



# CONDÉ NAST

## PRODUCT ANALYTICS DATA CHALLENGE

*Luyao (Doris) Liu*

[Luyao.Liu03@gmail.com](mailto:Luyao.Liu03@gmail.com)

+1 585 404 6403

Getting Comfortable  
With The Data  
Structure

Part I

Understanding  
Behavior

Part II

Data Exploration

Part III

*Agenda*

## *Part I: Getting comfortable with the data structure*

The logo for GLAMOUR magazine, featuring the word "GLAMOUR" in a red, stylized, serif font.

- 4087 pageviews happened in August for Glamour.

The logo for Condé Nast Traveler, featuring the words "Condé Nast" in a small, black, sans-serif font above the word "Traveler" in a large, bold, black, serif font.

- 7749 people saw a video on CNT.

The logo for SELF magazine, featuring the word "SELF" in a large, bold, black, sans-serif font.

- 67.09% of SELF's sessions are bounces.
- Tablet has the highest bounce rate, 68.08%.

Device_Type	Bounce_Rate
Desktop	66.93
Mobile	67.07
Tablet	68.08

Getting Comfortable  
With The Data  
Structure

Part I

Understanding  
Behavior

Part II

Data Exploration

Part III

*Agenda*

## Part II: India market has highest Subscription Rate

### Subscriptions/ Unique Users:

- India market has highest subscription rate, 27.13%.
- China is the second, and Italy is the third.
- Middle East has lowest subscription rate, 24.65%.



### Subscriptions/ Sessions:

- India market has highest subscription rate, 5.45%.
- Italy is the second, and Spain is the third.
- Middle East has lowest subscription rate, 4.74%.

Market	Subscription_Rate_1	Subscription_Rate_2	Unique_Users	Sessions
India	27.13	5.45	1183	5887
China	26.50	5.22	1147	5826
Italy	26.31	5.33	1144	5644
Spain	26.19	5.31	1111	5482
U.S.	25.53	5.04	1093	5535
U.K.	24.68	4.95	1082	5397
Middle East	24.65	4.74	1140	5923

Subscription\_Rate\_1: Subscriptions / Unique Users

Subscription\_Rate\_2: Subscriptions / Sessions



**Unique Users is not an accurate metric since it's cookie-based.**



*If he/she uses multiple browsers, devices or clears up cookies regularly, he/ she can be assigned with multiple User Ids.*



### Which market's subscription order form is the best?

- ❑ First, **we can't compare market's subscription order forms by subscription rate, and it doesn't help business decisions.**
  - Brand awareness, competitors of the market, and other factors that affect subscription rate **vary in different markets.**
- ❑ Second, if the true question is **how to improve a specific market's subscription rate by improving the order form.**
  - Design various versions of the order form by following conversion-centered improvement (CTA), UX rules of thumb, etc.
  - Conduct the **randomized experiment** to find out which order form contributes to a highest subscription rate.

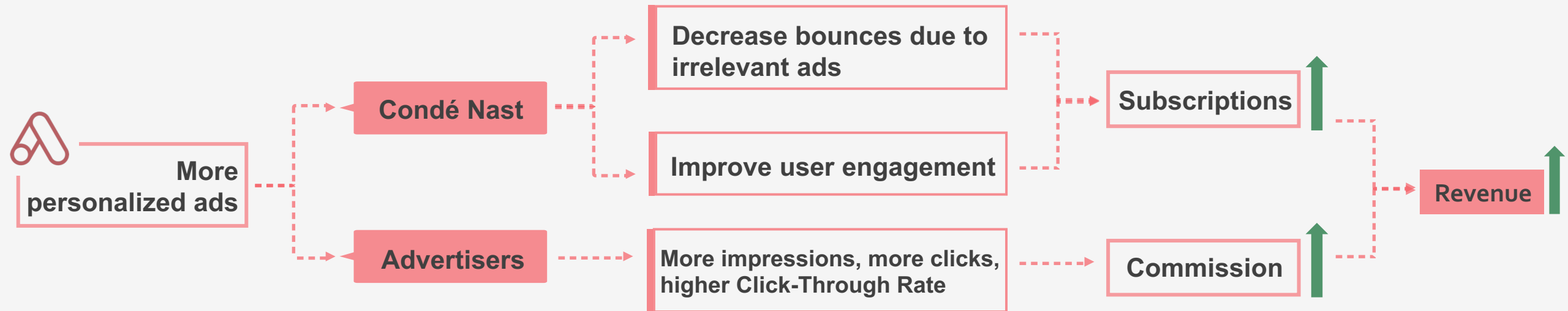
Select an Offer

<b>One-Year Subscription</b> 1 Year of Condé Nast Traveler + 2 free gifts! <a href="#">LEARN MORE</a> \$5	<b>Two-Year Subscription</b> 2 Years of Condé Nast Traveler + 2 free gifts! <a href="#">LEARN MORE</a> \$10
	
