BUSINESS ANALYTICS
LECTURE NOTES
UNIT 1

INTRODUCTION TO BUSINESS ANALYTICS

Analytics and data science- Analytics life cycle-Types of Analytics-Business Problem definition- Data collection- Data preparation-Hypothesis generation-Modeling-Validation and Evaluation-Interpretation-Deployment and iteration.

Introduction:

Every organization across the world uses performance measures such as market share, profitability, sales growth, return on investments (ROI), customer satisfaction, and so on for quantifying, monitoring, and improving its performance.

Organisation should understand the KPI's (Key performance Indicators) and the factors that have impact on KPI's.

1. Analytics:

Analytics is a body of knowledge consisting of statistical, mathematical and operations research techniques, and Artificial intelligence techniques such as machine learning and deep learning algorithms, data collection and storage, data management processes such as data extraction, transformation and loading (ETL).

Many companies use analytics as a competitive strategy. A typical data-driven decision making process uses following steps:

1. Identify the problem or opportunity for value creation.
2. Identify the sources of data (primary & secondary)
3. Pre-process the data for issues such as missing and incorrect data.
4. Divide the data sets into subsets training and validation.
5. Build analytical models and identify the best model using model performance in validation data.
6. Implement solution/Decision/Develop product.

1.1 Data Science:

Data Science is nothing short of magic, and a Data Scientist is a magician who performs tricks with the data in his hat. Now, as magic is composed of different elements, similarly, Data Science is an interdisciplinary field. We can consider it to be an amalgamation of different fields such as **data manipulation, data visualization, statistical analysis**, and **Machine Learning**. Each of these sub-domains has equal importance.

Data Manipulation:

With the help of data manipulation techniques, you can find interesting insights from the raw data with minimal effort. Data manipulation is the process of organizing information to make it readable and understandable. Engineers perform data manipulation using data manipulation language (DML) capable of adding, deleting, or altering data. Data comes from various sources.

While working with disparate data, you need to organize, clean, and transform it to use it in your decision-making process. This is where data manipulation fits in. Data manipulation allows you to manage and integrate data helping drive actionable insights.

Data manipulation, also known as data preparation, enables users to turn static data into fuel for business intelligence and analytics. Many data scientists use data preparation software to organize data and generate reports, so non-analysts and other stakeholders can derive valuable information and make informed decisions.

Importance of Data manipulation

Data manipulation makes it easier for organizations to organize and analyse data as needed. It helps them perform vital business functions such as analyzing trends, buyer behaviour, and drawing insights from their financial data.

Data manipulation offers several advantages to businesses, including:

- **Consistency:** Data manipulation maintains consistency across data accumulated from different sources, giving businesses a unified view that helps them make better, more informed decisions.
- **Usability:** Data manipulation allows users to cleanse and organize data and use it more efficiently.
- **Forecasting:** Data manipulation enables businesses to understand historical data and helps them prepare future forecasts, especially in financial data analysis.
- **Cleansing:** Data manipulation helps clear unwanted data and keep information that matters. Enterprises can clean up records, isolate, and even reduce unnecessary variables, and focus on the data they need.

Data visualization:

It is the practice of converting raw information (text, numbers, or symbols) into a graphic format. The data is visualized with a clear purpose: to show logical correlations between units, and define inclinations, tendencies, and patterns. Depending on the type of logical connection and the data itself, visualization can be done in a suitable format. So, it's dead simple, any analytical report contains examples of data interpretations like pie charts, comparison bars, demographic maps, and much more.

As we've mentioned, a data representation tool is just the user interface of the whole business intelligence system. Before it can be used for creating visuals, the data goes through a long process. This is basically a description of how Business Analytics works, so we'll break it down into the stages shortly:

1. First things first, you should define data sources and data types that will be used. Then transformation methods and database qualities are determined.
2. Following that, the data is sourced from its initial storages, for example, Google Analytics, ERP, CRM, or SCM system.
3. Using API channels, the data is moved to a staging area where it is transformed. Transformation assumes data cleaning, mapping, and standardizing to a unified format.
4. Further, cleaned data can be moved into a storage: a usual database or data warehouse. To make it possible for the tools to read data, the original base language of datasets can also be rewritten.
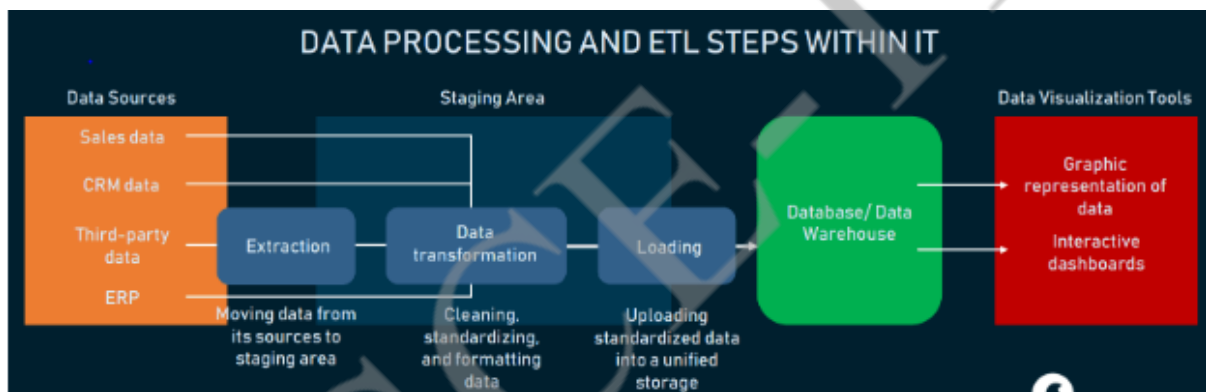
Business Intelligence Data processing in a nutshell

Common types of data visualizations

Each type of visual corresponds precisely to the idea of what data it can interpret, and what type of connection (relationship, comparison, composition, or distribution) it shows better. Let's look at the most common types of visualizations you encounter in Business Analytics in general.
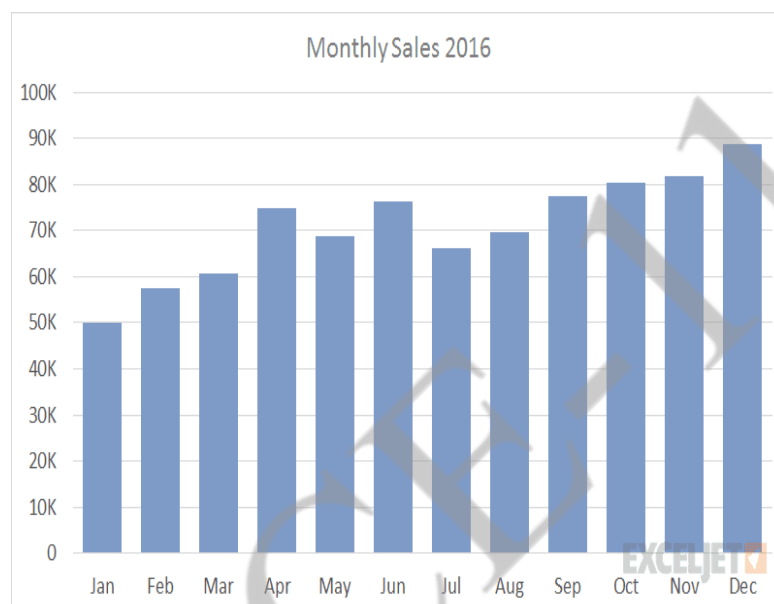
Bar chart

A bar chart is one of the basic ways to compare data units to each other. Because of its simple graphic form, a bar chart is often used in Business Analytics as an interactive page element.

Bar charts are versatile enough to be modified and show more complex data models. The bars can be structured in clusters or be stacked, to depict distribution across market segments, or

subcategories of items. The same goes for horizontal bar charts, fitting more for long data labels to be placed on the bars.

When to use: comparing objects, numeric information. Use horizontal charts to fit long data labels. Place stacks in bars to break each object into segments for a more detailed comparison.
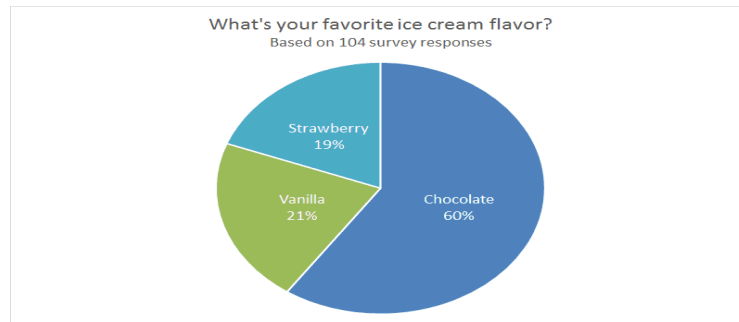

Monthly sales bar chart

Pie chart

One more common type of chart we see everywhere, is a pie chart.

This type of chart is used in any marketing or sales department, because it makes it easy to demonstrate the composition of objects or unit-to-unit comparison.

When to use: composition of an object, comparing parts to the whole object.

Pie chart showing percentage correlation of ice cream flavour preference

Line Graph

This type of visual utilizes a horizontal axis and a vertical axis to depict the value of a unit over time.

Line graphs can also be combined with bar charts to represent data from multiple dimensions.

When to use: object value on the timeline, depicting tendencies in behavior over time.
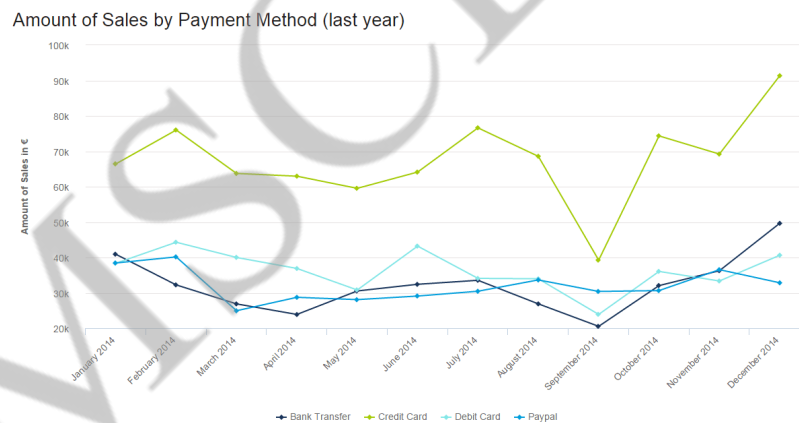

Sales analysis by payment methods

Box plot

At first glance, a box plot looks pretty complicated. But if we look closer at the example, it becomes evident that it depicts quarters in a horizontal fashion.

Our main elements here are minimum, maximum, and the median placed in between the first and third quartile. What a box shows is the distribution of objects, and their

deviation from the median.**When to use**: Distribution of the complex object, deviation from the median value.
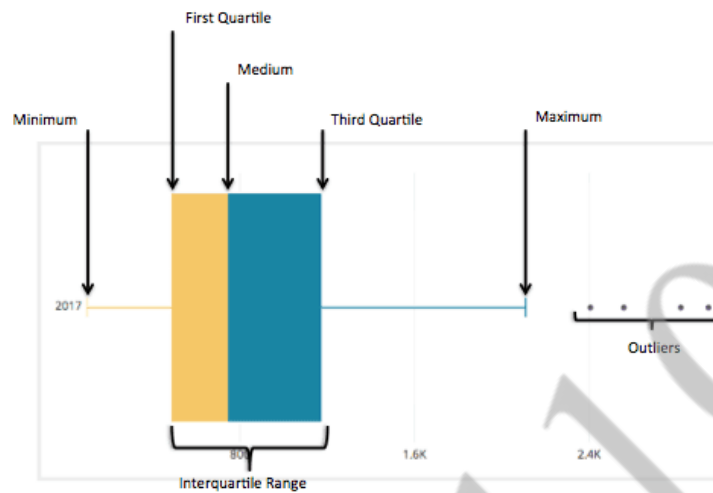
Box plot divided into 5 quartiles, while outliers are shown as object that fall out of distribution area

Scatter plot

This type of visualization is built on X and Y axes. Between them, there are dots placed around, defining objects. The position of a dot on the graph denotes which qualities it has.

As in the case of line graphs, dots placed between the axes are noticed in a split second. The only limitation of this type of visualization is the number of axes.

When to use: showing distribution of objects, defining the quality of each object on the
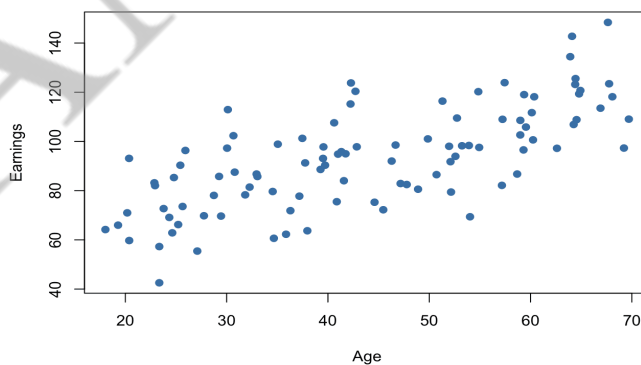
graph.

A sad scatterplot showing the inability of young people to earn money

Radar or spider chart

This type of chart is basically a line chart drawn in radial fashion. It has a spider web form that is created by multiple axes and variables.

Its purpose is the same as for a line chart. But because of the number of axes, you can compare units from various angles and show the inclinations graphically.

When to use: describing data qualities, comparing multiple objects to each other through different dimensions.
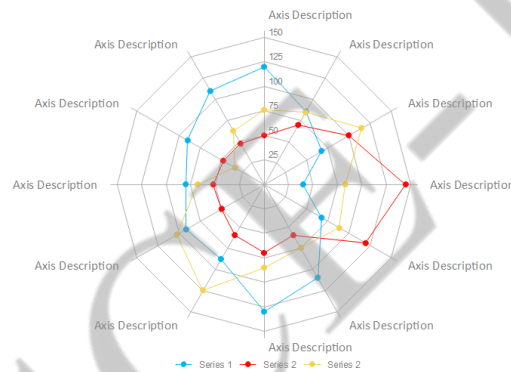
Spider chart structure

Dot map or density map

Superimposing a visualization over the map works for data's geographical domain. Density maps are built with the help of dots placed on the map, marking the location of each unit.

A simple representation of a dot map

Funnel charts

These are perfect for showing narrowing correlations between different groups of items. In most cases, funnels will utilize both geometric form and colour coding to differentiate items.

Website Conversions
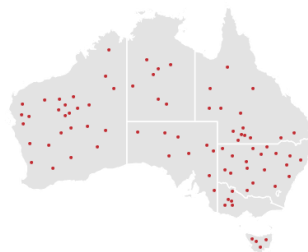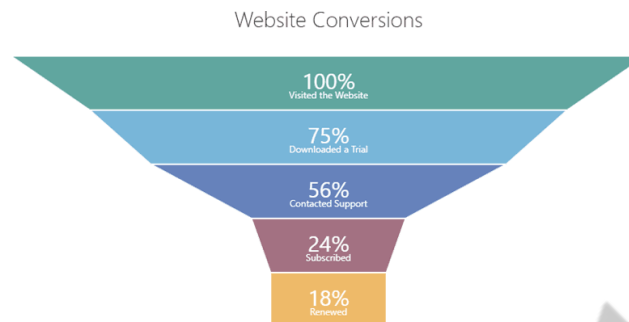
100%
Visited the Website

75%
Downloaded a Trial

56%
Contacted Support

24%
Subscribed

18%
Renewed

The example shows conversion results starting from total traffic number and the number of subscribers

This type of chart is also handy when there are multiple stages in the process. On the example above, we can see that after the "Contacted Support" stage, the number of subscribers has been reduced.

When to use: depicting processual stages with the narrowing percentage of value/objects

In choosing the type of visualization, make sure you clearly understand the following points:

1. **Specifics of your data set**: domain of knowledge or department in your company
2. **Audience**: people you want to present the information to
3. **Connection logic**: comparison of objects, distribution, relationship, process description, etc.
4. **Output**: simply, the reason for showing this information to somebody

1.1.3 What is statistical analysis?

Statistical analysis is the process of collecting and analyzing samples of data to uncover patterns and trends and predict what could happen next to make better and more scientific decisions.
Once the data is collected, statistical analysis can be used for many things in your business. Some include:

- Summarizing and presenting the data in a graph or chart to present key findings
- Discovering crucial measures within the data, like the mean
- Calculating if the data is slightly clustered or spread out, which also determines similarities.

- Making future predictions based on past behavior
- Testing a hypothesis from an experiment

There are several ways that businesses can use statistical analysis to their advantage. Some of these ways include identifying who on your sales staff is performing poorly, finding trends in customer data, narrowing down the top operating product lines, conducting financial audits, and getting a better understanding of how sales performance can vary in different regions of the country.

Just like any other thing in business, there is a process involved in business analytics as well. Business analytics needs to be systematic, organized, and include step-by-step actions to have the most optimized result at the end with the least amount of discrepancies.

Now, let us dive into the steps involved in business analytics:

- **Business Problem Framing:** In this step, we basically find out what business problem we are trying to solve, e.g., when we are looking to find out why the supply chain isn't as effective as it should be or why we are losing sales. This discussion generally happens with stakeholders when they realize inefficiency in any part of the business.
- **Analytics Problem Framing:** Once we have the problem statement, what we need to think of next is how analytics can be done for that business analytics problem. Here, we look for metrics and specific points that we need to analyze.
- **Data:** The moment we identify the problem in terms of what needs to be analyzed, the next thing that we need is data, which needs to be analyzed. In this step, not only do we obtain data from various data sources but we also clean the data; if the raw data is corrupted or has false values, we remove those problems and convert the data into usable form.
- **Methodology selection and model building:** Once the data gets ready, the tricky part begins. At this stage, we need to determine what methods have to be used and what metrics are the crucial ones. If required, the team has to build custom models to find out the specific methods that are suited to respective operations. Many times, the kind of data we possess also dictates the methodology that can be used to do business analytics. Most organizations make multiple models and compare them based on the decided-upon crucial metrics.
- **Deployment:** Post the selection of the model and the statistical ways of analyzing data for the solution, the next thing we need to do is to test the solution in a real-time scenario. For that, we deploy the models on the data and look for different kinds of insights. Based on the metrics and data highlights, we need to decide the optimum strategy to solve our problem and implement a solution effectively. Even in this phase of business analytics, we will compare the expected output with the real-time output. Later, based on this, we will decide if there is a need to reiterate and modify the solution or if we can go on with the implementation of the same.

2. Business Analytics Process

The Business Analytics process involves asking questions, looking at data, and manipulating it to find the required answers. Now, every organization has different ways to execute this process as all of these organizations work in different sectors and value different metrics more than the others based on their specific business model.

Since the approach to business is different for different organizations, their solutions and their ways to reach the solutions are also different. Nonetheless, all of the actions that they do can be classified and generalized to understand their approach. The image given below demonstrates the steps in Business Analytics process of a firm:
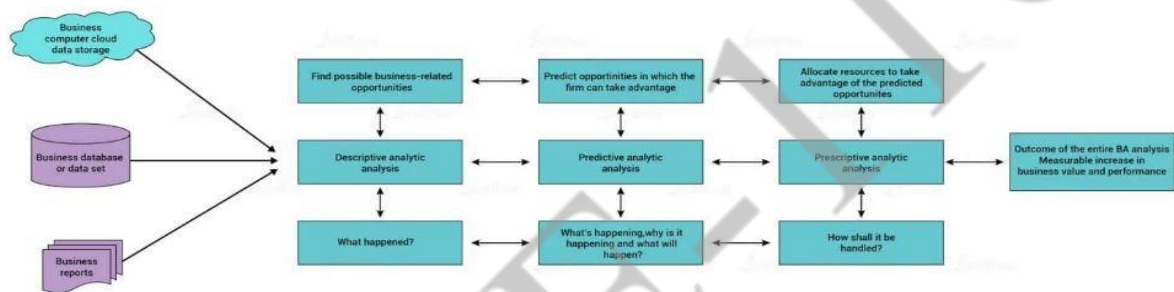
2.1 Six Steps in the Business Analytics Lifecycle

Step 1: Identifying the Problem

The first step of the process is identifying the business problem. The problem could be an actual crisis; it could be something related to recognizing business needs or optimizing current processes. This is a crucial stage in Business Analytics as it is important to clearly understand what the expected outcome should be. When the desired outcome is determined, it is further broken down into smaller goals. Then, business stakeholders decide the relevant data required to solve the problem. Some important questions must be answered in this stage, such as: What kind of data is available? Is there sufficient data? And so on.

Step 2: Exploring Data

Once the problem statement is defined, the next step is to gather data (if required) and, more importantly, cleanse the data—most organizations would have plenty of data, but not all data points would be accurate or useful. Organizations collect huge amounts of data through different methods, but at times, junk data or empty data points would be present in the dataset. These faulty pieces of data can hamper the analysis. Hence, it is very important to clean the data that has to be analyzed.

To do this, you must do computations for the missing data, remove outliers, and find new variables as a combination of other variables. You may also need to plot time series graphs as they generally indicate patterns and outliers. It is very important to remove outliers as they can have a heavy impact on the accuracy of the model that you create. Moreover, cleaning the data helps you get a better sense of the dataset.

Step 3: Analysis

Once the data is ready, the next thing to do is analyze it. Now to execute the same, there are various kinds of statistical methods (such as hypothesis testing, correlation, etc.) involved to find out the insights that you are looking for. You can use all of the methods for which you have the data.

The prime way of analyzing is pivoting around the target variable, so you need to take into account whatever factors that affect the target variable. In addition to that, a lot of assumptions are also considered to find out what the outcomes can be. Generally, at this step, the data is sliced, and the comparisons are made. Through these methods, you are looking to get actionable insights.

Step 4: Prediction and Optimization

Gone are the days when analytics was used to react. In today's era, Business Analytics is all about being proactive. In this step, you will use prediction techniques, such as neural networks or decision trees, to model the data. These prediction techniques will help you find out hidden insights and relationships between variables, which will further help you uncover patterns on the most important metrics. By principle, a lot of models are used simultaneously, and the models with the most accuracy are chosen. In this stage, a lot of conditions are also checked as parameters, and answers to a lot of 'what if...?' questions are provided.

Step 5: Making a Decision and Evaluating the Outcome

From the insights that you receive from your model built on target variables, a viable plan of action will be established in this step to meet the organization's goals and expectations. The said plan of action is then put to work, and the waiting period begins. You will have to wait to see the actual outcomes of your predictions and find out how successful you were in your endeavors. Once you get the outcomes, you will have to measure and evaluate them.

Step 6: Optimizing and Updating

Post the implementation of the solution, the outcomes are measured as mentioned above. If you find some methods through which the plan of action can be optimized, then those can be implemented. If that is not the case, then you can move on with registering the outcomes of the entire process. This step is crucial for any analytics in the future because you will have an ever-improving database. Through this database, you can get closer and closer to maximum

optimization. In this step, it is also important to evaluate the ROI (return on investment). Take a look at the diagram above of the life cycle of business analytics.
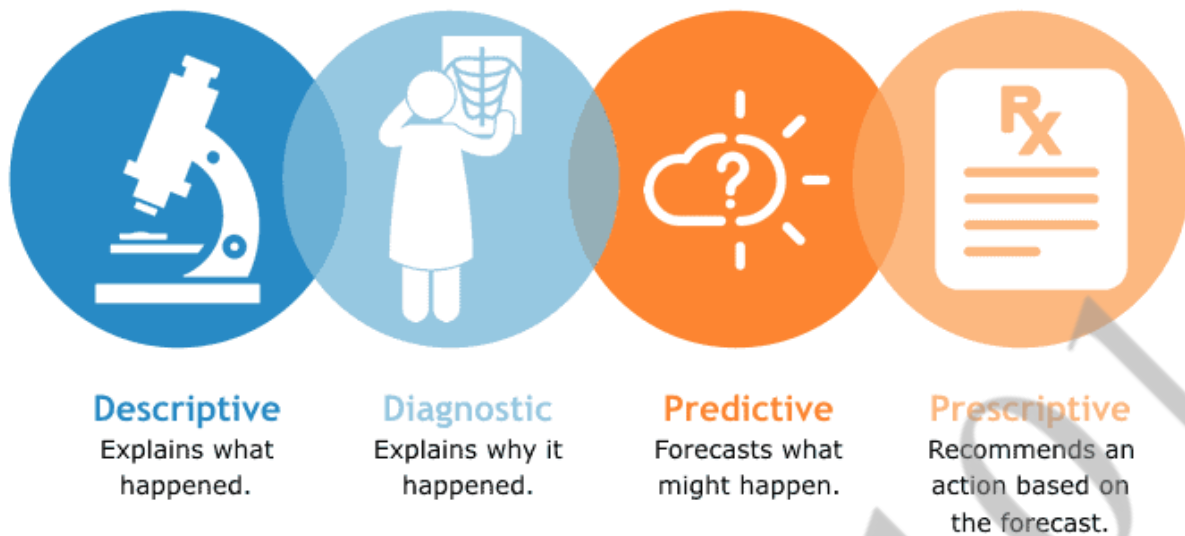
2.3 TYPES OF ANALYTICS :

For different stages of business analytics huge amount of data is processed at various steps. Depending on the stage of the workflow and the requirement of data analysis, there are four main kinds of analytics – descriptive, diagnostic, predictive and prescriptive. These four types together answer everything a company needs to know- from what's going on in the company to what solutions to be adopted for optimising the functions.

The four types of analytics are usually implemented in stages and no one type of analytics is said to be better than the other. They are interrelated and each of these offers a different insight. With data being important to so many diverse sectors- from manufacturing to energy grids, most of the companies rely on one or all of these types of analytics. With the right choice of analytical techniques, big data can deliver richer insights for the companies

Before diving deeper into each of these, let's define the four types of analytics:

1) **Descriptive Analytics**: Describing or summarising the existing data using existing business intelligence tools to better understand what is going on or what has happened.
2) **Diagnostic Analytics**: Focus on past performance to determine what happened and why. The result of the analysis is often an analytic dashboard.
3) **Predictive Analytics**: Emphasizes on predicting the possible outcome using statistical models and machine learning techniques.
4) **Prescriptive Analytics**: It is a type of predictive analytics that is used to recommend one or more course of action on analyzing the data.
Let's understand these in a bit more depth.

Descriptive — Explains what happened.
Diagnostic — Explains why it happened.
Predictive — Forecasts what might happen.
Prescriptive — Recommends an action based on the forecast.

2.3.1. Descriptive Analytics

This can be termed as the simplest form of analytics. The mighty size of big data is beyond human comprehension and the first stage hence involves crunching the data into understandable chunks. The purpose of this analytics type is just to summarise the findings and understand what is going on.

Among some frequently used terms, what people call as advanced analytics or business intelligence is basically usage of descriptive statistics (arithmetic operations, mean, median, max, percentage, etc.) on existing data. It is said that 80% of business analytics mainly involves descriptions based on aggregations of past performance. It is an important step to make raw data understandable to investors, shareholders and managers. This way it gets easy to identify and address the areas of strengths and weaknesses such that it can help in strategizing.The two main techniques involved are data aggregation and data mining stating that this method is purely used for understanding the underlying behavior and not to make any estimations. By mining historical data, companies can analyze the consumer behaviors and engagements with their businesses that could be helpful in targeted marketing, service improvement, etc. The tools used in this phase are MS Excel, MATLAB, SPSS, STATA, etc

2.3.2 Diagnostic Analytics

Diagnostic analytics is used to determine why something happened in the past. It is characterized by techniques such as drill-down, data discovery, data mining and correlations. Diagnostic analytics takes a deeper look at data to understand the root causes of the events. It is helpful in determining what factors and events contributed to the outcome. It mostly uses probabilities, likelihoods, and the distribution of outcomes for the analysis.

In a time series data of sales, diagnostic analytics would help you understand why the sales have decrease or increase for a specific year or so. However, this type of analytics has a limited ability to give actionable insights. It just provides an understanding of causal relationships and sequences while looking backward.

A few techniques that uses diagnostic analytics include attribute importance, principle components analysis, sensitivity analysis, and conjoint analysis. Training algorithms for classification and regression also fall in this type of analytics.

2.3.3 Predictive Analytics

As mentioned above, predictive analytics is used to predict future outcomes. However, it is important to note that it cannot predict if an event will occur in the future; it merely forecasts what are the probabilities of the occurrence of the event. A predictive model builds on the preliminary descriptive analytics stage to derive the possibility of the outcomes.

The analytics is found in sentiment analysis where all the opinions posted on social media are collected and analyzed (existing text data) to predict the person's sentiment on a particular subject as being- positive, negative or neutral (future prediction).

Hence, predictive analytics includes building and validation of models that provide accurate predictions. Predictive analytics relies on machine learning algorithms like random forests, SVM, etc. and statistics for learning and testing the data. Usually, companies need trained data scientists and machine learning experts for building these models. The most popular tools for predictive analytics include Python, R, RapidMiner, etc.

The prediction of future data relies on the existing data as it cannot be obtained otherwise. If the model is properly tuned, it can be used to support complex forecasts in sales and marketing. It goes a step ahead of the standard BI in giving accurate predictions.

2.3.4 Prescriptive Analytics

The basis of this analytics is predictive analytics but it goes beyond the three mentioned above to suggest the future solutions. It can suggest all favorable outcomes according to a specified course of action and also suggest various course of actions to get to a particular outcome. Hence, it uses a strong feedback system that constantly learns and updates the relationship between the action and the outcome.

The computations include optimisation of some functions that are related to the desired outcome. For example, while calling for a cab online, the application uses GPS to connect you to the correct driver from among a number of drivers found nearby. Hence, it optimises the distance for faster arrival time. Recommendation engines also use prescriptive analytics.

The other approach includes simulation where all the key performance areas are combined to design the correct solutions. It makes sure whether the key performance metrics are included in the solution. The optimisation model will further work on the impact of the previously made forecasts. Because of its power to suggest favorable solutions, prescriptive analytics is the final frontier of advanced analytics or data science, in today's term.

The four techniques in analytics may make it seem as if they need to be implemented sequentially. However, in most scenarios, companies can jump directly to prescriptive analytics. As for most of the companies, they are aware of or are already implementing descriptive analytics but if one has identified the key area that needs to be optimised and worked upon, they must employ prescriptive analytics to reach the desired outcome.

According to research, prescriptive analytics is still at the budding stage and not many firms have completely used its power. However, the advancements in predictive analytics will surely pave the way for its development.

3. Business Problem Definition:

Problem-solving in business is defined as implementing processes that reduce or remove obstacles that are preventing you or others from accomplishing operational and strategic business goals.

