1. **What is the project trying to achieve, why that is important and how is it different from previous work? (100-250 words)**

* This implementation-focused project is trying to produce a monitoring and forecasting software tool for a non-profit beverage recycling company in Switzerland called PET Recycling Switzerland (PRS). This will be achieved in two ways. Firstly, (1) building a model which predicts the PET recycling volumes in the coming weeks and months. Modelling relates to his null and alternative hypotheses, which centre on whether the PET volumes in Switzerland can be explained and forecasted from a set of independent variables – sales volumes, weather/seasonality and supply chains flows. Secondly, (2) building a user-friendly software tool and interface – that visualises the data and updates – displaying new predictions on fresh data.
* This project is important because it aims to make PRS’ sustainability efforts more effective. Indeed, PRS currently recycles 82% of the bottles sold in Switzerland. Consequently, this software tool will help reduce the environmental impact of Swiss society by making recycling plastic bottles more effective. Furthermore, demand for sustainably sourced plastic has increased - thus this tool is integral to enable PRS to mitigate (a) increased demand, (b) supply disruptions, (c) prevent PRS from facing contractual penalties and (d) offer a low-cost solution.
* This project differs from previous work because the tool is novel within the recycling industry in Switzerland and has the potential to inspire the development of more similar customised modelling tools within other recycling industries and different countries. The current system employed by PRS is an SAP ERP extension – however the project device differs as it is lower cost and customised.

1. **How is the student going to approach the project and how will the student know if the project met its goals? (100-250 words)**

* The student is going to approach the project in two key steps. Firstly, he is going to build a back-end statistical model using Python, Data Analysis and Machine Learning to determine whether his hypotheses are successful - that recycled PET volumes in Switzerland can be effectively forecasted from his independent variables. Secondly, once a model has been built, he will then devise a front-end user-friendly system that updates, displays predictions based on new data. This part is to a greater extent independent of the hypothesis. From the workplan we can see that the student already has the necessary data to get immediately started. Furthermore, he will employ an agile and test-driven approach, in which he will build the project across numerous sprints.
* The student will know if the project has met its goals by both quantitative and qualitative assessment. Quantitively, (a) the significance and strength of the correlating data produced by his models will be assessed in respect to his hypotheses – to determine if they are indeed effective and offer good weekly and monthly performance for predicting Swiss PET volumes. Additionally, (b) the student will have qualitive evaluation from the deployment of his product within the organisation. He will gain feedback from the CEO and CFO of PSR and will be able to determine if the product has been a success. If the tool is easy to install, use day-to-day, low maintenance and offers a good user experience and high value tool for PSR.

1. **How does the student plan to achieve the deliverables and mitigate risks, and is this planning approach appropriate? (100-250 words)**

* The student plans to achieve the deliverables (ie. a statistical model back-end and user-friendly front-end tool) utilising agile software methods. He plans to utilise a variety of opensource tools including Electron for the desktop app, GUI canvases and libraries, Python to form the basis of the system and lastly relational databases to model and manipulate the data. Furthermore, he plans to implement two rounds of user testing to ensure a successful product is delivered to the client PSR.
* He has initially mitigated a big risk by getting all necessary data upfront from the client, PSR – to enable him to get properly started. Furthermore, the student has effectively devised three layers of testing (unit tests, integration tests and end-to-end tests) to mitigate potential bugs and to lead to effective delivery of the products. Furthermore, his two sets of user testing should lead to sufficient evaluation and the role out of an effective UI and final product.
* This planning approach does seem appropriate – his Project Plan and Gantt chart is clear and effective. Furthermore, his plan to roll out alpha and beta versions seems incredibly logical. Machine Learning and Data Analysis has not been a part of the MSc Computer Science course; however, the student seems confident he can implement necessary data analysis and perhaps has sufficient experience in this field. Perhaps he could clarify his methods of performing data analysis and the mathematical approach for modelling the PET data or machine learning tools used.

1. **How well was the research proposal presented and formatted? (100-250 words)**

* The research proposal is well formatted. In particular, the Gantt Chart, Risk Table and Figures are excellently presented to a very high standard. The Figures representing system design and the recycling process are also a high standard and suitably help break down the text into easier visual information. Furthermore, the use of legends or pie charts to signify low, medium, high and very-high is clear and a nice feature.
* However, there are a couple of typos most glaringly “Gantt”, some missed capitalisation, and some indentation issues at the start of paragraphs. However, the student has effectively utilised LaTeX and the references have been done properly.
* Perhaps the abbreviated use of ERP ie. enterprise resource planning and SAP could have been defined before their use, I had to do a quick google to double check what they were accurately.
* Furthermore, I totally appreciate that the volume labels have been removed for confidentiality but it would be great to offer a bit more information regarding the significance of the Figure 1 – I was a tad confused by the x-axis and the significance of the fluctuating line graph.

