Project Handover Document

**Redback Operations**

Project 10: Workout Analysis Model and Data Visualisation

*Trimester 3, 2022*

Table of Contents

[1. Project Information 3](#_Toc122170536)

[1.1. Company Acting Director 3](#_Toc122170537)

[1.2. Project Team 3](#_Toc122170538)

[2. Project Overview 3](#_Toc122170539)

[3. User Manual 3](#_Toc122170540)

[4. Completed Deliverables 4](#_Toc122170541)

[5. Roadmap 4](#_Toc122170542)

[6. Open Issues 5](#_Toc122170543)

[7. Lessons Learned 5](#_Toc122170544)

[8. Product Development Life Cycle 5](#_Toc122170545)

[8.1. New Tasks 5](#_Toc122170546)

[8.2. Definition of Done 6](#_Toc122170547)

[8.3. Task Review 6](#_Toc122170548)

[8.4. Testing 6](#_Toc122170549)

[8.5. Branching Strategy 6](#_Toc122170550)

[9. Product Architecture 6](#_Toc122170551)

[9.1. UML Diagram 6](#_Toc122170552)

[9.2. Tech Stack 7](#_Toc122170553)

[10. Source Code 7](#_Toc122170554)

[11. Login Credentials 8](#_Toc122170555)

[12. Appendices 8](#_Toc122170556)

# Project Information

## Company Acting Director

**Dr Akansel Cosgun**

*Senior Lecturer, Emerging Technology Robotics*

[akan.cosgun@deakin.edu.au](mailto:akan.cosgun@deakin.edu.au)

## Project Team

**Project 10: Workout Analysis Model and Data Visualisation**

Redback Operations

# Project Overview

Whether a bike user is using an indoor trainer, cycling outdoors, or using Redback Operation’s VR (Virtual Reality) product, being able to provide post-workout analysis and succinct data visualisations to the user is a critical component in the user experience. Being able to develop analytical models and tools to analyse workouts will also benefit Redback Operation’s ability to develop popular/in demand structured workouts. Furthermore, by beginning to analyse user data, Redback Operations can begin ranking user vs users thus creating a competitive in-game environment.

The project is divided into three parts:

1. Data Acquisition
2. Data Analytics and Reporting
3. Implementation of ML algorithms to predict key cycling performance metrics

What are the deliverables?

1. Thorough Project documentation.
2. Thoroughly conduct competitor research specific to the project scope.
3. Organised data location of procured data.
4. Prototype of data visualisation within Tableau concerning post workout analysis.
5. Deployment (and testing) of one predictive learning model.
6. Thoroughly document key handover materials and upload to the DSA GitHub repo.

# User Manual

Full documentation can be found here:

<https://github.com/redbackoperations/data-analysis/blob/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/Redback%20Operations%20Project%205.pdf>

The manuals step through the process of data acquisition to the point data is organised into a data base in preparation for data analysis and ML applications.

A loom video has been produced to provide greater detail about key elements of the project.

# Completed Deliverables

1. Data Acquisition and Data Organisation – Able to pull fit files from a bike computer, convert into readable format, stored and organised within a databased in preparation for data analysis/handling: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/1.%20Dataset%20Procurment>
2. BigQuery Python API – Able to query the database (Google BigQuery) to streamline and optimised data analysis efforts: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/2.%20BigQuery%20Python%20API>
3. Dashboard for out workout/session activities – Using the acquired data which is now organised in BigQuery; an attempt to visualise a User’s single activity was made: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/3.%20Dashboard> and live dashboard: <https://datastudio.google.com/u/5/reporting/72bf3538-2b54-4ab7-95a7-c8517cc449e7>
4. Workout Data Analysis: Data analysis was conducted to further understand the data acquired and begin to articulate both relationships between numerous data points in the hop of forming conjectures that could be used and inform various ML model selections and implementations. <><><><><><
5. Machine Learning (ML) Algorithm Implementation: Three different models were explored, implemented and assessed: <><><><><><

# Roadmap

List of features to be completed in the future:

1. Further optimised the Random Forest implementation.
2. Attempt implications of LSTM and ANN models for predicting Power Output (wattage).
3. Included Predictive Analysis in Looker Studio reporting.
4. Pivot away from Looker Studio and deploy reporting with Tableau using custom python scripts in preparation to handover to the Web and UX teams to assess.
5. Further organise and streamline the data stored in BigQuery: The database needs refinement and restructure.
6. Research how Rider dynamics or session dynamics can vary from User to User, and how this could impact (if any) predictive models.
7. Streamline Google Cloud assets within a centralised Redback Operations account/

# Open Issues

List all of the issues and challenges that the team is still facing, and any progress that has been made so far to address them.