In business, a problem is a situation that creates a gap between the desired and actual outcomes. In addition, a true problem typically does not have an immediately obvious resolution.

Business problem-solving works best when it is approached through a consistent system in which individuals:

- Identify and define the problem
- Prioritize the problem based on size, potential impact, and urgency
- Complete a root-cause analysis
- Develop a variety of possible solutions
- Evaluate possible solutions and decide which is most effective
- Plan and implement the solution

3.1 Why Problem Solving Is Important in Business

Understanding the importance of problem-solving skills in the workplace will help you develop as a leader. Problem-solving skills will help you resolve critical issues and conflicts that you come across. Problem-solving is a valued skill in the workplace because it allows you to:

- Apply a standard problem-solving system to all challenges
- Find the root causes of problems
- Quickly deal with short-term business interruptions
- Form plans to deal with long-term problems and improve the organization
- See challenges as opportunities
- Keep your cool during challenges

3.2 How to Solve Business Problems Effectively

There are many different problem-solving skills, but most can be broken into general steps. Here is a four-step method for business problem solving:

1) Identify the Details of the Problem: Gather enough information to accurately define the problem. This can include data on procedures being used, employee actions, relevant workplace rules, and so on. Write down the specific outcome that is needed, but don't assume what the solution should be.

2) Creatively Brainstorm Solutions: Alone or with a team, state every solution you can think of. You'll often need to write them down. To get more solutions, brainstorm with the employees who have the greatest knowledge of the issue.

3) Evaluate Solutions and Make a Decision: Compare and contrast alternative solutions based on the feasibility of each one, including the resources needed to implement it and the return on investment of each one. Finally, make a firm decision on one solution that clearly addresses the root cause of the problem.

4) Take Action: Write up a detailed plan for implementing the solution, get the necessary approvals, and put it into action.

4 . WHAT IS DATA COLLECTION?

Data collection is the methodological process of gathering information about a specific subject. It's crucial to ensure your data is complete during the collection phase and that it's collected legally and ethically. If not, your analysis won't be accurate and could have far-reaching consequences.

In general, there are three types of consumer data:

- First-party data, which is collected directly from users by your organization
- Second-party data, which is data shared by another organization about its customers (or its first-party data)
- Third-party data, which is data that's been aggregated and rented or sold by organizations that don't have a connection to your company or users

Although there are use cases for second- and third-party data, first-party data (data you've collected yourself) is more valuable because you receive information about how your audience behaves, thinks, and feels—all from a trusted source.

Data can be qualitative (meaning contextual in nature) or quantitative (meaning numeric in nature). Many data collection methods apply to either type, but some are better suited to one over the other.

In the data life cycle, data collection is the second step. After data is generated, it must be collected to be of use to your team. After that, it can be processed, stored, managed, analyzed, and visualized to aid in your organization's decision-making.
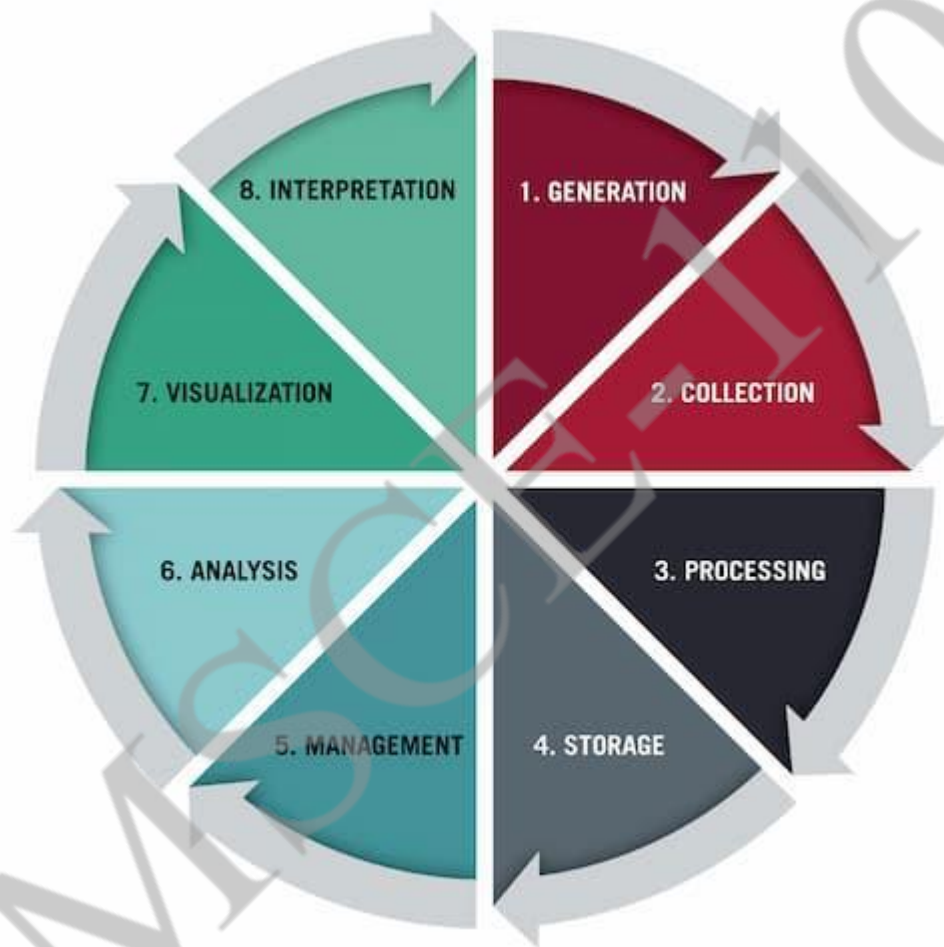
Before collecting data, there are several factors you need to define:

● The question you aim to answer
● The data subject(s) you need to collect data from
● The collection timeframe
● The data collection method(s) best suited to your needs

The data collection method you select should be based on the question you want to answer, the type of data you need, your timeframe, and your company's budget. Explore the options in the next section to see which data collection method is the best fit.

4.1 SEVEN DATA COLLECTION METHODS USED IN BUSINESS ANALYTICS

1. Surveys

Surveys are physical or digital questionnaires that gather both qualitative and quantitative data from subjects. One situation in which you might conduct a survey is gathering attendee feedback after an event. This can provide a sense of what attendees enjoyed, what they wish was different, and areas you can improve or save money on during your next event for a similar audience.

Because they can be sent out physically or digitally, surveys present the opportunity for distribution at scale. They can also be inexpensive; running a survey can cost nothing if you use a free tool. If you wish to target a specific group of people, partnering with a market research firm to get the survey in the hands of that demographic may be worth the money.

Something to watch out for when crafting and running surveys is the effect of bias, including:

- **Collection bias**: It can be easy to accidentally write survey questions with a biased lean. Watch out for this when creating questions to ensure your subjects answer honestly and aren't swayed by your wording.
- **Subject bias**: Because your subjects know their responses will be read by you, their answers may be biased toward what seems socially acceptable. For this reason, consider pairing survey data with behavioral data from other collection methods to get the full picture.

2. Transactional Tracking

Each time your customers make a purchase, tracking that data can allow you to make decisions about targeted marketing efforts and understand your customer base better.

Often, e-commerce and point-of-sale platforms allow you to store data as soon as it's generated, making this a seamless data collection method that can pay off in the form of customer insights.

3. Interviews and Focus Groups

Interviews and focus groups consist of talking to subjects face-to-face about a specific topic or issue. Interviews tend to be one-on-one, and focus groups are typically made up of several people. You can use both to gather qualitative and quantitative data.

Through interviews and focus groups, you can gather feedback from people in your target audience about new product features. Seeing them interact with your product in real-time and

recording their reactions and responses to questions can provide valuable data about which product features to pursue.

As is the case with surveys, these collection methods allow you to ask subjects anything you want about their opinions, motivations, and feelings regarding your product or brand. It also introduces the potential for bias. Aim to craft questions that don't lead them in one particular direction.

One downside of interviewing and conducting focus groups is they can be time-consuming and expensive. If you plan to conduct them yourself, it can be a lengthy process. To avoid this, you can hire a market research facilitator to organize and conduct interviews on your behalf.

4. Observation

Observing people interacting with your website or product can be useful for data collection because of the candour it offers. If your user experience is confusing or difficult, you can witness it in real-time.

Yet, setting up observation sessions can be difficult. You can use a third-party tool to record users' journeys through your site or observe a user's interaction with a beta version of your site or product.

While less accessible than other data collection methods, observations enable you to see first hand how users interact with your product or site. You can leverage the qualitative and quantitative data gleaned from this to make improvements and double down on points of success.

5. Online Tracking

To gather behavioural data, you can implement pixels and cookies. These are both tools that track users' online behaviour across websites and provide insight into what content they're interested in and typically engage with.

You can also track users' behavior on your company's website, including which parts are of the highest interest, whether users are confused when using it, and how long they spend on product pages. This can enable you to improve the website's design and help users navigate to their destination.

Inserting a pixel is often free and relatively easy to set up. Implementing cookies may come with a fee but could be worth it for the quality of data you'll receive. Once pixels and cookies are set, they gather data on their own and don't need much maintenance, if any.

It's important to note: Tracking online behavior can have legal and ethical privacy implications. Before tracking users' online behavior, ensure you're in compliance with local and industry data privacy standards.

6. Forms

Online forms are beneficial for gathering qualitative data about users, specifically demographic data or contact information. They're relatively inexpensive and simple to set up, and you can use them to gate content or registrations, such as webinars and email newsletters.

You can then use this data to contact people who may be interested in your product, build out demographic profiles of existing customers, and in remarketing efforts, such as email workflows and content recommendations

7. Social Media Monitoring

Monitoring your company's social media channels for follower engagement is an accessible way to track data about your audience's interests and motivations. Many social media platforms have analytics built in, but there are also third-party social platforms that give more detailed, organized insights pulled from multiple channels.

You can use data collected from social media to determine which issues are most important to your followers. For instance, you may notice that the number of engagements dramatically increases when your company posts about its sustainability efforts.

5. What Is Data Preparation?

Data preparation, also sometimes called "pre-processing," is the act of cleaning and consolidating raw data prior to using it for business analysis. It might not be the most celebrated of tasks, but careful data preparation is a key component of successful data analysis.

Doing the work to properly validate, clean, and augment raw data is essential to draw accurate, meaningful insights from it. The validity and power of any business analysis produced is only as good as the data preparation done in the early stages.

5.1 Why Is Data Preparation Important?

The decisions that business leaders make are only as good as the data that supports them. Careful and comprehensive data preparation ensures analysts trust, understand, and ask better questions of their data, making their analyses more accurate and meaningful. From more meaningful data analysis comes better insights and, of course, better outcomes.

To drive the deepest level of analysis and insight, successful teams and organizations must implement a data preparation strategy that prioritizes:

● **Accessibility:** Anyone — regardless of skillset — should be able to access data securely from a single source of truth
● **Transparency:** Anyone should be able to see, audit, and refine any step in the end-to-end data preparation process that took place
● **Repeatability:** Data preparation is notorious for being time-consuming and repetitive, which is why successful data preparation strategies invest in solutions built for repeatability.

With the right solution in hand, analysts and teams can streamline the data preparation process, and instead, spend more time getting to valuable business insights and outcomes, faster.

5.2 What Steps Are Involved in Data Preparation Processes?

The data preparation process can vary depending on industry or need, but typically consists of the following steps:

● **Acquiring data:** Determining what data is needed, gathering it, and establishing consistent access to build powerful, trusted analysis
● **Exploring data:** Determining the data's quality, examining its distribution, and analyzing the relationship between each variable to better understand how to compose an analysis
● **Cleansing data:** Improving data quality and overall productivity to craft error-proof insights

- **Transforming data:** Formatting, orienting, aggregating, and enriching the datasets used in an analysis to produce more meaningful insights

While data preparation processes build upon each other in a serialized fashion, it's not always linear. The order of these steps might shift depending on the data and questions being asked. It's common to revisit a previous step as new insights are uncovered or new data sources are integrated into the process.

The entire data preparation process can be notoriously time-intensive, iterative, and repetitive. That's why it's important to ensure the individual steps taken can be easily understood, repeated, revisited, and revised so analysts can spend less time prepping and more time analyzing.

Below is a deeper look at each part of the process.

5.2.1 Acquire Data

The first step in any data preparation process is acquiring the data that an analyst will use for their analysis. It's likely that analysts rely on others (like IT) to obtain data for their analysis, likely from an enterprise software system or data management system. IT will usually deliver this data in an accessible format like an Excel document or CSV.

Modern analytic software can remove the dependency on a data-wrangling middleman to tap right into trusted sources like SQL, Oracle, SPSS, AWS, Snowflake, Salesforce, and Marketo. This means analysts can acquire the critical data for their regularly-scheduled reports as well as novel analytic projects on their own.

5.2.2 Explore Data

Examining and profiling data helps analysts understand how their analysis will begin to take shape. Analysts can utilize visual analytics and summary statistics like range, mean, and standard deviation to get an initial picture of their data. If data is too large to work with easily, segmenting it can help.

During this phase, analysts should also evaluate the quality of their dataset. Is the data complete? Are the patterns what was expected? If not, why? Analysts should discuss what they're seeing with the owners of the data, dig into any surprises or anomalies, and consider if it's even possible to improve the quality. While it can feel disappointing to disqualify a dataset based on poor quality, it is a wise move in the long run. Poor quality is only amplified as one moves through the data analytics processes

5.2.3 Cleanse Data

During the exploration phase, analysts may notice that their data is poorly structured and in need of tidying up to improve its quality. This is where data cleansing comes into play. Cleansing data includes:

- Correcting entry errors
- Removing duplicates or outliers
- Eliminating missing data
- Masking sensitive or confidential information like names or addresses

5.2.4 Transform Data

Data comes in many shapes, sizes, and structures. Some is analysis-ready, while other datasets may look like a foreign language.

Transforming data to ensure that it's in a format or structure that can answer the questions being asked of it is an essential step to creating meaningful outcomes. This will vary based on the software or language that an analysts uses for their data analysis.

A couple of common examples of data transformations are:

- Pivoting or changing the orientation of data
- Converting date formats
- Aggregating sales and performance data across time

6.HYPOTHESIS GENERATION OR TESTING:

Hypothesis testing is the act of testing a hypothesis or a supposition in relation to a statistical parameter. Analysts implement hypothesis testing in order to test if a hypothesis is plausible or not.In data science and statistics, hypothesis testing is an important step as it involves the verification of an assumption that could help develop a statistical parameter. For instance, a researcher establishes a hypothesis assuming that the average of all odd numbers is an even number.

In order to find the plausibility of this hypothesis, the researcher will have to test the hypothesis using hypothesis testing methods. Unlike a hypothesis that is 'supposed' to stand true on the basis of little or no evidence, hypothesis testing is required to have plausible evidence in order to establish that a statistical hypothesis is true.

6.1 Types of Hypotheses
In data sampling, different types of hypothesis are involved in finding whether the tested samples test positive for a hypothesis or not. In this segment, we shall discover the different types of hypotheses and understand the role they play in hypothesis testing.
6.1.1 Alternative Hypothesis

Alternative Hypothesis (H1) or the research hypothesis states that there is a relationship between two variables (where one variable affects the other). The alternative hypothesis is the main driving force for hypothesis testing.

It implies that the two variables are related to each other and the relationship that exists between them is not due to chance or coincidence.

When the process of hypothesis testing is carried out, the alternative hypothesis is the main subject of the testing process. The analyst intends to test the alternative hypothesis and verifies its plausibility.

6.1.2 Null Hypothesis

The Null Hypothesis (H0) aims to nullify the alternative hypothesis by implying that there exists no relation between two variables in statistics. It states that the effect of one variable on the other is solely due to chance and no empirical cause lies behind it.

The null hypothesis is established alongside the alternative hypothesis and is recognized as important as the latter. In hypothesis testing, the null hypothesis has a major role to play as it influences the testing against the alternative hypothesis.

6.1.3 Non-Directional Hypothesis

The Non-directional hypothesis states that the relation between two variables has no direction. Simply put, it asserts that there exists a relation between two variables, but does not recognize the direction of effect, whether variable A affects variable B or vice versa.

6.1.4 Directional Hypothesis

The Directional hypothesis, on the other hand, asserts the direction of effect of the relationship that exists between two variables.

Herein, the hypothesis clearly states that variable A affects variable B, or vice versa.

6.1.5 Statistical Hypothesis

A statistical hypothesis is a hypothesis that can be verified to be plausible on the basis of statistics. By using data sampling and statistical knowledge, one can determine the plausibility of a statistical hypothesis and find out if it stands true or not.

6.2 Performing Hypothesis Testing

Now that we have understood the types of hypotheses and the role they play in hypothesis testing, let us now move on to understand the process in a better manner.

In hypothesis testing, a researcher is first required to establish two hypotheses - alternative hypothesis and null hypothesis in order to begin with the procedure.

To establish these two hypotheses, one is required to study data samples, find a plausible pattern among the samples, and pen down a statistical hypothesis that they wish to test.

A random population of samples can be drawn, to begin with hypothesis testing. Among the two hypotheses, alternative and null, only one can be verified to be true. Perhaps the presence of both hypotheses is required to make the process successful.

At the end of the hypothesis testing procedure, either of the hypotheses will be rejected and the other one will be supported. Even though one of the two hypotheses turns out to be true, no hypothesis can ever be verified 100%.

6.2.1 Seven steps of hypothesis testing

Let us perform hypothesis testing through the following 7 steps of the procedure:

Step 1 : Specify the null hypothesis and the alternative hypothesis

Step 2 : What level of significance?

Step 3 : Which test and test statistic to be performed?

Step 4 : State the decision rule

Step 5 : Use the sample data to calculate the test statistic

Step 6 : Use the test statistic result to make a decision

Step 7 : Interpret the decision in the context of the original question

To guide us through the steps, let us use the following example.

Assume a food laboratory analyzed a certified reference freeze-dried food material with a stated sodium (Na) content of 250 mg/kg. It carried out 7 repeated analyses and obtained a mean value of 274 mg/kg of sodium with a sample standard deviation of 21 mg/kg. Now we want to know if the mean value of 274 mg/kg is significantly larger than the stated amount of 250 mg/kg. If so, we will conclude that the reported results of this batch of analysis were of bias and had consistently given higher values than expected.

Step 1 : Specify the null hypothesis and the alternative hypothesis

The **null hypothesis** *Ho* is the statement that we are interested in testing. In this case, the null condition is that the mean value is 250 mg/kg of sodium.

The **alternative hypothesis** *H1* is the statement that we accept if our sample outcome leads us to reject the null hypothesis. In our case, the alternative hypothesis is that the mean value is not equal to 250 mg/kg of sodium. In other words, it can be significantly larger or smaller than the value of 250 mg/kg.

So, our formal statement of the hypotheses for this example is as follows:

Ho : $\bar{x} = 250$ mg/kg (i.e., the certified value)

H1 : $\bar{x} \neq 250$ mg/kg (i.e., indicating that the laboratory has a bias result.

Step 2 : What level of significance

The level of significance is the probability of rejecting the null hypothesis by chance alone. This could happen from sub-sampling error, methodology, analyst's technical competence, instrument drift, etc. So, we have to decide on the level of significance to reject the null hypothesis if the sample result was unlikely given the null hypothesis was true.

Traditionally, we define the *unlikely* (given by symbol ✔) as 0.05 (5%) or less. However, there is nothing to stop you from using ✔ = 0.1 (10%) or ✔ = 0.01 (1%) with your own justification or reasoning.

In fact, the significance level sometimes is referred to as the probability of a **Type I** error. A Type I error occurs when you falsely reject the null hypothesis on the basis of the above-mentioned errors. A **Type II** error occurs when you fail to reject the null hypothesis when it is false.

Step 3 : Which test and test statistic?

The test statistic is the value calculated from the sample to determine whether to reject the null hypothesis. In this case, we use Student's t-test statistic in the following manner:

$\mu = \bar{x} \pm (\alpha = 0.05, v = n-1)s\sqrt{n}$

or $t(\alpha = 0.05, v = n-1) = |\bar{x} - \mu|\sqrt{n}s$

By calculation, we get a t-value of 3.024 at the significance level of ✔ = 0.05 and $v = (7\text{-}1)$ or 6 degrees of freedom for $n = 7$ replicates.

Step 4 : State the decision rule

The decision rule is always of the following form:

Reject *Ho* if …..

We reject the null hypothesis if the test statistic is larger than a critical value corresponding to the significance level in step 2.

There is now a question in *H1* on either one-tailed (> or <) or two-tailed (≠ not equal) tests to be addressed. If we are talking about either "greater than" or "smaller than", we take the significance level at ✔ = 0.05 whilst for the unequal (that means the result can be either larger or smaller than the certified value), the significance level at ✔ = 0.025 on either side of the normal curve is to be studied.

As our *H1* is for the mean value to be larger or smaller than the certified value, we use the 2-tailed t-test for ✔ = 0.05 with 6 degrees of freedom. In this case, the t-critical value at ✔ = 0.05 and 6 degrees of freedom is 2.447 from the Student's t-table or from using the Excel function "=T.INV.2T(0.05,6)" or "=TINV(0.05,6) in older Excel version.

That means the decision rule would be stated as below:

Reject *Ho* if t > 2.447

Step 5 : Use the sample data to calculate the test statistic

Upon calculation on the sample data, we have got a t-value of 3.024 at the significance level of ✔ = 0.05 and $v = (7-1)$ or 6 degrees of freedom for $n = 7$ replicates.

Step 6 : Use the test statistic to make a decision

When we compare the result of step 5 to the decision rule in step 4, it is obvious that 3.024 is greater than the t-critical value of 2.447, and so we reject the null hypothesis. In other words, the mean value of 274 mg/kg is significantly different from the certified value of 250 mg/kg. Is it really so? We must go to step 7.

Step 7 : Interpret the decision in the context of the original question

Since hypothesis testing involves some kind of probability under the disguise of significance level, we must interpret the final decision with caution. To say that a result is "statistically significant" sounds remarkable, but all it really means is that it is more than by chance alone.

To do justice, it would be useful to look at the actual data to see if there are one or more high outliers pulling up the mean value. Perhaps increasing the number of replicates might show up any undesirable data. Furthermore, we might have to take a closer look at the test procedure and the technical competence of the analyst to see if there were any lapses in the analytical process. A repeated series of experiment should be able to confirm these findings.

7.MODELING

A model is an abstraction or representation of a real system, idea, or object. Models capture the most important features of a problem and present them in a form that is easy to interpret. A model can be as simple as a written or verbal description of some phenomenon, a visual representation such as a graph or a flowchart, or a mathematical or spreadsheet representation.

7.1 Decision Models

A decision model is a logical or mathematical representation of a problem or business situation that can be used to understand, analyze, or facilitate making a decision. Most decision models have three types of input:

1. Data, which are assumed to be constant for purposes of the model. Some examples would be costs, machine capacities, and intercity distances.

2. Uncontrollable variables, which are quantities that can change but cannot be directly controlled by the decision maker. Some examples would be customer demand, inflation rates, and investment returns. Often, these variables are uncertain.

3. Decision variables, which are controllable and can be selected at the discretion of the decision maker. Some examples would be production quantities, staffing levels, and investment allocations. Decision models characterize the relationships among the data, uncontrollable variables, and decision variables, and the outputs of interest to the decision maker.

Decision models can be represented in various ways, most typically with mathematical functions and spreadsheets. Spreadsheets are ideal vehicles for implementing decision models because of their versatility in managing data, evaluating different scenarios, and presenting results in a meaningful fashion. Using these relationships, we may develop a mathematical representation by defining symbols for each of these quantities:

TC = total cost

V = unit variable cost

F = fixed cost

Q = quantity produced

This results in the model TC = F + VQ

7.1.2 Model Assumptions:

All models are based on assumptions that reflect the modeler's view of the "real world." Some assumptions are made to simplify the model and make it more tractable; that is, able to be easily analyzed or solved. Other assumptions might be made to better characterize historical data or past observations. The task of the modeler is to select or build an appropriate model that best represents the behavior of the real situation. For example, economic theory tells us that demand for a product is negatively related to its price. Thus, as prices increase, demand falls, and vice versa (a phenomenon that you may recognize as price elasticity—the ratio of the percentage change in demand to the percentage change in price). Different mathematical models can describe this phenomenon.

7.2 Prescriptive Decision Models

A prescriptive decision model helps decision makers to identify the best solution to a decision problem. **Optimization** is the process of finding a set of values for decision variables that minimize or maximize some quantity of interest—profit, revenue, cost, time, and so on—called the **objective function**. Any set of decision variables that optimizes the objective function is called an **optimal solution**. In a highly competitive world where one percentage point can mean a difference of hundreds of thousands of dollars or more, knowing the best solution can mean the difference between success and failure.

Prescriptive decision models can be either **deterministic or stochastic**. A **deterministic model** is one in which all model input information is either known or assumed to be known with certainty. A **stochastic model** is one in which some of the model input information is uncertain. For instance, suppose that customer demand is an important element of some model. We can make the assumption that the demand is known with certainty; say, 5,000 units per month. In this case we would be dealing with a deterministic model. On the other hand, suppose we have evidence to indicate that demand is uncertain, with an average value of 5,000 units per month, but which typically varies between 3,200 and 6,800 units. If we make this assumption, we would be dealing with a stochastic model.

7.3 Uncertainty and Risks:

As we all know, the future is always uncertain. Thus, many predictive models incorporate uncertainty and help decision makers analyze the risks associated with their decisions. Uncertainty is imperfect knowledge of what will happen; risk is associated with the consequences and likelihood of what might happen.

For example, the change in the stock price of Apple on the next day of trading is uncertain. However, if you own Apple stock, then you face the risk of losing money if the stock price falls. If you don't own any stock, the price is still uncertain although you would not have any risk. Risk is evaluated by the magnitude of the consequences and the likelihood that they would occur. For example, a 10% drop in the stock price would incur a higher risk if you own $1 million than if you only owned $1,000. Similarly, if the chances of a 10% drop were 1 in 5, the risk would be higher than if the chances were only 1 in 100. The importance of risk in business has long been recognized.

8.Model Validation

Model validation is defined within regulatory guidance as "the set of processes and activities intended to verify that models are performing as expected, in line with their design

objectives, and business uses." It also identifies "potential limitations and assumptions, and assesses their possible impact."

Generally, validation activities are performed by individuals independent of model development or use. Models, therefore, should not be validated by their owners as they can be highly technical, and some institutions may find it difficult to assemble a model risk team that has sufficient functional and technical expertise to carry out independent validation. When faced with this obstacle, institutions often outsource the validation task to third parties.

In statistics, **model validation** is the task of confirming that the outputs of a statistical model are acceptable with respect to the real data-generating process. In other words, model validation is the task of confirming that the outputs of a statistical model have enough fidelity to the outputs of the data-generating process that the objectives of the investigation can be achieved.

8.1 The Four Elements

Model validation consists of four crucial elements which should be considered:

1. Conceptual Design

The foundation of any model validation is its conceptual design, which needs documented coverage assessment that supports the model's ability to meet business and regulatory needs and the unique risks facing a bank.

The design and capabilities of a model can have a profound effect on the overall effectiveness of a bank's ability to identify and respond to risks. For example, a poorly designed risk assessment model may result in a bank establishing relationships with clients that present a risk that is greater than its risk appetite, thus exposing the bank to regulatory scrutiny and reputation damage.

A validation should independently challenge the underlying conceptual design and ensure that documentation is appropriate to support the model's logic and the model's ability to achieve desired regulatory and business outcomes for which it is designed.

2. System Validation

All technology and automated systems implemented to support models have limitations. An effective validation includes: firstly, evaluating the processes used to integrate the model's conceptual design and functionality into the organisation's business setting; and, secondly,

examining the processes implemented to execute the model's overall design. Where gaps or limitations are observed, controls should be evaluated to enable the model to function effectively.

3. Data Validation and Quality Assessment

Data errors or irregularities impair results and might lead to an organisation's failure to identify and respond to risks. Best practise indicates that institutions should apply a risk-based data validation, which enables the reviewer to consider risks unique to the organisation and the model.

To establish a robust framework for data validation, guidance indicates that the accuracy of source data be assessed. This is a vital step because data can be derived from a variety of sources, some of which might lack controls on data integrity, so the data might be incomplete or inaccurate.

4. Process Validation

To verify that a model is operating effectively, it is important to prove that the established processes for the model's ongoing administration, including governance policies and procedures, support the model's sustainability. A review of the processes also determines whether the models are producing output that is accurate, managed effectively, and subject to the appropriate controls.

If done effectively, model validation will enable your bank to have every confidence in its various models' accuracy, as well as aligning them with the bank's business and regulatory expectations. By failing to validate models, banks increase the risk of regulatory criticism, fines, and penalties.

The complex and resource-intensive nature of validation makes it necessary to dedicate sufficient resources to it. An independent validation team well versed in data management, technology, and relevant financial products or services — for example, credit, capital management, insurance, or financial crime compliance — is vital for success. Where shortfalls in the validation process are identified, timely remedial actions should be taken to close the gaps.

Data Validation in Excel

The following example is an introduction to data validation in Excel. The data validation button under the data tab provides the user with different types of data validation checks based on the data type in the cell. It also allows the user to define custom validation checks

using Excel formulas. The data validation can be found in the Data Tools section of the Data tab in the ribbon of Excel:
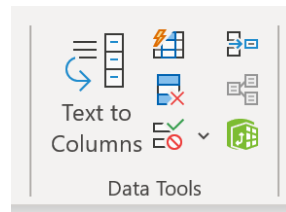
Fig 1: Data validation tool in Excel

Data Entry Task

The example below illustrates a case of data entry, where the province must be entered for every store location. Since stores are only located in certain provinces, any incorrect entry should be caught.

It is accomplished in Excel using a two-fold data validation. First, the relevant provinces are incorporated into a drop-down menu that allows the user to select from a list of valid provinces.

Fig. 2: First level of data validation

Second, if the user inputs a wrong province by mistake, such as "<u>NY</u>" instead of "NS," the system warns the user of the incorrect input.
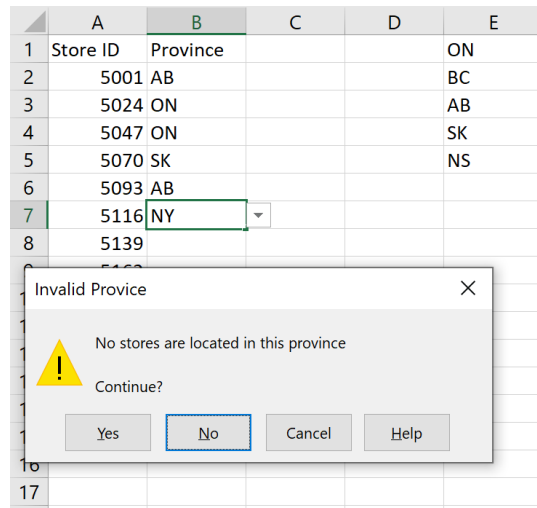
Fig. 3: Second level of data validation

Further, if the user ignores the warning, an analysis can be conducted using the data validation feature in Excel that identifies incorrect inputs.

