

# From Creative to Conversion: Examining the Relationship Between Online Display Advertising and Customer Conversion for an International Travel & Tourism Company

March 13, 2015

## Data at a Glance:

- Observation Window: Dec 2014 – Jan 2015
- 300 million individual impressions
  - includes details about creative, campaign, and site placement
- 22 million engagements
  - link to specific impression with which user interacted
- 300 million conversions
  - descriptions for ~40 types of conversions users can make

# WEBINAR WILL BEGIN SHORTLY

Presentation recording and slide deck will be made available on the WCAI website by Monday, March 16<sup>th</sup>

Submit your proposal by **Monday, March 30, 2015**. Learn more at <http://wcai.wharton.upenn.edu/research>



////////////////////////////////////  
KNOWLEDGE FOR ACTION  
////////////////////////////////////

# From Creative to Conversion: Examining the Relationship Between Online Display Advertising and Customer Conversion for an International Travel & Tourism Company

A WCAI Research Opportunity  
sponsored by Annalect

March 13, 2015



---

# Introductions

## **Erin Matts**

Chief Marketing Officer, Annalect

## **Anna Nicanorova, WG'11**

Data Scientist, Annalect

## **Melissa Hartz**

Assistant Research Director, WCAI

## **Colleen O'Neill**

Executive Director, WCAI

## **Elea Feit**

Senior Fellow, WCAI



# Data Solutions for Marketers, By Marketers



annalect

# Four Pillars of Data-Driven Marketing

We provide clients with a data driven marketing strategy, powered by a connected system of technology, tools, consultants and activation.



## New Technology

Automated systems creating smarter and more efficient processes



## Evolved Tools

Web-based reporting and visualization enabling deeper understanding and better informed decisions



## Established Consultants

Experts providing stronger and faster insights into audience, marketplace and media



## Dynamic Activation

Data informed activation delivering more effective investments and stronger audience insights

# Large Scale, Total Access and Secure

## Scale

**15  
Billion**

Cookies  
through our  
DMP

**450**

Analysts,  
developers,  
data scientists

**45**

Global  
Markets

## Access

**250+**

Private and 3rd  
party data  
providers

**\$1.6B**

Largest spend in  
programmatic

**1 of 2**

Data platforms  
accepting FB  
and Amazon

**24/7**

When you have  
access to your  
own data

## Security



Team of  
specialists are  
assigned to  
monitor, update  
and harden  
security



Congressional  
and governance  
team focused on  
protecting and  
informing clients  
around changes  
in data policy

---

# Possible Research Areas from Annalect

## Path-to-Purchase

- Improving attribution models
- Understanding advertising effects on path-to-purchase
- Predicting conversion based on browsing behavior

## Other Areas

- Detecting returning users who are browsing on a different device, or who have cleared their cookies
- Differential efficacy of creative by site



---

# Possible Research Areas from Annalect

## Path-to-Purchase

- Improving attribution models
- Understanding advertising effects on path-to-purchase
- Predicting conversion based on browsing behavior

## Other Areas

- Detecting returning users who are browsing on a different device, or who have cleared their cookies
- Differential efficacy of creative by site
- Something else *not on these lists...*



---

# Possible Research Areas from Annalect

## Path-to-Purchase

- Improving attribution models
- Understanding advertising effects on path-to-purchase
- Predicting conversion based on browsing behavior

## Other Areas

- Detecting returning users who are browsing on a different device, or who have cleared their cookies
- Differential efficacy of creative by site
- Something else *not on these lists...*

**NOTE:** With **hundreds of millions of exposures**, this dataset provides the opportunity for in-depth research. We encourage researchers to keep the considerable size of the data in mind as they submit proposals.



---

# Where Do These Data Come From?

Users are observed **viewing and clicking on ads for an expensive, highly considered product** (a trip) from an international travel & tourism company.

There are a **variety of “conversions”** users can carry out including, but not limited to, **landing on the homepage, exploring itineraries, and receiving a booking confirmation email.**

Devices are **tracked by cookies** over a period of **two months** from December 2014 to January 2015.



---

# How are the Data Collected?

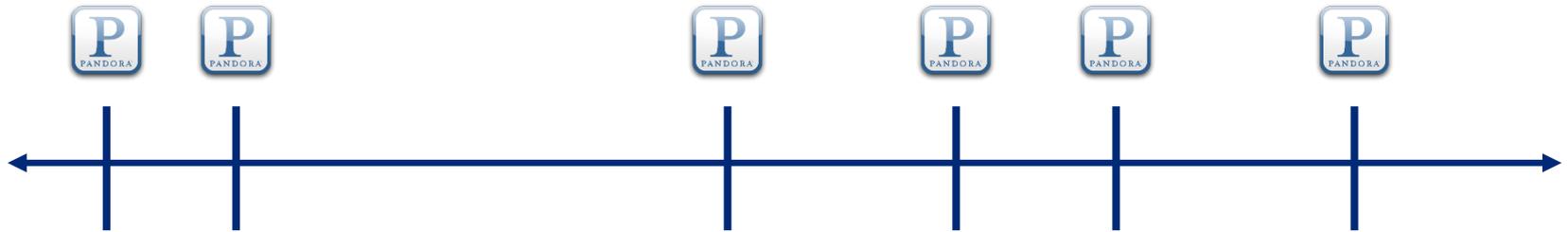
Laura listens to music on Pandora. We can see each time she is served a display ad on the site.



---

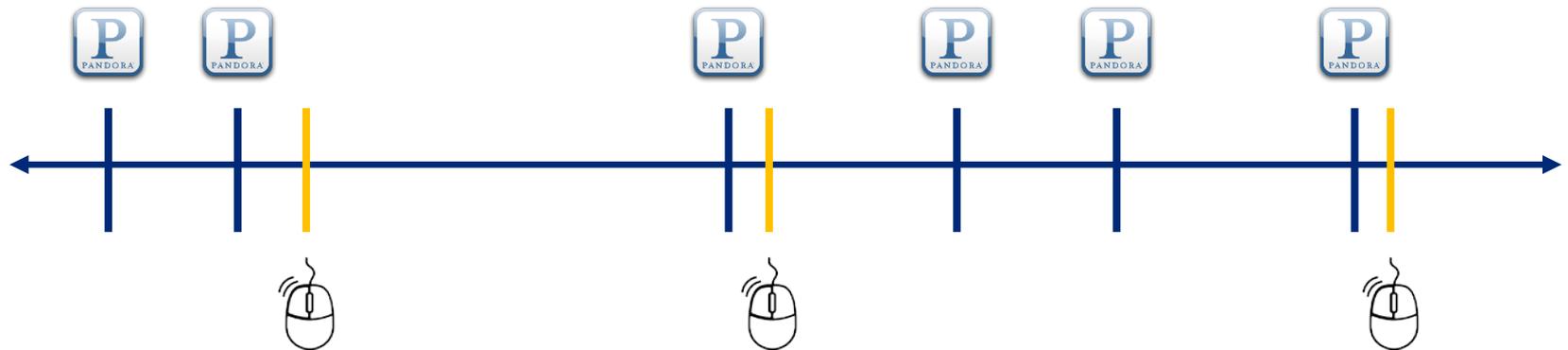
# How are the Data Collected?

Laura listens to music on Pandora. We can see each time she is served a display ad on the site.



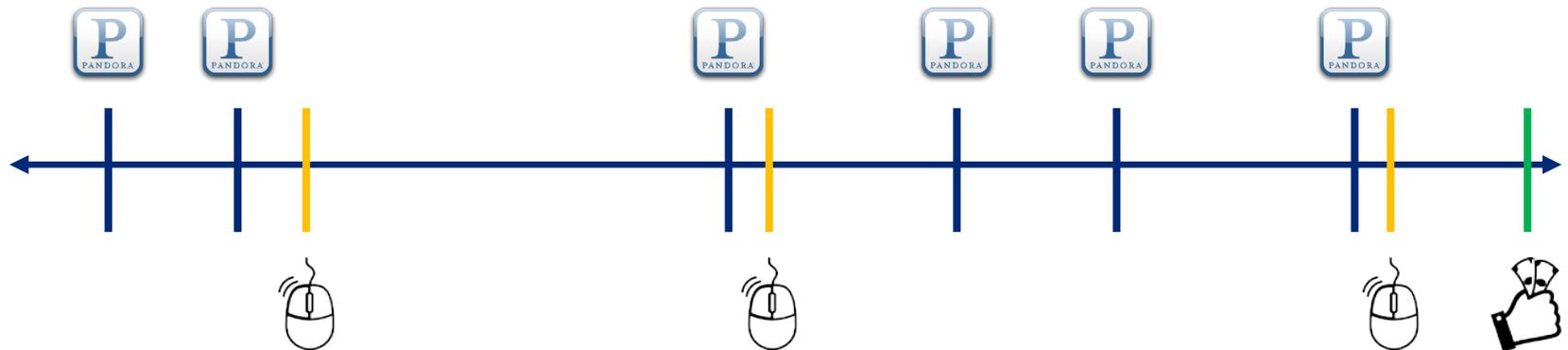
# How are the Data Collected?

