**Lecture 3 – Data Science Overview**

**Goal of Data Analytics**:

* Answer basic strategic business questions based on data a company has collected
* What is the best-selling beer brand?
* Does reducing classroom size (teach to student ratio) really improve performance?
* Discover not-so-obvious patterns from the data.

**BIDM Cycle:**

* Business is the act of doing something productive to serve someone’s needs, and this earn a living and make the world a better place.
* Business activates are recorded on paper or using electronic media, and then these records become data. There is more data from customers’ responses and on industry as a whole.
* All this data can be analyzed… (read on slides)

**Real-Life example of data Analytics Solutions:**

* Moneyball
* Most Expensive state to buy a house 🡪 Not NY or Cali but its Massachusetts
* **Note**: Don’t just look at one attribute but a few different attributes that may influence the situation (like for housing, property tax, pricing for square foot, on top of the average listing price).

**Different Type of Data:**

* Structured Data:
  + Tabular Data, typically stored in a relational database table
    - EX: Think in terms of managing computer system in keeping track of books in a library.
    - Borrower Table with borrower id, name, birth-date, address as columns
  + Storage Format
    - CSV file, an excel file or a JSON file, but more commonly in a SQL database
  + Type of data: Numerical fields, categorical field, date-time

**Structured Data (database and data modeling):**

* Data is organized in tables
  + Tables relate to entities and relationships among entities
  + Entities are nouns like Doctor, Patient **(The table itself)**
  + Relationships are verbs like ‘treats’ **(the connection between tables)**
* Data tables can be managed using SQL language in simple declarative mode

**Unstructured Data:**

* Blob (binary large objects)
* Satellite Data
* Graph to links friends on Facebook
* Image Data
* Video with annotation
* Twitter Feeds
* Time-Series Data
* Storage format: XML, JSON

**Type of Databases:**

* Relational Databases
  + mySQL
  + msSQL
  + Postgres
* NoSQL databases:
  + MongoDB
* Time Series Databases