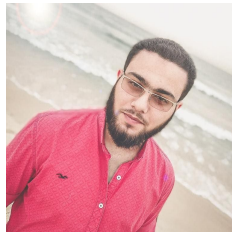


# Introduction to Data Science



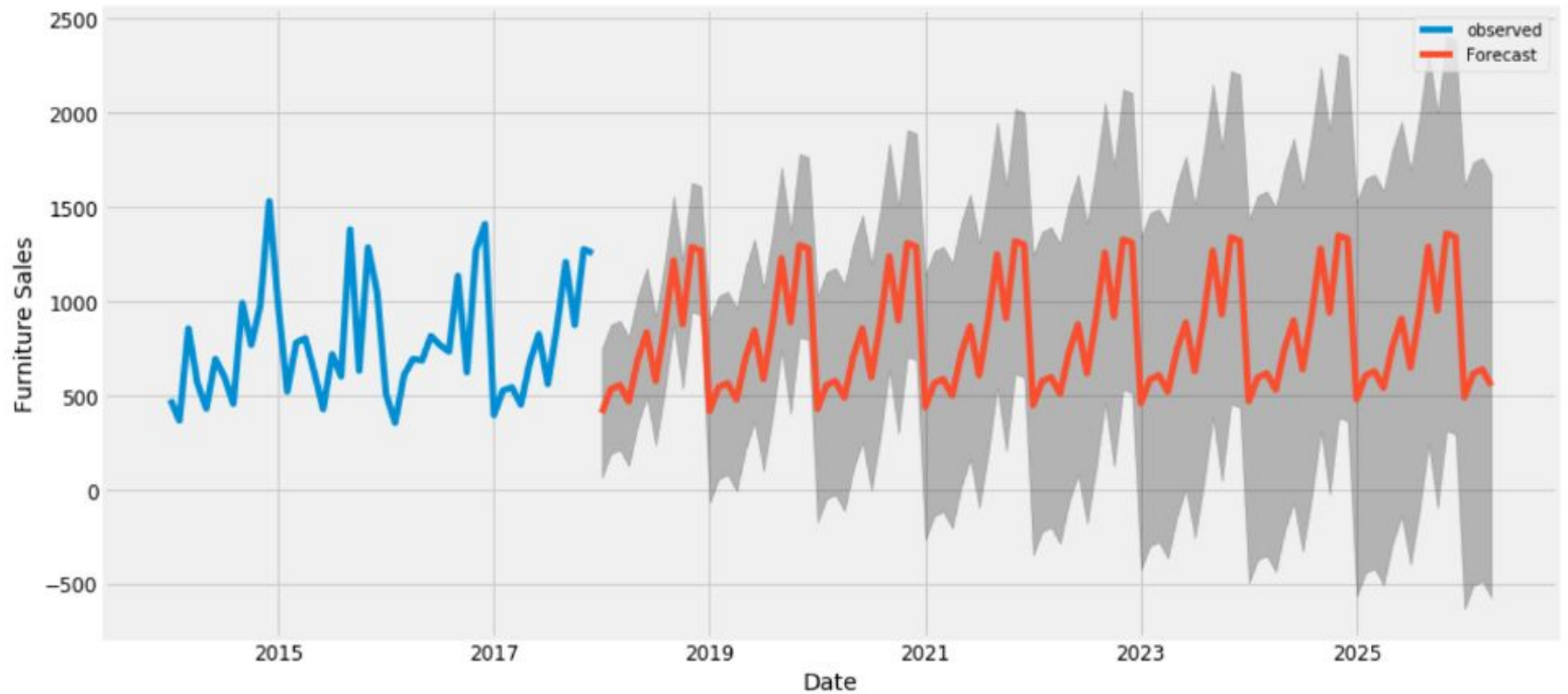
**Mohammed El-Agha**

Software Developer & Data Science Specialist

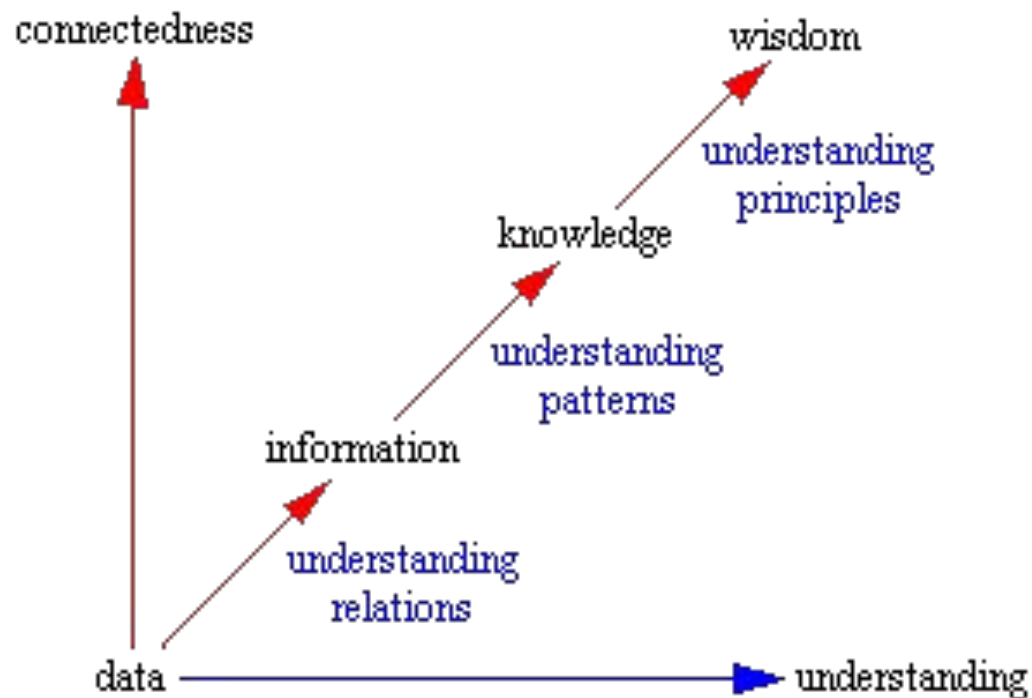
# Outline

- What's Data?
- What's Data Science?
- Data Science Tasks
- How to Work in Data Science?

# What's Data



# What's Data



# What's Data

- **Data** are sets of signs that represent empirical stimuli or perceptions
- **information** is a set of signs, which represent empirical knowledge
- **knowledge** is a set of signs that represent the meaning (or the content) of thoughts that the individual justifiably believes that they are true
- **Wisdom** is the ability to use your knowledge and experience to make good decisions and judgments