Because Laura is tracked by her cookies, we can see every time she clicks on an ad.



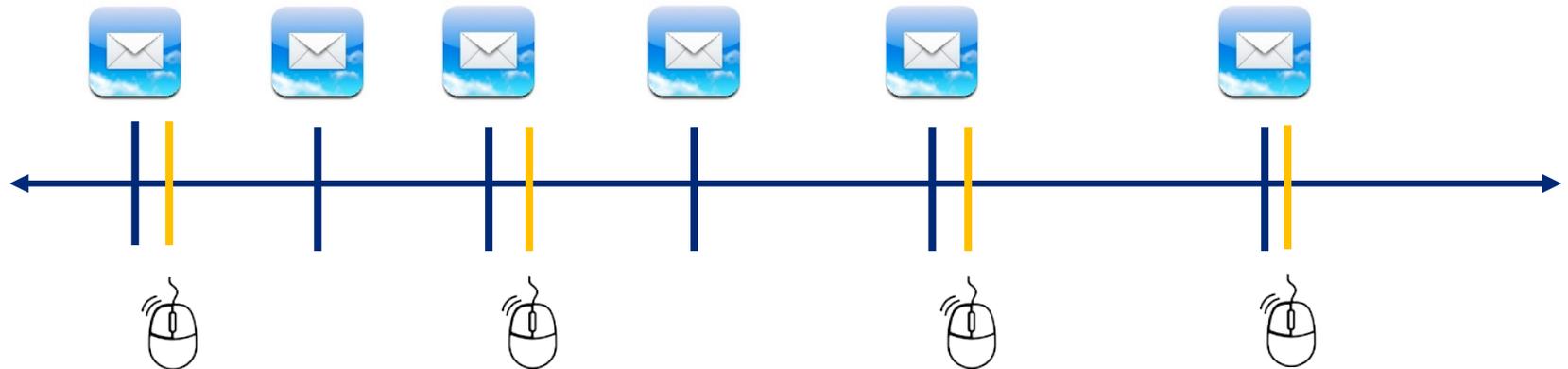
# How are the Data Collected?

We can also see when Laura ultimately makes a purchase at the company's website.



# How are the Data Collected?

Jack receives email ads. He often clicks on them, but has never purchased.



---

# How are the Data Collected?

Meanwhile, Martha is tracked making conversions without ever having been served an ad.



---

# Data Structure



---

# Data Structure

- impressions
- engagements
- conversions

- flags estimating past purchase behavior

---

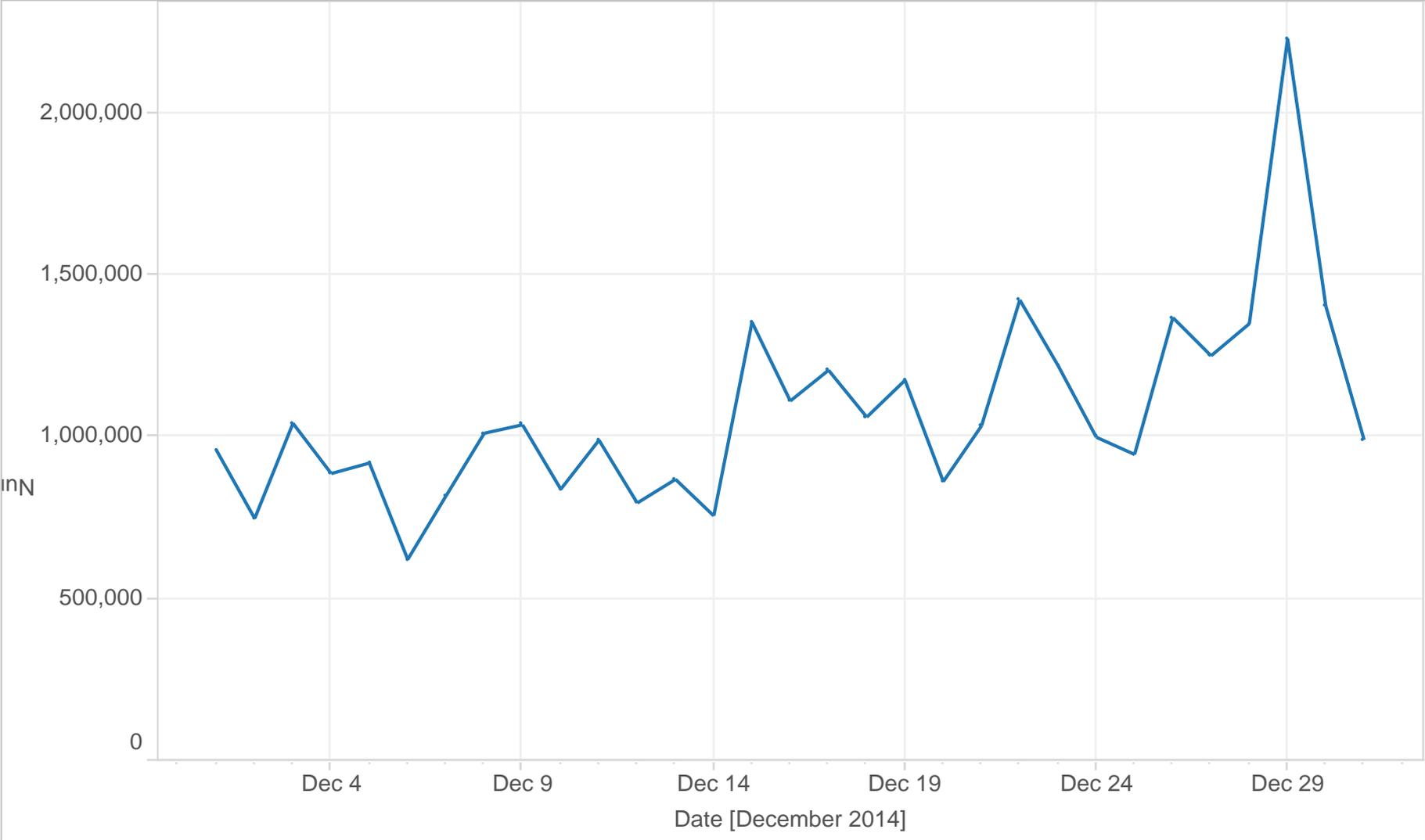
# Impression

An impression is **an online ad served to an end user**. Impressions can be banner ads, homepage takeovers, or emails. There are **~300 million individual impressions** recorded in the dataset.

