Blog on Data Science

By DEEPAK KAURA

This blog is about data science and it will revolve around it and also about data scientist

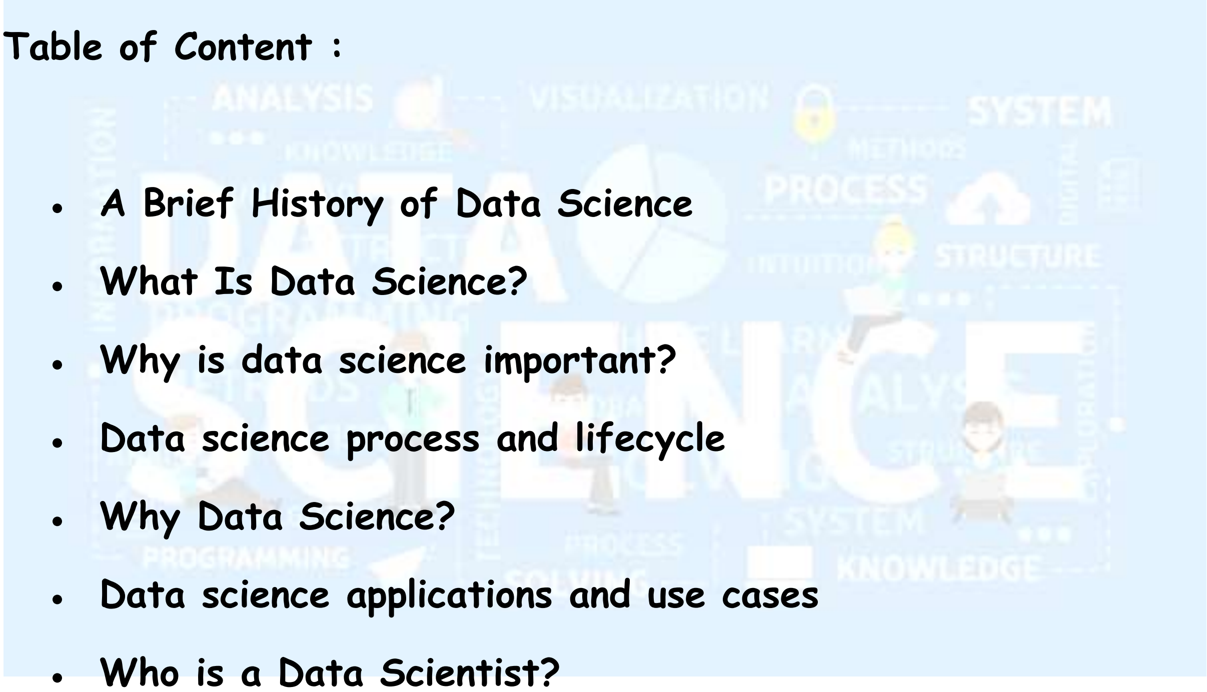


Table of Content :

- A Brief History of Data Science
- What Is Data Science?
- Why is data science important?
- Data science process and lifecycle
- Why Data Science?
- Data science applications and use cases
- Who is a Data Scientist?
- What does a Data Scientist do?
- Business Intelligence (BI) vs. Data Science (DS)
- Future of data science
- Case Study: Diabetes Prevention

A Brief History of Data Science

The term data science has existed for the better part of the last 30 years and was originally used as a substitute for "computer science" in 1960. Approximately 15 years later, the term was used to define the survey of data processing methods used in different applications. In 2001, data science was introduced as an independent discipline.

What Is Data Science?

Data science is the field of applying advanced analytics techniques and scientific principles to extract valuable information from data for business decision-making, strategic planning and other uses.

KEY POINTS:

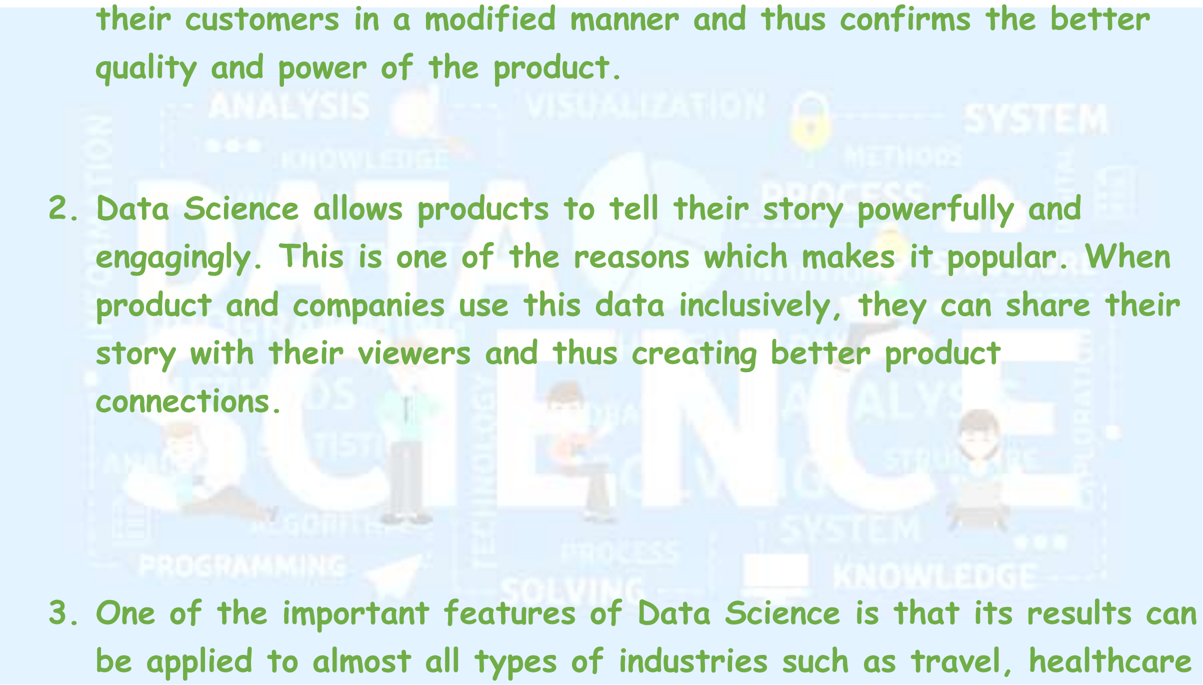
- Advances in technology, the Internet, social media, and the use of technology have all increased access to big data.
- Data science uses techniques such as machine learning and artificial intelligence to extract meaningful information and to predict future patterns and behaviors.
- The field of data science is growing as technology advances and big data collection and analysis techniques become more sophisticated.

The insights that data science generates help organizations increase operational efficiency, identify new business opportunities and improve marketing and sales programs, among other benefits.

Why is data science important?

Data science plays an important role in virtually all aspects of business operations and strategies. Below points will help you more to understand it's importance :

1. With the help of Data Science, the companies will be able to recognize their client in a more improved and enhanced way. Clients are the foundation of any product and play an essential role in their success and failure. Data Science enables companies to connect with their customers in a modified manner and thus confirms the better quality and power of the product.



2. Data Science allows products to tell their story powerfully and engagingly. This is one of the reasons which makes it popular. When product and companies use this data inclusively, they can share their story with their viewers and thus creating better product connections.

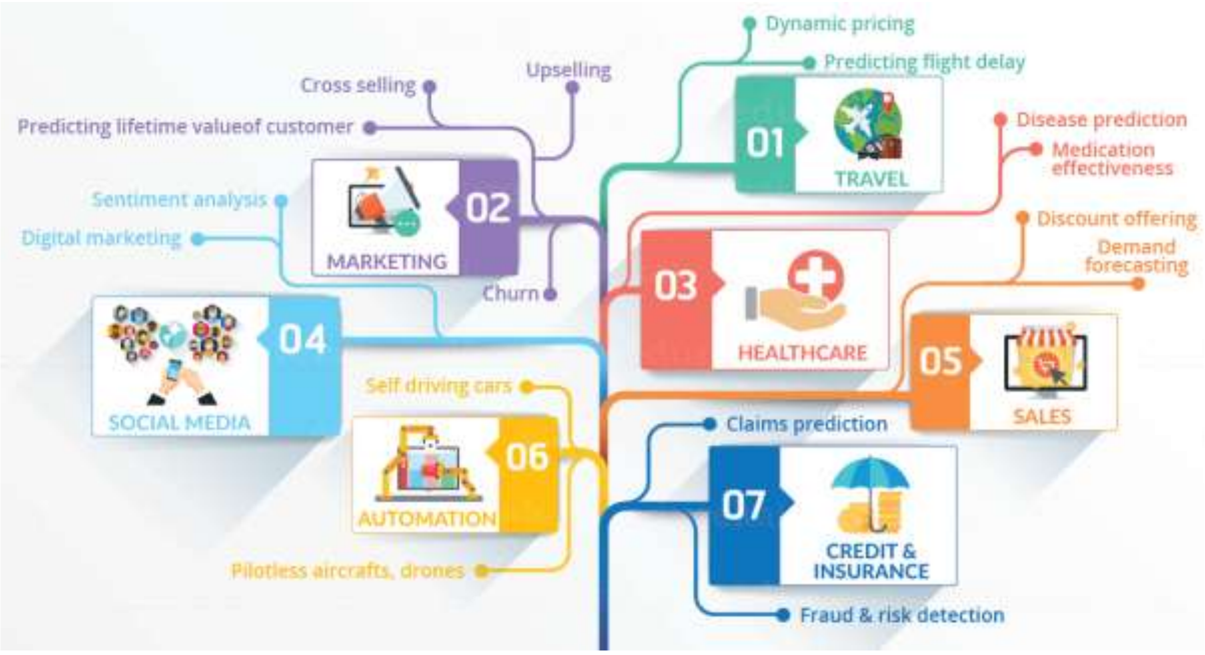
3. One of the important features of Data Science is that its results can be applied to almost all types of industries such as travel, healthcare and education. With the help of Data Science, the industries can analyze their challenges easily and can also address them effectively.

4. Data science is gaining popularity in every industry and thus playing a significant role in functioning and growth of any product. Therefore, the requirement of data scientist is also increased as they have to perform an important task of handling data and delivering solutions for the specific problems.

5. Data science also influenced the retail industries. Let's take an example to understand this, the older people were having a fantastic interaction with the local seller. The seller was also capable of fulfilling the requirements of the clients in a personalized way. But now due to the emergence and increase of supermarkets, this attention got lost. But with the help of data analytics, it is possible for the sellers to connect with their clients.

6. Data Science helps organizations to build this connection with the clients. With the help of data science, organizations and their products will be able to create a better and deep understanding of how customers can utilize their products.

Let’s have a look at the below infographic to see all the domains where Data Science is creating its impression.

