MET CS677 A2 (Fall 2020) - Data Science with Python (Mondays, 6 PM)

Instructor

Suresh Kalathur, Ph.D.
Assistant Professor, Computer Science Dept.
Boston Univeristy Metropolitan College
1010 Commonwealth Ave, Room 304
Boston, MA 02215

Email: kalathur@bu.edu
URL: http://kalathur.com/bu
Phone: 617-358-0006
Fax: 617-353-2367

Course Description

Students will learn major Python tools and techniques for data analysis. There are weekly assignments and mini projects on topics covered in class. These assignments will help build necessary statistical, visualization and other data science skills for effective use of data science in a variety of applications including finance, text processing, time series analysis and recommendation systems. In addition, students will choose a topic for a final project and present it on the last day of class.

Course Prerequisites

CS521 (*Information Structures with Python*) or instructor's consent.

Course Grading Policy

The course grade will be based on active class participation (10%), programming assignments (30%), final exam (30%), and a term project (30%). Assignments are expected to be submitted by their respective due dates. Late submissions carry a penalty.

Course Web Site

- **https://learn.bu.edu**

References

Text Books

- "*Python for Data Analysis: Data Wrangling with Pandas, Numpy, and IPython, 2nd edition*", by Wes McKinney, O'Reilly, 2017. ISBN13: 978-1-491-95766-0.**(Required book)**
- "*Python for Data Analytics: With Pandas, Numpy, and Matplotlib, 2nd edition*", by Fabio Nelli, APress, 2018. ISBN13: 978-1-4842-3913-1.**(Reference book)**

Student Conduct Code

Please review the academic conduct code

Tentative Course Schedule

- Module 1 -- Introduction and Review (9/14/2020, 9/21/2020)
    - Introduction to Data Science/Analytics
    - Software Setup (Python and Jupyter)
    - Review of Python Language
    - Numpy Basics, Visualization Basics
    - **Assignment1 Due: 9/27/2020**
- Module 2 -- Libraries for Data Analysis - Numpy and Pandas (9/28/2020, 10/5/2020)
    - Numpy
    - Pandas
    - **Assignment2 Due: 10/12/2020**
- Module 3 -- Data Visualization and Data Wrangling (10/13/2020 (Tue), 10/19/2020)
    - Visualization with Matplotlib, Seaborn, and Plotly
    - Data retrieval, Data wrangling, Data cleaning, and Data manipulation
    - **Assignment3 Due: 10/25/2020**
- Module 4 -- Data Aggregation, Time Series Analysis (10/26/2020, 11/2/2020)
    - Aggregation and Grouping
    - Regression, Time series analysis, Outlier detection
    - **Assignment4 Due: 11/8/2020**
- Module 5 -- Predictive Analytics (11/9/2020, 11/16/2020)
    - Scikit-learn library
    - Forecasting, Classification, Clustering, etc.
    - **Assignment5 Due: 11/22/2020**
- Module 6 -- Applications (11/23/2020, 11/30/2020)
    - Filtering and Recommender systems
    - Text and Social Media Analytics
    - Financial Applications
    - Other Applications
- **Final Project Presentations (Dec 7th)**
- **Final Exam (Dec 14th)**