

Introduction to Learning Analytics
Prof. Ramkumar Rajendran
Interdisciplinary Programme in Educational Technology
Indian Institute of Technology, Bombay

Lecture - 10
Descriptive Analytics

In this learning dialogue, we will talk about Descriptive Analytics. Descriptive analytics or data dashboard or dashboard analytics.

(Refer Slide Time: 00:26)



Data Dashboard

- Represent complex data in a user-friendly graphics – charts, graphs
- For a researcher to extract inferences
- To communicate the insights/findings
- Business intelligence tool
- Examples:
 - Product Review
 - Advertisements –Newspaper, TV
- Domain
 - Marketing
 - Health Sector
 - Finance

Introduction to Learning Analytics 2

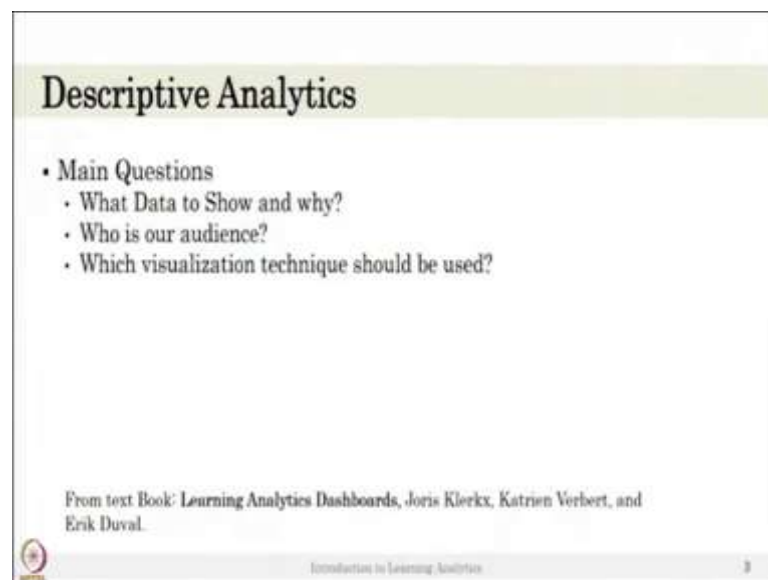
Data dashboard is representing the complex data in a user friendly graphics such as charts, graphs, pie charts or interactive graphs or some comparative graphs. The term dashboard is actually the pilot or the drivers see when they drive the vehicle in front of them is called dashboard. In a data dashboard we collect data from the students or the users or the employees based on the domain and we represent the data in a dashboard for the researches, learners or teachers or the other stakeholders.

So, in a learning analytics we will talk about data dashboard and how this learning data is used to represent in a dashboard. So, dashboard is very important for researchers to extract the inferences, to extract the insights to identify what happened; it is more important to look at the data and visually so, that they can compare two data and get new inferences from the data. So, this is the first step when you do the research on data that is collecting data and showing it in a digital format or graphics format. The dashboard is

also used to communicate the insights or something findings the researcher found to the other stakeholders like a teachers or the students.

In business intelligence tool dashboard is very important key metric. The examples of dashboard can be seen everywhere. You might have seen the dashboard in Amazon rating of the product or in a movie reviews where in Amazon you see number of users rated five star for a particular product number of user rated four star then (Refer Time: 02:04) bar chart will be shown or in the movie reviews you can see how many users reviewed this movie what is the ratings and of this with the stars. Dashboard analytics is very important for marketing, health sector and finance. Also in learning analytics for researchers it is very important to understand what is the data.

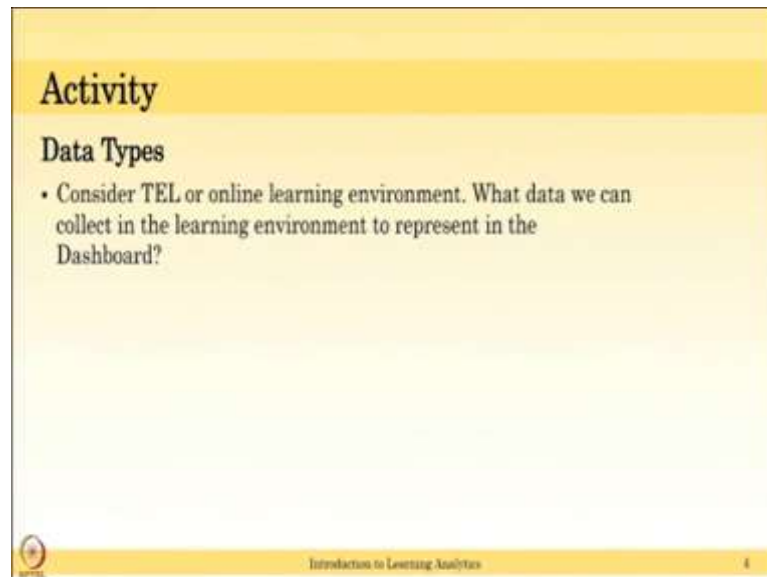
(Refer Slide Time: 02:25)



