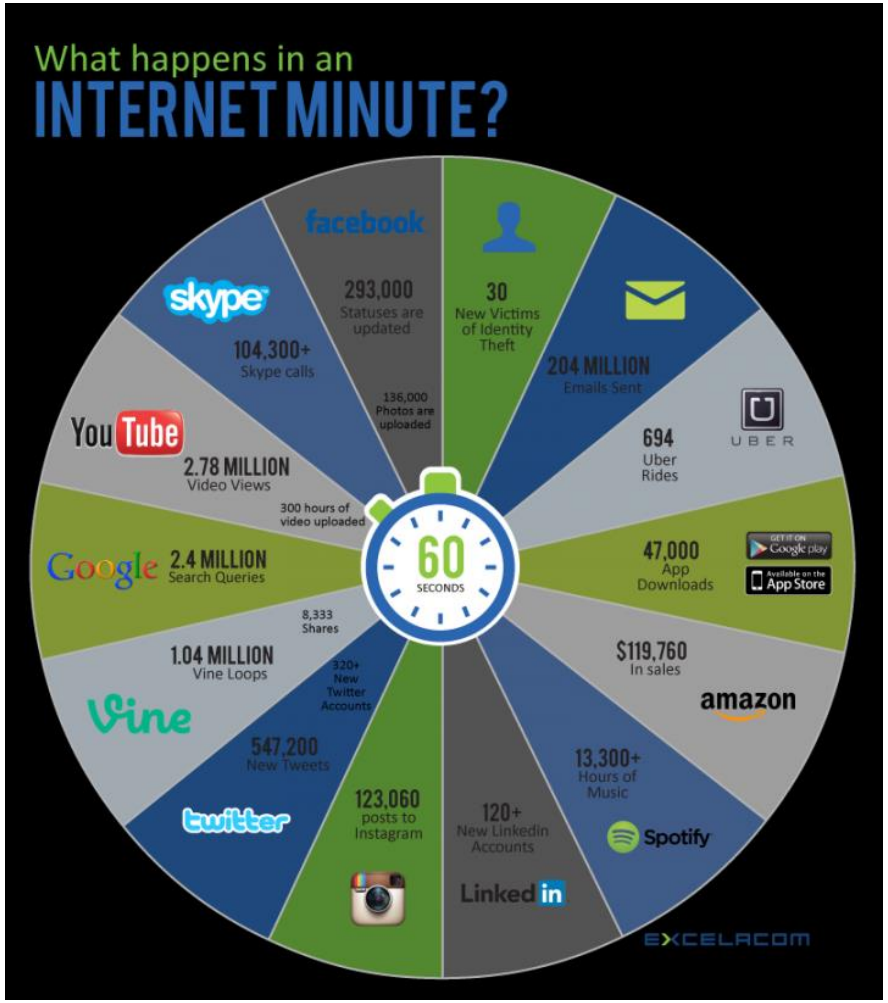
Big Data !!

What is Big Data?

- The data which can not be processed by traditional databases is called “Big data”

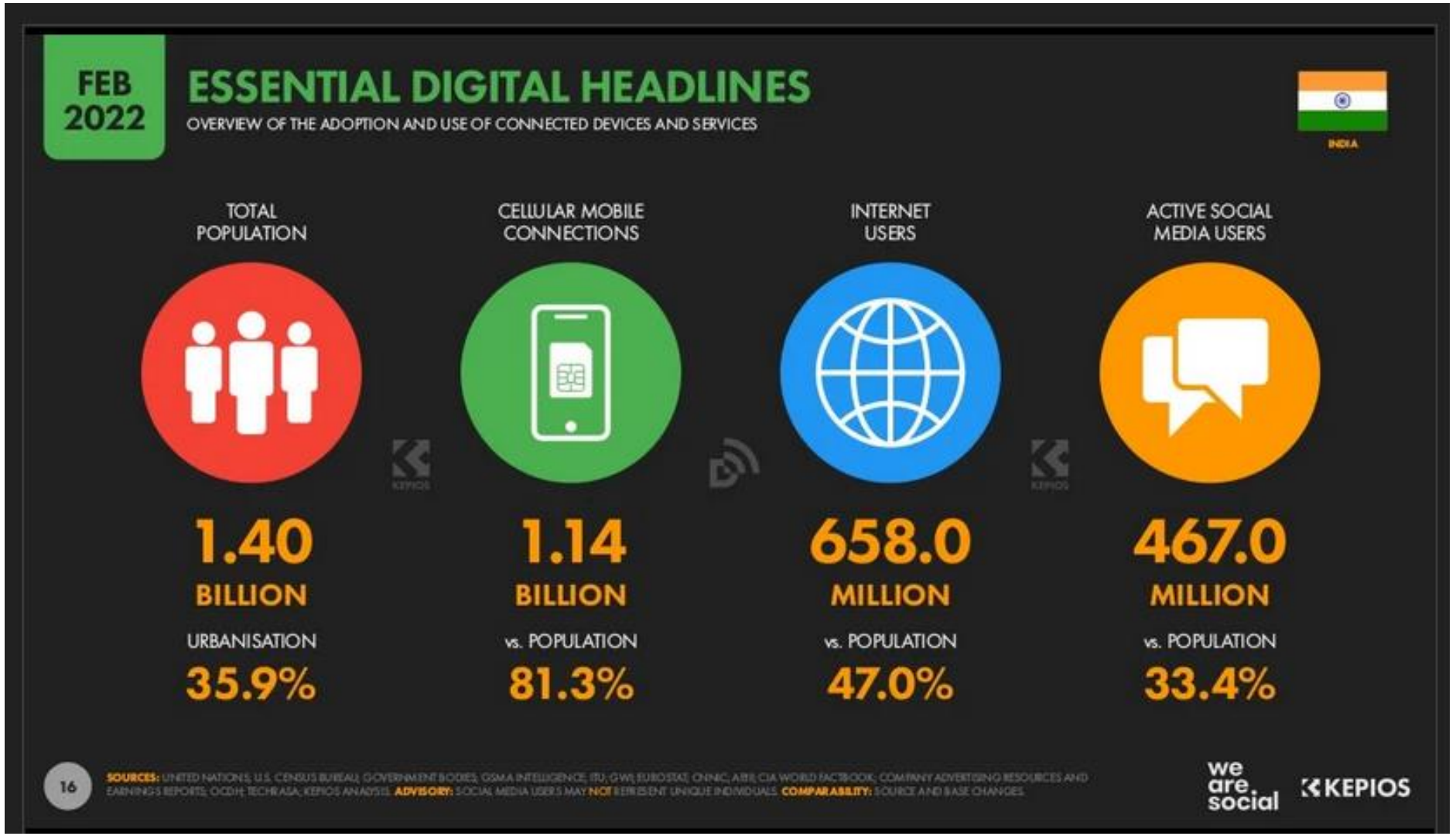
- MySql, Oracle, SQL etc are traditional databases.

The amount of data generated every minute is exploding



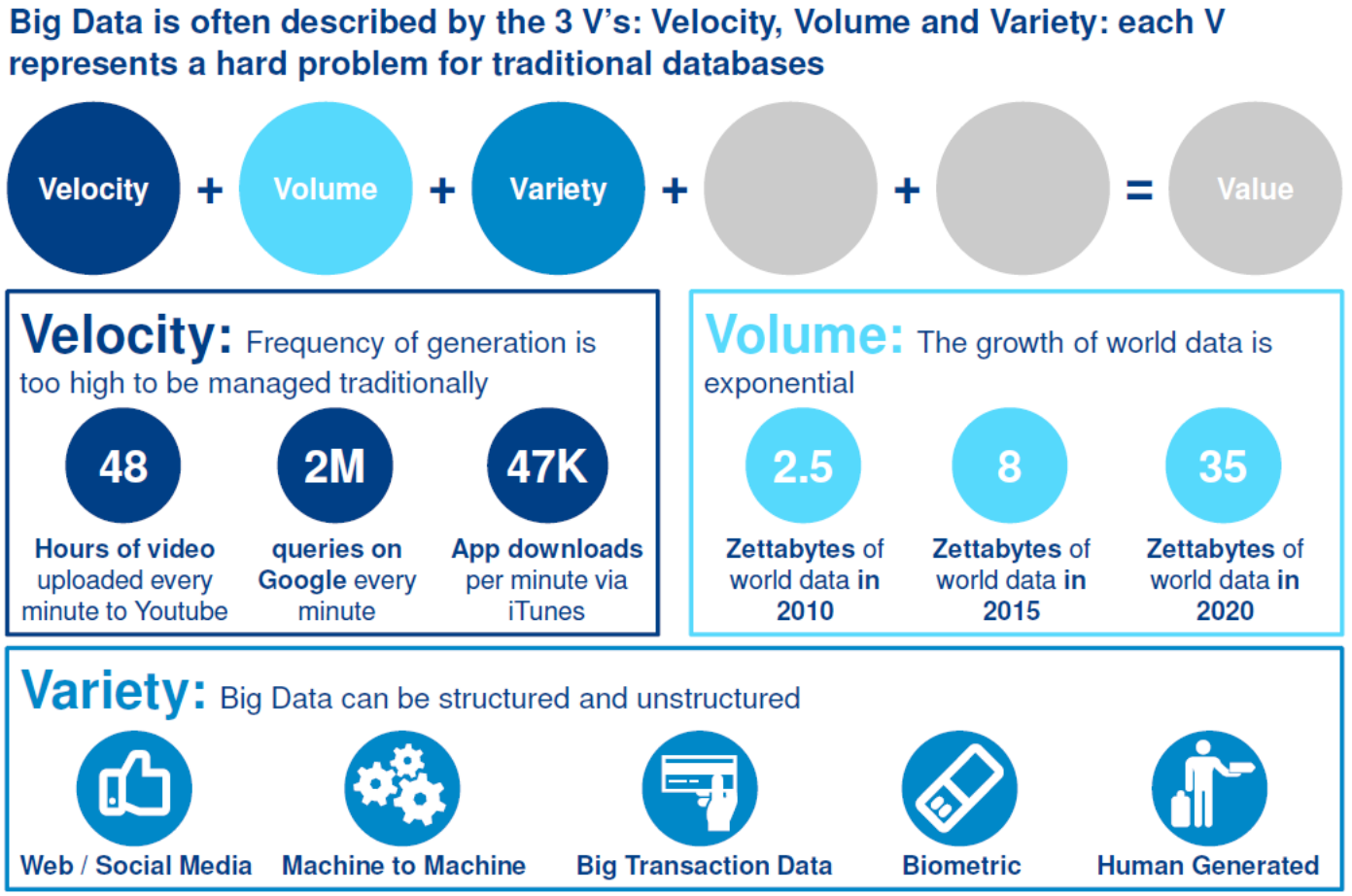
Can we handle it?

Hyper-connected India : Creating internet of people, things,… and everything



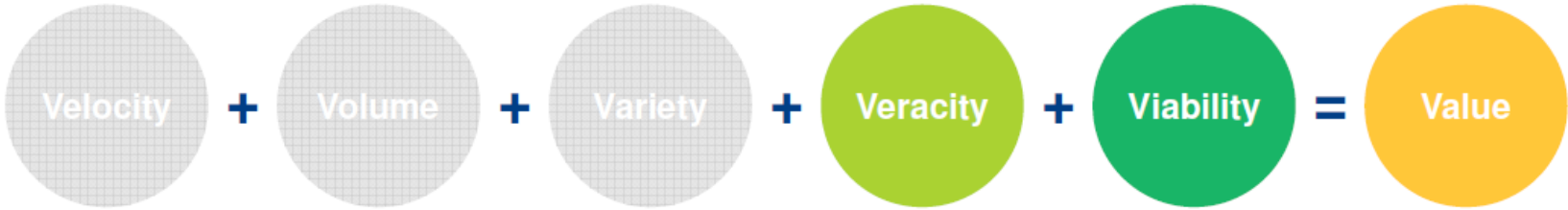
Are we ready for this change?

6 V's of Big Data



© 2013 Deloitte Touche Tohmatsu

However, additional V's are being proposed, to generate greater value: as the world of data grows, so does the challenge.



Veracity: Establishing trust in data

One Third of Business leaders **do not trust the information** they use

Uncertainty is due to **inconsistency, ambiguity, latency and approximation**

Viability: Relevance and Feasibility

Hypothesis - validation to determine **if the data will have a meaningful impact**

Long Term rewards and better outcomes from hidden relationships in data

Value: Measuring return on investments

Costs – there is a serious risk of simply creating Big Costs without creating strong value

Insights – Sophisticated queries, counterintuitive insights and unique learning

© 2013 Deloitte Touche Tohmatsu

6 V's of Big Data sum-up

- Volume - Vast amount of data
- Velocity - Speed at which data is generated/moves around.
- Variety - Different type of data
- Varacity- Trustworthiness of data. Quality
- Viability**:** Relevance and Feasibility
- Value - Ability to turn our data into value.

More reading: https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/

https://www.forbes.com/sites/oreillymedia/2012/01/19/volume-velocity-variety-what-you-need-to-know-about-big-data/

Types of data

Structured	Unstructured	Semi Structured
MySQL CSV XLX	Facebooks comments Tweets Image	HTML XML JSON

Unstructured

Internal

- Customer contact notes
- Customer
- survey text
- Scanned documents
- Web Cust. Experience
- More data sources
- Contracts

External

- Social media
- Web crawling
- Log Web
- Competitor
- scans

Structured

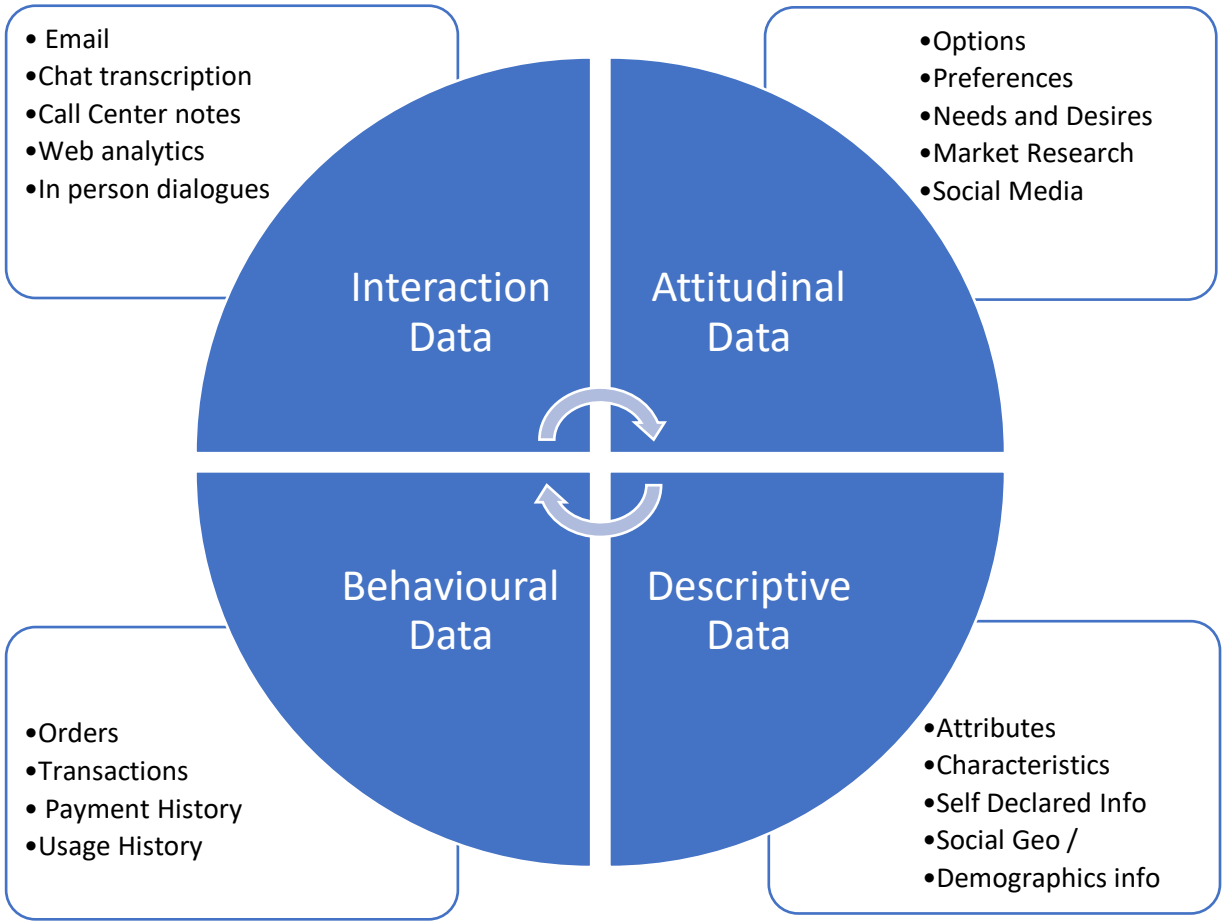
Internal

- Marketing data
  - P&L
  - Customer
  - survey results
- Customer contact logs
- Name, Address details
- Transaction history

External

- External (credit and risk) agency data
- Telephone directory
- Socio-demographic
- Price benchmark comparisons

Types of data



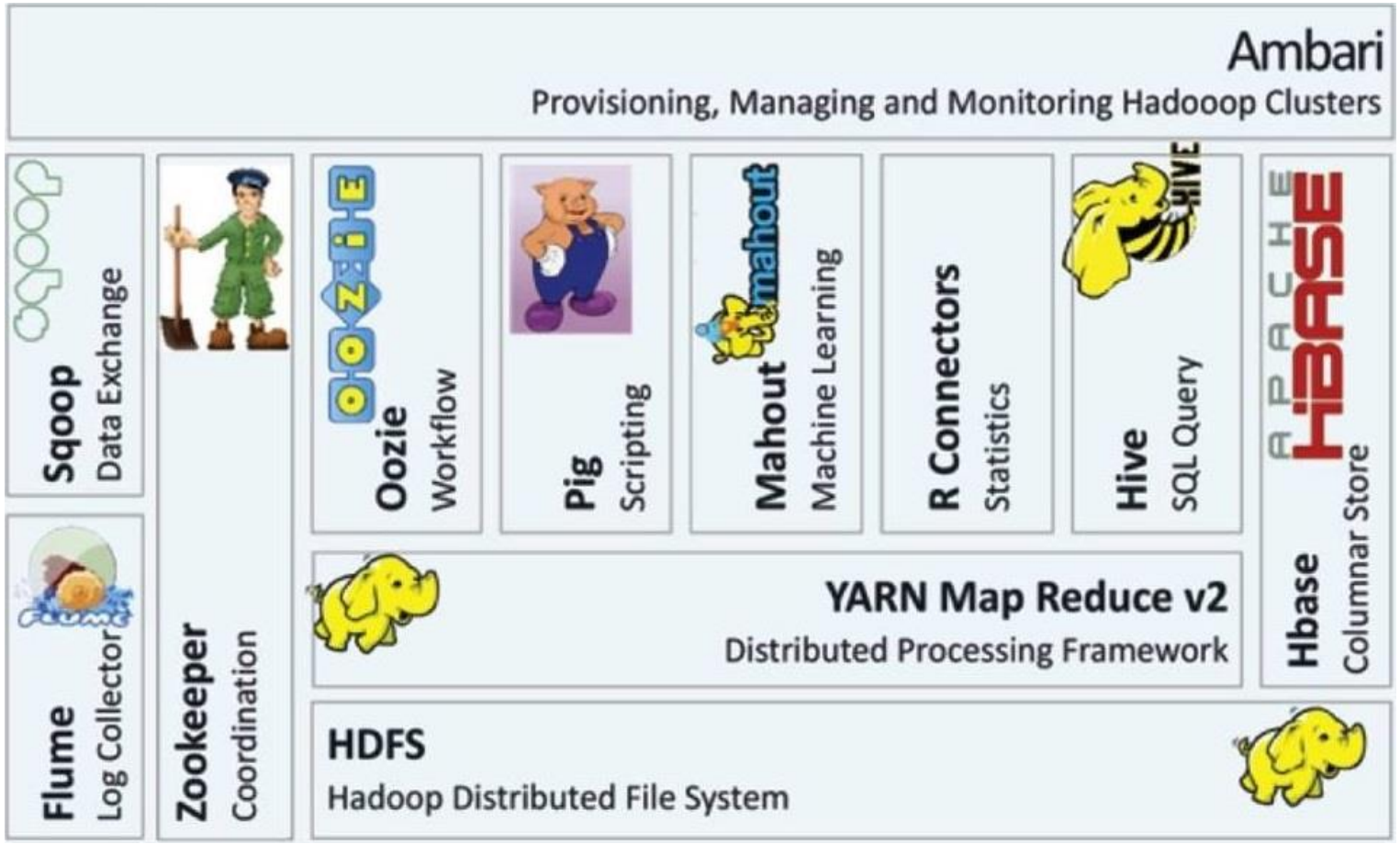
What Makes Big Data Valuable?

