**Data Scientist**

**N8101**

1. Which option denotes an example of a tool that can be used to validate imported data?
   1. **Spreadsheet Software**
   2. Image editing software
   3. Video editing software
   4. Word processing software
2. Which option denotes an example of metadata that can be used to describe a dataset?
   1. The size of the dataset in bytes
   2. The color of the data set’s header
   3. **The number of columns in the dataset**
   4. The name of the analyst who created the dataset
3. What is the purpose of populating metadata for imported data?
   1. To add new columns
   2. To rename the columns
   3. **To provide a description of each column**
   4. To remove duplicate rows
4. Consider that the marketing department of company wants to import data on customer demographics and purchasing habits. Which option denotes and example of a key variable to be imported?
   1. Number of employees
   2. Brand of the customer’s phone
   3. **Customer’s age and gender**
   4. Company’s previous year’s revenue generation
5. Assume that you want to import and store data on employee salaries and performance rating. Which is the best method to acquire and store the data?
   1. Manually entering the data into a spreadsheet
   2. **Pulling data from a cloud database**
   3. Hiring a data entry specialist
   4. Asking to submit the salary data via email.
6. Which is a suitable source of data for a study on the effects of air pollution on human health?
   1. Music history
   2. Student’s opinion on climate change
   3. **Air quality monitoring stations**
   4. Fast food restaurants
7. What is the first step you need to do while importing data for analysis?
   1. Identify the data type
   2. **Define the objective**
   3. Determining the volume
   4. Data verification
8. Why is it important for a company to identify the volume of data to be imported?
   1. To determine the data format
   2. To verify the imported data
   3. **To optimize the import process**
   4. To define the objective

**N9006**

1. How can you work a colleague who is better than you to deliver the project without any conflict?
   1. Don’t join this project
   2. Refer you colleague to a different project
   3. **Help each other**
   4. Talk to the client
2. Assume that a new intern is able to solve a major issue in a client project. How can you ensure that the intern gets his/her due credit?
   1. By naming it a team effort
   2. By telling other employees
   3. By taking the credit for being the leader
   4. **By acknowledging the efforts of the employee with seniors**
3. What will you do if you find out that one of your interns is not working according to expectations?
   1. Tell him/her that he/she is not working properly in team meetings
   2. **Understand the difficulty he/she is facing and give feedback**
   3. Tell other teammates about your opinion
   4. Give him/her bad ratings without talking to him/her
4. Consider that the team of data scientists at your organization comprises of a mix of freshers and experienced people. Which is NOT an appropriate way to treat the freshers in the team?
   1. Help them their tasks
   2. Make them feel safe in the working environment
   3. **Give them orders**
   4. Take inputs from them while making a decision
5. Which is the best way to recognize the contribution made by your colleagues at work?
   1. Take the credit for their work
   2. **Give them credit during team meetings**
   3. Appreciate them privately
   4. Criticize their contribution in public
6. What measures will you take with the current client project in order to ensure that you will be provided with more projects in the future?
   1. Setting up a meeting with them
   2. Engage with a campaign targeting this client
   3. **Prioritizing the client deliverables in terms of time & quality alike**
   4. Hoping for the best
7. How can you ensure that the teammate feels comfortable working together with colleagues that different communication styles?
   1. Ignore them
   2. Ask them to change
   3. Observer their communication style and try to adapt it
   4. **Find some common ground to talk and compromise**
8. How can you ensure that your colleague’s efforts are recognized properly?
   1. Ignore his/her contribution and work on your task
   2. Name it a team effort
   3. **Making sure the seniors know about his/her efforts and contribution**
   4. Take the credit