However, in dashboard analytics or descriptive analytics the main questions is what data to show and why we need to show this data. So, what data you want to look at is that number of students passed in last year or what is the passed percentage of students in last 3 years and what did I want to show and why you want to see; also who is the audience. Are we showing this data to teacher then you need to finer grained data. If you are showing this data to an administrative of the university or college, then you need a coarser grained or abstract level data showing that, there are like 30 percent of students passed all the exams, a few students did not pass the exam something like that.

So, you should be considering these questions when we are thinking about creating a dashboard analytics or descriptive analytics, that is what data you show and why you have to show this data, who is our audience and we decide what is the visualization technique you will be using it to show the data.


(Refer Slide Time: 03:22)



Activity

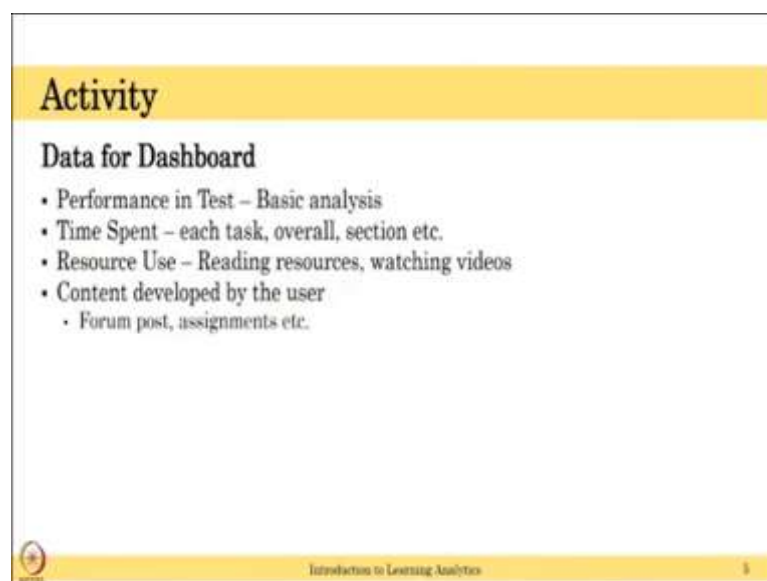
Data Types

- Consider TEL or online learning environment. What data we can collect in the learning environment to represent in the Dashboard?

 Introduction to Learning Analytics 4

In this activity consider you have a technology enhanced learning environment like METtLE we saw in last class or online learning environment MOOC or YouTube videos; you have developed the environment also you have a lot of data collected from these environments. What data is important in this environment to represent in the dashboard? Given that you have access to collect data in these two environments like a TEL or online learning environment, which data you will collect to show it in a dashboard. You can pause this video while answering these questions after completing your task you can resume to continue.


(Refer Slide Time: 04:02)



Activity

Data for Dashboard

- Performance in Test – Basic analysis
- Time Spent – each task, overall, section etc.
- Resource Use – Reading resources, watching videos
- Content developed by the user
 - Forum post, assignments etc.

