

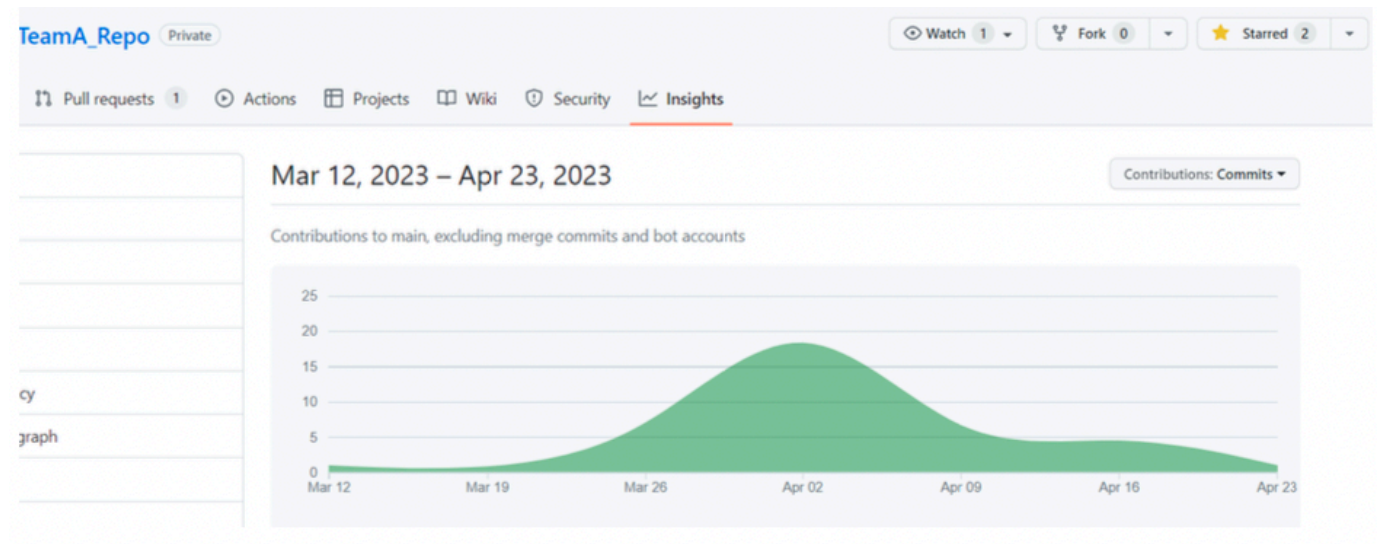
# Work Breakdown Report (Team A)

SAHELI DUTTA (21246513)

## Work Initiation

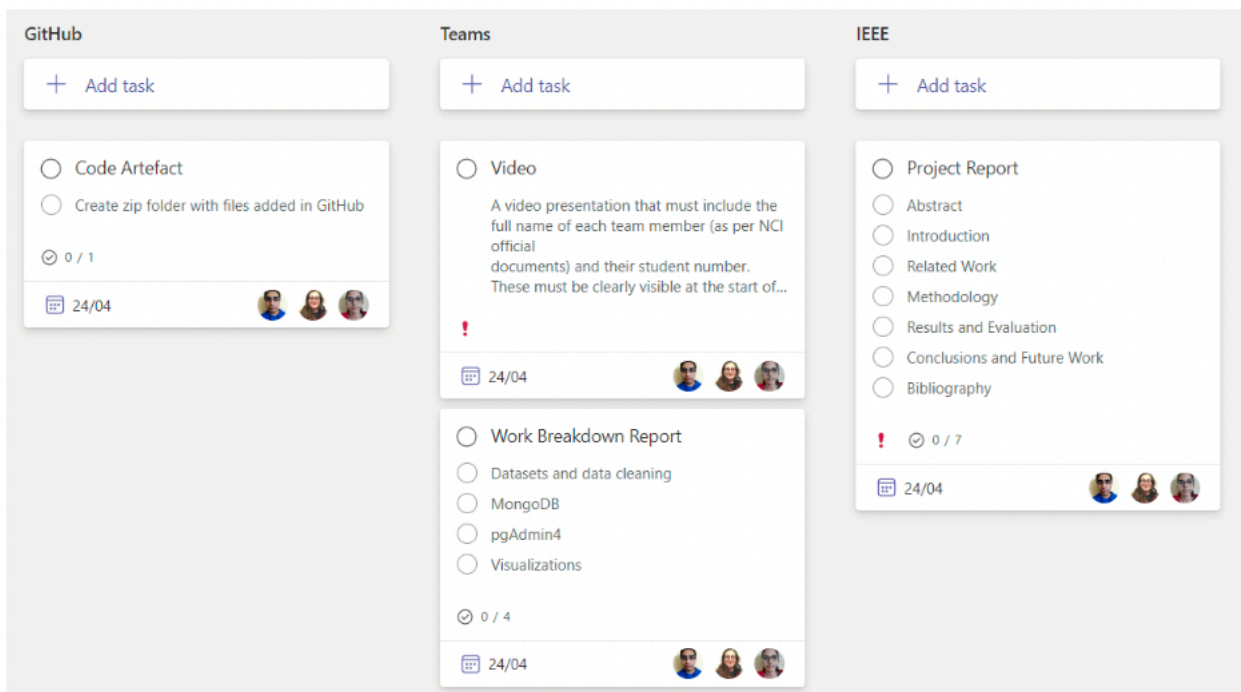
Our group project started on 12th March where we started to pitch our ideas. Finally, we choose to work on immigrant , homeless and rent issue in Ireland.