<a href="#">SELECT</a>	<a href="#">GREAT CHOICE!</a>

Is this for you or someone else?

<a href="#">FOR MYSELF</a>	<a href="#">FOR A GIFT</a>
----------------------------	----------------------------

## *Part II: The goal of more personalized ads is to increase revenue*



## *Observational data can't be used to evaluate the new marketing strategy*

□ *To evaluate the new marketing strategy with observational data, we have to make assumptions to continue the analysis:*

- There is no **seasonality** issue before and after 09/01/2019.
- The group of **users** is **similar** before and after 09/01/2019.
- The marginal effect of a marketing strategy **wouldn't diminish and be stable** in its lifecycle.



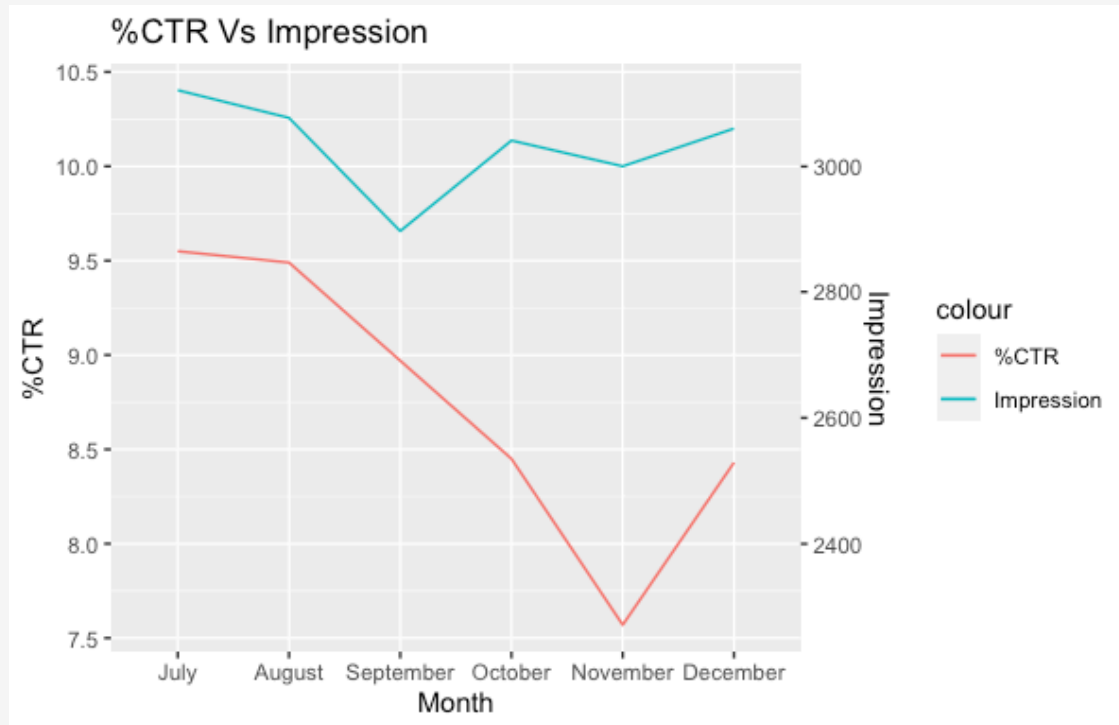
The **assumptions above can't all be true** in real life.  
We can't evaluate the new marketing strategy based on observational data.



We should run **A/B test** to find out how much effectiveness of the new marketing strategy improve or decrease.



*\*The analysis is only used to demonstrate analytical thinking since it's based on observational data, and it should be based on A/B test result data.*



07/01/2019 ~ 08/31/2019 Overall CTR: **9.88%**

09/01/2019 ~ 12/31/2019 Overall CTR: **8.35%**

- When based on A/B test data, we calculate overall CTR of each group. (control group, treatment group)
- Also, break down to compare MoM CTR of two groups, we'll know which months contribute most or least to the overall CTR.

