MODULE 1: Introduction

Components of an Information System





Series of tasks, activities, or operations to achieve a goal or an outcome

Technology



Combination of hardware and software to facilitate or automate processes

Data



Discrete measurement, fact, or observation representing a real-world process

Data Mining is an outdated terminology in the field of data.

Data Science

Statistics, Patterns, Machine Learning

CS35

Data Analytics

Business-focused, Reporting, Storytelling

CS39

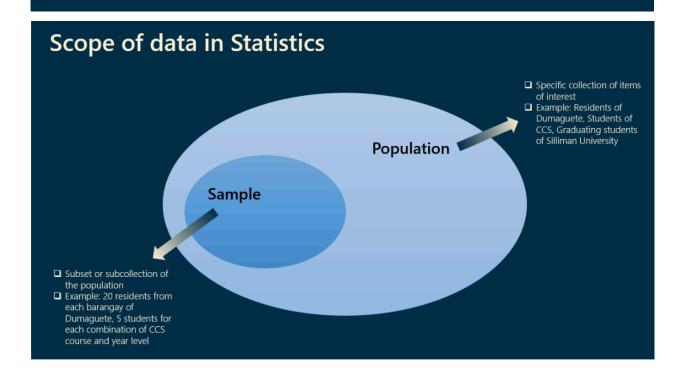
CompSci35_Data Mining Compiled and created by: Andrei Martin Diamante

Data Roles Today

	Data	Technology	Process	Statistics	Communication
Data Analyst	///	/ / /	///	✓	///
Data Engineer	///	/ / /	✓		✓
Business Analyst	✓		///	✓	///
Data Scientist	///	/ / /	///	///	///
Database Administrator	✓	/ / /			

MODULE 2: Statistics

STATISTICS is the mathematical discipline that studies the methods of collecting, analysing, and interpreting data.



Population

Sample Population

Mean

$$\mu = \frac{\sum x_i}{N}$$

$$\overline{x} = \frac{\sum x_i}{n}$$

Standard **Deviation**

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$
 $s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$

Range

 $min(x_i)$ to $max(x_i)$

 $min(x_i)$ to $max(x_i)$

Laboratory

Go to this week's page in mySOUL and download the spreadsheet



CCS and CAS students signed up for the swimming team of the university

Laboratory

NOW()
CONVERT()
AVERAGE()
STDEV.x()
AVERAGEIF()
Date Arithmetic
DATEDIF

Python Basics

Laboratory: Python Basics

Python is a scripting language that's heavily used in data science, as well as game development and web development.

