DATA STUDENT GUIDE

*Curriculum topics and dates are subject to change - Check BCS for updated guidance

Unit 1 - EXCEL

1.1 Dipping Into Data
1.2 Accelerating Through Excel
1.3 Charting a New Course in Excel

Objectives

- Gain perspective on the course structure and general direction of the program.
- Gain exposure to the high-level analytic strategies and tools to be covered in class.
- Feel fully proficient in basic Excel navigation and functionality
- Gain familiarity with the value of Pivot Tables and the steps for their utilization.
- Understand how to implement conditional formatting based on logical rules
- Be able to make, modify, and style bar charts, line graphs, pie graphs, and scatter plots.
- Understand how to filter data using Excel.
- Know how to calculate moving averages and regressions.
- Create, modify, and stylize basic charts from scratch using Microsoft Excel.
- Be comfortable creating scatter plots and trend lines.
- Have a firm understanding of how to create charts that contain filtered data.

Resources

- Storytelling With Data
- Excel Data Analysis
- Advanced Excel Essentials

Additional

- https://www.excelfunctions.net/Excel-VBA-Tutorial.html

Video Guide

- Lesson - Apples and Oranges - Use Excel functions to compare data from two different sheets. Watch the Video
- Lesson - Product Pivot - Use lookups to create a pivot table that visualizes the cost of recent orders of a small electronics company. Watch the Video
- Lesson - Game Sales - Create a series of scatter plots which will compare video game sales across regions. Watch the Video

Unit 2 - VBA

2.1 Very VBA
2.2 Vexing VBA
2.3 Getting Real with VBA

Objectives

- Understand the fundamental building blocks of all programming languages: variables, arrays, conditionals, loops, and functions.
- Create simple VBA macros to trigger pop ups and change cell values.
- Gain practice in writing VBA subroutines that utilize variables and conditionals.
- Begin to develop essential coding skills of syntax recollection, pattern recognition, problem decomposition, and debugging.
- Understand the basic syntax of a VBA for loop.
- Understand how to utilize for-loops in conjunction with conditionals to direct logic flow.
- Understand the value of a nested for-loop and gain basic proficiency in their use.
- Refine fundamental coding skills (syntax recollection, pattern recognition, problem decomposition, and debugging).
- Be comfortable formatting spreadsheets using VBA code.
- Understand how to loop through a table using VBA code and check for changes in values.

Resources

- Storytelling With Data
- Excel Data Analysis
- Advanced Excel Essentials

Additional

- Excel VBA Functions Tutorial

Video Guide

- Lesson - Checkerboard - Using VBA scripts, create an 8x8 grid with alternating red and black squares. Watch the Video
- Lesson - FizzBuzz - Create a VBA Script that populates the second column with the word "Fizz", "Buzz", or "Fizzbuzz" based on the value in the first column. Watch the Video
- Lesson - Hornets Nest - Create a VBA script to handle the growing Hornet infestation in your spreadsheet. Watch the Video

Unit 3 - Python

3.1 Introduction to Python I
3.2 Introduction to Python II
3.3 Introduction to Python III

Objectives

- Be able to navigate the desktop via the terminal.
- Create Python scripts and run them in terminal.
- Begin to understand programming concepts in Python.
- Feel confident reading data into Python from CSV files.
- Feel confident writing data from Python into CSV files.
- Know how to zip two lists together and when this is helpful.
- Have a firm understanding on how to create and use Python functions.
- Be able to create and use Python dictionaries.
- Be able to read data in from a dictionary.
- Have a firm understanding of coding logic and reasoning.

Resources

- Python - Beginner
- Python Scripting
- Python f-strings
- Python DataStructures
- Python CSV Module
- Git/Github
- Visual Git Guide
- Python 3's f-Strings

Additional

- Python Crash Course

Video Guide

- Lesson - House of Pies - Use Python to build a command line application to buy different types of pies. Watch the Video
- Lesson - Rock, Paper, Scissors - Create a RPS games that takes user input from the command line and plays against the computer. Watch the Video
- Lesson - Read Netflix - Find the info to some of Netflix's most popular videos Watch the Video