- Personalized Marketing
- Recommendation Engines
- Sentiment Analysis
- Mobile Advertising
- Consumer Growth
- Biomedical Applications
- Smart Cities

Data science tools and technology

- Open-Source Tools
  - Apache Hadoop
  - Apache Pig/Hive
  - Apache Spark
  - Apache Flume/Kafka
  - Matplotlib
- Paid/subscription-based Tools
- SAP HANA
- Tableau

Big data Infrastructure : Hadoop



WHAT IS DATA SCIENCE?

What's in it for you

- Need for Data Science
- What is Data Science?
- Data Science vs Business Intelligence
- The Prerequisites for learning Data Science
- What does a Data Scientist do?
- Data Science lifecycle with example
- Demand for Data Scientist

Need For Data Science

Need For Data Science

Need For Data Science

Need For Data Science

Dear Flyer, We regret to inform you that your flight has been cancelled due to delay from Airbus on account of engine delivery

1. Due to lack of data available, flights are often delayed or cancelled at the last minute
2. Due to improper route planning, customers don't get the flight for desired time and duration
3. Incorrect decisions in selection of right equipment leads to unplanned delays and cancellations

Need For Data Science

Using Data Science, we can achieve the following:

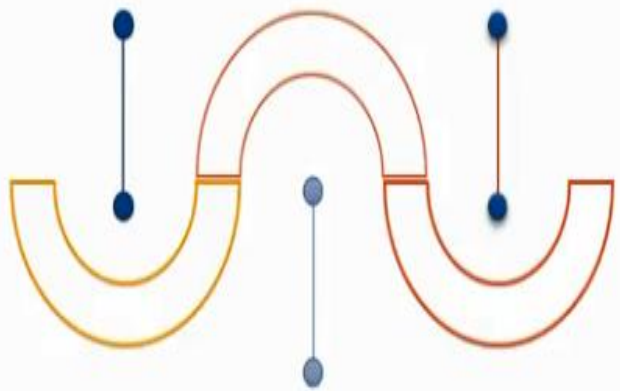
- Route Planning: Whether to schedule direct or connecting flights
- Predictive analytics model can be built to foresee flight delays
- Promotional offers depending on customer booking patterns
- Deciding which class of planes to purchase for better performance

Need For Data Science

Logistics companies like FedEx are using Data Science models for operational efficiency

Discover the best routes to ship

The best suited time to deliver



Need For Data Science

So Data Science is mainly needed for:

Better Decision Making

Whether A or B?

Predictive Analysis

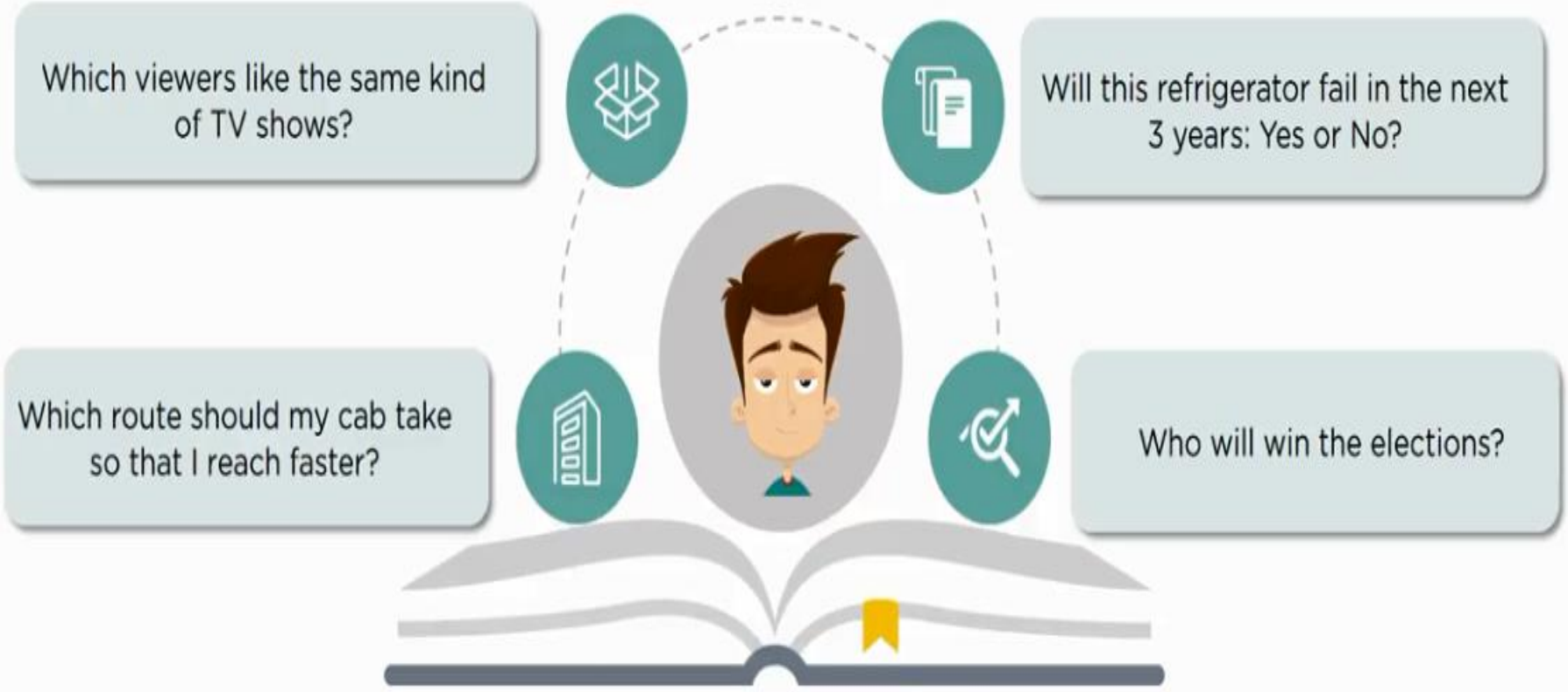
What will happen next?

Pattern Discovery

Is there any hidden information in the data?

What is Data Science?

Data Science can answer a lot of other questions as well!

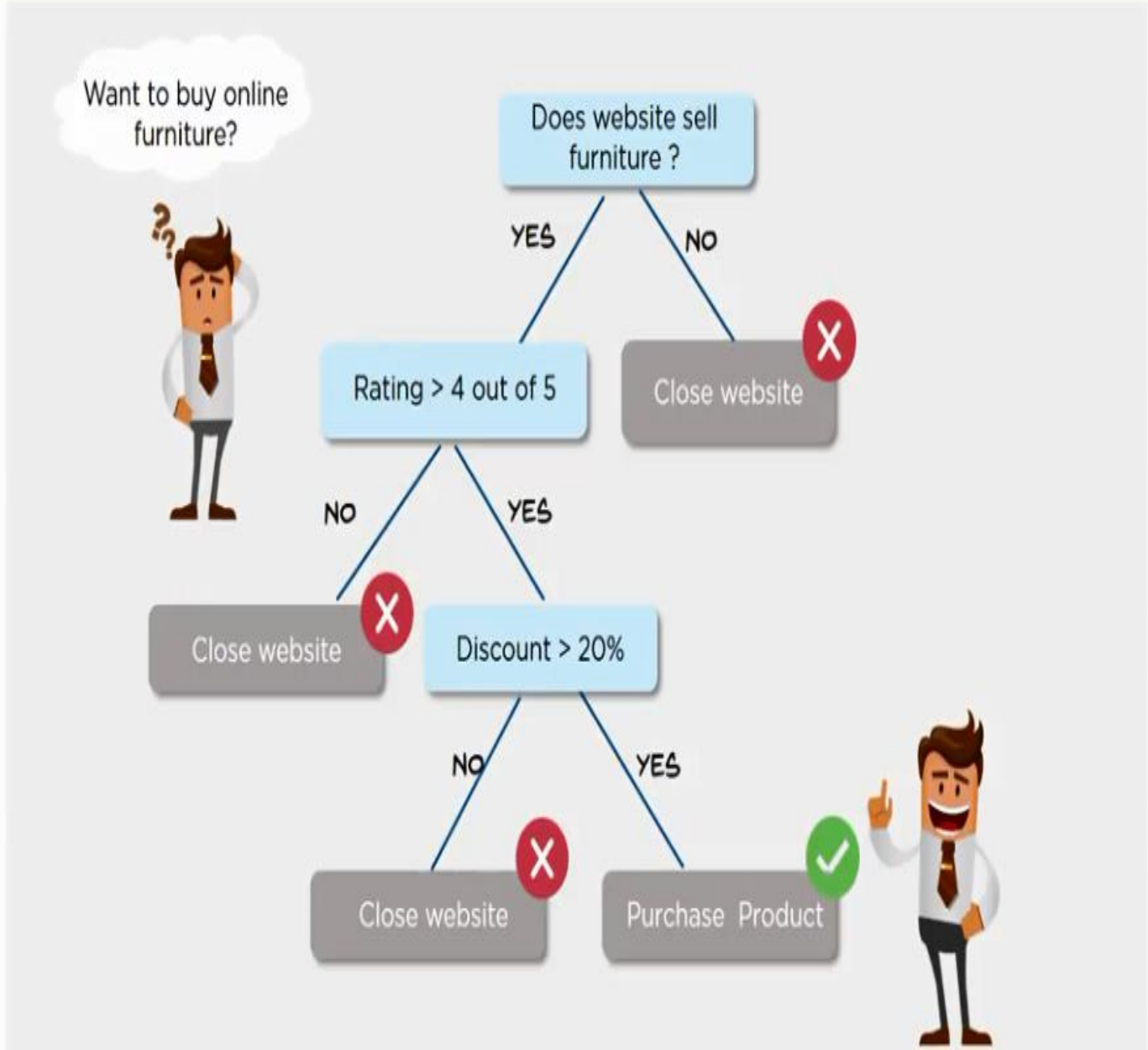


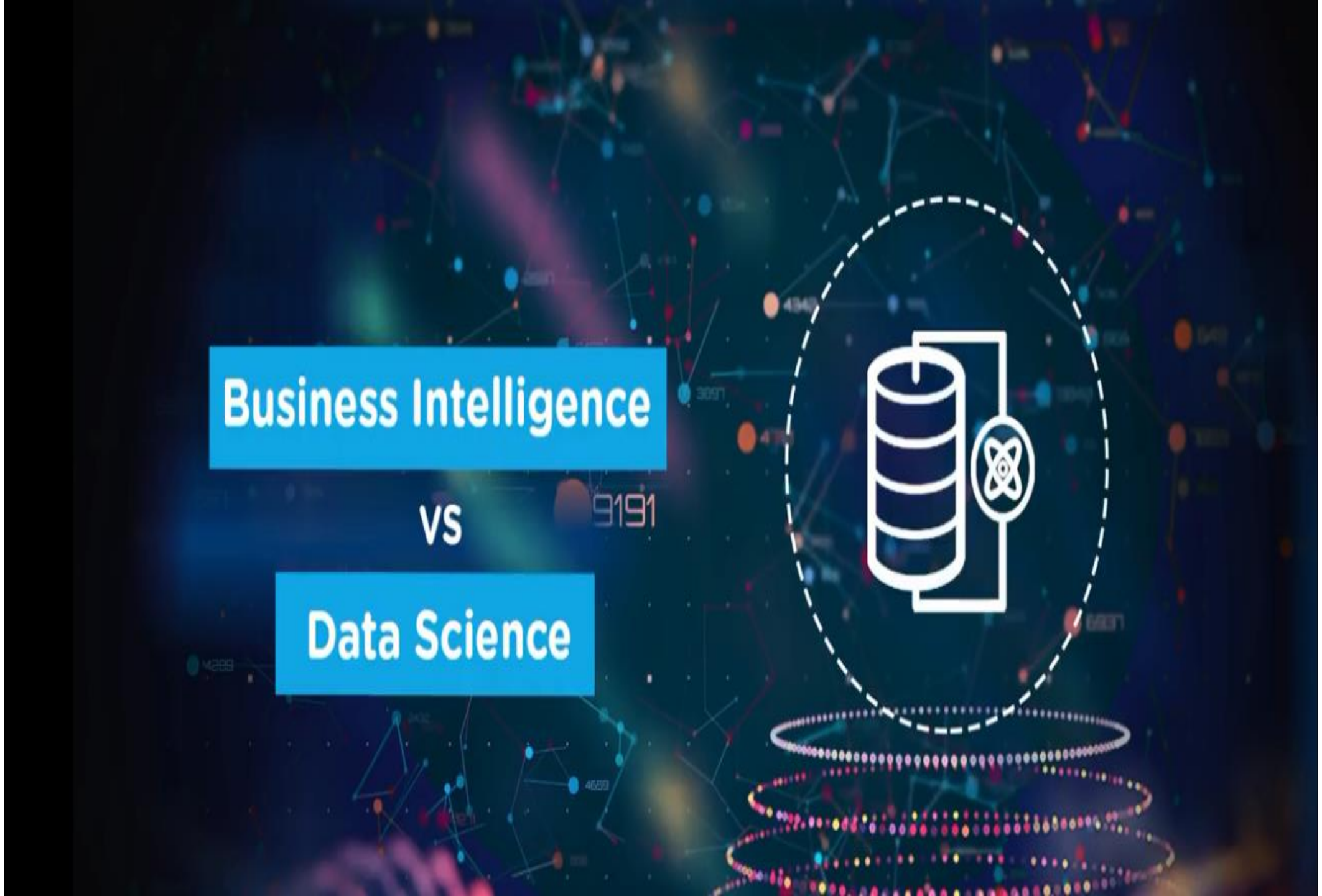
What is Data Science?

Suppose, you have decided to buy furniture online for your new office

How do you choose the right website?

What is Data Science?




Business Intelligence
vs
Data Science

Business Intelligence vs Data Science

Criterion	Business Intelligence	Data Science
Data Source	Structured data e.g. Data Warehouse	Unstructured data e.g. web logs

Business Intelligence vs Data Science

Criterion	Business Intelligence	Data Science
Data Source	Structured data e.g. Data Warehouse	Unstructured data e.g. web logs
Method	Analytical	Scientific

Business Intelligence vs Data Science

Criterion	Business Intelligence	Data Science
Data Source	Structured data e.g. Data Warehouse	Unstructured data e.g. web logs
Method	Analytical	Scientific
Skills	Statistics, Visualization	Statistics, Visualization, Machine Learning

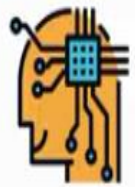
Business Intelligence vs Data Science

Criterion	Business Intelligence	Data Science
Data Source	Structured data e.g. Data Warehouse	Unstructured data e.g. web logs
Method	Analytical	Scientific
Skills	Statistics, Visualization	Statistics, Visualization, Machine Learning
Focus	Past and Present Data	Present Data and Future Predictions


Prerequisites For Data Science

Prerequisites for Data Science

1 MACHINE LEARNING



Machine learning is the backbone of Data Science. It is one of the many ways that Data Science uses to find solution to a problem

