KTP Project Work Plan

This Work Plan complements your application and allows Assessors to understand the project delivery activity to be undertaken by the Associate that will build the momentum towards project success and will be used by the Associate to deliver the project as well as by the Monitoring Officer to monitor the progress of the project.

The Work Plan is to be completed by Knowledge Base and Business Partner supervisors, collaboratively, as this will better align knowledge transfer expectations, overall project delivery, and anticipated outcomes for the project stages. It is important that your Work Plan delivers the knowledge gaps and outcomes of all partners (Business Partner, Knowledge Base and Associate) described in your application. The Work Plan should show where knowledge transfer is being delivered by the Knowledge Base and how knowledge transfer is being delivered to and exploited by the Business Partner and wider stakeholders.

For live projects, your Work Plan may require flexibility to change. This is welcomed and should be led by the Associate in collaboration with their supervisors and requires approval by LMC members.

Project Length

Once the workplan is complete, calculate the total effort (in months) of all the Activities and Stages, and enter it in the box below. This must match the length of the project declared on the application from.

Total effort (months) in Work Plan	28

Number of associates

If you are proposing a 2 Associates application, you must complete a work plan for each associate.

Associate number	1

Please ensure that the application number is on every page.

Upon completion, the Workplan (including the Risk Register) it should be uploaded as part of the application.

Guidance For Completion of Work Plan:
Set out your plan of work in with Standard Activities and Stages. Each Stage will have a number of Steps:

Standard Activity Table: Add appropriate time effort in table below; induction, T&D, holidays

Stages

- Typically, a 24-month project has 6 stages
- Name each stage and write a **concise** stage description (max 400 characters) to describe the key activity that builds towards delivery of the entire project
- The stages will normally progress through the following:
    - project exploration
    - developing testing and implementation of solutions
    - evaluation of benefits
    - post project planning

Steps

- Break each Stage into Steps, such as 1.1, 1.2, 1.3. Use as many Steps as required
- Each Step should require up to 1 month of effort. (Combine short tasks, break down long tasks)
- For each Step:
    - describe what the Associate will do and how they will do it
    - highlight what innovative new knowledge will be transferred by the academics
    - describe how new knowledge will be embedded into the business
- In the Outputs column, describe key outputs, milestones (M1, M2 etc) and key decision points
- There should be more detail in the steps for the first 6 months of the work plan and this can be recorded in 0.5 effort months. After 6 months use 1.0 effort months per step
- Highlight any activity planned to occur at the Knowledge Base
- Include and highlight opportunities for the Associate to gain commercial exposure
- Include Final Reporting and the development of a case study as an expected final stage of the workplan

Steps that must be included in Stage 1

- A Mini-Project: Associate to complete a 1-2 week duration mini-project which is of direct benefit to the Business and should be reported at LMC 1 or LMC 2
- A project and team effectiveness review to be reported at LMC 1 and updated at LMC 2
- The Associate to further develop the Risk Assessment for the entire project based on the workplan outlined below.
- The Associate should further develop the Project Plan. This typically will include a Gantt chart for the entire project including a detailed Step list for each of the project Stages defined below

Standard Activity	Description of activity	Effort (months)	Outputs and key decision points
1	Knowledge Base and Business Partner induction	0.25	KB and business induction completed
2	Training and development. This is calculated as 10% of the total project time. Include time for: • KTP residential modules 1 and 2 • KTP Associates conference	2.25	Personal development plan created, and implementation agreed with local management committee
3	Associate's holidays	2	Annual leave booked and approved
4	Mini project The mini project is detailed (and timed) in Stage 1	0	1-2 week duration which is of direct benefit to the business
	Total effort in this Stage	4.5	
	Cumulative effort	4.5	

Stage 1: Missing and Incorrect Key Values in the Data Identified by the Client (Academic Team: SY, PM, II, JB, FF, AL) (Max. 200 words)

Identification of missing and incorrect key/important values within a given data, as identified by the client as a pre-existing problem. Implementation of alerting systems for data managers when these values are observed, and machine learning solutions through imputation as a recommendation engine to reduce the data engineer workload.

