

## **Elements of Data Science: A First Course**

### **Fall 2020**

**Time:** Monday 7pm-9:30pm

**Instructor:** Bryan R. Gibson PhD

**Textbook:** Python Data Science Handbook by Jake VanderPas  
Python Machine Learning by Raschka and Mirjalili (3rd Ed Recommended)

**Prerequisite(s):**

- Introductory programming class as well as basic familiarity with Python 3.
- Basic familiarity with the command line.

**Course Description**

This course is designed as an introduction to elements that constitute the skill set of a data scientist. The course will focus on the utility of these elements in common tasks of a data scientist, rather than their theoretical formulation and properties. The course provides a foundation of methodology with applied examples to analyze large engineering, business, and social data for data science problems. Hands-on experiments with Python will be emphasized.

Topics include:

- Python Data Science Tools
- Data Cleaning, Exploration and Visualization
- Hypothesis Testing and Statistical Modeling
- Classification, Regression and Clustering
- Dimensionality Reduction and Topic Modeling
- Model Evaluation and Model Selection
- Feature Engineering and Feature Selection
- Natural Language Processing
- Data processing and delivery using ETL and APIs
- Dealing with Time Series Data
- Recommendation Engines

**Assignments and Grading**

Weekly Quiz	10%
Homework Assignments (Four, equally weighted at 10% each)	40%
Midterm Exam	25%
Final Exam	25%
TOTAL	100%

Quality of Performance	Letter Grade	Range %	GPA/Quality Pts.
Excellent - work is of exceptional quality	A+	99 - 100	4.33
	A	93 - 98.9	4.0
	A-	90 - 92.9	3.67
Good - work is above average	B+	87 - 89.9	3.33
Satisfactory	B	83 - 86.9	3.0
Below Average	B-	80 - 82.9	2.67
Poor	C+	77 - 79.9	2.33
	C	73 - 76.9	2.0
	C-	70 - 72.9	1.67
	D	65 - 69.9	1.0
	D-	60 - 64.9	0.67
Failure	F	< 60	0.0

## Weekly Outline

Week	Topic	Readings	ToDo
1	Introduction to Data Science Problems and Tools		
2	Data Exploration and Visualization in Python		
3	Hypothesis Testing and Statistical Modeling		
4	Prediction: Classification and Regression and Scikit-learn		
5	Data Cleaning, Feature Engineering		
6	Feature Selection and Dimensionality Reduction		
7	Midterm Exam		
8	Model Evaluation and Hyperparameter Tuning		

<b>9</b>	Clustering and Recommendation Engines		
<b>10</b>	Natural Language Processing and Topic Modeling		
<b>11</b>	Timeseries Data and Data Delivery via API		
<b>12</b>	Databases: SQL and NoSQL		
<b>13</b>	Review, summary and Final Exam		