Installation Guide – Python Notebooks

1. Notebook 1: 01_scraper_html_parser:

a. Purpose: The purpose of the notebook is to scrape Kununu for the companies of interest.

b. Python Dependencies:

- Selenium: pip install -U selenium
- Beautiful Soup: pip install beautifulsoup4
- Transformers: pip install transformers
- OS: pip install os-sys
- Regex: pip install regex
- Numpy: pip install numpy
- Pandas: pip install pandas
- Requests: pip install requests
- Time: pip install python-time
- Json: pip install jsons
- Tqdm: pip install tqdm
- Sys: pip install os-sys
- Warnings: pip install pytest-warnings
- Timeit: pip install pytest-timeit

2. Notebook 2: 02_sentiment_analysis:

a. Purpose: The purpose of the notebook is to extract the sentiment of textual data from Kununu along with with calculation of Net Promoter Score(NPS) and the aggregation and average of the scores that are provided in Kununu.

b. Python Dependencies:

- Pytorch: https://pytorch.org/get-started/locally/
    a. pip install torch torchvision torchaudio
- Scipy: pip install scipy
- Timeit: pip install pytest-timeit
- Pandas: pip install pandas
- Tqdm: pip install tqdm
- Time: pip install python-time
- Numpy: pip install numpy
- Sys: pip install os-sys
- Warnings: pip install pytest-warnings
- Matplotlib: pip install matplotlib
- DateTime: pip install DateTime
- Seaborn: pip install seaborn

3. Notebook 3: 03_topic_modelling:

a. Purpose: The Purpose of the code below is to dig deeper into the "'Gut am Arbeitgeber finde ich_plain_text' and 'Schlecht am Arbeitgeber finde ich_plain_text' colums and analyse the reviews. For this purpose, we utilise the Latent Dirichlet Allocation(LDA Model) for topic modelling and key topic extraction. LDA topic modeling is a probabilistic modeling method for classifying documents based on information related to the topic of the text in a large number of unstructured documents. By analysing the 2 columns/features, we are able to better understand the most spoken about topics within these columns.

b. Python Dependencies:

- Spacy:
    a. pip install -U spacy
    b. python -m spacy download en_core_web_sm
- Gensim: pip install genism
- pyLDAvis**:** pip install pyLDAvis
- Pandas: pip install pandas
- Scikit-Learn: pip install scikit-learn
- Numpy: pip install numpy
- Regex: pip install regex
- Pprint: pip install pprintpp

4. Notebook 4: 04_Word_Frequency:

a. Purpose: Calculate Bag of words and collaborate with the topic modelling words. Analyise the text and calculate the weights of these texts.

b. Python Dependencies:

- OS: pip install os-sys
- Numpy: pip install numpy
- Pandas: pip install pandas
- Json: pip install jsons
- Nltk: pip install nltk