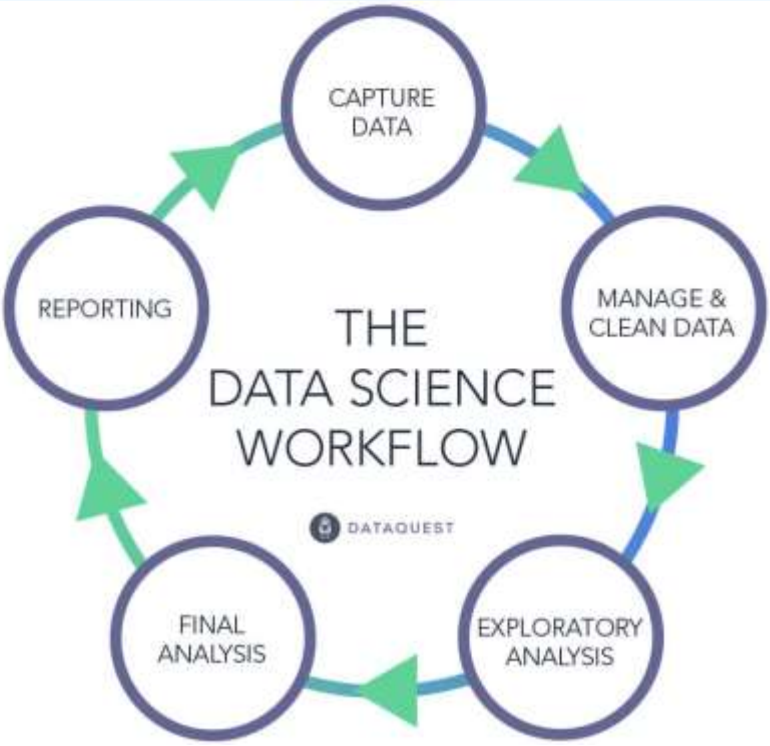


Data science process and lifecycle :

Data science projects involve a series of data collection and analysis steps.

Primary steps:

- Identify a business-related hypothesis to test.
- Gather data and prepare it for analysis.
- Experiment with different analytical models.
- Pick the best model and run it against the data.
- Present the results to business executives.
- Deploy the model for ongoing use with fresh data.



Why Data Science?

Interesting stats listed below prove Data Science career is the hottest profession in the new era:

- A Deloitte research says that by 2022, the income of data scientists will rise to as high as USD 130,176
- The job listings in data science in 2020 are supposed to grow by a whopping 364,000, with total number of openings skyrocket to 2,720,000
- The role of Advanced Analysts and Data Scientists will be most sought after in the job market in 2020, with an expected rise in demand by 28%
- The average salaries in data science and machine learning jobs are surging sky-high with a data scientist getting 105,000, & a data engineer taking home105,000, & a data engineer taking home117,000
- Forecasted demands for the much-crucial roles in data science industry – data developers, scientists, and engineers, indicate that in 2020, 700,000 new openings will emerge globally
- About 80% of the firms across the globe are investing a large part of their earnings into creating a skilful data analytics division, thus hiring the smartest of people in the industry domain

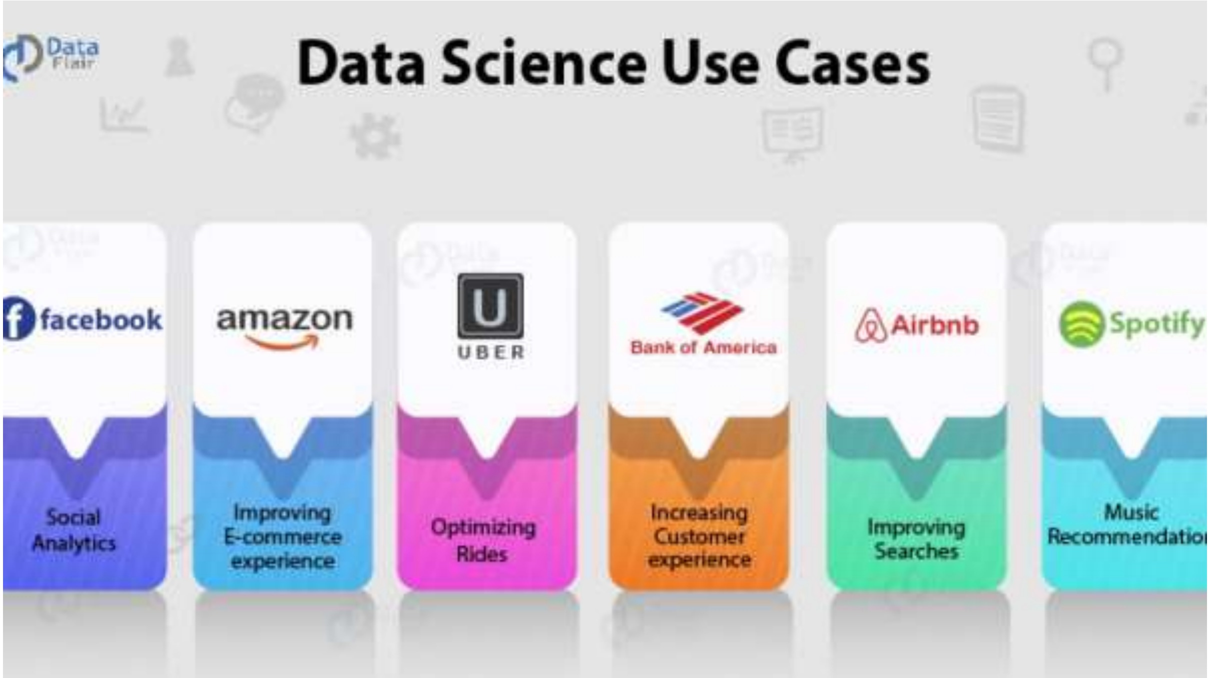
Top Skills in Data Science: Python, Data Mining, Data Analytics, Machine Learning, R.

Data science applications and use cases

Common applications that data scientists engage in include predictive modeling, pattern recognition, anomaly detection, classification, categorization and sentiment analysis, as well as development of technologies such as recommendation engines, personalization systems and artificial intelligence (AI) tools like chatbots and autonomous vehicles and machines.

Those applications drive a wide variety of use cases in organizations, including the following:

- customer analytics
- fraud detection
- risk management
- stock trading
- targeted advertising
- website personalization
- customer service
- predictive maintenance
- logistics and supply chain management
- image recognition
- speech recognition
- natural language processing
- cybersecurity
- medical diagnosis

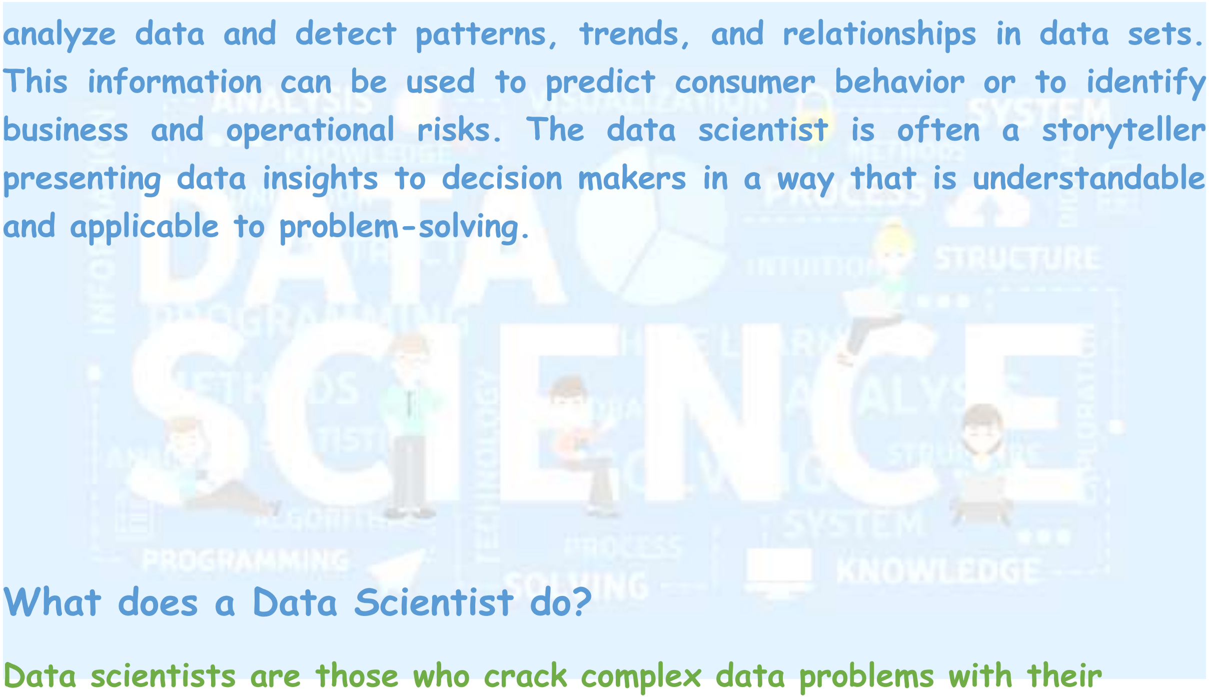


Challenges in data science

Data science is inherently challenging because of the advanced nature of the analytics it involves. The vast amounts of data typically being analyzed add to the complexity and increase the time it takes to complete projects.

One of the biggest challenges is eliminating bias in data sets and analytics applications. That includes issues with the underlying data itself and ones that data scientists unconsciously build into algorithms and predictive models. Such biases can skew analytics results if they aren't identified and addressed, creating flawed findings that lead to misguided business decisions. Even worse, they can have a harmful impact on groups of people