The purpose of this section is to flag things that may interfere with the future teams’s ability to work on the project, and to give advice as to how these issues could be fixed in future.

Here are some examples of Open Issues:

* Software compatibility issues that arise when members of the team use different version of software.
* An unclear process for reviewing completed tasks on Trello, leading to a backlog of work that is sitting somewhere between unfinished and finished.
* An essential team member had to leave the team with no notice, and there is currently a skill void in their place.
* ML test and train needs to be expand – Currently the approach is basic and needs to be opened up to allow for further refinement

# Lessons Learned

List key lessons learned from the project this Trimester and what you recommend future teams should do differently. You must also explain why you believe this to be the case.

In particular, try to think about processes or technology that you would recommend be changed in the future; things that an uniformed team may mistake for a good idea at first, but later learn to be ineffective.

For example, maybe your team had challenges communicating their progress during panel presentations, but towards the end of the Trimester, you developed an effective method for conveying progress accurately. This would be a great thing to talk about.

# Product Development Life Cycle

We work as a team and focus on delivering measurable values to the project and the company.

We have 3 daily stand-ups each week (start/mid/end of a week’s meeting) to sync the working progress between different team members. We discuss any issues and potential extra works we need to do for the project during the stand-ups. We’ve also had lots of ad-hoc chats in Teams app whenever we need a quick discussion.

We plan tasks in the [Trello board](https://trello.com/b/NSuF3z83/data-analytics) and indicate our progress from there. We create PRs frequently and get them merged timely by the team lead.

## New Tasks

We come up with new tasks along the way while we are working on existing planned tasks or from each stand-up meeting time. Any new tasks will be created in the [Trello board.](https://trello.com/b/NSuF3z83/data-analytics)

## Definition of Done

A DoD list is normally clearly defined in each Trello card, so the card assignee will be able to know exactly when a task is treated as completed by meeting all the DoD items. Additionally, we also have different status labels on each task on the Trello board to indicate their completeness.

## Task Review

Since only the team lead has the permission to merge a PR in GitHub, a task is treated as done when its associated PR is reviewed and merged by the team lead.

## Testing

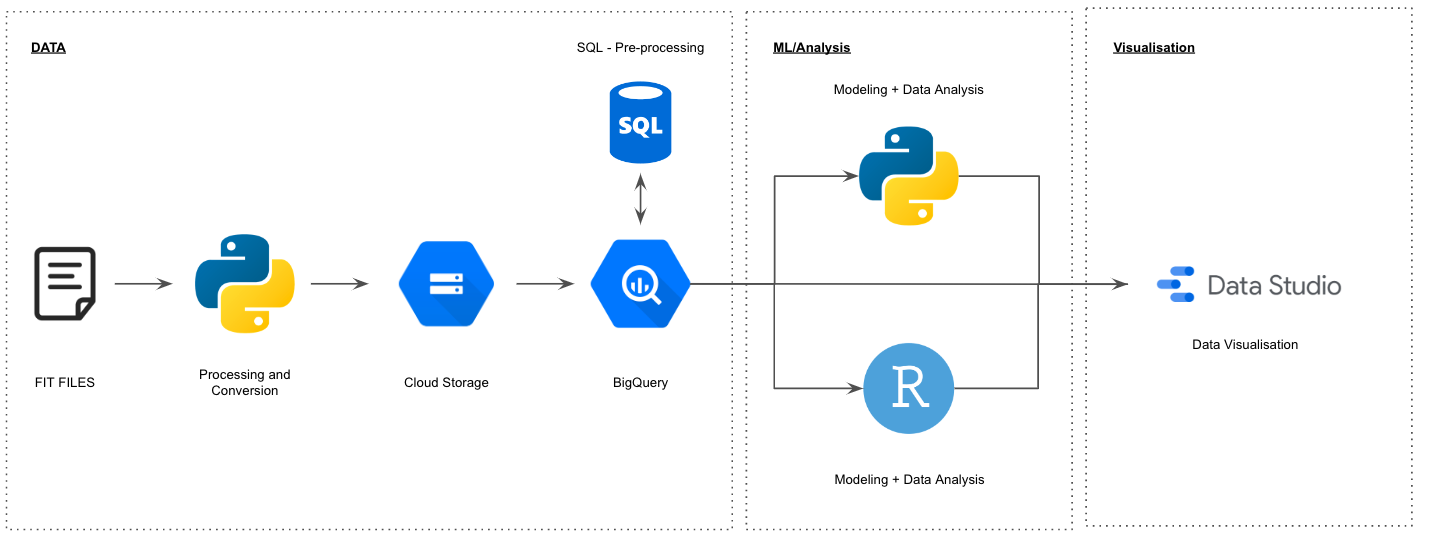
This Project was tested manual due to time constraint and the urgency (limited window of time) Trimester 3 bought While we doing manual tests for a specific task, We’d ensure all DoDs are met before marking it to be done.

