Week 01: Introduction:Python/SQL

Data Science Bootcamp
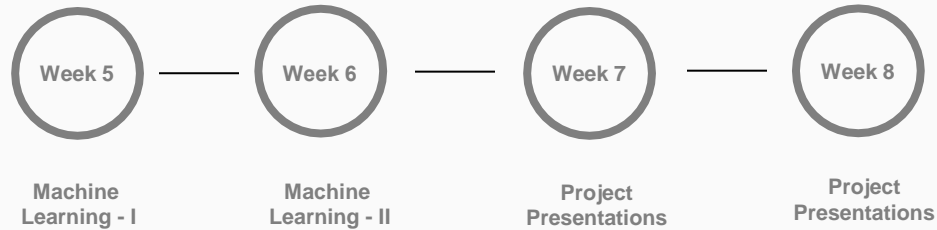Fall, 2022

Instructor: Vijayraj Gohil

About the Bootcamp

- We will start from scratch!
    - It is okay if you can not program in Python
- Preparation for **Technical** and **Personality** Assessment required for the corresponding roles in the industry
- Primarily focused on **Data Science**, with a little bit on **Python** and teamwork

Timeline

Week 1

Week 2

Week 3

Week 4

**Introduction:
Python & SQL**

Git &
Data Wrangling

Exploratory
Data Analysis

Data
Visualization

Timeline (Continued)

Week 5 — **Week 6** — **Week 7** — **Week 8**

Machine Learning - I

Machine Learning - II

Project Presentations

Project Presentations

Communities

- Join the **Slack** community to not miss out on any announcements and updates
  Link: https://join.slack.com/t/fall22datasci-djm1854/shared_invite/zt-1fvfwcl4i-14p7It~ijeQUftn0gASa~Q
  Share your **GitHub** Username on **#general** to be added to the NYU Data Science Bootcamp Organization where all the resources and tasks will be available after each session
  - If you do not have a GitHub account, please create one :)
- You can also email us at **datasciencebootcamp@nyu.edu**

Agenda

- Python
  - What makes Python so great?
  - Syntax

- SQL
  - Why SQL?
  - Introduction

Python, n.:

The best thing to happen to students and researchers

¯_(ツ)_/¯

What makes Python so great?

- Easy to read, learn and write
- Very **productive**
  - No need to spend time in understanding the syntax of the language
- Dynamically typed
  - The data type is assigned to the variable during execution
- Vast libraries support
  - Using **Python package manager** (pip) makes it easier to import external packages
- It's **FREE!**

But also…

- Python is very **slow**
  - The line by line execution of code often leads to slow execution
- Not memory-efficient
  - A **large** amount of **memory** is consumed during execution
- Runtime errors
  - Since Python is a **dynamically typed** language, the data type of the variable can change anytime
  - A variable containing integer may hold a string in the future

Python Syntax

- Python Identifier
  - Identifiers are names used to identify a variable, function, or class.
  - It can be anything from A(a) - Z(z) & can be combined with digits using _(underscores)
  - You cannot combine punctuation such as @, $, and % within identifiers.
  - Also, python is case sensitive e.g (Laptop and laptop are treated as different variables)
- Reserved Words
  - There is a list of words that cannot be used to define identifiers in python, namely (and, for, if, break, lambda, except)
  - We will explore more about them in the Interactive Google Colab notebooks...

Python Syntax (Continued)

- Identation
  - To have a clean code look, python has no braces to indicate blocks of code for function or flow control.
  - The blocks of code are denoted by line indentation

- More in the Colab Notebook...

SQL:

The lingua franca of the data professional...

Why SQL?

- Data lies at the core of Data Science
  - There needs to be a way to access and manage this data
  - SQL allows you to create, maintain and retrieve data from a database.
  - It allows insert, update, and retrieve data through various simple commands
- Easy to learn and use
  - The syntax is very similar to python and makes use of English language words
- Manage Large Volumes of Data
  - Th SQL comes with suitable commands and resources to efficiently handle and maintain extremely large chunks of data.
- SQL is Everywhere
  - It's the first choice of all the leading tech giants such as Facebook, Amazon, Netflix, etc..

Introduction

- RDBMS
  - RDBMS(Relational Database Management System) is the basis for SQL
- Easy Tables
  - The data in RDBMS is stored in database objects called tables.
  - A table is a collection of related data entries and it consists of columns and rows.
  - Each table is identified by a name that is used to access the entries of the table
- SQL Statements
  - Most of the actions you need to perform on a database are done with SQL statements called SQL queries.
  - SQL keywords are NOT case sensitive: select is the same as SELECT

Time for some hands-on...

That's all Folks...

See you in the next session :)