

Data Analytics with Python
Prof. Ramesh Anbanandam
Department of management studies
Indian Institute of Technology, Roorkee

Lecture No 1
Introduction to Data Analytics

Welcome students this course on data analytics with the Python today is the introduction class. This lecture is on introduction to data analytics.

(Refer Slide Time: 00:34)

Objective of the course

- The principle focus of this course is to introduce conceptual understanding using simple and practical examples rather than repetitive and point click mentality
- This course should make you comfortable using analytics in your career and your life
- You will know how to work with real data, and might have learned many different methodologies but choosing the right methodology is important

The objective of this course is to introduce the conceptual understanding using simple and practical examples rather than repetitive and point clique mentality, here most of the students generally they are how they are using the software for doing data analytics. Just they want to just click it they want to get the result, they do not want bother about exactly what is happening inside the software. This course should make you comfortable using analytics in her career and her life.

You will know how to work with a real data and you might have learnt the many different methodologies, but choosing the right methodology is important. This course will focus you will help you how to choose the right data analytics tools.

(Refer Slide Time: 01:17)

Objective of the course Contd...

- The danger in using quantitative method does not generally lie in the inability to perform the calculation
- The real threat is lack of fundamental understanding of:
 - Why to use a particular technique or procedure
 - How to use it correctly and,
 - How to correctly interpret the result



Objective of the course is when you look at this picture. How this person is using this tool there is a ladder. He was not knowing correctly how to use this ladder for the purpose it is intended. So the danger in using quantitative method does not generally lie in the inability to perform the calculation, because of the computer development in computer technology.

There are many packages are available for doing data analytics. But, the real threat is lack of fundamental understanding of why to use particular technique or procedures and how to use it correctly and, how to correctly interpret the result. This course will focus on how to choose the right technique and how to use it correctly and how to interpret the result.

(Refer Slide Time: 02:01)

Learning objectives

1. Define data and its importance
2. Define data analytics and its types
3. Explain why analytics is important in today's business environment
4. Explain how statistics, analytics and data science are interrelated
5. Why python?
6. Explain the four different levels of Data:
 - Nominal
 - Ordinal
 - Interval and
 - Ratio

So what was the learning objective of this class that is; after completing this lecture what you will learn one is you can define what is data and its importance. You can define what are data analytics and types. You can explain why analytics is in today's business environment is so important. Then we can see how statistics, analytics and data science are interrelated, there seems to be some overlap in this we will clarify that what is the difference, how these are overlapped how these are interrelated.

In this course we are going to use a package called Python. I will explain how why it is important to use the Python in this course, at the end of this session we will explain the four important levels of data that is nominal, ordinal, interval and ratio. Now we will go to the content;

(Refer Slide Time: 02:54)

1. Define Data and its importance

- Variable, Measurement and Data
- What is generating so much data?
- How data add value to the business?
- Why data is important?



We will define data and its importance. There are three term one is variable, measurement and data. Next we will see what is generating so much data. Next we will see how data add value to the business, and then we will say why data is important.

(Refer Slide Time: 03:11)

1.1 Variable, Measurement and Data

- Variables – is a characteristic of any entity being studied that is capable of taking on different values
- Measurements – is when a standard process is used to assign numbers to particular attributes or characteristic of a variable
- Data – data are recorded measurements

o

See the variable, measurement and data these are the terms which we are going to use frequently in this course. So what is a variable? Variable is a characteristics of any entity being studied that is capable of taking on different values. Say for example, X is the variable it can take any values it may be 1 it may be 2 or it may be 0 and so on. The measurement is when you standard processes used to assign numbers to your particular attributes or characteristics of variable are called a measurement.

For that X, you want to substitute some values for that value. You have to measure the characteristics of the variable that is nothing but your measurement. So then, what is the data? Data are recorded measurement. So there is a variable you measure the phenomena after measuring the phenomena you are substituting some value for the variable so that the variable will take a particular value that value is nothing but your data.

So X is the variable for example number 5 is the data. How you are measuring that 5, that is called measurement. Then what is generating so much of data.

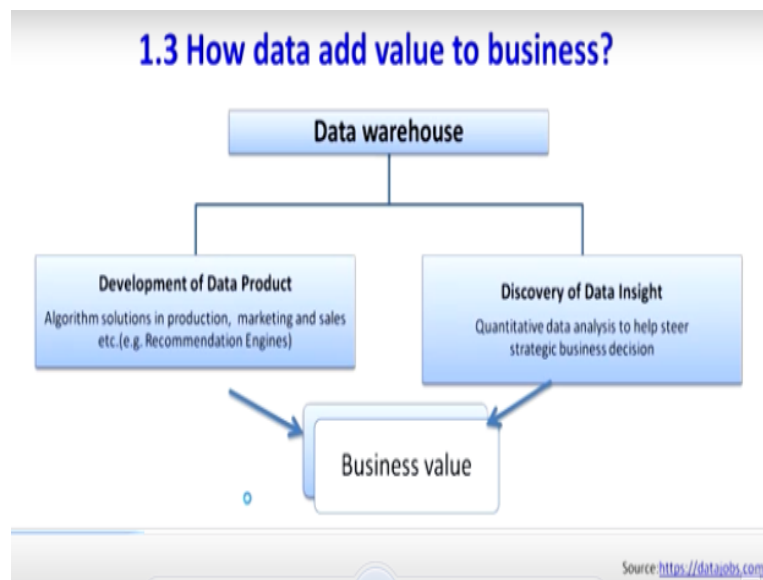
(Refer Slide Time: 04:33)

1.2 What is generating so much data?

- Data can be generated by
 - Humans,
 - Machines or
 - Humans-machines combines
- It can be generated anywhere where any information is generated and stored in structured or unstructured formats

Data can be generated different way humans, machines, and human - machines combines. The humans, machines and human - machines combines in the sense, now seen everybody is having the various Facebook account, we have LinkedIn account, we are in various social network sites. Now the availability of the data is not the problem. It can be generated anywhere where the information is generated and stored and structured or unstructured format.

