**1. What is the goal of EDA (exploratory data analysis)?**

The goal of EDA is process, modify and re-structure the data and make sure the quality is and integrity is augmented. It involves 3 classifications namely: data cleaning, data exploration and feature engineering.

**2. Suppose that you are given a dataset of customer product reviews for an e-commerce company. Each review is scored as a Likert-style survey item where 1 indicates a negative sentiment about the product and a 5 is positive. These reviews are collected on the company's website. a. What problems do you expect to find in the raw data? b. If your task is to build features that give information about customer sentiments, how would you approach this task and what kind of methods would you apply to accomplish it? c. Try to identify some potentially useful features that you might derive from the raw data. How would you derive them and how would you assess the usefulness of those features?**

**a. What problems do you expect to find in the raw data?**

-Possibility of null values.

-Numbers like “45” or 0 which don’t fall within the range.

-Emojis and abbreviations.

-Punctuations not properly placed.

**b. If your task is to build features that give information about customer sentiments, how would you approach this task and what kind of methods would you apply to accomplish it?**

-extract key words like “good”, “smell”, and correlate them with the rating.

-Extract most occurring words and sentiment.

**c. Try to identify some potentially useful features that you might derive from the raw data. How would you derive them and how would you assess the usefulness of those features?**

-Potential positive keywords and potential negative keywords.

-When a keyword appears more frequently in a sentiment and less frequently in another, that feature can be used to define the sentiment.