

Libraries used

- **Pandas** is a Python library for data manipulation and analysis, often used to work with tabular data like spreadsheets or databases. It's the backbone for managing data in the form of DataFrames.
- **Re (Regular Expressions)** allows us to find and manipulate specific patterns in strings, such as removing punctuation or matching specific text patterns.
- **NLTK (Natural Language Toolkit)** is used for text processing tasks like tokenization (splitting text into words) and removing common words (stopwords), making text analysis more efficient.
- **SpaCy** is another library for natural language processing, particularly useful for more advanced tasks like named entity recognition (NER), helping to identify and extract important elements from text.
- **TextBlob** is a simple library for sentiment analysis, allowing us to quickly determine the emotional tone (positive, negative, or neutral) of a given text.
- **Gensim** is a library for topic modeling and text similarity, with tools like Latent Dirichlet Allocation (LDA), useful for identifying topics and trends in large sets of text data.