## Branching Strategy

We never directly push any changes into the company’s `main` branch. To make any changes, we either create a new branch based off the latest `main` branch or fork the company’s `main` branch into our own repo. After we’ve finished the changes, we create a PR against the company’s `main` branch, and have it reviewed and merged by the team lead. We also ensure to resolve conflicts (if there’s any) before merging back to the latest `main` branch.

# Product Architecture

## UML Diagram



## Tech Stack

1. Google Console – Chosen due to infrastructure alignment requirements.
   1. BigQuery – Needed for database management and query abilities
   2. Cloud Storage – Needed to house large (200mb CSV) datasets
2. PyCharm / Python
   1. Scikit-Learn as the key library – Used for all algorithms
   2. Python 3.11.0 – Python was chosen ahead of R due to keeping things simple i.e., Python was used for the FIT file conversions.
3. Github – File management
4. Google Looker Studio – Dashboard and Visualisation: Selected to keep things simple and within the Google environment.
5. MS Excel – Basic data handling and csv output.

# Source Code

Refer to these locations for all source code:

Data Acquisition

1. .FIT conversion script: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/1.%20Dataset%20Procurment/Python%20Conversion%20Script\>
2. Table schema: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/1.%20Dataset%20Procurment/DataSchema_JSON>
3. Key SQL scripts: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/1.%20Dataset%20Procurment/SQL>

BigQuery API Python

1. Python Script: <https://github.com/redbackoperations/data-analysis/blob/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/2.%20BigQuery%20Python%20API/BigQueryPythonRedback.py>

Data Analysis

1. Initial BigQuery Python Data Analysis - <https://github.com/redbackoperations/data-analysis/blob/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/4.%20ML%20Models/1_bigquery_stat_analysis/main.py>
2. Further Data Analysis Python - <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/4.%20ML%20Models/2_workout_data_modelling>
3. Final Data Analysis Python - <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/4.%20ML%20Models/Data%20Analysis/3_WorkoutAnalysis_continued>

Machine Learning

1. Multi Linear Regression: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/4.%20ML%20Models/Models/Multi-Linear-Regression-Workout-Data>
2. Elastic Net Regression Models: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/4.%20ML%20Models/Models/Elastic-Net-Regression-Models-Workout-Data>
3. Random Forest: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022/Project%205%20Workout%20Analysis%20and%20Model/4.%20ML%20Models/Models/Random-Forest>

# Login Credentials

Details can’t be provided due to this repository being public

Contact: Mark Telley for access to:

1. BigQuery
2. Google Cloud
3. Looker Studio

# Appendices

Trimester Three – Trello Board:

* [Click Here](https://trello.com/b/NSuF3z83/data-analytics)

Trimester Three – Projects, Documentation and Everything Else:

* [Click Here](https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022)

All completed tasks/documents from T3 2022 can be found here: <https://github.com/redbackoperations/data-analysis/tree/main/Trimester%203%202022>