## Part II: Key metrics to evaluate the new marketing strategy

*\*The analysis is only used to demonstrate analytical thinking since it's based on observational data, and it should be based on A/B test result data.*



06/01/2019 ~ 08/31/2019



09/01/2019 ~ 12/31/2019

More personalized ads

Lift

### Primary Metrics

Click-Through Rate (of the ads):  
9.88%

Click-Through Rate (of the ads):  
8.35%

**-15.46%**  
*p-value: .9999*  
*not statistically significant*

Subscription Rate (by sessions):  
3.66%

Subscription Rate (by sessions):  
3.66%

0

### Secondary Metrics

% of Sessions to continue after  
seeing an ad:  
66.33%

% of Sessions to continue after  
seeing an ad:  
65.05%

**-1.93%**  
*p-value: .9448*  
*not statistically significant*

Average events per session after  
seeing an ad:  
2.43

Average events per session after  
seeing an ad:  
2.38

**-1.77%**  
*p-value: .2285*  
*not statistically significant*

# Pitchfork

## Part II: Monthly Cohort Retention Analysis

*\*Assumption1: From June to Dec 2019, each cohort users are new users, never visit CNT before June 2019.*

*\*Assumption2: The duplicated users are of same percentage in each cohort in each time period.*

*(Due to the inaccuracy of unique users metric)*

	cohort	cohort_size	month0	month1	month2	month3	month4	month5	month6
1	June 2019	3074	100	98.86	97.86	96.87	92.78	86.87	69.29
2	July 2019	1515	100	97.62	96.69	93.99	87.65	69.19	NA
3	August 2019	730	100	93.29	94.71	85.58	72.83	NA	NA
4	September 2019	340	100	90.59	85.71	67.80	NA	NA	NA
5	October 2019	174	100	74.14	72.87	NA	NA	NA	NA
6	November 2019	77	100	45.45	NA	NA	NA	NA	NA
7	December 2019	43	100	NA	NA	NA	NA	NA	NA

- *How to read the table?*
  - *Cohort Size: How many users start 1<sup>st</sup> session on each month.*
  - *Month0: The percentage of how many users stay on the month they start to get active.*
  - *Month1: The percentage of how many users stay on the second month compared to the previous month active users.*
  - ...