Unit 4 - PANDAS

4.1 Introduction to Numpy and Pandas
4.2 Data Munging with Pandas
4.3 Real-World Data Parsing with Pandas

Objectives

- Be able to serve Jupyter notebook files from local directories and connect to their development environment.
- Be able to create Pandas DataFrames from scratch.
- Understand how to run functions on Pandas DataFrames.
- Know how to read/write DataFrames from/to CSV files using Pandas.
- Understand how to navigate through DataFrames using Loc and Iloc.
- Understand how to filter and slice Pandas DataFrames.
- Understand how to create and access Pandas GroupBy objects.
- Understand how to sort DataFrames.
- Know how to merge DataFrames together whilst understanding the differences between inner, outer, left, and right merges.
- Be able to slice data using the cut() method and create new values based upon a series of bins.
- Feel more confident with fixing Python/Pandas bugs within Jupyter Notebook.
- Be able to use Google to explore additional Pandas functionality when necessary.

Resources

- Formatting
- 10 Minutes to Pandas
- Pandas Documentation
- Visual Guide to Joins
- Pandas Merging
- Pandas Cheat Sheet

Video Guide

- Lesson - Training Grounds - Use Pandas to perform analytical actions on the DataFrame provided. Watch the Video
- Lesson - Search for the Worst - Find the worst player in a position using Pandas. Watch the Video
- Lesson - Cryptocurrency - Analyze the given CSV files using merges. Watch the Video

Unit 5 - MATPLOTLIB

5.1 Intro to Graphing with Matplotlib
5.2 Matplotlib Advanced Customization
5.3 Real-World Data Visualization with Matplotlib

Objectives

- Understand Matplotlib's pyplot interface.
- Be able to create line; bar; scatter; and pie charts.
- Be familiar with basic plot configuration options, such as xlim and ylim.
- Feel comfortable creating plots using the DataFrame.plot() method.
- Understand the advantages and disadvantages of creating charts using the DataFrame.plot() method.
- Be able to work through a complex data set using Pandas and then chart some visualizations based upon the cleaned DataFrame.
- Be able to define mean, median, and mode, and choose which one is most appropriate to describe a given data set.
- Be able to explain the meaning of variance and standard deviation.
- Be able to describe standard error and the difference between a sample and a population.
- Be able to add error bars to their plots.
- Be able to fit lines to their data.

Resources

- Numpy
- Tutorials
- Matplotlib User's Guide
- Matplotlib Gallery
- Pandas Plotting
- Alternative Pie Charts
- Creating a Twitter Dev Account

Video Guide

- Lesson - Roller Coaster - Plot coaster speeds over time. Watch the Video
- Lesson - Winning Wrestling - Create charts to visualize a wrestler's wins and losses over the course of four years. Watch the Video
- Lesson - Battling Kings - Plot the total number of battles each king in the Game of Thrones data set participated in. Watch the Video

Unit 6 - PYTHON APIs

6.1 Working with Web Data (API Requests)
6.2 JSON Traversal and Data Parsing
6.3 Visualizing Financial Deserts with Google Places and Census

Objectives

- Be able to make GET requests with requests.
- Be able to convert JSON into a Python dictionary.
- Read and apply API documentation.
- Sign up for and use an API key.
- Create applications from scratch using nothing but knowledge of Python and an API documentation.
- Load JSON from API responses into a Pandas DataFrame.
- Be able to use try and except blocks to handle errors.
- Successfully use the Google Maps and Places API to obtain information about geographic areas.
- Understand how to use the Census API wrapper.
- Understand the concept of rate limits and the importance of creating "test cases" prior to running large scripts.
- Have a firmer understanding of how to dissect new API documentation.

Resources

- Using APIs
- Tweepy Documentation
- JSON Testing
- OMDb API
- New York Times API
- Open Weather Map API
- World Bank API
- Google Maps API
- Alternative Pie Charts
- Creating a Twitter Dev Account

Video Guide

- Lesson - Far Far Away - Collect Information from Star Wars API. Watch the Video
- Lesson - Bujumbura - Get the current temperature in Bujumbura. Watch the Video
- Lesson - Bank Deserts - Answer the question: "What is the relationship between poverty, age, and population with the number of banks in a given area?" Watch the Video

Unit 7 & 8 - PROJECT 1 (WOOH!)

