Reflective Essay

for

Research Practicum

Version 1.0

Prepared by Adam Ryan (14395076)

COMP47360

August 27, 2021

1 Reflections on Project Management, Technical Growth, and Teamwork

Coming into the research practicum I had prior professional experience developing a Django-based web application for a financial company, and in working as a customer analyst with a focus on data engineering for a leading luxury retailer. From these experiences, combined with my experience in I had a grounding in many of the aspects which I believed the project would touch upon particularly in the areas of software demonstrations, the importance of data architecture, and key project management principles for the delivery of a successful project. Even with this grounding, over the course of the project I feel I gained significantly in learning how to successfully manage software development projects from the scoping of requirements through to deployment, learning new technical and professional skills in the areas of data analytics and presenting by pulling together techniques from across the course, and learning techniques to effectively communicate with a team while working remotely towards a shared development goal.

A key area of growth over the course of the project was in learning how to effectively leverage a Scrum methodology to successfully deliver a project from a brief. When the project initially began, I believed the overall task presented to us was daunting in scope. Within the initial days of the project, I found it overwhelming how many different data sources were available to us of a varied quality and size, how many technical options were present in regards to how we could select a technical stack to build the application, and the lack of a 'focused' direction in building the application. In the initial week of the project, I felt there was infrequent communication and a poorly defined division of tasks which magnified the sensation that the project scope was significant. In hindsight, I believe these early days which to me felt somewhat aimless played a very important role in emphasising the importance of the project management skills which were initially touched upon in the usage of Agile methodology in COMP30830. This initial feeling was compoounded by the lack of daily standups in our initial sprint which made it challenging to have a unified view of what each team member was working on and in what areas progress was being made. Following this time period, the importance of establishing a Jira and Confluence board became very clear to clearly define what each project member would be working on over the course of each sprint. Spending the initial sprint as a discovery phase proved to be ultimately valuable in my opinion as I felt defining early requirements and user journeys as had been encouraged in COMP30830 helped to provide a much more focused direction to the project, and balancing sprints around the project presentations played a crucial role in guiding the aim of each sprint and making what initialy felt like a daunting project into a manageable and achievable goal. Having previously completed projects in a professional context using a Waterfall methodology, I found the Agile process was a much stronger method in continuously seeing notable progression as the project continued, particularly as we saw Epics become completed.

While I believe this division of the project into sprints and sub-dividing tasks into Epics and Tasks aided in the overall completion of the project and helping to alleviate worries over the scope of the project, in hindsight, one aspect of this division which I believe could have been improved upon was in more strongly encouraging deadlines and time estimates for the delivery of certain features. Initially we had scoped a variety of features which we intended to incorporate into the web application including Google Analytics integration, an administrator dashboard, and a more fully-developed points feature. I believe if we had more thoroughly researched how long we anticipated certain features to take in development, and to have prioritised more impactful features earlier we could have potentially delivered a more fully rounded points system and potentially minimised our final feature backlog. One aspect which played a particular role towards the end of the project was that as certain features were dependant on others, and as we had specialised into certain areas, there were times where some members of the team were waiting on development in areas before they could progress. This element in particular highlighted the importance of carefully aligning development tasks to allow each team member to sustain a consistent level of productivity over the course of the project. In the later stages, one of our team members was unavailable at times due to mental health reasons. Now that the project is over I believe if we had placed a greater emphasis on feature prioritisation early-on we could potentially have handled a hit in productivity better and rebalanced the workload to remove some of the burden on the affected team member and maintain productivity; the degree of specialisation which we engaged in also created a challenge in alleviating dependencies on certain team members if they were unavailable which leads me to believe that while specialisation and taking ownership of an area is important in managing a project's development, it is important when on a small team for everybody on the team to have a broad understanding of each component in the project so that if a problem occurs it is possible to support another area to a greater extent.

