

Dilan's Data Structure Documentation

In this data set, you will find the raw user activity data from Dilan's blog between 1 January 2018 and 31 March 2018. It's a log with ~600.000 rows.

```
2018-01-08 20:47:59;read;country_2;2458165868;SEO;North America
2018-01-08 20:48:15;read;country_2;2458165869;AdWords;South America
2018-01-08 20:48:33;subscribe;2458161488
2018-01-08 20:48:53;read;country_7;2458163241;South America
2018-01-08 20:48:54;read;country_5;2458155418;Asia
2018-01-08 20:49:17;read;country_7;2458165848;North America
2018-01-08 20:50:48;read;country_7;2458165870;AdWords;Asia
2018-01-08 20:51:01;read;country_7;2458165871;SEO;North America
2018-01-08 20:53:49;read;country_4;2458162459;Asia
2018-01-08 20:55:46;read;country_2;2458165872;AdWords;Asia
2018-01-08 20:56:25;read;country_5;2458165873;AdWords;North America
2018-01-08 21:00:24;read;country_8;2458165874;AdWords;Europe
2018-01-08 21:00:27;read;country_5;2458165875;AdWords;Europe
2018-01-08 21:01:19;read;country_7;2458165876;Reddit;Asia
2018-01-08 21:01:20;read;country_5;2458152183;North America
2018-01-08 21:01:37;read;country_8;2458165877;AdWords;Asia
2018-01-08 21:01:38;read;country_4;2458160893;Europe
2018-01-08 21:02:09;read;country_2;2458165878;Reddit;Asia
2018-01-08 21:02:22;read;country_4;2458162693;North America
2018-01-08 21:02:30;buy;2458159216;8
2018-01-08 21:03:36;read;country_6;2458165611;North America
```

1. The whole data log is one big file. You don't have to worry about automation and refreshing the data.
2. This is the `.csv` file. The field separator is a semi-colon (;).
3. The first column is the date and the time of the event, in this format: `YYYY-MM-DD HH:MM:SS` (eg. `2018-03-30 23:48:48`).
4. The second column is the event type. There are three different event types:
 1. `read`
 2. `subscribe`
 3. `buy`
5. The rest of the columns depend on the event type:
 1. For `read` there are two options:
 10. If this is the first visit from the given user, then the next fields are: `country; user_id; source; topic`
 11. If it's a returning reader, then these are the fields: `country; user_id; topic`
 2. For `subscribe` there is only one additional column: `user_id`
 3. For `buy` there are two more columns: `user_id` and the `price` of the product the user bought.

6. Examples for each event types:

- First time reader: 2018-03-30 23:59:56;read;country_5;2458361283;Reddit;Asia
- Returning reader: 2018-03-31 05:04:50;read;country_4;2458361208;Africa
- Subscriber: 2018-03-30 23:48:48;subscribe;2458173588
- Purchase: 2018-03-30 23:11:30;buy;2458339835;80

Find out more by downloading and exploring the data set by yourself.

7. The country column has 8 different values:

- country_1
- country_2
- country_3
- ...
- country_8

8. The source column has 3 different values:

- Reddit
- AdWords
- SEO

9. The topic column:

- Africa
- Europe
- South America
- North America
- Australia
- Asia

10. The course price column:

- 8
- 80