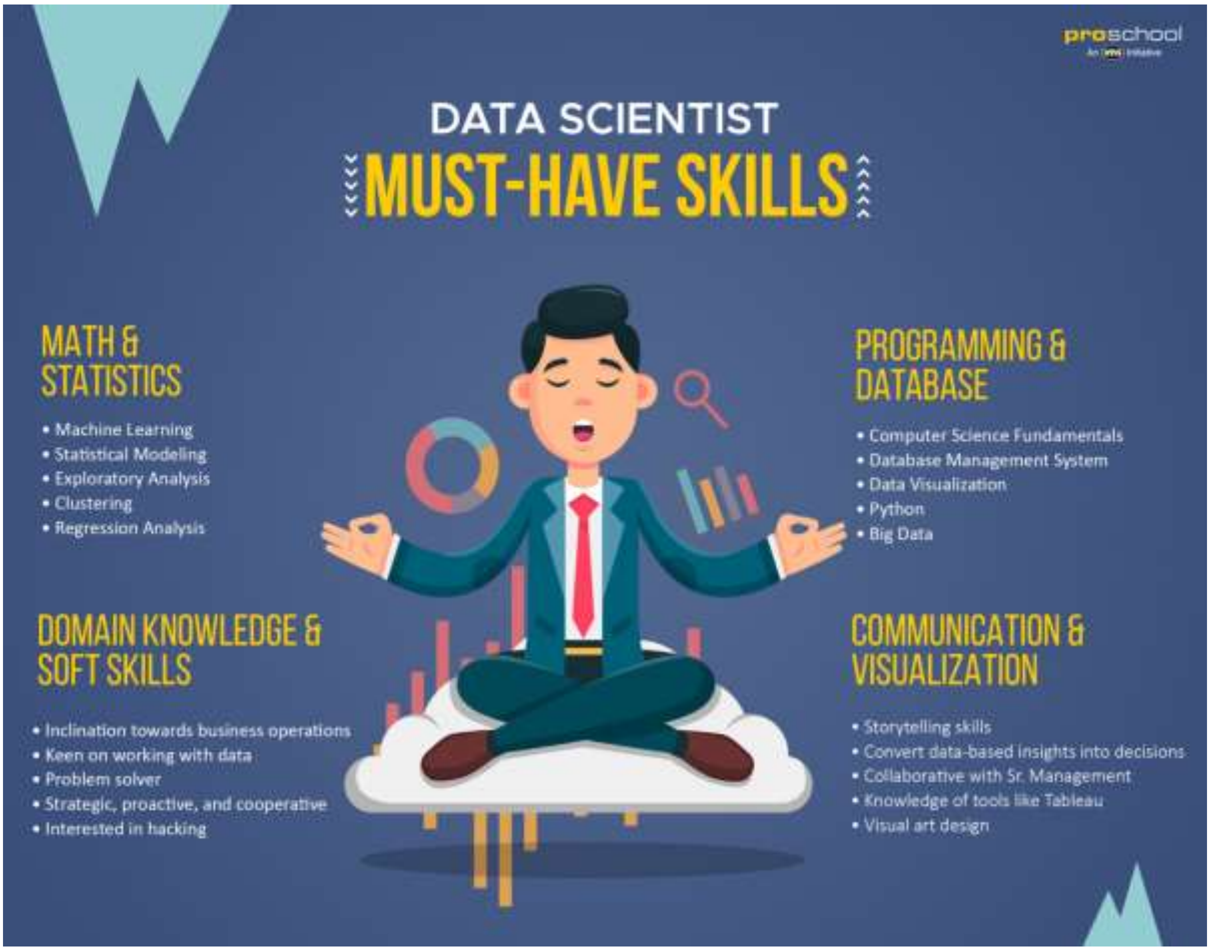
A data scientist collects, analyzes, and interprets large volumes of data, in many cases, to improve a company's operations. Data scientist professionals develop statistical models that analyze data and detect patterns, trends, and relationships in data sets. This information can be used to predict consumer behavior or to identify business and operational risks. The data scientist is often a storyteller presenting data insights to decision makers in a way that is understandable and applicable to problem-solving.



What does a Data Scientist do?

Data scientists are those who crack complex data problems with their strong expertise in certain scientific disciplines. They work with several elements related to mathematics, statistics, computer science, etc (though they may not be an expert in all these fields). They make a lot of use of the latest technologies in finding solutions and reaching conclusions that are crucial for an organization's growth and development. Data Scientists present the data in a much more useful form as compared to the raw data available to them from structured as well as unstructured forms.

A data scientist has a dual role – that of an "Analyst" as well as that of an "Artist"! Data scientists are very curious, who love a large amount of data, and more than that, they love to play with such huge data to reach important inferences and spot trends! This is what distinguishes a Data Scientist from a traditional Data Analyst. A Data scientist not only refers one particular source such as a social media site or a log file but various other sources with the aim to find out a hidden insight that can prove to be very significant for the organization.



DATA SCIENTIST

MUST-HAVE SKILLS

MATH & STATISTICS

- Machine Learning
- Statistical Modeling
- Exploratory Analysis
- Clustering
- Regression Analysis

PROGRAMMING & DATABASE

- Computer Science Fundamentals
- Database Management System
- Data Visualization
- Python
- Big Data

DOMAIN KNOWLEDGE & SOFT SKILLS

- Inclination towards business operations
- Keen on working with data
- Problem solver
- Strategic, proactive, and cooperative
- Interested in hacking

COMMUNICATION & VISUALIZATION

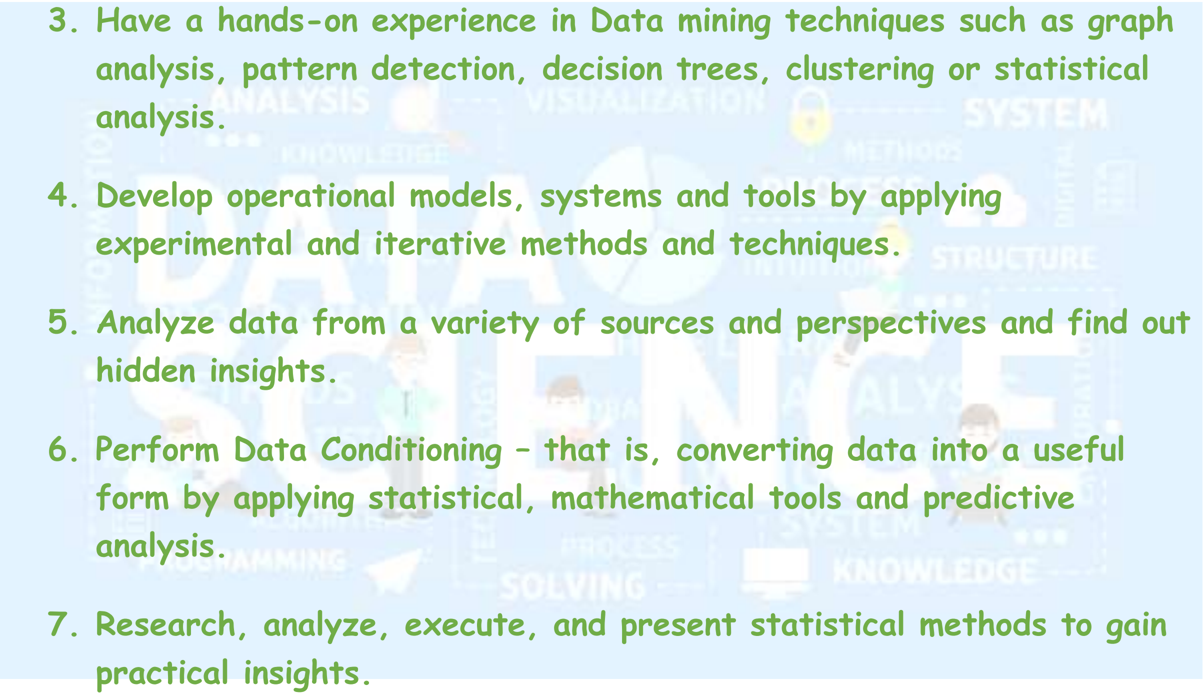
- Storytelling skills
- Convert data-based insights into decisions
- Collaborative with Sr. Management
- Knowledge of tools like Tableau
- Visual art design

What skills does a Data Scientist possess?

Role of a Data Scientist is indeed a challenging one! Though the skill-sets and competencies that Data Scientists employ differ extensively, to be an efficient Data scientist, he should:

1. Be very innovative and distinctive in his approach in applying various techniques intelligently to extract data and get useful insights in solving business problems and challenges.

2. Have the ability to locate and construe rich data sources.



3. Have a hands-on experience in Data mining techniques such as graph analysis, pattern detection, decision trees, clustering or statistical analysis.

4. Develop operational models, systems and tools by applying experimental and iterative methods and techniques.

5. Analyze data from a variety of sources and perspectives and find out hidden insights.

6. Perform Data Conditioning – that is, converting data into a useful form by applying statistical, mathematical tools and predictive analysis.

7. Research, analyze, execute, and present statistical methods to gain practical insights.

8. Manage large amounts of data even during hardware, software and bandwidth limitations.

9. Create visualizations that will help anyone understand the trends in data analysis with ease.

&

Be a team leader and communicate effectively with other business analysts, product Managers and Engineers.

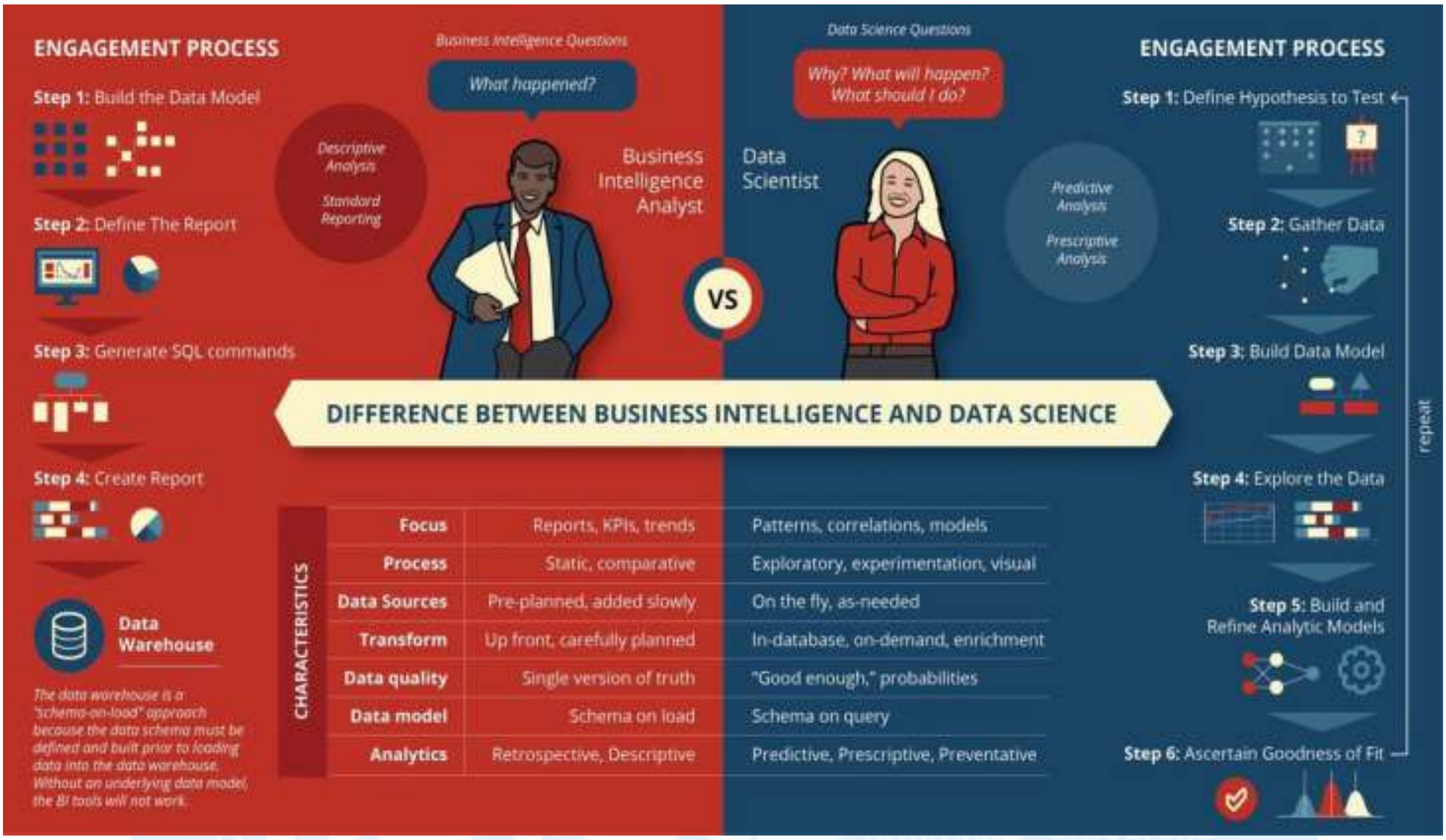
Business Intelligence (BI) vs. Data Science (DS)

Business Intelligence (BI) :

Basically analyzes the previous data to find hindsight and insight to describe business trends. Here BI enables you to take data from external and internal sources, prepare it, run queries on it and create dashboards to answer questions like quarterly revenue analysis or business problems. BI can evaluate the impact of certain events in the near future.

