NATIONAL UNIVERSITY OF SINGAPORE
NUS Business School
Department of Analytics & Operations

DAO2702X Programming for Business Analytics

Session: Semester 2, AY2022/2023
Instructor: Ryan Tan run@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, seminars 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

Scopes:

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 (**not tested**)
    1. Model assumptions and interpretations
    2. Package Statsmodels for regression analysis

Software:
Anaconda: https://www.anaconda.com/products/distribution#Downloads

Reference Books:
Python programming:

- Python data science handbook, by Jake VanderPlas

Data visualization:

- Storytelling with data, by Cole Nussbaumer Knaflic

Online Resources:

- Python tutor: https://pythontutor.com/
- Programming for Business Analytics: https://nus-biz-dao.herokuapp.com/
- Storytelling with data: https://www.storytellingwithdata.com/blog

Note to students: If you have any questions, please always check this document first as it contains many details that you may be asking for. If you are unable to find the answer here, please approach or contact me.

Class Materials:

- Offline lecture videos uploaded to Canvas
- Jupyter Notebook files as lecture notes uploaded to Canvas
- Jupyter Notebook files as exercises uploaded to Canvas
- Slides as supplementary uploaded to Canvas
- The folder "Advanced Topics" provides supplementary reading materials. They are **not tested** but may be helpful for your project.

Assessments:

Continuous Assessment:

Class Participation (20%)

- Participation in seminars.
- All questions are open-ended so that everyone has a chance to participate.
- Each student is expected to showcase his/her solutions to the exercises or advanced exercises 4 to 5 times, and contribute to discussions 8 to 10 times.
  Participation will be assessed based on the quality of the answers and discussions.
- **More details can be found in the FAQ and Appendices.**

Group Project (40%)

- Team work. Each team has 4 to 5 members.
- You may choose your own teammates from the same seminar session. One person from each team is required to submit to the instructor the names of all the team members and the team name. If you do not find a team to join before the deadline, you will be randomly assigned to a project group which may have fewer members.
- If you ask to change to another team after the deadline without valid reasons, there will be a 20% penalty on your project.
- A formal 15 to 20-minute presentation and an 8-page (4 pieces of double-sided paper) report.
- During each presenter's turn to present, ensure that the presenter's name and matriculation number are displayed on the bottom left of the slides the he/she is presenting, otherwise there will be a 20% penalty on the presenter's presentation.
- All files should be zipped into a zip file with the name Project_<seminar session name>_<team name>. The <seminar session name> is 1 for the 3pm seminar and 2 for the 630pm seminar. Please ensure that you follow this naming convention, otherwise there will be a 20% penalty on the project.
- The project peer evaluation is optional. However, your project grade would be affected by the peer evaluation of your teammates, if they choose to submit to the instructor. Every team member should contribute equally to the project**.**
- **More details can be found in the FAQ and Appendices.**

Final Examination (40%)

- Close-book examination using ExamSoft.
- A number of multiple-choice questions and a written question.
- A double-sided A4 cheat-sheet is allowed. Other notes and electronic devices are prohibited.
- Your student card and other relevant identity cards.
- All topics covered in the lecture content and exercises during the seminars (except week 13 and advanced topics) could be tested.
- Tuesday 25 April 1:00 PM
  (refer to your EduRec https://myedurec.nus.edu.sg/psp/cs90prd/?cmd=login for confirmed details)
- **More details can be found in the FAQ and Appendices.**

Schedule:

- Students are required to watch the offline videos and work on the exercises before attending the seminars.
- Seminars are used for discussing lecture content, going through exercise questions and presenting projects.
- Seminars are not recorded.

