1. Is data science new?(false)

2. Based on the symptoms of the patient,general disease prediction is possible.The doctor is trying to find out(B)

A. what the world was like. B.What the world is going to be like.

3. I have been collecting fainfall data for Ipoh since 2010.Then I plot them on a graph,I am trying to see(A)

A. what the world was like. B.what the world is going to be like.

4. Data product from malaysia include all these,except(B)

A. covid-19 Corona Tracker B.Mysejahtera

B. Aerodyne D.MoneyLion

5. Is data science the same as statistics or analysis?(B.C)

A. yes,the terms are synonymous

B. Can use statistics to do data science

C. Statistics &data analysis are parts of data science

D. The terms have no relathonships

6. Data driven opportunities includes the following except(C)

A. new research B.better innovation C.new data D.new business

7. ()is drawing useful conclusion from data using computation?

A. data mining B.data science C.data driven D.big data

8. which is the keyword in data science?(B)

A. data B.science C.both D.no

9. who is the data driven US presicent?(C)

A. Donald Trump B.George Bush C.Barrack Obama D.Bill Clinton

10. when you use data to plot a graph in your report,the report is considered to be (D)

A. data driven B.data indifferent C.data denial D.data informed

11. Data are not taken for museum purposes!Then,for what？（D）

A. basis for doing something B.provide basis for decision

B. provide basis for planning D.provide a basis or action or a recommendation for action

12. data is considered big when the size is greater than (C)

A.1PB B.1ZB C.1TB D.1GB

13. In simple terms,data science is looking for (C)from data.

A. value B.knowledge C.insights D.information

14. which is the hardest among these analysis?(A)

A. causal analysis B.predictive analysis C.inferential analysis D.exploratory analysis

15. Use of historial data&macine learning techniques to identify likelihood of future outcomes(C)

A. descriptive analysis B.causal analysis C.predictive analysis D.inferential alalysis

16. data analysis on textual data is known as(D)

A. signal processing B.machine learning C.econometrics D.natural language processing

17. song recommendation,new policy,targeted advertising and trend predictor are examples of (C)

A. data outcomes B.data output C.data products D.data system

18. data science projects start with (B)

A. problem area B.question C.dataset D.results

19. Infographics,word clouds,heat maps,graphs are examples of(A)

A. data visualization B.data story telling C.data graphs D.data product

20. Do you need big data to do data science?(B)

A. yes,definitely B.size of data does not matter C.Indifferent D.big and medium data

21. Hacking skill of data scientist means(B)

A. breaking the computers B.creative to build things&find clever solution to poblem C.expert in programming D.able to get data from different sources

22. Theory if you buy certain set of items,you're more or less likely to buy another set of items(D)

A. shopping cart model B.customer preferences C.association rules D.market basket analysis

1. data science typically is concerned with analyzing(B)data.

A. primary B.secondary C.primary and secondary D.big

2. The main componenet of IR 4.0 is(**B**)

A. computer based system B.cyber physical system C.mechanical

system D.robotic system

3. What are the effect of IR4.0?(C)

A. new jobs B.The end of certain jobs C.New jobs and the end of certain jobs D.more tools to work

4. The V that refers to the assurance of quality or credibility of the collected data.(C)

A. volume B.variety C.veracity D.velocity

5. which V denotes that most of the data is unstructed in nature?(C)

A. volume B.value C.variety D.velocity

6. How many credit card swipes being sent on telecom carrier every minute of daily,and you'll have a good appreciation(B)

A. veracity B.velocity C.volume D.variety

7. which V fits this 27% of business are not sure if the data they are working on is accurate?(D)

A. volume B.variety C.value D.veracity

8. phrases like,data smog,data asphyxiation,infobesity,infoxication are results of(D)

A. a lot of data,but little information B.information overload

B. big data D.data explosion

9. datafication is a resource and a tool,which is meant to (),and points toward(A)

A. inform,understanding B.explain,modelling C.question,sensing

D.quantify,measurement

10. Hadoop,MongoDB,Cassandra and NoSQL are technologies for big data(C)

A. processor B.analyzer C.storage D.visualizer

11. (A)is to put data in quantified,tabulated,analyzable format to enable prediction.

A. datafication B.digitization C.quantification D.tabulation

12. an uncertain concerning the value of all captured and yet-unexplored data in an organization.(B)

A. blind spot B.blind zone C.hidden zone D.unexpected area

13. data(bad,incorrect or defective)or data left behind during online activities by the users is called(C)

A. data reuse B.option value C.data exhaust D.data combination

14. (D)suggest that data can be used in future in ways not identified at the time of collection.