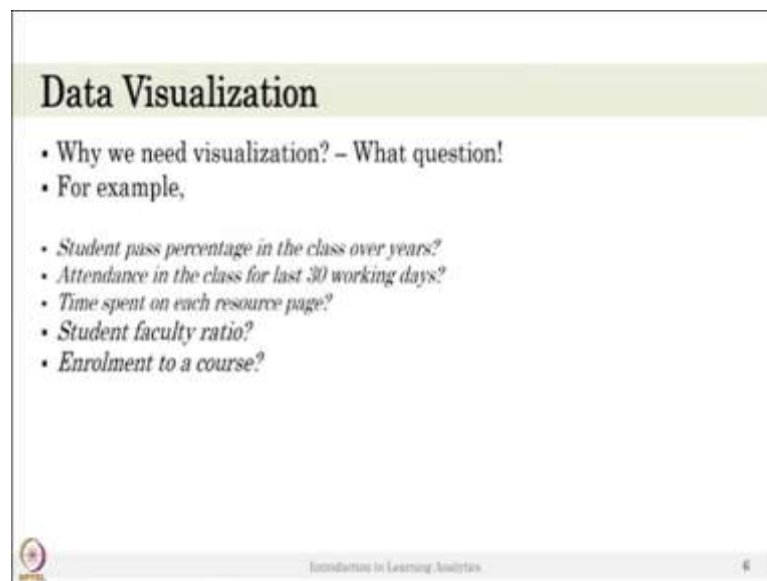
 Introduction to Learning Analytics 3

In general, for dashboard this data is very important for example, performance in the test if you are talking about the students' performance in a class or performance in the online environment or performance in the end semester, performance in some other learning environment this performance in the test pretest, post test or performance within the environment is very important to show. The other important thing is time spent on each task; is it the task within the learning environment a specific task like how much time a student spent on simulator or how much time students reading a particular page or what is an overall time spent on the task like if the student spends 80 minutes for a particular task and very less time for task 2.

Or you might say the student is focusing mostly reading chapter 1 and 2 lot of time compared to chapter 3 and 4 in this exams. So, time spent is important; time spent can be observed from the finer grained level to the coarse grained higher level. Also you have to note that what are the resource they use; the are they using the simulator, calculator or reading some resources which video they read, which video they watch, how long they watch the video, how many times they watch the video, what is speed of the watching the video all this information can be collected. It is important to show that the student is or the learner is actually using these resources to read multiple times and they are reading for this much time. So, if you have time and the resource used makes a good combination of data to represent in the dashboard analytics.

And the other important point is the content developed by the users. For example, in a discussion forum how many post the student is posting, how many forums the student is creating or in a Moodle, is the student is submitting assignment on time, how many assignments a student submitted or how many questions a student asked in the discussion forum so that you can identify that the student's engagement in the online learning environment.

(Refer Slide Time: 06:02)



Data Visualization

- Why we need visualization? – What question!
- For example,
 - *Student pass percentage in the class over years?*
 - *Attendance in the class for last 30 working days?*
 - *Time spent on each resource page?*
 - *Student faculty ratio?*
 - *Enrolment to a course?*

Introducing to Learning Analytics

So, the question is ok, we can collect all these information from the learning environments like online or the personalized and learning environment, but why we need visualization? As I mentioned earlier this is the form to show a complex data in a user friendly manner or in a graphic manner like pictures, graphs or a line charts, bar charts which is actually helping us to understand what has happened in the data or what has happened in the learning environment.

So, what questions can I ask? For example, what is the student pass percentage in a class over years. So, you can simply show a graph saying that the pass percentage over years is this so that we can look at it oh in the class compared to last 3 years last year class pass percentage might be less something like information can come out.

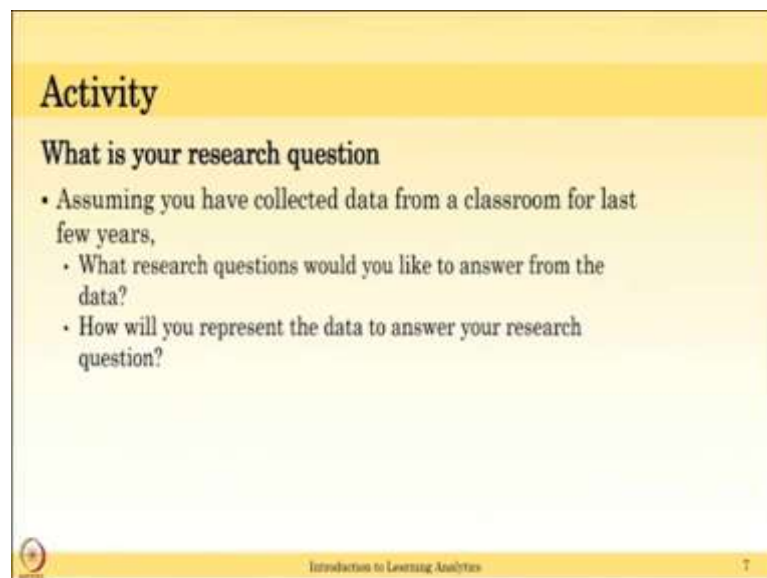
Or you want to know what is the attendance in the class for last 30 working days; is the attendance is reducing or is the attendance is constant over last 30 days or there is a huge variance in the attendance; if it is, why? Time spent on each resource page is also another

important key factor. For example, if you have five resources you can represent the average time spent by all students on each resource that makes the more insights for the researcher or you can simply (Refer Time: 07:24) on the student faculty ratio. In a class a student faculty ratio is 1 is to 15, in class b it is 1 is to 10 something like that. Then you can compare if the class b student faculty ratio is better, is the output or performance is better; you can compare those things in a correlation.

