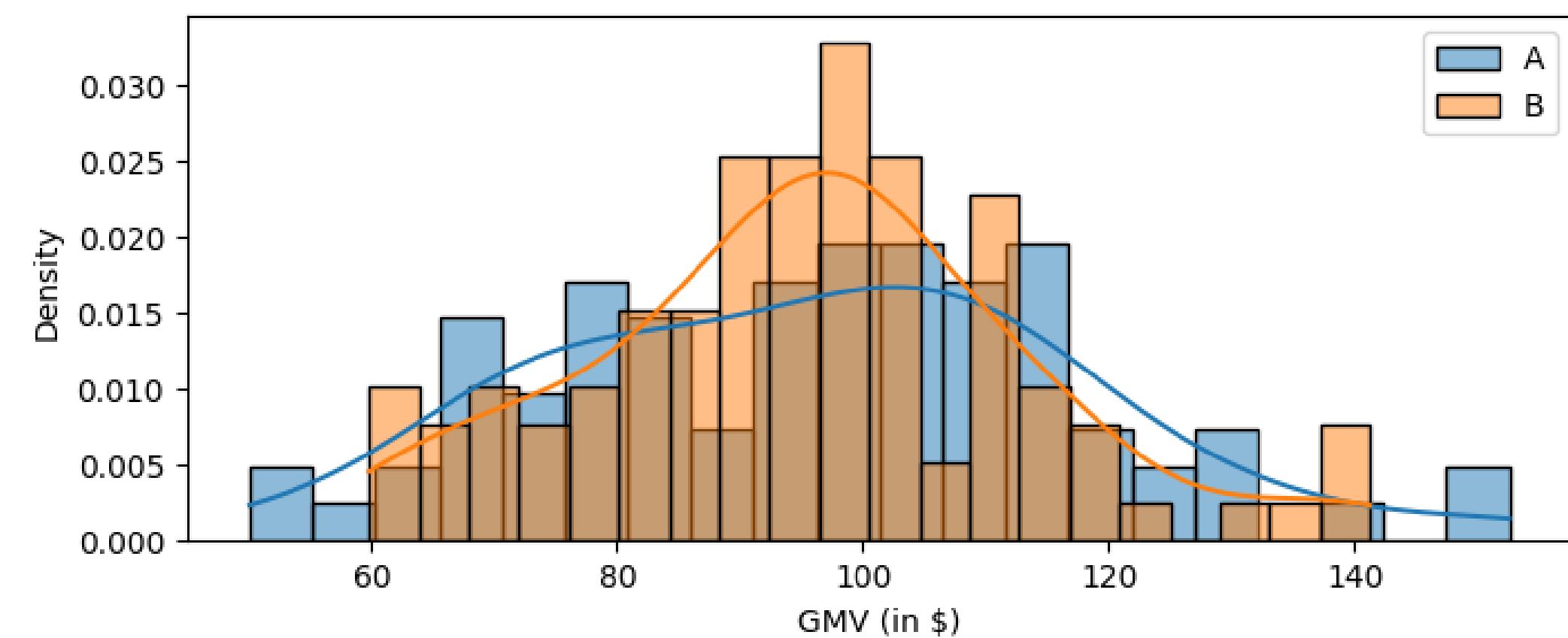


# Are columns distribution differs between A and B?

Distribution of GMV of All Users



Purchased Users GMV distribution



CHAPTER 5

# Statistical Test & Inference



# T-rrific !!!

---

- Two Sample Test
- Sample Size:  
1000 for each Variant
- Two-tailed Test

## ['Interacted']

### Chi-square Test

- statistic: 7.01139
- P-value: 0.00810
- Degrees of freedom: 1
- Expected frequencies table:
- [[246.754.] [246.754.]]

The association between 'Variant' and 'Interacted' is statistically significant.

p-value <  $\alpha$ , We reject  $H_0$ .



## ['Clicks on media']

### T- Test

- T-statistic: 3.17107
- P-value: 0.00154

The difference on 'Clicks on media' between A & B is statistically significant.

p-value <  $\alpha$ , We reject  $H_0$ .