A. reuse of data B.double value of data C.recombinant of data D.option value of data

15. which of these are not about customer delight(B)

A. small gifts,have a bigger impact when they are unexpected

B. Customers use the product they bought

C.TacoBell send truck with 10000 Tacos to remote Alaskan Town

D.loyal and actively promote your brand through word-of-mouth

16.which of these are disruptive technologies?(A,B,C,D)

A.smart phones B.mobile apps C.black chain D.social networking

17.data acquired by organization but not used in any manner to derive insights is (A)

A.dark data B.waste data C. D.dirty data

18.when designing a new product,creativity is important,but data cannot be forgotten entirely.This is (B)approach.

A.data driven B.data inspired C.data informed D.data indifferent

19.these are examples of (D)questions.

A.data informed B.data inspired C.data indifferent D.data driven

20.to be data driven,an organization must have(C,D)

A.reporting B.alert and dashboards C.move from BI to BA D.data culture

21.data driven culture include all these elements except(B)

A.require strong,top-down leadership B.decision making based on HiPPOs experiences,pre-conceived notion &gut C.promotion of data sharing practices D.more staff across the organization who are skilled at analytics

22.the diagram shows an anatomy of a (D)business.

A.data driven B.data company C.data culture D.data centric

1. if you are asking questions and using data to find answers,you are a(D)

A.data analyst B.data modeller C.data engineer D.data scientist

2.there is a shortage of talent in data science today.(A)

A.true B.false

3.data scientist is one who(A)

A.transform data into insights B.clean data C.tell data story D.uses scientific methods to create meaning from raw data

4.which of this does not constitute data science?(C)

A.computing and data skills B.substantive expertise C.thinking D.maths and statistics

5.Hacking skills means(C)

A.gain unauthorized access to data in a system or computer

B.Using computers to commit fraudulent acts.

C.Able to manipulate text files at the command-line.

D.restore the lost or forgotten information,e.g. passwords

6.this is a stereotypical perspective of (B)on data science

A.boss B.customer C.software engineer D.data scientist

7.To get real truth&useful answer from data, a data scientist must follow the (C)which mainly consist of 5 steps.

A.data pipe line B.data analysis steps C.data scientific method D.CRISP-DM

8.(D)are self-contained,and you do not need any special programs to view the data contained inside.

A.database B.APIs C.HTML D.Flat files

9.(AD)is the process of capturing the useful parts of raw and unkempt data under some bad conditions.

A.data munging B.data cleansing C.data jujitsu D.data wrangling

10.the kind of data scientist organization may not wish to hire(D)

A.unicorn data scientist B.detective data scientist C.generalist data scientist D.diva data scientist

11.the most important skill for a data scientist is(B)

A.communicate B.analytical C.programming D.creativity

12.the task to decide what data should be gathered require which area of expertise?(A)

A.substantive expertise B.computing and data skills C.maths and statistics D.technical skills

13.a unicorn data scientist is(D)

A.rare B.upgraded from”sexy data scientists” C.impossible D.an expert in each and everything of data science field/pillars.

14.a solution to replace unicorn data scientist is(C)

A.no replacement B.get a data superman C.have data science team D.train an experienced data scientist with new skills

15.an exception data scientist includes all these,except(C)

A.good communicator B.comfortable to learn new tools C.prefer doing data cleaning D.able to move between roles in a project

16.earns money by participating in data science competitions or by doing freelancing.They are(C)

A.fake data scientist B.extreme data scientist C.amateur data scientist D.self-made data scientist

17.data science portfolio should include all these,except(D)

A.one's particulars,like in CV. B.one's real world experience doing DS work.

C.one's contribution on DS,eg participating in competition D.one's success and also failure stories

18.The recipe for a useful answer in data science project exclude(B)

A.good question B.more data C.insightful analysis D.relevant data

19.these information are usually found in the(B)

A.code documentation B.user documentation C.developer documentation D.project documentation

20.A real world experience in data science can be gained from these except(C)

A.solving data science problem B.taking part in data science competition

C.reading case studies from data science books D.doing data science group project in a course

1. not a criteria for data-driven organizations(B)

A. must have analytics B.should have business intelligence C.leverage data as a strategic asset D.require a data culture

2. analytics is about(C),not changes means no(C)

A. result B.value C.impact D.computing

3. to demonstrate value,you must always connect action with(A)

A. outcome B.objective C.goal D.output

4. the most important thing in data science is the (),and then followed by()D

A. data,question B.problem,data C.goal,analysis D.question,data

5. which of these follow the right sequence in the data processing pipeline?(B)

A. ask question,get data,find data,clean data,analyze data

B. Ask quesion,find data,get data,clean data,analyze data

C. Ask question,get data,analyze data,present results

D. Find data,clean data,analyze data,ask question

6. In data science,most data analysis involve () and ()C

A. descriptive and explorative B.casual and mechanistic C.statistical inference and machine learning prediction D.exploratory and inferential

7. This type of analysis plays a significant role in strategic planning.(B)

A. exploratory B.predictive C.inferential D.descriptive

8. Annual reports provide an overview of firms past performance.Which

analysis is dominant in these reports?(D)

A. exploratory B.predictive C.inferential D.descriptive

9. which analysis is prevailing in a crime scence investigation?(A)

A. exploratory B.predictive C.inferential D.descriptive

10. key findings of a research that is published in a journal article largely based on (C)analysis.