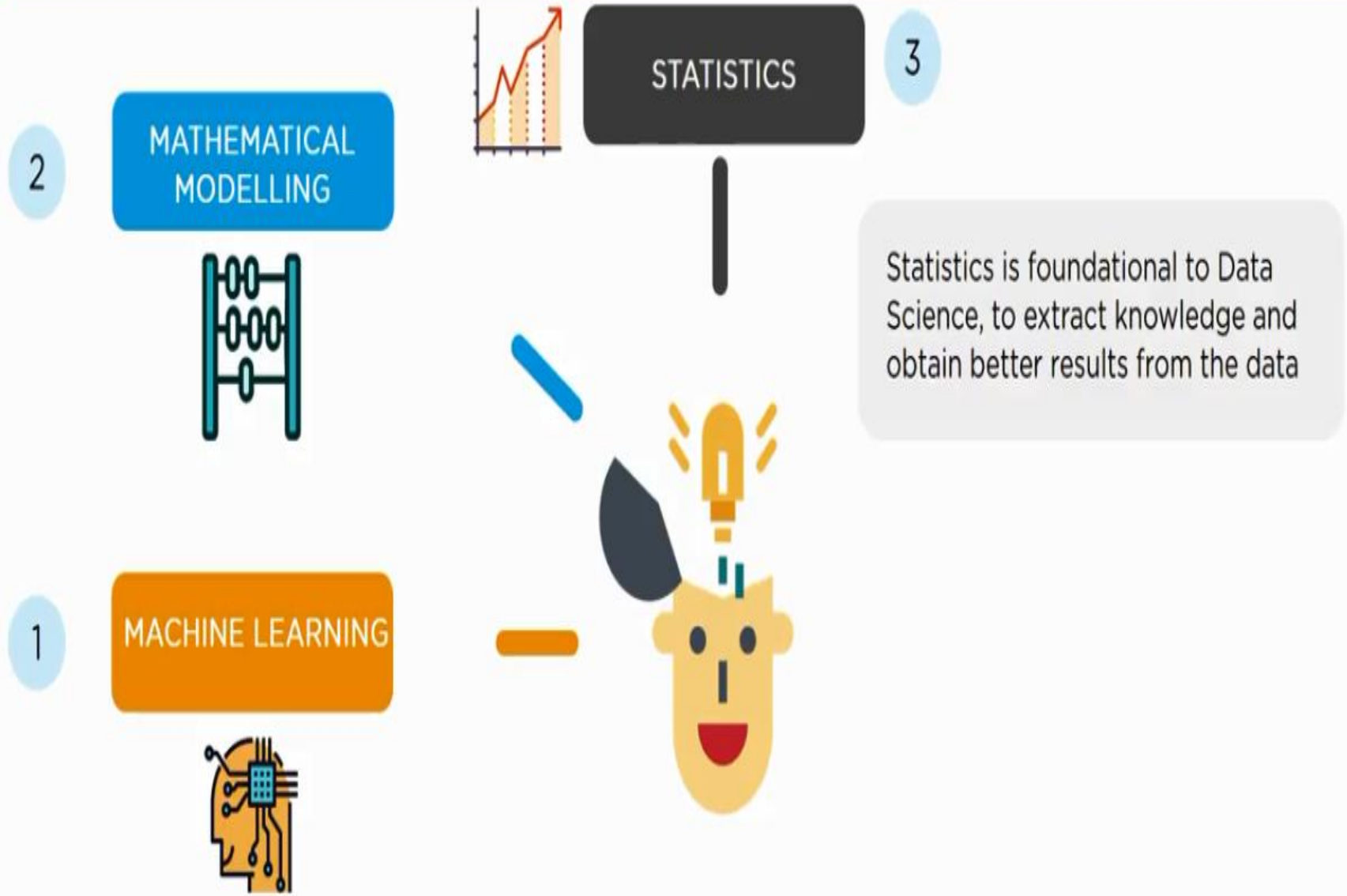
Prerequisites for Data Science

2 MATHEMATICAL MODELLING

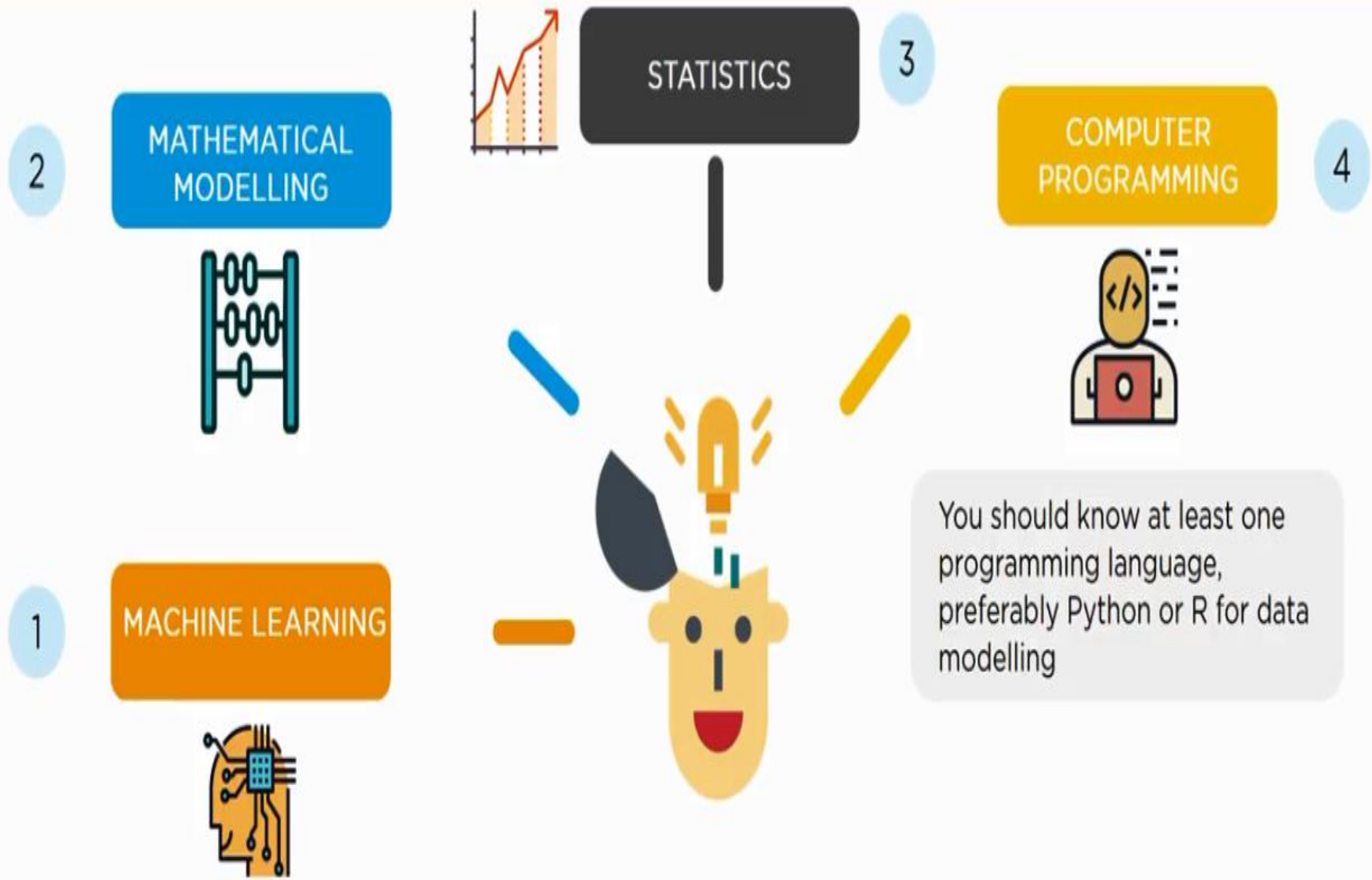
Mathematical Models can be extremely helpful to make fast calculations and predictions from what you know of your data

1 MACHINE LEARNING

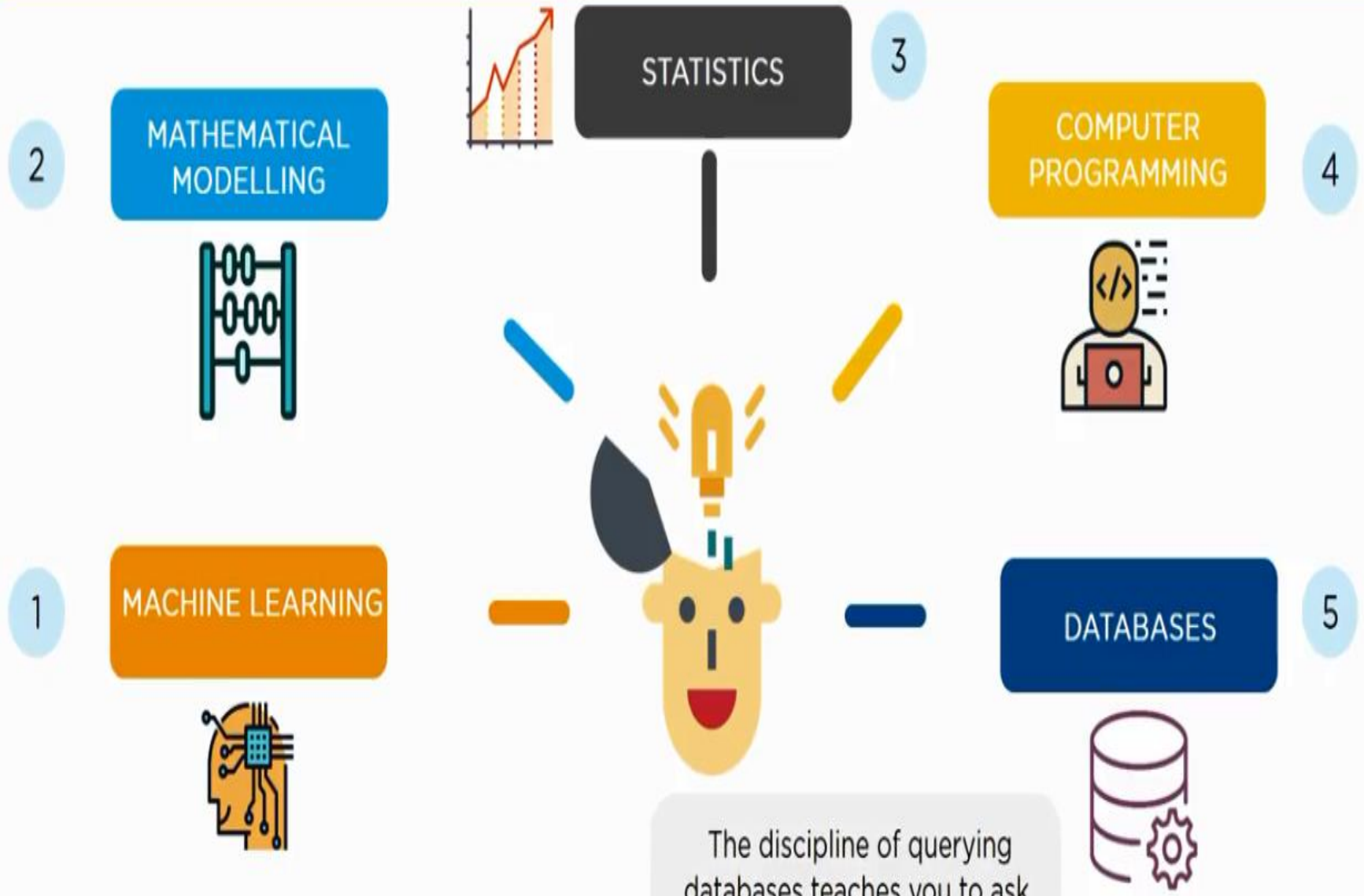
Prerequisites for Data Science



Prerequisites for Data Science



Prerequisites for Data Science

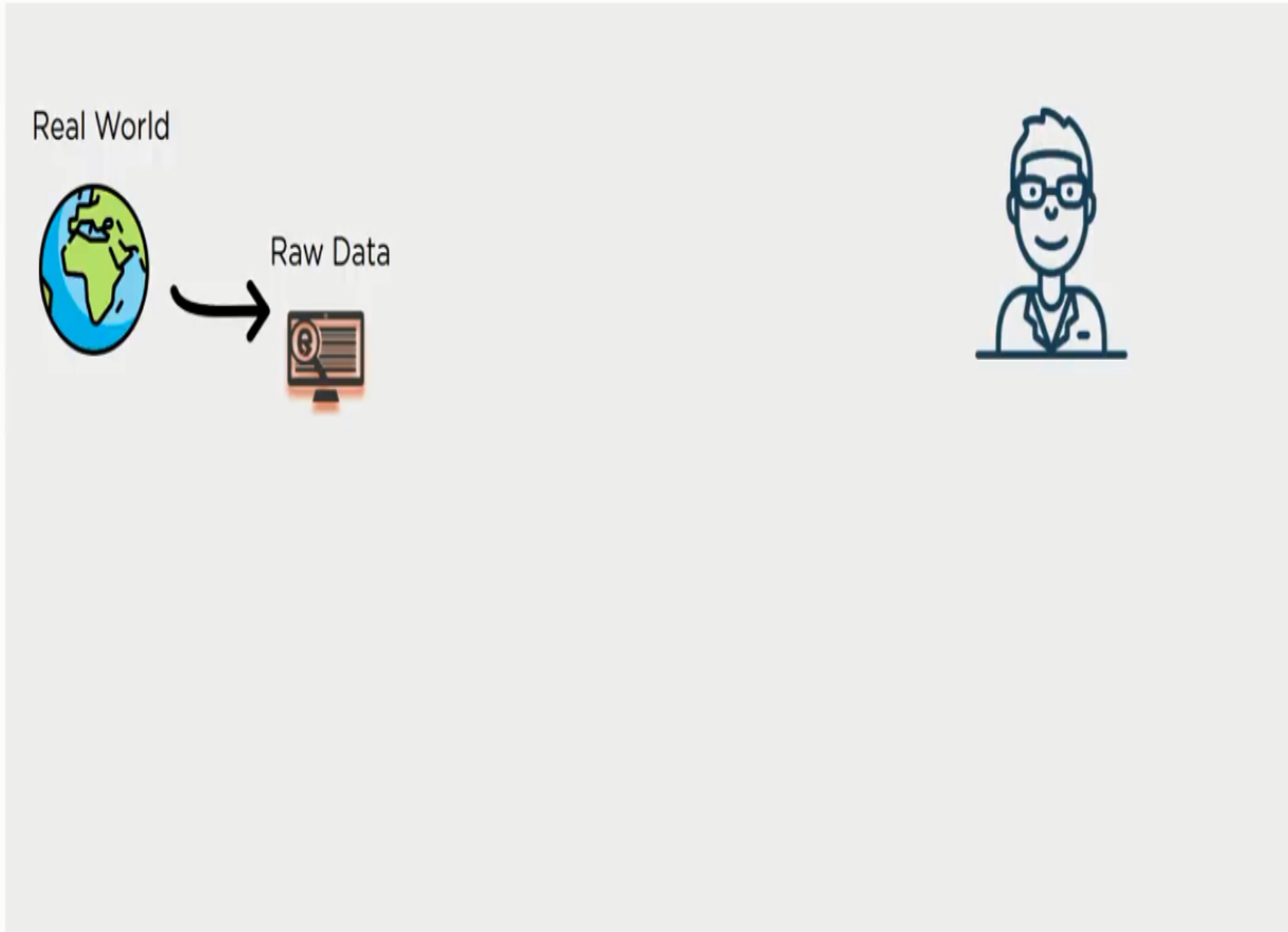



What Does A Data Scientist Do?

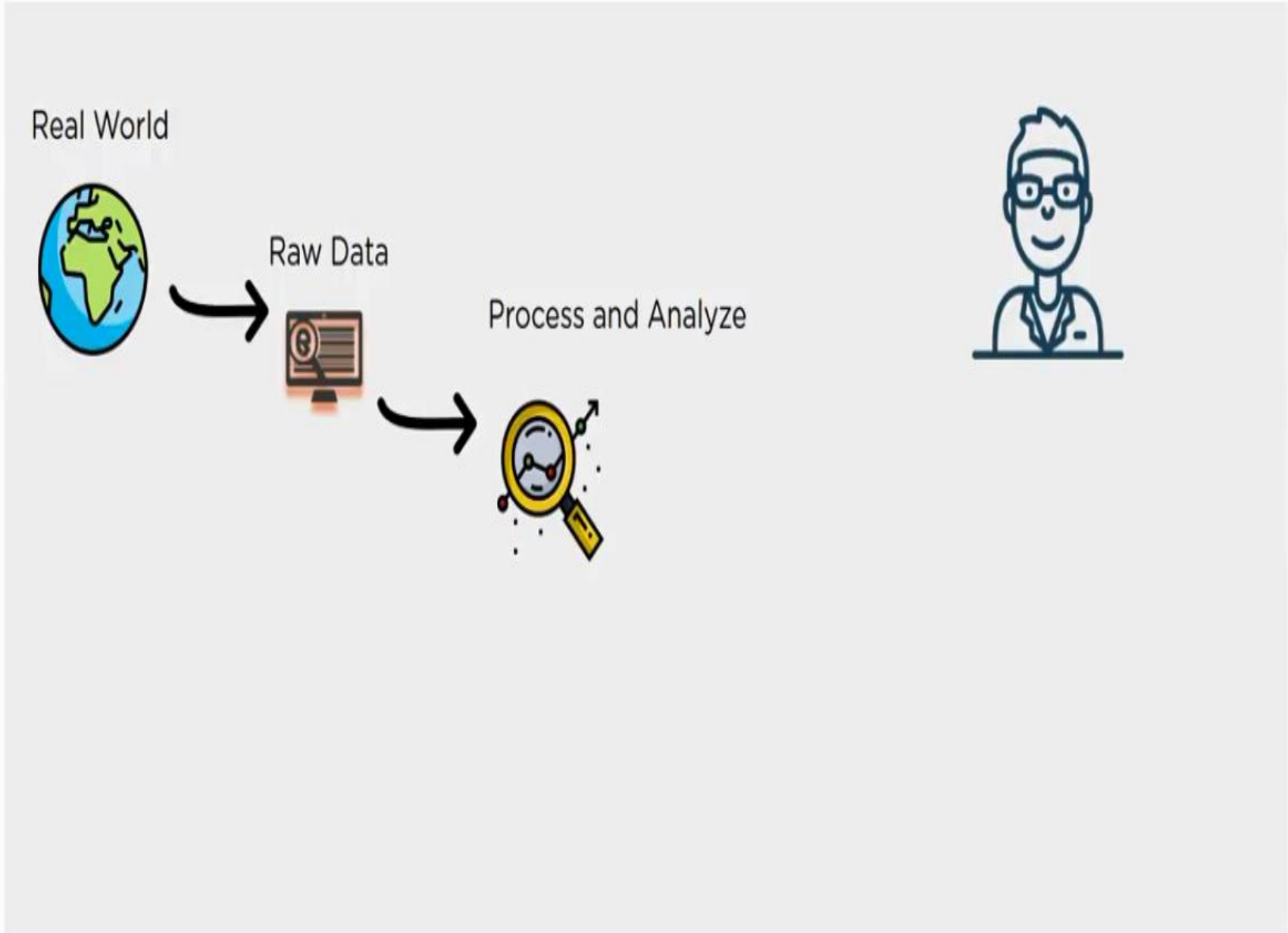
What does a Data Scientist do?

Real World

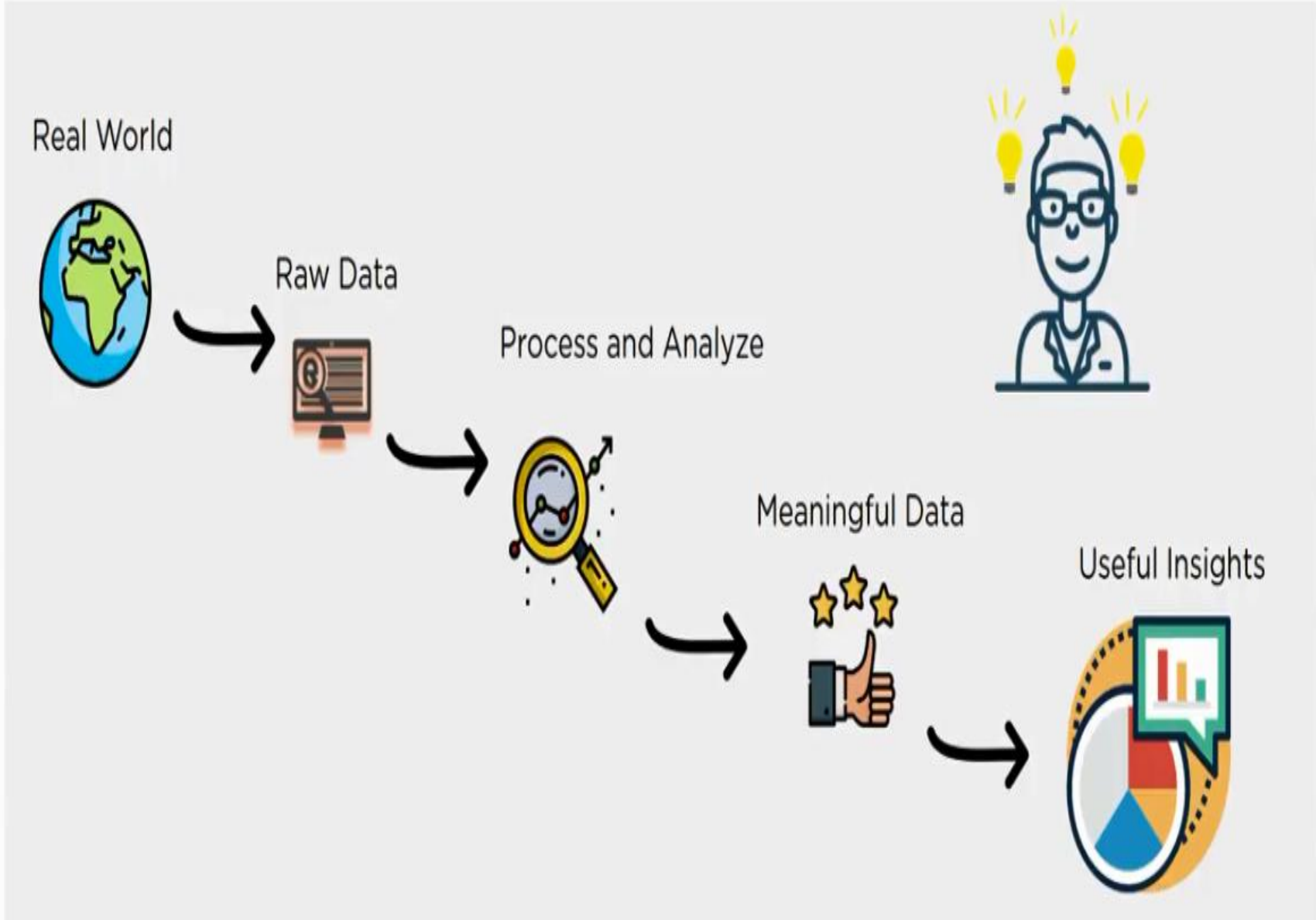
What does a Data Scientist do?



What does a Data Scientist do?

