

# ENGR 1330– Computational Thinking with Data Science

# Purpose of data science

The principal purpose of Data Science is to find patterns within data. It uses **various statistical techniques** to analyze and draw insights from the data. From data extraction, wrangling and pre-processing, a Data Scientist must scrutinize the data thoroughly.

Then, he has the responsibility of making predictions from the data. The goal of a Data Scientist is to derive conclusions from the data. Through these conclusions, he is able to assist companies in making smarter business decisions.

Not only business:

- Politics
- Economics
- Agriculture
- Medical science
- Sociology
- Weather
- Environment
- And many more....

# Probability

**Probability** is the branch of [mathematics](#) concerning numerical descriptions of how likely an is to occur, or how likely it is that a proposition is true. The probability of an event is a number between 0 and 1, where, roughly speaking, 0 indicates impossibility of the event and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur.

A simple example is the tossing of a fair (unbiased) coin. Since the coin is fair, the two outcomes ("heads" and "tails") are both equally probable; the probability of "heads" equals the probability of "tails"; and since no other outcomes are possible, the probability of either "heads" or "tails" is  $1/2$  (which could also be written as 0.5 or 50%).