**N9004**

1. What is an appropriate option to sort data for easy access to product information and related engages for a major analysis at the start of the month?
   1. Sorting product details
   2. Adding the data in a word format
   3. Present it in original format
   4. **Sorting when required**
2. What is NOT required when submitting the final report for a project with complete and accurate information to the senior architect at the end of the project?
   1. **Raw data**
   2. Graphical results
   3. Objective of research
   4. Cleaned data
3. What action is an employee performing on a database if he/she is asked to change the name of a person?
   1. Deletion
   2. **Updation**
   3. Insertion
   4. Change
4. Which method should you consider in order to ensure that the information is accurate and complete?
   1. Data sampling
   2. **Data cleaning**
   3. Data Visualization
   4. Data mining
5. What can be considered a best practice for good documentation when presenting the findings of a completed data mining project for XYZ organization?
   1. Submitting unfiltered data
   2. **Submitting relevant data**
   3. Submitting numeric data
   4. Submitting graphical data
6. Assume that you want to gather correct information related to the latest technologies on the market for analysis. What is the best practice for a data scientist to perform this task?
   1. Using unverified data
   2. Using popular data
   3. **Using verified data**
   4. Using low cost data
7. What steps should you take to ensure accurate results when faced with an incomplete and incorrect dataset from the client’s side during analysis?
   1. Continue working on the data without any actions
   2. Talk to your colleagues who are not working on the project
   3. **Reach out to a senior person working on the project**
   4. Ignore he issue and hope it will not affect the project
8. How will you go through the process to understand the fault in an algorithm you have just created for a project?
   1. Ask your interns
   2. **Go through each step**
   3. Ignore
   4. Ask for new algorithm
9. When should you report any unresolved anomalies in the data/information to appropriate people?
   1. **Immediately**
   2. After the problem have been resolved
   3. Only it the significant
   4. Never  
        
      **N9002**
10. What should you do if you face any problems while working with your colleagues?
    1. Ignore the problem
    2. Wait for your colleagues to solve the problem
    3. **Take the initiate to solve the problem**
    4. Complain to your supervisor about the problem
11. What should you do to ensure that you follow your organization’s policies and procedures for working with colleagues?
    1. Ignore the polices and procedures
    2. Create your own policies and procedures
    3. **Follow each and every policy**
    4. Follow policies relevant to you
12. How should you pass on essential information to colleagues?
    1. Use notice board
    2. **Email communication**
    3. Letters
    4. Newspaper Ad
13. When communicating with colleagues, what si an important quality to maintain?
    1. Being rude
    2. Being vague
    3. **Being concise**
    4. Being repetitive
14. What is an important thing to keep in mind when working with colleagues?
    1. Work against them
    2. Ignore them
    3. **Integrate your work with theirs**
    4. Criticize their work
15. How can you effectively work with your colleagues to integrate your work with theirs?
    1. Share all your work with them
    2. Work independently without consulting them
    3. **Communicate clearly and regularly**
    4. Ignore their work  
         
       **N9001**
16. Which is the most important factor to follow to get good quality work?
    1. Working as per your own schedule
    2. **Implementation of feedback received**
    3. Storing data in an arranged manner
    4. Cleaning workspace
17. At what stage/stages should you think of learning new skills as a part of following good delivery of work?
    1. At the requirement gathering phase
    2. At the requirement understanding phase
    3. At the development phase
    4. **At all phases of work completion**
18. What is NOT necessary to be carried out to effectively meet your work requirements?
    1. Asking questions
    2. Communicating with people
    3. **Taking breaks**
    4. Setting targets
19. Which aspect must NOT be taken into consideration to work as per the organization’s policies and procedures?
    1. The employees affected with the policy
    2. **Creators of the policy**
    3. Policy terms
    4. Penalties for violating the policy
20. What must be done to protect confidential data from your teammates?
    1. **Using data encryption**
    2. Using metadata
    3. Using data retention
    4. Using data visualization
21. As a data scientist which measure will NOT help to keep area clean and tidy?
    1. Using dustbin
    2. Using wet cloth for removing dust
    3. Bringing clean shoes to work area
    4. **Using best data analysis tools**
22. What is the term used for application of principles and techniques to manage resources effectively?
    1. Resource optimization
    2. **Resource management**
    3. Supervised learning
    4. Resource techniques
23. As a data scientist, what can be a limitation to the job role that you have?
    1. Being over emotional
    2. The workplace design
    3. **The data on which you are working**
    4. The confidential data of the company
24. What might be the result of failing to manage your time at work?
    1. Stress reduction
    2. **Missing deadlines**
    3. Great reputation
    4. Great efficiency

