Lecture 7

Introduction to Pandas

What is Pandas?

- Introducing a new python package
- One of the easiest ways to import and manipulate data

pandas

Importing Pandas

import pandas as pd

- This library imports data files with tremendous ease by storing them in a new data type called a **dataframe**

	count	names	percentage
0	19837	Liam	0.0102
1	18688	Emma	0.0101
2	18267	Noah	0.0097
3	17921	Olivia	0.0095
4	14924	Ava	0.0081

Uses: Reading in Data

- Reading in data from a csv file

  Data_frame = pd.read_csv('filename.csv')

- Reading in data from a excel file

  Data_frame = pd.read_excel('filename.excel')

- Reading in data from a hdf5 file

  Data_frame = pd.read_hdf('filename.hdf')

Uses: DataFrames to Arrays/Lists

- Now how do we extract a column from our data?

  Column_values = Data_frame[‘column name’]

- If you want to make the data values a numpy array:

  Column_values = Data_frame[‘column name’].to_numpy()

- If you want to make the data values a regular list:

  Column_values = Data_frame[‘column name’].to_list()

Making Your Own DataFrames

- You have some data lists or arrays:

```
names = ['Eleanor', 'Derek', 'Jason', 'Michael', 'Janet']

birth_count = [9837, 8688, 8267, 1921, 1924]
```

- Lets first construct a dictionary:

```
data = {'names': names, 'count': birth_count}
```

Data Conversion

- Now we can convert this data to a DataFrame:

  >>> df = pd.DataFrame(data)

- Lastly we can export our DataFrame to a .csv file:

  >>> df.to_csv('filename.csv')

	names	count
0	Jim	9837
1	Dwight	8688
2	Pam	8267
3	Michael	1921
4	Stanley	1924


Demo Time