

# Wrangle Report

Data wrangling is the process of gathering, selecting, and transforming data to answer an analytical question.

First of all, we made gathering for data, we obtained data (we downloaded this data set from Kaggle [Airbnb New User Bookings | Kaggle](#)).

We imported (test\_users, sessions, train\_users\_2) files to Jupyter Notebook Which is the programming environment we chose to code in it.

We named test\_users as users, sessions as users\_sessions, and train\_users\_2 as users\_cut.

Secondly, we assessed our data and founded 8 quality issues and 2 tidiness issues.

## Quality issues

We founded 13 quality issues: -

### ***1) In 'users' table we founded 6 issues which are: -***

- Ages are identified as float and this column must be identified as integers.
- We founded missing records in users table, there are 62096 records instead of 275547 records.
- We founded 186639 Missing values in column named 'date\_first\_booking', there are 186639 nan values in that column.
- Missing values in the column named 'age', There are 116866 nan values in that column.
- In column 'first\_affiliate\_tracked' we founded 6085 nan values which can't be solved because they are nan categorical values.

- In column 'country destination' we founded 62096 missing values which can't be solved because they are nan categorical values too.

## **II) In 'users\_sessions' table we founded 7 issues: -**

- We founded multiple records for all records.
- Column named 'secs\_elapsed' must be identified as integer not as a float.
- We founded missing values in column named 'user\_id'.
- We founded 36183 nan values in column named 'action'.
- We founded 456198 nan values in column named 'action\_type'.
- We founded one missing value in column named 'device type', there is 1 nan value.
- -We founded 50836 missing values in column named 'secs\_elapsed' they are nan values.

## **Tidiness issues**

We founded two tidiness issues: -

- Column named 'total\_secs\_elapsed' should be part of 'users' table.
- Columns named 'action\_type' and 'action\_details' in 'users\_sessions' table must be in one column called 'action\_type\_details'.