Python 101 Course Syllabus

# Course Summary

## Description

A 10-week Python training through the data science perspective, methodology, and instrument. The course is designed for analysts and those with little to no scripting experience and whose work consists of flat files and extracted data.

## Objectives

The course’s objective is to develop the technical skill of applying python for data science projects. The skills covered in the course will focus on data exploration, data wrangling, data analysis, and the dissemination of data-driven insights. The student will be instructed in the data science method using python and will demonstrate their mastery for the course in a midterm project and a capstone project.

## Information

|  |  |
| --- | --- |
| **Instructor**: | Nhan Tran |
| **TA**: | Angela Cao |
| **Course’s datetime**: | **Lecture**: Tuesday; 12p –2p (2hr)  **Seminar/ office-hour**: Thursday; 11a – 3p (4hr) |
| **Location**: | Virtual Zoom Call |

## Expectation and Conduct

Be respectful of each other and mindful of your actions and words. Students will have the opportunities to ask questions, collaborate on exercises, and expect a learning environment that is equitable and with minimal distraction. The student is expected to attend and participate in the course regularly and complete the required assignments. The student is also expected to complete quizzes, the midterm project, and the capstone project independently. Collaborating on quizzes, the midterm project, and the capstone project is prohibited and would be consider as cheating. A consequent of cheating will be an expulsion from the course.

# Course Outline

|  |  |  |
| --- | --- | --- |
| **Module** | **Objectives:** | **Topics:** |
| Python Basic  Week 1 | Learn coding basics, python syntax, concept, and terminology. Learn how to troubleshoot errors and loops. | * Introduction to Jupyter notebook * Understand & write python syntax * Python libraries * How to troubleshoot errors * Loops: inline & for-loops |
| Data Ingestion & Exploration  Week 2 | Learn data ingestion from different data sources and how to use it. Review Pandas API & method for data exploration, assess data quality issues, and review the capstone project. | * Flat file: csv, txt, Excel * Precleaning: Assigning datatype, dropna, etc. * Use built-in function to explore the data (e.g., .info(), describe(), etc.) * How to evaluate & identify data quality issues * Introducing the data & capstone challenge |
| Data Wrangling  Week 3-4 | Review Pandas API & method for data wrangling, transformation, join/merge/union, and how to work with the different data type. | * Working with numeric data (astype, etc.) * Working with string data (contains, startswith, etc.) * Working with datetime data (strftime, etc.) * Data transformation (iloc, loc,stack and unstack, shift) * Mapping (apply, map, applymap) * Data merge/join * Data merge validation * Midterm project |
| Data Analysis & Reporting  Week 5 | Review Pandas API & method for data validation, analysis, data visualization, and reporting. | * Data duplications & Missing data * Groupby aggregation * Pivot aggregation * Calculated field (Windowing Operations) * Format data (multiple-level index) * Simple data visualization * Export data * Data distribution |
| Capstone Project  Week 6-10 | Propose, develop, and present a python solution to a relevant problem/challenge. | * Proposal * Draft * Final Version * Presentation |

# Assessment of Learning

## Grading Rubric

To demonstrate your proficiency you’ll need a score of 350 or higher.

|  |  |  |
| --- | --- | --- |
| **Module** | **Assignments** | **Score** |
| Python Basic | Participation & attendance | 25 |
| Quiz 1 | 75 |
| Data Ingestion & Exploration | Participation & attendance | 25 |
| Data Wrangling | Participation & attendance | 25 |
| Midterm Project | 100 |
| Data Analysis & Reporting | Participation & attendance | 25 |
| Quiz 2 | 75 |
| Capstone Project | Proposal | 25 |
| Draft | 25 |
| Final Version | 25 |
| Presentation | 25 |
|  | **Total Grade** | **450** |

## Midterm project

* A sample dataset will be provided for the project.
* Complete the data ingestion, exploration, and wrangling process by:
  + Identifying 5 features in the data that are problematic and the solutions to resolve them
  + Provide 5 steps for standardizing the dataset
  + Create a short PowerPoint slide to showcase your work.
    - The PowerPoint should be between 5-15 slides. A PowerPoint template will be provided.
    - Provide a description and context of the dataset
    - Highlight the work done on the dataset
    - Provide an assessment of the data quality before and after
    - Provide your recommendation of the data, and whether others should or should not use the data.
* Submission: a working Jupyter notebook that showcases all the steps done to the data and a PowerPoint slide deck.
* The project will be graded by your team member. Each team member will grade 2 projects that are not their own. The midterm project grade will be an average of the 2 scores. A rubric will be provided to help with the scoring. Final project composite score = 50% come from the instructor, 50% from grading other projects (25% for each of the 2 project you will grade)

### Midterm Project Rubric

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria** | **Score** | | |
| Did the author submit a working Jupyter notebook? Yes = 5, No = 0 |  | / | 5 |
| Was there any syntactical error within the notebook? -1 point for each error found; 5 – number of errors found; Score is between 0-5. |  | / | 5 |
| Did the PowerPoint meet the length requirement of between 5-15 slides? Yes = 5, No = 0 |  | / | 5 |
| Did the PowerPoint include a description or context info about the dataset? Yes = 5, No = 0 |  | / | 5 |
| Did the PowerPoint highlight/outline the work done on the dataset? Yes = 10, No = 0 |  | / | 10 |
| Did the author identify features in the data that are problematic? 2 points for each feature identified; max score is 10. |  | / | 10 |
| Did the author provide solutions to the problem? 2 points for each solution identified; max score is 10. |  | / | 10 |
| Did the author provide steps to standardize the dataset? 2 points for each step identified; max score is 10. |  | / | 10 |
| Did the PowerPoint provide an assessment of the data quality before and after? Yes = 10, No = 0 |  | / | 10 |
| Did the PowerPoint make a recommendation of the data? Yes = 10, No = 0 |  | / | 10 |
| Is the PowerPoint clear and precise? In other word, were you able to follow along, or did it leave you unclear and confused? Yes = 10, No = 0 |  | / | 10 |
| Did the assessment make sense? In other word, was the assessment justified/valid? Yes = 10, No = 0 |  | / | 10 |
| **Total Score** |  | / | 100 |

## Capstone project

### A. Proposal

* Select a dataset to work with.
* Submit a PowerPoint slide with the following info:
  + Context and description of the dataset.
  + Identify the challenge and the goal of the project.
  + Describe what the final product will be.
* Graded for completion

### B. Draft

* Submit a draft of the project Jupyter notebook
* Graded for completion

### C. Final Version

* Submit:
  + A final draft of the project’s Jupyter notebook
  + The finished PowerPoint slides. Includes:
    - Context and description of the dataset.
    - Identify the challenge, impact, and goal of the project. Why is this important, and what good will come of this project?
    - Describe the work and the final product.
    - Discuss the implementation and scalability of the project. How will you deploy the project, and how long will it take to be in production?
* Graded for completion & accuracy

### D. Presentation

* Present the project to the team and leadership for evaluation
* Graded for content & quality