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 [(i)]	10%	n/a	n/a
Project Proposal [(g)]	15%	Feb 9th, 11:59pm	Feb 11th, 11:59pm
Project Progress Report [(g)]	10%	Mar 16th, 11:59pm	Mar 18th, 11:59pm
Project Presentation [(g)]	15%	Week 13	n/a
Final Project Report [(g)]	40%	Apr 20th, 11:59pm	Apr 22nd, 11:59pm
Code Reproducibility [(g)]	10%	Apr 22nd, 11:59pm	n/a

[(i)] individual assessment component; [(g)] 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.

ST5188 Summary

Students embark on applied research projects that emphasize practical applications and real-world solutions. Topics originate from faculty recommendations, challenges presented by industry and research entities, or student self-proposals. Throughout this course, students will work in teams and navigate all aspects of a data science project, from proposal formulation to delivery. This comprehensive approach enhances competencies in literature review, critical assessment of research papers, project proposal formulation, project planning, collaboration and teamwork, adept problem-solving, comprehensive report creation, and articulate presentations. As a core component of the MSc (Statistics) program, this course is typically chosen during the final semester.

During the course, 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 their approach and findings (typically, week 13).

Throughout the semester, project groups will have the opportunity to get direct feedback on their project progression via consultation sessions.

Learning Outcomes

A student will acquire many of the following skills:

- Literature Mastery: Systematically search and comprehensively review literature on a specific topic.
- Critical Analysis: Critically assess and evaluate academic papers and 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.

Pre-Requisite

- Pass at least 16 Units with 2 ST level 5000 core courses (i.e., 2 of ST5201X, ST5202X, ST5209X, ST5211X, ST5215, and ST5223).

Modular Credit

4

Workload

1-0-0-8-1

Teaching Modes

E-hybrid mode.

This is a project course with six lectures during weeks 1, 2, 3, 4, 5, and 7. Throughout the semester, students will work on their projects independently with the lecturer doubling as consultant. Consultation sessions run from week 3 all the way through to the end of week 12. Those sessions give students direct access to the lecturer and are a good way to seek clarification, guidance and/or feedback on project ideas, project progression or any challenges that you are facing. Each project group may book up to 5 consultation sessions.

Students will be allowed to complete the ST5188 Advanced Data Science Project course entirely online ***only if it is their final course required for graduation from the programme****.*

Grading Basis

Graded.

What to Expect: Example of Typical Activities (week-by-week)

Students are given the freedom to choose their own project groups, project topics, and approach to work towards tackling the project. However, with such freedom comes greater responsibility, accountability, and communication requirements (with your peers and teaching staff). Below, you find a sample timeline of core activities and course deliverables.

Week	Core Activities (Example)	Deliverables (Example)
1	• Read up on ST5188 • Attend lecture 1 • Form and register project group • Brainstorm / review project ideas	
2	• Attend lecture 2 • Explore project topics • Commence literature review	
3	• Attend lecture 3 • Formulate project scope, problem statement, objectives, approach, and success measures • Attend 1st consultation session with lecturer • Vetting of alternative project ideas	
4	• Attend lecture 4 • Formulate project plan • Complete 1st peer group evaluation form • Finalise project proposal	• Peer Group Evaluation 1 • Project Proposal
5	• Commence project work • Attend lecture 5 • Attend 2nd consultation session with lecturer	
6	• Continue project work	
Recess Week		
7	• Continue project work • Attend lecture 6 • Attend 3rd consultation session with lecturer • Complete 2nd peer group evaluation form	• Peer Group Evaluation 2
8	• Continue project work • Write project progress report	• Project Progress Report
9	• Continue project work • Attend 4th consultation session with lecturer	
10	• Continue project work	
11	• Continue project work • Attend 5th consultation session with lecturer	
12	• Continue project work • Plan for project presentation	
13	• Present your project to lecturer, TAs, and peers • Attend two peer project presentations • Complete project work • Finalise final project report	• Project Presentation • Project Report Submission
14	• Finalise project code to ensure reproducibility • Complete 3rd peer group evaluation form	• Code Submission • Peer Group Evaluation 3
15, 16	Examination Weeks 1 & 2	