As the project reached the final sprint, the usage of Jira and Confluence declined and a greater emphasis was placed on real-time communication in Discord. While this channel of communication was useful in helping to complete tasks immediately, I found this aspect again made it difficult to gain a complete understanding of the progress of the project as a whole, and it posed challenging in keeping track of how certain tasks were progressing. As I was trying to balance the project with employment, I found live chats were much more difficult to refer against later on, and this difficulty again re-affirmed the important role which Jira, Confluence, and similar project management tools play in helping members of a development team not only work towards their shared goal but also allows team members to more easily balance both internal and external commitments by providing a single and easily structured point of reference.

In completing the research practicum not only did I gain an appreciation for Agile processes, the importance of splitting a project into its constitutent tasks, and understanding the advantages which project management tools provide to support the usage of real-time communication tools, but I also grew in my technical ability as the project entailed technical challenges which I had not yet encountered in the course to date. The variety of data sources and difficulty in analysing and interpretting them, the usage of new technologies like Azure and Docker which had not previously been seen in modules prior to this project, and researching the optimal technology to use during the project all presented new avenues of technical development.

At the beginning of the project, one of the biggest decisions which initially faced the team was deciding what tools to use; this was something very much unique to the research practicum as most modules which we encountered enforced the usage of certain technologies while this was a unique opportunity to decide upon this as a team. Having previously worked with Django and React in a professional context, and having used Flask in COMP30830, I was initially hoping the team would use Django as I personally found Django to be much more suitable to the style of the project, being an MSc project where I believed the development should follow best industry practices which Django more-heavily reinforces and the stack of Django and React being very common within consulting projects I encountered while working in EY, however after discussing this with the team and learning of the team's preferences we ultimately settled on Flask and Vue due to the lightweight nature of this technology stack. While I still maintain that Django is a much more professional framework better suited to an MSc project, and while I would have used this back-end framework if we had the opportunity to complete the project individually, as the development of the application began it was true that the flexibility of Flask and its lightweight structure allowed for the team as a whole to quickly build the app during its infancy rather than becoming burdened with learning the intricacies of Django. The choice of whether to use Azure, Amazon Web Services, or a local MySQL database was another early choice which the team had debated. While I found Amazon Web Services very developer-friendly during the COMP30830 module, our student accounts had almost depleted their credits and as such we had to balance the familiarity we had with the service against the risk of a potential cost and we ultimately decided to seek alternatives. As we worried about potential size constraints on the UCD VM, we elected to use Azure despite this being a new technology to the team, and this decision was more easily justified with the generous student storage on offer. This element of balancing storage capacity, cost limitations, and performance was an aspect of the project which was particularly fascinating as it provided a key insight into a problem which is commonly encountered in businesses; in other modules thus far in the course, this typically hasn't been a significant factor of consideration and balancing these considerations to choose a final solution was valuable in helping me gain a perspective into an ever-present problem in industry.

Azure and Docker were technologies which I hadn't previously seen, and these posed unique challenges for me. While the rest of my team was using Windows 10, I was working on an Apple Silicon Macbook, and many aspects of the set-up which the team had agreed upon was not easily translatable to a Macbook and required me to independently research ways of aligning my set-up with the team's set-up to enable running the app correctly. While Linux and Windows contain a persistent storage volume for Docker, Macbooks hide this volume within a virtual machine which is created by Docker, and the mounting of storage volumes is the preferred method of persisting data. Because of how the web application was structured, I faced a unique challenge in having to work around this structure to develop my own custom version of the application which excluded this precise Docker set-up in order to run the application. While the code lead did not believe this to be a big deal, I regret not having vetoed this particular set-up early in the application development as it provided an initial struggle in getting the solution suggested by the code lead to work. A challenge pertaining to my set-up was also present in the usage of Azure. As the pyodbc module is not natively supported on Apple Silicon, and as SQL Developer is not available, I had to learn during the course of the project how to run items using Rosetta in order to use pyodbc and connect to an MSSQL Azure database. At the time, I found these aspects quite frustrating as I believed they added an unnecessary layer of difficulty to the project, but now I realise they instead emphasise the importance of aligning environments and ensuring solutions which are chosen are designed with portability and system neutrality in mind.

