



# ST5188 Advanced Data Science Project

## 2024/2025 Semester 2

Faculty of Science (Dept of Statistics & Data Science)

13<sup>th</sup> Jan 2025 – 10<sup>th</sup> May 2025

ST5188 @ Canvas: <https://canvas.nus.edu.sg/courses/72737>

© Copyright National University of Singapore. All Rights Reserved.

1

## Lecture Attendance & Participation

Join by QR code  
Scan with your camera app



### • For In-person and Zoom Live Stream Attendees

- You are expected to complete all activities during lecture time.
  - Activities: *Scan the QR codes* located in the top right corner of the respective slides.
- Your attendance will be tracked automatically.

### • For Recording Viewers (via Canvas' Panopto)

- You must watch the lecture from start to finish (→ tracking is done per second).
  - Activities: *Use the direct activity links* shared via Canvas Announcements and complete all activities by the end of Monday following the original lecture date.
    - Please note that QR codes do not work in lecture recordings at this time.

© Copyright National University of Singapore. All Rights Reserved.

2

2



3

## Content Overview

- ST5188 Facilitators
- **ST5188 Course Details & Course Guide**
- Organisational Matters (e.g., Leave of Absence)
- **Group Formation and Topic Selection**
- Consultation Sessions: Preparation Matters
- Q & A

5

## ST5188 Facilitators: Lecturer

Adjunct Assoc. Prof. **Markus Kirchberg**

- PhD (InfoSys) & MSc (CompSci)
- 20+ yrs Applied R&D (Data, Platforms, ML, AI)
  - Full innovation lifecycle mgmt. (ideation – research – incubation – production)
- 10+ yrs Payments, Fin/Reg/SupTech, Financial Crime
  - Incl. Head of Visa Labs, APAC
  - **Chief Executive & Technology Officer @ Wismut Labs**
- 10 yrs Entrepreneur
  - **Co-founder**, founding member, shareholder, advisor, ...
  - Singapore, Indonesia, US, and Europe
- 10<sup>th</sup>-yr **Adjunct @ NUS** (2010 – 2016 CompSci, since 2022 Stats & DS)
  - Primarily group project-based courses

© Copyright National University of Singapore. All Rights Reserved.



7

## ST5188 Facilitators: TAs

### Ms. **Anusri** Suresh Kumar

- 1<sup>st</sup>-year Masters by Research student
- Dept of Physics
- Research interests: Quantum computers and bosonic tomography
- Email: [a1353409@u.nus.edu](mailto:a1353409@u.nus.edu)

### Mr. **Pushkar** Kale

- 3<sup>rd</sup>-year Ph.D. student
- Dept of Stats & Data Science
- 2.5-years of industry experience
- Research interests: Bayesian inference & computation and machine learning
- Email: [pushkar.kale@u.nus.edu](mailto:pushkar.kale@u.nus.edu)

### Ms. **Shiqi** Wu

- 3<sup>rd</sup>-year Ph.D. student
- Dept of Mathematics
- Research interests: Machine learning techniques for scientific computing & model merging
- Email: [wushiqi@u.nus.edu](mailto:wushiqi@u.nus.edu)

### Mr. **Shen Tao**

- 3<sup>rd</sup>-year Ph.D. student
- Dept of Stats & Data Science
- Research interests: Decision making theory, causal inference, and statistical network analysis
- Email: [taoshen@u.nus.edu](mailto:taoshen@u.nus.edu)

All TAs are affiliated with the  
Faculty of Science, NUS

© Copyright National University of Singapore. All Rights Reserved.

8

8

# ST5188 Facilitator Contact Details

## Primary Point of Contact

- Each project group will be assigned to one of the TAs.
- TAs will delegate requests to the lecturer if / as necessary.  
→ Improves efficiencies / response time.
- **Expected Response Time:** Please allow up to two working days for a response to any queries or concerns.
  - If escalated, additional time might be needed.

## Secondary Point of Contact

- Lecturer
  - Consultation sessions
  - Feedback wrt. Project Proposal and Project Progress Report
  - Respond to enquiries delegated up by TAs
  - Compute support for NUS AWS resources
- Email direct when raising matters wrt. conflict resolution (should TAs not be able to assist in resolving the issues), code of conduct infringements, and any matters you feel need the lecturer's direct attention.
  - Email: [Markus.Kirchberg@nus.edu.sg](mailto:Markus.Kirchberg@nus.edu.sg)

© Copyright National University of Singapore. All Rights Reserved.

9

9

## ST5188 Advanced Data Science Project (AY 2024/25, Semester 2)

### – Course Guide –

Version 1.0

ST5188, which will be conducted in **e-hybrid mode**, is a **project course**. Throughout the semester, students will work (in groups) on their projects independently. However, there will be dedicated touch points with facilitators including **six lectures**, **five consultation sessions** (for each group) with the lecturer, and **ongoing TA support**.

Given that ST5188 is a project course, it is evaluated as such. Different students may contribute in different ways; according to their skills, abilities and project plan agreed upon by the whole group. However, **all students are expected to contribute similar efforts** to the project.