Date	Content
Week 1 (10 Jan)	**Course Overview + Introduction to Programming and Jupyter Notebook + Introduction to Python Programming** Lecture + Exercise
Week 2 (17 Jan)	**Control Flows of Python Programs** Lecture + Exercise
Week 3 (24 Jan)	**Public Holiday (Chinese New Year)** **No Seminar**
Week 4 (31 Jan)	**Built-in Data Structures I** Lecture + Exercise
Week 5 (7 Feb)	**Built-in Data Structures II** Lecture + Exercise
Week 6 (14 Feb)	**Functions, Modules, and Packages** Lecture + Exercise
Reading Week	**No Seminar**
Week 7	**No Seminar (Students to brainstorm and plan projects in their own groups)**
Week 8 (7 Mar)	**Lovely Pandas** Lecture + Exercise
Week 9 (14 Mar)	**Storytelling with Data** Lecture + Exercise
Week 10 (21 Mar)	**Sweet NumPy** Lecture + Exercise
Week 11 (28 Mar)	**Review of Probability + Random Sampling** Lecture + Project Presentation (Only presentation skills are graded)
Week 12 (4 Apr)	**Confidence Intervals and Hypothesis Testing** Lecture + Project Presentation (Only presentation skills are graded)
Week 13 (11 Apr)	**Regression Analysis (Not Tested)** Lecture + Project Presentation (Only presentation skills are graded) / Exercise

Note to students: If you have any questions, please always check this document first as it contains many details that you may be asking for. If you are unable to find the answer here, please approach or contact me.

FAQ

Seminar Registration

What if I want to change my seminar session?
Please contact the BBA office or ModReg as soon as possible. The instructor does not have access to the module registration system, so it will not help to email the instructor.

Class Participation

How do I showcase my solutions in the seminars?
You will be describing and explaining your solution (to the exercises or advanced exercises) by using your computer connected to the projector or by writing on the whiteboard. You may also respond to 1 to 2 questions from fellow classmates if there are any questions, which will result in more participation marks

How do I participate in discussions in the seminars?
There are many ways to participate in discussions. Examples are when an instructor ask for responses, responding to a question from your fellow classmate, giving a different response than the previous response for the same question, elaborating on a response that was not detailed enough, etc.

Project

Are the cover page, table of contents, references and appendices counted in the 8-page limit of the report?
The cover page, table of contents, references and appendices are not counted in the 8-page limit. Do not place your key findings and important diagrams in the appendices.

Can we place data visualizations in appendices?
Yes you can, but it is important to ensure that key data visualizations are placed in the main report so that graders can read and understand your report without referring to the appendices.

Are we supposed to include code in the project report or presentation?
Code is submitted in a separate file (refer to the appendices for details). Please avoid showing code in your project report or presentation.

Are there any requirements for the font, line spacing and margin of the report?
There is no specific requirement on the font, spacing, margin of the report. You could use any page layout as long as the report is clear and neat for reading (A4-size).

Is the peer evaluation optional or compulsory? Is it anonymous? What should I do if one or more of my teammates is not contributing to the project?
The peer evaluation is optional and anonymous. Only the instructor can see your peer evaluation. You can ignore it if you have nothing to report. If your teammates are not contributing, it is better to first communicate with them and solve the issue internally. If the issue cannot be settled internally and they still refuse to contribute, please keep all evidence and report in the peer evaluation. Your report will be taken into consideration for grading the project.

Final Exam

What are the topics to be tested in the final exam?
All topics covered in the lecture content and exercises during the seminars (except week 13 and advanced topics) could be tested.

Where can I find the time and venue for the final exam?
Please visit EduRec: https://myedurec.nus.edu.sg/psp/cs90prd/?cmd=login

Note to students: If you have any questions, please always check this document first as it contains many details that you may be asking for. If you are unable to find the answer here, please approach or contact me.

What are needed in the final exam?
- One double-sided A4-size cheat sheet. Other notes are prohibited.
- A laptop with ExamSoft installed. Other electronic devices are prohibited.
- Your student card and other relevant identity cards.

Appendix A: Important Dates of the Project

Task	Date	Consequence of missing the deadline
Deadline for emailing names of all the team members and the team name to the instructor. **The email should be titled "Project Group Details".**	17 Feb (Fri) 11:59 PM	You will be randomly assigned to a project team which may have fewer members.
Deadline of submitting the project.	14 Apr (Fri) 11:59 PM	20% penalty for late submission within 24 hours after the deadline. Submission more than 24 hours after the deadline will not be considered and receives zero mark.
Deadline of emailing project peer evaluation to the instructor (Optional). **The email should be titled "Project Peer Evaluation".**	21 Apr (Fri) 11:59 PM	Submission after the deadline will not be considered.

Appendix B: Project Guidelines

Project Submission

Each team must submit a zip file containing:
- A formal 8-page (4 pieces of double-sided paper) report saved as a PDF file (no code should appear in the formal report).
- Your Jupyter notebook source code saved as HTML files.
- Your project presentation Microsoft PowerPoint slides.

