Problem 1-a:

Create Tables: Tweet , User





Problem1-b:

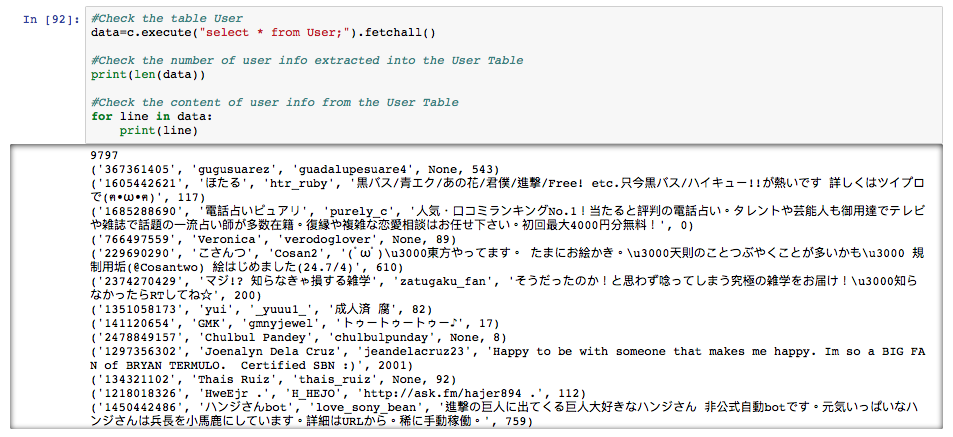




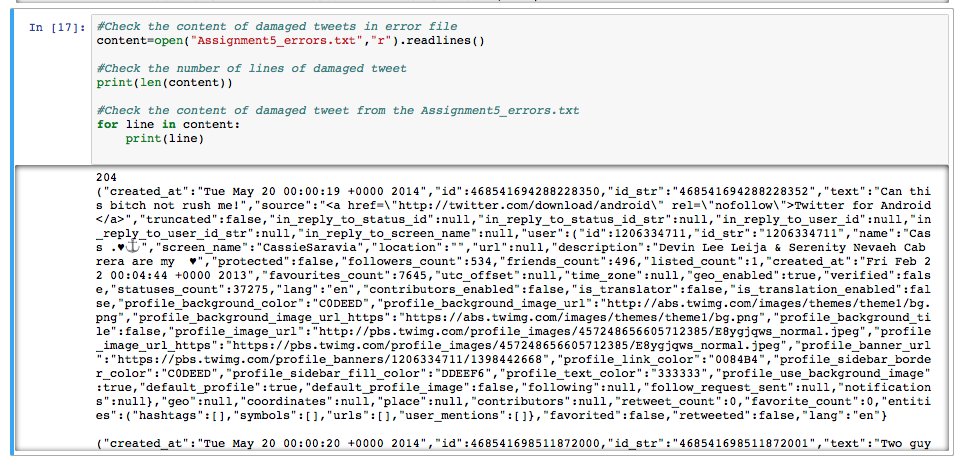
Table Tweet includes 9797 Records



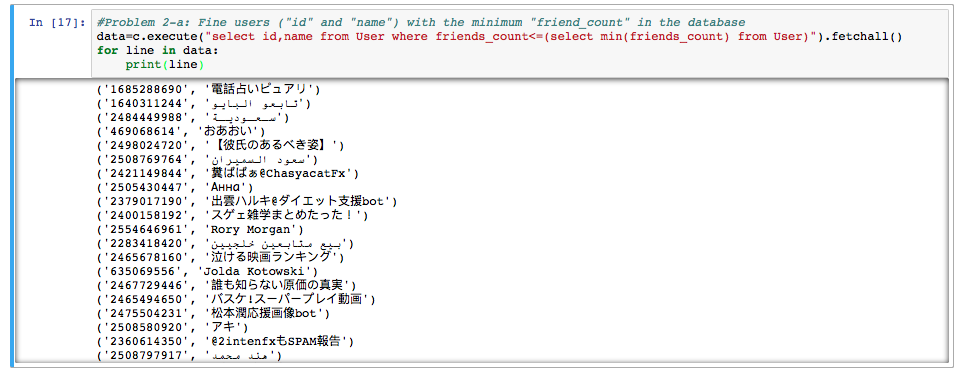
Table User includes 9797 Records



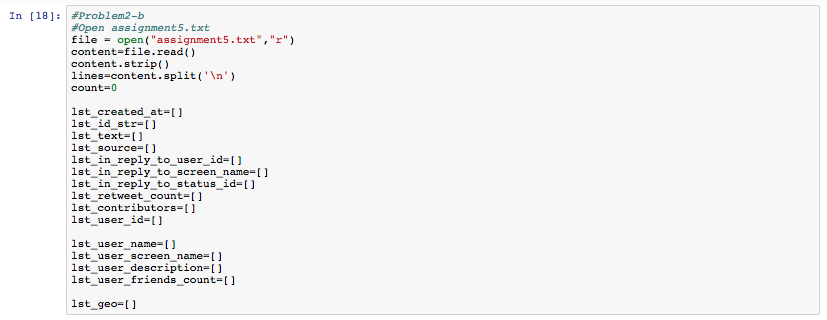
Assignment5\_errors.txt includes 204 damaged tweet