Fig. 4: Final level of data validation

8.2 Model Evaluation

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and overfitted models. There are two methods of evaluating models in data science, Hold-Out and Cross-Validation. To avoid overfitting, both methods use a test set (not seen by the model) to evaluate model performance.

- **Hold-Out:** In this method, the mostly large dataset is *randomly* divided to three subsets:

1. **Training set** is a subset of the dataset used to build predictive models.

2. **Validation set** is a subset of the dataset used to assess the performance of model built in the training phase. It provides a test platform for fine tuning model's parameters and selecting the best-performing model. Not all modelling algorithms need a validation set.

3. **Test set** or unseen examples is a subset of the dataset to assess the likely future performance of a model. If a model fit to the training set much better than it fits the test set, overfitting is probably the cause.

● **Cross-Validation:** When only a limited amount of data is available, to achieve an unbiased estimate of the model performance we use k-fold cross-validation. In k-fold cross-validation, we divide the data into k subsets of equal size. We build models k times, each time leaving out one of the subsets from training and use it as the test set. If k equals the sample size, this is called "leave-one-out".

Model evaluation can be divided to **two sections:**

● **Classification Evaluation**

● **Regression Evaluation**

9.Interpretation:

Data interpretation is the process of reviewing data and drawing meaningful conclusions using a variety of analytical approaches. Data interpretation aids researchers in categorizing, manipulating, and summarising data in order to make sound business decisions. The end goal for a data interpretation project is to develop a good marketing strategy or to expand its client user base.

There are certain steps followed to conduct data interpretation:

● Putting together the data you'll need(neglecting irrelevant data)

● Developing the initial research or identifying the most important inputs;

● Sorting and filtering of data.

● Forming conclusions on the data.

● Developing recommendations or practical solutions.

9.1 Types of data interpretation

The purpose of data interpretation is to assist individuals in understanding numerical data that has been gathered, evaluated, and presented.

9.1.1 Qualitative data Interpretation

To evaluate qualitative data, also known as categorical data, the qualitative data interpretation approach is utilized. Words, instead of numbers or patterns, are used to describe data in this technique. Unlike quantitative data, which can be studied immediately after collecting and sorting it, qualitative data must first be converted into numbers before being analyzed. This is due to the fact that analyzing texts in their original condition is frequently time-consuming and results in a high number of mistakes. The analyst's coding should also be defined so that it may be reused and evaluated by others.

Observations: a description of the behavioral patterns seen in a group of people. The length of time spent on an activity, the sort of activity, and the form of communication used might all be examples of these patterns.

Groups of people: To develop a collaborative discussion about a study issue, group people and ask them pertinent questions.

Research: Similar to how patterns of behavior may be noticed, different forms of documentation resources can be classified and split into categories based on the type of information they include.

Interviews are one of the most effective ways to get narrative data. Themes, topics, and categories can be used to group inquiry replies. The interview method enables extremely targeted data segmentation.

The following methods are commonly used to produce qualitative data:

- Transcripts of interviews
- Questionnaires with open-ended answers
- Transcripts from call centers
- Documents and texts
- Audio and video recordings are available.
- Notes from the field

Now the second step is to interpret the data that is produced. This is done by the following methods:

Content Analysis

This is a popular method for analyzing qualitative data. Other approaches to analysis may fall under the general category of content analysis. An aspect of the content analysis is thematic analysis. By classifying material into words, concepts, and themes, content analysis is used to uncover patterns that arise from the text.

Narrative Analysis

The focus of narrative analysis is on people's experiences and the language they use to make sense of them. It's especially effective for acquiring a thorough insight into customers' viewpoints on a certain topic. We might be able to describe the results of a targeted case study using narrative analysis.

Discourse Analysis

Discourse analysis is a technique for gaining a comprehensive knowledge of the political, cultural, and power dynamics that exist in a given scenario. The emphasis here is on how people express themselves in various social settings. Brand strategists frequently utilize discourse analysis to figure out why a group of individuals reacts the way they do to a brand or product.

It's critical to be very clear on the type and scope of the study topic in order to get the most out of the analytical process. This will assist you in determining which research collection routes are most likely to assist you in answering your query.

Your approach to qualitative data analysis will differ depending on whether you are a corporation attempting to understand consumer sentiment or an academic surveying a school.

9.1.2 Quantitative data Interpretation

Quantitative data, often known as numerical data, is analyzed using the quantitative data interpretation approach. Because this data type contains numbers, it is examined using numbers rather than words. Quantitative analysis is a collection of procedures for analyzing numerical data. It frequently requires the application of statistical modeling techniques such as standard deviation, mean, and median. Let's try and understand these;

Median: The median is the middle value in a list of numbers that have been sorted ascending or descending, and it might be more descriptive of the data set than the average.

Mean: The basic mathematical average of two or more values is called a mean. The arithmetic mean approach, which utilizes the sum of the values in the series, and the

geometric mean method, which is the average number of products, are two ways to determine the mean for a given collection of numbers.

Standard deviation: The positive square root of the variance is the standard deviation. One of the most fundamental approaches to statistical analysis is the standard deviation. A low standard deviation indicates that the values are near to the mean, whereas a large standard deviation indicates that the values are significantly different from the mean.

There are three common uses for quantitative analysis.

- For starters, it's used to compare and contrast groupings. For instance, consider the popularity of certain car brands with different colors.

- It's also used to evaluate relationships between variables.

- Third, it's used to put scientifically sound theories to the test. Consider a hypothesis concerning the effect of a certain vaccination.

Regression analysis

A collection of statistical procedures for estimating connections between a dependent variable and one or maybe more independent variables is known as regression analysis. It may be used to determine the strength of a relationship across variables and to predict how they will interact in the future.

Cohort Analysis

Cohort analysis is a technique for determining how engaged users are over time. It's useful to determine whether user engagement is improving over time or just looking to improve due to growth. Cohort analysis is useful because it helps to distinguish between growth and engagement measures. Cohort analysis is watching how individuals' behavior develops over time in groups of people.

Predictive Analysis

By examining historical and present data, the predictive analytic approach seeks to forecast future trends. Predictive analytics approaches, which are powered by machine learning and deep learning, allow firms to notice patterns or possible challenges ahead of time and prepare educated initiatives. Predictive analytics is being used by businesses to address issues and identify new possibilities.

Prescriptive Analysis

The prescriptive analysis approach employs tools like as graph analysis,

Prescriptive analytics is a sort of data analytics in which technology is used to assist organisations in making better decisions by analyzing raw data. Prescriptive analytics, in particular, takes into account information about potential situations or scenarios, available resources, previous performance, and present performance to recommend a course of action or strategy. It may be used to make judgments throughout a wide range of time frames, from the immediate to the long term.

Conjoint Analysis

Conjoint analysis is the best market research method for determining how much customers appreciate a product's or service's qualities. This widely utilized method mixes real-life scenarios and statistical tools with market decision models

Cluster analysis

Any organization that wants to identify distinct groupings of consumers, sales transactions, or other sorts of behaviors and items may use cluster analysis as a valuable data-mining technique.

The goal of cluster analysis is to uncover groupings of subjects that are similar, where "similarity" between each pair of subjects refers to a global assessment of the entire collection of features. Cluster analysis, similar to factor analysis, deals with data matrices in which the variables haven't been partitioned into criteria and predictor subsets previously.

10. Deployment and Iteration:
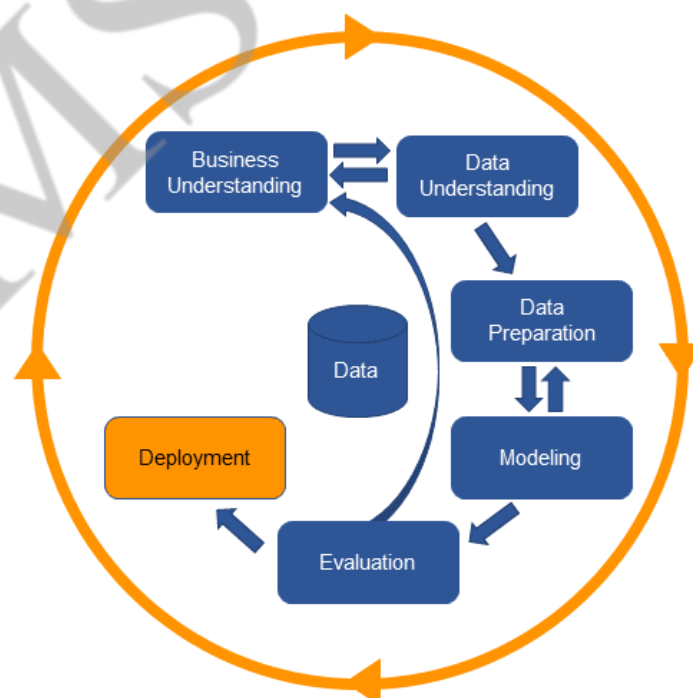
The iterative process is the practice of building, refining, and improving a project, product, or initiative. Teams that use the iterative development process create, test, and revise until they're satisfied with the end result. You can think of an iterative process as a trial-and-error methodology that brings your project closer to its end goal.

Iterative processes are a fundamental part of lean methodologies and Agile project management—but these processes can be implemented by any team, not just Agile ones. During the iterative process, you will continually improve your design, product, or project until you and your team are satisfied with the final project deliverable.

10.1 The benefits and challenges of the iterative process
The iterative model isn't right for every team—or every project. Here are the main pros and cons of the iterative process for your team.

Pros:

● Increased efficiency. Because the iterative process embraces trial and error, it can often help you achieve your desired result faster than a non-iterative process.
● Increased collaboration. Instead of working from predetermined plans and specs (which also takes a lot of time to create), your team is actively working together.
● Increased adaptability. As you learn new things during the implementation and testing phases, you can tweak your iteration to best hit your goals—even if that means doing something you didn't expect to be doing at the start of the iterative process.
● More cost effective. If you need to change the scope of the project, you'll only have invested the minimum time and effort into the process.
● Ability to work in parallel. Unlike other, non-iterative methodologies like the waterfall method, iterations aren't necessarily dependent on the work that comes before them. Team members can work on several elements of the project in parallel, which can shorten your overall timeline.
● Reduced project-level risk. In the iterative process, risks are identified and addressed during each iteration. Instead of solving for large risks at the beginning and end of the project, you're consistently working to resolve low-level risks.
● More reliable user feedback. When you have an iteration that users can interact with or see, they're able to give you incremental feedback about what works or doesn't work for them.

Cons:

● Increased risk of scope creep. Because of the trial-and-error nature of the iterative process, your project could develop in ways you didn't expect and exceed your original project scope.
● Inflexible planning and requirements. The first step of the iterative process is to define your project requirements. Changing these requirements during the iterative process can break the flow of your work, and cause you to create iterations that don't serve your project's purpose.
● Vague timelines. Because team members will create, test, and revise iterations until they get to a satisfying solution, the iterative timeline isn't clearly defined. Additionally, testing for

different increments can vary in length, which also impacts the overall iterative process timeline.

UNIT II   BUSINESS INTELLIGENCE

Data Warehouses and Data Mart – Knowledge Management – Types of Decisions –

Decision Making Process – Decision Support systems – Business Intelligence-

OLAP- Analytic Functions

1.Data Warehouses and Data Mart:

A Data Warehouse (DW) is an organised collection of integrated, subject-oriented databases designed to aid decision support functions. DW is organized at the right level of granularity to provide clean enterprise-wide data in a standardized format for reports, queries and analysis. DW is physically and functionally separate from an operational and transactional database. Creating a DW for analysis and queries represents investment in time and effort. It has to be constantly kept up-to-date for it to be useful.

A Data Warehousing (DW) is process for collecting and managing data from varied sources to provide meaningful business insights. A Data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting. It is a blend of technologies and components which aids the strategic use of data. It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

Data warehouse system is also known by the following name:

- Decision Support System (DSS)
- Executive Information System
- Management Information System
- Business Intelligence Solution
- Analytic Application
- Data Warehouse

1.1 How Data warehouse works?

A Data Warehouse works as a central repository where information arrives from one or more data sources. Data flows into a data warehouse from the transactional system and other relational databases.

Data may be:

1. Structured
2. Semi-structured
3. Unstructured data

The data is processed, transformed, and ingested so that users can access the processed data in the Data Warehouse through Business Intelligence tools, SQL clients, and spreadsheets. A data warehouse merges information coming from different sources into one comprehensive database.

By merging all of this information in one place, an organization can analyze its customers more holistically. This helps to ensure that it has considered all the information available. Data warehousing makes data mining possible. Data mining is looking for patterns in the data that may lead to higher sales and profits.

1.2 Types of Data Warehouse

Three main types of Data Warehouses are:

1. Enterprise Data Warehouse:

Enterprise Data Warehouse is a centralized warehouse. It provides decision support service across the enterprise. It offers a unified approach for organizing and representing

data. It also provide the ability to classify data according to the subject and give access according to those divisions.

2. Operational Data Store:

Operational Data Store, which is also called ODS, are nothing but data store required 1.2 when neither Data warehouse nor OLTP systems support organizations reporting needs. In ODS, Data warehouse is refreshed in real time. Hence, it is widely preferred for routine activities like storing records of the Employees.

3. Data Mart:

A data mart is a subset of the data warehouse. It specially designed for a particular line of business, such as sales, finance, sales or finance. In an independent data mart, data can collect directly from sources.

1.3 Components of Data warehouse

Four components of Data Warehouses are:

Load manager: Load manager is also called the front component. It performs with all the operations associated with the extraction and load of data into the warehouse. These operations include transformations to prepare the data for entering into the Data warehouse.

Warehouse Manager: Warehouse manager performs operations associated with the management of the data in the warehouse. It performs operations like analysis of data to ensure consistency, creation of indexes and views, generation of denormalization and aggregations, transformation and merging of source data and archiving and baking-up data.

Query Manager: Query manager is also known as backend component. It performs all the operation operations related to the management of user queries. The operations of this Data warehouse components are direct queries to the appropriate tables for scheduling the execution of queries.

End-user access tools:

This is categorized into five different groups like 1. Data Reporting 2. Query Tools 3. Application development tools 4. EIS tools, 5. OLAP tools and data mining tools.

1.4 Who needs Data warehouse?

Data warehouse is needed for all types of users like:

- Decision makers who rely on mass amount of data
- Users who use customized, complex processes to obtain information from multiple data sources.
- It is also used by the people who want simple technology to access the data
- It also essential for those people who want a systematic approach for making decisions.
- If the user wants fast performance on a huge amount of data which is a necessity for reports, grids or charts, then Data warehouse proves useful.
- Data warehouse is a first step If you want to discover 'hidden patterns' of data-flows and groupings.

1.5 What Is a Data Warehouse Used For?

Here, are most common sectors where Data warehouse is used:

Airline:

In the Airline system, it is used for operation purpose like crew assignment, analyses of route profitability, frequent flyer program promotions, etc.

Banking:

It is widely used in the banking sector to manage the resources available on desk effectively. Few banks also used for the market research, performance analysis of the product and operations.

Healthcare:

Healthcare sector also used Data warehouse to strategize and predict outcomes, generate patient's treatment reports, share data with tie-in insurance companies, medical aid services, etc.

Public sector:

In the public sector, data warehouse is used for intelligence gathering. It helps government agencies to maintain and analyze tax records, health policy records, for every individual.

Investment and Insurance sector:

In this sector, the warehouses are primarily used to analyze data patterns, customer trends, and to track market movements.

Retail chain:

In retail chains, Data warehouse is widely used for distribution and marketing. It also helps to track items, customer buying pattern, promotions and also used for determining pricing policy.

Telecommunication:

A data warehouse is used in this sector for product promotions, sales decisions and to make distribution decisions.

Hospitality Industry:

This Industry utilizes warehouse services to design as well as estimate their advertising and promotion campaigns where they want to target clients based on their feedback and travel patterns.

Steps to Implement Data Warehouse

The best way to address the business risk associated with a Data warehouse implementation is to employ a three-prong strategy as below

1. **Enterprise strategy**: Here we identify technical including current architecture and tools. We also identify facts, dimensions, and attributes. Data mapping and transformation is also passed.
2. **Phased delivery**: Datawarehouse implementation should be phased based on subject areas. Related business entities like booking and billing should be first implemented and then integrated with each other.
3. **Iterative Prototyping**: Rather than a big bang approach to implementation, the Datawarehouse should be developed and tested iteratively.

Here, are key steps in Datawarehouse implementation along with its deliverables.

Step	Tasks	Deliverables
1	Need to define project scope	Scope Definition
2	Need to determine business needs	Logical Data Model
3	Define Operational Datastore requirements	Operational Data Store Model

4	Acquire or develop Extraction tools	Extract tools and Software
5	Define Data Warehouse Data requirements	Transition Data Model
6	Document missing data	To Do Project List
7	Maps Operational Data Store to Data Warehouse	D/W Data Integration Map
8	Develop Data Warehouse Database design	D/W Database Design
9	Extract Data from Operational Data Store	Integrated D/W Data Extracts
10	Load Data Warehouse	Initial Data Load
11	Maintain Data Warehouse	On-going Data Access and Subsequent Lo

1,6 Best practices to implement a Data Warehouse

- Decide a plan to test the consistency, accuracy, and integrity of the data.
- The data warehouse must be well integrated, well defined and time stamped.
- While designing Datawarehouse make sure you use right tool, stick to life cycle, take care about data conflicts and ready to learn you're your mistakes.
- Never replace operational systems and reports
- Don't spend too much time on extracting, cleaning and loading data.
- Ensure to involve all stakeholders including business personnel in Datawarehouse implementation process. Establish that Data warehousing is a joint/ team project. You don't want to create Data warehouse that is not useful to the end users.
- Prepare a training plan for the end users.

1.7 Advantages of Data Warehouse:

- Data warehouse allows business users to quickly access critical data from some sources all in one place.
- Data warehouse provides consistent information on various cross-functional activities. It is also supporting ad-hoc reporting and query.
- Data Warehouse helps to integrate many sources of data to reduce stress on the production system.

- Data warehouse helps to reduce total turnaround time for analysis and reporting.
- Restructuring and Integration make it easier for the user to use for reporting and analysis.
- Data warehouse allows users to access critical data from the number of sources in a single place. Therefore, it saves user's time of retrieving data from multiple sources.
- Data warehouse stores a large amount of historical data. This helps users to analyze different time periods and trends to make future predictions.

1.8 Disadvantages of Data Warehouse:

- Not an ideal option for unstructured data.
- Creation and Implementation of Data Warehouse is surely time confusing affair.
- Data Warehouse can be outdated relatively quickly
- Difficult to make changes in data types and ranges, data source schema, indexes, and queries.
- The data warehouse may seem easy, but actually, it is too complex for the average users.
- Despite best efforts at project management, data warehousing project scope will always increase.
- Sometime warehouse users will develop different business rules.
- Organisations need to spend lots of their resources for training and Implementation purpose.

1.9 The Future of Data Warehousing

- Change in **Regulatory constrains** may limit the ability to combine source of disparate data. These disparate sources may include unstructured data which is difficult to store.
- As the **size** of the databases grows, the estimates of what constitutes a very large database continue to grow. It is complex to build and run data warehouse systems which are always increasing in size. The hardware and software resources are available today do not allow to keep a large amount of data online.
- **Multimedia data** cannot be easily manipulated as text data, whereas textual information can be retrieved by the relational software available today. This could be a research subject.
- retrieved by the relational software available today. This could be a research subject.

Data Warehouse Tools

There are many Data Warehousing tools are available in the market. Here, are some most prominent one:

1. MarkLogic:

MarkLogic is useful data warehousing solution that makes data integration easier and faster using an array of enterprise features. This tool helps to perform very complex search operations. It can query different types of data like documents, relationships, and metadata.

2. Oracle:

Oracle is the industry-leading database. It offers a wide range of choice of data warehouse solutions for both on-premises and in the cloud. It helps to optimize customer experiences by increasing operational efficiency.

3. Amazon RedShift:

Amazon Redshift is Data warehouse tool. It is a simple and cost-effective tool to analyze all types of data using standard SQL and existing BI tools. It also allows running complex queries against petabytes of structured data, using the technique of query optimization.

1.9.1 Differences between Data Warehouse and Data Mart
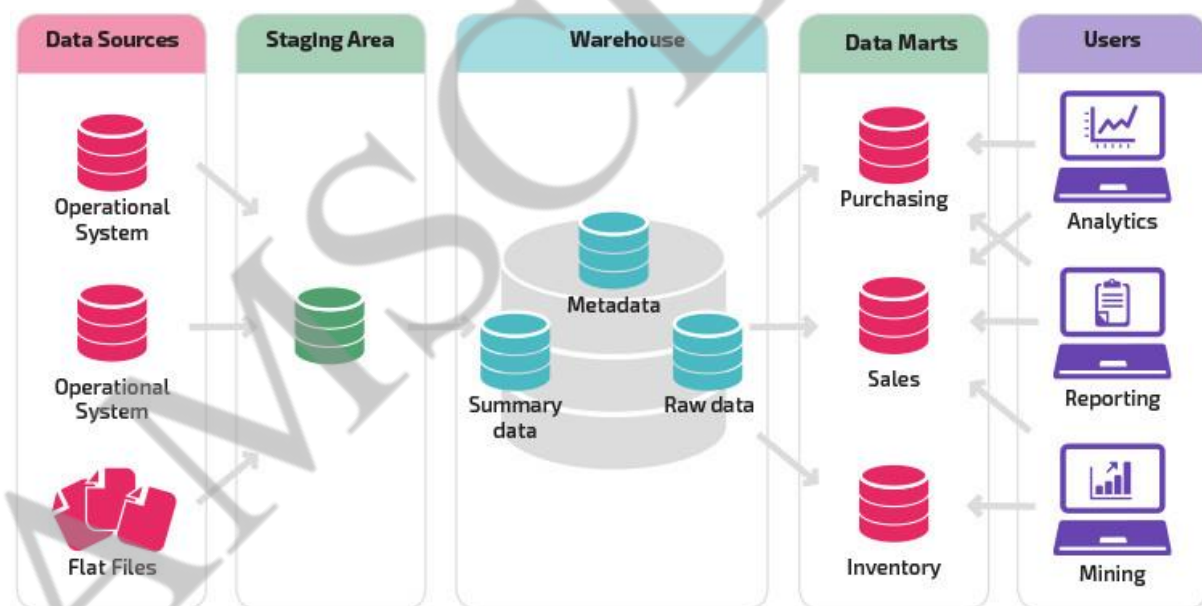
1.

Parameter	Data Warehouse	Data Mart
Definition	A Data Warehouse is a large repository of data collected from different organizations or departments within a corporation.	A data mart is an only subtype of a Data Warehouse. It is designed to meet the need of a certain user group.
Usage	It helps to take a strategic decision.	It helps to take tactical decisions for the business.
Objective	The main objective of Data Warehouse is to provide an integrated environment and coherent picture of the business at a point in time.	A data mart mostly used in a business division at the department level.
Designing	The designing process of Data Warehouse is quite difficult.	The designing process of Data Mart is easy.
	May or may not use in a dimensional model. However, it can feed dimensional models.	It is built focused on a dimensional model using a start schema.
Data Handling	Data warehousing includes large area of the corporation which is why it takes a long time to process it.	Data marts are easy to use, design and implement as it can only handle small amounts of data.
Focus	Data warehousing is broadly focused all the departments. It is possible that it can even represent the entire company.	Data Mart is subject-oriented, and it is used at a department level.
Data type	The data stored inside the Data Warehouse are always detailed when compared with data mart.	Data Marts are built for particular user groups. Therefore, data short and limited.
Subject-area	The main objective of Data Warehouse is to provide an integrated environment and coherent picture of the business at a point in time.	Mostly hold only one subject area- for example, Sales figure.
Data storing	Designed to store enterprise-wide decision data, not just marketing data.	Dimensional modeling and star schema design employed for optimizing the performance of access layer.
Data type	Time variance and non-volatile design are strictly enforced.	Mostly includes consolidation data structures to meet subject area's query and reporting needs.
Data value	Read-Only from the end-users standpoint.	Transaction data regardless of grain fed directly from the Data Warehouse.
Scope	Data warehousing is more helpful as it can bring information from any department.	Data mart contains data, of a specific department of a company. There are maybe separate data marts for sales, finance, marketing, etc. Has limited usage

Source	In Data Warehouse Data comes from many sources.	In Data Mart data comes from very few sources.
Size	The size of the Data Warehouse may range from 100 GB to 1 TB+.	The Size of Data Mart is less than 100 GB.
Implementation time	The implementation process of Data Warehouse can be extended from months to years.	The implementation process of Data Mart is restricted to few months.

2. Knowledge Management (KM): Concept, Features and Process

Concept of KM:

KM may be defined as follows:

Knowledge management is a process of acquiring, generating, accumulating and using knowledge for the benefit of the organisation to enable it to gain a competitive edge for survival, growth and prosperity in a globalized competitive economy.

According to some management experts, notably Peter F. Drucker, KM is a bad term; in as much as knowledge cannot be managed.

Rather, KM requires conditions for the emergence of a learning organisation; which is necessary for generation, sharing and use of knowledge residing in the minds of people.

2.1 Features of Knowledge Management

Some salient features of KM are described below:

(i) KM is a systematic process; consisting of standardized procedures to collect, store, distribute and use knowledge. The essence of KM is to get right knowledge to right people, at the right time.

(ii) Knowledge is of two types – explicit and implicit. Explicit knowledge is visible information available in literature, reports, patents, technical specifications, communication with customers, suppliers, competitors etc. It can be embedded in rules, systems, policies and procedures etc. of the organisation.

Tacit or implicit knowledge is personal knowledge residing in the minds of people as a result of their personal beliefs, values, perspectives and experience. There is a need for a learning organisation for enhancement, sharing and utilisation of tacit knowledge.

(iii) KM is a continuous process; as the world economy is dynamic and full of challenges. It requires constant creation of new skills and capabilities and improvement of existing ones.

(iv) KM requires whole-hearted support of top management, to provide cultural and technical foundation for the origination and implementation of KM practices.

(v) The objective of KM is improvement in organisational performance; to enable the organisation acquire, sharpen and utilize its competitive edge for survival and growth in the global economy of today.

2.2 Knowledge Management and Information Technology:

KM is not an outgrowth of IT. Rather, KM requires human skills, creativity and innovative capabilities of people; which are the base of KM. In fact I there are tools of IT like Intranets, Lotus Notes, MS-Exchange etc.; which provide an infrastructure for the free play of human creativity and innovative powers for the formulation of corporation strategy, in a competitive globalized environment.

The above ideas are illustrated with the help of the following diagram:

KM, IT and Corporate Strategy
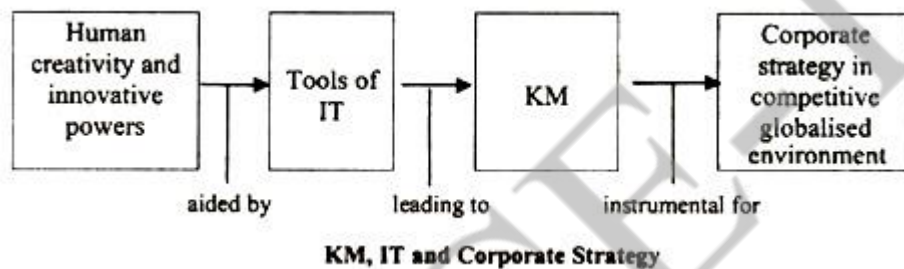
Knowledge Management IT and Corporate Strategy

An Overview of the Process of KM:

KM broadly consists of the following major steps:

(i) **Identification of Knowledge Needs:**

The first step in KM is an identification of what type of knowledge is required for the successful designing and implementation of corporate strategy.

(ii) **Determination of Knowledge Assets:**

The management must identify what are the knowledge assets of the organisation; which basically are competitors, suppliers, governmental agencies, products and processes, technology etc. Management must plan to get maximum returns out of knowledge assets.

(iii) **Generation of Knowledge:**

Generation of knowledge requires two sources:

(a) Acquisition of knowledge through knowledge assets e.g. knowledge about new products (from competitors), new technologies, social, economical, political changes. It

also requires transformation of raw information into knowledge, useful to solve business problems.

(b) Generation of knowledge, by creating conditions for the emergence of a learning organisation. This is the most important internal source of knowledge generation which makes tacit knowledge of individuals available for organisational purposes.

(iv) **Knowledge Storage:**

It includes preserving existing and acquired knowledge in knowledge repositories. (A knowledge repository is an on line computer based storehouse of organised information about a particular domain of knowledge).

(v) **Knowledge Distribution:**

It is a process which allows members of the organisation to have an access to the collective knowledge of the organisation.

(vi) **Knowledge Utilization:**

It requires embedding knowledge in products, processes, procedures etc. of the organisation. Best utilisation of knowledge takes place when managers utilize knowledge in organisational decision making. A learning organisation creates conditions for sharing and utilizing knowledge in organisational contexts.

(vii) **Feedback on Knowledge Management**

Feedback on KM implies evaluating the significance of knowledge assets. It also includes impact of KM on organisational performance; and devising techniques for betterment of KM in future.

An overview of the process of KM- at a glance

2.3 Significance of Knowledge Management

Significance of KM could be highlighted with reference the following advantages which KM provides to the organisation:

(i**) Building and Sharpening Competitive Edge:**

KM enables a corporation to build and sharpen its competitive edge, for survival and growth in the competitive globalized economy. In fact, KM aided by IT tools enables a corporation to design and implement most appropriate corporate strategies.

(ii) **Betterment of Human Relations:**

KM is basically built on the knowledge generated, shared and utilized through a learning organisation. There is no doubt that learning organisation provides the foundation on which the building of KM could be built. A learning organisation through facilitating interaction among people of the organisation, leads to betterment of human relations; which is a very big permanent asset an organisation can boast of to possess.

(iii) **Improvement in Organisational Efficiency:**

KM provides knowledge which can be embedded in organisational processes. It makes knowledge available for decision-making purposes. Thus it helps to improve organisational efficiency, resulting in reduced costs and increased profits, for the organisation.

(iv) **Enhancement of Human Capital Capabilities**

KM-its concept and practices – motivate people to enhance their intellectual capabilities, resulting in new skills, improvement of existing skills etc. Thus not only does KM enhance the intellectual elements of people; but also indirectly prevents depreciation of human capital.

