Undoubtedly, Data Science is the most revolutionary technology of the era.

It’s all about deriving useful insights from data in order to solve real-world complex problems.

this contains everything that you need to know in order to master data science.

Now before we get started let’s take a look at the agenda. The first module is a reduction to data science that covers all the basic fundamentals of data science followed by this.

where you all need to understand the statistics and math behind data science and machine learning algorithms.

and so on the next module is the supervised learning.

algorithms module.

Linear regression.

The next module is.

the logistic regression modulewhere we will see.

how logistic regressioncan be used to solve.

classification problems..

"After this well discussabout decision trees",

"and well see",

how decision treescan be used to solve.

complex data-driven problems..

The next module is random Foresthere will understand.

how random Forest can be usedto solve classification problems.

and regression problemswith the help.

of use cases and examples..

The next module.

will be be discussing isthe k-nearest neighbor module..

We will understand how gainand can be used to solve.

complex classification problemsfollowed by this..

We look at thenaive bias module,.

which is one of the mostimportant algorithms.

in the Gmail spam detection..

The next algorithmis support Vector machine.

where we will understand.

"how svms can be used",

to draw a hyperplane betweendifferent classes of data..

Finally..

We move on to the unsupervisedlearning module where we.

will understand how genescan be used for clustering..

And how you can performMarket Basket analysis by using.

Association rule mining..

The next moduleis reinforcement learning.

where we will understandthe different concepts.

of reinforcement learning.

along with a couple.

of demonstrationsfollowed by this bill..

Look at the Deep learning module.

where we will understand whatexactly deep learning is what.

our neural networks.

with different typesof neural networks..

And so on..

The last module is.

the data scienceinterview questions module.

where we will understandthe important concepts of data..

Along with a few tips in orderto Ace the interview now.

before we get startedmake sure you subscribe.

to Adorama YouTube channelin order to stay updated.

about the most trendingTechnologies data science is one.

of the most in-demandTechnologies right now..

Now this is probably.

"because were generating dataat an Unstoppable pace.",

And obviously we need to process.

and make sense outof this much data..

This is exactly.

"where data science comesin in todays session.",

"Well be talkingabout data science in depth.",

"So lets move ahead and takea look at todays agenda.",

"Were going to begin",

with discussing the varioussources of data and.

how the evolution of technology.

and introduction of IOD.

and social media have ledto the need of data sign next..

"Well discuss how Walmartis using insightful patterns",

from their database to increasethe potential of their business..

After that..

We will see whatexactly data science is,.

"then well move on and discusswho are data scientist is",

where we will also discussthe various skill sets..

Needed to becomea data scientist next.

we can move on to see.

the various datascience job roles.

such as data analyst dataarchitect data engineer.

and so on after this we.

will cover the data life cyclewhere we will discuss.

how data is extracted processedand finally use as a solution..

"Once were done with that.",

"Well cover the basicsof machine learning",

"where well see whatexactly machine learning is",

and the different typesof machine learning next..

We will move ontothe K means algorithm.

"and well discuss a use caseof the k-means clustering",

after which we Discussthe various steps involved.

in the k-means algorithm.

and then we will finally move onto the Hands-On part.

where we use the k-meansalgorithm to Cluster movies.

based on their popularityon social media platforms,.

"like Facebook at the endof todays session",

will also discuss about whata data science certification is.

and why you should take it up..

"So guys, theres a lot to coverin todays session.",

"Lets jump into the first topic.",

Do you guys remember the timeswhen we have telephones and we.

had to go to PC your bootsin order to make a phone call..

Call now those thingsare very simple.

"because we didnt generatea lot of data.",

"We didnt even storethe contacts and our phones",

or our telephones..

We used to memorize phonenumbers back then or you know,.

these have a diaryof all our contact.

but these dayswe have smartphones.

with store a lot of data..

"So theres everythingabout us in our mobile phones.",

We have images we have contacts..

We have various apps..

We have games..

Everything is storedon a mobile phones these days.

similarly the PCS that we usein the earlier times..

It used to processvery little data..

All right, there was A lotof data processing needed.

because technology wasan evolved that much..

So if you guys rememberwe use floppy disk.

back then and floppy..

