Nisha Bharakhada

**Learning Outcome:**

As part of the Capstone project, each team member was assigned individual task. My role was to perform the initial data analysis, provisional machine learning model and each task is captured and defined in the excel.

As part of group discussion and brainstorming sessions, I contributed to understand the data , define the steps in order for data cleaning, preprocessing and preparing the tableau dashboard for initial data analysis. Contributed to the below task/sub tasks:

* Looking for dataset which satisfies the conditions defined in the assignment (minimum 1000 records)
* Listing all the different dataset with description
* Having brainstorming session and capturing all inputs which dataset to select
* Updating readme file with source, description, and technologies to be used
* Ensuring simultaneously the points are captured in provisional flow diagram and presentation
* I also contributed along with other member of team to prepare provisional machine learning model to check the accuracy score and whether the prediction model provides good accuracy score
* Also, initially I prepared story on “Clickup”, however, being more time consuming, we decided not to use the same.
* The challenge encountered in first dataset was not achieving good score to answer the question we were hoping to answer.
* Before moving on to the different dataset, tried more machine learning models, however, with an agreement with other team members, we decided to switch to a different dataset.
* Ensured that flow diagram, presentation, readme and data analysis files committed to github.
* I ensured that all the links in readme file is working error free.
* During segment 3, I started working on dashboard to make sure that the individual features present in the clean dataset is captured visualized in the Tableau dashboard.
* I ensured that the datapoint is well understood and significance which feature has more weightage according to ML model.
* Further defined how to present story and dashboard so that graphical presentation captures individual feature analysis, multiple feature analysis and machine learning outcome with different models explored.
* I assisted in EDA for machine learning model and assisted in data cleaning process in order to identify which column to remove, to change the datatype, to select DB.
* Prepared process flow diagram so that it is easy to explain/ present high level process flow that team followed.

Our team used Google Drive, slack, zoom to set up meeting and have discussions to prepare for each task for each member. Believe we had more challenges in order to understand the first dataset (Youtube videos) and the other challenge personally I had in initial data analysis on Dashboard. However, reviewed the previous modules to get the answer and reached out to TA’s to understand the exact requirement.

The biggest strength for our team was that we were able to discuss any issues faced or technical challenges and everyone contributed to assist each other to resolve the same.

For any dataset to perform analysis and creating prediction model, it is important to understand the input data in relation to domain. Also, it is important to clean the data, use the correct datatype. Apart from that it is important to visualize and understand the significance of each datapoint. To select the machine learning model, it is important to use all the models to compare and check which model provide optimum result. In our model, since it was related to healthcare, it is important to check accuracy score as well as recall score.