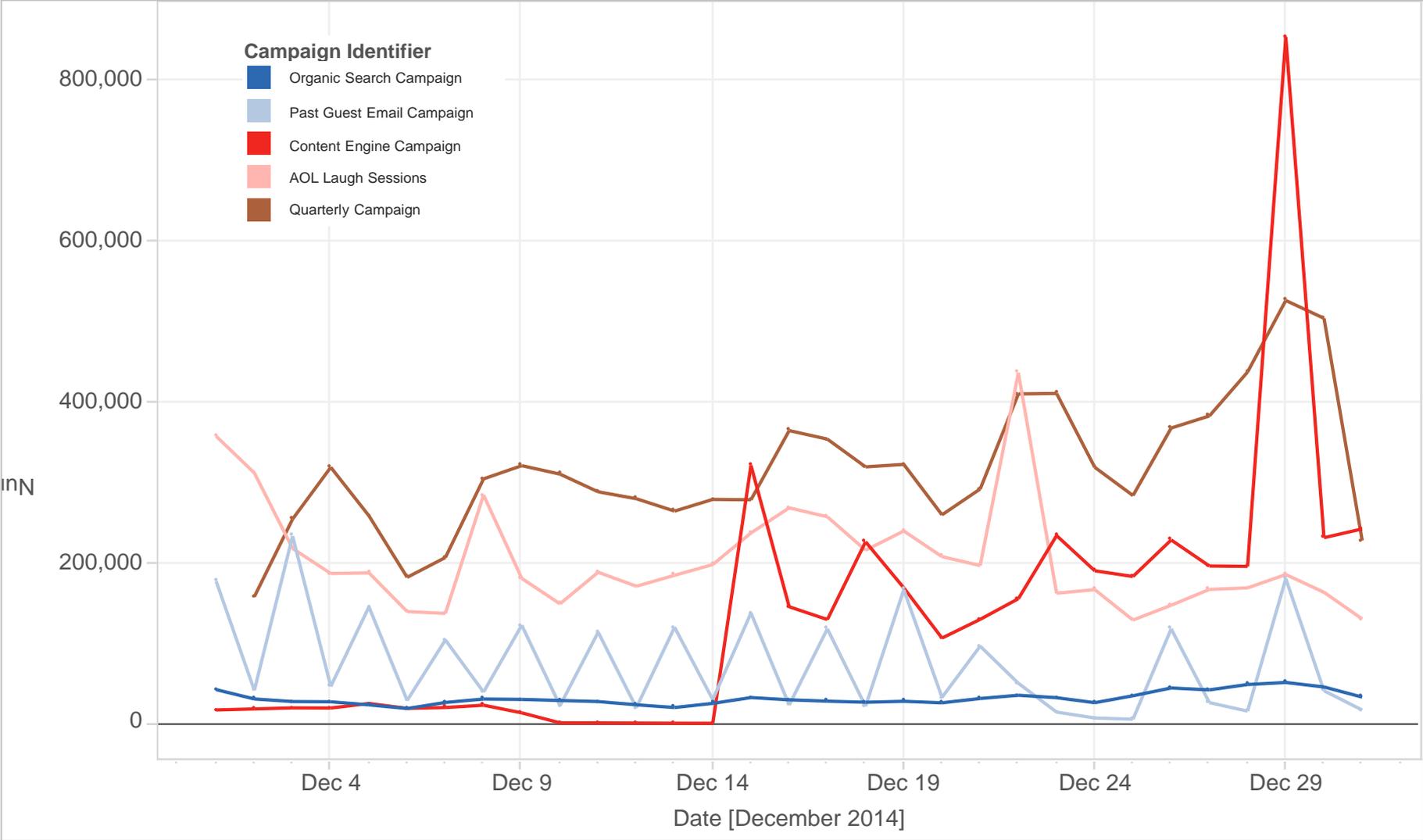
The impression data contains:

- **date and timestamp** of the exposure
- identifier for the **cookie**
- detailed information about **the creative**:
  - **creative ID**
  - **size and placement**
  - **site where the ad appeared**

# Impressions by Day, December 2014



# Impressions by Day for 5 Campaigns, December 2014



# Sneak Peek: Impression

record_date	ak_user_id	new_usr	tag_id	namespace_id	campaign_id	site_id	placement_id	creative_grp_id	creative_id
12/1/2014 0:02	588504455741313000	t	4024	19	1006	2005	3005	-1	4005
12/1/2014 0:02	1055398918702850000	t	2387	7	8265590	1474157	111542959	284781992	59605717
12/1/2014 0:03	756249074244711000	f	2387	7	8239032	1697774	110831087	284245006	59327538
12/1/2014 0:03	1141398937740480000	t	4024	19	1005	2005	3005	-1	4005
12/1/2014 0:03	758989254164948000	t	4024	19	1020	2010	3074	-1	4074
12/1/2014 0:03	338258528284510000	t	2387	7	8246217	1511424	111143515	285736707	60059531
12/1/2014 0:04	1055398918702850000	f	2387	7	8265590	1474157	111543120	284781980	59604926
12/1/2014 0:04	666604052365016000	f	2387	7	8265590	1474157	111542959	284781992	59605717
12/1/2014 0:04	666604052365016000	f	2387	7	8265590	1474157	111543024	284782218	59605380
12/1/2014 0:04	666604052365016000	f	2387	7	8265590	1474157	111543120	284781980	59604926
12/1/2014 0:05	88079954114333900	t	2387	7	8246217	1511424	111143515	285736707	60059531
12/1/2014 0:05	1055398918702850000	f	2387	7	8265590	1474157	111543024	284782218	59605380
12/1/2014 0:06	264299997242961000	t	2387	7	8265590	1414100	113007814	286130692	59606864
12/1/2014 0:06	1032353820773210000	t	4024	19	1020	2010	3074	-1	4074
12/1/2014 0:06	327945119409515000	f	4024	19	1002	2002	3002	-1	4002
12/1/2014 0:06	666604052365016000	f	2387	7	8265590	1474157	111543024	284782218	59605380
12/1/2014 0:06	666604052365016000	f	2387	7	8265590	1474157	111542959	284781992	59605717
12/1/2014 0:06	713002314786499000	t	2387	7	8246217	1511424	111143515	285736707	60059531
12/1/2014 0:08	296151931202952000	t	4024	19	1002	2002	3002	-1	4002
12/1/2014 0:08	787707947624478000	t	4024	19	1002	2002	3002	-1	4002
12/1/2014 0:08	666604052365016000	f	2387	7	8265590	1474157	111543024	284782218	59605380

---

# Engagement

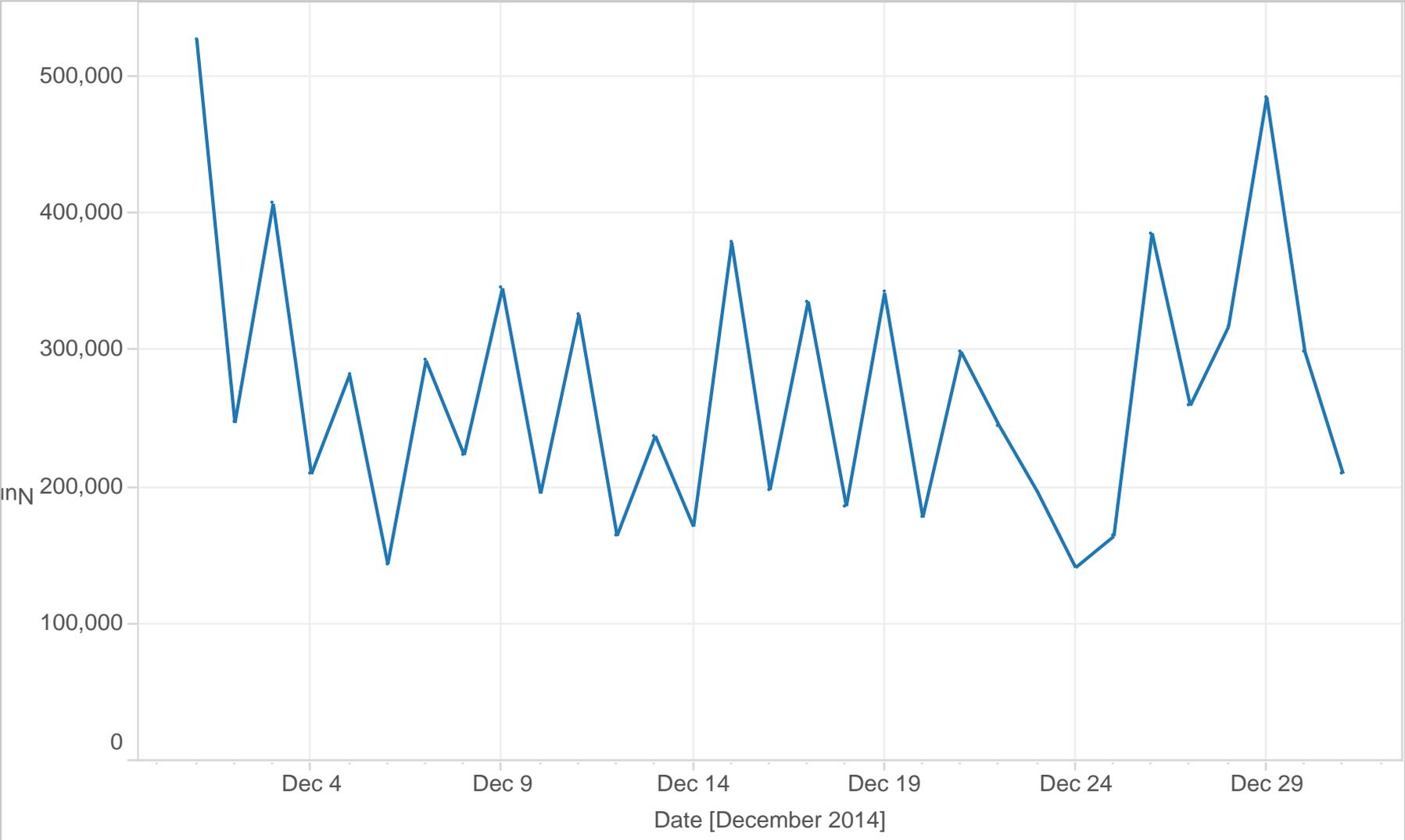
An “engagement” is defined as a **user’s interaction with an impression through an ad click**. The dataset contains more than **22 million engagements**.

