1. What is what if Analysis?

**What-If Analysis** in **Excel** allows you to try out different values (scenarios) for formulas.

1. List different types of What if Analysis.

What-If Analysis in Excel is a tool which helps us to create different models, scenarios, data table. In this article, we will look at the ways of using What-If Analysis.

We have 3 parts of What-If Analysis in Excel. They are as follows:

* 1. Scenario Manager
  2. Goal Seek in Excel
  3. Data Table in Excel

3. Do we always need a formulae cell to do what-if analysis?

4. What is the advantage of sorting?

it helps arrange all records in a table.

it is able to deal well with a huge list of items.

no additional storage is required.

5. What is the advantage of regression analysis?

**Predictive analytics**

**Operation efficiency**

**Supporting decisions**

**Correcting errors**

**New Insights**

6. What is sorting?

Sorting refers to ordering data in an increasing or decreasing fashion according to some linear relationship among the data items

7. What is the filter?

*The filter in excel helps display relevant data by eliminating the irrelevant entries temporarily from the view. The data is filtered as per the given criteria*

8. What is machine learning?

Machine learning is a branch of [artificial intelligence (AI)](https://www.ibm.com/cloud/learn/what-is-artificial-intelligence) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

9. Define statistics

Statistics is a collection of tools that you can use to get answers to important questions about data

10. Where is unsupervised learning used?

In supervised learning, the main idea is to learn under supervision, where the supervision signal is named as target value or label. In unsupervised learning, we lack this kind of signal. Therefore, we need to find our way without any supervision or guidance. This simply means that we are alone and need to figure out what is what by ourselves.

5 marks

1. Explain correlation with the help of an example.
   1. Correlation means association - more precisely it is a measure of the extent to which two variables are related. There are three possible results of a correlational study: a positive correlation, a negative correlation, and no correlation.
   2.  A **positive correlation** is a relationship between two variables in which both variables move in the same direction. Therefore, when one variable increases as the other variable increases, or one variable decreases while the other decreases. An example of positive correlation would be height and weight. Taller people tend to be heavier.
   3.  A **negative correlation** is a relationship between two variables in which an increase in one variable is associated with a decrease in the other. An example of negative correlation would be height above sea level and temperature. As you climb the mountain (increase in height) it gets colder (decrease in temperature).
   4.  A **zero correlation** exists when there is no relationship between two variables. For example there is no relationship between the amount of tea drunk and level of intelligence.

Some uses of Correlations

**Prediction**

* If there is a relationship between two variables, we can make predictions about one from another.

**Validity**

* Concurrent validity (correlation between a new measure and an established measure).

**Reliability**

* Test-retest reliability (are measures consistent).

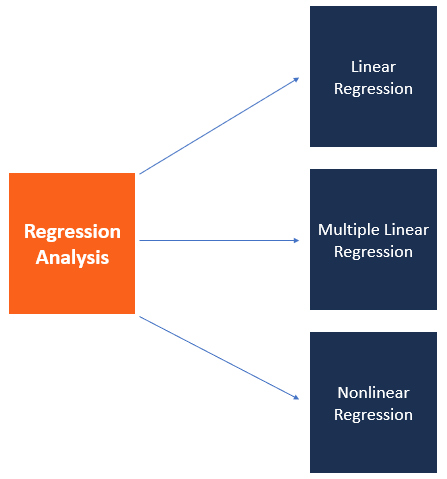
Inter-rater reliability (are observers consistent).

**Theory verification**

* Predictive validity.

1. Explain regression analysis.

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more [independent variables](https://corporatefinanceinstitute.com/resources/knowledge/modeling/independent-variable/). It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.



### Regression Analysis – Linear Model Assumptions

Linear regression analysis is based on six fundamental assumptions:

1. The dependent and independent variables show a linear relationship between the slope and the intercept.
2. The independent variable is not random.
3. The value of the residual (error) is zero.
4. The value of the residual (error) is constant across all observations.
5. The value of the residual (error) is not correlated across all observations.
6. The residual (error) values follow the normal distribution.

