

TransWorld Data Analysis - Bash Code

File name for the raw data: transworld_data.csv

*-Used the sed function to make separations between columns clearer.
Since the space between date and time needed to be replaced by a semi-colon, the space between the words 'North'/'South' and America needed to be replaced with an underscore (i.e., North America → North_America).
This is so that a semi-colon doesn't get placed between them by mistake.*

```
sed 's/North America/North_America/g' transworld_data.csv > tda.csv
sed 's/South America/South_America/g' tda.csv > tdb.csv
sed 's/ /;/g' tdb.csv > tdc.csv
rm tda.csv | rm tdb.csv
mv tdc.csv transworld_data_properformat.csv
```

*-Separated raw data into sections according to category
first (first visits), return (returning visits-for reading), subscribe (subscribers), purchase (purchases)*

-Kept only rows with SEO, AdWords and Reddit from the entire data set to create a new file for only first visits to the website

```
grep 'SEO\AdWords\Reddit' transworld_data_properformat.csv > first_visit.csv
```

-Took rows that are 'read' only and put them into a new file

```
grep read transworld_data_properformat.csv > read_only.csv
```

-Kept only rows that do not contain 'SEO', 'AdWords' and 'Reddit' from 'read only' and created a new file for returning visits/reading sessions

```
grep -v 'SEO\AdWords\Reddit' read_only.csv > return.csv
```

-Kept only rows that contain 'subscribe' to create a new file for subscribers

```
grep subscribe transworld_data_properformat.csv > subscribe.csv
```

-Kept only rows that contain 'buy' to create a new file for purchases

```
grep buy transworld_data_properformat.csv > purchase.csv
```

-Placed data into SQL, created SQL tables for each portion of the dataset

```
psql -U datafolder -d postgres
```

```
CREATE TABLE first_visit (
the_date DATE,
the_time TIME,
event_type TEXT,
country TEXT,
```

```
user_id BIGINT,  
source TEXT,  
topic TEXT);
```

```
CREATE TABLE return_visits (  
the_date DATE,  
the_time TIME,  
event_type TEXT,  
country TEXT,  
user_id BIGINT,  
topic TEXT);
```

```
CREATE TABLE subscribe (  
the_date DATE,  
the_time TIME,  
event_type TEXT,  
user_id BIGINT);
```

```
CREATE TABLE purchase (  
the_date DATE,  
the_time TIME,  
event_type TEXT,  
user_id BIGINT,  
price INTEGER);
```

-Placed data into the SQL tables

```
COPY first_visit FROM '/home/datafolder/first.csv' DELIMITER ';';  
COPY return_visits FROM '/home/datafolder/return.csv' DELIMITER ';';  
COPY subscribe FROM '/home/datafolder/subscribe.csv' DELIMITER ';';  
COPY purchase FROM '/home/datafolder/purchase.csv' DELIMITER ';';
```