d

MSIS 502: Business Data Analysis

Purple Cohort - Summer 2019

*“For every leader in the company, not just for me, there are decisions that can be made by analysis. These are the best kinds of decisions. They’re fact-based decisions.”* – Jeff Bezos, Amazon CEO.

*“Computers are able to see, hear and learn. Welcome to the future.”* – Dave Waters

In today’s business world, data is the most valuable resource. Many organizations have become data-rich yet still knowledge-poor. How to extract business insights and mine business value from data becomes a fundamental problem for every company. What can be learned from such “big data” to improve decisions? How can an organization leverage its massive data for strategic advantage?

Automation, Robotic and complex analytics together are shaping the future of the business, making them more sophisticated and smarter. The applications of machine learning are shining like a halo facilitating less equipment-failure, better on-time deliveries, advancement in quality, faster training cycles, advanced automation of design and production process. All this adds up to form endless possibilities for the industries opening the gates to a brand new world of highly efficient, customer-driven factories that seamlessly connects to the surrounding smart ecosystem.

The primary objective of this course is to teach you the skills you will need to perform basic data analysis for making informed business decisions and understanding a very commonly used statistical model for evaluating a dataset. We will use Python, the most widely used programming language in academia and industry for data analytics. The course will cover general concepts and techniques including Python basics, data visualization, regression analysis and interpretation and reporting of results.

The key objectives of this course are two-fold: (1) to provide you with the foundation of identifying a problem, setting a data analysis plan, performing data analyses, and interpreting the results; and (2) to provide you with hands-on experience in applying statistical and data analysis techniques to practical real-word business problems using Python, especially with data science packages including numpy, matplotlib, sklearn, etc.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Time** |  | See table below |
| **Location** |  | PCAR 292 |
| **Instructor** |  | Summer Elasaday |
|  |  | summe@uw.edu |
| **Software** |  | Anaconda with Python3, Jupyter Notebook |

**EXPECTATIONS, GUIDELINES AND POLICIES**

**Prerequisites**

Students are expected to have taken an intro-level Statistics class before, and have basic problem analyzing and solving skills such as mathematical calculation and algebra. Knowing at least one programming language (preferably Python) is a plus.

You are expected to have Anaconda with Python3 installed and working before class. You should be familiar with how to use Jupyter Notebook before class begins. A good review can be found at the following link:

<https://www.dataquest.io/blog/jupyter-notebook-tutorial/>

# Deliverables

You are responsible for completing each deliverable by its deadline. Deadlines are outlined in the syllabus or in supplementary documents accessible through Canvas. Assignments are due **before class** on their respective due dates. Unless under extreme circumstances, **make-up assignments and labs will not be accepted.**

|  |  |
| --- | --- |
| **Grading and Workload** |  |
| Individual Assignments | 20 points each (100 points total) |
| In-class Labs | 10 points each (100 points total) |
| Final Project (Analysis) | 100 points |
|  |  |

# Attendance Policy

Class attendance is very important. The labs and assignments will be based on actual class coverage. The students are expected to attend classes and are responsible for obtaining information about missed classes from other students such as changes to the due dates and contents of labs and assignments. They are also expected to actively participate in all class discussions. Arriving late to class will also likely detract from your participation grade, especially when disruptive.

# Late Policy

In class labs are a means of ensuring that you attend class on a regular basis. No late labs will be accepted.

Late individual assignments will be penalized 15% of the overall mark for each 24-hour period. (For instance, a homework turned in two days late will receive only 70% credit).

Exceptions to these rules will of course be made for serious illness or other emergency circumstances of a similar nature; in these cases, please contact me as soon as you are aware of the problem.

# Technology Courtesy Policy

Cellphones, tablets, e-readers, and etc. are expected to be turned off during class, unless instructed otherwise. Violation of the technology courtesy policy is likely to lower the participation portion of your grade.

# Foster Code of Conduct

By being a student in this course you acknowledge that you are a part of a learning community at the Foster School of Business that is committed to the highest academic standards. As a part of this community, you pledge to uphold the fundamental standards of honesty, respect, and integrity, and accept the responsibility to encourage others to adhere to these standards.

# Course Support

Because material in the course is cumulative, avoid falling behind. It will be difficult to do well in the course if you do not understand material as it is presented. If you do not understand material, please seek assistance as soon as possible.

# Tentative Course Schedule

**Bring your laptop to all classes**

|  |  |  |  |
| --- | --- | --- | --- |
| **Session** | **Date/Time** | **Topic** | **Due** |
| 1 | Mon, July 15th  6pm – 9:15pm | Introduction to Pandas for Data Analysis | HW1: Basic Pandas  Due July 19th |
| 2 | Sat, July 20th  9am - 12:15pm | Data Munging in Pandas | HW2: Data Munging  Due July 28th |
| 3 | Mon, July 29th  6pm – 9:15pm | Exploratory Data Analysis and visualization | HW3: EDA  Due Aug 4th |
| 4 | Mon, Aug 5th  6pm – 9:15pm | Linear Regression and visualization | HW4: LR Analysis  Due Aug 9th |
| 5 | Sat, Aug 10th  1:30 - 4:45pm | Logistic Regression and Reporting Analysis Results | HW5: Final Project-  Due August 20th |

## Disclaimer

I hope to cover all of the material on this schedule. We have a lot to learn on a tight schedule, so I reserve the right to speed up or slow down the material according to class needs. Accordingly, the above syllabus is subject to change as necessary.