### Regression Analysis – Simple Linear Regression

Simple linear regression is a model that assesses the relationship between a dependent variable and an independent variable. The simple linear model is expressed using the following equation:

##### Y = a + bX + ϵ

##### Where:

* **Y** – Dependent variable
* **X** – Independent (explanatory) variable
* **a** – Intercept
* **b** – Slope
* **ϵ** – Residual (error)

### Regression Analysis – Multiple Linear Regression

Multiple linear regression analysis is essentially similar to the simple linear model, with the exception that multiple independent variables are used in the model. The mathematical representation of multiple linear regression is:

Y = a + bX1 + cX2+ dX3 + ϵ

Where:

* **Y** – Dependent variable
* **X1, X2, X3**– Independent (explanatory) variables
* **a** – Intercept
* **b, c, d** – Slopes
* **ϵ** – Residual (error)

Multiple linear regression follows the same conditions as the simple linear model. However, since there are several independent variables in multiple linear analysis, there is another mandatory condition for the model:

* **Non-collinearity:**Independent variables should show a minimum correlation with each other. If the independent variables are highly correlated with each other, it will be difficult to assess the true relationships between the dependent and independent variables.

1. Discuss the role of data filter.

Data filtering is the process of choosing a smaller part of your data set and using that subset for viewing or analysis. Filtering is generally (but not always) temporary – the complete data set is kept, but only part of it is used for the calculation.

Filtering may be used to:

* Look at results for a particular period of time.
* Calculate results for particular groups of interest.
* Exclude erroneous or "bad" observations from an analysis.
* Train and validate statistical models.

Filtering requires you to specify a rule or logic to identify the cases you want to included in your analysis. Filtering can also be referred to as “subsetting” data, or a data “drill-down”. In this article we illustrate a filtered data set and discuss how you might use filtering.

1. What is the difference between filter and sorting?

**The difference between sorting and filtering are as follows-**

1. **Sorting** of data refers to the arrangements of data that add meaning to the data collected. It may involve re-arranging data in a specific order. For example, numerical data collected can be **sorted** in increasing or decreasing order. Whereas **filtering** the data means eliminating some information that is not applicable.

2. **Sorting** is done for better organization and easy analysis whereas **filtering** the data is done to narrow down the probabilities of the inferences.

Sorting and filtering offer two different ways to customize the view of data in your sheet. You can't do both at the same time so it's best to determine in advance which option will help more. Sorting lets you organize all or part of your data in ascending or descending order.

1. Name 5 programming languages most commonly used for AI.

**Popular AI Programming Languages**

* **Python**
* **Java**
* **C++**
* **R**
* **Prolog**

#### **Python:**

Python is an interpreted, high-level, and general-purpose programming language. It is one of the most popular programming languages widely used in AI and machine learning apps

**Java:**

Java has been consistently among the top programming languages in use for many years. Developed in 1995, Java is high-level, class-based, and object-oriented programming language that has created history with its write once, run anywhere (WORA) principle.

#### **C++:**

Known as an extension of the popular C programming language, C++ is a general-purpose programming language created by Bjarne Stroustrup.

#### **R Programming Language:**

When it comes to statistical programming, R is the language that comes first into the mind. R is very popular for data analysis and statistical computing.

#### **Prolog:**

Prolog stands for ‘Programming in Logic.’ It is mainly used for AI and computational linguistics.

1. Discuss descriptive statistics.
   1. Descriptive statistics are brief descriptive coefficients that summarize a given data set, which can be either a representation of the entire population or a sample of a population.
   2. Descriptive statistics summarizes or describes the characteristics of a data set.
   3. Descriptive statistics consists of two basic categories of measures: measures of central tendency and measures of variability (or spread).
   4. Measures of central tendency describe the center of a data set.
   5. Measures of variability or spread describe the dispersion of data within the set.
2. Discuss inferential statistics.
   1. Inferential statistics is one of the two main branches of statistics.
   2. Inferential statistics use a random sample of data taken from a population to describe and make inferences about the population. Inferential statistics are valuable when examination of each member of an entire population is not convenient or possible. For example, to measure the diameter of each nail that is manufactured in a mill is impractical. You can measure the diameters of a representative random sample of nails. You can use the information from the sample to make generalizations about the diameters of all of the nails.
3. What are Pivot Tables? How it helps in data interpretation and processing.