## Conversion/GMV

- T-statistic: 0.11826
- P-value: 0.906020

There is no statistically significant difference in purchases between A and B.



\*\*All presented data will be rounded to 5 decimal places.

## Page Views

- T-statistic: -3.18816
- P-value: 0.00145
- The difference in Page Views between A and B is statistically significant.



## Average Time Spent

- T-test for Time on page (sec):
- T-statistic: 0.55543
- P-value: 0.57866

There is no significant difference in 'Time on page (sec)' between A and B.

## Abandonment

- Z-statistic: -1.0357366825435268
- P-value: 0.30032500397917594

There is no statistically significant difference in abandonment between Variant A and Variant B.

## Average Revenue Per User

- T-statistic: 0.11826
- P-value: 0.90602

There is no statistically significant difference in ARPU between A and B.



\*\*All presented data will be rounded to 5 decimal places.

# Factors Influencing Test Result

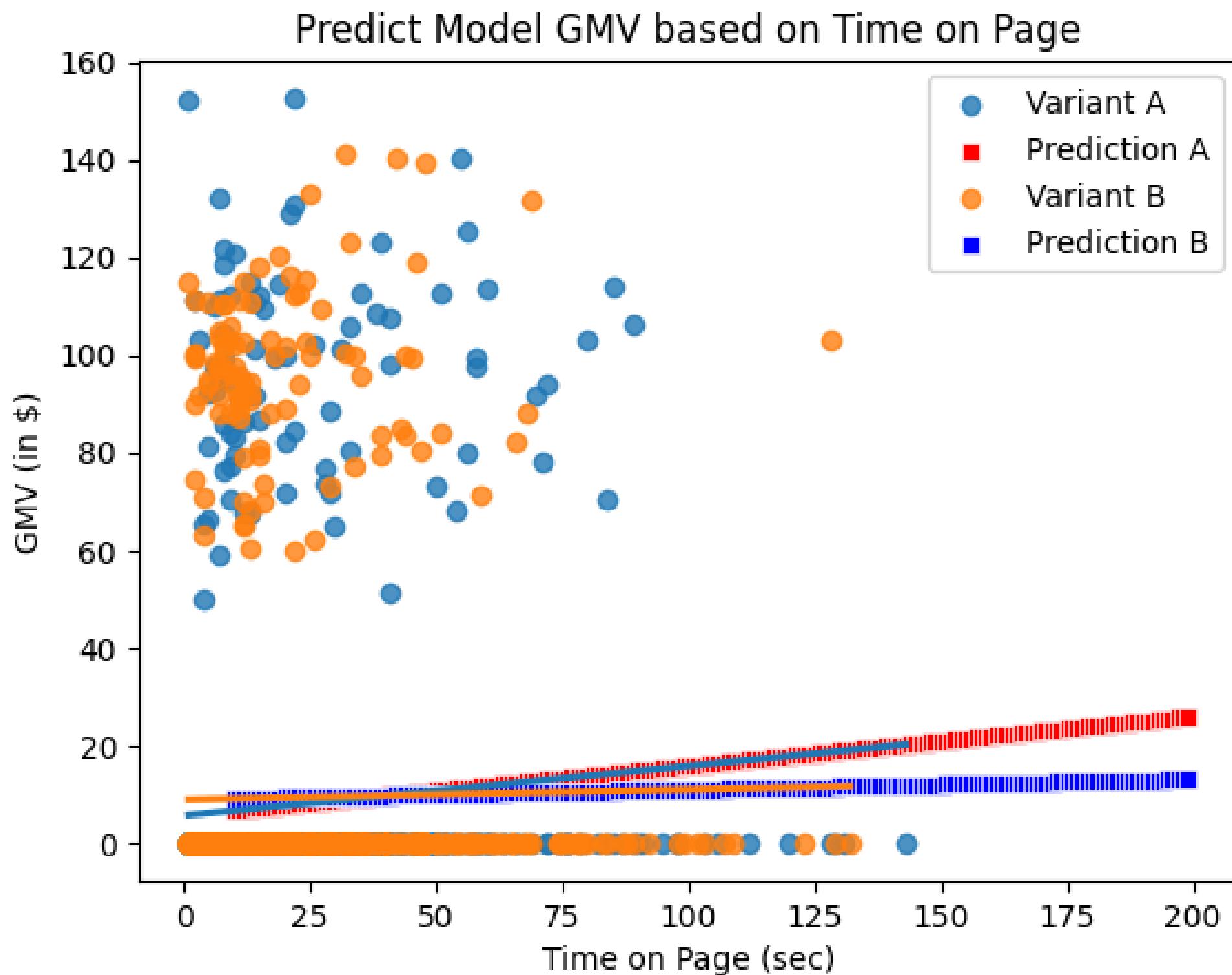
- Treatment effect
- Point estimate
- Primacy & Novelty Effect
- Heterogeneity
- Covariates
- Test duration
- Traffic volume





