**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 distribution are tracked 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.

Challenges faced by each member-

Waleed Bin Umer-

* **“**MongoDB timeout error**”**: This error is occurring while connecting MongoDB to my python environment this error was re-occurring after 30000ms after running establishing my MongoDB connection so, I create a new client and connect to the server and making sure that MongoDB server-status is working properly.
* **“**Uploading Data in MongoDB**”:** JSON file data was not uploading in MongoDB because data was more than 10000 records and I was using insert query for a single record so I search and troubleshoot and find “collection.insert\_many(data)” query that worked for me to upload my Json file data in to MongoDB atlas successfully
* **“**Inserting Data in SQL**”:** I was having error in inserting the data in SQL after creating the table and find it difficult to upload the data in SQL server I did research and finally find the way to upload my data on SQL server “my\_cursor.execute(datainsert, tuple(row))” this query works for me and successfully upload table data in the table