A pivot table is a table you can use to store the summary of a specific data set in a condensed manner. This data can be from a database, spreadsheet, or any other business intelligence program. Such a table consists of rows, columns, pages, and data fields. They usually include sums, averages, or other statistics, which the pivot table groups together in a meaningful way. It enables the user to view the differences in a vast set of information.

One of the reasons we make Pivot Tables is to transfer information.

**Simplicity.**

**Speed.**

**Flexibility**

**Accuracy**

**Formatting**

**Filtering.**

**Charts.**

9)How Machine learning helps in recommending your favourite movie/web series at Online Media or suggesting your products online.

**A product recommendation engine is essentially a solution that allows marketers to offer their customers relevant product recommendations in real-time. As powerful data filtering tools, recommendation systems use algorithms and data analysis techniques to recommend the most relevant product/items to a particular user.**

## **How does a Recommendation Engine Work?**

**One of the crucial components behind the working of a product recommendation engine is the recommender function, which considers specific information about the user and predicts the rating that the user might assign to a product.**

**Having the ability to predict user ratings, even before the user has provided one, makes recommender systems a powerful tool.**

**It uses specialized algorithms and techniques that can support even the largest of product catalogs. Driven by an orchestration layer, the recommendation engine can intelligently select which filters and algorithms to apply in any given situation for a specific customer. It allows marketers to maximize conversions and also their average order value.**

**Typically, a recommendation engine processes data through the below four phases-**

* **Collection**

**Data collected here can be either explicit such as data fed by users (ratings and comments o n products) or implicit such as page views, order history/return history, and cart events.**

* **Storing**

**The type of data you use to create recommendations can help you decide the kind of storage you should use, like the NoSQL database, a standard SQL database, or object storage.**

* **Analyzing**

**The recommender system analyzes and finds items with similar user engagement data by filtering it using different analysis methods such as batch analysis, real-time analysis, or near-real-time system analysis.**

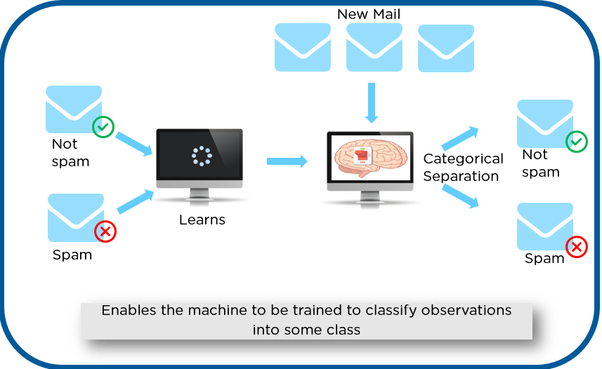
* **Filtering**

**The last step is to filter the data to get the relevant information required to provide recommendations to the user. And for enabling this, you will need to choose an algorithm suiting the recommendation engine from the list of algorithms explained in the next section.**

10)Discuss supervised learning with the help of an example.

In Supervised learning, an AI system is presented with data which is labeled, which means that each data tagged with the correct label.

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.



As shown in the above example, we have initially taken some data and marked them as ‘Spam’ or ‘Not Spam’. This labeled data is used by the training supervised model, this data is used to train the model.

Once it is trained we can test our model by testing it with some test new mails and checking of the model is able to predict the right output.

## Types of Supervised learning

* + 1. marks

1)What is the role of charts in data analysis? Explain with examples.