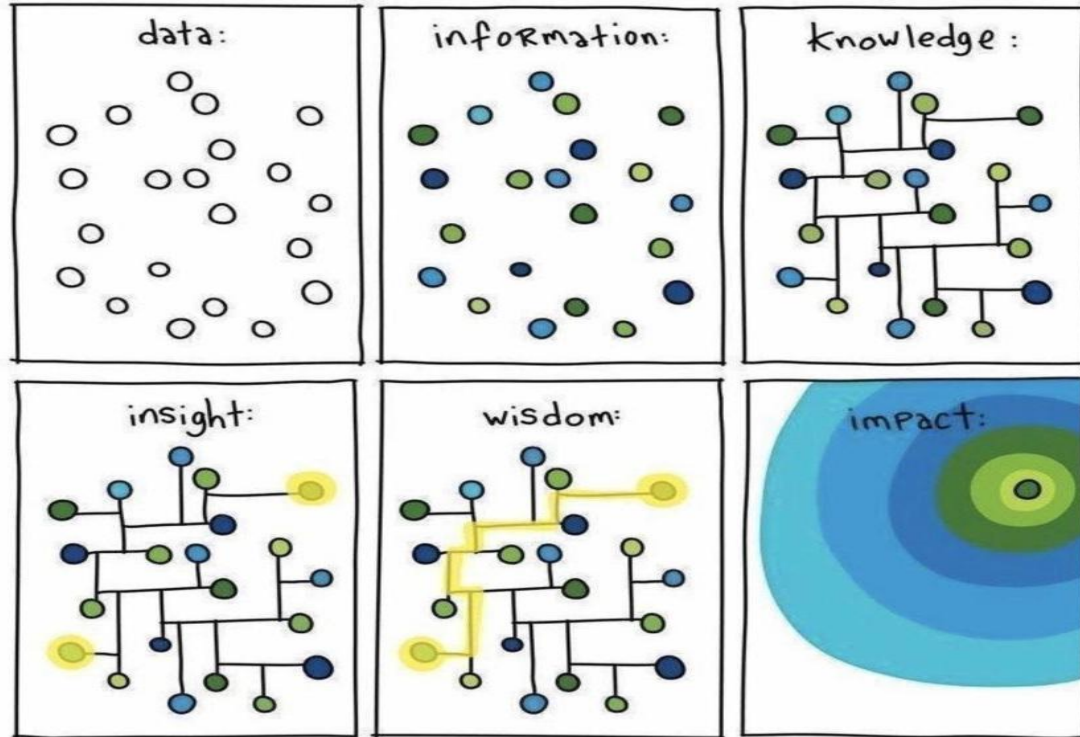
# What's Data

- **Data** represents a fact or statement of event without relation to other things
- **Information** embodies the understanding of a relationship of some sort, possibly cause and effect
- **Knowledge** represents a pattern that connects and generally provides a high level of predictability as to what is described or what will happen next
- **Wisdom** embodies more of an understanding of fundamental principles embodied within the knowledge that are essentially the basis for the knowledge being what it is. Wisdom is essentially systemic

# What's Data

- It is raining
- The temperature dropped 15 degrees and then it started raining
- If the humidity is very high and the temperature drops substantially the atmosphere is often unlikely to be able to hold the moisture so it rains
- It rains because it rains. And this encompasses an understanding of all the interactions that happen between raining, evaporation, air currents, temperature gradients, changes, and raining

