



The Analytics Edge (15.071x)

Monday, April 14, 2014

Knowledge • 1,685 teams

Finished

Monday, May 5, 2014

Dashboard

Home

Data

Make a submission

Information

Description

Evaluation

Rules

Timeline

Forum

Leaderboard

Public

Private

My Team

My Submissions

Leaderboard

1. Ed53
2. xboard
3. Xurfer
4. jcen
5. fernando nogueira
6. sl02
7. xaibeing
8. Peter 3
9. Zhuz
10. shaunbrophy

Forum (199 topics)

Final standings are in
6 hours ago

Getting 0.784
6 days ago

Brand new 0.78277 auc. Would
love to try to push further...
8 days ago

Request to share your code to all
the top 20s on leaderboard
8 days ago

Anyone here from Dubai?
12 days ago

Did -1, 0, 1 dummy code ERROR
actually put people in the top 10?
12 days ago

[Competition Details](#) » [Get the Data](#) » [Make a submission](#)

Data Files

File Name	Available Formats
train	.csv (1.90 mb)
test	.csv (831.13 kb)
Questions	.pdf (41.07 kb)
sampleSubmission	.csv (32.59 kb)

File descriptions

Here is a description of the files you have been provided for this competition:

- **train.csv** - the training set of data that you should use to build your models
- **test.csv** - the test set that you will be evaluated on. It contains all of the independent variables, but not the dependent variable.
- **sampleSubmission.csv** - a sample submission file in the correct format.
- **Questions.pdf** - the question test corresponding to each of the question codes, as well as the possible answers.

Data fields

- **UserID** - an anonymous id unique to a given user
- **YOB** - the year of birth of the user
- **Gender** - the gender of the user, either Male, Female, or not provided
- **Income** - the household income of the user. Either not provided, or one of "under \$25,000", "\$25,001 - \$50,000", "\$50,000 - \$74,999", "\$75,000 - \$100,000", "\$100,001 - \$150,000", or "over \$150,000".
- **HouseholdStatus** - the household status of the user. Either not provided, or one of "Domestic Partners (no kids)", "Domestic Partners (w/kids)", "Married (no kids)", "Married (w/kids)", "Single (no kids)", or "Single (w/kids)".
- **EducationLevel** - the education level of the user. Either not provided, or one of "Current K-12", "High School Diploma", "Current Undergraduate", "Associate's Degree", "Bachelor's Degree", "Master's Degree", or "Doctoral Degree".
- **Party** - the political party of the user. Either not provided, or one of "Democrat", "Republican", "Independent", "Libertarian", or "Other".

teams

players

entries

- **Happy** - a binary variable, with value 1 if the user said they were happy, and with value 0 if the user said that were neutral or not happy. This is the variable you are trying to predict.
- **Q124742, Q124122, . . . , Q96024** - 101 different questions that the users were asked on Show of Hands. If the user didn't answer the question, there is a blank. For information about the question text and possible answers, see the file Questions.pdf.
- **votes** - the total number of questions that the user responded to, out of the 101 questions included in the data set (this count does not include the happiness question).