Project Handover Document

City of Melbourne Open Data Playground

Trimester 3, 2022

Table of Contents

Project Information

Client/Product Owner/s

Mr Remi Cattier

Data Hub Lead

remi.cattier@melbourne.vic.gov.au

Mr. Will McIntosh

Senior Advisor to the CEO

will.mcintosh@melbourne.vic.gov.au

Academic Mentor/Supervisor

Dr Hourieh Khalajzadeh

Senior Lecturer, Software Engineering

hkhalajzadeh@deakin.edu.au

Project Team

Student ID	Full Name			Team Group/Role
216019732	HANNAH SMITH	Senior	Undergrad	Project Lead / Data Science & Engineering Team Leader
221377945	JACK PHAM	Senior	Postgrad	Data Science & Engineering
220410831	BARKHA JAVED	Senior	Postgrad	Data Science & Engineering
219274937	ADAM JAMES BULLIVANT	Junior	Undergrad	Data Science & Engineering
221491966	BASILIA CARMEL SETHU	Junior	Postgrad	Data Science & Engineering
222058299	MICK WIEDERMANN	Junior	Undergrad	Data Science & Engineering Assistant Team Leader
222053804	ANGIE HOLLINGWORTH	Junior	Undergrad	Data Science & Engineering
219297735	CALEB WEBB	Senior	Undergrad	Project Lead / Web & App Development Team Leader
222296654	MARCEL GEBARA	Senior	Undergrad	Web & App Development
219302799	JANITHA (JAY) THUSHARA NILAWEERA PATABANDIGE	Senior	Postgrad	Web & App Development
218322739	WILLIAM MCLINDEN	Senior	Undergrad	Web & App Development
220460361	ZHENGMIN ZHU	Senior	Postgrad	Web & App Development
221484111	TATE SEBASTIAN REMZI-JOHNSON	Junior	Undergrad	Web & App Development
221308586	PRAMODYA SATHSARANI SENANAYAKA	Junior	Postgrad	Web & App Development

Project Overview

In trimester 2 2021 the project's scope was redefined by the client. This introduced the current problem statement:

The City of Melbourne has been an Australian leader in Open Data since 2014. The latest research and local user engagement have identified a gap where users would like access to Open Data example tools so that our users can re-use these tools in their apps and city solutions.

In trimester 3 2021, the client identified two key benefits that this re-definition of the scope would deliver for this project:

- The toolkit examples created will be developed to maximise efficiency and time saving for the council's staff. This will save council time through the reduction of calls and emails relating to questions of how to use council's Open Data platform API.
- The use cases lists are the most common customer requests for assistance in creating a data solution using City of Melbourne's open data/API.

In Trimester 2 2022 this project continues to deliver a proof-of-concept example of how calls to Open Data API can be made to deliver a solution. Figure 2.1 illustrates the major components of the Melbourne Open Data Playground platform.

Building on the foundation of work completed in the previous trimesters, our project focussed on delivering five key outputs:

1. Implementing the use cases developed in trimester 2 2022
2. Creating new use cases
3. Preparing to migrate to the new APIs following the client transition
4. Implementing the new Melbourne Open Playground branding, including changes to how the use cases are styled

