

# NLP Interestship

Module 3 : Understanding problem statement, understanding the dataset, text preprocessing.

## Understanding the data first

In this project we will be predicting whether a question asked on Quora is sincere or not.

An insincere question is defined as a question intended to make a statement rather than look for helpful answers.



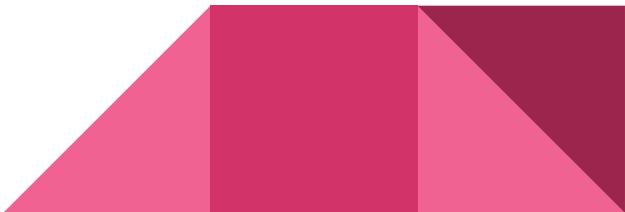
# Insincere Question Characteristics

1. Has a non-neutral tone
  - \* Has an exaggerated tone to underscore a point about a group of people
  - \* Is rhetorical and meant to imply a statement about a group of people



# Insincere Question Characteristics

## 2. Is disparaging or inflammatory

- \* Suggests a discriminatory idea against a protected class of people, or seeks confirmation of a stereotype
  - \* Makes disparaging attacks/insults against a specific person or group of people
- 

# Insincere Question Characteristics

## 3. Isn't grounded in reality

- \* Based on false information, or contains absurd assumptions



## Steps used

**Step 1: Punctutation Removal**

**Step 2: Lowering the text**

**Step 3: Tokenization**

**Step 4 : Stop Words Removal**

**Step 5: Lemmatization**



The background is a solid pink color. In the top right corner, there is a decorative pattern of overlapping geometric shapes, including triangles and squares, in various shades of pink and magenta.

THANK YOU