This was used to storesmall amounts of data,.

but later on hard diskswere created and those.

used to store GBS of data..

"But now if you lookaround theres data",

everywhere around us..

All right, we have a datastored in the cloud..

We have data in each and everyAppliance at our houses..

Similarly..

"If you look at smart carsthese days theyre connected",

to the internet they connectedto a mobile phones.

and this also generatesa lot of data..

"What we dont realize is",

that evolution of technologyhas generated a lot of data..

All right..

Now initially therewas very little data.

and most of it was evenstructured only a small part.

of the data was unstructuredor semi-structured..

And in those days you could useSimple bi Tools in order.

to process all of this dataand make sense out of it..

But now we have way.

too much data and orderto process this much data..

We need more complex algorithms..

We need a better process..

All right, and this is.

where data sciencecomes in now guys,.

"Im not going to getinto the depth of data science.",

"Yet Im sure allof you have heard of iot",

or Internet of things..

Now..

Did you guys know.

that we produce2.5 quintillion bytes.

of data each day..

And this is only acceleratingwith the growth of iot..

Now iot or Internetof Things is just a fancy term.

that we use for networkof tools or devices.

that communicate and transferdata through the internet..

So various devicesare connected to each other.

through the internet.

and they communicatewith each other right.

now the communication happensby exchange of data or by..

Generation of data now thesedevices include the vehicles..

We drive the include our TVsof coffee machines.

refrigerators washing machinesand almost everything else.

that we use in a daily basis..

Now, these interconnecteddevices produce an unimaginable.

amount of data guys iot datais measured in zettabytes.

and one zettabyte is equalto trillion gigabytes..

So according to a recentsurvey by Cisco..

"Its estimated that bythe end of 2019,",

which is almost here..

The iot will generate morethan five hundred zettabytes.

of data per year..

And this number will onlyincrease through time..

"Its hard to imagine datain that much volume,",

imagine processing analyzingand managing this much of data..

"Its only goingto cause as a migraine",

so guys having to dealwith this much data.

is not something thattraditional bi tools can do..

Okay..

We no longer can rely.

on traditional dataprocessing methods..

"Thats exactly whywe need data science.",

"Its our only hope right",

"now now lets not getinto the details here.",

Yet moving on..

"Lets see how socialmedia is adding on",

to the generation of data..

Now the fact.

that we are all in lovewith social media..

"Its actually generatinga lot of data for us.",

Okay..

"Its certainly one of the fuels",

for data creation Nowall these numbers.

that you see on the screenare generated every minute.

of the day..

Okay, and this numberis just going to increase so.

for Instagram it says.

that approximately1.7 million pictures uploaded.

in a minute and similarlyon Twitter approximately..

A hundred and forty eightthousand tweets are published.

every minute of the day..

So guys imagine in one are.

how much that would beand then imagine in 24 hours..

So guys, this isthe amount of data.

that is generatedthrough social media..

"Its unimaginable.",

Imagine processing this much.

data analyzing it and thentrying to figure out, you know,.

the important insightsfrom this much data analyzing.

this much data is going to bevery hard with traditional tools.

or traditional methods..

"Thats why data sciencewas introduced data science",

is a simple process.

that will just extract theuseful information from data..

"All right, its justgoing to process",

and analyze the entire data.

"and then its justgoing to extract",

what is needed now guys apartfrom social media and iot,.

there are other factors as well.

which contribute todata generation these days.

all our transactionsare done online, right?.

We pay bills online..

We shop online..

We even buy homes online.

these days you can even sellyour pets on oil excuses..

Not only thatwhen we stream music.

and Watch videos on YouTube allof this is generating a lot.

of data not to forget..

"Weve also brought Health Careinto the internet wall.",

Now there are variouswatches like bit fit.

which basically transour heart rate.

and it generates data abouta health conditions education is.

also an online thing right now..

"Thats exactly what youare doing right now.",

So with the emergenceof the internet,.

we now perform allour activities online..

Okay, obviously, thisis helping us,.

but we are unaware ofhow much data we are generating.

what can be done with Allof this data and what.

if we could use the data.

that we generatedto our benefit?.

"Well, thats exactly",

what data sciencedoes data science is all.

about extracting the usefulinsights from data and using.

it to grow your business..