In the era of information explosion, more and more data piles up. However, these dense data are unfocused and less readable. So we need [data visualization](http://www.finereport.com/en/data-visualization/data-visualization-2.html?utm_source=medium&utm_medium=media&utm_campaign=blog&utm_term=08) to help data to be easily understood and accepted. By contrast, visualization is more intuitive and meaningful, and it is very important to use appropriate charts to visualize data.

# 1. Column Chart

Column charts use vertical columns to show numerical comparisons between categories, and the number of columns

# 2. Bar Chart

Bar charts are similar to column charts, but the number of bars can be relatively large. Compared with the column chart, the positions of its two axes are changed.

# 3. Line Chart

A line chart is used to show the change of data over a continuous time interval or time span. It is characterized by a tendency to reflect things as they change over time or ordered categories.

# 4. Area Chart

The area chart is formed on the basis of the line chart. It fills the area between the polyline and the axis in the line chart with color. The filling of the color can better highlight the trend information.

# 5. Pie Chart

Pie charts are widely used in various fields to represent the proportion of different classifications, and to compare various classifications by the arc.

# 6. Scatter Plot

The scatter plot shows two variables in the form of points on a rectangular coordinate system. The position of the point is determined by the value of the variable. By observing the distribution of the data points, we can infer the correlation between the variables.

# 7. Bubble Chart

A bubble chart is a multivariatechart that is a variant of a scatter plot. Except for the values of the variables represented by the X and Y axes, the area of each bubble represents the third value.

# 8. Gauge

A gauge in data visualization is a kind of materialized chart. The scale represents the metric, the pointer represents the dimension, and the pointer angle represents the value. It can visually represent the progress or actual situation of an indicator.

2-Name and discuss types of Machine Learning.

According to Arthur Samuel, Machine Learning algorithms enable the computers to learn from data, and even improve themselves, without being explicitly programmed.

Machine learning (ML) is a category of an algorithm that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. The basic premise of machine learning is to build algorithms that can receive input data and use statistical analysis to predict an output while updating outputs as new data becomes available.

Types of Machine Learning?

Machine learning can be classified into 3 types of algorithms.

1. Supervised Learning

2. Unsupervised Learning

3. Reinforcement Learning

Overview of Supervised Learning Algorithm

In Supervised learning, an AI system is presented with data which is labeled, which means that each data tagged with the correct label.

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

As shown in the above example, we have initially taken some data and marked them as ‘Spam’ or ‘Not Spam’. This labeled data is used by the training supervised model, this data is used to train the model.

Once it is trained we can test our model by testing it with some test new mails and checking of the model is able to predict the right output.

Types of Supervised learning

• Classification: A classification problem is when the output variable is a category, such as “red” or “blue” or “disease” and “no disease”.

• Regression: A regression problem is when the output variable is a real value, such as “dollars” or “weight”.

Overview of Unsupervised Learning Algorithm

In unsupervised learning, an AI system is presented with unlabeled, uncategorized data and the system’s algorithms act on the data without prior training. The output is dependent upon the coded algorithms. Subjecting a system to unsupervised learning is one way of testing AI.

In the above example, we have given some characters to our model which are ‘Ducks’ and ‘Not Ducks’. In our training data, we don’t provide any label to the corresponding data. The unsupervised model is able to separate both the characters by looking at the type of data and models the underlying structure or distribution in the data in order to learn more about it.

Types of Unsupervised learning

• Clustering: A clustering problem is where you want to discover the inherent groupings in the data, such as grouping customers by purchasing behavior.

• Association: An association rule learning problem is where you want to discover rules that describe large portions of your data, such as people that buy X also tend to buy Y.

Overview of Reinforcement Learning

A reinforcement learning algorithm, or agent, learns by interacting with its environment. The agent receives rewards by performing correctly and penalties for performing incorrectly. The agent learns without intervention from a human by maximizing its reward and minimizing its penalty. It is a type of dynamic programming that trains algorithms using a system of reward and punishment.

There is a critic : Who will observer the agent form outside & give him indication whether the agent is doing wright or wrong.

