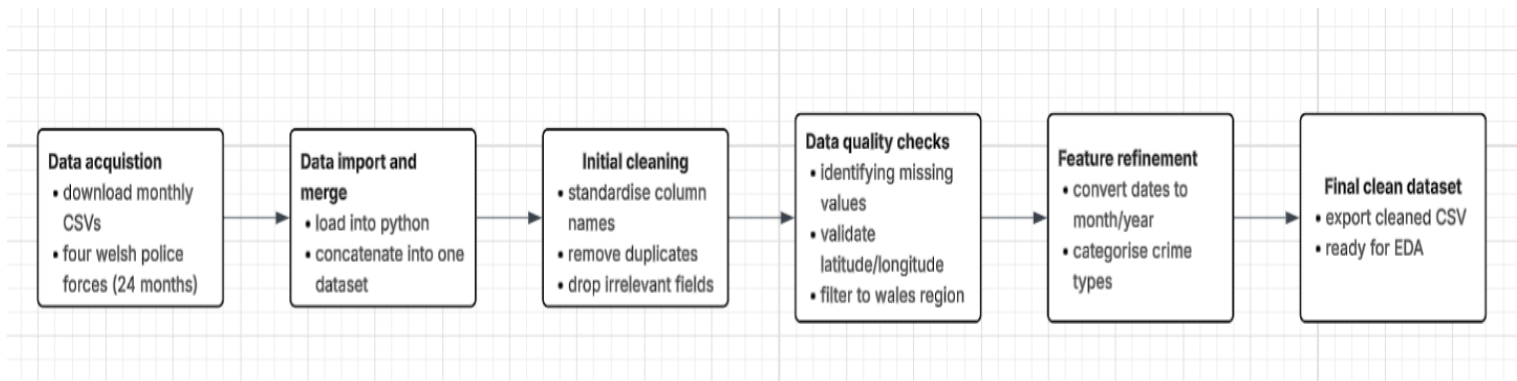
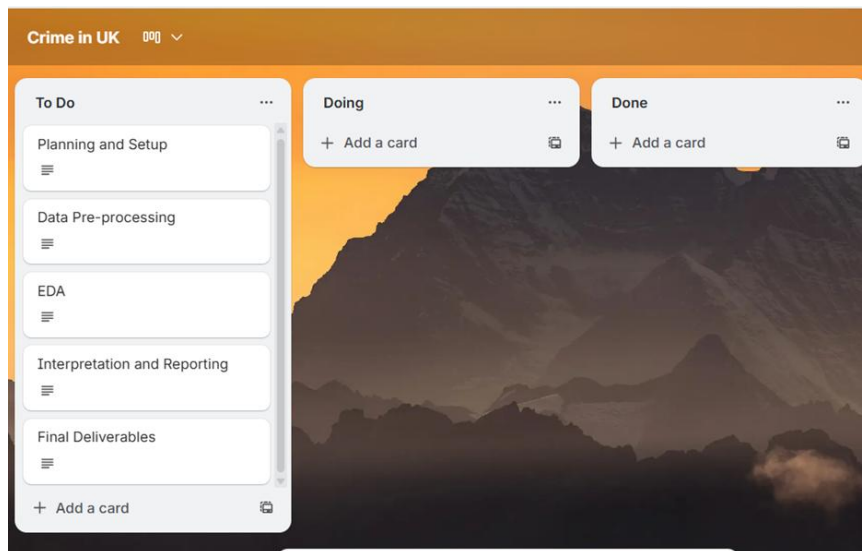


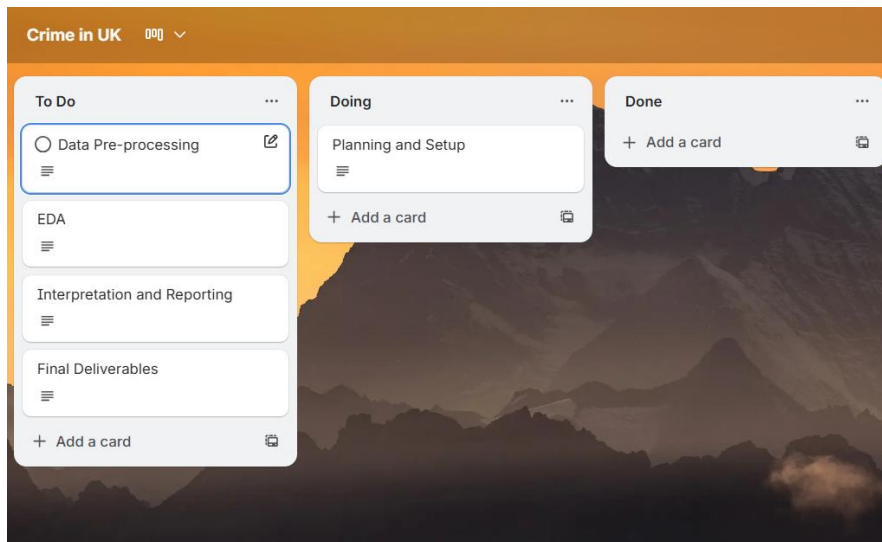
Data Preprocessing & Data Quality Workflow