Steps	Description of Step	Effort (m)	Outputs and key decision points
1.1	Mini project – *moved from standard activities for readability purposes.* M.1 - Identify data sources - data sources from current and historical clients will be identified, and the storage infrastructure. M.2 - Select and prepare data ingestion mechanism - Data ingestion mechanism will involves designing and implementing a streamlined process for efficiently collecting and transferring data from identified sources to the storage infrastructure,	0.5	

	ensuring a reliable and consistent flow of data for initial processing and analysis. M.3 - Initial data processing Entails the application of necessary transformations, cleaning, and structuring to the ingested data, ensuring its quality and consistency, and preparing data for initial analysis and integration into the data processing pipeline. This mini project will require the associate to travel to the client service offices to have a better understanding of the project requirements		
1.2	Exploratory data analysis - analysis of data containing examples of missing and incorrect key/important values.	0.25	KDP – review which data contains examples of the problem in terms of severity and frequency
1.3	Produce business intelligence report on missing and incorrect key value exemplars, detailing the most common problems as well as rare and unique cases.	0.5	O.1.1 - Report on findings based on observations of the problematic data
1.4	Review for strategy plan to alleviate issues surrounding missing and incorrect values of high importance with non-machine learning techniques. Given that many of the attributes are considered key and important for this problem, the likely solution is to alert the data manager to the problem. Incorrect key attributes may be detectable following a rule-based approach similar to that in Stage 2.	0.5	O.1.2 - Extend previous report with suggestions of solutions to alleviate the problem using non-ML techniques
1.5	Implementation of non-ML techniques. The findings from the report in 1.4 will be implemented based on simple rule based model without the need for extensive learning. Appropriate choice of programming language that suits with **bigspark's** stacks will be made	0.5	O.1.3 - Code for non-ML techniques
1.6	Validation and analysis of non-ML techniques. A similar dataset (provided by **bigspark**) will be used for evaluating the performance of the approach e.g. given a set of similar data with missing or incorrect entries, how accurate is the implementation in identifying those entries).	0.5	

1.7	Literature review and production of strategy plan to alleviate issues surrounding missing and incorrect key data with machine learning techniques. Examples of potential solutions include the use of machine-learning based imputation models to predict the correct value and suggest the prediction to the data manager. Dependent on the nature of the attribute, i.e. numeric and nominal data types, models trained on regression and classification may prove appropriate. A potential improvement via this ML technique would be to reduce the workload of data managers by providing an intelligent suggestion alongside the alert.	0.5	O.1.4 - Literature report on machine learning techniques appropriate for solving the problem.
1.8	Development and implementation of machine learning solution The output of stage 1.7 will be guiding the implementation of the ML solution. Outlier detection models such as One-Class SVM can help identify missing data and imputation and over-sampling approaches (e.g. SMOTE) can suggest values that can serve as a good replacement. ML libraries and programming languages that suits **bigspark** tech stack will be selected	1	O.1.5 - Code and model file for machine learning model
1.9	Validation and analysis of machine learning solution. As in 1.6, a similar dataset (provided by **bigspark**) will be used for evaluating the performance of the ML approach both in terms of identifying missing and incorrect values, as well as suggesting replacement. The data provided will allow for benchmarking of the result.	0.5	
1.10	Implementation of Explainable AI for interpretation of model predictions and suggestions - Adopted methodology will include a similarity measure (distance based) of the data entries using the best performing model selected in 1.9. Entries that have abnormally wide distance from the data clusters will be flagged. Further exploration of the individual features will reveal low level detail and explainable of the model prediction	0.25	

1.11	Comparison of non-ML and ML approaches to alleviate the problem The findings of stage 1.6 and 1.9 will evaluated and compared to the ground truth Since the evaluation data is provided by **bigspark**, it is easy to benchmark the performance of the different approaches. GPU will be utilised for running the experiments to speed up processing. Roboflow platform will be used as secondary validation mechanism of the obtained result	0.25	O.1.6 - short report unifying the non-ML and ML solutions with comparison between the two. KDP – decisions made based on improvements, and if they are either theoretical or practical improvements, to choose the best solution(s) to include in the overall pipeline.
1.12	Business strategy mini report for productization. This report will include implementation, deployment and usage guidelines	0.25	O.1.7 - Short report detailing which system features can be used for similar/all clients, and which system features will need to be managed bespoke prior to implementation for new clients.
	Total effort in this Stage	5.5	
	Cumulative effort	10.0	