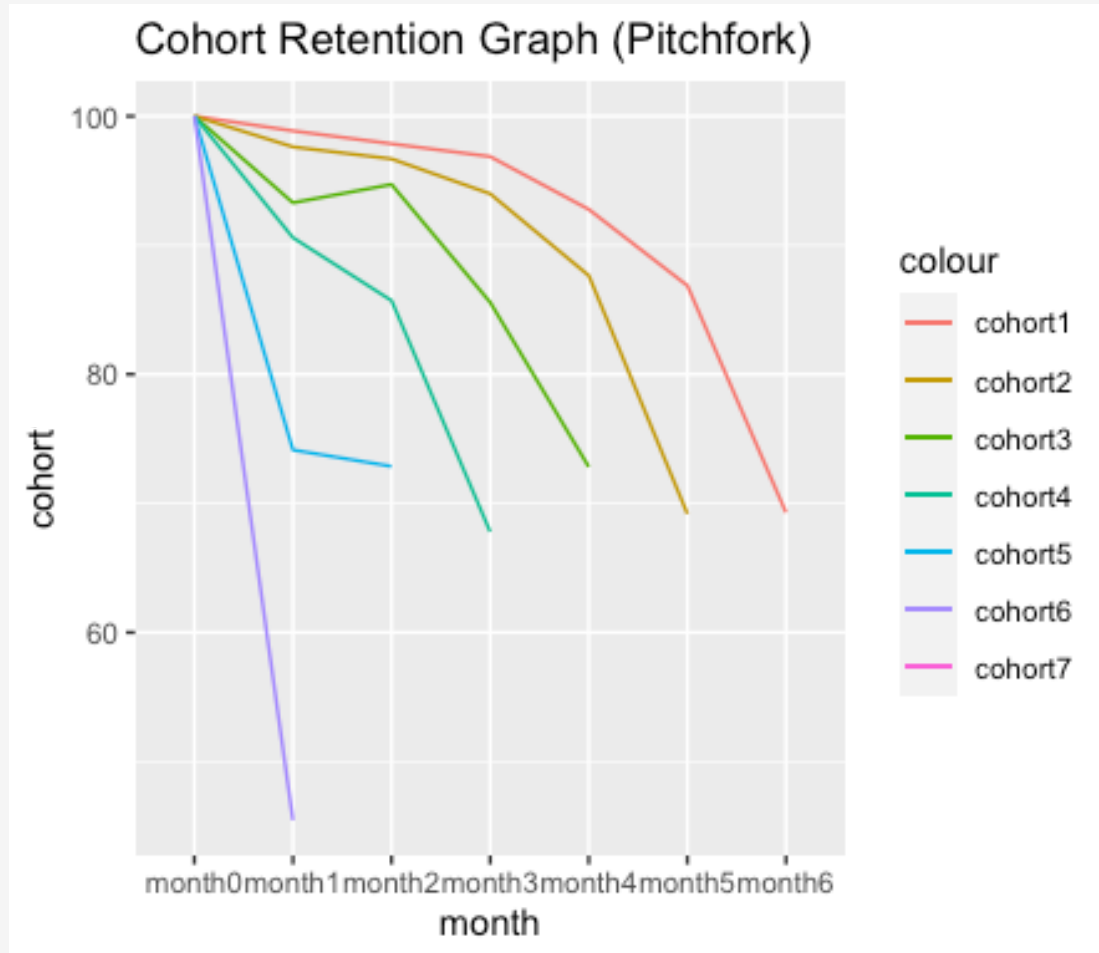
# Pitchfork

## Part II: Monthly Cohort Retention Analysis

*\*Assumption1: From June to Dec 2019, each cohort users are new users, never visit CNT before June 2019.*

*\*Assumption2: The duplicated users are of same percentage in each cohort in each time period.*

*(Due to the inaccuracy of unique users metric)*



- *How to read the graph?*

*For each month in which various cohort users start to active, how fast each cohort drop out users in the following months.*

- **Across cohorts, the dropout rate of cohort is getting faster.** The biggest reason in this situation is that assumption 1 can't be true in real life. We need more information (Boolean Variable: New/ Returning User) to verify it's true that the users we're getting are less loyal.
- **Within a cohort, the dropout rate is getting faster in month2/ month3.** Still, what the graph reflects may not be accurate because assumption 1 can't be true. But assume it's true, we should take more action during the 2nd or 3rd month to pull users back.

