MANGALORE UNIVERSITY

National Education Policy – 2020(NEP-2020)

QUESTION BANK

Of

Skill Enhancement Course

Artificial Intelligence

III/IV SEMESTER (OTHER THAN BCA)

COMMON TO ALL

QUESTION BANK

Skill Enhancement Course

Artificial Intelligence

III/IV SEMESTER (OTHER THAN BCA)COMMON TO ALL

Duration: 2 Hour **Max.Marks:30**

Section-A (1x10 = 10)

1. What is an AI?
   AI is the creation of software that imitates human behaviors and capabilities.
2. The capability for a computer to interpret written or spoken languageis Natural Language Processing.
3. Knowledge mining is extracting information from large volume of data.True or False ?
   True
4. What is Azure Cognitive Search?
   Azure Cognitive Search, is a private, enterprise, search solution that has tools for building indexes. Azure Cognitive Search can utilize the built-in AI capabilities of Azure Cognitive Services such as image processing, content extraction, and natural language processing to perform knowledge mining of documents.
5. Anomaly detection - a machine learning based technique is use to analyze data overtime and identifies unusualchanges.
6. What is conversational AI?
   Conversational AI is an umbrella term used to describe various methods of enabling computers to carry on a conversation with a human. This technology ranges from fairly simple NLP to more sophisticated machine learning models that can interpret a much wider range of inputs and carry on more complex conversations.
7. What is Descriptive analytics?
   Descriptive analytics help answer questions about what has happened based on historical data. Descriptive analytics techniques summarize large datasets to describe outcomes to stakeholders.
8. What is machine learning?
   This is the foundation for an AI system, and is the way we "teach" a computer model to make prediction and draw conclusions from data.
9. What is Azure Bot?
   This service provides a platform for conversational AI, the capability of a software agent to participate in a conversation. Developers can use the Bot framework to create a bot and manage it with Azure Bot service integrating back-end services like language and connecting to channels for web chat, email, Microsoft teams and others.
10. Define tuning.
    The process of finding the optimal configuration is called tuning.
11. Computer vision service is designed to perceive the world visually throughcameras, images and video.
    True or False?
    True.
12. Define data granularity.
    Data granularity is the detail that is represented within your data, meaning that the more granularity your data has, the greater the level of detail within your data.
13. Define Anomaly detection.
    Anomaly detection is an artificial intelligence technique used to determine whether values in a series are within expected parameters. A machine learning based technique that analyzes data over time and identifies unusual changes.

14. A language service to determine the key talking points in a text document is
    Key Phrase Extraction.
15. An automobile dealership wants tom use historic car sales data to train a machine learning model. The model should predict the price based on its make, model, engine size and mileage. What kind of machine learning model should the dealership use automated machine learning to create?
    Regression – a supervised machine learning model used to predict a continuous value; like a price, a sales total, or some other measure.
16. You want to create a model to predict sales of ice-cream based on historic data. Which Azure service should you use?
    Azure Machine Learning.
17. Which are the building blocks of Power BI?
    The five major building blocks of Power BI are: dashboards, reports, workbooks, datasets, and dataflows.
18. How is the admin workspace role different from other types of workspaceroles?
    - Admin is the only role that can remove users.
    - Configure data refreshes.
    - Can delete reports and content in a workspace.
19. Neural networks are widely known for its use in deep learning and modeling complex problems such as image recognition.
20. The process of finding the optimal configuration is tuning. True or False?
    True.
21. A predictive app provides audio output for visually impaired users.
    Inclusiveness - Principle of AI is reflected here.
22. Descriptive analytics help answer question about why event happened. True or False?
    False.
23. What is data analytics?
    Data Analytics is a term that is used to describe the technical aspects of analytics that have predictive capabilities and can be used to solve business problems and is the power to analyze and learn about large amounts of data from multiple sources and detect patterns to make future trend predictions.
24. A Business analyst is a specialist in interpreting the data that comes from the visualization.
25. Name any two statistical functions in Power BI.
    DAX functions to calculate average, sum, min, max, and so on.
    Histograms and bell curves
    Top N analysis
26. What is NLP?
    NLP is Natural language processing (NLP) is the area of AI that deals with creating software that understands written and spoken language.
27. What are the computer vision models and capabilities?
    Image classification, Object detection, Semantic segmentation, Image analysis, Face detection, analysis, and recognition and Optical character recognition (OCR).
28. Mention two uses of Computer vision.
    You can use this service to analyze images and video, and extract descriptions, tags, objects, and text.
29. What is cardinality of a table?
    Cardinality is the measure of unique values in a table.
30. What is Deep learning?
    Deep learning is a subset of machine learning, where artificial neural networks algorithms modeled to work like the human brain-learn from large amounts of data.