Figure 2.1: Melbourne Open Data Platform High Level Architecture

Melbourne Open Data Playground
High Level Architecture
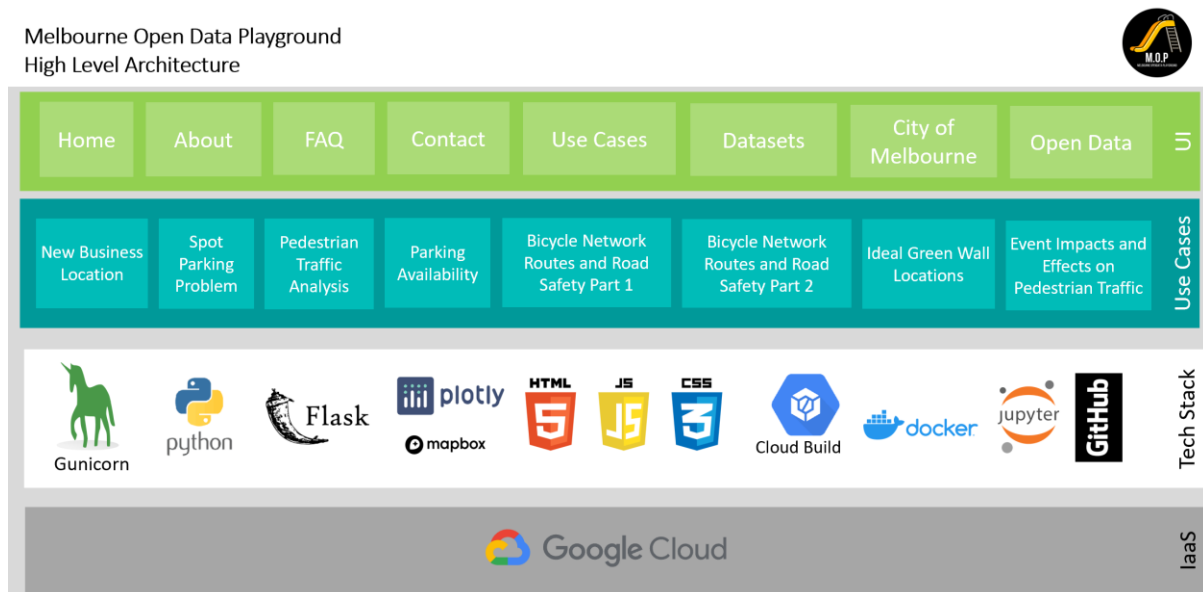
This trimester saw the data science and engineering team create multiple new use cases within three work streams: Transport and Safety, Environment and Wellbeing and Business and Activity. These work streams were altered slightly from Trimester 2 in line with the concepts underpinning the proposals for use cases this trimester. The use cases explored and analysed a variety of datasets available from Melbourne Open Data. Utilising python libraries such as Selenium, folium, Plotly, and mapbox, our data science team was able to deliver highly stylized, interactive tools to help in developing business insights using the City of Melbourne's Open Data API.

The web development team focused on implementing the design team's Figma wireframe that was produced in the previous trimester. This involved new key navigation pages such as home, about, FAQ, and contact us. Featuring a new light/dark mode functionality to improve accessibility. the codebase manageability, by centralizing all use case CSS into a single style sheet.

The design team's work was suspended for this trimester due to no seniors being available to continue the work. Juniors joining the design team in Trimester 1, 2023 should refer to the handover documents from Trimester 2 2022 in the company GitHub.

The work completed this trimester leaves the project and the students coming onto the project next trimester in a great place to continue with the project. This will allow next trimesters team to fast track the implementation of new use cases, implement site updates and further the project.

Public GitHub Repository

In addition to the Melbourne Open Data playground, the project has also curated a public-facing GitHub. In this environment we showcase python code that supports and extends the use cases integrated onto the web portal.

The Jupyter notebooks are shared for the benefit of those who may be more technically motivated to explore and build upon our examples.

Readme pages help users to navigate to and explore these works. This was created to aid in building a community around the playground and Melbourne Open Data more generally. Figure 4.1 below illustrates the home page of the GitHub site.

Figure 3.1: Public-facing GitHub repository
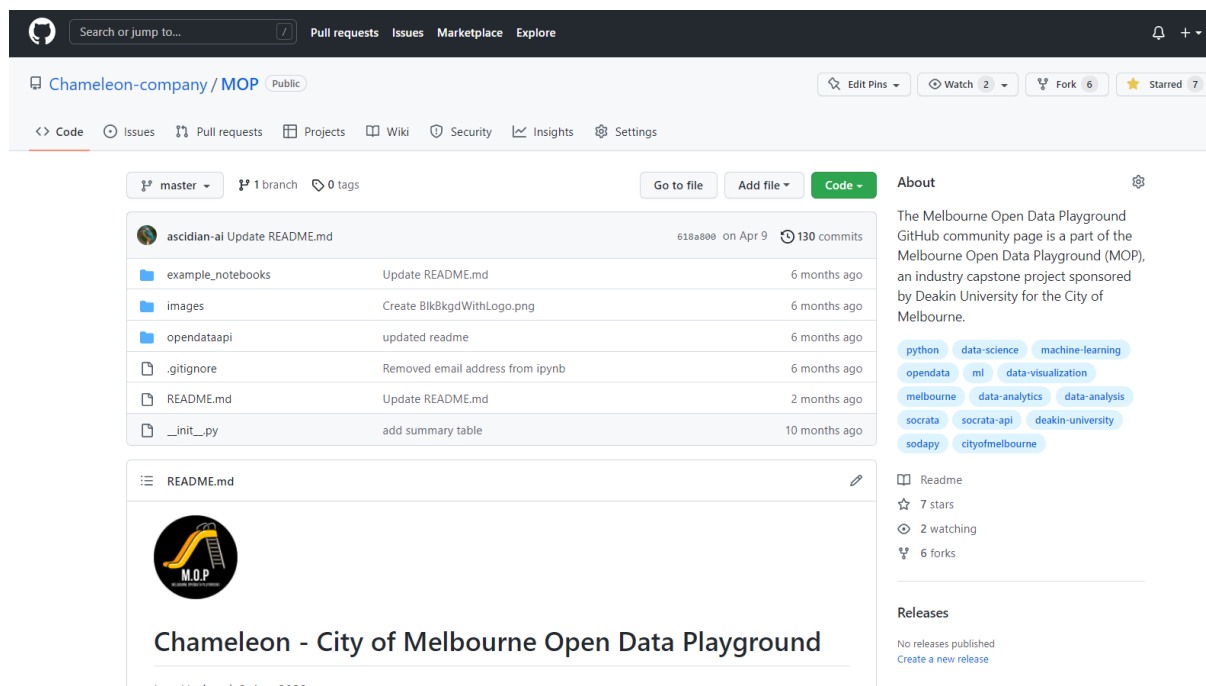
Completed Deliverables