The zip file name must follow the format Project_<seminar session name>_<team name>. Please ensure that you follow this naming convention, otherwise there will be a 20% penalty on the project.

Any files submitted after the deadline are considered late submission. There will be a 20% penalty for late submission within 24 hours after the deadline. Submission more than 24 hours after the deadline will not be considered and receives zero mark.

Project Scope

The team project requires students to solve ONE practical business problem by exploiting real datasets. Possible sources of real datasets are: 1) kaggle.com ; 2) data.gov.sg ; 3) data.world. You can also download real datasets from other websites, or even conduct your own survey to collect data. A few examples are given below:
- Make property investment decisions based on property price data or Airbnb data.
- Use a dataset of movies/video games made in recent years to decide what kind of movies/games to make in the future.
- Use a dataset of employees in a company to investigate if there is discrimination in the company.
- Based on the demand data of a perishable product, use Monte Carlo simulation to find the optimal production quantity.

Note to students: If you have any questions, please always check this document first as it contains many details that you may be asking for. If you are unable to find the answer here, please approach or contact me.

You are also allowed to work on other directions rather than analysing real datasets. For example, you could create a dashboard for data visualization, or to develop your own Python package. For such project topics, please first contact the instructor for approval.

Project Components

The project score is made up of two components, an individual project presentation score (37.5%) and a team project report score (62.5%).

Guidelines and Grading Criteria for Project Report

The main focus of the project should be on data visualization, but you are allowed to use whatever programming/analytics techniques you know (not restricted to this module). Please do not write your project report as an assignment with separate tasks. You are expected to have a good flow for each part of your report, and they are all used in solving ONE problem.

The project report of the whole team will be evaluated with the guidelines below:

Components	Suggestions
Problem Statements (25%)	• A clear and concise description of only ONE business problem • Emphasize how the dataset would help to 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 support your analysis well and be efficient in delivering information
Analysis and Discussion of the results (25%)	• Derive relevant information from data visualization and use such information to solve the business problem • You may use statistical models (linear regression or any other models you know that is not restricted to this module), 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 • Include references for all literatures/external sources you refer to • Good English and free of error/typos

Guidelines for Project Presentation

For the presentation,

- Every team member needs to participate
- The attire is smart casual attire or business attire
- The audience should be able to see your slides and your face when you are presenting
- Ensure that the presenter's name and matriculation number are displayed on the bottom left of the slides the he/she is presenting, otherwise there will be a 20% penalty on the presenter's presentation.
- Use proper data visualization for your slides
- Focus on the business problem and make it as friendly as possible for audiences with limited programming/math/statistics backgrounds
- Presentation is graded individually, according to each student's presentation skills. Please do not read off any script; students will be penalized if they are reading off scripts, including scripts on laptop screens or mobile phones.

Note to students: If you have any questions, please always check this document first as it contains many details that you may be asking for. If you are unable to find the answer here, please approach or contact me.

References

This book may help you on your project report and presentation:
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

Appendix C: Academic Honesty and Plagiarism

Academic integrity and honesty is essential for the pursuit and acquisition of knowledge. The University and School expect every student to uphold academic integrity & honesty at all times. Academic dishonesty is any misrepresentation with the intent to deceive, or failure to acknowledge the source, or falsification of information, or inaccuracy of statements, or cheating at examinations/tests, or inappropriate use of resources.

Plagiarism is 'the practice of taking someone else's work or ideas and passing them off as one's own' (The New Oxford Dictionary of English). The University and School will not condone plagiarism. Students should adopt this rule: You have the obligation to make clear to the assessor which is your own work, and which is the work of others. Otherwise, your assessor is entitled to assume that everything being presented for assessment is being presented as entirely your own work. This is a minimum standard. In case of any doubts, you should consult your instructor.

Additional guidance is available at:
https://nus.edu.sg/osa/docs/default-source/osa-doc/resources-and-policies/code-of-student-conduct.pdf

Online Module on Plagiarism: http://emodule.nus.edu.sg/ac/

Note to students: If you have any questions, please always check this document first as it contains many details that you may be asking for. If you are unable to find the answer here, please approach or contact me.