(v) **Enhancement of Enterprise Goodwill:**

Initiation and practices of KM help an enterprise enhance its goodwill in the global market; enabling it to acquire more success and prosperity.

3.**Types of Decisions in Business Intelligence**

The characteristics of decisions faced by managers at different levels are quite different. Decisions can be classified as structured, semi structured, and unstructured. Unstructured decisions are those in which the decision maker must provide judgment, evaluation, and insights into the problem definition. Each of these decisions is novel, important, and nonroutine, and there is no well-understood or agreed-on procedure for making them.

Structured decisions, by contrast, are repetitive and routine, and decision makers can follow a definite procedure for handling them to be efficient. Many decisions have elements of both and are considered semi structured decisions, in which only part of the problem has a clear-cut answer provided by an accepted procedure. In general, structured decisions are made more prevalently at lower organizational levels, whereas unstructured decision making is more common at higher levels of the firm.

Senior executives tend to be exposed to many unstructured decision situations that are open ended and evaluative and that require insight based on many sources of information and personal experience. For example, a CEO in today's music industry might ask, "Whom should we choose as a distribution partner for our online music catalog—Apple, Microsoft, or Sony?" Answering this question would require access to news,

government reports, and industry views as well as high-level summaries of firm performance. However, the answer would also require senior managers to use their own best judgment and poll other managers for their opinions.

Middle management and operational management tend to face more structured decision scenarios, but their decisions may include unstructured components. A typical middlelevel management decision might be "Why is the order fulfillment report showing a decline over the last six months at a distribution center in Minneapolis?" This middle manager could obtain a report from the firm's enterprise system or distribution management system on order activity and operational efficiency at the Minneapolis distribution center. This is the structured part of the decision. But before arriving at an answer, this middle manager will have to interview employees and gather more unstructured information from external sources about local economic conditions or sales trends.

Rank-and-file employees tend to make more structured decisions. For example, a sales account representative often has to make decisions about extending credit to customers by consulting the firm's customer database that contains credit information. In this case the decision is highly structured, it is a routine decision made thousands of times each day in most firms, and the answer has been preprogrammed into a corporate risk management or credit reporting system.

The types of decisions faced by project teams cannot be classified neatly by organizational level. Teams are small groups of middle and operational managers and perhaps employees assigned specific tasks that may last a few months to a few years. Their tasks may involve unstructured or semistructured decisions such as designing new products, devising new ways to enter the marketplace, or reorganizing sales territories and compensation systems.

SYSTEMS FOR DECISION SUPPORT

There are four kinds of systems used to support the different levels . We introduced some of these systems in *Management information systems (MIS)* provide routine reports and summaries of transaction- level data to middle and operational-level managers to provide answers to structured and semistructured decision problems.

1.Decision-support systems (DSS) are targeted systems that combine analytical models with operational data and supportive interactive queries and analysis for middle managers who face semistructured decision situations.

2.Executive support systems (ESS) are specialized systems that provide senior management making primarily unstructured decisions with a broad array of both external information (news, stock analyses, industry trends) and high-level summaries of firm performance. The purpose of ESS to help the C- level managers to focus on the information that really affect the overall profitability and success of the firm. The leading methodology for understanding the really important information needed by the firm's executive is called the Balanced Score Card Method, a frame work for operationalizing the firm's strategic plan by focusing on measurable outcomes on four dimensions of firm performance.Financial,business process ,customer, learning and growth. Performance on each dimension is measured using KPI's.

3.Group decision-support systems (GDSS) are specialized systems that provide a group electronic environment in which managers and teams can collectively make decisions and design solutions for unstructured and semistructured problems.GDSS guided meetings takes place in a conference rooms with special software and hardware tools to facilitate group decision making.It makes possible to increase the meeting size and increase in productivity.Because individuals contribute simultaneously at the same time rather than one at a time.
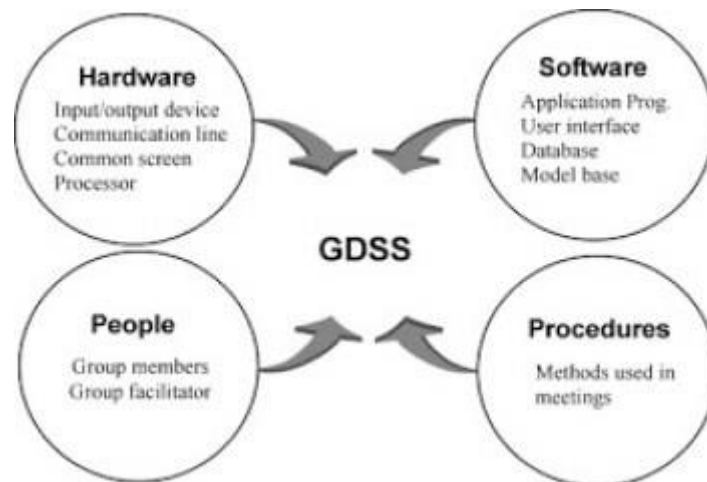
Organizational Level	Decision Type	Type of Decision-Support System	Examples
Senior management	Unstructured	Executive support systems (ESS)	Decide entrance or exit from markets Approve capital budget Decide long-term corporate objectives
Middle management/ project teams	Semistructured	Management information systems (MIS) Decision-support systems (DSS) Group decision-support systems (GDSS)	Allocate resources to managers and departments Design a new corporate Web site Develop a marketing plan Design a departmental budget
Operational management/ project teams Employees	Semistructured Structured	Decision-support systems (DSS) Management information systems (MIS) Group decision-support systems (GDSS)	Evaluate employee performance Restock inventory Routine credit decisions Determine special offers to customers

3.1.STAGES IN THE DECISION-MAKING PROCESS

Making decisions consists of several different activities. Simon (1960) describes four different stages in decision making: intelligence, design, choice, and implementation

Stages in Decision Making

Problem discovery: What is the problem? — Intelligence

Solution discovery: What are the possible solutions? — Design

Choosing solutions: What is the best solution? — Choice

Solution testing: Is the solution working? Can we make it work better? — Implementation

System Support

Management Information Systems (MIS): Routine reports and exception reporting

Decision-Support Systems (DSS): Using analytical models and spreadsheets to flexibly analyze data and design possible solutions

Decision-Support Systems (DSS): Using analytical models and larger data sets to identify the "optimal" solution

Group Decision-Support Systems (GDSS): Using online meeting and collaboration tools to choose the most acceptable and widely supported solution

Management Information Systems (MIS): Routine reports and exception reporting

Decision-Support Systems (DSS): Using analytical models and data to fine-tune the solution

 The decision-making process can be described in four steps that follow one another in a logical order. In reality, decision makers frequently circle back to reconsider the previous stages and through a process of iteration eventually arrive at a solution that is workable.

Intelligence consists of discovering, identifying, and understanding the problems occurring in the organization—why is there a problem, where, and what effects is it having on the firm. Traditional MIS that deliver a wide variety of detailed information can help identify problems, especially if the systems report exceptions.

Design involves identifying and exploring various solutions to the problem. Decisionsupport systems (DSS) are ideal in this stage for exploring alternatives because they possess analytical tools for modeling data, enabling users to explore various options quickly.

Choice consists of choosing among solution alternatives. Here, DSS with access extensive firm data can help managers choose the optimal solution. Also group decisionsupport systems can be used to bring groups of managers together in an electronic online environment to discuss different solutions and make a choice.

Implementation involves making the chosen alternative work and continuing

to monitor how well the solution is working. Here, traditional MIS come back into play by providing managers with routine reports on the progress of a specific solution. Support systems can range from full-blown MIS to much smaller systems, as well as project-planning software operating on personal computers.

In the real world, the stages of decision making described here do not necessarily follow a linear path. You can be in the process of implementing a decision, only to discover that your solution is not working. In such cases, you will be forced to repeat the design, choice, or perhaps even the intelligence stage.

For instance, in the face of declining sales, a sales management team may strongly support a new sales incentive system to spur the sales force on to greater effort. If paying the sales force a higher commission for making more sales does not produce sales increases, managers would need to investigate whether the problem stems from poor product design, inadequate customer support, or a host of other causes, none of which would be "solved" by a new incentive system.

3.2 Trends in Decision Support and Business Intelligence

Systems supporting management decision making originated in the early 1960s as early MIS that created fixed, inflexible paper-based reports and distributed them to managers on a routine schedule. In the 1970s, the first DSS emerged as standalone applications with limited data and a few analytic models. ESS emerged during the 1980s to give senior managers an overview of corporate operations. Early ESS were expensive, based on custom technology, and suffered from limited data and flexibility.

The rise of client/server computing, the Internet, and Web technologies has made a major impact on systems that support decision making. Many decision-support applications are now delivered over corporate intranets. We see six major trends:

- *Detailed enterprise-wide data.* Enterprise systems create an explosion in firmwide, current, and relatively accurate information, supplying end users at their desktops with powerful analytic tools for analyzing and visualizing data.
- *Broadening decision rights and responsibilities.* As information becomes more widespread throughout the corporation, it is possible to reduce levels of hierarchy and grant more decision-making authority to lower-level employees.
- *Intranets and portals.* Intranet technologies create global, company-wide networks that ease the flow of information across divisions and regions and delivery of near real-time data to management and employee desktops.
- *Personalization and customization of information.* Web portal technologies provide great flexibility in determining what data each employee and manager sees on his or her desktop. Personalization of decision

information can speed up decision making by enabling users to filter out irrelevant information.

- ***Extranets and collaborative commerce.*** Internet and Web technologies permit suppliers and logistics partners to access firm enterprise data and decision-support tools and work collaboratively with the firm.
- ***Team support tools.*** Web-based collaboration and meeting tools enable project teams, task forces, and small groups to meet online using corporate intranets or extranets. These new collaboration tools borrow from earlier GDSS and are used for both brainstorming and decision sessions.

4.Business Intelligence

Business intelligence combines business analytics, data mining, data visualization, data tools and infrastructure, and best practices to help organizations make more data-driven decisions. In practice, you know you've got modern business intelligence when you have a comprehensive view of your organization's data and use that data to drive change, eliminate inefficiencies, and quickly adapt to market or supply changes. Modern BI solutions prioritize flexible self-service analysis, governed data on trusted platforms, empowered business users, and speed to insight

Business Intelligence is a set of processes, architectures, and technologies that convert raw data into meaningful information that drives profitable business actions. It is a suite of software and services to transform data into actionable intelligence and knowledge.

BI has a direct impact on organization's strategic, tactical and operational business decisions. BI supports fact-based decision making using historical data rather than assumptions and gut feeling.

BI tools perform data analysis and create reports, summaries, dashboards, maps, graphs, and charts to provide users with detailed intelligence about the nature of the business.

4.1 Why is BI important?

- Measurement: creating KPI (Key Performance Indicators) based on historic data
- Identify and set benchmarks for varied processes.
- With BI systems organizations can identify market trends and spot business problems that need to be addressed.
- BI helps on data visualization that enhances the data quality and thereby the quality of decision making.
- BI systems can be used not just by enterprises but SME (Small and Medium Enterprises)

4.2 How Business Intelligence systems are implemented?

step 1) Raw Data from corporate databases is extracted. The data could be spread across multiple systems heterogeneous systems.

step 2) The data is cleaned and transformed into the data warehouse. The table can be linked, and data cubes are formed.

Step 3) Using BI system the user can ask quires, request ad-hoc reports or conduct any other analysis.

4.3 Examples of Business Intelligence System used in Practice

Example 1:

. In an Online Transaction Processing (OLTP) system information that could be fed into product database could be

- add a product line
- change a product price

Correspondingly, in a Business Intelligence system query that would beexecuted for the product subject area could be did the addition of new product line or change in product price increase revenues

In an advertising database of OLTP system query that could be executed

- Changed in advertisement options
- Increase radio budget

Correspondigly, in BI system query that could be executed would be how many new clients added due to change in radio budget

In OLTP system dealing with customer demographic data bases data that could be fed would be

- increase customer credit limit
- change in customer salary level

Correspondingly in the OLAP system query that could be executed would be can customer profile changes support support higher product price

Example 2:

A hotel owner uses BI analytical applications to gather statistical information regarding average occupancy and room rate. It helps to find aggregate revenue generated per room.

It also collects statistics on market share and data from customer surveys from each hotel to decides its competitive position in various markets.

By analyzing these trends year by year, month by month and day by day helps management to offer discounts on room rentals.

Example 3:

A bank gives branch managers access to BI applications. It helps branch manager to determine who are the most profitable customers and which customers they should work on.

The use of BI tools frees information technology staff from the task of generating analytical reports for the departments. It also gives department personnel access to a richer data source.

4.4 Four types of BI users

Following given are the four key players who are used Business Intelligence System:

1. The Professional Data Analyst:

The data analyst is a statistician who always needs to drill deep down into data. BI system helps them to get fresh insights to develop unique business strategies.

2. The IT users:

The IT user also plays a dominant role in maintaining the BI infrastructure.

3. The head of the company:

CEO or CXO can increase the profit of their business by improving operational efficiency in their business.

4. The Business Users"

Business intelligence users can be found from across the organization. There are mainly two types of business users

1. Casual business intelligence user
2. The power user.

The difference between both of them is that a power user has the capability of working with complex data sets, while the casual user need will make him use dashboards to evaluate predefined sets of data.

4.5 Advantages of Business Intelligence

Here are some of the advantages of using Business Intelligence System:

1. Boost productivity

With a BI program, It is possible for businesses to create reports with a single click thus saves lots of time and resources. It also allows employees to be more productive on their tasks.

2. To improve visibility

BI also helps to improve the visibility of these processes and make it possible to identify any areas which need attention.

3. Fix Accountability

BI system assigns accountability in the organization as there must be someone who should own accountability and ownership for the organization's performance against its set goals.

4. It gives a bird's eye view:

BI system also helps organizations as decision makers get an overall bird's eye view through typical BI features like dashboards and scorecards.

5. It streamlines business processes:

BI takes out all complexity associated with business processes. It also automates analytics by offering predictive analysis, computer modeling, benchmarking and other methodologies.

6. It allows for easy analytics.

BI software has democratized its usage, allowing even nontechnical or non-analysts users to collect and process data quickly. This also allows putting the power of analytics from the hand's many people.

4.6 BI System Disadvantages

1. Cost:

Business intelligence can prove costly for small as well as for medium-sized enterprises. The use of such type of system may be expensive for routine business transactions.

2. Complexity:

Another drawback of BI is its complexity in implementation of datawarehouse. It can be so complex that it can make business techniques rigid to deal with.

3. Limited use

Like all improved technologies, BI was first established keeping in consideration the buying competence of rich firms. Therefore, BI system is yet not affordable for many small and medium size companies.

4. Time Consuming Implementation

It takes almost one and half year for data warehousing system to be completely implemented. Therefore, it is a time-consuming process.

5.What is OLAP?

A core component of data warehousing implementations, OLAP enables fast, flexible multidimensional data analysis for business intelligence (BI) and decision support applications.

OLAP (for *online analytical processing*) is software for performing multidimensional analysis at high speeds on large volumes of data from a [data warehouse](#), data mart, or some other unified, centralized data store.

Most business data have multiple dimensions—multiple categories into which the data are broken down for presentation, tracking, or analysis. For example, sales figures might have several dimensions related to location (region, country, state/province, store), time (year, month, week, day), product (clothing, men/women/children, brand, type), and more.

But in a data warehouse, data sets are stored in tables, each of which can organize data into just two of these dimensions at a time. OLAP extracts data from multiple relational data sets and reorganizes it into a multidimensional format that enables very fast processing and very insightful analysis.

5.1What is an OLAP cube?

The core of most OLAP systems, the OLAP cube is an array-based multidimensional database that makes it possible to process and analyze multiple data dimensions much more quickly and efficiently than a traditional relational database.

A relational database table is structured like a spreadsheet, storing individual records in a two-dimensional, row-by-column format. Each data "fact" in the database sits at the intersection of two dimensions–a row and a column—such as *region* and *total sales.*

SQL and relational database reporting tools can certainly query, report on, and analyze multidimensional data stored in tables, but performance slows down as the data volumes increase. And it requires a lot of work to reorganize the results to focus on different dimensions.

This is where the OLAP cube comes in. The OLAP cube extends the single table with additional layers, each adding additional dimensions—usually the next level in the "concept hierarchy" of the dimension. For example, the top layer of the cube might organize sales by region; additional layers could be country, state/province, city and even specific store.

In theory, a cube can contain an infinite number of layers. (An OLAP cube representing more than three dimensions is sometimes called a hypercube.) And smaller cubes can exist within layers—for example, each store layer could contain cubes arranging sales by salesperson and product. In practice, data analysts will create OLAP cubes containing just the layers they need, for optimal analysis and performance.
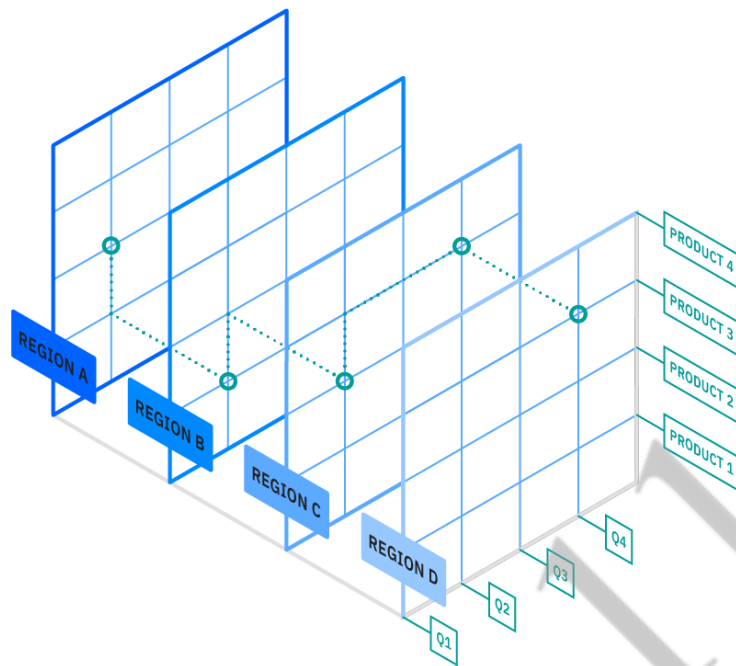
OLAP cubes enable four basic types of multidimensional data analysis:

Drill-down

The drill-down operation converts less-detailed data into more-detailed data through one of two methods—moving down in the concept hierarchy or adding a new dimension to the cube. For example, if you view sales data for an organization's calendar or fiscal quarter, you can drill-down to see sales for each month, moving down in the concept hierarchy of the "time" dimension.

Roll up

Roll up is the opposite of the drill-down function—it aggregates data on an OLAP cube by moving up in the concept hierarchy or by reducing the number of dimensions. For example, you could move up in the concept hierarchy of the "location" dimension by viewing each country's data, rather than each city.

Slice and dice

The slice operation creates a sub-cube by selecting a single dimension from the main OLAP cube. For example, you can perform a slice by highlighting all data for the organization's first fiscal or calendar quarter (time dimension).

The dice operation isolates a sub-cube by selecting several dimensions within the main OLAP cube. For example, you could perform a dice operation by highlighting all data by an organization's calendar or fiscal quarters (time dimension) and within the U.S. and Canada (location dimension).

Pivot

The pivot function rotates the current cube view to display a new representation of the data—enabling dynamic multidimensional views of data. The OLAP pivot function is comparable to the pivot table feature in spreadsheet software, such as Microsoft Excel, but while pivot tables in Excel can be challenging, OLAP pivots are relatively easier to use (less expertise is required) and have a faster response time and query performance.

MOLAP vs. ROLAP vs. HOLAP

OLAP that works directly with a multidimensional OLAP cube is known as *multidimensional OLAP,* or *MOLAP*. Again, for most uses, MOLAP is the fastest and most practical type of multidimensional data analysis.

However, there are two other types of OLAP which may be preferable in certain cases:

ROLAP

ROLAP, or *relational OLAP,* is multidimensional data analysis that operates directly on data on relational tables, without first reorganizing the data into a cube.

As noted previously, SQL is a perfectly capable tool for multidimensional queries, reporting, and analysis. But the SQL queries required are complex, performance can drag, and the resulting view of the data is static—it can't be pivoted to represent a different view of the data. ROLAP is best when the ability to work directly with large amounts of data is more important than performance and flexibility.

HOLAP

HOLAP, or *hybrid OLAP*, attempts to create the optimal division of labor between relational and multidimensional databases within a single OLAP architecture. The relational tables contain larger quantities of data, and OLAP cubes are used for aggregations and speculative processing. HOLAP requires an OLAP server that supports both MOLAP and ROLAP.

A HOLAP tool can "drill through" the data cube to the relational tables, which paves the way for quick data processing and flexible access. This hybrid system can offer better scalability but can't escape the inevitable slow-down when accessing relational data sources. Also, its complex architecture typically requires more frequent updates and

maintenance, as it must store and process all the data from relational databases and multidimensional databases. For this reason, HOLAP can end up being more expensive.

OLAP vs. OLTP

Online transaction processing, or *OLTP*, refers to data-processing methods and software focused on transaction-oriented data and applications.

The main difference between OLAP and OLTP is in the name: OLAP is analytical in nature, and OLTP is transactional.

OLAP tools are designed for multidimensional analysis of data in a data warehouse, which contains both transactional and historical data. In fact, an OLAP server is typically the middle, analytical tier of a data warehousing solution. Common uses of OLAP include data mining and other business intelligence applications, complex analytical calculations, and predictive scenarios, as well as business reporting functions like financial analysis, budgeting, and forecast planning.

OLTP is designed to support transaction-oriented applications by processing recent transactions as quickly and accurately as possible. Common uses of OLTP include ATMs, e-commerce software, credit card payment processing, online bookings, reservation systems, and record-keeping tools.

UNIT III   BUSINESS FORECASTING

Introduction to Business Forecasting and Predictive Analytics - Logic and Data Driven Models - Data Mining and Predictive Analysis Modeling - Machine Learning for Predictive Analytics.

1. Introduction to Business Forecasting

Business analysts may choose from a wide range of forecasting techniques to support decision making. Selecting the appropriate method depends on the characteristics of the forecasting problem, such as the time horizon of the variable being forecast, as well as available information on which the forecast will be based.

Three major categories of forecasting approaches are *qualitative and judgmental techniques*, *statistical time-series models*, and *explanatory/causal methods*. In this chapter, we introduce forecasting techniques in each of these categories and use basic Excel tools, *XLMiner*, and linear regression to implement them in a spreadsheet environment.

1.1 Qualitative and Judgmental Forecasting

Qualitative and judgmental techniques rely on experience and intuition; they are necessary when historical data are not available or when the decision maker needs to forecast far into the future. For example, a forecast of when the next generation of a microprocessor will be available and what capabilities it might have will depend greatly on the opinions and expertise of individuals who understand the technology. Another use of judgmental methods is to incorporate nonquantitative information, such as the impact of government regulations or competitor behavior, in a quantitative forecast. Judgmental techniques range from such simple methods as a manager's opinion or a group-based jury of executive opinion to more structured approaches such as historical analogy and the Delphi method.

1.1.1 The Delphi Method

A popular judgmental forecasting approach, called the **Delphi method**, uses a panel of experts, whose identities are typically kept confidential from one another, to respond to a sequence of questionnaires. After each round of responses, individual opinions, edited to ensure anonymity, are shared, allowing each to see what the other experts think. Seeing other experts' opinions helps to reinforce those in agreement and to influence those who did not agree to possibly consider other factors. In the next round, the experts revise their estimates, and the process is repeated, usually for no more than two or three rounds. The Delphi method promotes unbiased exchanges of ideas and discussion and usually results in some convergence of opinion. It is one of the better approaches to forecasting long range trends and impacts.

Indicators and Indexes

Indicators and indexes generally play an important role in developing judgmental forecasts.

Indicators are measures that are believed to influence the behaviour of a variable we wish to forecast. By monitoring changes in indicators, we expect to gain insight about the future behaviour of the variable to help forecast the future.

Example 1 Leading Economic Indicators

The Department of Commerce initiated an Index of Leading Indicators to help predict future economic performance.
Components of the index include the following:
• average weekly hours, manufacturing
• average weekly initial claims, unemployment
insurance
• new orders, consumer goods, and materials
• vendor performance—slower deliveries
• new orders, nondefense capital goods
• building permits, private housing
• stock prices, 500 common stocks (Standard & Poor)
• money supply
• interest rate spread
• index of consumer expectations (University of
Michigan)
Business Conditions Digest included more than 100 time series in seven economic areas. This publication was discontinued in March 1990, but information related to the Index of Leading Indicators was continued in *Survey of Current Business.* In December 1995, the U.S. Department of Commerce sold this data source to The Conference Board, which now markets the information under the title *Business Cycle Indicators*; information can be obtained at its Web site (www.conference-board.org). The site includes excellent current information about the calculation of the index as well as its current components.

1.2 Statistical Forecasting Models

Statistical time-series models find greater applicability for short-range forecasting problems.
Time Series
A **time series** is a stream of historical data, such as weekly sales. We characterize the values of a time series over T periods as A_t, t = 1, 2, c, T. Time-series models assume that whatever forces have influenced sales in the recent past will continue into the near future; thus, forecasts are developed by extrapolating these data into the future. Time series generally have one or more of the following components: random behavior, trends, seasonal effects, or cyclical effects.
Stationary Time Series
Time series that do not have trend, seasonal, or cyclical effects but are relatively constant and exhibit only random behavior are called **stationary time series**.
Many forecasts are based on analysis of historical time-series data and are predicated on the assumption that the future is an extrapolation of the past
Statistical time-series models find greater applicability for short-range forecasting problems.
A **trend** is a gradual upward or downward movement of a time series over time.

Time series may also exhibit short-term seasonal effects (over a year, month, week, or even a day) as well as longer-term cyclical effects, or nonlinear trends. A seasonal effect is one that repeats at fixed intervals of time, typically a year, month, week, or day.

At a neighborhood grocery store, for instance, short-term seasonal patterns may occur over a week, with the heaviest volume of customers on weekends; seasonal patterns may also be evident during the course of a day, with higher volumes in the mornings and late afternoons. Figure shows seasonal changes in natural gas usage for a homeowner over the course of a year (Excel file Gas & Electric). Cyclical effects describe ups and downs over a much longer time frame, such as several years. shows a chart of the data in the Excel file Federal Funds Rates.
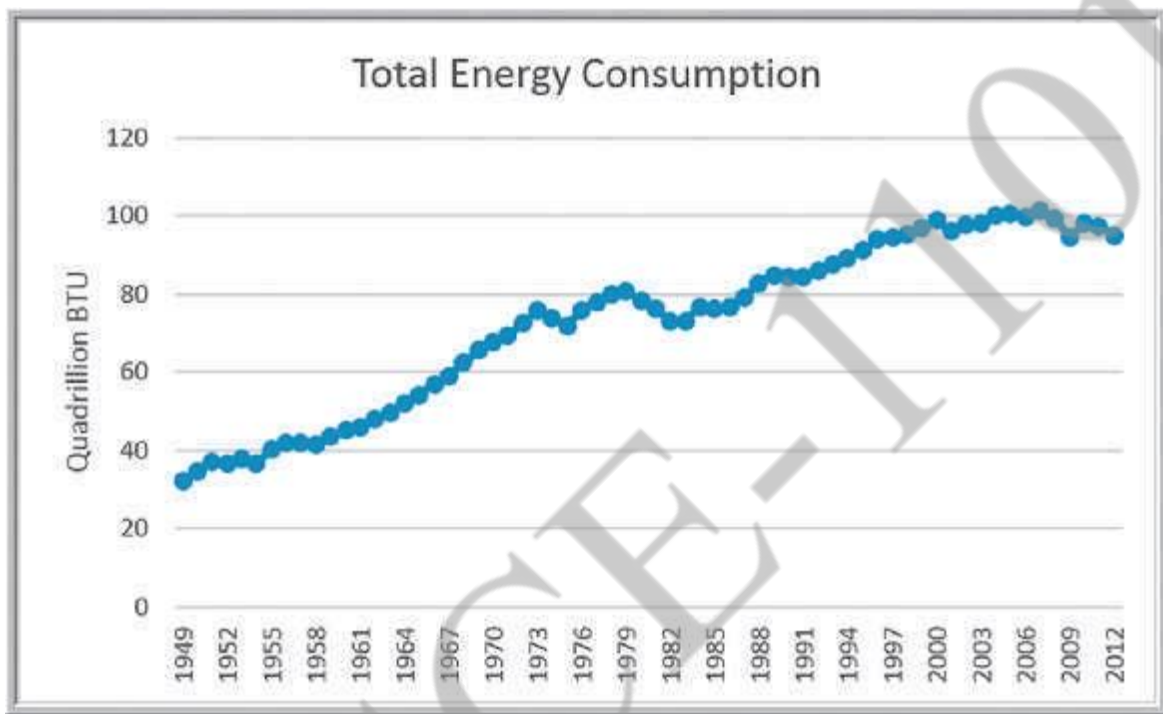
Total Energy Consumption Time Series
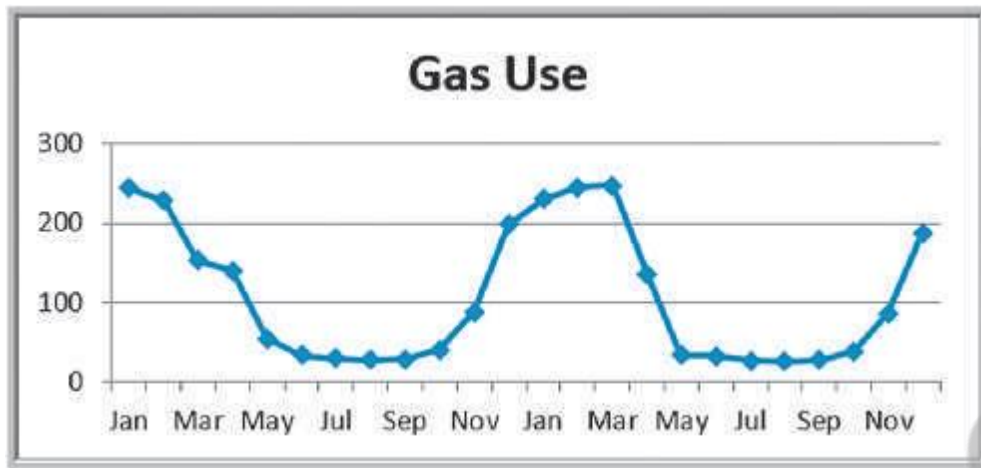
Seasonal Effects in Natural Gas Usage
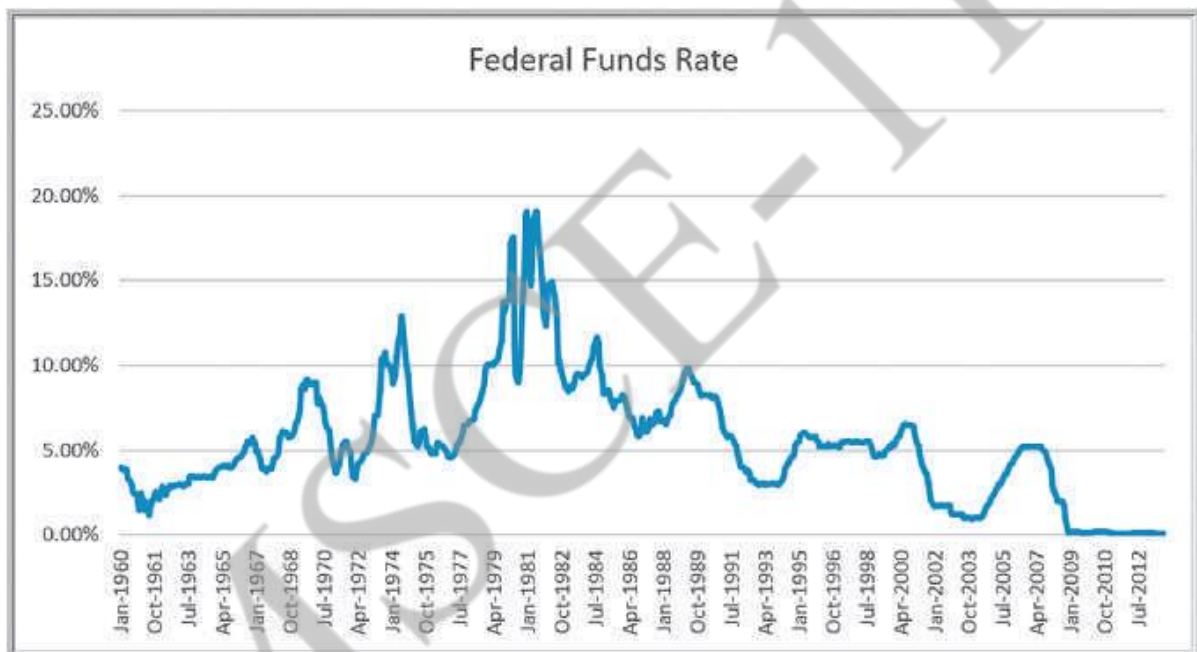
Cyclical Effects in Federal Fund Rates

1.3 Moving Average Models

simple moving average method is a smoothing method based on the idea of averaging random fluctuations in the time series to identify the underlying direction in which the time series is changing.

