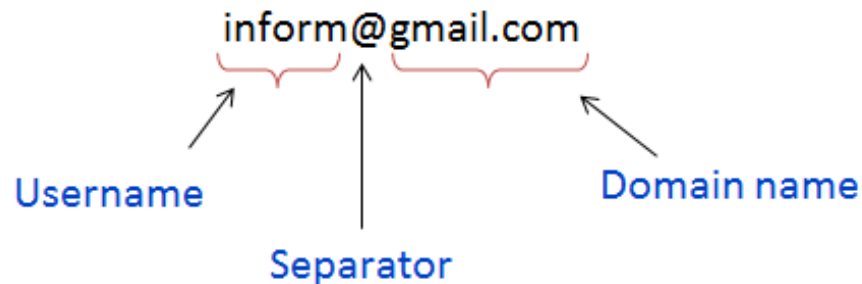# Assignment 2

DSSA 5001 – Introduction to Data Science | Due 12/11

For assignment 2, you are tasked with creating an email extrapolation program in Python. Email extrapolation is identifying the pattern of an email username for a specific company, then applying this pattern for other people who work for the same company to extrapolate their email.

For example, let's say you have the following data for a contact at AG Monitoring:

First Name: Cameron
Last Name: Richter
Email: crichter@agmonitoring.com

The program should identify that the email pattern for the above username as first initial, last name. This pattern can then be applied to AG Monitoring employees that do not have a provided email.

First Name: Matthew
Last Name: Nunes

The extrapolated email for Matthew Nunes would be mnunes@agmonitoring.com

Therefore, there are two components to this program:

1. A function that will identify and return an email pattern.
2. Apply this function to people at the same company to extrapolate an email.

Your dataset will consist of the following data fields and contain approximately 1000 records:
- Company Name
- Company Website
- Contact (Full Name)
- First Name
- Last Name
- Email (From third party provider)

We will be working with real world data, so this dataset will be inherently messy. Take into consideration the following:
- Punctuation in the email username
- Last names that have extraneous information, such as:
    o First Name: Cliff
    o Last Name: Baldwin, PhD
- If a company has more than one email pattern from our third-party email source, do not extrapolate any emails for that company.

Your deliverable is an additional column in the dataset, called Email Extrapolated, that contains the extrapolated email for each contact in the dataset, along with the Python code that executes the task.