Getting in Touch / Asking for Help / Receiving Feedback

Throughout group project activities, your lecturer will double up as your group's consultant and assist with brainstorming and decision-making activities; **plan ahead to ensure that you use this opportunity wisely**. Consultation sessions (up to 30 minutes each) commence in week 3 and are available until the end of week 12 (each group may book up to 5 such consultation sessions). In addition, **each group will be assigned a teaching assistant (TA) who should be your first point of contact wrt. any queries related to the project or the course in general**. If and as necessary, TAs will escalate enquiries to the lecturer.

Besides consultation sessions and TA support, you will also receive detailed written feedback for your first two assignments, i.e., your project proposal as well as your progress report, and verbal feedback during the Q&A session of your group's presentation during week 13.

Recommendations for consultation sessions

You are recommended to arrange for up to five consultation sessions with your lecturer; **the first such session MUST be completed before submitting your project proposal**.

- You can either book an online consultation session (conducted via Zoom) or an in-person consultation session (held in Blk S16, 6 Science Drive 2, Singapore 117546); please choose the mode that is most appropriate for all your group members. A booking form will be made available via the **ST5188 Canvas page**.
- For online consultation sessions, group leads should ensure that the meeting link is circulated to all group members well in advance of the agreed upon consultation slot.
- Consultation sessions are meant for you to get feedback and/or inputs from the lecturer. As such, **you should have an agenda** on what topic(s) to discuss / question(s) to raise. Ideally, you will run consultation sessions; your lecturer will then add-on and keep an eye on the time.
- Since every group will work on a different project, you will need to set the context. For this, **we strongly suggest that you send a summary** (limited to half a page or two paragraphs) **of your project idea, project progress, key discussion points or else** (at least 3 hours) **PRIOR to the consultation session**. That way, the lecturer can have a pre-read, and you have the full 30 minutes of your session available for discussion.
    - Prior to your first consultation session, your group's summary should include your chosen problem statement(s) for each project idea. If you have multiple ideas, you may submit half a page or two paragraphs per project idea.
- Be ready a few minutes BEFORE the start of your session. If some or all your group members join / arrive late, you will lose valuable discussion time.

Key Activity 1: Class Participation

Class participation is an essential part of ST5188 and contributes 10% (i.e., 10 marks) to your final grade. **Below is the breakdown of activities and their respective weightage** (3.0 + 1.5 + 1.5 + 0.5 + 3.5 = 10 marks):

1. Lecture Attendance (→ 3.0 marks):
    - You must attend or watch **all six** mandatory lectures in person or via Zoom.
    - If you cannot attend live, you must watch the recorded lecture **within 7 days of the live session** to fulfil this requirement.
2. Consultation Session Attendance (→ 1.5 marks):
    - You are required to (arrange and) attend **all three** mandatory consultation sessions (these occur during weeks **3–4**, **7–8**, and **11–12**).
    - These sessions serve as valuable **check-ins for progress tracking** and **obtaining guidance** on your capstone project.
3. Peer Group Evaluations (→ 1.5 marks):
    - You will evaluate your teammates' performance and contributions, as well as reflect on your own.
    - Evaluations take place at three points in the semester (at the beginning of weeks **4**, **7**, and **14**).
    - **Constructive and honest feedback** is crucial; large disparities in team contributions may affect final assessment scores.
4. Peer Project Evaluations (→ 0.5 marks):
    - In week **13**, you must attend **at least two** other groups' project presentations.
    - Afterwards, you will **submit a feedback form** for each presentation you attend, by the specified deadline.
5. Log Project Work in JIRA (→ 3.5 marks):
    - From week **7** to week **13**, you must regularly document your **individual project tasks and progress** in JIRA.
    - Further details on JIRA usage, tool support, and expectations will be covered in **Lecture 4**.

Each of these components is vital for **ensuring active participation, accountability, and transparency** in both individual and group efforts. **Ensure all activities are completed within the stipulated timelines** to maximize your participation score.

