



# Data Science

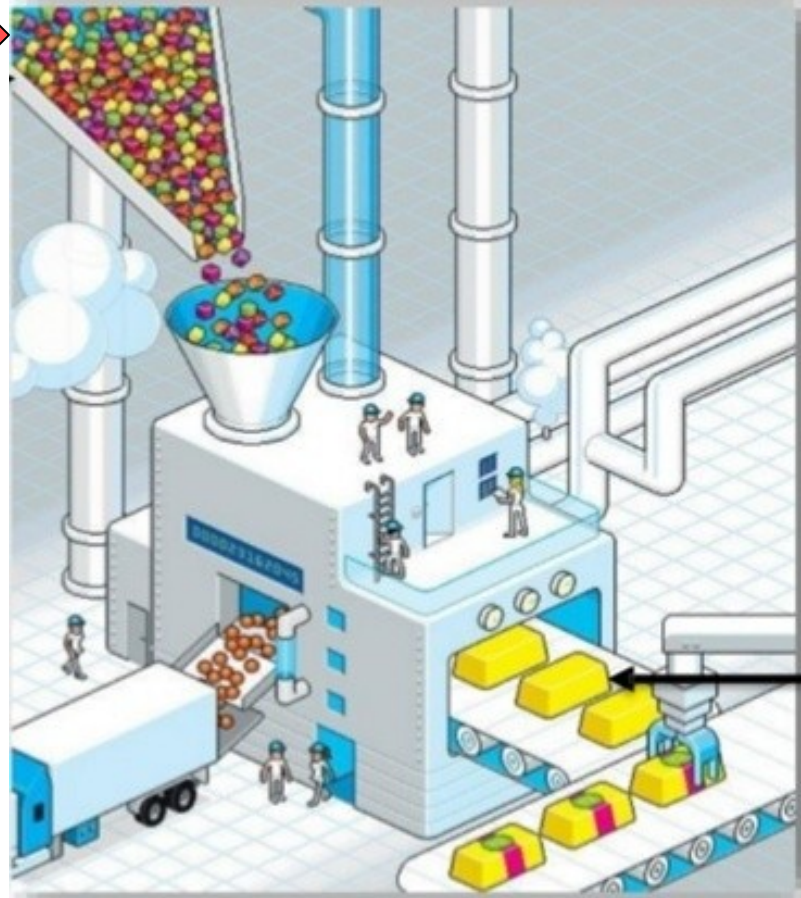
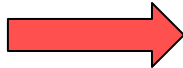
## IICT Lecture 11

Bismillah Jan



# What is Big Data?

Lots of  
Data  
(Terabytes  
or  
Petabytes)



Enterprises/  
Companies  
achieve related  
amount of  
information and  
business insights

# Big Data Scenarios



ESPN cricinfo 25 Live scores Series Teams Features Videos Stats

TOP EVENTS ▾

PAK v NZ **Lunch** NZ 153 & 249 T20 World Cup Result ENG-W 115/8 SA-W

PAK 227 & 130/4 (37.3 ov, target 176) WI-W 117/6 (19.3/20 ov, target 116) BD-W 79/5 (20 ov,