**N8105**

1. What is the purpose of validating implemented models?
   1. To make the model more complex
   2. To check the number of variables used
   3. To check the model type
   4. **To check accuracy**
2. What is the recommended approach to achieve a model’s desired performance or quality of output?
   1. Review the model
   2. **Iterate the model with feedback**
   3. Make a new model
   4. Implement the model
3. What is the process of identifying and selecting key features or variables that will be used to build a model known as?
   1. Data cleaning
   2. Data exploration
   3. **Feature engineering**
   4. Model evaluation
4. What is the term used for the process of applying a pre-trained model on one task to a different but related task?
   1. **Transfer learning**
   2. Recognition Training
   3. Related Learning
   4. Test Learning
5. What do you call the process of choosing statistical models based on analysis software packages, libraries or tools?
   1. Data Analysis
   2. Data Visualization
   3. **Algorithm Selection**
   4. Data Pre-processing
6. Which tool is most appropriate to develop algorithmic models in data science?
   1. Excel
   2. Illustrate
   3. Power point
   4. Word
7. Why is it important to identify the objective of an analysis?
   1. To select suitable libraries
   2. To gain inferences from the final output
   3. To iterate the model
   4. **To determine the goal of the analysis**  
        
      **N8104**
8. What is the purpose of defining connectors or combinations of key variables for each statistical model?
   1. To increase model complexity
   2. **To increase model accuracy**
   3. To speed up model training
   4. To reduce the volume of data
9. Consider that time series model is a category of statistical models based on volume and key variables. Which option represents a Time Series Model?
   1. **ARIMA(Autoregressive Integrated Moving Average)**
   2. Neural Network Model
   3. Kriging
   4. Deep Learning
10. Which type of model should be considered if the use case demands a more complex analysis?
    1. Excel Model
    2. Statistical Model
    3. Optimized Model
    4. **Deep learning Model**
11. What is the first step to be taken before building statistical models?
    1. Coding
    2. **Researching**
    3. Building the model
    4. Deriving model results
12. Which is the library that is NOT commonly used for data analysis?
    1. Pandas
    2. Matplotlib
    3. R
    4. **OpenSSL**
13. Which information for each designed algorithmic model will NOT be included in the documentation?
    1. Model Name
    2. Version
    3. Design Detail
    4. **User Detail**
14. What is the term you use for the process of defining and quantifying success metrics for an algorithmic model?
    1. Model Growth
    2. Model Validation
    3. **Model Assessment**
    4. Model Evaluation
15. What is the process of managing the weights and biases in a neural network to reduce the error between predicted and actual outputs known as?
    1. Gradient Boosting
    2. **Gradient Descent**
    3. Random Forest
    4. Linear Regression
16. What is the process of developing a working model of an algorithm based on the analysis of data known as?
    1. Design
    2. **Prototype**
    3. Testing
    4. Validation
17. Which option denotes the way that is used to fine-tune a model?
    1. **Iterate training**
    2. Validation
    3. Analyse
    4. Identifying the error  
         
       **N8103**
18. What is the key benefit of developing a hypothesis model?
    1. It helps to identify potential data errors
    2. It helps to count number of records
    3. It helps to perform data validation
    4. **It helps to understand the relationship between variables**
19. In order to gain meaningful ingereces from the final output of data analysis you need NOT have clear understanding of which factor?
    1. Objectives
    2. Limitations
    3. Assumptions
    4. **Graphics**
20. Which option denotes an example of data validation?
    1. Checking the number of rows
    2. **Checking that a numerical column only contains integers**
    3. Checking the number of columns
    4. Checking the source of the data
21. Which technique can be used to identify the strength of the association between the variables in a dataset?
    1. Clustering
    2. Dimension reduction
    3. **Correlation analysis**
    4. Statistical testing
22. In a dataset with 1000 variable, which technique can be used to reduce the number of variable for analysis
    1. Clustering
    2. **Dimension reduction**
    3. Correlation analysis
    4. Statistical testing
23. What is the objective of prescriptive action?
    1. **Turning data into actionable insights**
    2. Arranging the raw data
    3. Matching data to suggested ideas
    4. Identifying key variables
24. What is the first step you need to take while evaluating the results of an analysis and defining business outcomes?
    1. Collecting the data
    2. **Defining problem statement**
    3. Identifying key variables
    4. Defining goals
25. What is the first step you need to take while identifying the key variables required for modelling or analysis of a dataset with multiple variables?
    1. Sort the data based on a variable
    2. Remove variables with missing values
    3. **Identify the problem to be solved**
    4. Use statistical techniques to summarize variables  
         
