Onboarding Checklist for New CoM Members

The following documentation provides a checklist that new members of the CoM (Data Science) team can use to understand details and expectations about the CoM project. Juniors and Seniors can use this as a resource to onboard themselves into the team and best prepare their understanding of the project and scope before beginning technical work.

Checklist Instructions

An onboarding checklist is a simple yet effective way for companies to ensure new joiners or employees are engaged in their new role effectively from day one. Chameleon's CoM Open Data Project contains various resources and requirements for students to understand before beginning technical work. New Data Science Team joiners can follow this checklist and mark off tasks in their own time to ensure they hit the ground running. All required resources and tasks are listed below for onboarding into the CoM Data Science Team.

Simply copy this tick ✓ and paste into the "Status" column once the task has been completed.

Onboarding Checklist Part 1 (Administration & Comprehension)

Task	Instructions/Resources	Status
Join Chameleon Microsoft Teams Channels and Chats	Ensure you are in the Chameleon Teams channel as well as the Data Science and Engineering team private chat.	
Join Trello Board	Ensure you have access and have been added to the Melbourne City Trello board. • Melbourne City Trello Board	
Begin Team Worklog	Ask your project lead for access to the team worklog so you can begin logging hours and tasks for the project.	
Read Handover Documentation	Read the handover documentation from the previous trimester. For reference, here is T2 2022 handover: • T2 2022 Handover Documentation	
Familiarisation with GitHub	Read through the GitHub and its structure to understand the repository. Specifically, the MOP-Code contains the Melbourne Open Playground Code, and the "datascience" folder inside contains relevant files to the Data Science team. • Chameleon Repository	
Familiarisation with Client's Data Portal and Strategic Goals	Read through the CoM Open Data Portal, specifically, review different datasets inside "Browse All Data" and understand how to use the data in "Learn how to use our Data". • Open Data Portal Review the Strategic Goals of CoM to understand the future goals and plans for Melbourne. • Strategic Goals Try to brainstorm different ways we can use these datasets to help achieve the Strategic Goals for CoM.	

Onboarding Checklist Part 2 (Technical & Propositions)

Task	Instructions/Resources	Status
Installation of Tools	If you choose to use Conda, download these below. • Anaconda Distribution • Miniconda Download desired Git usage method. • GitHub Desktop • Git from Terminal	
Set Up Local Repository Environment and Use Git	Follow relevant instructions in the links below to setup and clone the MOP Repository onto your local machine. • GitHub Desktop – Tutorial • Git from Terminal – Tutorial	
Set Up Local Python Environment	If you're using Conda, the required environment can be made by following the instructions on the link below. • Setting Up CONDA Environment Furthermore, the entire process of setting up your local environment can be followed below if you are stuck with any of the above installations. • MOP-Code/datascience/documentation/Local Machine Setup There is also file that helps with utility methods to read over: • d2i_tools.py	
Review Data Science Team Documentation	Read and examine all of the Data Science Team documentation. • Data Science Documentation For reference, these are the important documents to review. *1. Guidance on how to generate use cases* *2. Peer review work practices* *3. Use case publishing guide*	
Analyse Existing Use Cases	To understand how previous students have implemented use cases, review and analyse already published notebooks. These can be found on the Melbourne Open Data Playground. • Melbourne Open Data Playground When implementing extensions or new ideas, you don't have to reinvent the wheel – examine how other students have used different libraries on CoM data is the best way to understand the core libraries and datasets you will most likely use.	
Propose Extensions and New Use Cases	Ask your project lead if there is an ideation document where you can propose extensions to existing use cases or propose your own new use case. Remember, carefully review the strategic goals along with available data from the portal to ensure your ideas align with the client's best interests.	

Onboarding Checklist Part 3 (Upskilling)

Task	Instructions/Resources	Status
Git	If you are new to Git, the following resource introduces you to the basics of Git and how to use it. This can be used as upskilling evidence. • Git Essential Training: The Basics If you are relatively familiar with Git, or have done the course above, the following resource provides a more in depth course on Git's features including Branches, Merges and Remotes. • Git: Branches, Merges, and Remotes	
Libraries and Packages Used by Data Science Team	If you are unfamiliar with any of the following libraries/packages, feel free to use upskilling courses online so you can generate evidence of upskilling and begin using them confidently. *Core* • Requests *Data Engineering* • NumPy • Pandas *Data Science/ML* • Scikit-learn *Visualisation and Image Processing* • Matplotlib • Seaborn • Plotly *Open Data* • SodaPy *Geospatial* • PyProj • Shapely • GeoPandas • GeoPy • Folium	

References

Hart D 2020 'How to Clone a Repo in GitHub Desktop, https://www.youtube.com/watch?v=8yqQeTbFZUg

Coding for Everybody 2020 'Git for Everybody: How to Clone a Repository from GitHub', https://www.youtube.com/watch?v=CKcqniGu3tA

Author

Adam Bullivant (version 1, T3 2022)