The engagement data contains:

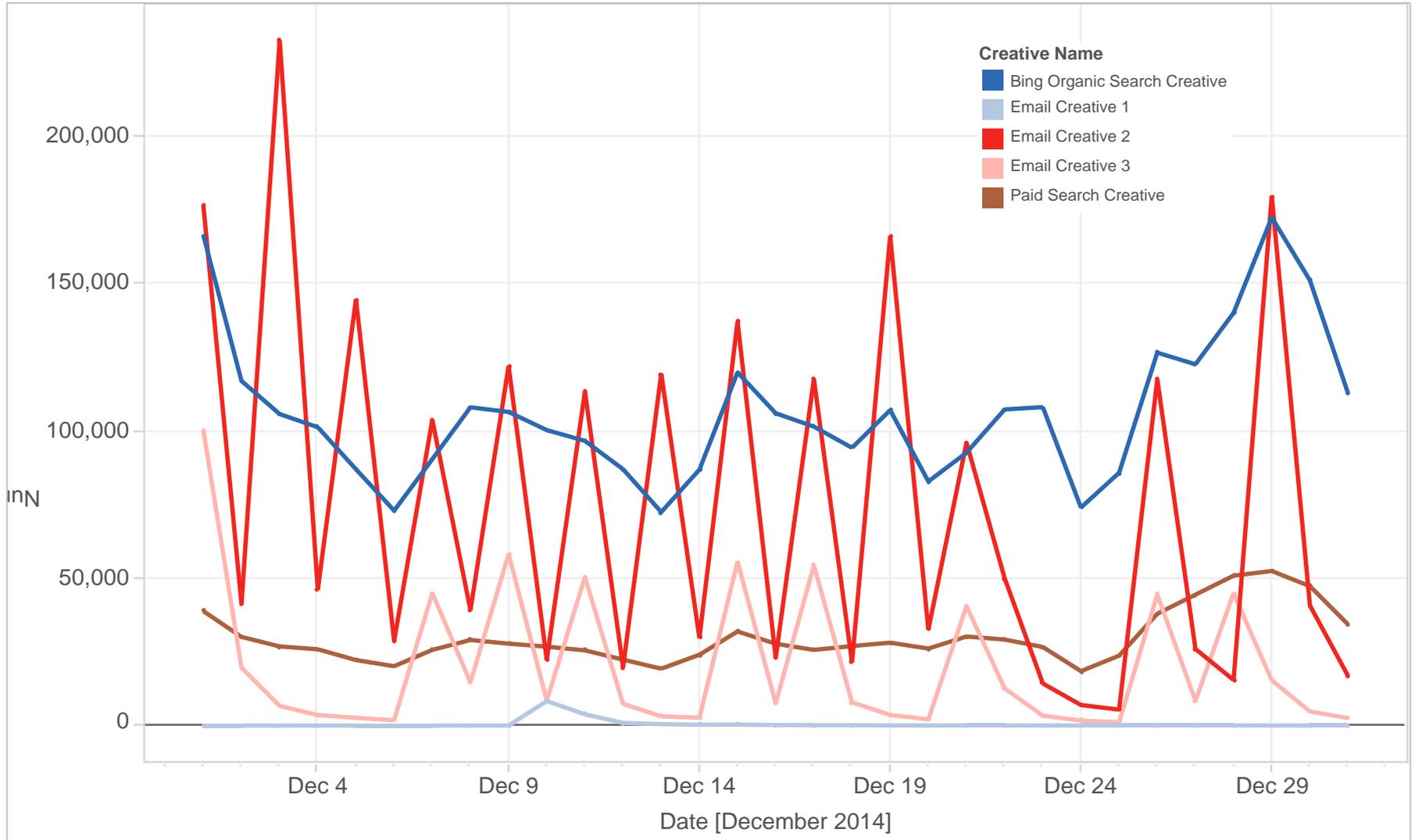
- **identifier for the impression** with which that user interacted
- **date and time** of the engagement
- **identifier for the user**
- **links to detailed information about the click**, such as:
  - **creative ID**
  - **campaign**
  - **site**
  - **physical ad placement**