Now before we get intothe details of data science,.

"lets see how Walmart uses datascience to grow that business.",

"So guys Walmart isthe worlds biggest retailer",

with over 20,000 storesin just 28 countries..

Okay..

"Now, its currently buildingthe worlds biggest.",

Good Cloud,.

which will be able to processtwo point five petabytes.

of data every hour now..

The reason behindWalmart success is.

how the user customer data.

to get useful insights aboutcustomers shopping patterns..

Now the data analyst andthe data scientist at Walmart..

They know every detailabout their customers..

They know thatif a customer buys Pop-Tarts,.

they might also buy cookies,how do they know all of this?.

Like how do they generateinformation like this now.

the user data that they getfrom their customers..

Hours and the analyze it.

to see what a particularcustomer is looking for..

Now..

"Lets look at a few cases",

where Walmart actuallyanalyze the data.

and they figured outthe customer needs..

"So lets consider the Halloween",

and the cookie sales example nowduring Halloween sales Analyst.

at Walmart tooka look at the data..

Okay, and he found out.

that a specificcookie was popular.

across all Walmart stores..

So every Walmart store wasselling these cookies very well,.

but he found out.

that they would to storeswhich are not selling..

A DOT..

Okay..

So the situation was immediatelyinvestigated and it was found.

that there wasa simple stocking oversight..

Okay, because of whichthe cookies were not put.

on the shelves for sale..

So because this issuewas immediately identified.

they prevented any further loss.

of sales nowanother such example,.

is that true Associationrule mining Walmart found out.

that strawberry Pop-Tart salesincreased by seven times.

before a hurricane..

So a data analyst at Walmartidentified the association.

between ha Hurricane.

and strawberry pop tartsthrough data mining now guys..

"Dont ask me the relationshipbetween Pop-Tarts",

and Harry Caine,.

but for some reason wheneverthere was a hurricane.

approaching people really wantedto eat strawberry Pop-Tart..

So what Walmart did.

was they place allthe strawberry Pop-Tarts?.

I will check outbefore a hurricane would occur..

So this way the increase salesof the Pop-Tarts Now,.

"wheres this is a natural thing.",

"Im not making it up.",

You can look it upon the internet..

Not only that Walmartis analyzing the data generated.

by Social media to find outall the training product so.

through social media..

You can find out the likesand dislikes of a person right?.

So what Walmart did isthey are quite smart.

the user data generated.

by social media to find outwhat products are trending.

or what productsare liked by customers..

Okay an example.

of this is 1 mod analyzesocial media data to find out.

that Facebook users were crazyabout cake pops..

Okay, so Walmartimmediately took a decision.

and they introduced cake popsinto the Walmart stores..

So guys the only reasonWalmart is so successful is.

because the huge amount of data.

"that they get they dont seeit as a burden instead.",

They process this data analyze.

it and then you try to drawuseful insights from it..

Okay, so they invest a lotof money a lot of effort.

and a lot of timeand data analysis..

Okay, they spend a lot.

of time analyzing data in orderto find any hidden patterns..

So as soon as they find outhidden pattern or association.

between any two products,.

these are giving out offers.

or Started having discountor something along that line..

So basically Walmart uses data.

in a very effective mannerthe analyzer very, well..

They process the data very well.

and they find outthe useful insights.

that they need in order to get.

more customers or in orderto improve their business..

So guys, this was allabout how Walmart uses.

data science now,.

"lets move ahead and lookat what is data set now",

guys data science is all aboutuncovering findings from data..

"Its all about surfacingthe hidden insights",

that can help..

Ponies to makesmart business decisions..

So all these hidden insights.

or these hidden patterns canbe used to make better decisions.

in a business now an exampleof this is also Netflix..

So Netflix, basically analyzesthe movie viewing patterns.

of users to understand.

what drives user interest.

and to see what users wantto watch and then.

once they find outthey give people.

what they want..

So guys actually datahas a lot of power..

You should just knowhow to process this data.

and how to extractthe useful information..

From data..

Okay..

"Thats what datascience is all about.",

So guys a big questionover here is.

how do data scientists getuseful insights from data..