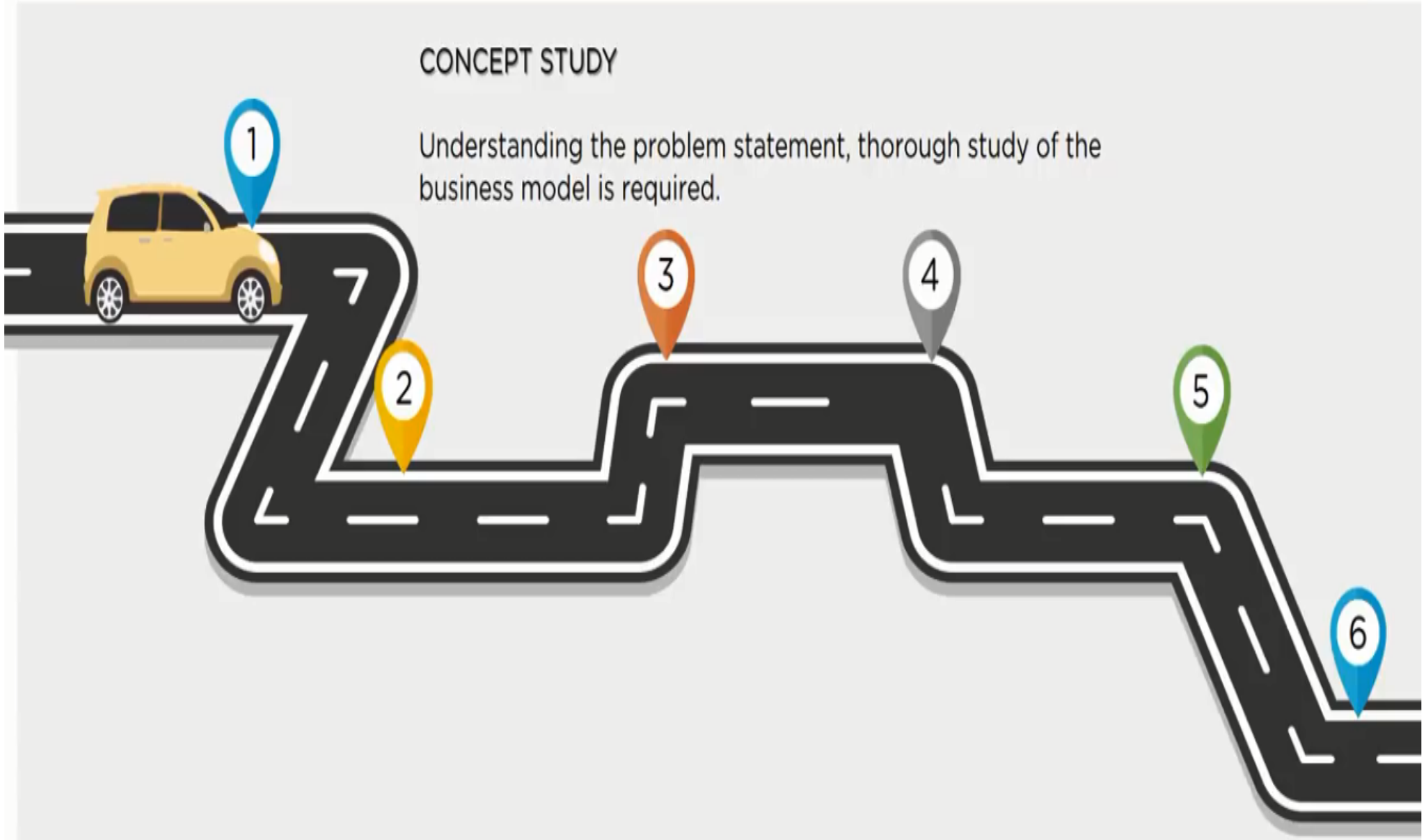


What does a Data Scientist do?




Data Science Lifecyle With Example

Concept Study - Life Cycle



Concept Study - Use Case

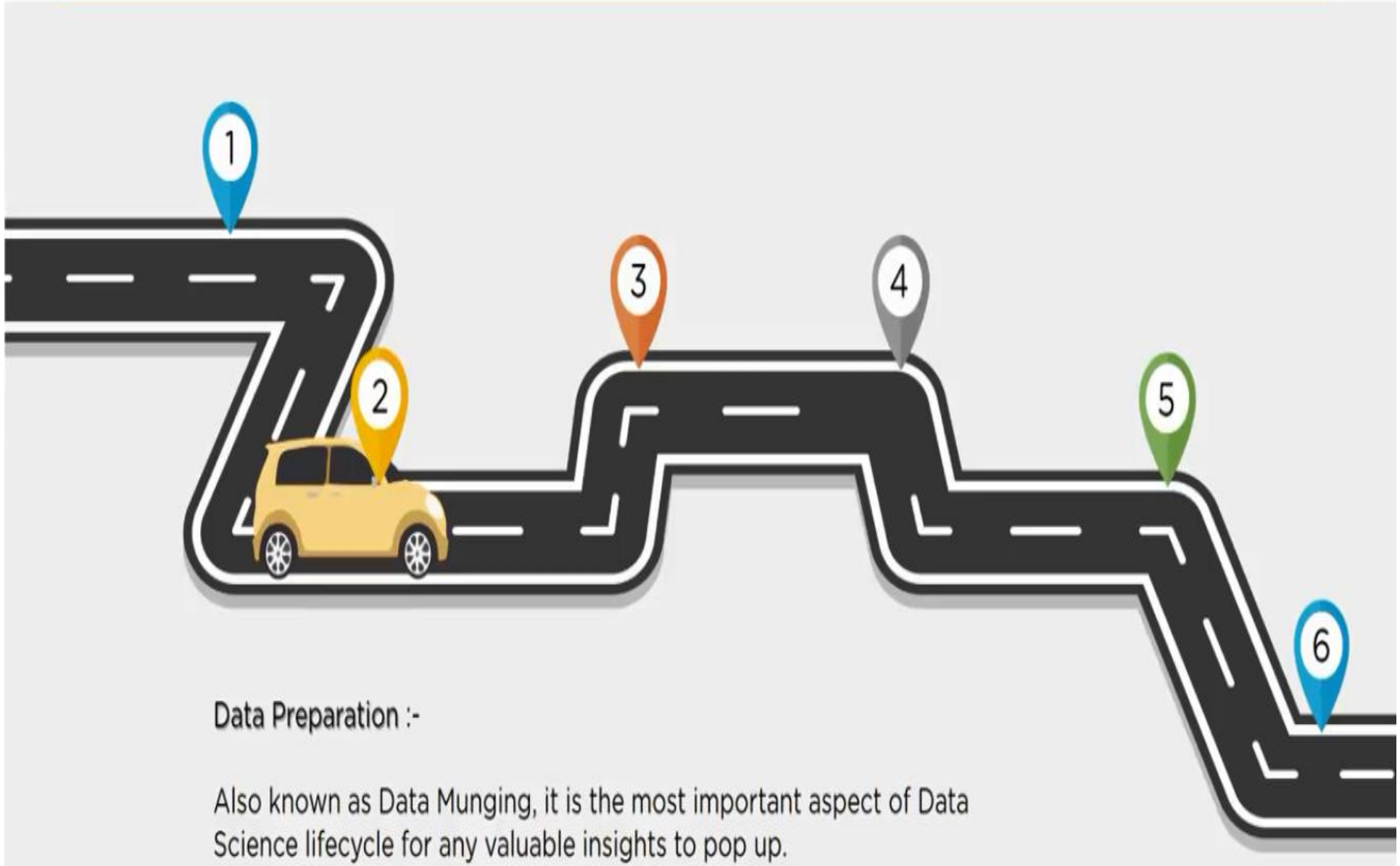
Concept of the task : Predict the price of 1.35 carat diamond

Get to know about the diamond industry, various terminologies used. Understand the business problem and collect RELEVANT and enough data

B	C
Carats	Price
1.01	7366
0.49	985
0.31	544
1.51	140
0.37	
0.73	3011
1.53	11413
0.56	1814
0.41	876
0.74	2690
0.63	
0.6	4172
Two	11764
1.1	4682
1.31	6171

Suppose, we get the price of diamonds from different diamond retailers. But we want to find out the price of 1.35 carat diamond.

Data Preparation - Life cycle



Data Preparation :-

Also known as Data Munging, it is the most important aspect of Data Science lifecycle for any valuable insights to pop up.

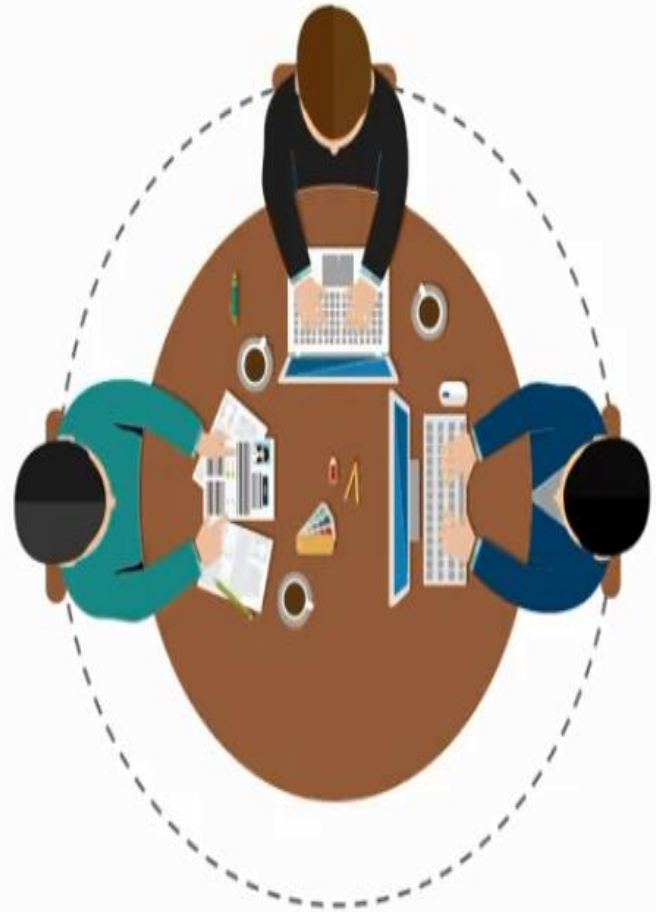
Data Preparation - Life cycle

Data Cleaning
Correcting inconsistent data by filling out missing values and smoothing out noisy data

Data Transformation
It involves normalization, transformation and aggregation of data using ETL methods

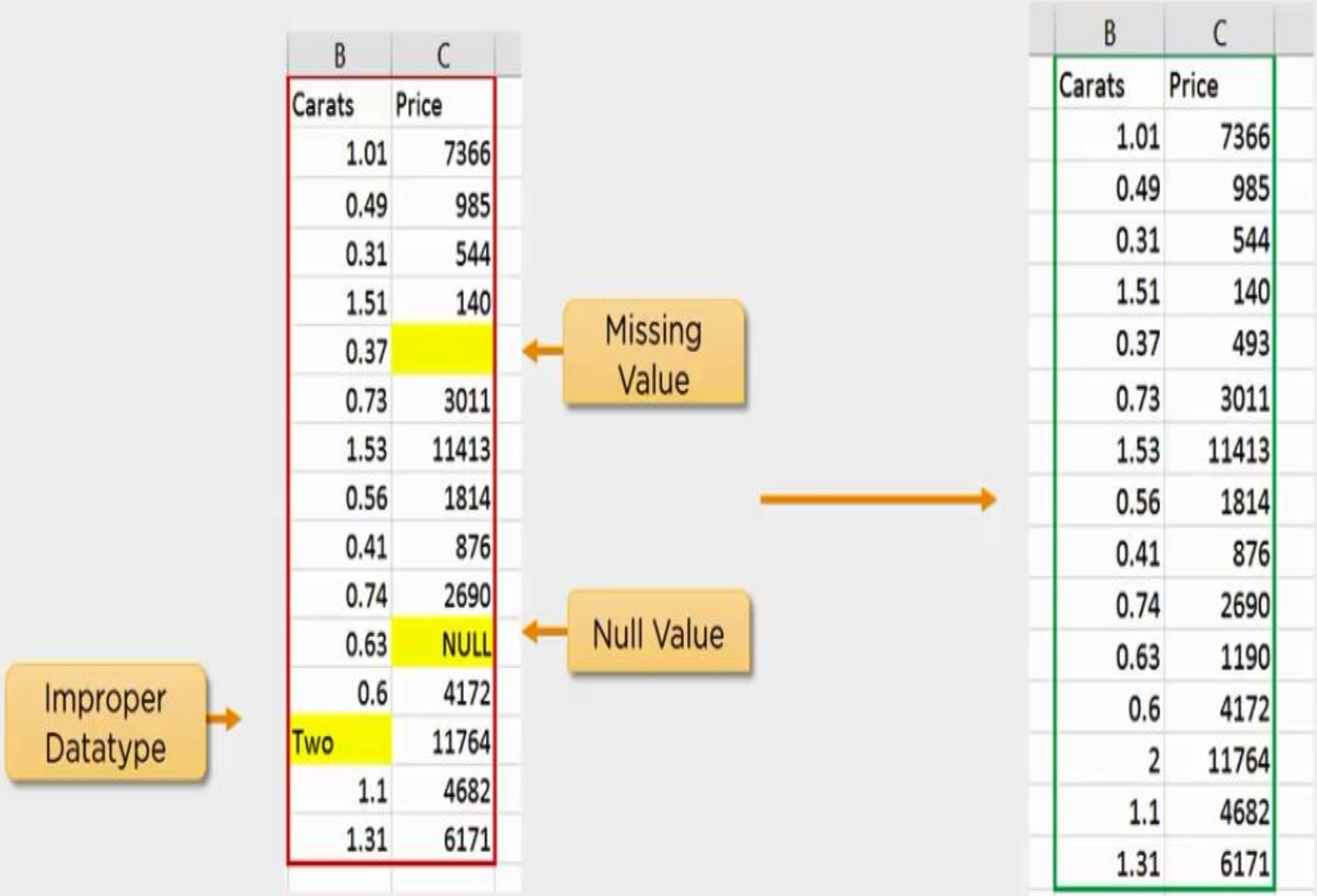
Data Reduction
Using various strategies, reducing the size of data but yielding the same outcome

Data Integration
Resolving any conflicts in the data and handling redundancies



Data Preparation - Use Case

Data preparation : Make the data clean and valuable.



Data Preparation - Use Case

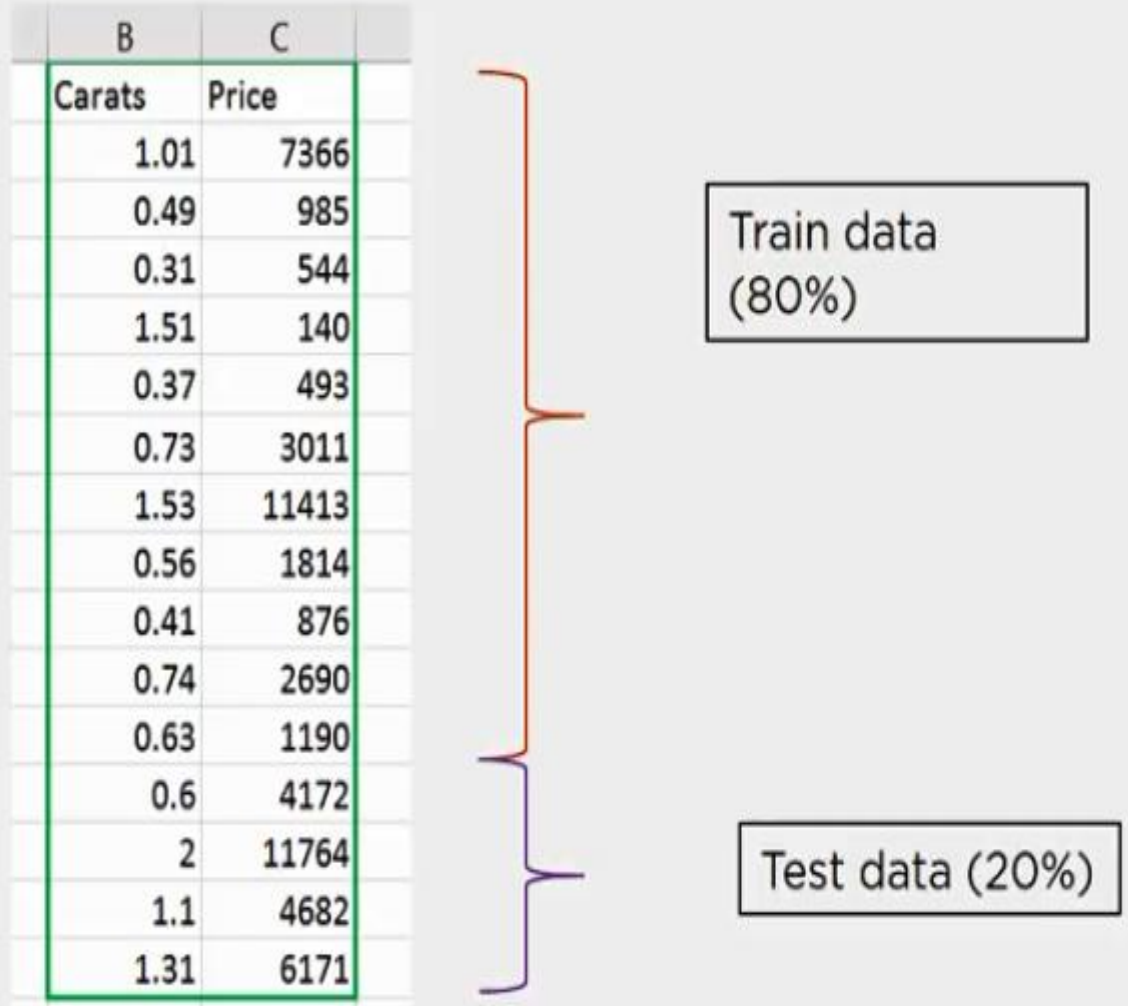
Ways to fill missing data values:

If dataset is huge, we can simply remove the rows with missing data vales. It is the quickest way.
i.e. we use the rest of the data to predict the values.

We can substitute missing values with mean of rest of the data using pandas' dataframe in Python.

i.e. df.mean()
df.fillna(mean)

- Split the data into train data and test data in the ratio of 80:20
- It is generally advised to divide the dataset into two random partition

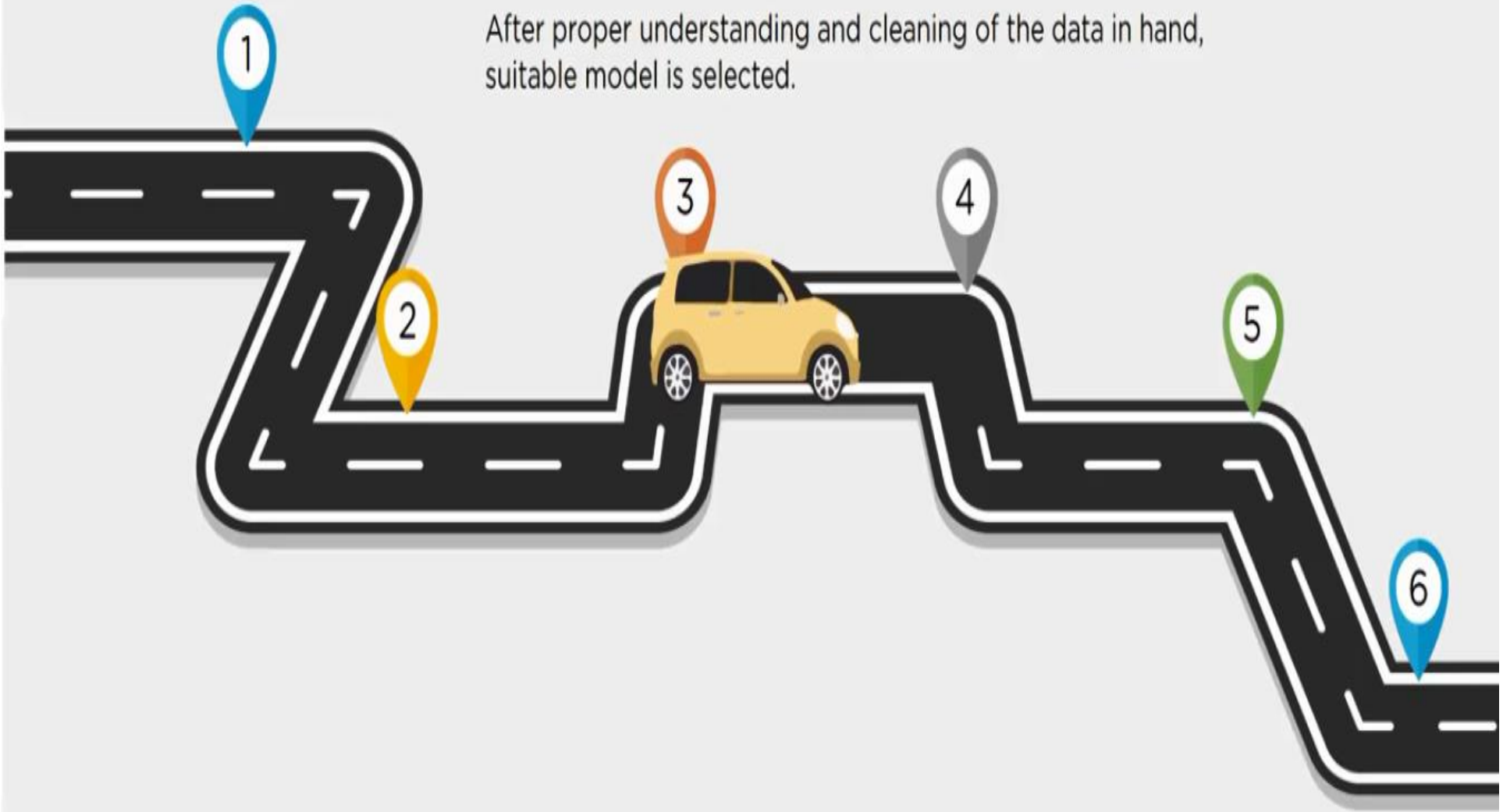


	B	C
	Carats	Price
	1.01	7366
	0.49	985
	0.31	544
	1.51	140
	0.37	493
	0.73	3011
	1.53	11413
	0.56	1814
	0.41	876
	0.74	2690
	0.63	1190
	0.6	4172
	2	11764
	1.1	4682
	1.31	6171

Model Planning - Life cycle

Model Planning:-

After proper understanding and cleaning of the data in hand, suitable model is selected.