A. exploratory B.predictive C.inferential D.descriptive

11. in general,movie recommendation systems are an example of(A)

A. clustering B.anomaly detection

12. you live in a gated housing society and your society has separate dustbins for different types of waste.This shows(C)

A.clustering B.anomaly detection C.classification D.regression

13. should the car be slowed or accelerated?(C)

A.clustering B.anomaly detection C.reinforcement learning D.regression

14.how much will MPH do in sales next month?(D)

A.clustering B.anomaly detection C.reinforcement learning D.regression

15.the identification of rare events which raise suspicions by differing significantly from the majority of the data.(B)

A.clustering B.anomaly detection C.reinforcement learning D.regression

16.(C)making digital information accessible&understandable to end user for decision-making.

A.open data B.free data C.data democratization D.data marketplace

17.FAIR research principles advocate for()()()and ()data?(C)

A.feasible,attainable,interorganization,and reusable data

B.free,approachable,interactive and relevant data

C.findable,accessible,interoperable and reusable data

D.favorable,ascertainable,important and reachable data

18.editing,cleaning or modifying the raw data(source data of atomic data)results in(B)

A.tidy data B.processed or cooked data C.clean data D.analyzable data

19.data scraping used for extracting data from websites are known as these,except(B)

A.web scraping B.web data pulling C.web harvesting D.web data extraction

20.where can you get global data on life expectancy and healthy life expectancy?

A.WHO B.WIPO C.UN D.World Bank

21.text from a scanned document can be extracted using Google(D)

A.Trend B.Analytics C.Assistant D.Tesseract OCR

22.(C)refers to downloading the entire website offline on hard disk for browsing it offline.

A.website monitoring B.website browsing C.website mirroring D.website scraping

23.which of these is not a human readable file?(B)

A.XML B.UPC barcodes C.JSON D.CSV

24.sentiment analysis,text analytics and topic modelling work on ()data.

A.True B.False

1. （C、D）are inaccurate,incomplete or inconsistent data,especially in a computer system of database.

A. bad data B.incomplete data C.messy data D.dirty data

2. These are reasons why data are dirty,except(B)

A. incomplete B.garbage in garbage out C.outdated D.inconsistent

3. cleaning data involved all these,except(D)

A. handle missing data B.rename columns C.convert data type D.improper encoding

4. your data is considered (C)when you did not remove any data from the data set.

A. clean B.original C.raw D.untouchable

5. (B)data are the data that is ready for analysis.

A. cooked B.processed C.formatted D.clean

6. dates normally has (C)issues.

A. naming convention B.different representations C.formatting D.redundancy

7. data is missing or not referenced violates which data quality dimensions?(D)

A. conformity B.consistency C.accuracy D.integrity

8. data is stored in a non-standard format violates which data quality diemensions?(A)

A. conformity B.completeness C.accuracy D.consistency

9. data that looks different but represents the same thing,e.g.'V.P''v.p'，'VP''Vice Pres'(B)

A. inaccurate data B.inconsistent(non-standardized)data C.incorrect data D.incomplete data

10. Geolocation variables include all these,except(B)

A. Zipcode B.property tax amount C.longitude D.latitude

11. data preparation or pre-processing involved these tasks.(B)

A. data cleansing,reshaping,conversion,encoding

B. Data cleaning,integration,transformation,reduction

C. Data cleaning,munging,wrangling,formatting

D. None of the mentioned

12. data scientists spend most of their time(C)

A. finding and getting data B.thinging about problems and questions to

ask

B. preparing and cleaning data D.interpreting analysis results

13. why must we worry about missing data?(A)

A. can add bias to model B.inaccurate results C.need to prepare training data D.analysis not perfect

14. which is a TRUE statement regarding imputation?(B)

A. imputation is making up data to artificially inflate results

B. For categorical data,impute with new category,'unknown'

C. I can just impute the mean for any missing data

D. Imputation is always the best way to deal with missing data

15. which of this pattern of missing data cause the least harm in your research?(B)

A. MNAR B.MCAR C.MAR D.None

16. data tidying is the process of structuring data in a way that makes it easy to analyze and use.(A)

A. true B.false

17. THe end goal of data tidying process is.(A)

A. tidy data B.instruction list C.code book D.clean data

18. This diagram refers to which data manipulation?(B)

A. filtering B.aggregating C.transforming D.sorting

19. data cleaning is closely related to(C)

A. data wrangling B.data munging C.data quality D.tidy data

20. (C)is the idea that the output of an algorithm is only as good as the quality of the input that it receives.

A.first in,first out B.quality in,quality out C.garbage in,garbage out D.garbage dump,garbage out