Getting Comfortable  
With The Data  
Structure

Part I

Understanding  
Behavior

Part II

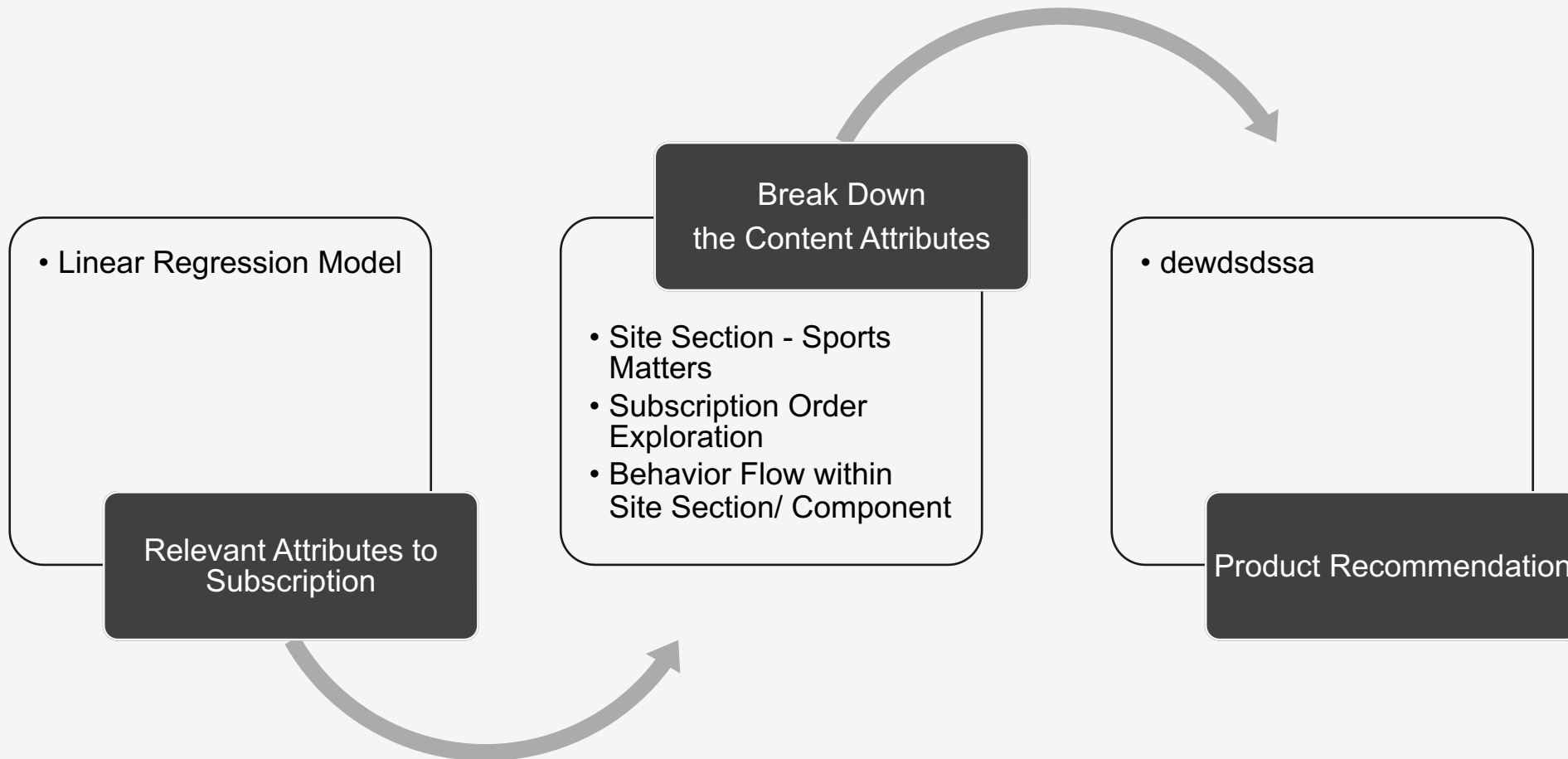
Data Exploration

Part III

*Agenda*

❑ Objective:

Explore the user behavior that's highly related to subscription event for Condé Nast Traveler and make product recommendations based on findings.