Key Activity 2: Project Formulation & Project Proposal

ST5188 group projects will require you to apply current or emerging statistical / data science concepts, methods or techniques to an interesting application or real-world data sets. Each project should include some form of analysis and some form of experimentation on real-world or synthetic data sets.

Throughout group project activities, your lecturer will double up as your group's consultant and assist with brainstorming and decision-making activities. **Each group MUST schedule and attend their first consultation session with the lecturer prior to submitting the project proposal.**

The first key activity is comprised of two parts:

1) **In-depth study of a statistical / data science concept, method, or technique** that falls under the scope of the MSc (Statistics) by Coursework Programme:
    - 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.
2) **Proposal of a group project** related to your concept study from part 1).

Guidelines for part 1: Study of a statistical / data science concept, method, or technique

You are expected to learn **BEYOND** what was covered in prior coursework. You may think of it as a **literature review** comprising of a technical concept summary, concept critique (pros vs. cons, limitations wrt. assumptions made, practical feasibility or scalability of approach, etc.), and **initial brainstorming of at least two potential project opportunities** (e.g., open questions, better model, better algorithm, test on different data set, reformulation / removal of assumptions, apply to different domain, etc.). Based on this, you will then work on the second part, your group's project proposal.

Guidelines for part 2: Proposal of a group project

Your project proposal (up to 6 pages using the provided project proposal report template, which will be made available via the **ST5188 Canvas page**) should cover the following areas:

- Project title: Choose a working title that captures the essence of your project.
- **Project introduction / motivation**:
  - Provide an overview of your project, highlighting its significance and relevance.
  - Clearly articulate the motivation behind choosing this specific project.
- **Problem statement** or **hypothesis**:
  - Clearly define and explore the problem your project aims to address or state the hypothesis you intend to test.
  - This section should lay the foundation for the objectives and methodology of your project.
- **Literature review** / concept study (2-3 pages):
  - Conduct a thorough review of existing literature or studies relevant to your project.
  - This review should not only summarize key findings but also identify gaps or areas your project will explore.
- **Project objective(s)**:
  - Clearly outline the objectives or goals of your project.
  - Ensure these objectives are specific, measurable, achievable, and aligned with your problem statement or hypothesis.
- **Requirements**:
  - Outline the requirements necessary to execute your project, including:
    - Data sets: How they will be obtained or generated.
    - Compute resources: Specify the computational tools and resources needed.
    - Tools: List any specific software, platforms, or other tools required.
- **Success measures**:
  - Define what success looks like for your project in terms of evaluation, experimentation, and testing.
  - These measures should be quantifiable and directly linked to your objectives.
- **Project plan** including key activities:
  - Outline a project plan that includes key activities and milestones.
  - Take into account the size of your group and the time allotted for this project (i.e., 9 weeks).
  - This plan should demonstrate a feasible approach to achieving your project objectives within the given timeframe.

Remember, **your project proposal is not just a plan, but a persuasive document that should convincingly articulate the importance and feasibility of your project**. It should be clear, well-structured, and demonstrate a deep understanding of the subject matter.

Lectures 2 and 3 are designed to provide guidance for crafting a coherent and compelling project proposal; they offer valuable insights into differentiating a robust project proposal from a less effective one. **Please thoroughly review these lectures to ensure your proposal aligns with the outlined expectations and embodies the qualities of a strong submission.**

Project proposals are due by Feb 9th, 11:59pm. You are highly recommended to submit your project proposals early (i.e., during week 4). We aim to return detailed project proposal feedback six to ten working days after the respective proposal has been received. That is, the earlier your submission is received, the earlier you will receive corresponding feedback!

Project proposal report templates, evaluation rubric as well as submission details will be made available via the **ST5188 Canvas page**. You are only required to submit one project proposal per group.

Key Activity 3: Project Progress Report

As a rough guideline, **your group should have completed about 50% of the work by the time you submit your progress report**. That being said, there may be circumstances under which this is not feasible (e.g., in case your group decided to significantly revise the project's focus or objectives). However, such circumstances should be made known to your assigned TA / lecturer as soon as they arise and not only during progress report submission. Thus, **please make sure that any significant project proposal amendments have been discussed during a consultation session prior to project progress report submission**.

