

Project Overview

Test Assignment

September, 29 2015

EXECUTIVE SUMMARY

One of our clients hosted a major event in 2014. Registrants used web-based software to register for the event, and for the ones who ended up attending the event, they participated in various sessions and visited various booths. We've been given the registration, session scan, and booth scan data for this event, and our job is to tell a story about the types of people who attended and identify any predictive patterns with which the marketing teams can use while preparing next year's event.

DELIVERABLES

For this test assignment, please help us do the following:

1. Create a clean, merged data set from all the files in the 'Session Activity Data' folder, all of the files in the 'Booth Scan Data' folder, and the single file in the 'Registration Data' folder. The new data set must include at least the fields in red text. The remaining fields are examples of other fields you can include in your data set. Many of the fields might require normalization.

Field	Description
Registration ID	Unique identifier in registration data
First Name	First name of attendee
Last Name	Last name of attendee
Status	Often called Status or Registration Status. Defines whether a

Project Overview

	registrant actually attended. Values will likely require normalization.
Registration Date	Date/time registered
Job Role	Field to provide summary of Titles. May not exist and need to be created. If it exists, will likely require normalization.
Registration Type	Registration type defines what kind of registrant (i.e., staff, customer, vendor, press, analyst, exhibitor)
Phone	Phone, cell, mobile
Salutation	Mr., Miss, Dr., etc...
Company	Company name
Email	Email address

- a. Note: Each of the files in the 'Session Activity Data' folder and 'Booth Scan Data' folder appear to follow the same structure, but contain columns not captured in the registration data file provided to us. Moreover, many columns appear to be superfluous to our end analysis.
- b. Note: The file names in the 'Session Activity Data' folder contain the names of the specific sessions that the registrants attended. This also appears to be the case for the data files in the 'Booth Scan Data' folder. In the new, clean data set you make, we'll also want to probably include a standardized field/column name of the session and booths each registrant attended.
- c. Note: We need to make sure there's a unique identifier for each attendee; however we might not be able to rely on the badge ID or Registrant ID alone. (One method to create a unique ID for each attendee is to concatenate their first name, last name, and email)
- d. Note: A quick glance at the structure of the data files in the 'Session Activity Data' and 'Booth Scan Data' folders indicate that the participants' registration data is what's being collected by the scanning device technology used at each session and booth; that is, when a registrant attended a session or visited a

Project Overview

booth and scanned his/her badge, the device recorded his/her registration data as a means to record attendance.

2. Using your new data set, tell a basic story about the attendees and their behavior.

Some motivating questions to consider are the following:

- a. What are the combinations of sessions that attendees went to and how does this correlate with booth scans?
- b. What can we tell about the Early Bird registration category? What can we tell about the others?
- c. Are there any predictors we can identify that will help us predict a session or booth a participant will attend? If so, what statistical/machine learning model would you employ?

Required format for results

For this assignment, please prepare a report using the structure in 'Report Template.pdf', located in your dropbox folder (see below).

COMMUNICATION/DATES

Method of file/data/document transfer

You will be given access to a shared Dropbox folder, where you will be able to access the data for this assignment, as well as upload your final report, clean data set .csv file, and source code.

1. Data Folder - This folder contains the data needed for this assignment. You'll also be required to output your final data set as a CSV file in this folder. Please title your csv file as, 'YourLastNameYourFirstName-final-data-set.csv'
2. Report Folder - This folder contains the report template, and this will be the folder in which you place your final report. Please title your final report, 'YourLastNameYourFirstName-test-assignment'
3. Source Code Folder - This folder is where you'll place the code you created for this project

Project Overview

Contact

For any questions or comments, your contact is:

Joshua Beach

Skype: abeachedjosh

Email: jbeach@insightxm.com

Check-in dates

Not required, but we are trying to get a sense of how you'll work and communicate. Please don't feel like you have to work in isolation. We work as a team, and we want you to actively communicate if you have any questions. We'd like to hear from you about how it's going.

Hours limitations

Work as many hours as you need/feel are appropriate with the fixed price amount.

There's no time limit, but please use the Odesk time tracker while working on this assignment.

Deadline

Please deliver your results before:

Monday, October 5, 2015, 5pm EDT