**Individual Project (Project) – Progress Report (Max 4 pages)**

|  |  |
| --- | --- |
| **Student name:** | Chidera Onwumbiko |
| **Student ID:** | 2321479 |
| **Programme name:** | MSc Applied Data Science |
| **Programme start date:** | Winter 2024 |
| **Project title** | Data-Driven Framework for Enhancing Student Applications, Acceptances, and Registrations at the University of Buckingham. |
| **Academic Supervisor(s):** | Prof. Harin Sellahewa |
| **Industry Supervisor: (if applicable)** |  |

|  |
| --- |
| **Project Overview (max 200 words):** Provide a summary of project’s overall aims and objectives (may refer to Project’s Terms of Reference / Proposal).  **Aim:**  The overall aim of this project is to provide the University of Buckingham with a data-driven analytical and predictive framework to gain insights into students’ applications, acceptances, and registrations at the university. This framework will facilitate a clear understanding of the university's application process in comparison to other educational institutions in the United Kingdom, as well as its acceptance and registration rates. Additionally, it will provide a predictive model to aid the university in enhancing student registration.  **Objectives:**   * **Data Collection and Preprocessing:**   Extract and preprocess historical data on students’ applications, acceptances, and registrations from university records.   * **Exploratory Data Analysis:**   Conduct exploratory data analysis to identify and analyse trends and patterns in application, acceptance, and registration rates over the past years.   * **Data Visualisation:**   Develop an easy-to-understand interactive dashboard to visualise key trends and patterns in the data.   * **Predictive Model:**   Build and validate a predictive model using machine learning techniques to forecast the quantity of student applications that need to be accepted to aid in enhancing student registration. |
| **Summary of significant changes, if any, to the Project’s scope:** Provide details of any changes to the project’s scope since approving the Project’s ToR and confirm whether the changes were approved by the supervisor(s).  Formulating data-driven strategies to enhance student application, acceptance, and registration rates and proposing geographic targeted marketing and outreach campaigns based on the insights gained from the analysis was removed from the scope of the project because the data will not include geographical information, and this was approved by my supervisor, Prof. Harin Sellahewa. |
| **Summary of tasks completed (minimum 250 words):** Provide a summary of key tasks completed by the time of submitting the progress report. These should include both practical work and report writing.   1. The student application, acceptance, and registration data has been extracted from university records. The data extraction was not a complex process, as my supervisor, Prof. Harin Sellahewa, had already achieved this. During the data preprocessing phase, the challenge was that the data format needed to be changed to the other various formats for visualisation and predictive modelling. This is also a repetitive stage, as mistakes might be noticed later while working on the project. No other challenges, such as missing values or inconsistent values, were found. The data is now prepared for subsequent phases, which has laid a solid foundation for the rest of the project. 2. The exploratory data analysis to identify and analyse trends and patterns in the data has been completed. This phase revealed several key insights into the trends in student application, acceptance, and registration rates over the years. Identifying these trends provided a deeper understanding of the university's performance to make easy comparisons with the performance of other institutions in the United Kingdom. This phase also involved generating descriptive statistics and visualisations, which allowed for easy interpretation of the data. 3. A literature review to gather a wide range of similar studies that explore using data-based decision-making to improve educational institutions was initiated. So far, 10 articles have been identified and reviewed to provide a foundation for the literature review section of the final project report, contextualising the project, and ensuring that the project is aligned with existing academic work. Although a proper summary of these articles needs to be done, and more articles need to be reviewed. The literature review task is ongoing, but early findings indicate a gap in studies directly relating to this project topic specifically, which strengthens the relevance of this project. To counter this, a wider range of research topics are considered. 4. A roadmap for the project tasks has also been outlined, ensuring that there is a clear structure and timeline for completing this project. There are several tasks that remain unfinished, and they will require significant attention in the coming weeks. |