Or you can go ahead and see how many students enrolled in a course or how many students enrolled or downloaded the particular app those kind of information can be represented in the dashboard. Why these (Refer Time: 07:56) of questions are important? These questions are important because this will help you to answer some research question. For example, student faculty ratio if there are two classes, if the class b has the very better student faculty ratio we expect their learner performance is better there; is it true or not. We can show the correlation analysis. The first step is looking at that data by showing this graph so that we can start asking questions.

If the performance is same we can understand there is no correlation. So, we may not ask this particular research question.



(Refer Slide Time: 08:29)



Activity

What is your research question

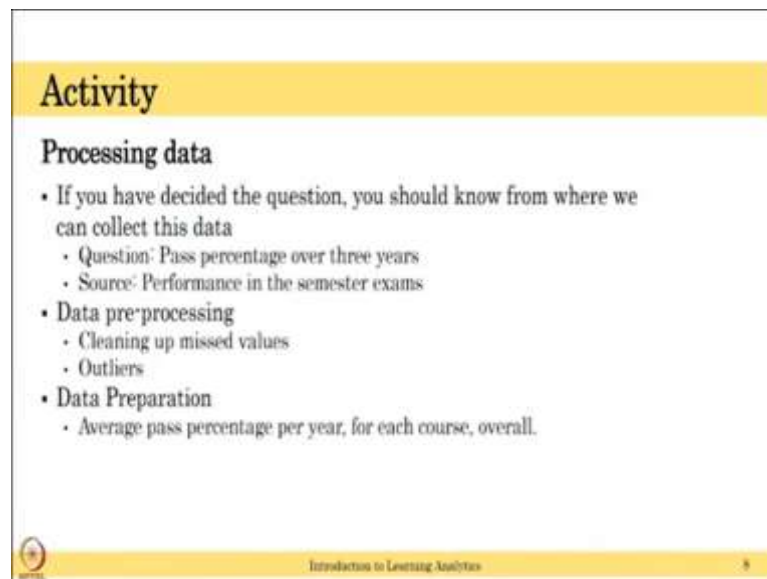
- Assuming you have collected data from a classroom for last few years,
 - What research questions would you like to answer from the data?
 - How will you represent the data to answer your research question?

 Introduction to Learning Analytics 

Let us assume that you have collected data from a classroom for last few years. This is same as the questions you were trying to answer for last two weeks. So, you have collected data from a classroom for last few years what research question would you like

to answer from the data? Think of it; also how will you represent the data to answer your research questions? So, you have to come up with the one or two research questions you want to ask from the classroom data which was not listed in the previous slide and how do you want to represent? You want to represent in a bar graph or pie chart or comparative chart or some other new methods.



(Refer Slide Time: 09:02)



Activity

Processing data

- If you have decided the question, you should know from where we can collect this data
 - Question: Pass percentage over three years
 - Source: Performance in the semester exams
- Data pre-processing
 - Cleaning up missed values
 - Outliers
- Data Preparation
 - Average pass percentage per year, for each course, overall.

 Introduction to Learning Analytics 

If you have decided the question, you should know from where we can collect this data; the data to answer the question. For example, I want to know pass percentage over 3 years, resource for this particular data is performance in a semester exams or the midterm exams or some other exams you conducted.

So, similarly if you have decided a question first you should think from where I can collect data to answer this question. List it down first, then you have to clean up the data like is there any value missed or is any outlier; it is like out of 100 students there are some data's like outlier say very less mark or really high mark; you might need to ignore them because they might be high performer we might need to look at the data which is where most students mark is distributed.

So, you have to consider the outlier or you have to also think about the missed values; there are some students' data might not have it so, you can ignore that particular data or you can recollect the data if you want. So, data preprocessing is very important manually checking the data and make sure the data is correct.