All the students in a group share **equal responsibility** for creating team spirit and making the group work as a whole. Should a problem arise, each student must be willing to work towards resolving the problem. **Do not hesitate to ask your assigned TA for mediation**; should problems persist, the TA will escalate the matter to the lecturer.

ST5188 assessment components are as follows:

	Contribution	Due Date	Late Submission (25% penalty applies)
Class Participation <sup>(i)</sup>	10%	n/a	n/a
Project Proposal <sup>(i)</sup>	15%	Feb 9 <sup>th</sup> , 11:59pm	Feb 11 <sup>th</sup> , 11:59pm
Project Progress Report <sup>(i)</sup>	10%	Mar 16 <sup>th</sup> , 11:59pm	Mar 18 <sup>th</sup> , 11:59pm
Project Presentation <sup>(i)</sup>	15%	Week 13	n/a
Final Project Report <sup>(i)</sup>	40%	Apr 20 <sup>th</sup> , 11:59pm	Apr 22 <sup>nd</sup> , 11:59pm
Code Reproducibility <sup>(i)</sup>	10%	Apr 22 <sup>nd</sup> , 11:59pm	n/a

<sup>(i)</sup> Individual assessment component; <sup>(ii)</sup> group-based assessment component

There will be **six MANDATORY lectures** (conducted in LT34 and streamed via Zoom).  
Lectures are held in weeks 1, 2, 3, 4, 5, and 7; Tuesdays 5–7pm.

For course details and updates, please refer to the [ST5188 Canvas page](#).

© Copyright National University of Singapore. All Rights Reserved.

1 | Page

Dept of Statistics and Data Science | Faculty of Science | National University of Singapore

Markus Kirchberg 10

10

# ST5188 Advanced Data Science Project [v3.0]

- **Project course** (with 6 supporting lectures).
- Students embark on **applied research projects** that emphasize practical applications / real-world solutions.
  - Topics originate from faculty recommendations, challenges presented by industry and research entities, or **student self-proposals**.
  - Navigate all aspects of a data science project, from proposal formulation to delivery.
  - [default] Students will work on projects in **groups of six**.
  - [industry projects] Smaller group sizes are permitted, but subject to approval by the lecturer.
- Modular Credits: 4
- Workload: 1-0-0-8-1
- Teaching Mode: E-hybrid
  - Students are allowed to complete ST5188 entirely online only if it is their final course required for graduation from the programme.
- Grading Basis: **Graded**
- **Learning Outcomes**
  - **Literature Mastery:** Systematically search and review literature on a specific topic.
  - **Critical Analysis:** Critically assess and evaluate academic papers & research findings.
  - **Project Formulation:** Develop and propose actionable projects based on literature insights.
  - **Technical Proficiency:** Demonstrate competency in coding using R, Python, or equivalent tools for data analysis and modelling.
  - **Project Planning:** Design and implement a strategic plan to guide the research project from initiation to completion.
  - **Collaboration and Teamwork:** Effectively work within teams, appreciating the value of diverse inputs and ensuring cohesive project progress.
  - **End-to-End Data Science Execution:** Navigate every stage of a data science project, ensuring integration of data, analysis, interpretation, and communication.
  - **Report Writing:** Compile and articulate findings in a well-structured, comprehensive report.
  - **Presentation Skills:** Effectively present research findings through oral presentations.
  - **Collaborative Discussion:** Engage in constructive discussions with peers and supervisors, showcasing an ability to both defend and adapt research perspectives based on feedback.

© Copyright National University of Singapore. All Rights Reserved.

11

11

## ST5188 Lectures

- Lecture 1: ***ST5188 Introductory Briefing Session***
- Lecture 2: ***Topic Selection and Foundations of Literature Review***
- Lecture 3: ***How to Write an Effective Capstone Project Proposal***
- Lecture 4: ***Project Planning, Management and Execution in Data Science***
- Lecture 5: ***Data Science Best Practises Part 1 (Data)***
- Lecture 6: ***Data Science Best Practises Part 2 (Model Development and Evaluation)***

All lectures are held Tuesdays 5 – 7pm (during weeks 1, 2, 3, 4, 5, and 7) and are conducted in LT34 / streamed via Zoom; recordings will be made available within 2 business days.

© Copyright National University of Singapore. All Rights Reserved.

13

13

## ST5188 Assessment & Grading

	Contribution	Due Date	Late Submission (25% penalty applies)
<b>Class Participation</b> <sup>(individual)</sup>	10%	not applicable	not applicable
<b>Project Proposal</b> <sup>(group)</sup>	15%	Feb 9 <sup>th</sup> , 11:59pm	Feb 11 <sup>th</sup> , 11:59pm
<b>Project Progress Report</b> <sup>(group)</sup>	10%	Mar 16 <sup>th</sup> , 11:59pm	Mar 18 <sup>th</sup> , 11:59pm
<b>Project Presentation</b> <sup>(group)</sup>	15%	Week 13	not applicable
<b>Final Project Report</b> <sup>(group)</sup>	40%	Apr 20 <sup>th</sup> , 11:59pm	Apr 22 <sup>nd</sup> , 11:59pm
<b>Code Reproducibility</b> <sup>(group)</sup>	10%	Apr 22 <sup>nd</sup> , 11:59pm	no

- Grading basis: **Graded**
  - Completing ST5188 with a Grade of at least C is a prerequisite for ST5290 Data Science Industry Project.