(Refer Slide Time: 05:06)



So how the data add value to the business? So the data after getting from various sources assume that it is a store in the form of data warehouse. So from the data warehouse the data can be used for development of a data product. Here we are using the word, data product and the coming slides I will explain exactly what the data product with some examples is. So the

same data, if we look at the right hand side that can be used to get more insights from the data.

Okay, what do you mean the data product? For example, algorithm solutions in production, marketing and sales, example of some data product. For example, recommendation engine one of the example for data product. Suppose, if you go for Flipkart or Amazon for buying a particular product in that package, that software itself, we will recommend to you what is the next product, possible product that you can buy. That is nothing but the recommendation engine.

Even if you watch some YouTube videos on particular topic, that YouTube itself will suggest to you what are the relevant videos are available. So that is a recommendation engine. That is one of the examples of your data product; with the help of data so that will help you to forming a data product or you can get an insight from the data. That will add your business value to you.

(Refer Slide Time: 06:27)

Data Products



See this is an example of your data products, this is the driverless car. Google car, so the whole concept of Google car is with the help of data. It is detecting all other requirements for driving the cars. The next example is for recommendation engine, as I told you previously when you buy any product they will suggest you that along with this product, the other product also can be purchased.

Another very common example for a data product is Google. The Google is lot of applications there one of the application of example for data product is Google mapping. So the Google mapping is helping you to find out what is the right route, which road there is a traffic, in which road there is a toll booth, so this kind of information we can get it from the Google map. So this Google map is the one of an example of your data production.

(Refer Slide Time: 07:20)

1.4 Why Data is important?

- Data helps in make better decisions
- Data helps in solve problems by finding the reason for underperformance
- Data helps one to evaluate the performance.
- Data helps one improve processes
- Data helps one understand consumers and the market

Now why data is important? The data helps in making better decisions, data helps in solve problem by finding the reason for underperformance. Suppose some company it is not performing properly by collecting the data we can identify what was the reason for this under performance. The data helps one to evaluate the performance. So what is the current performance, the data also can be used for benchmarking the performance of your business organization.

And after benchmarking data helps one improving the performance also, so data also can help one understand the consumers and the markets, especially the marketing context. You can understand who are the right consumers and what kind of preferences they are having in the market.

(Refer Slide Time: 08:16)

2. Define data analytic and its types

- Define data analytics
- Why analytics is important?
- Data analysis
- Data analytics vs. Data analysis
- Types of Data analytics

Next we will define what is a data analytics and its types? So in this coming two, three lecture or two, three slides we are going to discuss we will define what is data analytics? Then you say why analytics is important? Then we will see that data analysis? Then we will see how data analytics is different from data analysis? At the end will we see types of data analytics?

(Refer Slide Time: 08:40)

2.1. Define data analytics

- Analytics is defined as “the scientific process of transforming data into insights for making better decisions”
- Analytics, is the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions – James Evans
- Analysis = Analytics ?

We will define data analytics is the scientific process of transforming data into insights for making better decisions. See it is a scientific process for transforming the data into for making better decisions, even without the data also even without doing analytics also you can make the decision but you cannot make the better decision without analytics. By the virtue of your experience on intuitions you can take the decisions that also sometimes may be correct.

But about the help of data if you are making the decision then that will enable you to make the better decisions. Another professor James Evans, he has defined the data analytics in this way. it is the use of the data information technology, statistical analysis, quantitative methods and mathematical or computer-based models to help managers gain improved insight about their business operations and make better, fact-based decisions.

You see that there are many terms which are appearing here, one is IT, next one is a statistical analysis, and next one is the quantitative methods, then mathematical knowledge and computer-based models. So when we will see how these are interrelated in coming slides. Generally among the students, there is a confusion whether the analysis and analytics is same or different?

(Refer Slide Time: 10:13)

2.2 Why analytics is important?

- Opportunity abounds for the use of analytics and big data such as:
 1. Determining credit risk
 2. Developing new medicines
 3. Finding more efficient ways to deliver products and services
 4. Preventing fraud
 5. Uncovering cyber threats
 6. Retaining the most valuable customers



Why analytics is important. The opportunity abounds for the use of analytics and big data such as: for determining the credit risk, for developing new medicines, especially in healthcare. The healthcare analytics is an emerging, that is helping you to identify what is the correct medicines finding more efficient ways to deliver product and services. For example: in the banking context data analytics is used for preventing the fraud, and it is uncovering the cyber threats.

With the help of data analytics you can find out the possible cyber crimes and we can detect it we can prevent it. And data analytics are also important for retaining the most valuable customers. We can identify who is your valuable customer or non valuable customers. So we can focus on more on our valuable customers. Okay,

