**Capstone II: Project Ideas and Description**

1. **Heart failure prediction**

[**https://www.kaggle.com/andrewmvd/heart-failure-clinical-data/tasks?taskId=1177**](https://www.kaggle.com/andrewmvd/heart-failure-clinical-data/tasks?taskId=1177)

~ 17 million people die globally/year of cardiovascular diseases chiefly exhibiting as heart failures. This available data frame made of patients records quantifies symptoms, body features, and clinical laboratory test values. It can help perform prediction and analysis highlighting patterns and correlations otherwise undetectable by medical doctors. A machine learning model on this data set can help to predict patients’ survival outcomes. This data set looks self-sufficient and simplistic for my first Capstone and I am familiar with the subject background too.

1. **Bank Customers Churn:**

[**https://www.kaggle.com/santoshd3/bank-customers**](https://www.kaggle.com/santoshd3/bank-customers)

This is a bank costumer related dataset which contains records of who are withdrawing their account from this bank due to some loss and other issues. One aim of data analyses on this dataset can be understanding costumer behavior of account withdrawal based on the features mentioned in their account details. Developing a machine learning model to understand the set of features responsible for the withdrawal can help the bank in developing strategies to stop/help the migration of account holders. I choose this dataset because it provides me with scope of applying and practicing basic concepts of data science without being overcomplicated in nature.

1. **E-commerce shipping data**

[**https://www.kaggle.com/prachi13/customer-analytics**](https://www.kaggle.com/prachi13/customer-analytics)

An e-commerce enterprise is asking for insights from their customer database to understand costumer behavior which could help in increasing sales of electronic products. This dataset is taken from Kaggle and it mentions the inspiration of analyses to get the answers of questions like: What was Customer Rating? And was the product delivered on time? Is Customer query is being answered? If Product importance is high. having highest rating or being delivered on time? It is good dataset to start the analyses and a bit bigger with above two. It provides scope of applying data cleaning, exploratory, preprocessing and modelling concepts. It also makes me familiar with ecommerce and costumer behavior related basic and data analyses concepts