31. Name the three different types of storage modes in data analytics.
    - Import
    - DirectQuery
    - Dual (Composite)
32. Name the different roles in data.
    - Business analyst
    - Data analyst
    - Data engineer
    - Data scientist
    - Database administrator
33. Which data role enables advanced analytics capabilities specially throughreports and visualization?
    Data analyst.
34. What is a flat file?
    A flat file is a type of file that has only one data table and every row of data is in the same structure.
35. Who is a Business analyst?
    A business analyst is closer to the business and is a specialist in interpreting the data that comes from the visualization.
36. What is feature vector?
    A feature vector is an ordered list of numerical properties of observed phenomena. It represents input features to a machine learning model that makes a prediction. Humans can analyze qualitative data to make a decision.
37. What is Database?
    A database is a collection of data that is organized, which is also called structured data. It can be accessed or stored in a computer system.
38. Define DBMS.
    DBMS - A DataBase Management System is a software package designed to store, retrieve, query and manage data.
39. SQL stands for Structured Query Language.
40. NoSQL database is a relational database. True or False?
    False
41. What type of expression do you use to extract data from Microsoft SQLServer?
    T-SQL
42. DAX stands for Data Analysis Expression.
43. Do you need to import custom visuals each time you want to use them whenyou're developing a new report, not an existing report?
    Yes, custom visuals must be imported from AppSource each time you start developing a new report.
44. R or Python Programming languages can be used to visualize data within Power BI Desktop.
45. Power BI has the concept of directionality to a relationship. True or False?
    True.
46. What is image analysis?
    Image analysis techniques is used to extract information from images, including "tags" that could help catalog the image or even descriptive captions that summarize the scene shown in the image.
47. NLP stands for Natural Language Processing.
48. What is Azure Automated Machine Learning?
    Azure Machine Learning includes an automated machine learning capability that automatically tries multiple pre-processing techniques and model-training algorithms in parallel. These automated capabilities use the power of cloud compute to find the best performing supervised machine learning model for your data.

49. List the Challenges and Risk with AI.
    - Bias can affect results
    - Errors may cause harm
    - Data could be exposed
    - Solutions may not work for everyone
    - Users must trust a complex system
    - Who's liable for AI-driven decisions?
50. What are the uses of NLP?
    - Analyze and interpret text in documents, email messages, and other sources.
    - Interpret spoken language, and synthesize speech responses.
    - Automatically translate spoken or written phrases between languages.
    - Interpret commands and determine appropriate actions.
51. What is Relational Database?
    A relational database is a type of database that stores and provides access to data points that are related to one another.
52. What is hyper parameter?
    A hyperparameter is a machine learning parameter whose value is chosen before a learning algorithm is trained.
53. What is knowledge mining?
    Knowledge mining is the term used to describe solutions that involve extracting information from large volumes of often unstructured data to create a searchable knowledge store.
54. What is regression?
    Regression is used to predict a continuous value; like a price, a sales total, or some other measure.
55. What is the name of the app based on Computer Vision?
    Seeing AI.
56. What is Named entity recognition?
    Named entity recognition is where you can provide the Language service with unstructured text and it will return a list of entities in the text that it recognizes. For example: people, events, places, dates, and more.
57. What is the minimum number of Power BI datasets needed to support the reports?
    Two datasets are required.
58. What is dataset?
    In Azure Machine Learning, data for model training and other operations is usually encapsulated in an object called a dataset. A dataset is a collection of data that you import or connect to. Power BI lets you connect to and import all sorts of datasets and bring all of it together in one place. Datasets can also source data from dataflows.
59. What is text classification?
    Text classification is a machine learning technique that assigns a set of predefined categories to open-ended text. Text classifiers can be used to organize, structure, and categorize pretty much any kind of text.
60. What do you mean by language resource?
    A Language resource - choose this resource type if you only plan to use natural language processing services, or if you want to manage access and billing for the resource separately from other services.

Section-B: (2x5 = 10)

1. What are Computer Vision services in MS Azure?

Computer Vision - You can use this service to analyze images and video, and extract descriptions, tags, objects, and text.

Custom Vision - Use this service to train custom image classification and object detection models using your own images.

Face - The Face service enables you to build face detection and facial recognition solutions.

Form Recognizer - Use this service to extract information from scanned forms and invoices.

2. What is NLP? What is its use?
   Natural language processing (NLP) is the area of AI that deals with creating software that understands written and spoken language.
   Its uses are:
   - Analyze and interpret text in documents, email messages, and other sources.
   - Interpret spoken language, and synthesize speech responses.
   - Automatically translate spoken or written phrases between languages.
   - Interpret commands and determine appropriate actions.

3. What are hyperparameters? Give an example.
   A hyperparameter is a machine learning parameter whose value is chosen before a learning algorithm is trained.
   Examples of hyperparameters in machine learning include:
   - Model architecture.
   - Learning rate.
   - Number of epochs.
   - Number of branches in a decision tree.
   - Number of clusters in a clustering algorithm.

4. List any few tasks of computer vision.
   Tasks of computer vision are Image Classification, Object Detection, Semantic Segmentation, Image Analysis, Face Detection, analysis and recognition and Optical Character recognition.

5. Define AI. What are the key components of AI?
   AI is the creation of software that imitates human behaviors and capabilities.
   The key components of AI are Machine learning, Anomaly detection, Computer vision, Natural Language Processing and Knowledge mining.

6. List and explain two types of supervised machine learning task.
   Regression: used to predict a continuous value; like a price, a sales total, or some other measure.
   Classification: used to determine a binary class label; like whether a patient has diabetes or not.