© Copyright National University of Singapore. All Rights Reserved.

14

14

## ST5188 Project Evaluation

- ST5188 is a project course → It is evaluated as such.
  - Different students may contribute in different ways; according to their skills, abilities and project plan **agreed upon by the whole group**.
  - However, all students are expected to **contribute similar efforts** to the project.
    - Peer group evaluation feedback (weeks 4, 7 & 14) → May be taken into consideration should there be significant differences in contribution by the group members.
- All the students in a group share **equal responsibility** for creating team spirit and making the group work as a whole.
  - Should a problem arise, each student must be willing to work towards resolving it.
    - Do not hesitate to ask your assigned TA for mediation.
    - Should problems persist, the TA will escalate the matter to the lecturer.

© Copyright National University of Singapore. All Rights Reserved.

15

15

## What to Expect: Example of Typical Activities

Each group of students will progress through three stages:

1. **Formalise a project proposal** (typically, weeks 2 to 4);
2. **Conduct basic research activities** (typically, weeks 5 to 13); and
3. **Present approach and findings** (typically, week 13).

Throughout the semester, project groups will get direct feedback on their project progression (e.g., via TAs, **consultation sessions**, ...).

© Copyright National University of Singapore. All Rights Reserved.

Week	Core Activities (Example)	Deliverables (Example)
1	<ul style="list-style-type: none"> <li>Read up on ST5188</li> <li>Attend lecture 1</li> <li>Form and register project group</li> <li>Brainstorm / review project ideas</li> </ul>	
2	<ul style="list-style-type: none"> <li>Attend lecture 2</li> <li>Explore project topics &amp; commence literature review</li> </ul>	
3	<ul style="list-style-type: none"> <li>Attend lecture 3</li> <li>Formulate project scope, problem statement, objectives, approach, success measure, ...</li> <li>Participate in the project topic / problem statement pitching session</li> </ul>	
4	<ul style="list-style-type: none"> <li>Attend lecture 4</li> <li>Schedule and attend first consultation session with lecturer</li> <li>Formulate project plan</li> <li>Finalise project proposal</li> </ul>	Peer Group Evaluation 1 Project Proposal
5	<ul style="list-style-type: none"> <li>Commence project work</li> <li>Attend lecture 5</li> </ul>	
6	<ul style="list-style-type: none"> <li>Continue project work</li> <li>Schedule and attend second consultation session with lecturer</li> </ul>	
7	<ul style="list-style-type: none"> <li>Continue project work</li> <li>Attend lecture 6</li> </ul>	Peer Group Evaluation 2
8	<ul style="list-style-type: none"> <li>Continue project work</li> <li>Schedule and attend third consultation session with lecturer</li> <li>Finalise project progress report</li> </ul>	Project Progress Report
9	<ul style="list-style-type: none"> <li>Continue project work</li> </ul>	
10	<ul style="list-style-type: none"> <li>Continue project work</li> <li>Schedule and attend fourth consultation session with lecturer</li> </ul>	
11	<ul style="list-style-type: none"> <li>Continue project work</li> </ul>	
12	<ul style="list-style-type: none"> <li>Continue project work</li> <li>Schedule and attend fifth consultation session with lecturer</li> </ul>	
13	<ul style="list-style-type: none"> <li>Present your project to lecturer, TAs, and peers</li> <li>Attend two peer project presentations and write peer project evaluations</li> <li>Finalise final project report</li> </ul>	Project Presentation Final Project Report
14	<ul style="list-style-type: none"> <li>Finalise project code to ensure reproducibility</li> <li>Complete project work</li> </ul>	Finalise Code @ GitHub Peer Group Evaluation 3

16

## KA 1: Class Participation (individual)

### What activities count towards class participation?

- **Lecture attendance** (→ 3.0 marks).
  - Join in person or online via Zoom or watch lecture recordings within 7 days.
- **Consultation session attendance** (→ 1.5 marks).
  - Mandatory sessions: weeks 3 – 4, 7 – 8, and 11 – 12.
- **Complete peer group evaluations** during weeks 4, 7, and 14 (→ 1.5 marks).
- **Complete peer project evaluations** after project presentations (→ 0.5 marks).
- **Log project work in JIRA** during weeks 7 – 13 (→ 3.5 marks).
  - Lecture 4 will provide more details wrt. tool support and corresponding expectations.

© Copyright National University of Singapore. All Rights Reserved.

