**Peer Feedback Document**

We had our peer feedback directly from the TA, Haihan.

Here she gave a few suggestions to work on

1. Find the relationship between Injuries & Deaths

We have a computed correlation between total injuries, total deaths & damage. However, we didn’t find any critical correlation among these variables.

1. Google Maps Implementation

We have worked and implemented google maps to show more intuitively how the USA is affected by storms. We used heat maps to make the visualizations more comprehensive.

1. Use Log scale

We have used log scale while plotting damage, and here the damage cost is varying from 0 to Billions So, log scale helped to make the graphs more understandable.

1. Use the most recent data, instead of going for 10 or 20 years of data.

We initially thought of taking the past ten years of data. However, using the most recent years, data would be helpful because of the aspects of technology and improved disaster management. Taking older years would mislead the outcomes.

1. Empty Data

After merging three years of data, we have the right amount of data. For the null value evaluation, we came to know that removing the row is the best option. Because, based on our team intuition, imputing null values with the mean or median values has no impact. For example – in the case Total Damage column, most of the rows have a null value, so implementing any technique was generating improper results.

1. Adding Population Data

We have added a new column to add the population for each county. However, for this milestone, we are not sure about how to proceed. But we are confident that we will make some meaning full insights with this data. And we would like to take some suggestions with the staff.

1. Predictive Modelling

We were confident in our initial Project proposal that we would use regression techniques like Linear, SVM, Random forest and apply regularised models like Lasso & Ridge to reduce the number of columns. However, after digging deep into the data, we found that it is harder for us to perform regression - Initially, we assumed the magnitude is the best variable to perform regression. But, we see the magnitude column is different for different storms and cannot have a standard connection to estimate the target variable.

However, we are making sure to perform proper data analysis and best storytelling about the impact of storms on various factors.

We are also working to implement clustering for further important knowledge discovery. And we will schedule a meeting with the TA to seek suggestions about this task.