Report part-2

Cleaning and processing data

Firstly error showed up because maybe “date” word is reserved by hive so I can’t use that for creating column name , so I modified it to “tweet\_date” and it worked :

Text

Description automatically generatedGraphical user interface, text

Description automatically generated

For convenience I dropped this table and changed the table name to BigData\_tweets

Graphical user interface, text, application

Description automatically generated with medium confidence

I cleaned the tweets using the following 4 commands:

Text

Description automatically generatedText

Description automatically generated