17

17

## KA 2: Project Formulation & Project Proposal

### Guiding Expectations

- Apply current or emerging statistical / data science concepts, methods or techniques to an interesting application or real-world data set(s).
  - Learn **beyond** what was covered in prior coursework.
- Each project should include **some form of analysis** and **some form of experimentation** on real-world or synthetic data sets.
- **Support:**
  - **Your lecturer will double up as your group's consultant** and assist with brainstorming and decision-making activities.
  - **Each group will be assigned a teaching assistant (TA)** who should be your first point of contact wrt. any queries related to the project.
    - TAs will be able to escalate enquiries to the lecturer.

© Copyright National University of Singapore. All Rights Reserved.

### Core Activities

1. In-depth review / study of a statistical concept, method or technique covered by the MSc (Stats):
  - Statistical foundations of data science, applied regression analysis, design of experiments for product design and process improvements, nonparametric regression, analysis of time-series data, multivariate data analysis, sampling from finite populations, survival analysis, advanced categorical data analysis, advanced statistical methods in finance, stochastic processes and applications, statistical analysis of networks, spatial statistics, applied statistical learning, and deep learning in data analysis.
  - Recommendation: Initial **brainstorming of at least two potential project opportunities** (e.g., open questions, better model or algorithm, test on different data sets, reformulation / removal of assumptions, apply to different domain, etc.).
2. Proposal of a group project.

18

18

## KA 2: Project Formulation & Project Proposal (cont.)

### Project Proposal Content

Up to 6 pages (using the provided project proposal template):

- Project title (this can be a working title)
- **Project introduction and motivation** (incl. specific use case)
- **Problem statement or hypothesis**
- **Literature review / concepts study** (2-3 pages)
- **Project objective(s)**
- **Requirements** (in terms of data sets, tools, etc.)
- **Success measure(s)**
- **Project plan** (incl. key activities)

### Submission & Feedback

- **Project proposals are due by Feb 9<sup>th</sup>, 11:59pm.**
  - Highly recommended: Submit your project proposals early.
  - During week 4.
- **Feedback:** Will be given **6 – 10 working days** after the proposal has been received.
  - The later your submission is received, the later you will receive corresponding feedback.
- Project proposal submission: via [Canvas](#).
  - Submit one project proposal per group.

→ **Each group MUST attend their 1<sup>st</sup> consultation session prior to submitting the project proposal.**

© Copyright National University of Singapore. All Rights Reserved.

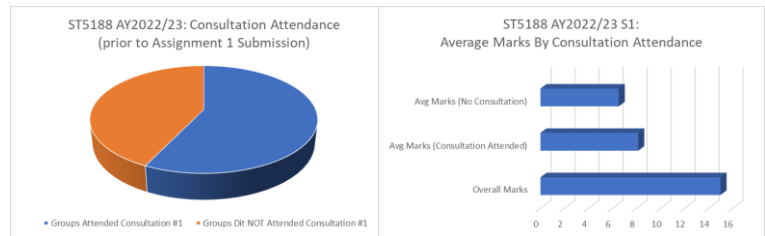
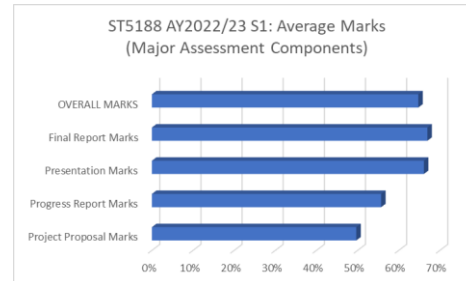
19

19



## Why Compulsory Consultation Sessions?

- Based on past marks, **assignments 1 (project proposal) & 2 (progress report) are the toughest** assessment component!
- Those students who attended a consultation session prior to submitting assignment 1 (project proposal) did **significantly better**.
  - This further amplifies for subsequent assessment components (overall 8 marks difference between those who attended all three consultation sessions and the rest).



© Copyright National University of Singapore. All Rights Reserved.

20

20

## KA 3: Project Progress Report

### Expectations & Content

- Your group should have completed ~50% of the work.
- Objectives: **Summarise your progress** and **discuss any challenges or difficulties encountered**.
  - Re-evaluate initial project proposal (based on feedback provided and challenges encountered) and present corresponding amendments.
- Content: (up to 5 pages):
  - Discussion of project progress:** Completed vs. pending tasks; challenges & difficulties; and exploratory data analysis.
  - Revised and refined project plan:** Weekly milestones and specific tasks for each group member.
  - Appendix:** Any modifications made to the initial problem statement, objectives, and/or success measures.

### Submission & Feedback

- Progress reports are due by Mar 16<sup>th</sup>, 11:59pm.**
  - Highly recommended: Submit your progress report early.
    - During week 8.
- Feedback: Will be given 6 – 10 working days** after the proposal has been received.
  - The later your submission is received, the later you will receive corresponding feedback.
- Project proposal submission: via [Canvas](#).
  - Submit one progress report per group.

© Copyright National University of Singapore. All Rights Reserved.