| **Summary of tasks to be completed (minimum 250 words):** Provide a summary of key tasks remaining to be completed. These should include both practical work and report writing.   1. The task of developing an interactive dashboard to visualise the key trends and patterns identified during the exploratory data analysis phase is still pending. This dashboard is essential for making the data accessible to the University of Buckingham in an easy-to-understand format. Tableau Desktop will be used to create this dashboard. To enhance the effectiveness of this task, I will focus on developing user-friendly visuals and include filtering options that allow users to interact with the data. 2. The next step involves building and validating a predictive model to forecast the quantity of student applications needed to be accepted to aid in enhancing registration outcomes. The success of this task will depend on the selected model's ability to accurately predict the number of applications that will result in successful registrations. A comparison of various models will be conducted, and the best-performing model will be selected based on the score of evaluation metrics. 3. While similar articles have been identified and reviewed, the task of summarising these articles for the literature review section of the final project report is still ongoing. To ensure the literature review adds value to the project, I will focus on drawing connections between existing studies and the project’s objectives. This will help demonstrate the newness of the approach being used, particularly because a wider range of research topics are considered due to the gap in studies directly relating to this project topic. 4. The final task involves compiling the findings and outcomes of the project into the final project report, following the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology. This report will include all phases of the CRISP-DM methodology. The report will be designed to be clear to both the technical and non-technical audiences. |
| **Risk assessment:** Provide a risk assessment for the key tasks remaining to be completed. Use work-place practice for risk assessment if applicable.   1. There are currently no risks that will hinder the completion of the remaining tasks. |
| **Reflections (minimum 500 words):** Provide a self-assessment of project’s progress and make suggestions to enhance the work over the remaining period.  The project has made steady progress, with several key tasks, including data extraction, data preprocessing, exploratory data analysis, and some aspects of the literature review, achieved. These tasks were done carefully and diligently, although so much time was spent on the data preprocessing since it is a repetitive stage because mistakes were made. However, there are still other important tasks to be completed, and adjustments will be made to ensure that the remaining work is completed within the project timeline. The incomplete tasks and suggested adjustments to be made are as follows:   1. **Dashboard Development**: This involves creating an interactive dashboard to visualise key trends and patterns identified during the exploratory data analysis. To enhance the efficiency of this task, I plan to use pre-built templates and draw inspiration from similar dashboard designs, which will save time and ensure a high-quality final dashboard framework. 2. **Predictive Model Development**: This involves building and validating a predictive model to forecast the number of student applications that will result in successful student registrations. The model's ability to successfully predict the quantity of student applications needed to be accepted to aid in enhancing registration outcomes is essential for achieving the project’s objectives. This phase may require several iterations of model hyperparameter tuning and validation to optimise performance. 3. **Literature Review Completion:** While the literature review is well underway, the process of summarising and synthesising the articles is ongoing. Using Zotero, a reference management tool, I will organise and streamline the literature review process, ensuring that all key findings are captured and aligned with the project’s framework. 4. **Final Project Report (Using CRISP-DM Methodology)**: The final stage involves compiling the findings and outcomes into the final project report, following the CRISP-DM methodology. One of the challenges in this phase is ensuring that the report is clear to both the technical and non-technical audiences. To address this, I will produce drafts of each section and seek feedback from my supervisor to refine the content. Additionally, I will ensure that the report adheres to the University of Buckingham's formatting guidelines and includes visual aids to support the findings.   A summary of the suggestions for enhancing the project work over the remaining period is as follows:   1. **Time Management**: With several important tasks still remaining, strict adherence to the project timeline is absolutely necessary. Allocating specific time frames to each individual task will help ensure steady and consistent progress. This structured approach will allow better tracking of milestones and ensure that no task is delayed. 2. **Feedback**: Regular and consistent communication with my supervisor, Prof. Harin Sellahewa, will be vital in receiving valuable feedback. This will help in identifying any issues or areas of improvement early on, allowing for necessary adjustments to be made to the project in a timely manner. 3. **Tools Mastery**: Gaining full proficiency in using Tableau for developing the interactive dashboard and experimenting with various machine learning models in Python will be key to delivering a high-quality predictive model and effective visualisations. Additionally, mastering Zotero will help organise and streamline the literature review process, speeding up progress. |

|  |  |
| --- | --- |
| **Signature (Learner\*)** | 2321479@buckingham.ac.uk |
| \*e-signatures or email confirmations are sufficient | |