1. Velocity is the speed in which data is process and becomes accessible.
   1. False
   2. **True**
2. Variability refers to the data which keeps on changing constantly.
   1. **True**
   2. False
3. Variety describes one of the biggest challenges of \_\_\_\_\_\_.
   1. **Big data**
   2. Data science
   3. Data integration
   4. None of the mentioned above
4. The data that can be processed, stored, and retrieved in a fixed format called \_\_\_\_\_,
   1. **Structured Data**
   2. Unstructured Data
   3. Semi-structured Data
   4. None of the mentioned above
5. Semi-structured data, also known as having a schema-less or self-describing structure
   1. False
   2. **True**
6. Select correct Elements of Big Data
   1. Volume
   2. Velocity
   3. Variety
   4. **All**
7. Text mining is process of transforming free of textual data into a structure format.
   1. False
   2. **True**
8. 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
9. The process of breaking out long-form text into sentences and words called?
   1. Stem
   2. Cluster
   3. Bag
   4. **Tokens**
10. 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
11. Typical text mining tasks include?
    1. text categorization
    2. text clustering
    3. entity relation modeling
    4. **All of the above**
12. 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
13. Most of the data in the world is in fact unstructured
    1. False
    2. **True**
14. 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
15. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Python libraries used to perform text analysis
    1. **Spacy**
    2. Pandas
    3. Numpy
    4. None
16. Tokenization refers to segmenting text into words, punctuations marks, numbers..etc
    1. **True**
    2. False
17. Select the right statement to install spacy library
    1. pip download spacy
    2. **pip install spacy**
    3. download spacy
    4. None
18. Big data refers to collection large data and strictly structure type data.
    1. True
    2. **False**
19. Select the sources of data in Big Data eco systems (Multiple Choice).
    1. In/Out
    2. **Internal**
    3. **External**
    4. None
20. Structured data can be defined as the data that has a defined repeating pattern.
    1. **True**
    2. False
21. Is a new data challenge that requires leveraging existing systems differently, then we classify data as Big Data.
    1. **True**
    2. False
22. Big Data usually in structured format and ~70-80% data available in structured format.
    1. **False**
    2. True