Error Metrics and Forecast Accuracy

The quality of a forecast depends on how accurate it is in predicting future values of a time series. In the simple moving average model, different values for k will produce different
forecasts.

To analyze the effectiveness of different forecasting models, we can define *error metrics*, which compare quantitatively the forecast with the actual observations. Three metrics that are commonly used are the *mean absolute deviation*, *mean square error*, and *mean absolute percentage error*.

1.Mean Absolute Deviation (MAD):

The **mean absolute deviation (MAD)** is the absolute difference between the actual value and the forecast, averaged over a range of forecasted values:

$$\text{MAD} = \frac{\sum\limits_{t=1}^{n}|A_t - F_t|}{n}$$

where A_t is the actual value of the time series at time t, F_t is the forecast value for time t, and n is the number of forecast values (*not* the number of data points since we do not have a forecast value associated with the first k data points). MAD provides a robust measure of error and is less affected by extreme observations.

2. **Mean square error (MSE):**

Mean square error (MSE) is probably the most commonly used error metric. It penalizes larger errors because squaring larger numbers has a greater impact than squaring smaller numbers. The formula for MSE is

$$\text{MSE} = \frac{\sum\limits_{t=1}^{n}(A_t - F_t)^2}{n}$$

n represents the number of forecast values used in computing the average.

3. **Root mean square error (RMSE):**

Sometimes the square root of MSE, called the **root mean square error (RMSE)**, is used. Note that unlike MSE, RMSE is expressed in the same units as the data (similar to the difference between a standard deviation and a variance), allowing for more practical comparisons.

$$\text{RMSE} = \sqrt{\frac{\sum\limits_{t=1}^{n}(A_t - F_t)^2}{n}}$$

4. **Mean absolute percentage error (MAPE)**:

MAPE is the average of absolute errors divided by actual observation values.

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|}{n} \times 100$$

The values of MAD and MSE depend on the measurement scale of the time-series data. For example, forecasting profit in the range of millions of dollars would result in very large MAD and MSE values, even for very accurate forecasting models. On the other hand, market share is measured in proportions; therefore, even bad forecasting models will have small values of MAD and MSE. Thus, these measures have no meaning except in comparison with other models used to forecast the same data. Generally, MAD is less affected by extreme observations and is preferable to MSE if such extreme observations
are considered rare events with no special meaning. MAPE is different in that the measurement scale is eliminated by dividing the absolute error by the time-series data value. This allows a better relative comparison. Although these comments provide some guidelines, there is no universal agreement on which measure is best.

1.4 Exponential Smoothing Models

Simple Exponential smoothing Model
A versatile, yet highly effective, approach for short-range forecasting is **simple exponential smoothing**. The basic simple exponential smoothing model is

$$F_{t+1} = (1 - \alpha)F_t + \alpha A_t$$
$$= F_t + \alpha(A_t - F_t)$$

where F_{t+1} is the forecast for time period $t + 1$, F_t is the forecast for period t, A_t is the observed value in period t, and a is a constant between 0 and 1 called the **smoothing constant**.

To begin, set $F1$ and $F2$ equal to the actual observation in period 1, $A1$. Using the two forms of the forecast equation just given, we can interpret the simple exponential smoothing model in two ways. In the first model, the forecast for the next period, F_{t+1}, is a weighted average of the forecast made for period t, Ft, and the actual observation in period t, At. The second form of the model, obtained by simply rearranging terms, states that the forecast for the next period, F_{t+1}, equals the forecast for the last period, Ft, plus a fraction a of the forecast error made in period t, $At - Ft$. Thus, to make a forecast once we have selected the smoothing constant, we need to know only the previous forecast and the actual value. By repeated substitution for Ft in the equation, it is easy to demonstrate that F_{t+1} is a decreasingly weighted average of all past time-series data. Thus, the forecast actually reflects *all* the data, provided that a is strictly between 0 and 1.

Double Exponential Smoothing
In double exponential smoothing, the estimates of a_t and b_t are obtained from the following
equations:

$$a_t = \alpha F_t + (1 - \alpha)(a_{t-1} + b_{t-1})$$
$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

In essence, we are smoothing both parameters of the linear trend model. From the first equation, the estimate of the level in period t is a weighted average of the observed value at time t and the predicted value at time t, $a_{t+1} + b_{t+1}$,based on simple exponential smoothing. For large values of a, more weight is placed on the observed value. Lower values of a put more weight on the smoothed predicted value. Similarly, from the second equation, the estimate of the trend in period t is a weighted average of the differences in the estimated levels in periods t and t - 1 and the estimate of the level in period t - 1.

Forecasting Time Series with Seasonality:

When time series exhibit seasonality, different techniques provide better forecasts,

- **Regression-Based Seasonal Forecasting Models**

    One approach is to use linear regression. Multiple linear regression models with categorical variables can be used for time series with seasonality.

- **Holt-Winters Forecasting for Seasonal Time Series**

    Holt-Winters models are similar to exponential smoothing models in that smoothing constants are used to smooth out variations in the level and seasonal patterns over time. For time series with seasonality but no trend, *XLMiner* supports a Holt-Winters method but does not have the ability to optimize the parameters

- **Holt-Winters Models for Forecasting Time Series with seasonality and Trend**

    Many time series exhibit both trend and seasonality. Such might be the case for growing sales of a seasonal product. These models combine elements of both the trend and seasonal models. Two types of Holt-Winters smoothing models are often used.

    The **Holt-Winters additive model** is based on the equation

$$F_{t+1} = a_t + b_t + S_{t-s+1}$$

and the **Holt-Winters multiplicative model** is

$$F_{t+1} = (a_t + b_t)S_{t-s+1}$$

The additive model applies to time series with relatively stable seasonality, whereas the multiplicative model applies to time series whose amplitude increases or decreases over time. Therefore, a chart of the time series should be viewed first to identify the appropriate type of model to use. Three parameters,$\propto$,β,γ, are used to smooth the level, trend,and seasonal factors in the time series. *XLMiner* supports both models.

Selecting Appropriate Time-Series-Based Forecasting Models

The table summarizes the choice of forecasting approaches that can be implemented by *XLMiner* based on characteristics of the time series.

	No Seasonality	Seasonality
No trend	Simple moving average or simple exponential smoothing	Holt-Winters no-trend smoothing model or multiple regression
Trend	Double exponential smoothing	Holt-Winters additive or Holt-Winters multiplicative model

Regression Forecasting with Causal Variables

In many forecasting applications, other independent variables besides time, such as economic indexes or demographic factors, may influence the time series. For example, a manufacturer of hospital equipment might include such variables as hospital capital spending and changes in the proportion of people over the age of 65 in building models to forecast future sales. Explanatory/causal models, often called **econometric models**, seek to identify factors that explain statistically the patterns observed in the variable being forecast, usually with regression analysis.

The Practice of Forecasting

Surveys of forecasting practices have shown that both judgmental and quantitative methods are used for forecasting sales of product lines or product families as well as for broad company and industry forecasts. Simple time-series models are used for short- and medium-range forecasts, whereas regression analysis is the most popular method for long range forecasting. However, many companies rely on judgmental methods far more than quantitative methods, and almost half judgmentally adjust quantitative forecasts.

In practice, managers use a variety of judgmental and quantitative forecasting techniques.
Statistical methods alone cannot account for such factors as sales promotions, unusual environmental disturbances, new product introductions, large one-time orders, and
so on. Many managers begin with a statistical forecast and adjust it to account for intangible factors. Others may develop independent judgmental and statistical forecasts and then combine them, either objectively by averaging or in a subjective manner.

It is important to compare quantitatively generated forecasts to judgmental forecasts to see if the forecasting method is adding value in terms of an improved forecast. It is impossible to provide universal guidance as to which approaches are best, because they depend on a variety of factors, including the presence or absence of trends and seasonality, the number of data points available, length of the forecast time horizon, and the experience and knowledge of the forecaster. Often, quantitative approaches will miss significant changes in the data, such as reversal of trends, whereas qualitative forecasts may catch them, particularly when using indicators.

2.Logic and Data Driven Models

Predictive modeling means the developing models that can be used to forecast or predict future events. Models can be developed either through logic or data.
Logic driven models remain based on experience, knowledge and logical relationships of variables and constants connected to the desired business performance outcome situation.
Data-driven Models refers to the models in which data is collected from many sources to qualitatively establish model relationships. Logic driven models is often used as a first step to establish relationships through data-driven models. Data driven models include sampling and estimation, regression analysis, correlation analysis, forecasting models and stimulation.

3. Mining and Predictive Analysis Modelling:

Data mining is a rapidly growing field of business analytics that is focused on better understanding characteristics and patterns among variables in large databases using a

variety of statistical and analytical tools. Many of the tools that we have studied in previous chapters, such as data visualization, data summarization, PivotTables, correlation and regression analysis, and other techniques, are used extensively in data mining. However, as the amount of data has grown exponentially, many other statistical and analytical methods have been developed to identify relationships among variables in large data sets and understand hidden patterns that they may contain

Some common approaches in data mining include the following

- ***Data Exploration and Reduction***.
  This often involves identifying groups in which the elements of the groups are in some way similar. This approach is often used to understand differences among customers and segment them into homogenous groups. For example, Macy's department stores identified four lifestyles of its customers: "Katherine," a traditional, classic dresser who doesn't take a lot of risks and likes quality; "Julie," neotraditional and slightly more edgy but still classic; "Erin," a contemporary customer who loves newness and shops by brand; and "Alex," the fashion customer who wants only the latest and greatest (they have male versions also).4 Such segmentation is useful in design and marketing activities to better target product offerings. These techniques have also been used to identify characteristics of successful employees and improve recruiting and hiring practices.
- ***Classification***. Classification is the process of analyzing data to predict how to classify a new data element. An example of classification is spam filtering in an e-mail client. By examining textual characteristics of a message (subject header, key words, and so on), the message is classified as junk or not. Classification methods can help predict whether a credit-card transaction may be fraudulent, whether a loan applicant is high risk, or whether a consumer will respond to an advertisement.
- ***Association***. Association is the process of analyzing databases to identify natural associations among variables and create rules for target marketing or buying recommendations.
  For example, Netflix uses association to understand what types of movies a customer likes and provides recommendations based on the data.
  Amazon.com also makes recommendations based on past purchases. Supermarket loyalty cards collect data on customers' purchasing habits and print coupons at the point of purchase based on what was currently bought.
  - ***Cause-and-effect modeling***. Cause-and-effect modeling is the process of developing analytic models to describe the relationship between metrics that drive business performance—for instance, profitability, customer satisfaction, or employee satisfaction. Understanding the drivers of performance can lead to better decisions to improve performance. For example, the controls group of Johnson Controls, Inc., examined the relationship between satisfaction and contract-renewal rates. They found that 91% of contract renewals came from customers who were either satisfied or very satisfied, and customers who were not satisfied had a much higher defection rate. Their model predicted that a one-percentage-point increase in the overall satisfaction score was worth $13 million in service contract renewals annually. As a result, they identified decisions

that would improve customer satisfaction.Regression and correlation analysis are key tools for cause-and-effect modelling.

3.1 Predictive Modeling

Predictive modeling is a commonly used statistical technique to predict future behavior. Predictive modeling solutions are a form of data-mining technology that works by analyzing historical and current data and generating a model to help predict future outcomes.

In predictive modeling, data is collected, a statistical model is formulated, predictions are made, and the model is validated (or revised) as additional data becomes available.

For example, risk models can be created to combine member information in complex ways with demographic and lifestyle information from external sources to improve underwriting accuracy. Predictive models analyze past performance to assess how likely a customer is to exhibit a specific behavior in the future. This category also encompasses models that seek out subtle data patterns to answer questions about customer performance, such as fraud detection models. Predictive models often perform calculations during live transactions—for example, to evaluate the risk or opportunity of a given customer or transaction to guide a decision. If health insurers could accurately predict secular trends (for example, utilization), premiums would be set appropriately, profit targets would be met with more consistency, and health insurers would be more competitive in the marketplace.

Predictive modeling is a method of predicting future outcomes by using data modeling. It's one of the premier ways a business can see its path forward and make plans accordingly. While not fool proof, this method tends to have high accuracy rates, which is why it is so commonly used. Predictive modelling uses statistics to predict outcomes. Most often the event one wants to predict is in the future, but predictive modelling can be applied to any type of unknown event, regardless of when it occurred.   For example, predictive models are often used to detect crimes and identify suspects, after the crime has taken place.

In many cases the model is chosen on the basis of detection theory to try to guess the probability of an outcome given a set amount of input data, for example given an email determining how likely that it is spam. Models can use one or more classifiers in trying to determine the probability of a set of data belonging to another set.

For example, a model might be used to determine whether an email is spam or "ham" (non-spam). Depending on definitional boundaries, predictive modelling is synonymous with, or largely overlapping with, the field of machine learning, as it is more commonly referred to in academic or research and development contexts. When deployed commercially, predictive modelling is often referred to as predictive analytics. Predictive modelling is often contrasted with causal modelling/analysis. In the former, one may be entirely satisfied to make use of indicators of, or proxies for, the outcome of interest. In the latter, one seeks to determine true cause-and-effect relationships. This distinction has given rise to a burgeoning literature in the fields of research methods and statistics and to the common statement that "correlation does not imply causation".

3.2 What Is Predictive Modeling?

In short, predictive modeling is a statistical technique using machine learning and data mining to predict and forecast likely future outcomes with the aid of historical and existing data. It works by analyzing current and historical data and projecting what it learns on a model generated to forecast likely outcomes.

Predictive modeling can be used to predict just about anything, from TV ratings and a customer's next purchase to credit risks and corporate earnings. A predictive model is not fixed; it is validated or revised regularly to incorporate changes in the underlying data. In other words, it's not a one-and-done prediction. Predictive models make assumptions based on what has happened in the past and what is happening now.

If incoming, new data shows changes in what is happening now, the impact on the likely future outcome must be recalculated, too. For example, a software company could model historical sales data against marketing expenditures across multiple regions to create a model for future revenue based on the impact of the marketing spend. Most predictive models work fast and often complete their calculations in real time. That's why banks and retailers can, for example, calculate the risk of an online mortgage or credit card application and accept or decline the request almost instantly based on that prediction.

Some predictive models are more complex, such as those used in computational biology and quantum computing; the resulting outputs take longer to compute than a credit card application but are done much more quickly than was possible in the past thanks to advances in technological capabilities, including computing power.

3.3 Top 5 Types of Predictive Models

Fortunately, predictive models don't have to be created from scratch for every application. Predictive analytics tools use a variety of vetted models and algorithms that can be applied to a wide spread of use cases.

Predictive modeling techniques have been perfected over time. As we add more data, more muscular computing, AI and machine learning and see overall advancements in analytics, we're able to do more with these models.

The top five predictive analytics models are:

1. **Classification model**:
   Considered the simplest model, it categorizes data for simple and direct query response. An example use case would be to answer the question "Is this a fraudulent transaction?"

2. **Clustering model**:
   This model nests data together by common attributes. It works by grouping things or people with shared characteristics or behaviors and plans strategies for each group at a larger scale. An example is in determining credit risk for a loan applicant based on what other people in the same or a similar situation did in the past.

3. **Forecast model**:
   This is a very popular model, and it works on anything with a numerical value based on learning from historical data. For example, in answering how much lettuce a restaurant should order next week or how many calls a customer support agent should be able to handle per day or week, the system looks back to historical data.

4. **Outliers model**:
   This model works by analyzing abnormal or outlying data points. For example, a bank might use an outlier model to identify fraud by asking whether a transaction is outside of the customer's normal buying habits or whether an expense in a given category is normal or not. For example, a $1,000 credit card charge for a washer and dryer in the cardholder's preferred big box store would not be alarming, but $1,000 spent on designer clothing in a location where the customer has never charged other items might be indicative of a breached account.

5. **Time series model:**

This model evaluates a sequence of data points based on time. For example, the number of stroke patients admitted to the hospital in the last four months is used to predict how many patients the hospital might expect to admit next week, next month or the rest of the year. A single metric measured and compared over time is thus more meaningful than a simple average.

3.4 Predictive Algorithms:

 Some of the more common predictive algorithms are:

1. Random Forest: This algorithm is derived from a combination of decision trees, none of which are related, and can use both classification and regression to classify vast amounts of data.

 2. Generalized Linear Model (GLM) for Two Values: This algorithm narrows down the list of variables to find "best fit." It can work out tipping points and change data capture and other influences, such as categorical predictors, to determine the "best fit" outcome, thereby overcoming drawbacks in other models, such as a regular linear regression.

3. Gradient Boosted Model: This algorithm also uses several combined decision trees, but unlike Random Forest, the trees are related. It builds out one tree at a time, thus enabling the next tree to correct flaws in the previous tree. It's often used in rankings, such as on search engine outputs.

 4. K-Means: A popular and fast algorithm, K-Means groups data points by similarities and so is often used for the clustering model. It can quickly render things like personalized retail offers to individuals within a huge group, such as a million or more customers with a similar liking of lined red wool coats.

 5. Prophet: This algorithm is used in time-series or forecast models for capacity planning, such as for inventory needs, sales quotas and resource allocations. It is highly flexible and can easily accommodate heuristics and an array of useful assumptions.

Predictive modeling is often performed using curve and surface fitting, time series regression, or machine learning approaches. Regardless of the approach used, the process of creating a predictive model is the same across methods.

3.5 Steps for Predictive Modeling:
 The steps are:
 1. Clean the data by removing outliers and treating missing data.
 2. Identify a parametric or nonparametric predictive modeling approach to use.
 3. Preprocess the data into a form suitable for the chosen modeling algorithm.
 4. Specify a subset of the data to be used for training the model.
 5. Train, or estimate, model parameters from the training data set.
 6. Conduct model performance or goodness-of-fit tests to check model adequacy.
 7. Validate predictive modeling accuracy on data not used for calibrating the model.
 8. Use the model for prediction if satisfied with its performance .

4.Machine Learning for Predictive Analytics

Machine learning is defined as an automated process that extracts patterns from data. To build the models used in predictive data analytics applications, we use supervised machine learning. Supervised machine learning techniques automatically learn a model of the relationship between a set of descriptive features and a target feature based on a set of historical examples, or instances. We can then use this model to make predictions for new instances. These two separate steps are shown in figure,
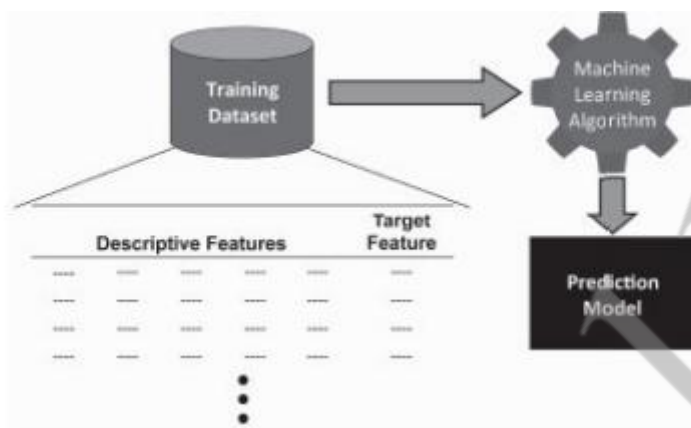
(a) Learning a model from a set of historical instances

(b) Using a model to make predictions

The two steps in supervised machine learning. Table1.1 lists a set of historical instances, or dataset, of mortgages that a bank has granted in the past. This dataset includes descriptive features that describe the mortgage, and a target feature that indicates whether the mortgage applicant ultimately defaulted on the loan or paid it back in full. The descriptive features tell us three pieces of information about the mortgage: the OCCUPATION (which can be professional or industrial) and AGE of the applicant and the ratio between the applicant's salary and the amount borrowed (LOANSALARY RATIO). The target feature, OUTCOME, is set to either default or repay. In machine learning terms, each row in the dataset is referred to as a training instance, and the overall dataset is referred to as a training data sets.

Table 1.1

A credit scoring dataset.

ID	OCCUPATION	AGE	LOAN-SALARY RATIO	OUTCOME
1	industrial	34	2.96	repay
2	professional	41	4.64	default
3	professional	36	3.22	default
4	professional	41	3.11	default
5	industrial	48	3.80	default
6	industrial	61	2.52	repay
7	professional	37	1.50	repay
8	professional	40	1.93	repay
9	industrial	33	5.25	default
10	industrial	32	4.15	default

An example of a very simple prediction model for this domain would be

if LOAN-SALARY RATIO > 3 then

OUTCOME = default

else

OUTCOME = repay

We can say that this model is consistent with the dataset as there are no instances in the dataset for which the model does not make a correct prediction. When new mortgage applications are made, we can use this model to predict whether the applicant will repay the mortgage or default on it and make lending decisions based on this prediction.

Machine learning algorithms automate the process of learning a model that captures the relationship between the descriptive features and the target feature in a dataset. For simple datasets like the one in Table , we may be able to manually create a prediction model, and in an example of this scale, machine learning has little to offer us.

Consider, however, the dataset in Table, which shows a more complete representation of the same problem. This dataset lists more instances, and there are extra descriptive features describing the AMOUNT that a mortgage holder borrows, the mortgage holder's SALARY, the type of PROPERTY that the mortgage relates to (which can be farm,house, or apartment) and the TYPE of mortgage (which can be ftp for first-time buyers or stb for second-time buyers).

The simple prediction model using only the loan-salary ratio feature is no longer consistent with the dataset. It turns out, however, that there is at least one prediction model that is consistent with the dataset; it is just a little harder to find than the previous one:

if LOAN-SALARY RATIO < 1.5 then

OUTCOME = repay

else if LOAN-SALARY RATIO > 4 then

OUTCOME = default

else if AGE < 40 and OCCUPATION =industrial then

OUTCOME = default

else

OUTCOME = repay

To manually learn this model by examining the data is almost impossible. For a machine learning algorithm, however, this is simple. When we want to build prediction models from large datasets with multiple features, machine learning is the solution.

4.1 How does Machine Learning Work?

Machine learning algorithms work by searching through a set of possible prediction models for the model that best captures the relationship between the descriptive features and target feature in a dataset. An obvious criteria for driving this search is to look for models that are consistent with the data.

There are, however, at least two reasons why just searching for consistent models is not sufficient in order to learn useful prediction models.

 First, when we are dealing with large datasets, it is likely that there will be noise in the data, and prediction models that are consistent with noisy data will make incorrect predictions.

Second, in the vast majority of machine learning projects, the training set represents only a small sample of the possible set of instances in the domain. As a result, machine learning is an ill-posed problem. An ill-posed problem is a problem for which a unique solution cannot be determined using only the information that is available.

Table 1.2

A more complex credit scoring dataset.

ID	AMOUNT	SALARY	LOAN-SALARY RATIO	AGE	OCCUPATION	PROPERTY	TYPE	OUTCOME
1	245,100	66,400	3.69	44	industrial	farm	stb	repay
2	90,600	75,300	1.20	41	industrial	farm	stb	repay
3	195,600	52,100	3.75	37	industrial	farm	ftb	default
4	157,800	67,600	2.33	44	industrial	apartment	ftb	repay
5	150,800	35,800	4.21	39	professional	apartment	stb	default
6	133,000	45,300	2.94	29	industrial	farm	ftb	default
7	193,100	73,200	2.64	38	professional	house	ftb	repay
8	215,000	77,600	2.77	17	professional	farm	ftb	repay
9	83,000	62,500	1.33	30	professional	house	ftb	repay
10	186,100	49,200	3.78	30	industrial	house	ftb	default
11	161,500	53,300	3.03	28	professional	apartment	stb	repay
12	157,400	63,900	2.46	30	professional	farm	stb	repay
13	210,000	54,200	3.87	43	professional	apartment	ftb	repay
14	209,700	53,000	3.96	39	industrial	farm	ftb	default
15	143,200	65,300	2.19	32	industrial	apartment	ftb	default
16	203,000	64,400	3.15	44	industrial	farm	ftb	repay
17	247,800	63,800	3.88	46	industrial	house	stb	repay
18	162,700	77,400	2.10	37	professional	house	ftb	repay
19	213,300	61,100	3.49	21	industrial	apartment	ftb	default
20	284,100	32,300	8.80	51	industrial	farm	ftb	default
21	154,000	48,900	3.15	49	professional	house	stb	repay
22	112,800	79,700	1.42	41	professional	house	ftb	repay
23	252,000	59,700	4.22	27	professional	house	stb	default
24	175,200	39,900	4.39	37	professional	apartment	stb	default
25	149,700	58,600	2.55	35	industrial	farm	stb	default

Table 1.3

A simple retail dataset

ID	BBY	ALC	ORG	GRP
1	no	no	no	couple
2	yes	no	yes	family
3	yes	yes	no	family
4	no	no	yes	couple
5	no	yes	yes	single

We can illustrate how machine learning is an ill-posed problem using an example in which the analytics team at a supermarket chain wants to be able to classify customer households into the demographic groups single, couple, or family, based solely on their shopping habits.

The dataset in Table 1.3 contains descriptive features describing the shopping habits of 5 customers. The descriptive features measure whether a customer buys baby food, BBY, alcohol, ALC, or organic vegetable products, ORG. Each feature can take one of the two values: yes or no. Alongside these descriptive features is a target feature, GRP, that describes the demographic group for each customer (single, couple, or family). The dataset in Table 1.3 is referred to as a labeled dataset because it includes values for the target feature.

Imagine we attempt to learn a prediction model for this retail scenario by searching for a model that is consistent with the dataset. The first thing we need to do is figure out many different possible models actually exist for the scenario. This defines the set of prediction models the machine learning algorithm will search. From the perspective of searching for a consistent model, the most important property of a prediction model is that it defines a mapping from every possible combination of descriptive feature values to a prediction for the target feature. For the retail scenario, there are only three binary descriptive features, so there are $2^3 = 8$ possible combinations of descriptive feature values. However, for each of these 8 possible descriptive feature value combinations, there are 3 possible target feature values, so this means that there are $3^8 = 6,561$ possible prediction models that could be used. Table illustrates the relationship between descriptive feature value combinations and prediction models for the retail scenario. The descriptive feature combinations are listed on the left hand side of the table and the set of potential models for this domain are shown as 1 to 6,561 on the right hand side of the table. Using the training dataset from Table1.3 , a machine learning algorithm will reduce the full set of 6,561 possible prediction models for this scenario down to just those that

are consistent with the training instances. Table 1.4(b) illustrates this; the blanked out columns in the table indicate the models that are not consistent with the training data.

Table 1.4

Potential prediction models (a) before and (b) after training data becomes available.

BBY	ALC	ORG	GRP	M$_1$	M$_2$	M$_3$	M$_4$	M$_5$	...	M$_{6561}$
no	no	no	?	couple	couple	single	couple	couple	...	couple
no	no	yes	?	single	couple	single	couple	couple	...	single
no	yes	no	?	family	family	single	single	single	...	family
no	yes	yes	?	single	single	single	single	single	...	couple
yes	no	no	?	couple	couple	family	family	family	...	family
yes	no	yes	?	couple	family	family	family	family	...	couple
yes	yes	no	?	single	family	family	family	family	...	single
yes	yes	yes	?	single	single	family	family	couple	...	family

(b)

BBY	ALC	ORG	GRP	M$_2$	M$_4$	M$_5$	...
no	no	no	couple	couple	couple	couple	...
no	no	yes	couple	couple	couple	couple	...
no	yes	no	?	family	single	single	...
no	yes	yes	single	single	single	single	...
yes	no	no	?	couple	family	family	...
yes	no	yes	family	family	family	family	...
yes	yes	no	family	family	family	family	...
yes	yes	yes	?	single	family	couple	...

Table1.4(b) also illustrates the fact that the training dataset does not contain an instance for every possible descriptive feature value combination and that there are still a large number of potential prediction models that remain consistent with the training dataset after the inconsistent models have been excluded. Specifically, there are three remaining descriptive feature value combinations for which the correct target feature value is not known, and therefore there are $3^3 = 27$ potential models that remain consistent with the training data. Three of these- M$_2$,M$_4$,M$_5$- shown in Table1.4(b). Because a single consistent model cannot be found based on the sample training dataset alone, we say that machine learning is fundamentally an ill-posed problem.

