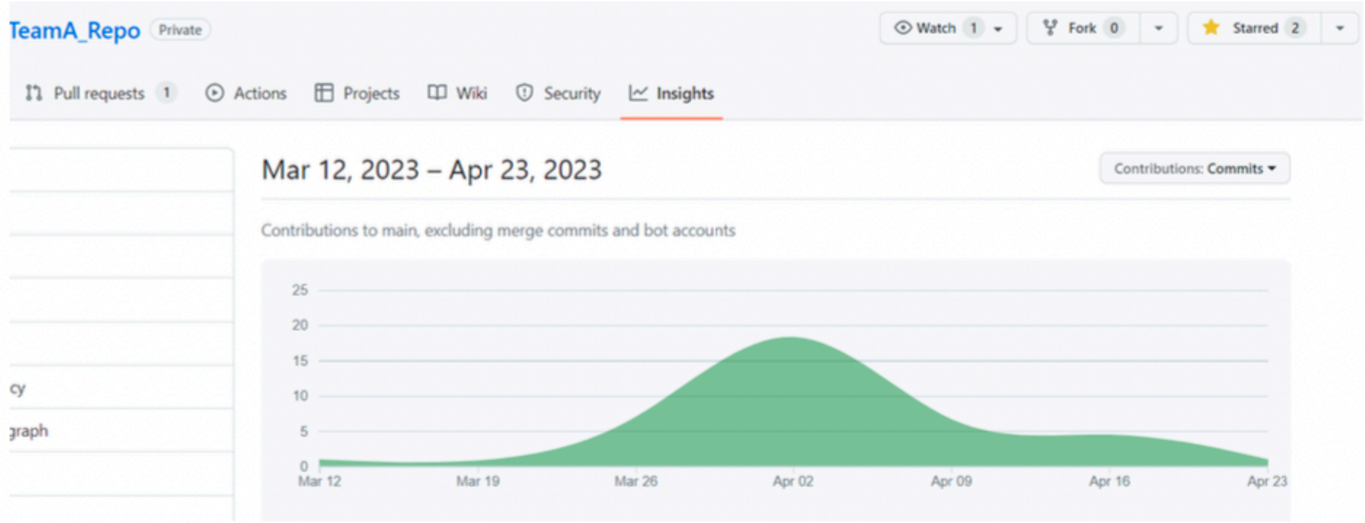
Work Breakdown Report(Team A)

Debmalya Deb
Msc in Data Analytics(School of Computing)
National College of Ireland
x21242101@student.ncirl.ie

Initiating Work

On March 12th, our group began the process of brainstorming project ideas. After careful consideration, we ultimately decided to focus on the topics of immigration, homelessness, and rental struggles in Ireland.

