Themes and Questions for Ticketmaster Data

By: Christian Gao

10/1/15

Section 1 Description of Datasets

Data Collection Description- The data presented to you are all web-based statistics. Even though the formatting is very different between the datasets they are generated through the same way. Whenever you visit a website, when the website loads, a package of javascript code is ran during the loading process and sends the data with a timestamp to service that collects the data.

Adwords Keywords Performance Report Description- The Keywords Performance report includes all statistics aggregated at the keyword level, one row per keyword. Within google's adword's hierarchical structure, an account contains many campaigns, a campaign contains many adgroups, and adgroup contains many ads and keywords. Since this report is at the keyword level, if you want to know how well a adgroup is doing, you must aggregate the data by adgroup.

User Behavior Data (GA Data Description)- Each row in this dataset represents a click from a user. T You can merge this dataset with the adwords dataset with adgroup_id or campaign_id to see what kind of ads attract what kind of people. You can also merge this dataset with the purchase dataset by event_id to include meta data for events such as the category of event to see if users behave differently to certain events. Be carful though, just because you can merge the two datasets by event_id does not mean every click was a purchase.

Purchase Data - Each row in this dataset represents a purchase. This dataset describes very basic things about the purchase such as what event it was purchased for, what type of event it was, how much the tickets cost and when the event was purchase for.

Final Comments - Since this data is generally in very raw form, it also gives you an opportunity for you to create your own aggregated fields. For example, if you wanted to add a field that describes the total amount of money a person spent on ticketmaster, you can make those fields because this data is in its raw form. Also to note that each purchase and action is recorded with a timestamp so the order in this data is not lost and you can do time series predictions.

Section 2 - Themes and Questions

Customer Segmentation - Here are ticketmaster we have many types of customers and potential customers. Having a better understand of our customers will help us market and improve our services.

Use Case- The group that we are most interested in identifying at ticketmaster are the true fans. True fans are people who buy tickets only from ticketmaster and who after buying tickets always go to the event instead of selling them on the resale market. Among our customers we also have other groups such as scalpers- people who buy tickets early and sell later at a much higher price. Come up with a way to segment ticketmaster user into groups for example nearest neighbor clustering, and try to identify a) who are the true fans b) what the other categories may be EX: one time buyers. c) once you identified your groups, based on the data, how shoudl we target market those groups?

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 Datascience division is very interested in insights as to how to price our tickets.

Use Case- Right now as an artist travels across the country to tour, the prices for seating are generally the same everywhere. Is there a way you can help us set the price? If we can rank geographic locations in general terms of popularity, or perhaps if you figure out certain genre of music is popular in a certain area, we can help our clients such as Selena Gomez price her tickets optimally across the United States. It would be a plus if you could also suggest a specific way or formula to price our tickets
such as: in location A with Genre B and Factor C we should general price tickets 20% higher.

Marketing Optimization - At Ticketmaster we recently started a new project that tries to automate the process of Ad generation. This process is very im-
portant to us because before every advertisement you see on Google was manually generated.

Use Case- Right now our SEM (Search Engine Marketing) automation process is still very primitive. First, we do not have a lot of logic with generating keywords. Can you come up with a way perhaps using third party resources to expand out keyword list for a specific event? For example, events with keywords Duke vs. UCLA Men Basketball can be expanded to include keywords such as : WSU, wazzu, Bruins

Section 3- Notes on Sampling

Purchase Data

The Purchase Data was sampled first across the years 2005-2016 by purchase. Then it was sampled by event, meaning only certain events were included. Then we added some special cases such as cases where we knew the distance to venue, cases where we knew the occupation and gender of a customer, and cases where the same purchaser had multiple transactions. Much of the data that was added was drawn from the years 2015 - 2016 so the data is NOT uniformly distributed through the years. The data is concentrated around the years 2015-2016 to make the datasets merge more easily. Just because you see more data for certain year does not mean there were more purchases.

Adwords Data

This dataset was simply a sample of data from the 2015 to the present.

User Behavior (GA Data)

This data was sampled from 1-2 months worth of data from late 2015 to early 2016.