Data Science (DS) :

Is a more forward-looking approach, an exploratory way with the focus on analyzing the past or current data and predicting the future outcomes with the aim of making informed decisions. It answers the open-ended questions as to "what" and "how" events occur.



DIFFERENCE BETWEEN BUSINESS INTELLIGENCE AND DATA SCIENCE

ENGAGEMENT PROCESS

Step 1: Build the Data Model

Step 2: Define The Report

Step 3: Generate SQL commands

Step 4: Create Report

Data Warehouse

The data warehouse is a "schema-on-load" approach because the data schema must be defined and built prior to loading data into the data warehouse. Without an underlying data model, the BI tools will not work.

Business Intelligence Questions

What happened?

Descriptive Analysis

Standard Reporting

Business Intelligence Analyst

VS

Data Scientist

Data Science Questions

Why? What will happen? What should I do?

Predictive Analysis

Prescriptive Analysis

ENGAGEMENT PROCESS

Step 1: Define Hypothesis to Test

Step 2: Gather Data

Step 3: Build Data Model

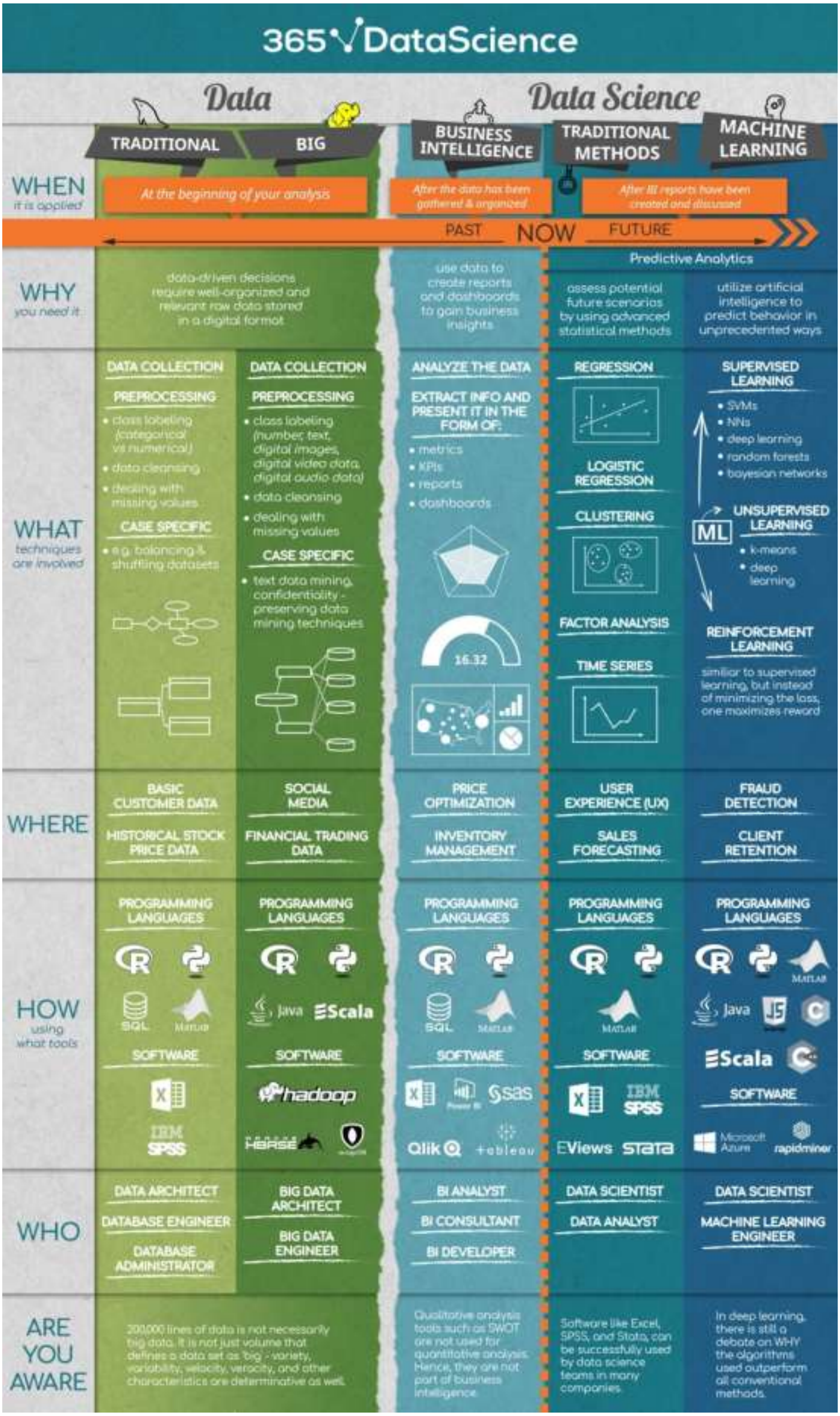
Step 4: Explore the Data

Step 5: Build and Refine Analytic Models

Step 6: Ascertain Goodness of Fit

repeat

CHARACTERISTICS	Business Intelligence Analyst	Data Scientist
Focus	Reports, KPIs, trends	Patterns, correlations, models
Process	Static, comparative	Exploratory, experimentation, visual
Data Sources	Pre-planned, added slowly	On the fly, as-needed
Transform	Up front, carefully planned	In-database, on-demand, enrichment
Data quality	Single version of truth	"Good enough," probabilities
Data model	Schema on load	Schema on query
Analytics	Retrospective, Descriptive	Predictive, Prescriptive, Preventative



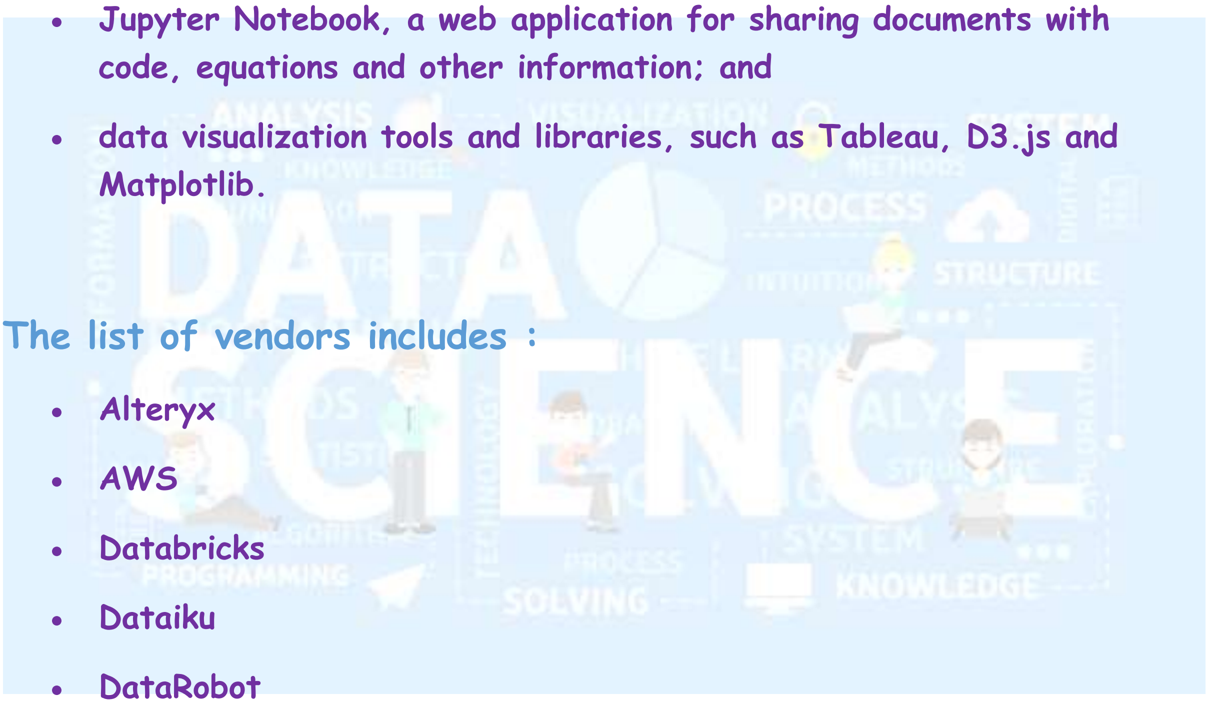
365√DataScience

	Data		Data Science		
	TRADITIONAL	BIG	BUSINESS INTELLIGENCE	TRADITIONAL METHODS	MACHINE LEARNING
WHEN *it is applied*	*At the beginning of your analysis*		*After the data has been gathered & organized*	*After BI reports have been created and discussed*	
	←		PAST	NOW	FUTURE →
				Predictive Analytics	
WHY *you need it*	data-driven decisions require well-organized and relevant raw data stored in a digital format		use data to create reports and dashboards to gain business insights	assess potential future scenarios by using advanced statistical methods	utilize artificial intelligence to predict behavior in unprecedented ways
WHAT *techniques are involved*	DATA COLLECTION PREPROCESSING • class labeling (categorical vs numerical) • data cleansing • dealing with missing values CASE SPECIFIC • e.g. balancing & shuffling datasets	DATA COLLECTION PREPROCESSING • class labeling (number, text, digital images, digital video data, digital audio data) • data cleansing • dealing with missing values CASE SPECIFIC • text data mining, confidentiality - preserving data mining techniques	ANALYZE THE DATA EXTRACT INFO AND PRESENT IT IN THE FORM OF: • metrics • KPIs • reports • dashboards 16.32	REGRESSION LOGISTIC REGRESSION CLUSTERING FACTOR ANALYSIS TIME SERIES	SUPERVISED LEARNING • SVMs • NNs • deep learning • random forests • bayesian networks ML UNSUPERVISED LEARNING • k-means • deep learning REINFORCEMENT LEARNING similiar to supervised learning, but instead of minimizing the loss, one maximizes reward
WHERE	BASIC CUSTOMER DATA HISTORICAL STOCK PRICE DATA	SOCIAL MEDIA FINANCIAL TRADING DATA	PRICE OPTIMIZATION INVENTORY MANAGEMENT	USER EXPERIENCE (UX) SALES FORECASTING	FRAUD DETECTION CLIENT RETENTION
HOW *using what tools*	PROGRAMMING LANGUAGES R SQL MATLAB SOFTWARE X IBM SPSS	PROGRAMMING LANGUAGES R Java Scala SOFTWARE hadoop HBASE	PROGRAMMING LANGUAGES R SQL MATLAB SOFTWARE X Power BI sas Qlik tableau	PROGRAMMING LANGUAGES R MATLAB SOFTWARE X IBM SPSS EViews Stata	PROGRAMMING LANGUAGES R MATLAB Java JS C Scala C SOFTWARE Microsoft Azure rapidminer
WHO	DATA ARCHITECT DATABASE ENGINEER DATABASE ADMINISTRATOR	BIG DATA ARCHITECT BIG DATA ENGINEER	BI ANALYST BI CONSULTANT BI DEVELOPER	DATA SCIENTIST DATA ANALYST	DATA SCIENTIST MACHINE LEARNING ENGINEER
ARE YOU AWARE	200,000 lines of data is not necessarily big data. It is not just volume that defines a data set as 'big' - variety, variability, velocity, veracity, and other characteristics are determinative as well.		Qualitative analysis tools such as SWOT are not used for quantitative analysis. Hence, they are not part of business intelligence.	Software like Excel, SPSS, and Stata, can be successfully used by data science teams in many companies.	In deep learning, there is still a debate on WHY the algorithms used outperform all conventional methods.