We might be tempted to think that having multiple models that are consistent with the data is a good thing. The problem is, however, that although these models agree on what predictions should be made for the instances in the training dataset, they disagree with regard to what predictions should be returned for instances that are not in the training dataset. For example, if a new customer starts shopping at the supermarket and buys baby food, alcohol, and organic vegetables, our set of consistent models will contradict each other with respect to what prediction should be returned for this customer, for example, M$_2$ will return GRP = single, M$_4$ will return GRP = family, and M$_5$ will return GRP = couple.

The criterion of consistency with the training data doesn't provide any guidance with regard to which of the consistent models to prefer when dealing with queries that are outside the training dataset. As a result, we cannot use the set of consistent models to make predictions for these queries. In fact, searching for predictive models that are consistent with the dataset is equivalent to just memorizing the dataset. As a result, no learning is taking place because the set of consistent models tells us nothing about the underlying relationship between the descriptive and target features beyond what a simple look-up of the training dataset would provide.

If a predictive model is to be useful, it must be able to make predictions for queries that are not present in the data. A prediction model that makes the correct predictions for these queries captures the underlying relationship between the descriptive and target features and is said to generalize well. Indeed, the goal of machine learning is to find the predictive model that generalizes best. In order to find this single best model, a machine learning algorithm must use some criteria for choosing among the candidate models it considers during its search.

Given that consistency with the dataset is not an adequate criterion to select the best prediction model, what criteria should we use? There are a lot of potential answers to this question, and that is why there are a lot of different machine learning algorithms. Each machine learning algorithm uses different model selection criteria to drive its search for the best predictive model. So, when we choose to use one machine learning algorithm instead of another, we are, in effect, choosing to use one model selection criterion instead of another.

All the different model selection criteria consist of a set of assumptions about the characteristics of the model that we would like the algorithm to induce. The set of assumptions that defines the model selection criteria of a machine learning algorithm is known as the inductive bias 6 of the machine learning algorithm.

There are two types of inductive bias that a machine learning algorithm can use, a restriction bias and a preference bias. A restriction bias constrains the set of models that the algorithm will consider during the learning process. A preference bias guides the learning algorithm to prefer certain models over others.

For example, we introduce a machine learning algorithm called multivariable linear regression with gradient descent, which implements the restriction bias of only considering prediction models that produce predictions based on a linear combination of the descriptive feature values and applies a preference bias over the order of the linear models it considers in terms of a gradient descent approach through a weight space. As a second example, we introduce the Iterative Dichotomizer 3 (ID3) machine learning algorithm, which uses a restriction bias of only considering tree prediction models where each branch encodes a sequence of checks on individual descriptive features but also utilizes a preference bias by considering shallower (less complex) trees over larger trees. It is important to recognize that using an inductive bias is a necessary prerequisite for

learning to occur; without inductive bias, a machine learning algorithm cannot learn anything beyond what is in the data.

In summary, machine learning works by searching through a set of potential models to find the prediction model that best generalizes beyond the dataset. Machine learning algorithms use two sources of information to guide this search, the training dataset and the inductive bias assumed by the algorithm.

HR & SUPPLY CHAIN ANALYTICS

Human Resources – Planning and Recruitment – Training and Development - Supply chain network - Planning Demand, Inventory and Supply – Logistics – Analytics applications in HR & Supply Chain

1. Introduction to Human Resource:

Every business is made up of people : its human resources. An organisation is nothing without human resources. Human resource management (HRM) is about managing these people effectively. It is aimed at achieving business objectives through the best use of an organisation's human resources. Effective management of human resources is vital in all types and sizes of organisations.

In fact, how effectively the human resources managed will have a major impact on how successful the business becomes. It is universally agreed that the quality of the human resources is the major factor in maintaining the competitiveness and profitability of the today's business.

1.1 What is HRM?

- Human Resource Management (HRM) involves all management decisions and practices that directly affect or influence the human resources, who work for the organisations.
- HRM is the set of organisational activities directed at attracting, developing, rewarding and maintaining an effective work force.
- **HRM Vs Personnel Management:**
  In general, the terms ' human resource management (HRM) and personnel management are used interchangeably.
- Proponents of HRM argue that it is different from personnel management.According to them, HRM incorporates practices developrd by practitioners of people management.
  That is ,
  Human Resource managers= Specialists in Personnel Management + Generalists in line and senior management.
- Others hold that both HRM and PM are little different and overlap in their techniques and range of interest.

Defining HRM:

(i)   **Effective HRM** benefits the individual, society, and the company.
(ii) Companies use HRM activities to manage their Human Resources.
(iii) The efficiency with which any organisation can be operated will largely depend upon how effectively its human resources are managed and utilized.
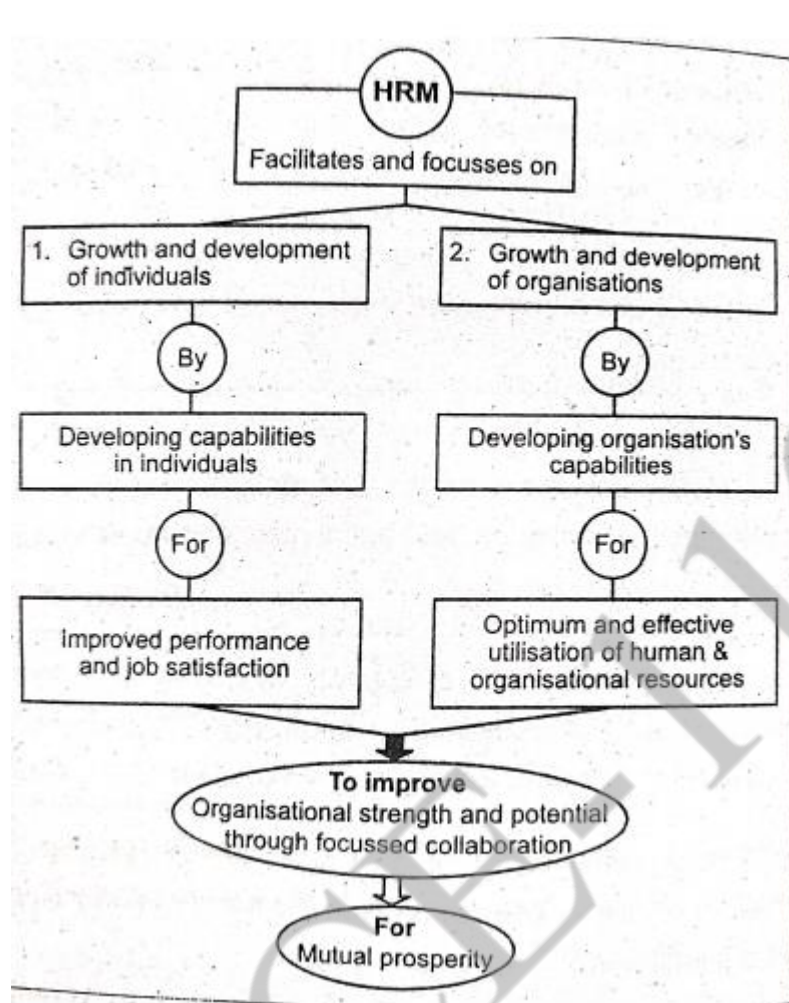
Figure 1.1 Essence of HRM

1.2 OBJECTIVES OF HRM:

Objectives are benchmarks against which actions are evaluated a broad objective of HRM is to optimise the usefulness (ie., productivity) of all workers in an organisation.However, there are four types of objectives that are common to Human Resource Management. They are:

1.Societal Objectives
2. Organisational Objectives
3. Personal Objectives
4. Labour Union Objectives

1. Societal Objectives:

Since an organisation is part of the society the main objective of HRM is to be responsive to the needs and challenges of society.

HRM's societal objectives include:

i)   To provide more employment oppurtunities.
ii)   To provide maximum productivity.
iii)   To provide material and mental satisfaction to workforce.
iv)   To control the wastage of effort.

v) To help help to maintain ethical policies and socially responsible behaviour.

vi) To encourage healthy human relations and social welfare.

vii) To manage change to the mutual advantage of individuals, groups ,the enterprise and the public

2. **Organisational Objectives** :

These objectives of HRM are based on the fact that human resource management exists to contribute to organisational effectiveness.

 HRM's organisational objectives include:

i) To help the organisation to reach its goals.

ii) To efficiently employ the skills and abilities of the workforce.

iii) To provide well-trained and well-motivated employees to the organisation.

iv) To develop and maintain a quality of work life that makes employment in the organisation desirable.

v) To communicate HRM policies to all employees.

3. **Personal (or employees) Objectives :**

The another important objective of HRM is to assist employees in achieving their personal goals.

HRM's Personal objective include:

i) To provide adequate remuneration to the employees.

ii) To provide job security

iii) To provide Facilities for proper Training and Development.

iv) To increase the employees's job satisfaction and self-actualisation.

v) To provide congenial working environment.

4. **Labour Union Objectives :**

The HRM is also concerned with labour unions and related issues.

 HRM's labour union related objectives include:

i) To recognise the labour unions.

ii) To establish the personnel policies in consultation with unions.

iii) To create congenial atmosphere with unions so as to maintain the spirit of self-discipline and co-operation with the management.

1.3 HUMAN RESOURCE MANAGEMENT MODEL

The figure 1.2 is a HRM model that illustrates how HRM activities come to bear on an organisation's environment, employees,jobs, job outcomes and organisational outcomes.As shown in the figure all of these forces are in turn affected by the organisation's external environment.
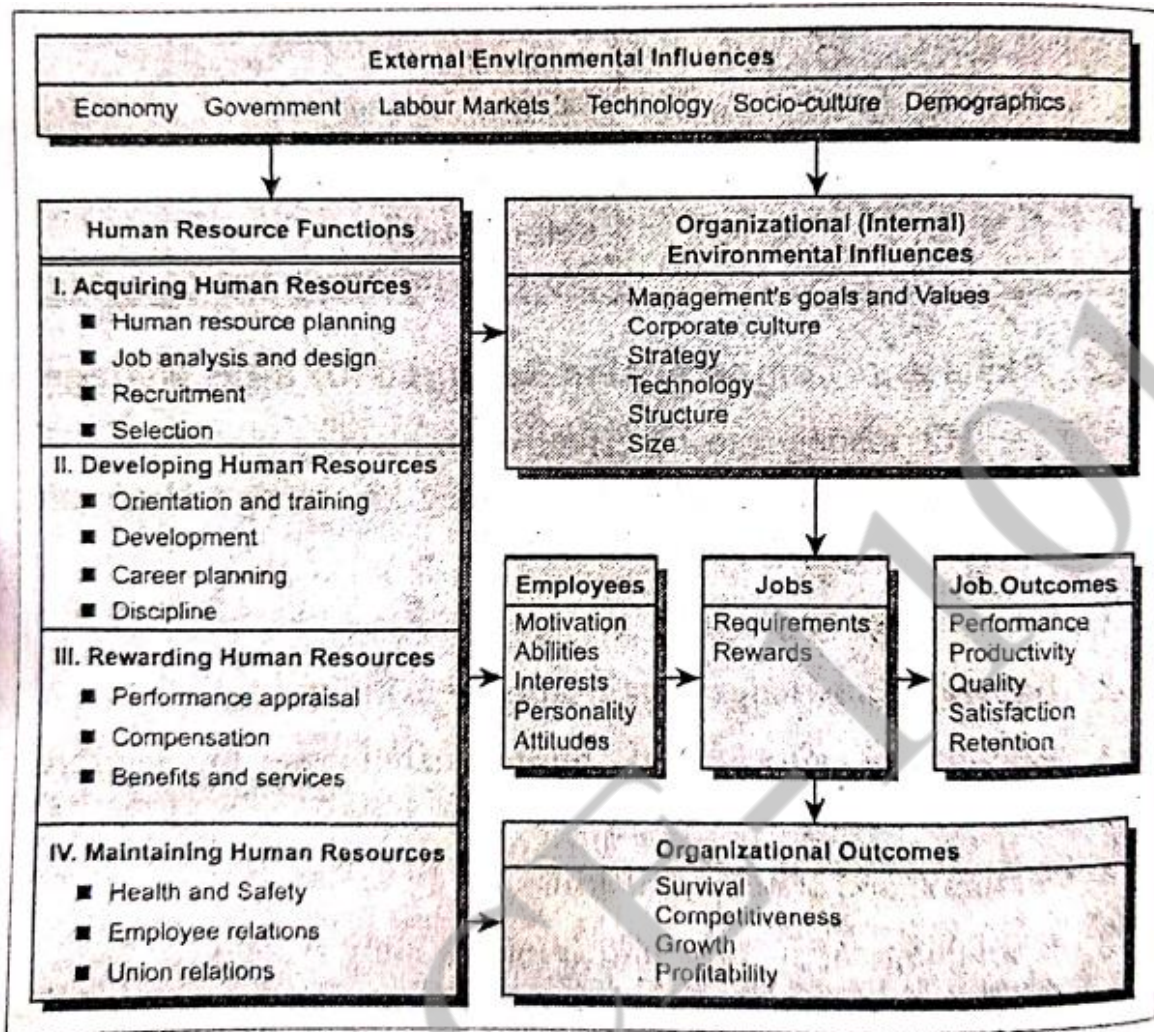
Fig 1.2 ; Human Resource Management Model

The model shown in the figure is also called as a diagnostic approach to HRM. Here the term diagnostic approach means that, in making human resource decisions, the HR manager must consider the employees, the jobs, the organisation (i.e., the internal environment), the external environment and the desired results.

Basic components of HRM:

From the figure 1.2 it may be noted that there are four basic components of HRM, each of which has general dimensions. The components are:

1. HRM activities/functions,
2. HRM outcomes,
3. Organisational (i.e, internal) environmental influences and
4. External environmental influences.

1. **HRM Activities/Functions:**
   a)Organisational Planning and Development:

- ✓ Determination of needs of the organisation based on long and short term objectives, technology selected, product feature and external environment.
- ✓ Design of organisational structure.
- ✓ Establishing a healthy organisational climate of mutual co-operation, trust and confidence.

b) Strategic Human Resource Planning
- ✓ Assessing current human resources.
- ✓ Assessing future human resources needs
- ✓ Developing a program to meet the future needs.

c) Job Analysis
- ✓ It is an assessment that defines jobs and the behaviour necessary to perform them.
- ✓ Preparation of Job descriptions and job specifications.

d) Staffing
- ✓ It concerns the recruitment and selection of human resources for an organisation.
- ✓ It includes: Man power planning, Recruitment, Selection and Placement,Induction,Promotion and Transfer and Seperation

e) Training and Development
It includes:

1. Orientation of new employees
2. Training of employees to perform their jobs.
3. Retraining of employees as their job requirements change .
4. Encouraging the development and growth of employees.

f) Performance Appraisal:
- ✓ It assesses how well employees are doing their jobs.
- ✓ Appraisals are useful:
  i) In making compensation decisions
- ✓ ii) In specifying areas in which additional development of employees is needed.
  Iii)In making Placement decisions.

g) Compensation Benefits:
- ✓ Compensation rewards people through pay, incentives and benefits for performing work within the organisation.
- ✓ Organisations must develop and refine their basic wage and salary to ensure that pay-for-performance policies are followed.

h) **Health and Safety:**

- ✓ Organisations should be more responsive to the concerns about the physical and mental health and safety of employees.
- ✓ Organisation should provide safer and healthy workshop conditions for employees.

 i)**Employee Relations:**

- ✓ The formal relationship between employees and their employers must be managed for the benefit of both.
- ✓ To facilitate good employee relations, it is important to develop and communicate HR policies and rules.

j) **Union Relations:**

- ✓ Union-related activities are important because they affect employees, managers, and the performance of many HR activities.
- ✓ At the formal organisation level, the union is the agent representing a group of employees in an organisation.
- ✓ The other activities of union include collective bargaining and grievance management.

k) **HR Information and Assessment Systems:**

- ✓ Information, communication and research systems are vital to the coordination of HR activities.
- ✓ Creating and maintaining HR Database and Systems are critical aspects of the strategic role of HR Management.
- ✓ Measuring HR effectiveness is done by evaluating how well HR activities are being performed in an organisation.

2.HRM Outcomes:

 The right hand side of the figure 1.2 indicates several outcomes that HR activities attempt to influence.

- ✓ HRM outcomes include:
  - i) Job Outcomes: Performance,Productivity, quality, satisfaction and retention.
  - ii) Organisational Outcomes: Survival, Competitiveness, Growth, ,Profitabilty.

3.Organisational (i.e internal) Environmental Influences

 The figure shows that forces inside the organisation affect the HRM activities. Some of the key Internal environmental factors that influence the HRM activities include:

- **i)** **Top management's goals and values**
- **ii)** **Corporate Culture**

iii) **Strategy**
iv) **Technology**
v) **Structure**
vi) **Size**

5.External Environmental Influences

- ✓ Organisations are surrounded by an external environment filled with many variable factors, as shown on the top of the figure.The forces outside the organisation greatly influence and restrict the organisation's HRM activities.
- ✓ The major external environmental factors that influence the HRM activities are:
  1. Economic Conditions
  2. Government Requirements
  3. Labour market conditions and union expectations.
  4. Technological Influences.
  5. Socio-cultural factors
  6. Demographic and competitiveness conditions.

  1. **Economic Conditions:**
     - ✓ Changing economic conditions directly influence the operation of any organisation and indirectly influence human resource actions.
     - ✓ A manager's decision to hire additional people, to lay off current employees and even how much to pay each job are all examples of HR decisions that are influenced by economic conditions.
     - ✓ Under Favorable economic conditions expansion of existing programs and creation of new programs are very likely.
     - ✓ Whereas with less favourable or deteriorating conditions, contraction or cancellation of programs maybe necessary.

  2. **Government (legal/political) Requirements:**
     - ✓ Government through the enforcement of laws has a direct and immediate impact on the human resource function.
     - ✓ Thus the laws and regulations of the central, state, and the local governments which are directed at HR issues influence and restrict objectives, strategies and HR actions.

  3. **Labour Market Conditions and Union Expectations:**
     - ✓ Changing conditions in the labour market,shortages of certain skilled workers and surpluses of others, changing market and expectations of people in the labour force influence the organisation's HR activities.
     - ✓ If the organisation is organised by the trade union, its expectations will restrict and influence how the organisation operates and what objectives it seeks.

  4. **Technological Influences:**
     - ✓ Technology influences HRM in two general ways.One way is for technology to change entire industries. Automation is the other way technology affects HRM.

- ✓ Another factor is increasing computerisation of major organisational functions.
- ✓ Thus the technological factor affects both positively and negatively the human resource activities.

5. Socio- Cultural Factors:

- ✓ The changing cultural values of society has direct impact on the human resource functions.
- ✓ The increased Participation in women in the labour forces is an example of a cultural change that influence HR activities.
- ✓ Changing attitudes towards work and leisure have confronted human resource departments with requests for longer vacations, more holidays and varied workweeks. Supervisors increasingly turn to HR managers for help with employee motivation.

6.Demographicc and Competitive conditions:

- ✓ The factors of geographic and competitive conditions influence the activities of human management.

7.Geographic Condition:

- ✓ One geographic factor affecting affecting the supply of Human Resources is the net migration into a particular region.
- ✓ The shift of population growth to the cities is an HR planning concern.
- ✓ Many workers are reluctant to accept geographic relocation as a precondition of promotion in the organisation. This trend has forced the organisations to change their development policies and practices.

8.Competitors:

- ✓ Competitors are another external force in staffing.
- ✓ Failure to consider the competitive labour market and to offer pay scales and benefits competitive to organisations in the same general industry and geographic locations may adversely affect the organisation's outcome.
- ✓ Underpaying or undercompeting may result in much lower quality workforce.

1.4 HUMAN RESOURCE POLICIES

What is meant by a policy?

A policy is a man-made rule of pre-determined course of action that is established to guide the performance of work towards the organisation objectives.

Policy is a type of standing plan that serves to guide subordinates in the execution of their tasks.

What are Human Resource Policies?

The Human resource policies provides guidelines for a wide variety of employment relationships in the organisation. These guidelines identify the identify the organisation's intentions in the recruitment,selection,development,compensation etc., people in the working organisation. HR policies serve as a road map for HR managers and line managers.

1.5 Essentials Characteristics of a sound human resource policy

1.The statement of HR policy should be defined, positive, clear and easily understood by everyone in the organisation so that what it proposes to achieve is evident.

2. It should be periodically revised, evaluated, assessed and revised.

3.It must be supplementary to the overall policy of an organisation.

4. It should be formulated with due regard for the interest of all concerned parties-the employers, the employees, and the public community.

5.It must provide a two way communication system between the management and the employees.

6. It should be consistent with the public policy.

7.It should be progressive and enlightened, and must be consistent with professional practice and philosophy.

8. It should be uniform throughout the organisation.

9. It should have a sound base in appropriate theory and should be translated into practices, terms and peculiarities of every department of an enterprise.

10. It must make a measurable impact , which can be evaluated and qualified for the guidance of all concerned.

1.6 New Trends in Human Resource Management:

Since HRM is the prime mover of the management of the people at work, therefore it has to encounter these challenges effectively in order to enable organisations to achieve their objectives.

Some of the important new trends that are emerging at the global level as well as in India are,

1.Globalisation of Economy

2.Corporate Restructuring

3.New Organisational Designs

4.Emphasis on Total Quality Management

5.Emphasis on Kaizen

6.Changing Job profile

7.Increasing diversity in the work force

8. Increasing role of Women Employees

9.Emphasis Knowledge Management

10. Increasing view on organisation as vehicles for achieving social goals.

2.0 WHAT IS HUMAN RESOURCE PLANNING?

Human Resource Planning is a process by which an organisation ensures that it has the right number and kinds of people , in the right places, and at the right times, who are capable of effectively and efficiently performing the assigned tasks.

2.1 HUMAN RESOURCE PLANNING PROCESS

HRP consists of forecasting future human resource needs, forecasting the availability of those human resources, and matching supply with the demand . Figure 1.3 Illustrates the model of human resource planning process.

Steps in Human Resource Planning:

As shown in the figure 1.3 the five major steps involved in Human Resource planning are

1.Collect Information

2.Forecast demand for Human Resources

3. Forecast supply of Human Resources

4. Identify Human Resource gap

5.Action Plans

1.Collect Information:

The first step in any form of HR planning is to collect information. A plan or a forecast cannot be any better than the data on which it is based.
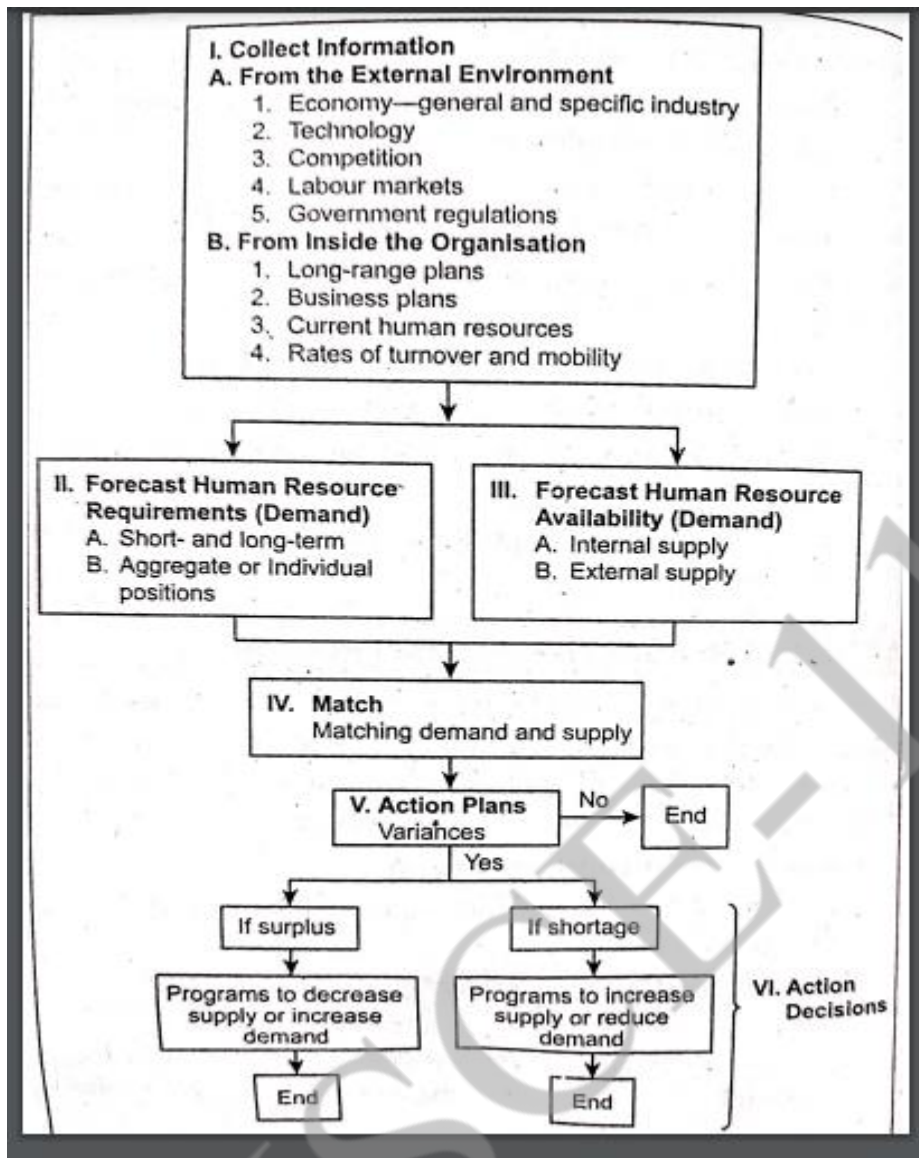
Figure 1.3 Human Resource Planning Process Model

- ✓ HR planning requires two types of information
  - A. Data from external environment
  - B. Data from inside the organisation

A.Data from external environment:

- These data include information on current conditions and predicted changes in the general economy, the economy of the specific industry, the relevant technology, and the competitons.
- Any of these factors may affect the organisations, business plans and thus the need for the human resources.
- Also, HR plnners must be aware of labour market conditions such as unemployment rates, skill availabilities and the age, and sex distributions of the labour force.

- Finally HR planners need to be aware of central and state government regulations: Those that directly affect the staffing practices.

B.Data from inside the organisation:

- Internal information includes short and long term organisational plans.
- The organisations plans to build, close, modify, or automates its facilities will have HR implications.
- Information is also needed on the current state of Human Resources in the organisations, such as how many individuals are employed in each job and location and how many are expected to leave or retire during the forecast period.

2.Forecast Demand for Human Resources:

- ✓ Once the HR planners have collected the information from both internal and external sources , they next forecast the future demand for the employees.
- ✓ The forecasting answers the question : how many and what type of employees will be needed to carry out the organisation's plans in the future.?
- ✓ The forecasts are grounded in information about the past and present and in assumptions about the future.

3.Forecast supply for Human Resources:

- ✓ Once the Human resource department makes the projections about the future resource demands, the next major concern is to forecast the supply of labour.
- ✓ There are two sources of supply: i) Internal supply ii) External supply
  - The internal supply of labour consists of all the individual currently employed by an organisation. It consists of present employees who can be promoted , transferred, or demoted to meet the anticipated needs.
  - The external supply of labour consists of people in the labour market who donot work for the organisation. These include employees of other organisation and those who are unemployed.

4.Identify Human Resource and Gap(Matching supply and Demand):

- ✓ Once HR planner has estimated the organisation's future demand and supply of Human Resources, the next step in HR planning is plan specific programs tto ensure that supply will match demand in the future.
- ✓ In this step , the gap between the human resource needed and their availability is identified.
- ✓ This human resource gap maybe in two forms: either surplus human resources or shortage of human resources.

5.Action Plans:

- ✓ Various action plans /decisions have to be devised to bridge the identified human resource gap.
- ✓ The two possible problems are either surplus human resource or shortage of human resources.
- ✓ If there is a shortage of human resources, the problem maybe resolved by discouraging retirements, hiring new people, transferring people from overstaffed areas, and installing labour saving equipments and processes.
- ✓ If there is a surplus of human resources, the problem, maybe resolved by utilizing attrition (i.e., not replacing people who leave), offering earlier retirements, Transfering people to understaffed areas and terminating people.
- ✓

2.3 Three ranges of Human Resource Forecasting:

The Human resource forecasting maybe categorized into three , based on the time frame as:

1. Short range of forecasting (0 to 2 years)
2. Intermediate Range forecasting (2 to 5 years)
3. Long Range forecasting (beyond 5 years)

Range \\ Forecast	Short range (0 – 2 years)	Intermediate range (2 – 5 years)	Long range (beyond 5 years)
Demand	Authorized employment (including growth, changes and turnover)	Operating needs from budgets and plans	In some organisations the same as "intermediate"; in others, an increased awareness of changes in environment and technology—essentially judgmental
Supply	Employee census less expected losses plus expected promotions from subordinate groups	Manpower vacancies expected from individual promotability data derived from development plans	Management expectations of changing characteristics of employees and future available manpower
Net needs	Number and kinds of employees needed	Numbers, kinds, dates, and levels of needs	Management expectations of future conditions affecting immediate decisions

2.4 Forecasting Human Resource Supplies

Types of forecasting techniques :

As with forecasting demand, two basic techniques help forecast internal labour supply. Thy are**:** 1. Judgemental forecasts 2.Statistical Techniques

1. **Judgemental forecasts**

   Organisation use two judgemental techniques to make supply forecasts

   i) Replacement analysis ii) Succession analysis

   i).Replacement Analysis:

   ✓ Replacement Analysis uses replacement charts. Replacement charts are a visual representation of who will replace whom in the event of a job opening. In other words, replacement charts are developed to show the names of the current occupants of positions in the organisations and the names of likely replacements.

   ✓ Replacement charts make potential vacancies readily apparent and indicate what types of positions most urgently need to be filled.
   Present performance levels of current employees can be used to estimate potential vacancies.

   ✓ On the replacement chart, the incumbents are listed directly under the job title.These individuals likely to fill the potential vacancies are listed listed directly under the incumbents.

   ✓ Such a listing can provide the organisations with a good estimate of what jobs are likely to become vacant and indicate if anyone will be ready to fill the vacancy.

   ii)Succession Analysis:

   Succession analysis is similar to replacement, except that succession planning tends to be a longer term and more developmental and tends to offer greater flexibility.

2. **Statistical Techniques:**

   ✓ With the advent of personal computers, organisations are using more sophisticated statistical model to forecast the supply of human resources.