Model Planning - Life cycle

Model Planning :

- This step involves Exploratory Data Analysis (EDA) to understand the relation between variables and to see what the data can tell us
- Key variables are selected

But what is Exploratory Data Analysis?

Definition : Deeper analysis of dataset to better understand the data.

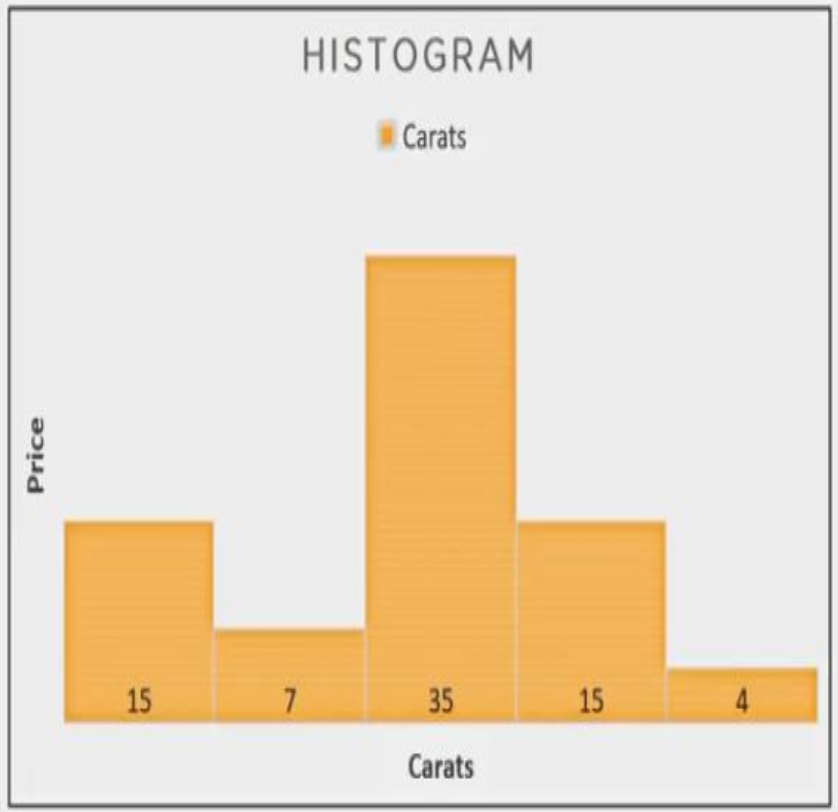
Goals :

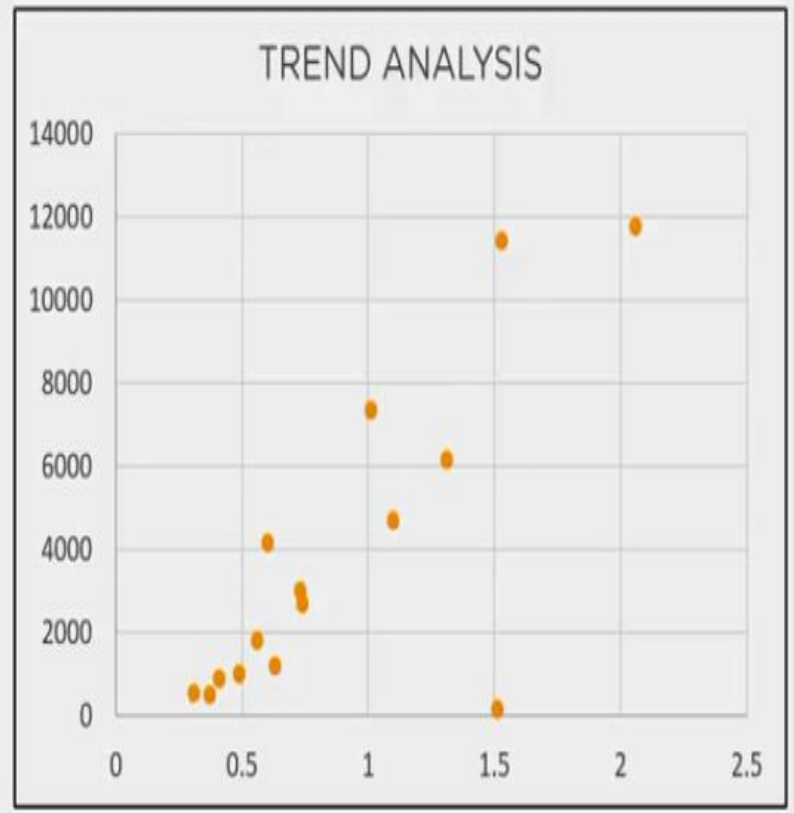
- Know the datatypes and answer questions with the data
- Understand how data is distributed
- Identify outliers
- Identify patterns, if any

Model Planning - Life cycle

Techniques:

- Histogram
- Trend Analysis

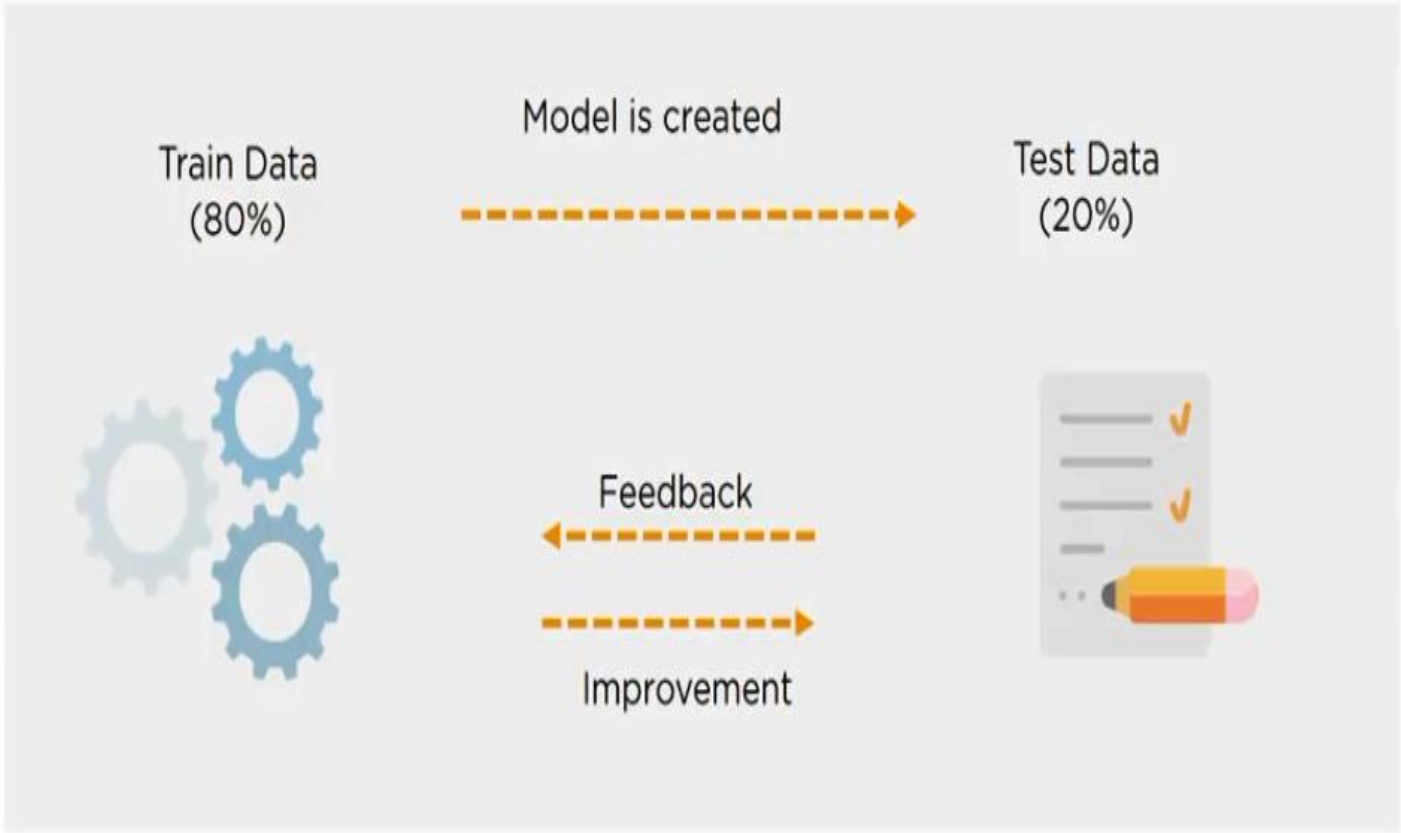




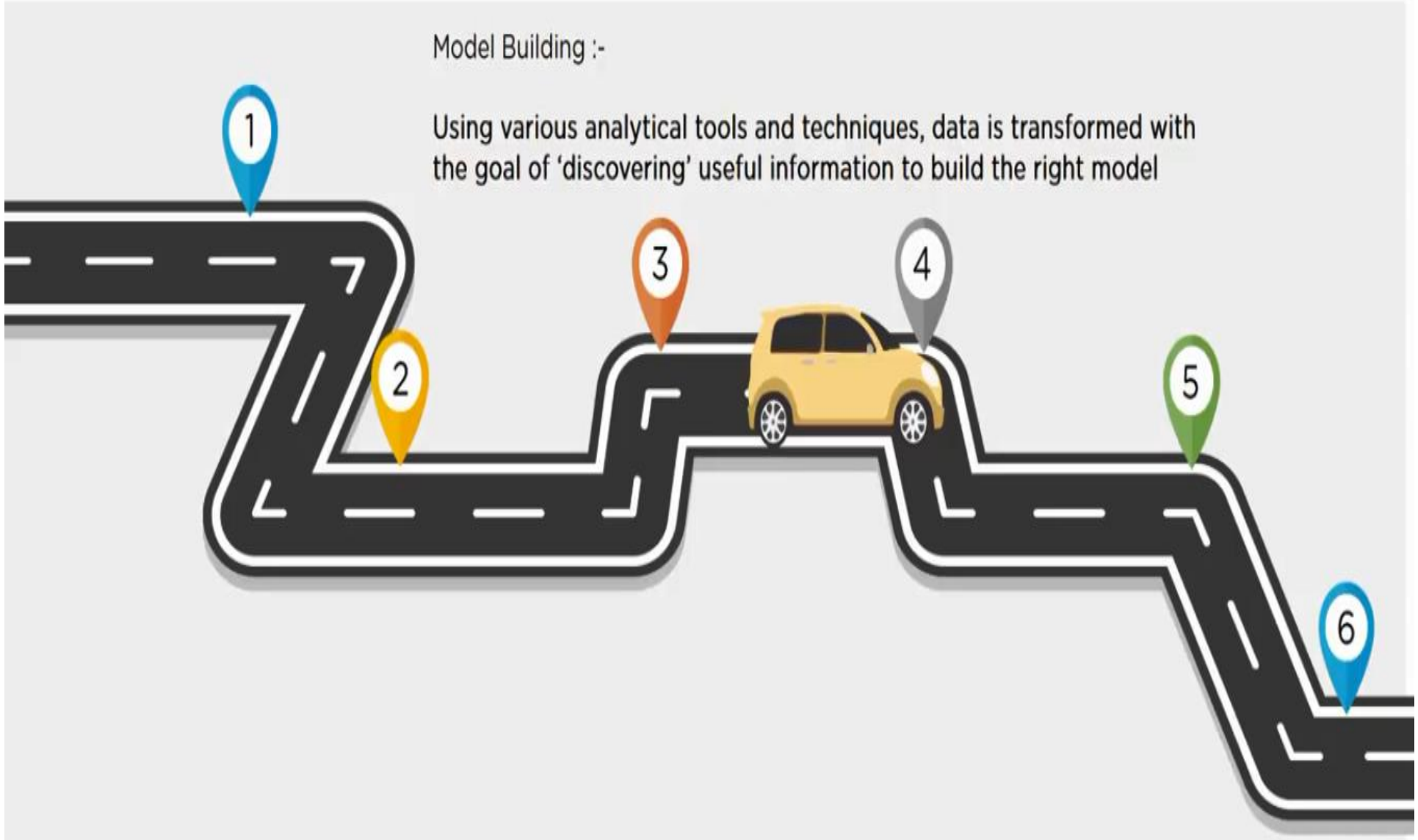
Model Planning - Use Case

Train Data vs Test Data

- Train Data is used to develop model
- Test Data is used to validate model



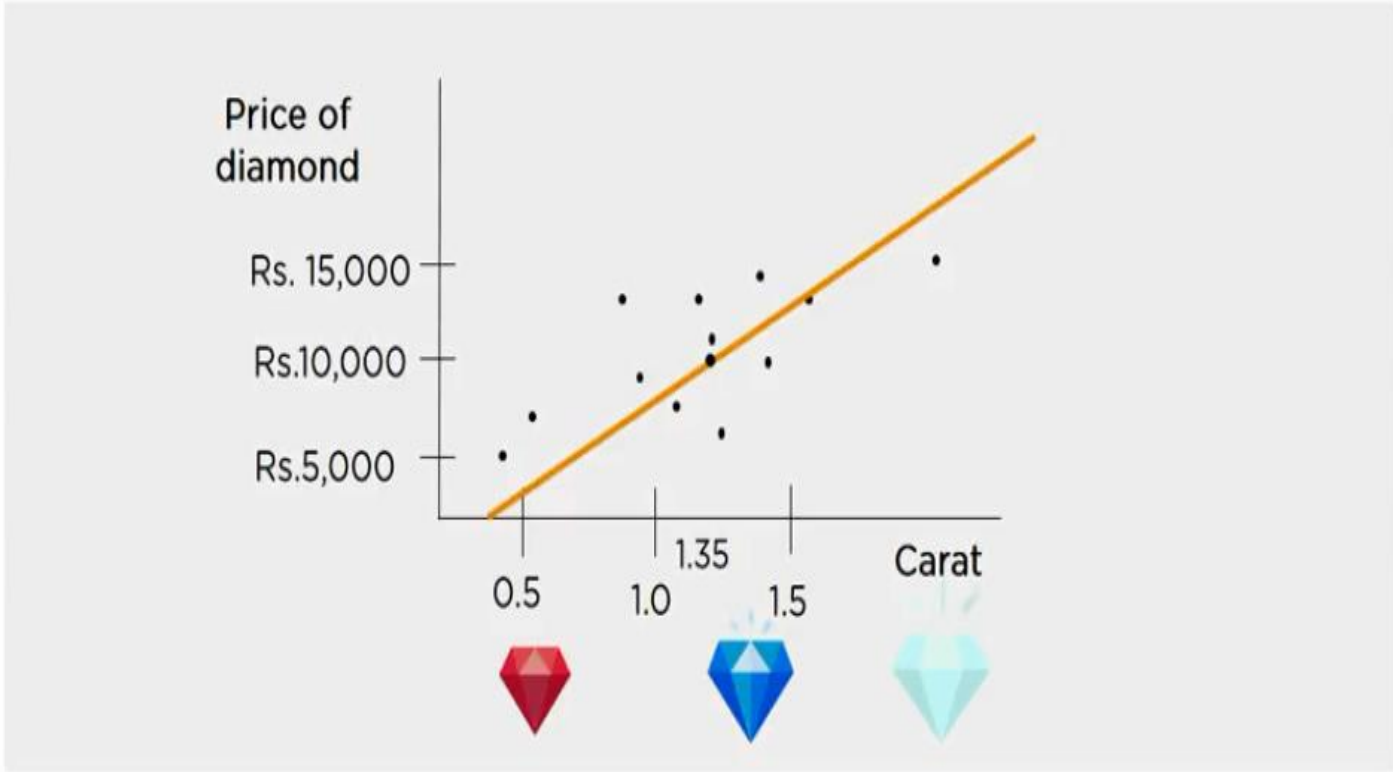
Model Building - Life cycle



Model Building - Example

Model Building:

On analyzing the data, we observe that the output is progressing linearly. Hence, we are using Linear Regression Algorithm for Model Building in this case



Model Building - Example

Linear regression describes the relation between 2 variables i.e. X and Y

After the regression line is drawn, we can predict Y value for a input X value using following formula: $Y = mX + c$