*For killing some curiosity,  
let's try fit some regressions...*

# Regression Analysis Results



Regression analysis helps in identifying which independent variables significantly impact the dependent variable, allowing businesses to allocate resources more effectively based on the most influential factors.

CHAPTER 6

# Strategy





# **Statistically Significant**



# **Practical Significance**

## Enhanced Engagement

- Variant B: More media interaction and lower bounce rate, suggesting increased user engagement.

## Potential Revenue Increase

- Variant B: Better conversion rate performance indicates users may make larger purchases, boosting average order value.

## Conversion Prospects:

- Variant B: Decreased abandonment rate hint at improved conversion rates.

## Customer Retention:

- Variant A: Users spend more time on the page, slightly higher average revenue per user, potentially enhancing customer retention.

**Insights**



# Average Revenue Per User

---

**A decrease in ARPU in Variant B would suggest that while the media rail may impact engagement, it does not positively influence, and may even detract from, revenue-generating actions.**

# Trade-offs Consideration

## A or B ?



# Recommendation

---

- Reposition the media layout.
- Continuously analyze data for customer journey.
- Gathering feedback from users and stakeholders.
- Conduct A/B testing on different website elements.
- Segment data to analyze varying user behaviors.
- Perform qualitative research, such as surveys and user interviews, for deeper insights
- Ensure the design meets the needs and preferences of the target audience.
- Analyze and optimize user engagement.



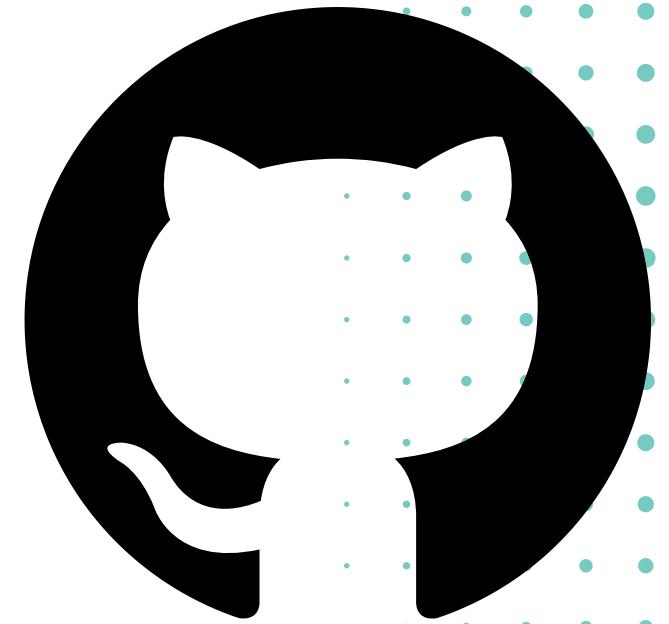
# Conclusion

---

**By implementing the vertical media rail, our website can potentially enhance the user experience, increase engagement, and drive higher conversions, leading to improved business outcomes.**



.ipynb  
available  
on  
**Github**



<https://github.com/HiYannnn/A-B-Test-Ecommerce>



# Thank You

---

PRESENTED BY

**Haiyan Qin - DA25**