# What's Data





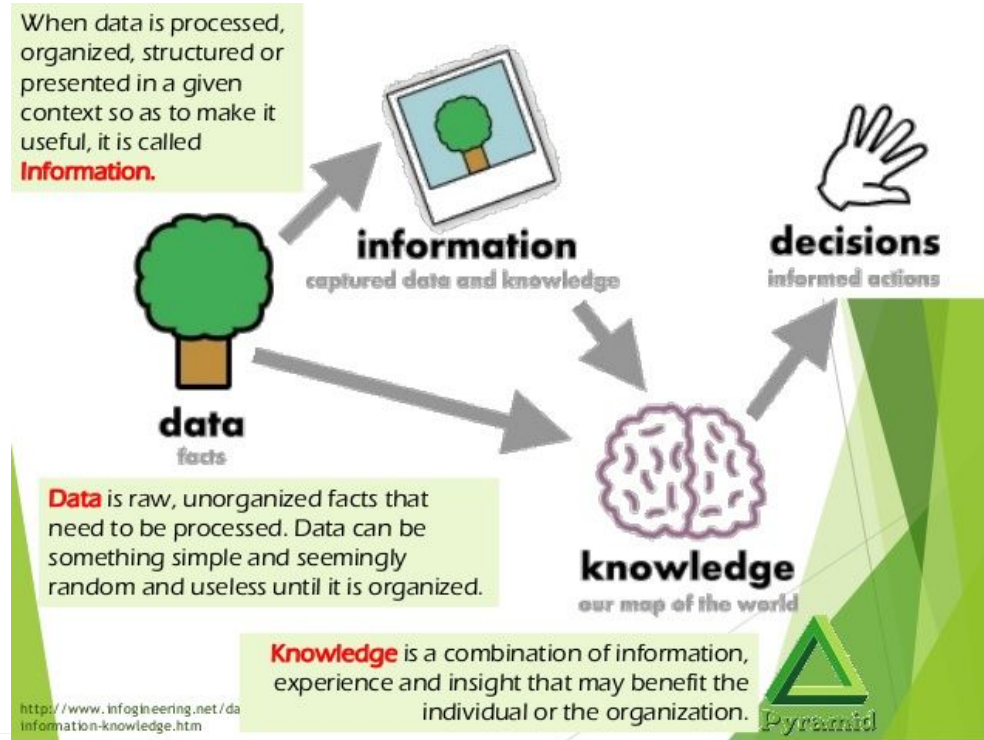
# What's Data

- **Data Product** is digital information can be purchased.
- **Data Product** is a software
  - a. Utilize input data
  - b. Process and handle it
  - c. Retrieve a useful, meaningful output
  - d. Which algorithmically-generated

# What's Data

- **Data Product** examples:
  - a. YouTube recommendation system
  - b. Amazon recommendation system
  - c. Gmail's spam filter
  - d. Sales forecasting