7. Explain any two computer vision services in Microsoft Azure.
   Computer Vision - You can use this service to analyze images and video, and extract descriptions, tags, objects, and text.
   Custom Vision - Use this service to train custom image classification and object detection models using your own images.
   Face - The Face service enables you to build face detection and facial recognition solutions.
   Form Recognizer - Use this service to extract information from scanned forms and invoices.

8. What is image analysis? Explain.
   Image analysis techniques is used to extract information from images, including "tags" that could help catalog the image or even descriptive captions that summarize the scene shown in the image.

9. What are the cognitive services provided by MS Azure to create Computer Vision solutions?
   Computer Vision - You can use this service to analyze images and video, and extract descriptions, tags, objects, and text.
   Custom Vision - Use this service to train custom image classification and object detection models using your own images.
   Face - The Face service enables you to build face detection and facial recognition solutions.
   Form Recognizer - Use this service to extract information from scanned forms and invoices.

10. Which are the two NLP services in Microsoft Azure

Language - Use this service to access features for understanding and analyzing text, training language models that can understand spoken or text-based commands, and building intelligent applications.

Translator - Use this service to translate text between more than 60 languages.

Speech - Use this service to recognize and synthesize speech, and to translate spoken languages.

Azure Bot - This service provides a platform for conversational AI, the capability of a software "agent" to participate in a conversation.

11. List two guiding principles of responsible AI and explain?

Fairness: AI systems should treat all people fairly. For example, suppose you create a machine learning model to support a loan approval application for a bank. The model should predict whether the loan should be approved or denied without bias.

Reliability and safety: AI systems should perform reliably and safely.

Privacy and security: AI systems should be secure and respect privacy. The machine learning models on which AI systems are based rely on large volumes of data, which may contain personal details that must be kept private.

Inclusiveness: AI systems should empower everyone and engage people. AI should bring benefits to all parts of society, regardless of physical ability, gender, sexual orientation, ethnicity, or other factors.

Transparency: AI systems should be understandable. Users should be made fully aware of the purpose of the system, how it works, and what limitations may be expected.

Accountability: People should be accountable for AI systems. Designers and developers of AI-based solutions should work within a framework of governance and organizational principles that ensure the solution meets ethical and legal standards that are clearly defined.

12. What are the types of machine learning?

The two types of machine learning are supervised and unsupervised machine learning.

The **supervised machine learning** approach requires you to start with a dataset *with* known label values. Two types of supervised machine learning tasks include regression and classification.

The **unsupervised machine learning** approach starts with a dataset *without* known label values. One type of unsupervised machine learning task is clustering.

13. What are the cognitive services provided by MS Azure to create NaturalLanguage Processing solutions?

Language - Use this service to access features for understanding and analyzing text, training language models that can understand spoken or text-based commands, and building intelligent applications.

Translator - Use this service to translate text between more than 60 languages.

Speech - Use this service to recognize and synthesize speech, and to translate spoken languages.

Azure Bot - This service provides a platform for conversational AI, the capability of a software "agent" to participate in a conversation.

14. List and explain potential uses for computer vision.

Content Organization: Identify people or objects in photos and organize them based on that identification. Photo recognition applications like this are commonly used in photo storage and social media applications.

Text Extraction: Analyze images and PDF documents that contain text and extract the text into a structured format.

Spatial Analysis: Identify people or objects, such as cars, in a space and map their movement within that space.

15. What are the two ways of creating neural network models?
    - Create a neural network model using the default architecture
    - Define a custom architecture for a neural network

16. Describe any two language services available in MS Azure.
    - Key Phrase Extraction - Key phrase extraction is the concept of evaluating the text of a document, or documents, and then identifying the main talking points of the document(s).
    - Entity recognition - You can provide the Language service with unstructured text and it will return a list of *entities* in the text that it recognizes. The service can also provide links to more information about that entity on the web. An entity is essentially an item of a particular type or a category; and in some cases, subtype.
    - Sentiment analysis - The text analytics capabilities in the Language service can evaluate text and return sentiment scores and labels for each sentence. This capability is useful for detecting positive and negative sentiment in social media, customer reviews, discussion forums and more.

17. Name the six principles in AI software development.

    The six principles in AI software development are Fairness, Reliability and Safety, Privacy and Security, Inclusiveness, Transparency and Accountability.

18. What is Azure? What are four processes of azure automated machinelearning?

    Azure Automated Machine Learning enables non-experts to quickly create an effective machine learning model from data. These automated capabilities use the power of cloud compute to find the best performing supervised machine learning model for your data.

    The four processes of azure automated machinelearning are:

    Prepare data: Machine learning models must be trained with existing data. Data scientists expend a lot of effort exploring and pre-processing data, and trying various types of model-training algorithms to produce accurate models, which is time consuming, and often makes inefficient use of expensive computer hardware.

    Train model: The automated machine learning capability in Azure Machine Learning supports *supervised* machine learning models - in other words, models for which the training data includes known label values. You can use automated machine learning to train models for:
    - **Classification** (predicting categories or *classes*)
    - **Regression** (predicting numeric values)
    - **Time series forecasting** (predicting numeric values at a future point in time)

    Evaluate performance: After the job has finished you can review the best performing model. In this case, you used exit criteria to stop the job. Thus the "best" model the job generated might not be the best possible model, just the best one found within the time allowed for this exercise.