In the above example, we can see that the agent is given 2 options i.e. a path with water or a path with fire. A reinforcement algorithm works on reward a system i.e. if the agent uses the fire path then the rewards are subtracted and agent tries to learn that it should avoid the fire path. If it had chosen the water path or the safe path then some points would have been added to the reward points, the agent then would try to learn what path is safe and what path isn’t.

It is basically leveraging the rewards obtained, the agent improves its environment knowledge to select the next action.

3-Discuss the various trending applications that are immensely influenced by Machine Learning.

**Applications of Supervised Learning**

Supervised Learning Algorithms are used in a variety of applications. Let’s go through some of the most well-known applications.

* **BioInformatics** – This is one of the most well-known applications of Supervised Learning because most of us use it in our day-to-day lives. BioInformatics is the storage of Biological Information of us humans such as fingerprints, iris texture, earlobe and so on. Cellphones of today are capable of learning our biological information and are then able to authenticate us bringing up the security of the system. Smartphones such as iPhones, Google Pixel are capable of facial recognition while OnePlus, Samsung is capable of In-display finger recognition.
* [**Speech Recognition**](https://www.edureka.co/blog/speech-recognition-python/) – This is the kind of application where you teach the algorithm about your voice and it will be able to recognize you. The most well-known real-world applications are virtual assistants such as Google Assistant and Siri, which will wake up to the keyword with your voice only.
* **Spam Detection** – This application is used where the unreal or computer-based messages and E-Mails are to be blocked. G-Mail has an algorithm that learns the different keywords which could be fake such as “You are the winner of something” and so forth and blocks those messages directly. OnePlus Messages App gives the user the task of making the application learn which keywords need to be blocked and the app will block those messages with the keyword.
* [**Object-Recognition**](https://www.edureka.co/blog/tensorflow-object-detection-tutorial/)**for Vision** – This kind of application is used when you need to identify something. You have a huge dataset which you use to teach your algorithm and this can be used to recognize a new instance. [Raspberry Pi](https://www.edureka.co/blog/raspberry-pi-tutorial/) algorithms which detect objects are the most well-known example.

## Applications of unsupervised machine learning

Some applications of unsupervised machine learning techniques are:

* Clustering automatically split the dataset into groups base on their similarities
* Anomaly detection can discover unusual data points in your dataset. It is useful for finding fraudulent transactions
* Association mining identifies sets of items which often occur together in your dataset
* Latent variable models are widely used for data preprocessing. Like reducing the number of features in a dataset or decomposing the dataset into multiple components

4)-Discuss different types of machine learning with examples.

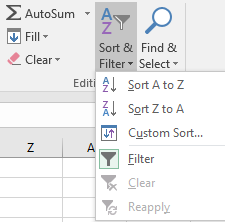
Classical machine learning is often categorized by how an algorithm learns to become more accurate in its predictions. There are four basic approaches:[supervised](https://www.techtarget.com/searchenterpriseai/definition/supervised-learning) learning, [unsupervised](https://whatis.techtarget.com/definition/unsupervised-learning) learning, semi-supervised learning and reinforcement learning. The type of algorithm data scientists choose to use depends on what type of data they want to predict.

* **Supervised learning:** In this type of machine learning, [data scientists](https://www.techtarget.com/searchenterpriseai/definition/data-scientist) supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is specified.
* **Unsupervised learning:** This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.
* **Semi-supervised learning:** This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled [training data](https://www.techtarget.com/searchenterpriseai/feature/Using-small-data-sets-for-machine-learning-models-sees-growth), but the model is free to explore the data on its own and develop its own understanding of the data set.
* **Reinforcement learning:**Data scientists typically use [reinforcement learning](https://www.techtarget.com/searchenterpriseai/definition/reinforcement-learning) to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

5. Describe what is the filter? Steps to use filters in excel. What is its purpose?

In addition to sorting, you may find that adding a filter allows you to better analyze your data. When data is filtered, only rows that meet the filter criteria will display and other rows will be hidden. With filtered data, you can then copy, format, print, etc., your data, without having to sort or move it first. To use a filter,

* Go to the Home ribbon, click the arrow below the Sort & Filtering icon in the Editing group and choose Filter.