Data science tools and platforms

Numerous tools are available for data scientists to use in the analytics process, including both commercial and open source options:

- data platforms and analytics engines, such as Spark, Hadoop and NoSQL databases;
- programming languages, such as Python, R, Julia, Scala and SQL;
- statistical analysis tools like SAS and IBM SPSS;
- machine learning platforms and libraries, including TensorFlow, Weka, Scikit-learn, Keras and PyTorch;
- Jupyter Notebook, a web application for sharing documents with code, equations and other information; and
- data visualization tools and libraries, such as Tableau, D3.js and Matplotlib.



The list of vendors includes :

- Alteryx
- AWS
- Databricks
- Dataiku
- DataRobot
- Domino Data Lab
- Google
- H2O.ai
- IBM
- Knime
- MathWorks
- Microsoft
- RapidMiner

- SAS Institute
- Tibco Software and others.

How industries rely on data science

Amazon were early users of data science and big data analytics for internal applications, along with other internet and e-commerce companies like Before they became technology vendors themselves, Google and Facebook, Yahoo and eBay. Now, data science is widespread in organizations of all kinds. Here are some examples of how it's used in different industries:

Entertainment.

Data science enables streaming services to track and analyze what users watch, which helps determine the new TV shows and films they produce. Data-driven algorithms are also used to create personalized recommendations based on a user's viewing history.

Financial services.

Banks and credit card companies mine and analyze data to detect fraudulent transactions, manage financial risks on loans and credit lines, and evaluate customer portfolios to identify upselling opportunities.

Healthcare.

Hospitals and other healthcare providers use machine learning models and additional data science components to automate X-ray analysis and aid doctors in diagnosing illnesses and planning treatments based on previous patient outcomes.

Manufacturing.

Data science uses at manufacturers include optimization of supply chain management and distribution, plus predictive maintenance to detect potential equipment failures in plants before they occur.

Retail.

Retailers analyze customer behavior and buying patterns to drive personalized product recommendations and targeted advertising, marketing and promotions. Data science also helps them manage product inventories and their supply chains to keep items in stock.

Transportation.

Delivery companies, freight carriers and logistics services providers use data science to optimize delivery routes and schedules, as well as the best modes of transport for shipments.

Travel.

Data science aids airlines with flight planning to optimize routes, crew scheduling and passenger loads. Algorithms also drive variable pricing for flights and hotel rooms.

Other data science uses, in areas such as cybersecurity, customer service and business process management, are common across different industries.

Is there any Scope of Data Science in India? Take Expert’s Opinion

At what point will we meet the equilibrium? A report stated that statistics even increased when it came to 2018; they estimated Data Science job vacancies were 2.9 million. The demand is shooting up like anything, and in the near future, it is said that the companies require more Data Scientist. The requirement will grow and can never decrease.

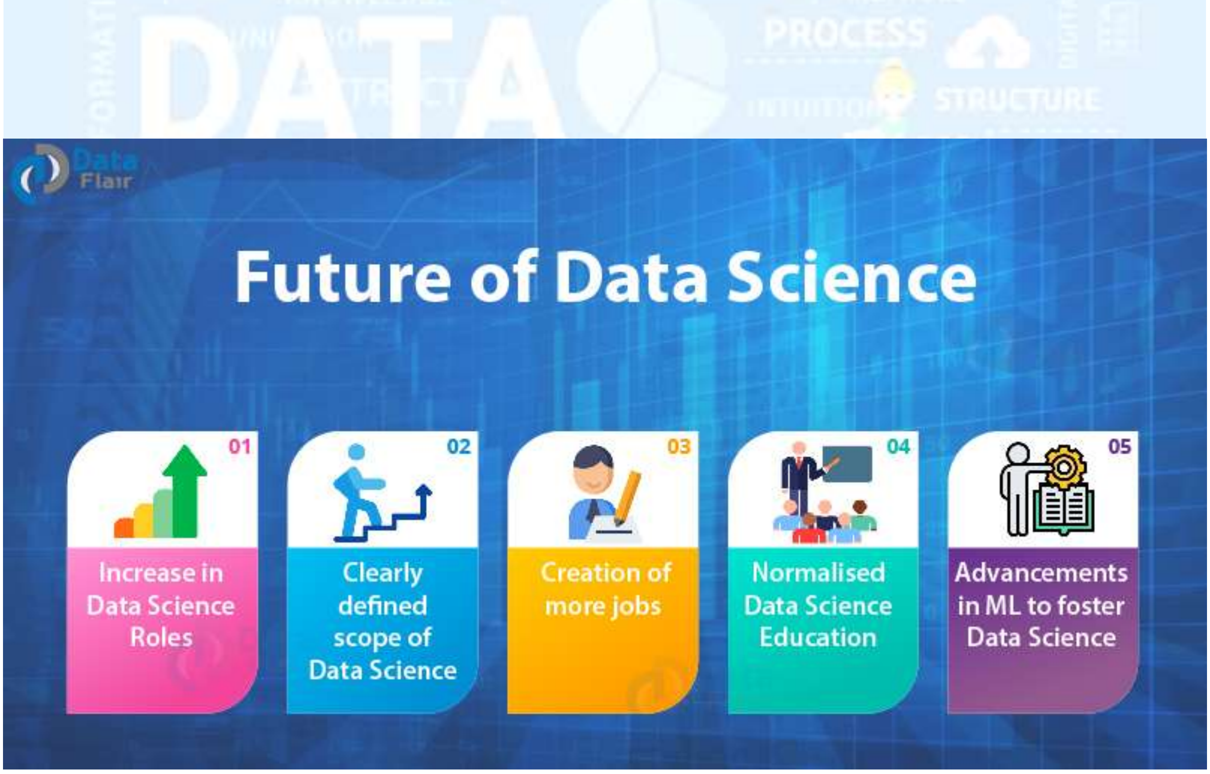
Do you know Flipkart has 700 job openings for Data science and other technical areas? Not only this, Amazon, Netflix and many such big companies have demand and massive job openings for data scientists.

Data Science is experiencing a surge in jobs across the World. India is one such country that is experiencing a data explosion. The scope of data science in India and the need for IT professionals to upgrade their skills in the field of Data Science is increasing.

India has been the center of software and IT industry. With the gradual degradation of traditional IT positions through automation, the Indian IT industry is experiencing a major transformation. This is the new age of data and it is a need of the hour for professionals to update themselves in order to sustain their relevancy.

