1. **Mount Google Drive**
   - Imports `drive` from `google.colab` and calls `drive.mount('/content/drive')`.
   - Allows read/write access to your Drive files.
2. **Configure Paths**
   - `TXT_FILE_PATH`: Full path to your input text file (e.g. `/path/to/your/data/`).
   - `OUTPUT_DIR`: Folder where output CSVs and plots will be saved (e.g. `/path/to/your/output/`).
3. **Install Dependencies**
   Installs all necessary Python libraries for NLP, ML, and visualization.
   Ensures the environment has the tools needed before analysis.
4. **Import Modules & Set Parameters**
   Loads core Python packages and NLP tools into the session.
   Defines key variables and thresholds used later in the pipeline.
5. **Download NLTK Resources**
   Fetches NLTK datasets (`punkt`, `stopwords`, `punkt_tab`) required for tokenization and filtering.
6. **Load spaCy & SBERT Models**
   Loads or downloads the Italian spaCy model for POS tagging and lemmatization.
   Loads the SBERT model for semantic similarity calculations.
7. **Read & Clean Text**
   Reads the input file and applies text cleaning: removes punctuation, extra whitespace, and lowercases text.
8. **Tokenization (NLTK)**
   Splits cleaned text into word tokens using NLTK's Italian tokenizer.
   Prepares the data for POS tagging and further processing.
9. **POS Tagging (spaCy)**
   Annotates each token with coarse and fine-grained part-of-speech tags.
   Enables subsequent filtering and linguistic analysis based on POS.
10. **Stopword Filtering**
    Combines custom and standard Italian stopwords for token filtering.
    Preserves verbs even if they appear in the stopword list.
11. **Semantic Segmentation**
    Divides the text into coherent segments by detecting semantic shifts using SBERT similarity.
    Uses syntactic chunk boundaries as potential split points.
12. **Sentence Reconstruction & DataFrame**
    Rebuilds sentences from tokens and assigns them to segments.
    Creates and saves a DataFrame with sentence texts and metadata.
13. **Lemmatization & TTR Calculation**
    Converts tokens to lemmas and calculates vocabulary diversity metrics.
    Computes both surface-form and lemma-based type–token ratios.
14. **TF-IDF Keyword Extraction**
    Identifies important terms by computing TF-IDF scores on filtered tokens.
    Exports a sorted list of keywords for analysis.
15. **Visualization**
    Creates a scatter plot of top TF-IDF scores and a word cloud of key terms.
    Helps visualize term importance and distribution.
16. **Psychological Phenomena Extraction**
    Searches segments for regex patterns of psychological verbs and expressions.
    Records matches with context and saves them for further inspection.

17. **Frequency Analysis & Export**
    Removes duplicate matches and calculates relative frequencies per pattern.
    Saves frequency summaries to CSV.