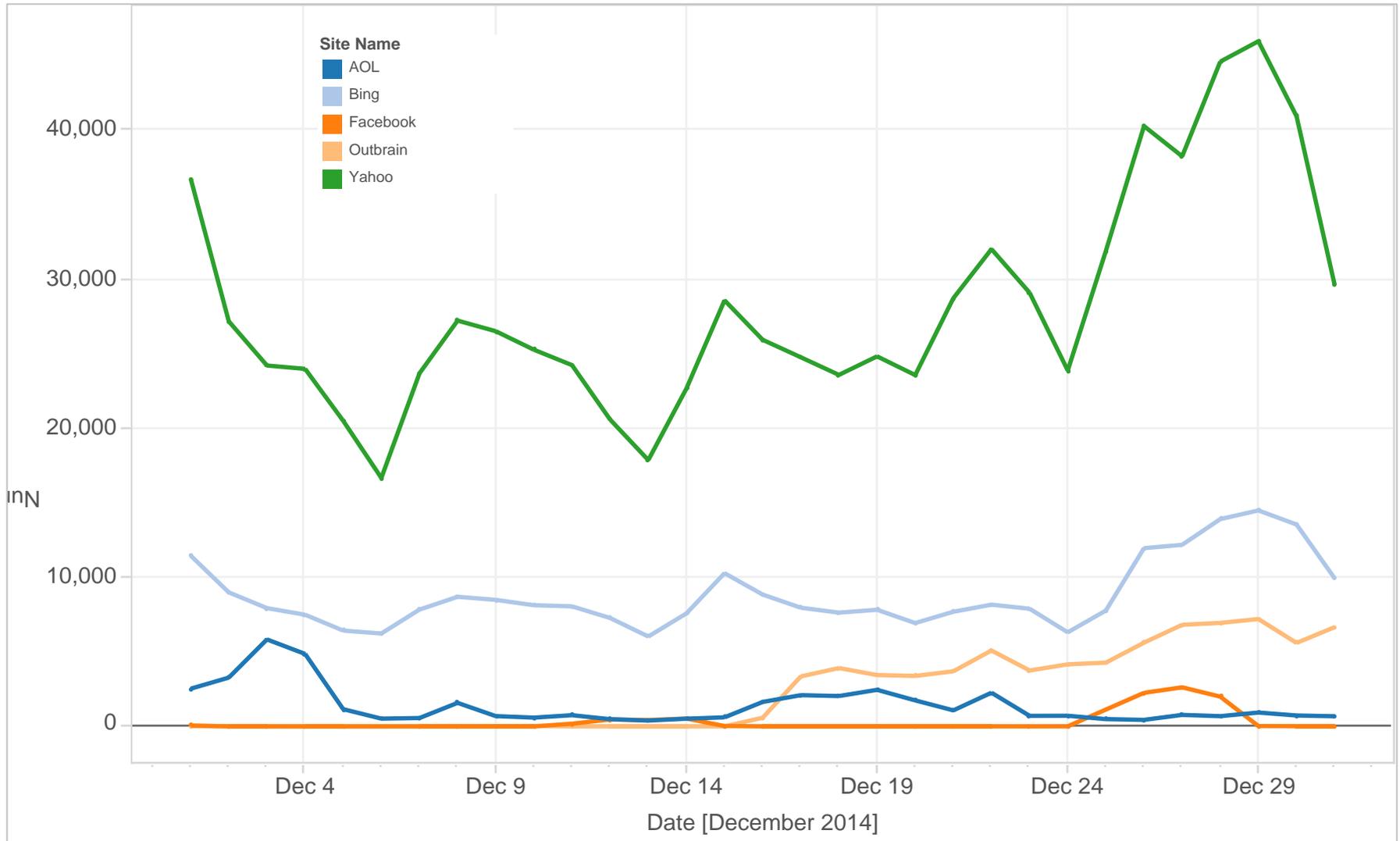
# Engagements by Day, December 2014



# Engagements by Day for 5 Creatives, December 2014



# Engagements by Day for 5 Sites, December 2014

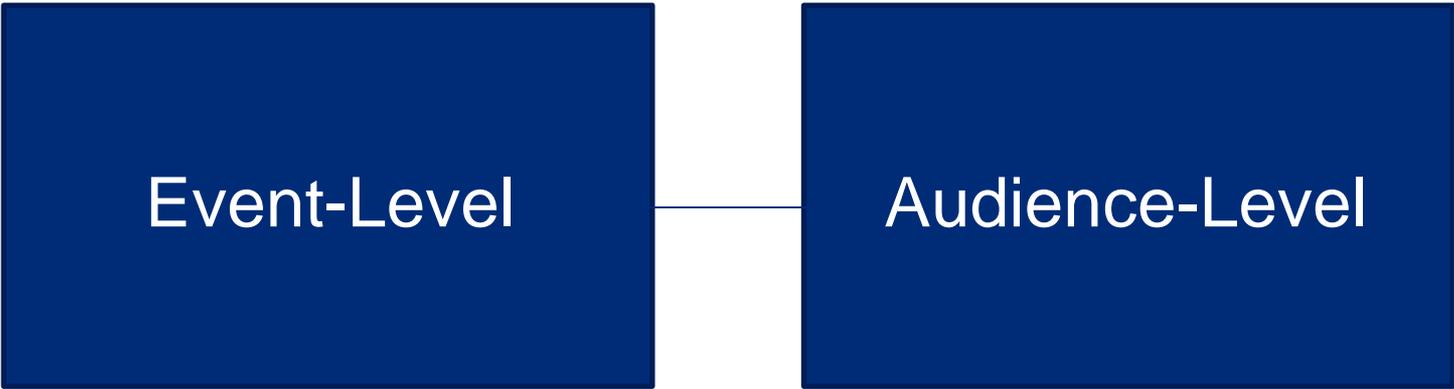


# Sneak Peek: Engagement

record_date	ak_user_id	new_usr	tag_id	namespace_id	campaign_id	site_id	placement_id	creative_grp_id	ad_id	impr_match	
12/1/2014 0:01	226420523945712000	f	4024	19		1002	2002	3002	-1	4002	1
12/1/2014 0:01	246691115148141000	t	4024	19		1012	2009	3066	-1	4066	1
12/1/2014 0:01	237017862468297000	t	4024	19		1020	2010	3074	-1	4074	1
12/1/2014 0:01	884146790315795000	t	4024	19		1020	2010	3074	-1	4074	1
12/1/2014 0:01	88815727434346000	f	4024	19		1020	2010	3074	-1	4074	1
12/1/2014 0:02	1139393168830280000	t	4024	19		1020	2010	3074	-1	4074	1
12/1/2014 0:04	493210741362066000	t	4024	19		1002	2002	3002	-1	4002	1
12/1/2014 0:05	617791063765617000	f	4024	19		1020	2010	3074	-1	4074	1
12/1/2014 0:05	547525863435406000	f	4024	19		1002	2002	3002	-1	4002	1
12/1/2014 0:06	557505489324347000	f	4024	19		1002	2002	3002	-1	4002	1
12/1/2014 0:06	226420523945712000	f	4024	19		1002	2002	3002	-1	4002	1
12/1/2014 0:07	557505489324347000	f	4024	19		1002	2002	3002	-1	4002	1
12/1/2014 0:07	980365464201380000	f	4024	19		1012	2009	3066	-1	4066	1
12/1/2014 0:07	980365464201380000	f	4024	19		1012	2009	3066	-1	4066	1
12/1/2014 0:08	786991230385464000	f	2389	7	8265590	1414100	111542500	284784342	0	0	0
12/1/2014 0:09	557505489324347000	f	4024	19		1002	2002	3002	-1	4002	1
12/1/2014 0:09	418361295943878000	f	4024	19		1004	2006	3005	-1	4005	1
12/1/2014 0:09	109711033357745000	f	4024	19		1005	2005	3005	-1	4005	1

---

# Questions?



---

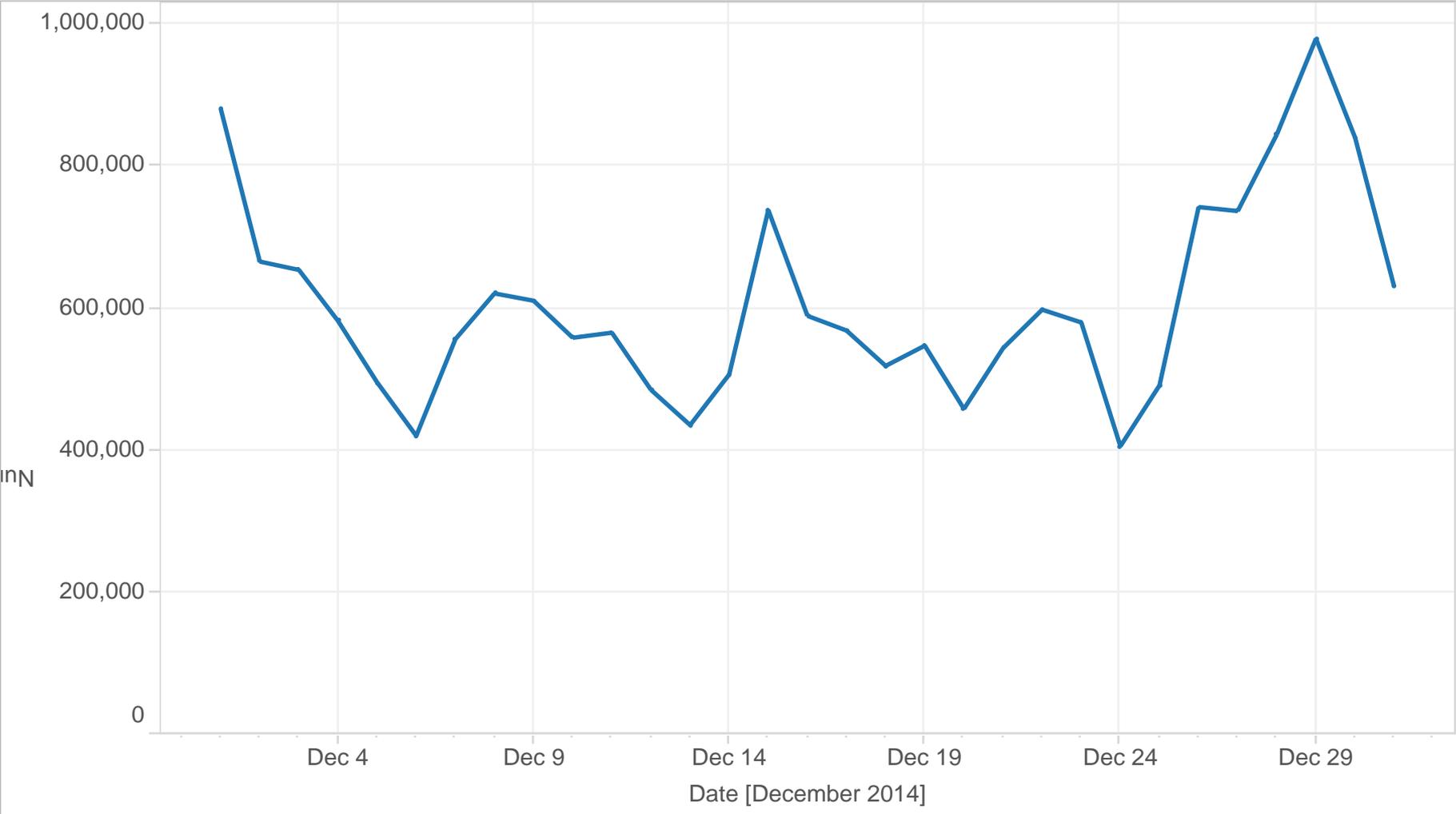
# Conversion

Conversions are events that occurred at the advertiser's website. There are approximately 300 million conversions in this dataset.