21

21

## KA 4: Project Presentation

- Week 13: 45-minutes presentation sessions
  - **25 mins group presentation followed by 15 mins of Q & A** (5 mins for set-up / contingencies).
  - All team members are required to attend.
  - Audience: members from two other project groups, one of the TAs, and the lecturer.
  - Venue, sign-up sheets, ... will be announced via the [ST5188 Canvas page](#).
- Objective: **Provide a compelling overview, highlighting key challenges, achievements, and engaging the audience's interest in your final report.**
  - Not expected to cover every detail of your project!
  - But: Since the audience may not be familiar with your project, begin with necessary context and background information.

© Copyright National University of Singapore. All Rights Reserved.

22

22

## KA 5: Final Project Report

- **A concise reflection of your group's project work**; example outline (up to 12 pages using the provided template):
  - Project title
  - Abstract
  - **Detailed problem description** (incl. context, problem statement, objectives, and assumptions)
  - **Related work**
  - **Methodology** (step-by-step process overview)
  - Description of **data collection and pre-processing processes**
  - Description of the **concept, method, model, algorithm, and/or technique that is/are at the core of your project**
  - **Evaluation of your experiments / results / findings** (including lessons learnt)
  - **Discussion of key challenges**
  - Recommendations for next steps or follow-on efforts
  - **Outline of each team member's contributions**
  - **List of references** (using either APA version 6 or 7, Chicago or IEEE formatted references).
- An appendix may cover additional content required for better understanding → Not considered during marking.
- Due date: Apr 20<sup>th</sup>, 11:59pm. Submission details will be made available via the [ST5188 Canvas page](#).

© Copyright National University of Singapore. All Rights Reserved.

23

23

## KA 6: Code Reproducibility

Ensuring your project code is easily replicable by others is a critical component of your capstone.

- **Repository & Submission**
  - Host all relevant source code on GitHub (details covered in Lecture 4).
  - Maintain a clear directory structure and informative commit messages.
- **Documentation & Replication Instructions**
  - Include a comprehensive README explaining environment setup, dependency installations, and execution steps.
  - Specify software versions, libraries, and configurations for consistent results.
  - Provide instructions on accessing or simulating data, respecting privacy and data-sharing agreements.
- **Referencing Code in Your Final Report**
  - In "Methodology" or "Appendix," briefly explain your code structure and link to your GitHub repo.
  - Key scripts / folders overview counts toward the 12-page limit; extensive code or environment files can go in an appendix.
- **Evaluation Criteria**
  - The quality, clarity, and completeness of documentation and repo organization are key.
  - Missing environment instructions or untracked data files will affect this grade.
- **Important Deadlines & Plagiarism**
  - Finalise your code repository by **Apr 22nd, 11:59pm**.
  - Properly attribute any borrowed or adapted code.

© Copyright National University of Singapore. All Rights Reserved.

24

24

## Organisational Matters ...

© Copyright National University of Singapore. All Rights Reserved.

25

25

## (Short-term) Leave of Absence

- Who is this for (primarily)?
  - Recently, China has implemented stringent measures regarding the overseas degree verification by the Chinese Service Center for Scholarly Exchange (CSCSE) for Chinese students returning to China for full-time jobs or PhD programmes after graduation.
- What has changed?
  - Until the end of AY23/24, NUS gave graduates retrospective permission for absences through a certification letter.
  - From AY24/25 onward, **students need to obtain permissions for absences beforehand** and not retrospectively.
- ST5188 is an anomaly → Students are allowed to take the ST5188 course entirely online.
  - BUT **only, if it is the last course required for graduation!**
  - In general, **NUS allows students to request overseas leave from courses for up to 2 weeks** with valid reasons.
- Point of contact for ST5188:
  - Your lecturer → [Markus.Kirchberg@nus.edu.sg](mailto:Markus.Kirchberg@nus.edu.sg)
  - Include evidence / valid reasons in your requests to avoid delays.
    - E.g., your unofficial transcript that you can generate on your own from myEduRec.

© Copyright National University of Singapore. All Rights Reserved.

26

26

## Student Conduct and Academic Integrity

### Code of Student Conduct

- All students have recently received an email from the Office of Student Conduct wrt. the updated [Code of Student Conduct](#).
- Key Points from the Code:
  - **Integrity & Honesty**
    - Uphold academic integrity in your coursework and assessments
    - Be truthful and transparent in all University-related matters
  - **Respect for Others**
    - Treat peers, staff, and all community members with respect
    - Maintain a safe and inclusive environment free from harassment or discrimination
  - **Responsible Behaviour**
    - Follow University policies and local laws
    - Use University resources responsibly (facilities, digital tools, etc.)
  - **Accountability**
    - Accept responsibility for your actions and decisions
    - Cooperate with investigations or inquiries conducted by the University

© Copyright National University of Singapore. All Rights Reserved.

### Academic Integrity

- **Plagiarism**
  - Plagiarism, including the uncredited use of work or ideas from others and from generative AI tools, is a serious academic offence.
  - Students must properly attribute all sources to avoid disciplinary action, as outlined in the updated [NUS Plagiarism Policy](#).
  - To prevent plagiarism, students should:
    - Seek guidance from TAs / faculty if unsure about correct citation practices.
    - Access [resources from NUS Libraries](#) on avoiding plagiarism.
    - Utilize the [Centre for Teaching, Learning and Technology's](#) services for detecting and preventing web-based plagiarism.
    - Refer to the Fundamentals of Academic Life e-package in the [Student Essentials e-course](#).
- **Cheating**
  - Cheating in any form is strictly prohibited.
  - Such offences are treated at least as "Moderate" in severity, with **the default penalty being a 'Fail' grade for the entire course**.