The preprocessing workflow began with acquiring monthly recorded crime datasets for four Welsh police forces over a two-year timeframe. The CSV files were imported into Python using pandas and combined into a single structured dataset. Initial cleaning steps involved standardising column names, removing duplicate records, and excluding non-relevant variables such as “Falls Within” and LSOA codes. Data quality checks were then conducted to identify missing values, validate geographic coordinates, and filter observations to ensure only locations within Wales were retained. Additional refinement included converting date fields into month and year formats and ensuring crime type categorisation was consistent across all entries. The final cleaned dataset was then exported for use in Exploratory Data Analysis (EDA).

Trello Board





Doing ▾

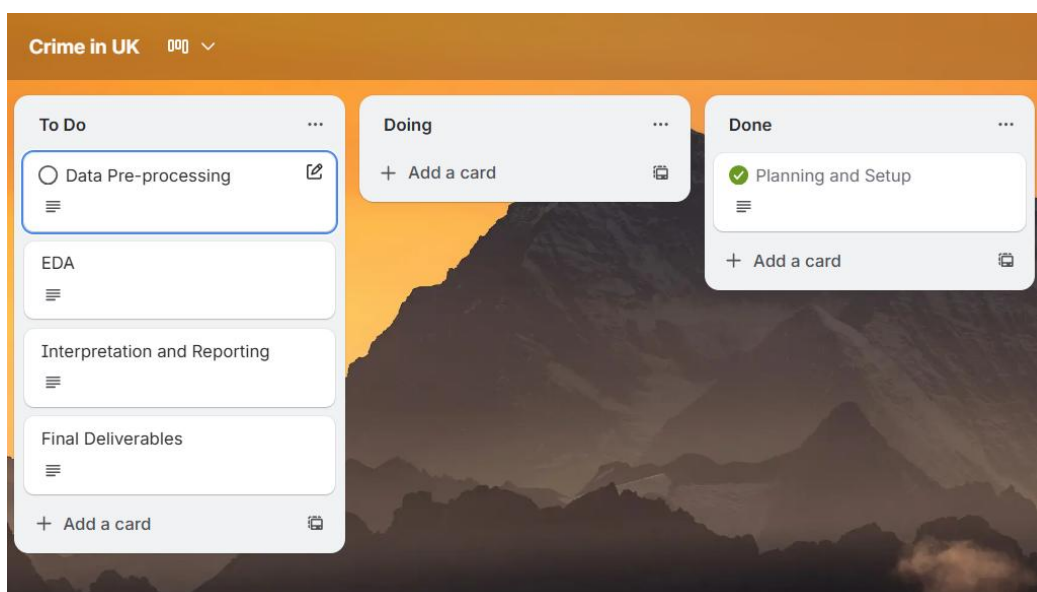
○ Planning and Setup

+ Add 🏷 Labels 🕒 Dates 📋 Checklist 👤 Members

☰ Description

Edit

- Write Project Statement of Work summarising goals, dataset, and deliverables
- Create folder structure
- select 4 police forces
- download 2 years of street level crime data (CSV)
- save files by force and year into folders
- set up anaconda environment (install pandas, numpy, seaborn, matplotlib, folium)
- create new Jupyter Notebooks



The same structured routine was maintained throughout the To-Do, Doing, and Done stages of the project. The entire project was completed over a period of five days, with each major section being addressed on a separate day. The final day was dedicated to reviewing all completed work, making refinements, and ensuring the overall quality and consistency of the project deliverables.

✔ Data Pre-processing

+ Add Labels Dates Checklist Members

Description

Edit

Goal: clean and unify datasets

- import pandas and load csvs from each police force
- combine monthly csvs
- add column for force name
- standardise column names
- convert month to datetime
- drop rows with missing latitude/longitude
- drop irrelevant columns if any
- save cleaned data
- verify number of rows and columns after cleaning
- document preprocessing workflow

✔ EDA

+ Add Labels Dates Checklist Members

Description

Edit

Goal: Explore and visualise trends

- Load processed data from all forces
- Create summary table (total crimes by force and year)
- Line plot: Crime trends over time per force
- Bar Chart: top 10 crime types overall
- pie chart: crime proportion by outcome...
- check regional differences in high/low crime areas
- save key plots to output/charts/ folder

✔ Interpretation and Reporting

+ Add Labels Dates Checklist Members

Description

Edit

Goal: Turn data into insights

- calculate average monthly crime per force
- identify top crime categories per force
- compare forces (high vs low crime)
- identify trend patterns (increasing/decreasing crime)
- recommend 2 police forces for deeper study

✔ Final Deliverables

+ Add

🏷 Labels

🕒 Dates

☑ Checklist

👤 Members

☰ Description

- write exec summary
 - objective
 - key insights
 - recommendations
- check all charts
- export files
- submit
- review