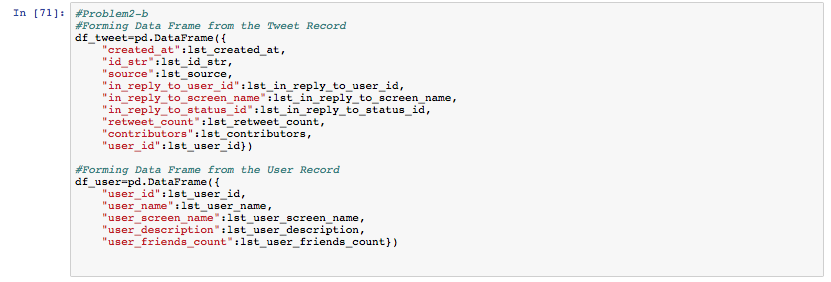
Problem 2a:

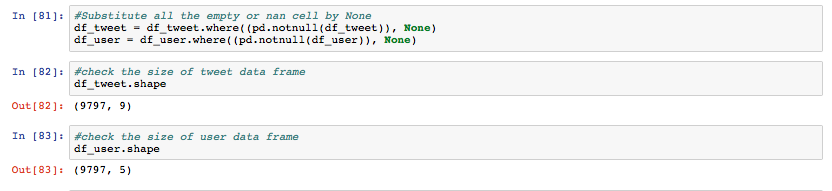


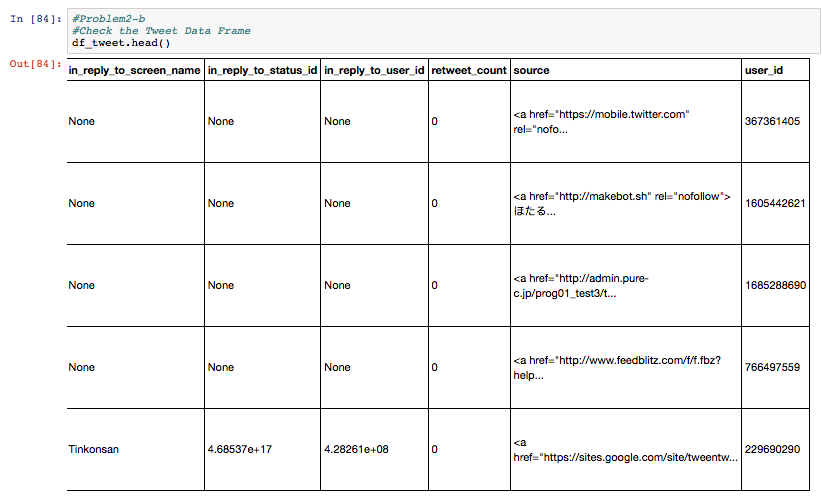
Problem 2b:

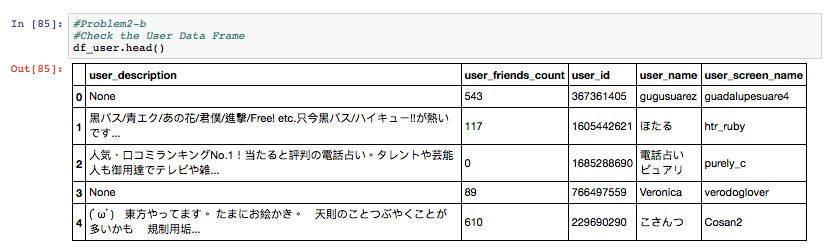




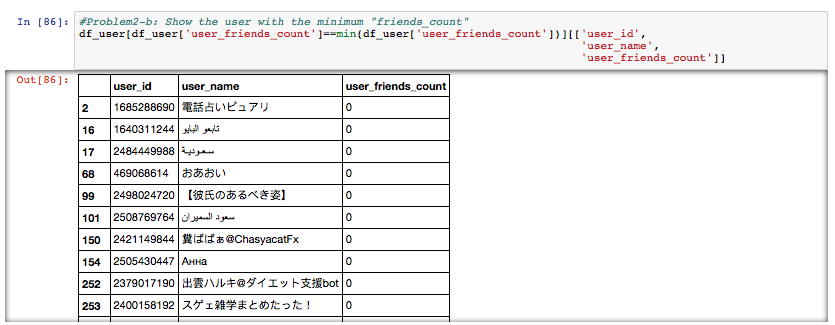








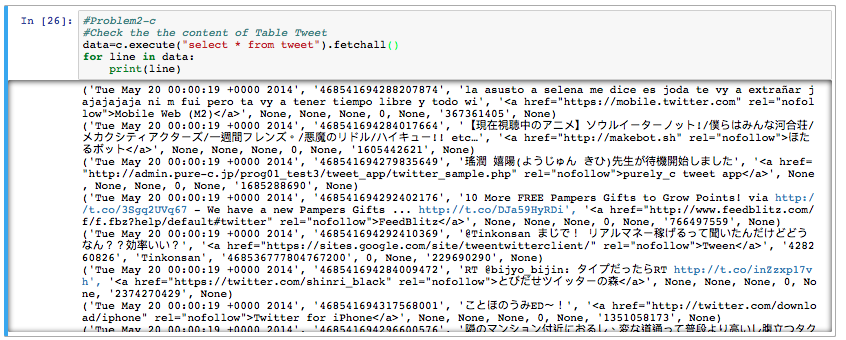
Question 2-b: Final Answer

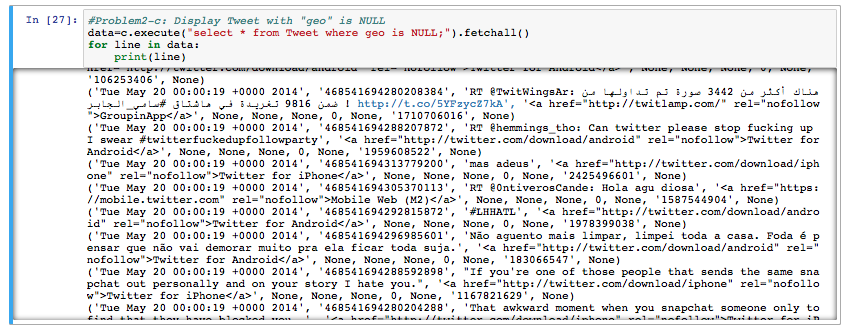


Problem 2-c: Add ‘geo’ attribute to the Table Tweet









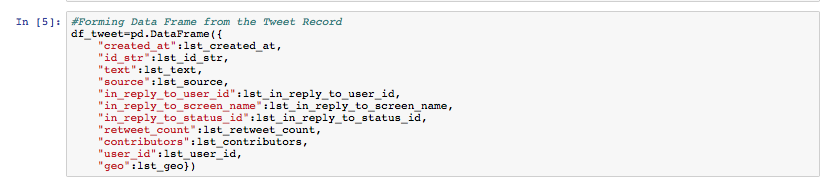
Problem 2-d: Add “geo” attribute to the Data Frame

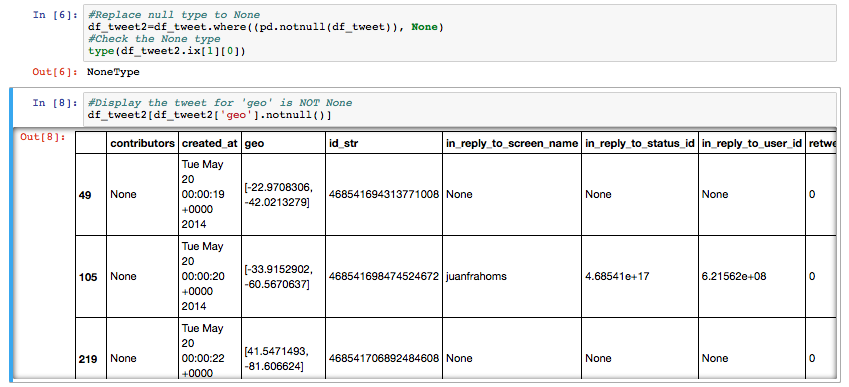




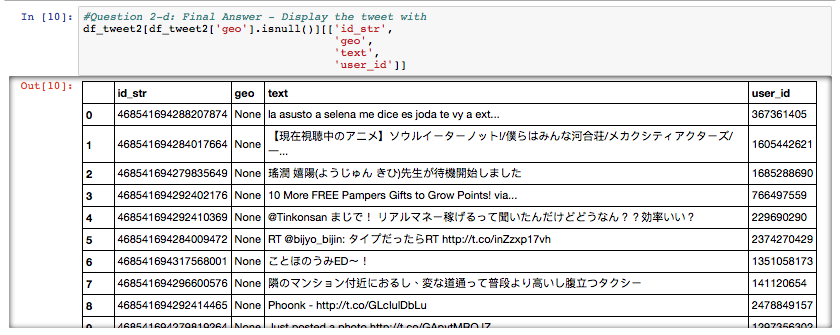
Create the Data Frame with ‘geo’ attribute:







Problem 2-d: Final Answer



Problem 3-a:

Longest Text in Tweet



Shortest Text in Tweet:



Problem 3-b: Find the top 5 most frequent terms from the ‘text’ of the tweets

Note: The original tweet2 data frame is formed from Problem 2-d