In the project progress report, you should provide a summary of your progress thus far (i.e., demonstrate that you have completed about 50% of the work) as well as a discussion of up to three unforeseen challenges or difficulties encountered. This report gives each group also a chance to re-evaluate initial problem statements and objectives (espc. for those groups who have been recommended to revise such sections in the project proposal feedback) and reassess and significantly refine their project scope and plan. The revised project plan should be detailed and precise, breaking down the remaining project timeline into weekly segments with clearly assigned tasks, ensuring that it aligns with the updated objectives and effectively tackles any unresolved challenges encountered.

Your project progress report (up to 5 pages – using the provided progress report template, which will be made available via the **ST5188 Canvas page**) should cover the following areas:

- Project title: Clearly state the title of your project at the beginning of the report.
- **Discussion of project progress**:
    - **Completed vs. pending tasks**: Provide a detailed demonstration (not just a mention) of the project components that have been completed and those that are still pending.
    - **Challenges and difficulties**: Highlight any difficulties or challenges encountered, especially focusing on data-related issues such as pre-processing. Describe how these challenges have impacted the project and any steps taken to address them.
    - **Exploratory data analysis**: Strongly consider including an exploratory analysis of your data set(s) to demonstrate their suitability for your project. This also forms a good baseline for the discussion of difficulties or challenges encountered (as they are often data pre-processing related in the early stages of project work).
- **Revised and refined project plan**:
    - Outline a detailed plan for the remaining work items (i.e., covering weeks 9 to 13). This plan should include weekly milestones and specific tasks to be completed by each group member, ensuring a clear path to project completion.
- Appendix (not subject to the page limit):
    - Any modifications made to the initial problem statement, objectives, and/or success measures since the project's inception should be thoroughly documented and included as an Appendix.

Lectures 4 and 5 provide guidance for writing an effective progress report; they not only guide you in effectively demonstrating your project's progress through practical examples but also outline the expectations for a meticulously formulated project plan. **Please thoroughly review these lectures to ensure your progress report aligns with the outlined expectations and embodies the qualities of a strong submission.**

Project progress reports are due by Mar 16th, 11:59pm. You are highly recommended to submit your project progress reports early (i.e., during week 8). We aim to return project progress report feedback six to ten working days after the respective report has been received. That is, the earlier your submission is received, the earlier you will receive corresponding feedback!

As a general guideline, you are encouraged to have completed your second consultation session with the lecturer prior to submitting your progress report.

Project progress report templates, evaluation rubric as well as submission details will be made available via the **ST5188 Canvas page**. You are only required to submit one project progress report per group.

Key Activity 4: Project Presentation

During week 13, each project group will be assigned a **45-minute slot for their project presentation**, with specifics regarding the venue and sign-up process to be announced via the **ST5188 Canvas page**. **Attendance by all group members is mandatory.**

- **Presentation format:**
    - Duration: Allocate approximately 25 minutes for the presentation, followed by 15 minutes for a Q&A session. Note that 5 minutes are reserved for setup and contingencies.
    - Content: There is no strict format for the presentation. You may use slides, walk through your project deliverables, or choose another presentation style. **The goal is not to cover every detail of your project but to provide a compelling overview, highlighting key challenges, achievements, and engaging the audience's interest in your final report.**
    - Timing: Be mindful of the time limit. We will not interrupt during the 25-minute presentation, but if you exceed this time by more than 5 minutes, we will have to move on to the Q&A session.
- **Audience considerations:**
    - Composition: Your presentation will be attended by members from (at least) two other project groups, one of the TAs (not necessarily your assigned TA), and the lecturer.
    - Context: Since the audience may not be familiar with your project, begin with necessary context and background information.
- **Q&A session:**
    - Duration: A 15-minute Q&A session will follow your presentation.
    - Participation: Questions may be asked by any audience member.