27

27

## Getting Started ...

© Copyright National University of Singapore. All Rights Reserved.

28

28

## Forming Project Groups (opens Jan 14th)

### Phase 1: Self-register

- Self-register via [Canvas](#) → ST5188 → People → Project Groups.
  - Option 1: Groups of 6 students.
    - Except for industry projects → seek approval.
    - First student to join a group becomes the group leader!
  - Option 2: Random group assignment
    - Join group "Request Random Assignment"
- **Self-registration closes on Monday (Jan 20<sup>th</sup>) at noon.**

### Phase 2: Randomly Assign / Fill

- On Tuesday (Jan 21<sup>st</sup>):
  - **Students who requested random assignment will be randomly assigned.**
  - **Any incomplete groups will be filled or merged.**
  - Any unassigned (non-participating) students will remain unassigned until they contact the lecturer.
  - **Each group will be assigned to one TA.**
    - Primary point of contact for all enquiries.

© Copyright National University of Singapore. All Rights Reserved.

29

29

# Guidelines for Forming Groups

- **Common Interests and Skills** → Form groups based on shared interests and expertise in specific areas of data science.
- **Diversity of Skills** → Ensure a mix of skills in each group, including statistical analysis, machine learning, programming, project management, domain expertise, and communication.
- **Group Size** → Maintain the default group size to balance workload and coordination.
- **Roles and Responsibilities** → Define clear roles based on individual strengths to cover all aspects of the project.
- **Communication Channels** → Decide on methods and frequency of communication. Schedule regular meetings for progress tracking.
- **Goals and Deadlines** → Set clear, achievable goals and internal deadlines to maintain project pace.
- **Conflict Resolution** → Establish a process for handling disagreements within the group.
- **Peer Review and Feedback** → Implement a system for regular feedback on each member's contributions.
- **Documentation and Record-Keeping** → Maintain detailed records of decisions, meetings, and project changes.
- **Respect and Inclusivity** → Ensure an environment where all ideas are valued, and everyone is treated with respect.

© Copyright National University of Singapore. All Rights Reserved.

30

30

## Group Composition (Example)

- Sample make-up of a balanced group:
  - Member 1 (group lead): Experienced in project management and organization.
  - Member 2: Strong in advanced statistical methods and analysis.
  - Member 3: Skilled in programming and data handling.
  - Member 4: Background in the specific industry or research area relevant to the project.
  - Member 5: Experienced in machine learning techniques and algorithms.
  - Member 6: Excellent communicator, adept at writing and presenting.
- Different skills do not imply rigid roles; members should share tasks to foster learning and collaboration.
- Ensure that your group can **handle the technical demands of a data science project, manage its workflow efficiently, and communicate its findings effectively.**
- **Motivation is Key:** Members should be genuinely interested and enthusiastic about the project.
- **Work Ethic:** Consistent effort, reliability in meeting deadlines, and proactiveness are essential.
- **Team Collaboration:** Respect for each other's contributions and effective collaboration enhances project success.
- **Positive Attitude:** Maintaining a positive approach, especially in facing challenges, is crucial.
- **Equal Contribution:** Fair and active participation from all members is necessary for balanced workload and success.

© Copyright National University of Singapore. All Rights Reserved.

31

31



# Industry Projects

## Qualifying Criteria

- An industry project is a project that requires either:
  - **Access to a data set not available in the public domain** (with the student having explicit permission to use the data set for ST5188); or
  - **Involves some commercially sensitive information that cannot easily be shared** (with the student having explicit permission to use such information for ST5188).
- If you plan to work on a project that falls into this category, *please contact your lecturer for approval ASAP.*

## Notables

- Most suitable for part-time students who plan to tackle a project sponsored by their current employer.
- Not suitable for internships!
- **Expression of interest for industry projects are due by Friday (Jan 17<sup>th</sup>), noon.**
  - Email: [Markus.Kirchberg@nus.edu.sg](mailto:Markus.Kirchberg@nus.edu.sg)

© Copyright National University of Singapore. All Rights Reserved.

33

33

# Tips for Choosing a Topic

## Brainstorm a Topic

- **Identify Interests and Strengths:** Have a brainstorming session to identify the common interests and strengths of the group members.
  - Projects that align with these are likely to be more engaging and successful.
- **Consult Faculty or Advisors.**
- **Consider Community Needs:** Look for projects that can benefit your local community or address a societal issue.
  - This not only provides practical experience but also contributes to social good.
- **Leverage Online Platforms** like GitHub, Kaggle, or academic journals to find ideas or datasets that could form the basis of a project.
- **Align with Career Goals:** Choose a project that aligns with the career goals of the group members.
  - This ensures that the project is not only an academic exercise but also a step towards professional development.
- **Seek Feedback:** Once a few ideas are shortlisted, seek feedback to gauge the viability and relevance of the ideas.
  - Discuss with your assigned TA and then during consultation session #1!

