

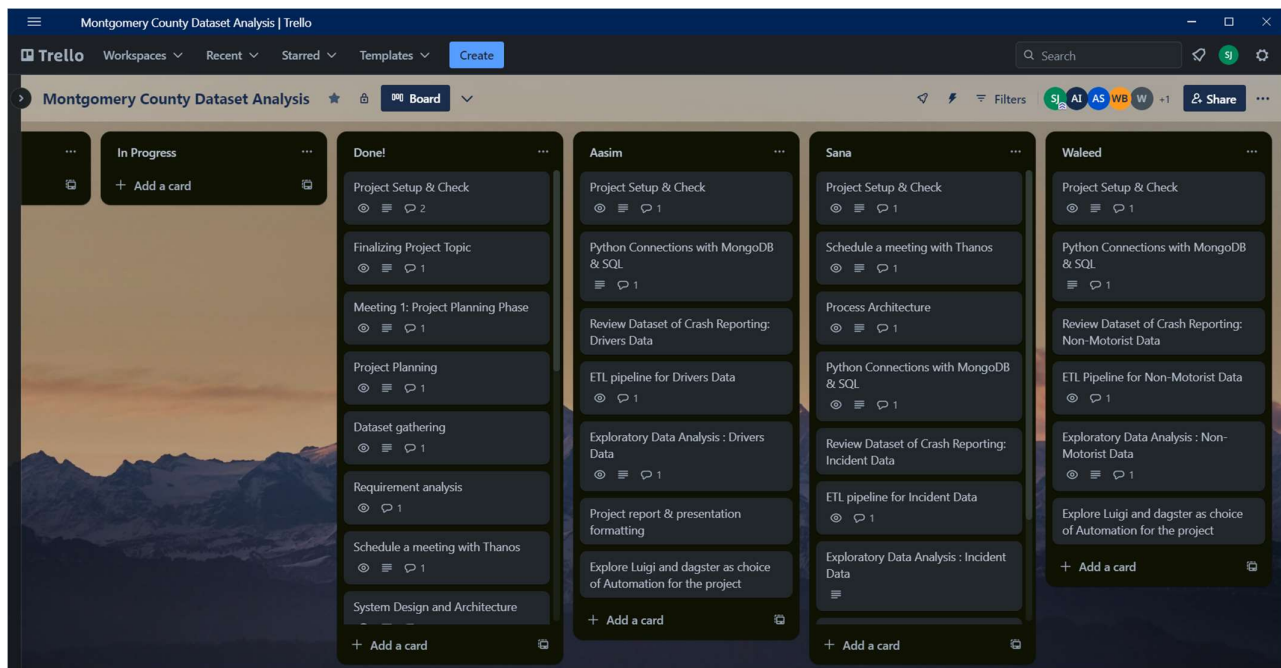
# Team I Work Breakdown Report

The team has been collectively working together right from data requirements phase to completing the lifecycle of the project. This report outlines the efforts made by the group and draws attention to the challenges that the team face altogether. The team consists of 3 members from Master of Science in Data Analytics, School of Computing, National College of Ireland. Below are the details of three members:

Name of the member	Student Id	Dataset	Role
Aasim Inamdar	23236108	Drivers	Member
Sana Jalgaonkar	22237941	Incidents	Team Lead
Waleed Bin Umer	23187956	Non - Motorists	Member

Project Management activities and task distribution are recorded at -

<https://trello.com/invite/b/RNWPGW83/ATTId0de96d75044461cd7bcd57597744d1c2FE3D115/montgomery-county-dataset-analysis>



Below are the individual responsibilities of each team member:

## Aasim Inamdar -

- Implemented the ETL build and visualization for drivers' dataset.
- Worked upon setting up the local installation and connection of relational database – MySQL in his and team member's machine.
- Collaborate with other team members to understand the requirements and iterate the solutions.
- Cleanse and preprocess the data to remove any inconsistencies and duplicates for the driver's dataset.
- Worked on the formatting of the final report and PowerPoint presentation.
- Provided the explanations to insights found in the analysis of driver's dataset in the final report.
- Carefully resolved disputes within the team, encouraging positive vibes during discussions and meetings.