## Part III: Device, Site Section and Month are important for subscription

- **Linear Regression Model:**

To find out what variables are important to Subscription, and whether it's statistically significant.

```
Response: Subscription
      Df Sum Sq Mean Sq F value    Pr(>F)
Month   6   0.42  0.069593   5.5054 1.035e-05 ***
Device   2   0.22  0.111418   8.8142 0.0001487 ***
Component_Name  4   0.01  0.001819   0.1439 0.9657221
Site_Section   3   0.21  0.071107   5.6252 0.0007500 ***
Residuals 140895 1781.02  0.012641
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- According to ANOVA Analysis Mean Sq, **Device**, **Site Section** and **Month** are three variables that are important for subscription event, and they're statistically significant.

*Dataset: Pick out the data whose brand name is CNT, and delete all events after subscription event.*

*Dependent variable: Subscription (dummy variable: 0,1)*

*Independent variables: Month, Device, Component, Site Section.*

## Part III: The portrait of high potential subscribers

- **Linear Regression Model:**

To create the portrait of the users who have highest potential of subscription.

Coefficients:

(Intercept)	MonthDecember	MonthJuly	MonthJune	MonthNovember
1.177e-02	1.584e-03	-3.865e-03	-1.656e-03	1.242e-03
MonthOctober	MonthSeptember	DeviceMobile	DeviceTablet	Component_NameArticle
-6.084e-04	-2.300e-04	1.969e-03	-2.361e-03	2.845e-04
Component_NameGallery	Component_NameHomepage	Component_NameVideo	Site_SectionFood	Site_SectionHomepage
-3.837e-04	-7.829e-04	1.833e-07	7.528e-04	1.201e-03
Site_SectionSports				
2.962e-03				

- A user visits CNT in **December**, on a **mobile** device, and interested in **Sports section**, then he/ she's of high possibility to subscribe.



## Part III: Sports Section is important for subscription

### ○ Deep dive into Site Section & Subscription:

From the content perspective, what's the difference between subscribers and all users.