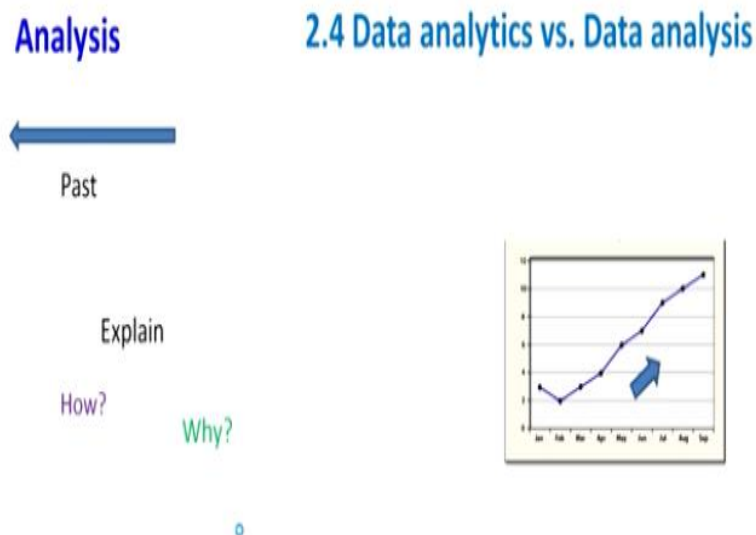
(Refer Slide Time: 11:08)

2.3 Data analysis

- Data analysis is the process of examining, transforming, and arranging raw data in a specific way to generate useful information from it
- Data analysis allows for the evaluation of data through analytical and logical reasoning to lead to some sort of outcome or conclusion in some context
- Data analysis is a multi-faceted process that involves a number of steps, approaches, and diverse techniques

Now what is the data analysis? Is the process of examining, transforming and arranging raw data in a specific way to generate useful information from it. So data analysis allows for the evaluation of data through analytical and logical reasoning to lead to some sort of outcome or conclusion in some context. Data analysis is a multi-faceted process that involves a number of steps approaches and diverse techniques. That we will see in coming lecture.

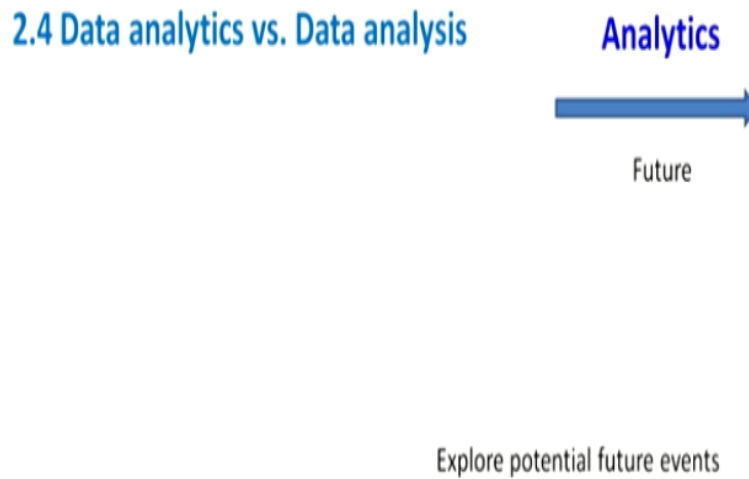
(Refer Slide Time: 11:41)



So now we will see what is the analysis is data analysis and data analytics. When you say analysis when you say data analysis it is something about what has happened in the past. So we will explain why that has happened? We will explain how it has happened?. We can explain why it has happened? For example, when we say data analysis that is nothing about

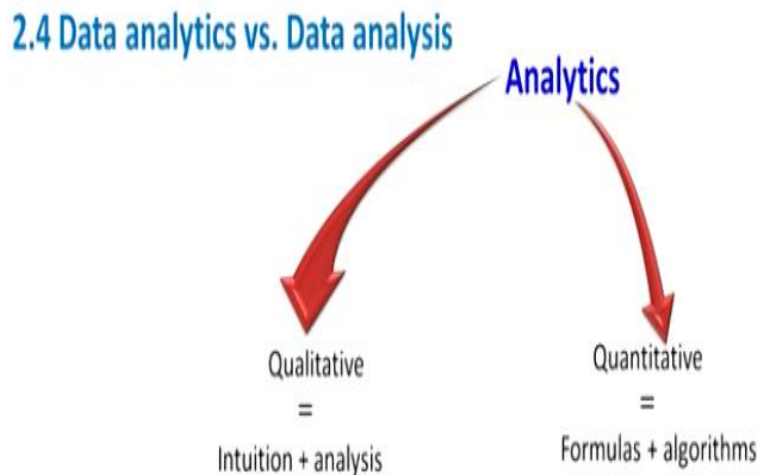
studying about what has happened it is like kind of a post-mortem analysis. What has happened in the past?

(Refer Slide Time: 12:13)



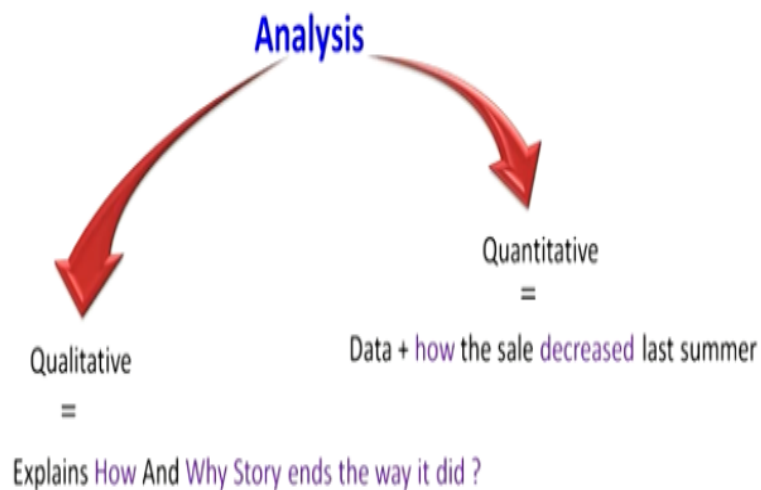
Okay, in the contrary the analytics is studying about what will happen in future and with the help of analytics. We can predict explore possible potential future events.

