



# Ticketmaster Data Guide

## [Description of Datasets](#)

### [Data Generation](#)

### [Datasets](#)

#### [Adwords Data](#)

#### [User Behavior Data \(GA Data\)](#)

#### [Purchase Data](#)

### [Final Comments](#)

## [Themes and Questions](#)

### [Customer Segmentation](#)

### [Pricing Optimization](#)

### [Marketing Optimization](#)

## Description of Datasets

### Data Generation

All of the data presented are web-based statistics. While the files look quite different from each other, the data are generated through the same process. When you visit a website and the website loads, several javascript scripts are run during the loading process. This collect of scripts (referred to internally as a pixel) records the user's actions on the website and sends the resulting data to an external database. The user behavior data (GA data) and the adwords data present this information to you in a tabular form. Note that the adwords data come from Google Adwords, a service provided by Google to TicketMaster.

# Datasets

## 1. Adwords Data

The adwords data includes all statistics associated with an adword aggregated to the keyword level. A keyword is a word or phrase that TicketMaster has placed a bid for with Google. If someone uses Google and types in that keyword, TicketMaster enters an auction to determine where, or whether, its ad will appear for that user. Within Google Adword's hierarchical structure, an account contains many campaigns, a campaign contains many adgroups, and an adgroup contains many ads / keywords. The unit of observation for this dataset is at the keyword level, each entry contains information on the ad / keyword as well as details on the adgroup, campaign and account. If you want to know how well an adgroup or campaign is doing, you will need to aggregate the data to that level.

## 2. User Behavior Data (GA Data)

Each row in the user behavior data represents a click from a user. This dataset contains information on the referral site (site from which user came to the Ticketmaster website), whether or not the referral site was a paid source (such as Google Adwords), the type of device the user is on, etc.

This dataset can be merged with the adwords dataset with `adgroup_id` and `campaign_id` fields. This merge can be useful for exploring what types of adwords attract what types of users.

This dataset can also be merged with the purchase dataset by `event_id` to capture meta-data for events, such as category of event. This merge can be useful for exploring whether users behave differently for different types of events. Caution: The fact that both datasets contain a field called `event_id` does not mean every click resulted in a purchase. Also note that the relationship between the GA and Purchase `event_id` is many-to-many and you should not directly merge the two data sets without first aggregating one or both data sets.

## 3. Purchase Data

Each row in the purchase dataset represents a purchase from Ticketmaster. This dataset contains information on what event the purchase was for, the type and date of event, cost of tickets, etc.

## Final Comments

These data are in a very raw form, which gives you an opportunity to create your own aggregated fields. For example, one can create a field that describes the total amount of money a person spent on Ticketmaster, total amount of time they spent on Ticketmaster, or the average number of tickets they buy per purchase.

Also note that each purchase and action is recorded with a timestamp, which means you have information on order of purchases made by the same user. This offers the opportunity to conduct a temporal analysis on broad categories or a single user.

## Themes and Questions

### Customer Segmentation

We at Ticketmaster are very interested in how to improve our services for our current users, and also expand our customer base. We know that our customers span a wide spectrum of various types of users. We want to have a better understanding of our customer base as this will help us market and improve our services.

The group of customers Ticketmaster we are most interested in identifying are the *true fans*. These are customers who buy tickets only from Ticketmaster and who, after buying tickets, always go to the event themselves, instead of selling them on the resale market. Among our customers we also have *scalpers* - they buy tickets early and sell them later at a much higher price on the resale market. Ticketmaster does not have data from the resale sites (such as StubHub), hence it is a challenge for us to distinguish between true fans and scalpers. We are very interested in doing customer segmentation so that we can

- identify the true fans,
- discover other segments among our user base (e.g. one time buyers), and
- devise a marketing plan to target these segments.

### Pricing Optimization

Ticket pricing is a topic that is largely unexplored at Ticketmaster because we are generally not the ones to set the price. The Data Science Division at Ticketmaster is very interested in insights into how to price tickets.

Currently the ticket prices for a touring artist are roughly the same at every performance location, i.e. there is little to no optimization done on pricing tickets based on location. We at Ticketmaster are very interested in determining the popularity of certain geographic locations for certain types of events and for certain types of people so that we can use this information for ticket price optimization.

### Marketing Optimization

At Ticketmaster we recently started a new project to automate the process of ad generation. Prior to the implementation of this process, every ad a user saw on Google was manually

generated by a human. However, for the many small shows, or shows that occur infrequently, it makes more sense to have a machine generate the keywords and determine the price to pay. Adgroups with the phrase "ZZ - Bob's Automation Shop" are ads generated by a machine. We would like to know how well these work.

Currently our SEM (Search Engine Marketing) automation process is still very primitive as we do not have a good system for generating keywords. We generally work with a template set of keywords that takes words from the event name and artist (attraction) name. We are interested in approaches that use third party resources to expand our keyword list for a specific event. For example, events with keywords like "Duke vs. UCLA Men's Basketball" can be expanded to include keywords such as: Blue Devils, Bruins, etc. We would like to know which ads will convert well and which will not and how this varies based on user behavior.