    Deploy a predictive service: In Azure Machine Learning, you can deploy a service as an Azure Container Instances (ACI) or to an Azure Kubernetes Service (AKS) cluster. For production scenarios, an AKS deployment is recommended, for which you must create an inference cluster compute target.

19. How NLP software works?

    NLP software can
    - Analyze and interpret text in documents, email messages, and other sources.
    - Interpret spoken language, and synthesize speech responses.
    - Automatically translate spoken or written phrases between languages.
    - Interpret commands and determine appropriate actions.

20. What are the Challenges and risks in AI?
    - Bias can affect results
    - Errors may cause harm
    - Data could be exposed
    - Solutions may not work for everyone
    - Users must trust a complex system
    - Who's liable for AI-driven decisions?

21. Machine Learning? Explain Briefly.

    This is often the foundation for an AI system, and is the way we "teach" a computer model to make prediction and draw conclusions from data. Machine learning is a technique that uses mathematics and statistics to create a model that can predict unknown values.

    How Machine learning works is, from data. In today's world, we create huge volumes of data as we go about our everyday lives. From the text messages, emails, and social media posts we send to the photographs and videos we take on our phones, we generate massive amounts of information. More data still is created by millions of sensors in our homes, cars, cities, public transport infrastructure, and factories. Data scientists can use all of that data to train machine learning models that can make predictions and inferences based on the relationships they find in the data.

22. What is azure Automated Machine Learning? Explain Briefly.

    Azure Machine Learning includes an automated machine learning capability that automatically tries multiple pre-processing techniques and model-training algorithms in parallel. These automated capabilities use the power of cloud compute to find the best performing supervised machine learning model for your data.

    Automated machine learning allows you to train models without extensive data science or programming knowledge. For people with a data science and programming background, it provides a way to save time and resources by automating algorithm selection and hyperparameter tuning.

23. What is the use of tune model Hyperparameter?

    Tune model Hyperparameter is used to determine the optimum hyperparameters for a machine learning model. Tune Model Hyperparameters can only be used to connect to built-in machine learning algorithm components.

24. Explain four compute resources available in azure machine learning.

    Compute Instances: Development workstations that data scientists can use to work with data and models.
    Compute Clusters: Scalable clusters of virtual machines for on-demand processing of experiment code.
    Inference Clusters: Deployment targets for predictive services that use your trained models.
    Attached Compute: Links to existing Azure compute resources, such as Virtual Machines or Azure Databricks clusters.

25. Define object detection and image classification.

    Object detection - Object detection machine learning models are trained to classify individual objects within an image, and identify their location with a bounding box. For example, a traffic monitoring solution might use object detection to identify the location of different classes of vehicle.

    Image classification - Image classification involves training a machine learning model to classify images based on their contents. For example, in a traffic monitoring solution you might use an image classification model to classify images based on the type of vehicle they contain, such as taxis, buses, cyclists, and so on.

26. List any two potential challenges and risks facing an AI application developer with an example.
    - **Bias can affect results** - A loan-approval model discriminates by gender due to bias in the data with which it was trained
    - **Errors may cause harm** - An autonomous vehicle experiences a system failure and causes a collision
    - **Data could be exposed** - A medical diagnostic bot is trained using sensitive patient data, which is stored insecurely
    - **Solutions may not work for everyone** - A home automation assistant provides no audio output for visually impaired users
    - **Users must trust a complex system** - An AI-based financial tool makes investment recommendations - what are they based on?
    - **Who's liable for AI-driven decisions?** - An innocent person is convicted of a crime based on evidence from facial recognition – who's responsible?

27. What do you mean by Cognitive services?
    Cognitive Services are cloud-based artificial intelligence (AI) services that help developers build cognitive intelligence into applications without having direct AI or data science skills or knowledge.

28. What are Computer Vision models and capabilities?
    Image classification, Object detection, Semantic segmentation, Image analysis, Face detection, analysis, and recognition and Optical character recognition (OCR).

29. When might you see NaN returned for a score in Language Detection?
    There may be text that is ambiguous in nature, or that has mixed language content. An ambiguous content example would be a case where the document contains limited text, or only punctuation. For example, using the service to analyze the text ":-)", results in a value of unknown for the language name and the language identifier, and a score of NaN is returned**.**

30. What is Neural Network Regression component?
    Neural network regression component is a supervised learning method, and therefore requires a tagged dataset, which includes a label column. Because a regression model predicts a numerical value, the label column must be a numerical data type.

31. What do you mean by conversational AI?
    Conversational AI is an umbrella term used to describe various methods of enabling computers to carry on a conversation with a human. This technology ranges from fairly simple NLP to more sophisticated machine learning models that can interpret a much wider range of inputs and carry on more complex conversations.

32. What is Azure machine learning studio? Explain briefly.
    Azure Machine Learning studio is a web portal for machine learning solutions in Azure. It includes a wide range of features and capabilities that help data scientists prepare data, train models, publish predictive services, and monitor their usage.