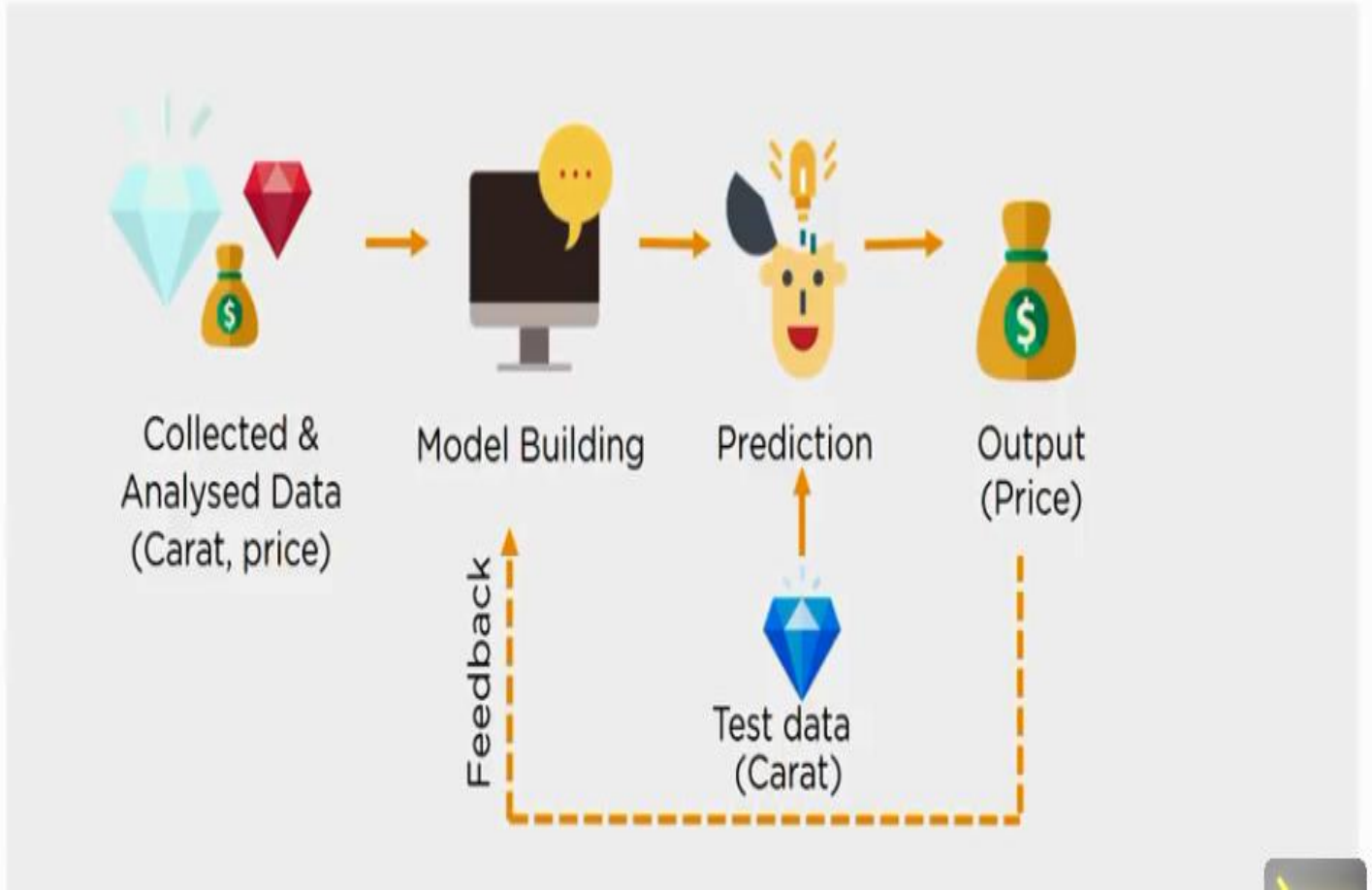
m = Slope of the line
c = Y intercept

X is Independent variable

Y is dependent variable

Model Building - Example

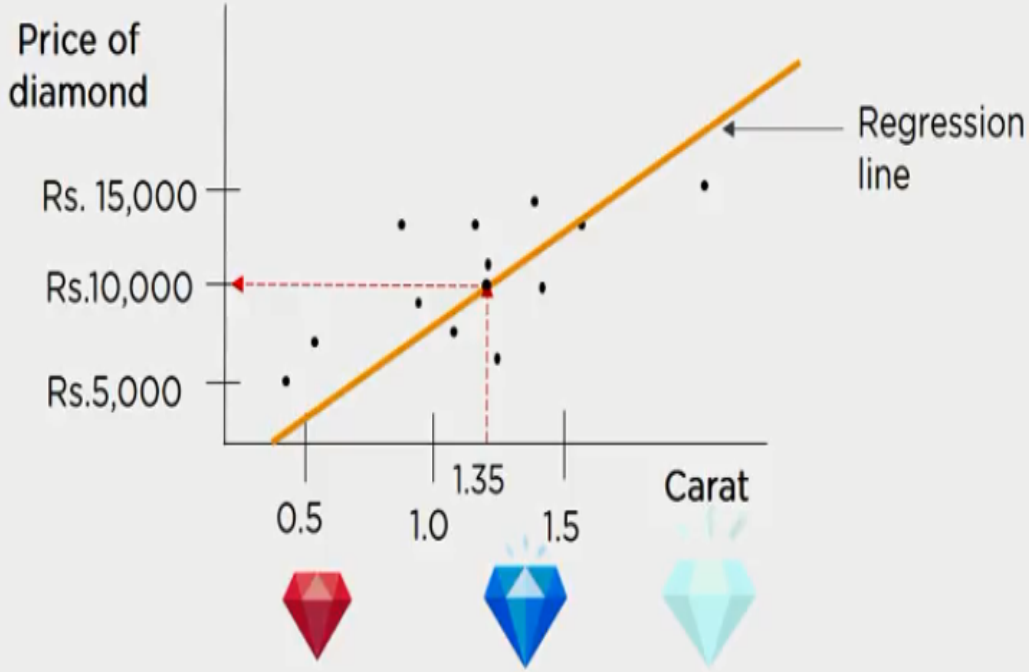
Using test data set, the built model is validated for the best accuracy



Model Building - Example

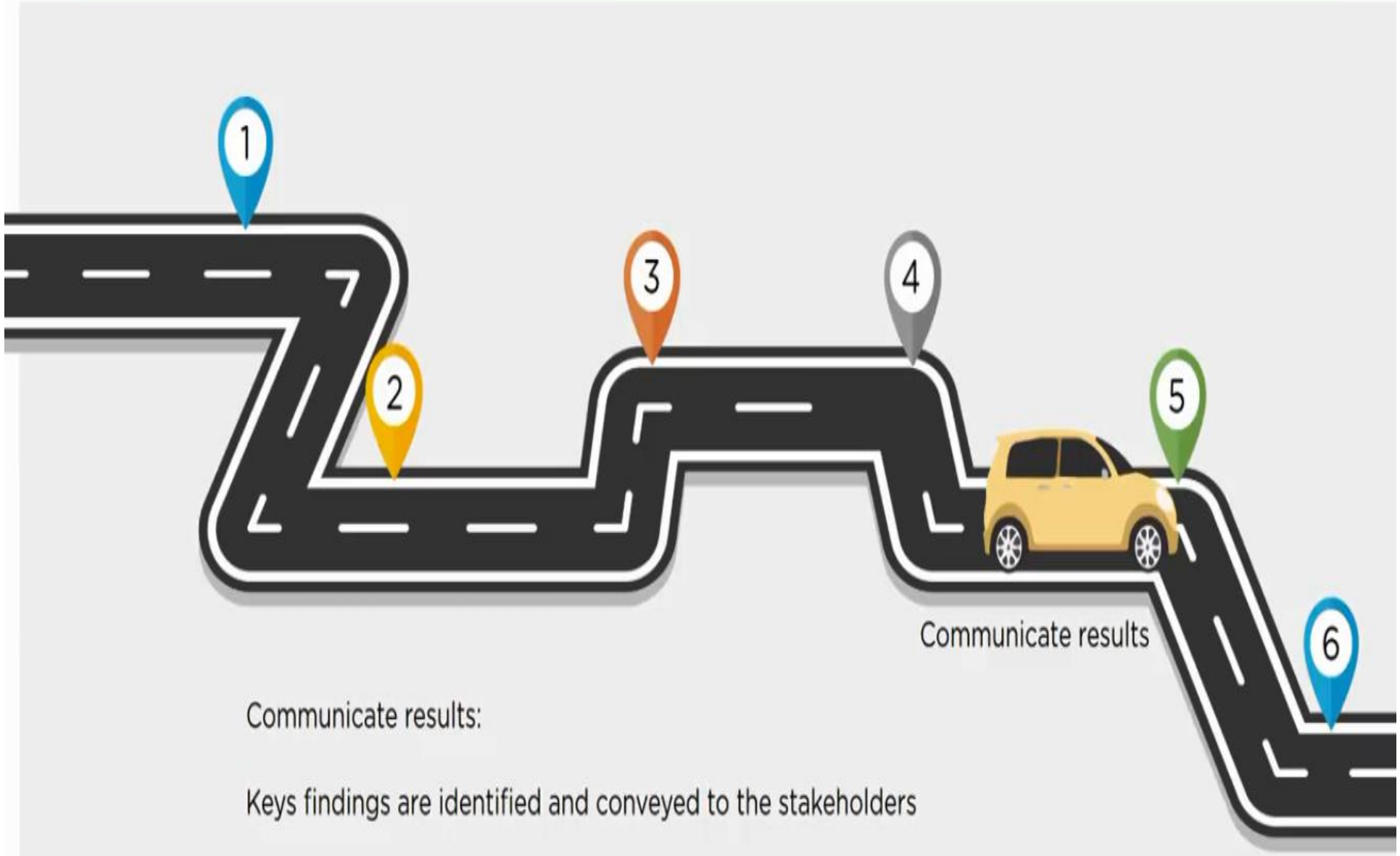
Prediction:

Thus, using Simple Linear Regression algorithm we have implemented a successful model and predicted the price of 1.35 carat diamond to be Rs. 10,000



©Simplilearn. All rights reserved

Communication - Life cycle



Communicate results:

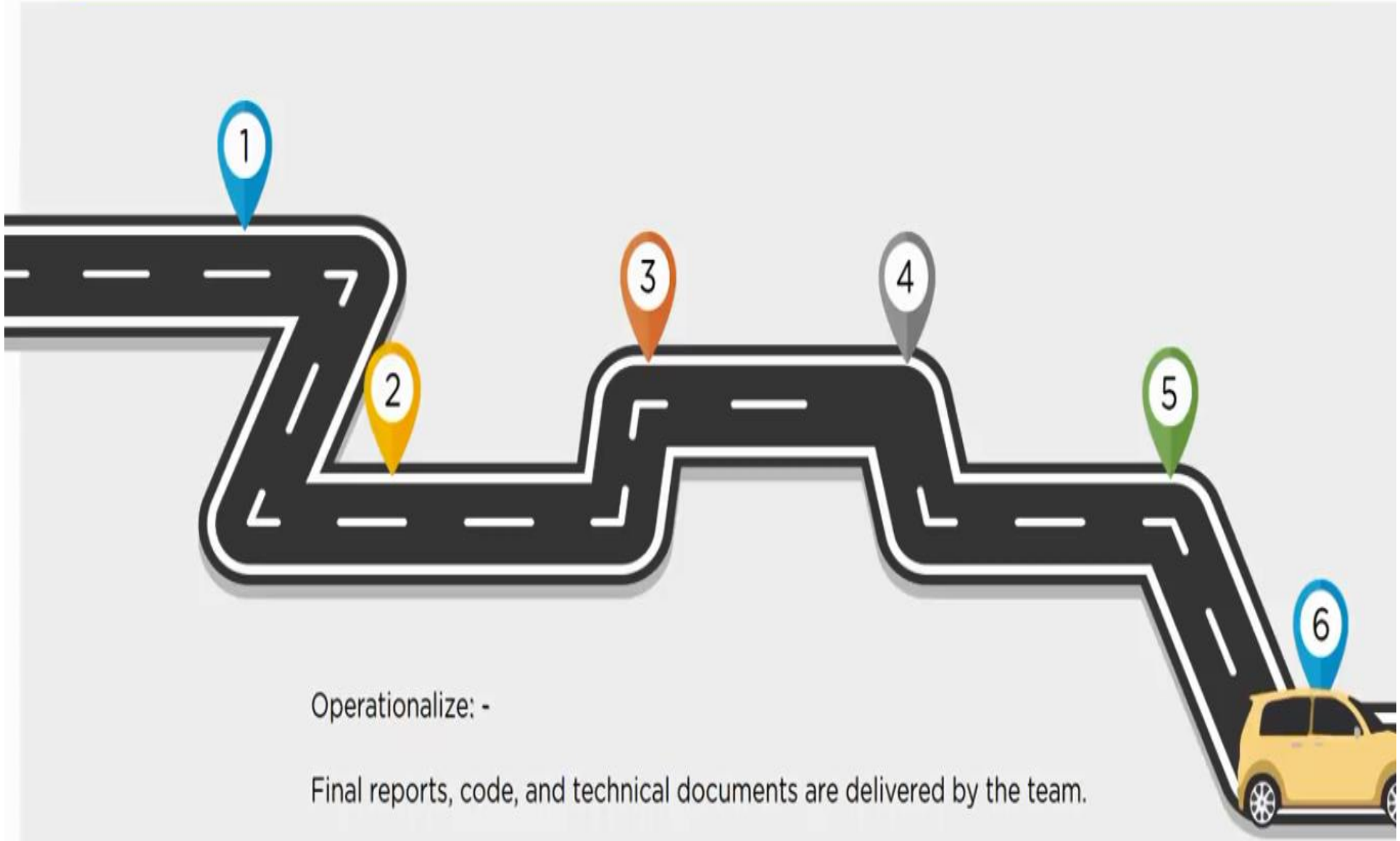
Keys findings are identified and conveyed to the stakeholders

Communication - Life cycle

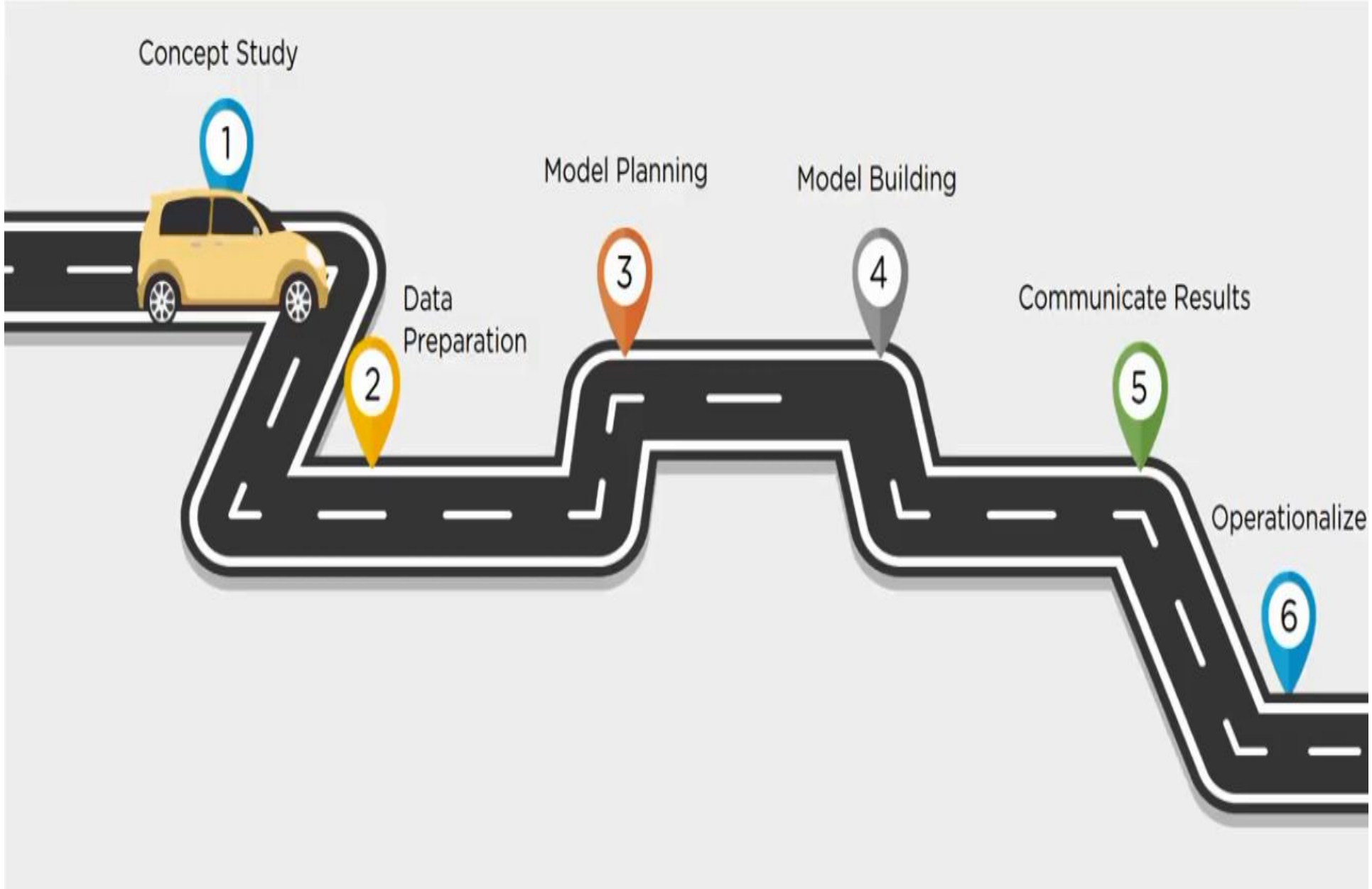
The Battle is not over yet!!

A good Data Scientist should be able to communicate his findings with the business team such that it easily goes into execution phase

Life cycle of Data Science project



Summary - Life cycle



Introduction to Data Analytics and Business Intelligence

Data Analytics

- **Data analytics** is the analysis of data, whether huge or small, in order to understand it and see how to use the knowledge hidden within it.
- **Business analytics** is the application of data analytics to business. An example is offering specific discounts to different classes of travelers based on the amount of business they offer or have the potential to offer
- **Data science** is an interdisciplinary field (including disciplines such as statistics, mathematics, and computer programming) that derives knowledge from data and applies it for predictive or other purposes.
- **Big data analytics** is the analysis of huge amounts of data (for example, trillions of records) or the analysis of difficult-to-crack problems.

Business User

- Someone who understands the domain area and usually benefits from the results.
- This person can consult and advise the project team on the context of the project, the value of the results, and how the outputs will be operationalized.
- Usually a business analyst, line manager, or deep subject matter expert in the project domain fulfills this role.

Project Sponsor

- Responsible for the genesis of the project. Provides the impetus and requirements for the project and defines the core business problem.
- Generally, provides the funding and gauges the degree of value from the final outputs of the working team.
- This person sets the priorities for the project and clarifies the desired outputs.

Project Manager

- Ensures that key milestones and objectives are met on time and at the expected quality.

Business Intelligence Analyst

- Provides business domain expertise based on a deep understanding of the data, key performance indicators (KPIs), key metrics, and business intelligence from a reporting perspective.
- Business Intelligence Analysts generally create dashboards and reports and have knowledge of the data feeds and sources.

Database Administrator (DBA):

- Provisions and configures the database environment to support the analytics needs of the working team.
- These responsibilities may include providing access to key databases or tables and ensuring the appropriate security levels are in place related to the data repositories.

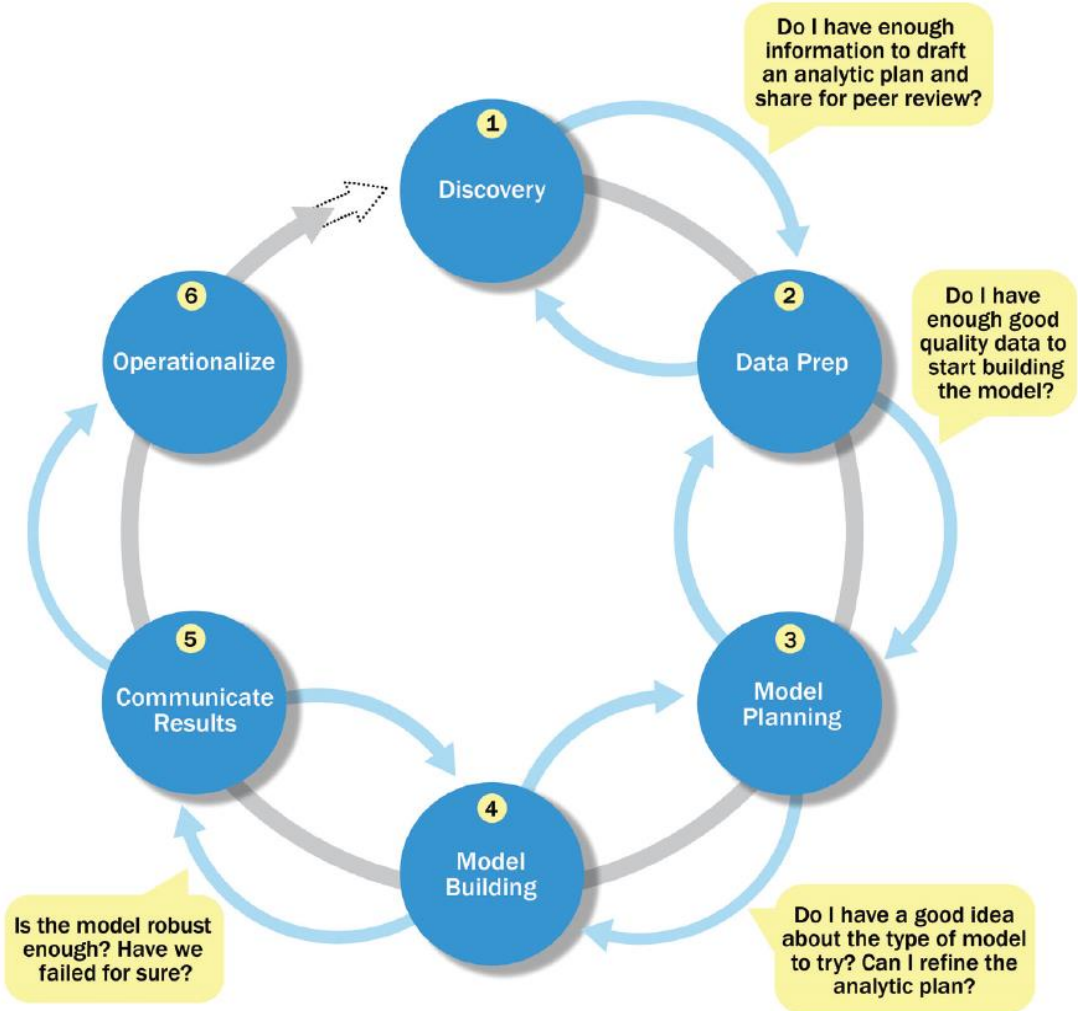
Data Engineer

- Leverages deep technical skills to assist with tuning SQL queries for data management and data extraction, and provides support for data ingestion into the analytic sandbox,
- Whereas the DBA sets up and configures the databases to be used, the data engineer executes the actual data extractions and performs substantial data manipulation to facilitate the analytics.
- The data engineer works closely with the data scientist to help shape data in the right ways for analyses.