       **N8102**
26. Which is the correct method to deal with data redundancy in a dataset?
    1. Normalizing the dataset
    2. Adding more data to the dataset
    3. Duplicating the dataset
    4. **Ignoring the redundancy**
27. In a dataset with customer information which way of storing customer data is an example of data redundancy?
    1. Name and address in separate columns
    2. Name and age in separate columns
    3. Name and email in separate columns
    4. **Date of birth and age in separate columns**
28. What action will you take considering that you notice that the age variable in your dataset is being stored as a string instead of an integer?
    1. Ignore the issue and move on the analysis the data
    2. Leave a note in the dataset explaining the issue
    3. Manually convert each value to an integer
    4. **Use a data transformation tool to convert the data type of the age variable**
29. For which option do you perform operations to transform the data type of variable?
    1. Incomplete data
    2. Missing data
    3. **Incorrect data**
    4. Confidential data
30. Which is a data type that can be used to represent variable in a dataset?
    1. Boolean
    2. Float
    3. String
    4. **All of them**
31. What is the item used for the process of identifying and fixing corrupted values in a dataset?
    1. Data validation
    2. **Data cleaning**
    3. Data pre-processing
    4. Data visualization
32. Why do you need to define the format and structure for a dataset?
    1. To add new rows to the dataset
    2. **To ensure that the dataset can be easily read and analyzed**
    3. To move columns from the dataset
    4. To change the color of the dataset

**N8108**

1. What is an example of a most popularly used visualization template?
   1. Scatter plot
   2. Line chart
   3. Heat map
   4. **Bar chat**
2. What is the importance of agreed language standards for visualization?
   1. **Better Understanding**
   2. Complex graphs
   3. Confusing language
   4. Avoid understanding of data
3. What is the best way to present a report to the company’s seniors?
   1. Word Document
   2. **PowerPoint presentation**
   3. Sending mail
   4. Paper printing report
4. Which is the best way to present data during company meetings?
   1. Long Reports
   2. Spreadsheets
   3. **Charts**
   4. Write-ups
5. What is the benefit of summarizing business outcomes into a narrative?
   1. **Comprehensive report**
   2. Report longer and more detailed
   3. Time consuming report
   4. Increases complexity
6. What is the importance of defining the delivery mode and format for reporting business outcomes?
   1. Looks visually appealing
   2. Looks more comprehensive
   3. **Reaches the target audience**
   4. Validate the accuracy
7. What should be the main focus while reporting business outcomes?
   1. Detailed methodology
   2. Raw Data
   3. **Key Insights**
   4. Technical Analysis  
        
      **N8107**
8. What is the role of grid search in hyper parameter tuning?
   1. **Identifying best parameters**
   2. Reducing overfitting
   3. Increasing size of the dataset
   4. Overloading code
9. What is the disadvantage of using grid search for hyper parameter tuning?
   1. Increased precision
   2. Less time consuming
   3. Increase the complexity
   4. **Computationally expensive**
10. Which theorem helps you in determining conditional probability?
    1. Pythagoras Theorem
    2. **Bayesian Theorem**
    3. Gauss Theorem
    4. Fermat’s Theorem
11. What are hyper parameter tunings used for?
    1. To reduce model complexity
    2. To improve model interpretability
    3. **To maximize model performance**
    4. To increase model complexity
12. What is the disadvantage of using force full methods for hyper parameter tuning?
    1. Overfitting
    2. Less time consuming
    3. Overloading
    4. **More time-consuming**
13. What is a potential risk of not implementing corrective actions for the model?
    1. Syntax errors
    2. No impact on the model
    3. **Reduced model functionality**
    4. Increased model complexity
14. How should evaluation criteria and metrics be selected?
    1. As per the convenience
    2. **Based on business**
    3. Based on data availability
    4. Based on the stakeholder
15. What is the first step in defining evaluation criteria and metrics?
    1. Collecting data
    2. **Identifying model objectives**
    3. Validating the model
    4. Building the model  
         
