GW | Business

Course Number and Title	**DNSC 6303 – Programming for Analytics II** (1.5 credits)
Instructor	Dr. Joseph Siryani Department of Decision Sciences *Phone*: 703 835 6883 *e-mail*: joesiryani@gwu.edu
Time & Place	7:10 pm - 9:40 pm, on Thursdays at **DUQUES 359** **Office Hours:** On-demand basis.
Course Description	The goal is to turn data into information, and information into insight. The world is now awash with data in all forms and kinds. Data generation and analysis are not sufficient. Programming is important for those who are aiming for a career in the Data Science/Business Analytics domain, then having a rendezvous with statistics is a necessity. In this analytical age, the turnaround time allowed for making decisions is decreasing. To be able to compete better, it necessitates the integration of machine learning algorithms into the decision-making process. This course builds upon the introductory programming course and prepares you for data mining and machine learning. Advanced programming skills are critical to completing data analytics projects. This course emphasizes and focuses on data management (ingestion, including web scraping, and cleaning), command line arguments and basic prescriptive and predictive workflows. After taking this course, you will have a working knowledge of descriptive, prescriptive and predictive approaches for analytics. This class is designed to help you 'learn by doing' - both individually as well as with others. We will use the Python programming language.
Learning Objectives	Students completing this course should: ▪ Be ready to power up their career with the best language for data science; ▪ Write scripts and complete programs in Python; ▪ Read in data (including unstructured text) and learn how to process for analysis; ▪ Design, develop, and deliver reproducible data products in Python; ▪ Use Python packages for prescriptive and predictive analytics; ▪ Participate in and get evaluated in competitive analytics platforms like Kaggle.

Class Material and Homework

All class material, including lecture slides, reading articles, practice exercises, cases, data files, software tools and tutorials, practice solutions, and links to other resources will be posted on Blackboard. The work to do in preparation for each session, as well as assignments due, will be indicated on Blackboard. For each session, you should inspect the Blackboard page at least once *before* class: to check for class preparation instructions and material; and then again *after* class: to review follow-up or updated documents for that session, such as exercise solutions or class notes.

Practice exercises, provided in all sessions, will be the main vector of learning in the course. Therefore, it is essential that you work on these exercises, which will be debriefed and discussed in class. Although the practice exercises will not be graded, you are expected to have prepared them for class as instructed on Blackboard.

Software Tools and Computer Use

Download, install, and configure Anaconda, which is the recommended version of Python to use during this course. If you do not already have it on your machine, the installation link is https://www.anaconda.com.

Textbooks

The material provided in the course will be self-sufficient; no textbook is required. Supplementary, optional references on various topics will be provided.

The following books are recommended (not required). They are available free online:

- Learning Python, 5th Edition by Mark Lutz, O'Reilly Media, 2013. ISBN 978-1-4493- 5573-9
- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinny, O'Reilly Media, 2012. ISBN 978-1-4493-1979-3
- Learn Python the Hard Way
- Python Data Science Handbook

Grading

Grade Distribution

One team assignment	20%
One individual assignment	30%
Two In-class quizzes (20% each)	40%
Class participation	10%
Total	100%

Letter Grade Distribution

Range	Grade	Range	Grade
96-100	A	71-75	C+
91-95	A-	66-70	C
86-90	B+	<66	Fail
81-85	B		
76-80	B-		

The groups will consist of 4 or 5 students. The students are expected to form their own groups. The groups should be formed by the second session.

Full details about the deliverables (format, due dates, etc.) will be specified with each assignment. All assignments will have to be submitted electronically on Blackboard.

ACADEMIC INTEGRITY

As a reminder, all students in the course are expected to fully adhere to the GW University's Code of Academic Integrity. Violations of the code entail substantial academic risks.

DISABILITY SUPPORT SERVICES

If you need disability accommodations, please register with Disability Support Services (DSS) at disabilitysupport.gwu.edu/registration. If you have questions about disability accommodations, contact DSS at 202-994-8250 or dss@gwu.edu or visit them in person in Rome Hall, Suite 102.

COURSE OUTLINE

Note: *All preparation material and assignments will be posted on Blackboard*

SESSION	TOPICS	READING/PREPARATION
Pre-class	**Get Python environment ready** Download, install, and configure Anaconda, which is the recommended version of Python to use during this course. Familiarize yourself with the Anaconda environment and Jupyter Notebooks IDE.	The Anaconda installation link is https://www.anaconda.com/
Session 1	**Introduction to Python** ❖ Understanding operators, Variables and Data Types ❖ Conditional statements, Looping constructs ❖ Functions and Data structures ❖ List and Dictionaries	Practice exercises
Session	**Pandas for Data Analysis in Python** ❖ Read and Write data using Pandas, Pandas Dataframes ❖ Data Exploration ❖ Data Manipulation, Aggregating Data, Merging Data	Practice exercises
Session 3	**Data Visualization using Python (matplotlib, NumPy)** ❖ Python Plotting with Matplotlib ❖ Case Study: Data Visualization and Exploration using Auto data	Practice exercises
Session 4	**Building Model for Prediction Problems** ❖ Simple Linear Regression ❖ Case Study: MBA salary vs GPA score ❖ Multiple Linear Regression ❖ Case Study: Indian Premier League - Player auction	Practice exercises
Session 5	**Introduction to web scraping** ❖ Popular Tools and Libraries used for Web Scraping in Python ❖ Components of Web Scraping ❖ Web Scraping: Procedure ❖ Case Study: Wikipedia page	Practice exercises
Session 6	**Prescriptive models: simple optimization, optimization with command line arguments** ❖ Sample linear optimization problem ❖ Canonical form of the problem ❖ Case Study - Bland Brewery Problem ❖ Result Interpretation and Decision making	Practice exercises
Session 7	**Machine learning workflow in Python** ❖ Clustering Techniques: K- Means ❖ Case Study – Beer market positioning using clustering **Class Wrap-up**	Practice exercises