## Look for Inspiration

- Check out faculty-proposed topics / projects.
- Do you have a 'bucket-list' of things you want to learn / explore hands-on "when you have the time"?
- Explore AI / ML / Data Science newsletters, magazines, journals & blogs.
- Search online / social media for recent developments related to topics of your interest.
- If you are a data person, look for interesting data-sets ...

© Copyright National University of Singapore. All Rights Reserved.

34

34

# Project Selection Considerations

- **Scope and Complexity:** The project should be achievable within 9 weeks (→ be ambitious and challenge yourselves but stay realistic).
- **Skills and Expertise:** Align with what you, as a group, already know while also providing opportunities to learn and to apply new skills.
- **Resource Availability:** This includes data, software, hardware, and any other necessary tools or materials.
- Align with the course's **Learning Objectives**.
- **Interest and Engagement:** A project that resonates with your passion or career goals will likely result in higher motivation and better outcomes.
- **Real-World Application:** The project should have a real-world application or relevance (e.g., a problem faced by an industry, a research question, or a social issue) → This not only enhances learning but also adds value to your portfolios.
- **Advanced Statistical Techniques:** Utilize methods beyond basic descriptive statistics.
  - Examples: Inferential statistics, regression analysis, time series analysis, Bayesian methods, or machine learning algorithms.
- **Large or Complex Dataset**
  - **Criteria:** Choose a dataset that has more than 1 million records or is high-dimensional, with over 100 features or use multi-media data.
- **Multivariate Analysis:** Analyse and interpret relationships between multiple variables.
- **Validation of Analysis:** Rigorously validate the performance of any analysis.
  - **Methods:** Include data splitting (training and test sets), cross-validation, and appropriate performance metrics.
- **Reproducibility and Documentation:** Maintain a strong emphasis on the ability to reproduce results.
  - **Expectation:** Document code, models, and analysis thoroughly for reproducibility.

→ Projects should be challenging, relevant, and align with the advanced level of the course / degree!

© Copyright National University of Singapore. All Rights Reserved.

35

35

# Faculty-proposed Topics

## Topics

1. Application of autoencoders to identify fraud
  - Targeting emerging fraud patterns or synthetic identities.
2. Application of generative adversarial networks (GANs) to generate brain-MRI scan images for dementia patients
3. Wrist joint x-ray image classification for rheumatoid arthritis
4. Fake speech detection
  - Given a voice recording, determine the likelihood of the recording being genuine or fake (in parts or as a whole).

## Project Idea

- Topic: Counterfactual evaluation for fraud models
- Description: Frame counterfactual evaluation from a recommendation system context into the fraud space. The high-level idea is that instead of having to run multiple online experiments (100% each commonly in AB tests), random switching of policies is done in various percentages, such that we can still obtain an unbiased estimator of the metrics that we want to compute.
- References:
  - Learning from Logged Implicit Exploration Data; <https://arxiv.org/abs/1003.0120>
  - Large-scale Validation of Counterfactual Learning Methods: A Test-Bed; <https://arxiv.org/abs/1612.00367>
  - Unbiased Offline Evaluation of Contextual-bandit-based News Article Recommendation Algorithms; <https://arxiv.org/abs/1003.5956>

© Copyright National University of Singapore. All Rights Reserved.

36

36



# Industry-proposed Projects I/II

## Object Tracking for Edge Cases

- Client: Early-stage Startup
- Problem set 1: Most object tracking methods struggle with rapid and highly non-rigid motion as well as thin structures.
  - Example: <https://omnimotion.github.io/> → Failure cases
- Problem set 2: Off-road racing car tracking use case
  - Data acquisition and sensor modality challenges: Dust clouds reduce visibility, limit annotated data, and demand multi-sensor setups (camera, LiDAR, radar) that must adapt to varied terrain and weather conditions.
  - Occlusion and noise challenges: Vehicles can be fully or partially hidden by dust, causing sensor noise (blur, "ghost" detections) and making it difficult to maintain consistent, accurate tracking.

## YOHO (You Only Hear Once)

- Client: Early-stage Startup
- Inspiration: The YOLO (You Only Look Once) framework.
- Aim: Classify and localize specific sound events within the audio track of a given video.
- Core features:
  - Sound classification: Detect & classify 6-10 sound categories (e.g., speech, music, background noise, sound effects, silence, laughter, applause, nature/ animal/ vehicle sounds).
  - Event localization: Provide basic timestamp localization for detected events.
  - Confidence scoring.
  - Noise handling.
- Primary application: Media analysis.

© Copyright National University of Singapore. All Rights Reserved.