After settling on a technology, many novel technical challenges were present in working with the key data-sets which were provided. Unlike the COVID-19 data-sets which we worked with in COMP47350 which were of a limited size and easily manipulatable using pandas, the data-sets which we encountered in this module were not only varied with key challenges such as differences in the schema between the current and historic data, but also the volume of the historic data was too significant to simply open in Pandas. These were problems which I had not previously encountered during the course, and identifying potential solutions to this involved bringing together techniques encountered not only in COMP47350 to assist in automating the analysis of a variety of data sources, but also leveraged aspects of data modelling which we saw in COMP20240 to understand the structure of the data and techniques from COMP20230 to consider how optimal our code was when working with such a significant data-set. The consideration of using MySQL to serve as a temporary store of the historic data until it was split by route, although ultimately not the solution which we went with once Dask was discovered by Danning, was a consideration which could only have been made through the technical growth which the course has enabled. Particularly in the context of a world where e-Commerce has grown significantly, and the size of data-sets is ever-increasing due to the ability to capture customer data-sets, omni-channel transactional data-sets, click-stream data-sets, social media data-sets, and user survey data-sets, with complicated links between various User IDs and Customer IDs which may not neatly align between sources, I think this experience will prove valuable to me as a data scientist and software developer going forward by providing me with an initial exposure in how to tackle a project where there are many data sources. Similarly, the experience in leveraging APIs to scrape data on a recurring basis and store it into our own database not only helped provide a valuable experience in working with JSON data, but gave a firm context in how data-sets can be enhanced by leveraging open-source technology and data sources.

One of the key challenges I faced during the project which I was particularly surprised by was not in the approach to project management or particular technical challenges, but in the difficulties in working together as a team and trying to balance differences in availability, disagreements in approaches to problems, and how to navigate when team members are not performing to a level which you expect from a professional environment.

Coming into the project with three years of professional experience, one aspect which I really struggled with at the beginning of the project was that I felt in some cases my previous exposure to certain problems was not taken seriously by the team which is something I had not previously encountered in the course. Some of the suggestions I had initially made regarding some of the approaches to the data analytics and the importance of data modelling was not, in my opinion, particularly acknowledged; in my view, I think this partly originated from my position as the customer lead and the team perhaps beliecing this meant I was not serious on the technical element. Upon reflection, perhaps I had not been convincing enough on the importance of these aspects and why they are significant, and going forward, I believe if I was to approach a similar project,I would put more emphasis on the supporting research to make a convincing argument for the team.

Differences between the standard I had expected in certain aspects of the project (such as presentations, the group report, or final delivered app) and what other team members were comfortable with was something I found particularly difficult to reconcile. While the data analyst on the project was content with the quality of the modelling and data exploration which was conducted, I had viewed it as very insufficient for the duration of the project. At the time while I didn't want to cause conflict and tried to emphasise how there were certain gaps by asking for key figures, I learned that it is important to confront aspects like this early and potentially rebalance team positions to try to encourage a greater level of engagement. Similarly, as I was working during the course of the project and had limited availability during working hours and had to concentrate my development in the 5pm to 12am during weekdays and all hours of the weekends, I was not quite expecting that other team members would be more resistent to working on the project as long as it takes including weekends. It made me very conscious in future to consider the different external factors that team members might have, and to understand the importance of compromising for the sake of team unity and a successful project delivery.

Overall, having completed the project, I truly feel I've grown significantly as a computer scientist, and through my learnings in project organisation, the acquisition of new technical skills by encountering previously unseen problems, and in gaining insight into some of the challenges in working with a team, I believe I am a more well-rounded software developer capable of completing complex projects.