33. Explain any two types of Machine learning.
    There are two general approaches to machine learning, supervised and unsupervised machine learning.
    The **supervised machine learning** approach requires you to start with a dataset *with* known label values. Two types of supervised machine learning tasks include regression and classification.
    - **Regression**: used to predict a continuous value; like a price, a sales total, or some other measure.
    - **Classification**: used to determine a class label; an example of a binary class label is whether a patient has diabetes or not; an example of multi-class labels is classifying text as positive, negative, or neutral.

    The **unsupervised machine learning** approach starts with a dataset *without* known label values. One type of unsupervised machine learning task is clustering.
    - **Clustering**: used to determine labels by grouping similar information into label groups; like grouping measurements from birds into species.

34. How Machine Learning works?

    Machine Learning works from data. In today's world, we create huge volumes of data as we go about our everyday lives. From the text messages, emails, and social media posts we send to the photographs and videos we take on our phones, we generate massive amounts of information. More data still is created by millions of sensors in our homes, cars, cities, public transport infrastructure, and factories.

    Data scientists can use all of that data to train machine learning models that can make predictions and inferences based on the relationships they find in the data.

35. What are the Responsibilities of AI? Explain.

    Fairness: AI systems should treat all people fairly. For example, suppose you create a machine learning model to support a loan approval application for a bank. The model should predict whether the loan should be approved or denied without bias.

    Reliability and safety: AI systems should perform reliably and safely.

    Privacy and security: AI systems should be secure and respect privacy. The machine learning models on which AI systems are based rely on large volumes of data, which may contain personal details that must be kept private.

    Inclusiveness: AI systems should empower everyone and engage people. AI should bring benefits to all parts of society, regardless of physical ability, gender, sexual orientation, ethnicity, or other factors.

    Transparency: AI systems should be understandable. Users should be made fully aware of the purpose of the system, how it works, and what limitations may be expected.

    Accountability: People should be accountable for AI systems. Designers and developers of AI-based solutions should work within a framework of governance and organizational principles that ensure the solution meets ethical and legal standards that are clearly defined.

Section-C: (2x5 = 10)

1. What are Power BI visuals? Explain

   Power BI visuals are attractive charts and graphics that you can use to revitalize your data. Visuals allow you to share data insights more effectively and increase comprehension, retention, and appeal. Visuals are a fundamental part of your report because they help your report audience connect and interact with the information to make informed decisions quickly.

2. What is an outlier? How Power BI allows you to identify outliers in yourdata?

   An outlier is a type of anomaly in your data, something that you didn't expect or that surprised you, based on historical averages or results.

   Power BI allows you to identify outliers in your data, but you need to first determine the logic behind what constitutes an outlier. You can use trigger points, such as calculations, around what you would consider the outlier to be.

   The process of identifying outliers involves segmenting your data into two groups: one group is the outlier data and the other group is not. You could use calculated columns to identify outliers, but the results would be static until you refresh the data. A better way to identify outliers is to use a visualization or DAX formula because these methods will ensure that your results are dynamic.

   When you have identified the outliers in your data, you can then use slicers or filters to highlight those outliers. Additionally, you can add a legend to your visuals so that the outliers can be identified among the other data. You can then drill in to the outlier data for more detailed analysis.

3. Explain the components of analytics.

   Descriptive analytics - Descriptive analytics help answer questions about what has happened based on historical data. An example of descriptive analytics is generating reports to provide a view of an organization's sales and financial data.

   Diagnostic analytics - Diagnostic analytics help answer questions about why events happened. Diagnostic analytics techniques supplement basic descriptive analytics, and they use the findings from descriptive analytics to discover the cause of these events.

   Predictive analytics - Predictive analytics help answer questions about what will happen in the future. Predictive analytics techniques use historical data to identify trends and determine if they're likely to recur.

   Prescriptive analytics - Prescriptive analytics help answer questions about which actions should be taken to achieve a goal or target. By using insights from prescriptive analytics, organizations can make data-driven decisions.

   Cognitive analytics - Cognitive analytics attempt to draw inferences from existing data and patterns, derive conclusions based on existing knowledge bases, and then add these findings back into the knowledge base for future inferences, a self-learning feedback loop. Cognitive analytics help you learn what might happen if circumstances change and determine how you might handle these situations.

4. What process occurs in Diagnostic analytics?

   The process that occurs in Diagnostic analytics are:

   - Identify anomalies in the data. These anomalies might be unexpected changes in a metric or a particular market.
   - Collect data that's related to these anomalies.
   - Use statistical techniques to discover relationships and trends that explain these anomalies.

5. Explain Predictive analytics with example?

   Predictive analytics help answer questions about what will happen in the future. Predictive analytics techniques use historical data to identify trends and determine if they're likely to recur. Predictive analytical tools provide valuable insight into what might happen in the future. Techniques include a variety of statistical and machine learning techniques such as neural networks, decision trees, and regression.

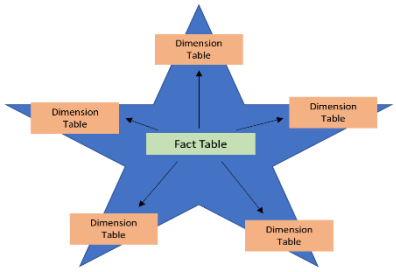
6. Explain the role of data analysts.

