ASSIGNMENT – GENERAL RULES

Text Mining and Sentiment Analysis – 17711-ENG

- The assignment accounts for a maximum of 3 extra points on the grade of the final exam based on the quality of the work.

- Students should work in groups of **2 or 3 people**.

- The assignments should be submitted through the e-learning page and should consist of two files (for each group):
    - A **PDF** with the report for the case study (up to **5 pages**, including figures);
    - The **R script code** for the case study.

- Only one member of the group should submit the assignment by **July 4th** through the dedicated Moodle submission page. The name(s) of the file(s) is: SURNAME1_SURNAME2(_SURNAME3). Additionally, write the full names at the beginning of each file.

- The report for the case study should contain the following sections:
    1. **Introduction** (where the research question is discussed)
    2. **Data** (a short description of the data)
    3. **Results and Discussion** (text mining)
    4. **Conclusions** (about the results, soundness of methodologies and implications)

ASSIGNMENT – INSTRUCTIONS
Text Mining and Sentiment Analysis – 17711-ENG

- **Objective**: choose a product from https://www.amazon.co.uk/, for at least 200 reviews scrape the review title, the corresponding text, and stars. Analyse the sentiment of the reviews.

- **Follow these steps:**
    - **Clean** the data.
    - Provide **one or two visualizations** you think are more relevant to describe the data.
    - Implement dictionary-based analyses: choose **one lexicon** and compare the **tidy approach** with the **udpipe** one. Compare the differences of the two approaches, considering the number of stars in the reviews too. Also discuss the effect of stop-words removal, and dictionary composition with respect to the context.
    - Report the results using **some** of the visualization tools studied during the course.