Stage 2: Similar and Duplicate Data Entries Identified by the Client (Academic Team: SY, PM, II, JB)
(Max. 200 words)
Another key problem identified by the client is that of overly similar and duplicate data entries. While duplicate data is easy to detect, the notion of "similar" is data-dependent, and often an "overly similar" pair of data objects may in fact be duplicates which have been entered slightly differently. To solve this, this stage will first implement duplicate detection before exploring ML-based improvements to the solution by clustering data and measuring the Euclidean distance between a given data object and other objects plus cluster centroids. The goal of this approach is to quantify what "overly similar" means within a given dataset, and identify those cases as anomalies that should be observed by an engineer.

Steps	Description of Step	Effort (m)	Outputs and key decision points
2.1	Exploratory data analysis - analysis of data containing examples of overly similar or duplicate values.	0.25	KDP – review which data contains examples of the problem in terms of severity and frequency
2.2	Produce business intelligence report on duplicated and overly similar exemplars, detailing the most common problems as well as rare and unique cases.	0.5	O.2.1 - Report on findings based on observations of the problematic data
2.3	Review for strategy plan to alleviate the problem with non-machine learning techniques	0.5	O.2.2 - Extend previous report with suggestions of solutions to alleviate the problem using non-ML techniques

	Non-ML techniques could include a similarity metric between two data types. Given a similarity metric of 1.0, the data may be a duplicate of a pre-existing entry.		
2.4	Implementation of non-ML techniques The report produced in 2.3 will be implemented in a manner agreed on with **bigspark**, i.e., appropriate choice of programming language to best suit company workflow. This is a rule-based approach that does not require learning or training.	0.5	O.2.3 - Code for non-ML techniques
2.5	Validation and analysis of non-ML techniques Given a set of similar and duplicate entries provided by **bigspark**, the non-ML techniques will be tested and scored based on their performance in detection.	0.5	
2.6	Literature review and production of strategy plan to alleviate the problem with machine learning techniques. While comparison is apt for detection of duplicate data, more intelligent techniques may be required for other data entries. For example, if two entries are technically duplicate but are seen as dissimilar due to typographic differences or formatting etc., such as when two different data engineers may enter the same data in a different way. Potential solutions to this problem include clustering algorithms, where a data object is compared in terms of Euclidean distance from others in n-dimensional space (where n is the number of input features). Furthermore, the distance between a data object and a cluster centroid, such as from DBSCAN, may also present as evidence that a data object is overly similar or identical to a given other data. The improvement from the addition of ML techniques surrounds the detection of “overly similar” data, which is a much more difficult problem than duplicate detection.	0.5	O.2.4 - Literature report on machine learning techniques appropriate for solving the problem.

2.7	Development and implementation of machine learning solution The report produced in stage 2.6 will be used to inform the implementation of the ML solution. As before, appropriate selection of programming languages and libraries will be agreed with bigspark (but will not affect the implementation itself). For example, the Scikit Learn Python library or Weka Java libraries could be suggested (e.g. DBSCAN is available within both), and decided upon to best suit company workflow.	1	O.2.5 - Code and model file for machine learning model
2.8	Validation and analysis of machine learning solution Similarly, to stage 2.4, the same example set of similar and duplicate data entries provided by **bigspark** will be benchmarked given the trained ML solution.	0.5	
2.9	Implementation of Explainable AI for interpretation of model predictions and suggestions – methodology will be dependent on the best model selected during stages 2.7 and 2.8. For example, decision trees fitted to clusters would reveal explainable features as to why items of data are closely related (according to: Dasgupta, S., Frost, N., Moshkovitz, M. and Rashtchian, C., 2020, July. Explainable k-means and k-medians clustering. In Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria (pp. 12-18).)	0.25	
2.10	Comparison of non-ML and ML approaches to alleviate the problem The findings of stages 2.5 and 2.8, given that they are tested on the same data, will be compared and critically analysed given ability and efficiency. This enables a direct comparison between the ML and non-ML approaches which will be used for choice of implementation into the final workflow. The adaptability of the two approaches will be evaluated on new dataset to ascertain the efficiency of the solutions. Similarly, GPU will be utilised for experiment and Roboflow platform will be used as secondary validation mechanism of the obtained result	0.25	O.2.6 - short report unifying the non-ML and ML solutions with comparison between the two. KDP – decisions made based on improvements, and if they are either theoretical or practical improvements, to choose the best solution(s) to include in the overall pipeline.