Future of data science

Future Scope Of Data Science



DATA
PROCESS
STRUCTURE
Data Flair

Future of Data Science

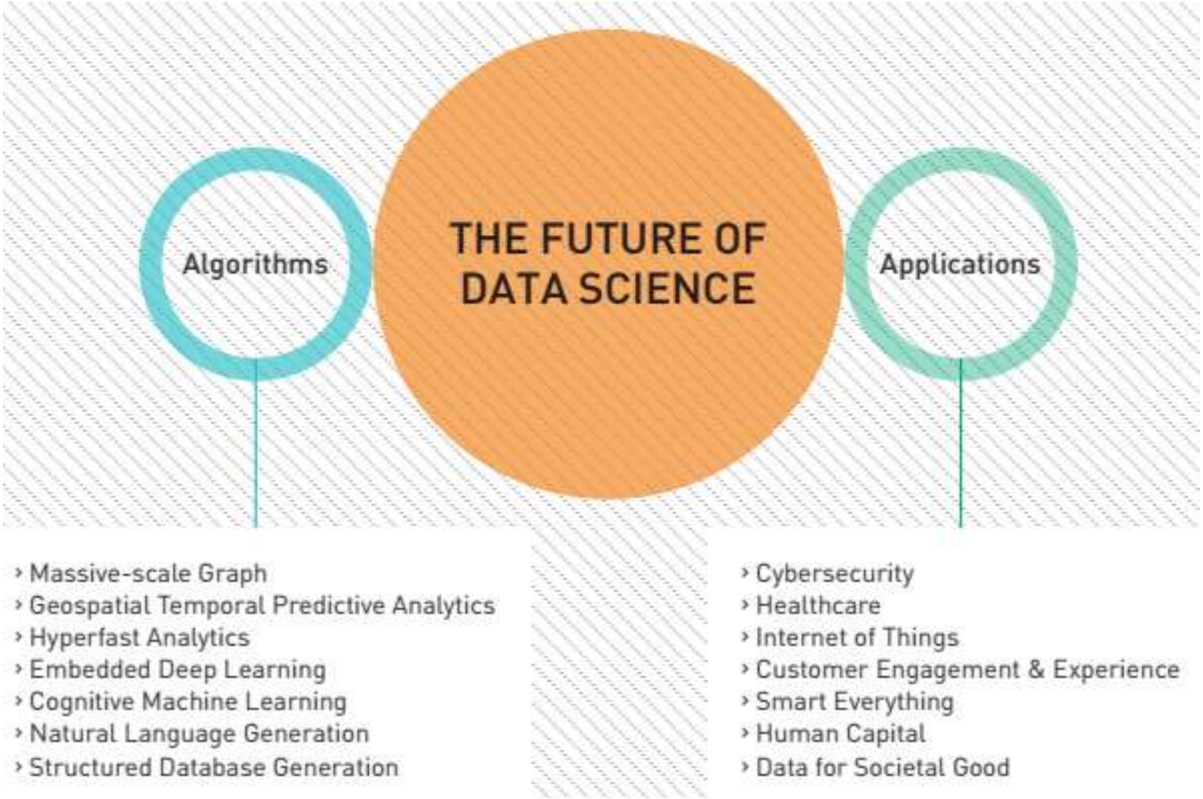
01 Increase in Data Science Roles

02 Clearly defined scope of Data Science

03 Creation of more jobs

04 Normalised Data Science Education

05 Advancements in ML to foster Data Science



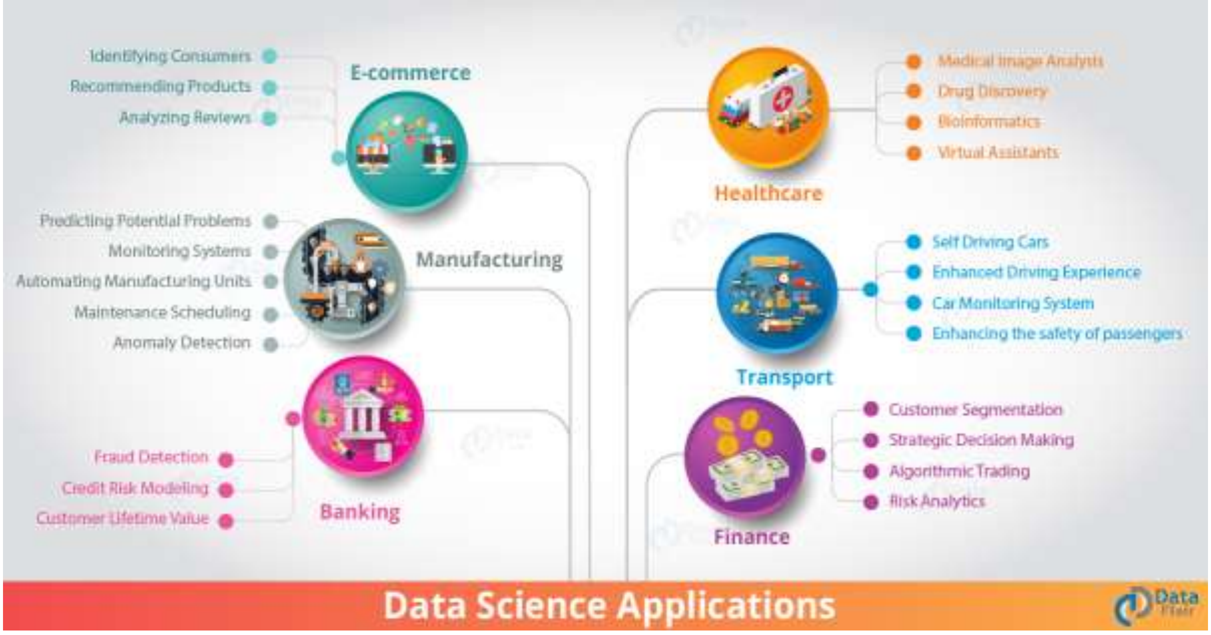
THE FUTURE OF DATA SCIENCE

Algorithms

- Massive-scale Graph
- Geospatial Temporal Predictive Analytics
- Hyperfast Analytics
- Embedded Deep Learning
- Cognitive Machine Learning
- Natural Language Generation
- Structured Database Generation

Applications

- Cybersecurity
- Healthcare
- Internet of Things
- Customer Engagement & Experience
- Smart Everything
- Human Capital
- Data for Societal Good



Data Science Applications

E-commerce
- Identifying Consumers
- Recommending Products
- Analyzing Reviews

Manufacturing
- Predicting Potential Problems
- Monitoring Systems
- Automating Manufacturing Units
- Maintenance Scheduling
- Anomaly Detection

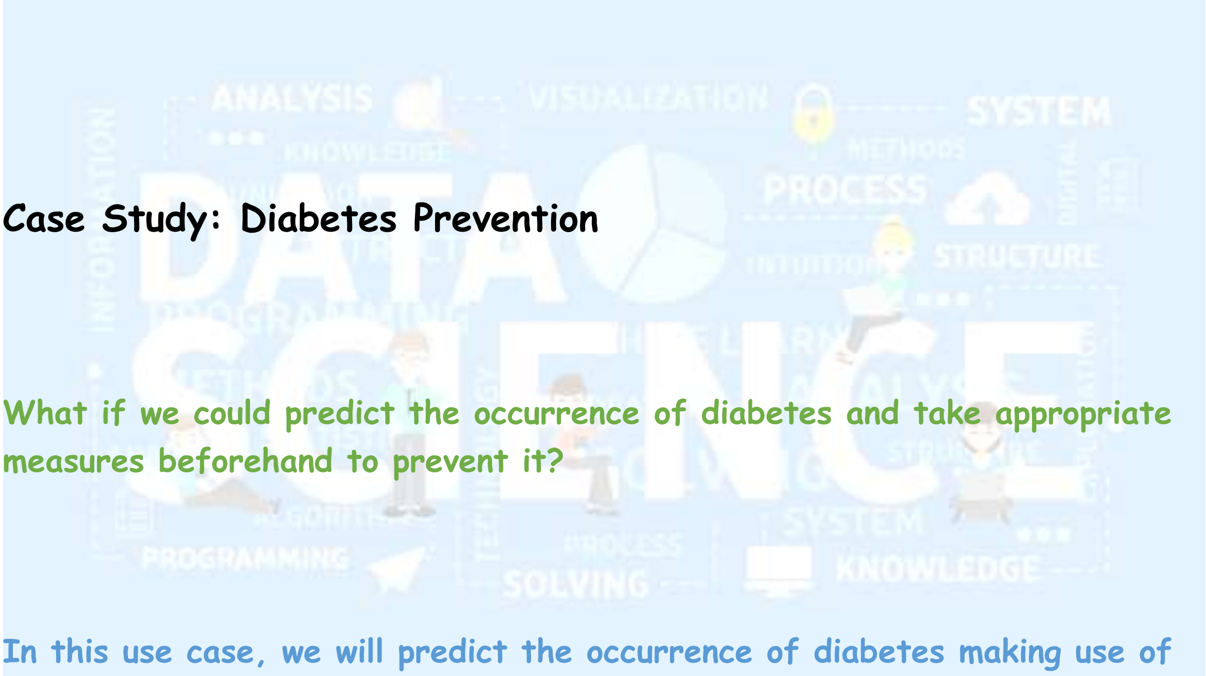
Banking
- Fraud Detection
- Credit Risk Modeling
- Customer Lifetime Value

Healthcare
- Medical Image Analysis
- Drug Discovery
- Bioinformatics
- Virtual Assistants

Transport
- Self Driving Cars
- Enhanced Driving Experience
- Car Monitoring System
- Enhancing the safety of passengers

Finance
- Customer Segmentation
- Strategic Decision Making
- Algorithmic Trading
- Risk Analytics

Future of Data Science and Data Scientist : After the next 5 years, they will develop the ability to use all sorts of data in real-time. For the needs of the future, it will spark the emergence of new data science paradigms. We can use more data to drive key business decisions. We will enable innovations like "Deep Learning". it allows for accurate predictions and decision making. Further, modern applications have brought to fore new statistical paradigms. The most important thing: Skilled data scientists; statisticians, and, business analysts will be the key to unlocking the endless possibilities of big data.



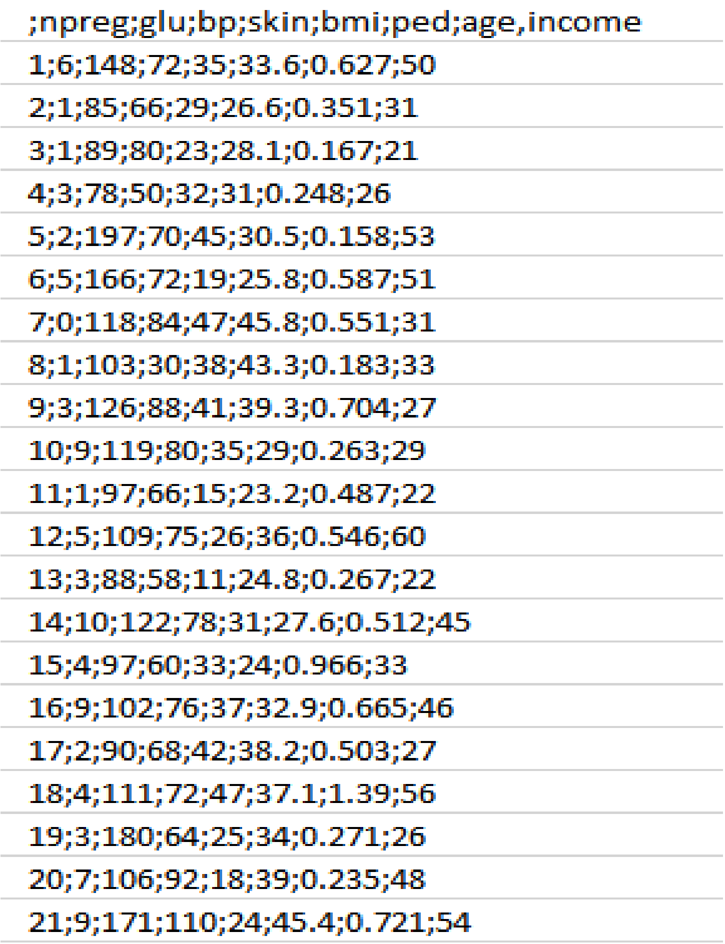
Case Study: Diabetes Prevention

What if we could predict the occurrence of diabetes and take appropriate measures beforehand to prevent it?

In this use case, we will predict the occurrence of diabetes making use of the entire lifecycle that we discussed earlier. Let's go through the various steps.

Step 1:

- First, we will collect the data based on the medical history of the patient as discussed in Phase 1. You can refer to the sample data below.



```
;npreg;glu;bp;skin;bmi;ped;age,income
1;6;148;72;35;33.6;0.627;50
2;1;85;66;29;26.6;0.351;31
3;1;89;80;23;28.1;0.167;21
4;3;78;50;32;31;0.248;26
5;2;197;70;45;30.5;0.158;53
6;5;166;72;19;25.8;0.587;51
7;0;118;84;47;45.8;0.551;31
8;1;103;30;38;43.3;0.183;33
9;3;126;88;41;39.3;0.704;27
10;9;119;80;35;29;0.263;29
11;1;97;66;15;23.2;0.487;22
12;5;109;75;26;36;0.546;60
13;3;88;58;11;24.8;0.267;22
14;10;122;78;31;27.6;0.512;45
15;4;97;60;33;24;0.966;33
16;9;102;76;37;32.9;0.665;46
17;2;90;68;42;38.2;0.503;27
18;4;111;72;47;37.1;1.39;56
19;3;180;64;25;34;0.271;26
20;7;106;92;18;39;0.235;48
21;9;171;110;24;45.4;0.721;54
```

- As you can see, we have the various attributes as mentioned below.