(Refer Slide Time: 12:25)



So the analytics is maybe qualitative or quantitative. For example in analytics if we say qualitative analytics. So it is the decision mostly based on the intuition. But if you say in quantitative where with the help of formulas with the help of algorithms will make the decisions.

(Refer Slide Time: 12:44)



So in the analysis data analysis also we can go for qualitative. We can explain how and why a story ends in that way it did? When we say in quantitative we can say, how the sales decreased the last summer. When I say as I am repeating and you say analysis is something studying about what has happened in the past.

(Refer Slide Time 13:12)

Analysis ≠ Analytics

Data Analysis ≠ Data analytics

Business Analysis ≠ Business analytics

Okay, so it is not exactly analysis equal to analytics. Similarly when you say data analysis is different data analytics is different.

Similarly business analyst is different business analytics. When you say analytics is nothing but studying about the future events with the help of the past data.

(Refer Slide Time 13:34)

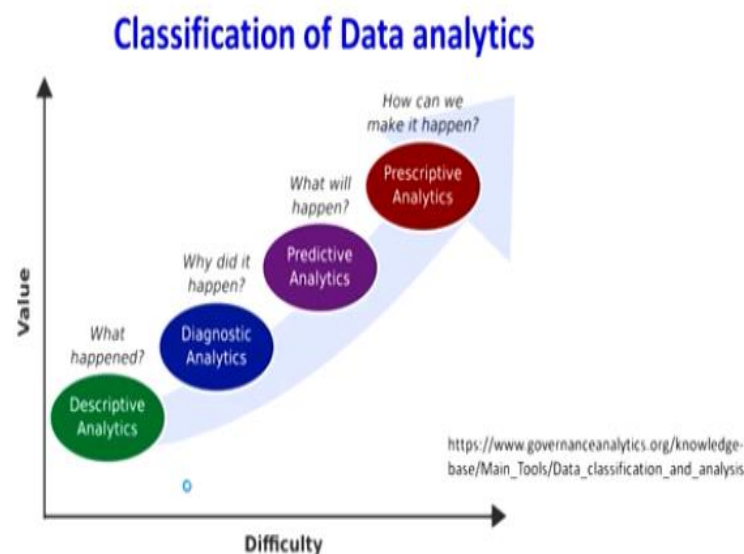
2.5 Classification of Data analytics

Based on the phase of workflow and the kind of analysis required, there are four major types of data analytics.

- Descriptive analytics
- Diagnostic analytics
- Predictive analytics
- Prescriptive analytics^o

Next we will go for classification of data analytics, based on the phase of workflow and the kind of analysis required, there are four major types of data analytics. One is descriptive analytics, diagnostic analytics, predictive analytics and prescriptive analytics. We will see these four types of analytics in detail in coming classes:

(Refer Slide Time 13:57)



If we look at the difficulty and the kind of value which we can get from different types of analytics; this picture shows for example: when you see the descriptive analytics that will answer what happened? Diagnostic analytics, will help you to answer why did it happen? Predictive analytics will help you what will happen? Prescriptive analytics will help you to answer how can we make it happen?. There is one context when you look at the level of difficulty you see that the descriptive analytics is the level of difficulty is very less.

And the contrary when you look at the prescriptive analytics the difficulty level is more and the value also, value in the sense business value which adds to you also more. so when there is a more difficulty there is a more value. Okay,

(Refer Slide Time 14:54)

Descriptive Analytics

- Descriptive Analytics, is the conventional form of Business Intelligence and data analysis
- It seeks to provide a depiction or “summary view” of facts and figures in an understandable format
- This either inform or prepare data for further analysis
- Descriptive analysis or statistics can summarize raw data and convert it into a form that can be easily understood by humans
- They can describe in detail about an event that has occurred in the past

Then we listen what is the descriptive analytics? Descriptive analytics is the conventional form of business intelligence or data analysis. It seeks to provide the depiction or summary view of facts and figures in an understandable format. These either inform or prepare data for further analysis. so descriptive analysis or we can say another way in statistics can summarize raw data and convert it into your form that can be easily understood by humans. They can describe in detail about an even that has occurred in the past. Okay,

(Refer Slide Time 15:40)

Example

A common example of Descriptive Analytics are company reports that simply provide a historic review like:

- Data Queries
- Reports
- Descriptive Statistics
- Data Visualization
- Data dashboard



Source: <https://www.linkedin.com/learning/478e9692-d13d-338f-907e-d76d0724d773>

Some of the examples of descriptive analytics is a common example of descriptive analytics are company reports that simply provide the historic review like: data queries, reports, data descriptive statistics, data visualization and data dashboard. Okay,

(Refer Slide Time 16:00)

Diagnostic analytics

- Diagnostic Analytics is a form of advanced analytics which examines data or content to answer the question "Why did it happen?"
- Diagnostic analytical tools aid an analyst to dig deeper into an issue so that they can arrive at the source of a problem
- In a structured business environment, tools for both descriptive and diagnostic analytics go parallel

The next one will go to the diagnostic analytics. Diagnostic analytics is a form of advanced analytics which examines data are content to answer the question why did it happen? So we are diagnosing, suppose we are meeting a doctor for consulting, so he will try to understand why this has happened? Okay so that kind of analytics nothing but diagnostic analytics. So the diagnostic analytical tools aid and analyst to dig deeper into an issue.

So that, they can arrive at the source of the problem. So doctor also will identify you somebody has got some disease what was the sources of the problem. Similarly the diagnostic analytics also if something has happened for example the company's not performing well that diagnostic abilities will help you to identify what was the core reason for that. In a structured business environment tools for both descriptive and diagnostic analytics go parallel.

So when you look at the whether it is a prescriptive or diagnostic analytics, the tools analytical tools which are using can be same only the purpose may be different.

