The project method that will be used is CRISP-DM at the first stage of data analytics and will add Scrum together with it when software development process is involved.

The final project outcome is to develop and test a web app data visualization dashboard and a web app that deploys a simple machine learning algorithm which tackle the gender pay gap problem based on analysis on the gender pay gap dataset or different size companies from UK Government (GOV.UK, undated). As the project requires both data analytics and software development processes, the selection of the project management methodology started by looking at the most popular three frameworks combining both data science and software development according to a poll (Saltz, 2022) from Data Science Process Alliance, which are CRISP-DM, Scrum and Kanban respectively. Then the three methodologies were compared and contrasted in parallel with criteria below:

1. Level of volatility in the requirement — expected to change frequently as no specific requirement is given at the start and no specific client / demand group;
2. Expertise on software engineering and data science process methodologies — exposed to scrum before but limited knowledges on others;
3. Time frame of the project — fixed deadline and each section of the project for an outcome to be delivered will take about 4-5 weeks;
4. Ease of access to the client — minimal to no access or interaction as there is no specific client;
5. Experience of the team in working together — all team members are quite familiar with each other and used to work together for several times.

The evaluation process can be shown by the table below. By looking at whether each kind of method is suitable to conduct the project with respect to each selected criteria, it can be concluded that the most suitable methodology is CRISP-DM for data science and Scrum should be added in the later stage of software development.

Methods: Criteria:	CRISP-DM	Scrum	Kanban
Level of volatility in the requirement — expected to change frequently as no specific requirement is given at the start and no specific client / demand group	As there is no specific client or requirement for the project, CRISP-DM becomes a good method choice as it involves the process of business and data understanding at first. It also involves the data preparation process so that the team may generate many insights on which direction to go given a broad topic from the progressive process. There are also iterations included in this method so that the team is able to review and change the plan. (think insights.net, undated)	Scrum becomes a suitable tool to use as there is no specific requirements, where the project goals will be set by team members. The breakdown of the big project into small Sprints has made it adjustable for the project goal after each Sprint. The team is able to be flexible to change after testing of each Sprint. However, as there is no specific client, it becomes hard to test.	Kanban fits the flexible project as the team is able to add or delete work-in-process at any time from the board. However, with the high flexibility and extremely high volatility in the requirement, the team may find it difficult to have a clear direction and change a lot, which makes the project more complex. This drawback also reflects that Kanban is a service-oriented approach, which needs the requirement to be clarified at first.

Expertise on software engineering and data science process methodologies — exposed to scrum before but limited knowledges on others	CRISP-DM is a friendly methodology to use for beginners in data science as it follows the natural and typical logics that someone should think when looking at a data science problem. Meanwhile, the clear framework has made it easy to follow and its popularity has already proven it can be a solid tool in some way. It is easy to learn by reading documentations without extensive training.	Scrum is the most familiar tool for all team members as it was the only tool that team members were exposed before in other projects. Meanwhile, it is easy to find loads of documents describing Scrum available online as a popular and wide used methodology, which makes it easy to check when there is any concern.	It is quite easy to learn the process of Kanban method as it is simply putting every to-do task on the board to the position regarding its completeness. However, without experience of doing such project, it becomes difficult to figure out what need to be put on, thus it might be normal for bottlenecks to appear and delay the whole project process.
Time frame of the project — fixed deadline and each section of the project for an outcome to be delivered will take about 4-5 weeks	It is easy to follow the steps in CRISP-DM, which minimize the time cost of learning the methodology process. Meanwhile, by first understanding the business needs and dataset, it provides a clear direction for the team to pursue. In addition, it involves iterations between processes, which allows quick corrections when the problem is found. Therefore, CRISP-DM is a good method in such small projects.	Scrum has claimed to be a framework that reduces project time significantly as it has not burdensome documentation and quick prototyping. (Guadarrama, 2019) Meanwhile, it divides a big project into a series of small projects called 'sprints', which fits the timeline of the coursework requirements as in a fixed time a deliverable should be created.	Kanban provides a clear visualization on the project process, thus enables the team to concentrate on unfinished tasks. It also limits the maximum work-in-process and can dynamically show the current bottleneck of the project, so that can be helpful in a short time constrained project. However, without sufficient experience, it may be difficult to choose the most suitable to-do list in Kanban and bottlenecks may appear frequently, thus delay the project.
Ease of access to the client — minimal to no access or interaction as there is no specific client	It is not a problem for teams adopting CRISP-DM method when there is no access to or interaction with the clients as the whole process relies on the first understanding by the team. The feedbacks from clients are not mandatory in this case.	After each Sprint, feedbacks and reflections are required for improvements on the next Sprint. However, as there is no specific client, such feedbacks are nearly impossible to obtain. However, with the PGTA session, the TA can be assumed as a consumer to test the outcome of each Sprint.	Kanban aims at developing a service-oriented approach. (kanbanize.com, undated) It requires to profoundly understand the customers' needs at first and regularly managing the network of services. Therefore, as there is minimal access to the client, Kanban seems not to be a good choice in this aspect.