Attributes:

1.npreg – Number of times pregnant

2.glucose – Plasma glucose concentration

3.bp – Blood pressure

4.skin – Triceps skinfold thickness

5.bmi – Body mass index

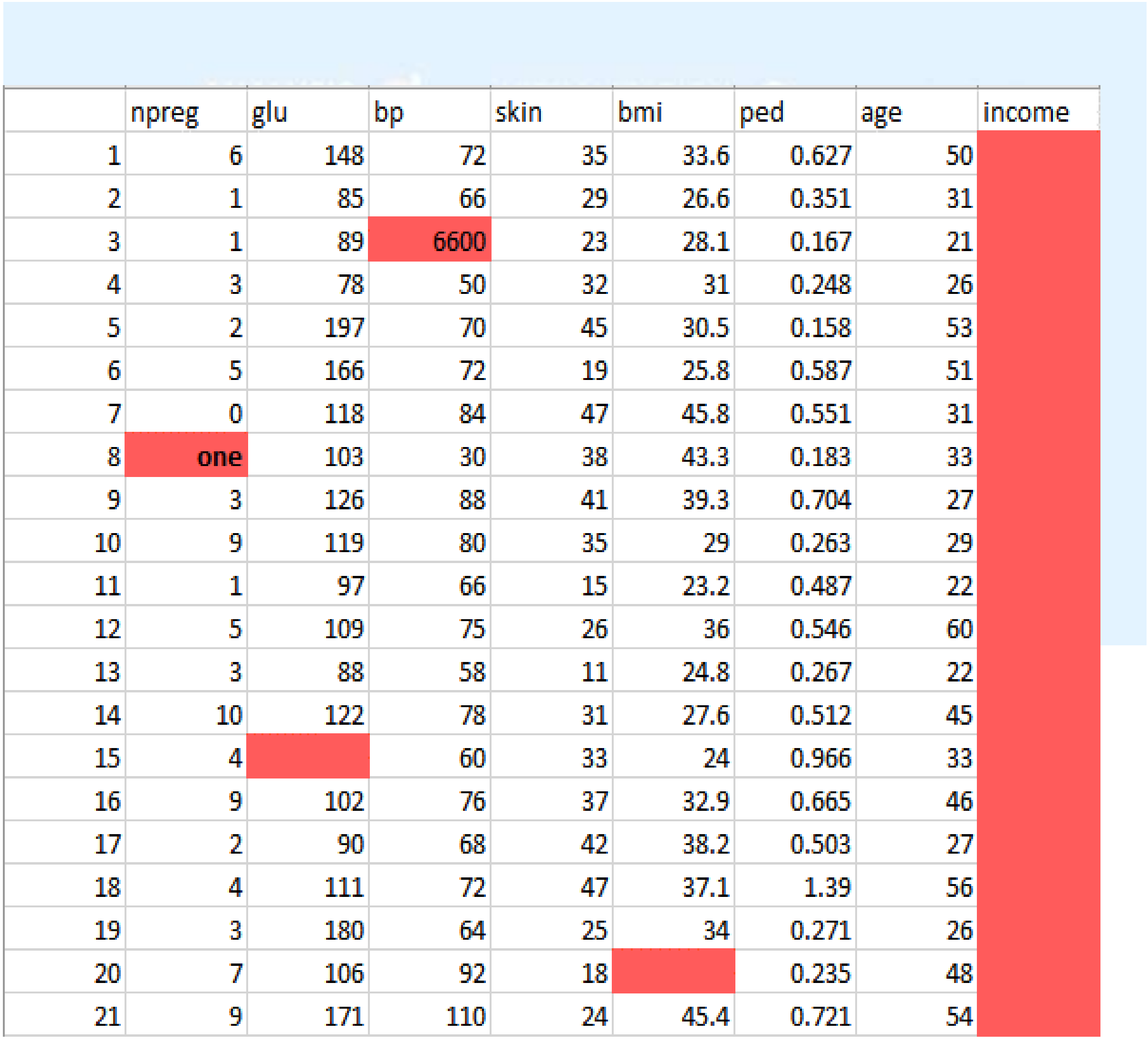
6.ped – Diabetes pedigree function

7.age – Age

8.income – Income

Step 2:

- Now, once we have the data, we need to clean and prepare the data for data analysis.
- This data has a lot of inconsistencies like missing values, blank columns, abrupt values and incorrect data format which need to be cleaned.
- Here, we have organized the data into a single table under different attributes – making it look more structured.
- Let's have a look at the sample data below.



	npreg	glu	bp	skin	bmi	ped	age	income
1	6	148	72	35	33.6	0.627	50	
2	1	85	66	29	26.6	0.351	31	
3	1	89	6600	23	28.1	0.167	21	
4	3	78	50	32	31	0.248	26	
5	2	197	70	45	30.5	0.158	53	
6	5	166	72	19	25.8	0.587	51	
7	0	118	84	47	45.8	0.551	31	
8	one	103	30	38	43.3	0.183	33	
9	3	126	88	41	39.3	0.704	27	
10	9	119	80	35	29	0.263	29	
11	1	97	66	15	23.2	0.487	22	
12	5	109	75	26	36	0.546	60	
13	3	88	58	11	24.8	0.267	22	
14	10	122	78	31	27.6	0.512	45	
15	4		60	33	24	0.966	33	
16	9	102	76	37	32.9	0.665	46	
17	2	90	68	42	38.2	0.503	27	
18	4	111	72	47	37.1	1.39	56	
19	3	180	64	25	34	0.271	26	
20	7	106	92	18		0.235	48	
21	9	171	110	24	45.4	0.721	54	

This data has a lot of inconsistencies.

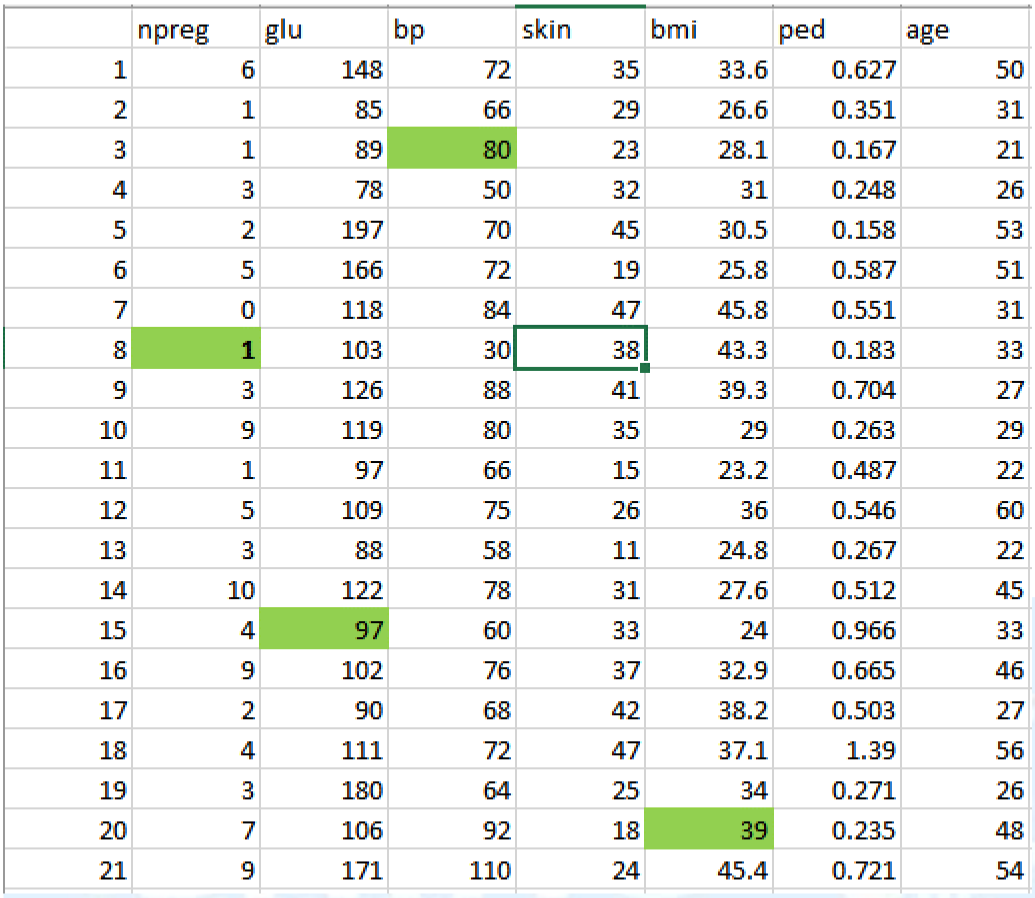
1. In the column npreg, “one” is written in words, whereas it should be in the numeric form like 1.

2. In column bp one of the values is 6600 which is impossible (at least for humans) as bp cannot go up to such huge value.

3. As you can see the Income column is blank and also makes no sense in predicting diabetes. Therefore, it is redundant to have it here and should be removed from the table.

- So,we will clean and preprocess this data by removing the outliers, filling up the null values and normalizing the data type. If you remember,this is our second phase which is data preprocessing.

- Finally, we get the clean data as shown below which can be used for analysis.



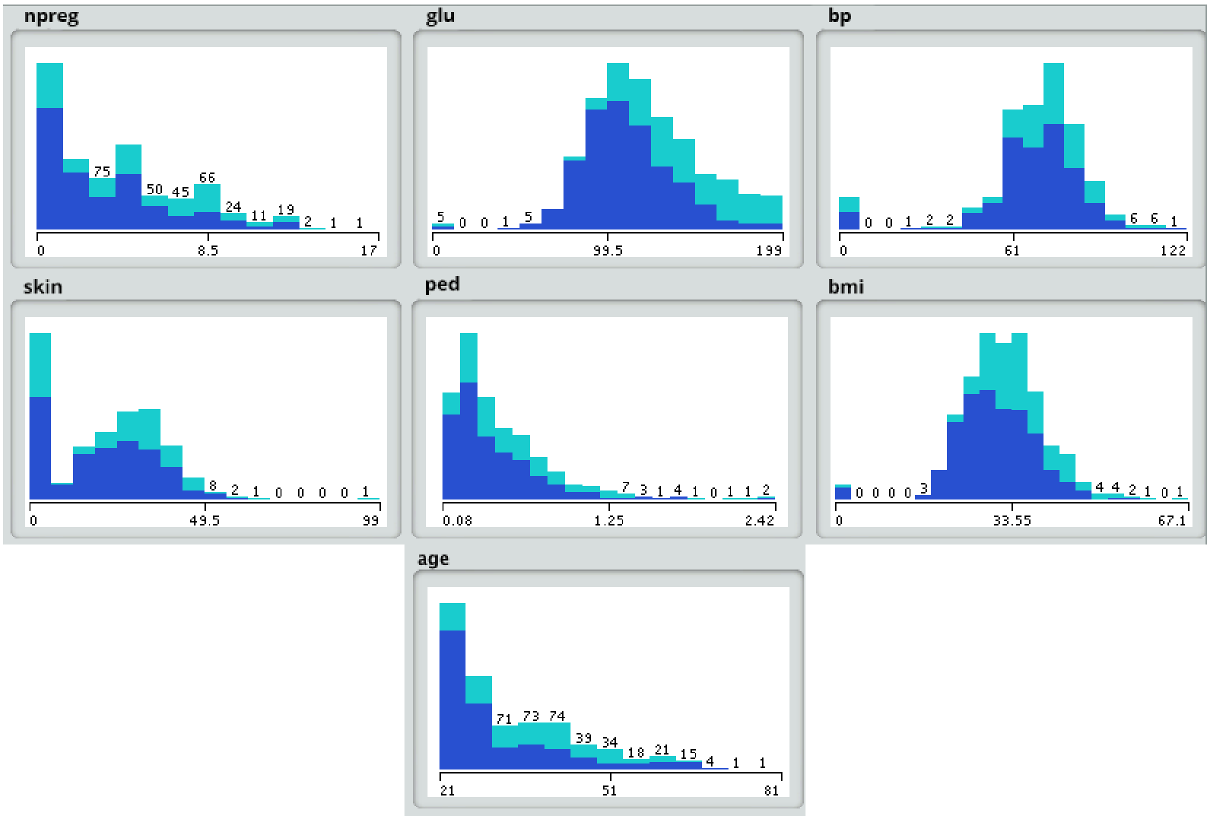
	npreg	glu	bp	skin	bmi	ped	age
1	6	148	72	35	33.6	0.627	50
2	1	85	66	29	26.6	0.351	31
3	1	89	80	23	28.1	0.167	21
4	3	78	50	32	31	0.248	26
5	2	197	70	45	30.5	0.158	53
6	5	166	72	19	25.8	0.587	51
7	0	118	84	47	45.8	0.551	31
8	1	103	30	38	43.3	0.183	33
9	3	126	88	41	39.3	0.704	27
10	9	119	80	35	29	0.263	29
11	1	97	66	15	23.2	0.487	22
12	5	109	75	26	36	0.546	60
13	3	88	58	11	24.8	0.267	22
14	10	122	78	31	27.6	0.512	45
15	4	97	60	33	24	0.966	33
16	9	102	76	37	32.9	0.665	46
17	2	90	68	42	38.2	0.503	27
18	4	111	72	47	37.1	1.39	56
19	3	180	64	25	34	0.271	26
20	7	106	92	18	39	0.235	48
21	9	171	110	24	45.4	0.721	54