2.11	Business strategy mini report for productization containing implementation and deployment guidelines	0.25	O.2.7 - Short report detailing which system features can be used for similar/all clients, and which system features will need to be managed bespoke prior to implementation for new clients.
	Total effort in this Stage	5.0	
	Cumulative effort	15.0	

Stage 3: Discrepancies in Attribute Conventions and Entries Identified by the Client (Academic Team: SY, PM, II, JB) (Max. 200 words)

While standardisation is important and often solves discrepancies, some may remain in data, especially with systems that have existed and evolved over a long period of time (e.g., within the banking sector). This problem has been identified by the client and the goal of this stage is to first implement automated standardisation prior to identifying discrepancies through probabilistic record linkage and entity resolution models, which are machine learning techniques aiming to clarify whether two values should belong to the same attribute.

Steps	**Description of Step**	**Effort (m)**	**Outputs and key decision points**
3.1	Exploratory data analysis - analysis of data containing examples of discrepancies of conventions and entries.	0.25	KDP – review which data contains examples of the problem in terms of severity and frequency
3.2	Produce business intelligence report on discrepancy exemplars, detailing the most common problems as well as rare and unique cases.	0.5	O.3.1 - Report on findings based on observations of the problematic data
3.3	Review for strategy plan to alleviate the problem with non-machine learning techniques Examples include the automated standardisation of data, such as an overarching system which will correct naming conventions and change formats (e.g. date formatting, number formatting, floating point)	0.5	O.3.2 - Extend previous report with suggestions of solutions to alleviate the problem using non-ML techniques
3.4	Implementation of non-ML techniques as proposed by the report produced in Stage 3.3. Rule based solutions based on the data quality criteria will be developed in a suitable programming language that fits into **bigspark** workflow.	0.5	O.3.3 – Implementation of the non-ML solution utilising the outcome of O.3.2
3.5	Validation and analysis of non-ML techniques, as prior – a set of example data with discrepancies to be provided by **bigspark** wherein the non-ML approach will be benchmarked, scored on the number of issues corrected and notes on which	0.5	

	issues were unable to be corrected through non-ML approaches.		
3.6	Literature review and production of strategy plan to alleviate discrepancies with machine learning techniques. Examples include probabilistic record linkage or entity resolution models, which can learn to clarify whether two values should belong to the same attribute. While standardisation is important, the addition of this ML technique will improve the process by consideration of statistical relationships between values.	0.5	O.3.4 - Literature report on machine learning techniques appropriate for solving the problem.
3.7	Development and implementation of machine learning solution This stage will physically implement the findings of Stage 3.6 given the strategy plan. As prior, use of programming language and machine learning libraries will be agreed with **bigspark** to best suit company workflow and evaluation metrics.	1	O.3.5 - Code and model file for machine learning model
3.8	Validation and analysis of machine learning solution As prior, concrete examples of discrepancies will form a testing dataset for the model to apply learnt rules to.	0.5	
3.9	Implementation of Explainable AI for interpretation of model predictions and suggestions As prior, explainable techniques will rely on the best method of ML approach (revealed in Stages 3.6, 3.7, and 3.8). The level of explainability possible (if at all) will aid in informing business strategy decision on which approach is best suited to the final workflow.	0.25	
3.10	Comparison of non-ML and ML approaches to alleviate the problem Given that the ML and non-ML solutions are tested on the same example data, a direct comparison can be implemented. Given ability and	0.25	O.3.6 - short report unifying the non-ML and ML solutions with comparison between the two. KDP – decisions made based on improvements, and if they are either

