**XTRACT 2.0 : Murder Videos Sentiment Analysis**

The YouTube links arrived in your inbox like whispers in the night, innocuous at first glance. Little did you know that they would unravel a web of intrigue so dense, it would threaten to consume you whole. Alone in your office, you clicked on the first link, expecting mundane chatter. Instead, you were met with cryptic messages, encoded in the rhythms of everyday speech. Your curiosity piqued, You delved deeper, each transcript revealing layers of deception.As the hours stretched into days, you became immersed in the mystery, your obsession growing with each passing moment. The case took hold of you, driving you to the brink of madness as you deciphered the hidden meanings behind the innocuous words.

Then, a breakthrough. A connection between the cryptic messages and a recent murder, sending a chill down your spine. The victim, a prominent figure in data science, lay lifeless, surrounded by enigmatic symbols that whispered of dark secrets.

With no one to trust but yourself, you embark on a solitary journey into the heart of the conspiracy. Each step forward felt like wading through treacherous waters, uncertainty gnawing at your resolve. But you press on, driven by a relentless pursuit of justice. With every clue you uncover, the puzzle pieces fell into place, painting a picture of betrayal and deceit. Alone once more, you watch as the city sleeps, knowing that even in the silence of the night, the truth would always find a way to surface

# **Objective**

The objective of this assignment is to extract textual transcript from the given YouTube Links and perform text analysis to compute variables that are explained below.

# **Data Extraction**

Input.csv

For each of the articles, given in the input.xlsx file, extract the article text and save the extracted article in a text file with URL\_ID as its file name.

While extracting text, please make sure your program extracts only the transcripts.

# **Data Analysis**

For each of the extracted texts from the article, perform textual analysis and compute variables, given in the output structure excel file. You need to save the output in the exact order as given in the output structure file, “Output Data Structure.xlsx”

**NOTE: YOU MUST USE PYTHON PROGRAMMING FOR THE DATA ANALYSIS**

# **Variables**

The definition of each of the variables given in the “Text Analysis.docx” file.

Look for these variables in the analysis document (Text Analysis.docx):

1. POSITIVE SCORE
2. NEGATIVE SCORE
3. POLARITY SCORE
4. SUBJECTIVITY SCORE
5. AVG SENTENCE LENGTH
6. PERCENTAGE OF COMPLEX WORDS
7. FOG INDEX
8. AVG NUMBER OF WORDS PER SENTENCE
9. COMPLEX WORD COUNT
10. WORD COUNT
11. SYLLABLE PER WORD
12. PERSONAL PRONOUNS
13. AVG WORD LENGTH

# **Output Data Structure**

**Output Variables:**

1. All input variables in “Input.xlsx”
2. POSITIVE SCORE
3. NEGATIVE SCORE
4. POLARITY SCORE
5. SUBJECTIVITY SCORE
6. AVG SENTENCE LENGTH
7. PERCENTAGE OF COMPLEX WORDS
8. FOG INDEX
9. AVG NUMBER OF WORDS PER SENTENCE
10. COMPLEX WORD COUNT
11. WORD COUNT
12. SYLLABLE PER WORD
13. PERSONAL PRONOUNS
14. AVG WORD LENGTH

Check out the output data structure spreadsheet for the format of your output, i.e. “Output Data Structure.xlsx”.

# Output:

**Your final answer would be an integer 👍  
  
Answer:**

* **Calculate the average value of each metric**
* **Final Answer = Ceiling to nearest ten’s value of {( Sum of all average values of each metric ) / No. of Metrics }**

**eg: Your answer of average of all metrics is 45.7 then your final password is 50**