In this section we describe the key deliverables as well as the main teams and contributors responsible.

Domain 1: Transport & Safety

Use Case 1 Impacts of high-rise residency and demand for public transport (Basilia and Jack)

This use case provides an analysis of the existing public transport infrastructure and their locations relative to high-rise dwellings as well as identifying where future public transport infrastructure should be developed to sustainably support population growth within the City of Melbourne. The analysis identified Docklands does not have easy access to any form of public transport, with the nearest stops being more than a 10-minute walk. Docklands is also expected to have the highest number of new dwellings in 2040 based on current available forecasts. Future work for this use case could include identifying if income levels have an impact on demand for high-rise dwellings and public transport.

Trello link | **GitHub link**

Use Case 2 Small Area Population Growth & Active Transportation Needs Analysis (Angie and Mick)

This use case analysis looked at the predicted population growth in the City of Melbourne suburbs over the next 20 years alongside the Super Tuesday bike route usage counts. They identified areas with high population growth (percentage) in five-year buckets, and the active transportation traffic passing through those areas. This analysis can be used by the council when considering upgrading facilities or future residents assessing suburb liveability. There were also enough unused data points in both datasets to provide a more in-depth analysis such as; traffic direction, and residents' age, or to include Super Sunday survey data for a weekday versus weekend analysis.

Trello link | **GitHub link**

Use Case 3 Walkability map of the City of Melbourne (Hannah)

This use case provides a detailed analysis of the walkability of streets within the City of Melbourne using existing research metrics: population density, ease of movement, and the density of essential amenities and public transport. It extends these metrics to include an

analysis of traffic conditions for adjacent roads, tree cover, and footpath steepness. This use case could be updated to combine data from the Pedestrian Safety Use Case, to provide a more detailed picture of how changes to roads can be made to impact pedestrian safety.

Trello link | **GitHub link**

Domain 2: Environment & Wellbeing

Use Case 1 Liveability analysis by location, identifying current and future factors that impact residents' wellbeing (Adam & Basilia)

The liveability analysis use case was an insightful investigation into the specific suburbs and blocks around Melbourne regarding their liveability characteristics. Such characteristics included access to housing, education, health services, as well as job opportunities. The robust analysis provides users with visualisations of each liveability characteristic by each suburb and block around Melbourne, ultimately allowing them to dive deeper into which areas are better to live in Melbourne. To extend this use case, further datasets could be incorporated based on extra liveability components, such as crime rates by suburb, or access to public transport.

Trello link | **GitHub link**

Domain 3: Business & Activity

Use Case 1 Entertainment location projections (Barkha & Jack)

Utilise the 2021 CLUE survey data recently released in Dec 2022, for bars pubs and taverns and café restaurants and bistros. The entertainment venue capacity based on number of seats and number of patrons, and the planned works and activities data, is used to provide statistics on clue small areas. In addition, pedestrian traffic is used to evaluate usage of entertainment locations during day and night, and to predict growth of entertainment locations.