	efficiency, a decision can then be made on which approach best suits **bigspark's** workflow. As in Stage 2, a GPU will be utilised for running the experiments while Roboflow will be used for validation		theoretical or practical improvements, to choose the best solution(s) to include in the overall pipeline.
3.11	Business strategy mini report for productization container implementation and deployment guidelines	0.25	O.3.7 - Short report detailing which system features can be used for similar/all clients, and which system features will need to be managed bespoke prior to implementation for new clients.
	Total effort in this Stage	5.0	
	Cumulative effort	20.0	

Stage 4: AI Framework Implementation and Scalability/Maintenance Strategy (Academic Team: SY, PM, II, JB)
(Max. 200 words)
The previous stages address the data quality problem using both ML and non-ML approaches by considering one data related problem at a time. However, there are often multiple quality problem associated with a dataset. This stage aims to develop a framework to unify the individual solution developed previous through model fusion and enhances the outcome using explainable AI to help the data engineer understand the problem. A client software tool will also be developed for easy utilisation of the tool.

Steps	Description of Step	Effort (m)	Outputs and key decision points
4.1	Development of the AI framework Given the findings of all previous approaches, the "AI framework" is enabled by utilising all prior chosen models into a unified system. The AI Framework presents as an analysis tool where data is input, and the findings revealed by each model in-turn will then be presented to the data manager in a unified fashion. The software developed in this stage can be thought of as an interface to easily access all produced models.	1.0	O.4.1 - An AI framework is developed for fusing the different models developed in stage 1 -3
4.2	Validation and analysis of the developed framework in terms of effectiveness and scalability This stage is an analysis of Big Data Infrastructure – that is, running in-house tests to see how the system would perform given x number of	0.5	O.4.3 - The developed framework is validated against the results of the individual approaches developed in stage 1 - 3

	customers and their frequency of processing requirements. The results of these tests will enable future decision-making and business strategy given the time and resources required for efficient processing of the data. It will also present an opportunity to evaluate cost and computational requirements of running the singular vs consolidated model for different use cases (e.g. should a single model be used for x problem as against the fused model?) Similarly, the use of GPU for experiments and Roboflow for secondary validation will be adopted		
4.3	Presentation of XAI (eXplainable Artificial Intelligence) for each problem-solving ML model This step bares similarly to step 4.1. Given that XAI is explored for each of the individual solutions, this step focuses on how the individual XAI outputs can be concatenated and presented in such a way that they are easy to understand. For example, a dashboard or report.	1.0	O.4.3 – A tool will be developed to aggregate the XAI results for the fused models
4.4	Client software development and UI design for interfacing with the developed solutions Given the outputs of steps 4.1 and 4.3, this stage focuses on the unification of all “presentations” towards an easy-to-use client software. The aim of this stage is to translate the model outputs and XAI explanations to information that can be understood by both technical and non-technical users.	0.5	O.4.4 - A software will be developed that unifies the XAI outputs for each model into one easy-to-use interface for the data engineer to observe.
4.5	Software evaluation, documentation and user manual	0.5	O.4.5 - Evaluation of O.6.4 and the generation of a short user manual
	Total effort in this Stage	3.5	
	Cumulative effort	23.5	