This presentation is a valuable opportunity to showcase the essence of your project and engage with an audience that includes your peers and faculty. **It is important to balance the technical details with a clear, accessible narrative that captures the significance and impact of your work.**

Key Activity 5: Final Project Report

The final project report **must not exceed 12 pages** (using the provided final report template, which will be made available via the **ST5188 Canvas page**). It should be **a self-contained and concise reflection of your group's project work**; content should cover the following aspects:

- **Project title:** Clearly state the title of your project.
- **Abstract:** Provide a succinct summary of your project, including the key objectives, methodology, results, and conclusions.
- **Detailed problem description:**
  - Context: Set the stage for your project, explaining the background and relevance.
  - Problem statement: Define the specific problem your project addresses.
  - Objectives: List the goals you aimed to achieve.
  - Assumptions: Mention any assumptions made during the project.
- **Related work:** Discuss relevant literature and how your work relates to or differs from these existing studies or projects.
- **Methodology (step-by-step process overview):**
  - Describe your methodology in detail, ensuring it is understandable for both technical and non-technical audiences.
- **Data collection and pre-processing:**
  - Detail the processes of data collection and any pre-processing steps taken.
- **Core concept / method / model:**
  - Describe the central concept, method, model, algorithm, or technique your project is based on.
- **Evaluation of results / findings:**
  - Present an analysis of your experiments, results, or findings, including lessons learned.
- **Discussion of key challenges:**
  - Address the major challenges faced during the project and how they were managed or overcome.
- **Recommendations for future work:**
  - Suggest next steps or potential follow-on efforts based on your project's outcomes.
- **Group member contributions:**
  - Outline each group member's specific contributions to the project.
- **List of references:**
  - Compile references using either APA (version 6 or 7), Chicago, or IEEE formatting. While the reference list can extend beyond page 12, ensure the first reference appears on page 12.
- Appendix and supplementary details:
  - You may include supplementary details, such as an extended description of dataset variables or larger versions of tables / figures, in the appendix. This section is not subject to the 12-page limit.

Your final report should **stand on its own**, meaning it should be comprehensible and complete without reliance on previous submissions. It is crucial to **present your information clearly, coherently, and in a manner that demonstrates the depth and scope of your project work**.

Final project reports are due by Apr 20th, 11:59pm. Submission details will be made available via the **ST5188 Canvas page**. You are only required to submit one project report per group.

Note: **Final report submissions will be examined for plagiarism!** You are strongly encouraged to self-inspect your final report documents prior to submission. As a general guidance, your final report's similarity score must be below 20% overall with no individual paper contributing more than 2% to the overall score.

Key Activity 6: Code Reproducibility

Ensuring your project code is easily replicable by others is a critical component of your capstone work. This activity will be evaluated separately from your written report, but it **must** be referenced in your **Final Project Report** (Key Activity 5).

1. **Repository & Submission**
    - Host all relevant source code on **GitHub** (details on repository creation and management will be covered in **Lecture 4**).
    - Maintain a clear directory structure and informative commit messages.
2. **Documentation & Replication Instructions**
    - Include a comprehensive **README** or similar document **within your repository** explaining how to set up the environment, install dependencies, and run your code.
    - Specify **software versions**, **libraries**, and **configurations** needed to ensure consistent results.
    - Provide instructions for accessing or simulating any data used in your project, while respecting data sharing agreements and privacy concerns.
3. **Referencing Code in Your Final Report**
    - In the "**Methodology**" and/or "**Appendix and supplementary details**" sections of your **Final Project Report**, include a brief explanation of your code structure and a link to your GitHub repository.
    - You may provide a short overview of key scripts or folders in the main body of the report (counted toward the 12-page limit), while additional details — such as extensive code samples or environment files — can be placed in an appendix that does not count toward the page limit.
4. **Evaluation Criteria & Weighting**
    - This activity accounts for **10%** of your overall grade. The **quality**, **clarity**, and **completeness** of your documentation and repository organization are key.
    - Inconsistencies or omissions (e.g., missing environment instructions, untracked data files) will impact this portion of your grade.
5. **Important Deadlines & Plagiarism**
    - Your code repository must be finalized by **Apr 22nd, 11:59pm**.
    - While your report will be subjected to plagiarism checks, ensure that any borrowed or adapted code is properly attributed and cited where applicable.