   A data analyst enables businesses to maximize the value of their data assets through visualization and reporting tools such as Microsoft Power BI. Data analysts are responsible for profiling, cleaning, and transforming data. Their responsibilities also include designing and building scalable and effective data models, and enabling and implementing the advanced analytics capabilities into reports for analysis. A data analyst is also responsible for the management of Power BI assets, including reports, dashboards, workspaces, and the underlying datasets that are used in the reports.

7. What are the benefits of the good data model?

   - Data exploration is faster.
   - Aggregations are simpler to build.
   - Reports are more accurate.
   - Writing reports takes less time.
   - Reports are easier to maintain in the future.

8. What do you mean by Star schemas?
   Star schemas - You can design a star schema to simplify your data. It's not the only way to simplify your data, but it is a popular method; therefore, every Power BI data analyst should understand it. In a star schema, each table within your dataset is defined as a dimension or a fact table.



   Fact tables contain observational or event data values: sales orders, product counts, prices, transactional dates and times, and quantities. Fact tables can contain several repeated values. For example, one product can appear multiple times in multiple rows, for different customers on different dates.
   Dimension tables contain the details about the data in fact tables: products, locations, employees, and order types. These tables are connected to the fact table through key columns. Dimension tables are used to filter and group the data in fact tables. The dimension tables, by contrast, contain unique values, for instance, one row for each product in the Products table and one row for each customer in the Customer table.

9. Define data granularity. Explain
   Data granularity is the detail that is represented within your data, meaning that the more granularity your data has, the greater the level of detail within your data.

   Consider a scenario where your company manages 1,000 refrigerated semi-trucks. Every few minutes, each truck uses a Microsoft Azure IoT application to record its current temperature. This temperature is important to your organization because, if the refrigeration were to malfunction, it could spoil the entire load, costing thousands of dollars. With so many trucks and so many sensors, extensive data is generated every day. Your report users don't want to sift through numerous records to find the ones that they are particularly interested in.

   In this scenario, you might want to import the data by using a daily average for each truck. That approach would reduce the records in the database to one record for each truck for each day. If you decide that the approach was acceptable enough for tracking costs and errors, then you could use that data granularity.

10. What is the role of Deployment pipeline feature in the development lifecycle?
    The deployment pipeline feature in Power BI manages content in dashboards, reports, and datasets between different environments in the development life cycle. With this feature, you can develop and test Power BI content in one centralized location and streamline the process before deploying the final content to your users. This Power BI Premium feature requires you to be a Capacity admin.

    The advantages of using the deployment pipeline are:

    - **Increased productivity** - Through this feature, you can reuse previous deployment pipelines, ensuring that efforts aren't duplicated.
    - **Faster delivery of content** - Report development becomes more streamlined, meaning that it takes less time to get to production.
    - **Lower human intervention required** - Having the ability to reuse deployment pipelines means a decreased chance of error associated with moving content from one environment to another.

11. Who is a data analyst? Name the tasks of data analyst.

A data analyst is one of several critical roles in an organization, who help uncover and make sense of information to keep the company balanced and operating efficiently.

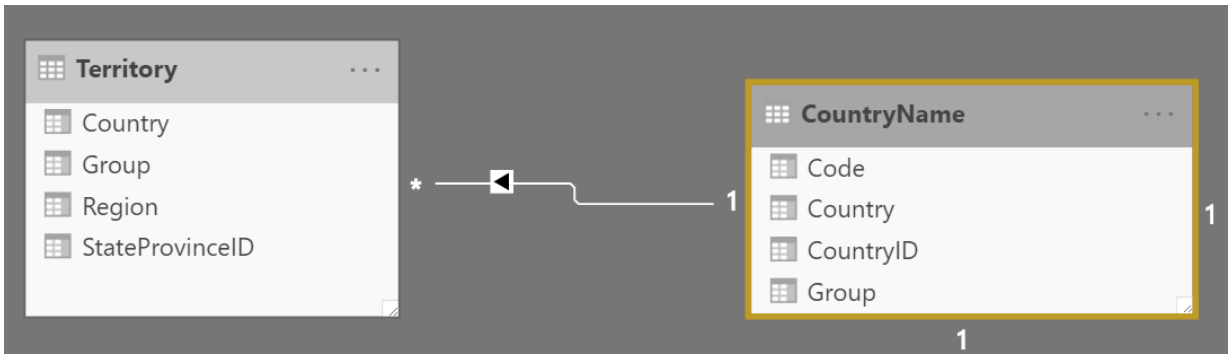
The tasks of Data Analyst are Prepare, Model, Visualize, Analyze and Manage.

12. What are Many-to-One and One-to-Many relationship in Power BI?

Many-to-one (*:1) or one-to-many (1: *) relationship

- Describes a relationship in which you have many instances of a value in one column that are related to only one unique corresponding instance in another column.
- Describes the directionality between fact and dimension tables.
- Is the most common type of directionality and is the Power BI default when you are automatically creating relationships.

An example of a one-to-many relationship would be between the CountryName and Territory tables, where you can have many territories that are associated with one unique country.



13. What is smart narrative visualization? What is its use?

The smart narrative visualization helps you quickly summarize visuals and reports. It provides relevant innovative insights that you can customize.

