

Rayanna Harduarsingh

August 26th, 2020

Module 1 Notes

IST 687

Lecture 1.1- Many Skills

- Data Science is more than analyzing data and solving problems of data users, there are certain skills associated with completing those tasks. It's important to have perception when analyzing data in order to find the most valuable insights.
- What is the difference between structured and unstructured data?
 - Structured data is well-defined quantitative data that aligns with that of a relational database. It's highly organized and easy to analyze. Some examples of structured data would be SQL databases, excel files, phone numbers, zip codes, product inventories, customers names, and dates (easy identifiable data). Unstructured data is the complete opposite and this data does not easily align into a spreadsheet or database. It is not pre-defined nor does it follow a data model. and can be human or machine- generated. Some examples include audio and video files, images, text files, (word docs or powerpoint), email, social media data

Lecture 1.2- Data Science

- When purchasing a product, or any activity generates data that creates a cycle. This data is shared and stored (Data Architecture). As we saw in the example, lots of data was acquired (Data Acquisition). From the data, we can analyze certain points that will tell us valuable insights. Data Archiving is where we store the results of our analysis.
- Where might there be data analysis in this process?
 - I think you would analyze the data in the sales flow portion or when the cookies was scanned. When the cookies were scanned, a bunch of data was released such as the item number, quantity, the inventory, and the coupon amount. At this point, we can

see how well the cookies are selling, maybe in that particular location. If there are a lot of cookies in the inventory, we can see how long they've been there and if it is taking up shelf space.

Lecture 1.3- Data Science

- There is a lot of value we can get from data, such as information, knowledge, intelligence, and wisdom. Data Science focuses on data problems and how we can solve them smartly by analyzing it and coming up with insights. When it comes to thinking through the problem, we look at its scope, boundaries, focus area, its context, and mitigation. Data is supposed to tell a story.

Lecture 1.4- Getting Started with R

- R is an open-source software program; it is constantly being developed. Strings are statements in quotations: "this is a piece of text".
- Strings
 - R code: `myText <- "this is a piece of text"`.
- Type/ Mode
 - List of integers 43, 42, 12, 8, 5
 - Integer list is all the same "type/mode."
 - R code: `c(43, 42, 12, 8, 5)` —> data set

R Coding 1.4 - Conditionals

- Creating a vector (list of numbers)
 - `agesTest <- c(43, 38, 40)`
 - `ages <- c(Jeff, Pat)`
- To get sum, mean: `sum(ages)` or `mean(ages)`

Is R similar to more of Python or Java in terms of commands, formatting. Also, is there a list of some sort where we can get a list of commands or will we just learn on the way? Thanks!