Experience of the team in working together— all team members are quite familiar with each other and used to work together for several times	CRISP-DM does not include formal roles. As a team formed up with friends, it may not be a problem as tacit understandings exist. However, too much familiarity in personal life may cause problems as people tend not to point out mistakes made by others. Nevertheless, CRISP-DM provides a retrospective phase and can iterate when problems being found together.	Scrum method requires to have a Scrum Master which facilitates the Scrum process as a leader. This role will have formal power to monitor the whole process to push team members working. (Hotz, 2022) Meanwhile, all team members are restricted by the team charter. Therefore, it minimizes the risk of sluggish process due to familiarity between team members.	There is no formal roles in Kanban methodology. Though team members are quite familiar with each other and had previous collaboration experiences, all team members are not familiar with software development. Without a clear individual responsibility, the team may be sluggish as everyone is too familiar with each other to point out problems.

By looking at the table above, it is clear that CRISP-DM has absolute advantage among three methodologies in data science projects. However, as the project also involves software development process, a pure data science method like CRISP-DM will not be enough. Therefore, we plan to adopt CRISP-DM method in the beginning data science part of the project to help better understanding the problem and data. Then we will add Scrum on the software development phase to minimize adverse variances and maximize beneficial opportunities. However, there are also potential challenges of the two methods need to be considered. To begin with, the CRISP-DM method may face problems of insufficient understandings at the first stage. In this case, it requires us to do more research on the project topic and be more careful when dealing with the dataset. Meanwhile, it is difficult to test the proposed solution without a target group. The team may ignore the drawbacks made on ourselves. In addition, all of the team members are expected to be extremely busy at the time Scrum needs to be added, whereas Scrum requires meeting overhead. It is necessary to think about how to make meetings succinct and maintain effective and efficient discussions at meetings.

References:

GOV.UK (undated) *Download gender pay gap data.* GOV.UK Available at: https://gender-pay-gap.service.gov.uk/viewing/download [Accessed: Oct 2022]

Guadarrama, C. (Sep, 2019) *How to develop a product with unclear requirements.* Medium. Available at: https://cesarguadarrama.medium.com/how-to-develop-a-new-type-of-product-with-unclear-requirements-491027684a72 [Accessed: Oct 2022]

Hotz, N. (June, 2022) *Scrum for Data Science*. Data Science Process Alliance. Available at: https://www.datascience-pm.com/scrum/ [Accessed: Oct 2022]

kanbanize.com (undated) *What Is Kanban? Explained for Beginners.* Kanbanize. Available at: https://kanbanize.com/kanban-resources/getting-started/what-is-kanban [Accessed: Oct 2022]

Saltz, J. (May, 2022) *CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects*. Data Science Process Alliance. Available at: https://www.datascience-pm.com/crisp-dm-still-most-popular/ [Accessed: Oct 2022]

thinkinsights.net (undated) *CRISP-DM – A framework for Data Mining & Analysis*. Think Insights. Available at: https://thinkinsights.net/data-literacy/crisp-dm/ [Accessed: Oct 2022]