Data Scientist

- Provides subject matter expertise for analytical techniques, data modeling, and applying valid analytical techniques to given business problems. Ensures overall analytics objectives are met.
- Designs and executes analytical methods and approaches with the data available to the project.

Data Analytics Lifecycle:



Phase 1—Discovery

- In Phase 1, the team learns the business domain, including relevant history such as whether the organization or business unit has attempted similar projects in the past from which they can learn.
- The team assesses the resources available to support the project in terms of people, technology, time, and data.
- Important activities in this phase include framing the business problem as an analytics challenge that can be addressed in subsequent phases and formulating initial hypotheses (IHs) to test and begin learning the data.

Phase 2—Data preparation

- Phase 2 requires the presence of an analytic sandbox, in which the team can work with data and perform analytics for the duration of the project.
- The team needs to execute extract, load, and transform (ELT) or extract, transform and load (ETL) to get data into the sandbox.
- The ELT and ETL are sometimes abbreviated as ETLT.
- Data should be transformed in the ETLT process so the team can work with it and analyze it.
- In this phase, the team also needs to familiarize itself with the data thoroughly and take steps to condition the data

Phase 3—Model planning

- Phase 3 is model planning, where the team determines the methods, techniques, and workflow it intends to follow for the subsequent model building phase.
- The team explores the data to learn about the relationships between variables and subsequently selects key variables and the most suitable models.

Phase 4—Model building

- In Phase 4, the team develops datasets for testing, training, and production purposes.
- In addition, in this phase the team builds and executes models based on the work done in the model planning phase.
- The team also considers whether its existing tools will suffice for running the models, or if it will need a more robust environment for executing models and workflows (for example, fast hardware and parallel processing, if applicable).

Phase 5—Communicate results:

- In Phase 5, the team, in collaboration with major stakeholders, determines if the results of the project are a success or a failure based on the criteria developed in Phase 1.
- The team should identify key findings, quantify the business value, and develop a narrative to summarize and convey findings to stakeholders.

Phase 6—Operationalize:

- In Phase 6, the team delivers final reports, briefings, code, and technical documents.
- In addition, the team may run a pilot project to implement the models in a production environment

Phase 1: Discovery

- **Learning the Business Domain**
- **Resources**
- **Framing the Problem**
- **Identifying Key Stakeholders**
- **Interviewing the Analytics Sponsor**
- **Developing Initial Hypotheses**
- **Identifying Potential Data Sources**

Phase 2: Data Preparation

- Preparing the Analytic Sandbox
- Performing ETLT
- Learning About the Data
- Data Conditioning
- Survey and Visualize
- Common Tools for the Data Preparation Phase

Phase 3: Model Planning

- Data Exploration and Variable Selection
- Model Selection
- Common Tools for the Model Planning Phase
- R [14] has a complete set of modeling capabilities and provides a good environment for building interpretive models with high-quality code
- SQL Analysis services [15] can perform in-database analytics of common data mining functions, involved aggregations, and basic predictive models.
- SAS/ACCESS [16] provides integration between SAS and the analytics sandbox via multiple data onnectors such as OBDC, JDBC, and OLE DB.

Phase 4: Model Building

- The data science team needs to develop datasets for training, testing, and production purposes.
- These datasets enable the data scientist to develop the analytical model and train it (“training data”), while holding aside some of the data (“hold-out data” or “test data”) for testing the model.
- During this process, it is critical to ensure that the training and test datasets are sufficiently robust for the model and analytical techniques.
- A simple way to think of these Datasets is to view the training dataset for conducting the initial experiments and the test sets for validating an approach once the initial experiments and models have been run.
- During this phase, users run models from analytical software packages, such as R or SAS, on file extracts and small datasets for testing purposes
- Common Tools for the Model Building Phase
- SAS Enterprise Miner, SPSS Modeler, STATISTICA [20] and Mathematica, Matlab
- R, Python,SQL

Phase 5: Communicate Results

- After executing the model, the team needs to compare the outcomes of the modeling to the criteria established for success and failure.
- In Phase 5,The team considers how best to articulate the findings and outcomes to the various team members and stakeholders, taking into account caveats, assumptions, and any limitations of the results.
- Because the presentation is often circulated within an organization, it is critical to articulate the results properly and position the findings in a way that is appropriate for the audience

Phase 6: Operationalize

- In the final phase, the team communicates the benefits of the project more broadly and sets up a pilot project to deploy the work in a controlled way before broadening the work to a full enterprise or ecosystemof users.
- In Phase 4, the team scored the model in the analytics sandbox.
- Phase 6 represents the first time that most analytics teams approach deploying the new analytical methods or models in a production environment.
- Rather than deploying these models immediately on a wide-scalebasis, the risk can be managed more effectively and the team can learn by undertaking a small scope, pilot deployment before a wide-scale rollout.

Phase 6: Operationalize

- This approach enables the team to learn about the performance and related constraints of the model in a production environment on a small scale and make adjustments before a full deployment.
- During the pilot project, the team may need to consider executing the algorithm in the database rather than with in-memory tools such as R because the run time is significantly faster and more efficient than running in-memory, especially on larger datasets

Introduction business analytics

Imagine the following situations: Individual perspective

- You visit a hotel in Switzerland and are welcomed with your favorite drink and dish; how delighted you are!
- A **probable riot** at your **planned travel destination**. Based on **this warning, you cancel the visit**; you later learn from **news reports** that a **riot** does happen at that destination!
- Your preferred airline reserves tickets for you well in advance of your vacation travels and at a lower rate compared to the market rate.
- You are planning to travel and are forewarned of a possible cyclone in that place. Based on that warning, you postpone your visit. Later, you find that the cyclone created havoc, and you avoided a terrible situation.

Scenarios from a business perspective

- You are in the taxi business and are able to repeatedly attract the same customers based on their earlier travel history and preferences of taxi type and driver.
- You are in the fast-food business and offer discounted rates to attract customers on slow days. These discounts enable you to ensure full occupancy on those days also.
- You are in the human resources (HR) department of an organization and are bogged down by high attrition. But now you are able to understand the types of people you should focus on recruiting, based on the characteristics of those who perform well and who are more loyal and committed to the organization.

Business analytics

- All these scenarios are possible by analyzing data that the businesses and others collect from various sources.
- There are many such possible scenarios.
- The application of data analytics to the field of business is called business analytics.

Drivers for Business Analytics

- Increasing numbers of relevant computer packages and applications. One example is the R programming environment with its various data sets, documentation on its packages, andready-made algorithms.
- Feasibility to consolidate related and relevant data from various sources and of various types (data from flat files, data from relational databases, data from log files, data from Twitter messages, and more). An example is the consolidation of information from data files in a Microsoft SQL Server databasewith data from a Twitter message stream.
- Growth of seemingly infinite storage and computing capabilities by clustering multiple computers and extending these capabilitiesvia the cloud. An example is the use of Apache Hadoop clusters to distribute and analyze huge amounts of data.
- Availability of many easy-to-use programming tools, platforms, and frameworks (such as R and Hadoop).
- Emergence of many algorithms and tools to effectively use statistical and mathematical concepts for business analysis. One example is the k-means algorithm used for partition clustering analysis.

Growth of Computer Packages and Applications

- Many computer packages and applications that collect a lot of data about us.
- This data is used by businesses to make them more competitive, attract more business, and retain and grow their customer base.
- With thousands of apps on platforms such as Android, iOS, and Windows, the capture of data encompasses nearly all the activities carried out by individuals across the globe (who are the consumers for most of the products and services).
- This has been enabled further by the reach of hardware devices such as computers, laptops, mobile phones, and smartphones even to remote places.

Feasibility to Consolidate Data from Various Sources

- Technology has grown by leaps and bounds over the last few years. It is now easy for us to convert data from one format to another and to consolidate it into a required format.
- The growth of technology coupled with almost unlimited storage capability has enabled us to consolidate related or relevant data from various sources—right from flat files, to database data, to data in various formats.
- This ability to consolidate data from various sources has provided a great deal of momentum to effective business analysis.

Growth of Infinite Storage and Computing Capability

- The memory and storage capacity of individual computers has increased drastically, whereas external storage devices have provided a significant increase in storage capacity.
- This has been augmented by cloud-based storage services that can provide a virtually unlimited amount of storage.
- The growth of cloud platforms has also contributed to virtually unlimited computing capability

Easy-to-Use Programming Tools and Platforms

- Open source tools or platforms such as R and Hadoop are available. These powerful tools are easy to use and well documented.
- They do not require high-end programming experience but usually require an understanding of basic programming concepts.
- Hadoop is especially helpful in effective and efficient analysis of big data

Survival and Growth in the Highly Competitive World

- Businesses have become highly competitive. With the Internet easily available to every business, every consumer has become a target for every business.
- Each business is targeting the same customer and that customer's spending capability. Each business also can easily reach other dependent businesses or consumers equally well.
- Using the Internet and the Web, businesses are fiercely competing with each other; often they offer heavy discounts and cut prices drastically.
- To survive, businesses have to find the best ways to target other businesses that require their products and services as well as the end consumers who require their products and services.
- Data or business analytics has enabled this effectively

Business Complexity Growing out of Globalization

- Economic globalization that cuts across the boundaries of the countries where businesses produce goods or provide services has drastically increased the complexities of business.
- Businesses now have the challenge of catering to cultures that may have been previously unknown to them.
- With the large amount of data now possible to acquire (or already at their disposal), businesses can easily gauge differences between local and regional cultures, demands, and practices including spending trends and preferences.

Applications of Business Analytics

- Marketing and Sales
- Human Resources
- Product Design
- Service Design
- Customer Service and Support Areas

Marketing and Sales

- Marketing and sales teams are the ones that have heavily used business analytics to identify appropriate approaches to marketing in order to reach a maximum number of potential customers at an optimized or reduced effort.
- These teams use business analytics to identify which marketing channel would be most effective (for example, e-mails, web sites, or direct telephone contacts).
- They also use business analytics to determine which offers make sense to which types of customers (in terms of geographical regions, for instance) and to specifically tune their offers.

Human Resources

- Retention is the biggest problem faced by an HR department in any industry, especially in the support industry.
- An HR department can identify which employees have high potential for retention by processing employee data.
- Similarly, an HR department can also analyze which competence (qualification, knowledge, skill, or training) has the most influence on the organization’s or team’s capability to deliver quality within committed timelines.

Product Design

- Product design is not easy and often involves complicated processes. Risks factored in during product design, subsequent issues faced during manufacturing, and any resultant issues faced by customers or field staff can be a rich source of data that can help you understand potential issues with a future design.
- This analysis may reveal issues with materials, issues with the processes employed, issues with the design process itself, issues with the manufacturing, or issues with the handling of the equipment installation or later servicing.
- The results of such an analysis can substantially improve the quality of future designs by any company.
- Another interesting aspect is that data can help indicate which design aspects (color, sleekness, finish, weight, size, or material) customers like and which ones customers do not like

Service Design

- Services are also carefully designed and priced by organizations.
- Identifying components of the service (and what are not) also depends on product design and cost factors compared to pricing.
- The length of warranty, coverage during warranty, and pricing for various services can also be determined based on data from earlier experiences and from target market characteristics.
- Some customer regions may more easily accept “use and throw” products, whereas other regions may prefer “repair and use” kinds of products.
- Hence, the types of services need to be designed according to the preferences of regions. Again, different service levels (responsiveness) may have different price tags and may be targeted toward a specific segment of customers (for example, big corporations, small businesses, or individuals).

Customer Service and Support Areas

- After-sales service and customer service is an important aspect that no business can ignore.
- A lack of effective customer service can lead to negative publicity, impacting future sales of new versions of the product or of new products from the same company.
- Customer service is an important area in which data analysis is applied significantly.
- Customer comments on the Web or on social media (for example, Twitter) provide a significant source of understanding about the customer pulse as well as the reasons behind the issues faced by customers.
- A service strategy can be accordingly drawn up, or necessary changes to the support structure may be carried out, based on the analysis of the data available to the industry.

Skills Required for a Business Analyst

- The business and problems of the business
- Data analysis techniques and algorithms that can be applied to the business data
- Computer programming
- Data structures and data-storage or data-warehousing Techniques, including how to query the data effectively
- Statistical and mathematical concepts used in data analytics (for example, regression, naïve Bayes analysis, matrix algebra, and cost-optimization algorithms such as gradient descent or ascent algorithms)

Understanding the Business and Business Problems

- clear understanding of the business and business problems is one of the most important requirements for a business analyst.
- If the person analyzing the data does not understand the underlying business, the specific characteristics of that business, and the specific problems faced by that business, that person can be led to the wrong conclusions or led in the wrong direction.
- Having only programming skills along with statistical or mathematical knowledge can sometimes lead to proposing impractical (or even dangerous) suggestions for the business.
- These suggestions also waste the time of core business personnel.

Understanding Data Analysis Techniques and Algorithms

- Data analysis techniques and algorithms must be applied to suitable situations or analyses.
- For example, linear regression or multiple linear regression (supervised method) may be suitable if you know (based on business characteristics) that there exists a strong relationship between a response variable and various predictors.
- You know, for example, that geographical location, proximity to the city center, or the size of a plot (among others) has a bearing on the price of the land to be purchased.
- Clustering (unsupervised method) can allow you to cluster data into various segments.
- Using and applying business analytics effectively can be difficult without understanding thesetechniques and algorithms.

Having Good Computer Programming Knowledge

- Good computer knowledge is required for a capable business analyst, so that the analyst doesn't have to depend on other programmers who don't understand the business and may not understand the statistics or mathematics behind the techniques or algorithms.
- Having good computer programming knowledge is always a bonus capability for business analysts, even though it is not mandatory because analysts can always employ a computer programmer.
- Computer programming may be necessary to consolidate data from different sources as well as to program and use the algorithms.
- Platforms such as R and Hadoop have reduced the pain of learning programming, even though at times we may have to use other complementary programming languages (for example, Python) for effectiveness and efficiency.

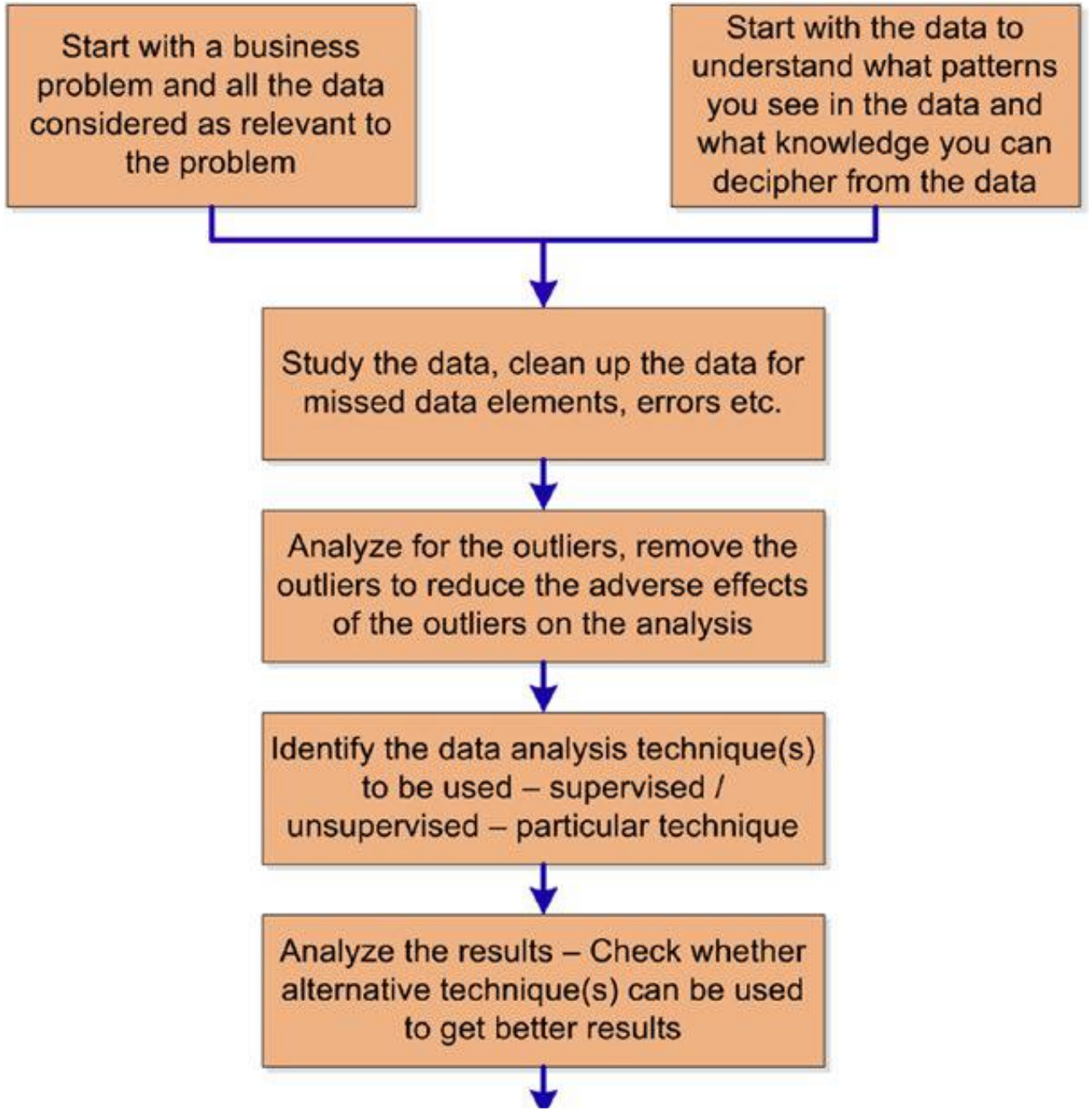
Understanding Data Structures and Data Storage/ Warehousing Techniques

- Knowledge of data structures and of data storage/warehousing techniques eases the life of a business analyst by eliminating dependence on database administrators and database programmers.
- This enables business analysts to consolidate data from varied sources (including databases and flat files), put them into a proper structure, and store them appropriately in a data repository.
- The capability to query such a data repository is another additional competence of value to any business analyst.
- This know-how is not a must, however, because a business analyst can always hire someone else to provide this skill.

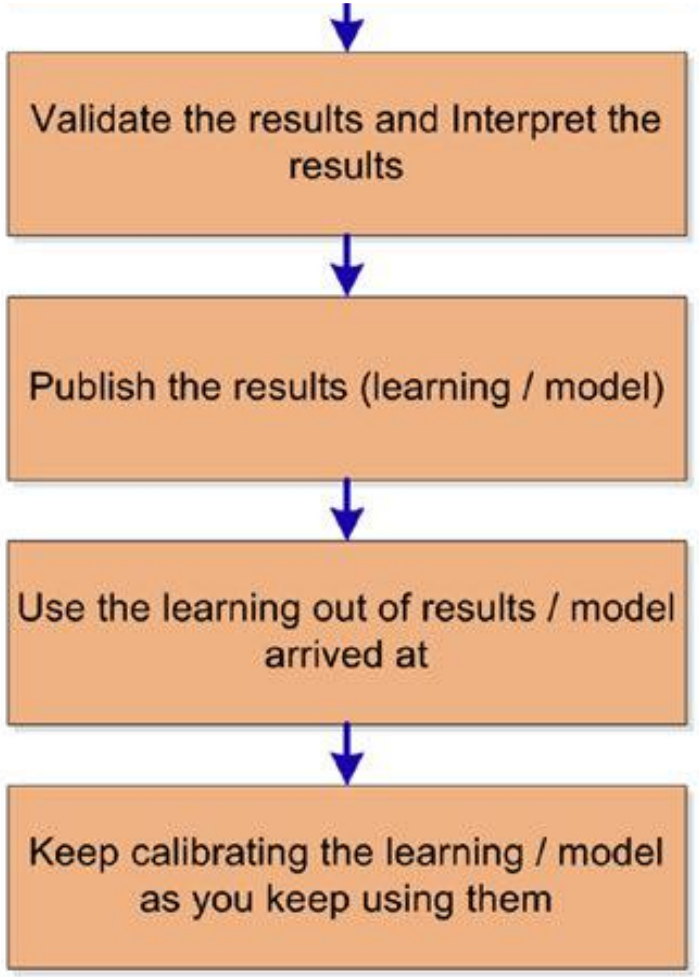
Knowing Relevant Statistical and Mathematical Concepts

- Data analytics uses many statistical and mathematical concepts on which various algorithms, measures, and computations are based.
- A business analyst should have good knowledge of statistical and mathematical concepts in order to properly use these concepts to depict, analyze, and present the data and the results of analysis.
- Otherwise, the business analyst can lead others in the wrong direction by misinterpreting the results because the application of the technique or interpretation of the result itself was wrong.

