Foundations of Data Science

Course Code - 23DS3PCFDS

PREPARED BY - LAKSHMI SHREE K

AI & DS

Overview of UNIT 1

PART 1

- **Introduction to Data Science**
  - Describing Data science
  - The data science Venn diagram
  - Python for Data Science
  - Data science case studies

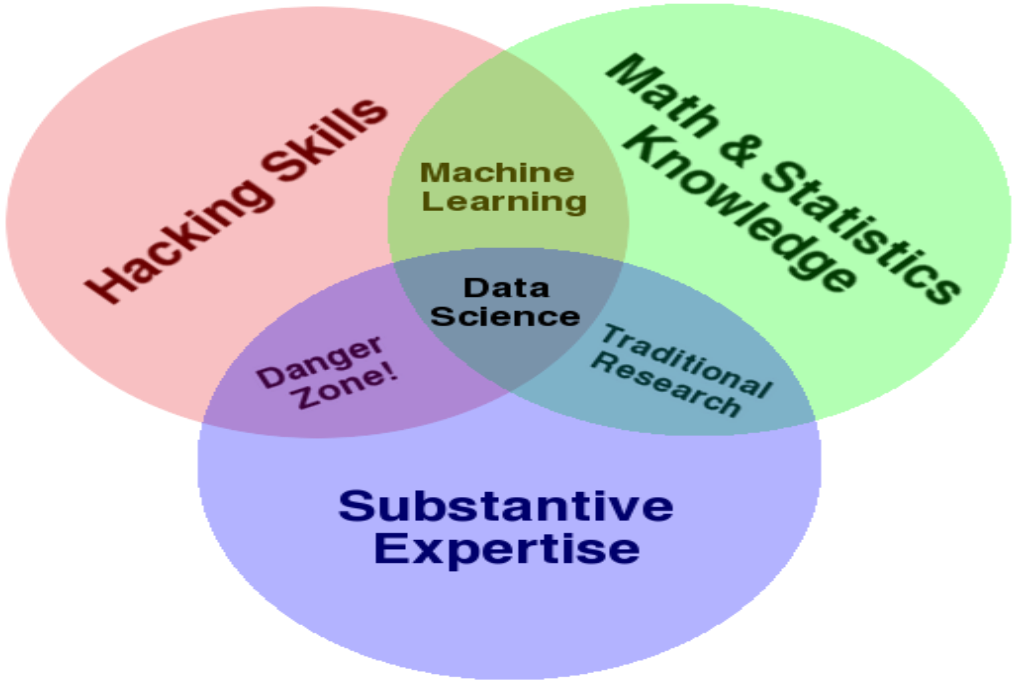
Data

- Industrial Age to Information Age
- Estimates around 64 zettabytes
- Data is created when you send a message, tweet, like , share, create a MS word doc and so on.
- SO much data!!!! -In every industry
- Data leaks
- Make sense of the data – Data Age!!!
  - Create insights and sources of knowledge that every human can benefit from.

History of Data Science

- The art of uncovering insights and trends in data has been around since ancient times.
- The ancient Egyptians used census data to increase efficiency in tax collection and accurately predicted the Nile River's flooding every year.
- People have continued to use data to derive insights and predict outcomes.

Drew Conway's Data Science Venn Diagram



Data Scientist's role in an organization

- Clarify the problem
- Data Collection
- Analysis

- Recognition
- Storytelling
- Visualization

Data Science: The Sexiest Job in the 21st Century

- The digital revolution has touched every aspect of our lives
  - Opportunity to benefit from learning about our behaviors is more so now than ever before.
  - Given the right data, marketers can take sneak peeks into our habit formation.

Introduction to Data Science

- Data is collection of information.
- Organized data
  - Data is sorted into a row/column structure, where every row represents a single observation and the columns represent the characteristics of that observation.
- Unorganized data
  - Data is in free form, usually text, raw audio/video signals

Introduction to Data Science

- Data Science is all about how we take data, use it to acquire knowledge, and then use that knowledge to do the following:
  - Make decisions
  - Predict the future
  - Understand the past/present
  - Create new industries/products
  - Gain’s new insights which would have missed.

Why Data Science?

- Parsing the huge volume of data in a reasonable time frame with previous forms of analysis is difficult
- Data can be missing, incomplete or wrong
- Data on different scales making it tough to compare
- Analytics on generated data decisions over stick-to-your-gut decisions