DATA 101

Lecture 2: January 25, 2021

Homework Questions...Not Much To Review But...

Policy and Nits

Jupyter Notebook

Conda Prompt

conda install ...

git clone ...

Let's have a discussion about your datasets...

- What was in them? In data terms
- What might that mean? In data terms
- What problems might you be able to or NOT be able to address? Why?

What are some questions a Data Scientist answers?

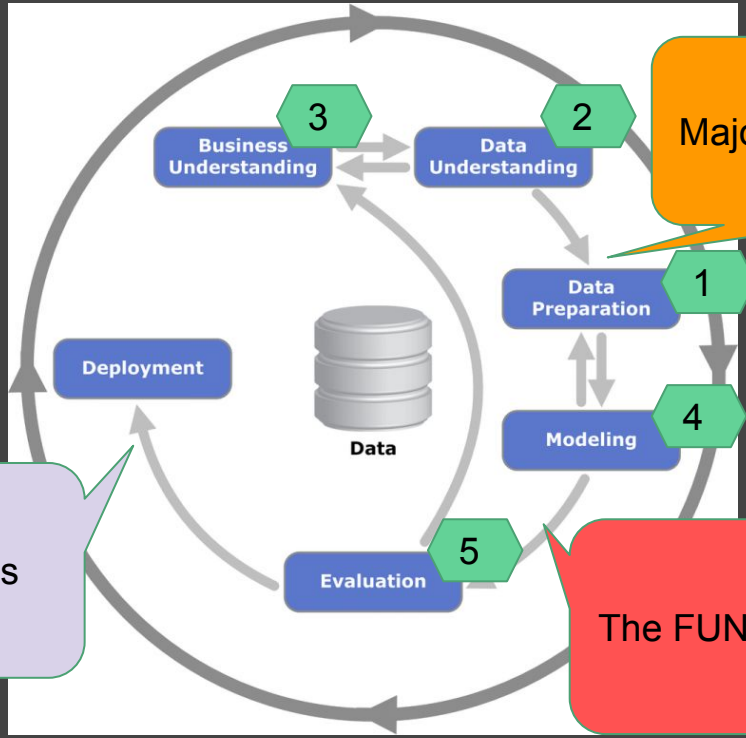
- Trend analysis
- How well do certain columns/rows compare to each other
- Use data to find a solution
- Risk
- Efficiency
- Which ‘features’ are important (application features)
- Which ‘features’ are important (data features)

Review of resources (available .ipynb)

https://github.com/wesm/pydata-book

https://github.com/jakevdp/PythonDataScienceHandbook/tree/master/notebooks

How do we do this in practice? (Also this course!)



Exploratory Data Analysis (EDA)

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide
count	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000
mean	6.854788	0.278241	0.334192	6.391415	0.045772	35.308085
std	0.843868	0.100795	0.121020	5.072058	0.021848	17.007137
min	3.800000	0.080000	0.000000	0.600000	0.009000	2.000000
25%	6.300000	0.210000	0.270000	1.700000	0.036000	23.000000
50%	6.800000	0.260000	0.320000	5.200000	0.043000	34.000000
75%	7.300000	0.320000	0.390000	9.900000	0.050000	46.000000
max	14.200000	1.100000	1.660000	65.800000	0.346000	289.000000

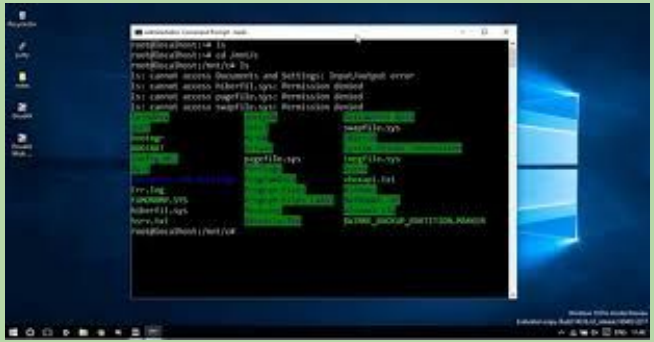
March Project

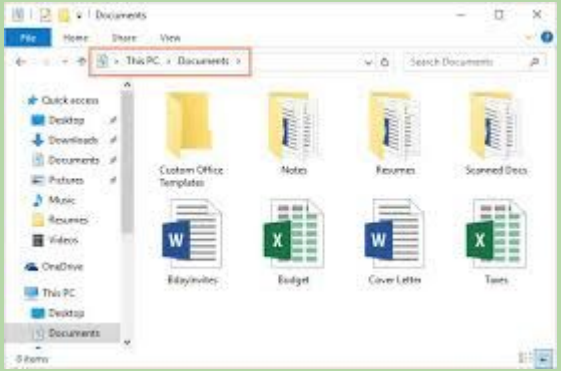
5 Minute Helpers

Brief Programming History

Terminals and “Spaces”

File Structure - what is behind that expensive GUI





5 More Minutes

What is data?

How is it stored typically (locally)?

Do we care how *its specifically stored*?

Some common formats - .csv, .xls, .json, .sql...others...?

Quick Agenda Check

Today:

- Learn to code...basics

February 1st:

- ⅓ Statistics
- ⅓ Principles of Logic Statements
- ⅓ Exploring data in code

February 8th:

- Exploratory Data Analysis *Really* Begins!
- Wrangling Data
- Organizing data with our thoughts (and our code)
- Begin March Project

February 15th:

- If all goes well...we will start talking about our first Algorithm (confusion sets in)

Jupyter Notebook

PYTHON INTRODUCTION DAY