Class 1 notes

Make sure you check in for attendance

Reach out to Angelica for who to reach out to for missing class on 3/10/and 3/12

Greg Patrick is the interim instructor

* Career computer scientist
* Technical Director for a software company

Erin Wills – TA

* Chemical Engineering and Education
* Completed the DV bootcamp in Feb 2020

Sambo Amaza – TA

* Economist in school and Data Scientist by profession

Jessica Zhang – TA

* 5 years in healthcare, extracting data from multiple sources

Career services team

* Career Material Advisor
  + Works in a more day to day fashion
* Career Director
  + A more holistic approach

Support strategies

* Live chat on BCS
* Ask BCS Learning assistants
* Attend Office Hours
* Weekly Feedback Survey
* Schedule a check-in with your SSM

**The Zen of Data**

The rise of data

Why is Data in such demand?

* Explosive growth in digitized Data (Creation)
* Explosive growth in Analytic Tools (Synthesis)
* Accelerating Search for Actionable Insight (Value)

What does the term data science mean?

* Involves spreadsheets and formulas

Fundamentally, data science is about storytelling and truth-telling

Data as Storytelling

Data as Truth-telling

Making Predictions

**Course Overview**

Our Means

* Microsoft Excel, python, pandas, matplotlib/Seaborn, APIs, Beautiful Soup, Machine Learning, SQL, MongoDB HTML/CSS, JavaScript, D3.js, Leaflet.js/Google, Maps, Tableau, hadoop

Each Class will include the following

* Overview of Lesson Topics
* Lecture
* Demonstration
* Discussions
* Activities
* Project Work

Weekly Breakdown by week

* 1 & 2 - Excel and VBA
* 3-8 – python analytics and visualization
* 9-13 – Deep dive into Databases (SQL and others)
* 14-17 – Web based visualization (HTLM, CSS, JavaScript)
* 19-24 – Final Projects and Advance Topics (Tableau)

Each one will have an example activity

Projects

Project 1 – Exploratory Data Analysis

Project 2 – Extract, transform and load, (ETL)

Project 3 – Data Visualizations

Project 4 – Machine Learning final

Helpful Tips

* Embrace your inner toddler
* Brace yourself for doubt, challenge, and confusion
* Got to put in the hours
* Form a community with your classmates
* Enjoy the lightbulb moments
* Celebrate your successes

The great debate

What data will you attempt to gather?

* Immigration of Americans
* Restaurant number of each type of food across the nation
* Yelp Reviews
* Search by hashtags for food

What relationships will you be looking for?

* Does the food traffic per restaurant mean there is a stronger connection to that type of food?
* Does a higher review for certain restaurants mean that the entire population of that area agrees?

How will you ensure your answer is most likely "true"?

How to break down the great debate

Step 1:

* Decompose the Ask
  + Who is an America, is it just homeowners, just big cities, just millennials?
* How cam we get a preference?
* Both kinds of foods are broad
  + How to do you categorize them?

Step 2:

* Identify Data Sources

Step 3:

* Define the strategy and metrics

Step 4:

* Build data retrieval plan
* Yelp Fusion API to programmatically run our queries

Step 5:

* Retrieve the Data
* Python equation
  + Randomly select a zip code – create an APO request – Save the output to a data frame

Step 6:

* Assemble and clean the data

Step 7:

* Analyze for trends

Step 8:

* Acknowledge limitations

Step 9:

* Make the Call

An analytics Paradigm

Prepare for Next Class

By next class:

1. Have Excel
2. Have Slack
3. Have the Git repo for the class
4. Figure out where the class videos will be posted

Homework

Charting Crowdfunding

* Organize and analyze a database of 1,000 generated sample projects to find any hidden trends