Brand	Conde Nast Traveler			
Count of Session_Id	Event_Name			
Site_Section/ Component_Name	Impression	Pageview	Click	Grand Total
<b>Food</b>	<b>25713</b>	<b>17113</b>	<b>2515</b>	<b>45341</b>
Advertisement	12969		1318	14287
Article		8571		8571
Gallery		8542		8542
Video	12744		1197	13941
<b>Entertainment</b>	<b>25561</b>	<b>17016</b>	<b>2501</b>	<b>45078</b>
Advertisement	12798		1247	14045
Article		8561		8561
Gallery		8455		8455
Video	12763		1254	14017
<b>Sports</b>	<b>25539</b>	<b>16979</b>	<b>2435</b>	<b>44953</b>
Advertisement	12791		1190	13981
Article		8614		8614
Gallery		8365		8365
Video	12748		1245	13993
<b>Homepage</b>	<b>3153</b>	<b>2087</b>	<b>299</b>	<b>5539</b>
Advertisement	1559		151	1710
Homepage		2087		2087
Video	1594		148	1742
<b>Grand Total</b>	<b>79966</b>	<b>53195</b>	<b>7750</b>	<b>140911</b>



- From the previous slide, the importance of section for subscription:  
Sports > Homepage > Food > Entertainment
- While the count of events of section:  
Food > Entertainment > Sports > Homepage

### - Why the difference?

Users who read Sports section are more likely subscribers than those who read Food section.

### - Hypothesis:

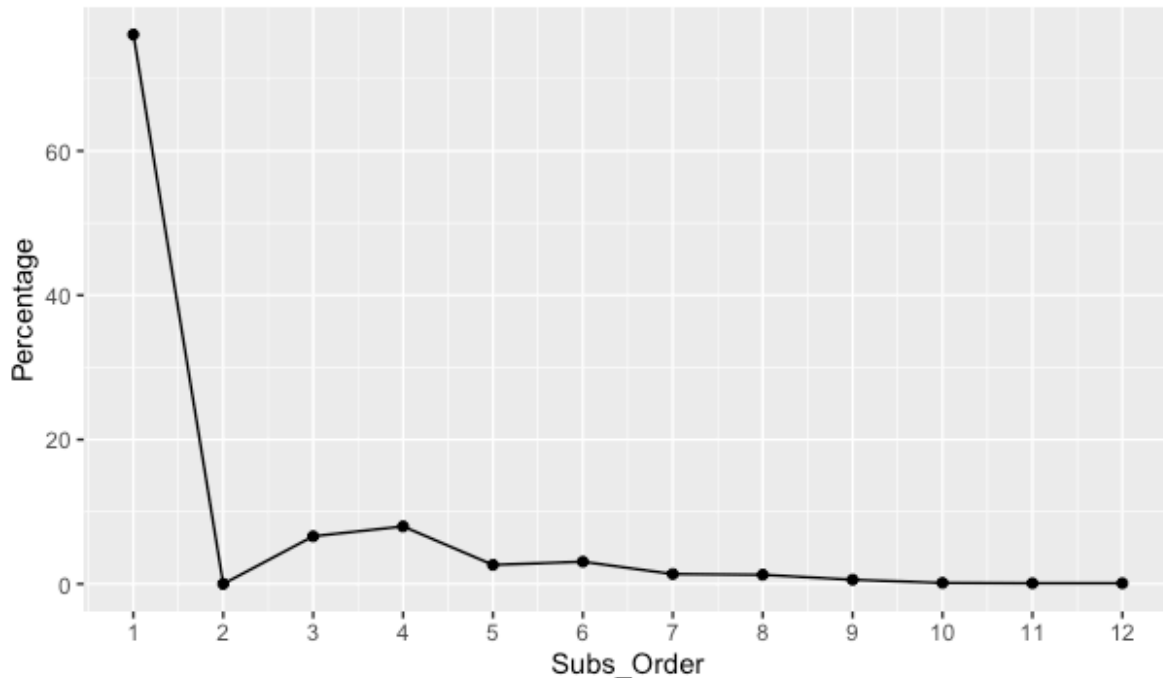
- (1) Sports is the differentiation selling point for a travel magazine.
- (2) Sports section reader has a higher propensity.
- (3) ...

Dataset: Pick out the data whose brand name is CNT, and delete all events after subscription event.

## Part III: 76% Subscription occur in the 1st event of a session

- **Deep dive into Site Section & Subscription:**  
*Explore in what order Subscription event normally occur.*

The percentage of Event Order for Subscription (CNT)

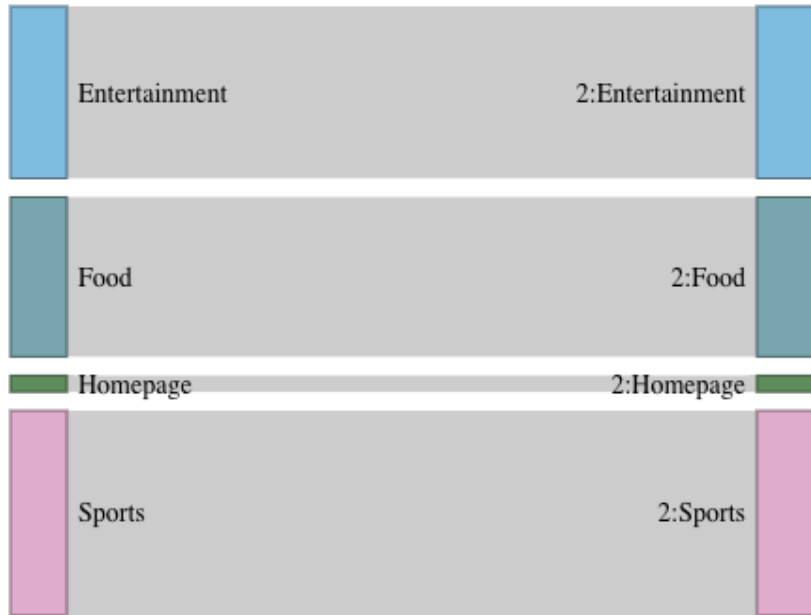


- **76% Subscription occur in the 1st event of a session.**  
*(Hypothesis: Call-To-Action campaigns target effectively and directly convert the target into customers.  