"So its all startswith data exploration.",

Whenever a data scientist comesacross any challenging question.

or any sortof challenging situation,.

they become detectives sothe investigative leads.

and they try to understandthe different patterns.

or the differentcharacteristics of the data..

Okay..

They try to getall the information.

that they can from the dataand then Then they use it.

for the bettermentof the organization.

or the business..

"Now, lets look atwho is a data scientist.",

So guys the data scientists.

has to be able to view datathrough a quantitative lengths..

So guys knowing math is oneof the very important skills.

of data scientists..

Okay..

So mathematics is importantbecause in order to find.

"a solution youre going to builda lot of predictive models",

and these predictive models aregoing to be based on hard math..

So you have to be ableto understand all.

the Underlying mechanicswith these models most.

of the predictive models most.

of the algorithmsrequire mathematics..

"Now, theres amajor misconception",

that data science isall about statistics..

"Now, Im not sayingthat statistics is an important.",

It is very important,.

"but its not the only typeof math that is utilized",

in data science..

There are actuallymany machine learning algorithms.

which are basedon linear algebra..

So guys overall you needto have a good understanding.

of math and apartfrom that data scientist..

"Elis technology,",

so data scientists have to bereally good with technology..

Okay..

So their main work isthey utilize all the technology.

so that they can analyze.

these enormous data sets andwork with complex algorithms..

So all of this requires tools,.

which are much moresophisticated than Excel.

"so theres data scientist needto be very efficient",

with coding languages.

and few of the core language.

has associated with data scienceinclude SQL python R & sass..

It is also importantfor a data scientist..

Be a tacticalbusiness consultant..

So guys business problems can beon a sword by data scientist.

since our data scientistswork so closely with data.

they know everythingabout the business..

If you have a businessand you give the entire data set.

of your businessstored data scientist,.

he know each and every aspectof your business..

Okay?.

"Thats how data scientists work.",

They get the entire data set..

They study the data setthe analyze it and then we see.

where things are going wrong.

or what needs to bedone more or what?.

Needs to be excluded..

So guys having this businessAcumen is just as important.

as having skills.

in algorithms or being goodwith math and technology..

So guys business isalso as important as.

these other fields now,.

you know whoour data scientist is..

"Lets look at the skill setsthat a data scientist names.",

Okay, it always starts.

with Statistics statisticswill give you the numbers.

from the data..

So a good understandingof Statistics is very important.

for becoming a data scientist..

You have to be familiarwith satisfaction..

Contest distributions maximumlikelihood estimators and all.

of that apart.

from that you should alsohave a good understanding.

of probability Theory.

and descriptive statistics..

These Concepts will help youmake Better Business decisions..

So no matter what type.

"of company or roleyoure interviewing for.",

"Youre going to beexpected to know",

how to use the toolsof the trade..

Okay..

This means that you have.

to know a statisticalprogramming language like our.

"or Python and also youll needto know or database.",

Wiring language like SQL now.

the main reason whypeople prefer our.

and python is because ofthe number of packages.

that these languages have.

and these predefinedpackages have most.

of the algorithms in them..

"So you dont haveto actually sit down",

and code the algorithms instead..

You can just load oneof these packages.

from their libraries and run it..

So programming languagesis a must at the minimum..

You should know our.

or python and a databasequery language now,.

"lets move on to dataextraction and processing.",

So guys That you have.

multiple data sources likemySQL database Mongo database..

Okay..

So what you have to dois you have to extract.

from such sourcesand then in order to analyze.

and query this database you haveto store it in a proper format.

or a proper structure..

Okay, finally, then you can loadthe data in the data warehouse.

and you can analyzethe data over here..

Okay..

So this entire process is calledextraction and processing..

So guys extraction.

and processing is allabout getting data..

From these differentdata sources and then.

putting it in a format.

so that you can analyze itnow next is data wrangling.

and exploration nowguys data wrangling is one.

of the most difficult tasksin data science..

This is the mosttime-consuming task.

because data wrangling is allabout cleaning the data..

There are a lot of instances.

where the data setshave missing values.

or they have null values.

or they have inconsistentformats or inconsistent values.

and you need to understandwhat to do with such values..

This is Data wrangling.

or data cleaning comesinto the picture then.