       **N8106**
16. Who is responsible for recommending and implementing corrective actions for the model?
    1. **Data Scientists**
    2. AI tools
    3. CEO
    4. Back-end codes
17. What type of mitigation measure can be introduced to mitigate the risk of model complexity?
    1. Feedback loop
    2. Increase data volume
    3. **Model simplification**
    4. Nothing can be done
18. What is a potential benefit of evaluating the model for all possible use cases/scenarios?
    1. Reduced accuracy
    2. **Enhanced functionality**
    3. Reduced complexity
    4. Increased performance
19. Which option denotes an example of an evaluation metric used in data science?
    1. **Mean squared error (MSE)**
    2. Logistic regression
    3. R-squared
    4. Random forest
20. Which is a commonly used technique for testing a model with different inputs in data science?
    1. Sensitivity Analysis
    2. Precision-Recall analysis
    3. **Cross validation**
    4. Hypothesis testing
21. How can risk be reduced in cases where the algorithmic model deviates from the expected outcomes?
    1. Improve precision
    2. **Feedback loop**
    3. Increase complexity
    4. Hyperparameter set
22. Which is a commonly used algorithm for dimension reduction in data science?
    1. Decision tree
    2. **Principal component analysis**
    3. Random forest
    4. Meta Data

**N9014**

1. What is the most important benefit of appropriate verbal and nonverbal communication schemes for people with disabilities (PwD) at the workplace?
   1. No benefit
   2. Improved company image and company profits
   3. **Enhanced communication and collaboration with PwD**
   4. Faster promotion
2. What is the most effective resource to promote workplace awareness among colleagues about using appropriate verbal and non-verbal communication with PWD?
   1. Google the correct way of talking
   2. **Use disability language style guide**
   3. Use normal language which you do with friends
   4. Address them in the form of their disability
3. What is the appropriate way to address people with disabilities (PWD) in order to ensure respectful communication in your organization?
   1. Differently-abled
   2. Disability
   3. Special
   4. Physically-challenged
4. What is the primary purpose of segregating waste into recyclable, non-recyclable, and hazardous categories at the workplace?
   1. To increase the quantity of non-recyclable waste generated
   2. To minimize the cost of waste disposal
   3. **To ensure safe and efficient waste management**
   4. To make the workplace clean
5. As a part of an annual data clean up consider that you are required to clean up the excess irrelevant data in storage devices such as CDs , hard drives, etc. How will you dispose off such devices safely?
   1. **Break down all the disks and dispose it**
   2. Sell the disks to some other firm
   3. Born all disks and storage devise
   4. Give the disks to employees
6. How can you ensure that your workplace has optimized electricity usage and zero energy waste?
   1. Keeping the computers switched on all the time
   2. **Installing lights with auto switch off**
   3. Not disposing waste properly
   4. Always keeping fans switched on

**N9010**

1. What is the importance of adapting arguments to consider diverse needs as an AI data scientist?
   1. It helps in creating partiality free models
   2. It ensures that the model is accurate for a diverse range of users
   3. It reduces the risk of mismatch of data
   4. **All the given options**
2. How should an AI data scientist facilitate a structured discussion to reach a consensus on the tools to be use for a project?
   1. Checking on Google
   2. Letting team members make decisions without guidance
   3. **Facilitating discussions with everyone**
   4. Asking the CEO
3. What aspect should you consider to ensure that you get project from a new client?
   1. Ask them to trust you
   2. Let them decide on their own
   3. **Tell them about your previous projects and end results achieved**
   4. Plead them to give you a chance
4. What is the primary purpose of segregating waste into recyclable, non-recyclable, and hazardous categories at the workplace?
   1. To increase the quantity of non-recyclable waste generated
   2. To minimize the cost of waste disposal
   3. **To ensure safe and efficient waste management**
   4. To make the workplace clean
5. Which option will be appropriate in choosing the correct algorithm to meet client requirements efficiently?
   1. Choosing the algorithm on your own
   2. Choosing the algorithm which one employee likes
   3. **Choosing the algorithm which is supported by the team**
   4. Choosing the algorithm which is complex
6. Which is a great team building practice to engage with the employees and motivate them to be better?
   1. Giving them leaves
   2. Incentives for everyone
   3. **Recognition for extra efforts**
   4. Flexibility in work hours
7. What is an effective way to persuade stakeholders about the benefits of implementing an AI solution?
   1. Sharing your personal beliefs
   2. **Presenting concrete examples of successful AI implementation**
   3. No need to present the idea
   4. Telling them there is no other option but to implement AI