Trello link | **GitHub link**

Published Use Case: New Business, Pedestrian Traffic Day and Night

Use case from last trimester published to the Open Data Playground by the Web Dev team. We validated and verified during the release.
Trello link

Technical Documentation

Onboarding for new members, including checklist and environment set up (Adam)

Two new documentation pieces were added for use within the data science team that help juniors in their first few weeks. The 'Onboarding Checklist' is a single source document that contains all relevant links, documents, and tutorials for juniors to understand Chameleon, as well as the CoM project. The 'Setting Up Local Machine' is a tutorial that students can follow to setup their environment and execute previous use cases as well as create their own through Git.

Trello link | **GitHub link – Onboarding Checklist** | GitHub link – Setting Up Local Machine

Style guide for Python notebooks (Template and addition to Use Case Publishing Guide) (Hannah)

Created a use case template with new style guidelines for colours (for accessibility and compatibility with the site styling) and also instructions for how to implement the site-wide styling using the CSS classes. The style guide was also added to the "Use Case Publishing Guide" created by Brendan in Trimester 2.

Trello link | **(template)** | **(updated use case publishing guide)**

Cross team (Web Development and Data Science)

CSS class-based styling for use cases (Caleb, Hannah, Mick)

A single style document for all use cases was created, and using this a style update matching the new look for the site has been implemented using CSS classes, updating the previous styling in the HTML tags of individual notebook cells. This will allow future styling updates to be implemented without the need for the data science team to re-export the notebook HTML of published use cases. The new classes are included in the style guide and template for use cases, and all published use cases have had the HTML files updated on GitHub (yet to be updated on the live site).

Trello link | **GitHub link**

Web Development Team

Publish Pedestrian Network Routes and Road Safety Part 1 (Caleb)

In the previous trimester, the data science team had created this use-case template for the web development to publish to the web-app. This document is exported from Jupyter as HTML and handed over to us. We utilized the markdown guide created by earlier cohorts as a guide. This was completed over a 1-2 week period.

Trello link | **GitHub link**

Publish Ideal Green Roof Locations (Marcel, Tate & William)

In the previous trimester, the data science team had created this use-case template for the web development to publish to the web-app. This document is exported from Jupyter as HTML and handed over to us. We utilized the markdown guide created by earlier cohorts as a guide. This was completed over a 1-2 week period.

Trello link | **GitHub link**

Publish Evaluate Business Locations Using Pedestrian Traffic Day / Night (Marcel, Tate & William)

In the previous trimester, the data science team had created this use-case template for the web development to publish to the web-app. This document is exported from Jupyter as HTML and handed over to us. We utilized the markdown guide created by earlier cohorts as a guide. This was completed over a 1-2 week period. We then made some further amendments at the request of the data science team.

Trello link | **GitHub link**

Centralize use-case CSS (Caleb)

This was implemented to reduce the complexity of adding new use-cases to the production web app. Removing all inline CSS and creating a static stylesheet for a more centralized code base, eliminating several steps for future implementations.

Trello link | **GitHub link**

Implement Figma Wireframes (Caleb, Marcel, William & Tate)

Last trimester, the design team produced new site wireframes in Figma. At the start of Sprint 2, the team spent some time testing ideas and different methods of implementation. Once we had decided on a format and style, we implemented to each of the navigation pages (home, about, contact page and FAQ). This was the largest deliverable for trimester 3.

Trello link | **Home, FAQ, Contact Us GitHub link** | **About GitHub link**

Roadmap

Progress in 2022

The Melbourne Open Data Playground is the result of three years of development and industry collaboration with the City of Melbourne to support greater community and industry use of its published open data sets. The current implementation of the Melbourne Open Data Playground web application (site) and its associated GitHub repository came to fruition in Trimester 2 of 2021 and has experienced significant enhancements this trimester (T3 2022). T3 2022 saw multiple new use cases created in addition to a re-styling of the web app.

Future Direction

Having established a process for converting Jupyter notebooks into web application use cases, the platform is now well positioned to realise rapid implementation of future use cases. Future new use cases for implementation will be decided on and created in each new trimester and delivered iteratively over multiple sprints across the remainder of each trimester.