"after youre done with that.",

You are goingto analyze the data..

"So wheres after data wranglingand cleaning is done.",

"Youre going to start exploring.",

This is where you try to makesense out of the data..

Okay, so you can do thisby looking at the different.

patterns in the datathe different Trends outliers.

and various unexpected resultsin all of that..

Next..

We have machine learning..

"So guys if yourea large company",

"or with huge amounts of data orif youre working at a company.",

See where the productis data driven,.

"like if youre workingin Netflix or Google Maps,",

then you have to be.

familiar with machinelearning methods, right?.

You cannot processlarge amount of data.

with traditional methods..

"So thats why you needa machine learning algorithms.",

So there are few algorithms..

Like knok nearest neighbordoes random Forest.

this K means algorithmthis support Vector machines,.

all of these algorithms..

You have to be aware of all.

of these algorithmsand let me tell you.

that most of these algorithmscan be implemented..

Using our or python libraries..

Okay, you need tohave an understanding.

of machine learning..

If you have large amountof data in front of you.

which is going to be the casefor most of the people right now.

because data is being generated.

at an Unstoppable Pace earlierin the session we discussed.

how much of data is generated..

So for now knowingmachine learning algorithms.

and machine learning Conceptsis a very required skill.

if you want to becomea data scientist,.

"so if youre sittingfor an interview as",

a data scientist,.

you will be askedmachine learning..

Seems you will be asked.

how good you arewith these algorithms.

and how well youcan Implement them..

Next we have bigdata processing Frameworks..

So guys, we know.

"that weve been generatinga lot of data and most",

of this data can be structuredor unstructured as well..

So on such data,.

you cannot use traditionaldata processing system..

"So thats why you need",

to know Frameworkslike Hadoop and Spark..

Okay..

These Frameworks can be usedto handle big data lastly..

We have data visualization..

So guys data visualization is Isone of the most important part.

of data analysis,.

it is always very importantto present the data.

in an understandableand Visually appealing format..

So data visualizationis one of the skills.

that data scientistshave to master..

Okay, if you want to communicatethe data with the end users.

in a better way thendata visualization is a must.

so guys are a lot of tools.

which can be used for datavisualization tools like Diablo.

and power bi are few the mostpopular visualization tools..

So with this we sum upthe entire skill set.

that is needed to becomea data scientist apart from this.

you should also have data-drivenproblem solving approach..

You should also bevery creative with data..

So now that we know the skills.

that are needed to becomea data scientist..

"Lets look at the differentjob roles just data science is",

a very vast field..

There are many job rolesunder data science..

"So lets take a lookat each role.",

"Lets start offwith a data scientist.",

"So theres data scientistshave to understand.",

The challenge is over business.

and they have to offer the bestsolution using data analysis.

and data processing..

So for instance.

if they are expectedto perform predictive analysis,.

they should also be ableto identify Trends and patterns.

that can have the companiesin making better decisions.

to become a data scientist..

You have to be an expert inour Matlab SQL Python and other.

complementary Technologies..

It can also help.

if you have a higherdegree in mathematics.

or computer engineeringnext we have data..

An analyst so a dataanalyst is responsible.

for a variety of tasks,.

including visualizationprocessing of massive amount.

of data and among them..

They have to also performqueries on databases..

So they should be awareof the different query languages.

and guys one of the most.

important skills ofa data analyst is optimization..

This is because they haveto create and modify algorithms.

that can be used to pullinformation from some.

of the biggest databaseswithout corrupting the data.

so to become Be done..

You must know Technologiessuch as SQL our SAS and python..

So certification in any.

of these Technologiescan boost your job application..

You should also havea good problem solving quality..

Next..

We have a data architect..

So a data architectcreates the blueprints.

for a data management.

so that the databasescan be easily integrated.

centralized and protectedwith a best security measures..

Okay..

They also ensure.

that the data Engineershave the best tools.

and systems to work with Soto become a data architect,.

you have to have expertiseand data warehousing.

data modeling extractiontransformation and loan..

Okay..

You should also bewell versed in Hive Pig.

and Spark now apart from thisthere are data Engineers..

So guys,.

the main responsibilities ofa data engineer is to build.

and test scalableBig Data ecosystems..

