1. Text Mining/Analysis can be used in:
   1. Detecting spam model
   2. Predicting stock Movements
   3. News stories categorization
   4. **All the above**
2. \_\_\_\_\_\_\_\_\_\_\_ is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights.
   1. Data mining
   2. **Text Mining**
   3. File Mining
   4. Deep Mining
3. The process of breaking out long-form text into sentences and words called?
   1. Stem
   2. Cluster
   3. Bag
   4. **Tokens**
4. Text mining is being used by large media companies, to clarify information and to provide readers with greater search experiences.
   1. **TRUE**
   2. FALSE
   3. Can be true or false
   4. Can not say
5. Typical text mining tasks include?
   1. text categorization
   2. text clustering
   3. entity relation modeling
   4. **All of the above**
6. Stemming: This refers to the process of separating the prefixes and suffixes from words to derive the root word form and meaning.
   1. **TRUE**
   2. FALSE
   3. Can be true or false
   4. Can not say
7. Most of the data in the world is in fact unstructured
   1. False
   2. **True**
8. Which of the following techniques can be used for the purpose of keyword normalization, the process of converting a keyword into its base form(Multiple Choice)?
   1. **Lemmatization**
   2. Levenshtein
   3. **Stemming**
   4. Soundex
9. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Python libraries used to perform text analysis
   1. **Spacy**
   2. Pandas
   3. Numpy
   4. None
10. Tokenization refers to segmenting text into words, punctuations marks, numbers..etc
    1. **True**
    2. False
11. Select the right statement to install spacy library
    1. pip download spacy
    2. **pip install spacy**
    3. download spacy
    4. None
12. Read Excel file in Pandas, which functions used to read file and store as DataFrame
    1. **read\_excel**
    2. read.excel
    3. read\_file
    4. None
13. Identify the valid statement for installing pandas packages.
    1. **pip install pandas**
    2. python install pandas
    3. pip setup pandas
    4. None
14. Function which used to display summary of Pandas DataFrame
    1. summary
    2. **describe**
    3. head
    4. None
15. value\_counts() function gives the frequency count of any selected columns from Pandas DataFrame.
    1. **True**
    2. False
16. Most common used graphics library to display graphs with Pandas DataFrame
    1. Seaborn
    2. Ploty
    3. **Matplotlib**
    4. None
17. Identify the most common libraries used for performing text mining/analysis in Python? (Multiple choice)
    1. NLTK
    2. Spacy
    3. Sklearn
    4. **All of the above**
18. \_\_\_\_\_\_\_\_\_\_\_\_\_ Function to get the frequency of categorical/labeled variable.
    1. count
    2. freq
    3. **value\_counts**
    4. None
19. To view no of rows & columns for Pandas data, which is best option to get (assume data is panda DataFrame)
    1. **data.shape**
    2. print(data)
    3. data
    4. None
20. To visualize graphs in Pandas which is most common library used
    1. Ploty
    2. **Matplotlib**
    3. Seaborn
    4. None
21. A \_\_\_\_\_\_\_\_\_\_\_ is a sequence of observations over a certain period.
    1. Theano
    2. Scikit
    3. **Time Series**
    4. Tensor
22. \_\_\_\_\_\_\_\_\_\_\_ is the process of transforming unstructured text into a structured format to identify meaningful patterns and new insights.
    1. Data mining
    2. **Text Mining**
    3. File Mining
    4. Deep Mining
23. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Python libraries used to perform text analysis
    1. **Spacy**
    2. Pandas
    3. Numpy
    4. None
24. Function which used to display summary of Pandas DataFrame
    1. summary
    2. **describe**
    3. head
    4. None
25. To view no of rows & columns for Pandas data, which is best option to get (assume data is panda DataFrame)
    1. **data.shape**
    2. print(data)
    3. data
    4. None
26. A \_\_\_\_\_\_\_\_\_\_\_ is a sequence of observations over a certain period.
    1. Theano
    2. Scikit
    3. **Time Series**
    4. Tensor
27. \_\_\_\_\_\_\_\_\_\_\_\_\_\_ plot used to display analysis of two or more variables.
    1. Bar graph
    2. **Pair plot**
    3. Histogram
    4. Line chart
28. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ function used to rename the columns in pandas DataFrame, and always used axis=1 while performing operations.
    1. colname
    2. **rename**
    3. modify
    4. alias
29. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ functions used to get the summary, min, max, sum…etc
    1. **Aggregate**
    2. Sum
    3. Mix
    4. Max
30. Most common function to display top 40 records in Python.
    1. tail(40)
    2. display(40)
    3. **head(40)**
    4. print(40)
31. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ functions to import comma separated values files in Python.
    1. **read\_csv**
    2. read\_excel
    3. read.csv
    4. None