Methodology Selection

In this project, we intend to use the offline data of China's population distribution to develop a website application. This application aims to draw an interactive map of population distribution based on China's population data. I will first analyse three different project management methods in this report: CRISP-DM, Scrum and Waterfall.
According to four different selection criteria, a criterion table has been created as shown below:

Criterion table

Process Model/ Criterion	CRISP-DM	Scrum	Waterfall
Expected to change frequently	The method is implemented in stages. Making changes at any stage along the way is relatively expensive. By accepting that a project starts with significant unknowns, the user can cycle through steps, each time gaining a deeper understanding of the data and the problem.[1]	The main concept of the method is to build MVP (Minimum Viable Product), which means it is easily to be changed during the process. The team is constantly refining the model as they collaborate. Even overturning the initial model is less expensive than other methods.	Similar to CRISP-DM, the prerequisite of utilising waterfall model is that the clients seem unlikely to change the scope of the project once it is underway.
Ease of access to the client/Interaction with management and customers	CRISP-DM lacks communication strategies with stakeholders. So be sure to set expectations and communicate with them frequently.	Scrum focuses on delivering customer value. It structures development efforts from the customer's perspective and encourages stakeholders to provide continuous feedback through practices such as sprint reviews.	Customers often don't fully know what they want on the front end, and when it's more difficult for them to adapt, they make requests for changes and new features later in the process. The client is not involved in the design and implementation phases.[3]
Timeframe of the project	The timeframe of this method depends on the initial plan at the beginning. It depends on the needs of clients. Since it generally does not change halfway through, its timeframe is generally relatively fixed	The time it takes to implement a solution for most data science problems is ambiguous and depends on unknown factors such as data access, questionable data quality, or an untested ability to distinguish signal from noise.	The timeframe for utilising this method is relatively short compared to other methods, since it doesn't allow clients to make changes on their needs after beginning of the project.
Ease of learning the method/availability of document and support	Compared to other methods, this method is quite feasible. It can be implemented without much training, organizational role changes, or controversy.	According to user feedback, this method is difficult to learn. Teams working on data science projects in controlled experiments performed the worst with Scrum, mainly because of their inability to understand the methodology and create clear sprints.[2]	This approach does not require team members to spend too much time learning but requires managers to do their best to specify the work of each team member.

It is not difficult to see from the above analysis that both CRISP-DM and Waterfall need to clarify the purpose and requirements of the project at the very beginning of the project. We need to strictly follow the plan we made at the beginning step by step. The Scrum method has great flexibility. It needs to establish an MVP at the beginning stage, and continuously improve the project based on this prototype. In this project, we need to draw an interactive population distribution map based on the Chinese population to help the government better formulate medical policies. In this process, we cannot determine which factor (such as urbanization, male-to-female

ratio, geographical area) to analyze before the project starts. I personally insist that scrum is more suitable for this project. Using this method allows us to continuously mine data in multiple dimensions during the development of applications, so that the established population model can better help customers judge medical needs.

Reference

[1] https://www.datascience-pm.com/crisp-dm-2

[2] https://www.datascience-pm.com/scrum/

[3] https://www.datascience-pm.com/waterfall/