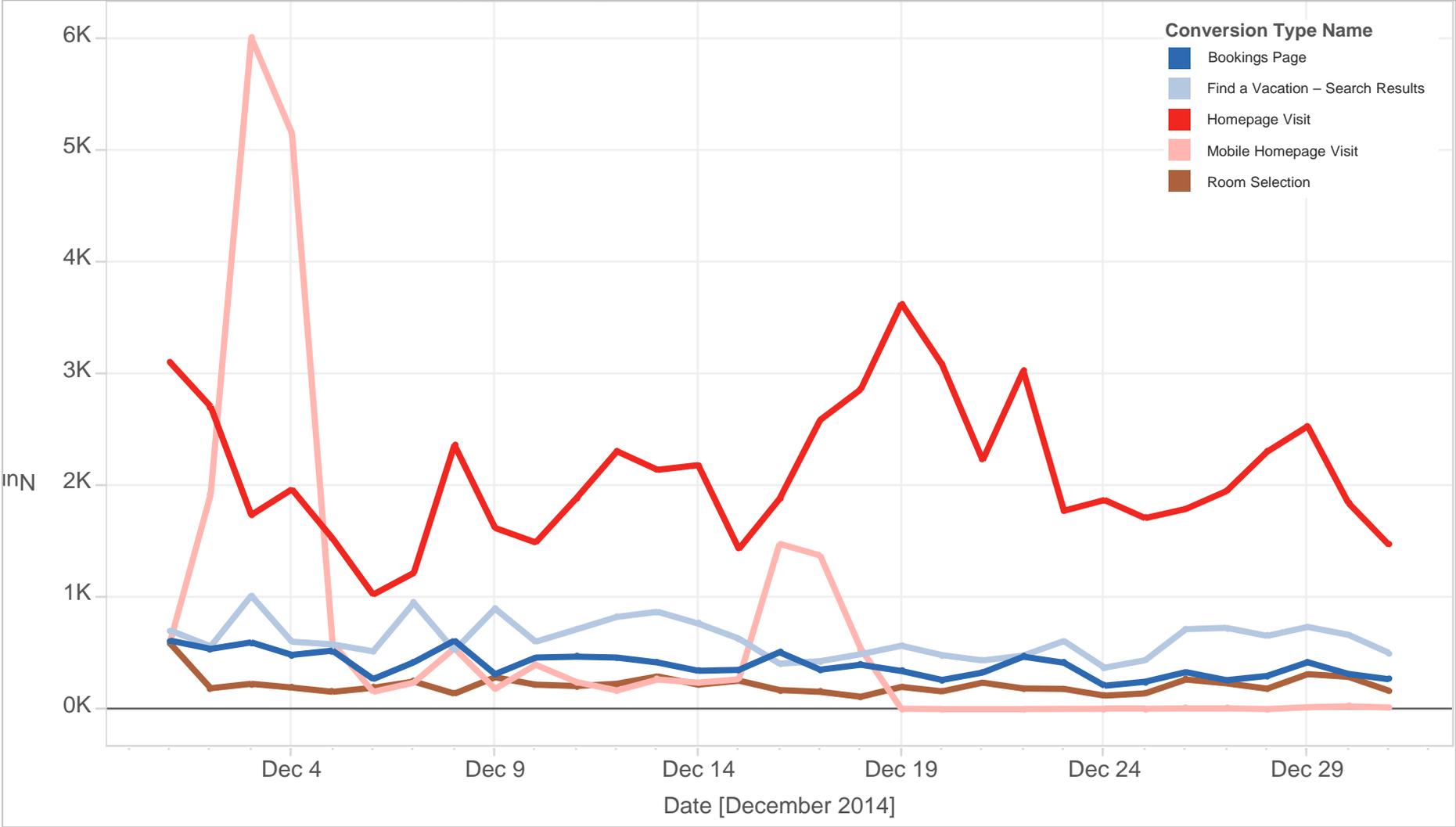
The conversion data contains:

- user/cookie identifier
- the type of conversion, such as landing on the homepage or receiving a booking confirmation
- the date/time on which the conversion occurred
- the impression and/or engagement that led to conversion, if applicable