## Sana Jalgaonkar –

- Maintained a project management plan in place keeping in mind the project lifecycle right from requirements gathering to delivery of the project.
- Responsible for communication between the team and professor on the project.
- Distributed team's tasks to each member based on their strengths and weaknesses.
- Designed the data architecture of the entire ETL process.
- Worked upon collecting data via API endpoints for Incidents and driver's datasets.

- Implemented the ETL build and visualizations for Incidents datasets.
- Worked upon the common content along with explanation to insights to Incidents dataset in the final report and PowerPoint presentation.
- Tried to implement automation of the ETL pipeline using Luigi.
- Refactorized the code snippets to optimize and make the static versions of code dynamic.
- Configured MongoDB atlas cluster with proper Database and network access for the team.

#### *Waleed Bin Umer-*

- Implemented the ETL build and visualization for non-motorist's dataset.
- Collaborate with other team members to understand the requirements and iterate the solutions.
- Cleanse and preprocess the data to remove any inconsistencies and duplicates for the non-motorist's dataset.
- Worked upon understanding to establish the connection with MongoDB.
- Provided the explanations to insights found in the analysis of non-motorist's dataset in the final report.
- Urged team members to work together to promote creative thinking and harmony.

#### *'Few from the Many' challenges faced while working on the project-*

- **Learning Things:** Since the whole team was new to idea of implementing ETL leveraging python instead of any other ETL tool it was great chance to use the opportunities to revise and learn the programming languages – python, SQL, and NoSQL in whole before beginning with requirement gathering phase.
- **Configuration of MongoDB Compass in systems:** team faced problem with establishing a connection via MongoDB compass on Sana's and Waleed's system. Thus, to overcome this the team moved to use of MongoDB Atlas and getting the configurations done.
- **Connection time-out error in MongoDB:** While establishing a connection with MongoDB using python, the team got stuck and spent quite a lot of time to troubleshoot the problem with timeout and dns error. Finally, after some tweaks to code snippet and discovery that college's network might be the reason causing the issue there was some profound relief.
- **Loading csv into MySQL Database:** Creating a table as simple as it may seem and was, tougher the underlying brainstorming to insert the data into it. This must be next persistent brainstorming round for all the 3 members to stress their brains out as to figure out why the team's facing every now and then parameters mismatch errors. Finally, the solution lied in simply converting the data frame that was used for insertion into a str object before converting it to a csv file and having a default datatype 'Text' for all the columns in the table at the time of table creation.
- **Handling NaN values:** Empty values in the csv generated by API and extracted from the official government site of the Montgomery County were another reason as to why the data wasn't fed into MySQL tables. Thus, replacing 'NaN' and empty/ missing values from the file with "Missing" keyword was a great that simplified the feeding process. So that in analysis this "Missing" keywords would be the proof that ACRS failed to keep track of parameters there.
- **Analysis of important features in all the 3 datasets:** All three datasets had more then 30 features so it was important to decide and rank the features that would be involved in the analysis and decision-making process. The team brainstormed on this issue by discussing and classifying the features into 5 main categories- Very High Importance, High Importance, Moderate to High Importance, Moderate, and varies.
- **Automation using Luigi:** The idea of implementing automation using dagster an luigi at the final moments was something sudden and new, yet a good chance to step out of the comfort zone and learn new. After learning more about it, the team implemented automation with luigi on the Incidents dataset – though the automation somewhat succeeded in whole that is ETL process is getting completed successfully but there's an error that needs future attention for creating further implementation of extracting back the data from Target to perform visualizations.
- **Team bonding and collaboration:** 3 individuals with different mindsets collaborating in whole was quite challenging. Thus efforts made by individuals to co-operate and co-ordinate each other's view and facilitating brainstorming sessions, group discussions and meetings attended together made the project delivery possible giving their best.