Okay, they are also neededto update the existing systems.

with newer or upgraded versions.

and they are also responsiblefor improving the efficiency..

For database now..

If you are interestedin a career as a data engineer,.

then technologies.

that require hands-onexperience include Hive nosql.

are Ruby Java C++ and Matlab,.

it would also help.

if you can workwith popular data apis.

and ETL tools next..

We have a statistician..

So as the name suggests you haveto have a sound understanding.

of statistical theoriesand data organization..

Not only do they extractand offer valuable insights..

They also create new..

Methodologies for engineersto apply now..

If you want to becomea statistician then you have.

to have a passion for logic..

They are also good variety.

of database systemssuch as SQL Data Mining.

and other various machinelearning Technologies by that..

I mean, you should be goodwith math and you should also.

have a good knowledge.

about the weight isdatabase system such as SQL.

and also the variousmachine learning Concepts.

and algorithms isthe most next we have.

the database administrator..

So guys the job profile ofa database administrator.

is Much self-explanatory,.

they are basically responsiblefor the proper functioning.

of all the databases.

and they are also responsiblefor granting permission.

or the working in services tothe employees of the company..

They also have to take careof the database backups.

and recoveries..

So some of the skills.

that are needed to becomea database administrator include.

database backup and Recoverydata security data modeling.

and design next..

We have the business analyst nowthe role of a business analyst.

is a little It different.

from all of the otherdata signs job now..

"Dont get me wrong.",

They have a very.

good understanding of the dataoriented Technologies..

They know how to handle a lotof data and process it.

but they are also very focusedon how this data can be linked.

to actionable business inside..

So they mainly focuson business growth..

Okay..

Now a business analystacts like a link.

between the data engineersand the management Executives..

So in order to become.

a business analyst you haveto have an understanding.

of business financesbusiness intelligence..

And also I did acknowledge,.

"hes like data modeling datavisualization tools and Etc",

at last we have a data.

and analytics managera data and analytics.

manager is responsiblefor the data science operations..

Now the main responsibilitiesof a data and analytics.

manager is to overseethe data science operation..

"Okay, hes responsiblefor assigning the duties",

to the team accordingto their skills.

and expertise now their strengthshould include Technologies.

like SAS our SQL..

And of course,.

they should have good managementskills apart from that..

They must have excellent socialskills leadership qualities.

and and out-of-the-boxthinking attitude..

And like I said earlier.

you need to have a goodunderstanding of Technologies..

Like pythons asour Java and Etc..

So Guys, these werethe different job roles.

in data science..

I hope you all foundthis informative..

"Now, lets move aheadand look at the data lifecycle.",

So guys are basically six stepsin the data life cycle..

It starts witha business requirement..

Next is the data acquisition.

after that youwould process the data.

which is called data processing..

Then there isdata exploration modeling.

and finally deployment..

So guys before you even starton a data science project..

It is important.

"that you understand the problemyoure trying to solve.",

So in this stage,.

"youre just going to focus",

on identifying the centralobjectives of the project.

and you will do thisby identifying the variables.

that need to bepredicted next up..

We have data acquisition..

Okay..

"So now that you haveyour objectives I find its time",

for you to startGathering the data..

So data mining is the process.

of gathering your datafrom different sources.

at this stage some.

of the questions youcan ask yourself is.

what data do I needfor my project?.

Where does it live?.

How can I obtain it?.

And what is the mostefficient way to store.

and access all of it?.

Next up there is data processingnow usually all the data.

that you collectedis a huge mess..

Okay..

"Its not formatted.",

"Its not structured.",

"Its not cleaned.",

So if Find any data set.

"that is cleanedand its packaged well for you,",

"then youve actuallywon the lottery",

because finding the right datatakes a lot of time.

and it takes a lot of effort.

and one of the majortime-consuming task.

in the data science processis data cleaning..

Okay, this requiresa lot of time..

It requires a lot of effort.

because you have to gothrough the entire data set.

to find out any missing values.

or if there areany inconsistent values.

or corrupted data,.

and you also findthe unnecessary data..

Over here and youremove that data..

So this was all.

about data processing nextwe have data exploration..