It's uses are:

- Use smart narrative summaries in your reports to address key takeaways, to point out trends, and to edit the language and format for a specific audience.
- In PowerPoint, instead of pasting a screenshot of your report's key takeaways, you can add narratives that are updated with every refresh.
- Your audience can use the summaries to understand the data, get to key points faster, and explain the data to others.

14. What can you do to improve performance when you're getting data in Power BI?

Using query folding within Power Query Editor helps you increase the performance of your Power BI reports. Power Query takes advantage of good performance at the data source through a technique called Query Folding. You can use Power Query to load data into Power BI. Then use Power Query Editor to transform your data, such as renaming or deleting columns, appending, parsing, filtering, or grouping your data.

You can also Process as much data as possible in the original data source, Use native SQL queries and Separate date and time, if bound together to optimize query performance in Power BI.

15. What is the use of Line and area charts?

The line chart and area chart visualizations are beneficial in helping you present trends over time. The basic area chart is based on the line chart, with the area between axis and line filled in. The main difference between these two chart types is that the area chart highlights the magnitude of change over time.

16. Write the difference between prescriptive analytics with predictiveanalytics.

Prescriptive analytics	Predictive analytics
Prescriptive analytics help answer questions about which actions should be taken to achieve a goal or target.	Predictive analytics help answer questions about what will happen in the future.
By using insights from prescriptive analytics, organizations can make data-driven decisions.	Predictive analytics techniques use historical data to identify trends and determine if they're likely to recur.

17. What are the benefits of using a good data model?
    - Data exploration is faster.
    - Aggregations are simpler to build.
    - Reports are more accurate.
    - Writing reports takes less time.
    - Reports are easier to maintain in the future.

18. What do you mean by statistical summary in data analytics? Explain.
    The statistical summary is the information that provides a quick and simple description of your data. Exploring the statistical summary gives the user a high-level view of the available data, where they can see clusters, patterns on behavioral data, data averages, and more. They can gain insights about their data that will help drive business decisions.

19. How does Power BI clustering technique useful?
    Clustering allows you to identify a segment (cluster) of data that is similar to each other but dissimilar to the rest of the data. The Power BI clustering feature allows you to quickly find groups of similar data points in a subset of your data. It analyzes your dataset to identify similarities and dissimilarities in the attribute values, and then it separates the data that has similarities into a subset of the data.

    These subsets of data are referred to as clusters. For example, you might want to look for patterns in your sales data, such as the behavior of customers overall. You can segment the customers into clusters according to their similarities, such as age or location.

20. What are table and matrix visualizations in power BI? Explain.
    The **table** is a grid that contains related data in a logical series of rows and columns. The table supports two dimensions and it can also contain headers and a row for totals.
    The **Matrix** visualization looks similar to the table visualization; however, it allows you to select one or more elements (rows, columns, values) in the matrix to cross-highlight other visuals on the report page.

21. What is statistical summary? List any two Power BI functions used toconduct statistical analysis?
    The statistical summary is the information that provides a quick and simple description of your data.
    Power BI has many functions that help you to conduct a statistical analysis, such as Data Analysis Expressions (DAX) functions, visuals such as histograms and bell curves, advanced analytics visuals, and statistical programming languages such as Python and R.

22. Explain workspace?
    A workspace is a centralized repository in which you can collaborate with colleagues and teams to create collections of reports and dashboards.

    Workspaces offer the following benefits:
    - Focused collaboration efforts. You can use workspaces to house reports and dashboards for use by multiple teams.
    - Ability to share and present reports and dashboards in a single environment.
    - Assurance that the highest level of security is maintained by controlling who can access datasets, reports, and dashboards.

23. Explain the features of key influencers visual?

- **Tabs**: Select a tab to switch between views. **Key influencers** shows you the top contributors to the selected metric value. **Top segments** shows you the top segments that contribute to the selected metric value. A *segment* is made up of a combination of values. For example, one segment might be consumers who have been customers for at least 20 years and live in the west region.
- **Drop-down box**: The value of the metric under investigation. In this example, look at the metric **Rating**. The selected value is **Low**.
- **Restatement**: It helps you interpret the visual in the left pane.
- **Left pane**: The left pane contains one visual. In this case, the left pane shows a list of the top key influencers.
- **Restatement**: It helps you interpret the visual in the right pane.
- **Right pane**: The right pane contains one visual. In this case, the column chart displays all the values for the key influencer **Theme** that was selected in the left pane. The specific value of **usability** from the left pane is shown in green. All the other values for **Theme** are shown in black.
- **Average line**: The average is calculated for all possible values for **Theme** except **usability** (which is the selected influencer). So the calculation applies to all the values in black. It tells you what percentage of the other **Themes** had a low rating. In this case 11.35% had a low rating (shown by the dotted line).
- **Check box**: Filters out the visual in the right pane to only show values that are influencers for that field. In this example, the visual is filtered to display usability, security, and navigation.

24. Which data role enables advanced analytics capabilities specificallythrough reports and visualizations?

A data analyst enables businesses to maximize the value of their data assets through visualization and reporting tools such as Microsoft Power BI. Their responsibilities also include designing and building scalable and effective data models, and enabling and implementing the advanced analytics capabilities into reports for analysis.