(Refer Slide Time 17:09)

Example

- It uses techniques such as:
 1. Data Discovery
 2. Data Mining
 3. Correlations

For example: data discovery, data mining, and correlations. These tools can be used for your prescriptive analytics also. Okay,

(Refer Slide Time 17:20)

Predictive analytics

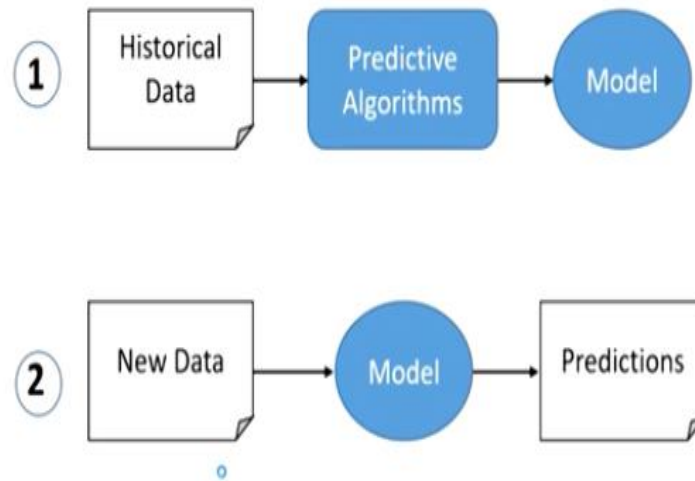
- Predictive analytics helps to forecast trends based on the current events
- Predicting the probability of an event happening in future or estimating the accurate time it will happen can all be determined with the help of predictive analytical models
- Many different but co-dependent variables are analysed to predict a trend in this type of analysis

Now we will go for predictive analytics, predictive analytics helps to forecast trends based on the current events. When you say predicting obviously it say that it is discussing about what will happen in future? Predicting the probability of an event happening in future are estimating accurate time it will happen can all be determined with the help of predictive analytical models. Many different but co-dependent variables are analysed to predict a trend in this type of analysis.

So in the predictive analytics one of the tool is the regression analysis. There may be some independent variables, some dependent variables, sometimes more dependent variable, more

than one dependent variable and how these variables are inter-related. So that kind of study is nothing but your predictive analytics.

(Refer Slide Time 18:11)



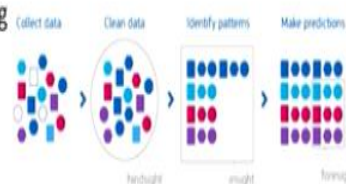
Source: <https://www.logianalytics.com/wp-content/uploads/2017/11/predictive-1.png>

When you look at this picture, you see that with the help of historical data by using different algorithm, predictive algorithms you can come with a model. Once the model is developed a new data can be fit into this model so we can get some predictions about the past events.

(Refer Slide Time 18:35)

Example

- Set of techniques that use model constructed from past data to predict the future or ascertain impact of one variable on another:
 1. Linear regression
 2. Time series analysis and forecasting
 3. Data mining



Source: <https://bigdata.madesimple.com/5-examples-predictive-analytics-travel-industry/>

Example is linear regression, time series analysis and forecasting and data mining. These are the techniques for predictive analytics.

(Refer Slide Time 18:46)

Prescriptive analytics

- Set of techniques to indicate the best course of action
- It tells what decision to make to optimize the outcome
- The goal of prescriptive analytics is to enable:
 1. Quality improvements
 2. Service enhancements
 3. Cost reductions and
 4. Increasing productivity

The last one is the prescriptive analytics. A set of techniques to indicate the best course of action. It tells what decision to make to optimize the outcome. The goal of prescriptive analytics is to enable: quality improvements, service enhancements, cost reductions and increasing productivity. Okay,

(Refer Slide Time 19:13)

Prescriptive analytics: Example

- Optimization Model
- Simulation
- Decision Analysis

In the prescriptive analytics, some of the tools which we can use is optimization models, simulation model, and decision analysis. These are the tools under prescriptive analytics.

(Refer Slide Time 19:27)

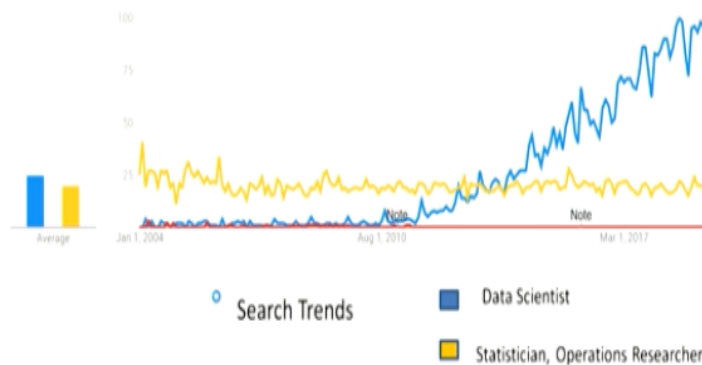
3. Explain why analytics is important

- Demand for Data Analytics
- Element of data Analytics

Next is we are going to see, why the analytics so important? In this section we will see what is happening the demand for data analytics and we look at the different elements of data analytics.

(Refer Slide Time 19:44)

3. Explain why analytics is important



This picture shows, Google Trends, this was up to 2017. See for example, the blue represents the data scientist; this orange represents the statistician operation researchers. You see the trend is it is increasing that means people are searching in the Google search engine the word data scientist more number of times. See the search count is increasing. That means there is a demand for that particular say job.

(Refer Slide Time 20:19)



You see, if you look at this is the newspaper clipping from Times of India. There are so many news are coming about data scientists and the future requirement of data scientists. You see the data scientist earning more than CA's and engineers. You can look at this link for further.

