**Self-Assessment**

My role in this project changed several times throughout its development. My part at the start of the process was to serve as a resource for my teammates, helping them to make sense of data analytics problems and explaining how we might solve them. By working with my team to focus on problem-solving instead of data analytics theory, I helped them take small steps toward solving their problem and develop a deeper understanding of how data analytics can be applied. I also advocated for my group members by encouraging them to share their ideas and listening carefully to what they said so that I could help them build on their ideas and contribute even more than they realized they could.

As the project developed, my role shifted from one of theory to one of application. As the group began focusing on finding a dataset and writing an analysis plan, I became more involved in the research process by leading my team through brainstorming exercises and helping them break down large topics into smaller pieces. Because I had been part of these kinds of projects before joining the cohort, I knew what it would take to accomplish our goal within the available time frame, which allowed me to set realistic expectations for myself and my teammates. This also helped me find ways for us to work together.

I ensured everyone was on track with their commitments during the team meetings. When other interns were struggling with a task, I would assume the role of a facilitator and help them get unstuck. I was also the scribe for our meetings, which meant that I maintained the minutes during these meetings and sent the minutes to the team members afterwards to ensure we had shared expectations and a common understanding. I also worked on the schedule for these meetings and stayed in contact with the other interns regarding commitments. Finally, when there was an urgent issue that needed to be addressed immediately, I would assume a leadership role so that we could make decisions without wasting time.

**Project and Team Summary**

**The Objective** was to clean, transform, analyze bike crash data in the state of North Carolina state and generate patterns to visualize frequency of Serious & Fatal accidents. We utilized supervised Machine Learning techniques to identify the root causes of these accidents and develop strategies for prevention.

**We wanted to** determine what causes the crashes, level of fatality to help the road safety departments improvise safety measures and policies pertaining to reduce the occurrence of crashes.

The Machine Learning classifier algorithms that we used

* Bagging Classifier (Sklearn)
* Adaboost Classifier (sklearn)
* Random Forrest Classifier (sklearn)

Based on the performance metrics, Random Forest Classifier was the best model in our case. Although all the models had high accuracy, Random Forest had the lowest log loss and error rate compared to Bagging Classifier and AdaBoost Classifier. Also, the recall and F1 score was similarly high for all three models.

The possible limitations of the data set were:

* The data was highly imbalanced, with most of the accidents being non-fatal. This affected the model performance, and under sampling was done to improve accuracy.
* Some crucial factors related to accidents were not included, such as information on speeding by drivers, cell phone usage, time of arrival of emergency units, and information on passengers.
* The low correlation values indicated weak relationships between the features and the target variable.

Some suggestions to overcome these limitations could be:

* Collecting data on driver behaviour, such as speeding and cell phone usage, to improve the understanding of factors contributing to accidents.
* Collecting information on the arrival time of emergency units and passengers to get a more comprehensive view of the situation.
* To improve the model performance, consider alternative techniques to handle imbalanced data, such as oversampling or synthetic data generation.

We decided to communicate via email or WhatsApp when there needed higher attendance at meetings. In this case, my role was to ensure that everyone followed up on their commitments as soon as possible and resolved any conflicts or issues ASAP.

When we had video calls via zoom, my roles included ensuring everyone was prepared for it and calling out those who needed to pay attention during the meeting. As a moderator, I would ensure that everyone's voice was heard and keep discussions focused on what we had to achieve in this meeting.

Finally, I served as an available resource when it came to questions.