We can test the hypothesis by analyzing acquisition channels/ campaigns of the sessions.)*
- **0 % Subscription occur in the 2nd event of a session.**  
*(There could exist data issue, we should follow up to QA.)*
- **6% occur in the 3rd, and 8% occur in the 4th.**  
*(Please refer to the next slide to see the flow of Site section/ Component before Subscription occur.)*

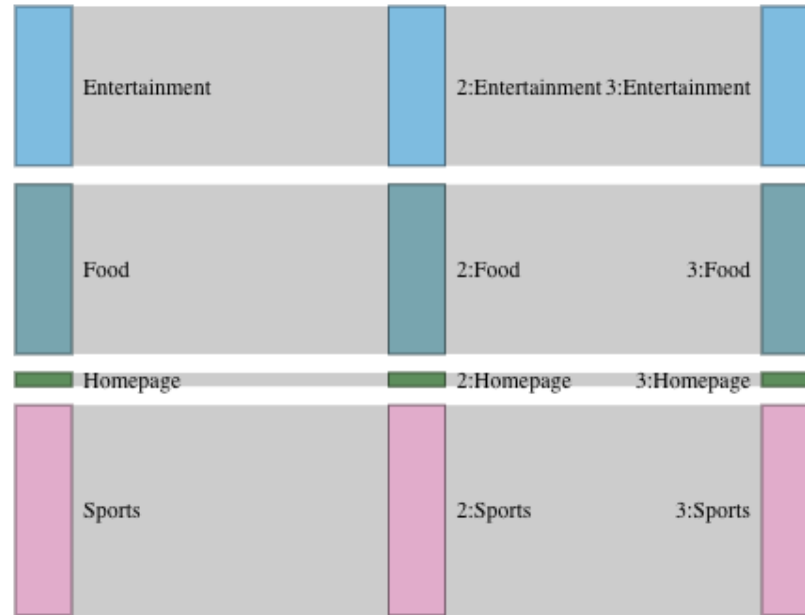
## Part III: Subscribers are clear about what they want in CNT

- How's the flow of Site Section like before Subscription happen on 3<sup>rd</sup> or 4<sup>th</sup> event?

6% Subscription occur in the 3<sup>rd</sup> event of a session.



8% Subscription occur in the 4<sup>th</sup> event of a session.

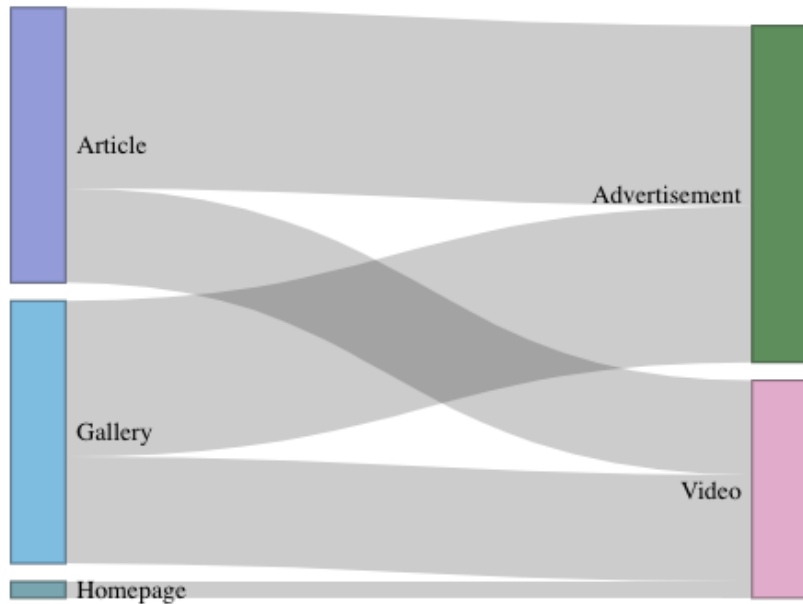


- Before users subscribe, they tend to **stay in the section where they enter in**. Our users are clear about what they want in CNT.
- Most subscribers read Sports section, and then Entertainment, food.

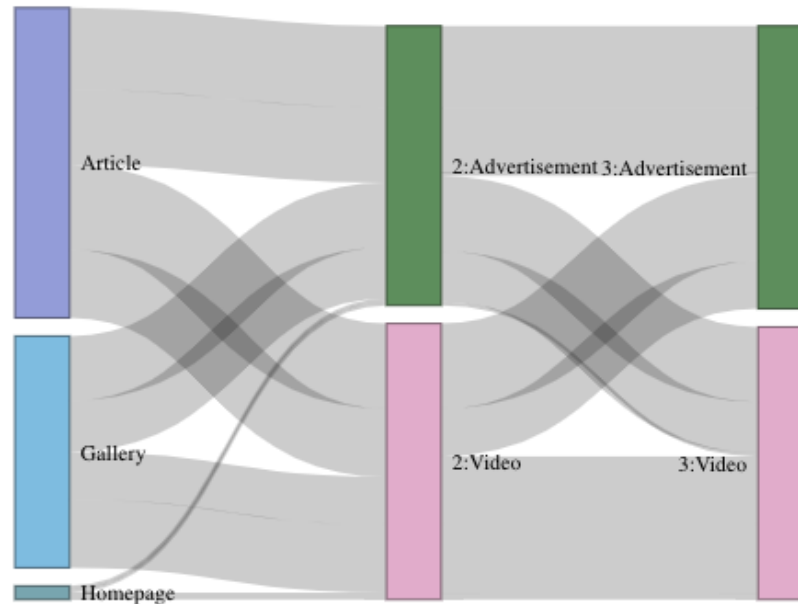
## Part III: Subscribers normally read articles/gallery in CNT

- How's the flow of Component like before Subscription happen on 3<sup>rd</sup> or 4<sup>th</sup> event?

6% Subscription occur in the 3<sup>rd</sup> event of a session.



8% Subscription occur in the 4<sup>th</sup> event of a session.



As for content component perspective, users normally read article or gallery first.

### Relevant Attributes

- Device: Mobile
- Site Section: Sports



- ❑ Improve mobile UI/UX
- ❑ A/B test: Sports – Selling Point

### Break Down the Content Attributes

- Subscription usually happen in 1<sup>st</sup>, 3<sup>rd</sup> and 4<sup>th</sup> event of a session.
- Users tend to stay in the site section where they enter in before subscription.
- Users normally enter the site in article/ gallery component before subscription.



- ❑ Site Section: Sports, Entertainment, Food
- ❑ Content Component: Article/ Gallery