Design Team (from T2, 2022)

1. Complete the Re-branding Guide
   - Finalise the Re-branding Guide and add to handover documentation
   - Provide to client for feedback and assessment
2. Conduct Functionality, Usability, Testing of the Site after new Use Cases are added to ensure all functionality works as intended.
   - Utilize framework report templates created during sprint 1 T1 2022
   - Mobile device testing
   - Using a virtual machine to test different operating systems and browsers
3. Penetration testing of the new GCP deployment of the live site
   - Focus on Python Vulnerabilities
4. Run stress testing on the new GCP deployment of the live site
   - Focus on Python Vulnerabilities

Web Development Team

1. Implementation of use cases created in T3 2022
2. Domain name for the production site

3. Improving mobile device accessibility

4. Further minor cosmetic changes & database fetching time

Data Science and Engineering Team

1. Continue to develop Domain 1: Transport and Safety

2. Continue to develop Domain 2: Environment and Wellbeing

3. Continue to develop Domain 3: Business and Activity

4. Migration of open data to the new APIs provided by the client

5. Migration of flat file data sources into a relational database for improved performance of the search function

It should be noted that all scope shown for 2022 has been proposed by the City of Melbourne Open Data Project team but has not yet been ratified by the City of Melbourne. This roadmap will serve as a starting point for discussion and agreement of scope with the City of Melbourne. Figure 6.1 below illustrates the most significant deliverables completed and anticipated across T1 2022 and into T2 & T3 2022.

Figure 6.1: Roadmap for T2 2022-T3 2022
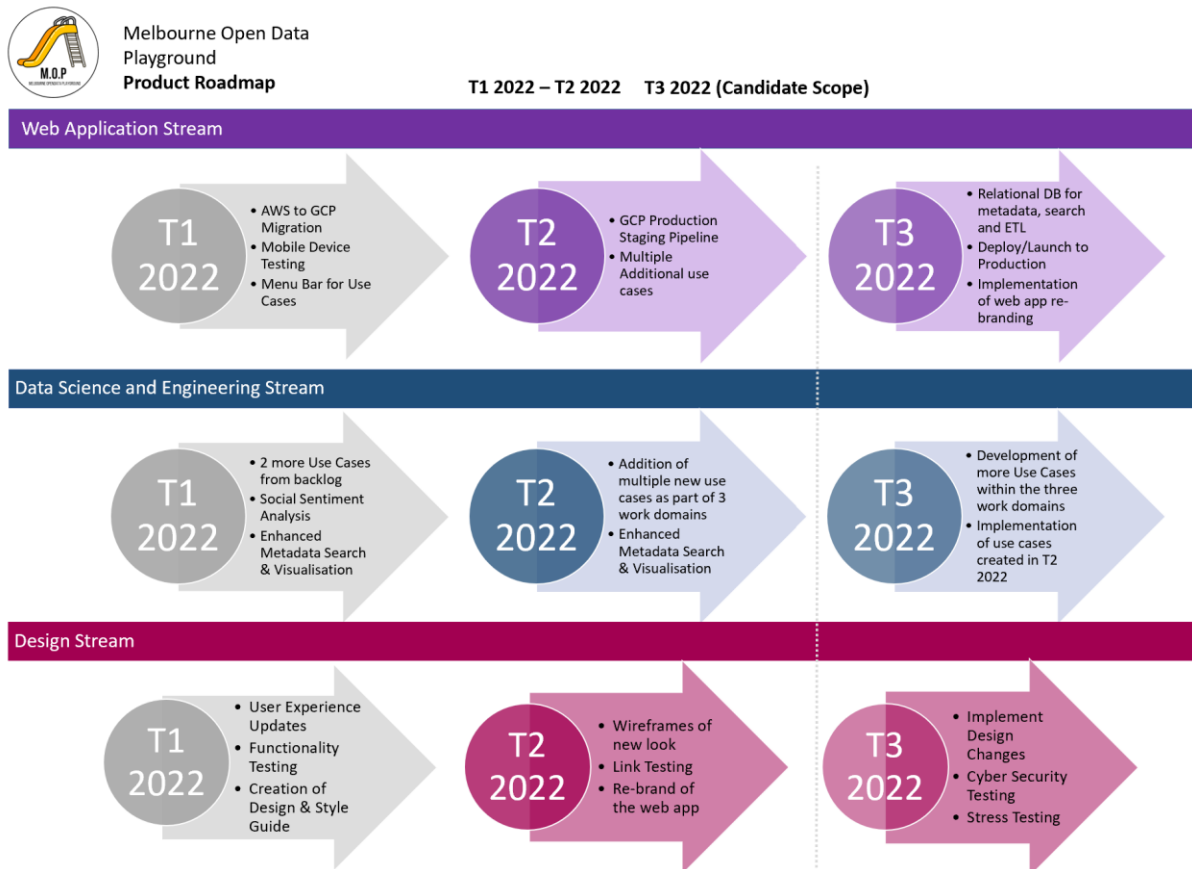
Open Issues

API migration

In mid-December the client began transitioning their open data APIs to a new set of end points, which have been shared with the team. The legacy system will end on December 21, 2022 so it should be a priority for the data science team to upgrade the published use cases to use these new endpoints where possible before re-exporting the HTML to send to the web development team.