Below is the Github repository

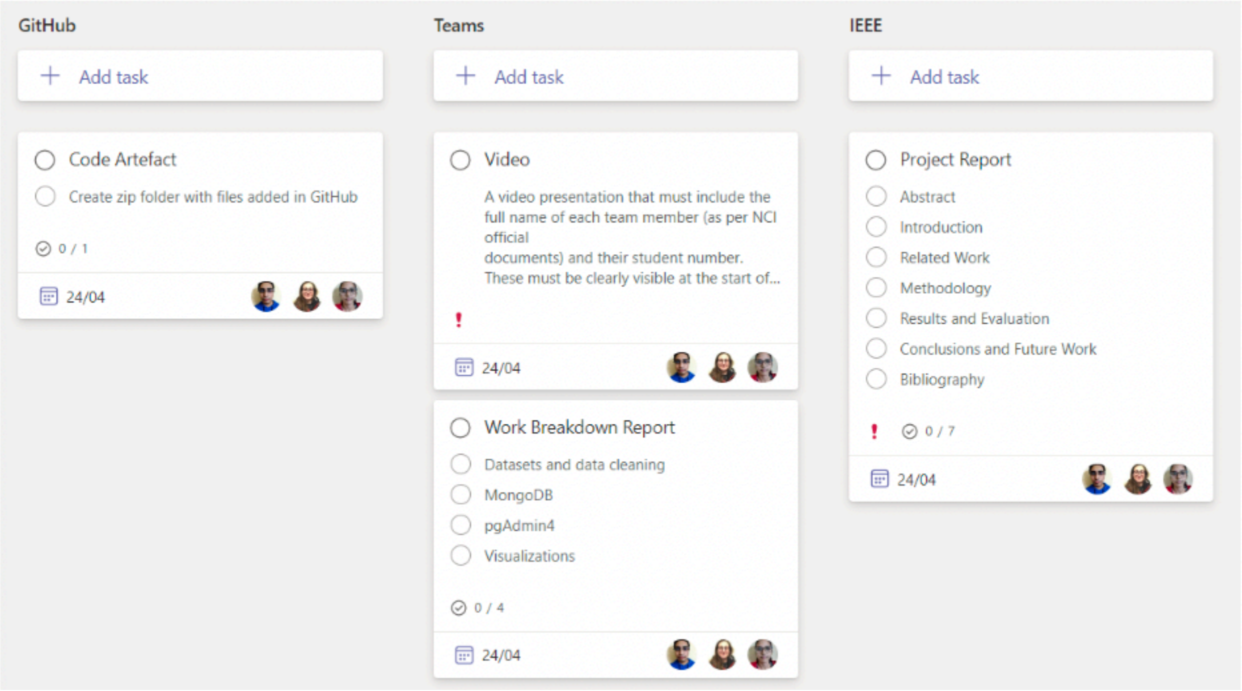


Mar 12, 2023 – Apr 23, 2023

Contributions to main, excluding merge commits and bot accounts

Planner

We utilised a planner to map out our tasks and divide them into various components. Each member of the team was allocated their respective responsibilities.



GitHub

+ Add task

○ Code Artefact

○ Create zip folder with files added in GitHub

0 / 1

24/04

Teams

+ Add task

○ Video

A video presentation that must include the full name of each team member (as per NCI official documents) and their student number. These must be clearly visible at the start of...

!

24/04

○ Work Breakdown Report

○ Datasets and data cleaning

○ MongoDB

○ pgAdmin4

○ Visualizations

0 / 4

24/04

IEEE

+ Add task

○ Project Report

○ Abstract

○ Introduction

○ Related Work

○ Methodology

○ Results and Evaluation

○ Conclusions and Future Work

○ Bibliography

! 0 / 7

24/04

Work Roles and Responsibilities

Initialising a Docker Instance for MongoDB -To create a container for MongoDB and establish a connection with specific credentials, researchers used a combination of two files: docker-compose.yml and mongo-init.js. The former was used to establish the container and the latter to initialise a volume within it by creating a user. The docker-compose.yml file specified the necessary parameters for the container, such as the image to use, ports to expose, and the name and credentials for the MongoDB instance. Meanwhile, the mongo-init.js file was used to create a user for the MongoDB instance, which was specified in the docker-compose.yml file. Together, these two files were used to initialise a Docker instance for MongoDB, allowing the researchers to work with the database in a secure and controlled environment.

Setting up PostgreSQL Docker Instance and Creating Tables for ETL -

- Composed docker-compose.yml file with details such as version, database, image, port, volumes, and credentials
- Created Docker instance within Docker Desktop environment via terminal/command prompt
- Created server in pgAdmin 4 using PostgreSQL Docker container with appropriate user credentials
- Created "TeamAHousingCrisisProject" database for storing mixed structured and semi-structured data using ETL technique
- Created "homeless", "immigration", and "rent" tables under "TeamAHousingCrisisProject" database for storing extracted datasets after ETL process.

Created whole ETL workflow -

- Extract: Python libraries (such as pymongo and pandas) are used to extract data from MongoDB and create three Pandas DataFrames (ImmigrationDataFrame, RentDataFrame, and HomelessDataFrame)..
- Load: CSV files are loaded into a PostgreSQL database using Pandas and SQLAlchemy, with error handling and logging.
- ETL Workflow: The ETL job extracts data, applies transformations, and loads the data into the final target using multiple functions.