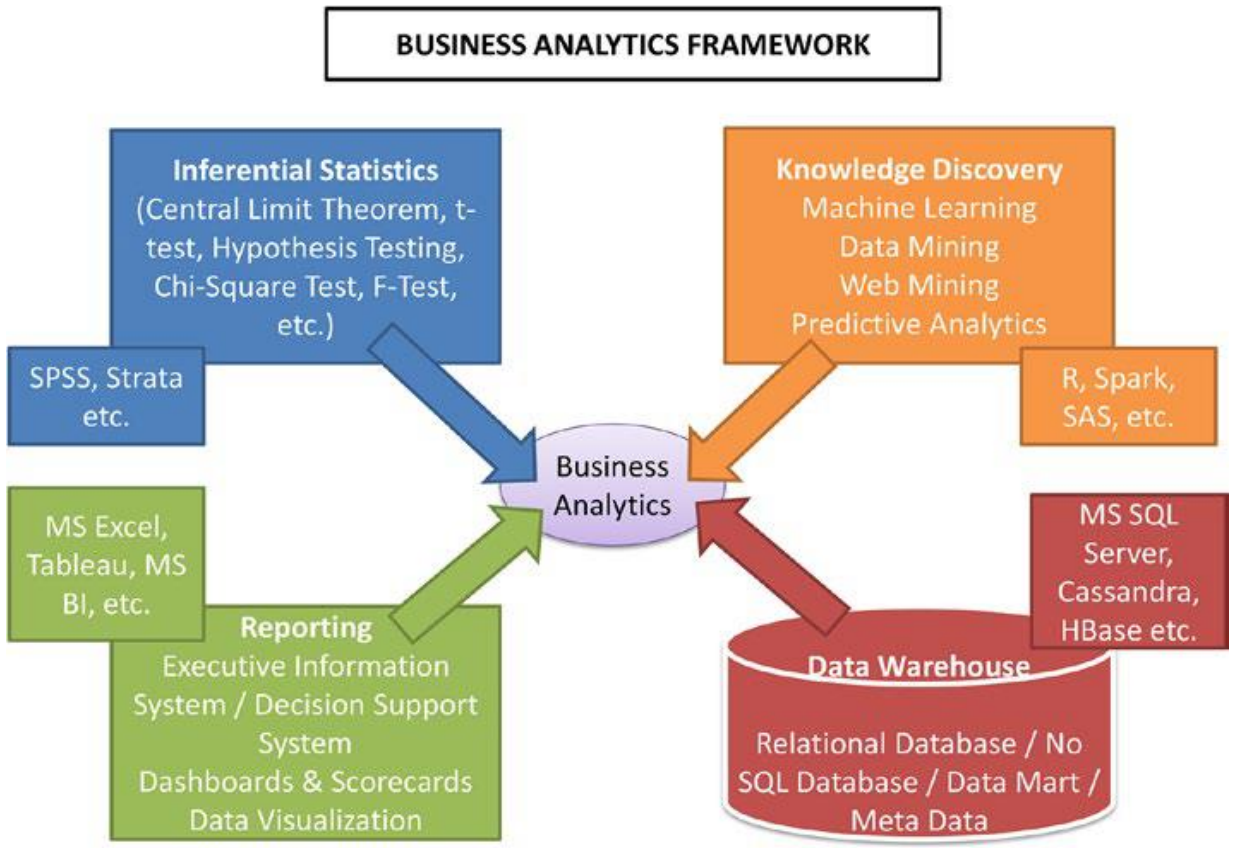
Life Cycle of a Business Analytics Project



Life Cycle of a Business Analytics Project

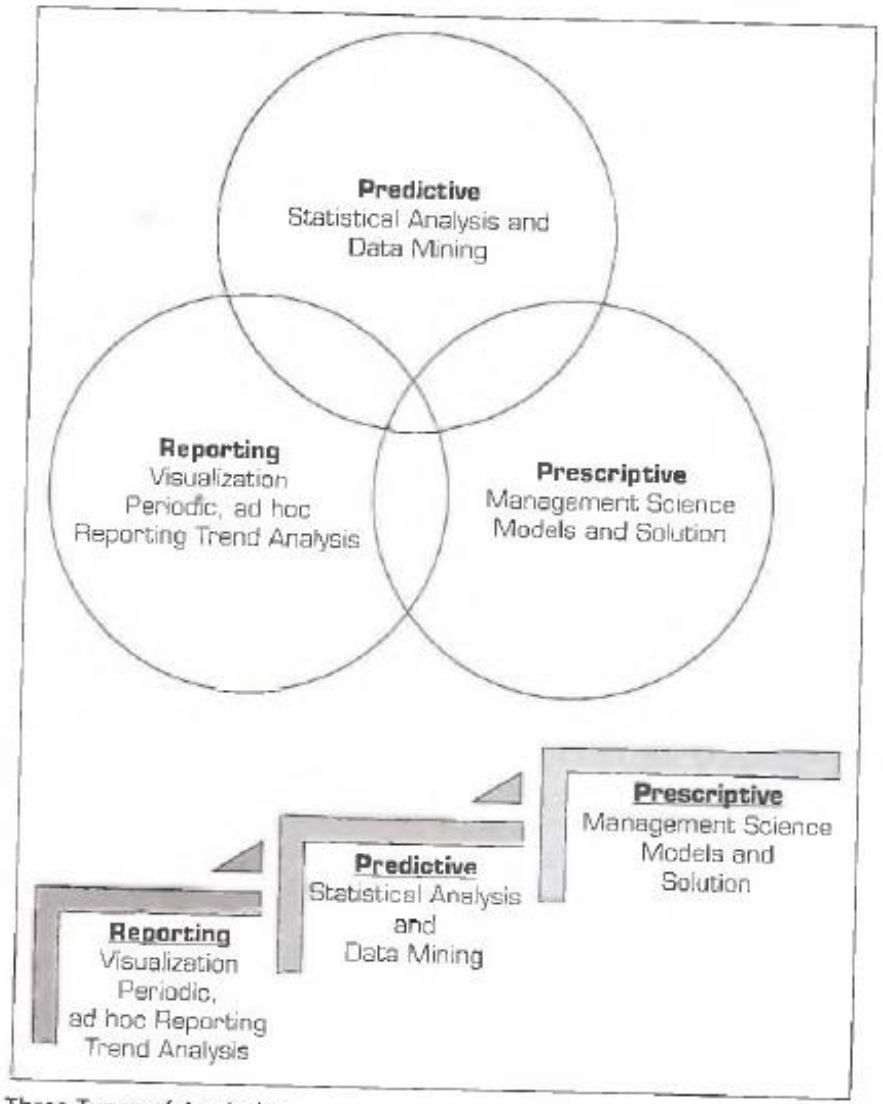


Features of Business analytics



Types of Business Analytics?

- Descriptive Analytics
- Diagnostic Analytics
- Predictive Analytics
- Prescriptive Analytics



GURE 1.5 Three Types of Analytics.

Descriptive Analytics

- Concise analyses or reports may simplify and transform raw data into a type that humans may readily comprehend.
- They will describe an incident that has existed in the past. This type of business analytics is useful in deriving some trend, if any, from past events or in drawing conclusions from them such that better potential plans can be presented.
- Descriptive analytics juggles raw data from many sets of data to provide useful perspectives into the past.
- Such results, though, clearly indicate something is incorrect or right, without specifying why.

Diagnostic Analytics

- Diagnostic business analytics is a natural counterpart to descriptive analytics.
- Diagnostic analytical techniques help an investigator dive further into a question at hand so they can get to the root of an issue.
- This type of business analytics offers in-depth observations into a complex issue.
- At the same time, an organization should have comprehensive details at its fingertips, because with each problem and time-consuming, data collection can turn out individual.

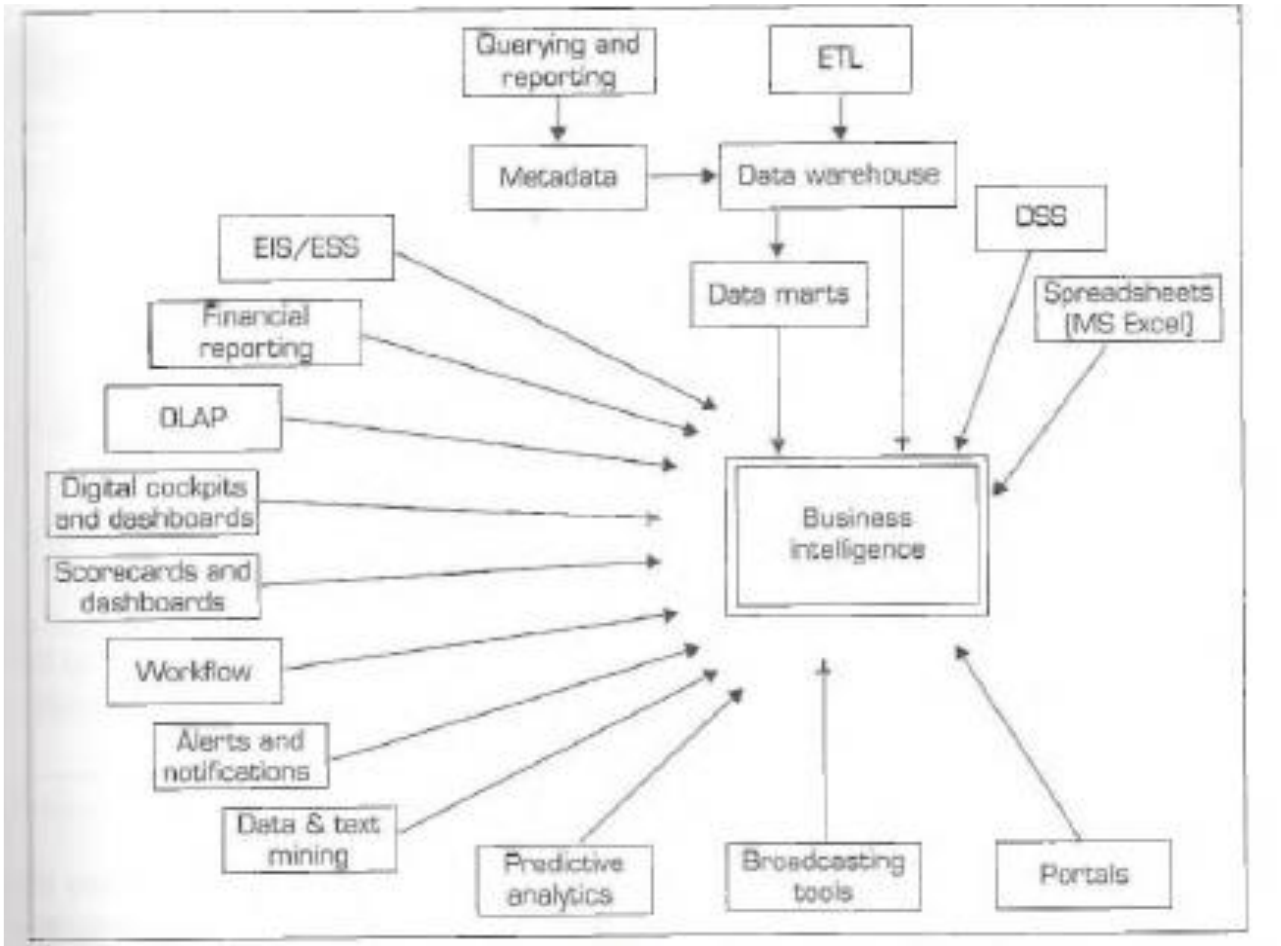
Predictive Analytics

- Every productive company will have foresight. Predictive analytics lets businesses identify patterns dependent on real events.
- Whether it’s forecasting the probability of an occurrence happening on in the future or determining the precise moment it’s likely to happen may both be calculated through predictive computational models.
- Typically, in this type of business analysis, several separate yet co-dependent variables are tested to forecast a pattern.
- One must note that forecasting is only an estimation, the precision of which depends heavily on the data quality and the situation’s stability, so it needs diligent evaluation and constant optimization.

Prescriptive Analytics

- This type of business analytics illustrates the mechanism in a scenario, step-by-step.
- For eg, when your e-hailing driver gets the simpler route from online maps a prescriptive approach is what comes into action.
- Considering the width of each open path from your pick-up path to the destination and the traffic restrictions on each lane, the best route was selected.
- Prescriptive analytics utilizes specialized techniques and technology, such as artificial intelligence, company law, and algorithms, allowing the design and administration sophisticated.

Evolution of Business Intelligence



Evolution of Business Intelligence

- The term B I was coined by the Gartner Group in the mid-1990s.
- The concept is much older; it has its roots in the MIS reporting systems o f the 1970s.
- During that period, reporting systems were static, two dimensional, and had no analytical capabilities.
- In the early 1980s, the concept o f executive information systems (EIS) emerged. This concept expanded the computerized support to top-level managers and executives.
- Some o f the capabilities introduced w ere dynamic multidimensional (ad hoc or on-demand) reporting, forecasting and prediction, trend analysis, drill-down to details, status access, and critical success factors.
- These features appeared in dozens o f commercial products until the mid-1990s.
- Then the same capabilities and some new ones appeared under the name BI
- Today a good BI-based enterprise information system contains all the information execute s need. So, the original concept of EIS was transformed into BI.
- By 2005, BI systems seined to include artificial intelligence capabilities as well as powerful analytical capabilities.

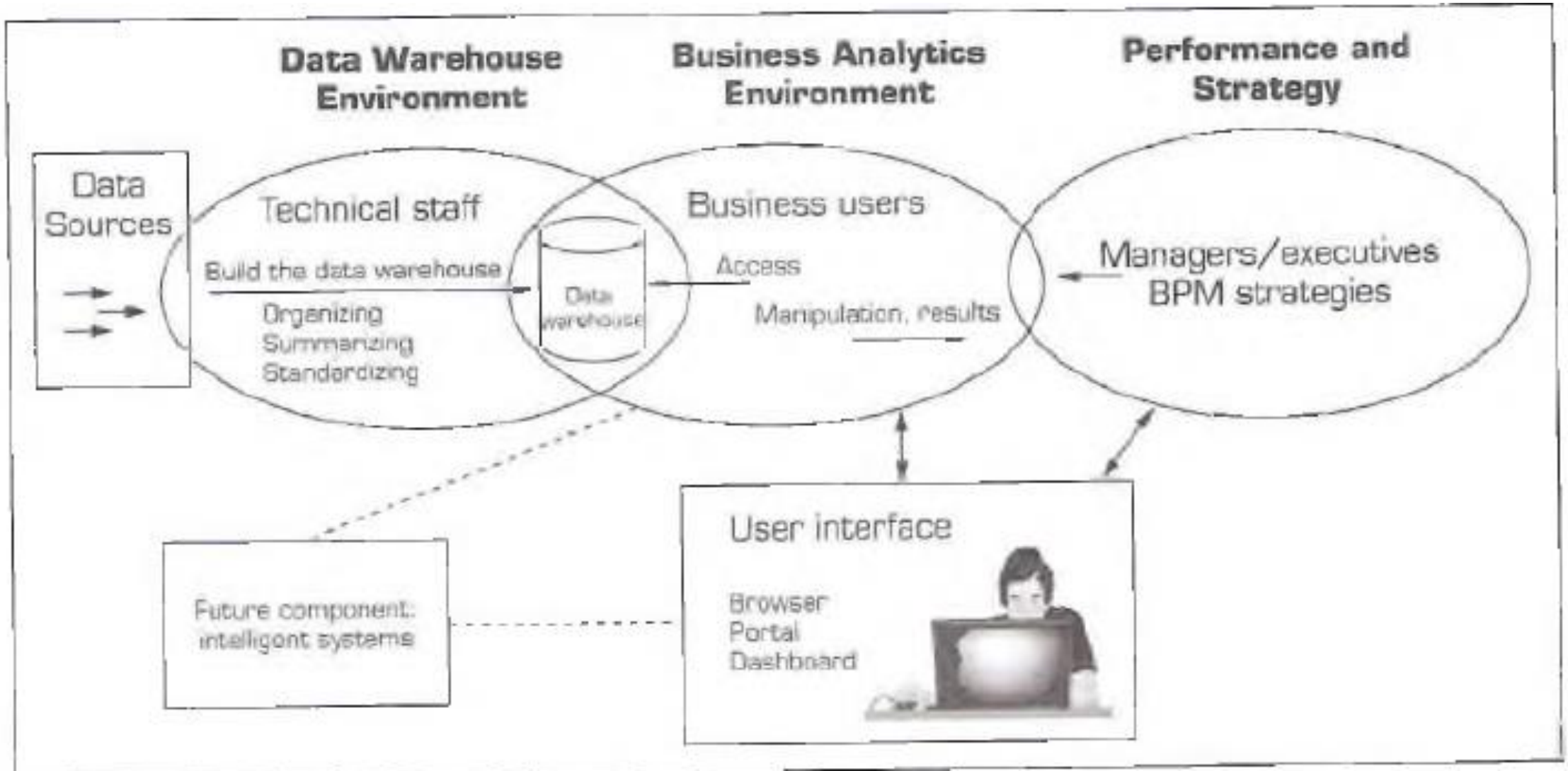
Definition of Business Intelligence

- Business intelligence (BI) is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies.
- B I's major objective is to enable interactive access (sometimes in real time) to data, to enable manipulation o f data, and to give business managers and analysts the ability to conduct appropriate analyses.
- By analyzing historical and current data, situations, and performances, decision makers get valuable insights that enable them to make more informed and better decisions.
- The process o f BI is based on the *transform action* o f data to information, then to decisions, and finally to actions.

The Architecture of BI

- A BI system has four major components:
- **A *data warehouse***, with its source data;
- ***Business analytics,*** a collection o f tools for manipulating, mining, and analyzing the data in the Datawarehouse
- ***Business performance management*** *(BPM)* for monitoring and analyzing performance;
- ***Yser interface*** (e.g., a dashboard).

The Architecture of BI



Business Value of BI Analytical Applications

Analytic Application	Business Question	Business Value
Customer segmentation	What market segments do my customers fall into, and what are their characteristics?	Personalize customer relationships for higher satisfaction and retention.
Propensity to buy	Which customers are most likely to respond to my promotion?	Target customers based on their need to increase their loyalty to your product line. Also, increase campaign profitability by focusing on the most likely to buy.
Customer profitability	What is the lifetime profitability of my customer?	Make individual business interaction decisions based on the overall profitability of customers.
Fraud detection	How can I tell which transactions are likely to be fraudulent?	Quickly determine fraud and take immediate action to minimize cost.
Customer attrition	Which customer is at risk of leaving?	Prevent loss of high-value customers and let go of lower-value customers.
Channel optimization	What is the best channel to reach my customer in each segment?	Interact with customers based on their preference and your need to manage cost.

Source: A. Ziama and J. Kasher, *Data Mining Primer for the Data Warehousing Professional*. Teradata, Dayton, OH, 2004.

Thank You