25. What is the difference between a fact table and a dimension table?

Fact table	Dimension table
Fact tables contain observational or event data values: sales orders, product counts, prices, transactional dates and times, and quantities.	Dimension tables contain the details about the data in fact tables : products, locations, employees, and order types.
Fact tables can contain several repeated values. For example, one product can appear multiple times in multiple rows, for different customers on different dates.	The dimension tables, by contrast, contain unique values, for instance, one row for each product in the Products table and one row for each customer in the Customer table.

26. What Power BI feature can give an in-depth analysis of the distribution ofdata?

The Analyze feature can give an in-depth analysis of the distribution of data. The Analyze feature allows a user to understand why the distribution looks the way that it does.

27. What feature in Power BI service can you use to troubleshoot the flow ofdata from its source to its destination?

Lineage view allows you to view and troubleshoot the data flow from source to destination.

28. What are the core components of data analytics?
    The core components of data analytics are:
    - Descriptive
    - Diagnostic
    - Predictive
    - Prescriptive
    - Cognitive

29. What are the tasks of a data analyst?
    - Prepare - Data preparation is the process of taking raw data and turning it into information that is trusted and understandable.
    - Model - The model is another critical component that has a direct effect on the performance of your report and overall data analysis.
    - Visualize - The visualization task is where you get to bring your data to life. The ultimate goal of the visualize task is to solve business problems.
    - Analyze - The analyze task is the important step of understanding and interpreting the information that is displayed on the report.
    - Manage - The management of your content helps to foster collaboration between teams and individuals.

30. Name the different visualization options available in Power BI Desktop.
    The different visualization options available include charts, maps, cards, a table, a matrix.

31. What are the different types of storage modes in data analytics? Explain.
    - **Import mode** - The Import mode allows you to create a local Power BI copy of your datasets from your data source.
    - **DirectQuery mode** - The DirectQuery option is useful when you don't want to save local copies of your data because your data won't be cached.
    - **Dual (Composite mode)** - In Dual mode, you can identify some data to be directly imported and other data that must be queried. Any table that is brought in to your report is a product of both Import and DirectQuery modes. Using the Dual mode allows Power BI to choose the most efficient form of data retrieval.

32. What do you mean by Time series analysis? Explain.
    Time series analysis involves analyzing a series of data in time order to identify meaningful information and trends and make predictions. The result of time series analysis is the best data that you can use for forecasting activities.

    Time series analysis often involves the use of visuals such as Gantt charts, project planning, and stock movement datasets. To conduct a time series analysis in Power BI, you need to use a visualization type that is suitable for displaying trends and changes over time, such as a line chart, area chart, or scatter chart.

33. Explain briefly on NoSQL.
    A NoSQL database (also referred to as non-SQL, not only SQL or *non-relational*) is a flexible type of database that does not use tables to store data. Some organizations don't use a relational database but instead use a *NoSQL* database.

34. What is Slicer visualization?
    The **slicer** visualization is a standalone chart that can be used to filter the other visuals on the page. Slicers provide a more advanced and customized way of filtering, in comparison to the **Filters** pane, which is suited to more basic filtering operations. You can learn more about these two filtering options in another module.

35. What are predictive and cognitive analytics?

Predictive analytics - Predictive analytics help answer questions about what will happen in the future. Predictive analytics techniques use historical data to identify trends and determine if they're likely to recur.

Cognitive analytics - Cognitive analytics attempt to draw inferences from existing data and patterns, derive conclusions based on existing knowledge bases, and then add these findings back into the knowledge base for future inferences, a self-learning feedback loop. Cognitive analytics help you learn what might happen if circumstances change and determine how you might handle these situations.

36. Differentiate SQL and NoSQL.

SQL	**NoSQL**
SQL stands for Structured Query Language and is a standardized programming language that is used to manage relational databases and perform various data management operations.	A NoSQL database (also referred to as non-SQL, not only SQL or *non-relational*) is a flexible type of database that does not use tables to store data.
Examples of SQL: MySQL, PostgreSQL, Oracle, MS-SQL Server, etc.	Examples of NoSQL: MongoDB, GraphQL, HBase, Neo4j, Cassandra, etc.

37. What is Query Folding?

Query folding is the process by which the transformations and edits that you make in Power Query Editor are simultaneously tracked as native queries, or simple **Select** SQL statements, while you're actively making transformations. The query folding within Power Query Editor helps you increase the performance of your Power BI reports.

38. What is Query Diagnostics?

Query diagnostics is another tool that you can use to study query performance. You can determine what bottlenecks may exist while loading and transforming your data, refreshing your data in Power Query, running SQL statements in Query Editor, and so on.

39. What are the benefits to Query Folding?

The benefits to query folding include:

- **More efficiency in data refreshes and incremental refreshes.** When you import data tables by using query folding, Power BI is better able to allocate resources and refresh the data faster because Power BI doesn't have to run through each transformation locally.
- **Automatic compatibility with DirectQuery and Dual storage modes.** All DirectQuery and Dual storage mode data sources must have the back-end server processing abilities to create a direct connection, which means that query folding is an automatic capability that you can use. If all transformations can be reduced to a single **Select** statement, then query folding can occur.