# Conversions by Day, December 2014

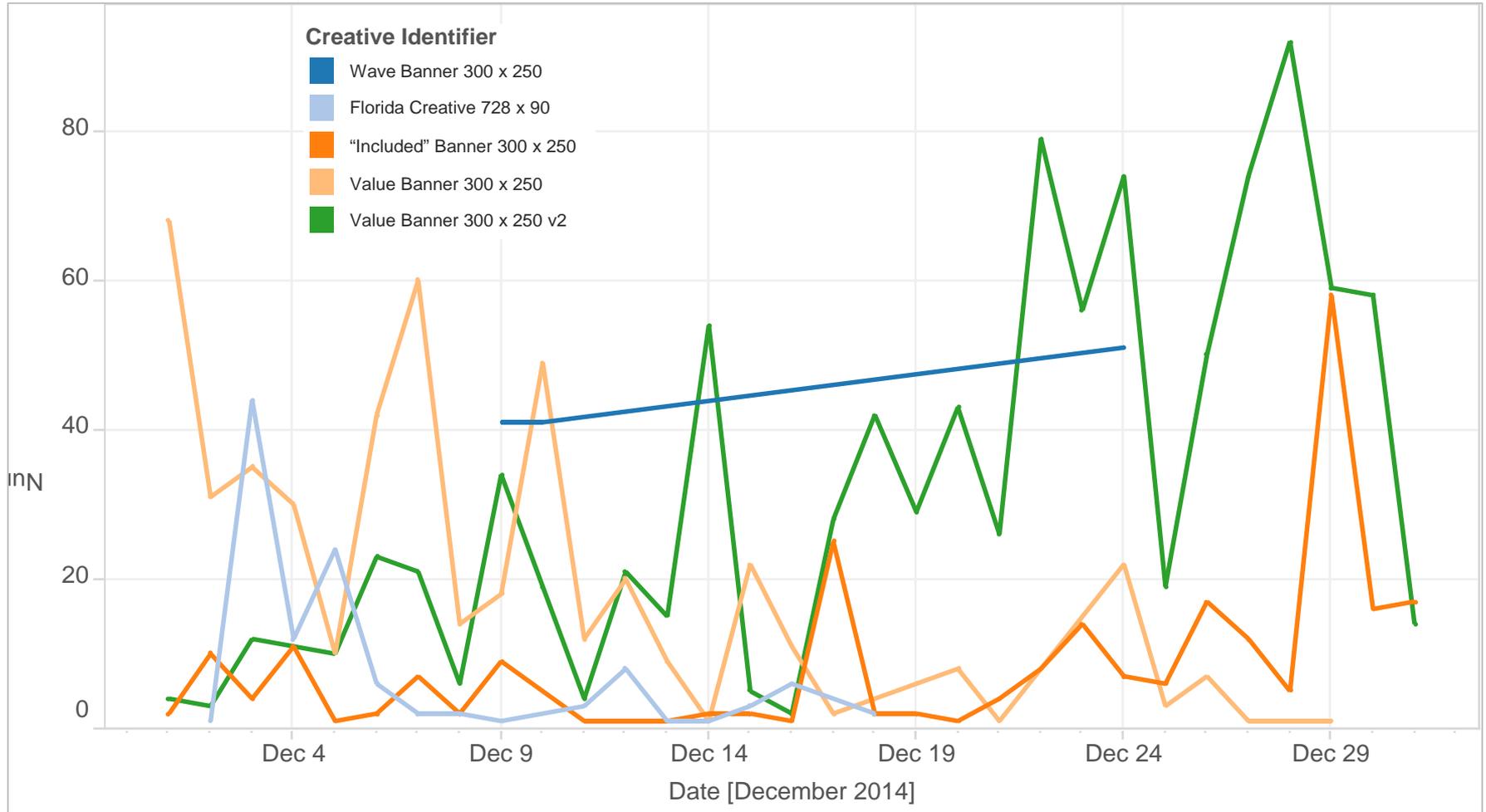


# Five Conversion Types by Day\*, December 2014

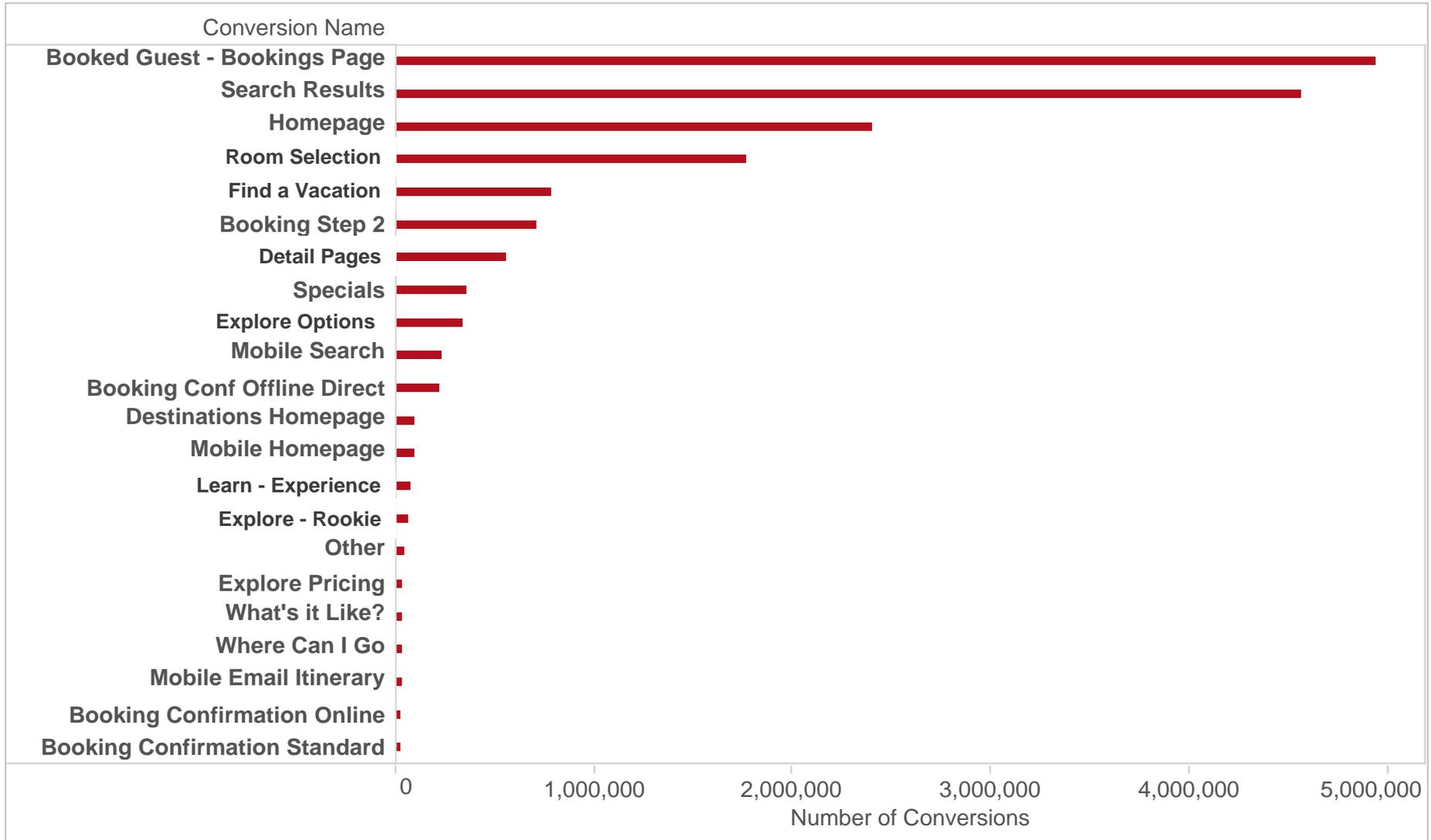


\*For conversions attributed to clicks  
 \*\*The dataset contains ~40 conversion types.

# Conversions by Day for 5 Creatives, December 2014



# Conversions by Type, December 2014



# Sneak Peek: Conversion

AK User ID	record_date	Namespace ID	Placement ID	Site ID	Campaign ID	Creative ID	Creative GRP ID	Conv Type Code
228631182196104997	12/28/2014 10:57:54 AM	19	7003	6000	5003	8004	-1	bcextdata
228631182196104997	12/28/2014 10:56:34 AM	19	7003	6000	5003	8004	-1	bcextdata
983667734324705936	12/31/2014 9:48:04 AM	19	7003	6000	5003	8004	-1	bcextdata
983667734324705936	12/31/2014 7:22:35 AM	19	7003	6000	5003	8004	-1	bcextdata
559264314627948743	12/27/2014 11:00:33 AM	19	7003	6000	5003	8004	-1	bcextdata
889323064810348921	12/14/2014 12:21:21 AM	19	7003	6000	5003	8004	-1	bcextdata
1146735536605933779	12/28/2014 9:49:42 PM	19	7003	6000	5003	8004	-1	bcextdata
36520717092142310	12/13/2014 6:50:36 PM	19	7003	6000	5003	8004	-1	bcextdata
6605862545831885	12/26/2014 10:44:11 AM	19	7003	6000	5003	8004	-1	bcextdata
494611021147406659	12/1/2014 5:32:50 PM	7	108352577	1506425	8086081	0	281759120	bcstandard
405855757220231216	12/1/2014 11:15:29 AM	7	108352578	1506425	8086081	0	281759120	bcstandard
350687737870491286	12/6/2014 2:22:10 PM	7	112700588	1698582	8239032	0	285824949	bcstandard
1036872557227241145	12/24/2014 4:42:29 AM	7	112700588	1698582	8239032	0	285824949	bcstandard
885576390321358257	12/15/2014 11:07:02 PM	7	113311252	1506425	8382714	0	286455006	bcstandard
918208769295577023	12/11/2014 9:51:52 PM	7	113311435	1506425	8382714	0	286455006	bcstandard
877845669015955868	12/17/2014 10:47:12 AM	7	113350763	1506425	8385609	59604926	286525847	bcstandard
937367955117553160	12/15/2014 10:05:24 PM	19	3001	2001	1001	4001	-1	bcstandard
707568154971585228	12/10/2014 9:16:33 PM	19	3001	2001	1001	4001	-1	bcstandard
948287417304898955	12/27/2014 5:14:53 PM	19	3001	2001	1001	4001	-1	bcstandard
49595471820664532	12/1/2014 1:20:46 PM	19	3001	2001	1001	4001	-1	bcstandard
80306078155818166	12/8/2014 11:15:59 PM	19	3001	2001	1001	4001	-1	bcstandard
682111099754084222	12/23/2014 4:32:48 PM	19	3001	2001	1001	4001	-1	bcstandard
530753271874244940	12/21/2014 9:05:06 PM	19	3001	2001	1001	4001	-1	bcstandard
194874050575294097	12/20/2014 5:30:08 PM	19	3001	2001	1001	4001	-1	bcstandard
841480777051711310	12/27/2014 4:00:44 PM	19	3001	2001	1001	4001	-1	bcstandard

---

# Questions?



---

# Estimated Purchase Behavior Data

For some cookies who have made conversions, the dataset also contains estimated information from a financial data provider about the cookie's purchasing behavior.

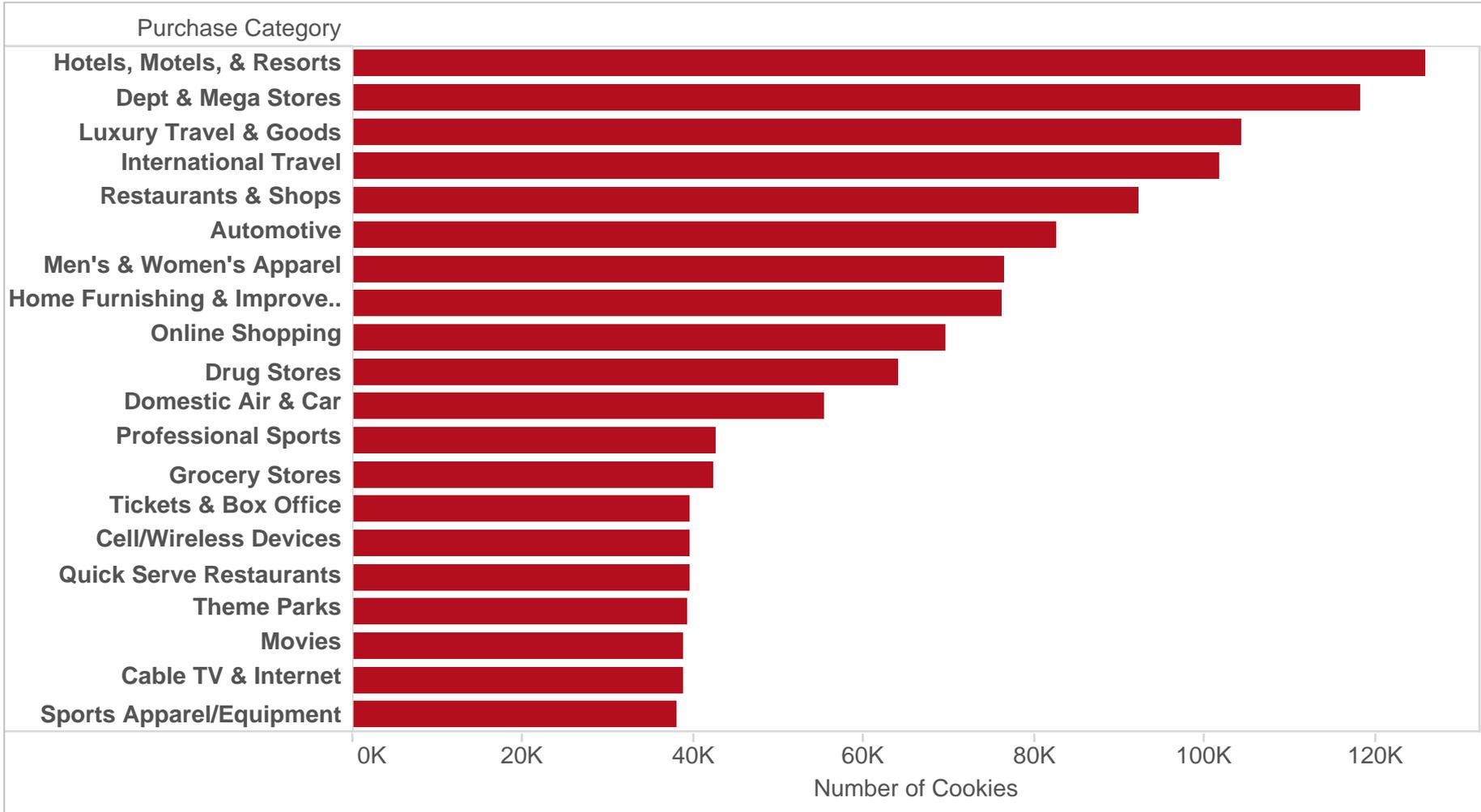
The purchase behavior data designates “Top Tier Spenders”, “Frequent Transactors”, and “Highly Likely to Purchase in the Next 30 Days” for categories including, but not limited to\*:

- department stores
- cruise and domestic/international air travel
- economy hotels & motels
- fine dining
- travel to Asia/Europe
- luxury retailers

\*this data includes approximately 65 purchase categories.



# Distribution of Top 20 Purchase Categories

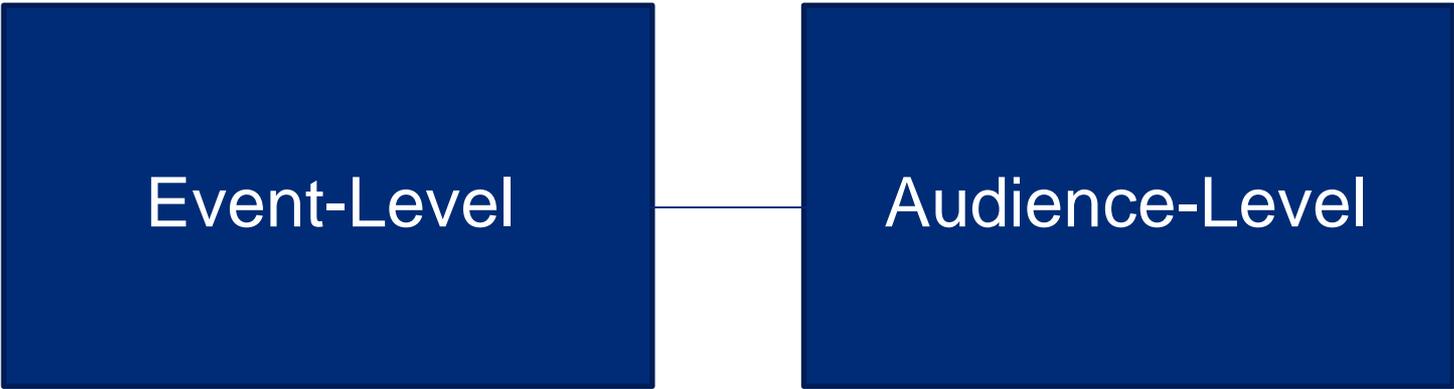


# Sneak Peek: Purchase Behavior Data

Cookie_ID	Purchase Category/Description
934072111371294000	Pizza Restaurants - Frequent Transactor
603222227657465000	Mid-Range Restaurants / Non-Chain - Frequent Transactor
427227865375447000	Travel to Africa - Highly likely to travel in the next quarter
931945891975821000	Cleaning and Exterminating Services - Top Tier Spender
432673646687730000	Travel to Southern Europe - Highly likely to travel in the next quarter
341930166854965000	Family Apparel - Top Tier Spender
317525583498476000	Camera & Photography - Frequent Transactor
353810032733232000	Home Improvement - Top Tier Spender
692728519226952000	College or University Education - Top Tier Spender
677607782884760000	Camera & Photography - Frequent Transactor
200889128451454000	Travel to Northern Europe - Highly likely to travel in the next quarter
1054440667962750000	Quick Serve Restaurants - Frequent Transactor
868012946136007000	Online Video Subscription Services - Top Tier Spender
533155699790606000	Gift Stores - Highly likely to purchase next 30 days
562244871947634000	Camera & Photography - Highly likely to purchase next 30 days
795870001948742000	Grocery Stores - Frequent Transactor
1138478797612570000	Train Travelers - Top Tier Spenders
144067774656819000	Casinos & Resorts - Highly likely to purchase next 30 days
701878529993868000	Online Shopper - Frequent shopper
174402304413439000	Theme Parks - Highly likely to purchase next 30 days
247790931853395000	Luxury Women's Apparel - Frequent Transactor
788071835092139000	Grocery Stores - Frequent Transactor

---

# Questions?



---

# Possible Research Areas from Annalect

- Analyzing the efficacy of creative elements, particularly by site
- Improving attribution models
- Understanding advertising effects on path-to-purchase
- Predicting conversion based on browsing behavior
- Detecting returning users who are browsing on a different device, or have cleared their cookies
- Something else *not on this list...*



---

# Possible Research Areas from Annalect

- Analyzing the efficacy of creative elements, particularly by site
- Improving attribution models
- Understanding advertising effects on path-to-purchase
- Predicting conversion based on browsing behavior
- Detecting returning users who are browsing on a different device, or have cleared their cookies
- Something else *not on this list...*

**What would YOU study with a dataset like this?**



---

# A Few Operational Notes

This webinar is intended for **researchers & scholars**. It should not be published or presented without permission from WCAI.

- Questions on program: [wcai-research@wharton.upenn.edu](mailto:wcai-research@wharton.upenn.edu)
- Press contacts: [wcai-press@wharton.upenn.edu](mailto:wcai-press@wharton.upenn.edu)

To get access to the data, research teams should submit a proposal to WCAI and the sponsor for approval.

Please note that the data sponsor **cannot be named in any publication or presentation.**



---

# Proposal Format

Proposals should be in PDF format, no more than 2,000 words, and include the following information:

- **Title**
- **Author(s) name, title, affiliation and e-mail address:** Please designate a corresponding author.
- **Summary information:** a single “slide” that visually summarizes team & project.
- **Abstract**
- **Introduction:** Describe expected contribution(s), covering both the academic and practical aspects. Please keep it concise, and cite relevant work as necessary to explain your academic contribution. There is no need to include a lengthy literature review.
- **Detailed project proposal:** Please include enough detail that we can assess the feasibility & merit of the proposed approach. For example, modeling projects should include at least a sketch of the model. In addition, include a rough estimate of how long the project will take. Also include the business relevance of your research and the impact for the sponsor as well.
- **Data Needs:** Bulleted list of data required or requested for analysis not explicitly mentioned during this webinar. While we can't guarantee the inclusion of these items, we are happy to investigate the availability.
- **Biographies:** Include up to a paragraph-long biography highlighting what each team member will contribute to the project.



---

# Proposal Submission Process

- Read more about the proposal format at <http://www.wharton.upenn.edu/wcai/Proposals.cfm>
- Submit proposals at <http://www.wharton.upenn.edu/wcai/proposalform.cfm> no later than **Monday, March 30, 2015, 12 noon US Eastern**.
  - Proposals will only be accepted **online**.
  - A **single PDF** combining the written portion and the single slide.
- Proposals will be evaluated both on **academic contribution** and **potential to significantly improve the research sponsor's marketing practice** by the proposal review committee:

**P.K. Kannan (University of Maryland), Anindya Ghose (New York University),  
Elea Feit (Drexel University), Eric Bradlow (WCAI),  
Pete Fader (WCAI), and representatives from Annalect.**
- Contact [wcai-research@wharton.upenn.edu](mailto:wcai-research@wharton.upenn.edu), if you have questions prior to submitting your proposal.



---

# Other WCAI Opportunities for Researchers

If you registered for this webinar, you will receive regular announcements regarding:

- Research Opportunities like this one:
  - Online Reviews for a Large Scale Retailer
  - Supply Side and Demand Side Data for a Fast Food Chain

Also find us at:

- SSRN Research Paper series:  
<http://www.ssrn.com/link/Wharton-Cust-Analytics-Initiative-RES.html>
- Announcements: <http://wcai.wharton.upenn.edu>





////////////////////////////////////  
KNOWLEDGE FOR ACTION  
////////////////////////////////////