

## 数据说明

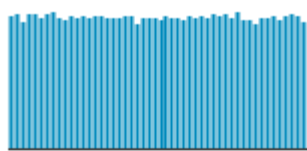
当你走过街角咖啡馆的门时，没有什么比你最爱的饮料迎接更令人欣慰了。一个体贴的咖啡师知道你每周三早上 8 点 15 分就要喝一杯香草拿铁，但在数码空间里，产品要满足你的个性化体验要困难得多。

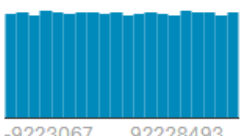
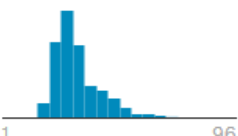
从中国最大的第三方移动数据平台收集到中国每天活跃的 5 亿移动设备中 70% 以上的数据，请你尝试着对此数据集开展数据分析。

数据收集自集成在移动应用程序中的 SDK，已获得这些应用程序的个人用户的完全认可和同意，并已进行了适当的匿名以保护隐私。由于保密性的原因，不会提供有关如何获取性别和年龄等数据的详细信息。

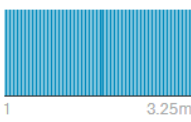
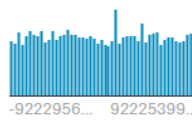
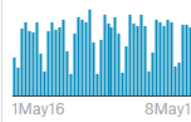
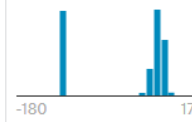
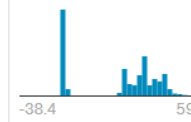
文件说明：

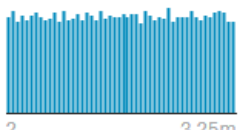



- **gender\_age\_train.csv, gender\_age\_test.csv**: 性别、年龄训练数据和测试数据。文件预览如图：

gender_age_test.csv (2.18 MB)	
#	device_id
	
-9223321...	9223069...
1	1002079943728939269
2	-1547860181818787117
3	7374582448058474277
4	-6220210354783429585
5	-5893464122623104785
6	-7560708697029818408
7	289797889702373958
8	-402874006399730161
9	5751283639860028129
10	-848943298935149395
11	6873889408535437611
12	9097600484609173263
13	7918118984690678938
14	-5723359257508971431
15	-5045999078783569038

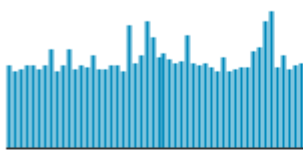
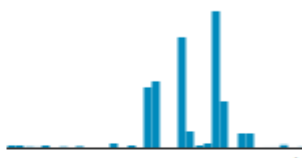
gender_age_train.csv (2.26 MB)						4 of 4 columns		V
	# device_id	A gender	# age	A group				
		M 64% F 36% Other (1) 0%		M23-26 13% M32-38 13% Other (11) 74%				
	-9223067... 92228493...		1 96					
1	-8076087639492063270	M	35	M32-38				
2	-2897161552818060146	M	35	M32-38				
3	-8260683887967679142	M	35	M32-38				
4	-4938849341048082022	M	30	M29-31				
5	245133531816851882	M	30	M29-31				
6	-1297074871525174196	F	24	F24-26				
7	236877999787307864	M	36	M32-38				
8	-8098239495777311881	M	38	M32-38				
9	176515041953473526	M	33	M32-38				
10	1596610250680140042	F	36	F33-42				

- group:性别和年龄分组，作为 target.
- events.csv, app\_events.csv: 当用户通过 SDK 访问应用程序时，就会在 events.csv 中记录此次事件，每个事件有对应的"event\_id,device\_id,timestamp,longitude,latitude"信息，分别为事件 id，设备 id，使用的时间和地点信息。在 app\_events.csv 文件中有事件对应的 app\_id. 文件预览如图：

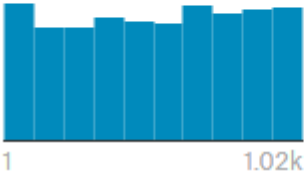
events.csv (186.38 MB)						5 of 5 columns		Views				
	# event_id	# device_id	timestamp	# longitude	# latitude							
												
	1 3.25m	-9222956... 92225399...	1May16 8May16	-180 175	-38.4 59.9							
1	1	29182687948017175	2016-05-01 00:55:25	121.38	31.24							
2	2	-6401643145415154744	2016-05-01 00:54:12	103.65	30.97							
3	3	-4833982096941402721	2016-05-01 00:08:05	106.60	29.70							
4	4	-6815121365017318426	2016-05-01 00:06:40	104.27	23.28							
5	5	-5373797595892518570	2016-05-01 00:07:18	115.88	28.66							
6	6	1476664663289716375	2016-05-01 00:27:21	0.00	0.00							
7	7	5990807147117726237	2016-05-01 00:15:13	113.73	23.00							
8	8	1782450055857303792	2016-05-01 00:15:35	113.94	34.70							

app_events.csv (989.22 MB)					4 of 4 columns ▾		Vie
	# event_id	# app_id	# is_installed	# is_active			
							
	2	3.25m	-92211569... 92224881...	1	1	0	1
1	2	5927333115845830913	1	1			
2	2	-5720078949152207372	1	0			
3	2	-1633887856876571208	1	0			
4	2	-653184325010919369	1	1			
5	2	8693964245073640147	1	1			
6	2	4775896950989639373	1	1			
7	2	-8022267440849930066	1	0			
8	2	9112463267739110219	1	0			
9	2	-3725672010020973973	1	0			
10	2	7167114343576723123	1	1			

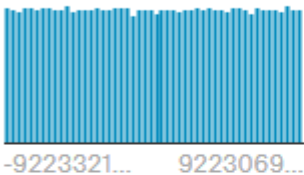
- app\_labels.csv: app\_id,label\_id. 根据 label\_id 可以在 label\_categories.csv 中查看 app 的分类。文件预览如图：

app_labels.csv (10.67 MB)		
	# app_id ▾	# label_id ▾
		
	-9223281... 92233179...	2 1.02k
1	7324884708820027918	251
2	-4494216993218550286	251
3	6058196446775239644	406
4	6058196446775239644	407
5	8694625920731541625	406
6	8694625920731541625	407

- label\_categories.csv: label\_id,category。文件预览如图：

label_categories.csv (16.06 KB)		
#	label_id	A category
		unknown 3% Financial Informa... 0% Other (834) 97%
1	1	
2	2	game-game type
3	3	game-Game themes
4	4	game-Art Style
5	5	game-Leisure time
6	6	game-Cutting things

- phone\_brand\_device\_model.csv: device\_id,phone\_brand,device\_model, 分别是设备 id, 手机品牌, 设备型号。文件预览如图:

phone_brand_device_model.csv (6.4 MB)			
#	device_id	A phone_brand	A device_model
		小米 23% 三星 18% Other (130) 59%	红米note 4% MI 3 3% Other (1598) 93%
1	-8890648629457979026	小米	红米
2	1277779817574759137	小米	MI 2
3	5137427614288105724	三星	Galaxy S4
4	3669464369358936369	SUGAR	时尚手机
5	-5019277647504317457	三星	Galaxy Note 2
6	3238009352149731868	华为	Mate
7	-3883532755183027260	小米	MI 2S
8	-2972199645857147708	华为	G610S

小建议:

- 可以预测年龄和性别
- 可以分析最受欢迎的 app
- 可以分析最受欢迎的手机品牌

.....