   ✓ Two commonly used statistical techniques for forecasting human resource supplies are i) Markov Analysis ii) Goal Programming

   i)   Markov Analysis

   ✓ Markov Analysis is a fairly simple method of predicting the internal supply of labour at some future time.

   ✓ The heart of Markov Analysis is the transition probability matrix.The transition matrix shows the probability of an employee staying in his present job, moving from one position to another or leaving the organisation, for forecast time period.

   ✓ When this transition matrix is multiplied by the number of people beginning the year in each job, the results show how many people are expected to be in each job by the end of the year.

✓ Markov analysis can help to identify the lower retention probability, but it does not suggest any particular solution to the potential problem.

3. **Goal Programmimg:**
   ✓ Goal programming is a further extension of Markov Analysis.
   ✓ The objective of goal programming is to optimise goals- in this case, a desired staffing pattern—given a set of constraints concerning such thimgs as the upper limits on flows, the percentages of new recruits and the total salary budget.

2.5 Statistical Techniques used to forecast staffing demand needs

Name	Description
1. Regression analysis	✓ Past levels of various work load indicators, such as sales, production levels, and value added, are examined for statistical relationships with staffing levels. ✓ If sufficiently strong relationships are found, a regression (or multiple regression) model is derived. ✓ Forecasted levels of the retained indicator(s) are entered into the resulting model and used to calculate the associated level of human resource requirements.
2. Productivity ratios	✓ Historical data are used to examine past levels of a productivity index (P) : $$P = \frac{\text{Work load}}{\text{Number of people}}$$ ✓ If constant, or systematic, relationships are found, human resource requirements can be computed by dividing predicted work loads by P.
3. Personnel ratios	✓ Past personnel data are examined to determine historical relationships among the employees in various jobs or job categories. ✓ Regression analysis or productivity ratios are then used to project either total or key group human resource requirements, and personal ratios are used to allocate total requirements to various job categories or to estimate requirements for non-key groups.
4. Time series analysis	✓ Past staffing levels (instead of work load indicators) are used to project future human resource requirements. ✓ Past staffing levels are examined to isolate seasonal and cycloidal variations, long-term trends, and random movement. ✓ Long-term trends are then extrapolated or projected using a moving average, exponential smoothing, or regression technique.
5. Stochastic analysis	✓ The likelihood of landing a series of contracts is combined with the personnel requirements of each contract to estimate expected staffing requirements. ✓ Potential applications are mostly in government contractors and construction industries.

3.0 Recruitment

Defnition:

Recruitment maybe defined as the process of discovering potential candidates for actual and anticipated organisational vacancies.

Recruitment can be described as those activities in HRM which are undertaken inorder to attract sufficient job candidates who have the necessary potential, competencies and traits to fill job needs and to assist the organisation in achieving its objectives.
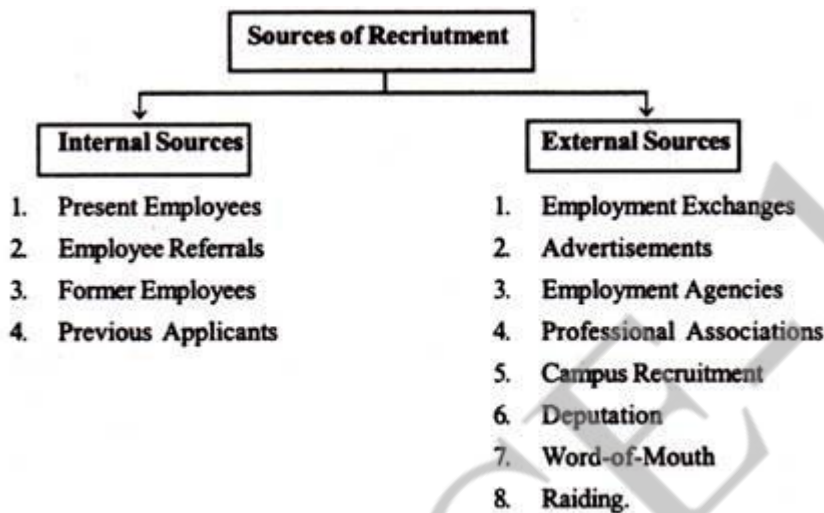
3.1 Recruitment Sources :

```
                    ┌──────────────────────────┐
                    │  Sources of Recriutment  │
                    └──────────────────────────┘
                          │              │
            ┌─────────────────────┐  ┌──────────────────┐
            │   Internal Sources  │  │  External Sources │
            └─────────────────────┘  └──────────────────┘
```

Internal Sources	External Sources
1. Present Employees	1. Employment Exchanges
2. Employee Referrals	2. Advertisements
3. Former Employees	3. Employment Agencies
4. Previous Applicants	4. Professional Associations
	5. Campus Recruitment
	6. Deputation
	7. Word-of-Mouth
	8. Raiding.

Fig. 6.1: Recruitment Sources

A. Internal Sources:

1. Present Employees:

Promotions and transfers from among the present employees can be a good source of recruitment. Promotion implies upgrading of an employee to a higher position carrying higher status, pay and responsibilities. Promotion from among the present employees is advanta-geous because the employees promoted are well acquainted with the organisational culture, they get motivated, and it is cheaper also.

Promotion from among present employees also reduces the require-ment for job training. However, the disadvantage lies in limiting the choice to a few people and denying hiring of outsiders who may be better qualified and skilled. Furthermore, promotion from among present employees also results in inbreeding which creates

frustration among those not promoted. Transfer refers to shifting an employee from one job to another without any change in the position/post, status and responsibilities. The need for transfer is felt to provide employees a broader and varied base which is considered necessary for promotions. Job rotation, involves transfer of employees from one job to another on the lateral basis.

2. Former Employees:

Former employees are another source of applicants for vacancies to be filled up in the organisation. Retired or retrenched employees may be interested to come back to the company to work on a part-time basis. Similarly, some former employees who left the organisation for any reason may again be interested to come back to work. This source has the advantage of hiring people whose performance is already known to the organisation.

3. Employee Referrals:

This is yet another internal source of recruitment. The existing employ-ees refer their family members, friends and relatives to the company as potential candidates for the vacancies to be filled up in the organisation. This source serves as one of the most effective methods of recruiting people in the organisation because employees refer to those potential candidates who meet the company requirements known to them from their own experience. The referred individuals are expected to be similar in type in terms of race and sex, for example, to those who are already working in the organisation.

4. Previous Applicants:

This is considered as internal source in the sense that applications from the potential candidates are already lying with the organisation. Sometimes, the organisations contact through mail or messenger these applicants to fill up the vacancies particularly for unskilled or semi- skilled jobs.

B. External Sources:

External sources of recruitment lie outside the organisation. These outnumber internal sources.

1. Employment Exchanges:

The National Commission on Labour (1969) observed in its report that in the pre-Independence era, the main source of labour was rural areas surrounding the industries. Immediately after Independence, National Employment Service was established to bring employers and job seekers together.

In response to it, the compulsory Notification of Vacancies Act of 1959 (commonly called Employment Exchange Act) was instituted which became operative in 1960. Under Section 4 of the Act, it is obligatory for all industrial establishments having 25 workers or more, to notify the nearest employment exchange of vacancies (with certain exceptions) in them, before they are filled.

The main functions of these employment exchanges with their branches in most cities are registration of job seekers and their placement in the notified vacancies. It is obligatory for the employer to inform the outcome of selection within 15 days to the employment exchange.

Employment exchanges are particularly useful in recruiting blue-collar, white-collar and technical workers. A study conducted by Gopalji on 31 organisations throughout the country also revealed that recruitment through employment exchanges was most preferred for clerical personnel i.e., white-collar jobs.

2.Employment Agencies:

Generally, these agencies select personnel for supervisory and higher levels. The main function of these agencies is to invite applications and short list the suitable candidates for the organisation. Of course, the final decision on selection is taken by the representatives of the organisation. At best, the representatives of the employment agencies may also sit on the panel for final selection of the candidates

3.Advertisement:

Advertisement is perhaps the most widely used method for generating many applications. This is because its reach is very high. This method of recruitment can be used for jobs like clerical, technical and managerial. The higher the position in the organisation, the more specialized the skills or the shorter the supply of that resource in the labour market, the more widely dispersed the advertisements is likely to be.

For example, the search for a top executive might include advertise-ments in a national daily like 'The Hindu'.

4. Professional Associations:

Very often, recruitment for certain professional and technical positions is made through professional associations also called ' headhunters'. Institute of Engineers, Indian Medi-cal Association, All Indian Management Association, etc., provide placement services for their members. For this, the professional associations prepare either list of job seekers or publish or sponsor journals or magazines containing advertisements for their members. The professional associations are particularly useful for attracting highly skilled and professional personnel. However, in India, this is not a very common practice and those few that provide such kind of service have not been able to generating a large number of applications.

5. Campus Recruitment:

This is another source of recruitment. Though campus recruitment is a common phenomenon particularly in the American organisations, it has made its mark rather recently Of late, some organisations such as HLL, HCL. L &T, Citi Bank, ANZ Grindlays, Motorola, Reliance etc., in India have started visiting educational and training institutes/campuses for recruitment purposes.

Ex-amples of such campuses are the Indian Institutes of Management, Indian Institutes of Technology and the University Departments of Business Management. For this purpose, many institutes have regular placement cells/offices to serve as liaison between the employers and the students. Tezpur Central University has, for example, one Deputy Director (Training and Placement) for the purposes of campus recruitment and placement.

The method of campus recruitment offers certain advantages to the employer organisations. First, the most of the candidates are available at one place; Second, the interviews are arranged at short notice; third, the teaching faculty is also met; and Fourth, it gives them opportunity to sell the organisation to a large student body who would be graduating subsequently. However, the disadvantages of this type of recruitment are that

organisations have to limit their selection to only "entry" positions and they interview the candidates who have similar education and experience, if at all.

6. Deputation:

Another source of recruitment is deputation, i.e., sending an employee to another organisation for a short duration of two to three years. This method of recruitment is practiced, in a pretty manner, in the Government Departments and public sector organisations. Deputation is useful because it provides ready expertise and the organisation does not have to incur the initial cost of induction and training.

However, the disadvantage associated with deputation is that the deputa-tion period of two/three years is not long enough for the deputed employee to prove his/her mettle, on the one hand, and develop commitment with the organisation to become part of it, on the other.

7. Word-of-Mouth:

Some organisations in India also practice the 'word-of-mouth' method of recruitment. In this method, the word is passed around the possible vacancies or openings in the organisation. Another form of word-of-mouth method of recruitment is "employee-pinching" i.e., the employees working in another organisation are offered an attractive offer by the rival organisations. This method is economic, both in terms of time and money.
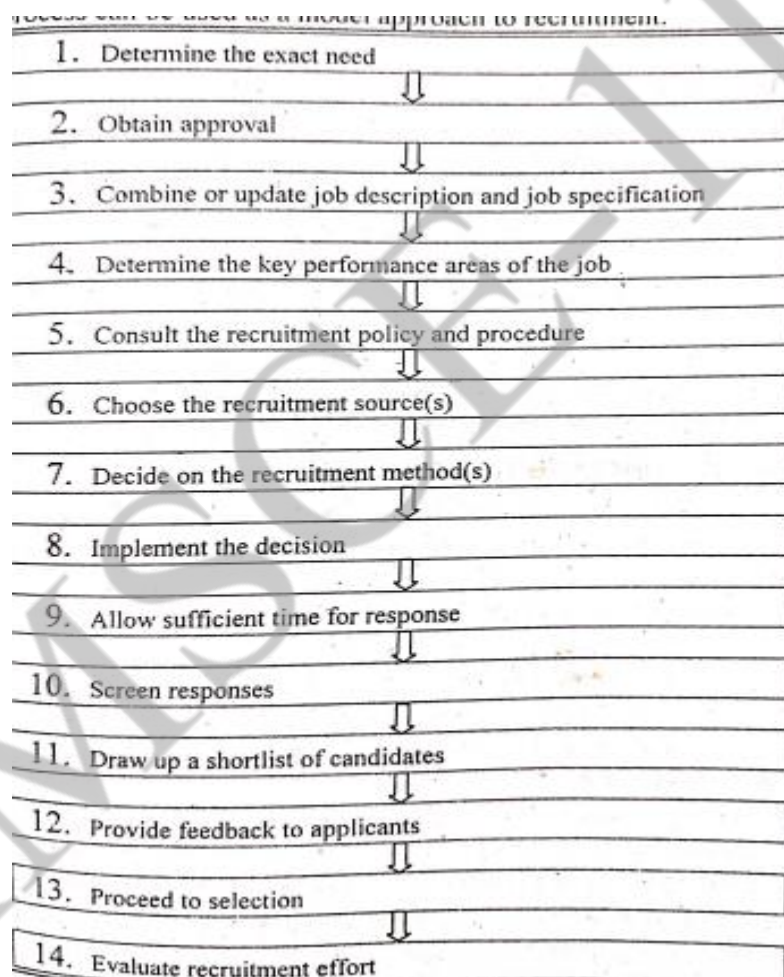
Some organisations maintain a file of the applications and bio-data sent by job-seekers. These files serve as very handy as and when there is vacancy in the organisation. The advantage of this method is no cost involved in recruitment. However, the drawbacks of this method of recruitment are non-availability of the candidate when needed and the choice of candidates is restricted to a too small number.

8. Raiding or Poaching:

Raiding or poaching is another method of recruitment whereby the rival firms by offering better terms and conditions, try to attract qualified employees to join them. This raiding is a common feature in the Indian organisations.

For example, several executives of HMT left to join Titan Watch Company, so also exodus of pilots from the Indian Airlines to join private air taxi operators. Whatever may be the means used to raid rival firms for potential candidates, it is often seen as an unethical practice and not openly talked about. In fact, raiding has become a challenge for the human resource manager. Besides these, walk-ins, contractors, radio and television, acquisitions and mergers, etc., are some other sources of recruitment used by organisations.

3.2 Various steps involved in Recruitment process

process can be used as a model approach to recruitment.

1. Determine the exact need
   ⇓
2. Obtain approval
   ⇓
3. Combine or update job description and job specification
   ⇓
4. Determine the key performance areas of the job
   ⇓
5. Consult the recruitment policy and procedure
   ⇓
6. Choose the recruitment source(s)
   ⇓
7. Decide on the recruitment method(s)
   ⇓
8. Implement the decision
   ⇓
9. Allow sufficient time for response
   ⇓
10. Screen responses
    ⇓
11. Draw up a shortlist of candidates
    ⇓
12. Provide feedback to applicants
    ⇓
13. Proceed to selection
    ⇓
14. Evaluate recruitment effort

3.3 Realistic Job Previews(RJP):

One technique to improve the recruitment process is known as the Realistic Job Previews.

RJP refers to a description provided by the organisation to applicants and new employees that gives both the positive and negative aspects of a job.

RJP can improve the recruitment process by giving each candidate all the pertinent and realistic information about the job and the organisation. In this the positive and negative sides of the job and firm are included.

In this manner , the candidate can make a more-informed choice and select jobs for which he or she is better suited. In the long run, the RJP helps to achieve overall job satisfaction and performance. It also avoids situations where dissatisfaction and poor performance results from a person finding that the job and its environment were not as advertised.

UNIT V MARKETING AND SALES ANALYTICS

Marketing Strategy, Marketing Mix, Customer Behavior – selling Process – Sales Planning – Analytics applications in Marketing and Sales

1. Marketing Strategy

• Analytical marketing strategies help to measure the effectiveness of marketing tools and the development of any brand. An advertising campaign and marketing initiatives require a huge amount of investment. So how to understand the budget? Which channels are most effective? How much profit is to be obtained? Marketing analytics strategies provide the answers to all these questions.

1.1 Significance of Marketing Analytics Strategies

• Marketing analytical strategies are critical in addressing and resolving marketing issues. The prime motive behind implementing these strategies is to evaluate the effective marketing programs in terms of return on investments in a business.

• The outcomes of adopting these strategies:

• Comparisons with competitors

• Recommendations for effective allocation of budget and resources

• Data processing analysis

• Collection of data through all the channels of communication and units in the company

• Creating a structured template for reporting the purpose of effective analysis of units

• Thus the marketing analysis strategies help in the given aspects:

• Having a holistic view of business

• Improving the management of the company and finance

• Forecasting and planning marketing initiatives

• Increasing the effectiveness of existing marketing programs through the allocation of resources

• Increase the profitability and return on investments

• Thus you should set a proper marketing analytics framework within the organization to have the right processes along with the right technology platforms to

capture data-driven strategy and deliver consistent information about this.

1.2 Analytics Marketing Strategies

•   Marketing analytics is the practice of combining and analyzing databases, identifying patterns and then coming up with actionable insights that improve the return on investments of marketing efforts. Modern marketing analytics provides a holistic picture of the business and lets you plan and optimize the whole process based on revenue attribution. Besides this you should follow a new marketing strategy to survive long term with competition from customer-focused services and products. Today almost all businesses are following self-service, AI-powered analytics to analyze and visualize the data and design the dashboards. Here are some prime analytics marketing strategies as described below:

1. Exploring The Top Marketing Analytics Resources

You need to explore top marketing analytics resources, some of these are as

follows:

•   Hear From Peers

•   Get the Buyers Guide

•   Know the Trends

These all are prime tools that can be used by you to explore marketing techniques. You should get the buyer's guide, this will enable you to meet the expectations and requirements of customers. Besides this, you should also follow the latest marketing trends to navigate today's fast-paced world.

2. Website Marketing Analysis

As far as digital marketing tools are considered, the website is the best tool for it. Here you should understand the top pages of the websites to generate a high amount of conversion and traffic. Besides this, you should also identify the pages that are receiving high traffic but not conversions. Heatmaps can help you in analyzing the audience interaction with each element on your page. This will enable you to identify pages that are getting a high bounce, identify the audience, their demographics, devices they are using to access your content and the ranking of keywords on your web pages.

3. Social Media Analytics

The social media platform has become the most accessible and diverse tool from the perspective of marketing. This marketing strategy can help you to understand the sentiments of people and how they are responding and engaging with you. This will enable you to take decisive action and approach the right audience of the target market. To implement a successful analytics marketing strategy you would have to reach more people and engage with the followers to understand the improvements they are looking for.

4. Campaign Analytics

This strategy helps you in tracking your campaign, like how these are performing, getting the leads or not. So what you can do is understand the lead conversion rates from multiple channels and sources. After doing this, you have to identify the opportunities by product category and the source of lead. With this, you need to identify the content and platform that is majorly resonating with your audience. This will enable you to optimize the messaging and target of your content strategy.

5. Link Analytics

• Link is the most crucial aspect of searching algorithms. By taking the assistance of link analytics you can view the link of the site, the domain, and page authority of referring domains, like the total number of inbound links, top pages by link, anchor text and many more.

• Thus having transparency is the prime motive of marketing managers. For this, they have to set a common agreement on different KPIs. in today's competitive age it is essential to opt for effective marketing strategies by learning the art of positioning your brand, as it can help to win over the competitors. In addition to this, another important element of a reliable marketing analytics framework is to build an effective analytics dashboard. This dashboard should represent KPIs by unifying data strategies from different marketing data sources.

6. Keyword Research–

With keyword research, you can obtain very detailed insights into how your business is appealing to your potential customers and if there are areas that you can optimise. View how competitive your target keywords are, the average monthly search volume for that particular keyword, the estimated CPC's if you decide to bid on those keywords, the number of clicks that you are getting for that keyword and the click-through rates.

Thus the marketing analytics strategies are necessary for any business to obtain timely, reliable, complete and operative information.

Tools

• It is the practice of studying the data of Marketing efforts of various channels and campaigns and form models in order to report the metrics like ROI, Channel Performance, etc. to identify parameters for improvement. Marketers will be able to provide answers

to the analytics questions that are most important to their stakeholders by monitoring and reporting on business performance results, diagnostic metrics, and leading indicator metrics

• The intelligence derived from marketing analytics allows you to spend each dollar as effectively as possible.

• However, despite the emergence of several platforms and technologies that can streamline the marketing analytics workflows, it remains a challenge for companies to build concrete, actionable data analytics solutions for marketing efforts. According to a survey of senior marketing executives published in the Harvard Business Review, "more than 80% of respondents were dissatisfied with their ability to measure marketing ROI."

• To set up a practical marketing analytics framework within your organisation, you must have the right processes along with investing in the right technology platforms to capture data-driven strategy and deliver unified and consistent information on your measurement metrics.

1.3 Marketing Analytics Strategies Process

• With marketing analytics, you can gather intelligence into several different areas of your marketing strategy. It will help you understand how your programs are performing against the cost and which programs are delivering the best ROI. It will help you to segregate your efforts and identify the area that you need to focus on the most.

• Analytics strategy will help you to realise how your programs are working in conjunction to nurture your leads. With this, you can build a solid base upon which you can qualify them and pass the leads on to your sales reps as opportunities.

• With marketing analytics, you can also identify laggards, i.e. the programs that are not providing adequate return based on efforts invested at them. You can then choose to redefine your data-driven strategy at them or remove them from your focus altogether.

• Market and competitor analysis will give you crucial insights into your competitor data-driven strategy and which channels/ programs are working for them. Learning from your competitors is an old business principle and marketing analytics can give you a powerful arsenal to use and base your actions on the digital platforms.

• Even better! Advanced analytics can provide insights into trends, make forecasts and capitalise on opportunities before anyone else.

• This will help grow your bottom line and avoid wastage on marketing spending, optimising the dollar spend and viewing campaign performance in real-time. It helps you to measure the impact of your strategies and compare it against the cost. •

Marketing strategies and tactics are normally based on explicit and implicit beliefs about consumer behavior. Decisions based on explicit assumptions and sound theory and research are more likely to be successful than the decisions based solely on implicit intuition.

• Knowledge of consumer behavior can be an important competitive advantage while formulating marketing strategies. It can greatly reduce the odds of bad decisions and market failures. The principles of consumer behavior are useful in many areas of marketing, some of which are listed below –

✓ **Analyzing Market Opportunity**

Consumer behavior helps in identifying the unfulfilled needs and wants of consumers. This requires scanning the trends and conditions operating in the market area, customer's lifestyles, income levels and growing influences.

✓ **Selecting Target Market**

The scanning and evaluating of market opportunities helps in identifying different consumer segments with different and exceptional wants and needs. Identifying these groups, learning how to make buying decisions enables the marketer to design products or services as per the requirements.

Example – Consumer studies show that many existing and potential shampoo users did not want to buy shampoo packs priced at Rs 60 or more. They would rather prefer a low price packet/sachet containing sufficient quantity for one or two washes. This resulted in companies introducing shampoo sachets at a minimal price which has provided unbelievable returns and the trick paid off wonderfully well.

✓ **Marketing-Mix Decisions**

Once the unfulfilled needs and wants are identified, the marketer has to determine the precise mix of four P's, i.e., Product, Price, Place, and Promotion.

✓ **Product**

A marketer needs to design products or services that would satisfy the unsatisfied needs or wants of consumers. Decisions taken for the product are related to size, shape, and features. The marketer also has to decide about packaging, important aspects of service, warranties, conditions, and accessories.

Example – Nestle first introduced Maggi noodles in masala and capsicum flavors.Subsequently, keeping consumer preferences in other regions in mind, the company introduced Garlic, Sambar, Atta Maggi, Soupy noodles, and other flavours.

✓ **Price**

The second important component of marketing mix is price. Marketers must decide what price to be charged for a product or service, to stay competitive in a tough market. These decisions influence the flow of returns to the company.

✓ **Place**

The next decision is related to the distribution channel, i.e., where and how to offer the products and services at the final stage. The following decisions are taken regarding the distribution mix –

•   Are the products to be sold through all the retail outlets or only through the selected ones?

•   Should the marketer use only the existing outlets that sell the competing brands? Or, should they indulge in new elite outlets selling only the marketer's brands?

•   Is the location of the retail outlets important from the customers' point of view?

•   Should the company think of direct marketing and selling?

   ✓ **Promotion**

•   Promotion deals with building a relationship with the consumers through the channels of marketing communication. Some of the popular promotion techniques include advertising, personal selling, sales promotion, publicity, and direct marketing and selling.

•   The marketer has to decide which method would be most suitable to effectively reach the consumers. Should it be advertising alone or should it be combined with sales promotion techniques? The company has to know its target consumers, their location, their taste and preferences, which media do they have access to, lifestyles, etc.

2.   Marketing Mix

•   Marketing mix modeling is a marketing analytics strategy that can help your brand maximize on return and get a deeper understanding of how your business actually functions. Let's look into the benefits that this strategy can provide for your brand. As the world of digital marketing has exploded, the rise of big data and incredibly technical and complex data sets has been both a blessing and a curse to brands big and small.

•   While it's true that detailed data can help businesses understand their consumers and grow their businesses, it's often the case that the data is overwhelming.

•   With technology platforms and analytics tools being able to collect enormous amounts of data, brands are often left struggling to get through it all and understand what it is that they've gathered.

•   In order to address the issue of how to manage incoming data and then use that information to make impactful decisions, a clear analytics strategy is necessary for all brands.

•   Picking the right strategy for your business is the key to making sure you are getting the most out of your planning and marketing activity.

•   Marketing mix modeling is one example of a marketing analytics strategy that can really help your brand manage data and learn the best places to invest your budget and time on.

•   Keep reading this post to learn more.

2.1 What is Marketing Mix Modeling?

• Marketing mix modeling is a statistical marketing method that attempts to determine the effectiveness of marketing campaigns and initiatives by taking apart data and attributing contributions to different marketing tactics and factors to better predict future success.

• Put another way, marketing mix modeling looks at different pre-determined factors and the data that has been gathered from marketing campaigns to see which factors have had the biggest impact on return and which factors have contributed the most to success.

• Once this data has been collected and organized, the marketing mix modeling system will use the past and historical data to predict or forecast future marketing and sales success.

• By looking at the trends that have worked before, the marketing mix modeling will theoretically be able to forecast with more accuracy than other analytical methods.

2.2 The 5 P's of Marketing Mix Modeling

As stated above, marketing mix modeling distributes success from data to different pre-determined factors.

Those factors are often referred to as the 5 P's of marketing, which are derived from other marketing research and studies. Let's look at those 5 P's now.

- ✓ **Product :** Product refers to the actual products or services that are created and offered to customers by a brand.
- ✓ **Price :** The price takes into consideration any deals, sales, pricing models, and methods of payment involved in a sale.
- ✓ **Place :** Place refers to the channels through which products are available to consumers and how consumers are able to find the offers that the brand has. Promotion
- ✓ **Promotion** : is the method by which products or services are marketed and shared among audiences.
- ✓ **People:** People is the final P, and is sometimes left off of marketing mix modeling. People refers to both the internal staff and the customers that drive sales in a brand.

2.3 Marketing Mix Modeling vs. Attribution Modeling

• Marketing mix modeling is often compared to another popular model of marketing analytics, attribution modeling.

• Attribution modeling is the process of setting up different touchpoints that trigger events on the customer's journey.

• Each touchpoint is assigned a value to help determine which points in the customer's journey are responsible for bringing in revenue. • While attribution modeling can be helpful to understand data and provide context for ROI, it also has a few major drawbacks.

- The biggest problem is that not every touchpoint in a customer's journey can possibly be tracked and analyzed through collected data.

- Another drawback of attribution modeling is that it functions mainly through clicks and clicks alone — other potential data points are put aside in favor of clicks that can "prove" a conversion has taken place at a touchpoint.

- Attribution modeling also doesn't prove the effectiveness of a campaign. After all, a customer will have to pass through the same touchpoints whether they were convinced through an advertisement to make a conversion or not.

- That makes it difficult to assign return to specific touchpoints.

2.4 Benefits of the Marketing Mix Modeling

Let's take a deeper dive into the benefits that it can provide to your brand's analytics and reporting models.

✓ Prove the ROI of Marketing Initiatives

Marketing mix modeling allows marketers to really prove the ROI of their initiatives. By relating data insights back to the factors in each campaign that provided success, it can help brands understand the full impact of their efforts.

✓ Gather Insights

Marketing mix modeling is also great for understanding key insights from business initiatives. Those insights can be used to drive effective budget allocations within marketing and sales departments and convince stakeholders of the benefits of the model.

✓ Create Better Sales Forecasting

Sales forecasting refers to the practice of estimating how much revenue can be generated in the future based on the impact that your sales and marketing efforts have had in the past.

By allocating success to key factors, marketing mix modeling allows brands to have more accurate forecasting.

✓ Understand Historical Data and Trends

Marketing mix modeling is based on understanding the past data that has been collected during initiatives and campaigns.Many other analytics models will ignore this valuable data or only look at parts of it. The marketing mix system ensures historical data and trends are examined closely for value.

✓ Account for Negative Impacts

Just as marketing mix modeling allows brands to see the positive impacts that their efforts have created, it can also be used to see negative impacts on different marketing factors.That helps brands know which areas of the business need work and where serious corrections need to take place.

2.5 What are the Limitations of Marketing Mix Modeling?

Like all marketing analytics methods, there are drawbacks and limitations to this system.The amount of data collected means that there isn't one method of analytics that can address every data set. Here are some of the major drawbacks of marketing mix modeling:

• Infrequent reporting, meaning no real-time data analytics.

• Does not analyze the customer's experience or journey.

• Doesn't provide the 1:1 analysis of attribution modeling.

• Doesn't examine relationships between channels.

• Doesn't look into brand awareness, messaging, or reach.

• Requires a large marketing analysis budget.

• Harder to implement in B2B businesses than B2C brands.

2.6 How to Build the Marketing Mix Modeling

While there are some setbacks, marketing mix modeling can provide major benefits to your brand. Let's take a look at how you can go about building this system in your own organization.

1. **Establish Your Goals**

• The end goal of any marketing analytics strategy is to parse through and gather insights from your data sets.

• That means that marketing mix modeling is meant to help organize your data and your analytics methods.

• Therefore, it makes sense that the first step is to establish the specific goals you want to attain through your strategy.

• Your goals might center around budgets, marketing campaigns, product pricing, or your brand in comparison to competitors.

2. **Create Internal Alignment**

In order to succeed, you need to have clear alignment across your organization.

• As with most data analytics, marketing mix modeling requires you to pull data from many different systems from different departments.

• That requires compliance across different teams and with the key stakeholders in your organization, such as:

CMO, Media agencies, Marketing agencies, Marketing executives and managers,CRM managers,Sales leads

3.0 Consumer Behavior

Consumer behavior is about the approach of how people buy and the use merchandise and services. Understanding consumer behavior will assist business entities to be more practical at selling, designing, development of products or services, and every other different initiative that impacts their customers. In this tutorial, it has been our endeavor to cover the multidimensional aspects of Consumer Behavior in an easy-to-understand manner.