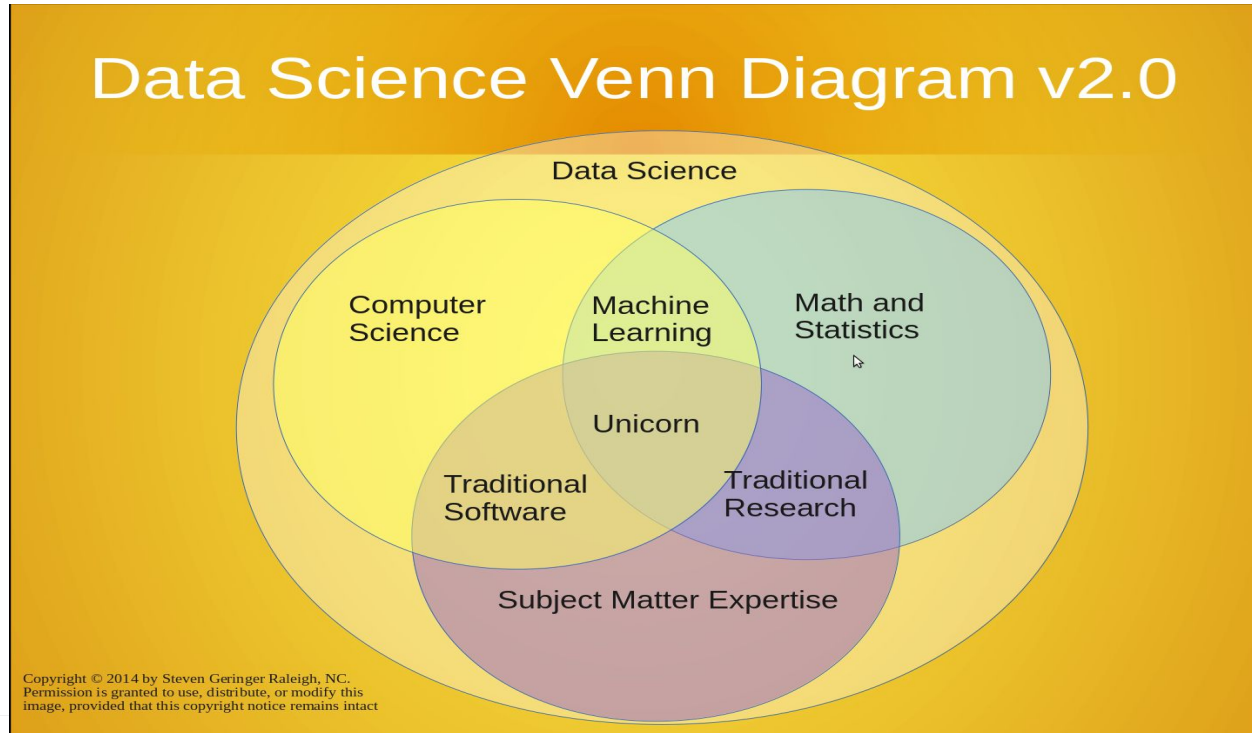
# What's Data Science



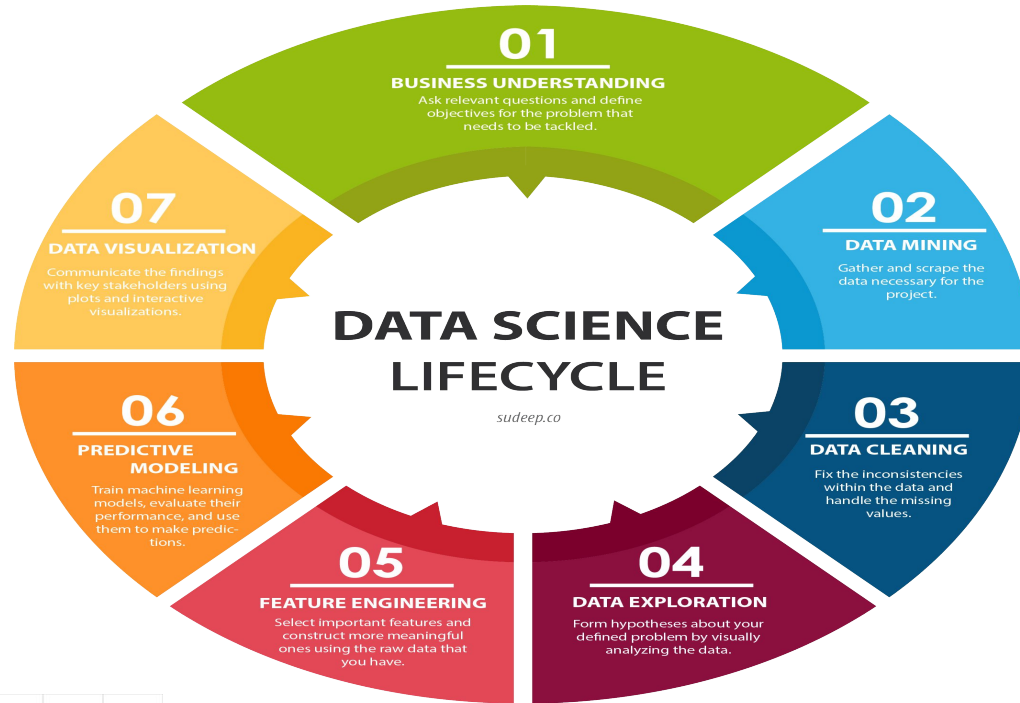
# What's Data Science

- **Data Science** extremely intersect with Statistics
- **Data Science** is a science which uses **Computer Science** to automatically apply models, methods and approaches of **Statistics, Data Analysis, Visualization** and **Decision-Making**; by a complete process of collecting, cleaning, integrating, analyzing, visualizing and interacting data; to create **data product**

# What's Data Science



# What's Data Science



# Data Science Tasks

In general Data Science tasks can be classified into two categories:

- **Descriptive** tasks characterize the general properties of the data.
- **Predictive** tasks perform deduction on the current data in order to make predictions.

# Data Science Tasks

## Descriptive

### Clustering

Detect natural groupings

### Outlier detection

Detect anomalies

### Affinity Analysis

Co-occurrence patterns

## Predictive

### Classification

Predict a category

### Regression

Predict a value

### Recommendation

Predict a preference



# Data Science Tasks

We typically use data science to answer five types of questions:

- How much or how many? (regression)
- Which category? (classification)
- Which group? (clustering)
- Is this weird? (anomaly detection)
- Which option should be taken? (recommendation)

# Data Science Tools

- **Scripting/Programming Languages:** R, Python
- **Applications:** Excel
- **Software Tools:** RapidMiner, Weka, Orange
- **Storage:** MySQL, PostgreSQL, MongoDB
- **Processing Tools:** Hadoop, Spark

# Any Questions



# Thank You!



**Mohammed El-Agha**  
**Data Science Instructor**