OR

* Go to the Data ribbon, and then click Filter in the Sort & Filter group.



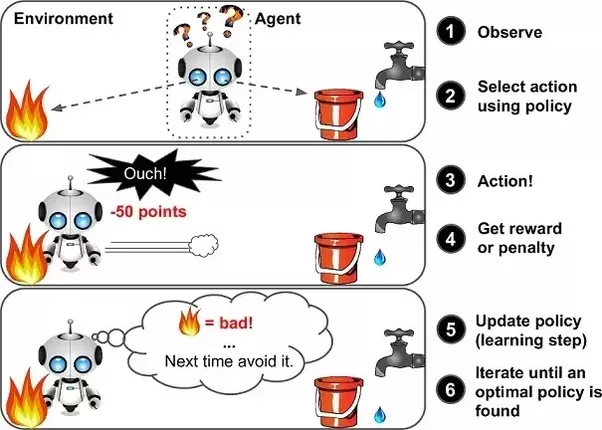
You will notice that all of your column headings now have an arrow next to the heading name. Click on the arrow next to the heading with which you want to filter, and you will see a list of all the unique values in that column. Check the box next to the criteria you wish to match and click OK. Click on the arrow next to another heading to further filter the data.  
  
To clear the filter, choose one of these options:

* Click on the Filter icon next to the heading and choose Clear Filter from “Name of Heading”.
* Go to the Data ribbon and click the Clear icon in the Sort & Filter group.
* Go to the Home ribbon, click the arrow below the Sort & Filter icon in the Editing group and choose Clear.

1. What is reinforcement machine learning? Explain with the help of an example. List 2 area of applications.

A reinforcement learning algorithm, or agent, learns by interacting with its **environment**. The **agent** receives rewards by performing correctly and penalties for performing incorrectly. The agent learns without intervention from a human by maximizing **its reward and minimizing its penalty**. It is a type of dynamic programming that trains algorithms using a system of reward and punishment.

**There is a critic :** Who will observer the agent form outside & give him indication whether the agent is doing wright or wrong.



In the above example, we can see that the agent is given 2 options i.e. a path with water or a path with fire. A reinforcement algorithm works on reward a system i.e. if the agent uses the fire path then the rewards are subtracted and agent tries to learn that it should avoid the fire path. If it had chosen the water path or the safe path then some points would have been added to the reward points, the agent then would try to learn what path is safe and what path isn’t.

It is basically leveraging the rewards obtained, the agent improves its environment knowledge to select the next action.

With **reinforcement learning**, the computer observes its environment and uses that data to identify the ideal behaviour that will minimise risk and/or maximise reward. This is an iterative approach that requires some kind of reinforcement signal to help the computer better identify its best action

1. What are soring and filters? How sorting and filter are useful for AI?

Sorting and filtering offer two different ways to customize the view of data in your sheet. You can't do both at the same time so it's best to determine in advance which option will help more.

Sorting lets you organize all or part of your data in ascending or descending order. Note that you cannot undo a sort after it has been saved so you'll want to make sure that all of your rows in your sheet, including parent rows in a hierarchy, are ordered the way you want before saving.

Filters allow you to show or hide information on your sheet based on selected criteria. They're useful because they don't change the overall layout of your sheet. You can also save filters and share them with anyone who is shared to the sheet. You can even set default filters on your sheet so that when shared users open that sheet, they see the the same view.

A sorting algorithm will put items in a list into an order, such as alphabetical or numerical order. For example, a list of customer names could be sorted into alphabetical order by surname, or a list of people could be put into numerical order by age.

Sorting a list of items can take a long time, especially if it is a large list. A computer program can be created to do this, making sorting a list of data much easier.

There are many types of sorting algorithms. Two of them are **bubble sort** and **bucket sort**.