(Refer Slide Time 20:37)

3.1 Demand for Data Analytics

With companies across industries striving to bring their research and analysis (R&A) departments up to speed, the demand for qualified data scientists is rising.

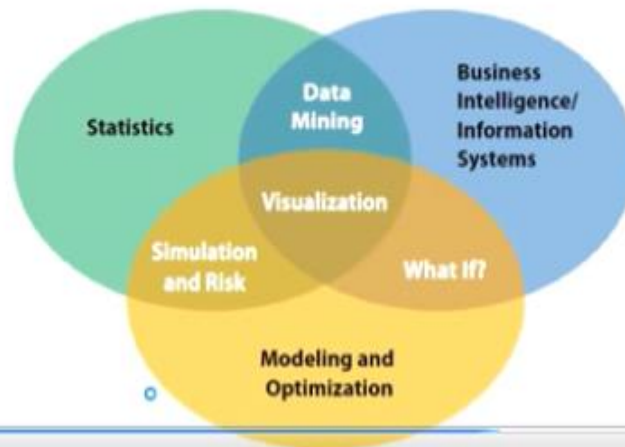
"India will face a demand-supply gap of 2,00,000 analytics professionals over the next three years. Even in the US, only 40 out of 100 positions for analytics professionals can be filled," said Rituparna Chakraborty, co-founder & senior VP of TeamLease Services.

http://timesofindia.indiatimes.com/articleshow/52171064.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst

And you see the demand for data analytics. This also newspaper clipping with companies across industries striving to bring their research and analysis department up to speed, the demand for qualified data scientist is rising. So there is an emerging field. so many companies is looking for the qualified data scientist. So if you take this course and end up the course that you may be qualified for getting into these companies.

(Refer Slide Time 21:07)

3.2 Element of data Analytics



Many times you see what is data analytics, Statistics, data mining, optimizations. These are students having different understanding on that. So when we say data analytics, there are different element one is statistics, next one is the business intelligence information systems, then modelling and optimizations, then simulation and risk. We can say if you are able to do what if analysis? That is nothing but sensitivity analysis, visualization data mining. These are the components of data analytics and how these different domains are interrelated?

(Refer Slide Time 21:47)

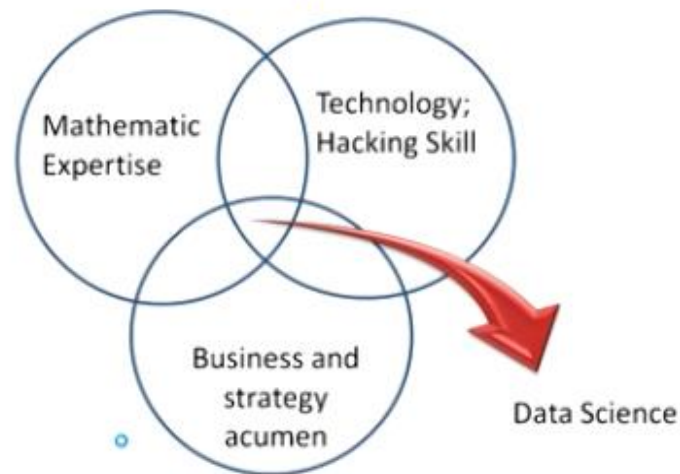
4. Data analyst and Data scientist

- The requisite skill set
- Difference between Data analyst and Data Scientist

Next we will see, what kind of skill set is required to become a data analyst? then we will see the small difference between data analyst and data scientist?

(Refer Slide Time 21:59)

4.1 The requisite skill set



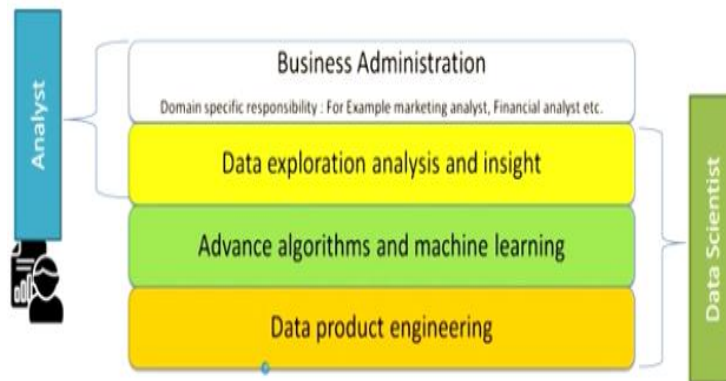
See to become a data analyst is the basic fundamental knowledge is you need to have knowledge of mathematics. Next you need to have the knowledge of technology is nothing but hacking skill. Hacking skills the sense, if the data is given hacking is done in the look at the positive way. How to use the data to get more information? The next skill is business and strategy acumen; you should have the knowledge of the domain knowledge of the business and you knew to the strategy equipment.

So these three skills are required for a good data scientist. It is very difficult to have a one person will have all these three skills that's why availability of good data analytics is becoming very difficult. Because somebody may be very good at mathematics but they may not have very good knowledge and business, some people may have very well at technology they may not technology in the sense information technology they may not have good knowledge on the business knowledge.

So we need to have the combination of all these three skills otherwise the group of people some people from mathematics department or mathematics area, some people from computer science, some people from the domain knowledge. They were to work together to form a good data scientist team, so these forms data science.

(Refer Slide Time 23:31)

4.2 Difference between Data analyst and Data Scientist



Source <https://datajobs.com/>

Now what is the difference between data analysts and data scientists and the difference is what kind of role they are doing? For example; the role of your data analyst is see in your business context he may have the knowledge of business domain. For example; if he is good at doing analytics in the area of marketing, he can be called as a marketing analyst. If the person is from finance area, he can be called it as a finance analyst. So he is the analyst, data analyst.

