



## TalkingData Mobile User Demographics

Get to know millions of mobile device users

\$25,000 · 1,689 teams · a year ago

[Overview](#)[Data](#)[Kernels](#)[Discussion](#)[Leaderboard](#)[Rules](#)[Late Submission](#)

### Competition Data

[Edit](#)

app\_events.csv.zip

**gender\_age\_test.csv.zip** 1.05 MB[Download](#)

app\_labels.csv.zip

events.csv.zip

gender\_age\_test.csv...

gender\_age\_train.csv...

label\_categories.csv...

phone\_brand\_device\_m...

sample\_submission.cs...

### Data Description

In this competition, you are going to predict the demographics of a user (gender and age) based on their app download and usage behaviors.

The Data is collected from TalkingData SDK integrated within mobile apps TalkingData serves under the service term between TalkingData and mobile app developers. Full recognition and consent from individual user of those apps have been obtained, and appropriate anonymization have been performed to protect privacy. Due to confidentiality, we won't provide details on how the gender and age data was obtained. Please treat them as accurate ground truth for prediction.

The data schema can be represented in the following chart:

### File descriptions

- **gender\_age\_train.csv, gender\_age\_test.csv** - the training and test set
  - group: **this is the target variable you are going to predict**

- **events.csv, app\_events.csv** - when a user uses TalkingData SDK, the event gets logged in this data. Each event has an event id, location (lat/long), and the event corresponds to a list of apps in app\_events.
  - timestamp: when the user is using an app with TalkingData SDK
- **app\_labels.csv** - apps and their labels, the label\_id's can be used to join with label\_categories
- **label\_categories.csv** - apps' labels and their categories in text
- **phone\_brand\_device\_model.csv** - device ids, brand, and models
  - phone\_brand: note that the brands are in Chinese (translation courtesy of user [fromandto](#))
    - 三星 samsung
    - 天语 Ktouch
    - 海信 hisense
    - 联想 lenovo
    - 欧比 obi
    - 爱派尔 ipair
    - 努比亚 nubia
    - 优米 youmi
    - 朵唯 dowe
    - 黑米 heymi
    - 锤子 hammer
    - 酷比魔方 koobee
    - 美图 meitu
    - 尼比鲁 nibilu
    - 一加 oneplus
    - 优购 yougo
    - 诺基亚 nokia
    - 糖葫芦 candy
    - 中国移动 ccmc
    - 语信 yuxin
    - 基伍 kiwu
    - 青橙 greeno
    - 华硕 asus
    - 夏新 panosonic
    - 维图 weitu
    - 艾优尼 aiyouni
    - 摩托罗拉 moto
    - 乡米 xiangmi
    - 米奇 micky
    - 大可乐 bigcola
    - 沃普丰 wpf
    - 神舟 hasse
    - 摩乐 mole

- 飞秒 fs
- 米歌 mige
- 富可视 fks
- 德赛 desc
- 梦米 mengmi
- 乐视 lshi
- 小杨树 smallt
- 纽曼 newman
- 邦华 banghua
- E派 epai
- 易派 epai
- 普耐尔 pner
- 欧新 ouxin
- 西米 ximi
- 海尔 haier
- 波导 bodao
- 糯米 nuomi
- 唯米 weimi
- 酷珀 kupo
- 谷歌 google
- 昂达 ada
- 聆韵 lingyun

**sample\_submission.csv** - a sample submission file in the correct format