Step 3:

Now let's do some analysis as discussed earlier in Phase 3.

- First, we will load the data into the analytical sandbox and apply various statistical functions on it. For example, R has functions like describe which gives us the number of missing values and unique values. We can also use the summary function which will give us statistical information like mean, median, range, min and max values.

- Then, we use visualization techniques like histograms, line graphs, box plots to get a fair idea of the distribution of data.

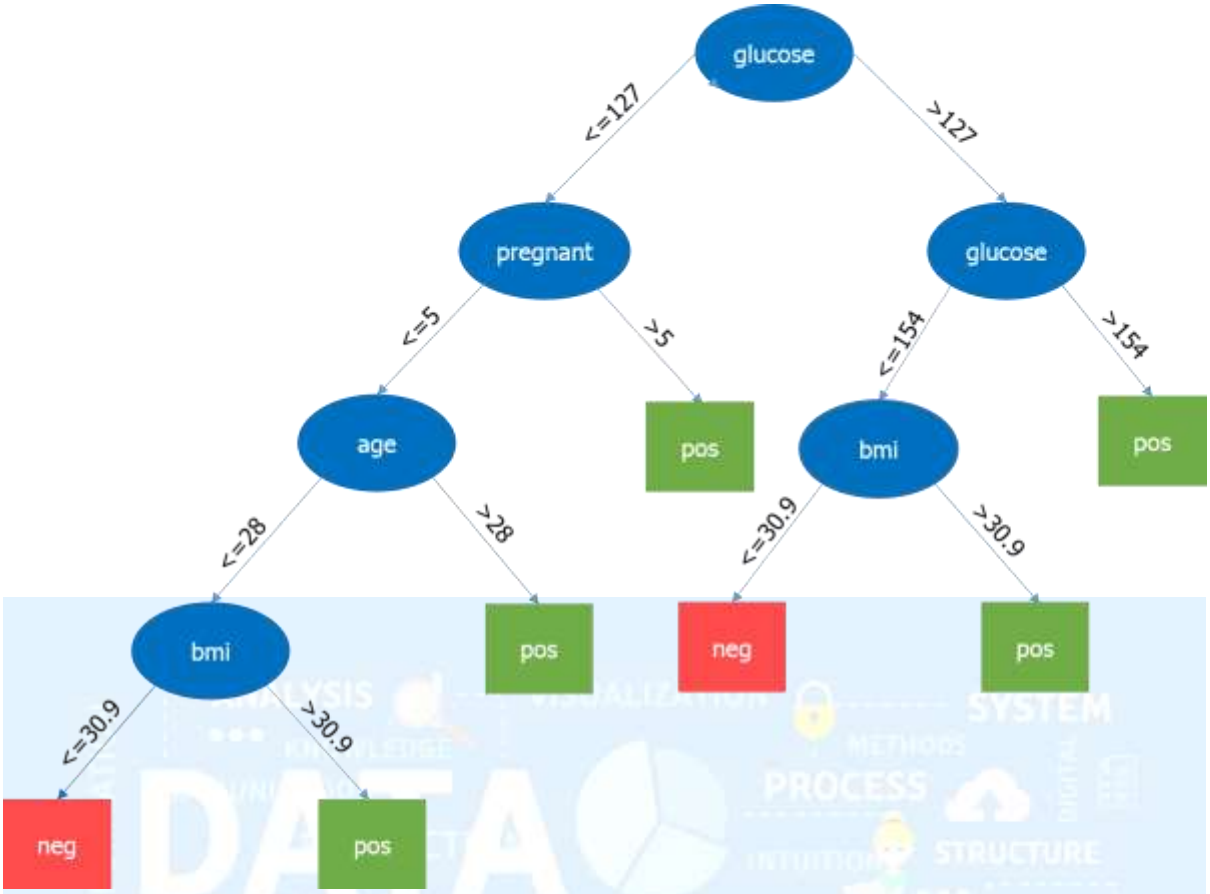


Step 4:

Now, based on insights derived from the previous step, the best fit for this kind of problem is the decision tree. Let's see how?

- Since, we already have the major attributes for analysis like npreg, bmi, etc., so we will use supervised learning technique to build a model here.
- Further, we have particularly used decision tree because it takes all attributes into consideration in one go, like the ones which have a linear relationship as well as those which have a non-linear relationship. In our case, we have a linear relationship between npreg and age, whereas the nonlinear relationship between npreg and ped.
- Decision tree models are also very robust as we can use the different combination of attributes to make various trees and then finally implement the one with the maximum efficiency.

Let's have a look at our decision tree.



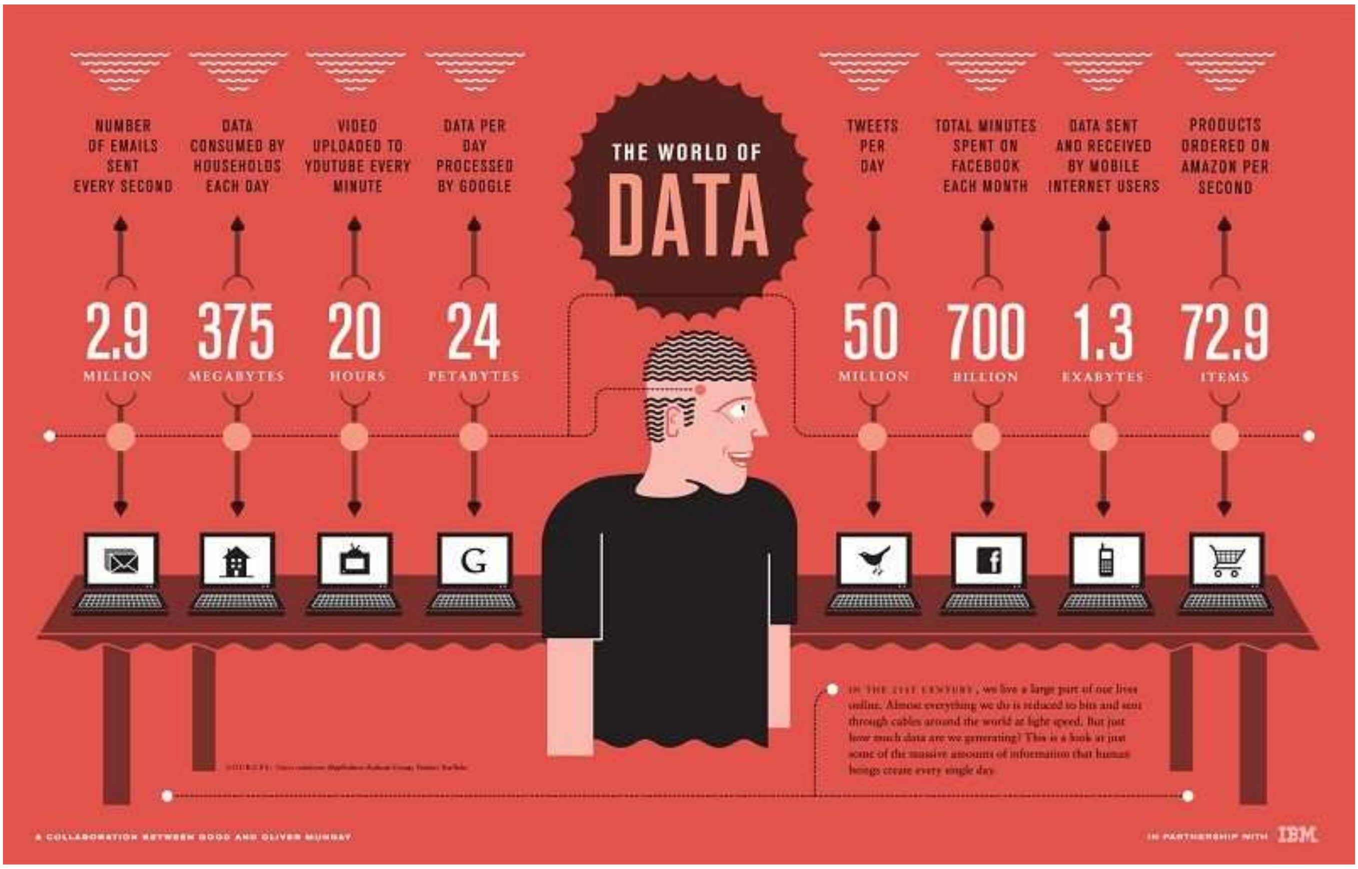
Here, the most important parameter is the level of glucose, so it is our root node. Now, the current node and its value determine the next important parameter to be taken. It goes on until we get the result in terms of pos or neg. Pos means the tendency of having diabetes is positive and neg means the tendency of having diabetes is negative.

Step 5:

In this phase, we will run a small pilot project to check if our results are appropriate. We will also look for performance constraints if any. If the results are not accurate, then we need to replan and rebuild the model.

Step 6:

Once we have executed the project successfully, we will share the output for full deployment.



IN THE 21ST CENTURY, we live a large part of our lives online. Almost everything we do is reduced to bits and sent through cables around the world at light speed. But just how much data are we generating? This is a look at just some of the massive amounts of information that human beings create every single day.

Thank You

!!!!!!!!