

## Logbook - Assignment 2 (Mastering Big Data Handling)

Name	Matric No
Nik Zulaikhaa Binti Zuraidi Afandi	A22EC0232
Wong Qiao Ying	A22EC0118

Date	Task	PIC	Search Engine Keywords	Sources/Link
23/5/25	Setting up first meeting	Nik, Qiao Ying		
	Look up category data type	Qiao Ying		
	Find a Dataset in Kaggle	Qiao Ying	Filter: Min 700 MB	<a href="https://www.kaggle.com/datasets/devdope/900k-spotify">https://www.kaggle.com/datasets/devdope/900k-spotify</a>
	Setting up IDE and logbook	Nik		Google Colab: <a href="https://colab.research.google.com/drive/195WtU92RuX6so3w6lBtXNdQntwDj_rmk?usp=sharing">https://colab.research.google.com/drive/195WtU92RuX6so3w6lBtXNdQntwDj_rmk?usp=sharing</a>
	Load and calculate performance of dataset using traditional method (full loads)	Qiao Ying		
	Perform brief inspection on traditional method dataset	Nik		
	Do optimized strategies 1-3 for Pandas library	Nik		1. Load less data 2. Chunking 3. Optimize data type
	Do optimized strategy 4 for Pandas library	QiaoYing		1. Sampling
28/5/25	Do optimized strategies 1-4 for Polars library	Nik		1. Load less data 2. Chunking 3. Optimize data type 4. Sampling
	Create GitHub folder for report and submission	Nik		<a href="https://github.com/Jingyong14/HPDP02/tree/main/2425/assignment/asgn2/submission/Group_Colab">https://github.com/Jingyong14/HPDP02/tree/main/2425/assignment/asgn2/submission/Group_Colab</a>

	Do optimized strategies 1-4 for Dask library	Qiao Ying		<ol style="list-style-type: none"> <li>1. Load less data</li> <li>2. Chunking</li> <li>3. Optimize data type</li> <li>4. Sampling</li> </ol>
29/5/25	Update readme.md in GitHub	Qiao Ying		<a href="https://github.com/Jingyong14/HPDP02/blob/main/2425/assignment/asgn2/submission/Group_Colab/readme.md">https://github.com/Jingyong14/HPDP02/blob/main/2425/assignment/asgn2/submission/Group_Colab/readme.md</a>
	Update 1.1 and 1.2 in report	Qiao Ying		1.1 Dataset Details 1.2 Features Included
1/6/25	Update Task 4 in report	Qiao Ying		
1/6/25	Update Task 3 in report	Nik		
1/6/25	Update Task 5 in report	Nik		
1/6/25	Update Task 5 (reflection) in report	Qiao Ying		