So now that you have sparklingclean set of data,.

you are finally ready to getstarted with your analysis..

Okay, the data exploration stageis basically the brainstorming.

of data analysis..

So in order to understandthe patterns in your data,.

you can use histogram..

You can just pull upa random subset of data.

and plot a histogram..

You can even createinteractive visualizations..

This is the point.

where you Dive deepinto the data.

and you try to explorethe different models.

that can be appliedto your data next up..

We have data modeling..

So after processing the data,.

"what youre going to dois youre going to carry",

out model training..

Okay..

Now model training is basicallyabout finding a model.

that answers thequestions more accurately..

So the process of model traininginvolves a lot of steps..

"So firstly youll startby splitting the input data",

into the training data setand the testing data set..

"Okay, youre going to takethe entire data set",

"and youre going to separate itinto Two two parts one is",

the training and oneis the testing data.

after that your build a model.

"by using the training data setand once youre done with that,",

"youll evaluate the training",

and the test data set nowto evaluate the training.

and testing data..

"So youll be using series",

of machine learningalgorithms after that..

"Youll find out the model",

which is the most suitablefor your business requirement..

So this wasmainly data modeling..

Okay..

This is where you build a modelout of your training data set.

and then you evaluate this modelby using the testing data set..

You have deployment..

So guys a goal of this stageis to deploy the model.

into a production or maybea production like environment..

So this is basically donefor final user acceptance.

and the users have to validatethe performance of the models.

and if there are any issues.

with the model or any issueswith the algorithm,.

then they have to befixed in this stage..

So guys with thiswe come to the end.

of the data lifecycle..

I hope this was clear statisticsand probability are essential.

because these disciplesform the basic Foundation.

of all machine.

learning algorithms deeplearning artificial intelligence.

and data science..

In fact, mathematicsand probability is.

behind everything around usfrom shapes patterns.

and colors to the count.

of petals in a flowermathematics is embedded.

in each and every aspectof our lives with this in mind..

"I welcome you allto todays session.",

"So Im going to go aheadand Scoffs the agenda for today",

with you all now going to beginthe session by understanding.

what is data after that..

"Well move on and look at thedifferent categories of data,",

like quantitativeand qualitative data,.

"then well discuss whatexactly statistics is",

the basic terminologies instatistics and a couple.

of sampling techniques..

"Once were done with that.",

"Well discuss the differenttypes of Statistics",

which involve descriptiveand inferential statistics..

Then in the next sessionwill mainly be focusing.

on descriptive statisticshere will understand.

the different measuresof center measures.

of spread Information Gain.

and entropy will alsounderstand all of these measures.

"with the help of a use caseand finally well discuss",

what exactly aconfusion Matrix is.

"once weve covered the entiredescriptive statistics module",

will discuss the probabilitymodule here will understand what.

exactly probability isthe different terminologies.

in probability will also.

study the Differentprobability distributions,.

"then well discuss the typesof probability which include",

marginal probability jointand conditional probability..

Then we move onand discuss a use case.

"where and well seeexamples that show us",

how the different typesof probability work.

and to betterunderstand Bayes theorem..

We look at a small example..

Also, I forgot to mention.

that at the end of thedescriptive statistics module.

will be running a small demoin the our language..

So for those of you.

"who dont know muchabout our Ill be explaining",

every line in depth,.

but if you want to havea more in-depth understanding.

"about our Ill leavea couple of blocks.",

And a couple of videosin the description box.

you all can definitelycheck out that content..

"Now after weve completed theprobability module will discuss",

the inferential statisticsmodule will start this module.

by understanding.

what is point estimation..

We will discusswhat is confidence interval.

and how you can estimatethe confidence interval..

We will also discuss marginof error and will understand all.

of these concepts by lookingat a small use case..

"Wed finally end the inferentialReal statistic module by looking",

at what hypothesistesting is hypothesis..

Testing is a very important partof inferential statistics..

"So well end the sessionby looking at a use case",

that discusses howhypothesis testing works.

and to sum everything up..

"Well look at a demo",

that explains howinferential statistics Works..

"Alright, so guys,theres a lot to cover today.",

"So lets move ahead",

and take a lookat our first topic.

which is what is data..