1. **What should be improved before the student embarks on their project? (50-250 words, can use bullet points)**

* Firstly, I think the student’s project is fantastic. To have a large client and implement something that can offer real-world change via plastic bottle recycling is excellent. The student has clearly used his networking as an elite management consultant to great effect.
* Yet, I have minor suggestions to improve his project:

1. The Forecasting/Machine Learning is not particularly explicitly discussed in the proposal – I felt the student could have given this more emphasis. Utilising computer science data techniques for how he plans to model, rationalise, and predict data. Personally, I have little knowledge in this research area but am unsure of what statistical models and methods he will be implementing with the three variables to correlate, visualise and predict future PET bottle quantities.
2. Considering the user-friendly software tool forms the secondary part of his research. Potentially, he could reflect more on the interface of the tool. For instance, he could have explained his approach for delivering the user interface (UI) of the – this would provide another area (exclusive of data) to analyse the effectiveness of the product.
3. Lastly, although this is an implementation project the student could potentially have used more academic sources and honed on his academic contribution. Although a novel product, he has not fully reviewed academic literature. Despite much academic material within data science, machine learning, recycling, UI design etc. – all relevant to his project.
4. **Give a mark out of 100 using the research proposal marking scheme and justify the mark given (50-250 words)**

* 63/100
* This proposal satisfies the assessment criteria to a good standard with few areas of weakness. There appears to be a good understanding of knowledge relating to PET bottle recycling and applying his research for a large commercial client is excellent. Furthermore, the Report is well organised – with all boxes being ticked from a formatting standpoint.
* This proposal comes up just short of a distinction grade because the student has not reflected enough on academic context, his academic contribution, and the mathematical/ data-science component of his proposal. This is arguably quite important as the MSc Computer Science at the University of Bristol did not cover this field of research at all – therefore I felt the student could have spent more time justifying his knowledge and approach to forecasting PET bottle outputs. Furthermore, this would have enabled the student to utilise and grapple with more academic sources.
* Regardless of this weakness, I feel that the student will certainly deliver the tool – considering he is working for a corporate client and appears very competent based on his background as a consultant. Lastly, perhaps the student could have applied more to his second deliverable which focused on developing a User Interface tool for his client PRS. This would provide another area (outside of data) – to analyse the effectiveness of his research.

1. **What is the project trying to achieve, why that is important and how is it different from previous work? (100-250 words)**

* This project is extending research conducted by Dhiman et al. on content-based filtering for email providers. Consequently, the student is trying to determine how effective three major email providers ie. Hotmail, Gmail, Yahoo and one niche safety-focused provider ie. ProtonMail are at blocking two types of emails. Firstly, obfuscated scam emails. Secondly, obfuscated normal emails. Both types of email will be deceptively modified to homograph versions of the same message. After which, the student will measure the success of each provider via determining how many emails were successfully delivered and evaluate the ability of different email providers to withstand homograph attacks.
* This project is important because it extends Dhiman’s research, which staggeringly found 96% of the scam emails tested were successfully delivered. Furthermore, homograph-adapted obfuscated emails are indistinguishable to human eyes and thus could be a potent method for email scammers to effectively attack unsuspecting email users. Indeed, the students section Related Work notes multiple effective application and phishing attacks exploiting homograph protection weaknesses.
* The project differs from Dhiman et al.’s study as the student plans to extend the study. Firstly, by utilising additional niche email provide, ProtonMail – which offers a more secure and encrypted email platform than mainstream providers. Secondly, the student plans to send two types of emails, rather than one type - as was the case in Dhiman’s original study. Furthermore, this is a growing research area and thus the student’s project seeks to add to the limited previous work on this topic.

1. **How is the student going to approach the project and how will the student know if the project met its goals? (100-250 words)**

* The approach of the project is divided into three key steps. Firstly, selecting confusable characters manually. This will involve selecting the most visually similar and confusable characters based on their glyphs to obtain homographs. Furthermore, this process will involve significant hands-on tweaking as sometimes the characters may have difficult spacing. Secondly, the student will build an obfuscator utilising the 58 homographs identified in Figure 1. The obfuscator will be written in Java and determine the best character to choose within each word. Thirdly and lastly, the student will experiment by implementing and running their scam approach. This will involve selecting three types of emails: (a) known scam emails which have been blocked by at least two mailbox providers, (b) scam emails which have been obfuscated and (c) obfuscated normal emails. From this, contrasting analysis can be achieved. Moreover, these emails will be sent over 5 days – to best simulate the low-and-slow behaviour of scammers and to avoid triggering rate-limits enforced by email providers.
* Not much has been established by the student regarding key project goals. Nevertheless, the student will know if the project has met its goals by successful execution of the three key steps outlined in the methodology. Furthermore, from the Gantt Chart and emphasis placed on extending Dhiman et al.’s research – surely means that the project will be successful if the student is able to collect comparative data, make important comparisons and reach academic conclusions to contribute to further research into scamming, spamming and content-based filtering.