37

37

# Industry-proposed Projects II/II

## Re-identifying Objects (across space and/or time)

- Client: Early-stage Startup
- Context: Many systems, such as in surveillance, sports analysis, or autonomous vehicles, need to re-identify objects across different locations and times despite changes in lighting, perspective, or environment.
- Challenge: To ensure privacy, the system must avoid using PII, like faces or personal identifiers, and instead rely on non-PII features such as shape, colour, size, motion patterns, or other physical characteristics.
- Questions to consider:
  - How can the system accurately re-identify objects across different environments or time frames while accounting for changes in lighting, angle, or appearance?
  - What non-PII features are most reliable for consistent re-identification across space and time?
  - How can we balance the need for accuracy in identification with privacy concerns in environments like sports arenas or public spaces?

## Non-Intrusive Speech Quality Assessment (NISQA)

- Client: Seed-funded Startup
- Context: The NISQA framework, with its model for synthetic speech assessment, is a significant advancement in evaluating speech quality, particularly for synthesized or artificial voices like those produced by text-to-speech (TTS) systems.
  - Reference: <https://github.com/gabrielmittag/NISQA>
- Challenges:
  - Commercial usability: Many training datasets are restricted to R&D use only, limiting their commercial applicability. A solution should use datasets with licenses that allow commercial use.
  - Emotional & expressive quality: Current models focus mainly on clarity and intelligibility. Develop a module to assess emotional and expressive qualities in synthetic speech, crucial for virtual assistants, audiobooks, and entertainment.
  - Multilingual & dialect assessment: Extend evaluation to diverse languages and dialects, ensuring speech quality across under-resourced languages and varying dialects.
  - Noisy environment evaluation: Synthetic speech should be tested in noisy real-world environments. Develop a module to simulate and assess speech intelligibility and quality under various background noises.

© Copyright National University of Singapore. All Rights Reserved.

38

38

# Getting Feedback / Asking for Help

© Copyright National University of Singapore. All Rights Reserved.

39

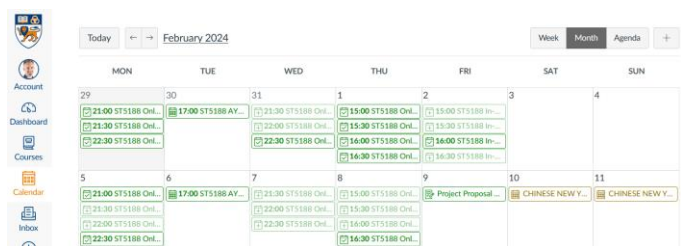
39

## Consultation Sessions: *Preparation Matters!*

- Each group can book **up to 5 consultation sessions** (30 mins each).
- Objectives:
  - Seek clarification / guidance / feedback to help with your project's progress or decisioning.
- What **not** to expect / ask:
  - The lecturer will not run those session; **you are expected to do so!**
  - We will not preview / pre-mark assignments / submission documents.
  - We will not make decisions for you!
- Consultations run from week 3 to week 12; booking opens on Jan 22<sup>nd</sup>.

© Copyright National University of Singapore. All Rights Reserved.

40



[in advance]

- Send a summary (limited to half a page / two paragraphs) of your project idea, project progress, key discussion points, ... **PRIOR to the consultation session to the lecturer.**

[during the session]

- Be on time
- Have a (written) agenda prepared
  - What do you want to ask or seek clarification / guidance / feedback on?
  - Provide sufficient context.
  - Prioritise your agenda items!
- Have supplementary information readily available
- Respect the meeting time

40

# Support Beyond ST5188 Project Work

## Office Hours (one-on-one)

- **Purpose**
  - Discuss individual **academic concerns, communication issues, or alternative learning needs.**
  - Note: Office hours are not meant for project-related discussions. Separate consultation sessions are provided for project work.
- **Preparation**
  - Note down specific questions or concerns beforehand.
  - Gather any relevant materials (e.g., notes, readings) you wish to review.
- **Scheduling**
  - Refer to the Canvas Calendar / Announcements for office hour timings.

© Copyright National University of Singapore. All Rights Reserved.

41

## FoS Support Services

- **Student Support**
  - The Student Life Team offers assistance with personal, family, or other concerns → Contact a [Student Support Manager](#).
  - [Peer Student Supporters](#) are fellow students who meet one-on-one to provide listening ears and general support (personal, relationship, academic, social).
- **Academic Support**
  - Speak to your tutors, academic staff or support staff if you face course-related challenges (e.g., workload, enjoyment, or difficulty) → Email: [asksci@nus.edu.sg](mailto:asksci@nus.edu.sg).
  - They can guide you on options such as module adjustments or a leave of absence.
- **Key Contacts**
  - **Student Support Managers:** [scisupport@nus.edu.sg](mailto:scisupport@nus.edu.sg)
  - **University Counselling Services:** [uhc\\_counselling@nus.edu.sg](mailto:uhc_counselling@nus.edu.sg) | 6516 2376
  - **Lifeline NUS (24-hr Hotline):** 6516 7777

41

## Questions?

© Copyright National University of Singapore. All Rights Reserved.

43

43

 **Thank You**

Markus Kirchberg



Markus.Kirchberg@nus.edu.sg



[www.markuskirchberg.net](http://www.markuskirchberg.net)



© Copyright National University of Singapore. All Rights Reserved.

44