Now, this isa quite simple question.

if I ask any of Youwhat is data?.

"Youll see that itsa set of numbers",

or some sort of documents.

that have stored in my computernow data is actually everything..

All right, look around you thereis data everywhere each click.

on your phone generatesmore data than you know,.

now this generated dataprovides insights for analysis.

and helps us makeBetter Business decisions..

This is why data isso important to give you.

a formal definition data refersto facts and statistics..

Collected togetherfor reference or analysis..

All right..

This is the definitionof data in terms.

of statistics and probability..

So as we know datacan be collected it.

can be measured and analyzed.

it can be visualized byusing statistical models.

and graphs now data is dividedinto two major subcategories..

Alright, so first wehave qualitative data.

and quantitative data..

These are the twodifferent types of data.

under qualitative data..

We have nominal and ordinal dataand under quantitative data..

We have discreteand continuous data..

"Now, lets focuson qualitative data.",

Now this type of data deals withcharacteristics and descriptors.

"that cant be easily measured",

but can be observed subjectively.

now qualitative datais further divided.

into nominal and ordinal data..

So nominal data isany sort of data.

"that doesnt haveany order or ranking?",

Okay..

An example of nominaldata is gender..

Now..

There is no ranking in gender..

"Theres only male femaleor other right?",

There is no one two,.

three four or any sortof ordering in gender race is.

another example of nominal data..

Now ordinal data is basically anordered series of information..

"Okay, lets saythat you went to a restaurant.",

Okay..

Your information is storedin the form of customer ID..

All right..

So basically you are representedwith a customer ID..

Now you would have ratedtheir service as.

either good or average..

"All right, thatshow no ordinal data is",

"and similarly theyll havea record of other customers",

who visit the restaurantalong with their ratings..

All right..

So any data which hassome sort of sequence.

or some sort of orderto it is known as ordinal data..

All right, so guys,.

this is pretty simpleto understand now,.

"lets move on and lookat quantitative data.",

"So quantitative databasically these Hes",

with numbers and things..

Okay, you can understand.

that by the word quantitativeitself quantitative is.

basically quantity..

Right Saudis will numbers.

a deals with anything that youcan measure objectively..

All right, sothere are two types.

of quantitative data there isdiscrete and continuous data.

now discrete data is alsoknown as categorical data.

and it can hold a finite numberof possible values..

Now, the number of studentsin a class is a finite Number..

"All right, you canthave infinite number",

of students in a class..

"Lets say in your fifth grade.",

They have a hundred studentsin your class..

"All right, there werentinfinite number but there",

was a definite finite numberof students in your class..

"Okay, thats discrete data.",

Next..

We have continuous data..

Now this type of datacan hold infinite number.

of possible values..

Okay..

So when you say weightof a person is an example.

of continuous data.

what I mean to see is my weightcan be 50 kgs or it NB 50.1 kgs.

or it can be 50.00 one kgs.

or 50.000 one or is50.0 2 3 and so.

on right thereare infinite number.

of possible values, right?.

So this is what I meanby a continuous data..

All right..

This is the difference betweendiscrete and continuous data..

"And also Id like to mentiona few other things over here.",

Now, there are a coupleof types of variables as well..

We have a discrete variable.

and we have a continuousvariable discrete variable.

is also known asa categorical variable.

or and it can hold valuesof different categories..

"Lets say that you havea variable called message",

and there are two typesof values that this variable.

"can hold lets say",

that your messagecan either be a Spam message.

or a non spam message..

"Okay, thats when you calla variable as discrete",

or categorical variable..

All right, because itcan hold values.

that represent differentcategories of data.

now continuous variablesare basically variables.

that can store infinitenumber of values..

So the weight of a personcan be denoted as.

a continuous variable..

"All right, lets say there is a variable called weight",

and it can store infinite number of possible values..

"Thats why we will call it a continuous variable.",

So guys basically variable is anything.

that can store a value right?.

So if you associate any sort of data with a Able,.

then it will become either discrete variable.

or continuous variable..

There is also dependent and independent type of variables..

"Now, we wont discuss all of that in death because",

"thats pretty understandable.",

"Im sure all of you know,",

what is independent variable and dependent variable right?.

...]