7.1 Project Work + Mini-Session
7.2 Project Work + Mini-Session
7.3 Project Work + Mini-Session
8.1 Project Work
8.2 Project Work
8.3 Project Presentations

Objectives

- Students will be able to articulate the requirements for Project 1.
- Students will be able to draw and interpret diagrams of Git branching workflows.
- Students will be able to create new branches with Git.
- Students will be able to push local branches to GitHub.
- Students will be able to pull a branch from GitHub.
- Students will be able to merge branches with Git.
- Students will be able to open, review, and merge PRs with GitHub.
- Students will resolve merge conflicts in their working copy.
- Students will push branches to GitHub.
- Students will be able to open a PR against a given branch.
- Students will be able to use Git's stash feature to save "dirty" work.

Unit 9 - SQL

9.1 Introduction to MySql I
9.2 Introduction to MySql II
9.3 SQL Application Building #1

Objectives

- Create a localhost connection to a MySQL server and have successfully connect to it.
- Create, use, and populate a MySQL database with data.
- Create, populate, and select data from a MySQL table.
- Import large CSV datasets into MySQL Workbench using the import wizard.
- Use MySQL to select specific rows/columns of data out from a table.
- Understand the different kinds of joins and how to use them to create new tables in MySQL.
- Solidify the foundations of writing basic- to intermediate-level MySQL statements.
- Develop an introductory understanding of table design and database management.

Resources

- MySQL Tutorial
- MySQL Documentation
- MySQL Command Lines

Unit 10 - Advanced Data Storage and Retrieval

10.1 Introduction to SQLAlchemy
10.2 Advanced Usage of the SQLAlchemy ORM
10.3 Introduction to Flask & Serving Data with APIs

Objectives

- Connect to a SQL database using SQLAlchemy.
- Perform basic SQL queries using engine.execute().
- Create Python classes and objects.
- Create, read, update, and delete data from a SQL database using SQLAlchemy's ORM.
- Reflect existing databases.
- Use the SQLAlchemy ORM to create classes that model tables.
- Use the ORM define relationships and foreign key constraints.
- Use joins to query related data.
- Use Flask to create and run a server.
- Define endpoints using Flask's @app.route decorator.
- Extract query variable path values from GET requests.
- Use variable paths to execute database queries on behalf of the client.
- Return JSONified query results from API endpoints.

Resources

- Essential SQLAlchemy Book
- Introduction to SQLAlchemy
- Flask Mega-Tutorial

Unit 11 - Web

11.1 Intro to HTML
11.2 Intro to GitHub Pages and CSS
11.3 Bootstrap

Objectives

- Gain a high-level understanding of HTML, CSS, and JavaScript and what their roles are when creating websites.
- Understand the basic parts of an HTML web page and how to create one from scratch.
- Learn to cover and utilize some of the most common HTML tags and selectors.
- Understand how to deploy HTML webpages to the internet using GitHub Pages.
- Understand the basics of CSS styling.
- Position HTML elements on a webpage using CSS.
- Be able to discuss media queries, the technology that is used to create the responsive Bootstrap grid.
- Understand the Bootstrap Grid and discover how to utilize it to position the elements on the page.
- Discover how to quickly and easily build web pages using pre-built Bootstrap components.

Resources

- Bootstrap 4 Tutorial
- Codecademy HTML & CSS
- Mozilla HTML Docs
- Github Pages
- Bootstrap

Unit 12 - Web Scraping and Document Databases

12.1 Mastering MongoDB
12.2 Kiss My Fist and Scrape the Sky
12.3 Rendering Your Data With Flask

Objectives

- Create and connect to local MongoDB databases.
- Create, read, update, and delete MongoDB documents using the Mongo Shell.
- Create simple Python applications that connect to and modify MongoDB databases using the PyMongo library.
- Use Beautiful Soup to scrape their own data from the web.
- Save the results of web scraping into MongoDB.
- Become comfortable rendering templates with Flask using data retrieved from a Mongo database.
- Use Beautiful Soup to scrape data.
- Use PyMongo to save data to a Mongo database.
- Use Flask to render templates.