- **Audience**

This tutorial will help management students as well as industry professionals who work in a product development environment, or in packaging, or for that matter, any part of a company that has an interface with the customers.

- **Prerequisites**

To understand this tutorial, it is advisable to have a foundation level knowledge of basic business and management studies. However, general students and entrepreneurs who wish to get an understanding about consumer behavior may find it quite useful.

- **Consumer Behavior - Consumerism**

Consumerism is the organized form of efforts from different individuals, groups, governments and various related organizations which helps to protect the consumer from unfair practices and to safeguard their rights. The growth of consumerism has led to many organizations improving their services to the customer.

- **Consumerism**

Consumer is regarded as the king in modern marketing. In a market economy, the concept of consumer is given the highest priority, and every effort is made to encourage consumer satisfaction. However, there might be instances where consumers are generally ignored and sometimes they are being exploited as well. Therefore, consumers come together for protecting their individual interests. It is a peaceful and democratic movement for self-protection against their exploitation. Consumer movement is also referred as consumerism.

3.1 Features of Consumerism

Highlighted here are some of the notable features of consumerism –

Protection of Rights – Consumerism helps in building business communities and institutions to protect their rights from unfair practices.

Prevention of Malpractices – Consumerism prevents unfair practices within the business community, such as hoarding, adulteration, black marketing, profiteering, etc.

Unity among Consumers – Consumerism aims at creating knowledge and harmony among consumers and to take group measures on issues like consumer laws, supply of information about marketing malpractices, misleading and restrictive trade practices.

Enforcing Consumer Rights – Consumerism aims at applying the four basic rights of consumers which are Right to Safety, Right to be Informed, Right to Choose, and Right to Redress.

Advertising and technology are the two driving forces of consumerism –

• **The first driving force of consumerism is advertising**. Here, it is connected with the ideas and thoughts through which the product is made and the consumer buys the product. Through advertising, we get the necessary information about the product we have to buy.

• **Technology** is upgrading very fast. It is necessary to check the environment on a daily basis as the environment is dynamic in nature. Product should be manufactured using new technology to satisfy the consumers. Old and outdated technology won't help product manufacturers to sustain their business in the long run.

3.2 Consumer Behavior – Significance

• Consumer behavior covers a broad variety of consumers based on diversity in age, sex, culture, taste, preference, educational level, income level, etc. Consumer behavior can be defined as "the decision process and physical activity engaged in evaluating, acquiring, using or disposing of goods and services."

• With all of the diversity to the surplus of goods and services offered to us, and the freedom of choices, one may speculate how individual marketers actually reach us with their highly definite marketing messages. Understanding consumer behavior helps in identifying whom to target, how to target, when to reach them, and what message is to be given to them to reach the target audience to buy the product.

The following illustration shows the determinants of consumer behavior.

• The study of Consumer Behavior helps in understanding how individuals make decisions to spend their available resources like time, money, and effort while purchasing goods and services. It is a subject that explains the basic questions that a normal consumer faces – what to buy, why to buy, when to buy, where to buy from, how often to buy, and how they use it.
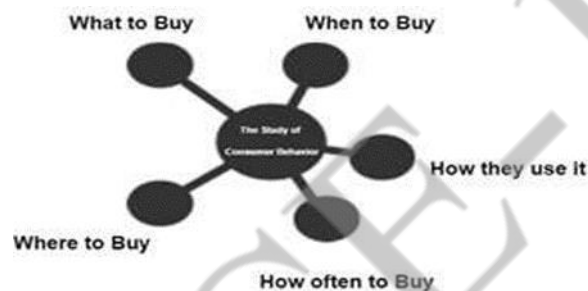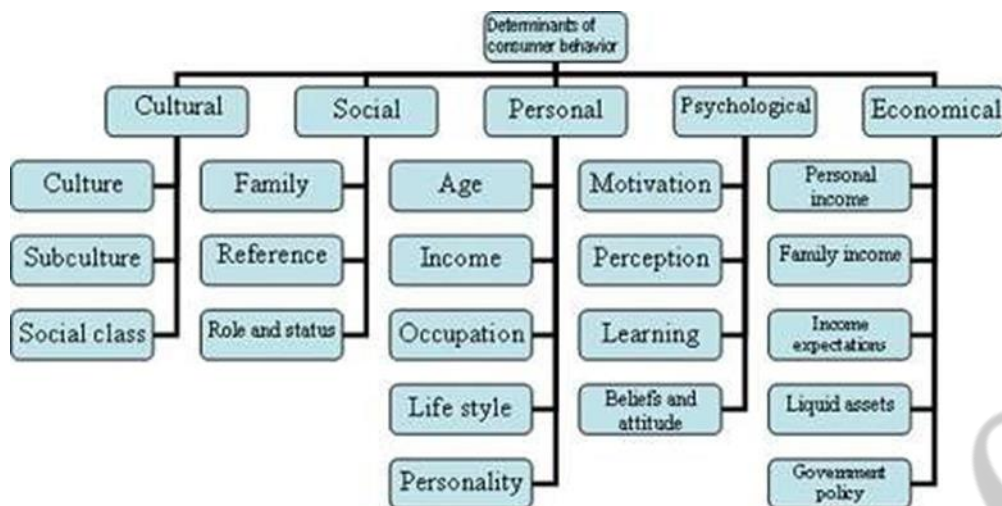
• Consumer behavior is a complex and multidimensional process that reflects the totality of consumer decisions with respect to acquisition, consumption, and disposal of goods and services.

3.3 Dimensions of Consumer Behavior

Consumer behavior is multidimensional in nature and it is influenced by the following subjects –

• **Psychology** is a discipline that deals with the study of mind and behavior. It helps in understanding individuals and groups by establishing general principles and researching specific cases. Psychology plays a vital role in understanding how consumers behave while making a purchase.

• **Sociology** is the study of groups. When individuals form groups, their actions are sometimes relatively different from the actions of those individuals when they are operating individually.

• **Social Psychology** is a combination of sociology and psychology. It explains how an individual operates in a group. Group dynamics play an important role in purchasing decisions. Opinions of peers, reference groups, their families and opinion leaders influence individuals in their behavior.

12

•   **Cultural Anthropology** is the study of human beings in society. It explores the development of central beliefs, values and customs that individuals inherit from their parents, which influence their purchasing patterns.

3.4 How Consumer Behavior affects Marketing Strategy ?

•   Business organizations across globe try to influence consumer by encouraging them to buy products and services. This is done by studying about the needs of the consumer and creating appropriate strategies so that consumer buys products. There are several marketing strategies used for influencing consumer behavior which affects the buying decision.

•   The first thing to be kept in mind while building strategies for marketing products is communicating with consumers emotionally. This can be done by giving promotional material in order to get attention of consumer. It has been found that consumers are attracted to products that create emotions in the form of joy and surprise.

•   All businesses throughout the world are seeking for solutions to assure long-term sales and profitability, as well as market sustainability. To do so, companies must pay close attention to their source of profit – consumers – and, more crucially, their behaviour.

•   Consumer behaviour is the study of consumer demands and how consumers (customers and organizations) meet these needs, as well as their motivation for using and purchasing a certain product or service.

•   This is an exceptionally helpful study for corporations looking seeking strategies to stay relevant in the market since it assists them in determining the best marketing plan for their items.

•   After rigorously analyzing consumer behaviour, only a relevant marketing plan can be established to advertise the service/product to the correct segment of the audience by finding a market gap or demand; failing to do so exposes the firm to

product/service failure. Businesses are expected to research all the criteria listed below to effectively analyze their customers.

• A successful marketing strategy is critical to a company's success since it assists the company in developing a product or service that has the potential to sell and provide high levels of profit yield   A marketing strategy is a company's plan for selling its product, which includes considering the four variables listed below.

Consumer behaviour and marketing strategy are inextricably linked:

• Consumer behaviour assists firms in determining whether what they are selling will be lucrative, as well as in tailoring their marketing plan to the appropriate target population for their product/service.

• Catering a product/service to the wrong audience may be detrimental to a business, whereas, Catering the appropriate product/service to the right consumers by observing their behaviour, on the other hand, might be invaluable to a company.

• Many organizations look for the most cost-effective way to do consumer research. By using technologies like Google Analytics, Google Survey, CRM, and the social networking sites listed above, businesses may keep track of their customers' web activity, making it easier to determine client preferences. Keepingtrack of consumer behaviour is critical for ensuring profitability

With the recent change towards the Covid-19 crisis, businesses must monitor customer behaviour more now than ever.Observably Covid-19 has bought drastic changes in consumer behaviour.Consumers are also less likely to make large purchases during an economic- financial crisis such as the recession; therefore, businesses must study and analyze consumer behaviour to ensure sustainability through having the right marketing strategy catered to the consumer's financial and emotional preferences. Failure to do so may result in the suspension of operations or bankruptcy.

In conclusion, consumer behaviour has a significant influence on marketing strategy and is important to the success of a product; so, the marketing strategy must be determined through analyzing consumer behaviour to understand what customers want. Meeting consumer demand is the quickest method to make profits – the ultimate objective of any and every firm.

4. Selling process

The sales process – also known as a sales cycle – is the method your company follows to sell your product or service to customers. It involves a series of steps, from initial contact with a lead to the final sale.

The sales process is similar to developing a relationship with someone new. When you first meet, you get to know each other, learn what they like, and determine their goals. Along the way, you decide if you can work together and whether you are a match. If this is the case, the relationship can proceed and grow.

4.1 Importance of building a sales process

These are some benefits of building a sales process for your business:

You can optimize the structure of your sales team to support the sales process and identify the main challenges in the sales cycle.

It will be easier to onboard new sales personnel.

It helps you identify short-term and long-term goals and how each step in the sales process supports the next one.

It highlights where time and resources are being wasted, so you can remove activities with low return on investment and focus your efforts on activities with more positive returns.

It identifies the steps that need to be improved. This allows you to invest in training, education, and practice to get better in areas of weakness, which will help match your success in other parts of the sales process.

4.2 The 7-step sales process

Prospecting

Preparation

Approach

Presentation

Handling objections

 Closing

Follow-up

If you are one of the 2.5 million employees in the United States working in sales, you know that even for the most natural salesperson, it can sometimes be difficult to turn potential leads into closed sales. Across industries, you need different skills and knowledge to prove to your potential customers that your solution is best for their particular problem.

The seven-step sales process outlined in business textbooks is a good start, especially since leading sales ops teams attribute to 60% or more of their total pipeline in any quarter to actively designed and deployed sales plays. The seven- step sales process is not only a good start to customizing it to your particular business but more importantly, customizing it to your target customers as you move them through the sales funnel.

As the old adage goes, "Learn the rules like a pro so you can break them like an artist." Once you've mastered the seven steps of the sales process you might learn in a business class or sales seminar, then you can break the rules where necessary to create a sales process that may not necessarily follow procedure but gets results.

The textbook 7-step sales process

What are the seven steps of the sales process according to most sales masters? The following steps provide a good outline for what you should be doing to find potential customers, close the sale, and retain your clients for repeat business and referrals in the future.

THE SELLING PROCESS
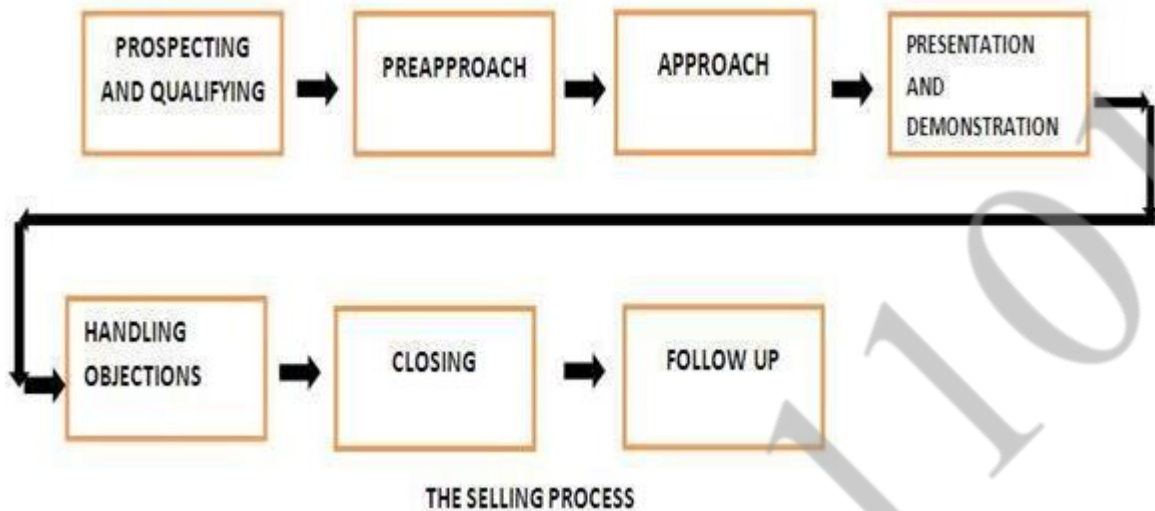
1. Prospecting

The first step in the sales process is prospecting. In this stage, you find potential customers and determine whether they have a need for your product or service— and whether they can afford what you offer. Evaluating whether the customers need your product or service and can afford it is known as qualifying.

Keep in mind that, in modern sales, it's not enough to find one prospect at a company: There are an average of 6.8 customer stakeholders involved in a typical purchase, so you'll want to practice multi-threading, or connecting with multipledecision-makers on the purchasing side. Account maps are an effective way .

2. Preparation

The next step is preparing for initial contact with a potential customer, researching the market and collecting all relevant information regarding your product or service. Develop your sales presentation and tailor it to your potential client's particular needs. Preparation is key to setting you up for success. The better you understand your prospect and their needs, the better you can address their objections and set yourself apart from the competition.

3. Approach

Next, make first contact with your client. This is called the approach. Sometimes this is a face-to-face meeting, sometimes it's over the phone. There are three common approach methods. Premium approach: Presenting your potential client with a gift at the beginning of your interaction

Question approach: Asking a question to get the prospect interested.

Product approach: Giving the prospect a sample or a free trial to review and evaluate your service

4.   Presentation

In the presentation phase, you actively demonstrate how your product or service meets the needs of your potential customer. The word presentation implies using PowerPoint and giving a salesy spiel, but it doesn't always have to be that way—you should actively listen to your customer's needs and then act and respond accordingly.

5.   Handling objections

Perhaps the most underrated step of the sales process is handling objections. This is where you listen to your prospect's concerns and address them. It's also where many unsuccessful salespeople drop out of the process—44% of salespeople abandoning pursuit after one rejection, 22% after two rejections, 14% after three,and 12% after four, even though 80% of sales require at least five follow-ups to convert. Successfully handling objections and alleviating concerns separates good salespeople from bad and great from good.

6. Closing

In the closing stage, you get the decision from the client to move forward. Depending on your business, you might try one of these three closing techniques.

Alternative choice close: Assuming the sale and offering the prospect a choice, where both options close the sale—for example, "Will you be paying the whole fee up front or in installments?" or "Will that be cash or charge?"

Extra inducement close: Offering something extra to get the prospect to close, such as a free month of service or a discount.

Standing room only close: Creating urgency by expressing that time is of the essence—for example, "The price will be going up after this month" or "We only have six spots left"

7.Follow-up

Once you have closed the sale, your job is not done. The follow-up stage keeps you in contact with customers you have closed, not only for potential repeat business but for referrals as well. And since retaining current customers is six to seven times less costly than acquiring new ones, maintaining relationships is key.

4.3 Prospect for potential customers

The first step is to prospect for customers, which requires some research. This stage has three components.

1.Create an ideal customer profile (ICP). The goal is to identify and understand your ideal customers. This helps you determine whom to contact and why you are contacting

them as potential customers. The ICP uses real data to create a fictional characterization of a client who:

Can provide your company with value (e.g., revenue, influence)

Your company can provide value to (e.g., return on investment, better service)

2.Identify potential leads. Use the ICP to create a list of potential leads that fit this profile. Use a variety of sources (e.g., online databases, social media) to develop a list of ideal client companies. Then create a list of prospects from these companies that your sales team can contact and qualify.

3.Perform initial qualification. First, qualify the company by conducting research to see if it meets the criteria that matter to you (e.g., company size, geography, industry, growth phase). Then qualify the prospects with an interview to determine if they are a good fit as a customer. Determine if the prospect has:

- ✓ A need for your product or service.
- ✓ The budget to purchase your product or service.
- ✓ The authority to make the purchasing decision.
- ✓ The timing to make the purchase

4. Make contact with prospects

- ✓ After identifying the ideal prospect, reach out to contact them. This step has two parts:
- ✓ Determine the best way to contact the prospect (e.g., telephone, email, social media).
- ✓ Reach out to the prospect. Make sure you are prepared (e.g., with a script, introduction and questions) before making contact. Introduce yourself and work on building trust, not making a sale.

5. Qualify prospects.

- ✓ Although you have already done your research to qualify the prospect before making contact, you still need to determine if they would make an ideal customer. This can only happen in a direct conversation with the prospect (either over the phone or in person).
- ✓ To qualify the prospect, learn more about them. Ask about their goals, budget, challenges and other issues that will help you to make your decision. Make sure that the person you are speaking with has the power to make decisions on doing business with you. When speaking with the prospect, identify opportunities to provide value.
- ✓ Qualifying the prospect involves confirming whether they meet the criteria of a good customer. If they are not a good fit, tell the prospect why. If they are still interested, determine why.

6..Nurture prospects.

Once you have qualified the prospect, demonstrate the relevance of your solution to them. This typically involves answering questions about your unique offer, the benefits you provide, and the problems you solve.

When answering the prospect's questions and learning about their needs, you have to nurture them along the process of making a decision. This involves: Moving the prospect along the stages of awareness

Unaware: The person does not know they have a problem.
Problem aware / pain aware: The person knows they have a problem but is not aware of a solution.
Solution aware: The person knows there is a solution but does not know about your product.
Product aware: The person knows about your product but does not know if it can solve their problem.
Most aware: The person knows a lot about your product but needs to know about its benefits.

- Educating the prospect about the product, service or industry Personalizing your communications
- Responding to common challenges

- Building your reputation with the prospect as someone who is helpful, responsible and reliable in your area of expertise.
  Some prospects may be both interested in your offering and qualified, but might not be ready or able to become a customer at this time. To nurture this type of prospect, stay in touch going forward and demonstrate your ability to help. This will help to keep you top of mind when they are ready to buy.

7.. Present your offer.

Use the information you have collected to this point to present the prospect with your best possible offer. Make the offer personalized, targeted and relevant to your prospect's needs. Craft the offer to address their challenges, budget and goals.

While the content of your offer is very important, how you present the offer can be the difference between success and failure. Consider your audience and the situation when deciding how to present your offer. Creativity can be very effective, but you should also focus on what works best for you given the experience of previous presentations.

8.. Overcome objections.

You've made the best possible offer – now it's up to the prospect to make the next move. The most common response is some type of objection to your offer, such as:

Price (e.g., too expensive for the value provided)
Risk (e.g., too "dangerous" to switch to a new solution)

Content of offer (e.g., offer does not provide enough detail) Contract terms (e.g., term is too long)

Ideally, you addressed the common objections during the nurturing phase or when creating the offer. However, you cannot always address every objection before the prospect makes it.

To overcome or address objections:

Be patient and measured in your response. Listen to the prospect's concerns objectively. Do not rush or pressure the prospect to move forward. Address objections that are related to each other. For example, if the prospect questions the value and price, go over everything you've included in the offer to show how the value you provide exceeds the price.

When you have explained your reasoning, ask the prospect if you have properly addressed their objection.

Read between the lines of generic objections (e.g., "We are not interested").

Ask more questions to determine the real reasons behind each objection. Listen carefully to the answers before responding.

9 .Close the deal.

Once you have overcome all objections, you can close the deal to make the sale.

First, work on sealing the deal. The goal is to confirm the prospect's engagement and work toward the next steps. The key is to make it easier for the prospect to say yes to the deal. Prime the prospect by reminding them how they will achieve a specific goal in purchasing your product or service.

To close the deal:

Ask a direct question or make a direct statement (e.g., "Would you like to sign the deal now?").

Ask an indirect question (e.g., "Are you satisfied with what is included in the offer?").

Provide an incentive to close the deal (e.g., add a sign-up bonus). Offer a free trial period (e.g., "Try it for one week").

Emphasize the urgency or scarcity of the offer (e.g., "This is a limited-time offer"). Ask what else the prospect requires to make a decision. When the prospect has committed to the purchase, answer any additional questions they have and give them details on the next steps. Provide a written agreement and summary of the conversation so that their supervisor or other stakeholders can review it for accuracy.

If the prospect still responds with "not yet" or "not now" for reasons beyond your control (or theirs), then return the prospect to the nurturing stage. Stay in touch and follow up with prospects who are not ready to purchase.

4.4 How to implement a sales process

Consider the following approach to implement the sales process in your organization.

1.Understand the customer.

The sales process begins with the buyer. To implement an effective sales process, you must understand the buyer and then design your sales process to address their goals, motivations, and needs. This requires identifying and then answering their "why" question. For instance, why is the buyer looking for a solution? Why are they looking to you for the solution?
Build a sales process to help your salespeople find the answer to the key question. Conduct interviews with buyers and salespeople and perform industry research to find the answers to include in the process.

2.Set milestones.

Once you've defined the stages of your sales process, establish the key steps and milestones within those stages. A milestone could be identifying where the buyer is in the sales process or engaging with stakeholders within a certain time period. Score each milestone to determine how many resources to invest into that part of the sales process. When you set a milestone for each stage, train salespeople to meet that milestone at the assigned stage. This will prevent them from skipping steps or taking the wrong approach at the wrong time (such as talking about the price too soon). Instructing salespeople on when and how to do handoffs will also help correct problems in the sales process. This simplifies the process of helping buyers move from one stage to the next.

3.Apply skills and resources.

Build skills, resources and activities into the sales process to help your salespeople move to the next milestone. Resources could include brochures, case studies and whitepapers for a salesperson to share with customers. Provide your salespeople with specific training for particular milestones or have them engage in activities for other milestones.

4.Iterate and improve.

A sales process is not static; it should be refined and improved over time. Get feedback from salespeople, measure buyer behavior, and track and analyze sales data to evaluate the effectiveness of your sales process. Use the results to solidify the successful activities and resources within the sales process, implement activities and processes to prevent negative outcomes, and remove activities and resources that do not advance the sales process. This will keep the sales process relevant, actionable and efficient.
By constantly iterating and improving your sales process, you will: Reduce the time it takes to onboard new salespeople.

5.Increase the percentage of successful sales.

- Minimize costly mistakes.
- Improve sales forecasting.
- Reach sales targets on a more consistent basis.
- Align your technology and systems with the sales process.
  It's important to equip your salespeople with technology (such as CRM software) that enables them to perform each step of the sales process efficiently. However, software tools alone won't make salespeople more effective or encourage them to follow best practices. You need to combine the technology with supportive systems, guidance and resources.
- Provide technology that streamlines the sales process, collects and organizes information on customers, and lists the required activities for salespeople to follow.
- Create systems and resources to support the sales team's use of the technology during the sales process, such as these:
- Checklists to make sure all steps are performed in order
- Content and video to demonstrate the importance of the stages and milestones Buyer-focused content tied to where they are in the sales process
- Reminders to prevent salespeople from skipping steps
- Training content for each step in the sales process

5. Sales Planning

Sales planning is a set of strategies that are designed to help sales teams reach their target sales quotas and help the company reach its overall sales goals. Sales planning helps to forecast the level of sales you want to achieve and outlines a plan to help you accomplish your goals. A sales plan covers past sales, risks, market conditions, your target personas, and plans for prospecting and selling. Sales planning occurs at various stages of the sales cycle. Generally, businesses set monthly or quarterly sales goals. Sales don't happen all on their own just because your sales manager sets goals. By defining the steps in a sales plan, sales managers can help their teams reach their targets and enjoy the rewards that come with collective success.

Another important part of the sales planning process is evaluating the company and understanding its position in the marketplace. Market conditions are ever-changing, so it's important to study them and to adjust your sales plan accordingly.

Sales plans typically account for short- and long-term planning. Goals without rewards aren't sufficient to incentivize each salesperson to reach for the sky. The right tools and sales strategies go a long way toward motivating salespeople to reach their targets.

As salespeople reach their goals, you'll want to set new ones. Every time you set new targets, it's appropriate to amend your sales plan. Changes to your sales

plan may also mean that you need to change how your company allocates resources to ensure that your salespeople have the resources they need. If you haven't already invested in a cloud-based phone system and VoIP integrations, you might consider how setting up a sales call center, complete with call center software, could help streamline your sales activities and help you reach your goals more easily.

5.1 The Role of a Sales Plan for Your Business

In case there's any doubt about the important role that your sales plan plays in your business, you may be interested to know that a little more than half of sales professionals annually miss their sales quotas. Sales experts attribute this underwhelming percentage to the lack of strategic planning and failure to align sales goals in accordance with conditions in the marketplace.

Top sales performances only come about after proper planning and preparation. A well-thought-out plan streamlines sales tasks, which increases the efficiency and productivity of your sales teams.

For the best results, develop your sales plan well in advance. The best plans account for multiple levels. A common approach is to start with annual targets and break them down by the quarter, month, and week. Also, you'll need to pre-plan your resources, logistics, and activities for every part of your sales plan. These activities will give you a road map that leads to sales success.

Short-term planning and monitoring are important activities because they give you the opportunity to make changes to your sales plan based on weekly or monthly sales results. If your salespeople are way ahead of – or way behind on – your projections, short-term planning will ensure that sales goals are reasonable and attainable.

A good sales plan means that your sales teams can function as efficiently as possible. Inside sales reps and call center agents can easily use call center software for sales call planning, freeing up outside salespeople to focus on making in-person calls and closing sales.

5.2 Sales Planning & Aligning Your Sales Strategy

- When your business is experiencing a downturn in revenue generation, this is usually a sign of poor sales and marketing alignment. Misalignment between sales and marketing reduces revenue, negatively impacts customer experience, and makes it tougher for salespeople to meet their quotas.
- Here some interesting things about marketing and sales alignment:
- Sales and marketing productivity decline when there is a misalignment between them.

- Alignment between sales and marketing improves the customer experience because it helps to improve customer service and to create a single customer journey.
- Salespeople don't always use marketing content when there's no alignment.

- Marketing software and sales automation software make it possible to develop data-driven sales and marketing plans.
- Alignment ensures that marketing and sales teams develop profiles of the same audience segments and target personas.
- Strong alignment means that marketing and sales messaging to customers are consistent and tell the same story.
- Sales and marketing alignment also has a positive impact on post-sale growth, retention, and brand loyalty.
- Overall, when sales and marketing teams align with each other, it positions your company to get the most value from prospects and customers. It's the best path to take your company to new heights.

5.3 How to Use Sales Planning Templates

A proven sales plan template should be part of your brand strategy because it will guide your business growth every step of the way. You could think of it as telling your sales story. Every story tells the who, what, why, where, when, and how from beginning to end.
Let's break the strategic process down into five parts:

1.Goal setting
2.Sales forecasting
3.Market and customer research
4. Prospecting
5. Sales

- One process seamlessly dovetails with the next. Start with your high-level goals and then factor in the various market factors. Set realistic goals as a benchmark for forecasting reasonable goals in the future. You'll need to base your goals on several things, including the size of the market, your annual company goals, your sales teams' experience, and the resources that you have available.
- A cloud-based phone system offers dashboard analytics that gives you metrics such as the number of inbound calls and outbound calls and the average call length. This will allow you to set standards for your call agents. Also, it will help you to scale your contact center so that it's not over- or understaffed.
- Marketing and customer research is an important activity that helps you position your company properly for business growth. The right data will determine your niche markets so you can start building traction with a receptive audience. Your niche encompasses your products, content, culture, and branding.

- The next step is to identify the most likely sources for finding high-quality leads so that you can start building a quality prospect list. It's also a great idea to leverage current client relationships as you build your prospecting plan.

5.4 Importance of sales planning

- Sales planning is an important aspect of business that identifies current issues, such as a lack in sales, and seeks to find solutions or develop strategies. Sales planning takes advantage of new opportunities, such as when a company develops a new product, to create brand awareness or interest. Sales plans address various sales opportunities and the plan's objectives may vary depending on whether the company sells directly to the consumer, or to another business.
Ideally, a sales plan:

- Define targets
- Creates strategies
- Identifies tactics
- Motivates teams
- Sets budgets to achieve targets
- Reviews goals and suggests improvements

6.0 Analytics applications in Marketing and Sales

- **Measure Performance of Marketing Campaigns**

  The most basic form of marketing analytics is to provide marketers with the tools to understand what business impact their marketing campaigns have. This task can range from something as straightforward as providing standard metrics (click- through rate, ROI, etc..) at the campaign level to an analysis as complex as developing a Market Mix Model to come up with the optimal marketing strategy to maximize profit.
- **Find Opportunities in Marketing Performance**
  While marketing performance analytics will let you know on the whole how a campaign performs, it isn't until someone digs in to many cuts of data to uncover whether there are certain types of users that respond better to particular marketing treatment - perhaps some campaigns work better in certain markets or on mobile. Marketing analysts mine and model your data to uncover nuggets that can be acted on by marketers.
- **Understand Your Customers**

Diving deep into customer demographics and behaviors can help you understand which are more likely to be successful. This information can then be used by marketers when selecting their target audiences. Through data mining and statistical modeling, marketing analysts can provide a rich understanding of your customers and what drives success.

- **Understand Your Competition**
  Market research is often within the domain of marketing analytics and it can help marketers understand the competition better and adjust their strategy accordingly.

Sales analytics applications
The full list of applications we have seen are:
- ✓ Sales forecasting
- ✓ Sales force management Sizing
- ✓ Geo-distribution
- ✓ Predictive/prescriptive lead scoring Customer contact analytics
- ✓ Sales rep compensation improvements
- ✓ Sales attribution between marketing and sales
- ✓ Sales process improvements
- ✓ Performance management

Additionally sales analytics enables numerous applications we listed above. Some of these applications have dramatic benefits:

- **Reduction of sales support activities**

  Sales reps spend more time on non-sales activities according to most research on the topic. These include making sales forecasts, prioritizing leads, deciding how to approach leads which can all be automated with sales analytics applications. To perform such tasks, sales reps can use behavioural analytics.

- **Improved prioritization**

  There are several levels of improved prioritization thanks to sales analytics: Predictive/prescriptive lead scoring techniques enable improved prioritization by sales reps.Sales rep compensation can be improved with advanced analytics enabling company to focus on successful sales reps.

- **Sales attribution models** allow the company to focus its resources appropriately between sales and marketing.
  Improved sales processes and practices
- Insights can lead managers to learn from top performers, improve their coaching and sales processes .