*Project: Medical cost prediction:*

**Business Objective**: The insurance.csv dataset contains 1338 observations (rows) and 7 features (columns). The dataset contains 4 numerical features (age, bmi, children and expenses) and 3 nominal features (sex, smoker and region) that were converted into factors with numerical value desginated for each level.

The purposes of this exercise to look into different features to observe their relationship, and plot a multiple linear regression based on several features of individual such as age, physical/family condition and location against their existing medical expense to be used for predicting future medical expenses of individuals that help medical insurance to make decision on charging the premium.

**Acceptance Criterion:**

Need to deploy the end results using Streamlit.

**Milestones:**

**30 days to complete the Project**

|  |  |  |
| --- | --- | --- |
| **Milestone** | **Duration** | **Task start - End Date** |
| Kick off and Business Objective discussion | 1 day |  |
| EDA/Student interactive | 1 Week |  |
| Model Building/ Evaluation | 1 Weeks |  |
| Deployment | 1 Week |  |
| Final presentation | 1 week |  |

Protocols:

1. All participants should adhere to agreed timelines and timelines will not be extended.
2. All the documentation – Final presentation and R/python code to be submitted before the final presentation day.
3. All the participants must attend review meetings.