1ST TEST, ABU DHABI

**Lunch**  
1st Test, New Zealand tour of United Arab Emirates at Abu Dhabi, Nov 16-20 2018

**New Zealand** 153 & 249  
**Pakistan** 227 & 130/4 (37.3 ov, target 176)

Day 4: Pakistan require another 46 runs with 6 wickets remaining

Pakistan v New Zealand

Australia v South Africa

Australia v India

Sri Lanka v England

Bangladesh v Zimbabwe

Bangladesh v West Indies

Pakistan A v England Lions

Mzansi Super League

World Cup

QEA Trophy

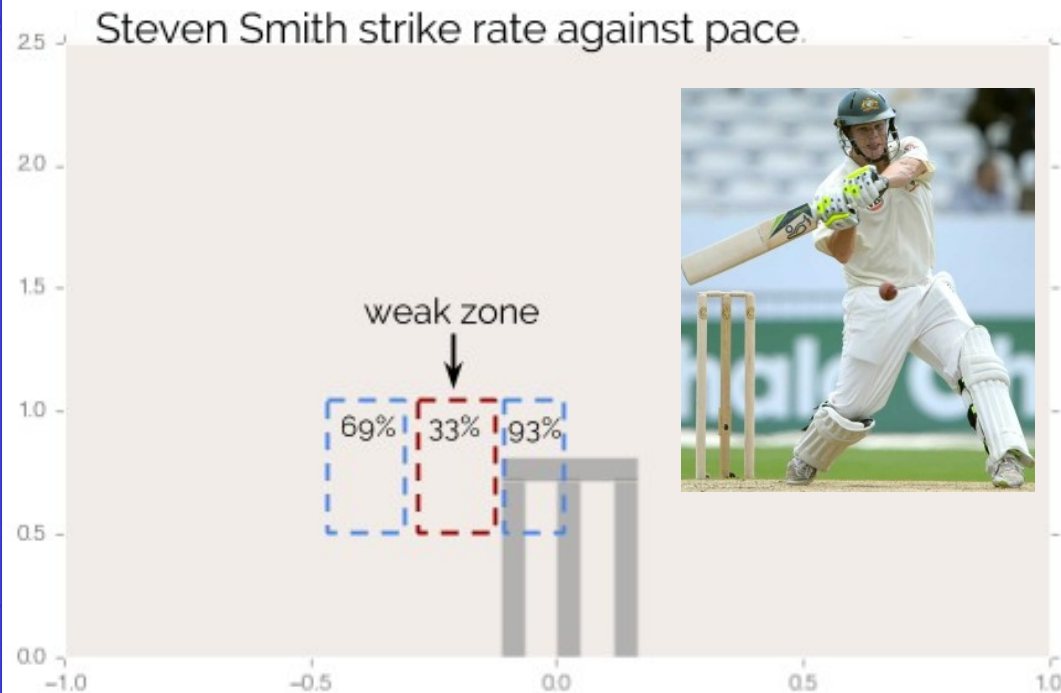
ICC Women's Championship

Wagner gets Shafiq as Paki

# Info: Sports

- Ticketing companies used big data to track ticket Sales
- Sports Teams are using data for tracking team strategies

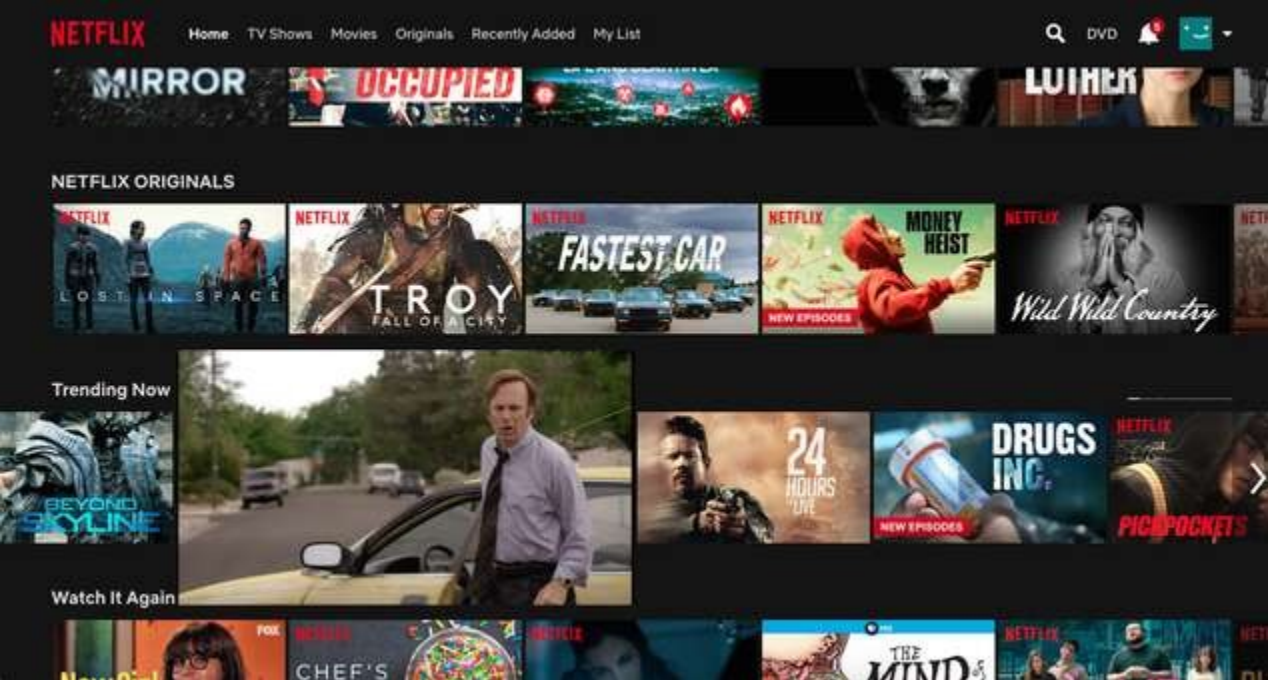
FAST



# Hospital Care

- Hospitals are analyzing medical data and patient records to predict those patients that are likely to seek readmission within a few months of discharge.
- Medical diagnostics company analyzes previous existing data to develop non-intrusive test for predicting coronary artery diseases, lung cancers, brain tumors and so on