Stage 5: Realtime Data Ingestion and Quality testing (Academic Team: SY, PM, II, JB, FF) (Max. 200 words) This stage performs a holistic end to end validation of the AI framework by looking at the problem from the data engineer's perspective starting from ingestion. Efficient approach will be developed for the different quality control stages using a fresh dataset.			
Steps	**Description of Step**	**Effort (m)**	**Outputs and key decision points**
5.1	Development of a real-time mechanism for data ingestion pipeline - This step would involve the implementation of real-time data ingestion pipeline to allow for continuous and efficient ingestion of fresh/new data from various sources as they are received into the BS data infrastructure enabling seamless integration of data for real-time data processing. Cloud services and storage utilisation such as Amazon Deequ, storage buckets and other related services will be considered.	1.0	O.5.1 Functional and scalable real-time data processing infrastructure
5.2	Integration of data streams from sources to AL/ML model for real time real-time batch deviation detection by utilising the ingestion pipeline developed in 5.1	0.25	O.5.2 - Final product in the form of a software.
5.3	A test run of real-time batch deviation detection - A series of tests of the final product to validate its ability to identify significant deviations in newly ingested data batches to ensure the effectiveness of the automated data quality validation process. As with previous stages, the GPU will be used for running the experiments and Roboflow for secondary validation	0.25	O.5.3 Test report of the performance of the final product which should also include the identification of previously unnoticed data consistencies and anomalies. KDP – decisions made based on improvements to be included in the final product.
5.4	Realtime Validation of Pipeline and final report on real time system, to be used for the final report delivered to InnovateUK.	0.5	O.5.4 Final report of the products functionalities highlighting the pipelines capacity to enhance data quality and provide timely alerts for corrective actions.
	Total effort in this Stage	2.0	
	Cumulative effort	25.5	

Stage 6: Embedding of the Knowledge Exchange - Post-project Training and Exploitation (Academic Team: SY, FF, AL) (Max. 200 words) This stage revolves around supporting KDP to build capacity in taking ownership of the project. A Knowledge Base (KB) will be developed and populated with relevant documentations and guides on the usage and adaptation of the developed tools			
Steps	**Description of Step**	**Effort (m)**	**Outputs and key decision points**
6.1	Consolidation of **final** report Includes all previously written reports within user manuals and the development of screencast collection as user guides	0.5	O.6.1 - merging of reports and user guides developed for the different stages. A screencast will be recorded to compliment the documents
6.2	Deployment of a product knowledge base dashboard using standard CMS tools (such as WordPress)	0.25	O.6.2 - A CMS will be deployed as a KB for the project for easy access to the product details and guidance
6.3	Populating the knowledge based with documentation and user guides for easy adaption to customers' needs	0.25	O.6.3 - Populating the KB will the content from stage 1 – 4 documentations. A proper indexing and search methods will be implemented
6.4	Hands-on staff training on usage and adaptation methods This includes training relating to the usage of the solution, modifying and adapting it for different customers’ needs and extending the functionality when needed.	0.5	O.6.4 - A face to face hands-on training will be carried out as part of the embedding process on how to use the tools and adapt it to different customers’ need
6.5	Development of productisation framework. This covers the workflow for onboarding new clients, adapting the solution to the client need, deployment and validation processes.	0.25	O.6.5 - A framework for the productisation of the solution will be developed starting from client onboarding to deployment
6.6	Development and attendance of conference presentations and a high-impact academic journal article. Potential examples include *the International Conference on Big Data (2026)* and Springer’s *Journal of Big Data* or Elsevier’s *Big Data Research*.	0.75	O.6.6 - ML and non-ML approaches proposed for the data quality improvement and explainable AI and the corresponding findings will be published
	Total effort in this Stage	2.5	
	Cumulative effort	28.0	

Use the table below to add additional stages as required

Time Plan

Use Time Plan Table below. We have entered the tasks from the project tasks table. Please use the empty rows to add each of the project stages. For each stage show the duration, milestones (M1, M2 etc) to match the ones you've used in the project stages table

Project task or stage	Project months									
	0 to 3	3 to 6	6 to 9	9 to 12	12 to 15	15 to 18	18 to 21	21 to 24	24 to 27	27 to 30
Standard Activities (including associate training, holiday etc)										
Stage 1: Missing and Incorrect Key Values in the Data Identified by the Client		O.1.1 - O.1.7								
Stage 2: Similar and Duplicate Data Entries Identified by the Client				O.2.1 - O.2.7						
Stage 3: Discrepancies in Attribute Conventions and Entries Identified by the Client						O.3.1 - O.3.7				
Stage 4: AI Framework Implementation and Scalability/Maintenance Strategy							O.4.1 - O.4.5			
Stage 5: Realtime Data Ingestion and Quality testing								O.5.1 - O.5.4		
Stage 6: Embedding of the Knowledge Exchange - Post-project Training and Exploitation									O.6.1 - O.6.6	

