EDA and Data Visualization 11/02/2022

You will find here some insights about the data set you gave us.

General description

Size of the data:	10M lines, 6M unique tweets ID
date range:	2021-02-01 12:00:03 - 2021-09-01 11:59:5
latitude range:	-54.69 : 12.42
longitude range:	-91.17 : -32.18

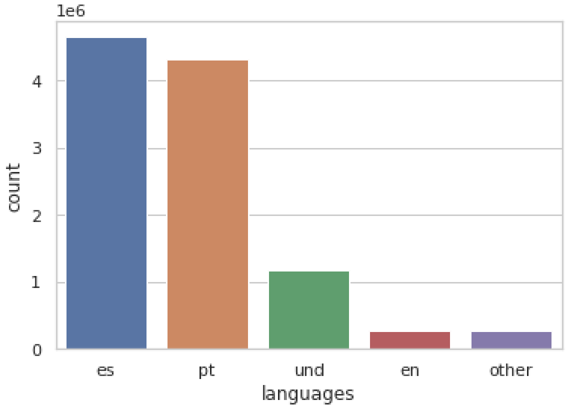
There were 2M tweets who were presents in various CSV with different locations.

Categorical features

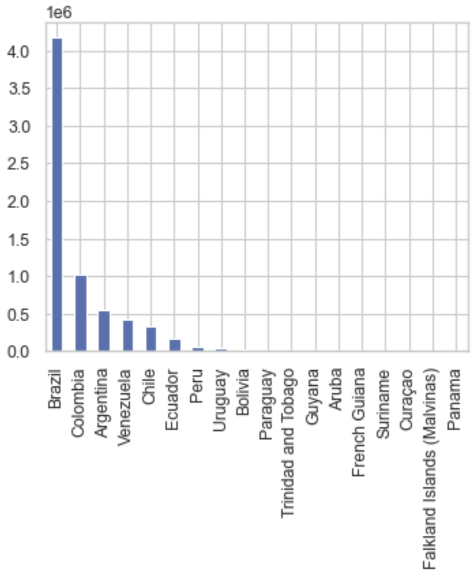
Languages:

Other: various languages

Und languages: undetermined languages

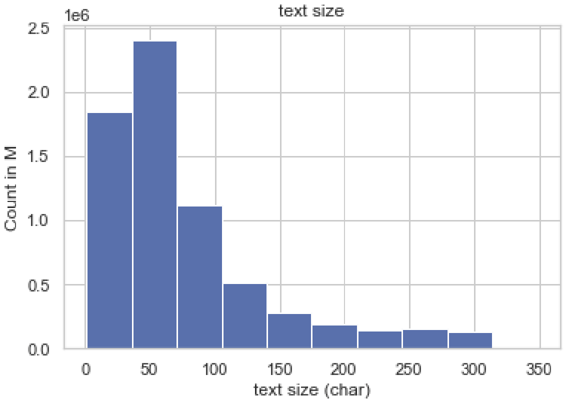


Countries:



Text length:

Some tweets mentions a lot of people and have size longer than 350 but there are a negligible part of the data.

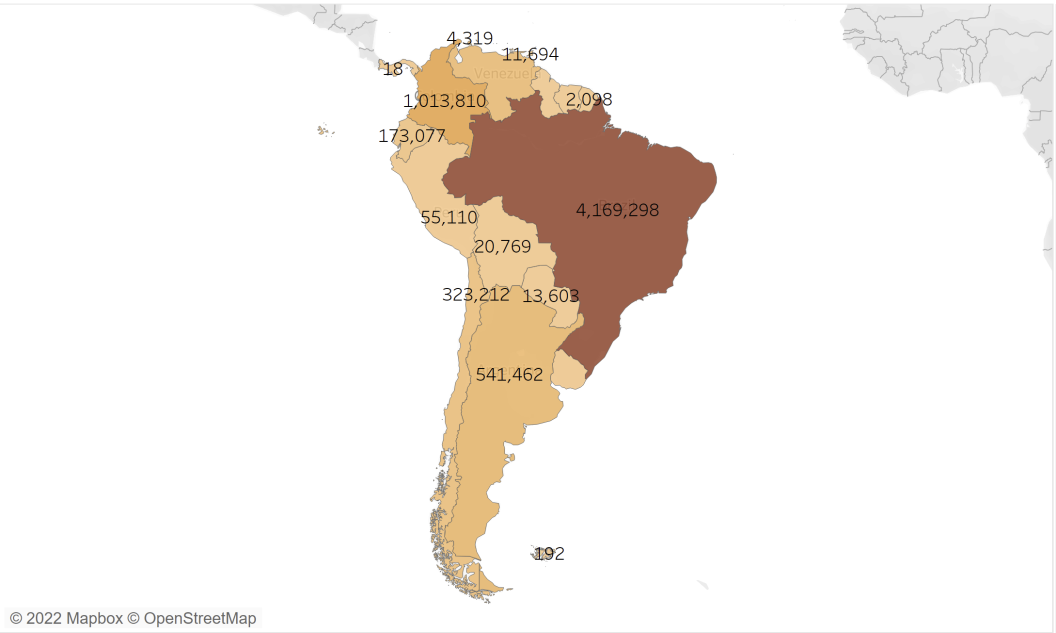


Location:

In **99%** of the cases, users send their tweets from the **same bounding box**.

In **47%** of the cases, users send their tweets from the **same geolocation** (long,lat).

#id by Country


4,319
11,694
18
Venezuela
1,013,810
2,098
173,077
4,169,298
55,110
20,769
323,212
13,603
541,462
192
© 2022 Mapbox © OpenStreetMap