# NetFlix

- Netflix, Inc. is an American media-services provider headquartered in Los Gatos, California
- Uses 1 Petabyte to store the videos for streaming  
(According to the Motion Picture Association of America, the number fluctuates, but the average seems to be around **600** movies created in the US every year)
- 1 Petabyte of average MP3-Encoded songs (for mobile, roughly one megabyte per minute), would require 2000 years to play
- Movies recommendation





# Large Hadron Collider (CERN Franco-Swiss)

- The **Large Hadron Collider (LHC)** is the world's largest and most powerful **particle collider** and the largest machine in the world.
- The LHC consists of a 27-kilometre ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles along the way.
- The experiments in the LHC produce about 15 petabytes of data per year, which are distributed over the Worldwide LHC Computing Grid

# And many more

- Predict Traffic given a time and location
- Facebook's friends suggestions, Image automatic annotations....etc.
- Real time soccer Analytics
- ...



# What can Data do?



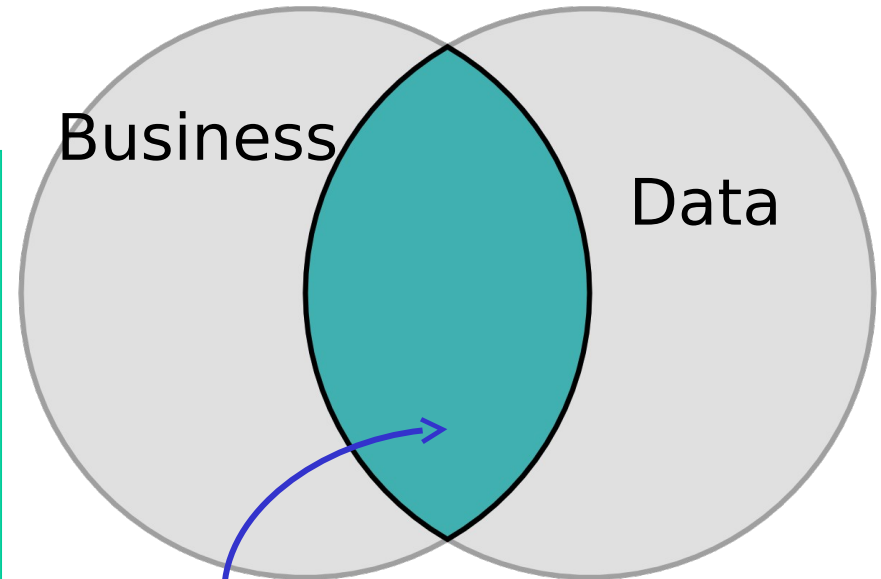


# What is Data Science?

- Data science is a multifaceted field used to gain insights from complex data. [MITx]
- **Data science** is an interdisciplinary field that uses **scientific methods**, **processes**, **algorithms** and **systems** to extract **knowledge** and **insights** from **data** in various forms, both structured and unstructured. [Wikipedia]
- **Data science** is the study of where information comes from, what it represents and how it can be turned into a valuable resource in the creation of business and IT strategies

# What can Data Science do for business?

Data Science! Extracting useful information and knowledge from large volumes of data in order to improve business decision-making or providing the business insights to make data-driven decisions



# Other applications of Data Science

fitbit



amazon

amazon.com

Recommended for You

Amazon.com has new recommendations for you based on items you purchased or told us you own.



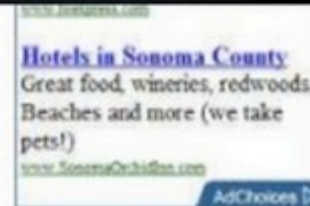
flu tracker



traffic



ads





# What Data Science Involve?

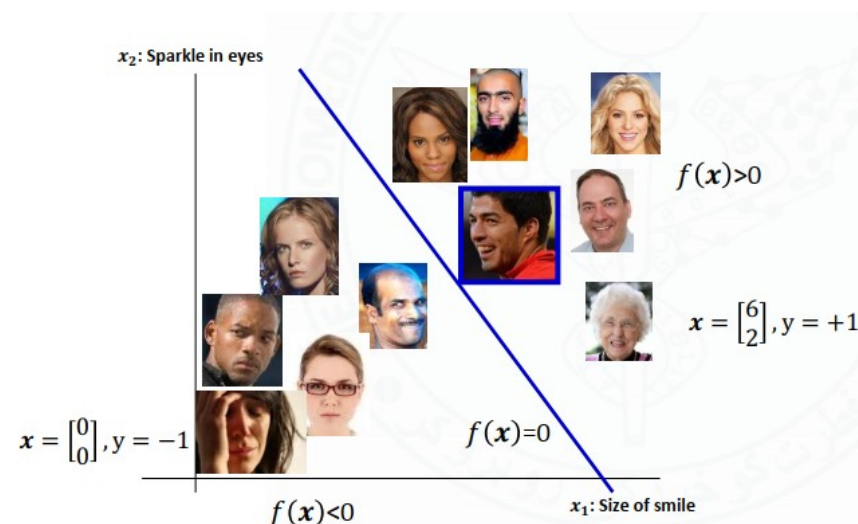
- Math and Statistics Knowledge
- Machine Learning
- Data Analyses
- Moreover, It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