Interactivity Concerns in Notebooks

The spot parking problem use case utilises a third-party service (Binder) to initialise interactivity. The time for initialisation ranges from a few seconds up to a minute. The delay has a negative impact on the user experience and is antithetical to the streamlined product that the tribe is hoping to develop. Future squads may wish to explore alternative techniques

to enable this level of interactivity including cloud service hosting of notebooks or other possible solutions. Trimester 1 2022 project team decided that this use case should be retired.

Handling user feedback for website visitors

During the client feedback session at the end of Trimester 2 2022 the client suggested developing a way to handle feedback or queries made by visitors to the MOP website. Future teams may wish to investigate avenues such as building a feedback form into the website, or utilising GitHub's Issues feature.

Product Architecture

Technology Stack

The teams have different workflows and goals along with different technologies and tools. Some common technologies, such as Python and HTML, help with the integration of Data Science team generated use cases into the web application by the Web Development Team.

The technologies used by each team are listed below.

DATASCI Team

- **General - Python, SodaPy, Pandas, NumPy**: The DS Team works with the powerful, open-source Python language and its well-known utility libraries such as NumPy and Pandas. SodaPy is also used to access Melbourne Open Data through the Socrata API.

- **NLP Libraries - PyLDAvis, NLTK, Spacy, Genism:** Some Natural Language processing libraries may be used for future work in the backlog.

- **Geospatial Libraries - GeoPy, Shapely, GeoPandas, Folium, Mapbox & Plotly:** These libraries are used for the various mapping tools and visualisations used in our use-cases.

- **IDE - Jupyter Notebooks:** Jupyter Notebooks are the tool of choice for data scientists working with Python, allowing us to create and share documents that integrate live code, equations, computational output, visualizations, and other multimedia, along with explanatory text in a single document. Notebooks are created by the DS Team for each use-case, which can be converted to HTML by the WEB Team to display on the site.

DESIGN Team

- **Front end - JavaScript, HTML, CSS:** The website content, styling, and functionality such as navbar, search input, and clickable buttons are all coded using these ubiquitous tools.

- **Back end - Python, Flask:** In the backend, the webapp runs on Flask, a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or other components where pre-existing third-party libraries provide these functions.

- **Security - Flask-Talisman:** Talisman is a small Flask extension that handles setting HTTP headers that can help protect against a few common web application security issues.

NMAP, SQLMap, Nikito: Penetration testing tools used with Kali Linux to test the security of the website.

GCP Cyber Security Features: The use to several in-built GCP features; API gateway, Cloud Build, Web app, Google Risk Protection Program and container registry.

WEBDEV Team

- **Frontend - JavaScript, HTML, CSS:** The website content, styling, and functionality such as navbar, search input, and clickable buttons are all coded using these ubiquitous tools.

- **Backend - Python, Flask:** In the backend, the web app runs on Flask, a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or other components where pre-existing third-party libraries provide these functions.

- **Security - Flask-Talisman:** Talisman is a small Flask extension that handles setting HTTP headers that can help protect against a few common web application security issues.

- **Deployments - Docker, AWS, Google Cloud Platform (GCP):** Docker is the most popular solution for creating and working with containers. Currently the team employees the use of GCP and their *Cloud Run* service. There are many other services available in these Cloud Providers that we can use as we expand the project.

- **IDE - Visual Studio Code:** Visual Studio Code is a lightweight code editor which the WEB Team uses for coding in the webapp. It allows for running the Flask server in a local development environment so we can see changes live in a browser, and it also has several useful extensions for working with Git, debugging, etc.

Project Links

City of Melbourne MS Teams Channels Files

Link to the project showcase video

Melbourne City | Trello

Link to the live website

Chameleon Company GitHub

MOP Community repository

MOP Code repository

Chameleon Handover Documentation repository