1. **How does the student plan to achieve the deliverables and mitigate risks, and is this planning approach appropriate? (100-250 words)**

* From the Gantt Chart, the student plans to achieve deliverables outlined in the methodology via writing their literature review first, devise an experiment plan, invite experiment candidates, performing experimentation, followed by data analysis, and lastly writing their thesis. The student has emphasised that they will replicate common scam approaches across 90 emails and devised three different email strategies for executing differing homograph attacks.
* The student has identified three key risks. The student plans to mitigate these risks by starting early, being disciplined in sending emails slowly and in small amounts. Furthermore, as the student is working abroad in China, the student plans to use a state-approved VPN to help with internet access. Lastly, the writer suggests that they may have trouble accessing candidates – which may result in improper experimentation.
* I think that their project is certainly achievable based on the Bristol MSc curriculum – I think developing a parser/Obfuscator tool in Java utilising Unicode variables should be manageable. However, the planning approach is lacking in several areas. Initially, the Gantt Chart and Risk Analysis tables are both very short. There has been little evaluation or analysis of both tables, significant milestones stated and acknowledgement of risks concerning the global pandemic. I think the Risk Table would be improved with a severity column. Lastly, I struggled to understand the Gantt Chart - I am unsure why the student has decided to invite experiment candidates, this is not discussed in the body of the report and Dhiman et al. did not use candidates.

1. **How well was the research proposal presented and formatted? (100-250 words)**

* The research proposal seems to be formatted according to ACM guidelines in LaTeX. Although not many, the in-text references have been properly referenced. Furthermore, the Figures are graphically clear and simple. It would have been a nice to have the figures referenced within the text – to add more coherence as to their relevance and place within the body of the report. However, the Bibliography has not been properly presented utilising ACM criteria – websites and DOI should be bold and reference numbers should be in square brackets e.g. [2].
* The student should more closely use formatting advice on the criteria and structure of the proposal presented by faculty. In particular, the student could have for instance used the four large and clear key headings (Introduction, Literature Review, Proposed Approach, Gantt Chart and Risk table) and made greater use of subheadings such as ‘Evaluation’, ‘Objectives’, ‘Requirements’ and ‘Hypothesis’. This would have resulted in a much clearer formatted Research Proposal satisfying all the criteria emphasised during our Research Skills lectures. Presently, the formatted structure of the proposal is slightly ad-hoc and has potentially resulted in essential things such as testable hypotheses not being written by the student.
* The Gantt Chart is quite basic - it would have been better to have milestones and more information. Presently, the Gantt Chart is do work and finish, where are the risks, dependencies, concurrencies – it is overall not a suitable level of detail.

1. **What should be improved before the student embarks on their project? (50-250 words, can use bullet points)**

* I believe the student has chosen an important project – I agree with their contention that 96% of homograph scam messages successfully delivering to recipients is far too high. Furthermore, I believe the project is certainly deliverable.
* However, this proposal requires greater clarity and detail to reach a higher MSc level:
* I would implore the student to read Dr. Ray’s slides on Research Proposal structure. In short, this proposal is missing some key sections – particularly precise, focused, and testable hypotheses. With scientific hypotheses in hand, e.g.: “Within the devised experiment, ProtonMail will offer the best performance at defending against homograph attacks compared to mainstream providers” – the student can then start forging coherent arguments, debate, testing and evaluation. There is too much description and not enough argumentation.
* The title and replication of Dhiman et al.’s study is insufficient – you should make your own independent research title/project and instead thoroughly evaluate Dhiman et al.’s research – what did they did right/wrong and how your project will be better. I appreciate the methodology is different, but you need to make your proposal more argumentative/original within the academic literature – with clear lines of argument formed from the hypotheses.
* I would advise the student to get their supervisor, friends, peers, and family to proofread their work because it is very helpful. Some of the proposal suffers from incoherence leading to avoidable mistakes. Particularly, the Gantt Chart and introduction of experiment candidates – without this method of experimentation being clearly expressed within the body of the report.

1. **Give a mark out of 100 using the research proposal marking scheme and justify the mark given (50-250 words)**

* 53/100
* The project assessment criteria are satisfied with no significant errors – the student is replicating an already published study, thus the research should meet academic criteria. The student lacks independence/originality in their approach and analysis – to achieve a higher grade and MSc standard. Equally, there is an admission of focused and testable hypotheses to form the foundation for the research. However, the research area is important and growing in prominence with rises in cyber-crime and potential homograph-driven attacks. The report is organised and well presented for the most part, but limited in scope, evaluation, and reflection – particularly concerning the lack of detail and evaluation for their Gantt Chart and Risk Table. Indeed, the Gantt Chart lacks coherence with the rest of the proposal because it includes experiment candidates and the Risk Table should be implemented with greater detail resulting in more columns and rows.
* The student has done well to work in some additional and relevant literature. Therefore, perhaps armed with more clear line of arguments derived from hypotheses – the student can start to write more argumentatively and critically. As emphasised earlier, the student should evaluate other researchers approaches – to justify their own approach, experimentation, and key academic contribution. Furthermore, the student’s explanation and evaluation of their research is slightly convoluted and confusing. To easily avoid this, the student should (a) get other individuals to proofread and (b) re-structure the proposal along Dr Ray’s suggested criteria from the lecture slides. Consequently, this would result in no key sections being missed.