# Machine Learning

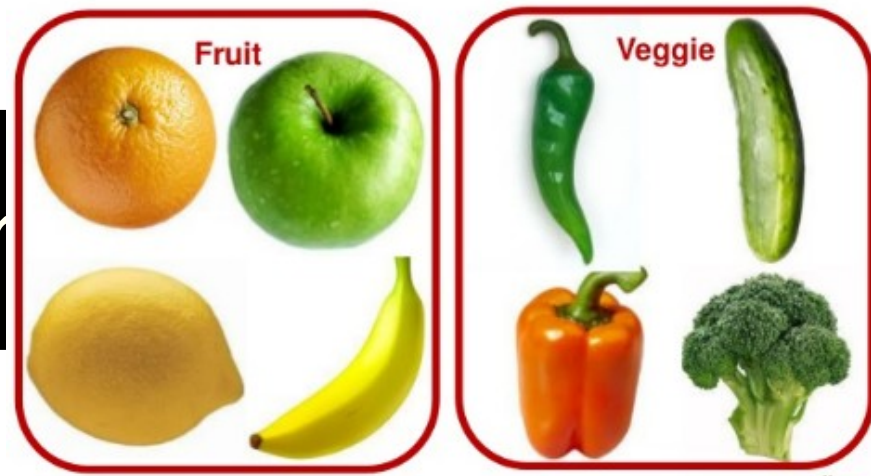
- **Machine learning** focuses on the development of computer programs that can access data and use it to learn for themselves.
- Machine Learning programs are also designed to learn and improve over time when exposed to new data

# Supervised Learning

- Definition:
  - Where a program is “trained” on a pre-defined dataset. Based off its training data the program can make accurate decisions when given new data.
- Business Application
  - Classifying Twitter sentiment
  - Recommender systems



# Unsupervised Learning

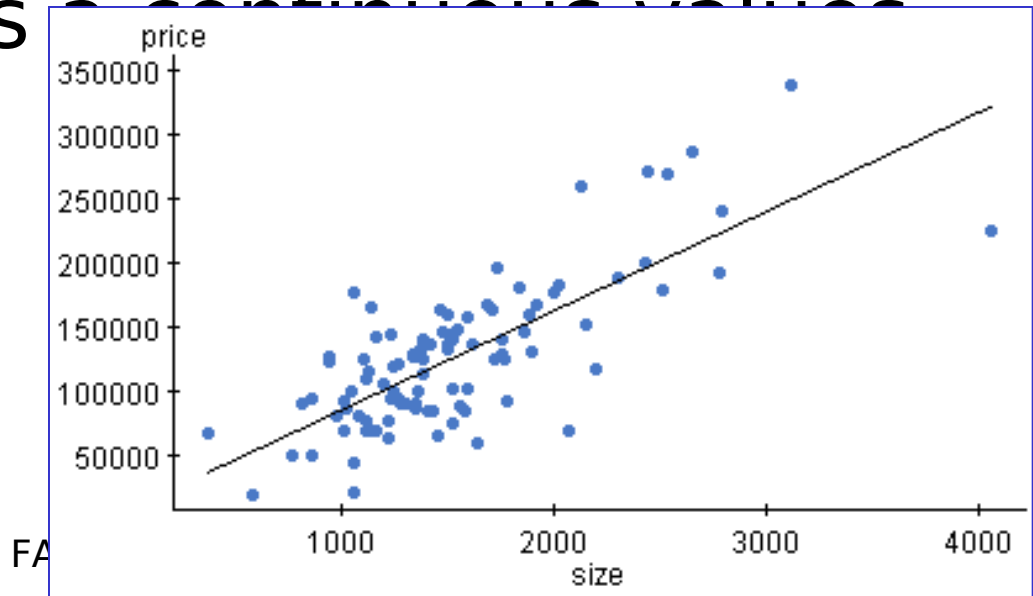


- Definition
  - Where a program, given a dataset, can automatically find patterns and relationships within the dataset.
  - Clustering or grouping of like data.
- Business Application
  - Customer segmentation
  - Understanding users and Behaviors
  - Classifying unknown and predefined images into categories

# Regression

- Sub-category of Supervised Learning
- Regression is a type of algorithm that predicts a continuous value

Plots prices  
with relative to  
increase in area





# End of the Lecture