But the role of data scientist is little bigger, because the data scientist need to have the knowledge of advanced algorithms and machine learning and able to come out with a data product. Which I told you in the previously, so the data scientist can come out with a data product. Okay,

(Refer Slide Time 24:30)

5. Why python?



In this course we are going to use Python. In this in the next lecture, I will tell you the basic introduction about the Python. Here we will see why we are going to use the Python?. Because python is very simple and easy to learn. Most importantly it is a free software and open source. It uses interpreted, it is not the compiler. Suppose what do you my compiler and interpreter is you need a compiler to solve the whole program but interpreter need not be in that way.

It can solve even you can interpret one sentence also, one line in the programming line also. it is dynamically typed, dynamically type in the sense in some other programs every time you have to declare the variable. What is the nature of the variable? Whether it is integer? Whether it is a float? But here you need not do. It is dynamically takes the value. it is extensible, extensible in the sense if you make a code in some other language that can be extended with the help of Python.

And can be embedded, embedded in the sense you have made some program in Python it can be embedded with the some other platforms and it has extensive library.

(Refer Slide Time 25:45)

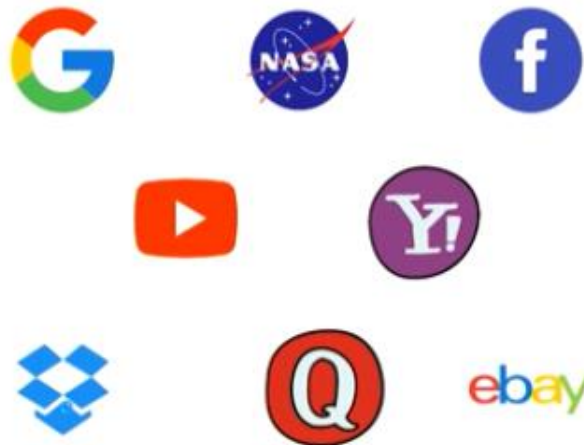
5. Why python?



The usability of Python is it is a desktop and web applications, it can be used for data applications, it can be used for networking applications, most importantly it can be used for data analyst, data science can be used for machine learning, it can be used for IoT Internet of Things and artificial intelligence applications and can be used for games.

(Refer Slide Time 26:05)

Companies using Python



Another reason for choosing Python is most of the companies, they use Python is a language in their company. Like for example; Google, Facebook, NASA, Yahoo and eBay. They use Python is a programming language.

(Refer Slide Time 26:23)

Why Jupyter Notebook?



In this Python also we are going to use Jupyter notebook. In the next class I will explain you because it is the client-server application is edit code on web browser. It is easy in documentation, easy in demonstration and user friendly interface.

(Refer Slide Time 26:39)

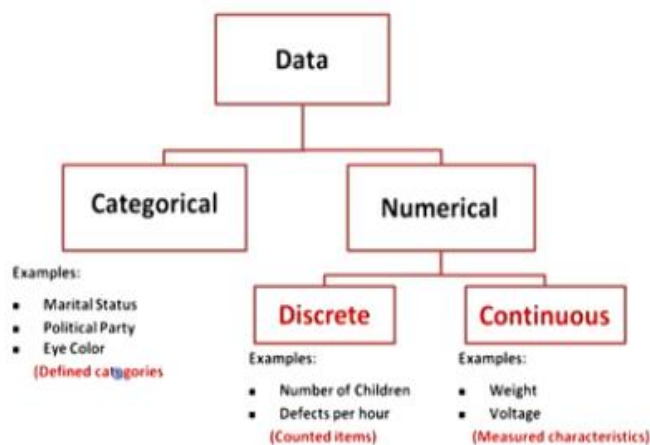
6. Explain the four different levels of Data

- Types of Variables
- Levels of Data Measurement
- Compare the four different levels of Data:
 - Nominal
 - Ordinal
 - Interval and
 - Ratio
- Usage Potential of Various Levels of Data
- Data Level, Operations, and Statistical Methods

This was the last session of this lecture; we will explain four different levels of the data. What is the type of variables?, Levels of data measurement? Compare for different level of data: will say nominal, ordinal, interval and ratio. We will see that why and what is the usage of knowing this different level of data?

(Refer Slide Time 27:03)

6.1 Types of Variables



The one way for classifying the data is the categorical data, one is a numerical data. In categorical data; you see marital status, political party, and eye color. These are categorical data. Numerical data; it can be discrete or continuous. Discrete data may be a number of children and defects per hour. So this is the discrete data. In the continuous data may be weight and voltage. These are the example of continuous data.

So what is the difference between discrete and continuous is, you say a number of children you may say two children or three children 2.5 children was not possible but in continuous, if you look at between 0 & 1 the numbers are continuing there are infinite number of values that are there between 0 & 1. So it is a continuous variable.

(Refer Slide Time 27:56)

6.2 Levels of Data Measurement

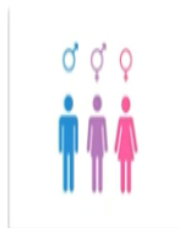
- Nominal – Lowest level of measurement
- Ordinal
- Interval
- Ratio – Highest level of measurement

Next will you see the different level of data measurement? Easily we have seen the classification of data. We classified as the categorical data and numerical data. There is another way of classification is, classifying into nominal data, ordinal data, interval data and ratio data.

(Refer Slide Time 28:14)

6.3.1 Nominal

- A **nominal scale** classifies data into distinct categories in which no ranking is implied
- Example : Gender, Marital Status



We will look at, what is a nominal data? Nominal scale classifies data into distinct categories in which no ranking is implied. The example of nominal data is gender, marital status. For

example; gender suppose you are conducting a questionnaire. Suppose you captured the gender male 0, female 1. This 0 1 represents just the gender. You cannot do any arithmetic operations with the help of the 0 & 1.

