MGMT 590: Computing for Analytics (CRN 28349)

Professor:	Yaroslav Rosokha
E-mail:	yrosokha@purdue.edu
Office Hours:	TBA
Office:	Krannert 410
Phone:	765-496-3668

I. Class Description

The main goal of this course is to introduce students to the tools and methods for data analytics. The course complements other courses in BAIM program with a bottom-up, programmatic approach of how to retrieve, manipulate, visualize, and analyze the data. The course will introduce relevant programming techniques in Python.

II. Tentative Schedule

Date	Time	Class	Contents	Due
7/6	8.15am	1	Introduction	
7/6	10.00am	2	Data Structures	
7/6	1.00pm	3	Data Visualization I	
7/8	8.15am	4	Computing Summary Statistics	HW1
7/8	10.00am	5	Data Visualization II	
7/8	1.00pm	6	Network Data Analysis and Visualization	
7/12	8.15am	7	Computational Complexity	HW2
7/12	10.00am	8	Gradient Based Algorithms	
7/12	1.00pm	9	Computational Methods and Algorithms Continued	
7/14	8.15am	10	Data Structures Continued	HW3
7/14	10.00am	11	Supervised versus Unsupervised Learning	
7/14	1.00pm	12	Introduction to NoSQL	
7/20	10.00am	13	Introduction to Large Scale Data Sets	HW4
7/20	1.00pm	14	Parallel Computing	
7/22	**3.00pm**		**Final Project Due No Later Than 3pm**	

III. Course Materials

- Slides from class will be posted on Blackboard
- Other relevant links and examples will be posted on the course resources page: http://web.ics.purdue.edu/~yrosokha/comp_tools_2016.html

IV. Course Requirements

1. Homework (50%)

Each student will be required to complete four homework assignments. In most cases HW will involve programming exercises based on the material covered in lectures. When submitting the assignment via email please use the subject line "MGMT590 HW[#]".

Grading criteria for programming assignments:

- 5 – Code submitted, but there are errors running/executing the code.
- 6 – Code runs, but results are incomplete.
- 7 – Code runs, results are complete but some results are incorrect.
- 8 – Complete results, but there is room for improvement.
- 9 – Good analysis, all results well presented.
- 10 – Excellent, with interesting issues identified (i.e. doing more than what has been asked).

2. Class Attendance and Participation (10%)

This will be determined based on your attendance and your overall contribution to the class. You are expected to come to class on time. Your instructor may cold call you to answer questions in class, and you are expected to be prepared to answer these. If you are not attentive in class (for example distracted by electronic devices) then you will not be able to answer questions when called upon.

Attendance will be taken in several class sessions. Each student will be expected to contribute to class discussion. The discussion may include questions about the material, the important points in the lectures/assignments, or suggestions for alternative approaches.

3. Final Project (40%)

A more extensive final project, along with written report, will be due on the last day of class. Students will be expected to agree with the instructor on the topic of the project by about halfway through the module. Please email a copy of your code, your final report, and any relevant data by May 5th.

Summary

Final Project	40%
4 HWs	50%
Class Participation	10%

V. Academic Honesty

Purdue University and Krannert School of Management guidelines on Academic Honesty will be strictly enforced: http://www.purdue.edu/odos/academic-integrity

VI. Class Details

1. **Introduction**
    - Course Overview
    - Data Analytics Problems
    - Why Need Computational Tools?
    - Why Python?
        - Python Setup

2. **Data Structures**
    - Lists
    - Tuples and Sequences
    - Dictionaries and Hash Tables
    - Data Frames
        - Pandas Library

3. **Data Visualization I**
    - Drawing Informative (and attractive) graphs
    - Bar Charts and Histograms
    - Scatter and Bubble Plots
    - Heatmaps
        - Matplotlib and Seaborn Libraries

4. **Computing Summary Statistics**
    - Calculating Mean, Median, Std. Dev, etc.
    - Random Sampling
    - Scipy and Numpy Libraries

5. **Data Visualization II**
    - Dynamic and interactive data visualization in web browsers
        - D3
        - Bokeh Library

6. **Introduction to Network Data Analysis and Visualization**
    - Eigen Centrality
    - Tools
        - Gephi

7. **Computational Complexity**
    - Order Notations
    - Linear versus quadratic computational complexities

8. **Gradient based algorithms: Estimating regression coefficients**
    - Recursive procedures

9. **Computational Methods and Algorithms Continued**
    - Binary Search Tree

10. **Data Structures Continued**
    - Sparse Matrix Storage

11. **Supervised versus Unsupervised Learning**

12. **Introduction to NoSQL**
    - MapReduce

13. **Introduction to Large Scale Data Sets**
    - Hadoop

14. **Parallel Computing**
    - MPI