

ISOM6670D
Learning Statistics with Python
SRING , 2021
Department of Information Systems,
Business Statistics and Operation Management
(Preliminary Draft)

Instructor: Prof. Xuhu Wan, LSK Building, Room 4072, Ext.7731, imwan@ust.hk.

Tutor: Mak Chun Cheong, Isaac, imccmak@ust.hk

Class meets: Saturday 2:00 pm-5:20 pm, Jan 30, Feb 6, Feb 13, 20,27, March 6, 13,20.

Course Description

Main Contents: This course is about how to preprocess real business data using python(mainly using pandas), identify and predict the patterns of data and finally explore the real-world application of these patterns.

Objectives: •

- Manipulation and Munging of business data
- Identifying and predicting the patterns
- Evaluating the performance of statistical models

Outcomes: •

- Know how to import and export data, how to merge or concatenate data.
- Know how to deal with missing values
- Know how to build new variables from the data with all kinds of pandas methods
- Know how to evaluate the performance of models with practical standards
- Know how to do basic modeling in different business scenarios .

Course Materials

- A. Daniel Y. Chen. "Pandas for Everyone: Python Data Analysis ", 1st Edition, Addison-Wesley Data & Analytics Series
- B. All topics and assignments require the installation of python package "Anaconda".
- C. A course website (<http://canvas.ust.hk>) is maintained which contains lectures notes, assignments, and links of data.

Evaluation

Your overall grade will be based on the following:

- A. 2 assignments(40%): Students are allowed to work individually or in group (group size ≤ 3) on the problem sets.
- B. Final(50%): The final must be done independently.
- C. Participation (10%)

Course Organization

- **Unit I. Introduction to Python**

- Lecture 1: Basics of Python
 - * Installing Anaconda

- * Jupyter notebook
- * Data types
- * Basic python language
- * Installing and importing modules
- * Numpy
- * Matplotlib
- * pyplot and Dash

- **Unit II: Data Preprocessing with Pandas**

- Lecture 2: Data Manipulation and Munging
 - * DataFrame and Series
 - * Making changes to Series and DataFrames
 - * Exporting and importing data
 - * Plotting data
 - * Assembly
 - * Missing data
 - * Tidy data
 - * Apply
 - * Group operations
 - * Datetime

- **Unit III: Data Modeling**

- Lecture 3: Reshaping statistics with Python
 - * Random Variable, Random Sample and Distribution
 - * Confidence Interval
 - * Association between Variables
 - * Extreme Value Distribution
- Lecture 4: Clustering : Customer Segmentation with K-means Clustering
 - * Customer segmentation
 - * Grouping customers with clustering method
 - * Evaluation of segmentation
- Lecture 5: Regression: Quantitative Trading with Regression Models
 - * Prediction with multiple linear regression
 - * Paper trading with python
 - * Evaluation of performance
 - * Regression after clustering
- lecture 6: Application of Time Series Model
 - * Stationary v.s. nonstationary time series
 - * Predicting the mean and variance of time series
 - * Cointegration
 - * Kalman-Bucy filter
- lecture 7: Rethinking Statitics with Bayesian Method
 - * Motivation and Bayesian Theorem
 - * Markov Chain Monte Carlo approximation, MCMC
 - * Robust regression with fat tails
 - * A Bayesian structural time series model

- **Unit IV: Final**

- Lecture 8: in-class final

Grievance Procedure

If you disagree with grades that have been assigned to your work, you have the possibility to meet instructors within one week after the grades have been published on the course website. Be specific about what it is that you don't agree with.

Academic Integrity

Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating of information facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of other groups, or tampering with the academic work of other groups. All exam answers must be your own, and you must not provide any assistance to other students during exams.