**NATIONAL UNIVERSITY OF SINGAPORE**

**NUS Business School**

**Department of Analytics & Operations**

**DAO2702 Programming for Business Analytics**

**Session**: Semester 1, 2020/2021

**Instructor**: Xiong Peng [bizxio@nus.edu.sg](mailto:bizxio@nus.edu.sg)

**Description**:

This module is an introductory course to business analytics and data science. It covers basic Python programming and preliminary statistics, with a great emphasis on addressing practical business problems and real datasets.

Data science is an interdisciplinary field that requires business insights and expertise, proficiency in programming, as well as a strong background in mathematics and statistics. Therefore, lectures and tutorials in this semester would focus on trainings in the following perspectives:

* Python programming and Pythonic coding styles
* Analytical and visualization packages
* Math and statistics
* Practical business insights and problem solving skills

**Objective**:

With the training of programming, statistics, and business insights, students are supposed to gain a big picture of business analytics, and enhance their skills in using software tools and practical problem-solving.

**Syllabus**:

1. Basics of Python programming
   1. Data structures and flow control
   2. Functions and packages
2. Data analysis with Python
   1. Analytical tools: NumPy, SciPy, Pandas
   2. Data visualization: Matplotlib
   3. Data collection and cleaning
3. Statistical inference
   1. Sampling and inference
   2. Confidence intervals
   3. Hypothesis testing
4. Linear regression
   1. Model assumptions and interpretations
   2. Categorical variables and modelling nonlinearity

**Learning Outcomes:**

Through this course, students would strengthen their skills in

1. Programming in Python;
2. Basic statistics;
3. Practical business insights.

After learning this module, students should be able to apply Python in managing, visualizing data and draw conclusions from real-world datasets via statistical models.

**Schedule**:

*Week 1.*  
**Course Overview  
Introduction to Programming and Business Analytics**  
  
*Week 2.*

**Variables, data types, and control flow I**  
  
*Week 3.*  
**Control flow II and strings**  
  
*Week 4.*  
**Built-in compound data types**

*Week 5.*  
**Functions, modules, and packages**

*Week 6.*

**Data arrays and data visualization**

*Week 7.*

**Basics of Pandas**

*Week 8.*

**Facts from data**

*Week 9.*

**Confidence intervals and hypothesis testing**

*Week 10.*

**Introduction to regression analysis**

*Week 11.*

**Regression analysis for explanatory modeling**

*Week 12.*

**Nonlinearity and categorical variables**

*Week 13.*

**Review**

**Class Materials:**

* Jupyter Notebook files as lecture notes
* Jupyter Notebook files as case studies
* Jupyter Notebook files as exercises
* Slides as supplementary (uploaded after the lecture)
* The folder "Advanced topics" provides supplementary reading materials. They are note tested but may be helpful for your project.

**Class Preparations:**

* Students are required to review lectures covered in previous weeks.
* Students are supposed to work on exercises of the previous lectures before coming to the class.
* All case studies are related to topics covered in previous lectures. Please make sure that you have worked on them before coming to the class.

**Software:**

**Anaconda:** [installation](https://docs.anaconda.com/anaconda/install/)

**Reference Books**:

Python programming:

* Python for data analysis, by Wes Mckinney
* The hitchhiker’s guide to Python, by Kenneth Reitz and Tanya Schlusser
* Python data science handbook, by Jake VanderPlas

Statistics:

* Introductory econometrics, by Jeffrey M. Wooldridge
* An introduction to statistical learning, by Trevor Hastie et al.
* Storytelling with data, by Cole Nussbaumer Knaflic

**Assessments**:

**Continuous Assessment**:

Class Participation 15%

* Answering questions posted on the forum
* Posting questions will be considered

Group Project 35%

* Team work (Team to be announced in week 2)
* Analysing real-world dataset with Python
* A report and a formal 10 to 15-minute presentation
* Grading standard would be slightly lowered due to the tight schedule
* More details can be found from the Appendix

**Final Examination**: 50%

* Open book
* Two hours
* Coding test
* All topics covered in lectures/case studies/exercises could be tested, except: 1) data visualization; 2) functions given by external links; 3) advanced topics.

# Appendix: Project Guidelines

# Scope of the Project

The team project requires students to solve a practical business problem by exploiting real datasets. You may use whatever programming/analytics techniques you know (not restricted to this module). Possible sources of real datasets: 1) [Data.world](https://data.world/datasets/open-data); 2) [Kaggle](https://www.kaggle.com/datasets?tagids=13207); 3) [Data.gov.sg](https://data.gov.sg/). You can also download datasets from other websites, or even conduct your own survey to collect data. A few examples are given below:

1. Make property investment decisions based on property price data or Airbnb data;
2. Use a dataset of movies/video games made in recent years to decide what kind of movies/games to make in the future;
3. Use a dataset of employees in a company to investigate if there is discrimination in the company.

# Submission

Each team must submit a zip file containing:

* A formal report written in Jupyter Notebook files.
* A video presentation(about 10~15 minutes)where each team member participates in the presentation. It can be recorded by ordinary smart phones, as long as it is able to show the presenters and the slides clearly.

The zip file name must follow the format: **Project\_<session name>\_<team name>**.

Late submission would be penalized, and no excuse is accepted, so please plan earlier.

# Detailed Guidelines and Grading Criteria for Project Report

The project report of the whole team would be evaluated following the guidelines below:

|  |  |
| --- | --- |
| Components | Suggestions |
| Problem statements (25%) | * A Clear and concise description of only **one** business problem * Emphasize how the dataset would help solve the problem. * Justify any assumptions/approximation/missing data. * Novelty and creativity in the selected problem/dataset is highly valued. |
| Data visualization (40%) | * **Data visualization is the key component of this project** * Use proper graphs to present the information of the dataset * Clear presentation and accurate description of your graphs * Do not create irrelevant graphs; good data visualization should well support your analysis and be efficient in delivering data information. |
| Analysis and Discussion of the results (25%) | * Derive relevant information form data visualization and use such information to solve the business problem. * You may use statistical models (KNN, linear regression, or any other models you know) covered in the second half of the semester, but this is not compulsory. * Finalize your report with necessary discussion/conclusion/recommendations. * Possible discussion on the influence of missing data and model limitations; give suggestions on solutions or future research. |
| Writing (10%) | * Well-structured flow and clear logic. * Better to keep a list of reference for all literatures/external sources you refer to. * Good English and free of error/typos |

Please note that your grade of the project report would also be affect by the peer review of your teammates. **Zero mark for zero contribution! Big trouble for plagiarism!**

# Detailed Guidelines for Project Presentation

For the presentation,

* Every team member needs to participate.
* You may simply use your cell phone to record the video.
* Due to the current situation, you are allowed to make separate video clip at home and combine them as a presentation.
* The audience should be able to see your slides and your face. The preferred format is shown as the following picture.

A screenshot of a cell phone

Description automatically generated

* No need to dress professionally
* Use proper data visualization for your slides
* Try to focus more on the business problem itself and make it as friendly as possible for audience with a limited programming/math/statistics background.
* Presentation is graded individually, according to each student's presentation skills.
* **DO NOT read a script!**

# Reference

This book may help to work on your project report/presentation: [Storytelling with Data](https://www.amazon.com/Storytelling-Data-Visualization-Business-Professionals/dp/1119002257)

Some videos from the same author:

<https://www.youtube.com/watch?v=8EMW7io4rSI>

<https://www.youtube.com/channel/UCjhGlILWNloXJdR2NTCBMlA/videos?view=0&sort=da&flow=grid>