Risk Statement & Register

Guidance
In this section outline the potential risks to the project succeeding and the actions which you can take to mitigate them. There are many risks that may occur during the execution that may affect the successful delivery of a KTP. Some of these are common to most KTP's such as the timely appointment of a KTP Associate, but the aim here is to identify only the most significant six risks that may impede the partnership from delivering the required outcome. These top six risks, with their impact on the partnership, the likelihood of them occurring and proposed mitigation should be documented in tabular in Risk Register. The Risk Register will be regularly reviewed and updated as required during the execution of the KTP and also presented at the LMC meetings.

- **Risk Factor**
  This will be a brief description that identifies the risk factor. These should be the greatest 6 risks to project success. Please avoid common risks like associate recruitment.

- **Impact on the outputs**
  What the consequential impact is, should the risk occur to the delivery of the KTP.

- **Likelihood**
  This is the estimated likelihood or probability that the risk will occur at some point during the duration of the KTP and then become a KTP impacting issue. This will be purely qualitative and described as High coloured red, Medium coloured amber, or Low coloured green

- **Scope Mitigation**
  This will be a brief description of what actions will be taken to address the risk

Please use the template provided below to create the initial Risk Register, replacing the example risks with your own:

No.	Risk identified	Severity	Mitigation
Technical risks			
1	Data are not available for the experimentation and evaluation	Low	**bigspark** already have access to relevant dataset for testing the hypothesis of this project. Additional client specific datasets will be provided in the course of the project, but an existing dataset is already acquired.
2	Integration of the developed solution with cloud providers and data injections pipelines fail	Medium	To ensure that the solution is scalable, adaptable and commercially viable, integration with existing data ingestion pipelines must be ensured. A discussion with the business partner will be held at the early stage of the project to identify relevant tools to integrated.
3	Proposed approach for data quality and correction is not working and the results are not optimal	Low	Preliminary experiment has been conducted to evaluate approaches for data quality and the results are promising. The Business partner also has extensive experience in working with faulty data. Both the academic team and business partner has significant background in the field and therefore are able to support the associate with early and iterative model validation.
Commercial risks			
4	**bigspark** faces finance issues that could affect their commitment to the project	Medium	**bigspark** is an established company with a track record of delivery and is financially stable. The company has been trading profitably for the past 4 years,
5	User acceptability reveals the quality of the results cannot be commercialised	Low	The team have experience of fully assessing the effectiveness of research study and the business partner has years of industrial experience and privy to the challenges relating to data quality. During the development, several iterations and validation of the approaches will be carried out to ensure the commercial viability and quality of the output, thereby minimising the chances of failure.
Project Management risks			
6	Unsuitable candidate being found within the hiring window	Low	Make clear the skills, experience, and education required for the position. This can help narrow down the candidate pool and ensure that only suitable candidates are considered. Once potential candidates have been identified through an initial screening process, then conduct multiple rounds of interviews to evaluate their skills and fit for the role. This can help weed out candidates who may look good on paper but don't have the necessary skills or experience.
7	Research lead or business partner supervisors are absent for long periods	Low	The project team is composed of qualified individuals from the academia and industry to support the project in the interim.
8	Research Associate fails to perform	Low	Weekly supervision will ensure close monitoring and if performance is lacking a replacement will be easily obtained from our Research Group at NTU. However, to address the chance of this occurring the quality of the associate will be rigorously assessed

			during the application and interview stage to help ensure we attract the best candidate. In addition, the relationships between the academic and industry supervisors will be strengthened which will help mitigate the associate failing to deliver as required.
9	Project fails to progress sufficiently quickly	Low	Close monitoring of project progress by the supervisory team and Associate should ensure that there is minimal creep. The Workplan is also structured such that at any one time, there is more than one step in progress to ensure that delays outside of the control of the project (unplanned lead times, supplier failure etc) have minimal impact.