**Question 2: Describe the data processing chain.**

To get new value and insight, we need to analysis exiting data. For analyzing data, there is a data processing cycle, which was suggested to follow:

Data 🡺Database 🡺 Data warehouse 🡺 Data Mining 🡺 Data visualization

**Data:**  Data might be come from different source like from database of an organization, customer and owner interaction, government body where data type can be also different. In data analytic, data can be classified into different type such as nominal, order, interval, ratio data. apart from, another kind of data which does not lend to that much mathematical analysis. for instance – audio, video and graph files.

**Database:** A database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a database management system (DBMS). It is modeled collection of data which is accessible in many ways. A data model can be designed to integrate the operational data of the organization. Now a days, many database management tools are available like Oracle , MySQL, MongoDB,Cassandra to help store, manage and model data.

**Data warehouse:** Data warehouse is designed to help organization to decision making which is simpler version of operational database. it is an organized data store where at first, data can be extracted from operational database to find out answer of some queries. this data, might be combined with other data and rolled up to a consistent granularity. Finally, it is uploaded to a separated data store which is named data warehouse.

**Data Mining:** Data Mining is a part of data analysis where it is defined as a process of extracting unused data from a large data set. it is a method of recognizing and discovering hidden pattern throughout whole data set. There are several data mining technique for data analytics. Such as (a)Decision tree - which used different algorithm to make decision tree. (b)Regression - used for figuring out best fitting curve through the many data point (c) Artificial Neural Networks (d) Cluster analysis and so on.

**Data visualization:** Data is being gradually increased as business is growing every day. This data seems to be piling up faster than processing. As a result, it is challenging to make sense or decision making from this vast amount of data. data visualization provides us opportunity to convert raw data to visual representation which can be easily understood. Normally, visualization come in the form of charts, diagram, graph and map which allow users to trace and analyze trends and patterns.

**Question 4. What are the different data mining techniques? Which of these would be relevant in your current work?**

Data mining is the way to extract or excavate knowledge or patterns from large data set, where different techniques are followed. Bellow popular techniques are used for data mining:

**Decision trees:** it is one of the popular data mining techniques which provide assistance to organization for figuring out pattern from business data set. Decision trees actually particular types of predictive model which extensively used for machine learning technique, because of its boldly straightforward flow of natural. Numbers of algorithms are available to make decision tree which can be used for different circumstances.

**Regression:** it is a technique which plays role to figure out relation amongvariables of dataset. It is a simple and straightforward way to expose, how variables are connected with other. Besides, Regressions model find out best fitting curve among several data points.

**Artificial Neural Network:** it is specific types of model of Machine and deep learning. ANN (Artificial Neural Network) is a technique of non- linear data processing that gain knowledge from old data and provide prediction of future benefits.

**Cluster Analysis:**  Cluster analysis is another technique of data mining which depend on visual approach to discern data. It is the way of data mining for dividing and concurring vast data set that assist to depict similarities and dissimilarities within data.

Apart from above mentioned data mining technique, there are some other data mining techniques such classification, association rule, outliner detection and so on.

**Question 6 :**

**Answer:**

By reading number of blog posts, I found several ways to display weather information together. we can use separate technology and display real-time weather information. For instance, if we can have any exiting API which read temperature, humidity , wind flow , rain and snow , we can get weather data from that API and send this to real-time dashboard like **Grafana** for visualization . If we need to perform any extra analysis based on API data, for that, we can use **Apache Kafka**.

Apart from, We can use Microsoft excel according to following blog post:

<https://www.visualcrossing.com/resources/blog/using-microsoft-excel-to-analyze-gardening-weather-and-climate/>