# KHOA HỌC DỮ LIỆU – CHƯƠNG I

[***lecture***](https://inferentialthinking.com/chapters/01/what-is-data-science.html)

## Data science là gì?

* + Đặc điểm: Dữ liệu lớn và đa dạng (large and diverse data)
  + Gồm 3 bước chính:
  + **(Exploration)**: Đây là quá trình phát hiện các mẫu và mối quan hệ trong dữ liệu. Các công cụ chính cho việc khám phá là các kỹ thuật trực quan hóa dữ liệu và thống kê mô tả, giúp người phân tích tìm ra các xu hướng, mẫu và sự phân bố trong dữ liệu
  + (Prediction): sử dụng các dữ liệu hiện có để đưa ra những dự đoán về các giá trị mà chúng ta muốn biết
  + **Suy luận (Inference)**: Đây là quá trình xác định mức độ chắc chắn của các kết luận chúng ta rút ra từ dữ liệu. Liệu các mẫu đã phát hiện có xuất hiện trong các quan sát mới không? Mức độ chính xác của các dự đoán của chúng ta ra sao? Các công cụ chính cho việc suy luận là các kiểm định và mô hình thống kê, giúp đánh giá độ tin cậy của các giả thuyết hoặc kết luận.

## Problem solving

* Xác định có vấn đề không
* Formulate (xây dựng)
* Quan sát, thu thập dữ liệu (Khảo sát)
* Phân tích -> Tìm giải pháp -> Thực hiện, kiểm tra