For example you cannot find out the average, software will give you some number but there is no meaning for that. Similarly marital status, whether it is married or unmarried. This is the example of nominal data.

(Refer Slide Time 29:01)

6.3.2 Ordinal scale

- An **ordinal scale** classifies data into distinct categories in which ranking is implied
- Example:
 - Product satisfaction → Satisfied, Neutral, Unsatisfied
 - Faculty rank → Professor, Associate Professor, Assistant Professor
 - Student Grades → A, B, C, D, F

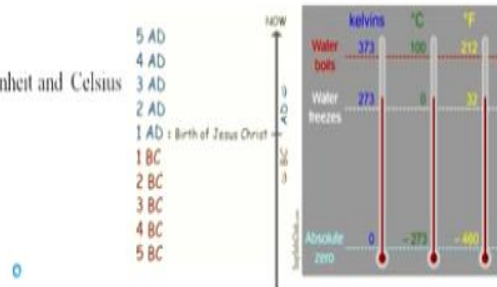
The next level of data is the ordinal scale. It classifies data into distinct categories in which the ranking is implied. Here the numbers are the ranked. For example; you may ask the customer to give a ranking about their level of satisfaction. For example, satisfied, neutral, unsatisfied. The faculty ranking, for example; professor, associate professor, assistant professor.

You see that their rank is followed for example 1 professor, 2 associate professor, 3 three professor. Student grades, A, B, C,D, E, F. These are ranking, because the numbers 1, 2, 3 represents the rank.

(Refer Slide Time 29:45)

6.3.3. Interval scale

- An **interval scale** is an ordered scale in which the difference between measurements is a meaningful quantity but the measurements do not have a true zero point.
- Example
 - Temperature in Fahrenheit and Celsius
 - Year



The next level of data is interval scale. The interval scale is ordered scale, in which the difference between measurements is a meaningful quantity but the measurement do not have to zero point. The example of interval scale is, for example year. Say now this here is 2019, you can add and subtract something. You can add another five years 2024 or you can subtract another nine years 2010.

But you cannot multiply, if you multiply that number for example 2019 and 2020 you will end up with the big number there is no meaning for that. Because, there is no meaning for zero. Another example of interval scale is your Fahrenheit temperature. For example in the Fahrenheit scale, the zero represents freezing point but it is not the absence of the seat but absence of the temperature but at the same time in the Kelvin for example minus 273 it is absence of heat. So Kelvin will be the some other scale. That you will see the next one,

(Refer Slide Time 30:52)

6.3.4 Ratio scale

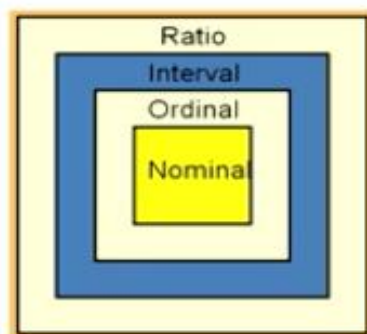
- A **ratio scale** is an ordered scale in which the difference between the measurements is a meaningful quantity and the measurements have a true zero point.
- Example
 - Weight
 - Age
 - Salary

The ratio scale is the ordered scale in which the difference between the measurements is a meaningful quantity and the measurements have the true zero point. Weight, age, salary and the Kelvin temperature comes under ratio scale. Because 0 Kelvin that means the absence of the heat. So in the ratio scale, he can do all kinds of arithmetic operation. For example the nominal, you cannot do any arithmetic operation. In ordinal you cannot do in arithmetic operation.

In the interval you can add and subtract but you cannot multiply. But in the ratio data, you can do all kinds of arithmetic operations. You can subtract, you can multiply, and you can divide.

(Refer Slide Time 31:35)

6.4 Usage Potential of Various Levels of Data



You see the usage potential various level of data. For example the usage potential of nominal data is not that much. The next one is ordinal; next one is interval, next one ratio. So the ratio data is having the highest to use its potential. The nominal data is having the least usage potential.

(Refer Slide Time 31:56)

6.5 Impact of choice of measurement scale

Data Level	Meaningful Operations	Statistical Methods
Nominal	Classifying and Counting	Nonparametric
Ordinal	All of the above plus Ranking	Nonparametric
Interval	All of the above plus Addition, Subtraction	Parametric
Ratio	All of the above plus multiplication and division	Parametric

This is more important, why we are to still know the different types of data. Because this types of data is helping to choose the right analytical tools for doing analysis. For example; if the data is the nominal data. You can do only nonparametric tests. For example the data is ordinal, here also you can do only nonparametric test. But if the data is interval, you can do parametric test. You see that interval; you can do all above plus addition and subtraction.

In the ratio, if you can do all of the above plus multiplication and division and statistical methods. You can go for parametric methods. So the purpose of classifying the data into nominal, ordinal, interval, ratio is to choose the right analytical tools with it whether it is a parametric or non parametric. The other reason is sometime for if we want to do a non parametric analysis that is used only for nominal data.

Sometime the students they will the data may be nominal but they may go for a parametric test that should not be done. That is the purpose of knowing what kind of what is the nature of this data. So in this class we have seen the introduction for data analytics. We have seen the importance of data analytics. We have seen the classification of data analytics. Then we can we have seen what is the analytics and analyst and we have seen different types of data.

The next class we will learn about what is Python? How to install the Python and what kind of descriptive analysis we can do with the help of Python? So the next class will meet you with another lecture. Thank you very much.