Resources

- Mongo in 30 minutes
- Python Requests
- Webscraping with BeautifulSoup
- Python Splinter

Unit 13 - ETL Case Study Project

13.1 A Case Study of Extract, Transform, Load
13.2 A Case Study of Extract, Transform, Load
13.3 A Case Study of Extract, Transform, Load

Objectives

- Gain a high-level understanding of HTML, CSS, and JavaScript and what their roles are when creating websites.
- Understand the basic parts of an HTML web page and how to create one from scratch.
- Learn to cover and utilize some of the most common HTML tags and selectors.
- Understand how to deploy HTML webpages to the internet using GitHub Pages.
- Understand the basics of CSS styling.
- Position HTML elements on a webpage using CSS.
- Be able to discuss media queries, the technology that is used to create the responsive Bootstrap grid.
- Understand the Bootstrap Grid and discover how to utilize it to position the elements on the page.
- Discover how to quickly and easily build web pages using pre-built Bootstrap components.

Resources

- Bootstrap 4 Tutorial
- Codecademy HTML & CSS
- Mozilla HTML Docs
- Github Pages
- Bootstrap

Unit 14 - Intro to JavaScript

14.1 Intro to JavaScript
14.2 Objects, ES6, and Tables
14.3 A Case Study of Extract, Transform, Load

Objectives

- Understand JS fundamentals: arrays, conditionals, loops, functions, objects.
- Understand functional programming with map, forEach.
- Work with common data structures.
- Be introduced to data driven documents (d3.js).
- Understand how to select elements using d3.select.
- Use d3 for basic DOM manipulation.
- Understand how to use callbacks.
- Understand the structure of html tables.
- Populate a table using static data structures.
- Understand events.
- Use d3 to attach events to DOM elements.
- Dynamically manipulate the DOM through events.
- Filter data with JavaScript.

Resources

- Interactive JavaScript Sheet
- Scrimba Intro to JavaScript
- Scrimba ES6+
- You Don't Know JS (book series)
- JavaScript Tutorial

Unit 15 - Interactive Visualizations & Dashboards

15.1 Javascript APIs
15.2 Javascript Web Charting
15.3 Interactive Data Visualizations

Objectives

- Use Plotly to create the fundamental charts: Box, scatter, bar, pie, and line plots.
- Use Plotly's `layout` object to customize the appearance of their charts.
- Annotate charts with labels; text; and hover info.
- Create and manipulate advanced Plotly charts.
- Create bubble charts to visualize three-dimensional data.
- Use Flask to serve data to a Plotly frontend.

Resources

- Plotly.js Getting Started Guide

Video Guide

- **Lesson - Dynamic Pie Chart with Dropdown:** Use a dropdown selector to choose from one of three datasets to update a pie chart. Watch the Video
- **Lesson - D3 Song Lyrics:** Create a pie chart of the frequency of song lyrics using Plotly and D3. Watch the Video
- **Lesson - Pet Pals on Heroku:** Deploy your Pet Pals app on Heroku. Watch the Video

Unit 16 - D3

16.1 Fundamentals of D3 1
16.2 Fundamentals of D3 2
16.3 Advanced Concepts in D3

Objectives

- Gain a high-level understanding of SVG elements and how to append/modify them using D3.
- Understand how to bind data to SVG elements using D3 so as to create basic bar charts from scratch.
- Create a bar chart with axes using D3 so as to visualize data.
- Create different types of charts and graphs using D3.
- Cover scales in greater depth.
- Plot multiple columns from a dataset, either simultaneously or in alternation.
- Gain a better understanding of reusable code.

Resources

- Scrimba D3.js Tutorial
- D3 Official Website
- D3 Galleryhttps://scrimba.com/g/gd3js
- D3 Tutorial

Video Guide

- **Lesson - Binding Data:** Bind data values to HTML list elements using D3. Watch the Video
- **Lesson - D3 Table:** Create a D3 Table using data binding. Watch the Video
- **Lesson - Enter, Exit, Update:** Manipulate images on a page to represent data using D3. Watch the Video

Unit 17 - Mapping the Web

17.1 Geomapping 101 w/ Leaflet.js
17.2 Geomapping 102 w/ Leaflet.js
17.3 Primer on CartoDB

Objectives

- Understand the benefits that visualizing data with maps can provide.
- Learn the basics of creating maps and plotting data with the Leaflet.js library.
- Gain an understanding of the GeoJSON format.
- Understand the concept of layers and layer controls and how we can use them to add interactivity to our maps.
- Gain a firm grasp of mapping with GeoJSON.
- Learn about and practice using Leaflet plugins and third-party libraries.
- Learn how different maps can effectively visualize different datasets.
- Gain a Leaflet mastery by completing an in-class project.
- Learn the basics of creating maps with CARTO, including writing custom CSS and SQL queries to style and filter data, while also incorporating multiple data sets within the same map.
- Understand how different types of maps are better for visualizing different datasets.