Below is our Github repository



## Planner

We used planner to chalk out our work and segment it to several parts. Each one of us assigned



to their respected roles.

# Work Roles and Responsibilities

- A. Review Data Cleaning Code and debug lines** - I have helped to review and debug in data cleaning file.
- B. Uploaded the three datasets to MongoDB** - I wrote code to create a MongoDB client and connects to a database named "housing" using the username "dap" and password "dap". It then creates a collection named "immigration" within the database. The code reads the contents of a CSV file named "new\_emigration\_and\_imigration.csv" and inserts each row from the CSV file into the "immigration" collection in MongoDB using the `insert_one()` method. Same process has been repeated for 2 other datasets. For the Rent data set, after cleaning and filtering the data was stored in new rent dataset. The rent data set was loaded into "housing" database.
- C. Wrote Sql Query to Merge three datasets** - This SQL query selects data from three tables: "homeless", "immigration", and "rent". It joins the tables using the common identifier between "homeless" and "immigration", and between "homeless" and "rent". The query returns data for each record where an immigrant's nationality and gender match a homeless person's record. The data includes the region where the homeless person is located, the number of male and female homeless adults, the immigrant's nationality, gender, and year of immigration, as well as the location and cost of the rented place.
- D. Visualisation and Analysis** -
- In the first visual, The code creates a scatter plot using Python's Plotly Express library. It visualizes the relationship between rent cost, immigrant population, and homeless adults, with marker size representing the number of homeless adults. The plot also uses color mapping and a hover tooltip to display rented location names. The title and axis labels are customized, and the final plot is displayed using the `show()` method.
  - In the second visual, The code groups a dataset by year and region, calculates the mean rent cost for each group, and creates a line chart using Plotly Express to visualize the average rent cost over time for each region. The chart is color-coded and includes axis labels and a title, and is displayed using the `show()` method.
  - In the third visual, The code creates a box plot of rent costs by region using data from a dataframe. The plot's x-axis represents regions, and the y-axis shows the average rent cost in euros.
  - In the fourth visual, The code creates a histogram of rent costs by homeless region using data from a Pandas DataFrame called "housing\_df". The histogram has 20 bins and uses a color map to represent the homeless region. The resulting plot is displayed using the `show()` method.
  - In the fifth visual, The code groups data by region and calculates the sum of homeless male and female adults for each region in a pandas dataframe. It then creates a grouped bar chart with a title, labels, and legend
  - In the sixth visual, The Python code uses OpenCageGeocode API to get latitude and longitude for rent locations, adds it to a new DataFrame and concatenates with original. It checks for missing values, converts `rent_cost` to numeric, and creates a scatter mapbox visualization using Plotly Express library to compare rent costs across locations.