

Coding Challenge: Important Phrase Extraction

- Imagine you are product manager for a smartphone manufacturer who uses our dashboard to analyze the smartphone market.
- By doing that you realize that the attribute 'Camera' is very important, and you want to understand what people are talking about when evaluating this attribute.
- For doing this your task will be to **research and test different approaches to extract important phrases** from sentences of an attribute.
- The ultimate goal is to **find an approach that reveals interesting insights to a product manager** (e.g., the most important phrases within an attribute). There is no right or wrong to this question, so feel free to move in any direction that supports this goal.
- A phrase does not need to be a single word, it could also related to a larger chunk within a sentence.

The rules



Three datasets with different attributes for testing.



Use Python or R in the IDE of your choice.



No slides needed. Just present us your code together with your findings and results.



Do you have questions or are you stuck? Please ask us via email – we are here to help.

Information about the datasets

- Together with this challenge you will receive 3 Excel files – each one contains sentences where users express opinions about the attribute
 - **attribute_name**: The name of the attribute
 - **indicator_phrase**: The indicator phrase within the sentence that triggered the algorithm to detect the attribute
 - **sentence**: Sentence text of the attribute
 - **sentiment**: -1 for negative, 0 for neutral and 1 for positive opinions. This column is not needed but could potentially be interesting
- Datasets (we recommend starting with **Smartphones** before moving to other datasets):
 - **Smartphones_Camera.xlsx**: 10k sentences where the attribute ‘Camera’ of smartphones is evaluated
 - **Nutritiondrinks_Ingredients.xlsx**: 10k sentences where the attribute ‘Ingredients’ of probiotic capsules is evaluated
 - **Cafes_Atmosphere.xlsx**: 10 sentences where the attribute ‘Atmosphere’ of cafes in Germany is evaluated

Schedule



8 AM: You have just received the coding challenge. We hope you are excited and ready for the day.



9:30 AM: 15 minutes Q&A if something is unclear.



During the day: Takes us much time as you want to work on the problem independently. Don't forget to take some time to also think of presenting the results.



6:00 PM: We are looking forward to your results. No slides needed but make sure to structure your thought to explain us your approach and solution.