Resources

- Leaflet Documentation
- MapBox API
- Leaflet.js Tutorial

Unit 18 - R

18.1 Introduction to R + Project Work
18.2 Data Wrangling in R + Project Work
18.3 Project Work

Objectives

- Learn the basics of R syntax.
- Learn the fundamental R data types.
- Gain familiarity with RStudio.
- Learn how to create tibbles.
- manipulate data in tibbles.
- Compare and contrast the features of Python and R.
- Load data into tibbles.
- Use the pipe operator to sequentialize operations.
- Create tibbles.
- Manipulate data in tibbles.

Resources

- RStudio
- RStudio tutorial
- R for Data Science

Unit 19 - Project 2

19.1 Project Work
19.2 Project Work
19.3 Project Work

Objectives

- This week is all project days and project presentations.

Unit 20 - Tableau

20.1 Business Intelligence w/ Tableau
20.2 Interconnected Data w/ Tableau
20.3 Real-World Tableau Dashboarding

Objectives

- Use Tableau to rapidly manipulate tables of data and create visualizations using a drag-and-drop style interface.
- Connect various data formats such as CSV and Excel Workbooks to Tableau.
- Perform exploratory data analysis using Tableau.
- Create groups and sets.
- Create maps and use built-in U.S. Census data.
- Create custom calculations.
- Understand what LOD calculations entail.

Resources

- Tableau Learning Site
- Tableau Certification

Unit 21 - Machine Learning

21.1 Introduction to Machine Learning & Regression
21.2 Classification & Clustering
21.3 Introduction to Neural Networks & Deep Learning
21.4 Convolutional Neural Networks & Deep Learning

Objectives

- Calculate and apply regression analysis to datasets.
- Understand the difference between linear and non-linear data.
- Understand how to quantify and validate linear models.
- Understand how to apply scaling and normalization as part of the data preprocessing step in machine learning.
- Understand how to calculate and apply the fundamental classification algorithms: logistic regression, SVM, KNN, decision trees, and random forests.
- Understand how to quantify and validate classification models including calculating a classification report.
- Understand how to apply `GridSearchCV` to hyper tune model parameters.
- Understand unsupervised learning and how to apply the kmeans algorithm.
- Articulate specific problems on which neural nets perform well.
- Use sklearn's to build and train a deep neural network.
- Use Keras to build and a train a deep neural network.

Resources

- Scikit-Learn
- Machine Learning with Python Cookbook
- Deep Learning with Python
- MNIST and Neural Networks

Unit 22 - Big Data

12.1 Introduction to Big Data & Hadoop
12.2 Natural Language Processing w/ Spark

Objectives

- Identify the pieces of the Hadoop ecosystem.
- Identify the differences and similarities between Hadoop and Spark.
- Write MapReduce jobs locally with MRjob.
- Manipulate data using PySpark dataframes.
- Explain why NLP is necessary in a big data toolkit.
- Apply transformations resulting from NLP data processing to PySpark dataframes.
- Explain and utilize PySpark text processing methods like tokenization, stop words, n-grams, term and document frequency.
- Utilize a NLP data processing pipeline to create a spam filter.

Resources

- Recommended Book: Python Natural Language Processing Techniques
- The NLTK Toolkit
- NLP for Big Data: What Everyone Should Know
- What is Natural Language Processing?
- 7 Applications of Deep Learning for Natural Language Processing
- The Art of Tokenization
- What does tf-idf mean?
- TD-IDF Explained
- TD-IDF in Apache Spark
- 6 Easy Steps to Learn Naive Bayes Algorithm (with codes in Python and R)
- Feature Extraction and Transformation in Adobe Spark - Documentation

Units 23-24 - Final Project

23.1 Project Work
23.2 Project Work
23.3 Project Work
24.1 Project Work
24.2 Project Work
24.3 Final Presentations

Objectives

- This week is all project work days and Final Project Presentations.