1. Big data is also changing at a faster rate,
   1. **False**
   2. True
2. Text Mining/Analysis can be used in:
   1. Detecting spam model
   2. Predicting stock Movements
   3. News stories categorization
   4. **All the above**
3. Semi-structured data, also known as having a schema-less or self-describing structure
   1. False
   2. **True**
4. Select wrong Elements of Big Data
   1. Volume
   2. Velocity
   3. Variety
   4. **VolPro**
5. \_\_\_\_\_\_\_\_\_\_\_ 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
6. In which database, data is a blend between structured and unstructured data formats?
   1. Full-structured data
   2. Partial-structured data
   3. **Semi-structured data**
   4. Uni-structured data
7. The process of breaking out long-form text into sentences and words called?
   1. Stem
   2. Cluster
   3. Bag
   4. **Tokens**
8. 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
9. Typical text mining tasks include?
   1. text categorization
   2. text clustering
   3. entity relation modeling
   4. **All of the above**
10. 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
11. Most of the data in the world is in fact unstructured
    1. False
    2. **True**
12. The structured and unstructured data that is collected and analyzed in social media analytics programs includes all of the following, EXCEPT:
    1. Comments/reviews
    2. Demographic information
    3. **Transaction data**
    4. Number of likes and followers
13. 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
14. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Python libraries used to perform text analysis
    1. **Spacy**
    2. Pandas
    3. Numpy
    4. None
15. Tokenization refers to segmenting text into words, punctuations marks, numbers..etc
    1. **True**
    2. False
16. Select the right statement to install spacy library
    1. pip download spacy
    2. **pip install spacy**
    3. download spacy
    4. None
17. Big data refers to collection small and strictly structure type data.
    1. True
    2. **False**
18. Select the sources of data in Big Data eco systems (Multiple Choice).
    1. In/Out
    2. **Internal**
    3. **External**
    4. None
19. Structured data can be defined as the data that has a defined repeating pattern.
    1. **True**
    2. False
20. Is a new data challenge that requires leveraging existing systems differently, then we classify data as Big Data.
    1. **True**
    2. False
21. Big Data usually unstructured and ~70-80% data available in unstructured format.
    1. False
    2. **True**